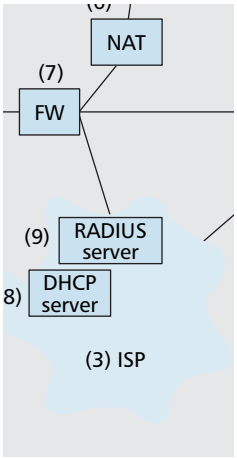
ACCEPTED FROM OPEN CALL

# IP CONNECTIVITY FOR GATEWAY GPRS SUPPORT NODE

YUAN-KAI CHEN, NATIONAL CHIAO TUNG UNIVERSITY
YI-BING LIN, PROVIDENCE UNIVERSITY




The authors focus on the GGSN functions for IP connection including the Access Point Name processing, IP address allocation, tunneling technologies and Quality of Service management.

## ABSTRACT

In the Universal Mobile Telecommunications System, the gateway GPRS support node (GGSN) provides IP connection between the mobile telecommunications network and external packet data networks (e.g., the Internet). Specifically, the GGSN exercises session management to transfer user packets between mobile stations and external data networks. In this article we focus on the GGSN functions for IP connection including access point name processing, IP address allocation, tunneling technologies, and QoS management. Based on our experience as a mobile operator, we give several examples to show how these functions can actually be implemented in a commercial mobile network.

## INTRODUCTION

The Universal Mobile Telecommunications System (UMTS) is a third-generation mobile network evolved from GSM and General Packet Radio Service (GPRS) [1, 2]. UMTS enables high-speed packet switched data (up to 2 Mb/s) with quality of service (QoS) to access versatile multimedia services anytime and anywhere. Figure 1 shows the UMTS (Release 99) architecture [1]. In this figure the dashed lines represent signaling links, and the solid lines represent data and signaling links. The *core network* consists of two service domains, the circuit-switched (CS) and packet-switched (PS) service domains. In the CS service domain, UMTS connects to the public switched telephone network (PSTN; Fig. 1a) through the mobile switching center (MSC; Fig. 1b). In the PS service domain, UMTS connects to the external packet data network (PDN; Fig. 1c) through the serving GPRS support node (SGSN; Fig. 1d) and gateway GPRS support node (GGSN; Fig. 1e). The SGSN in the PS domain plays a similar role as the MSC in the CS domain. The GGSN provides interworking to the external PDN, and is connected with SGSNs via an IP-based GPRS backbone network. The UMTS Terrestrial Radio Access Network (UTRAN) consists of *Node Bs* (the UMTS term for base stations; Fig. 1f) and radio network controllers (RNCs; Fig. 1g) connected by an asynchronous transfer mode (ATM) network. A mobile station (MS) or *user equipment* communicates with one or more Node Bs through radio interface *Uu* based on wideband code-division multiple access (WCDMA) radio technology [3].

General design guidelines for a GPRS core network have been intensively studied [1, 4, references therein]. However, IP connectivity offered by the GGSN is seldom elaborated on in depth in the literature. This article focuses on GGSN functionality. Several interfaces based on IP are defined for the GGSN. The Gn interface between the SGSN and GGSN uses the GPRS Tunneling Protocol (GTP) to transport user data and control signaling [5]. The GGSN connects to the PDN through the Gi interface. The Ga interface between the GGSN and charging gateway (CG; Fig. 1j) uses the GTP protocol to transfer call detail records (CDRs). The GGSN also provides an interface to the network management system (NMS; Fig. 1k) using protocols such as Simple Network Management Protocol (SNMP).

Before an MS can access any mobile data service, the packet data protocol (PDP) context for the service must be activated, which specifies the application-layer PDP and routing information for the communication session. The PDP context is maintained in the MS, SGSN, and GGSN, which will be elaborated on later.

This article concentrates on the design of the GGSN functionalities for IP connection, including IP address allocation, IP packet routing, and transfer. We describe general GGSN functionalities and elaborate on access point name (APN) and IP address allocation. We show the interworking techniques between the GGSN and the external PDN, and address GGSN QoS issues. In these sections we describe how the GGSN functions are implemented based on our experience as a mobile operator (i.e., Chunghwa Telecom.). For readers who are not familiar with GPRS/UMTS, background information is available in the literature. For example, general GPRS/UMTS network architecture was described in [4]. In [6] the GPRS/UMTS network was introduced from the viewpoint of wireless Internet access. The functionality


1536-1284/05/$20.00 © 2005 IEEE

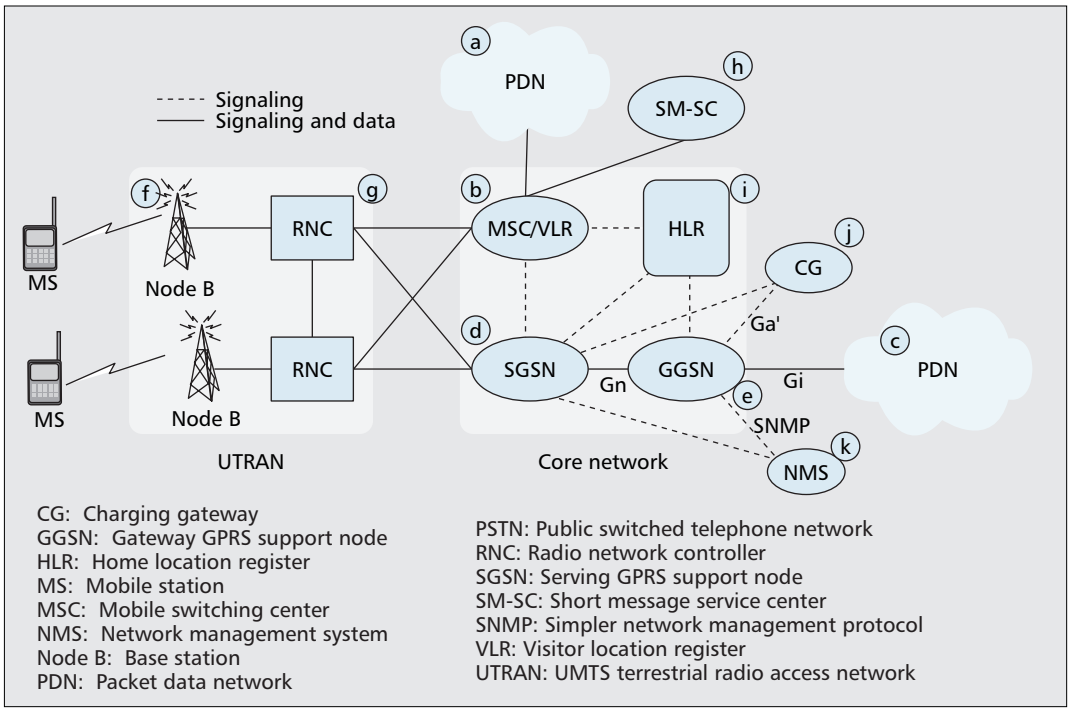



**Figure 1.** *The UMTS network architecture.*

partitioning of GPRS applications and a possible implementation on a CompactPCI-based platform were presented in [7].

