Masters Thesis
Information and Communication Technology

GPRS for GSM Upgrade Strategies using Network Redundancy

Erlend Larsen

Grimstad, Norway
Spring 2001

Abstract

The Serving GPRS Support Node (SGSN) is the most fundamental node in GPRS. Ericsson produces and manages an increasing number of SGSN nodes in the world. SGSN nodes require upgrades on software and hardware regularly, forcing restarts, which gives service interrupts for subscribers.

This report presents four different upgrade strategies intended to avoid or minimize service interrupts, while being cost-efficient: Hot standby, Subscriber based PCU routing, RA based PCU routing, and Redundant links solutions.

The RA based PCU routing is considered the best, because it provides a smooth transfer of subscribers from one SGSN to another on a routing area basis, but not under all circumstances. It employs the Inter SGSN Routing Area Update (ISRAU) procedure to transfer subscribers. The routing mechanism is based on existing technology, namely Point-To-Point BVC connections bound to different Network Service Entities (NSEs) leading to different SGSN nodes from each cell through the Packet Control Unit (PCU).

Preface

The thesis has been done on assignment from Ericsson, Grimstad. Though the title has changed since I got the assignment 'Network Redundancy Upgrade Strategies in GPRS', the idea has stayed the same, to find an efficient way to upgrade an SGSN node. The learning curve has been exponential, starting out with absolutely no knowledge about the GPRS system. Thanks to a two day GPRS introduction course in January I got an insight in the system that since then has grown bigger and bigger, as I have dug further into it.

During three months from January to June I have been given invaluable help and feedback from Product Manager Steinar Strømsvåg, my supervisor. Together we have developed the various solutions, and with help from other people at Ericsson, namely Ole Jonny Gangsøy, Kjell Lindborg, Terje Urfjell and Bent Vale, we have made these solutions probable. I would also like to thank Stein Bergsmark, who was the contact at Ericsson giving me the assignment. He has also been a great help through his thesis seminars, held at HiA, and his review of my report.

As this thesis marks the end of 17 years of school, I want to take the opportunity to thank some of my fellow students, especially Morten Thorsen, Arnfinn Andersen, Sigurd N. Daland and Stian Haugen. Finally, Kikki has been a great support to me, as she always is.

Erlend Larsen

Grimstad, May 28, 2001

Table of Contents

Table of figures

Table of tables

Definitions and abbreviations

See also GSM 01.04 “Abbreviations and acronyms” [1].

BSC	Base Station Controller
BSS	Base Station System
BSSGP	Base Station System GPRS Protocol
BTS	Base Transceiver Station
BVC	BSSGP Virtual Connection
BVCI	BVC Identifier
CN	Core Network
CS	Circuit Switched
ETSI	European Telecommunication Standard Institute
GGSN	Gateway GPRS Support Node
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
GSN	GPRS Support Node
HLR	Home Location Register
IMSI	International Mobile Subscriber Identity
IP	Internet Protocol
ISRAU	Inter SGSN Routing Area Update
LCC	Life Cycle Cost
MSC	Mobile-services Switching Centre
NS	Network Service
NSE	Network Service Entity
NSEI	Network Service Entity Identifier
NS-VC	Network Service Virtual Connection
NS-VCI	Network Service Virtual Connection Identifier
O&M	Operations and Maintenance
PCU	Packet Control Unit
PS	Packet Switched
PSPDN	Public Switched Packet Data Network
PSTN	Public Switched Telephone Network
PTP	Point-To-Point
QoS	Quality of Service
RA	Routing Area
RAC	Routing Area Code
RAI	Routing Area Identity
RAM	Random Access Memory
RAPR	Routing Area based PCU Routing
RL	Redundant Links
RP	Regional Processor
RPP	Type of RP that the PCU is built on
SBPR	Subscriber based PCU routing
SGSN	Serving GPRS Support Node
TMSI	Temporary Mobile Subscriber Identity
UMTS	Universal Mobile Telecommunication System
VLR	Visitor Location Register

Chapter 1 Introduction

1.1 Background

GPRS introduces IP in mobile networks. The benefits are among others that the subscriber is always connected, and the system is very widespread. It is being rolled out in almost every country in the world. Ericsson has high market share on production and management of several different nodes in the General Packet Radio Service (GPRS) system. One of these nodes is the Serving GPRS Support Node (SGSN). I.a. it manages subscribers connections through the network, and is essential to transfer traffic to/from subscribers. When the rollout of the GPRS system is finished, Ericsson could very well have management responsibility for 10.000 SGSN nodes. This will put high demands on the quality of the Operations and Maintenance (O&M) in order to minimize the Life Cycle Cost (LCC). As of now, Ericsson's SGSN nodes have an average uptime of 99.9986%. This equals about minute per month. Ericsson's goal is 99.999% uptime.

An SGSN node is not without flaws, and the software needs to be upgraded roughly once every three months. There is also often a need to replace or add hardware. A restart takes from three to twelve minutes where the SGSN is down (out of service), and this means there is only room for one restart every three months. This leaves no time for larger software and hardware upgrades. Failures come in addition. Ericsson's wish is to be able to do an upgrade on any node when necessary, and at least once a month. They want this for a number of reasons:

1. To minimize the number of different software releases in the marked at the same time.
2. To satisfy the customers demands for shorter time to service and thereof the operators possibility to increase the revenue.
3. To improve cash flow, reduce operational cost and use the opportunity to create new business opportunities and more sales.

The uptime can also be defined as In Service Performance (ISP). This is what is most important from the operator's and the end-users' viewpoint, as the operators only make money when subscribers are using their service, and the end-users want to use the service they subscribe to. Therefore it is important to cut down any downtime for the SGSN nodes to minimize the system downtime related to software and hardware upgrade.

1.2 Problem and thesis description

The GPRS system consists of a number of different nodes. The most fundamental node in GPRS is the SGSN. It is among other things responsible for all packet-switched data going to and from the mobile stations. From time to time the SGSN node has to be upgraded. Upgrade of a SGSN node is important for a number of reasons:

- To keep the node's software as bug-free as possible.
- To introduce additional services to operators and end-users.
- Add and change hardware in the node

When an upgrade is ongoing on a SGSN node, the node cannot at the same time serve subscribers. It has to be taken out of traffic. In contrast to failures, upgrades can be prepared for. Therefore it should be possible to upgrade an SGSN node without affecting the subscribers as much as at present. Both operators and producers of SGSN nodes would benefit greatly from an interrupt-free upgrade strategy:

- The subscribers can use the service all the time, which gives increased revenue for the operators
- SGSN nodes can be upgraded when wanted, without having to accumulate several operations to make the most out of one upgrade interrupt.

Because upgrades are vitally important to keep an SGSN node in operation, it would be natural to open for changes in already defined standards, if these changes would give sufficient benefits. Also, investments up to a certain point are considered acceptable, if the outcome is a better upgrade strategy.

The original assignment as it was decided in February is quoted below. It has not been necessary to change it in any way, but it should be seen in relation to the limitations in chapter 1.3:

> "The General Packet Radio Service (GPRS) system consists of a number of different nodes. The most central node in GPRS is the Serving GPRS Support Node (SGSN). It is among other things responsible for all packet-switched data going to and from the mobile stations.
>
> The assignment is in two parts where the first is to find different strategies for upgrade of software and hardware on the SGSN so the end-users or subscribers experience as few and short occurrences of cut-offs and interrupts as possible.
>
> The second part consists of elaborating on the strategy recognized as the best one in the preceding analysis, giving a clear picture of what has to be done to implement this solution in the GPRS network. Also, the consequences of implementing the strategy will be discussed."

1.3 Goals and limitations

The goals of this thesis are firstly to find different usable strategies to enable upgrade on an SGSN node so that the subscribers connected through the node experience as few and short occurrences of cut-offs and interrupts as possible. Next, the strategy considered the best is described in further detail, with weight on the implementation and consequences of the strategy.

The limitations on the goals of this thesis are on the detail level, as the complexity of the GPRS system is very high. All implementation descriptions and function descriptions should be held at a network level. In the same way, consequences of the strategies should be discussed at an executive level.

1.4 Methods

1.4.1 Literature reviews

I have studied several of the GPRS for GSM standards published by the European Telecommunication Standard Institute (ETSI). I have obtained these over the Internet, at ETSI's web pages with address: http://www.etsi.org/. In addition I have had access to many of Ericsson's GPRS documents.

The document covering the GPRS for GSM system on an overall level is the GSM 03.60 [3]. It covers all parts of the GPRS system in a service description.

Next, the GSM 03.03 [2] contains the descriptions of the cell, routing area and other identities. Here, documentation about Cell Identifiers and Routing Area Identifiers can be found

The GSM 08.16 [6] gives valuable information of the service layer on the Gb interface between SGSN and BSS. It documents the use of Network Service Entities and Point to Point BSSGP Virtual Connections.

The 3GPP TS 08.18 [7] contains the BSSGP standardization. Among others it describes the different data packets that can be sent over the BSS GPRS Protocol.

The Administration and use of Network Configuration Data in SGSN [10] is a function description of the configuration of an Ericsson SGSN node. From this it can be gathered how an implementation of one of the strategies will take place

Finally the Gb Interface Configuration [11] describes the configuration of the Gb interface between the BSS and SGSN on the SGSN node. It is specific for Ericsson's SGSN node

1.4.2 Practical work

The first task in my work was to find out in what ways upgrades are done at present. Next, I went through the different parts of the GPRS system, and especially the links and nodes directly connected to the SGSN node. Then I gathered information from employees at Ericsson in Grimstad to find out what thoughts they had concerning upgrade strategies. Through all this I had frequent meetings with my supervisor, Product Manager Steinar Strømsvåg, where we discussed the various solutions.

When I could start sketching the different proposed solutions, I again asked numerous employees of their opinions and comments concerning possible faults in the solution arguments. After more research, Strømsvåg and I decided the Routing Area based PCU routing (RAPR) solution was the best solution of the four I ended up with.

As I focused on the RAPR solution, it became evident that there exists a routing mechanism that can be used in the way the RAPR solution is dependent on. I then investigated further to confirm this, but unfortunately there was not time to complete the investigation.

1.5 Report outline

The report starts with documenting the GPRS system. It contains information about circuit switched and packet switched communication and a history on the GSM system, on which the GPRS system is based. Then comes an overview of the GPRS system, with description of the nodes important to the solutions following this. Next, several interfaces in the GPRS system is described. After this, important identities are documented. Finally, the last descriptions in chapter 2 are introductions to the Routing Area Update Procedures and GPRS Attach.

The next chapter (3) documents the general work process to come up with the solutions, criteria to build the discussion of the solutions on, the present solution, a transmission-based approach, and then the various upgrade solutions developed. Chapter 3.8 discusses the different solutions and concludes with the best of them.

Chapter 4 describes the implementation of an RA based PCU routing solution. It covers the different parts of the GPRS system that will be affected by the implementation, and presents a new routing mechanism for the solution. Also, it discusses the various consequences the solution will provoke.

Chapter 5 is a summary of the results achieved. Chapter 6 discusses the various results, and chapter 7 presents the conclusions of the report. Finally, chapter 8 contains references to the documents this report is based on.

Chapter 2 GPRS overview

2.1 Introduction

To be able to develop the various update strategies, I have had to do study of the GPRS system. In the following pages I will describe the various parts of the system, how they communicate, and some examples of communication between the mobile station and the core network.

I start with the basic understanding of circuit switched and packet switched communication to show what GPRS adds to the GSM system. The historical presentation of GPRS is important to get an overview of the thesis subject. Then the nodes essential to the thesis are presented. Thereafter several interfaces are documented, the most important being the Gb interface between SGSN and BSS. Explanation of PDP Context is also documented, as it clarifies what connections the SGSN has to take care of.

Next, I present several important identities, where the Routing Area Identity and the Cell Identity is considered the most important. Finally, the Routing Area Update Procedures, which the Routing Area based PCU routing solution is based upon, and also GPRS Attach, are documented.

2.2 Circuit switched and packet switched communication

Circuit-switched networks were designed primarily for voice communications. The network establishes a connection between multiple switching points and creates a circuit for the duration of the call. Even if no conversation is taking place, the circuit is in use. This is not an efficient use of resources available in the network, unless there is a constant flow of information going through the circuit while connected.

Packet-switched networks, on the other hand, use capacity only when there is something to be sent. This way, all available capacity can be filled with packets from the same or different senders. Address information is placed in the header of each information packet.

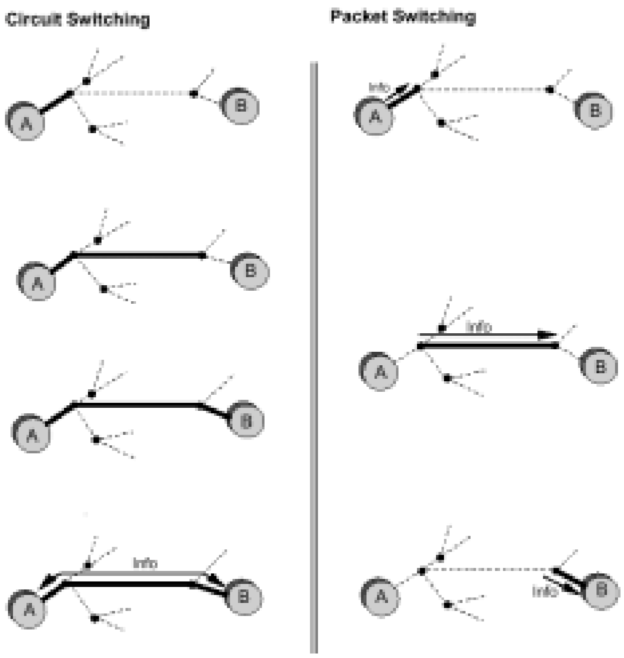

Figure 2.1 Circuit switched and packet switched communication

2.3 GSM

In the early 80's the analogue mobile phone market grew rapidly in Europe. Because there was no set standard each country decided to develop their own system. This caused major problems with coverage and roaming as the systems were not compatible. Realising the mistake with the analogue systems and also the potential of mobile telecommunications the Conference of European Posts and Telegraphs (CEPT) formed Groupe Spécial Mobile in 1982 to develop a system that would work across Europe. The European Telecommunications Standards Institute (ETSI) gained responsibility for the development of the Global System for Mobile Communications (GSM) in 1989 and published phase 1 of the GSM specifications in 1990.

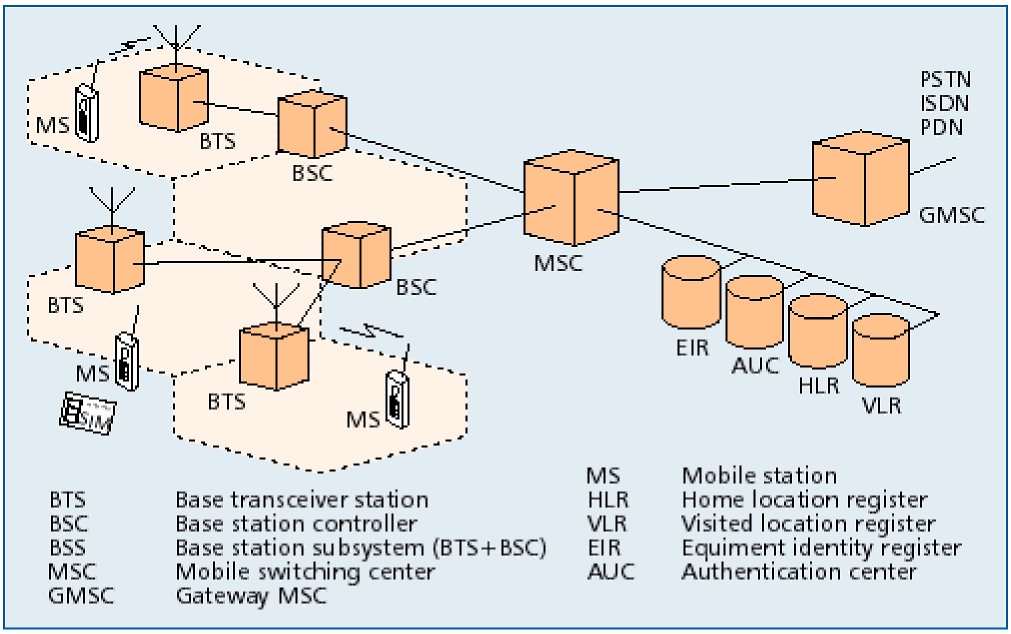

Figure 2.2 GSM system architecture

The past decade we have seen, and are seeing, a convergence of technologies, and this is also the case between datacom and telecom technologies. The GSM network allows data transmission services. However, transmission is only possible in circuit switched mode. That is, once the connection between the two users is established, the network allots resources to the connection until this is terminated, even if the two users are not exchanging information during the whole of the connection.

