Evaluation of Mobile Communications: From Voice Calls to Ubiquitous Multimedia Group Communications

Vasco Pereira[φ], Tiago Sousa[φ], Paulo Mendes[θ], Edmundo Monteiro[φ]
[φ] *Department of Informatics Engineering of the University of Coimbra*
{vasco, tmas, edmundo}@dei.uc.pt
[θ] *DoCoMo Euro-Labs*
mendes@docomolab-euro.com

Abstract

Today, mobile communications play a central role in the voice/data network arena. With the deployment of 3G just around the corner, new directions are already being researched. In this paper we address the evolution of mobile communications, from its first generation, 1G, to the latest 3G, and give a glimpse of the foreseeable future of Beyond 3G (B3G). In the latter case, we evaluate B3G scenarios in terms of network heterogeneity, migration to B3G, mobility over heterogeneous networks, and real-time multimedia communications in mobile environments.

Key Words: Mobile Networks, Beyond 3G networks, Heterogeneous networks, Real-time communications

1. Introduction

From the early analog mobile generation (1G) to the last deployed Third Generation (3G) the paradigm has changed. The new mobile generation not only tries to improve the quality of voice communications, but also tries to give the user access to a new global communication reality. The aim is to reach communication ubiquity (anytime, anywhere) and to provide users with a new set of services.

The growth in the number of mobile subscribers over the last years led to a huge utilization of voice-oriented wireless telephony. From a number of 214 million subscribers in 1997 to 1.162 million in 2002 [1], it is predicted that by 2010 there will be more than 1.700 million subscribers worldwide [2]. At same time, mobile multimedia is also growing at a fast rate, as new terminals, with color screens and digital cameras, gain popularity. It is now time to explore new demands and to find new ways to extend the mobile concept. The first steps have already been taken by the 2.5G, which gave users access to a data network, e.g., Internet access, and *Multimedia Message Service* (MMS). However, users and applications are constantly demanding more communication power. As a response to this demand a new generation with new standards has been developed - 3G, based on new mobile technologies like the *Universal Mobile Telecommunications System* (UMTS). In spite of the big initial euphoria that evolved this technology, commercial use of 3G networks is still today very limited. The first deployment was called *Freedom of Mobile Multimedia Access* (FOMA) and was released by NTT DoCoMo in Japan in 2001, using international standard IMT-2000, with great success. Nowadays some other providers are starting to make 3G services available, namely Hutchinson in Austria and Italy, Vodafone in Portugal, Germany, Spain and Italy, and TMN in Portugal.

In the last years, benefiting from 3G constant deployment delays, many new mobile technologies gained popularity. Now, all this new technologies (e.g. UMTS, Wi-Fi, Bluetooth) claim for a convergence that can only be achieved by a new mobile generation - *beyond 3rd generation* (B3G). This new mobile generation must allow the coexistence of different mobile technologies and provide a differentiate set of services to the end user, which should be kept agnostic of all the network diversity. However, the provision of differentiated services over heterogeneous mobile networks encompasses several challenges. One of these challenges is the deployment of multimedia group communications in a mobile environment with different access technologies.

The main contribution of this paper is twofold. First, we evaluate, in Section 2, the evolution of mobile communications to its present generation – 3G, focusing the reasons that motivated the successive development of different generations. Second, we analyze, in Section 3, future B3G mobile communication systems, namely the coexistence of heterogeneous mobile networks, and the migration to B3G systems.

Moreover, Section 3 also analyzes two core B3G functionalities: the mobility in a heterogeneous environment, and the support for real-time multimedia group communications. Finally Section 4 concludes our contribution.

2. Evolution from Homogeneity to Heterogeneity

The first generation of mobile networks, with its focus on voice communication, started a new era. Today voice and data live together in mobile networks with the latter gaining more and more importance. The use of a specific technology to communicate is also giving place to terminals that support many access methods. Analyze how this path was built is the aim of this section.

Clearly, in the history of mobile communications occurred some specific evolutions that stimulated and transformed the way we communicate, leading to new communication paradigms and recently to new generations of mobile communications. This can be viewed if we analyze the key evolutions that mainly contribute to the emergence of new generations.

In what concerns mobile communications, the key transition from the first to the second generation was due to digitalization, i.e., while 1G was based on analog technology, 2G was based on digital signal processing techniques. The transition from 2G to 2.5G, was based upon the introduction of data service and packet switching methods. This new methods enabled the introduction of new services, namely the Internet. With 3G this scenario was further improved with a focus in new services, instead of only improving technology to provide higher data rate and broader bandwidth.

All this evolution will be better explained in the remaining of this section.

2.1. The First Mobile Generations – from 1G to 2.5G

The first operational cellular communication system was deployed in the Norway in 1981 and was followed by similar systems in the US and UK. These first generation systems provided voice transmissions by using frequencies around 900 MHz and analog modulation.

The *second generation* (2G) of the wireless mobile network was based on low-band digital data signaling. The most popular 2G wireless technology is known as *Global Systems for Mobile Communications* (GSM). The first GSM systems used a 25MHz frequency spectrum in the 900MHz band. The *Frequency Division Multiple Access* (FDMA), which is a standard that lets multiple users access a group of radio frequency bands and eliminates interference of message traffic, is used to split the available 25MHz of bandwidth into 124 carrier frequencies of 200 kHz each. Each frequency is then divided using a *Time Division Multiple Access* (TDMA) scheme into eight timeslots and allows eight simultaneous calls on the same frequency. This protocol allows large numbers of users to access one radio frequency by allocating time slots to multiple voice or data calls. TDMA breaks down data transmission, such as a phone conversation, into fragments and transmits each fragment in a short burst, assigning each fragment a time slot. With a cell phone, the caller does not detect this fragmentation. Today, GSM systems operate in the 900MHz and 1.8 GHz bands throughout the world with the exception of the America Continent where they operate in the 1.9 GHz band. Within Europe, the GSM technology made possible the seamless roaming across all countries.