## GGSN Functionalities

The GGSN plays the role of a gateway, which controls user data sessions and transfers the data packets between the UMTS network and the external PDN. Table 1 shows the four meta functions defined in UMTS [8]. The functions implemented in the GGSN are described as follows:

**Network access control** (item 1 in Table 1) enables a user to access UMTS services, which include the following GGSN functions:

- Admission control (item 1.3 in Table 1): The GGSN admits the incoming service requests and assigns the IP address for the MS depending on requested APN and its available resources (in terms of bandwidth and QoS). The APN and IP address are stated in later sections. We also elaborate on the GGSN QoS mechanism. Additionally, the GGSN coordinates with the SGSN and UTRAN to determine the end-to-end resources allocated for a user.
- Message screening (item 1.4 in Table 1): This function filters out unauthorized and unsolicited messages. In Third-Generation Partnership Project (3GPP) TS 23.060, message screening is defined in the GGSN. In most GPRS products, however, this function is implemented in a firewall server. Feature interactions between the message screening function and new services such as voice over IP (VoIP) and video streaming may cause unexpected results. The traffic of these services use specific UDP ports and thus might be filtered out when enabling the message screening function at the GGSN. Therefore, the message screening function must be carefully configured based on the mobile operation environments.
- Charging data collection (item 1.6 in Table 1): The GGSN generates CDRs when a predefined event (e.g., PDP context deactivation) occurs. The CDRs are included in the charging packets, which are sent to the corresponding charging gateway (Fig. 1j).

Several UMTS network access control functions are not performed by the GGSN. For example, the registration function (item 1.1 in

| Function | MS | UTRAN | SGSN | GGSN | HLR |
|---|---|---|---|---|---|
| **1. Network access control** | | | | | |
| 1.1 Registration | — | — | — | — | v |
| 1.2 Authentication and authorization | v | — | v | — | v |
| 1.3 Admission control | v | v | v | v | — |
| 1.4 Message screening | — | — | — | v | — |
| 1.5 Packet terminal adaptation | v | — | — | — | — |
| 1.6 Charging data collection | — | — | v | v | — |
| **2. Packet routing and transfer** | | | | | |
| 2.1 Relay | v | v | v | v | — |
| 2.2 Routing | v | v | v | v | — |
| 2.3 Address translation and mapping | v | v | v | v | — |
| 2.4 Encapsulation | v | v | v | v | — |
| 2.5 Tunneling | — | v | v | v | — |
| 2.6 Compression | v | v | — | — | — |
| 2.7 Ciphering | v | v | — | — | v |
| **3. Mobility management** | v | — | v | v | v |
| **4. Radio resource management** | v | v | — | — | — |

**Table 1.** *Functions of UMTS network elements.*

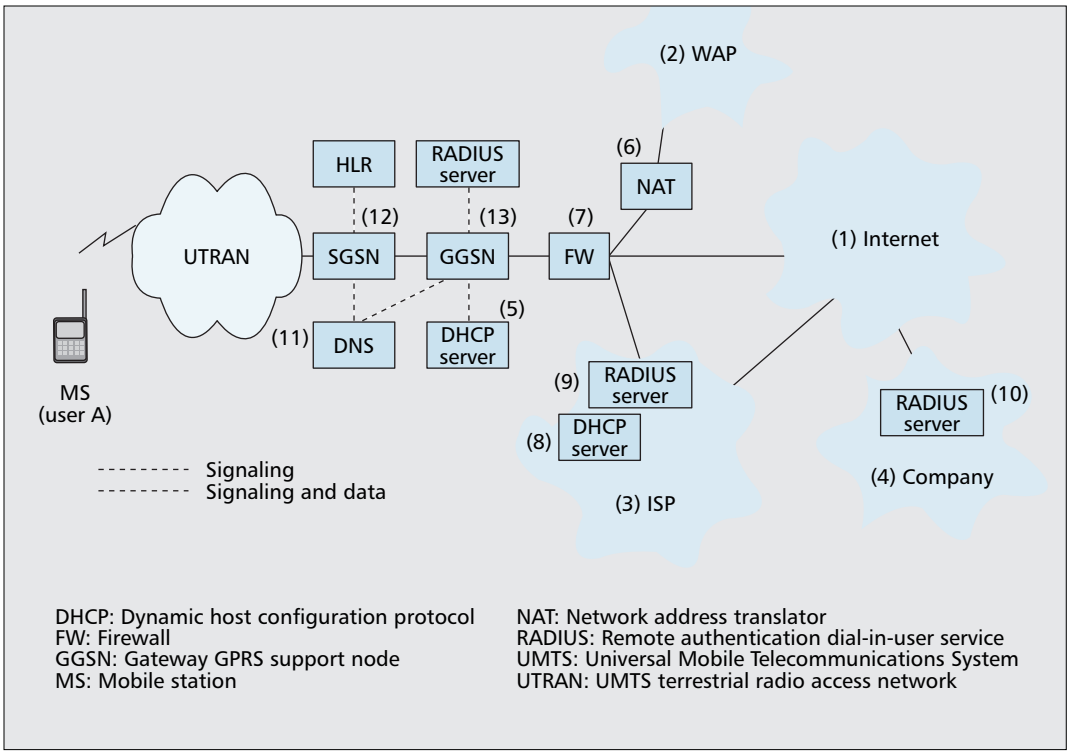



**Figure 2.** *APN configuration examples.*

Packet routing and transfer determines the route to transfer the user data packets within and between the UMTS networks. Within the UMTS backbone network, packets are delivered between the GGSN and the SGSN by the GPRS Tunneling Protocol in the Gn interface.

Table 1) associates a user with his/her subscription and routing information stored in the home location register (HLR; Fig. 1i). The authentication and authorization functions (item 1.2 in Table 1) are exercised by the SGSN and MS to authenticate the user and authorize the requested services. The packet terminal adaptation function (Item 1.5 in Table 1) is performed by the MS to adapt a user's data packets to be delivered across the UMTS network.

**Packet routing and transfer** (item 2 in Table 1) determines the route to transfer user data packets within and between UMTS networks. Within the UMTS backbone network, packets are delivered between the GGSN and SGSN by GTP in the Gn interface. The following functions are defined in the GGSN:

- Relay (item 2.1 in Table 1): The GGSN relays the data packets between the SGSN (GTP packets) and the external PDN (IP packets). The GGSN exercises the tunneling technique to transfer the data packets between the UMTS and the external PDN. Details of the tunneling techniques are described later.
- Routing (item 2.2 in Table 1): The GGSN utilizes the APN to route packets to the external PDN. Details of the APN are elaborated on in the next section.
- Address translation and mapping (item 2.3 in Table 1): The address translation function converts an address from one address type (e.g., GTP tunnel endpoint identifier, TEID, or IP address) to another. The address mapping function translates a network address to another network address of the same type. These functions are performed by the GGSN, SGSN, and UTRAN.
- Encapsulation (item 2.4 in Table 1): The GGSN encapsulates the incoming data packets from an external PDN into GTP packets by adding the GTP headers (including the SGSN address, sequence number, etc.) to user data packets. The GGSN decapsulates GTP packets by removing the GTP headers to reveal the original data packets, and relays them to the external PDN.
- Tunneling (item 2.5 in Table 1): Through GTP, this function transfers the encapsulated data packets within and between the UMTS networks. GTP tunneling is performed by the GGSN, SGSN, and UTRAN. GTP related issues are not discussed in this article, but can be found in [5].

