Paving the Road to Systems Beyond 3G – The IST BRAIN and MIND Projects

Dave Wisely and Enric Mitjana

Abstract: **Wireless LAN technology is complementary to 3G systems and could be used to provide high bandwidth hotspot coverage, for example in railway stations and offices, in order to provide the high bandwidth video and broadband services such as those emerging on DSL fixed access. The IST Projects BRAIN and MIND have investigated a number of key technical enablers for such a "system beyond 3G." These include scenarios and business models, design of an all-IP access network, consideration of ad hoc network extensions, enhancing Wireless LAN efficiency and compatibility with IP and, finally, terminal middleware and signalling for rapid adaptations to network QoS changes.**

Index Terms: **Systems beyond 3G, ad hoc networking, wireless LAN, HIPERLAN/2, mobile networks, UMTS.**

I. INTRODUCTION

A. Background

There is a great deal of activity taking place in the mobile arena at the present time. 3G systems have been operating in the Far East with steadily growing user numbers and are being trialled in Europe; large scale deployment is expected by the end of 2003. In the mean time GRPS (GSM Packet Radio Service) has, for the first time, offered mobile users per-packet billing and an "always on" capability. GPRS has also raised the data rate available to 64kbit/s with some operators.

Wireless LAN (WLAN) hotspots are also growing very rapidly across the world – in North America these have been available for some time and, following regulatory changes, look set to become widespread in Western Europe over the next few years. In Japan, also, a public WLAN service offering data and voice is planned. Home users are also recognizing the capabilities of WLAN for forming community networks, where broadband connection can be shared, and forming ad hoc networks at conferences and other events.

In the fixed network the emphasis is on broadband, where millions of users worldwide have Internet access at data rates of over 500kbit/s and are becoming accustomed to faster connections and data-rates that can support a reasonable video image. Napster showed that network traffic patterns can change rapidly and that the trend for bandwidth requirements is likely to continue upwards as video applications come within the range of DSL bandwidths.

Manuscript received July 31, 2002.

D. Wisely is with UMTS Research Group, Rigel House, B28/2B, BTexact Technologies, Adastral Park, Martlesham Heath, Ipswich, e-mail: dave.wisely@bt.com.

E. Mitjana is with the Research and Concepts group of the Mobile Networks division, Siemens AG, Hofmannstr. 51, 81359, Munich, e-mail: enric.mitjana@siemens.com.

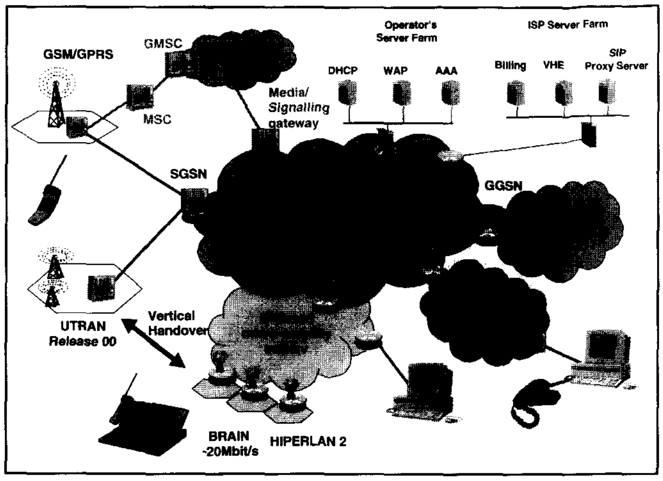


Fig. 1. Typical architecture for a "system beyond 3G."

In the application space, we are seeing network operators abandoning the search for the 'killer application' in favour of an open interface to their network - offering independent application developers access to location, payment and personalization services, for example.

There has also been consolidation within the telecommunications industry and a realization that sound business models and cash flow are required to sustain new services. For network operators the original Internet model of complete transparency has given way to more vertically integrated service offerings with the emphasis on value add and "ownership" of customers, as opposed to bit transport or commodity services.

It is yet uncertain where are all these developments leading. Some commentators believe that within some years we will have what are called "systems beyond 3G" where 3G, WLAN, ADSL, and other access technologies are complementary and users are able to access common services using TCP/IP running over all of them [1]. Fig. 1 depicts the introduced concept of systems beyond 3G with several access technologies, a common IP backbone, and service creation and intelligence pushed to the edge of network.

B. The BRAIN and MIND Projects

In order to investigate and research some of the key technical challenges that systems beyond 3G represent a consortium of some of the world's leading operators, manufacturers and academia formed the BRAIN (Broadband Radio Access for IP based Networks) project [2]. BRAIN has been succeeded by a follow-on project called MIND.

The MIND (Mobile IP based Network Developments) [3] project is partially funded by the European Commission in

1229-2370/02/$10.00

the frame of the Information Society Technologies (IST) Programme. The project partners of MIND are drawn from three key areas:

- Manufacturers: Ericsson AB (Sweden), Nokia Corporation (Finland), Siemens AG (Germany), Sony International (Europe) GmbH (Germany), and Infineon Technologies (Germany);
- Network operators: British Telecommunications plc (UK), France Telecom S.A. (France), NTT DoCoMo, Inc. (Japan), and T-Systems Nova GmbH (Germany); and
- Small enterprise and academia domain: Agora Systems S.A. (Spain), Universidad Politecnica de Madrid (Spain), and King's College London (UK).