While GSM technology was developed in Europe, *Code Division Multiple Access* (CDMA) technology was developed in North America. CDMA uses spread spectrum technology to break up speech into small, digitized segments and encodes them to identify each call. CDMA distinguishes between multiple transmissions carried simultaneously on a single wireless signal. It carries the transmissions on that signal, freeing network room for the wireless carrier and providing interference-free calls for the user. Several versions of the standard are still under development. CDMA promises to open up network capacity for wireless carriers and improve the quality of wireless messages and users access to the wireless airwaves. Whereas CDMA breaks down calls on a signal by codes, TDMA breaks them down by time. The result in both cases is an increased network capacity for the wireless carrier and a lack of interference for the caller. While GSM and other TDMA-based systems have become the dominant 2G wirelesses technologies, CDMA technology is recognized as providing clearer voice quality with less background noise, fewer dropped calls, enhanced security, greater reliability and greater network capacity.

The 2G wireless networks mentioned above are also mostly based on circuit-switched technology, are digital and expand the range of applications to more advanced voice services. 2G wireless technologies can handle some data capabilities such as fax and short message service at the data rate of up to 9.6 kb/s, but it is not suitable for web browsing and multimedia applications.

So-called ‘2.5G’ systems enhance the data capacity of GSM and mitigate some of its limitations. These systems add packet data capability to GSM networks, and the most important technologies are *General Packet Radio Service* (GPRS) and *Wireless Application Protocol* (WAP). WAP defines how Web pages and similar data can be passed over limited bandwidth wireless channels to the small screens being built into new mobile telephones. At the next lower layer, GPRS defines how to add IP support to the existing GSM infrastructure. GPRS provides both the means to aggregate radio channels for higher data bandwidth and the additional servers required to off-load packet traffic from existing GSM circuits. It supplements today's circuit switched data and short message service. Theoretical maximum speeds of up to 171.2 kb/s are achievable with GPRS using all eight timeslots at the same time, about ten times as fast as current circuit switched data services on GSM networks. However, it should be noted that it is unlikely that a network operator will allow all timeslots to be used by a single GPRS user. Additionally, the initial GPRS terminals (phones or modems) are supporting only one to four timeslots. The bandwidth available to a GPRS user will therefore be limited.

Meanwhile, developers are focusing on the much-hyped *third generation* (3G) of wireless systems, where beyond increased data rates a complete new set of services will be available. All these wireless technologies are summarized in Table 1.

Table 1 – Transport Technologies (adapted from [3])

	Transport Technology	**Description**	**Typical Use / Data Transmission Speed**	**Pros/cons**
2G	TDMA	Time Division Multiple Access	Voice and data up to 9.6 kb/s.	+ Low battery consumption − One-way transmission; slow speed
	GSM	Global System for Mobile Communications	Voice and data. This European system uses the 900 MHz and 1.8 GHz frequencies. In the United States it operates in the 1.9 GHz PCS band up to 9.6 kb/s.	+ Popularity; roaming in about 180 countries − GSM's short messaging service only transmits one-way, with maximum 160 characters long
	CDMA	Code Division Multiple Access is a 2G technology developed by Qualcomm that is transitioning to 3G	TIA/EIA IS-95 (Telecommunications Industry Association / Electronic Industries Association Interim Standard - 95) defines the first CDMA. Supports voice and data up to 14.4 Kb/s.	+ More capacity than TDMA − Fewer subscribers than TDMA
2.5G	GPRS	General Packet Radio Service - supports data packets	Data Up to 115 kb/s; the AT&T Wireless GPRS network will transmit data at 40 kb/s to 60 kb/s.	+ Messages not limited to 160 characters
3G	EDGE	Enhanced Data Rates for Global Evolution	Data Up to 384 kb/s.	+ Temporary solution for operators unable to get W-CDMA licenses; offers higher-speed mobile-data access, serve more mobile-data customers, and free up GSM network capacity to accommodate additional voice traffic
	W-CDMA (UMTS)	Wideband CDMA (also known as Universal Mobile Telecommunications System -UMTS).	Voice and data. UMTS is being designed to offer speeds from 144 kb/s (for users in fast-moving vehicles) to 2 Mb/s, initially. Up to 10 Mb/s by 2005, according to designers.	+ Likely to be dominant outside the United States, and therefore good for roaming globally − Commitments from U.S. operators are currently lacking
	CDMA2000 1xRTT	1xRTT is the first phase of CDMA2000	Voice and data Up to 144 kb/s.	+ Proponents say migration from TDMA to CDMA2000 is simpler than to W-CDMA, and that spectrum use is more efficient − W-CDMA will likely be more common in Europe
	CDMA2000 1xEV-DO	Delivers data on a separate channel	Data up to 2.4 Mb/s.	(see CDMA2000 1xRTT above)
	CDMA2000 1xEV-DV	Integrates voice and data on the same channel	Voice and data up to 2.4 Mb/s.	(see CDMA2000 1xRTT above)

2.2. The Third Mobile Generation - 3G

All 2G wireless systems are voice-centric. GSM includes *Short Message Service* (SMS), enabling text messages of up to 160 characters to be sent, received and viewed on the handset. Most 2G systems also support some data over their voice paths, but at painfully slow speeds usually 9.6 kb/s or 14.4 kb/s (CDMA). So in the world of 2G, voice remains king while data is already dominant in wired communications. However, both wired and wireless communications, are affected by the rapid growth of the Internet.