Other packet routing and transfer functions such as compression (item 2.6 in Table 1) and ciphering (item 2.7 in Table 1) are only performed by the RNC and MS.

- Mobility management (item 3 in Table 1) keeps track of the current location of an MS. The GGSN is involved in this function only when there are data sessions (PDP contexts) between the MS and the GGSN. When the MS moves from one SGSN to another, the GGSN updates the SGSN address and TEID in the PDP context (i.e., the GTP link is switched from the old SGSN to the new SGSN).
- Radio resource management (item 4 in Table 1) does not involve the GGSN. This function is concerned with the allocation and maintenance of radio links, performed by the MS and UTRAN.

In the remainder of this article we focus on the packet routing and transfer meta function for the GGSN. Specifically, we describe how an IP connection is established in the GGSN. Other meta functions are outside the scope of this article, and the reader is referred to 3GPP TS 23.060 [8] for details.

In either the transparent or the non-transparent access modes, the GGSN may negotiate with a DHCP server to allocate an IP address from the address pool maintained by this DHCP server. Alternatively, the IP address of an MS may be assigned by a RADIUS server.

## APN Allocation

An APN is used in UMTS/GPRS as a reference point to an external PDN that supports the services to be accessed by an MS. The APN information is permanently distributed and maintained in the HLR, GGSN, and domain name server (DNS). A set of APN labels is defined in the HLR. Each mobile user can subscribe to one or more APNs from this set. The labels of these subscribed APNs are then stored in the MS at subscription time. Among the subscribed APNs, there is one default APN. If a user attempts to access a service without specifying the APN, the default APN is used. Additionally, the HLR may also define a wild card APN "*" which allows an MS to access any unsubscribed APNs. For each APN, the DNS keeps an IP address list of the GGSNs associated with this APN label. Figure 2 illustrates an architecture for various APN configurations. In this architecture the DNS (Fig. 2, 11) is connected to the SGSNs (Fig. 2, 12) and GGSNs (Fig. 2, 13). When an SGSN performs an APN query, the DNS will convert the APN label to the IP address of a GGSN in this list. In existing mobile networks, all SGSNs are fully connected to all GGSNs. During a GPRS session, when a user moves from an old SGSN to a new SGSN, the new SGSN connects with the original GGSN, and the same IP connectivity is maintained. Note that the Mobile IP mechanism is applied to provide IP-level mobility when the user changes GGSN.

Every GGSN maintains the configurations for a subset of the APNs defined in the HLR. Each configuration includes the IP address allocation method (static or dynamic), IP address type (IPv4 or IPv6), and destination external PDN that offers the service through this APN.

Consider the example in Fig. 2, where the HLR defines four APNs. The default APN is **INTERNET**. The mobile operator uses this APN to provide Internet services (Fig. 2, 1). The **WAP** APN is used to provide WAP access services (Fig. 2, 2). The GGSN may transfer the MS's MS ISDN number (MSISDN, i.e., the mobile telephone number) to the corresponding WAP content server for accounting purposes. With this GGSN feature, the WAP content server can provide customized personal services based on the received MSISDN. The **ISP** APN provides Internet services offered by an Internet service provider (ISP) other than the mobile operator (Fig. 2, 3). This APN typically is not offered in a UMTS network due to business considerations. We include this APN in Fig. 2 for demonstration purposes. The **COMPANY** APN provides mobile office services (Fig. 2, 4), which allows a corporate user to access the intranet services of his/her company through the UMTS network.

Suppose a user subscribes to the **INTERNET** and the **WAP** APNs in Fig. 2. If the user requests a service through the **COMPANY** APN, the SGSN will reject the request. On the other hand, if the HLR defines a wild card APN for the user, the request will be accepted. Clearly, the use of a wild card APN may cause security problems. For example, with a wild card an unauthorized user can access the intranet of a corporation through the **COMPANY** APN, and retrieve the data or attack the database server of the company.

## IP Address Allocation

Based on the APN setting specified in 3GPP TS 29.060 [5], there are two access modes for IP address allocation in the GGSN: *transparent* and *nontransparent*. In transparent access mode, the mobile operator acts as an ISP, and an MS is given an IP address from the operator's IP address space. The IP address can be allocated statically at subscription time or dynamically at activation of the PDP context. In Fig. 2 transparent access mode is exercised if the requested APN is **INTERNET**.

In nontransparent access mode, the mobile operator only provides a user with the access channel of an ISP (if the requested APN is **ISP** in Fig. 2) or a company (if the requested APN is **COMPANY** in Fig. 2). The IP address pool is owned by the ISP or corporation, and the IP address for an MS is dynamically allocated. Therefore, in this access mode the UMTS operator only serves as the access service provider.

The IP addresses can be allocated by either the GGSN, or a Dynamic Host Configuration Protocol (DHCP) or RADIUS server. In transparent access mode the GGSN may allocate the IP address for a user using its own address pool. This address pool is maintained by the UMTS operator. In the current implementation, IPv6 addresses can only be allocated by this alternative.

In either transparent or nontransparent access mode, the GGSN may negotiate with a DHCP server to allocate an IP address from the address pool maintained by this DHCP server. Alternatively, the IP address of an MS may be assigned by a RADIUS server, where the IP address pool is maintained by this RADIUS server.

We use the example in Fig. 2 to illustrate the setup of four APN configurations. Consider a mobile network with 1,000,000 subscribers. Based on our experience, 20 percent of users are expected to simultaneously access mobile data services. The average number of PDP contexts activated by each active user is typically dimensioned to two. Each PDP context is assigned an IP address. One PDP context is used for WAP service, where the MS keeps the allocated IP address to browse the WAP service. Another PDP context is used for wireless Internet service, where the notebook/PDA connecting to the MS keeps another allocated IP address to access the services provided by the APNs **INTERNET**, **ISP**, and **COMPANY**. Thus, the dimensioned network capacity is 200,000 simultaneous active users and 400,000 PDP contexts activated by these users. Note that a typical GGSN is designed to support 300,000 simultaneous users. In a mobile network with 1,000,000 users, 200,000 active users are expected. Therefore, two GGSNs are expected to be deployed in the network ( one for operation and one for standby; or both operate in the manner of load sharing). The configurations of these APNs are listed in Table 2. In our example, the IP addresses for the **INTERNET** APN

| APN label | INTERNET | WAP | ISP | COMPANY |
|---|---|---|---|---|
| GGSN access mode | Transparent | Transparent | Nontransparent | Nontransparent |
| IP address allocator | GGSN | DHCP server | DHCP server | RADIUS server |
| IP address type | IPv6 | IPv4 | IPv4 | IPv4 |
| DHCP server's IP address | — | 192.168.30.1 | 140.113.214.1 | — |
| RADIUS server's IP address | — | — | 140.113.214.2 | 192.168.70.1 |
| Starting MS IP address in the IP address pool | 2002:1234:5678:1111:2222:3333::0001 | 10.144.1.1 | 168.100.1.1 | 10.100.1.1 |
| Maximum number of active PDP contexts | 160,000 | 200,000 | 36,000 | 4000 |