MIND is scheduled to run from June 2001 to November 2002. The overall aim is to facilitate the rapid creation of broadband multimedia services and applications that are fully supported and customized when accessed by future mobile users from a wide range of wireless access technologies.

C. Key Technical Challenges

Some of the key technical challenges in developing a system beyond 3G, as described above are:

1) Delivering services over more flexible network topologies including ad hoc segments.
2) Adapting applications in a generic way to the highly variable quality of service (QoS) that heterogeneous wireless environments are likely to present.
3) Providing real-time services over WLAN access allowing voice and video on both WLAN and cellular systems.
4) Increasing the range, efficiency of transporting IP packets, support for higher speeds and lower cost of WLANs.
5) Providing seamless mobility, so that users simply access services and pay their bill without any concern for adaptation or service definition.
6) Developing a flexible and dynamic service creation environment.

In addition, there is the challenge of producing realistic scenarios and business models to support the concept of systems beyond 3G.

This paper sets out to look at how the BRAIN and MIND projects have been tackling these issues. In Sections II and III, the proposed architecture for an all-IP access network providing seamless mobility and QoS support to radio access technologies is described. Section IV, then presents an end terminal architecture capable of initiating services and very rapidly renegotiating them in response to changes in network QoS. The next two sections examine how WLANs (especially HIPERLAN/2) can be enhanced to efficiently carry IP with a guaranteed QoS, how the performance of the basic system can be increased and, finally, discuss options for 3G to WLAN interworking. The final sections describe scenarios and business models that have been developed to support the concept of systems beyond 3G. Section IX presents the results of a techo-economic modelling of the viability of rolling out UMTS and WLAN as complementary technologies developed in conjunction with the IST project TONIC [4].

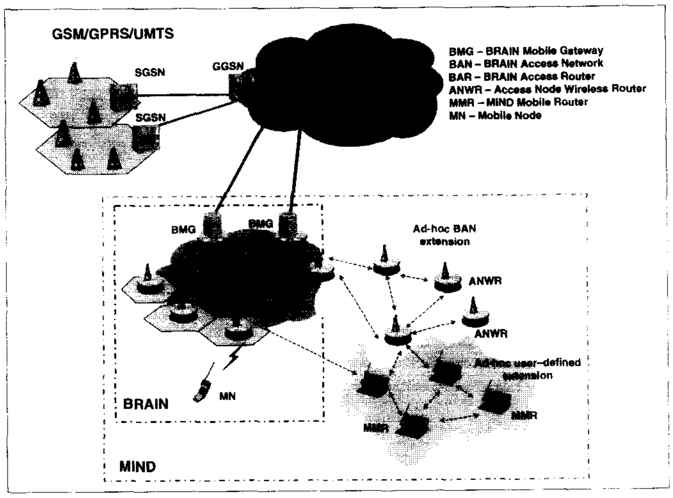


Fig. 2. Basic network architecture.

II. AN ALL–IP RADIO ACCESS NETWORK

A. What does All-IP Access Network Mean?

The term "all-IP" is often used in conjunction with mobile networks and has come to mean a number of different things to different people. At its weakest, the phrase really only implies that all the traffic is encapsulated, but not necessarily transported, within IP packets in the network core as in UMTS Release 5, where IP packets from the mobile are only reconstructed at the SGSN and then tunnelled to the Internet proper.

In contrast, in BRAIN and MIND we have taken a number of principles – mostly based on well-known Internet architectural concepts – to design an all-IP access network that transports IP packets natively and provides QoS and mobility support.

Our first design principle is network transparency: The network's primary role is to deliver packets and all other intelligence functions are provided at the edge of the network. This is an extension of the end-to-end principle [5] that really states that as much functionality as possible should be provided within the terminals, as opposed to within the network, since the end terminal is best placed to determine exactly what service is required.

Secondly, we wanted to design a network that enabled and encouraged further evolution implying that the components were independent. To this end we tried to follow a strict layering model. For example, only the IP layer processes had access to the link layer.

Our final precept was to only solve the special problems of mobile wireless access. We adopted existing and emerging solutions from evolving standards, such as the IETF, and contributed to developing these to take mobility, as well as QoS aspects, into account.

