TCP/IP in Cellular Mobile Environment

Pekka Toppila
Pekka.Toppila@hut.fi

Abstract

This article handles the mobility issues in TCP/IP data transmission. The study focuses on the cellular mobile environment, which is used in the largest mobile telephone networks. The articles discusses the properties of the mobile IP and some basic TCP/IP protocols needed in mobile environment.

Introduction

The rapid growth of both the Internet usage and the number of mobile telephones form two major trends at telecommunications field. This creates need for bringing the Internet services to the mobile terminals. The second generation mobile systems are designed for speech traffic, and don't therefore effectively support data traffic transfer. The situation is different with the third generation mobile systems, since the Internet access is on of their main purposes.

Mobility of the TCP/IP host is a problematic issue to the basic routing protocols. In the mobile cellular environment the host location may change during the connection. In addition to the mobility also other cellular network aspects must be considered. These include initializing the IP connection over the wireless link as well as handling the high bit error rate and the fading effects caused by the radio network.

Additional protocols are needed to enable TCP/IP transmission in mobile cellular environment. This document studies the point-to-point protocol (PPP) and the authentication protocol RADIUS as well as its successor DIAMETER.

1 Mobile IP

In this article mobility means the capability of a node to change its attachment point from one link to maintaining its IP address and all old communications [1]. This is not possible with normal IP routing, where IP addresses and routing are based on subnetwork prefixes. Mobile IP is thus needed to solve two major problems: when a node moves from one link to another without changing its IP address, it cannot receive packets in its new address, and if the node changes its IP address when it moves, it must terminate and restart any communications.

The Mobile IP introduces three new functional entities[2]:

- Mobile Node
- Home Agent
- Foreign Agent

A mobile node is a host that changes its attachment point from one network to another. It may do this without changing its IP address and without terminating its ongoing connections.

Home agent is a router on a mobile node. It maintains location information for the mobile node and tunnels datagrams to the mobile node when the node is away from the home network.

Foreign agent is a router on a network that a mobile node visits. It provides routing services to the mobile node, and detunnels and delivers the datagrams that were sent by the home agent to the mobile agent.

A mobile node has a long-term IP address on a home network. This works the same was a normal, "permanent" IP address. When the mobile node is in a foreign network, it is given a temporary "care of address". The mobile node normally uses the home address as the source address of all IP datagrams it sends. An example of mobile IP functionality is presented in Figure 1.

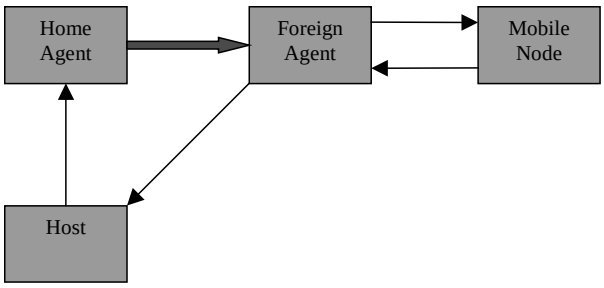

Figure 1: Mobile IP architecture [2]

In the illustrated example the mobile node is located in a foreign network. When a host sends packets to the mobile node the packets fist arrive to the home agent via standard IP routing. The datagrams are then tunneled to the care-of address, in this case to the foreign agent. The datagram is detunneled in the foreign agent node and delivered to the mobile node. The datagrams sent by the

mobile node are delivered using standard IP routing. In this case the foreign router is the default router of the mobile node.

In the illustrated case the care-of address was the IP address of the foreign agent. The foreign agent was the end-point of the tunnel and it delivered the inner datagrams to the mobile host. Another alternative is to use the mobile nodes local IP address as the care-of address. In this case the mobile node itself serves as the enpoint of the tunnel.

Agent discovery and registration support services are defined for Mobile IP. Foreign and home agents may advertise their availability to the mobile nodes in their networks. When a mobile node arrives into a foreign network it has to register its care-of address to the home agent.

1.1 Tunneling

Tunneling plays an important role in Mobile IP. The only tunneling method that should always be supported is IP in IP encapsulation[3]. Encapsulation methods that may optionally be supported [2] are Generic Routing Encapsulation (GRE) [4]and Minimal Encapsulation within IP[5].