**Table 2.** *Four APN configuration examples.*

The Network Address Translator server will perform address translation between private IP address realm and public IP address realm. In most GGSN products, the NAT function is implemented in the firewall server.

are dynamically allocated from the GGSN's IP address pool. When an MS requests an IP address through the GGSN, no authentication is required. In this scenario, the MS has already been authenticated by the standard UMTS authentication procedure in mobility management [4, 9]. Table 2 assumes that the address type is IPv6. The IP addresses range from 2002:1234:5678:1111:2222:3333:4444:0001 to 2002:1234:5678:1111:2222:3333:4444:fffe, and the maximum number of the supported PDP contexts is 200,000. Note that the maximum numbers of PDP contexts and simultaneous active users are limited by the capacity of the GGSN, which is planned based on user requirements. The size of the IP address pool is determined by the number of simultaneous active users, which is assumed to be 20 percent of total subscribed users in our example.

For the **WAP** APN, the IP addresses are dynamically allocated from a DHCP's IP address pool (Fig. 2, 5). Similar to the **INTERNET** APN, authentication for the MS is performed in UMTS mobility management, and no extra authentication procedure is required when the MS requests the IP address. Due to the shortage of IPv4 address space, each MS is allocated a private IP address. The Network Address Translator (NAT) server (Fig. 2, 6) performs address translation between the private and public IP address realms. In most GGSN products, the NAT function is implemented in the firewall server (Fig. 2, 7). The IP addresses range from 10.144.1.1 to 10.144.15.254, and the maximum number of PDP contexts is 20,000.

For the **ISP** APN, the IP address of a user is allocated from a DHCP server's IP address pool (Fig. 2, 8). Before the allocation, the user is authenticated by a separate RADIUS server (Fig. 2, 9). Both the RADIUS and DHCP servers are owned by the ISP. The MS is allocated a public IP address. The connection between the UMTS network and the external ISP PDN can be established by using either dedicated leased lines or a virtual private network (VPN). VPN interworking is implemented by the tunneling technologies described in the next section. In our example the IP addresses range from 168.100.1.1 to 168.100.15.254, and the maximum number of PDP contexts is 10,000.

For the **COMPANY** APN, the user is authenticated by a RADIUS server (Fig. 2, 10), and is allocated an IP address from the RADIUS IP address pool. The RADIUS server is maintained by the corporation. In the existing implementations, an IP address in the company intranet is allocated from a private IP address pool. Since the IP addresses range from 10.100.1.0 to 10.100.1.254 in Table 2, the maximum number of PDP contexts is 1024.

## PDP Context Activation

For a subscribed service, the corresponding APNs are recorded in a list in the MS at subscription time. When the MS attempts to access this service, it initiates the PDP context activation procedure defined in 3GPP TS23.060 [8]. During this procedure, the MS specifies a requested APN from the APN list. Then the SGSN uses this requested APN to select a GGSN. If the user does not specify any requested APN in the activation procedure, the default APN is chosen by the SGSN. Figure 3 shows the message flow of the PDP context activation procedure. Note that the MS must attach to the UMTS network before it can activate a PDP context [4]. In the attach procedure, the corresponding SGSN obtains the subscriber data of the MS from the HLR, which will be used in Step 3 of the PDP context activation procedure described below.

**Step 1.** The MS specifies the APN in the `Activate PDP Context Request` message and sends it to the SGSN.

**Step 2**. The SGSN negotiates with the UTRAN to allocate the radio bearer bandwidth for the data session.

**Step 3**. The SGSN checks if the requested APN (obtained from the `Activate PDP Context Request` message sent by the MS) is specified in the APN list of the subscription data for the MS. If not, the default APN is used. Then the SGSN creates the PDP context for the user, and sends the requested APN to the DNS server. The DNS server uses this APN to derive the GGSN's IP address.

**Step 4**. Based on the GGSN's IP address obtained from the DNS, the SGSN sends the `Create PDP Context Request` message to

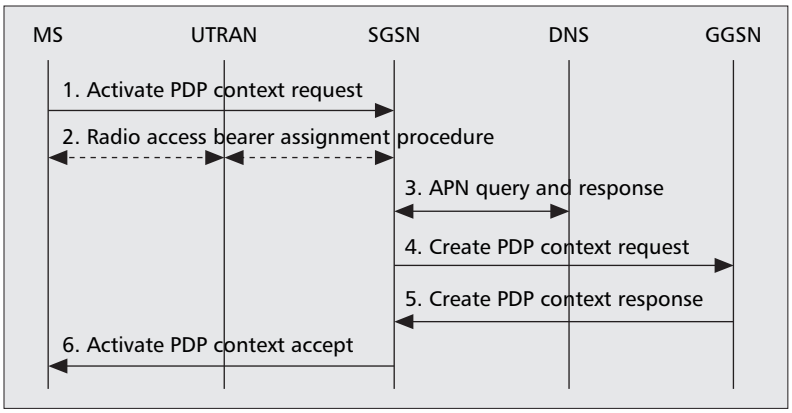



Figure 3. *PDP context activation procedure.*

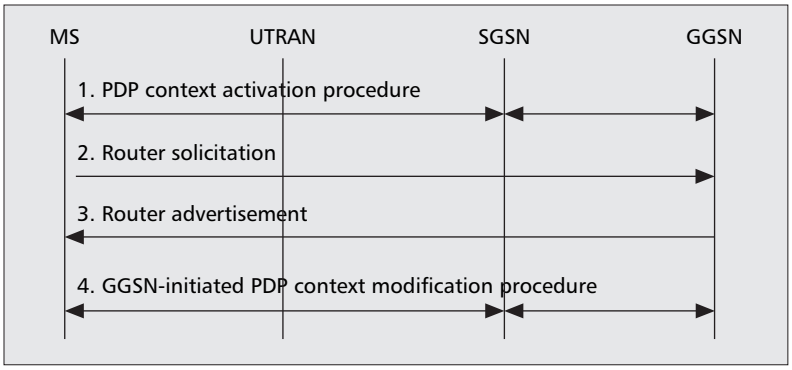



Figure 4. *IPv6 stateless address autoconfiguration procedure.*

the GGSN to establish a GTP tunnel between the SGSN and the GGSN, which will be used as the packet routing path between the GGSN and the MS.

