IP Technology in WCDMA/GSM core networks

Heino Hameleers and Christer Johansson

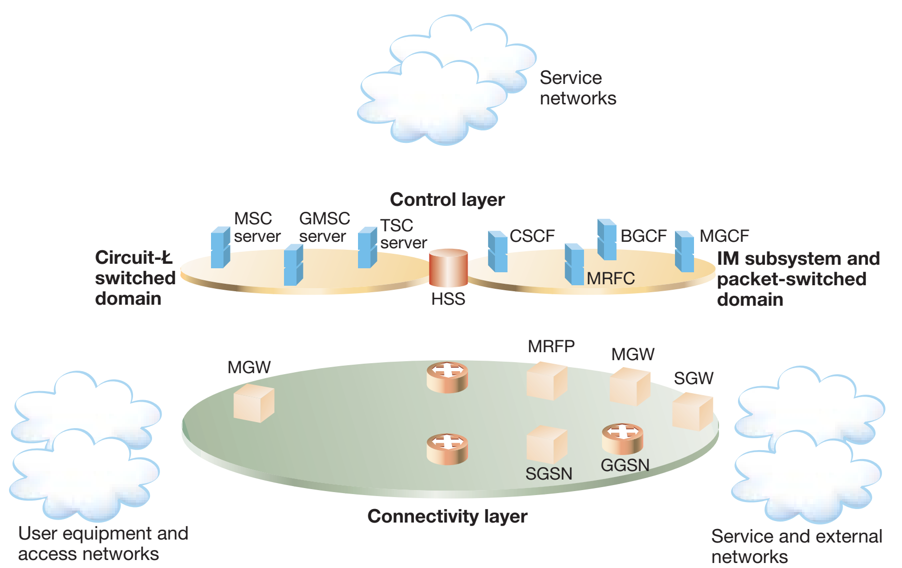
Mobility and the Internet, the two most dynamic forces in communications today, meet in the design and implementation of the mobile core network. Support for new end-user services and a common transport technology are the main drivers of integration of IP technology into our systems. The IP multimedia application plays a special role in providing these new end-user services. IP transport technology addresses the vision of multi-service backbone networks, based on a single network layer technology. Ericsson provides complete solutions and products to support deployment of new IP-based services and transport networks. Moreover, Ericsson's flexible core network architecture allows operators to address these drivers in an independent way.

The main parts of this article describe how the requirements for the two main drivers for IP technology are met in the mobile core network. Paying special attention to support for the IP multimedia application, the authors describe how support for IP applications is implemented. They then describe how IP transport technology can be supported, including site configurations and specific issues like quality-of-service and network redundancy.

Introduction

The marriage of mobile communication and the Internet has the potential to produce a revolution whose scope far outpaces that associated with the advent of the personal computer (PC). In parallel with the fantastic worldwide growth of mobile subscriptions, the fixed Internet and its service offerings have grown at a rate far exceeding all expectations. And this is only the beginning—the future will be even more spectacular given that the number of people connected to the Internet will continue to increase and that GPRS and WCDMA mobile networks will enable connectivity virtually everywhere and at any time with any device. A new form of interactive communication behavior is emerging from the combination of different media with applications that are randomly invoked by multiple users and end-systems.

Figure 1
Different core network functions in the layered network architecture.



Drivers of IP technology

There are two main arguments that drive the integration of IP technology into mobile core networks:

- support for (new) IP applications to generate (new) revenues; and
- a common transport technology to reduce costs.

The entire mobile telecommunications industry is funded out of the end-user's pocket. Therefore, to ensure future growth in the industry, end-user value needs to be enhanced. Service and application offerings are the prime drivers of the entire network and terminal evolution. End-to-end IP solutions must have access to the dynamic IP applications industry. This is the main prerequisite for success in the world of third-generation mobile networks. In this world, visualization and real-time behavior are fundamental components for bridging distance and for giving the end-user experience of "presence" a broader spectrum of human senses.

IP is gradually becoming a dominating transport technology thanks to recent advances in optics and routing technology and the impact that these have had on price/performance. When combined with other key technologies, such as IP-based virtual private networks (VPN), IP enables a new generation of advanced multiservice networks. The use of a common infrastructure based on a single technology simplifies network implementation and operation and helps reduce costs. Ericsson's core network architecture and product solutions allow operators to address these two main drivers of IP technology individually.

Core network overview

The mobile core network is at the heart of present-day mobile communication networks. It provides support for network features and telecommunications services, including essential functions such as session and call control, charging, mobility, and security. Because of its central role in the overall architecture, the mobile core network interfaces with and coordinates other network elements, including user equipment, radio access networks and service networks.

Core network architecture

Ericsson's core network solution is based on the separation of functionality into a control layer, a connectivity layer, and an application layer. The control layer hosts network control servers that are in charge of call or session set-up, modification, and release. The control servers might also handle mobility management, security, charging and interworking functions that relate to external networks at the control plane level.

The connectivity layer hosts routers, switches, signaling gateways, media gateways and other user-plane functions. Its routers and switches provide transport capabilities for traffic on the control and user planes. The media gateways facilitate interworking on the user plane. This includes interworking between different transmission technologies and media formats.

The interface between the control layer and the connectivity layer mainly consists of gateway control protocols. The network control servers use these interfaces to manipulate media gateway resources in the connectivity layer.

The application layer, which is implemented as part of the service network, hosts application and content servers. There are two interfaces between the core network and the service network: a horizontal interface and a vertical interface.

The horizontal interface between the core network and the service network refers to regular peer-to-peer or client/server mode of operation for typical end-user applications, such as Web browsing, e-mail and audio/video services. These applications are normally invoked by an end-user but might also be invoked by an application server.

The vertical interface allows applications that reside on specific application servers to complement or modify the normal procedures for setting up calls or sessions through the core network. These applications interwork with the core network through a set of standardized application program interfaces (API).

The layered architecture allows each layer to evolve independently and in pace with the evolution of the market and technology. It also supports the migration to new transport technologies, since the upper layers are independent of the transport technology deployed in the connectivity layer. The layered architecture also allows different, optimized technologies to be deployed in the connectivity layer (which is payload-processing-intensive) instead of in the control layer (which is transaction-oriented). Figure 1 shows how the different core network functions fit into the layered architecture.

In the circuit-switched domain, the mobile services center (MSC), gateway MSC (GMSC) and transit services switching center (TSC) servers are part of the control layer.