B. Basic Network Architecture

Fig. 2 shows our basic network architecture resting on the above principles. We have assumed an IP core network, where IP packets are highly aggregated and QoS is provided only for a few classes of traffic. The interface to the BRAIN Access Network (BAN) is through a number of BRAIN Mobile Gateways

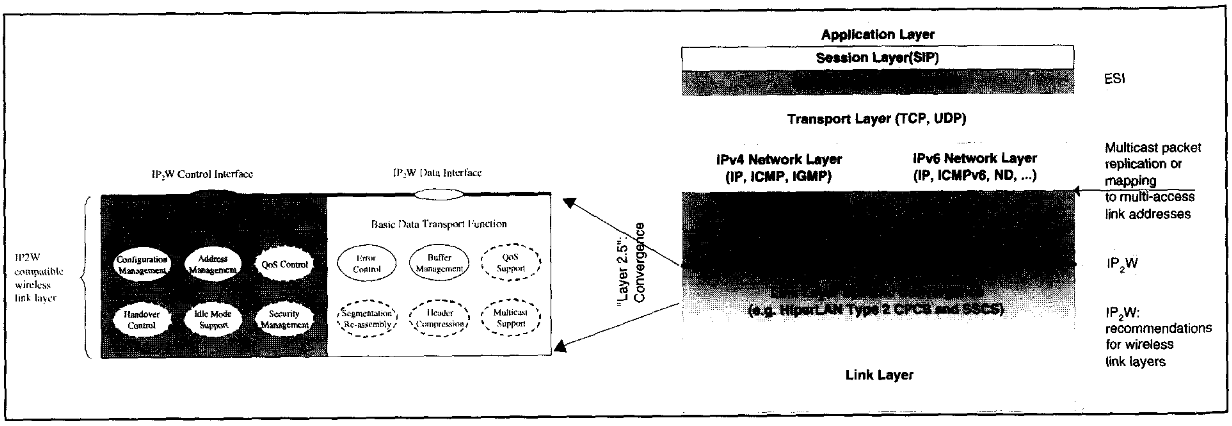


Fig. 3. IP2W and ESI interfaces.

(BMG). These act as normal IP gateways running exterior routing protocols, firewalls and so forth. More than one BMG is recommended for resilience. At the other end of the network are BRAIN Access Routers (BARs) having a radio link to mobile nodes (MN) and providing IP level mobility, security and QoS functions.

The BAN hides the fact that the nodes are mobile, and all details of the particular radio access technologies, from the core. Mobile access as seen by the IP backbone, at the network level, is very similar to access through DSL or UMTS (packet switched domain).

The BAN is responsible for allocating one or more globally addressable IP addresses to a visiting terminal and has the basic role of delivering packets to that terminal as it moves within the access network. This gives rise to three fundamental functions that are required within the network: Mobility management, QoS, and security. In this paper, we concentrate largely on mobility management and QoS because security issues have much more in common with those on fixed networks and mobile solutions are expected to develop from these.

The basic architecture can be further enhanced by networks of self-organizing unmanaged routing nodes. These are categorized as either Access Network Wireless Routers (ANWR), which could be portable or rooftop basestations, or MIND Mobile Routers (MMR) that are envisaged as end user devices. This part of the network we have called the ad hoc fringe. Typical examples of the ad hoc fringe are, for the former, an operator wishing to provide temporary mobile coverage for an exhibition, and, for the latter, a University campus where WLAN coverage is patchy and extended by routing through terminals that are within range of a fixed basestation.

The unique features of this architecture are:

- Simplicity – Only a few new elements are introduced for mobility and QoS support. Fundamentally, the network just delivers packets within the QoS requirement.
- Robustness – The network is designed to be robust against a single point of failure.
- Expandability – The use of strict layering and IP design principles means that the network is highly extensible.
- Scalability – Modelling has shown that the mobility management and QoS frameworks within the BAN can scale to cellular network sizes.

The IETF defines a very loose IP architecture and largely concentrates on protocol standardization where one protocol is developed for each function that must be supported. However, the design of network elements and interfaces between network elements is considered an implementation issue. In 3GPP (the body responsible for standardising UMTS), by contrast, a complete, fully integrated, architecture has been defined with clear network elements, well-defined interfaces between them for equipment interworking purposes and tight integration between the layers. The BRAIN/MIND architecture identifies the key network elements and some of the protocols that provide support for mobility, QoS, and security.

C. *Interfaces*

In fixed IP networks the interface between layers within the stack (e.g., IP to link layer) contains very little functionality. For mobile networks, where QoS, handover and other advanced capability is required, much more functionality and coordination between the layer 2 and layer 3 is needed. The projects have designed an IP to Wireless interface (IP2W) [6], see Fig. 3, which provides a way for the IP layer to interface to a number of different wireless link layer technologies. The IP2W interface is only present in the mobiles and the Access Routers because the wireless link layer needs to provide much greater functionality than typical wired links such as Ethernet. In particular QoS must be addressed at the link layer, a MAC is required and link layer addresses must be mapped to IP layer addresses.

The IP2W interface supports discovery of link layer capabilities. Many functions, such as handover, can be entirely handled at layer 3. However, if layer 2 information, such as signal strength is available, then better performance is possible. Layer 2 QoS must also be coordinated with layer 3 QoS to avoid inefficiency or instability; for example, at which layer does queuing and QoS differentiation take place.

In order to implement the functionality of the IP2W interface a convergence layer is necessary for WLAN technologies to build on the underlying functionality. Within the project we have specified a convergence layer for HIPERLAN/2 [7].

