

GPRS support nodes

Lars Ekeröth and Per-Martin Hedström

Telecommunications and data communications are converging, due in no small part to the increasingly prominent role of the Internet protocol (IP). Also, users want access to the Internet while they are away from their offices and homes.

Packet-switched services present new opportunities for operators and users. They allow operators to capitalize on the rapid growth of Internet usage and to position cellular service as mobile Internet access. The introduction of general packet radio service (GPRS) in today's cellular networks is a key step in the evolution toward third-generation mobile networks.

GPRS makes the Internet mobile. It allows users to access corporate intranets or Internet service providers (ISP) from a mobile device. Its users can remain online without continuously occupying a specific radio channel. Each channel is shared by several users and is used only when data packets are sent or received.

In this article, the authors describe Ericsson's GPRS support nodes (GSN)—the core network nodes at the heart of Ericsson's GPRS solution that provide packet data capability to GSM, UMTS and TDMA cellular networks.

Introduction

GPRS support nodes

The GPRS support nodes constitute the parts of the Ericsson cellular system core network that switch packet data. The two main nodes are the serving GPRS support node

(SGSN) and the gateway GPRS support node (GGSN). Figure 1 shows an example of the architecture of Ericsson's GPRS solution in a GSM cellular network. GSNs are also used for GPRS domains within a UMTS or TDMA system.

Hardware and software redundancy have been designed into the platform, which enables operators to upgrade individual modules without disturbing traffic. Because the payload-carrying devices and control devices are kept separate in the platform, software upgrades usually have only a minimal effect on ongoing end-user payload transfer. Also, the $n+1$ redundancy of hardware in the platform makes it possible to upgrade most hardware devices without affecting traffic.

Functionality

The SGSNs route packets to and from the geographical SGSN area, while GGSNs interface with external IP packet networks. The SGSN and GGSN are physically separate from the circuit-switched part of the Ericsson cellular system core network.

The functionality of the SGSN and GGSN can be combined in the same physical node (combined GPRS support node, CGSN) or reside in different physical nodes. Both the SGSN and GGSN contain GPRS backbone network protocol (IP) routing functionality, and can be interconnected with IP routers.

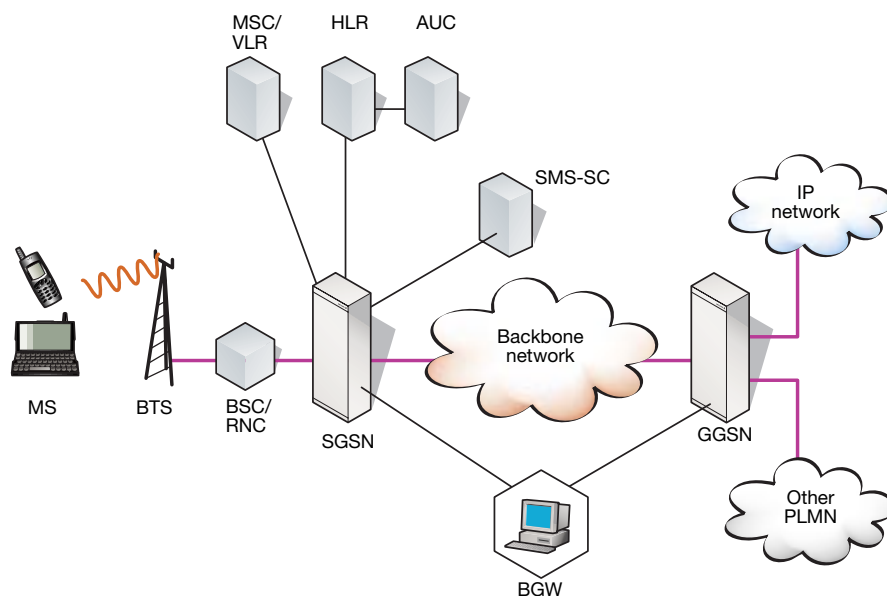
In other respects, Ericsson's GPRS architecture uses existing cellular network elements, such as subscriber databases and radio transmission systems.

Architecture

The GSNs are based on the wireless packet platform (WPP), a new general-purpose, high-performance packet-switching platform. The WPP, which is used for GPRS, EDGE and UMTS, combines features usually associated with data communications (such as compactness and high functionality) with features from telecommunications (such as robustness and scalability).

Ericsson's middleware solution consists of object-oriented components that use the common object-request broker architecture (CORBA) and interface definition language (IDL). CORBA and IDL are also used for interfaces to the application layer. The solution provides a framework for building robust, real-time applications for processing transactions in a distributed multiprocessor environment using software modules written in C/C++, Java, or Erlang (Figure 2).

Figure 1
Overview of the Ericsson packet-data core network in a GSM system.



GSN components and features

Serving GPRS support node

The SGSN is a primary component of cellular networks that employ GPRS. Via the radio network, the SGSN routes incoming and outgoing IP packets addressed to or from any GPRS subscriber physically located within the geographical area served by that SGSN. Each SGSN provides

- ciphering (encryption and decryption) and authentication;
- session management and communication set-up to the mobile subscriber;
- mobility management—that is, support for roaming and handover within and between mobile networks;

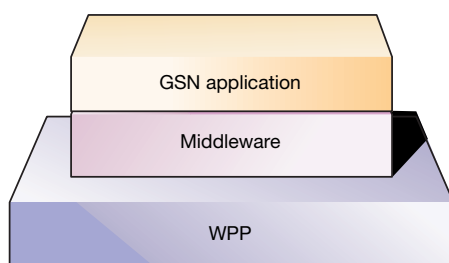


Figure 2
The GSN architecture.