This transfer mode is most efficient when the two users have to exchange a significant amount of data (e.g. file transfer or other traffic that comes in bursts); it becomes inefficient when the data to be exchanged is minimal or, more typically, when the data exchange is interactive or transactional, that is when the actual use of resources is only a part of the total connection time (e.g. when navigating on the Internet through the World Wide Web).

The GSM network has the same problem that has troubled for some years the PSTN (Public Switched Telephone Network): to create a packet transfer mode, where the users' data, packed as self contained protocol units with the indication of sender and recipient, can be transported by the network without having to be associated with a physical circuit.

An intermediate stage towards that goal was reached with phase 2 GSM, allowing for services with access to appropriate gates to the PSPDN network (Public Switched Packet Data Network). Nonetheless it is still necessary to have a physical connection (in circuit switched mode) on the radio link, even though we are accessing a virtual channel of the packet network. It follows that radio resources are underused, as the user engages a data channel to access a network where data does not travel on a continuous throughput.

The GPRS system (General Packet Radio Service), introduced by the ETSI (European Telecommunication Standard Institute) for phase 2+ of the GSM system, brings access to the packet network to the mobile phone user without having to use intermediate circuit switched connections.

2.4 GPRS for GSM

With GPRS comes packet switching with IP end to end in mobile telecom. It is a new technology to provide high capacity end to end IP packet services over the GSM network. GPRS is also a preparation for the UMTS, which is based on packet services.

In contrast to the data transfer service in circuit switched mode, where every connection established is solely dedicated to the user requesting it, the GPRS service allows the transmission of packets in link by link mode, where data packets are routed in separate stages through the various nodes supporting the service, nodes referred to as GSN (GPRS Support Node). For example, once the packet has been transmitted by the radio interface (Um), Um resources are freed, so that another user can use them, and the packet is sent node by node towards its destination.

In GSM services resources are managed in resource reservation mode, in other words they are engaged until the service requested has been performed. In GPRS, instead, the context reservation technique is used, where we tend to maintain the necessary information to support both the active service requests and those that are momentarily quiescent. In this way radio resources are engaged only when data has actually to be received or sent. All mobile stations (MS) share the radio resources of one given cell, and this greatly increases the efficiency of the system.

For Ericsson's GSM/GPRS solution, the only adjustments needed to enable GPRS are the new GPRS Support Nodes (GSN) and the Packet Control Unit (PCU) card in the BSC.

Below we see the GPRS system architecture.

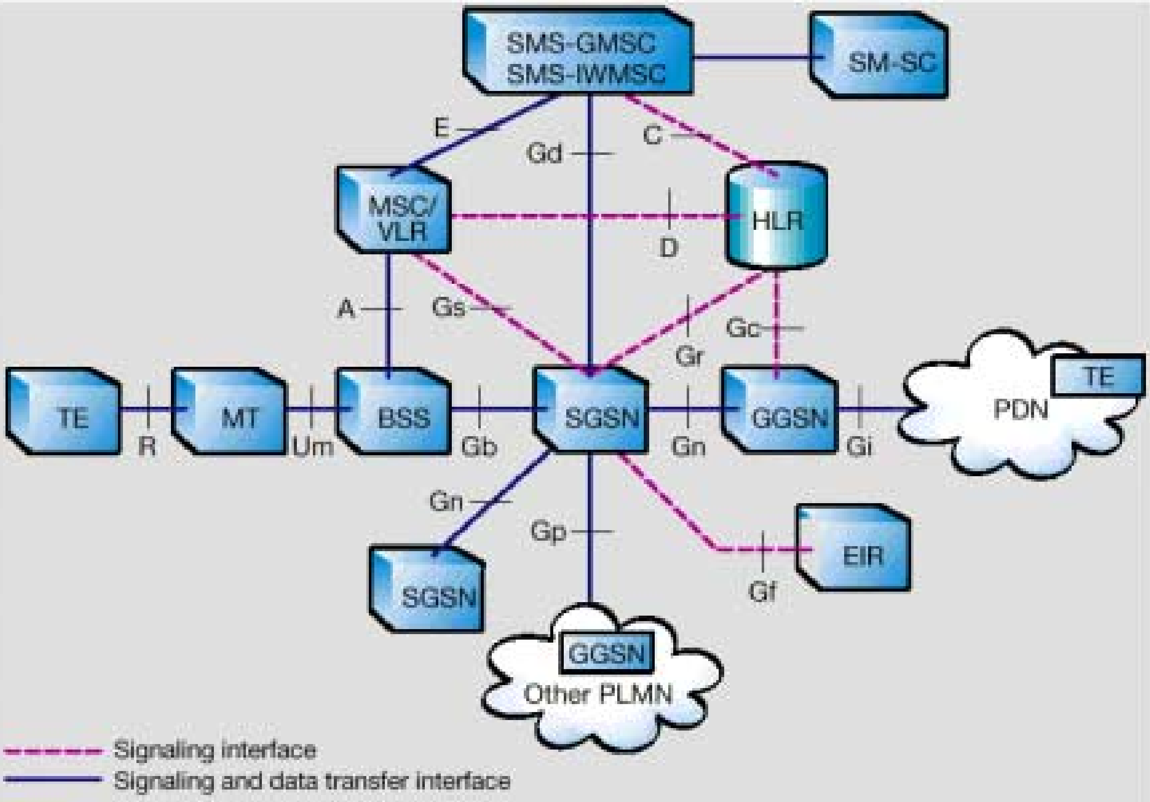

Figure 2.3 GPRS system architecture

2.5 SGSN

The Serving GPRS Support Node (SGSN) is a primary component in the GSM network using GPRS and a new component in GSM. The SGSN forwards incoming and outgoing IP packets addressed to/from an MS that is attached within the SGSN area. The SGSN provides:

- Packet routing and transfer to and from the SGSN area. It serves all GPRS subscribers that are physically located within the geographical SGSN area. A GPRS subscriber may be served by any SGSN in the network, all depending on location. The data traffic is routed from the SGSN to an MS via a BSS.
- Authentication and ciphering.
- Session management.
- Mobility management.
- Logical link management toward the MS.
- SMS forwarding.
- Charging management. The SGSN collects charging information for each MS related to the radio network usage. Both the SGSN and the GGSN collect charging information on usage of the GPRS network resources.

There is redundancy internally in the SGSN node. Both router boards and links can be configured to be redundant. Also the power distribution is redundant.

2.6 GGSN

Like the SGSN, the Gateway GPRS Support Node (GGSN) is a primary component in the GSM network using GPRS and is a new component. The GGSN provides:

- The interface toward the external IP networks. The GGSN therefore contains access functionality that interfaces external Internet Service Provider (ISP) functions like routers and Remote Access Dial-In User Service (RADIUS) servers, which are used for security purposes. From the external IP network's point of view, the GGSN acts as a router for the IP addresses of all subscribers served by the GPRS network. The GGSN thus exchanges routing information with the external network.
- Session management, i.e. connection toward an external network and dynamic IP address allocation.
- Functionality for associating the subscribers to the correct SGSN
- Charging management. The GGSN collects charging information for each MS, related to the external data network usage. Both the GGSN and the SGSN collect charging information on usage of the GPRS network resources.

2.7 BSS

2.7.1 The BSS in general

The Base Station System (BSS) is the system of base station equipment (transceivers, controllers, etc.), which is viewed by the MSC/SGSN through a single interface as being the entity responsible for communicating with Mobile Stations in a certain area. A cell is formed by the radio area coverage of a base transceiver station (BTS). Several BTSs together are controlled by one base station controller (BSC). The BTS and BSC together form the base station subsystem (BSS).

The BSS adds the Cell Global Identity including the RAC and LAC of the cell where the message was received before passing the message to the SGSN.

2.7.2 BTS

The Base Transceiver Station (BTS) is a network component, which serves one cell, and is controlled by a Base Station Controller. It houses the radio transceivers that define a cell and handles the radio-link protocols with the Mobile Station. The BTS can consist of one or more transceivers (TRXs) with or without common control equipment.

The BTS is the radio equipment, which transmits and receives information over the air to enable the BSC to communicate with MSs in the BSC area. A group of BTSs is controlled by a BSC.

2.7.3 BSC

The Base Station Controller (BSC) forwards the circuit-switched calls to the MSC/VLR and packet data to the SGSN. It has functions for control of one or more BTSs. The BSC provides all the radio-related functions. The BSC has the functionality to set up, supervise and disconnect circuit-switched calls and packet-switched data sessions. It is a high capacity switch that provides functions such as handover, cell configuration data, and channel assignment.

2.8 PCU

Both the BTS and the BSC can contain a Packet Control Unit (PCU), but not at the same time. Also the SGSN can contain the PCU.

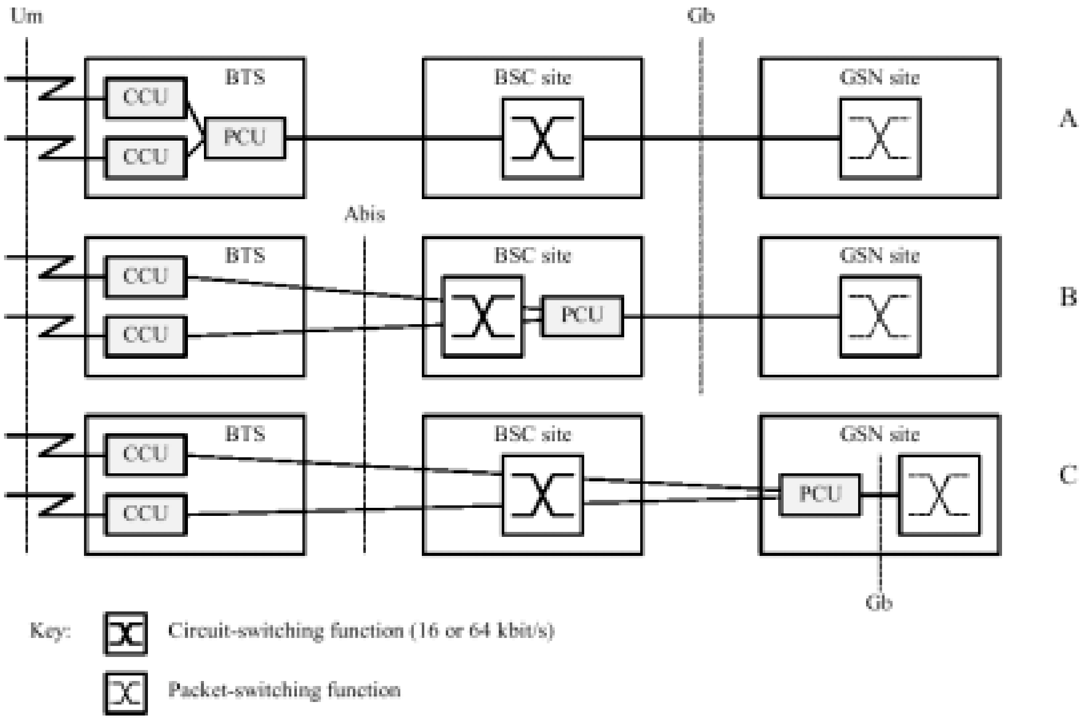

Figure 2.4 The placement of the PCUs in the BSS

The PCU is responsible for the GPRS packet data radio resource management in BSS. In particular the PCU is responsible for handling the MAC and RLC layers of the radio interface and the BSSGP and Network Service layers of the Gb interface. The Gb interface is terminated in the PCU.

The PCU consists of both central software and hardware devices with regional software. It will have one or more Regional Processors (RP). An RPP (type of RP that the PCU is built on) can work with both the Gb and the Abis interfaces or with Abis alone. The function of the RPP is to distribute PCU frames between Gb and Abis.

Where there is just one RPP in the PCU, it will work with both the Gb and the Abis interfaces. Where there is more than one RPP, each RPP may work with either Abis or with both Gb and Abis.

Where more than one RPP is used (except for the two RPPs in an active/standby configuration) they will communicate with each other using Ethernet. A cell cannot be split between two RPPs. If an RPP does not handle the cell to which the message is destined, the message is forwarded via the Ethernet to the right RPP. A duplicated Ethernet connection is provided in the backplane of the PCU magazine. In addition some HUB boards are needed to connect the RPPs via the Ethernet. The HUB boards are doubled for redundancy reasons.

2.9 HLR

The Home Location Register (HLR) is the database that holds subscription information for every MS subscribed to the GSM/GPRS operator. The HLR stores information for circuit-switched and packet-switched communication. The HLR contains information about, for example, supplementary services, authentication parameters, and whether or not packet communication is allowed. In addition, the HLR includes information about the location of the

MS. For GPRS, subscription information is exchanged between HLR and SGSN, and the HLR is used for the authentication procedure of the MSs.

The HLR stores the subscription information and location information regarding MS-terminated SMS, which includes the option of SMS message transfer via the SGSN to the MS. The Short Message Service Gateway MSC (SMS-GMSC) interrogates the HLR for location information and SMS information.

The information going from HLR to SGSN is what the operator sets up for the subscriber. This information transfer is done when the operator changes the subscription information, or when a new SGSN needs to have data for a subscriber after attachment or roaming. The old SGSN is also informed of the roaming. The information going from SGSN to HLR is the location information that is transferred upon MS action, for example attachment or roaming. For a roaming MS, the HLR may be in a different PLMN than the SGSN serving the MS.

2.10 MS

2.10.1 The MS in general

The combination of a Terminal Equipment (TE) and a Mobile Terminal (MT) forms a Mobile Station (MS). The term MS is often used when discussing the GPRS features. It can be concluded from the context, which parts would relate to the MT or the TE parts. Note that the MT and TE parts could actually be in the same piece of equipment.

GPRS MSs can operate in the following modes of operation:

- Class A mode of operation, i.e. the MS is attached to both GPRS and other GSM services, and the MS can operate both sets of services at a time.
- Class B mode of operation, i.e. the MS is attached to both GPRS and other GSM services, but the MS can only operate one set of services at a time.
- Class C mode of operation, i.e. the MS is exclusively attached to GPRS services or GSM services.

The Subscriber Identity Module (SIM) card for a GPRS MS has GPRS-specific functionality regarding ciphering, location information, and GPRS service.

2.10.2 TE

The TE is the computer terminal that the end-user works on. This is the component used by GPRS to transmit and receive end-user packet data. The TE could for example be a laptop computer. GPRS provides IP connectivity between the TE and an ISP or Corporate LAN connected to the GPRS network. The IP connection is static from the TE point of view, i.e. the TE is not aware of being mobile and retains its assigned IP address until the MT detaches.

