4G Mobile Network Architecture

Kalle Ikkelä, Marko Myllynen, Juha Heinanen and Olli Martikainen
Lappeenranta University of Technology, P.O.Box 20, 53851 Lappeenranta, Finland

{kalle.ikkela, rnarko.myllynen}@lut.fi, juha.heinanen@telia.fi, olli.martikainen@icon.fi

Key words: 4G, SIP, mobility, GSM

Abstract: The convergence of telecommunications, computing, and content industries has been one of the major trends during recent years. The convergence has led to the creation of a wide range of multimedia services that are available in digital form through the Internet. There is a need for user and terminal mobility supporting both personal communication and multimedia services on top of the Internet. To help meet this need, this paper presents ongoing research on the Fourth Generation (4G) roaming and mobility model. The 4G model is compatible with existing IP standards. The basic elements of the model are described and the model is compared to the GSM (2G) model.

1. INTRODUCTION

Traditional telecom and content services are vertically integrated. Each service depends on a dedicated network and corresponding terminals. Examples of such vertical services are the fixed telephone services, traditional data services and GSM services. Internet changes the vertical structure to a horizontal one: all terminals and services will be Internet compatible. Instead of vertical service "pipes" there will be a horizontal structure of services, network and access, as illustrated in Figure 1.

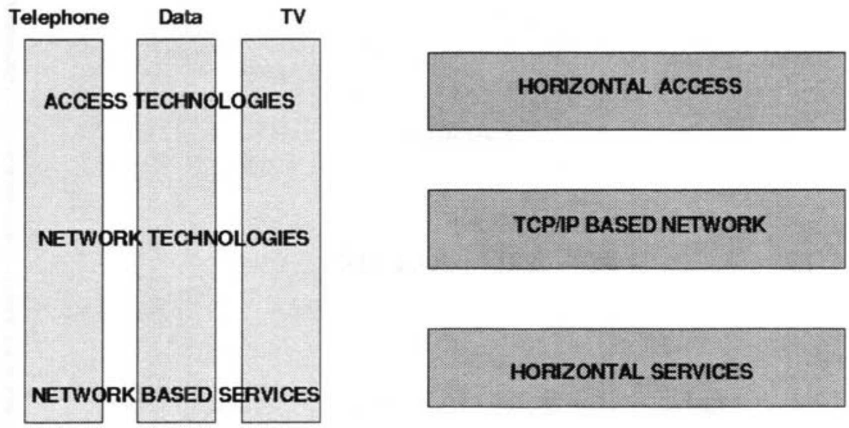

Figure 1. Vertical service "pipes" compared to horizontal service structure

The horizontal structure will change terminals, services, and the way services are managed. The horizontal structure will allow different combinations of service functionalities in the terminal equipment as depicted in Figure 2. The ovals show possible combinations of services in single terminals.

Horizontal networks will not only make existing services easier and more widely applicable, but also create a platform for the integration of various new services and applications into the same terminals.

New terminal applications in horizontal networks can be divided into **simple fixed-purpose terminals and intelligent terminals.** Possible fixed-purpose terminals can be wearables (watches, eyeglasses, clothes) or appliances (light switches, doors, micro-ovens). Intelligent terminals include Personal Digital Assistants (PDA), Smart Phones or Media Terminals. Simple terminals will connect to Personal Area Networks (PANs) or Domestic Area Networks. Intelligent terminals apply to Local Area Networks (LAN) or public access networks. They will have software and content-defined functionalities that allow various applications within one device.