BOX A, ABBREVIATIONS

3GPP	Third-generation Partnership Project	HTTP	Hypertext transfer protocol	PASOS	Packet-switched operation support system
AAL5	ATM adaptation layer 5	IBAM	155 Mbit/s interface for ATM providing a PMC module for multimode fiber	PCI	Peripheral component interconnect
AP	Application processor	IBE1	E1 interface board with a PowerPC processor and E1 PMC modules	PDCH	Packet data channel
AP/C	Application processor control	IBEN	Ethernet interface board with a PowerPC processor and Ethernet PMC modules	PDP	Packet data protocol
APN	Access point name	IBT1	E1 interface board with a PowerPC processor and T1 PMC modules	PDU	Packet data unit
ASN.1	Abstract syntax notation one	ICMP	Internet control message protocol	PEB	Power and Ethernet board
ATM	Asynchronous transfer mode	IDL	Interface definition language	PLMN	Public land mobile network
BER	Basic encoding rules	IIOB	IP-based inter-ORB protocol	PMC	PCI mezzanine card
BG	Border gateway	IMEI	International mobile equipment identity	PPP	Point-to-point protocol
BGP	Border gateway protocol	IMSI	International mobile subscriber identity	PTM	Point-to-multipoint
BNSI	Basic network surveillance interface	IP	Internet protocol	PVC	Permanent virtual circuit
BSC	Base station controller	IPsec	IP security protocol	PXM	Packet exchange manager
CGF	Charging gateway function	ISP	Internet service provider	QoS	Quality of service
CGSN	Combined GSN	L1, L2, L3	Layer 1, layer 2, layer 3	RADIUS	Remote access dial-in user service
CORBA	Common object request broker architecture	LCT	Local craft terminal	RANAP	Radio access network application part
cPCI	Compact PCI	LDAP	Lightweight directory access protocol	RD	Resource deployment
DES-CBC	Data Encryption Standard, cipher block chaining (USA)	LLF	Link layer forwarding	RIP	Routing information protocol
DHCP	Dynamic host configuration protocol	MAC	Medium access control	SCCP	Signaling connection control part
DP	Device processor	MAP	Mobile application part	SCF	Software configuration file
DPE	Distributed process environment	MD5	Message digest algorithm 5	SGSN	Serving GSN
E1	ETSI 2 Mbit/s interface	MIB	Management information base	SMS	Short message service
E3	ETSI 34 Mbit/s interface	MSC	Mobile switching center	SMS-GMSC	SMS gateway MSC
EIR	Equipment identity register	MTP	Message transfer part	SMS-IW MSC	SMS interworking MSC
EM	Element manager	NCS	Network control system	SNMP	Simple network management protocol
ESP	Encapsulating security payload	NE	Network element	SS7	Signaling system no. 7
ETSI	European Telecommunications Standards Institute	NOC	Network object control	T1	ANSI 1.5 Mbit/s interface
FPGA	Field programmable gate array	O&M	Operation and maintenance	T3	ANSI 45 Mbit/s interface
FTP	File transfer protocol	ORB	Object request broker	TC	Traffic control
GGSN	Gateway GSN	OSPF	Open shortest path first	TCAP	Transaction capabilities application part
GMM	GPRS mobility management	OTP	Open telecom platform	TCP	Transmission control protocol
GPB	General processor board			TDMA	Time-division multiple access
GPRS	General packet radio service			UDP	User datagram protocol
GSN	GPRS support node			UMTS	Universal mobile telecommunications system
GUI	Graphical user interface			WPP	Wireless packet platform
HLR	Home location register				
HTML	Hypertext markup language				

- logical link management to the mobile subscriber; and
- connection to other nodes (HLR, MSC, BSC, SMS-GMSC, SMS-IW MSC, GGSN).

The SGSN also collects charging data for each mobile subscriber, such as the actual use of the radio network and GPRS network resources.

Gateway GPRS support node

The GGSN is also a primary component of cellular networks that employ GPRS. The GGSN serves as the interface to external IP packet networks, accessing external ISP functions such as routers and remote access dial-in user service (RADIUS) servers. In terms of the external IP network, the GGSN routes the IP addresses of subscribers served by the GPRS network, exchanging routing information with the external network.

In the Ericsson GGSN, a border gateway (BG) shares the GGSN's physical interfaces to external networks and the backbone network. One border gateway can handle multiple public land mobile networks (PLMN).

The GGSN sets up communication with external networks and manages GPRS sessions. It also includes functionality for associating subscribers to the appropriate SGSN. For each mobile subscriber, the GGSN also collects charging data—use of the external data network and use of GPRS network resources.

IP router

Each Ericsson GSN has an integrated router which

- serves as a primary or secondary IP router in IP networks, although only temporarily—a bigger IP router is strongly recommended for routing to other IP networks; and
- provides redundancy for the *Gn/Gi* interface.

The router supports open shortest path first (OSPF) and the border gateway protocol (BGP) as well as other routing protocols. It can also filter IP packets in all IP interfaces—for example, from one PLMN to another. The filter, whose configuration data can be set during operation and maintenance (O&M), applies to transmission control protocol/Internet protocol (TCP/IP) header information, and a combination of IP source address, IP destination address, protocol type, TCP flags, Internet control message protocol (ICMP) message type, TCP/user datagram protocol (UDP) source port, and

TCP/UDP destination port and physical interface.

Charging

As mentioned above, both the SGSN and GGSN can produce charging data records. Combined with a mediation device, such as the Ericsson Billing Gateway, this gives operators a wide range of options. Charging can be based on data volume, duration of call, type of service, destination point, or some other factor.

The European Telecommunications Standards Institute (ETSI) has specified both a centralized and a distributed alternative for the charging gateway function (CGF). In Ericsson's solution, the basic CGFs are distributed and the enhanced CGFs are centralized.

The basic CGFs—collecting, storing, and transferring charging data—are implemented in the GSNs. The charging data records are buffered in nodes, which provides greater security against network or transmission problems. Charging data output from SGSNs and GGSNs is encoded for abstract syntax notation one/basic encoding rules (ASN.1/BER) and transferred via the file transfer protocol (FTP) using either a push or pull mechanism.

The enhanced CGFs—consolidation, filtering, pre-processing, and formatting of charging data records—are implemented centrally in the Ericsson Billing Gateway. Centralization reduces the number of interfaces to the billing system and provides various post-processing options. Also, centralized mass media storage is better than distributed storage on GSNs.

Allocation of IP addresses

Dynamic IP addresses

The allocation of dynamic IP address enables operators, ISPs, and corporate networks to reuse IP addresses from a pool allocated to the PLMN or some other network. It also significantly reduces the total number of IP addresses required per PLMN.

A dynamic IP address can be allocated by or via a GGSN in the visited network or by a GGSN in the home network. The dynamic IP address can be provided by the GGSN itself or by a RADIUS server chosen by the GGSN.

The GGSN contains a RADIUS client that supplies the external RADIUS server with authentication information from the mobile subscriber. Thus, the RADIUS serv-

er can return an IP address if authentication is correct. The configuration can specify that the GGSN should contact a specific RADIUS server for each access point name (APN)—that is, each corporate network or ISP. The RADIUS server can be located either at the ISP or at a corporate site. The GGSN will also include a dynamic host configuration protocol (DHCP) client. Ericsson's GSN also supports the overlaying of private IP addresses.

Static IP addresses

The use of static IP addresses is not recommended, primarily due to a shortage of IP addresses (this situation will change when IPv6 is introduced). The static IP address is defined for the subscription by the HLR—as an option, a RADIUS server can be contacted for authentication purposes. When the terminal is in the attached state, the subscriber's IP addresses are copied to the SGSN. Accordingly, when it sends a packet data protocol (PDP) context activation request, the terminal either

- provides an IP address—which is checked against the subscription information; or
- allows the SGSN to allocate the IP address—provided that the subscription contains only one IP address.