IP to IP encapsulation is a simple method for encapsulating an IP packet inside the payload portion of another IP packet. IP packets are transferred between the tunnel endpoint so that the tunnel appears as a single virtual link to packets that pass through it. The IP to IP encapsulation method also maintains state information at tunnel entry-points, which enables tunnel entry-points to relay ICMP messages from inside the tunnel to the original source.

The Minimal Encapsulation method minimizes the number of additional bytes needed in implementing the tunnel. Tunnel endpoints modify the original IP packet header and add a minimal forward header between the IP header and payload. The additional header is normally 12 bytes long, saving 8 bytes over IP in IP Encapsulation. However, it cannot be used if the original packet is fragmented.

The Generic Routing Encapsulation method is the heaviest of the three tunneling methods. In this method the incoming IP packet is packet inside another IP packet, and also an additional GRE header is added between the new IP header and the original IP packet. The GRE method has the advantage that it supports also other network layer protocols in addition to IP.

In order to provide security secured tunneling methods should be used. Secured encapsulation as well as other security issues are considered in the next section.

1.2 Security

Mobile IP is often used over wireless links or in otherwise unsecured environments. It is therefore important to provide additional security mechanisms when Mobile IP is used. Another aspect that should be considered is that encrypting IP traffic limits intermediate node's access to the IP packets and may thus make some Mobile IP functions impossible.

Secure connections can be provided end-to-end or using trusted mediator entities, security proxies. Encryption is used to make data unintelligible during transformation and to provide confidentiality. Encryption keys, digital codes, are needed to encrypt, decrypt, and sign transferred data packets. A security association (SA) is formed between secure connection end points. SA is a relationship formed between two or more entities that describes how the entities use security services to communicate securely.

The IP Security (IPSEC) protocol suite forms a complete family of TCP/IP security protocols. IPSEC is an IETF standard for IP level packet protection. It provides algorithm independent base specifications offering privacy and authentication services to all IP-based communication. Unidirectional SAs are created between IPSEC entities. An IPSEC SA is a data structure that specifies the authentication and encryption algorithms, needed keys and their lifetimes as well as the lifetime of the SA and a sequence number for replay prevention. Each SA is identified by an IPSEC protocol identifier, the destination host address and a Security Parameter Index (SPI), which is a 32-bit pseudo random number assigned to the SA upon SA negotiations.

IPSEC has two main transformation types: the Authentication Header (AH) transformation type and the Encapsulated Security Payload (ESP) transformation type[6]. AH provides authentication, including connectionless datagram integrity, but offers no confidentiality. An IPSEC AH transformation calculates a Message Authentication Code (MAC) to each outgoing datagram. The MAC value is used to verify that the datagram is not changed. The ESP transformation provides, in addition to authentication and integrity, also confidentiality. This is achieved by encrypting portions of datagrams.

Both the AH and ESP transformation types can operate in transport or tunnel mode. The transport mode in commonly used in host-to-host communications whereas the tunnel mode is typically used in security gateways.

The transport mode uses the original packet's IP header as the transformed packet's IP header, whereas in the tunnel mode the original IP header information is included in the transformed packet and a new IP header is added to the transformed packet. ESP transformation in different transport modes is illustrated in Figure 2.

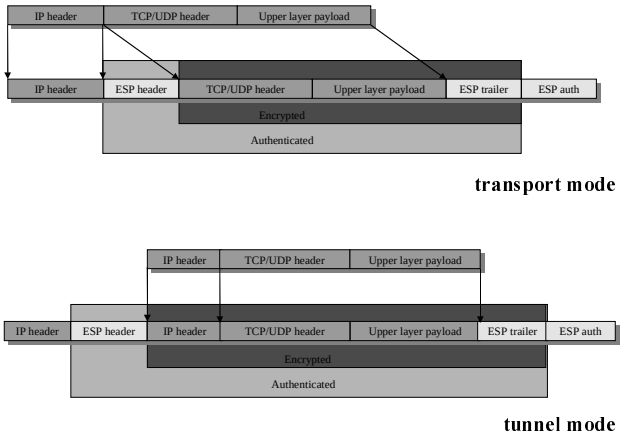

Figure 2: ESP transformation in tunnel and transport modes

When ESP transport mode is used the packet's IP header is copied from the original IP header. The TCP/UDP header and the upper layer payload are included in the encrypted message part. An ESP header is added before the encrypted data part, and ESP trailer is added to mark the end of the part. An ESP authentication trailer ends the transport packet. It authenticates the encrypted data part and the ESP header. In the ESP tunnel mode the original IP header is transported in the encrypted data part and a new IP header is created for the transport packet.