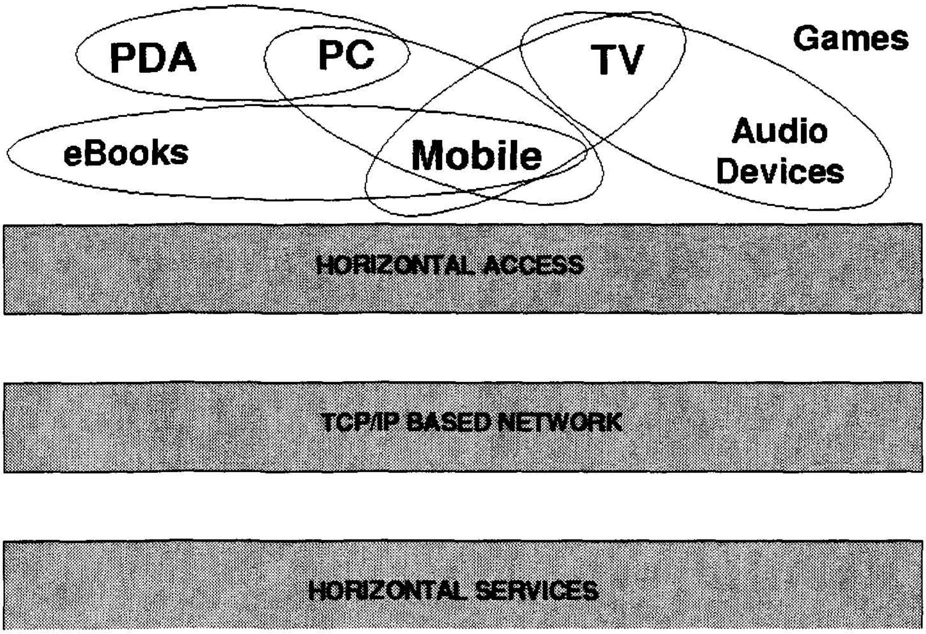

Figure 2. Restructuring to multifunction terminals and horizontal services

Fourth Generation Mobile (4G) means broadband mobile wireless services, which are based on IEEE 802.11 Wireless LAN (WLAN) [1] or Bluetooth (Bt) [3] access, IP mobility and Web type services. The radio access can be connected to private corporate LANs, public administration LANs, mobile WLANs installed in trains, airplanes, buses or cars and to single DSL connections. Corporate offices, shopping centers, hotels, airports, home networks, and personal area networks (PAN) will be the leading adopters of these technologies. The client devices for the 4G applications can be categorized as follows:

- Laptop PC with WLAN
- Laptop PC with Bt
- PDA with WLAN
- PDA with Bt
- Dual-mode wireless phone with GSM and WLAN or Bt
- Wireless phone with Bt
- Other specialized Bt devices.

2. ARCHITECTURE OF GSM AND UMTS NETWORKS

An overview of a GSM network architecture is presented in Figure 3. A GSM network is composed of several functional entities, whose functions and interfaces are defined. The GSM network can be divided into three parts: The Mobile Station (MS), carried by the subscriber; the radio link, controlled by the Base Station Subsystem (BSS) with the mobile station; the Mobile services Switching Center (MSC). The MSC is the main part of the Network Subsystem and performs the switching of calls and management of mobile services, authentication, for example. The operation and setup of the network is managed by the Operations and Maintenance Center (OMC). Each component dealing with mobility is described in more detail next.

