The Evolution and Future of Mobile Communication Systems

Written by David G Ainscough

Chapter 5 : UMTS (Universal Mobile Telecommunication System)

This page is intentionally left blank.

5.1 UMTS Introduction

This chapter provides an overview of the GSM/UMTS core network, as specified by the 3GPP release 99 standards. There is still no strict formal definition of the core network, but the position of the core network for the purpose of this chapter and how it relates to other parts of the PLMN (Public Land Mobile Network) and co-operating networks.

Originally decided by the ETSI SMG, the 3GPP release 99, the UMTS core network standard is evolved from the earlier GSM standards releases. The core network forms the central part of a PLMN and accommodates important functions for intra and inter PLMN roaming, the provision of IP-connectivity and internet access, ISDN services and interworking with other networks.

The core network may be looked upon as consisting of two different parts :

Managing Circuit Switched mode communication services. (ISDN type services)

Managing Packet Switched mode communication services. (GPRS type services)

Although logically very different the two parts share certain core network functionality, such as the HLR, Auc, EIR etc, and may also run over a common transport infrastructure.

5.2 Horizontal Layering

Today's modern telecommunications and data communications environment consists of a variety of networks. Most of these networks are highly specialized and optimized to serve a specific purpose such as PLMN, ISDN/PSDT and Data/IP Network. To a larger extent these networks are all vertically integrated in a sense they combine functionality in one and the same network element.

Due to the fact that most networks are vertical in structure, they have evolved separately and therefore differ in many respects. However, this has the effect of limiting the network operators ability to create synergies among their networks.

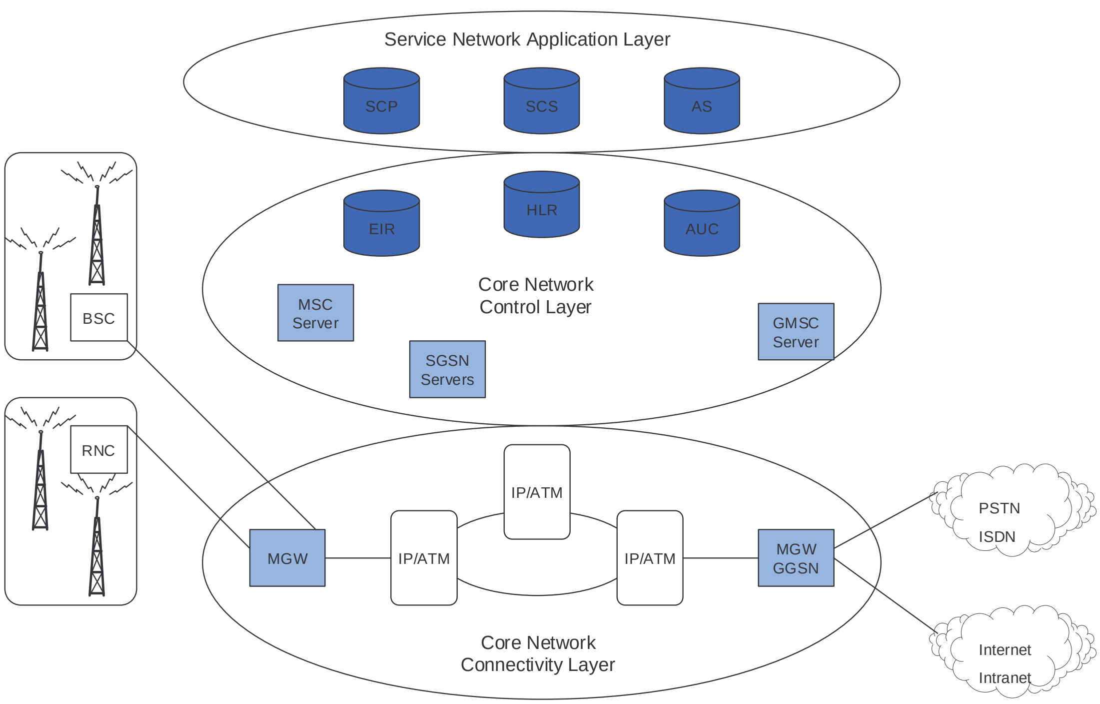

Figure 1 Three Layered UMTS Core Network Architecture

(Adapted Edlin, B 2001)

The architecture of the core network for UMTS is shown is Figure 35, this illustrates the way to evolve this multi-network situation in order to overcome some of the existing defects. In short the UMTS solution is based on a horizontalisation of the core network into a number of independent networks. Basically UMTS is a “*Network of Networks*” (Personal Comment). Special Media Gateways (MGW), controlled by specific network servers, adapt and connect different access types to a common backbone network. End user applications are provided by a specific service layer, common to different access types.