BOX A, ABBREVIATIONS

3GPP	Third-generation Partnership Program
AAA	Authentication, authorization and accounting
AF	Assured forwarding
AMR	Adaptive multirate
ATM	Asynchronous transfer mode
AUC	Authentication center
BE	Best effort
BGCF	Breakout gateway control function
BGP	Border gateway protocol
BICC	Bearer independent call control
BSC	Base station controller
CN	Core network
CPP	Cello packet platform
CS	Circuit-switched
CSCF	Call session control function
CSE	Customized applications for mobile network-enhanced logic (CAMEL) service environment
DNS	Domain name server
DS	Differentiated services
DSCP	DS code point
ECMP	Equal cost multipath routing
EF	Expedited forwarding
FTP	File transfer protocol
GGSN	Gateway GPRS support node
GMSC	Gateway MSC
GPRS	General packet radio service
GSM	Global system for mobile communication
GSN	GPRS support node
GSTN	General switched telephone network
GTP	GPRS tunneling protocol
HLR	Home location register
HSS	Home subscriber server
HTTP	Hypertext transfer protocol
I-CSCF	Interrogating CSCF
IETF	Internet Engineering Task Force
IMS	IP multimedia subsystem
IP	Internet protocol
IPSec	IP security
ISC	IP multimedia service control
ISDN	Integrated services digital network
ISP	Internet service provider
ISUP	ISDN user part
LAN	Local area network
LER	Label edge router
LSP	Label switched path
MAP	Mobile application part
MGCF	Media gateway control function
MGW	Media gateway
MPLS	Multiprotocol label switching
MRF	Media resource function
MRFC	MRF control part
MRFP	MRF processing part
MSC	Mobile services center
O&M	Operation and maintenance
OSA	Open service architecture
OSPF	Open shortest path first
PBN	Public backbone network
PCM	Pulse code modulation
P-CSCF	Proxy CSCF
PHB	Per-hop behavior
PLMN	Public land mobile network
PS	Packet-switched
QoS	Quality of service
RAN	Radio access network
RFC	Request for comment
RNC	Radio network controller
RTSP	Real-time streaming protocol
SCS	Service capability server
S-CSCF	Serving CSCF
SDH	Synchronous digital hierarchy
SDP	Session description protocol
SGSN	Serving GPRS support node
SGW	Signaling gateway
SIP	Session initiation protocol
SLF	Subscriber location function
SSF	Service switching function
STM	Synchronous transfer mode
TDM	Time-division multiplexing
TSC	Transit services switching center
UE	User equipment
UMTS	Universal mobile telecommunications system
VLAN	Virtual LAN
VPN	Virtual private network
WAN	Wide area network
WAP	Wireless application protocol
WDM	Wavelength division multiplexing

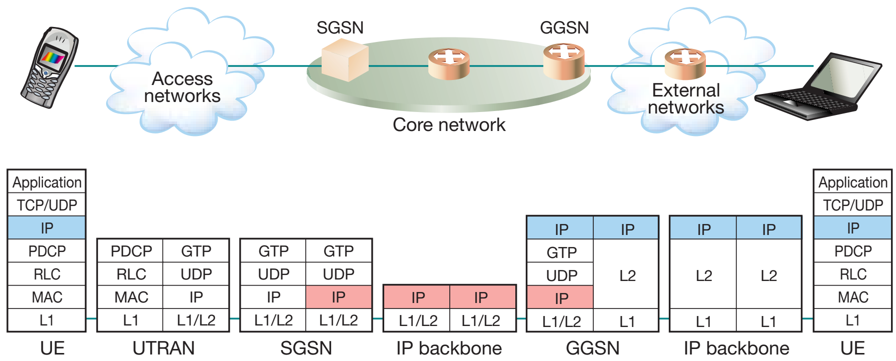


Figure 2
End-to-end protocol stack for an IP application that runs on top of a WCDMA/GSM packet-switched network.

The corresponding media gateway belongs to the connectivity layer.

In the packet-switched domain, both the serving GPRS support node (SGSN) and the gateway GPRS support node (GGSN) are considered to be part of the connectivity layer—they contain some control functionality, but the dominant functionality lies in providing IP connectivity.

With 3GPP Release 5, one more "domain" is being added to the mobile core network: the IP multimedia subsystem (IMS). The principal network entities of the IMS are the

- call/session control function (CSCF);
- media gateway control function (MGCF);
- breakout gateway control function (BGCF);
- media resource function (MRF) control part, or MRFC;
- MRF processing part (MRFP);
- media gateway (MG); and
- signaling gateway (SG).

The master subscriber database, called the home subscriber server (HSS), is common to the circuit-switched domain, the packet-switched domain, and the IP multimedia subsystem.

The two roles of IP

The two roles that IP plays in the mobile core network are also expressed in the core network protocol stacks. Figure 2 shows an end-to-end protocol stack for an IP application running on top of a WCDMA/GSM packet-switched network. The traffic leaves the mobile core network at the GGSN. There are also other scenarios in which traffic from the GGSN remains in the core network. However, for the sake of describing the roles of IP technology in the mobile core network a simplified scenario is used.

Two separate IP layers can be identified. The upper layer (drawn in blue) denotes the IP application layer, which runs between the user equipment (UE) and an external entity with which the UE is communicating. This would typically be an IP application server or another UE.

The lower layer (drawn in red) denotes the IP transport layer, which has only local significance to the public land mobile network (PLMN). The IP transport layer is needed to transport (control and user plane) traffic within the mobile network. In this particular case, the layer is terminated at the GGSN, where the traffic leaves the PLMN; routing is performed directly on the IP application layer. In other cases, IP application-layer traffic might continue to be carried over an IP transport layer even after it leaves the GGSN.

Support for IP applications

Role of the core network in providing IP applications

With respect to the IP application layer, the role of the mobile core network has traditionally been limited to providing a tunnel that allows the UE to communicate with another IP host. This support is implemented in the GPRS support nodes (GSN).[3]

In general, IP applications are transparent to the core network. This is true for all IP applications but one: the session initiation protocol (SIP) application. The main difference compared with other IP applications is the communication model. Most IP applications target a client-server model. The file transfer protocol (FTP) allows a client to download files from a server, the hypertext transfer protocol (HTTP) and

BOX B, HISTORY OF IP MULTIMEDIA STANDARDIZATION

Support for IP multimedia applications in mobile core networks was first discussed in 1999 in the 3G.IP forum. The 3G.IP forum was an industry consortium initially consisting of eight of the main operators and vendors, including Ericsson. It had set itself the goal of defining an IP technology based architecture for the next generation of mobile networks that would support voice, data and multimedia services.

This network architecture proposal was brought into the 3GPP forum. 3GPP has accepted the proposal and has since spent considerable effort to define a complete end-to-end architecture, including solutions for essential functions such as security, charging and QoS. 3GPP selected SIP as the session control protocol and has also mandated IPv6 for IP multimedia applications.

wireless application protocol (WAP) allow clients to download electronic pages with content from a server, and the real-time streaming protocol (RTSP) allows clients to "stream" content from a server. By contrast, SIP primarily targets a client-to-client communication model.

How is SIP supported in the mobile core network? The main parts are defined as the IP multimedia subsystem (IMS). Together with the circuit-switched domain and the packet-switched domain, the IMS builds the 3GPP mobile core network.

IMS: providing IP session control

To understand the IMS architecture, one must first understand the basic concepts on which the architecture has been built: the home-visited-interworking architecture and functional entities, and services.

Home-visited-interworking

In second-generation mobile systems, such as GSM, services are provided by the PLMN in which the subscriber is roaming. Personalized service information is transferred from the home PLMN to the visited PLMN. This approach requires that both the home PLMN and the visited PLMN support the service to be provided. That is, the services that can be provided to the end-user represent the lowest common denominator of what the home PLMN and the visited PLMN can support.