Security functions in GSNs

Security in SGSNs

Authentication is always performed for attach and inter-SGSN routing area updates, for both home and visiting subscribers. A log file of failed authentication attempts is kept. The log contains the time and date, and the international mobile subscriber identity (IMSI), international mobile equipment identity (IMEI), SGSN ID and cell identity of mobile subscribers or handsets that failed authentication.

The GSN supports selective authentication settings for all home network subscribers in the node. The settings dictate the number of attach procedures and inter/intra-SGSN routing-area-update procedures that can occur between each authentication procedure. Authentication is always performed for visiting subscribers.

Security triplets are fetched from the HLR. However, for inter-SGSN routing area updates, unused triplets are fetched from the previous SGSN (if known).

Secure connections can be provided on layer 1 (L1), using a dedicated physical line; on layer 2 (L2), using asynchronous transfer

mode (ATM) permanent virtual circuits (PVC), frame relay PVC, or the point-to-point protocol (PPP); and on layer 3 (L3), using IP security (IPsec). These techniques can also be used in combination.

Security in GGSNs

The GGSN ensures that traffic for a specific mobile subscriber comes from the ISP to which the mobile subscriber was connected during PDP context activation. The GGSN can access RADIUS servers that are located in the external data network or operated by an ISP.

The GGSN provides an IPv4 IPsec authentication header using the keyed message digest algorithm five (MD5), and encapsulating security payload (ESP) using the

BOX B, INTERFACES

Standards from ETSI and the Third-generation Partnership Project (3GPP) specify several logical interfaces to and from the GSNs. Some of these are described below (see also Figure 3).

Interface	Used for
<i>Gn</i> and <i>Gp</i>	control signaling (for mobility and session management) between the SGSNs and GGSNs, and tunneling of end-user data payloads in the backbone network.
<i>Iu</i>	carrying IP traffic between the core network and the radio network. SGSN control signaling between the radio network and the core network. (The RANAP protocol, transported on SCCP/MTP3-B/SSCF/SSCOP/AAL5/ATM, is used over this interface to support mobility and session management signaling between mobile terminals and the core network.)
<i>Gb</i>	SGSN signaling with the BSCs in GSM or TDMA packet-access networks.
<i>Gi</i>	transportation of end-user IP data between the mobile network and external IP networks, and GGSN control signaling with ISP servers located in IP networks (including end-user authentication and IP address allocation via RADIUS).
<i>Gr</i>	MAP signaling to support storage and retrieval of subscriber data between the SGSN and HLR.
<i>Gd</i>	MAP signaling to support the SMS service over packet-switched radio channels between the SGSN and the SMS-C.
<i>Gm</i>	signaling between the PTM-SC, the GGSN, and the SGSN, and carrying messages between these nodes after a request has been made by a PTM server application to send data to a group with or without geographical filtering. The <i>Gm</i> interface is currently being specified by the 3GPP.
<i>Gf</i>	MAP signaling to support identity-check procedures between the SGSN and EIR servers when a user is attaching.
<i>Gs</i>	The SGSN server supports the standard <i>Gs</i> interface to the MSC server, in order to provide mobility management for subscribers who are attached both to packet-switched and circuit-switched channels. These combined procedures cover, for example, location updates and paging. For TDMA, this interface serves to convey ANSI signaling messages to and from the GPRS network (for transactions including registration, paging, and teleservices bearer information). The <i>Gs</i> interface is also used for SMS, since TDMA does not employ the <i>Gd</i> interface.

cipher block chaining mode of the Data Encryption Standard (DES-CBC). The system is also ready for the introduction of new encryption algorithms, such as an asymmetrical public-key authentication protocol.

Various packet-filtering options are available to protect the GGSN against intrusion or denial-of-service attacks, including source, destination, protocol, and port number.

Security for maintenance access

Maintenance commands are issued from the element manager in charge of the packet exchange manager (PXM). IPsec tunneling protects the link to the PXM.

To prevent unauthorized access, the management commands for the network element are assigned one of several command categories. Individual operator profiles can be set up with privileges that operate in one or more command categories.

Operators must identify themselves with a password to gain access to a specific, pre-configured set of command categories. All service requests are logged. Read-only access from external nodes can be gained—for example, to read alarms—via the simple network management protocol (SNMP). The access is restricted due to security flaws in SNMP v1 and the lack of standardized

SNMP management information bases (MIB) for GPRS.

IPsec

IPsec is an optional feature for the *Gi*, *Gn*, and *Gp* interfaces (Box B). IPsec can provide a secure intra-PLMN backbone and interface to external networks, such as ISPs, corporate networks, and other PLMNs.

Payload handling

Packets are divided into different QoS delay classes according to assigned priority. Within a given time period, all packets from a QoS delay class with high priority are delivered before packets from a class with lower priority. Traffic to and from mobile subscribers with the same QoS delay class can be queued in a first-in first-out (FIFO) fashion.

Overload situations will set off an alarm. The SGSN systematically discards packet data units (PDU) in order to preserve committed QoS levels: QoS class 1 PDUs take precedence over class 2 PDUs, and so forth.

Quality of service

The GPRS QoS profile is based on GSM standard 03.60. However, only reliability classes 2 and 3 are supported, because they are suitable for IP data. Similarly, only delay classes 1 through 4 are supported for subscriber data.

The SGSN applies an admission control function to each PDP context activation request. The function results in further processing of the request, negotiation of the QoS with the mobile subscriber, or rejection of the PDP context activation request.

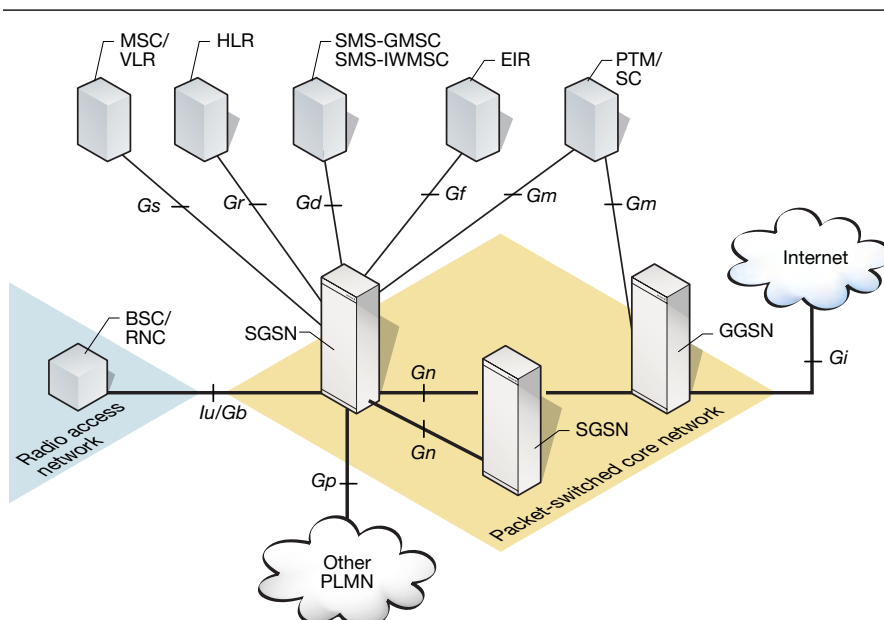
The SGSN negotiates QoS with the mobile subscriber when the level requested by the subscriber cannot be supported or when the QoS level negotiated from the previous SGSN cannot be supported at an inter-*SGSN* routing area update. The response to the mobile subscriber depends on the stored subscriber data, the requested QoS, and the statistically averaged bandwidth for each cell.