In addition to the IP2W interface the projects have also defined the Enhanced Socket Interface (ESI), see Fig. 3. The ESI takes the UNIX or Windows network socket concept further by including QoS support for applications or middleware allowing them to access the underlying QoS facilities provided by the network layer. Actual QoS might be actually achieved through, for example, DiffServ or overprovision without the applications being aware of how this is achieved.

III. DESIGN SOLUTION FOR AN ALL - IP RAN WITH AD HOC EXTENSIONS

A. *Mobility Management*

Mobile IP has been developed over a number of years as an IP mobility solution. Users have a Home Agent and packets sent to this agent are tunnelled to a local Foreign Agent. As terminals move they acquire new care-of-addresses and the tunnel endpoint is moved. Mobile IP, however, does not provide seamless handover and has other disadvantages such as triangular routing and makes QoS support difficult [8]. Within the project we have taken the view that Mobile IP might be used for IP macromobility (i.e., between domains) but that an IP micromobility protocol should run within the mobile access network. This is not only to overcome some of the listed disadvantages of Mobile IP but also because our business models suggest that the mobile has to be authenticated by the access network (although this may be with a remote operator) and then receives a globally routable address for the time that it is active within the network. This approach allows mobility-related issues to be confined within the access network. When a terminal receives an IP address from the access network it can then register this with a Mobile IP Home Agent but it could also use this address as part of a Session Initiation Protocol (SIP) registration process [9].

The role of the micromobility management function is to support terminal mobility allowing sessions to be continued as terminals move between access routers as well as supporting signaling of incoming sessions (e.g., paging). Some of the requirements on the micromobility management solution are:

1. Minimise mobility signalling traffic.
2. Provide seamless handovers (with minimum delay and without loss of packets).
3. Be scalable.
4. Be robust, i.e., support multiple routes or rapid rerouting.
5. Be compatible with other Internet protocols.

There are a very large number of IP micromobility protocols available for this type of mobility support. In order to classify and evaluate them we decomposed the micromobility function into four parts:

1. Mobile Node to Access Router protocol - how the mobile signals interacts with the network.
2. Handover Framework - describing how the old and new access routers buffer and transfer packets to achieve seamless handover (e.g., by temporary tunnel).
3. Path updates - how the routers alter their routing tables to take account of the mobile handover.
4. Paging - such that nodes can report their location with a lower degree of accuracy than that of the cell.

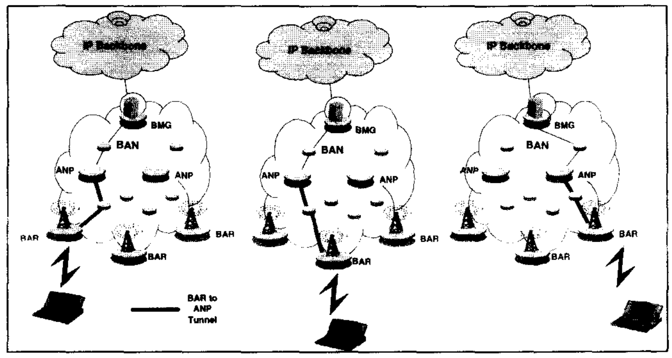


Fig. 4. BRAIN candidate mobility protocol (BCMP).

From a full analysis and evaluation [10] it was found that no existing protocol was able to provide all the requirements for a micromobility protocol. Using concepts from existing protocols the BRAIN Candidate Mobility Protocol (BCMP) [11] was developed to specifically provide such a solution. Essentially the BCMP is an IP micromobility protocol that allows users to connect to a wireless network, be allocated a globally routable address and then move around the mobile network without changing IP address, together with providing full support for real-time handover as well as paging for idle nodes. In this protocol users are authenticated either locally or remotely by an Anchor Point (ANP). If accredited the mobile is allocated a globally routable address from a range pointing at the ANP. The ANP tunnels received packets to the local access router and reverse tunnels packets from the mobile in the opposite direction, see Fig. 4.

The BCMP has the advantages that it will run over existing router networks such as OSPF, offers good QoS control since the packets are tunnelled and reverse tunnelled from access router to anchor point, it supports a fast handover framework (planned and unplanned handover) and paging.

Within the ad hoc fringe there are really two very different types of network extension. The wireless routers (the ANWRs introduced in Section II-B) might well be owned and controlled by the network operator and used, as an example, to provide temporary coverage at an exhibition. In this case the ANWRs must perform exactly like normal access routers in terms of authentication, handover and so forth. The key issue is how the ANWRs are securely added to an existing network without in any way compromising the existing system. We have explored several architectures for this attachment the most promising being a gateway ANWR that attaches directly to a BRAIN Access Router.