2.10.3 MT

The MT communicates with a TE, and over the air with a BTS. The MT is associated with a subscriber in the GSM system, and establishes a link to an SGSN. Channel reselection is provided at the radio link between the MT and the SGSN. The MT can be compared to a modem, connecting the TE to the GPRS network. The MT must contain GPRS-specific software.

2.11 Interfaces

2.11.1 The GPRS interfaces in general

The GPRS system is mainly a net for transportation of information placed in packets. The various parts of the system communicate via interfaces. In the next chapters I present the interfaces important in respect to the various strategies that I have developed. In the figure below, the GPRS protocol stack in the transmission plane is presented.

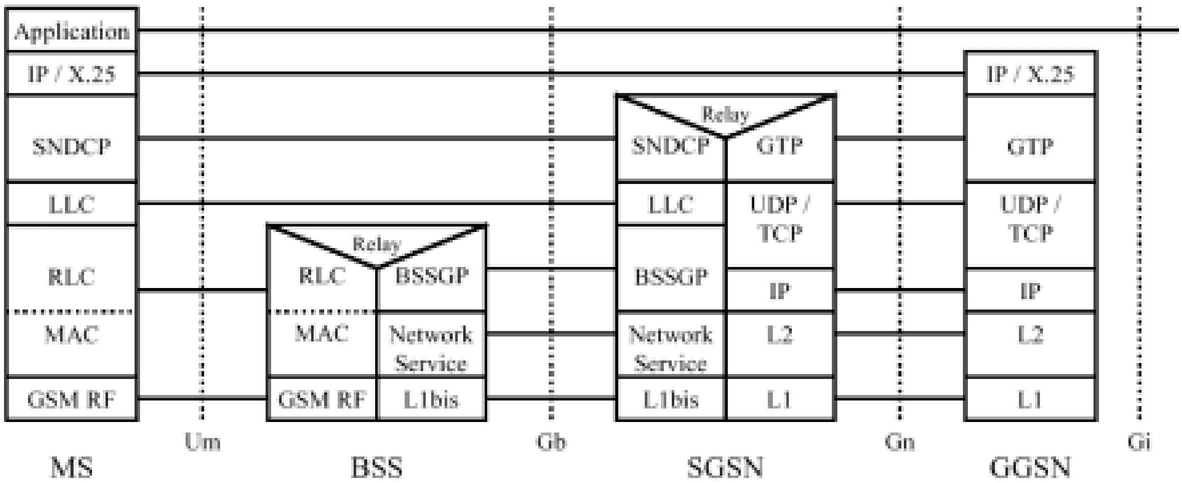

Figure 2.5 The GPRS protocol stack in the transmission plane

2.11.2 SGSN-BSS Gb

The Gb interface connects the SGSN to BSSs and MSs, allowing exchange of signalling information and data. There is one Gb interface per PCU in the BSS. If a BSS contains more than one PCU, the SGSN has a distinct Gb interface to each PCU in the BSS.

The protocol architecture on the Gb interface is divided into a lower part, which interconnects the BSS and the SGSN, and an upper part, which interconnects the MS and the SGSN. As seen in figure x, the protocol layers on top of the BSSGP are used for transfer of user or signalling data between the SGSN and the MS. They are transparently passed through the BSS.

BSSGP	
Network Service Control (NS Control)	
Frame Relay	
E1	T1

Figure 2.6 SGSN-BSS Interface Protocols

E1/T1 is the physical layer where data is transported. The SGSN can use one or many physical links (E1/T1 fractions) to connect to a BSS. Each physical link supports one or more Network Service Virtual Links (NS-VLs). Each NS-VL is supported by one physical link.

The Gb interface link layer is based on Frame Relay. One or many Frame Relay Permanent Virtual Circuits are established between the SGSN and the BSS for the transfer of signalling and user data. A Data Link Connection Identifier identifies the Frame Relay PVCs. Frame Relay and Network Service Control (NS Control) together make up the Network Service (NS) layer.

The NS Control protocol adds GPRS specific node management functionality and support for congestion and recovery handling on top of Frame Relay. Peer NS Control protocol instances communicate via such enhanced Frame Relay PVCs, which are called Network Service Virtual Connections (NS-VCs). At each side of the Gb interface there is a one-to-one correspondence between NS-VCs and NS-VLs. The NS-VC Identifier (NS-VCI) identifies each NS-VC end-to-end. Figure 2.7 describes the relationship between the NS-VCs and the NS-VLs.

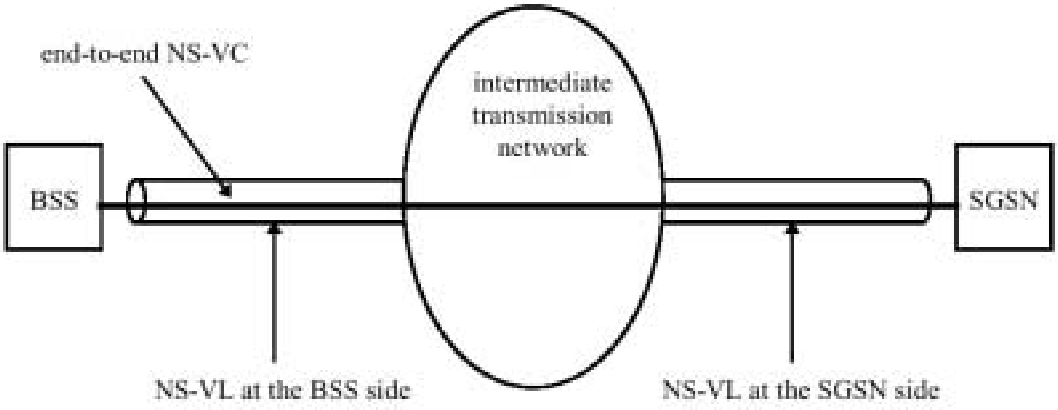

Figure 2.7 Relationship between NS-VCs and NS-VLs

The BSS GPRS Protocol (BSSGP) provides means to transport payload and signalling data for the mobile stations between the BSS and the SGSN. The peer BSSGP protocol instances communicate via one or more BSSGP Virtual Connections (BVCs). Each BVC is identified end-to-end by the BVC Identifier (BVCI). A load sharing function in the NS layer distributes the BVC traffic on the available NS-VCs. Figure 2.8 describes the relationship between BVCs and NV-VCs.

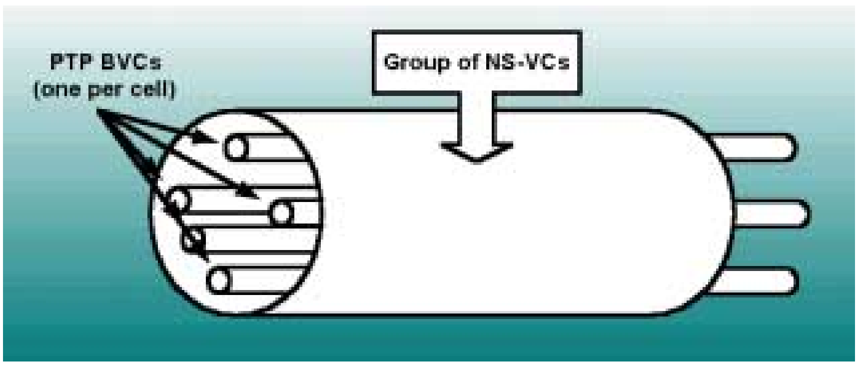

Figure 2.8 Relationship between BVCs and NS-VCs

For each GPRS supporting cell in the BSS area, BSSGP uses one Point-To-Point BVC (PTP BVC). All cell specific signalling data and user data or user signalling data for an MS residing in the specific cell is sent on the respective PTP BVC. All other signalling data is sent via the Signalling BVC. The BSSGP also provides means for BVC management and for flow control between the BSS and the SGSN for each cell and for each MS.

2.11.3 SGSN-MS

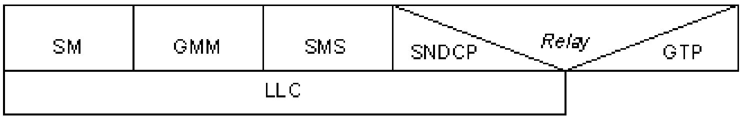

Figure 2.9 SGSN-MS Interface Protocols

The Logical Link Control (LLC) protocol provides a highly reliable ciphered logical link between the MS and the SGSN. It provides services for user data transfer used by the following three protocols. The Subnetwork Dependent Convergence (SNDCP) protocol implements segmentation and re-assembly of the IP payload packets. GPRS Mobility Management (GMM) uses the services of LLC to transfer mobility management related signalling messages, for example an attach request or a routing area update request between the MS and the SGSN.

The Session Management (SM) uses the services of LLC to transfer session management related signalling messages, like a PDP context activation request between the MS and the SGSN. The GPRS Tunnelling Protocol (GTP) tunnels user data and signalling between GSNs in the GPRS backbone. The relay function relays payload packets between the Gb and the Gn interfaces. The Short Message Service (SMS) uses the services of the LLC layer to transfer short messages between the MS and the SGSN.

2.11.4 PDP Context

In order to exchange data packets with external Public Data Networks (PDNs) after a successful GPRS attach, a mobile station must apply for one or more addresses used in the PDN, e.g., for an IP address in case the PDN is an IP network. This address is called Packet Data Protocol address (PDP address). For each session, a PDP context is created, which describes the characteristics of the session. It contains the PDP type (e.g., IPv4), the PDP address assigned to the mobile station (e.g. 129.187.222.10), the requested QoS, and the address of a GGSN that serves as the access point to the PDN. This context is stored in the MS, the SGSN, and the GGSN. With an active PDP context, the mobile station is "visible" for the external PDN and is able to send and receive data packets.

The mapping between the two addresses, PDP and International Mobile Subscriber Identity (IMSI), enables the GGSN to transfer data packets between PDN and MS. A user may have several simultaneous PDP contexts active at a given time. The allocation of the PDP address can be static or dynamic. In the first case, the network operator of the user's home-PLMN permanently assigns a PDP address to the user. In the second case, a PDP address is assigned to the user upon activation of a PDP context. The PDP address can be assigned by the operator of the user's home-PLMN (dynamic home-PLMN PDP address) or by the operator of the visited network (dynamic visited-PLMN PDP address). The home network operator decides which of the possible alternatives may be used. In case of dynamic PDP address assignment, the GGSN is responsible for the allocation and the activation/deactivation of the PDP addresses.

2.12 Important identities

2.12.1 General

Here are only the identities essential to the understanding of the different update strategies discussed in this thesis. Other identities important to the GPRS system can be found in GSM 03.60 [3].

2.12.2 Routing Area Identity

Routing Area Identity (RAI), defined by an operator, identifies one or several cells. RAI is broadcast as system information and is used by the MS to determine, when changing cell, if an RA border was crossed. If that was the case, the MS initiates the RA update procedure.

The location of an MS in STANDBY state is known in the SGSN on an RA level. Cells that do not support GPRS within an LA are grouped by the SGSN and BSS into a null RA. The MS is paged for packet services in the RA where the MS is located when mobile-terminated traffic arrives in the SGSN. The MS is paged for circuit-switched services by the SGSN in the last known RA plus in the null RA.

A Routing Area is a subset of one, and only one, Location Area (LA), meaning that an RA cannot span more than one LA. An RA is served by only one SGSN. The Routing Area Code (RAC) is only unique when presented together with LAI. RAI = MCC + MNC + LAC + RAC;

2.12.3 Cell Identity

The BSS and cell within the BSS is identified within a location or routing area by adding a Cell Identity (CI) to the location or routing area identification. CI is only unique when presented together with LAI or RAI.

The Cell Global Identification is the concatenation of the Location Area Identification and the Cell Identity. Cell Identity must be unique within a location area. CGI=LAI+CI

2.12.4 BVCI

BSSGP Virtual Connections (BVCs) provide communication paths between BSSGP entities. Each BVC is used in the transport of BSSGP PDUs between peer point-to-point (PTP) functional entities, peer point-to-multipoint (PTM) functional entities and peer signalling functional entities. The BSSGP Virtual Connection Identifier (BVCI) is used to enable the lower network service layer to efficiently route the BSSGP PDU to the peer entity. This parameter is not part of the BSSGP PDU across the Gb interface, but is used by the network service entity across the Gb interface.

A PTP functional entity is responsible for PTP user data transmission. There is one PTP functional entity per cell. A PTM functional entity is responsible for PTM user data transmission. There are one or more PTM functional entities per BSS. A signalling functional entity is responsible for other functions e.g. paging. There is only one signalling entity per Network Service Entity (NSE). There are one or more NSEs per BSS.

Each BVC is identified by means of a BSSGP Virtual Connection Identifier (BVCI), which has end-to-end significance across the Gb interface, Each BVCI is unique between two peer Network Service Entities.

In the BSS, it is possible to configure BVCIs statically by administrative means, or dynamically. In case of dynamic configuration, the BSSGP accepts any BVCI passed by the underlying Network Service entity.

At the SGSN side, BVCIs associated with PTP functional entities are dynamically configured. The BVCIs associated with signalling functional entities and PTM functional entities are statically configured.

2.13 Routing Area Update

2.13.1 General

A routing area update takes place when a GPRS-attached MS detects that it has entered a new RA, when the periodic RA update timer has expired, or when a suspended MS is not resumed by the BSS. The SGSN detects that it is an intra SGSN routing area update by noticing that it also handles the old RA. In this case, the SGSN has the necessary information about the MS and there is no need to inform the GGSNs or the HLR about the new MS location. A periodic RA update is always an intra SGSN routing area update.

Here, only the Inter SGSN RA Update procedure and the periodic RA and LA updates are presented, as these are important to the strategies discussed in the thesis. The definitions of the routing area update procedures may be found in GSM 03.60 [3]

2.13.2 Inter SGSN RA Update

This is a short version of the Inter SGSN RA Update based on the definition found in GSM 03.60 [3].