5.3 General Principles

The layered core network architecture is derived from the current standards reference model by separating the control plane functions, thus turning these nodes into Servers and Media Gateways.

5.3.1 User Plane

The user plane, sometimes referred to as the connectivity layer (see Figure 35), could be seen as a layer of distributed resources managing user data (and signaling) flows. The user plane’s functions are primarily handled by the GGSNs and the MGWs, located at the edges of the core network. The MGW carries out the processing of end user data such as speech coding and also acts as an access switch/router to the backbone network. The MGW also sets up the bearer connections carrying the user data flows in the user plane.

The MGWs are controlled by the MSC and SGSN servers via the H.248 gateway control protocol. The resources needed for a call/session may be distributed over multiple MGWs. For Circuit mode communications the user data processing is primarily allocated to the MGW on the boarder of the ISDN/PSTN.

For interconnecting the network elements in the user plane, different transport technologies may be used.

5.3.2 Control Plane

The control plane houses a number of network servers and databases of different types (MSC Server, SGSN Server, HLR, Auc, EIR etc). These servers are responsible for handling subscriber data, security, mobility management, setup and release of calls and sessions, requested by the end users, circuit mode supplementary services and similar functions. The network servers communicate among themselves and with other network elements by means of standard layer 3 protocols such as the following, Iu “Mobility Management”, Iu “Call/Session control”, ISUP, MAP etc. The MSC and the SGSN servers can also determine which MGW functions and resources are required by the call/session and control these functions and resources in the MGW by means of the H.248 gateway protocol.

5.3.3 Application Layer

The application layer is a simplified abstraction of the layer where most of the end-user applications reside. To a large extent these applications are implemented partly in the terminals and partly in the specific application servers, which reside within the network*. Applications exist in a broad variety, from the simple end-to-end client server solutions to complex applications involving multiple interactive end-systems, networks, media communication models etc.

* Some services handled by the network elements in the control plane, particularly the more traditional supplementary services for circuit mode communication.

5.4 Core Network Standards and Interfaces

(This section refers to the Ericsson UMTS Core Model)

The Ericsson core network architecture fully aligns with the 3GPP release 99 standards and complies with the interfaces worked out by the 3GPP. The separation of the control and user plane functions can be seen as an enhancement to the release 99 reference model and aligns with a number of important standardization initiatives in IETF, ETSI, IT-U and other different industry forums.

As a result of the various control plane/user plane separation initiatives in different standardization forums, a gateway control protocol (H.248) has now been standardized by ITU-T.

5.5 Key Benefits of the Layered Architecture

The key principle of modern networking is the separation of network functionality into independent layers; this is also true in the tele and datacoms industry. The 'layered' thinking is also a very fundamental and visible aspect in a number of standardization initiatives and industry forums, such as Multiservice Switching Forum, led by several of the largest operators and manufactures.

3G mobile communication systems are designed with these principles in mind, and hence offer a number of advantages and possibilities. Special attention needs to be put on meeting the requirements for time-to-market, low cost of ownership, openness and future evolution towards an "all-IP" solution.

Key characteristics :

Offers an open and versatile architecture capable of meeting the current and future demand in a fast changing telecommunications enviroment.

Provides and inherent flexibility for coping with growth and/or changing traffic patterns and traffic mixes.

Independence between layers allows each layer to evolve independently.

Provides great transport flexibility and allows different transport technologies, both existing and new. To be deployed without impacting the control or services/application layers.

Allows common transport arrangements for multi-services network.

Allows access independent and seamless services through a common service/applications layer.

Provides efficient use of network resources.

Relies on proven and stable protocols and design.

Allow a very flexible re-use of investments in the GSM infrastructure.

5.6 Core UMTS Network Elements

The following section sets out the elements of the GSM/UMTS core network, and some of the reasoning behind it. Figure 36 shows how these core elements fit together and how they fit into the overall system picture. However Figure 36 is not a true physical representation of the core network, nor does it show all the interfaces. As the core network has not been standardized there still exist other possibilities for bundling different functions into physical network elements.