A campus or large office, where WLAN coverage is patchy, typifies the other type of ad hoc fringe network extension we have investigated. Peripheral users can only reach the access routers via one or more relaying hops through other user's terminals. Under these circumstances the intermediate nodes should not impersonate access routers and so the end node must establish a direct security association with one of the access routers. In one architecture that has been advanced from within the project, called the Virtual Radio Link (VRL), IP data and control packets are transported across the ad hoc fringe by an ad hoc routing function, see Fig. 5. The end terminal is authenticated

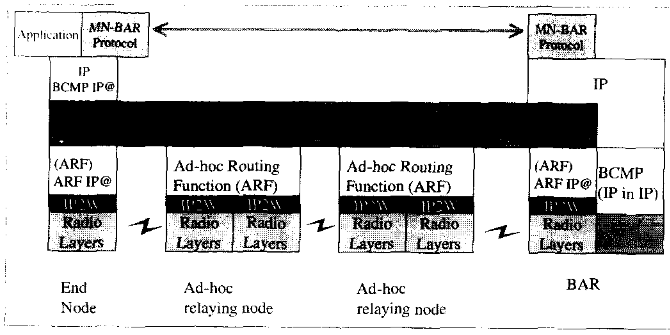


Fig. 5. Virtual radio layer (VRL) in the user-defined ad hoc fringe.

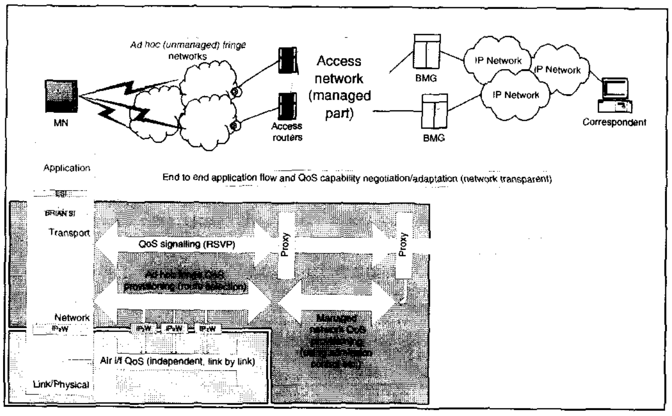


Fig. 6. MIND QoS architecture.

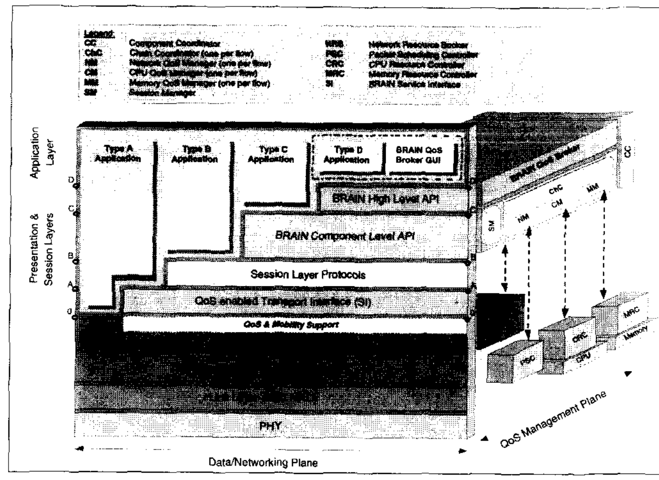


Fig. 7. BRAIN end terminal architecture (BRENTA).

by the access network and obtains a globally routable IP address from the Anchor Point. However it also obtains another IP address (possibly self-configured) for the ad hoc routing function. IP packets from the end node to the BAR are either tunnelled, IP in IP, or the access routers maintains an IP address mapping table to alter the headers appropriately. The Mobile Node to BAR protocol runs over the VRL as shown in Fig. 5.

B. Quality of Service

In our QoS architecture, see Fig. 6, the Extended Socket Interface provides an Applications Programming Interface (API) to RSVP for applications to signal QoS requirements to both the access network and the correspondent node. IntServ, per-flow, reservations are made over the wireless hop and then mapped to DiffServ Per Hop Behaviors (PHB) at the ingress to the access network. The access network, therefore, holds state only at the edge but can also be enhanced, for example, to provide better support for real-time services with the Bounded Delay PHB; this gives a hard guarantee on the maximum delay a packet will experience at each router. Another advantage of this architecture is that, even if end-to-end QoS is not available, localized QoS within the access network is still supported. Finally, nodes that only support DiffServ packet markings can obtain a better than best effort service.

Within the user-defined ad hoc fringe QoS routing [12] in which the routing protocol takes the requested QoS and computes a route that can deliver that level of QoS given the current state of the network has been studied. QoS routing avoids the need for the access network to try and organize QoS within a part of the network that it does not control. It is likely that only DiffServ classes would be supported in the ad hoc fringe but RSVP would still be used to signal application QoS requirements.

Another important aspect of QoS is Radio Resource Management (RRM). In traditional cellular networks this refers to such functions as layer 2 admission control, congestion control, handover management and power control. An IP access network must provide similar functionality and it is clear that some of the IP RRM functions reside below the IP2W interface and some, such as those involving more than one router (handover, non local congestion, etc.) must reside above the interface.