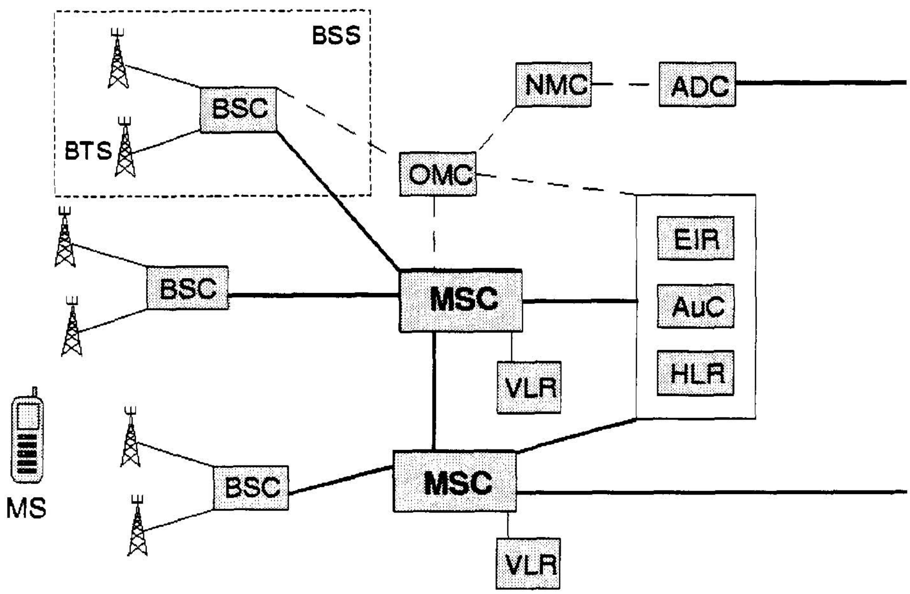

Figure 3. GSM network overview

Mobile Station (MS): The mobile station is the user terminal, which consists of a radio transceiver, signal processors, display, and a Subscriber Identity Module (SIM card). The SIM card enables the usage of services and personal mobility.

Base Station Subsystem (BSS): The Base Station Subsystem has two components, the Base Transceiver Station (BTS) and the Base Station Controller (BSC). The radio transceivers are located in BTSs. The BTS also manages the radio-link with MS. The Base Station Controller manages one or more BTSs and handles radio resources, such as radio-channel setup, frequency hopping, and handovers.

Mobile Services Switching Center (MSC): The Mobile services Switching Center is the central component of the Network Subsystem. MSC is like a normal switch in PSTN or ISDN. MSC also handles all the functions needed to manage mobile subscribers, including registration, authentication, location updating, handovers and call routing. The roaming functionality of GSM is provided by the Home Location Register (HLR) and Visitor Location Register (VLR) together with the MSC. The MSC has no information about particular mobile stations; this information is stored in VLRs and HLRs.

Home Location Register (HLR): This functional entity is a data base in charge of the management of mobile subscribers. A Public Land Mobile Network (PLMN) may contain one or several HLRs. The number of HLRs depends on the number of mobile subscribers, on the capacity of the equipment and on the organization of the network. The HLR contains two kinds of information: subscription information and some location information that enables the charging and routing of calls toward the MSC where the MS is located (e.g., the MS roaming number, the MSC address).

Visitor Location Register (VLR): This controls mobile stations roaming in the MSC area it is in charge of. When a mobile station enters a new location area it starts its registration procedure. The MSC in charge of that area notices this registration and transfers the identity of the location area where the MS is situated to the Visitor Location Register. If this MS is not yet registered, the VLR and the HLR exchange information to allow the proper handling of calls involving the MS.

2.1 Areas of the GSM network

Different areas of GSM network and the functions associated with them are described below.

Location area (LA): The location area is defined as an area in which a mobile station may move freely without updating the location register. A location area may consist of several cells.

Network coverage area: An area in which mobile stations are able to communicate with the network.

Service area (SA): An area in which a mobile station is obtainable by another PLMN, PSTN or ISDN subscriber without the subscribers knowledge of the actual location of the mobile station within its area.

System area: The system area consists of one or more service areas with fully compatible MS-BS interfaces. The location registers of the individual service areas remain autonomous; updating of the location information is not performed when roaming mobile station moves from one service area to another.

2.2 Roaming and Handover

Roaming is the movement of the mobile terminal from one part of the network area to another part, while retaining the capability of making or receiving calls.

The handover is the action of switching a call in progress from one cell to another (or between radio channels in the same cell). The handover is used to allow established calls to continue when mobile stations move from one cell to another.

The handover can be carried out in several ways:

Intracell handover: The mobile unit is switched from one channel to another within the cell area.

BTS-BTS handover: The mobile station is switched from one Base Transceiver Station to another under the control of the same Base Station Controller.