A request for a specific QoS level might be rejected when the number of subscribers simultaneously attached to a particular *SGSN* exceeds a predefined limit.

Wireless packet platform

Software for the GSN runs on the wireless packet platform (WPP), which is a combined processor and communications plat-

Figure 3
Packet-switching core network interfaces.



form designed to support mobile Internet products (Figure 4). The software consists of middleware and the GSN application.

General description

The WPP is built around a backplane that provides a redundant Ethernet backplane for interprocessor communication and a duplicated power supply to all the circuit boards. The redundant Ethernet switch provides fully switched Ethernet with full 100 Mbit/s to each circuit board position.

Power distribution in the cabinet provides a duplicated 48V power feed to each magazine. The power feed to each circuit board position in the magazine is distributed by two power and Ethernet board (PEB) units at each side of the magazine. Each PEB unit also contains an Ethernet switch.

Multiple magazines can be connected to one another using a duplicated gigabit Ethernet link. Each magazine is equipped with a fan for forced cooling.

WPP circuit boards

The circuit boards used in the WPP are designed to accommodate the use of standard components, including the enhancement of full redundancy and telecom-grade support. Each circuit board consists of three parts: a carrier board, a compact peripheral component interconnect (cPCI) module circuit board, and PCI mezzanine card (PMC) modules (Figure 5). The architecture of the circuit boards allows multiple boards to be introduced easily and quickly by combining a carrier board with different cPCI module circuit boards and PMC modules.

The carrier board provides access to the redundant Ethernet backplane and power feed. It manages the duplicated Ethernet access and hides this complexity from the PMC modules and cPCI module circuit board.

The cPCI module circuit board is mounted as a daughter board on the carrier board—if production volumes are sufficiently large, the carrier board and the cPCI module circuit board can also be designed as a single unit. The standard cPCI gives access to many off-the-shelf circuit boards.

A cPCI bus provides access to the cPCI module circuit board and the PMC modules. PMCs using cPCI are common in the open-standards market. They provide different link access modules and processing modules. Several PMC modules can be mounted on the carrier board (with cPCI). A spe-

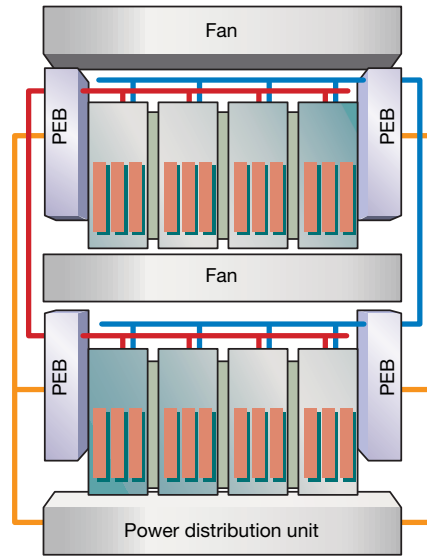


Figure 4
WPP cabinet—schematic view.

cial field programmable gate array (FPGA) PMC module provides encryption support for GPRS mobile payload and IPsec.

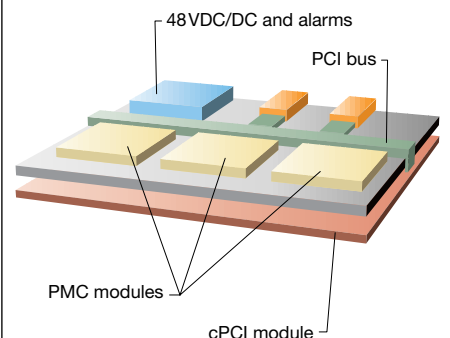
At present, two different types of circuit board are provided in five different configurations.

- GPB—a general processor board providing an ultraSPARC processor with a hard disk drive in one of the PMC module positions;
- IBEN—an Ethernet interface board with a PowerPC processor and Ethernet PMC modules;
- IBAM—a 155 Mbit/s interface for ATM providing a PMC module for multimode fiber;
- IBE1—an E1 interface board with a PowerPC processor and E1 PMC modules; and
- IBT1—an E1 interface board with a PowerPC processor and T1 PMC modules;

New types of processor board can be added when needed. Interface boards currently under development or in the research and development (R&D) plan include interfaces for E3/T3, 155 Mbit/s single-mode fiber, 155 Mbit/s electrical interface, and gigabit Ethernet.

All current circuit boards require two circuit-board positions. A new interface board combining the cPCI PowerPC module with a carrier board is under development to provide the same features with a single circuit-board position.

Figure 5
WPP circuit board—schematic view.



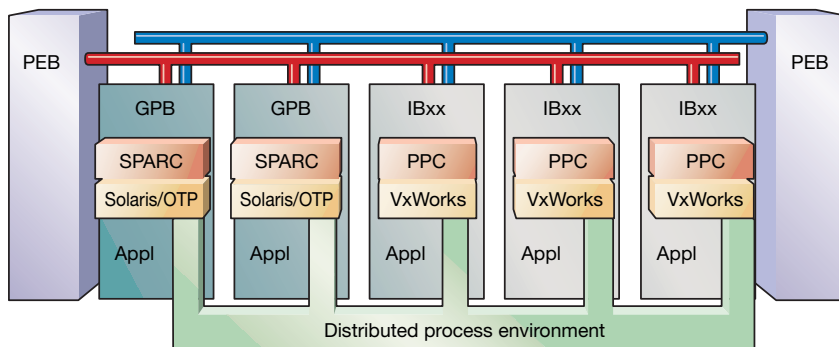


Figure 6
Distributed process environment (DPE).

WPP software

Three different types of operating system are provided:

- Solaris (on the UltraSPARC processors) for control tasks;
- Open telecom platform (OTP), for Erlang virtual machine support; and
- VxWorks (on the PowerPC processors), for real-time characteristics.

In addition, Java support is being developed for the UltraSPARC processors.