In addition to the IPSEC IP packet securing protocol family, additional protocols are needed to determine and handle the keys needed for authentication and encryption. The key handling mechanisms have no significant effect on the Mobile IP and these protocols are therefore not studied further.

From the security point of view the ESP tunnel mode transport would be a recommended method for tunneling between the agents. Since the IP address of the mobile node is not visible to the communicating host, it makes forming end-to-end encrypted connections more difficult. A natural solution would be to use the foreign and home agents as secure proxies.

1.3 IPv6

IP version 6[7] offers many enhancements compared to the current version. Firstly, the IPv6 address space is much lager; 128 bit address solves the current problem of lacking free addresses. It also supports external headers and has some new functions, which can be used by Mobile IP.

The most important improvement is the larger IP address space. The IPv6 addresses enable giving each mobile host an own IP address and foreign agents are no longer needed. The IPv6 protocol includes router discovery functionalities for determining the current location of the host. These functionalities can be used for agent discovery in the Mobile IP.

1.4 Real-Time Traffic

Real-time traffic can be defined to be traffic, where the transmission delay requirements are high and keeping the delay low is more important than low bit error rate. The basic TCP/IP protocol family cannot support real-time traffic. Protocols like RSVP [8] are defined for receiving resources and for providing guaranteed QoS.

Using RSVP in a mobile environment has some additional problems[1]. Since the route to the mobile node may change during the session new reservation must be done within a short period of time. Two options for the new routes are to use the old reserved path and make new reserved path from the old location to the new, and to make a new path from the source to the new location.

Another problem is with the tunneling used in the Mobile IP. In order to provide QoS control during the tunnel the intermediator nodes in the tunnel may have to access the IP data transferred. In case of the basic IP tunneling methods presented this is possible, although it requires additional resources, but when ESP encrypted connection is used this cannot be done. A solution to the secured tunnel is to receive own tunnels for each quality class. This of course makes the tunneling functionality more complicated.

2 Cellular Mobile Systems

All the most common mobile telephone systems today base on the cellular network structure. Each base station has a certain area, a cell, inside which it handles the mobile users. The base stations are then further divided into groups that are each controlled by an own controller node. The cellular network structure is illustrated in Figure 3.

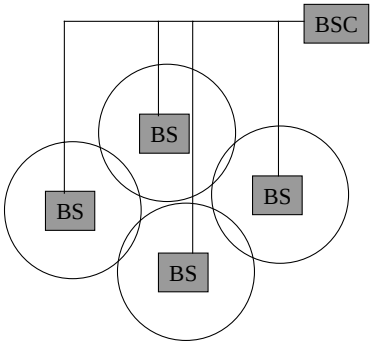

Figure 3:Cellular network structure

When a mobile host moves from one cell to another the control of the connection also changes from one base station to another. This change of location cell is called handover. Another mobility issue in cellular networks is roaming, which means location handling in larger scale movement, possibly when between networks of different operators.

Handover happens in shorter time period and during a connection, and is therefor more interesting from the TCP/IP mobility's point of view. A part of the handover functionality is handled on the link-layer but some parts also require IP mobility functionality. TCP/IP traffic handling in some important cellular systems is handled in the next sections.

2.1 GSM

Work for the GSM cellular digital standard began in ETSI in the 1980's. First commercial GSM service started in 1991. With over 130 million users in almost 130 countries (end of '98) it is the most popular digital cellular system.

Original GSM frequency band is 900 MHz but also systems at 1800 and 1900 MHz have been implemented. Current GSM radio systems are based on radio cells with diameters from about 300 meters to 30 kilometers. The cell size depends on the subscriber density. Each subscriber is identified by a SIM card, which makes the identification independent from the equipment used.

GSM multiplexing is a combination of the Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA) methods. The frequency band is divided into 200 kHz sub-bands each including 8 timeslots. When the original 25 MHz band is used, totally 25 000 / 200 * 8 = 1000 timeslots are available for speech and signaling. The actual number of timeslots available in a cell is smaller since neighbor cells are not allowed to use the same frequency bands.