BSC-BSC handover: The mobile station switches between BTSs as well as between the BSCs at the same time. The handover is controlled by the target BSC.

MSC-MSC handover: When switching between the BTSs and BSCs, the mobile station may also switch from one MSC (Mobile Services Switching Center) area to another. The handover is controlled by the target MSC.

2.3 Packet radio (GPRS and UMTS)

GSM in phase 2+ will be capable of handling both the conventional circuit switched transmission already introduced in GSM Phase 1 and the packet switch transmission provided by the GPRS (General Packet Radio Service). GPRS and UMTS are both based on broadband radio technologies. as distinct from GSM Phase 1. The evolution from GSM towards UMTS (3G) happens slowly, with GPRS gradually adding SGSN (Serving GPRS Support Node) and the GGSN (Gateway GPRS Support Node) elements to the GSM architecture (Figure 4). The circuit switched transmission path between the GSM BSS (Base Station Subsystem) and external networks is routed through the GSM network via the MSC (Mobile Services Switching Center) and the GMSC (Gateway MSC), while the packet switched transmission is routed via the GPRS components SGSN (Serving GPRS Support Node) and the GGSN (Gateway GPRS Support Node). The UTRAN (UMTS Terrestrial Radio Access Network) will be interconnected to this core network via two NUs (Inter Working Units), one between the Iu interface and the GSM A interface, and another between the Iu interface and the GPRS Gb interface (see Figure 4). This architecture makes it possible for both GSM and UMTS customers to be connected both to circuit switched networks (e.g., PSTN and N-ISDN) and packet switched networks (e.g. the Internet and intranets). Additionally, users should also be able to roam between GSM and UMTS networks. [8]

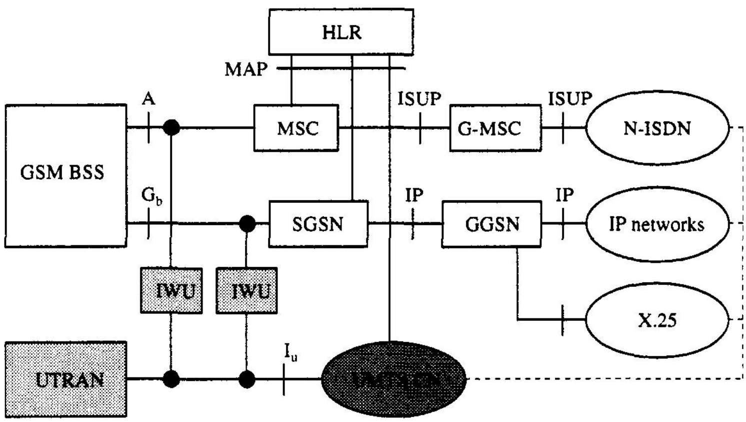

Figure 4. Evolution of GSM platform towards UMTS

3. REQUIREMENTS FOR IP MOBILITY

4G terminals will be used to access various services, including current Internet based services as well as new services targeted for 4G terminal users. A factor common to these services is that they all use IP. Thus, clearly, 4G terminals must be based on IP too.

Hybrid terminals, which support IP and some other communications technologies, such as 3G UMTS, can be used but because these hybrid terminals are also able to communicate by means of IP, 4G service providers are required to offer only IP based access to their services.

The problem with current IP schemes is that hosts are expected to be stationary, using one fixed P address all the time. In the 4G networks terminals must be able to move around without limitations set by the network architecture. The following sections describe requirements and solutions for IP Mobility.

Goals for the New Architecture

No additional software required in terminals

Some terminals will have limited memory and processing capability, and power consumption by any kind of terminal should be minimized by avoiding any unnecessary activity. Therefore, all possible mobility related functions should be executed by servers leaving as few functions as possible for terminals to execute. Terminals will cost less when no additional software and processing capability is required, and only servers will need to be upgraded.