Planning for 3G started in the 1980s. Initial plans focused on multimedia applications such as videoconferencing for mobile phones. When it became clear that the real killer application was the Internet, 3G thinking had to evolve. As personal wireless handsets become more common than fixed telephones, it is clear that personal wireless Internet access will follow and users will want broadband Internet access wherever they go.

Today's 3G specifications call for 144 kb/s while the user is moving quickly, 384 kb/s for pedestrians, and up to 2 Mb/s for stationary users. This is a big step up from 2G bandwidth using 8 to 13 kb/s per channel to transport speech signals.

The second key issue for 3G wireless is that users will want to roam worldwide and stay connected. Today, GSM leads with almost global roaming. Because of the pervasiveness of GSM, users can get comprehensive coverage in Europe, parts of Asia and some U.S. coverage. A key goal of 3G is to make this roaming capacity universal.

A third issue for 3G systems is capacity. As wireless usage continues to expand, existing systems are reaching limits. Cells can be made smaller, permitting frequency reuse, but only to a point. The next step requires new technology and new bandwidth.

The *International Mobile Telecommunications-2000* (IMT-2000) is the official *International Telecommunication Union* (ITU) name for 3G and it is an initiative intended to provide wireless access to global telecommunication infrastructure through both satellite and terrestrial systems, serving fixed and mobile phone users via both public and private telephone networks. GSM proponents put forward UMTS, an evolution of GSM, as the road to IMT-2000. Alternate schemes have come from the US, Japan and Korea. Each scheme typically involves multiple radio transmission techniques in order to handle evolution from 2G. Agreeing on frequency bands for IMT-2000 has been more difficult and the consensus included five different radio standards and three widely different frequency bands. They are now all part of IMT-2000. To roam anywhere in this “unified” 3G system, users will likely need a quintuple-mode phone able to operate in an 800/900 MHz band, a 1.7 to 1.9 GHz band and a 2.5 to 2.69 GHz band.

Third-generation wireless also requires new infrastructure. There are two mobility infrastructures in wide use. While GSM has the mobile access protocol, GSM-MAP, the North American infrastructure uses the IS-41 mobility protocol. These protocol sets define the messages passed between home location registers and visitor location registers when locating a subscriber, and the messages needed to deal with hand-offs as a subscriber moves from cell to cell. 3G proponents have agreed on an evolution path so that existing operators, running on either a GSM-MAP or an IS-41 infrastructure, can interoperate. But, the rest of the landline infrastructure to support IMT-2000 is ready, and may be used in the near future. The IMT-2000 family is illustrated in Figure 1.

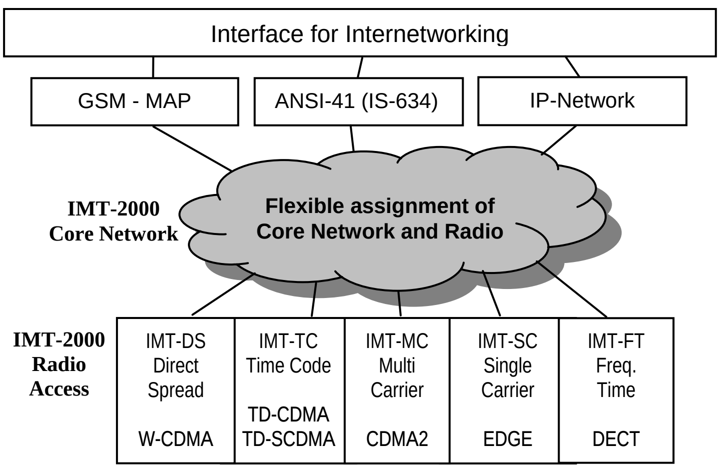

Figure 1 – IMT 2000 Family (adapted from [4])

UMTS uses the radio technology called *Wideband Code Division Multiple Access* (W-CDMA). W-CDMA is characterized by the use of a wider band than CDMA. W-CDMA has additional advantages of high transfer rate, and increased system capacity and communication quality by statistical multiplexing. W-CDMA utilizes efficiently the radio spectrum to provide a maximum data rate of 2 Mb/s.

With the advent of mobile Internet access, suddenly the circuit-based backhaul network from the base station and back has to significantly change. 3G systems are IP-centric and will justify an all-IP infrastructure.

There will be no flip to 3G, but rather an evolution and, because of the practical need to re-use the existing infrastructure and to take advantage of new frequency bands as they become available, that evolution will look a bit different depending on where you are.

The very definition of 3G is now an umbrella, not a single standard. However, the industry is moving in the right direction towards a worldwide, converged, network. Meanwhile, ever-improving *Digital Signal Processing* (DSP) will allow multi-mode, multi-band telephones that solve the problem of diverse radio interfaces and numerous frequency bands. When one handset provides voice and data anywhere in the world that will be 3G no matter what is running behind the scenes, although it is expectable that another air interface, more powerful than 3G, will arise.

3. Evaluation of Heterogeneous Mobile Communications Systems

In this section we introduce heterogeneous mobile networks B3G, stating their objectives and primary characteristics. Essential questions as the migration beyond 3G, mobility in heterogeneous environments and real-time multimedia communications are also addressed.