In the future, the differentiation between network operators will be made at the application and service levels, instead of at the access and network levels. Roaming subscribers will no longer be the exception, but will have become the norm. This implies that the provision of seamless services for roaming subscribers is increasingly important.

Ericsson has proposed the concepts of *home* and *visited* to describe the architecture of modern communication networks. *Home* denotes user data and services, whereas *visited* denotes connectivity and mobility. This implies that the main task for the visited network is to provide a subscriber with (mobile) connectivity to the home network (Figure 3). The home network hosts user data, session control and services.

Acceptance of these concepts implies that subscribers are always roaming in a visited network. However, the services are controlled from the home network, regardless of which visited network the subscriber is roaming in.

This approach limits the functional and protocol dependency between the home and visited networks, thereby

- minimizing the restrictions put on the services that can be deployed in the home network; and
- increasing the rate at which services can be deployed.

In addition to the personalized services provided by the home network, the home and the visited networks can each provide local services—however, these services are not tied to the user profile in the home network.

Although this implies that control signaling must always go through the home domain, the actual payload must not. Payload is routed independently of control signaling and can follow the optimal (shortest) path for efficient transmission and optimal quality of service (QoS).

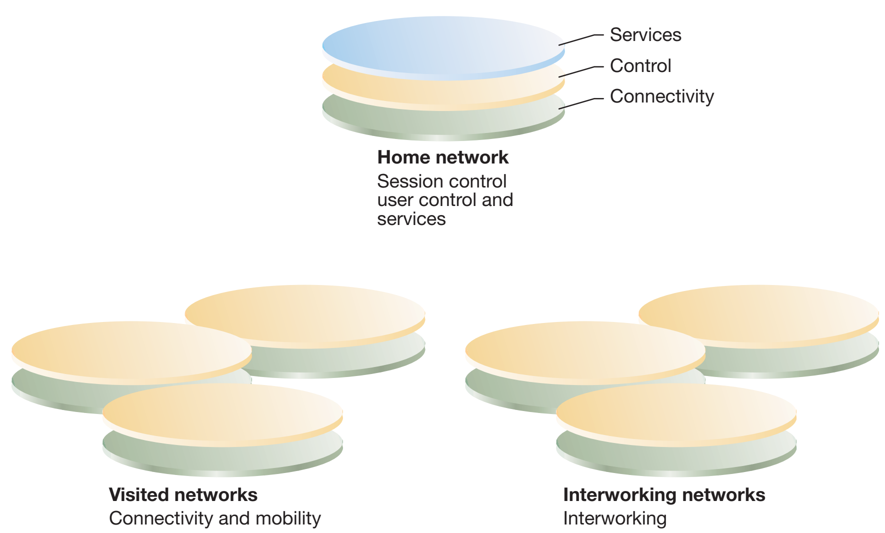
If the IMS is to interconnect with legacy networks, some functionality must be provided to enable the systems to interwork. To allow for the optimal user-plane transport path (for example, by keeping the session in its own network as long as possible), this functionality can be placed in an interworking network outside of the home and visited networks. Home, visited and interworking networks can be physically different networks or they can be implemented in one and the same network.

BOX C, IP MULTIMEDIA PROTOCOLS

The session initiation protocol (SIP) has been defined in IETF RFC 2543. It describes a way of supporting session-based applications over IP networks that involve one or more participants. In mobile terms this means that it allows a mobile client to set up an IP session to another mobile client. An updated version of SIP is expected in 2002.

SIP "uses" the session description protocol (SDP) to describe the nature of the multimedia sessions—that is, which media are included in the session, in which format the media will be transported, and so on. SDP has been defined in IETF RFC 2327.

Figure 3
The visited network provides connectivity to the home network.



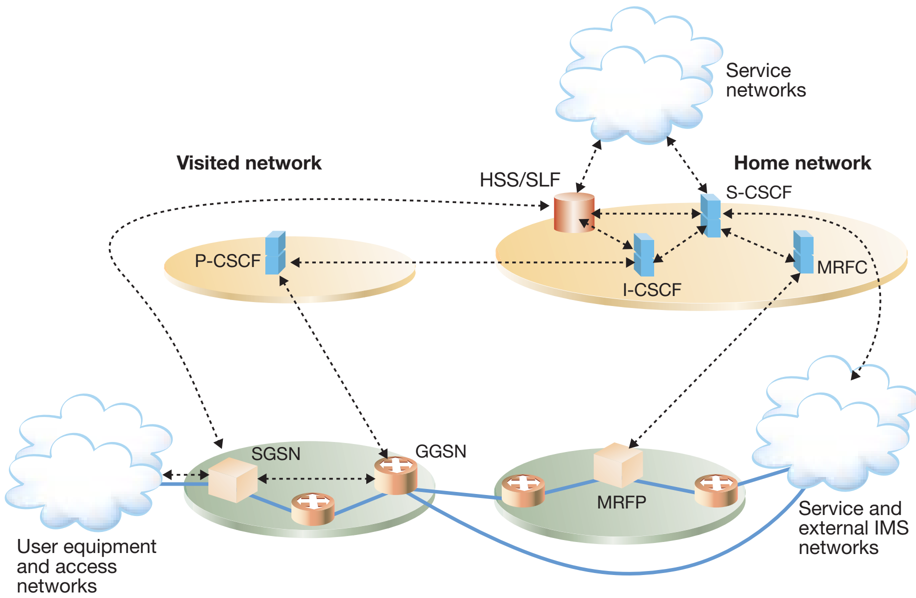


Figure 4
Basic IMS architecture.

Architecture and functional entities

The IP multimedia system is built around the call/session control function, of which there are three different kinds:

- the interrogating CSCF (I-CSCF);
- the proxy CSCF (P-CSCF); and
- the serving CSCF (S-CSCF).

The P-CSCF is the UE's first point of contact with the IMS. The P-CSCF forwards SIP messages received from the UE to a SIP server in the home network (and vice versa). The P-CSCF might also modify an outgoing request according to a set of rules defined by the network operator (for example, address analysis and potential modification).

The I-CSCF function, which forms the entrance to the home network, hides the inner topology of the home network from other networks and provides flexibility for selecting an S-CSCF.

The S-CSCF performs the session control services for the UE. This includes routing originating sessions to external networks and routing terminating sessions to visited networks. The S-CSCF also decides whether or not an application server is required to receive information on an incoming SIP session request to ensure appropriate service handling. This decision is based on information received from the HSS (or other sources, such as an application server).

All CSCF functions can generate call detail records for input to the charging process.

The HSS, which is an evolution of the home location register (HLR) and authentication center (AUC), holds the subscriber profile and keeps track of which core network node is currently handling the subscriber. It also supports subscriber authentication and authorization functions (AAA).

In networks with more than one HSS, the subscriber location function (SLF) provides information on the HSS that contains the profile of a given subscriber.

The media resource function (MRF), which contains the functionality for manipulating multimedia streams, supports multiparty multimedia services, multimedia message playback and media conversion services. The Third-generation Partnership Project (3GPP) has split the MRF into a control part (MRFC) and a processing part (MRFP). Figure 5 depicts a scenario in which a multimedia session interworks with a general switched telephone network (GSTN).