As simple base stations as possible

As some terminals, also base stations (access points) will have limited processing capability. Increasing processing capability would cause base stations to cost more, which is highly undesirable. Bluetooth and WLAN access points have maximum coverage of roughly about 100m, depending on power and surroundings. With such a short distance coverage, a relatively large number of base stations are needed. Thus, expensive base stations could dramatically increase the overall cost of a network.

Standard RFC based protocols

To avoid service dispersion and interoperability problems all communications should be based on open, standardized, protocol definitions. Closed, proprietary protocols could easily lead to a situation where a user cannot access all the services needed because of communication protocol incompatibilities. This would cause additional costs for both service providers (providers should support many protocols instead of just one global standard) and customers. Higher costs and incompatibilities could reduce general interest in the whole 4G concept.

Scalability down to Pico cells and locations up to 10^10 users globally

New architecture should be scalable both downwards and upwards. The less coverage one base station has, the more accurately the location of terminals is known. Therefore, some operators use low coverage base stations even if high coverage base stations are available. This means that the new architecture has to operate smoothly with both low and high coverage base stations.

Although it is unfortunate, it is nevertheless a fact that at the beginning of the 4G era most of the population of the world are not immediately able to use the new technology. The new architecture should be ready for them to avoid scalability and other problems in the future.

VoIP and Web compatibility

New 4G terminals and the whole 4G architecture will support both VoIP (Voice over IP) and current World Wide Web (WWW) based services. Currently WWW provides a vast amount of services that will not be modified for 4G terminals. Therefore 4G terminals must be able to communicate with current protocols to achieve Web based service access.

Most handhelds used today are cellular phones used to direct conversation between humans. Undoubtedly, the need for direct conversations remain and 4G should support this. This can be done with the help of VoIP.

4. 4G ARCHITECTURE

The fourth generation (4G) of mobile networks will offer mobile services based on high-speed wireless connections, IP mobility, intelligent terminals, and World Wide Web type services. 4G operators are the most likely service and content providers to use different kinds of radio access technologies. Radio access can be based on private corporate LANs, public wireless LANs or mobile LANs installed on trains, airplanes, and so on. Handhelds, laptops, and mobile phones will be used to access the Internet and local services.

4G location area (4GLA) diameter can be from 100m to 1 kilometer. Figure 5 presents the 4G-network architecture. The idea is to use Session Initiation Protocol (SIP) [2]. Every home location area contains a SIP redirect server, which is responsible for maintaining the current location of users. The home SIP redirect server is analogous to HLR in GSM network architecture. When a call is made, the home SIP redirect server returns the current address of called party. The SIP client of the caller then makes another call to this particular address (or addresses - SIP redirect server can return several addresses).