3.1. Heterogeneous Networks beyond 3G

The objective of the 3G systems was to develop a new protocol and new technologies to further enhance the mobile experience. In contrast, a new framework B3G will try to accomplish new levels of user experience and multi-service capacity by integrating the different mobile technologies, such as GSM, GPRS, IMT-2000, Wi-Fi and Bluetooth (Figure 2). In spite of different approaches, each resulting from different visions of the future platform currently under investigation, the main objectives of B3G networks can be stated as being ubiquity, multi-service platform, and low cost per bit.

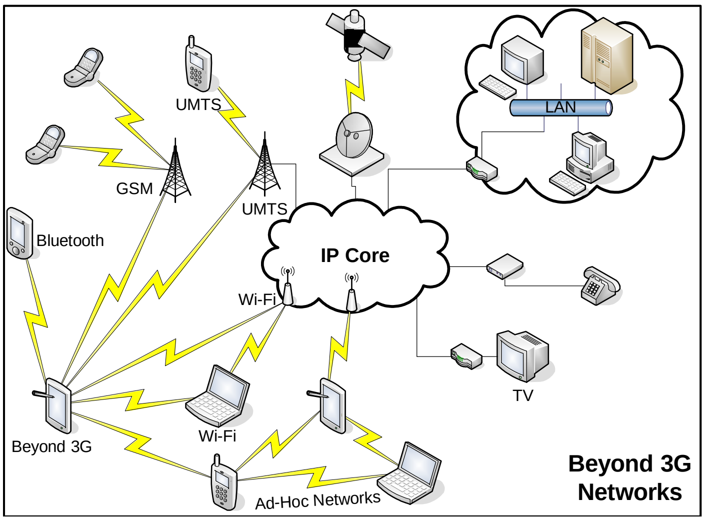

Figure 2 – B3G Networks

Ubiquity means that mobile networks must be available to the user, any time, anywhere. To accomplish this objective, services and technologies must be standardized in a worldwide scale. Furthermore, the services to be implemented should be available not only to humans as have been the rule in previous systems, but also to everything that needs to communicate. In this new world we can find transmitters in our phone to enable voice and data communications (e.g. high bandwidth Internet access, multimedia transmissions), in our wrist, to monitor our vital signs, in the packages we send, so that we always know their location, in cars, to always have their location and receive alerts about an accident, in remote monitor/control devices, in animals to track their state or location, or even in plants. Based on this view, NTT DoCoMo, that has already a wide base of 3G mobile users, estimates the number of mobile communication terminals to grow in Japan from the actual 82.2 million to more than 500 million units by 2010 [5].

A **multi-service platform** is an essential property of the new mobile generation, not only because it is the main reason for user transition, but also because it will give telecommunication operators access to new levels of traffic. Voice will loose its weight in the overall user bill with the raise of more and more data services.

Low cost per bit is an essential requirement in a scenario where high volumes of data are being transmitted over the mobile network. With the actual price per bit, the market for the new high demanding applications, which transmit high volumes of data, such as video, is not possible to be established. In networks with high volume of traffic, cost per bit should be between 1/10 and 1/100 of 3G systems [5].

To achieve the proposed goals, a flexible network architecture that allows the convergence of different radio access technologies, must be created. The overall network must provide high bandwidth, from 50-100 Mb/s for high mobility users, to 1 Gb/s for low mobility users [2], fast handoffs, an efficient delivery system over different wireless technologies, and an intelligent method of choosing the most suitable wireless access from the available ones. Also necessary is a *Quality of Service* (QoS) framework that enables fair and efficient medium sharing among users with different QoS requirements, supporting the different priorities of the services to be deployed. The core of this network should be based upon the Internet Protocol version 6 – IPv6. IP protocol is the only that has already demonstrated its validity via a worldwide acceptance as the basic technology for the Internet (with IPv4) and has now been improved with a suitable number of Internet addresses (in opposition to IPv4), together with self-configuration and neighbor discovery capabilities.

3.2. Migration beyond 3G

The fact that B3G mobile networks intend to integrate almost every wireless standard already in use, enabling its simultaneous use and interconnection poses many questions not yet answered. The research areas that present key challenges to migrate current systems B3G are many but can be summarized in the following: Mobile Station, System and Service [6].

To be able to use B3G mobile networks a new type of mobile terminals must be conceived. These terminals must adapt seamlessly to multiple wireless networks, each with different protocols and technologies. Auto reconfiguration will also be needed so that terminals can adapt to the different available services. This adaptation may imply that it must download automatically configuration software from networks in range. Moreover, terminals must be able to choose from all the available wireless networks the one to use with a specific service. To do this it must be aware of specifications of all the networks in terms of bandwidth, QoS supported, costs and respect to user preferences.

Terminal mobility will be a key factor to the success of B3G networks, as it was for previous generations. The new terminals must be able to provide wireless services anytime, everywhere. This implies that roaming between different networks must be automatic and transparent to the user. There are two major issues in terminal mobility, location management and handoff management [6]. Location management deals with tracking user mobility, and handling information about original, current and (if possible) future cells. Moreover, it must deal with authentication issues and QoS assurances. Handoff management primary objective is to maintain the communications while the terminal crosses wireless network boundaries. In addition, B3G networks, in opposition to the other mobile generations, must deal with vertical and horizontal handoffs, i.e., a B3G mobile client may move not only between cells of the same wireless network, but also between different types of wireless networks, e.g., GSM and Wi-Fi. Furthermore, many of the services available in this new mobile generation like videoconference and multimedia streaming have tight time constraints and QoS needs that must not be perceptible affected by handoffs. To avoid these problems new algorithms must be researched and a prevision of user mobility will be necessary, so as to avoid broadcasting at the same time to all adjacent antennas what would waste unnecessary resources. Another major problem relates to security, since B3G pretends to join many different types of mobile technologies. As each standard has its own security scheme, the key to B3G systems is to be highly flexible.