The BGCF selects the network in which the interworking is to be performed. If the interworking is performed in the home network, the BGCF selects an MGCF. If the interworking is to be performed in another network, the BGCF selects another BGCF or an MGCF.

The MGCF provides interworking func-

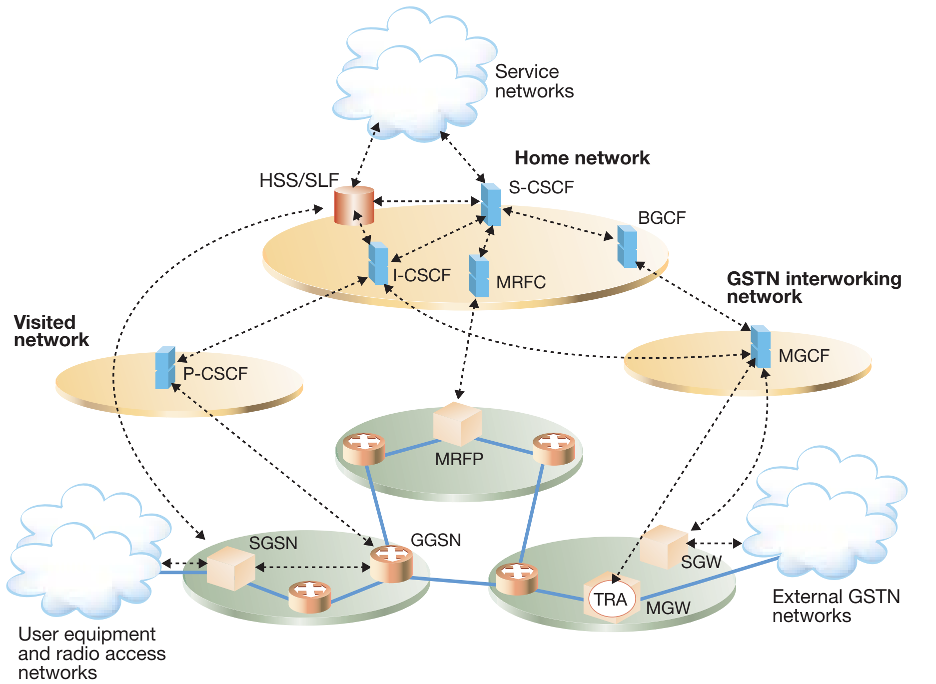


Figure 5
A multimedia session interworks with a general switched telephone network (GSTN).

tionality between SIP session control signaling from the IMS and ISUP/BICC call control signaling from the external GSTN networks. It also controls the media gateway that provides the actual user-plane interworking functionality (for instance, for converting between AMR- and PCM-coded voice). The signaling gateway provides bearer interworking functionality for the control signaling (ISUP/IP – ISUP/TDM).

Services

The 3GPP is working to define the IMS architecture, but its concepts and protocols were derived from the Internet Engineering Task Force (IETF). The IMS enables the convergence of, and access to, voice, video, messaging, data and Web-based technologies for the wireless user. The main tools that the IMS provides to build these services are

- peer-to-peer addressing architecture for IP-based sessions;
- flexible integration of "any" type of media into sessions;
- integration with other IP applications, such as RTSP, HTTP, and so on; and
- end-to-end QoS charging and security architecture.

Although the IMS mainly targets client-to-client services, its tools can be used to facilitate or enhance client-server-based services, such as those found in gaming. Moreover, besides providing real-time services (such as video-conferencing), the IMS provides non-real-time services (such as instant messaging).

Unlike second-generation mobile systems (such as GSM), specific IMS services will not be standardized. The 3GPP is merely defining the architecture framework and service capabilities that can be used to build services. The actual services are implemented on top of these capabilities by network vendors, operators or third parties.

In keeping with IETF principles, the endpoints of the system contain considerable intelligence for supporting services with little or no assistance from the network. However, there are also scenarios in which the network provides value-added services—for instance, to provide a presence service or to optimize the use of resources for conferencing services.

To support these network-controlled or network-assisted services, the 3GPP is defining an IMS service creation environment. The S-CSCF and the HSS each feature one or more vertical service creation interfaces, and the 3GPP is deliberating whether or not it will also include them in the MRFC.

Horizontal interfaces might also be considered. A horizontal interface would apply to a scenario in which the S-CSCF routes control of the user plane to an external ap-

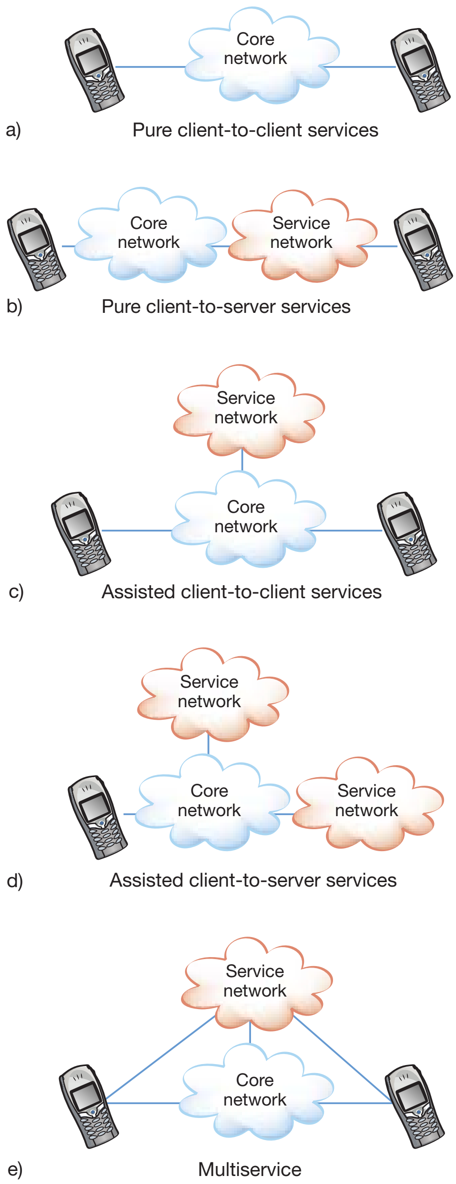


Figure 6
Overview of five service execution scenarios.

plication server. In that case, the application server would have full control over session routing and media. Figure 6 gives an overview of five service execution scenarios:

a) The main service functionality lies in the user equipment (UE). The core network can provide assistance for some value-added services. Example: a video-phone session between two clients.
b) The core network passes the session via a horizontal interface to an application server. Example: a gaming session.
c) The core network passes session control via a vertical interface to an application server. After session control has been returned to the core network, the session is established to the terminating user equipment. Example: a voice session with session forwarding.
d) The core network passes session control via a vertical interface to an application server. When session control has been returned to the core network, the session is passed to another application server. Example: a gaming session in which the service network determines the closest gaming server.
e) The main session runs between user equipment. In addition, there may be one or more sessions running between the user equipment and the service network. The user equipment coordinates these multiple sessions. Example: a voice session during which both clients view the same streaming video.