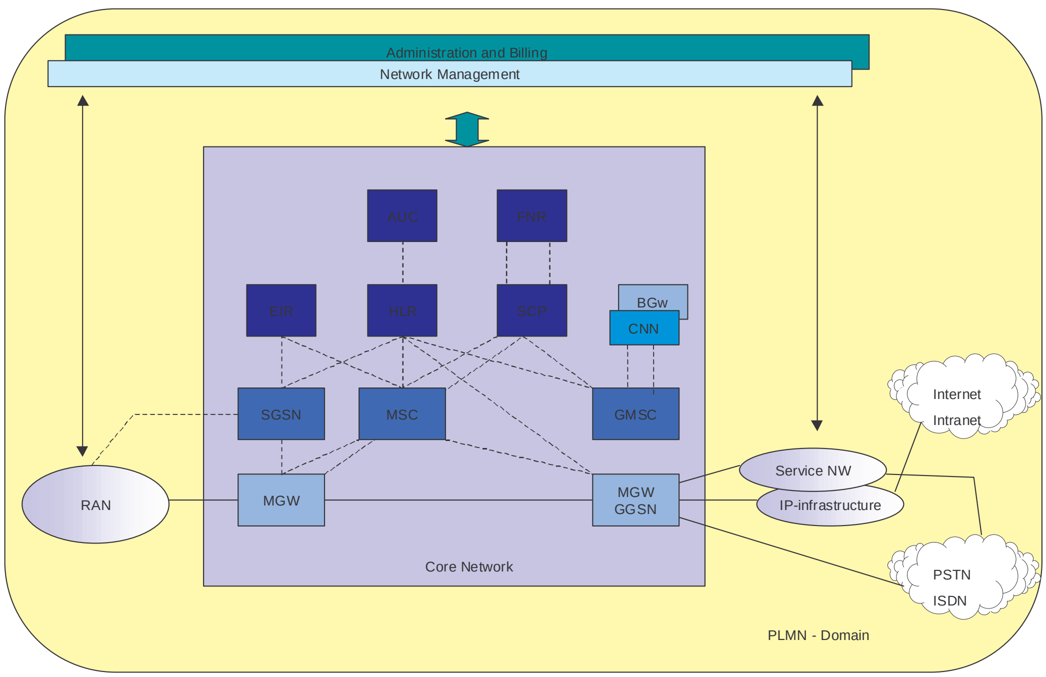

Figure 2 UMTS Core Network Elements

(Adapted from Ericsson Document EN/LZT 123 5374 R1B)

MSC Server

The MSC Server handles control layer functions related to circuit mode communication services at the UTRAN and PSTN/ISDN borders and performs among others the following functions:

- Media gateway control
- ISDN services control
- Mobility management
- Authentication
- Charging data collection/output
- Services switching function (5SF)
- Internet dial-in services (RAS)
- Element management

In addition to these functions the MSC Server also houses the interworking and gateway functionality necessary to act as an SMS-IWMSC and SMS-GMSC for the Short Message Service.

Serving GPRS Support Node Server (SGSN Server)

The SGSN Server handles control layer functions related to packet mode communication services at the border between UTRAN and the basic core network and performs among others the following functions:

- Media gateway control
- Session management
- Mobility management
- Authentication
- Charging control
- Relaying of SMS
- Element management

Media Gateway (MGW)

The MGW handles transport and user plane functions for both packet and circuit mode communication at the borders between networks/network segments. These lower layer functions primarily concern the user data handling and includes e.g.:

- Media processing (speech coding. conference call bridging etc)
- Media generation (tones etc)
- Setup/release of user data bearers
- Provision of traffic/charging info for packet mode communication
- Security management
- Routing and switching QoS management
- Element management

Most MGW resources are shared between packet and circuit communication services or can easily be re-configured from one communication mode to the other. This offers a very cost efficient and flexible solution for managing future changes in the circuit and packet mode traffic balance.

Gateway GPRS Support Node (GGSN)

The GGSN constitutes the tunnel end-point for the GPRS specific GTP-tunnel for packet mode communication and is situated on the border between the basic UMTS core network and the ISP-POP". The GGSN is (indirectly) selected by the end-user at setup of the PDP-context. From an addressing point of view the GGSN represents the point of presence for 'logged on' end-users, i.e. end-users with an established PDP-context. Addresses can be dynamically assigned (fetched from an external server or a pool of own addresses) or statically assigned (fetched from the HLR). In order to fulfill its role in the network the GGSN performs the following functions among others:

- Tunnel management
- IP-address management
- Charging data collection/output
- Security management
- Packet filtering
- Packet routing/tunneling
- QoS management
- Element management

Optionally, a GGSN may also include certain non-GPRS specific functions such as a Foreign Agent (FA).