GSM offers relatively good transfer security with encryption algorithms, digital transfer, discontinuous transmission and frequency hopping. GSM codecs work at 13 kbit/s, which gives a flow of 22 kbit/s, when equipped with error correction data. Data transfer is normally available at 9600 bit/s, but different codings are used to enable data transfer at speed 14.4 kbit/s or even faster.

GSM system is originally designed for speech traffic. A connection with constant bit rate is created between communicating hosts. Data delivery is clumsy and often expensive. Data connections are mainly created with a dial-in connection protocol like in the fixed telephone networks. Some enhancements are made to the original GSM specification to enable more efficient data connections. These include HSCSD for faster switched data transfer and GPRS for packet data delivery.

2.2 GPRS

The generic Packet Radio Service (GPRS) [9]operates in packet mode to efficiently transfer mobile data and signaling information. GPRS enables dynamical radio resource allocation between users in order to optimise the network and radio resource usage[10]. The main packet protocol to be used in GPRS is IP, but also interworking with X.25 is defined. Radio and network subsystems are strictly separated in order to allow GPRS network subsystems to be used with various radio access technologies.

GPRS defines a number of new flexibly allocable radio channels. In GPRS one TDMA frame may include from one to eight time slots. These time slots are shared between active users, separately on uplink and downlink directions. The coding schemes defined at the starting phase allow bitrates from 9 to over 150 kbit/s per user. GPRS can temporarily transfer large volumes of data and supports also bursty and intermittent data transfer.

A Quality of Service profile is defined by five different class values. These are precedence, delay, and reliability classes as well as peak and mean throughput classes. The five QoS attributes allow a large number of different quality profiles to be created in order to cover separate communication cases. The GPRS network supports only a limited subset of all these available profiles.

GPRS transports user data transparently between the MSs and external data networks using tunneling and encapsulation. This transparent method makes it easier to add new interworking protocols in the future. The GPRS system handles user mobility and mapping between IP and GSM internal addressing. In this way the GPRS system hides user mobility from external networks and from the Internet it looks like a common subnetwork.

The GPRS network is designed for packet data transport and especially for Internet connections. Therefore it

effectively supports TCP/IP transfer. The IP mobility is handled inside the GPRS network in a way much like the Mobile IP.

3 Protocols Needed in Mobile TCP / IP access

Additional higher-layer protocols may be needed to provide TCP/IP connection over cellular network. This chapter studies the PPP protocol that is a protocol often used for creating a data connection over a telephone line. Also two protocols for authentication, authorisation, and accounting, RADIUS and DIAMETER are handled.

3.1 Point-to-Point Protocol

The Point-to-Point Protocol (PPP) [11]offers a standard encapsulation for IP protocol over point-to-point IP links. PPP is the most commonly used protocol in serial line Internet connections. It also includes mechanisms for IP address assignment and management, asynchronous and synchronous encapsulation, network protocol multiplexing, link quality testing and configuration, error detection and a negotiation option.

In comparison to the older serial line protocol, SLIP, PPP offers support for also other protocols than TCP/IP, such as Appletalk and IPX. Also automatic login and configuration negotiation, as well as support for using several protocols simultaneously on same connection are properties not supported by SLIP.

The Point-to-Point protocol frame structure as well as the principles and terminology are adapted from the ISO standardised High-level Data Link Control (HDLC). The PPP frame format is presented in Figure 4.

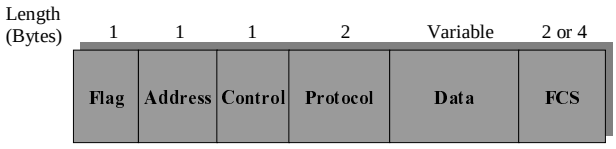

Figure 4: PPP frame format

The *flag* field is a binary sequence 01111110 that indicates the beginning or end of a frame. The *address* is an octet of ones indicating the standard HDLC broadcast address. The *control* field is a sequence 00000011 to indicate transmission of HDLC user data in an unsequenced frame. The *protocol* field identifies the protocol packed in the *data* part of the frame. The data field is maximum 1500 bytes long. A closing flag sequence is added to the data field to indicate the end of the field. The PPP frame is ended by a frame check sequence *(FCS)* flag.

The PPP protocol has three main components. The encapsulation component is based on HDLC. An extensible Link Control Protocol (LCP) handles data link establishment, configuring and testing. The third component is a group of Network Control Protocols (NCPs), which are used for network-layer protocol establishment and configuring. The NCPs enables using multiple network-layer protocols simultaneously.