**Step 5**. The GGSN creates a PDP context for the MS. This PDP context records the requested APN, PDP type, MSISDN, IP address, and so on [8]. The GGSN allocates an IP address for the MS using either transparent or nontransparent access mode, and determines the tunneling mechanism to the destination external PDN (described later).

**Step 6**. Finally, the SGSN informs the MS that session setup is completed.

The above procedure assumes IPv4 IP address allocation. For IPv6, IP address allocation is different. Support of public IP addresses is a major difference between UMTS address allocation in IPv4 and IPv6. For IPv4, the MS is typically allocated a private address because of limited IPv4 address space. For IPv6, the MS is always allocated a public address.

At step 5 of the PDP context activation procedure, the GGSN allocates a complete IP address for IPv4. For IPv6, there are two alternatives for dynamic address allocation [8, 10] : *stateless* address allocation and *stateful* address allocation. Like IPv4, a stateful IPv6 address is allocated by the DHCP server at step 5. On the other hand, in *stateless address autoconfiguration*, the GGSN allocates a part of the IPv6 address called a link-local address for the MS by using its own IPv6 address pool at step 5. Then the MS generates the public IP address by combining the link-local address and a network-prefix address. The details are given as follows. As shown in Fig. 4, the MS first obtains the link-local address in the PDP context activation procedure (step 1 in Fig. 4, which is steps 1–6 in Fig. 3). Then the MS activates the IPv6 address autoconfiguration by sending the `Router Solicitation` message to the GGSN (step 2 in Fig. 4). The GGSN replies with the `Router Advertisement` message (step 3 in Fig. 4), which includes the network-prefix address. After the MS has received the `Router Advertisement` message, it obtains the IPv6 address by concatenating the link-local address and the network-prefix address. Then the GGSN updates the IPv6 address of the PDP contexts in the SGSN and MS (step 4 in Fig. 4). To avoid conflicts in link-local address assignment, the GGSN shall exercise neighbor discovery with other GGSNs. Note that in traditional IPv6 stateless address allocation, neighbor discovery is conducted by the mobile host. In UMTS, neighbor discovery is exercised by the GGSNs. Also note that the existing UMTS core network was developed based on the IPv4 transport network. Therefore, IPv6 packets are carried on top of the IPv4-based GTP tunnel and are invisible to the UMTS core network.

## Tunneling between UMTS and an External PDN

The GGSN interworks with the external data network through the Gi interface. The interworking mechanisms may be different for various APN configurations. Consider the example in Fig. 2 . For the **INTERNET** and **WAP** APNs, the GGSN connects to the external PDN directly through Ethernet or leased lines. For the **ISP** APN, the external PDN can be connected to the GGSN through either leased lines or a VPN. If the ISP connects to the GGSN through a VPN, tunneling is required. For the **COMPANY** APN, tunneling is always required for interworking between the GGSN and the corporate intranet. For each corporation, two tunnels are manually provisioned between the GGSNs and the VPN gateway in our example. Three tunneling methods that commonly used in existing PDNs have been proposed for UMTS.

**IP-in-IP tunneling.** Figure 5 shows the protocol stack of IP-in-IP tunneling [11]. In this method, a user IP packet (Fig. 5, 2) is encapsulated within a tunnel IP packet (Fig. 5, 1). The source and destination IP addresses in the tunnel IP packet are used to identify the endpoints of the tunnel (i.e., the GGSN and VPN gateway in the external PDN). The time-to-live and type-of-service parameters of the tunnel IP packet are the same as those of the encapsulated user IP packet. IP-in-IP tunneling incurs the smallest overhead compared to the other two tunneling methods, described next.

**Generic routing encapsulation (GRE) tunneling.** GRE supports multiprotocol tunneling where GRE packets can carry user packets of various protocols such as IP and Point-to-Point Protocol (PPP) [12]. GRE tunneling that uses PPP to carry user packets is also called Point-to-Point Tunneling Protocol (PPTP) [13]. PPTP

uses an enhanced GRE mechanism to provide a flow and congestion-controlled encapsulated datagram service for carrying PPP packets. Figure 6 shows an example where a user PPP packet (Fig. 6, 3) is encapsulated within a GRE packet. The GRE packet (Fig. 6, 2) is transferred over the IP (Fig. 6, 1). Compared to IP-in-IP tunneling, GRE tunneling requires an extra IP layer, and therefore more overhead is expected.

**Layer 2 Tunneling Protocol (L2TP) tunneling.** L2TP (Fig. 7, 3) supports PPP sessions (Fig. 7, 4) over UDP or other lower-layer protocols such as frame relay (FR) and ATM [14]. The user IP packets (Fig. 7, 5) are transferred over the PPP session. For the example in Fig. 7, an L2TP packet is encapsulated within a UDP packet (Fig. 7, 2). The PPP session is established on a per-session basis when the user PDP context is created. The user data packets are encapsulated within PPP packets.

Each of the above three methods can be used together with IPsec to provide end-to-end protection for packet delivery. IPsec protection is established between the firewall server and the VPN gateway, where the firewall function can be implemented within the GGSN or a separate firewall server collocated with the GGSN. Table 3 summarizes the characteristics of the three tunneling methods. Note that if an MS supports both PPP and IP, all three tunneling methods can be used to provide data sessions to this MS. IP-in-IP tunneling has better network efficiency because of short protocol overhead. However, if the user applications can only be carried over PPP, GRE tunneling must be used. Furthermore, if the lower-layer transport network is FR or ATM, L2TP tunneling must be selected.