Services also pose many questions as B3G users may have different operators to different services and, even if they have the same operator, they can access data using different network technologies. Actual billing using flat rates, time or cost per bit fares, may not be suitable to the new range of services. At the same time it is necessary that the billing can be well understood by operator and client. A broker system would be advisable to facilitate the interaction between the user and the different service providers.

Another challenge is to know, at each time, where the user is and how he can be contacted. This is very important to mobility management. A user must be able to be reached wherever he is, no matter the kind of terminal that is being used. This can be achieved in various ways one of the most popular being the use of a mobile-agent infrastructure. In this framework, each user has a unique identifier served by personal mobile agents that make the link from users to Internet.

3.3. Mobility in Heterogeneous Environments

The IP protocol was designed to provide communication services to fixed stations or, at most, to nomadic stations, which have discrete mobility as it is provided by the *Point-to-Point Protocol* (PPP) or by the *Dynamic Host Configuration Protocol* (DHCP), where stations reconnect at different network attachment points, not requiring a continuous session between them. Now, the goal is to make IP the basic technology platform for networks B3G. The main requirement is to extend the concept of IP mobility to a scenario where users are able to travel between different networks attachment points while maintaining a continuous communication session. Some steps have already been made in this direction, like the solutions that where proposed by the *Internet Engineering Task Force* (IETF) for different mobility scenarios, i.e., for macro-mobility as for micro-mobility.

A **macro mobility** framework is required to allow roaming between different access networks. The aim is to provide users with a continuous session, while traveling between different access networks and between different technologies. Users should not need to restart applications nor manually update any information, and applications should not need to be mobile aware.

To enable this feature in an IP network, IETF has created Mobile IP, as defined in RFC3344 for IPv4 [7] and proposed in an Internet Draft for IPv6 [8]. Basic IP requires a different IP address for a different network attachment point, to enable the correct routing of data packets. But, to enable mobility and maintain existing transport layer connections, while traveling between different networks attachment points, the IP address must remain the same. Mobile IP provides a seamless macro mobility solution that permits moving between different networks that can have different types of access technologies. To do this, in Mobile IP each mobile node has two different addresses, a fixed Home Address and a dynamic Care-of Address. In the home network, a mobile agent – Home Agent, receives all the traffic for the mobile client from the Correspondent Node (the node with which the mobile node communicates). When the client exits its home network the Home Agent tunnels all the traffic to another mobile agent belonging to the new network where the mobile node has attached – Foreign Agent, using the current Care-of Address. The tunneling is made by rebuilding the IP header of the packet received by the Home Agent with the new address, encapsulating the original data packet. A new Care-of Address is registered in the client's Home Agent whenever a handoff takes place, with the mobile node moving to a different foreign network. The process described for Mobile IP works both for IPv4 and IPv6, with some differences (e.g. use of IPv6 routing header instead of encapsulation of the packet).

The big strength of Mobile IP is that it enables an easy and transparent way to deploy mobility in IP networks, by the use of a simple protocol. However, this simplicity and transparency don't come without some costs. The first is the suboptimal routing implied, as the Home Agent is a fixed redirection point for every IP packet received by the mobile node, even if there is a shorter routing path between source and destination. This can also happen with packets sent by the mobile node in case of a bidirectional tunnel is needed, as the foreign network may not permit the direct forwarding of packets from the mobile node (using its home address as source address), because the source address does not belong to the network. Suboptimal routing leads to delays and overload of some links connecting to the home network (where the tunnel is created), even if other paths are available and free. Also, the need to redirect packets to the home network, that can be far away, can also lead to performance issues like latency and overhead. Furthermore, as the home network is an essential item for any communication of the mobile node, it is also a point of failure, decreasing the robustness of the entire protocol. Handoff problems can also occur if a mobile node is in constant movement, changing its point of attachment frequently. The latency and overhead of being always updating the Home Agent may lead to a severe amount of packet losses while the operation occurs. Other problems are the additional tunneling in IPv4 and routing headers in IPv6, that lead to additional per packet overheads. Also, tunneling can conflict with firewall and IPSec Security Policies (IPv4) [9]. Finally Mobile IP does not guarantee mobile node privacy, as it is possible for the correspondent node to know where the mobile node is.

All these problems led to several enhancements and extensions to the original Mobile IP protocol, such as route optimization to prevent suboptimal routing, and Micro Mobility extensions (addressed next) that were designed to mask the rapid mobility from distant nodes.

Another approach to macro mobility is presented by the *IP-based IMT network Platform* (IP^2) [10] mobility management scheme to be implemented B3G, which forms a full IP transport network to efficiently handle increasing multimedia traffic. IP^2 addresses the performance and transmission quality of the network, by taking advantage of mobility characteristics, choosing optimal paths, and efficiently using network resources while providing low loss and latency. Moreover, IP^2 enhances the privacy of mobile devices, since correspondent nodes are not aware of the Care-of Address used by mobile devices.

Micro mobility technologies act in a LAN scope or in a local domain. They present some solutions to extend and enhance the Mobile IP technology, by offering fast and seamless handoff control in limited geographical areas, and IP paging in support of scalability and power conservation. The general purpose of micro mobility protocols is to ensure that packets arriving from the Internet and addressed to mobile hosts are forwarded to the appropriate wireless access point in an efficient manner. To do this, micro mobility protocols maintain a location database that maps mobile host identifiers to location information. This is an important issue in micro mobility, the management of location information, whereby a Mobile Node is able to continue receiving packets in a new subnet before the corresponding changes in either the Home Agent or Correspondent Node binding [11].