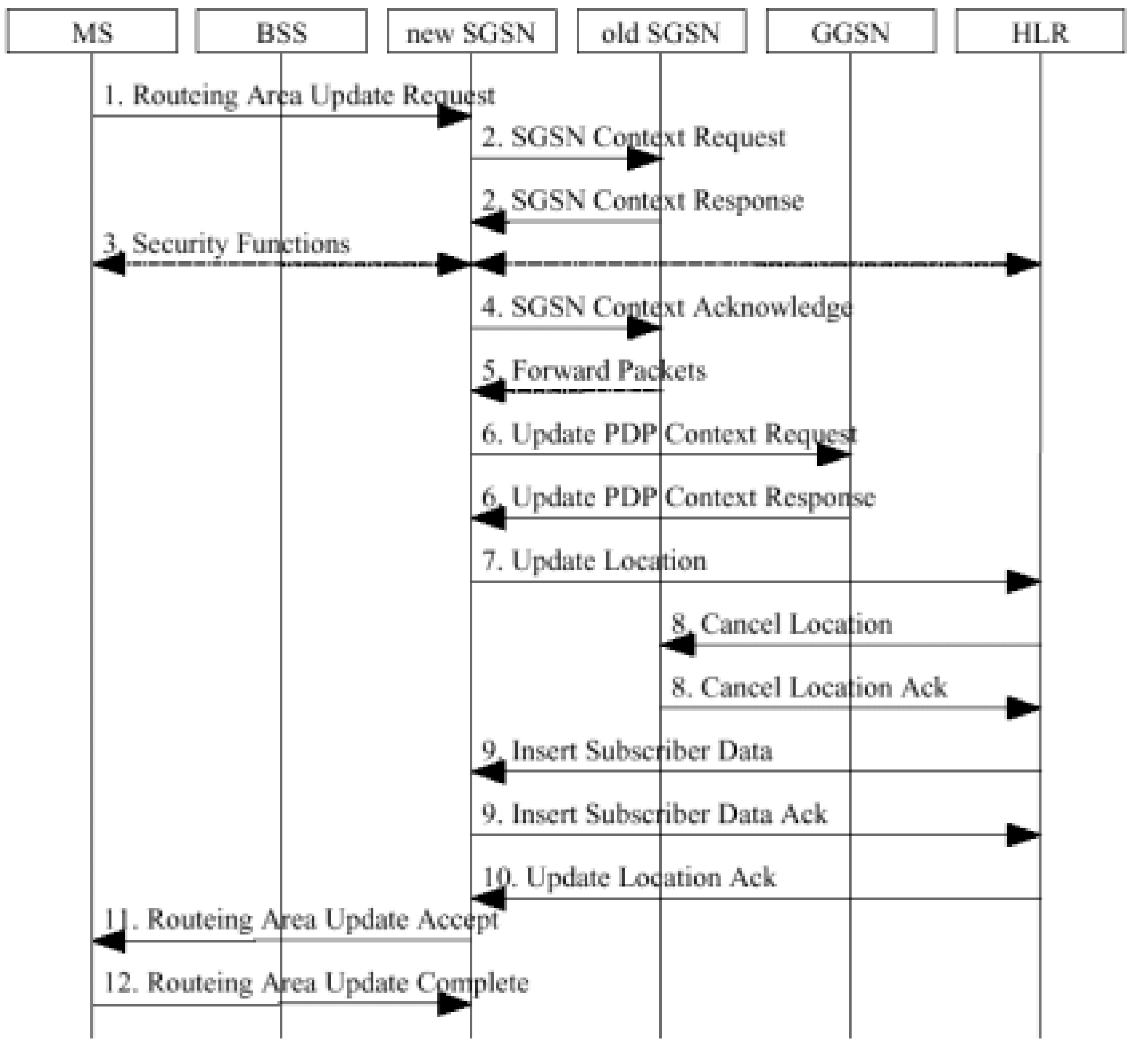

Figure 2.10 Inter SGSN Routing Area Update Procedure

1. The MS sends a Routeing Area Update Request (old RAI, old P-TMSI Signature, Update Type) to the new SGSN.
2. The new SGSN sends SGSN Context Request (old RAI, TLLI, old P-TMSI Signature, New SGSN Address) to the old SGSN to get the MM and PDP contexts for the MS. Provided the security functions in the new SGSN authenticate the MS correctly, the new SGSN sends an SGSN Context Request (old RAI, TLLI, MS Validated, New SGSN Address) message to the old SGSN.
3. Security functions may be executed. Ciphering mode is set, if ciphering is supported.
4. The new SGSN sends an SGSN Context Acknowledge message to the old SGSN.
5. The old SGSN duplicates the buffered N-PDUs and starts tunnelling them to the new SGSN.
6. The new SGSN sends Update PDP Context Request (new SGSN Address, TID. QoS Negotiated) to the GGSNs concerned. The GGSNs update their PDP context fields and return Update PDP Context Response (TID).
7. The new SGSN informs the HLR of the change of SGSN by sending Update Location (SGSN Number, SGSN Address, IMSI) to the HLR.
8. The HLR sends Cancel Location (IMSI, Cancellation Type) to the old SGSN with Cancellation Type set to Update Procedure. The old SGSN acknowledges with Cancel Location Ack (IMSI).
9. The HLR sends Insert Subscriber Data (IMSI, GPRS subscription data) to the new SGSN. The new SGSN validates the MS's presence in the (new) RA. If all checks are successful, then the SGSN constructs an MM context for the MS and returns an Insert Subscriber Data Ack (IMSI) message to the HLR.
10. The HLR acknowledges the Update Location by sending Update Location Ack (IMSI) to the new SGSN.
11. The new SGSN validates the MS's presence in the new RA. If all checks are successful, then the new SGSN constructs MM and PDP contexts for the MS. A logical link is established between the new SGSN and the MS. The new SGSN responds to the MS with Routeing Area Update Accept (P-TMSI, P-TMSI Signature, Receive N-PDU Number).
12. The MS acknowledges the new P-TMSI by returning a Routeing Area Update Complete (Receive N-PDU Number) message to the SGSN.

2.13.4 Periodic RA and LA Updates

All GPRS-attached MSs, except MSs in class-B mode of operation engaged in CS communication, performs periodic RA updates. MSs that are IMSI-attached performs periodic LA updates. Periodic RA updates are equivalent to intra SGSN RA updates with Update Type indicating periodic RA update. For MSs that are both IMSI-attached and GPRS-attached, the periodic updates depend on the mode of operation of the network.

The periodic RA update timer in the MS is stopped when an LLC PDU is sent since all sent LLC PDUs set the MM context state to READY. The periodic RA update timer is reset and started when the state returns to STANDBY.

If the MS could not successfully complete the periodic RA procedure after a retry scheme while the MS was in GPRS coverage, then the MS waits a backoff time equal to periodic LA update timer broadcast by the network before re-starting the periodic RA update procedure.

2.14 GPRS Attach

GPRS attach and PDP context activation must be executed in order for GPRS users to connect to external packet data networks.

The mobile terminal makes itself known to the network by means of GPRS attach – GPRS attach corresponds to IMSI attach, which is used for circuit switched traffic. Once the terminal is attached to the network, the network knows its location and capabilities. If the unit is a class A or class B terminal, then circuit switched IMSI attach can be performed at the same time (Figure 2.11).

1. The mobile terminal requests that it be attached to the network. The terminal's request, which is sent to the SGSN, indicates its multi-slot capabilities, the ciphering algorithms is supports, and whether it wants to attach to a packet switched service, a circuit switched service, or both.
2. Authentication is made between the terminal and the HLR.
3. Subscriber data from the HLR is inserted into the SGSN and the MSC/VLR.
4. The SGSN informs the terminal that it is attached to the network.

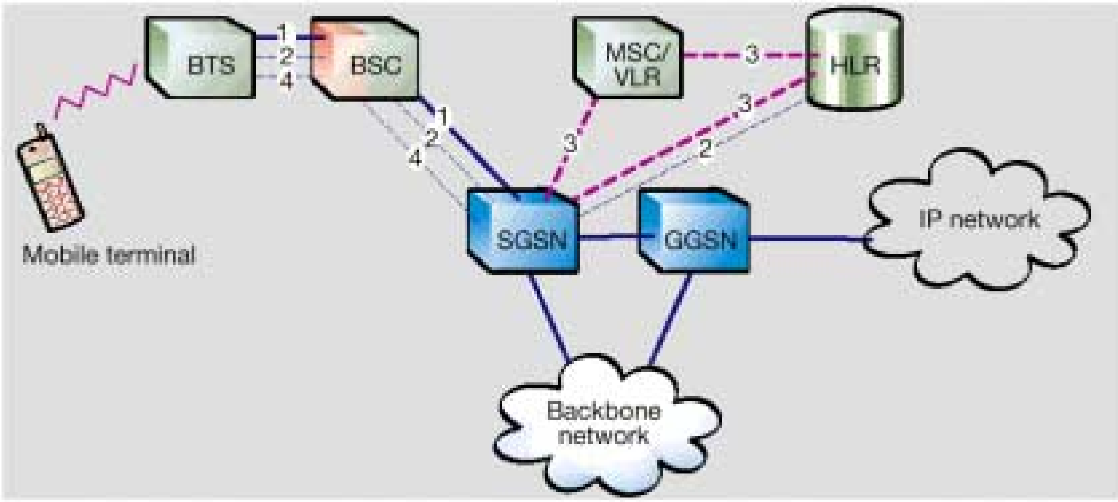

Figure 2.11 GPRS attach

Chapter 3 Possible upgrade strategies

3.1 Introduction

3.1.1 General

The assignment's first part is finding possible strategies for upgrading an SGSN node without the subscribers experiencing great loss of service by being offline while the upgrade process is executed. I will do this using network redundancy.

Redundancy means explicitly to have more than needed. Redundant data is often seen as unnecessary and inconvenient (except in backups), and redundant resources are not cost efficient. Therefore it is important that the system makes it possible to exploit the resources available as much as possible.

The SGSN nodes have, as previously described, connections to several other nodes in the GPRS system, both for data and signalling traffic. When I examined the different connections, I discovered that the link between the SGSN and PCU was allegedly a one-to-many relation. I.e. a PCU could only be connected to one SGSN, while the SGSN could be connected to many PCUs. I found, however, nothing in the ETSI standards to support this. This discovery is important for three of the solutions I have described, as they are based on the possibility of the PCU being connected to more than one SGSN node.

The reason why the Gb interface has gotten as much attention as it has, is due to the fact that the links from the MS to the SGSN are based on geographical location, while the rest of the network topology is not. Further, the SGSN node's primary task is to administrate subscribers' connections to the Core Network. Therefore it is natural to focus the solutions around the incoming connections from the subscriber.

On this background the Gb interface seems to be the bottleneck in exploiting redundancy on the SGSN nodes, as traffic cannot automatically be transferred to a different SGSN node when we need to upgrade (and thus take out of traffic) our SGSN.

The solutions below are structured in the way I have found best suitable. There are many elements that are part of several solutions, and the boundaries between the different solutions are not as distinct as one might wish. They are, however, fairly distinct from each other.

3.1.2 Evaluation criteria

There are some criteria for each solution making it better or lesser suitable for development. In the discussion of each solution I have used these criteria to point out good and bad qualities in the solutions. The criteria I have selected are the cost elements of each solution.

- Service interrupts:

The effect the upgrade strategy has on the service being provided to the users of the system is possibly the most important factor.

- Increased Life Cycle Cost (LLC):

LCC is the total cost of implementing the solution. E.g. how much new equipment is needed to realize the strategy.

- Change of standards:

This factor concerns whether the strategy makes it necessary to change any ETSI GPRS standards. This is important, as a change in the standards already agreed on, can drag an implementation of the strategy out in time. It is not regarded impossible to change a standard providing the benefits are so large they can justify this.

- Affected parts of the system:

What parts of the system are affected by the implementation of the strategy, and in what ways.

- Development problems:

The problems that have to be solved to enable an implementation of a strategy in the system.

3.1.3 Scenario

To simplify the presentation of each of the solutions, the descriptions are based on the scenario outlined below.

The SGSN we want to upgrade has capacity to handle 100.000 subscribers. We want to upgrade hardware and OS software on the node. The node has connections to, and is serving, 10 different BSSs.

The upgrade operation should take place when traffic in the network is low, to avoid breaks of traffic as much as possible. Therefore the process should start at approximately 01:00 and be over at 05:00 at the latest.

At this hour we picture around 50.000 subscribers are attached through the SGSN node with several PDP-contexts configured. Around 10% of these are engaged in traffic, transferring files or other information, while the others are dormant. Dormant subscribers are typically subscribers that are sleeping and not able to actively maintain PDP-contexts, but want to be reached by the network, for instance to be told when new mail arrives.

3.2 The present upgrade method

3.2.1 Description

The present method used for upgrading a particular SGSN node gives interrupt of service except when installing software patches not affecting the operating system (OS). If the node has to be restarted, one shuts down the SGSN, leaving all subscribers attached to this node through the applying BSC without service. They remain without service until the SGSN has finished its restart and they are able to reattach.

An estimated loss of revenue for a 12 minutes stoppage, assuming each subscriber in average uses GPRS for 35,- NOK, could represent 0,8 million NOK in lost revenues. The aspects are loss off traffic and charging data, additional ISP related revenues and extra operational cost. Also, there may be loss of subscribers, if another operator can offer better In Service Performance.

3.2.2 Discussion

The present method has an advantage over the other strategies, as it requires no change in procedures. No new investments have to be made to achieve this grade of in-service performance.

When the SGSN node is taken down, the subscribers attached to it will loose their connection, and while the SGSN node is down the subscribers have no connection until the SGSN has restarted and the MS's can reattach. Though the SGSN may be up and running in 10 minutes, the SGSN cannot process all reattach-requests simultaneously, and it may take as much as 40 minutes before all previously attached MSs are reconnected.

The methodology internally in the node related to upgrade can be improved and possibly reduce downtime, but OS upgrades and hardware upgrades will demand downtime.

3.2.3 Conclusion

While the solution has no cost, with large upgrades it leaves the subscribers without service for the duration of the restart, or until the subscribers move out of the area covered by the SGSN node.

3.3 Transmission-based approach

As I tried to sort out the various solutions, I found there was a relation in the solutions between the addressing of the SGSN nodes and technology required by the PCU. At an early stage this relation helped me categorize the solutions I ended up with, and it clarifies a main difference between the solutions. I have summarized this relation in the table below.

The term 'Identical SGSN nodes' means the nodes are perceived as one node from the outside. Only one node can be operative at once, and it should be possible to switch between the two without affecting the network. They have identical addresses and configuration data. The term 'Different SGSN nodes' means all nodes are identified from the outside. Both nodes are known in the network and both can be in operation simultaneously. They have different addresses and configuration data.

Table 3.1 Relations between SGSN-addresses and PCU-technology

	Identical SGSN nodes	**Different SGSN nodes**
PCU routes packets between several SGSN nodes	Not applicable	Subscriber-based PCU routing Routing Area-based PCU routing
PCU does not route packets	Hot standby SGSN	Redundant links

3.4 Hot standby SGSN

3.4.1 Description

We have two SGSN nodes, SGSN1 and SGSN2. SGSN2 is a hot standby node for SGSN1, i.e. it has an exact copy of both configuration data for the SGSN and connections data for the subscribers from SGSN1. The nodes have also identical connections outward. The SGSN2 is continually updated with changes in SGSN1s connections and configuration data. At a given time we shut down the SGSN1, and at the same time we tell the SGSN2 to act as the executing SGSN node. We then have the SGSN1 out of traffic and can carry out upgrades.

3.4.2 Discussion

As lined out above, we change executing SGSN nodes from SGSN1 to SGSN2 without interrupting user traffic or detaching Mobile Stations. There will probably be some packet loss at the switching moment, but this will depend on the switching procedure developed and will in any case not give breaks or long delays. We will need a dedicated standby SGSN2 node for each SGSN1 node. This is very expensive and does not take advantage of the redundancy already in the system. Another problem is the design of the SGSN node. Though there is much redundancy using hot standby built into the SGSN, this is not enough when the whole node has to be taken down. It has not been designed with mirror functionality (i.e. sharing RAM), and a new node architecture will probably have to be developed.

The part of the GPRS system affected by this strategy is the SGSN node. All other parts of the system view the two SGSN nodes as one, and all changes will be on the SGSN nodes. The development problem that is most obvious, while doing this initial study of the various strategies, is to have the SGSN nodes share the subscribers' connection-data in real-time (i.e. mirroring an SGSN node), and thus make it possible for the standby SGSN node to take over the execution at an instant. Another problem will be how to switch between the two SGSN nodes.

The upgrade does not have to be done in slack hours e.g. after midnight, as it is not affected by high traffic load, and there will be full redundancy if hardware or software fails.

3.4.3 Conclusion

Though this is a solution that provides a practically non-interruptive upgrade, it depends on new development before it can be utilized. It will also be expensive to have dedicated standby nodes only for upgrade purposes. All the time an SGSN2 node is not in use, will be a waste of resources. However, there will be full node redundancy in case of hardware and software failures, and upgrades can be done at all hours without traffic load regards.