The PPP protocol offers a way to create an IP point-to-point link over a telephone connection. It is usable in speech optimised network like fixed telephone network or GSM but restricts the connection control too much to be efficiently used in packet transfer optimised systems like GPRS or UMTS.

3.2 RADIUS

The Remote Authentication Dial-In User Service (RADIUS) [12] is a protocol for centralising authentication, authorisation, and accounting of remote access users. RADIUS enables a network manager to maintain only one database for all remote user authentication, which greatly simplifies administration management with large numbers of dial-in users.

The RADIUS protocol works on the top of UDP. RADIUS operations follow the client-server model. A Network Access Server (NAS) operates as a RADIUS client to which RADIUS servers offer authentication and configuration information service. A transaction between the client and a user can be secured with shared secret recognition and password encryption, which improves RADIUS security. The RADIUS protocol is an extensible protocol, to which new attributes can be easily added and it supports several authentication methods, including PPP, PPP Authentication Protocol (PAP), Unix login, and Challenge Handshake Authentication Protocol (CHAP).

A typical RADIUS authentication process is illustrated in Figure 5. A user contacts a RADIUS client presenting authentication information either through a prompted login or an authentication packet. If the client uses RADIUS authentication, it sends an access request packet to the RADIUS server. This packet usually contains the user's name information, the user's password, which is encrypted using the MD5 encryption algorithm, and identification numbers of the user and the port the user is accessing. If no response from the server is received within a certain time, the packet is resent or sent to an alternative server.

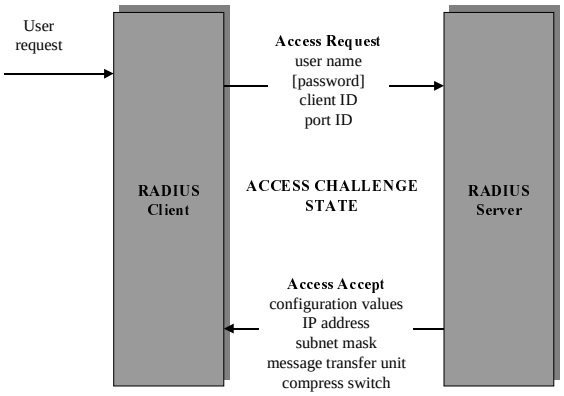

Figure 5: RADIUS authentication flow [12]

Once the server receives a request it validates it and consults a user database for the user's access requirements. These requirements may include, in addition to the required password, client or port numbers the user is allowed to access. If some requirements are not fulfilled, an access reject-response is sent to the client. If all conditions are met and no access challenge is performed, the server responses with an access accept message, which includes a list of configuration values for the user. These include the service type used and required values to support the service, which for PPP or SLIP service may include an IP address, a subnet mask, message transfer unit size, and desired compression.

If the server requires access challenge authentication, it sends an access challenge as a response to the access request message. This response includes a challenge to the user. The client sends the challenge to the user and responses with a new access request, in which the user password attribute is replaced with an encrypted user's response. After receiving this new request the server may answer with an accept, reject, or another challenge message.

A RADIUS accounting server collects user's usage information from RADIUS clients. This information may contain user name, types of service, address used, and access port. The communication with the accounting server uses accounting request and response messages. Since UDP is used, accounting server acknowledges successfully received and recorded requests to the client. The account information may be used for user charging or usage monitoring.

The RADIUS concept is not depended on the location of the nodes and is therefore very suitable for mobile environment. RADIUS is a functional standard widely used among radio access providers. The RADIUS software is also available as shareware. However, protocol development is concentrated on the more advanced DIAMETER protocol, which is planned to replace RADIUS.

3.3 DIAMETER

DIAMETER specifies a protocol for authentication, authorisation, accounting, and policy and resource control. It is based on the widely used RADIUS protocol. DIAMETER can be used over TCP or UDP. It is lightweight and simple to implement. Even though it is a request/response protocol, it allows a server to send unsolicited messages to a client, which wasn't possible in the RADIUS client-server model.

A DIAMETER message consists of a header followed by objects encapsulated as Attribute-Value-Pairs (AVPs). These AVPs include an identifier, length and data field. The AVP space is large enough to make the protocol extensible for future needs and to enable vendor specific AVPs. DIAMETER framework structure is discussed in [13].