Home Location Register (HLR)

The Home Location Register is a network database for mobile telecommunications. The HLR holds all mobile specific subscriber data and contains a number of functions for managing these data, controlling services and enabling subscribers to access and receive their services when roaming within and outside their home PLMN. The HLR communicates with the GSNs, MSCs and other network element via the MAP-protocol.

The Authentication Centre (AUC) contains functions for secure storage of individual subscriber identifiers and keys. AUC also includes algorithms necessary for generating authentication and ciphering data based on the subscriber keys. The authentication and ciphering data, provided by the AUC upon request, are used by different network elements to protect the network, users and operators against unauthorized use of the system.

Service Control Point (SCP)

The SCP is a part of the IN concept and contains the service logic and its execution environment. The SCP works in close co-operation with the service switching functions in the MSCs and provides IN-services such as Virtual Private Network (VPN), Number Portability etc.

Flexible Number Register (FNR)

The FNR accommodates translation functions necessary to de-couple end-user identities from the actual network databases holding end-user data. These translation functions are essential in order to be able to e.g.:

Easily reconfigure some of the network databases, e.g. as a result of a growing number of end-users.

Allow end-users to keep their identities when changing from one administration to another.

Equipment Identity Register (EIR)

The Equipment Identity Register is a network database holding status information on mobile station equipment. The EIR is interrogated by means of the MAP-protocol from e.g. network elements providing access into the PLMN (MSC and SGSN servers), in order to ensure that the mobile station equipment is not black-listed for any reason.

Cost Control Node (CCN)

The Cost Control Node is a central part of the 3G charging environment. CCN contains centralized rating functions and constitutes the charging determination point for services provided to subscribers at both into and inter PLMN roaming. CCN also handles credit limits for accumulated subscriber charges. These on-line charging mechanism are used for provision of features such as credit control (e.g. pre-paid), user cost information (advice of charge), fraud control etc.

CCN receives traffic information on-line and off-line from the different traffic handling/service nodes and instructs these nodes to proceed with the calls/services according to the outcome of the operator defined cost control analyses. CCN also has the ability to receive CDR's deriving from the Serving network providing near-real-time revenue charging.

Billing Gateway (BGw)

The Billing Gateway is a key component for the off-line (CDR based) charging in

UMTS. The BGw collects information from the different traffic handling/service nodes (GSN, MSC. SMS-C. voice mail, application servers. etc) and forwards it to the operator's administrative systems for off-line billing, accounting. traffic analyses and similar functions. The BGw also acts as an intermediate storage and pre-processor for formatting the CDR-information into the specific formats used by the operator.

The BGw also provides an interface to CCN and can be used to relay information for online charging from the traffic handling/service nodes not having direct interfaces to CCN.

BGw also acts as an accounting/settlement broker by providing the ability to distribute CDR’s from the Serving network to the Home environment BGw or CCN in near real time.

5.8 Routers/Switches

For routing and switching of user data and signalling flows in the core network connectivity layer, standard switches and routers are used. Most infrastructures offer a wide range of routers and switches with different functions and performance characteristics for most transport needs in all segments of the network (access, aggregation, core) when undertaking total or turnkey contracts.

5.9 Traffic Handling in a Layered Architecture

In order to facilitate the understanding of call/session handling in a Server/Media Gateway architecture, with separated control and user plane functions there are four traffic cases in the following sections. Please note that these traffic cases have been simplified.

5.9.1 Traffic Cases

5.9.1.1 Mobile Originated Call

An AAL2-switche core network has been assumed.