The Mobile IPv6 technology has two important drawbacks: the latency of handover and data loss. In view of that, several proposals have been developed such as Cellular IP [12], Hawaii [13], *Hierarchical Mobile IPv6* (HMIPv6) [14], and the *Fast Handovers for Mobile IPv6* (FMIPv6) [15]. Others recent approaches are the Localized Mobility Management Goals [11] and the Mobile IPv6 Fast Handovers for 802.11 Networks [16]. Here, only the HMIPv6 and FMIPv6 protocols are discussed.

Both protocols, HMIPv6 and FMIPv6, constitute enhancements or extensions to the standard Mobile IPv6 Protocol. HMIPv6 is intended to reduce the handoff latency by reducing the amount of signaling between the Mobile Node, its Correspondent Nodes and its Home Agent. The development of a new network node, called *Mobility Anchor Point* (MAP), which acts like a local Home Agent, helps to improve the performance of Mobile IPv6 in terms of handoff speed by preventing the global handoff signaling, outside of local domain. In practice, the MAP introduces a new level of mobility in any place in the network hierarchy. On the other hand, FMIPv6 reduces losses with the tunneling and buffering mechanism. However, does not reduce overall latency, since the bindings of the Home Agent and correspondent node are still needed. In addition, FMIPv6 provides support for pre-configuration of link information (such as the subnet prefix) in the new subnet while the mobile node is still attached to the old subnet.

Despite these enhancements, both technologies suffer from some limitations. In the HMIPv6 protocol the routing optimization is sacrificed, namely due the tunneling existent between the MAP and the mobile node. In small map domains this is not a serious problem but for larger domains this solution introduces QoS deteriorations and node congestion. For this, it is said that hierarchical architectures compromises the routing optimization. With FMIPv6, the problems are related to scalability and QoS support. Moreover, reducing packet loss and delay during handovers is of major importance in mobile networks. However, either HMIPv6 or FMIPv6 have failed to minimize both of these factors.

Several studies [17,19] have proposed hybrid approaches, where the FMIP and HMIP integration were implemented and evaluated. The Fast HMIPv6 scheme (F-HMIPV6) [17] aims to integrate the low delay and low loss characteristics of HMIPv6 and FMIPv6, respectively. However, a simple integration of the two mentioned schemes will result in packet disorder [18] as some of the forwarded packets from the previous access router to the new Care-of Address may arrive after the new packets arriving directly from MAP to the new access router. This is because the propagation time from MAP to the new access router is shorter than from the previous access router to the new one. Hence, in the F-HMIPv6 approach the tunneling between the previous and new access routers is replaced by a tunneling between the MAP and the new access router, which eliminates packet disorder. However, this approach involves high signaling overheads during the handover epoch.

Gwon et al.[19] proposes a hybrid of FMIPv6 and HMIPv6, called FFMIPv6, to address both micro-mobility and macro-mobility issues. At the micro-mobility level, the scheme is similar to FMIPv6. At the macro-mobility level, it employs FMIPv6-like scheme between old MAP and the new access router. Thus, the tunnel extends from the old MAP to the new access router. However, they fail to reduce the signaling overheads and temporary tunnel overheads as experienced by FMIPv6 and HMIPv6.

The purpose of the questions presented in this section is to clarify some problems that current mobility technologies still have to solve. Important evidence is that the efficiency of global mobility depends upon a well-engineered integration of macro and micro mobility schemes. Moreover, micro mobility schemes have still some limitations to provide mobility with low delay and losses. It is our perception that seamless micro mobility can be achieved by the combination of some key features, namely the ability to collect and process information about the bandwidth requirements of applications and the load of access networks in order to make the most suitable handover decision.

3.4. Real-time Communications in Mobile Environments

Seamless mobility depends not only upon the capability of the mobility management schemes to support handovers with low packet losses and delays, but also upon the capability of ensuring always good end-to-end quality levels to multimedia sessions. The major requirement to achieve such a goal is the implementation of a QoS framework that enables fair and efficient medium sharing among users with different QoS requirements, supporting the different priorities of the services to be deployed.

The most referred models to ensure QoS are the *Integrated Services* (IntServ) [20] and *Differentiated Services* (DiffServ) [21] IETF models. Some work has been done to adapt the DiffServ model to provide QoS in heterogeneous networks [22,23]. It is expected that in a heterogeneous environment these two models will co-exist, which requires a flexible end-to-end control of resources. This is, a signaling protocol capable of operate in IntServ and DiffServ domain is required. Although the *ReSerVation Protocol* (RSVP) [24] was developed to provide a control plane for the IntServ model, there have been some proposals to adapt its behavior to signal in a DiffServ domain [25].

In a mobile environment, mobile hosts need to reserve resources in several end-to-end paths that they will use during the lifetime of their multimedia session. However, RSVP is not adequate to make such reservations for mobile hosts. To overcome this problem, Talukdar et al. propose a new reservation protocol, called MRSVP [26], to support integrated services in a network with mobile hosts. However, MRSVP is not suitable to support differentiated services.

The *Next Steps in Signaling* (NSIS) working group of IETF aims to standardize an IP signaling protocol with QoS signaling as the first use case. The intention is to re-use RSVP, while simplifying it and applying a more general signaling model. Although this working group does not aim to propose NSIS for mobile scenarios, some analyses have been made to understand how NSIS can be designed in order to support mobility [27].

One of the decisions made within the NSIS working group, to make the proposed generic signaling protocol simpler than RSVP, was to exclude the support for multicast. This characteristic makes NSIS unsuitable to support multimedia group communications.