Core network products

Given the complex functional architecture of the IMS, careful consideration ought to be given to the mapping between functions and actual products. The result of implementing each function as a separate physical box would be a complex system that is extremely hard to operate.

In all probability, the first commercial IMS products will be deployed on a small scale. Therefore, cost-efficient, entry-level solutions should exist that provide enough capacity to handle the initial IMS traffic. As the market evolves, more flexible and powerful configurations will also be required.

The Ericsson IMS product line will fulfill near-term and long-term requirements by implementing the IMS functions in an integrated yet modular way, by

- adhering to a horizontally layered architecture;
- providing scalability from entry-level (integrated) to high-end (distributed) configurations;
- using a modular software architecture (for example, separate S-CSCF, I-CSCF and P-CSCF modules) that allows for distributed solutions;
- implementing servers on TSP; and
- implementing gateway and user-plane functions on CPP.[4]

Figure 7 shows some example IMS product configurations. The trial site provides full IMS functionality, including support for video-conferencing and interworking with GSTNs. It also provides a local service execution environment on the CSCF.

The figure also shows some long-term

configurations that enable handling larger volumes of IMS traffic. In this particular scenario, we can differentiate access, service and gateway sites. The access sites can host P-CSCF and possibly also I-CSCF functionality to handle the interface to the UE and to forward sessions to the appropriate service sites.

The service sites host the functions that provide the actual session control and services. The gateway site handles the interface to external GSTNs.

Figure 8 shows only a handful of many possible configurations. Given the flexibility of the IMS product offering, many other configurations can be created to suit specific network deployment scenarios.

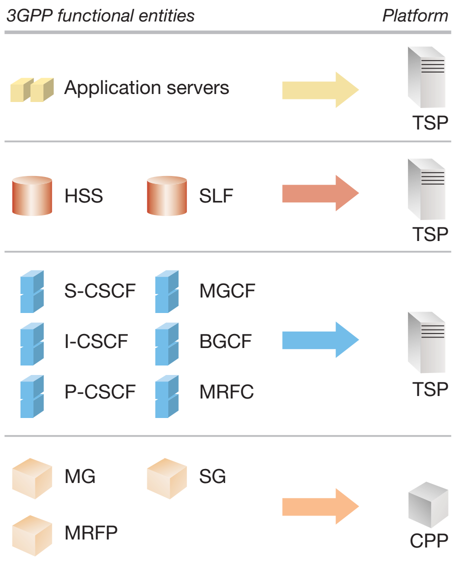


Figure 7
Example IMS product configurations.

IP connectivity

General

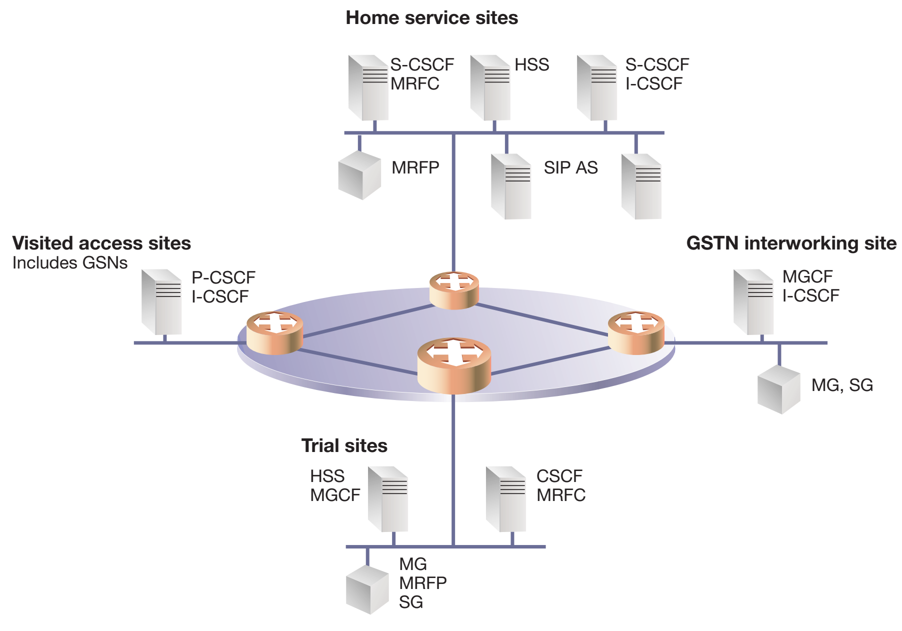
Most mobile operators have introduced (or are in the process of introducing) packet-based transport technologies into their networks. IP networks are being introduced for the intranet when WAP servers are introduced—following the introduction of GPRS—and when operators become Internet service providers (ISP). At the same time some operators are introducing asynchronous transfer mode (ATM) backbones to decrease transport costs for traditional voice services. Ordinarily, this is done at the transit level of the network.

More and more operators are seeing a need to harmonize the introduction of packet-based networks and have concluded that the best move is to invest in a single multiservice network. But the question remains: which basic technology should be used, ATM or IP? ATM is a mature technology for traditional voice services and for running IP. But for operators to run IP over ATM they must still invest in routers.

IP is mature technology for best-effort traffic. An operator who wants to deploy IP technology as the basis for a multiservice network that also supports traditional voice services (real-time) must consider several new aspects. At this time, there is no well-established dominant network design for this deployment. Emerging technologies, such as IP-VPNs, differentiated services, and resilience mechanisms, are now available in router products for these kinds of carrier-class IP network. These technologies also provide wire-speed routing for acceptable delay and jitter.

Ericsson's work with these issues has resulted in the multiservice IP network presented in this article. The basic assumption is that an operator who is building a multiservice IP network is not willing to sacrifice the characteristics associated with traditional services that are provided in TDM and ATM networks. We must thus solve the

Figure 8
Example configurations.



problem of running telephony over IP multiservice networks while preserving the characteristics of present-day networks, such as quality of service, congestion control, security, and resilience.

The objective is to run all services over the same IP network, which makes the IP layer the converging layer.

In the near future we expect to see a shift from synchronous digital hierarchy (SDH) to IP on wavelength division multiplexing (WDM). Today there is a lot of surplus capacity in fiber networks. The deployment of solutions based on, for example, Ethernet technologies, instead of SDH on layer 2 can give a dramatic reduction in transmission costs. To realize this potential, the services must be transported in a packet-based form.