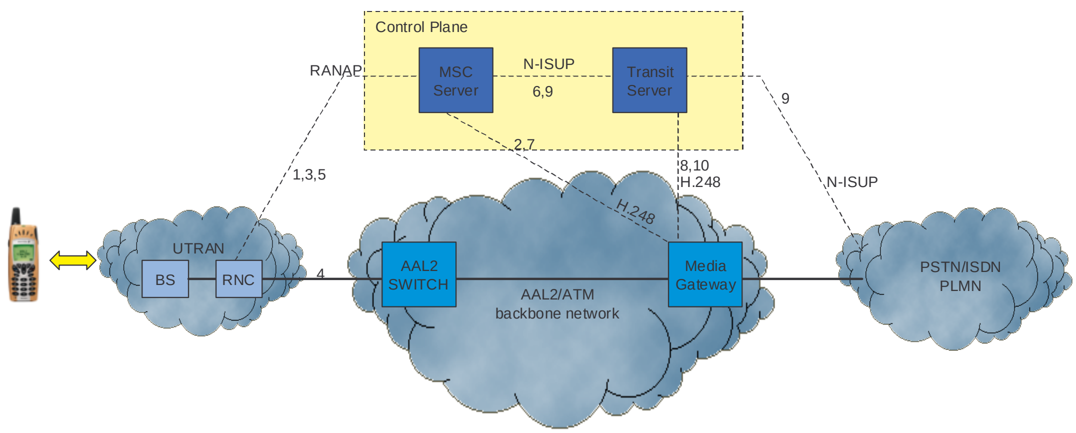

Figure 3 Mobile Originated Call (Traffic Case)

1. A request for call setup is received from the UE. Selection of an applicable MGW
2. Resources controlled by the MSC-server are reserved in the selected MGW
3. The MSC-server orders the RNC to setup a radio access bearer
4. The User plane is setup from the RNC to the MGW
5. The MSC server is informed that the user plane setup to the MGW has been completed

6 The call setup to the transit server is initiated by the MSC

7 Backward through connection of the equipment controlled by the MSC server is ordered in the MGW

8 Resources controlled by the transit server are requested from the MGW

9 Call setup to the selected PSTN switch is initiated by the transit server

10 Through connection of the transit server controlled path in the MGW is ordered

5.9.1.2 Mobile Terminated Call

A simplified procedure for a mobile terminated call

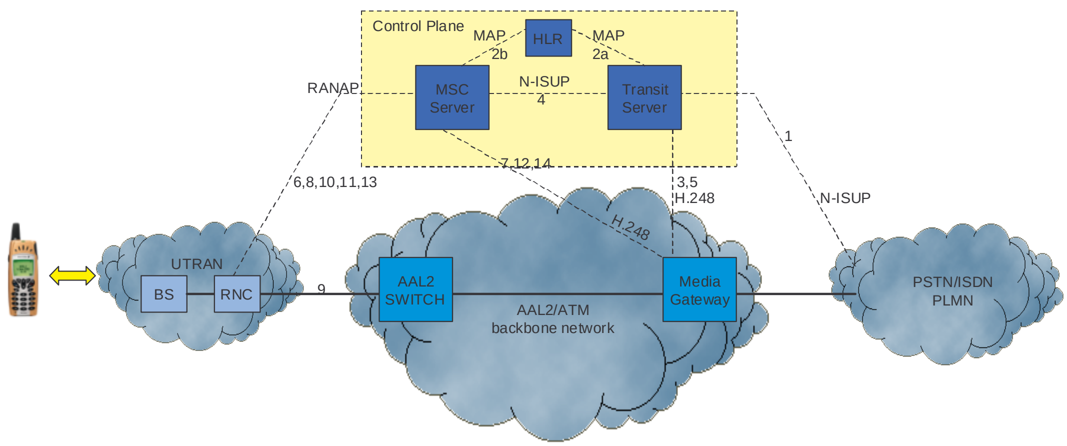

Figure 4 Mobile Terminated Call (Traffic Case)