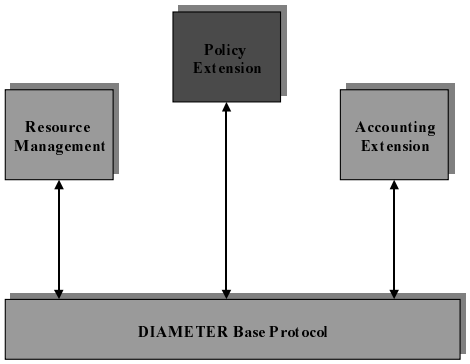

Figure 6: DIAMETER basic architecture

The basic DIAMETER architecture is presented in Figure 6. DIAMETER consists of modules that define primitives for DIAMETER entities. The base protocol defines header formats, security extensions, requirements, and some mandatory commands as well as AVPs. Resource management extension provides capability to maintain session state information in client to server and server to server communication. This session identification can be used to assign QoS options to a certain session by reserving certain resources to the session. The accounting extension is used by most of the DIAMETER users. This extension must be scalable and secure. Many applications use RADIUS accounting although it is only an informational protocol, which indicates that a standard accounting protocol is required in DIAMETER. The policy extension presented is an example of an additional service's ability to use DIAMETER.

DIAMETER offers both hop-by-hop and end-to-end security. The hop-by-hop security can be realised relying on IP security or by shared secret method analogous to the system used in RADIUS. DIAMETER uses public key cryptography to provide end-to-end security. This ensures that information integrity, confidentiality, and non-repudiation are preserved when a message is transferred through a DIAMETER proxy chain.

A quality of service extension to DIAMETER is presented in [14]. The extension adds a simple client/server model support for RSVP and differentiated services schemes. It introduces bandwidth request and response commands as well as a group of AVPs for QoS flow control and parameter configuration.

4 Conclusions

Basic TCP/IP protocol family cannot handle IP host mobility. Mobile IP extension is developed to allow an IP node to move and still maintain its IP address and ongoing connections. This is done by using a permanent home address and a temporary visitor address given by a foreign agent.

Most of the current cellular networks are designed for speech traffic. Mobile IP allows TCP/IP connections over these networks, but the data connections are not effectively supported in these connected mode, constant bit rate networks. On the other hand, new cellular systems like GPRS and the third generation cellular network UMTS, are optimised for IP traffic.

5 References

[1] Solomon, J.D.: "*Mobile IP: The Internet Unplugged*", Prectice-Hall, USA, 1998, ISBN_0-13-856246-6

[2] Perkings, C.:"IP Mobility Support", *Request for Comments 2002,* October 1996

[3] Perkings, C.: "IP Encapsulation within IP", *Request for Comments 2003*, October 1996

[4] Hanks, S., Li, R., Farinacci, D., Traina, P.: "Generic Routing Encapsulation (GRE)", *Request for Comments 1701*, October 1994

[5] Perkings, C.: "Minimal Encapsulation within IP", *Request for Comments 2004*, October 1996

[6] Kent, S., Atkinson, R.: "IP Encapsulating Security Payload (ESP)", *Request for Comments 2406,* November 1998

[7] Deering, S.,Hinden, R.: "Internet Protocol, Version 6 (IPv6) Specification", *Request for Comments 2460,* December 1998

[8] Braden, R., Zhang, L., Berson, S., Herzog, S., Jamin, S.: "Resource Reservation Protocol (RSVP) Version 1 Functional Specification ", *Request for Comments 2205*, September 1997

[9] "General Packet Radio Service (GPRS) Service Description", *ETSI specification GSM 03.60 version 6.2.0*, October 1998

[10] Cai, J., Goodman, D. J.:"General Packet Radio Service in GSM", Rutgers University, *IEEE Communications Magazine,* October 1997

[11] Simpson, W.:"The Point-to-Point Protocol (PPP) ", *Request for Comments 1661*, July 1994

[12] Rigney, C., Rubens, A., Simpson, W., Willens, S.: " Remote Authentication Dial In User Service (RADIUS) ",*Request for Comments 2138*, April 1997

[13] Calhoun, P.R., Zorn, G., Pan, P.: " DIAMETER Framework Document ",*Work in progress*, August 1998

[14] Calhoun, P., Speer, M., Peirce, K.:"DIAMETER QoS Extension ",*Work in progress,* May 1998