IV. BRAIN END TERMINAL ARCHITECTURE (BRENTA)

End-to-end QoS is very difficult to deliver across different IP domains. User's data may have to traverse an ad hoc fringe, be subject to DiffServ Per Hop Behavior in the BRAIN Access Network and, perhaps, little QoS support in a distant access network. Mobility, with sudden deteriorations of link quality and handover, makes any guarantees that are given liable to be broken several times during a session. Only in really homogeneous systems, where end domains are similar, services tightly defined and radio resource continuously allocated, as it is in GSM, can mobile QoS be delivered by the network alone.

In the BRAIN/MIND architecture the network provides IP transport with a target QoS that is not a hard guarantee even over the access network. If applications were directly exposed to this kind of rapidly changing network QoS during a session then the user-perceived QoS could be very poor indeed. What is needed is a way for nodes to measure and adapt to the QoS that the network is providing. Given the rapid nature of QoS fluctuation some QoS information and an adaptation strategy will typically need to be negotiated and agreed by the end nodes of a multimedia session in advance.

There are other QoS considerations that must be taken into account at the end nodes as well as QoS adaptation. Firstly, it is no good launching a video application with a computationally

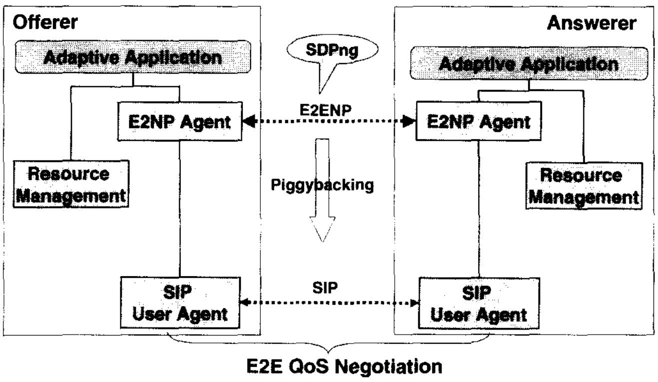


Fig. 8. Implementation of QoS adaptation.

intensive codec if the processor does not have spare capacity. Secondly, all the parties to the session must negotiate in advance details about the session such as codecs, data rates and data formats.

Fig. 7 shows the complete BRENTA architecture which is capable of varying levels of these functions to different application types: legacy (type A); those that utilize session protocols (type B); those that can make use of a component API providing frame grabbers and packetizers (type C), and those that can make use of a full blown QoS broker to deal with all connection issues (type D). Full details of all the functions and components can be found in [13].

Currently SIP is the IETF protocol for session initiation, negotiation and termination. Session Description Protocol (SDP) is normally used to describe the type of session and the details such as codec type and bitrate. In a mobile, heterogeneous, environment using SIP to renegotiate each time a QoS violation took place would be too slow and generate too many signaling messages. The project has designed a model for the end-to-end negotiation of terminal capabilities and a model for the end-to-end negotiation of QoS. Fig. 8 shows an implementation of this functionality.

Central to the implementation is a new protocol, the End-to-End Negotiation Protocol (E2ENP) [14]. E2ENP allows end peers reaching an early agreement on what capabilities to use and which QoS levels to enforce. End terminal use the protocol to agree an Adaptation Path (AP) defining in advance which alternative to choose upon QoS violations, out of the negotiated ones. The Extended Socket Interface (ESI) is then responsible for actually enforcing the QoS.

E2ENP uses next generation SDP (SDPng) and extends it to take account of its use to describe terminal capabilities, the adaptation path and so on. E2NP itself is carried within SIP messages and a SIP User Agent is present at both end terminals.

When handover takes place between different networks, such as from WLAN to UMTS, then a major renegotiation and adaptation to QoS is required. The MIMD/BRAIN architecture tackles this at several levels:

- Terminal middleware – interprets the user's preferences and QoS requirements into underlying reservation and performs adaptation (e.g., using the End-to-End Negotiation Protocol) when QoS changes abruptly.

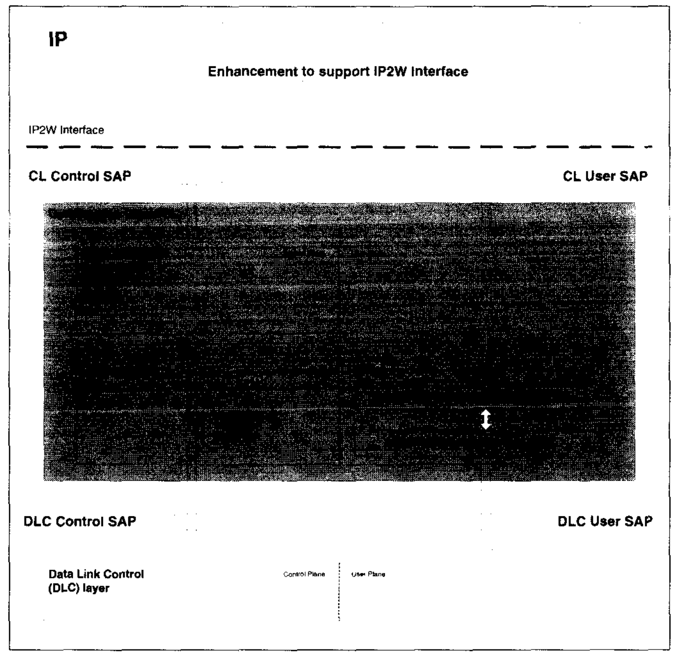


Fig. 9. HIPERLAN/2 convergence layer for IP.