3.5 Subscriber-based PCU routing

3.5.1 Description

We have two SGSN nodes, SGSN1 and SGSN2. The SGSN1 is a permanent node, which we from time to time want to upgrade. While upgrading we have to let traffic go through a different node, the SGSN2.

Initially all traffic goes through the SGSN1 node. At a certain point of time we connect a second SGSN node (SGSN2) to the BSSs (via the PCUs), the GGSNs and the other parts of the network. When this is done, we let the PCUs control each packet of data and route it to the suitable SGSN node, based on which node the subscriber that sends the packet is attached through. We make the PCUs send traffic from all new subscribers to the SGSN2. The new subscribers are in this case either new to the network or coming from a different SGSN node. At the same time, the PCUs send traffic from the 'old' subscribers to the SGSN1.

A movement out of the area operated by the SGSN1 and SGSN2 nodes would trigger an 'Inter SGSN RA Update'-procedure, initiated by a different SGSN node than SGSN1 and 2, and create no breaks. If this subscriber re-enters the area covered by the SGSN1 and 2, it will be treated as a new subscriber. And the SGSN2, now receiving all packets from new subscribers, will initiate a new 'Inter SGSN RA Update'-procedure.

With time, the number of subscribers still attached to the SGSN1 will move towards zero, as any new subscribers will be connected through the SGSN2. After a while there will be none or very few subscribers still attached through the SGSN1, and we can shut down this and carry out the upgrade. The few subscribers that get detached from the SGSN1 will try to reattach, and the PCUs will treat them as new subscribers, because the MS will send an 'Attach Request' to the SGSN.

After the upgrade, the whole procedure is reversed to move all subscribers back to the SGSN1. Figure 3.1 shows a simplified version of the SBPR solution where all 10 BSSs are represented as one (BSS1).

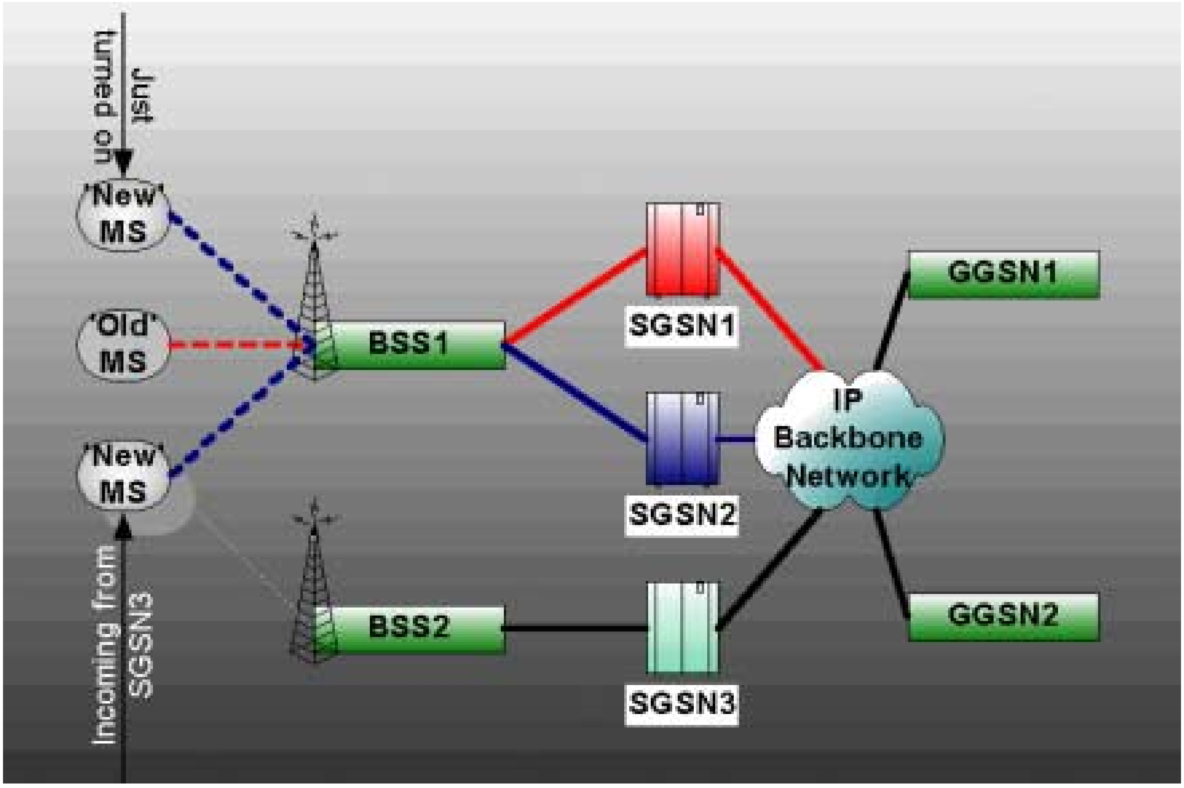

Figure 3.1 Subscriber-based PCU routing

3.5.2 Variations

There are some variations of this solution. Variation 1 is to take down the SGSN1 node although there still are some subscribers attached. These will experience a break in service and have to reattach. Variation 2 of the solution above is to, at a suitable time, move the remaining subscribers to the SGSN2 by an Inter SGSN RA Update (ISRAU) or a similar procedure that updates the network, as the subscriber changes SGSN. In this way none of the subscribers will experience service interrupts.

3.5.3 Discussion

This solution stands and falls upon whether it can be made possible for the PCUs to recognize which subscriber a packet belongs to and thereupon pass it on to the right SGSN node. There are three main features that do not exist in the PCU and need to be developed, to make this work:

- The PCU must be connected to, and able to differentiate between several (at least two) SGSN nodes.
- The PCU must be able to recognize the dispatcher of every packet passing through it.
- The PCU must know which dispatchers that are 'old' subscribers and route these packets to the SGSN1 node.

The PCU has to open every packet passing through, to find which subscriber it comes from. This will produce a huge workload on the PCU, and if it cannot take care of this the result will be delays or thrown packets.

Yet another development problem is in the event that an MS moves out of the area covered by the SGSN1 and SGSN2. The SGSN node covering the area the MS enters must know which of the SGSN1 or 2 to ask for subscriber information.

The variation 2 implies the development of a procedure to move the remaining subscribers to the new SGSN from SGSN1. This must be done without service interrupts for the subscribers. Although a GPRS attach (Figure 3.4) is demanding on the SGSN node, an ISRAU is much mode demanding both on CPU in the SGSN and on the rest of the core network.

During the four hour long scenario we cannot expect many subscribers to change SGSN nodes themselves by crossing SGSN coverage zones. Therefore this solution should go over a longer period of time, so at least 20% change nodes themselves. With 20% of 50.000 subscribers over on the new node, we are left with 40.000 subscribers that have to be moved otherwise. A moving procedure must therefore be developed in a way that one avoids extreme loads on the node and the network. It then becomes a matter of judgement whether the 40.000 should be cut off or moved, hopefully without breaks. It is probable that there is a majority of those who have changed SGSN nodes themselves that are also engaged in traffic. This is good, because it is the active subscribers that will suffer the most from the service interrupts.

Any service interrupts will happen twice, as the subscribers also have to be moved back to the SGSN1 when the upgrade is over.

Any detached MSs that are not actively trying to communicate over the network will, when the periodic RA timer expires after maximum 54 minutes, discover they are detached and try to reattach. The most unfortunate subscriber will experience a break of service lasting 2 x 53 minutes, as the MS has to discover it is detached twice. The MSs will in some cases be dependent on the user to set up the PDP-contexts again, and as the upgrade happens at night, these will mostly be created in the morning.

In addition to the development cost, there will be need for at least one dedicated SGSN2 node in the network to replace the SGSN1 while these are being upgraded one by one. It will probably be normal with one SGSN2 at every geographical location where there is one or more SGSN1 nodes. This means the Life Cycle Cost could be greatly increased, as we need to have an extra node at each SGSN1 location.

The ‘Subscriber-based PCU routing’ solution will demand some change in the standards for the Gb interface. There is a considerable difference between using the PCU only to forward incoming packets and using the PCU as an intelligent router, and the PTP BVCs that logically connect each cell to the SGSN must be changed to make the routing function possible. Also variation 2 will demand a change or appendage to the standards with a new procedure to move subscribers to a new SGSN.

The benefits of this solution are big, however. Only very few subscribers need to experience interrupts in service, as most subscribers will be moved smoothly to the SGSN2 from other SGSN nodes via the ‘Inter SGSN RA Update’-procedure. This is provided the routing function in the PCU is given enough time, so as many subscribers as possible can get moved to the new SGSN2. With Variation 2 of the solution the remaining ‘old’ subscribers may also be moved to the SGSN2 without interrupts.

3.5.4 Conclusion

The solution is very demanding on the PCU, and the development costs will be high. Many subscribers may experience interrupts of service, unless one uses the Variation 2. But with Variation 2 there must be calculated with even more development costs. The benefits of the Variation 2 are none or very few breaks, provided the procedure to move remaining subscribers works.

3.6 Routing Area-based PCU routing

3.6.1 Description

As the title suggests, this solution uses the Routing Area Code (RAC) to smoothly transfer subscribers from a SGSN node we want to upgrade, to a new SGSN node.

A number of cells are served through a PCU. All the cells have the same RAC, but can change this when needed. For simplicity, the RAC for these cells is RA1. Two SGSN nodes (SGSN1 and SGSN2) are also connected to the PCU. The SGSN nodes have been assigned two different RAs, and the SGSN1 covers the RA1 while the SGSN2 has been assigned RA2. As no cells attached through the PCU initially has Routing Area Identity RA2, the SGSN2 receives no traffic.

The PCU has functionality to route all traffic coming from cells with RA1 to the SGSN node assigned the RA1, and it can in the same way route the traffic coming from cells with RA2 to the SGSN2 node. We can now change the RAC for one or several cells from RA1 to RA2. This forces any MSs inside these cells to send a ‘Routing Area Update Request’ as it is led to believe it has moved into a new RA. This request comes from a cell with RA2 identity and the traffic is thus sent to the SGSN2. In order we change the RAC of all cells served by the SGSN1 to RA2, so they are served by the SGSN2. When all traffic has been moved to SGSN2 we can upgrade the SGSN1. Afterwards, the operation is reversed, as we change RAC of all the cells back to RA1 from RA2.

Below is a figure showing a simplified illustration of how Cell2 changes RAC from RA1 to RA2. Immediately all traffic from that cell is sent to the SGSN2 instead of the SGSN1. SGSN2 receives a Routing Area Update Request from all the MSs in Cell2, and initiates accordingly an Inter SGSN RA Update Procedure towards the SGSN1 for all MSs in Cell2.

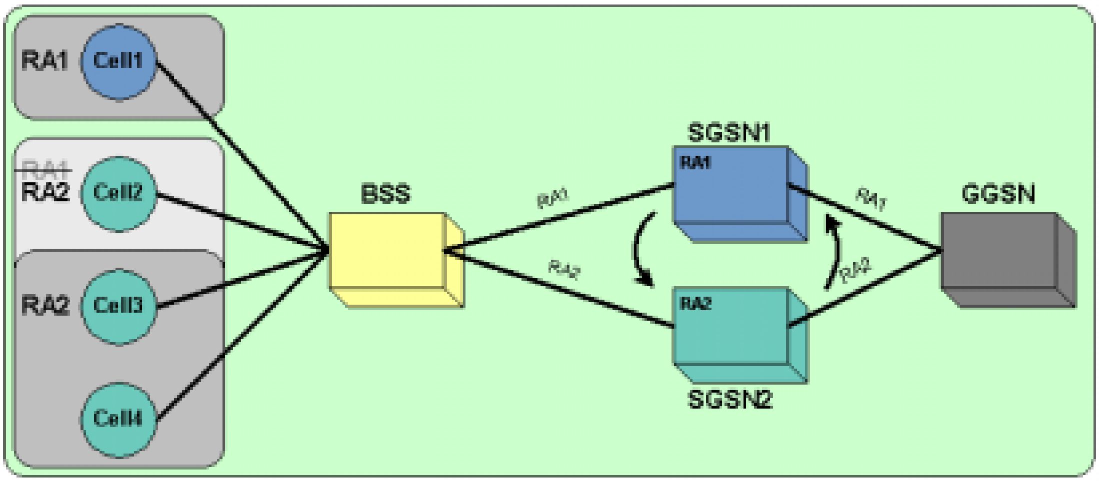

Figure 3.2 Routing area change forces an Inter SGSN RA Update

3.6.2 Discussion

This solution depends on functionality in the PCUs to route traffic to different SGSN nodes. The PCUs need to know what Routing Area the packet comes from, and based on this be able to route it to the correct SGSN node.

The development problems most obvious here are:

- The PCUs must be able to be connected to several SGSN nodes simultaneously, and route packets to either SGSN.
- The RAC of a cell must be changeable without implying downtime for the cell or the BSS.
- The PCUs must know what RA a packet comes from, without having to open it.

There is need for at least one new SGSN node in the network. It is probably practical with one extra SGSN node at every location where there is one or more SGSNs.

The solution represents changes mainly to the PCU equipment, but the ETSI standards for the PCU will not necessarily have to be altered when implementing this solution. This is because the standards describe the communication interface between the PCU and the other nodes, rather than the functions performed by the PCU. However, the Gb interface with its PTP BVCs could become a problem with respect to routing functionality on the PCU, as they link each cell point-to-point to the SGSN.

Initially, neither breaks nor delays were expected with this solution, because it depends on a standard Inter SGSN RA Update to transfer the subscribers' connection data between the SGSN nodes. However, the possibility to change the Routing Area of one and one cell at a time does not necessarily ensure that there will not be an overload of traffic generated by too many Inter SGSN RA Update procedures at once. Also, there is a question whether the RAC of cells may be changed without implying downtime for the cell or BSS.

3.6.3 Conclusion

Though there is anticipated a great deal of development costs on the PCU RA routing function, the solution bases itself on functions already implemented in the system. Probably one will avoid changes of standards, and although the impact the solution will have on service interrupts is uncertain, it is seemingly very positive.

3.7 Redundant links

3.7.1 Description

The interface between PCU and SGSN is also the subject in this last solution. However, while the other solutions have either been based on a PCU aware of there being more than one SGSN connected to it, or the SGSN nodes having the same configuration and address, this solution does not. Here the PCU is not aware of the existence of more than one SGSN, i.e. the PCU just relays the packets while the logical links provide the routing functionality. Also, the SGSN nodes have different addresses.

We have two SGSN nodes, SGSN1 and SGSN2. The SGSN1 is the operating node, while the SGSN2 is a replacement node to take over traffic from the SGSN1 while this is being upgraded. We have a connection between each of the SGSN nodes and the PCU. These are both physically and logically different links, and only one of them is open at a time.

We start by having the link between the SGSN1 and the PCU open. Then, at a certain time, we open the link between the PCU and the SGSN2 while we at the same time close the PCU-SGSN1 link. This is done by means of operations and maintenance (O&M) on the Gb interface. Thereby all new packets coming from the PCU ends up at the SGSN2, and the SGSN1 is free for upgrade procedures. The SGSN2 will ignore all packets except Attach Requests and RA update requests. These will be sent by the MSs when they have timed out waiting for response from the SGSN1. When an Attach Request is received, the MS will be connected to the Core Network through the SGSN2, and the SGSN1 node is history.

When upgrade of the SGSN1 is finished, the procedure is done over again, with switched SGSN1 and 2 to transfer all subscribers back to the SGSN1.