Each application executes locally on a local processor and its operating system. To create one system, the loosely coupled processors are held together by means of distributed process environment (DPE) middleware. The DPE supports redundancy and the distribution of functions, detects application failure, and can activate redundant applications in different ways.

The WPP platform supports various protocols to provide access to traditional telecommunications and data communication networks. Over the interfaces in the node, the platform supports signaling system no. 7 (SS7), frame relay, and IP. The IP interfaces support full redundancy.

The IP functions set provides limited transit routing, including support for OSPF, BGP, and the routing information protocol (RIP). IP access to applications is provided through the logical router.

The GGSN requires connections to numerous intranets and the Internet. The address constraints of the present IPv4 networks are commonly handled using private address ranges with corporate intranets. Consequently, the node has been designed to connect many intranets that use conflicting address ranges.

The platform also supports basic O&M interfaces, to provide access to existing telecommunications and data communication networks. A thin-client concept, using the hypertext markup language (HTML) and the CORBA IP-based inter-ORB protocol (IIOP), has been implemented to provide local and remote management via standard Web browsers and Java applets.

An SNMP agent provides access to standard datacom management systems. A basic network surveillance interface provides access to Ericsson's fault management systems. A fault management application is available, as are auxiliary applications that support the collection and storage of performance and charging data.

Hardware accelerated encryption support

The use of the public Internet or IP networks exposes systems to security risks. In addition, the use of wireless mobile systems exposes systems to eavesdropping and information security risks. To neutralize these risks, encryption seems the obvious solution, perhaps augmented by authentication. IPsec is the industry standard for IP transmission but requires high-capacity processing for both the encryption and the authentication steps. To support IPsec, a special module has been developed that contains an FPGA circuit.

All interface circuit boards have been designed to include one FPGA module and two PMC modules. For an IP router interface, the FPGA module provides IPsec security. Similarly, for a frame-relay interface, it provides mobile subscriber encryption—for the GPRS *Gb* interface.

Distributed process environment

As mentioned above, the wireless packet platform uses two types of processor: the UltraSPARC and PowerPC. The UltraSPARC processor runs Solaris/OTP—a combination well suited to transaction and control tasks. The PowerPC processor runs VxWorks, a real-time operating system that is well suited to critical, real-time tasks. Depending on the architecture of future mobile networks, the two processors can work in unison, for example, in a combined server and gateway node, or apart, as separate servers or gateways.

The design of the application can be independent of the number of different processors used in a specific node configuration (Figure 6).

DPE middleware functions

The DPE middleware holds the loosely coupled processors together and supports the application with process supervision, distribution, redundancy, hardware management, and software upgrades. All processes are supervised and a notification service informs applications that a certain application has stopped.

Software is distributed in the node by means of the DPE and a software configuration file (SCF) system, which employs four distribution methods according to

- circuit board—software modules can be distributed to certain circuit boards. For example, applications that apply to frame boards can be distributed to all circuit boards configured as frame relay boards;
- number of instances—the distribution module can specify that a certain number of instances of a software application must be made available in the node;
- circuit board position—software modules can be loaded onto circuit boards that are located at a particular position in the magazine; and
- redundancy—redundancy can be distributed either by two instances with a hot standby relation or in accordance with $n+1$ redundancy with one standby for a number of similar applications.

The DPE supports three types of redundancy. With hot standby, two applications are always loaded; one is active and one is on hot standby. To a limited extent, the DPE can replicate data between the two applications. According to $n+1$ redundancy, several applications are running, and one standby application is activated on standby in case any of the applications fails. Finally, the functional distribution can be set up to guarantee that one instance of an application is always available. If an application fails, the DPE activates a new instance of the application.

The DPE also manages equipment, allowing “hot” insertion of circuit boards, and the controlled as well as uncontrolled swapping of circuit boards. Any change in equipment will cause a notification to be sent to the applications, which can then redistribute functions. The functions can reallocate applications when boards are removed or fail. They also support plug-and-play, say, when a new circuit board is added to the node or a faulty circuit board is replaced by a new one.

The DPE supports smooth software upgrades as well as controlled patch handling

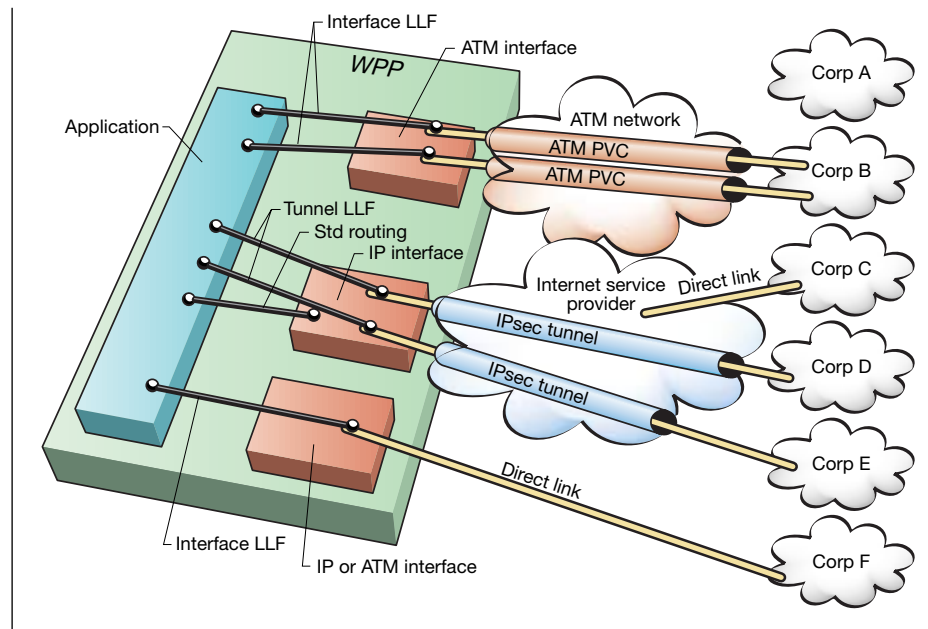


Figure 7
Logical routing.

for emergency corrections. Its support to the application is the same, regardless of whether the application executes on an UltraSPARC or PowerPC processor. Thus the applications need not be adapted to different sizes or configurations of nodes.