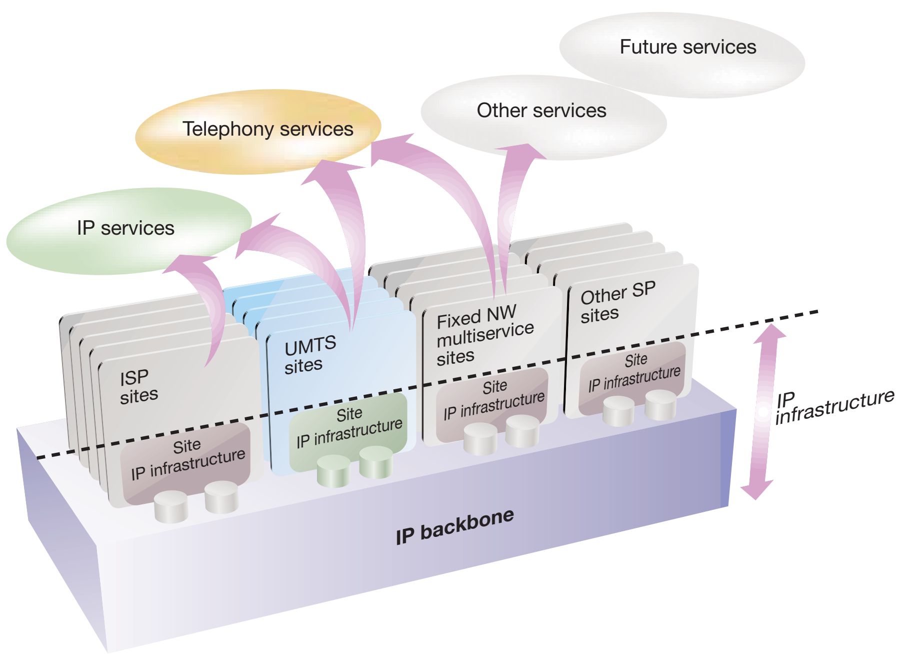
IP-based core network architecture

General

In an IP-based WCDMA/GSM core network, all core network elements use the connectivity services of a common IP infrastructure to interconnect user traffic and internal signaling.

A key component of the IP infrastructure is an IP backbone network that is used as a common backbone for WCDMA/GSM services, ISP services or fixed network services (Figure 9). The challenge is to find an IP network solution which integrates security, resilience, QoS, dual-stack IPv4/IPv6, and bandwidth efficiency, and which can handle connection-oriented services in a "connectionless" network. One possible network solution is described below, but variations of this solution are possible (depending on the operator's specific prerequisites).

Figure 9
Service provisioning over an IP infrastructure.



The structure of the IP infrastructure

The IP infrastructure is made up of two main tiers: a backbone tier, which is used to carry all traffic between sites; and a site infrastructure tier, in which the site IP infrastructure extends IP connectivity to the core network elements at the site. Each site IP infrastructure is attached to the backbone tier through one or more edge routers, which serve as traffic aggregation points and demarcation points between the local IP network and the backbone IP network domain.

The backbone IP network is a shared network, which also interconnects sites belonging to other service networks (non-WCDMA/GSM networks, such as ISPs or fixed service network providers). The IP backbone tier provides wide-area IP connectivity between sites. It is designed for simple high-speed packet transport and is typically built with large backbone routers interconnected with fast links, such as Ericsson's AXI 520/580 routers interconnected with Gigabit links.

In this article, we assume that the backbone IP infrastructure is run by a single operator. The routers support different classes of service and layer-3 VPNs. They also support layer-2 encapsulation, which can be used in parallel to IP traffic to carry Frame Relay or ATM between sites.

The site IP infrastructure tier extends IP connectivity to the various WCDMA/GSM network elements. Each site has a local IP infrastructure that is connected to the various elements at the site and serves as a liaison to the IP backbone through edge routers. The site IP infrastructure is adapted to low-cost, high-capacity traffic capabilities within the site, typically using Fast Ethernet, Gigabit Ethernet, and LAN switches that can make use of virtual LAN (VLAN) techniques. The site IP infrastructure is duplicated to guarantee full availability.

The edge routers connect the site IP network to the IP backbone. In terms of structure, they belong to the site IP network domain and to the IP backbone domain, and participate in routing protocols in both domains. The edge routers contain advanced functions (for instance, MPLS LER function, 2547*bis*, BGP, and filtering) for defining a service agreement between the site IP network and the IP backbone. A site typically uses a pair of edge routers connected to different core routers in the IP backbone.

Different types of site

In a practical network design the physical equipment is grouped together in sites. Sites can be grouped by type, depending on their role in the network. For the core network, three types of site can capture the needs of an operator. Other types can also be defined to suit specific operator conditions.

- The *primary site* includes a complete set of functions needed for a WCDMA/GSM network (control servers, media gateways, GPRS support nodes, and radio access network controllers). A primary site might also include a service network configuration. To distribute redundant load, a network can have several primary sites.
- The *secondary site* contains media gateways, GPRS support nodes, and radio access network controllers. If necessary, secondary sites can also have peering connections to other networks.
- The *concentrator site*, which includes media gateways and radio access controllers, is used for concentrating load far out in the network.

Peering connections to other networks can be made from any site. Figure 10 shows the mapping of different site types on the IP backbone. The primary site is the most important type. In fact, a complete network can be built exclusively from primary sites. Figures 11 shows example configurations of a primary site.

Logical networks and VPNs

Logical networks

Different kinds of information are exchanged between sets of network elements. Different types of network (STM/TDM, ATM, IP and SS7) can be used to handle different kinds of information flows, each with well-defined quality of service and little or no connectivity between the networks. It is thus possible to have complete separation of traffic between information flows.

In the context of a multiservice IP network, the IP infrastructure must be able to

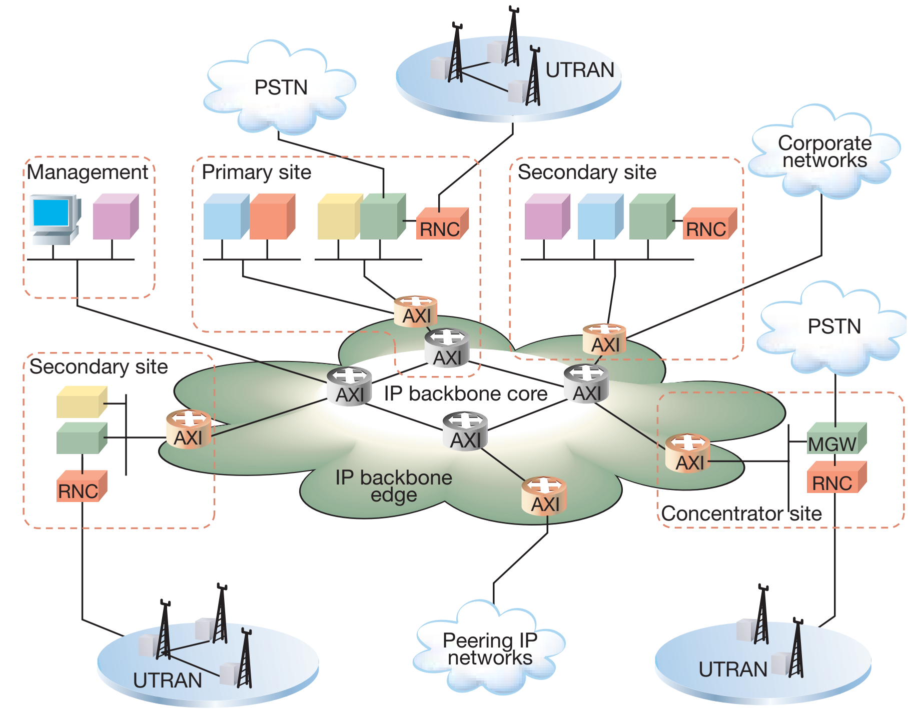
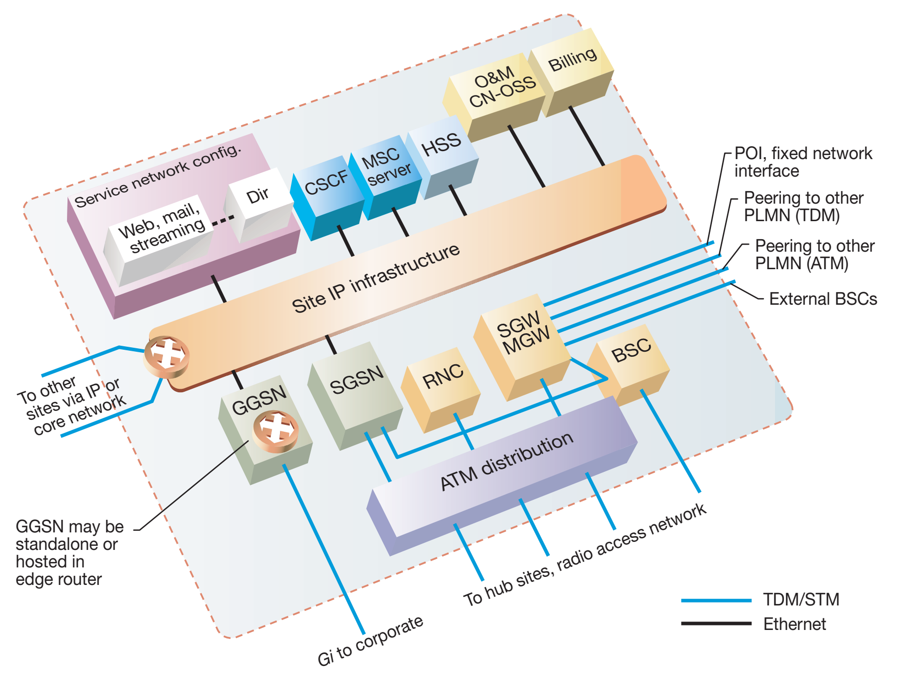