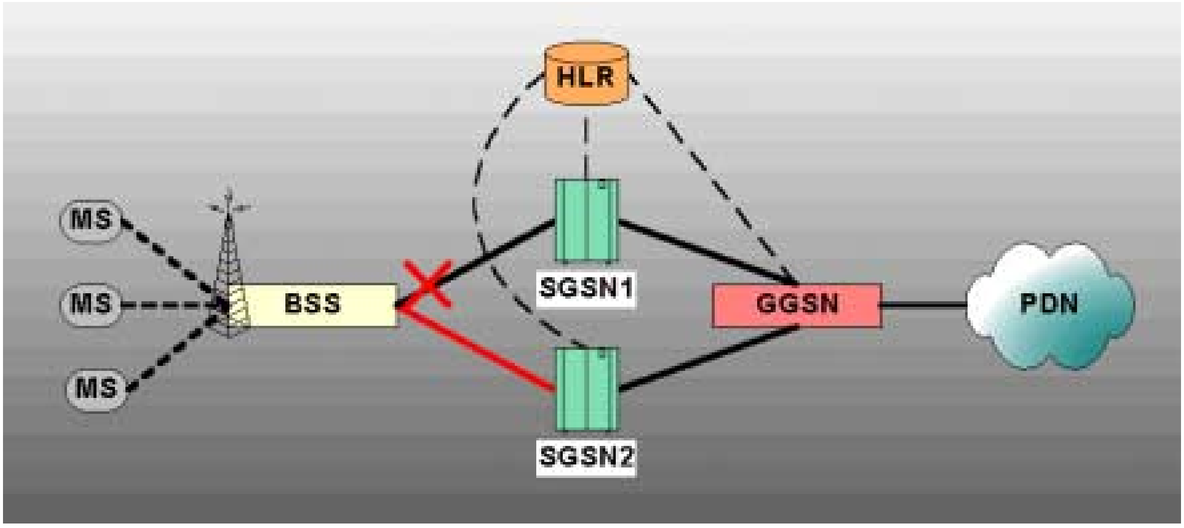

Figure 3.3 Redundant links on the PCU SGSN interface

3.7.2 Discussion

The investment costs developing the solution will probably be small, as everything concerning this solution already is standardized. The increase in LCC will consist mainly of new replacement SGSN nodes and cables to link the PCUs to the SGSNs.

What is the major drawback in this solution is the service performance. The solution is based upon breaking all the subscribers' connections, and then let every subscriber self find out that it has been detached. All MSs will discover the break simultaneously, except for the ones not using the service at the time. These will believe they are still attached to the network without being this. As long as they neither are attached to the CN nor trying to get attached, they are unreachable for anyone trying to reach them.

There is a periodic RA update timer in the MS that makes the MS check the network every 54 minutes to confirm that all connections are alright. All the MSs that were not engaged in traffic when the SGSN1 went down will discover that their connections are gone when they send the periodic RA update. Then they will do a GPRS attach. In worst case a subscriber will believe all is ok for 53 minutes after the SGSN1 node is shut down.

The solution takes only care of routing in the MS – SGSN perspective and leaves the SGSN – GGSN routing (i.e. incoming traffic to the MS) to take care of itself. This means that when the SGSN1 node has been taken down, the PDP contexts will still point towards the SGSN1, and packets will be sent until the GGSN discovers that the SGSN1 no longer can be reached. It discovers this by means of a Path Echo Request timeout, as described in GSM 09.60 [8]. Then, all PDP-contexts to the SGSN1 node are deleted.

Worse still is the vast number of Attach Request that will be received by the SGSN2 node as the MSs try to reconnect. In a worst case scenario an unlucky subscriber will have to wait 20 minutes after sending an attach request before getting reattached, but first it has to discover it needs to reattach. The reason why it could take this long is that the SGSN has to treat requests from the already connected subscribers, while attaching new subscribers. The attach curve below illustrates the problem. However, it should be pointed out that this curve only visualizes the problem. Figures and scale have not been validated.

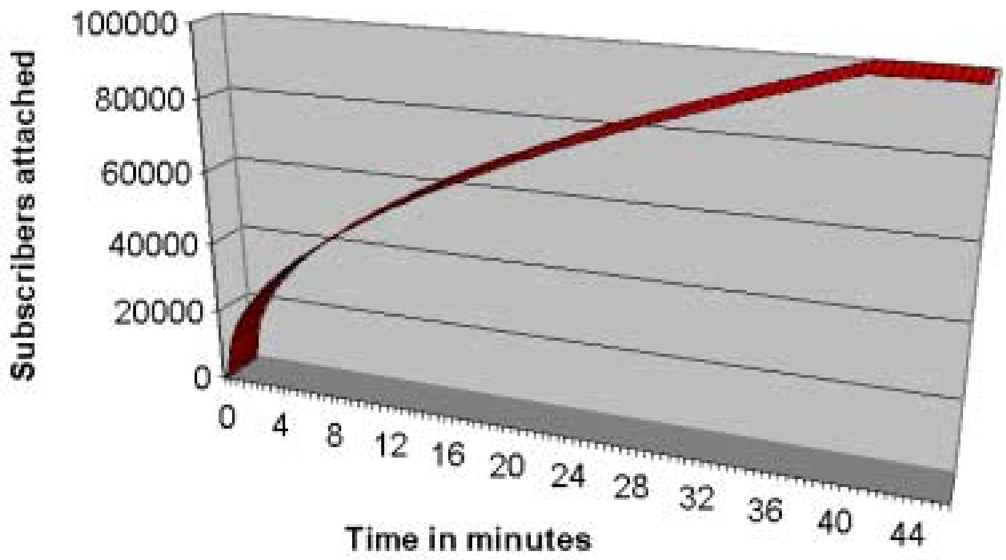

Figure 3.4 Attach curve for reattaching 100.000 subscribers

Finally, as the SGSN2 node is a temporary node, traffic has to be transferred back to the SGSN1 node when the upgrade procedure is done. This means another break-and-reattach for all subscribers. Some subscribers will experience breaks for up to 106 minutes if they are unlucky enough to do a periodic RA update request moments before the SGSNs are taken down. (These will not experience long attach delays, as the rush of attaches will be over by then.) This is an upgrade highly dependent on being carried out at slack hours.

The average downtime for the subscribers, taken into consideration that 10% are engaged in traffic and discovers the break at once, will be:

45.000 subscribers that are not engaged in traffic will, based on how long it has been since the periodic RA timer was reset, have an average break of 26 minutes.
5.000 subscribers engaged in traffic discover the break at once and has an average break of 1 minute.

45.000 subscribers with an average break of	26 min
5.000 subscribers with an average break of	1 min
Total average break ((45000*26min + 5000*1min)/50000)	~24 min

This is, however, based on only one reattach, and since the solution requires two reattaches the total average break is 2 x 24 minutes.

3.7.3 Conclusion

Though the cost of implementing this solution will be minor, the average subscriber will experience breaks for 2 x 24 minutes with a scenario where 50.000 subscribers are attached to the SGSN.

3.8 Comparison of the solutions

3.8.1 General

When comparing the different upgrade strategies, I will use the same criteria used in the discussion of each strategy. These can be found in chapter 3.1.2.

Table 3.2 The various solutions with important criteria

Upgrade strategy	Service interrupts	Increased Life Cycle Cost	Change of standards	Affected parts of the network	Development problems
Hot standby	Probably some packet loss at the switching moment, but not affected by high traffic load	Demands a dedicated second SGSN for each SGSN	None	SGSN	• How to mirror an SGSN • How to switch between the SGSNs
Subscriber based PCU routing	Var1: Some subscribers experience a break. This happens twice. Var2: Hopefully none, but probably some.	Needs at least one dedicated new SGSN node in the network. Probably one at each SGSN location.	The Gb interface must be changed to allow the PCU to route packets. With var2 the new procedure to move subscribers should be added to the standards	PCU and SGSN	• The PCU must be able to handle several SGSN nodes and route packets to each, based on subscribers. • The PCU must know new and old subscribers • There must be capacity in the PCU to open every packet • Routing Area problems with ISRAU • Var2: A procedure to move the remaining subscribers to the new SGSN
RA based PCU routing	Dependent on the ability to change RAC in a cell without breaks and a limited number of subscribers in each cell, so the number of ISRAUs is possible to handle. (Hopefully no service interrupts)	Needs at least one dedicated new SGSN node in the network. Probably one at each SGSN location.	Maybe on the Gb interface to enable PCU routing	PCU, SGSN, BTS and cells	• The PCU must be able to handle several SGSN nodes and route packets to each, based on Routing Area. • Var2: The SGSN must share a PCU node with another SGSN • The RAC of a cell must be changeable without implying downtime for the cell or the BSS. • The PCU must know what RA a packet comes from, without having to open it.
Redundant links	All subscribers must reattach twice, and they must also discover that they need to reattach. (Average break: 2x24 min)	Needs at least one dedicated new SGSN node in the network. Probably one at each SGSN location.	No	SGSN	No problems needed to be solved before utilizing the solution.

3.8.2 Service interrupts

The service interrupts should be kept at a minimum, and preferably be eliminated. The Hot standby solution offers an almost interrupt free upgrade. There will probably be some packet loss at the switching moment, but compared to Ericsson's goal of 99.999% In Service Performance this will be well inside the limit. The Subscriber based PCU routing (SBPR) solution with variation 1 allows that some subscribers be cut off from the SGSN1 node without knowing it. How many is hard to say, but if we refer to the scenario of 50.000 subscribers attach through the SGSN and the fact that only a few will cross a SGSN border in the four hours of upgrade, this number could be very high. Compared to this, the Hot standby solution is clearly better.

The variation 2 of the SBPR solution, on the other hand, provides a solution to the problem with variation 1, and should give an interrupt-free upgrade. This is only true, however, if the SGSN1 and 2 are able to cope with the sudden transfer of all subscribers remaining on the SGSN1 node. If this holds, then the SBPR solution with variation 2 is better than both the Hot standby and SBPR var1. However, it is more probable that there will be some interrupts also with SBPR var2, and the Hot standby solution will be better.

The Routing Area based PCU routing (RAPR) solution offers a smooth transfer of subscribers to the SGSN2, and should also give no service interrupts. This is only true if RAC change of the cells can be done without downtime. In that case, it will provide the best In Service Performance along with the Hot standby solution. Offering the largest amount of service interrupts is the Redundant Links (RL) solution. Here we anticipate 2 x 24 minutes for the average subscriber.

From this we conclude that both the RAPR and the Hot standby solutions offer the best In Service Performance, with the SBPR solution as number three. The RL solution comes out very unfavourable here.

3.8.3 Life Cycle Cost

Though the favourite solution should be as low as possible on the increase of LCC, obviously there has to be enough capacity in the system to let one SGSN node go out of traffic.

The Hot standby solution demands a dedicated second SGSN node. This is a very high demand. The SBPR solution does not require more than one extra SGSN node in the network in theory. In practice, however, one extra SGSN on every location where there is already a SGSN will probably be best. This is also the case with the RAPR and RL solutions.

Consequently, the solution having the largest LCC increase is the Hot standby solution with an extra SGSN node for every existing SGSN in the network. Ergo, the other three variations come out better than the Hot standby here.

3.8.4 Change of standards

As a change of already accepted standards would imply a lot of extra time to implement, this should be avoided to the extent possible. However, if the benefits of a change in the standards are great enough, then the change should be carried out.

The Hot standby solution demands a change in the architecture on the SGSN node, since the node hasn't been designed with a mirror function in mind. However, there will not be need for a change in the ETSI standards, as these do not describe node architecture. The Gb interface probably has to be changed to meet the demands of the SBPR solution, as the PCU must be able to differ between packets coming from different subscribers, and route these to different SGSN nodes. In Variation 2 also the SGSN requires a change in standards, as a function for moving the remaining subscribers needs to be implemented. The RAPR solution will possibly require a change in the Gb standards, as it must be able to route packets from different RAs to different SGSN nodes.

The only solution clearly avoiding a change of standards is the RL solution, which is founded only on technology offered by existing standards, but also the Hot standby avoids standards change. Thereby the RL and the Hot standby solutions come best out here.

3.8.5 Affected parts of the network

The solution is considered better if it affects as few parts of the system as little as possible.

The Hot standby solution will affect only the SGSN nodes. All other parts regard the paired SGSN nodes as one node, and thus are not affected. The SBPR solution affects the PCU, as this has to route packets to different SGSNs based on dispatching subscriber. Also the Gb interface must probably be changed. This would in worst case affect SGSN and the entire BSS. The RAPR solution is probably the solution that affects most parts of the system. The PCU has to route packets based on RA to different SGSNs and is thus affected. Probably, the Gb interface must also be altered. Further, this solution calls for the cells to change RAC. The RL solution affects the operations and maintenance (O&M) of the Gb interface.

The SGSN node is not considered affected by other than the Hot standby solution, as it automatically configures itself with respect to issues introduced by the solutions, other than configuration of a new SGSN in the network.

The RL solution along with the Hot standby solution comes best out of this evaluation, both only affecting one part of the network, the Gb interface and the SGSN node respectively.

3.8.6 Development and cost

It is preferable to have as simple and low cost solutions as possible.

There are two main problems that need to be solved with the Hot standby solution. First of all, the possibility to mirror an SGSN must be attained. This has been done on AXE centrals, and is thus proven not to be impossible. Second, there must be some way of switching between the SGSN nodes, provided the mirror problem has been solved, to change operating SGSN node. Here one probably faces packet loss. Compared to the other solutions' development problems, the Hot standby has the most extensive, as it requires redesigning the node-architecture.

The SBPR solution presents several challenges. It must be possible for the PCU to handle several SGSN nodes and route packets to each, based on subscribers. This requires the PCU to differ between old and new subscribers and have capacity in the PCU to open every packet. Another problem demanding attention is in the event of an MS moving out of the area covered by the SGSN1 and SGSN2. The new SGSN node has to know which of the SGSN1 or 2 to ask for subscriber information. Finally the Variation 2 of the solution depends on a function to transfer subscribers smoothly from SGSN1 to 2, much in the same way as the ISRAU procedure, being developed.

The RAPR solution has problems with the PCU, much in the same way as the SBPR solution. The PCU must be able to handle several SGSN nodes and route packets to each, based on Routing Area. Further, the PCU must know what RA a packet comes from, without having to open it. Another problem is that the RA of a cell must be changeable without implying downtime for the cell or the BSS.

The RL solution has no development problems needed to be solved, as it is founded only on technology already in use. Thus it comes best out of the development problems discussion.

3.8.7 Conclusion

It is the best solution of the four discussed in length above that I will elaborate further on, in the second part of this thesis. Each of the criteria used for discussion will not be equally weighted when evaluating these, though they are all important. The main factor is the service interrupts, which should be as small and few as possible.

Reviewing the discussions of the solutions compared against each other based on the factors above, we have the RL solution being the best in four areas. However, it would be wrong to pronounce this as the best solution, as it probably would give more total service interrupts than the present solution. The Hot standby solution is better than the RL on an overall basis, when weighting the service interrupts high, but it has a big increase in LCC, and its development problems are the greatest of all the solutions.

We are therefore left with the RAPR and the SBPR solution var2, as these offer a very low degree of service interrupts, are low on increase in LCC and are considered feasible with respect to the development problems. The solutions are similar, but the routing mechanism is different in the two. The RAPR routes on the RAI, while the SBPR routes on subscribers. There is a big problem with the SBPR solution’s routing mechanism, as the PCU would have to open each and every packet to see which subscriber it belongs to. This is because the subscriber info lies in the BSSGP layer. The RA information, however, is known by the BSS as it is defined there. The RAPR solution gives the impression of being a solution exploiting existing functions in the GPRS system instead of depending on future functions. It is unlikely that the SBPR solution will be any use without the ISRAU like function to help move subscribers, who have not themselves provoked an ISRAU, over to the new SGSN.