Logical routers

Packet-handling nodes address different kinds of applications using IP. Address conflicts usually arise when an IP-based platform is connected to numerous different IP networks. Many networks do not normally use public IP addresses but rather one of the private address ranges set aside by the Internet society. Consequently, the packet-handling node must be able to handle conflicting address ranges. It does so using a combination of methods (Figure 7):

- ordinary IP routing—the node can handle one address range series and perform standard routing from the interface boards up to the application. However, because of address conflicts, only one such address range can exist at any one time;
- the interface link-layer-forwarding (LLF) function (APN routing) permits the node to bypass IP forwarding and connect all packets on one interface directly into an application. The interface types can be ei-

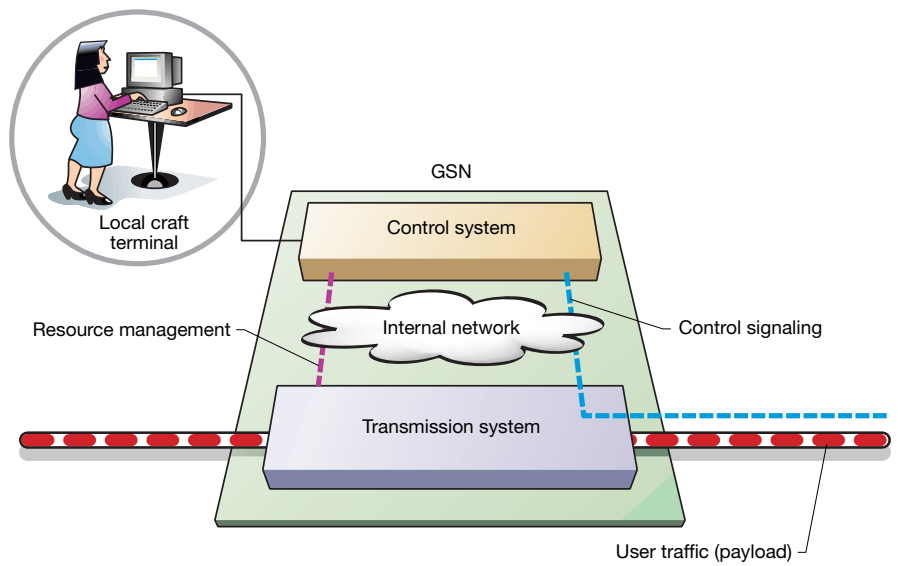


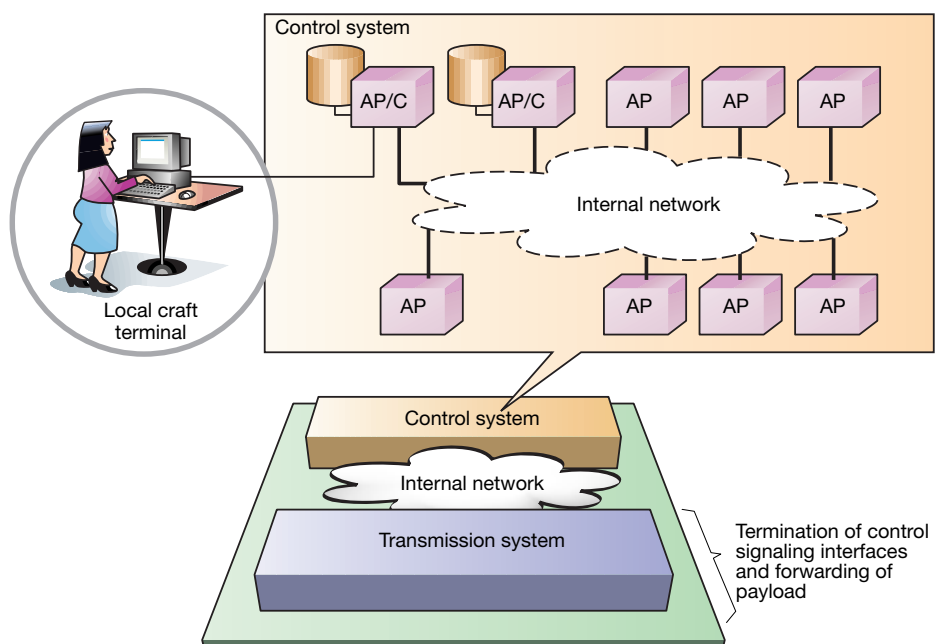
Figure 8
The GSN software architecture.

ther physical interfaces, such as Ethernet and E1, or PVC logical interfaces on an ATM interface; and

- the tunnel LLF function permits the node to terminate several IPsec tunnels on the incoming physical interface and

to cause each tunnel termination to bypass IP forwarding, instead connecting all packets from the tunnel end-point to an application. The LLF design allows thousands of different intranets and the Internet to connect to the same node

Figure 9
The control system architecture.



even if they all use the same IP address range.

GSN software architecture

The basic premises of the GSN software architecture are openness to third-party hardware and software, layering, and robustness (Figure 8). Management is based on the thin-client principle defined in the WPP, which only requires that the local craft terminal (LCT) should support an HTTP/CORBA interface with the GSN.

The software architecture defines two separate and loosely coupled computing entities: the control system (for mobility management), and the transmission system (for user traffic). One reason for the distinct separation of control and transmission is the need for flexible and independent scaling. Another important aspect is the totally different focus of the two systems.

The main focus of the control system is on

robustness and an efficient design environment, whereas that of the transmission system is on performance and low manufacturing costs. The GSN software architecture facilitates the physical separation of the SGSN into a server node, which handles the control parts, and a media gateway node, which handles the transmission of data.

Control system

The control system (Figure 9) consists of

- traffic control functions, such as GPRS mobility management and higher-level protocols (MAP);
- object control functions, such as start/restart, distribution, and communication (the network-element object-control middleware, NOC);
- O&M functions; and
- adaptation functions (drivers) for the transmission system.

The O&M functions are written in Erlang and Java. The rest of the control system is in Erlang.

BOX C, NCS COMPONENTS

The network control system (NCS) is a subsystem that implements the NOC middleware layer and some services that are not part of the NOC. A fundamental third-party building block of the middleware platform is Erlang/OTP. Other third-party components, such as the WPP, are interfaced by a few products and can thus be replaced by third-party components that offer the same functionality.

Control system hardware redundancy

If an AP breaks down, the NCS fetches a new AP from the standby pool. No redundancy functionality is implemented in the application—all redundancy is handled by the NCS. If the standby pool is exhausted, the NCS merges the traffic onto an AP in operation.

Transmission system hardware redundancy

If a device processor (DP) breaks down, the NCS fetches a new DP from a suitable standby pool. Under standard conditions, no redundancy functionality is required in the application. Callbacks are made to the control module/device configuration modules.

Load balancing

When a new connection is created in the system, the NCS selects the most suitable AP and DPs.

Start/Restart

The NCS coordinates all restarts including a start-phase-driven callback to the application.

The NCS manages four restart levels: connection, minor local restart, minor restart, and major restart. The first three restart levels maintain connections, whereas all connections are released at the major restart level. The NCS also handles mapping to the WPP's restart levels.

Resource control and supervision

The NCS controls and supervises APs and DPs that use any available WPP or OTP functionality. In the event of failure, the NCS takes appropriate action (correct restart level or redundancy).

Programming model framework

The NCS defines how functionality is to be implemented in Erlang. Major support for the programming model is offered, such as persistent storage, process handling, and the supervision of data access.