The need to support QoS for multimedia group communications in B3G is a reality, since services such as streaming are already supported in 3G networks. For instance, NTT DoCoMo already provides live video streaming to its *Freedom of Multimedia Access* (FOMA) video-enabled mobile phones. The used platform enables video and data, transported using the *Realtime Transport Protocol* (RTP) [28]. The system uses MPEG4 encoding, which enables distribution to heterogeneous devices. Currently these streaming services are based on unicast, being the use of multicast only dependent on the standardization of the *Multimedia Broadcast Multicast Service* (MBMS) in 3GPP.

To support QoS for multimedia group communications in B3G networks, it is assumed that streaming services in core networks will be based on IP multicast, since this technology is more suitable for streaming than unicast, mainly due to its better usage of network resources. Moreover, the combination of multicast and unicast to cope with access networks and devices that may not be multicast aware is required. Therefore, B3G QoS will require networks capable of providing self-configuration of resources and distributed admission control to cope with different patterns of usage in a heterogeneous environment, and multicast of multimedia content that fulfils QoS requirements. In the University of Coimbra we are working on the PHB-D3 (*Per-Hop-Behaviour Dynamic Degradation Distribution*) [29] and UC-QoSR (*UC QoS Routing*) [30] proposals to overcome this limitation. Moreover, future mobile networks must be able to provide a fair allocation of resources among multimedia streams in order to increase the number of clients with a good level of quality reception in a cost effective manner for the operator. [31,32].

In order to fulfill all the enumerated requirements to support QoS for multimedia group communications in B3G networks, an architecture called *QoS Architecture for Mobile Multicast Multimedia Services* (Q_3M) [33] is being developed. Q_3M aims to allow users to seamlessly move while sending and receiving multimedia streams, without having to include new functionalities on terminals.

Q_3M also aims to bring some benefits to mobile operators, namely by providing robust and reliable streaming services, and self-organized control, with efficient usage of network resources.

4. Conclusion

Mobile communications are growing in a fast pace, with increasing weight in the communications industry, and in the daily life of millions of users. Also increasing is the number of technologies that promises to achieve a new level of services, extending the mobile experience to new limits.

In this paper we described the mobile evolution to its present generation (3G), and analyze the future B3G networks. In the first topic, the paper presents a comparison between the successive mobile technologies, and explains the technological limitations that triggered the mobile evolution to overcome current technological limitations. In the second topic we analyze the essential characteristics of future B3G networks, migration problems, as well as foreseeable new applications, of which we focused real-time applications and group communication services.

Based on the characteristics and applications analyzed, the future B3G networks can be expected to be highly heterogeneous (integrating most of the different wireless access technologies available), ubiquitous, multi-service and enabling a low per bit cost. Ubiquity will be expressed as the capacity to connect to all available wireless networks, using different access technologies such as WLAN, UMTS, GSM and Bluetooth. As for the multi-services platform, the objective will be the increase of data services and the extension of existing one-to-one services (real-time or not) to group communications.

In the future other needs are expected to emerge, as the quest for empathetic communications continuous. All in-one terminals that provide all kind of services, linking to different networks, including the new ad-hoc and sensor networks, as well as *Body Area Networks* (BANs), will be able to support a new range of innovative services. In this scenario, new services and technologies will surely start to emerge, with communications becoming more human oriented, achieving new levels of reality, like with the addition of virtual reality and inclusion of other physical sensations besides vision and sound. Also, there will be a raise of implicit communication activations (e.g. monitoring health parameters) versus explicit communications.

However, the migration to these new scenarios pose many problems namely how to approach the migration from 3G to beyond, problems of mobility over heterogeneous networks, and the requirements of real-time communications in mobile environments. In what concerns migration we can outline problems such as the transition to an all-IP core network, security and accounting issues. In mobility the main concerns are the support of seamless services with different types of handoffs to be made while the user crosses different access points and technologies.

Based on our evaluation of B3G characteristics and needs, the actual technologies and standards do not fulfill all the needs for B3G networks. QoS guarantees, multicast support, efficient self-managed network with self-configuration of resources, group communications support, seamless and fast handoffs between different technologies are identified as the main strong needs. To achieve the described B3G scenario an improved architecture solution must be developed and deployed.

5. References

[1] ITU, “Mobile cellular, subscribers per 100 people”, International Telecommunication Union Statistics, 2002
http://www.itu.int/ITU-D/ict/statistics/at_glance/cellular02.pdf

[2] Kim, Y., Jeong, B.J., Chung, J., Hwang, C., Ryu, J.S., Kim, K., Kim, Y.K., “Beyond 3G: Vision, Requirements, and Enabling Technologies”, IEEE Communications Magazine, March 2003, pp. 120-124

[3] “2G – 3G Cellular Wireless data transport terminology”, Arc Electronics
www.arcelect.com/2G-3G_Cellular_Wireless.htm

[4] Schiller, J., “Mobile Communications”, slides 2002
http://www.jochenschiller.de/

[5] Tachikawa, Keiji, “A perspective on the Evolution of Mobile Communications”, IEEE Communications Magazine, October 2003, pp. 66-73.

[6] Hui, Suk Yu, and Yeung, Kai Hau, “Challenges in the Migration to 4G Mobile Systems”, IEEE Communications Magazine, December 2003, pp. 54-59.

[7] Perkins, C., “IP Mobility Support for IPv4”, Request for Comments (Standards Track) RFC 3344, IETF Network Working Group, August 2002.

[8] Johnson, D., Perkins, C., Arkko, J., “Mobility Support in IPv6”, Internet-Draft, IETF Mobile IP Working Group, June 2003.