Therefore I have chosen to elaborate further on the RAPR solution.

Chapter 4 Implementing an RA based PCU routing solution

4.1 Introduction

In this chapter, which is the second part of my Masters Thesis, I will elaborate on the strategy recognized as the best one in the preceding analysis, namely the RA based PCU routing solution. I will try to give a clear picture of what has to be done to implement this solution in the GPRS network.

Though it seemed evident that a change in standards was required in order to achieve a routing function from the PCU to several SGSNs, this might not be the case after all. It may be possible to exploit the PTP BVC to route packets from cells, based on RA.

4.2 Implementing

4.2.1 General

The implementation of the RAPR solution affects several parts of the GPRS system. Below I will try to present a structured view on the implementation both in different parts of the system and at different levels. As I do this thesis on assignment from Ericsson, I have used function specifications for Ericsson's SGSN nodes. The functions may differ from what other companies' nodes have.

The subchapters below cover: configuration of the SGSN nodes required to change because of the implementation of the solution, routing functionality from the PCU to the SGSNs, change of RA in cells, discussion of these aspects and a conclusion.

4.2.2 SGSN node configuration

The variation 1 of the solution demands a new SGSN2 node to be used as a replacement node for the SGSN1. For the other variations there will be need to configure the cooperating RAs and SGSNs in the same way as for a new SGSN in variation1. I will not go into all network configuration attributes, but emphasise the ones affecting RA-configuration.

A new SGSN node must be configured with its own address and knowledge of its surrounding, or cooperating, SGSN nodes. In order to allow a mobile station to perform inter SGSN routing area updates while travelling all over the PLMN, at least those other SGSNs which own routing areas that are adjacent to any local routing area should be defined as cooperating SGSNs. In the same way at least those routing areas that reside within the service area of another SGSN, and that are adjacent to any local routing area should be defined as cooperating routing areas. This is to allow a mobile station to perform routing area updates. Local routing areas are the RAs served by SGSN2.

The local RAs are created automatically in the SGSN by the system, through BSSGP node management and cannot be modified by the operator. A local RA exists if and only if there exists at least one PTP BVC, which supports a cell within the local RA, and PTP BVCs are dynamically created on the SGSN without any possible interaction from the operator. Thus the operator cannot modify the support of RAs on the SGSN. However, the operator can statically create PTP BVCs from the BSS.

4.2.3 Routing in the PCU

The packets going between the MS and the SGSN nodes need to be routed from the PCU to the different SGSNs. In the previous chapter the routing function of the PCU was considered craving a change in the standards, allowing new software in the PCU to route the packets. However, it may be possible to use the BVC connections for this.

Exploiting the fact that there is a PTP BVC between each GPRS cell and the SGSN node can solve the routing of packets to different SGSNs, based on RA. One can set up two PTP BVCs between each cell and the SGSN, where the BVC1 is set to support the RA1 which the cell is currently set up with. The BVC1 is a connection between the SGSN1 node and the cell. The other BVC2 is set to support the RA2, which the cell will be configured to support when we want to move traffic from SGSN1 to SGSN2. This BVC2 is a connection between the cell and the SGSN2.

In effect only one of the two PTP BVCs we have created per cell will be active at a time. The reason is: The short cell id used when defining a PTP BVC. This short cell id consists of the CI and the RAI. If we create one PTP BVC between the SGSN1 to the CI+RA1 and another between the SGSN2 and the CI+RA2, then the first PTP BVC will be used when the cell is defined with RA1, and the second will be used when the cell is defined with RA2.

The PTP BVCs are dynamically created in the SGSN, but can be statically created from the BSS side. Ergo all PTP BVCs that point to a CI+RAI that does not exist at the given time, must be created on the BSS.

4.2.4 Change of cell RA

When it is time to transfer subscribers from the SGSN1 to the SGSN2, we can do this on the cell-level. We change the cell's RAC, with the result that the MS discovers it is in a new RA. I have not been able to confirm that the cell's RAC can be changed without restart of the cell. In that case the subscribers in the serving cell will experience a detachment and need to attach again to the GPRS network.

4.2.5 Discussion

Although this solution is seemingly without problems, there are some aspects that need discussion.

The load may get too high on the SGSN nodes and the backbone. Although we change RA for one cell at a time, there are no limitations to how many subscribers the cell can hold at a given time. For instance, at a concert or a football match many thousands may be inside the coverage zone of one cell. If that cell all of a sudden changes RA, all MSs inside the cell will send, or try to send, a RA Update Request, resulting in equally many ISRAU procedures. Although the SGSN most lightly has overload protection, many subscribers will be offline until they are properly moved to the new SGSN.

One SGSN node serves thousands of cells. Changing the RA of all the cells must be done automatically and remotely. If the change is to be done manually and locally at each cell, the operation will use much more time and manpower. This may make the cost of the solution to high on Ericsson's part.

It must be possible to change the RA of a cell without implying downtime for the cell. In case the cell has to be restarted, this would give a break or delay in service for the subscriber.

4.2.6 Conclusion

Though the solution offers an upgrade without service interrupts, this is only true for a limited number of subscribers inside a cell. Also, the cost of the solution may get very high if change of the RA of each cell must be done manually. Finally, the solution's claimed interrupt freeness depends on the possibility to change RA on a cell without a restart of the cell.

4.3 Consequences

4.3.1 Discussion

The consequences of implementing this solution are several. The subscriber will have a more stable provision of service than at present, because of the absence of service interrupts caused by upgrades.

Ericsson's, or any other producer or manager of SGSN nodes', benefits are, because upgrades are done without interrupts, upgrades can be done whenever desired, instead of once every three months. In other words upgrades can be done earlier after a patch release than today. This minimizes the number of releases in the marked and thereof the cost related to SW maintenance. This again gives improved cash flow and reduced operational cost.

Operators' benefits are the possibility to increase the revenue as customers can use the service all the time. They also get an advantage over other operators, as they can offer a more reliable service. This may help them to draw more customers and avoid loosing others. Also, they can get a patch for a SGSN node the day it is ready, instead of evaluating whether it is better to wait or not.

Although the benefits are clear, there are also drawbacks. The new upgrade process should be automated as much as possible to avoid human errors. Management of the connections between SGSN and BSS must be handed carefully as these make up the entire routing mechanism of the solution.

When a new RA is introduced in a Location Area, a new border is created, that causes ISRAUs when crossed. This will contribute to higher load in the backbone network and between SGSN nodes. If two cells not neighbouring each other are defined as having the same RA, two new RA borders are created. These will also cause ISRAUs when crossed, and it is recommended to avoid such a constellation.

4.3.2 Conclusion

The benefits overshadow the drawbacks in my opinion. There will be extensive load on the system when introducing a new RA, but the SGSN nodes have overload protection and the impact on service will be limited. The benefits can be, provided the service interrupts can be eluded, improved cash flow, reduced operational costs, more reliable service, and earlier upgrades.

Chapter 5 Results

5.1 introduction

In this chapter I summarize and point out the results that I have described in chapters 3 and 4. I begin with the RA based PCU solution, but present also the other solutions, as they represent a considerable part of my thesis work.

The Gb interface, connecting the SGSN nodes to the BSS, is the interface connected to the SGSN that causes most problems when wanting to move traffic from a SGSN to another. Allegedly this Gb link is a one-to-many relation, between SGSN and BSS respectively. There is nothing in the ETSI standards that support this, and the PTP BVCs can be constructed between more than one SGSN and one BSC. There were some difficulties concerning what way to structure the different solutions, but the resulting four cover most of the upgrade solutions possible, while being fairly distinct from each other.

5.2 Routing Area based PCU routing

5.2.1 The preliminary view

The Routing Area based PCU routing (RAPR) was the solution considered the best of the four solutions evaluated in the first part of this thesis. This solution uses the Routing Area Identity (RAI) to smoothly transfer subscribers from a SGSN node we want to upgrade, to a new SGSN node.

A number of cells are served through a PCU. All the cells have the same RAI, but can change this when needed. For simplicity, the RAI for these cells are RA1. Two SGSN nodes (SGSN1 and SGSN2) are also connected to the PCU. The SGSN nodes have been assigned two different RAs, and the SGSN1 covers the RA1 while the SGSN2 is assigned RA2. As no cells attached through the PCU initially has Routing Area Identity RA2, the SGSN2 receives no traffic. The PCU has functionality to route all traffic coming from cells with RA1 to the SGSN node assigned the RA1, and it can in the same way route the traffic coming from cells with RA2 to the SGSN2 node.

We can now change the RAI for one or several cells from RA1 to RA2. This forces any MSs inside these cells to send a ‘Routing Area Update Request’ as it is led to believe it has moved into a new RA. This request comes from a cell with RA2 identity and the traffic is thus sent to the SGSN2.

5.2.2 Elaboration

The routing mechanism in this solution was imagined to be situated in the PCU, or at least in the BSC. However, as research and development of the solutions progressed, a twist and mix of the RAPR and the RL solutions became a better solution. The only thing changed compared to the original RAPR solution is the routing mechanism. Now it seems we can avoid a change in standards to achieve a routing function from the PCU to several SGSNs, as we exploit the Network Service Entities (NSEs) from the Redundant link solution. It may be possible to exploit the PTP BVC to route packets from cells, based on RA.

A new SGSN node must be configured with its own address and knowledge of its surrounding, or cooperating, SGSN nodes. In order to allow a mobile station to perform inter SGSN routing area updates while travelling all over the PLMN, at least those other SGSNs which own routing areas that are adjacent to any local routing area should be defined as cooperating SGSNs. In the same way at least those routing areas that reside within the service area of another SGSN, and that are adjacent to any local routing area should be defined as cooperating routing areas. This is to allow a mobile station to perform routing area updates. Local routing areas are the RAs served by SGSN2.

The local RAs are created automatically in the SGSN by the system, through BSSGP node management and cannot be modified by the operator. A local RA exists if and only if there exists at least one PTP BVC, which supports a cell within the local RA, and PTP BVCs are dynamically created on the SGSN without any possible interaction from the operator. Thus the operator cannot modify the support of RAs on the SGSN. However, the operator can statically create PTP BVCs from the BSS.

The fact that there is a PTP BVC between each GPRS cell and the SGSN node can solve the routing of packets to different SGSNs, based on RA. One can set up two PTP BVCs between each cell and the SGSN, where the BVC1 is set to support the RA1 which the cell is currently set up with. The BVC1 is a connection between the SGSN1 node and the cell. The other BVC2 is set to support the RA2, which the cell will be configured to support when we want to move traffic from SGSN1 to SGSN2. This BVC2 is a connection between the cell and the SGSN2.

It is important to point out that the reason why the PTP BVCs lead to different SGSNs is because they are bound to different Network Service Entity Identifiers (NSEIs) leading to different SGSN nodes from the BSC. The NSEIs are configured both on the SGSN side and the BSC side, and this configuration has to be synchronized to achieve contact on the BSSGP level.

In effect only one of the two PTP BVCs we have created per cell will be active at a time. The reason is: The short cell id used when defining a PTP BVC. This short cell id consists of the CI and the RAI. If we create one PTP BVC between the SGSN1 to the CI+RA1 and another between the SGSN2 and the CI+RA2, then the first PTP BVC will be used when the cell is defined with RA1, and the second will be used when the cell is defined with RA2. The BVC1 is bound to one NSEI1 and the other to a NSEI2. The PTP BVCs are dynamically created in the SGSN, but can be statically created from the BSS side. Ergo all PTP BVCs that point to a CI+RAI that does not exist at the given time, must be created on the BSS.

When it is time to transfer subscribers from the SGSN1 to the SGSN2, we can do this on the cell-level. We change the cell’s RAC, with the result that the MS discovers it is in a new RA. I have not been able to confirm that the cell’s RAC can be changed without restart of the cell. In that case the subscribers in the serving cell will experience a detachment and need to attach again to the GPRS network.

5.2.3 Effects

This subchapter is an overview of the effects by implementing the solution.

Benefits:

- Upgrades can be done earlier after a patch release, which minimizes the number of releases in the market and thus the cost related to SW maintenance. This again gives improved cash flow and reduced operational cost.
- Possibility for the operators for increased revenue as customers can use the service all the time.
- An advantage for the operators over other operators, as they can offer a more reliable service. This may help draw more customers and keeping them.

Drawbacks:

- When a new RA is introduced in a Location Area, a new border is created, that causes ISRAUs when crossed. This will contribute to higher load in the backbone network and between SGSN nodes.
- If two cells not neighbouring each other are defined as having the same RA, two new RA borders are created. These will also cause ISRAUs when crossed, and it is recommended to avoid such a constellation.
- The change of RA on a cell may cause downtime for some subscribers.

5.3 Hot standby

One SGSN2 is a standby node for another SGSN1 and mirrors this. In other words the SGSN1 updates the SGSN2 with change in subscribers' connection data. At some point in time we shut down the SGSN1 and tell the SGSN2 to start executing incoming traffic and signalling. Then the SGSN1 is free to be upgraded without traffic interrupts.

5.4 Subscriber based PCU routing

Based on subscribers we control and route packets to the wanted SGSN from the BSS. The BSS/PCU is connected to two SGSN nodes or more, where SGSN1 is a node we want to do upgrades on and SGSN2 is a replacement node for the SGSN1. Initially all traffic goes through the SGSN1 node. We then make the PCU send traffic from all subscribers that are new to the SGSN1, to SGSN2. This way, all subscribers that are new inside the coverage zone of the SGSN1&2 send attach requests or routing area update requests to the SGSN2. At the same time, the PCU sends traffic from the 'old' subscribers to the SGSN1.

Any subscribers moving out of the area operated by the SGSN1 and SGSN2 nodes would experience no service interrupts as an ISRAU would move data to the new SGSN node. If a subscriber enters or re-enters the area, the PCU would route all packets to the SGSN2 as they are considered new subscribers. The number of subscribers still attached to the SGSN1 will move towards zero as people move in and out of the area covered by the SGSN1&2. New subscribers will be connected through the SGSN2. After a while there will be none or very few subscribers still attached through the SGSN1, and we can shut down this and carry out the upgrade. The few subscribers that get detached from the SGSN1 will try to reattach, and the PCU will treat them as new subscribers, because the MS will send an 'Attach Request' to the SGSN.

Different variations of this solution are:

1. Take down the SGSN1 node although there still are some subscribers attached. These will experience a break in service and have to reattach.
2. At a suitable time one moves the remaining subscribers to the SGSN2 by an 'ISRAU' or a similar procedure. In this way none of the subscribers will experience service interrupts.

5.5 Redundant links

One SGSN1 is an operating node, while a SGSN2 is a replacement node to take over traffic from the SGSN1 while this is being upgraded. We have a connection between each of the SGSN nodes and the PCU. These are both physically and logically different links, and only one of them is open at a time. Initially, the link between the SGSN1 and the PCU is open and the link connecting SGSN2 and the PCU is blocked by O&M. Then, at a certain time, we open the link between the PCU and the SGSN2, while we at the same time close the PCU-

SGSN1 link. Thereby all new packets coming from the PCU ends up at the SGSN2, and the SGSN1 is free for upgrade procedures. The SGSN2 will ignore all packets except Attach Request packets. These will be sent by the MSs when they have timed out waiting for response from the SGSN1. When an Attach Request is received, the MS will be connected to the Core Network through the SGSN2, and the SGSN1 node is history.