Overload protection and recovery

The NCS can grant new load in the system to avoid restarts due to overload. However, if a restart is caused by overload, the NCS might release connections in order to recover.

Software upgrades

The NCS coordinates all software upgrade activities and performs the required callbacks to the application to convert data structures. A software upgrade is always accompanied by a major or minor restart.

Event handler

Highly optimized event handlers are part of the dynamic processes in the control system. Anyone can subscribe to an event view defined by the event suppliers. There are many event views, such as those for charging or event recording. The event suppliers are located in the dynamic traffic control worker process.

Event recording

The event-recording function enables operators to record events associated with an application on a per-connection basis. The event-recording function is initiated from a management terminal. Operators can also view logged events from the management terminal.

Performance management

The WPP offers a central API for performance management. To support a distributed environment, the NCS implements a distributed performance-management framework in the control system by collecting distributed counters.

Charging support

The WPP offers a central API for the transparent forwarding of call data records. The NCS implements a distributed framework for managing the charging devices, which collect charging information from the payload and forward it to the AP/C.

The main purpose of the control system is to process high-level protocols and to control payload routing in the transmission system. The software control system does not put any requirements on the implementation environment of the transmission system.

The control system software features a distributed architecture that is based on the OTP. Several application processors (AP) are interconnected via an internal network. In this context, an AP is any computing resource that can run the OTP and WPP. The internal network is defined by the WPP. In the current release, it is switched Ethernet.

Two APs, denoted AP/C, are dedicated for central O&M functionality (one is on standby). Another AP is assigned the task of running global traffic control functionality. Its main purpose is to distribute jobs to the APs. The remaining APs run local traffic control functionality, such as mobility management.

The control system scales in a linear fashion, from a node consisting of only one AP/C (central, global, and local functions), to a node that consists of numerous APs, where each local AP is itself a scalable entity.

Distribution model

Apart from the AP/C, which contains software for node-level functions, such as O&M, every AP contains the same software. This software defines a simple distribution model in which the processing of one context is handled by a single AP. For example, if there are 70 mobile stations in the area covered by seven APs (Figure 2), each AP will handle the context of 10 mobile stations. This approach simplifies design and results in a highly scalable system.

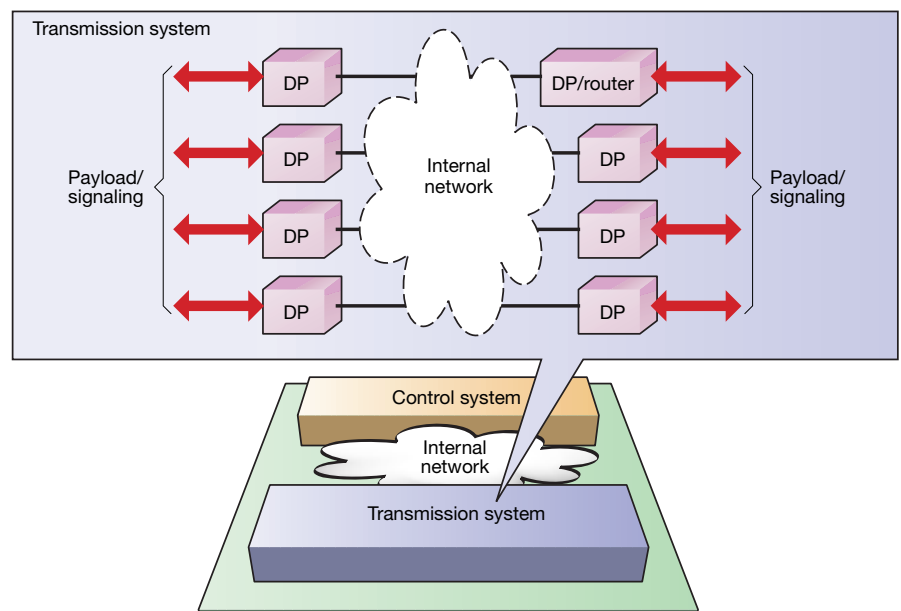
Transmission system

The transmission system (Figure 10) consists of the

- transport, routing, and processing of user traffic (payload); and
- termination of the lower layer of signaling protocol stacks, such as the message transfer part (MTP), signaling connection control part (SCCP), and transaction capabilities application part (TCAP) in SS7. Apart from the management interface, every interface is terminated by the transmission system.

A framework has been designed for the

Figure 10
Transmission system architecture.



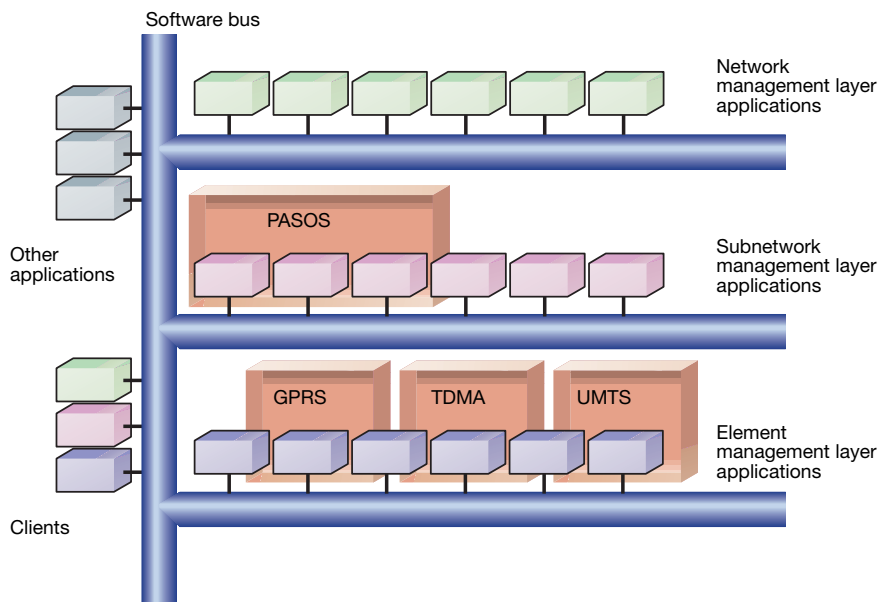


Figure 11
Logical management architecture.

development of GPRS applications for a transmission system based on Vx-Works/WPP. In this system, the payload protocol stacks are implemented in a STREAMS environment.

Software layers

The GSN software is divided into three layers that span the control and transmission systems:

- the traffic control (TC) layer;
- the network element object control layer; and
- the resource deployment (RD) layer.

These layers facilitate the active use of plug-in technology, which simplifies the addition of new application components.