- IP layer – with end-to-end IP QoS signalling it is possible to set up common reservations across different underlying networks.
- IP2W - the IP2W interface is able to support IP QoS classes and frameworks directly.
- Multihoming – terminals can be connected to multiple Access Routers, multiple Access Networks or multiple ISPs.

V. WIRELESS LANS

Wireless LANs (WLANs) are very much in the news at the moment with many network providers announcing plans recently to install access points in hotspots in Japan and Western Europe; in America this service has been available for some time. These offerings are an example of what has been termed a first generation WLAN. The services that are offered are Internet access and Virtual Private Network (VPN) connection to a corporate Intranet. At hotspots (coffee shop, airport, etc.) a WLAN base station is connected to a DSL back-haul to a hub where authentication takes place. Initially these services will be on a subscription basis and authentication will be by user name and password. Operating in the ISM 2.4GHz spectrum the systems will offer a maximum of about 500kbit/s for each individual user.

Within BRAIN and MIND we have been investigating so called second generation WLANs that will offer:

1. Higher data rates – up to 10Mbit/s for each user.
2. Support for real-time data services such as voice and video.
3. Horizontal handover between WLAN cells.
4. Vertical handover between WLANs and cellular technologies, such as UMTS.
5. A wide range of IP-based services.

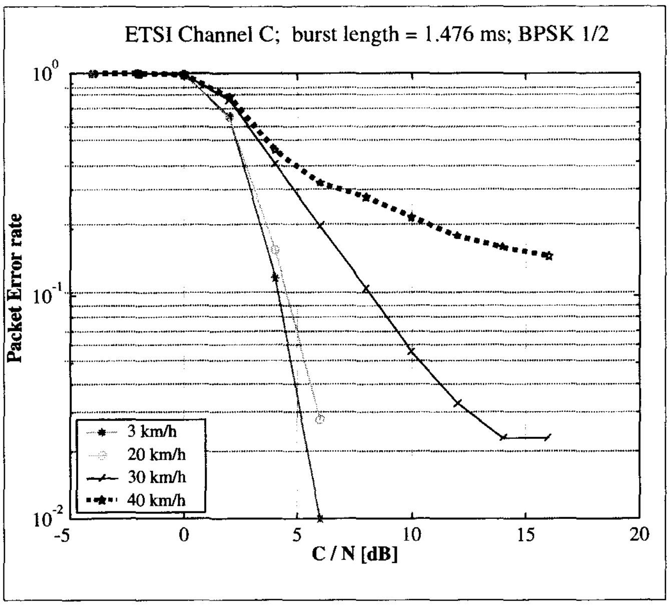


Fig. 10. HIPERLAN/2 degradation with terminal speed.

Within the projects we have taken HIPERLAN/2 as a candidate second generation WLAN. HIPERLAN/2 operates within the 5GHz band where it is expected significant spectrum will soon be globally available for WLAN operation. HIPERLAN/2 also offers a high level of functionality when compared to 802.11b, for example, in the areas of QoS, connection management and dynamic frequency selection.

The project has designed and specified a convergence layer to enable HIPERLAN/2 to efficiently transport IP packets with defined QoS, see Fig. 9, adding to the convergence layers already available for Ethernet and ATM over HIPERLAN/2.

The convergence layer specification, that has been submitted to ETSI BRAN for standardization [7], provides support for:

1. IPv4 and IPv6.
2. IntServ and DiffServ.
3. Detailed address resolution schemes.
4. The IP2W interface.
5. QoS mapping and scheduling.
6. Paging complying with IETF drafts.
7. Unicast, anycast and multicast.

Another important aspect of WLANs is that they cover only a relatively small area, typically 50m indoors and 100m outdoors. This is limited by the permitted transmit power. Obviously, if this can be extended then costs can be reduced and new scenarios covered. Also the spectrum can become crowded. The 2.4GHz ISM band is unlicensed and also used by Bluetooth short range radio links and at a typical hotspot it is not hard to imagine that the capacity will be greatly reduced by non-operated WLAN and Bluetooth links. In addition to increasing capacity, mitigating interference and extending range is the need to support higher mobile speeds as the ad hoc fringes will sometimes include moving vehicles, see Fig. 10.

Within the project physical layer enhancements to HIPERLAN/2 for extended range and higher speeds have been studied including: Multiple antennas, turbo coding and adaptive modulation. Efficient MAC strategies for vehicle-to-vehicle communications and multipoint-to-multipoint were studied as well [15].

VI. 3G – WLAN INTERWORKING

Since the preparation stages of the BRAIN project back in 1999, the projects' partners have considered 3G and Wireless LAN complementary technologies. This combination caters for wide area coverage through UMTS and allows the delivery of broadband services in hotspot areas via WLANs.