Chapter 6 Discussion

6.1 General

In this chapter I bring to attention the problems with each of the solutions, the solution I found best on what grounds, and the implementation problems with the RAPR solution.

To enable an upgrade of the SGSN node by removing the SGSN from traffic, the Gb link has to be altered. It is easier to exploit the functionality in the GPRS system, where the SGSN informs the HLR of which IP address the subscriber has, and at which SGSN/GGSN node it can be reached, instead of adding functionality at the Gn, Gp and Gi link. All the solutions have to do with the Gb link because the Gb link connects the subscribers to the SGSN.

6.2 Routing Area PCU routing

6.2.1 General solution

The solution depends on functionality in the PCU to route traffic to different SGSN nodes. The PCU needs to know what Routing Area the packet comes from, and based on this be able to route it to the correct SGSN node. The development problems most obvious here are:

- The PCU must be able to be connected to several SGSN nodes simultaneously, and route packets to either SGSN.
- The RA of a cell must be changeable without implying downtime for the cell or the BSS.
- The PCU must know what RA a packet comes from, without having to open it.

With this solution there is need for at least one new SGSN node in the network. It is probably practical with one extra SGSN node at every location where there is one or more SGSNs.

The solution represents possible changes to the PCU equipment, but the ETSI standards for the PCU will not necessarily have to be altered when implementing this solution. This is because the standards describe the communication interface between the PCU and the other nodes, rather than the functions performed by the PCU. Also, when it is possible to exploit the functionality already available on the Gb interface, there should not be need for any standard changes.

The solution depends on a standard Inter SGSN RA Update to transfer the subscribers' connection data between the SGSN nodes, and initially neither breaks nor delays were expected.
Initially, neither breaks nor delays were expected with this solution, because it depends on a standard Inter SGSN RA Update to transfer the subscribers' connection data between the SGSN nodes. However, the possibility to change the Routing Area of one and one cell at a time will not necessarily ensure the absence of traffic overload generated by too many Inter SGSN RA Update procedures at once. Also, there is a question whether the RAC of cells may be changed without implying downtime for the cell or BSS.

6.2.2 Implementation and effects

The load may get too high on the SGSN nodes and the backbone. Although we change RA for one cell at a time, there are no limitations to how many subscribers the cell can hold at a given time. For instance, at a concert or a football match many thousands may be inside the coverage zone of one cell. If that cell all of a sudden changes RA, all MSs inside the cell will send, or try to send, a RA Update Request, resulting in equally many ISRAU procedures. Although the SGSN most lightly has overload protection, many subscribers will be offline until they are properly moved to the new SGSN.

One SGSN node serves thousands of cells. Changing the RA of all the cells must be done automatically and remotely. If the change is to be done manually and locally at each cell, the operation will use much more time and manpower. This may make the cost of the solution to high on Ericsson's part.

It must be possible to change the RA of a cell without implying downtime for the cell. In case the cell has to be restarted, this would give a break or delay in service for the subscriber. This does not mean the solution cannot or should not be implemented, but it lowers the wanted effect of the strategy, namely an interrupt-free upgrade.

The benefits will be, if the solution is implemented:

- Upgrades can be done earlier after a patch release, which minimizes the number of releases in the market and thus the cost related to SW maintenance. This again gives improved cash flow and reduced operational cost.
- Possibility for the operators for increased revenue as customers can use the service all the time.
- An advantage for the operators over other operators, as they can offer a more reliable service. This may help draw more customers and keeping them.

The drawbacks are:

- When a new RA is introduced in a Location Area, a new border is created, that causes ISRAUs when crossed. This will contribute to higher load in the backbone network and between SGSN nodes.
- If two cells not neighbouring each other are defined as having the same RA, two new RA borders are created. These will also cause ISRAUs when crossed, and it is recommended to avoid such a constellation.
- The change of RA on a cell may cause downtime for some subscribers.
- If the number of subscribers in a cell gets too high, the SGSNs will not be able to move all simultaneously to the new SGSN. Some subscribers may experience this as an interrupt of service, as it takes time from they loose contact with the old SGSN node, until they are moved properly to the new SGSN.

6.3 Hot standby solution

We want to change executing SGSN nodes from SGSN1 to SGSN2 without interrupting user traffic or detaching Mobile Stations. We will need a dedicated standby SGSN2 node for each SGSN1 node. This is very expensive and does not take advantage of the redundancy already in the system. Another problem is the design of the SGSN node. It has not been designed with mirror functionality (i.e. sharing RAM), and a new node architecture will probably have to be developed.

The part of the GPRS system affected by this strategy is the SGSN node. All other parts of the system view the two SGSN nodes as one, and all changes will be on the SGSN nodes. The development problem that is most obvious, while doing the initial study of the various strategies, is to have the SGSN nodes share the subscribers' connection-data in real-time (i.e. mirroring an SGSN node), and thus make it possible for the standby SGSN node to take over the execution at an instant. Another problem will be how to switch between the two SGSN nodes.

6.4 Subscriber based PCU routing solution

The solution stands and falls upon whether it can be made possible for the PCU to recognize which subscriber a packet belongs to, and thereafter be able to pass it on to the right SGSN node. There are three main features that do not exist in the PCU and need to be developed, to make this work:

- The PCU must be connected to, and able to differentiate between several (at least two) SGSN nodes.
- The PCU must be able to recognize the dispatcher of every packet passing through it.
- The PCU must know which dispatchers that are 'old' subscribers and route these packets to the SGSN1 node.

The PCU has to open every packet passing through, to find which subscriber it comes from. This will produce a huge workload on the PCU, and if it cannot take care of this, the result will be delays or thrown packets. Yet another development problem is in the event that an MS moves out of the area covered by the SGSN1 and SGSN2. The SGSN node covering the area the MS enters must know which of the SGSN1 or 2 to ask for subscriber information.

The variation 2 implies the development of a procedure to move the remaining subscribers to the new SGSN from SGSN1. This must be done without service interrupts for the subscribers. In addition to the development cost, there will be need for at least one dedicated SGSN2 node in the network to replace the SGSN1 while these are being upgraded one by one. It will probably be normal with one SGSN2 at every geographical location where there is one or more SGSN1 nodes. This means the Life Cycle Cost could be greatly increased, as we need to have an extra node at each SGSN1 location.

The SBPR solution will demand some change in the standards for the PCU. There is a considerable difference between using the PCU only to forward incoming packets and using the PCU as an intelligent router.

The benefits of this solution are big, however. Only very few subscribers need to experience interrupts in service, as most subscribers will be moved smoothly to the SGSN2 from other SGSN nodes via the 'Inter SGSN RA Update'-procedure. This is provided the routing function in the PCU is given enough time, so as many subscribers as possible can get moved to the new SGSN2. With Variation 2 of the solution the remaining 'old' subscribers may also be moved to the SGSN2 without interrupts

6.5 Redundant links solution

The RL solution bases itself on already standardized equipment, and the investment costs developing the solution will probably be small. The Life Cycle Cost will be increased mainly by new replacement SGSN nodes and cables to link the PCUs to the SGSNs.

The major drawback in this solution is the service performance. The solution is based upon breaking all the subscribers connections, and then let every subscriber self find out it has been detached. All MSs will discover the break simultaneously, except for the ones not using the service at the time. These will believe they are still attached to the network without being this. As long as they neither are attached to the CN and don't try to get attached, they are unreachable for anyone trying to reach them

All the MSs that were not engaged in traffic when the SGSN1 went down will discover that their connections are gone when they send the periodic RA update. Then they will do a GPRS attach towards the new SGSN. In worst case a subscriber will believe all is ok for 53 minutes after the SGSN1 node is shut down. The solution takes only care of routing in the MS – SGSN perspective and leaves the SGSN – GGSN routing (i.e. incoming traffic to the MS) to take care of itself. There will be a vast number of Attach Request received by the SGSN2 node as the MSs try to reconnect. In a worst case scenario an unlucky subscriber will have to wait 20 minutes before getting reattached, but first it has to discover it needs to reattach.

Finally, as the SGSN2 node is a temporary node, traffic has to be transferred back to the SGSN1 node when the upgrade procedure is done. This means another break-and-reattach for all subscribers. The average subscriber will experience two breaks lasting 24 minutes each. This is an upgrade highly dependent on being carried out at slack hours.

6.6 Best recognized solution

The comparison between the different solutions can be seen in chapter 3.8.1-6 and will not be repeated here. Each of the criteria used for discussion are not equally weighted when comparing and evaluating these, though they are all important. The main factor is the service interrupts, which should be as small and few as possible.

Reviewing the discussions of the solutions compared against each other based on the factors above, we have the RL solution being the best in four areas. It is wrong, however, to pronounce this as the best solution, as it will probably give more total service interrupts than the present solution. The Hot standby solution is better than the RL on an overall basis, when weighting the service interrupts high, but it has a big increase in LCC, and its development problems are the greatest of all the solutions.

We are therefore left with the RAPR and the SBPR solution var2, as these offer a very low degree of service interrupts, are low on increase in LCC and are considered feasible with respect to the development problems. The solutions are similar, but the routing mechanism is different in the two. The RAPR routes on the RAI, while the SBPR routes on subscribers. There is a big problem with the SBPR solution's routing mechanism, as the PCU would have to open each and every packet to see which subscriber it belongs to. This is because the subscriber info lies in the BSSGP layer. The RA information, however, is known by the BSS as it is defined there. The RAPR solution gives the impression of being a solution exploiting existing functions in the GPRS system instead of depending on future functions. It is unlikely that the SBPR solution will be any use without the ISRAU like function to help move subscribers, who have not themselves provoked an ISRAU, over to the new SGSN.

Therefore I have chosen to elaborate further on the RAPR solution.

Chapter 7 Conclusion

7.1 Introduction

The solutions I present below are valid in the GPRS system for GSM. They have been developed on basis of the ETSI GSM Digital cellular telecommunications system (Phase 2+) standards, and are not implementation-specific. However, the description of implementation of the RAPR solution is based on Ericsson's SGSN specifications and the implementation may be different on other companies' SGSN nodes.

7.2 Different strategies for upgrade

The upgrade strategies are all based on the need for the SGSN node to be out of traffic while being upgraded. I came up with four different solutions that all tries to reduce cut-offs and interrupts as much as possible.

The Hot standby solution provides a virtually non-interruptive upgrade. But, as it depends on the possibility for SGSN nodes to share RAM, it craves much development before it can be used. The Life Cycle Cost will be high, as the solution demands an extra SGSN node for each existing node. All the time the extra node is not in use, this will be waste of resources. However, the solution will provide full software and hardware failure protection.

The Subscriber based PCU routing solution can provide a non-interruptive upgrade under perfect conditions. However, it depends on a procedure similar to the Inter SGSN RA Update to move all subscribers without breaks to the new SGSN; and if there are many subscribers that have to be moved, then breaks will occur regardless of such a procedure. Development cost will be extensive as a new routing functionality based on subscribers has to be developed for the PCU.

The Life Cycle Cost could get greatly increased, as there is a practical need for an extra SGSN node at every existing SGSN location. There is a problem concerning subscribers exiting the area covered by these SGSN nodes, because when a new SGSN node wants to ask the SGSN in the area the subscriber comes from, it does not know which SGSN to ask.

The Redundant links solution is based on existing equipment and standards except for additional needed capacity to replace the upgrade-SGSN. Therefore the cost of implementing this solution will be minor, but the average subscriber will unfortunately experience breaks for 2 x 24 minutes, as they all have to discover they are detached and reattach to the network twice. Compared to the present upgrade solution the RL solution is unfavourable.

The RA based PCU routing solution is the solution recognized as the best one of the four presented here. Though there is anticipated a great deal of development costs on the PCU RA Routing function, the solution bases itself on functions already implemented in the system, mainly the Inter SGSN RA Update. It also enables a more effective use of the SGSN resources, as several SGSN nodes already in traffic can replace one SGSN taken down for upgrades. And most importantly, it offers an upgrade procedure without service breaks, provided the cells RAC can be changed without downtime.

7.3 Implementation and effects of the 'RAPR' solution

Though it seemed evident that a change in standards was required in order to achieve a routing function from the PCU to several SGSNs, this is not the case. Instead of having a routing function implemented and processed in the PCU, we take advantage of the Point to

Point BSSGP Virtual Connections to route packets to different SGSN nodes based on RA. This makes the RAPR solution possible to implement directly in the GPRS system, without change of standards.

The solution offers an upgrade without service interrupts, but this is only true for a limited number of subscribers inside a cell. Also, the cost of the solution may get very high if change of the RA of each cell must be done manually. Finally, the solution's claimed interrupt freeness depends on the possibility to change RA on a cell without a restart of the cell

It is unknown whether a change of a cell's RAC can be done without a break in service. It is also uncertain where a cell's RAC is configured.

Concerning the effects of the solution, the benefits overshadow the drawbacks. There will be extensive load on the system when introducing a new RA, but the SGSN nodes have overload protection and the impact on service will be limited. The benefits can be: improved cash flow, reduced operational costs, more reliable service, and earlier upgrades. All provided the service interrupts can be avoided or kept at a minimum.

Chapter 8 References

[1] ETSI GSM 01.04:"Digital cellular telecommunications system (Phase 2+); Abbreviations and acronyms (GSM 01.04 version 8.0.0 Release 1999)"

[2] ETSI GSM 03.03:"Digital cellular telecommunications system (Phase 2+); Numbering, addressing and identification (GSM 03.03 version 7.5.0 Release 1998)"

[3] ETSI GSM 03.60:"Digital cellular telecommunications system (Phase 2+); General Packet Radio Service (GPRS); Service description; Stage 2 (GSM 03.60 version 7.4.1 Release 1998)"

[4] ETSI GSM 04.01:"Digital cellular telecommunications system (Phase 2+); Mobile Station – Base Station System (MS - BSS) interface; General aspects and principles (GSM 04.01 version 8.0.0 Release 1999)"

[5] ETSI GSM 08.14:"Digital cellular telecommunications system (Phase 2+); General Packet Radio Service (GPRS); Base Station System (BSS) – Serving GPRS Support Node (SGSN) interface; Gb interface Layer 1 (GSM 08.14 version 8.0.0 Release 1999)"

[6] ETSI GSM 08.16:"Digital cellular telecommunications system (Phase 2+); General Packet Radio Service (GPRS); Base Station System (BSS) – Serving GPRS Support Node (SGSN) interface; Network Service (GSM 08.16 version 8.0.0 Release 1999)"

[7] ETSI 3GPP TS 08.18:"Digital cellular telecommunications system (Phase 2+); General Packet Radio Service (GPRS); Base Station System (BSS) – Serving GPRS Support Node (SGSN) interface; BSS GPRS Protocol (BSSGP) (3GPP TS 08.18 version 8.6.0 Release 1999)"

[8] ETSI GSM 09.60:"Digital cellular telecommunications system (Phase 2+); General Packet Radio Service (GPRS); GPRS Tunnelling Protocol (GTP) across the Gn and Gp Interface (GSM 09.60 version 7.5.1 Release 1998)"

[9] "GSM Phase 2+ General Packet Radio Service GPRS: Architecture, Protocols, and Air Interface". Christian Bettstetter, Hans-Jörg Vögel, and Jörg Eberspächer. Technische Universität München (TUM). (IEEE Communications Surveys, Third Quarter 1999, vol. 2 no. 3)

[10] "Administration and use of Network Configuration Data in SGSN; Function Description" Ericsson Radio Systems AB 2000, 11/155-17-CSA 250 08 Uen Rev A 2000-04-18

[11] "Gb Interface Configuration" Ericsson Radio Systems AB 1999, 1543-AXB 250 01/1 Uen Rev A 00-05-18