The traffic-control and resource-deployment layers are application-specific layers, in which GPRS, UMTS, and TDMA services are implemented. Application functionality requiring robustness and extensive support is implemented in the traffic-control layer. Similarly, application functionality that requires high performance is implemented in the resource-deployment layer, as are external interfaces and the low-level protocols for signaling.

The NOC is a middleware layer that supports traffic-control and resource-deployment layers. It is a generic layer in the sense that it can support any packet-processing application and requires no mod-

ifications. In other words, applications are developed in the traffic-control and resource-deployment layers; the NOC serves as the communication channel and glue between the application components.

GSN management

The management system (Figure 11) for the packet-switched core network focuses on customer needs. The solution is made up of logical applications that can be accessed by any desktop computer running a Web browser. The following management applications ensure clear separation between element, network, and subnetwork-level management:

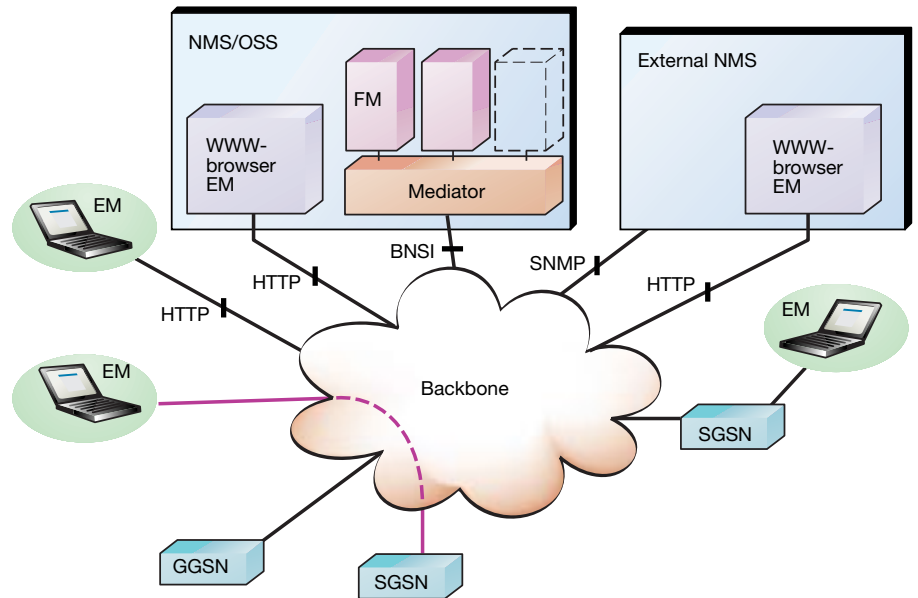
- network-level management mediation;
- subnetwork manager (PASOS); and
- embedded element manager (EM).

Network-level management mediation

Because they support open interfaces for fault and performance management, Ericsson's GSNs can be migrated to a network-management system by means of

- a packaged solution that is integrated into the GSM OSS system;
- a package containing the network-management system for the highest level of supervision offered by Ericsson; or
- adaptation units for integrating common network-management packages.

Figure 12
Embedded element manager.



TRADEMARKS

PowerPC is a trademark of International Business Machines Corporation.

Sun, Sun Microsystems, the Sun Logo, Solaris, and Java are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States and other countries. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. in the United States and other countries. Products bearing SPARC trademarks are based upon an architecture developed by Sun Microsystems, Inc.

VxWorks is a registered trademark of Wind River Systems, Inc.

Subnetwork manager

The subnetwork management system (called the packet-switched operation support system, PASOS) is a task-oriented, portable software application whose main role is to manage several nodes with very few commands. PASOS provides the operator with configuration-management, software-management, and equipment-management applications whose powerful plug-and-play capabilities guarantee the integrity of software, configurations, and data. The aim is to improve functionality and ease-of-use while enabling cost-effective administration of GSN configurations in a clear and consistent manner.

The management systems provide a task-oriented user interface that can be operated either locally or remotely. All management documentation is available online.

Embedded element manager

The embedded element manager is a fundamental component of the packet-switching core network O&M system. All software required for management tasks is contained in the GSNs. The element-management solution is implemented using a client/server architecture in which the client can be installed on any desktop computer that supports a Web browser and a Java

virtual machine. The server part executes on the network element. The GSNs support the hypertext transfer protocol (HTTP), SNMP, lightweight directory access protocol (LDAP), IIOP, BNSI, and FTP.

An element manager can be connected locally or remotely from different entities (Figure 12). The actual connection to the GSN is transparent to the user.

Fault management

Fault management software provides functions for detecting and isolating improper behavior within a GSN. The GSN always indicates the severity of an alarm, and provides a procedure that assists in correcting each fault. This procedure is automated (with hypertext links) and allows control operations, or recovery actions, to be launched from the fault management application.

Performance management

Performance management software provides support for collecting statistics and events generated by the node relative to the quality and availability of services, optimization within the node or the subnetwork, and planning. Measurement groups can be created, modified or deleted through the application interface. A range of mea-

surements apply to trend analysis and forecasting.

Configuration management

Element management and subnetwork management should be configured through the LDAP server. That is, central configuration of the nodes ensures consistency in the GSN network. Nevertheless, it is possible to configure a node remotely without using LDAP, or locally from a task-oriented GUI through which parameters associated with the SGSN can be set and modified. Configuration management applies to

- software management—for example, loading, installing and uninstalling software, checkpointing software configurations, and software dump handling;
- equipment management—for instance, listing equipment with administrative and operational states, and changing the administrative state (block or deblock) or operational state (reset);
- execution management—for example, listing applications within the node and their execution states, starting, stopping, killing, or restarting applications; and
- setting parameters—for example, configuration parameters for SS7, routers, and physical interfaces.

Security management

Security management software at the network-element level provides functions that protect the resources of network elements against intentional and unintention-

al destruction. Hence, it includes common security functions, such as the administration of end-user profiles and user-access authorization, as well as logging facilities that record user activities and all attempts to access the network elements.

Multilingual support

The packet-switched management solution provides operators with multilingual support. PASOS supports English-, Japanese- and Chinese-language documentation, help texts, and menus. System-generated information, however, such as alarms, events, and the names of devices, are only given in English.

Conclusion

Ericsson's solution for introducing GPRS into a GSM system—as well as the GPRS domains within a UMTS or TDMA system—is based on two new nodes: the SGSN and GGSN.

Initially, these nodes can be combined in the same physical node. At a later stage, the centralized GPRS node can be separated into a dedicated SGSN and GGSN.

If future expansion involves other access networks, the nodes can work together, using central parts.

The architecture allows for easy separation of the SGSN into a server node and a media gateway node, thereby allowing more flexible allocation of power between control and data throughput.

REFERENCES

1. Granbohm, H. and Wiklund, J.: GPRS- General packet radio service. Ericsson Review Vol. 76 (1999):2, pp.82-88.