[9] Henderson, T., “Host Mobility for IP Networks: A Comparison”, IEEE Network, November/December 2003, pp. 18-26
[10] H. Yumiba, K. Imai, and M. Yabusaki, “IP-Based IMT Network Platform”, IEEE Personal Communication Magazine, Vol. 8 (5), October 2001
[11] Williams, C., “Localized Mobility Management Goals”, draft-ietf-mipshop-lmm-requirements-02, IETF, February 2003.
[12] Valkó, A., “Cellular IP: A New Approach to Internet Host Mobility,” ACM SIGCOMM Comp. Commun. Rev., vol. 29, no. 1, Jan. 1999, pp. 50–65.
[13] Ramjee, R., et al., “HAWAII: A Domain-Based Approach for Supporting Mobility in Wide-area Wireless Networks,” Proc. IEEE Int’l, Conf. Network Protocols, 1999.
[14] Soliman, H., et al., “Hierarchical Mobile IPv6 Mobility Management (HMIPv6)”, draft-ietf-mobileip-hmipv6-07.txt, IETF. October 2002
[15] Koodli, R., et al., “Fast Handovers for Mobile IPv6”, draft-ietf-mobileip-fast-mipv6-06.txt, a work in progress, IETF, April 2003
[16] McCann, P., “Mobile IPv6 Fast Handovers for 802.11 Networks,” draft-ietf-mipshop-80211fh-00.txt, IETF, February 2004
[17] Jung, H., et al., “Fast Handover for Hierarchical MIPv6 (F-HMIPv6)”, draft-jung-mobileip-fastho-hmipv6-01.txt, IETF, June 2003
[18] Robert Hsieh and Aruna Seneviratne and Hesham Soliman and Karim-El-Malki, Performance Analysis on Hierarchical Mobile IPv6 with Fast-handoff over End-to-End TCP, in Proc. of GLOBECOM, 2002
[19] Gwon, Y., Kempf, J. and Yegin, A., “Scalability and Robustness Analysis of Mobile IPv6, Fast Mobile IPv6, Hierarchical Mobile IPv6, and Hybrid IPv6 Mobility Protocols Using A Large-scale Simulation”, ICC, Paris 2004
[20] Braden, R., Clark, D., Shenker, S., “Integrated Services in the Internet Architecture: an Overview”, Request for Comments (Informational) RFC 1633, IETF Network Working Group, June 1994.
[21] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and Weiss, W., “An architecture for differentiated service”. Request for Comments 2475, Internet Engineering Task Force. December 1998
[22] Mendes, P., Schulzrinne, H., and Monteiro, E., “Session-Aware Popularity Resource Allocation for Assured Differentiated Services”.
[23] Walsh, R., Xu, L., Paila, T., “Hybrid networks – a step beyond 3G”, third international symposium on Wireless Personal Multimedia Communications (WPMC’00), pp. 109-114, Bangkok, Thailand, November 2000.
[24] Braden, R., Zhang, L., Berson, S., Herzog, S., Jamin, S., “Resource ReSerVation Protocol (RSVP)”, Request for Comments (Standards Track) RFC 2205, IETF Network Working Group, September 1997.
[25] Bernet, Y., Ford, P., Yavaktar, R., Baker, F., Zhang, L., Speer, M., Braden, R., Davie, B., Wroclawski, J., Felstaine, E., “A Framework for Integrated Services Operation over Diffserv Networks”, Request for Comments (Informational) RFC 2998, IETF Network Working Group, November 2000.
[26] A. Talukdar, B. Badrinath, A. Acharya, “MRSVP: A Resource Reservation Protocol for an Integrated Service Network with Mobile Hosts”, In Wireless Networks, vol. 7 (1), 2001.
[27] Bless. R., Fu, X., Hancock, R., Jeong, S., Kappler, C., Lee, S., Manner, E., Mendes, P., Tschofenig, H., “Mobility and Internet Signaling Protocols” <draft-manyfolks-signaling-protocol-mobility-00.txt>, IETF, January 2004.
[28] Schulzrinne, H., Casner, S., Frederick, R., Jacobson, V., “RTP: A Transport Protocol for Real-Time Applications”, Request for Comments (Standards Track) RFC 3550, IETF Network Working Group, July 2003.
[29] Quadros, G., Alves, A., Monteiro, E., Boavida, F., “An Approach to Support Traffic Classes in IP Networks”, QofIS’2000. http://eden.dei.uc.pt/~boavida/papers/2000QofIS_PHB_final.pdf
[30] Marília Oliveira, João Brito, Gonçalo Quadros, Edmundo Monteiro, "Evaluation of a Quality of Service Routing Strategy for the Differentiated Services Framework", in Proceedings of the 2001 International Conference on Internet Computing (IC'2001), Monte Carlo Resort, Las Vegas, Nevada, USA, June 25-28, 2001.

[31] Paulo Mendes, Henning Schulzrinne, and Edmundo Monteiro, “Session-Aware Popularity Resource Allocation for Assured Differentiated Services", IEEE Communications Magazine, feature topic on Evaluation of Internet Architecture and Protocols, Vol. 40 (9), September 2002

[32] Paulo Mendes, Henning Schulzrinne, and Edmundo Monteiro, "Signaling Protocol for Session-Aware Popularity-based Resource Allocation", In Proc. of the fifth IFIP/IEEE International Conference on Management of Multimedia Networks and Services, Santa Barbara, California, USA, October 2002.

[33] Q_3M – Quality of Service for Mobile Multicast Multimedia http://q3m.dei.uc.pt