Figure 10
Mapping of different site types on the IP backbone.

Figure 11
Example configurations of a primary site.



handle this exchange of information flows. To facilitate traffic separation and ensure quality of service for the different traffic types, the WCDMA/GSM core network is conceptually divided into a number of logical networks.

Each logical network encompasses a particular kind of information flow between a designated set of functional entities in the WCDMA/GSM network elements. Furthermore, each logical network has a set of requirements with respect to connectivity, QoS, network availability, and so on.

The functional entities reside in WCDMA/GSM network elements at different sites. To support the logical networks, the IP infrastructure is configured into several virtual networks. Logical networks are implemented as virtual networks. Depending on the requirements for the supported logical network, the virtual networks are implemented using an appropriate set of capabilities in the site and backbone IP infrastructures (Box D).

Mapping of logical networks to VPNs

Logical networks can be grouped according to their specific characteristics into virtual networks (security, redundancy and resilience, addressing, QoS, and scaling). This helps operators to decide which VPN technology to apply for each network.

Figure 12 shows one possible implementation. In this example, BGP/MPLS IETF RFC 2547 *bis* is used to separate logical networks into layer-3 VPNs in the backbone. Virtual LAN tagging gives characteristics that are similar in nature to those of an ATM network and gives the operator good tools for controlling the different traffic flows through the IP network. Another simple mapping would be to avoid the use of VPNs and instead use filtering and BGP communities in the routers and client nodes.

Information flows can be encrypted using IPSec in the client nodes or IPSec VPNs between edge routers.

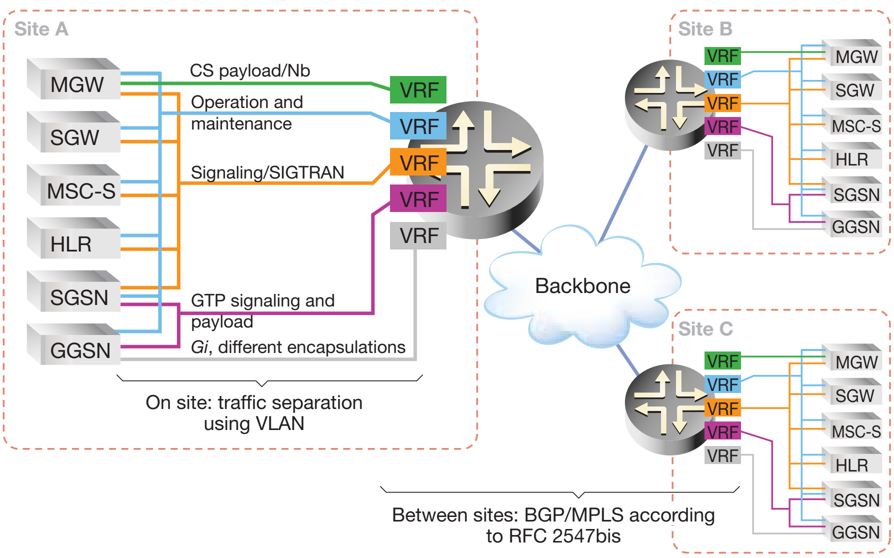
Network redundancy and resilience

The network redundancy principles are based on the assumption that the network should be able to withstand single-failure situations and resume service to its users with very short restoration time. Furthermore, it is assumed that

- the IP infrastructure—that is, the site IP network and the IP backbone network—can be configured to support alternative paths through the IP infrastructure; and
- redundant access can be provided to the infrastructure.

The main routing protocols in the IP backbone will be OSPF or IS-IS and BGP-4. At present, large networks require anywhere from a few seconds to tens of seconds to converge link state protocols, such as open shortest path first (OSPF) and IS-IS. This is too long for time-critical WCDMA/GSM

Figure 12
Possible implementation of grouping logical networks into VPNs.



BOX D, EXAMPLES OF LOGICAL NETWORKS

Signaling network
Used to carry UMTS signaling traffic such as H.248 and BICC signaling, making use of SCTP as transport protocol. The signaling network has application level awareness of alternative addresses to signaling recipients. Traffic volume is relatively low (a few Mbit/s to a site). Signaling traffic must be well protected from external traffic to avoid possibilities of network intrusion.

ISDN voice network
Carries circuit voice and data traffic between media gateways. The traffic is characterized by short packet traffic, with high QoS requirements.

***Gi* toward ISP**
Carries user traffic between GGSN and the Internet. The traffic must be well separated from the internal UMTS traffic. Traffic is currently of best-effort class only, but future traffic can be of any class of service.

***Gi* toward corporate networks**
Carries user traffic between GGSN and an appropriate access point to the corporate network. Separate virtual networks, thus allowing for overlapping addresses, will be required for each corporate network with several access points in the IP infrastructure.

***Gn*, *Gp* traffic**
User traffic carried in GTP tunnels between GSN nodes. QoS requirements are depending on the carried user traffic.

O&M network
Used for operation and maintenance of UMTS core network components. Very high availability and security requirements exist.

traffic. Therefore, to reduce fail-over times to 50 ms or less, an operator should not solely rely on layer-3 redundancy in the IP backbone. One solution is to rely on underlying SDH mechanisms. Another is to use MPLS with redundant secondary LSPs and the MPLS fast reroute mechanism.