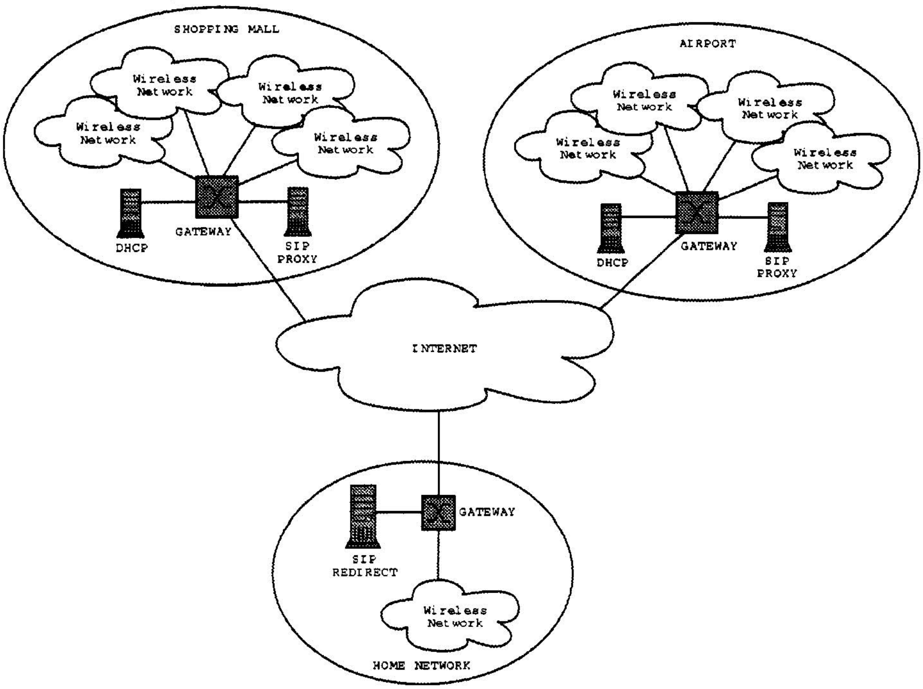

Figure 5. 4G-network architecture

Every location area where visitors are allowed, contain a visitor SIP proxy (Outbound proxy), which is analogous to VLR in GSM network. Every visitor in a foreign network registers with the home SIP redirect server if they want to be reachable. This registration can be done in several ways as described in [7]. We propose that only Outbound proxy intercept and User-initiated proxy registrations are used. Both these registration methods use Outbound proxy to forward SIP registration messages. In the case of Outbound proxy intercept, terminals send registration messages to their home network, but Outbound proxy intercepts these messages and changes the visitor address to point itself. In the case of the User-initiated proxy registration, the terminal recognizes it is visiting a foreign network and sends registration messages to the Outbound proxy that then forwards the messages to the terminal's home network. Thus, all incalls and outcalls involving a visiting user travel through the outbound SIP proxy.

The use of Outbound proxy for all registrations allows network operators to collect statistics and possible billing information and also to reduce the number of SIP registrations. In this way, several location areas can be combined to one **4G service area (4GSA)**. By using hierarchical registration [9], the home location area SIP redirect server is updated only when a terminal moves from one service area to another. SIP redirect server redirects SIP calls to the Outbound proxy that forwards calls to the terminal in its current location in the visited network.

With SIP, users can be reached globally. The same effect could be achieved using Mobile IP [5], but the triangle routing creates problems. The data transferred from a server to the user's terminal is routed via a home agent, which is a non-optimal situation. The usage of SIP enables direct point-to-point data transfer.

5. MICRO MOBILITY

The term micro mobility here refers to mobility inside a single location area. Micro mobility is needed when terminals move around using different base stations and IP sub-networks. The requirement for micro mobility means that the terminal is able to use the same IP address all the time to keep TCP (or other higher layer) connections alive when the terminal moves to another base station cell or IP sub-network. Micro mobility is controlled by the gateway router responsible for the location area.

In the 4G architecture we are working with, base stations are invisible to the IP layer of terminals. When a terminal sends an IP packet, it sends it to the gateway that all base stations are connected to. Base stations only relay layer 2 packets between terminals and the gateway. Link layer mobility is managed is done by the device drivers of terminals and base stations. Thus, if all base stations belong to the same IP sub-network, no additional IP layer mobility management is required for any part of the network. However, we are experimenting with multicast to neighbouring base stations to ensure smooth handovers.

In the case of several IP sub-networks additional mobility management is needed. Several IP sub-networks are required for scalability, thus using a single IP sub-network is not a solution in the case of large networks. The problem of using several IP sub-networks is the following: when a terminal, using IP subnet A address moves to IP subnet B area, all IP packets destined to it are routed to IP subnet A. Thus, the terminal will not receive any of these packets. The IP address can not be changed without breaking live TCP connections, so we will need a solution which allows a terminal to use an IP sub-network address of IP subnet A in all other IP subnets, too.