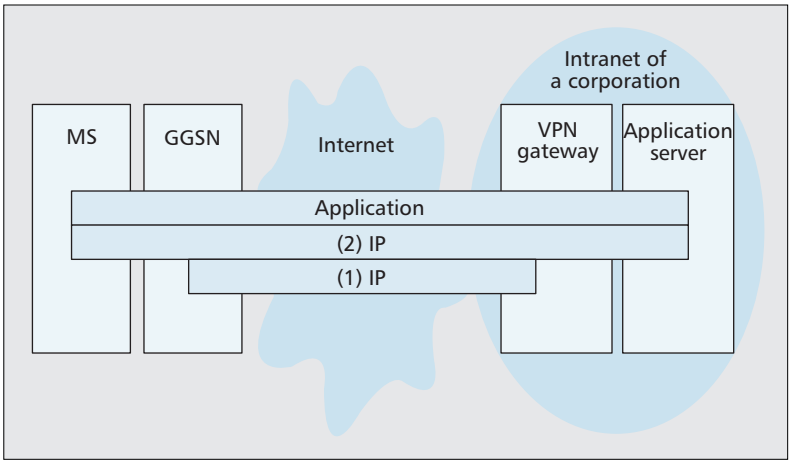



**Figure 5.** *IP-in-IP tunneling.*

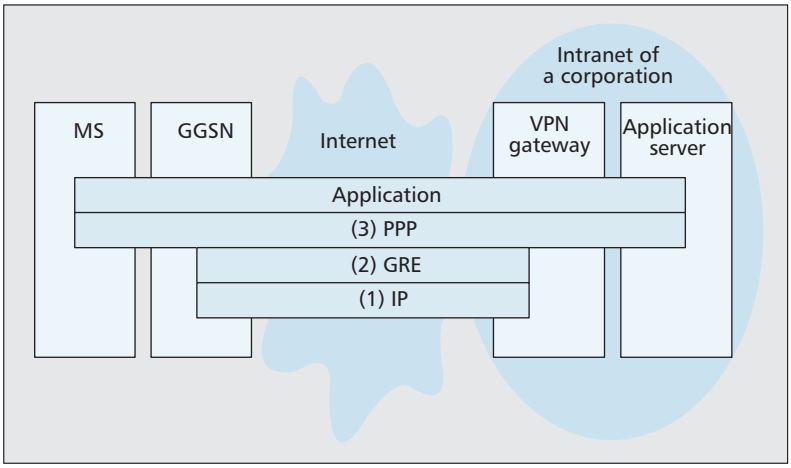



**Figure 6.** *GRE (PPTP) tunneling.*

## Quality of Service

UMTS defines four QoS classes for user data traffic: *conversational*, *streaming*, *interactive*, and *background* [4, 8]. The conversational and streaming classes support real-time traffic for services such as voice and streaming video. The interactive and background classes support non-real-time traffic for services such as Web browsing and email. Each class defines parameters including maximum bit rate, guaranteed bit rate, bit error ratio, and transfer delay.

Table 4 shows major QoS parameters for VoIP and Internet access services. In this table, VoIP is a conversational-class service with the maximum bit rate of 16 kb/s. Internet access is an interactive-class service with the maximum bit rate of 128 kb/s. The transfer delay (100 ms) for VoIP is significantly lower and more stringent than that of (unguaranteed) Internet access. On the other hand, the Internet access traffic requires lower bit error rate than that of VoIP.

Table 5 shows five scenarios for the end-to-end IP QoS conceptual model specified in 3GPP TS 23.207 [15]. The end-to-end QoS for PS service is negotiated among the MS, GGSN, and remote host located in the external PDN. 3GPP TS 23.207 assumes that the external PDN supports the differentiated services (DiffServ) QoS mechanism, and the GGSN is required to perform the DiffServ edge function in all scenarios. Within the UMTS network (MS-UTRAN-SGSN-GGSN), the IP QoS is translated and maintained by the UMTS QoS mechanism where the QoS parameters are set in the PDP contexts. In scenario 1, the GGSN performs the DiffServ edge function to support the IP QoS mechanism (between the GGSN and the remote host). Scenario 2 is similar to scenario 1 except that the MS also supports the DiffServ edge function to control the IP QoS over the external PDN. In this scenario, the GGSN DiffServ edge function may overwrite the IP QoS setup received from the MS. In scenarios 3 and 4, the end-to-end QoS is controlled by Resource Reservation Protocol (RSVP) signaling performed in the MS and remote host. In scenario 5, the MS and GGSN support the service-based local policy (SBLP) QoS mechanism to support 3GPP R5 IP multimedia service [16, 17]. In the above scenarios, either DiffServ marking or RSVP signaling is carried through the PDP context transparently over the UMTS network, and the GGSN performs interworking between the UMTS QoS and the IP QoS of the external PDN.

For illustration purposes, Fig. 8 shows a possible GGSN QoS architecture design for scenarios 1, 2, and 3, where the GGSN exercises the UMTS PDP context and DiffServ edge function. In this figure the dashed lines represent signal-

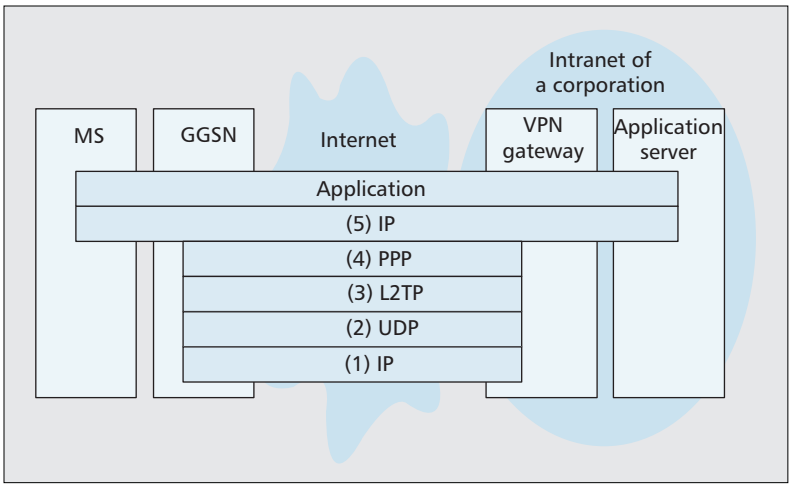



**Figure 7.** *L2TP tunneling.*

| Tunneling method | Overhead | Multiprotocol support | Transport network protocol support | MS protocol support |
|---|---|---|---|---|
| IP-in-IP | Low | No | IP | IP |
| GRE (PPTP) | Medium | Yes | IP | PPP |
| L2TP | High | Yes | IP/UDP, FR, ATM | IP |

**Table 3.** *Characteristics of the tunneling methods.*

| QoS parameter | VoIP (conversational class) | Internet access (Interactive class) |
|---|---|---|
| Maximum bit rate | 16 kb/s | 128 kb/s |
| Guaranteed bit rate | 12.2 kb/s | 100 kb/s |
| Bit error ratio | $10^{-4}$ | $10^{-6}$ |
| Transfer delay | 100 ms | Unguaranteed |

**Table 4.** *Major QoS parameters for VoIP and Internet access services.*

ing links, and the solid lines data links. The QoS management in the GGSN includes the following functions.

The **resource manager** (Fig. 8, 1) is responsible for *bearer management* and *resource monitoring*. The bearer management function interrogates the `Admission Controller` to determine whether the GGSN supports the specific requested service, and checks if the resources are available. It allocates the resources requested for each individual bearer service. The negotiated resources are then specified in the PDP context. The available resources are tracked by the resource monitor function.

The **admission controller** (Fig. 8, 2) determines if a new or modified request of a PDP context can be accepted based on the available resource information provided by the `Resource Manager`.

The **packet classifier:** (Fig. 8, 3) maps each incoming user packet to the corresponding PDP context. The `Packet Classifier` may enable multiple PDP contexts to share one IP address using the traffic flow template (TFT) technique, which employs IP header fields and higher-level headers (UDP/TCP) to differentiate PDP contexts. For example, the user may activate a PDP context with interactive class for Web browsing, and then activate the secondary PDP context with conversational class for VoIP. These two PDP contexts share one IP address, but are managed with different QoS classes. With TFT, UMTS can efficiently manage IP address allocation. The 2.5-generation GPRS can also simultaneously support multiple PDP contexts for a mobile user, but each PDP context shall be assigned an individual IP address. Therefore, without TFT, the IP address space will be an issue when the number of mobile users becomes large.