The OSPF protocol with equal cost multipath routing (ECMP) is the recommended method of applying the network-provided multipath principle in the site IP infrastructure. ECMP distributes the traffic between multiple paths between routers and allows fast fail-over.

Quality of service

IP quality-of-service capabilities are implemented using overprovisioning, admission control, and differential services (DiffServ or DS).

Overprovisioning

Within the site, overprovisioning gives simple management and is the cheapest way of guaranteeing QoS. In the backbone, however, other means must be added due to the cost of bandwidth.

The extent to which overprovisioning can be reduced depends on how sophisticated the congestion control mechanisms are. The amount of overprovisioning needed is determined by fault situations (link or router failure), traffic concentration in conjunction with "abnormal" events, and the provision of best-effort capacity (capacity must never be completely starved).

Admission control

Various mechanisms and policies are used for controlling the amount of traffic that is injected into the IP backbone. Admission control is exercised at the edges of the network and serves to protect the backbone from being overloaded. An overloaded backbone results in packet loss and increased delays. There are three main ways of controlling admission:

- admission control in client nodes;
- policing of external interfaces; and
- policing of internal interfaces.

In principle, every node that generates a substantial amount of non-best-effort traffic should perform admission control—for example, GSN, MGW, data-intensive applications servers, such as streaming servers and O&M nodes.

BOX E, USE OF DIFFERENTIAL SERVICES CODE POINTS

EF PHB will be used for traffic that has requirements on lowest delay, the UMTS conversational QoS class.
AF4 is reserved for interactive traffic.
AF3 PHB is proposed for UMTS streaming QoS class.
AF2 PHB is proposed for CS and PS signaling.
BE is proposed for UMTS background QoS traffic.
Network control signaling is assumed to have a separate code point for PHB.
Different kinds of O&M traffic may have different requirements on the DiffServ PHB. Thus multiple DiffServ PHBs will be used for O&M.

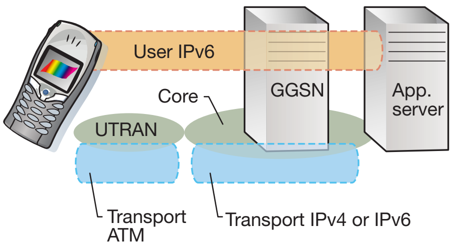


Figure 13
Coexistence of IPv6 and IPv4.

Admission can be controlled in different ways and according to different principles/algorithms. For instance, a trunk-based model can be used for voice traffic between media gateways. The media gateway controls the amount of traffic on the routes to other media gateways. Before new sessions are accepted, the originating node must check that the required bandwidth is available for the destination.

DiffServ

In the proposed QoS solution, differential services constitute a cornerstone for handling quality of service at the IP layer. Various applications in the client nodes and end-user applications or clients in terminals mark IP packets. Examples of applications in client nodes are the ISDN application in the media gateway, the GTP encapsulation function in the SGSN/GGSN, and the SIGTRAN application in the HLR. To differentiate between independent traffic flows, several DiffServ per-hop behaviors (PHB) have been proposed (Box E). When tunneling traffic over GTP, the DSCP in the SGSN and GGSN is marked as follows:

- The DSCP in the end-to-end IP header can be set by the UE.
- In the uplink this setting can be overwritten by the GGSN in accordance with the PDP context (APN) when the DSCP is forwarded over *Gi*.
- The DSCP in the outer IP header is set according to the PDP context (APN) in the SGSN (uplink and downlink) and for the GGSN (downlink).

The different routers are configured to schedule and prioritize traffic packets according to their DSCP. Ericsson's AXI routers and the embedded router in the media gateway and GSN provide rich mechanisms for this.

IPv6

Ericsson's IP solutions support IPv4 and IPv6. To have enough IP addresses for every connected terminal, operators will need to use IPv6 between end-users and applications. IPv4 and IPv6 will coexist in the IP backbone for a long time. Therefore, Ericsson products include dual-stack IPv4/IPv6 implementations.

Ericsson products for the IP infrastructure

The IP solution described above can be supported using Ericsson products. The RXI 820 real-time router capabilities are

TRADEMARKS

NetScreen is a trademark of NetScreen Technologies, Inc.

Extreme Networks is a registered trademark of Extreme Networks, Inc, Santa Clara, California in the United States, and may be registered in other countries.

being embedded in the media gateway (AXM 101) together with SIGTRAN capabilities and signaling gateway functions. The media gateway will also include transformation functions for payload between ATM, TDM and IP. In addition, IP and SIGTRAN capabilities are being included in server nodes, such as AXE.

The packet platform for GSN nodes continues to evolve to include new functionality and improved performance.

For the site IP infrastructure, Ericsson uses products from partners: NetScreen for firewalls, and Extreme for LAN switches. These products are well suited for carrier-class networks with high-availability architecture and hardware-supported filtering and forwarding. Ericsson's partnerships include the identifying and implementing of specific mobile network requirements.

Ericsson IP-Works develops DNS and DHCP products adapted to mobile networks.

Ericsson's partnership with Juniper Networks aims at providing a carrier-class router family with enhancement for the mobile core network. Ordinarily, edge routers are based on the AXI 520 series (equal to Juniper M20, M40 series). A new GGSN product J20 has been developed on the Juniper router platform. Either the AXI 520 or AXI 580 (equal to Juniper M20, M40, M160) can be used for the core routers in the backbone. The partnership between Ericsson and Juniper combines unique competence in mobile systems with that of building carrier-class routers.

Conclusion

Mobility and the Internet, the two most dynamic forces in communications today, converge in the design and implementation of the Ericsson mobile core network. Support for new services and a common transport technology are the main drivers for the integration of IP technology into the core network.

Ericsson is fully committed to the introduction of

- IP-based technologies in its products; and
- the mobile core network solutions that address mobility and the Internet.

Ericsson's flexible core network architecture allows operators to address mobility and the Internet independently.

The introduction of GPRS is the first step toward supporting IP-based applications. Support for new IP multimedia services is now being implemented according to the ongoing 3GPP standardization work outlined in this article.

New IP-based connectivity solutions are being introduced step by step with the objective of decreasing transport costs. The basis of these solutions are IP-based multiservice networks based on carrier-class routers. Other IP technologies, such as VPN, resilience, quality of service, and security mechanisms, are needed to provide the characteristics we associate with carrier networks.

The new IP-based multiservice network also opens up the way for much cheaper transmission techniques in the long-haul network.

REFERENCES

1 IP Multimedia Subsystem Stage 2, 3GPP technical specification 3G TS 23.228
2 Peck, M., Granström, P and Olson, S.: The future of communication using SIP, Ericsson Review Vol. 80(2002):1, pp. 30-37.
3 Ekeroth, L. and Hedström, P-M.:GPRS Support Nodes, Ericsson Review Vol. 77(2000):3, pp. 156-169
4 Reinius, J.: Cello—An ATM transport and control platform, Ericsson Review Vol. 76(1999):2, pp. 48-55
5 Session Description Protocol, RFC 2327
6 Session Initiation Protocol, RFC 2543