We are experimenting with a Proxy ARP [6] and DHCP [4] based scheme. Firstly, a terminal is granted an IP address by the means of DHCP with short lease time (lifetime) when the terminal enters to the local network. The terminal starts sending DHCP Renewing Requests when more than half of the lease time is passed. Secondly, when the gateway receives Renewing Requests it updates its routing caches. When IP packets destined to a terminal arrive to the gateway, packets are routed to the correct IP subnet based on the updated routing cache. In the handover phase the packets can be multicasted to both old and new sub-networks. The address requested to be renewed will be renewed even if the terminal has moved to another IP sub-network. Thus, the terminal is able to use several IP sub-networks without changing its IP address. Lastly, when other hosts in the network are communicating with the terminal that has moved to another IP subnet, gateway uses Proxy ARP to relay ARP requests to the moved terminal. Again, the terminal will be able to communicate even if it has moved to other IP sub-network.

This solution requires an intelligent gateway that can perform these actions. The evaluation of the scalability of the proposed solution is work for the future.

6. ROAMING

Roaming to a new 4G location area requires several functions to be accomplished in order to be able to use the services of the local network and the global network, the Internet. These functions are described in this section.

As a new terminal roams to a location area, the terminal must acquire an IP address provided by the new network. This can be done, for example, with DHCP or Mobile IP. Without such an address a terminal cannot communicate with the network.

If the previous step succeeds and no additional restrictions apply on the network, a user will then be able to use both global and local services as well as to make VoIP calls to other parties (outcalls) on the local or other networks. In order to be able to receive incoming VoIP calls (incalls) and other SIP based contacts, the terminal must also be registered with the local SIP proxy.

A user with a terminal may roam to another location area whilst having open TCP or SIP initiated connections. A TCP connection is established between two IP addresses and is used to transmit data between those two endpoints. If either of the used IP addresses becomes unreachable, then the TCP connection will die. Therefore, when roaming with open TCP connections, the terminal's IP address must remain unchanged in order to preserve the TCP connections. This is usually done with Mobile IP. However, in our experiment we manage IP addresses granted by the location area DHCP so that they can be kept in the neighbouring location areas. The smooth location update is then carried out by multicasting the IP packets arriving at the terminal to both location areas involved.

SIP initiated connections use a suitable protocol for actual data transmission. For example, VoIP uses Real-Time Protocol (RTP). In order to retain SIP initiated connections when roaming, the data stream must be redirected to the new IP address of the terminal with SIP handoff messages [10]. In the case of VoIP, for example, a new RTP connection is created using the terminal's new IP address. Since the IP address may be changed when roaming, no Mobile IP is needed.

7. CONCLUSIONS

In this paper we have presented our research on the fourth generation mobile network architecture. The architecture is analogous to GSM. SIP is used to enable incalls and outcalls even when the terminal is visiting a foreign network. Our future work will include scalability testing of the proposed micro mobility management scheme and analyzing the issue of roaming between different types of mobile networks.

REFERENCES

[1] ISO/IEC 8802-1 1:1999(E), ANSI/IEEE Std 802.11, 1999 edition.
[2] M. Handley, H. Schulzrinne, E. Schooler, and J. Rosenberg, "SIP: Session Initiation Protocol", RFC 2543, March 1999.
[3] Bluetooth SIG, Specification of the Bluetooth System – Version 1 .0 B, Specification Volume 1 & 2, December 1999.
[4] Droms, R., Dynamic Host Configuration Protocol, RFC 2131, March 1997.
[5] Perkins, C., IP Mobility Support, RFC 2002, October 1996.
[6] Postel, J., Multi-LAN Address Resolution, RFC 925, October 1984.
[7] Schulzrinne, H., SIP Registration, Internet Draft, October 2000.
[8] Martikainen O., Nyman E., Räsänen J., Koponen P., Terminal Mobility in Broadband Networks, PWC'1998, 1998.
[9] Wedlund, E., Schulzrinne, H., Mobility Support using SIP. Second ACM/IEEE International Conference on Wireless and Mobile Multimedia, August 1999.
[10] Biggs, Dean, SIP Call Control: Call Handoff, Internet Draft, January 2001.