The **traffic conditioner** (Fig. 8, 4) provides conformance to the negotiated QoS for the data traffic of a service. The incoming data packets from the external PDN may result in bursts that do not conform to the negotiated QoS. If the incoming user data traffic exceeds the maximum bit rate specified in the corresponding PDP context, these data packets will be queued in the buffer and transferred later. The queued packets may be dropped due to network congestion.

The **packet mapper** (Fig. 8, 5) marks each incoming data packet with a specific QoS indication related to the bearer service, and translates the QoS parameters of the outgoing data packet into those of the external PDN. For example, if the external PDN supports the DiffServ control point (DSCP) mechanism [18–20], the conversational class packets are marked with the expedited fforward (EF) codepoint, which specifies the priority for delivery over the external PDN. The mapping between UMTS QoS classes and DSCP codepoints is given in Table 6.

The **packet scheduler** (Fig. 8, 6) delivers incoming data packets based on the priority specified for QoS classes. The `Packet Sched-`

| Network elements | Scenario 1 | Scenario 2 | Scenario 3 | Scenario 4 | Scenario 5 |
|---|---|---|---|---|---|
| MS | — | DS | DS and RSVP | RSVP | SBLP |
| GGSN | DS | DS | DS | DS and RSVP | DS and SBLP |
| External PDN | DS | DS | DS | DS and RSVP | DS |
| Remote host | DS | DS | DS and RSVP | DS and RSVP | DS and SBLP |

DS: DiffServ
RSVP: Resource Reservation Protocol
SBLP: Service-based local policy

**Table 5.** *Five conceptual end-to-end IP QoS models.*

uler also checks the available resources with the Resource Manager. If the available resources cannot support delivery of all packets, the Packet Scheduler queues the packets, and first transfers the-high priority packets when resources are available.

The **GTP/IP packet converter** (Fig. 8, 7) encapsulates incoming IP packets into GTP packets, and decapsulates outgoing GTP packets into IP packets.

Note that the Resource Manager and Admission Controller are involved in PDP context activation. The Packet Classifier, Traffic Conditioner, Packet Mapper, and Packet Scheduler are involved in packet delivery. Consider the QoS processing for an incoming IP packet from an external data network to the SGSN, which is illustrated in the nine steps of Fig. 8 . Assume that the external data network supports the DSCP QoS mechanism, and the user activates two PDP contexts. The primary PDP context (PDP-1) is used for VoIP service (conversational class). The secondary PDP context (PDP-2) is used for Internet access service (interactive class). Both PDP contexts share one IP address of the MS. Suppose that the IP address is 140.150.220.110. PDP-1 utilizes the UDP port 8010 and PDP-2 utilizes the TCP port 80.

**Step 1.** The incoming IP packet from the external data network is categorized to the corresponding PDP context by the Packet Classifier. The Packet Classifier first checks if the destination IP address of the incoming IP packet can be found in any PDP contexts created in the GGSN database. If not, the IP packet is filtered out. In our example, since the IP address is mapped to multiple PDP contexts (PDP-1 and PDP-2), the Packet Classifier exercises the TFT to determine the corresponding PDP context. That is, if the IP address of a data packet is 140.150.220.110 and the UDP port number is 8010, the data packet is classified as traffic for PDP-1. If the TCP port number is 80, this data packet will be classified as traffic for PDP-2.

| QoS class | DSCP codepoint | Delivery priority |
|---|---|---|
| Conversational | Expedited forwarding | 1 (high) |
| Streaming | Assured forwarding, class 1 | 2 |
| Interactive | Assured forwarding, class 2 | 3 |
| Background | Best effort | 4 (low) |

**Table 6.** *Mapping between UMTS QoS classes and DSCP codepoints.*

**Step 2.** Based on the maximum bandwidth parameter set in the PDP context, the Traffic Conditioner examines to see if the traffic volume of the incoming data packets with specific PDP context exceeds the allocated bandwidth. For example, the maximum allocated bandwidth is 128 kb/s for PDP-2 (Internet access; Table 4). Therefore, the exceeded packets are queued when the traffic volume of PDP-2 exceeds 128 kb/s.

**Steps 3 and 4.** For each data packet, the Packet Mapper specifies the corresponding UMTS QoS parameters such as bit error ratio and transfer delay. Based on the specified QoS, the Packet Scheduler checks the available resources and determines the delivery priority of the IP packet. In our example, the data traffic of PDP-1 has higher delivery priority than PDP-2.

**Steps 5 and 6.** The GTP/IP Packet Converter encapsulates the IP packet into a GTP packet. Then the GTP packet is transferred to the SGSN through the Gn interface.

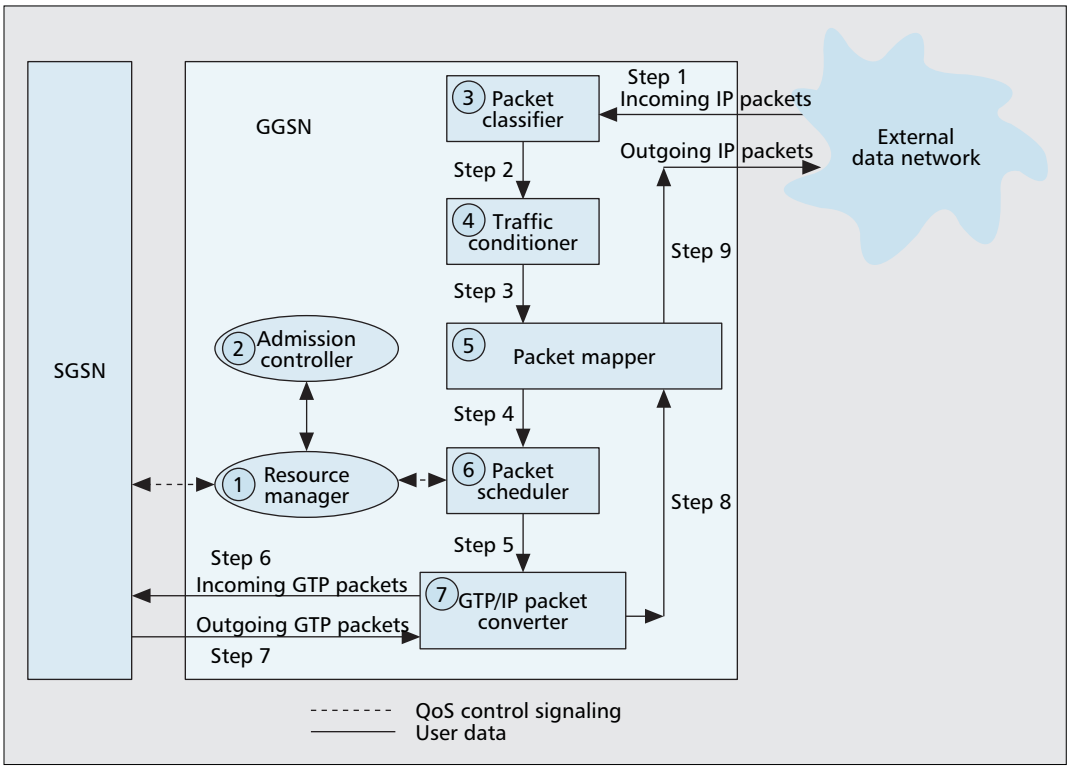



**Figure 8.** *GGSN QoS architecture.*

Based on our experience as a mobile operator, we described how the IP address can be dynamically allocated by the GGSN, a DHCP server or a RADIUS server, and how the GGSN can interwork with the external PDN through either the leased lines or the VPN.