1 An incoming call is received in the GMSC server

2 Routing data is obtained from the MSC server (VLR) via the HLR

3 Resources controlled by the GMSC server are reserved in the MGW

4 The call setup towards the visited MSC server is initiated by the

GMSC server

5 Through connection of the MGW resources is ordered by the GSMC server

6 Paging of the called party is ordered by the MSC server, the call setup is confirmed by the UE.

7 Resources controlled by the MSC server are reserved in the MGW

8 The MSC server orders the RNC to setup a radio access bearer

9 The user plane is setup from the RNC to the MGW

10 The RNC informs the MSC server that the user pane setup to the MGW has been completed

11 The MSC server is informed by the UE that altering of the called party has started

12 The MSC server orders the MGW to start sending ringing tone towards the calling party

13 Answer is received from the called party

14 Through connection of the MGW is ordered by the MSC server

5.9.1.3 PDP-context Activation

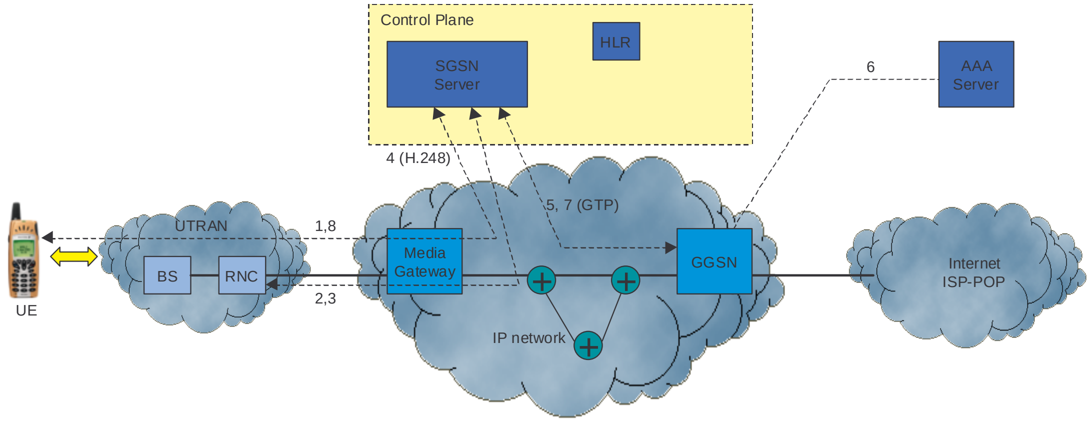

Figure 5 PDP-context Activation (Traffic Case)

1 An "active PDP context request" is sent by the UE to the SGSN server

2 The SGSN server requests the UTRAN to setup a radio access bearer (RAB)

3 The request for a RAB is acknowledged by UTRAN

4 The SGSN server requests the MGW to create a GTP tunnel transfer point

5 A "create PDP context request" is sent from the SGSN server to the GGSN

6 The GGSN server requests an IP address fro the AAA server (or similar)

7 GGSN sends a "create PDP context response" back to the SGSN server

8 An "active PDP context response" is sent back to the UE from the SGSN server

5.9.1.4 Packet Forwarding

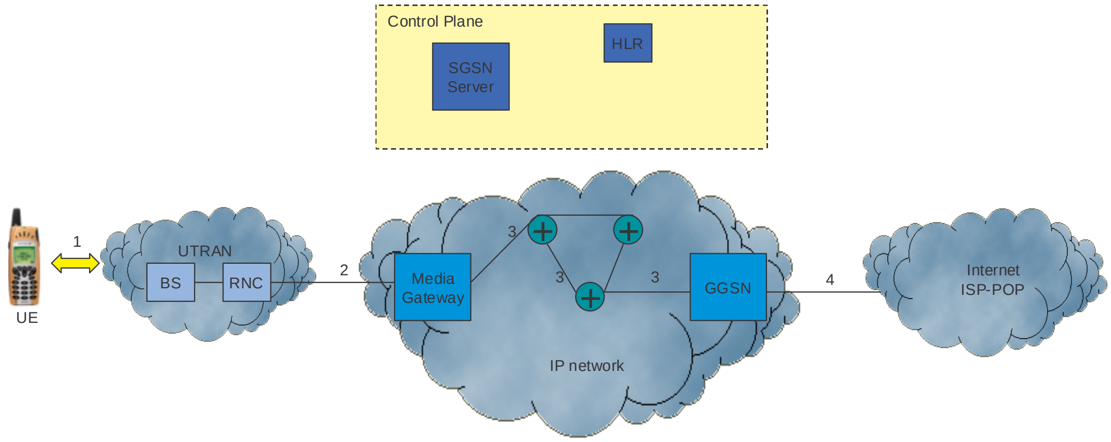

Figure 6Packet Forwarding (Traffic Case)

1 A PDU is sent from the UE

2 The PDU is forwarded over the GTP-U tunnel between the RNC and the MGW

3 The PDU is forwarded over the core network GTP-U tunnel to the GGSN

4 The PDU is forwarded to the Internet/ISP-POP

Chapter Summary and Key points

As technology stands at this time UMTS is the best solution for modern mobile communications. This chapter on UMTS is mainly based on the Ericsson view of what UMTS will be, this is down to the fact that there have been no standards agreed.

One of the main advantages of UMTS is that it has a layered architecture, this is different to other existing systems for mobile communications which are vertical is structure.

The great advantage of UMTS is that it uses the existing systems for mobile communication, UMTS is not necessarily a speaking a new technology, but rather the application of existing technology. Eventually the UMTS system will be able to detect what service the user requires be it Voice or Data communication.

The mobile phone industry is already starting to work on new phones which combine the functionality of a personal computer and the size of a mobile phone.