The QoS processing for an outgoing GTP packet from the SGSN to the external data network is described as follows.

**Step 7.** The `GTP/IP Packet Converter` decapsulates a GTP packet into an IP packet by removing the GTP header.

**Step 8.** The `Packet Mapper` marks the IP packet with the corresponding codepoint of the external data network. For data traffic of the VoIP service (PDP-1), the IP packet is labeled with the EF codepoint. For data traffic of the Internet access service (PDP-2), the IP packet is labeled with the BF codepoint.

**Step 9.** Finally, the IP packet is transferred to the external data network through the Gi interface. The external data network transports the IP packet depending on its QoS specified in the GGSN.

## Concluding Remarks

This article describes the GGSN functions for IP connection including APN and IP address allocation, tunneling technologies, and QoS management. In GPRS/UMTS, the GGSN serves as a gateway node to support IP connection with external PDNs. Based on our experience as a mobile operator, we describe how the IP address can be dynamically allocated by the GGSN, a DHCP server, or a RADIUS server, and how the GGSN can interwork with the external PDN through either leased lines or a VPN. For VPN interworking, tunneling is required to connect the GGSN with external PDNs. We show the tunneling alternatives including IP-in-IP tunneling, GRE tunneling, and L2TP tunneling. Finally, we elaborate on QoS processing of a user data packet at the GGSN.

For further reading, readers are encouraged to understand UMTS radio technology (i.e., WCDMA) [3]. For UMTS core network architecture, the reader is referred to [2]. The UMTS protocol stacks are introduced in [8]. Details of UMTS mobility management can be found in [1]. The complete 3GPP specifications can be found at http://www.3gpp.org.

## Acknowledgments

The three anonymous reviewers have provided useful comments that significantly improve the quality of this article. This work was sponsored in part by the MOE Program for Promoting Academic Excellence of Universities under grant number 89-E-FA04-1-4, IIS, Academia Sinica, CCL/ITRI, and the Lee and MTI Center for Networking Research, NCTU.

## References

[1] Y.-B. Lin *et al.*, "Mobility Management: From GPRS to UMTS," to appear, *Wireless Commun. and Mobile Comp,*, 2001.
[2] 3GPP, Services and Systems Aspects; Architectural Requirements for Release 1999," Tech. rep. 3G TS 23.121 v. 3.4.0 (2000-10), 2000.
[3] H. Holma and A. Toskala, Eds., *WCDMA for UMTS*, Wiley, 2000.
[4] Y.-B. Lin, and I. Chlamtac, *Wireless and Mobile Network Architectures*, Wiley, 2001.
[5] 3GPP, "Core Network; General Packet Radio Service (GPRS); GPRS Tunneling Protocol (GTP) across the Gn and Gp Interface (Release 99)," 3G TS 29.060 v. 3.7.0 (2000-12), Technical report, 3GPP, 2000.
[6] J.-H. Park, "Wireless Internet Access for Mobile Subscribers Based on the GPRS/UMTS Network," *IEEE Commun. Mag.*, Apr. 2002.
[7] A. Mishra, "Performance and Architecture of SGSN and GGSN of General Packet Radio Service (GPRS)," *IEEE GLOBECOM 2001*, 2001.
[8] 3GPP, "Services and Systems Aspects; General Packet Radio Service (GPRS); Service Description; Stage 2," Tech. spec. 3G TS 23.060 v. 3.6.0 (2001-01), 2000.
[9] Y.-B.Lin and Y.-K. Chen, "Reducing Authentication Signaling Traffic in Third Generation Mobile Network," to appear, *IEEE Trans. Wireless Commun.*
[10] S. Thomson and T. Narten, "IPv6 Stateless Address Autoconfiguration," IETF RFC 2462, Dec. 1998.
[11] W. Simpson, "IP in IP Tunneling," IETF RFC 1853, Oct. 1995.
[12] S. Hanks *et al.*, "Generic Routing Encapsulation over IPv4 Networks," IETF RFC 1702, Oct. 1994.
[13] K. Hamzeh *et al.*, "Point-to-Point Tunneling Protocol (PPTP)," IETF RFC 2637, July 1999.
[14] W. Townsley *et al.*, "Layer Two Tunneling Protocol 'L2TP,'" IETF RFC 2661, Aug. 1999.
[15] 3GPP, "Services and System Aspects; End-to-End Quality of Service (QoS) Concept and Architecture (Release 5)," Tech. spec. 3G TS 23.207 v. 5.6.0 (2002-12), 2002.
[16] 3GPP, "Services and Systems Aspects; IP Multimedia Subsystem (IMS) Stage 2 (Release 5)," 3G TS 23.228 ve. 5.7.0 (2002-12), 2002.
[17] Y.-B. Lin *et al.*, "All-IP Approach for UMTS Third Generation Mobile Networks," *IEEE Network*, vol. 16, no. 5, 2002, pp. 8–19.
[18] K. Nichols *et al.*, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers," IETF RFC 2474, Dec. 1998.
[19] J. Heinanen *et al.*, "Assured Forwarding PHB Group," IETF RFC 2597, June 1999.
[20] V. Jacobson, K. Nichols, and K. Poduri, "An Expedited Forwarding PHB," IETF RFC 2598, June 1999.

## Biographies

YUAN-KAI CHEN (ykchen@cht.com.tw) received his B.S.C.S.I.E. and M.S.C.S.I.E. degrees from National Chiao Tung University (NCTU), Hsinchu, Taiwan, R.O.C., in 1989 and 1991, respectively, and his Ph.D. degree in computer science and information engineering from NCTU in 2003. In 1991, he joined the Telecommunication Laboratories, Chunghwa Telecom Co., Ltd. and was involved in the implementation of a SONET/SDH multiplexer and the development of an ADSL transceiver. In 1998 he worked on the 3G trial team. Since then, he has been involved in the design of the radio access network , mobile packet switched data and multimedia services, and the study of mobile network evolution. His research interests include design and analysis of personal communications services (PCS) networks, 3G networks, wireless Internet, mobile computing, and performance modeling.

YI-BING LIN [F] (liny@csie.nctu.edu.tw) received his B.S.E.E. degree from National Cheng Kung University in 1983, and his Ph.D. degree in computer science from the University of Washington in 1990. He is a Chair Professor at Providence University. He is an ACM Fellow.