Through this approach users can enjoy multimedia services with high mobility and large throughput in low mobility scenarios. Depending on the architectural choices, seamless hand over between access technologies and service continuity can be provided. Furthermore, a convenient common billing can be offered to the user.

This complementary approach is also beneficial to network providers. They can reduce infrastructure costs in 2.5G or 3G equipment by shifting traffic to WLANs in high traffic areas. One could also expect that getting users acquainted with high bit rate WLANs will have a postive effect on the demand for mobile multimedia services, thus promoting the early uptake of 3G.

Service providers should also benefit from this technology independent access to the services they have to offer by increasing the reachability of their services. The possibility of accessing the services via different networks (possibly at different cost) leaves room for designing creative billing packages. Beyond that, since users may not be tied to a single network provider for wireless access, service providers can make a step towards establishing direct relations with the final user. This facilitates the concept of mobile virtual network operators (MVNOs) that sell services over others in 3G infrastructure but have contracts with customers and could, for example, issue SIM chips.

A. 3G-WLAN Coupling Types

BRAIN developed a number of ways in which roaming between a BRAIN access network and a 3G network could be supported:

- No coupling: Users must reauthenticate when changing network and would be charged in two separate ways (e.g., two bills or bill and credit card) but might use a third party Mobile IP or SIP server to maintain session continuity albeit with a long delay during handover. End terminal software, such as the E2ENP presented in Section IV, would be expected to adapt to any change in bandwidth or QoS. This is a flexible and provider-independent way of accessing services.
- Loose coupling: Where a common network or service provider ensures that users would not need to reauthenticate and could be billed in a common way for access to the two different networks. The common operator might provide a way of maintaining session connectivity, but this would not be seamless and, again, end terminals would be responsible for adapting to altered bandwidth or QoS. A

typical example might be a WLAN to UMTS handover where the bandwidth drops from 500kbit/s to 144kbit/s.

- Tight coupling: Where the BRAIN access network gateway is connected directly to network elements within the 3G network, such as the UMTS Radio Network Controller (RNC) or Serving GPRS Support Node (SGSN). Only with tight coupling would a seamless (i.e., fast and smooth) handover be possible. By coupling resource management in both networks it might be possible to support existing QoS reservations and, where this was not possible, to perform adaptation within the network itself (e.g., compressing a streamed video).

Within the MIND trial activities, the loose and no coupling approaches are being evaluated to quantify the disruptions of inter-system handovers and to see whether end terminal adaptation is effective.

ETSI BRAN has done extensive anaylsis of the different coupling options [16]. At the beginning of 2002 it endorsed the loose coupling approach for the interworking of HIPERLAN/2 and 3G. Recently, the system architecture group (SA2) of 3GPP has created an ad hoc group on WLAN interworking and acknowledged the work done by ETSI BRAN. 3GPP suggests that this work is applicable to other WLAN technologies like the MMAC HiSWAN and IEEE802.11.

It has to be mentioned that coupling options and restrictions result from the business relations between the different involved parties in the provision of mobile services. For example, all options are open to an incumbent 3G network provider that is at the same time a service provider and deploys WLAN in hotspots, although tight or loose coupling seem the most reasonable choices. On the other hand, only a no coupling approach is realistic for a pure WLAN network provider. Thus, it is clear that the architectural decisions and the business models pursued need to be aligned.

VII. SCENARIOS AND BUSINESS MODELS

A. *The Relevance of Scenarios*

In a user centric approach to the definition of upcoming mobile systems it seems reasonable to start looking at the needs and requirements of the user. For this reason, having a set of 'usage scenarios' was considered crucial in the BRAIN and MIND projects when addressing that goal.

A scenario definition provides freedom of imagination on the possible future user needs and wishes independently from any constraints resulting from present technical or economical viability. Scenarios can help in the determination and specification of the issues to be solved. They allow to identify the involved elements and players, and assign functionalities and constraints to them. This top down analysis permits to derive clear technical requirements to support a user centric vision and to spot flows of services and money.

B. *An Example Scenario*

Based on the original BRAIN work, MIND has defined three scenaros called "leisure time," "nomadic worker," and "medical care." A section from the "nomadic worker" follows [17]:

"Stephanie Jones is a member of a company based in Frankfurt. She constantly travels, often abroad. One morning, Stephanie travels to Munich to attend an important meeting with a customer at her company's facilities in Munich. Stephanie takes the train to get there. She knows that some colleagues attending the meeting will also take the same train in some intermediate stations. She has earlier invited them to an on-the-road meeting in the train, in order to prepare the meeting. When entering the train, Stephanie's terminal informs her that the train offers a high-performance mobile communication network to its passagers. Stephanie logs into the train network and connects to her company's Virtual Private Network (VPN). When her colleagues gradually join her in the train, they attach their terminals in an ad hoc fashion to Stephanie's one. They use Stephanie's terminal to access the internet and their company's VPN as well . . ."

Based on scenarios like the one presented here, the basic requirements for the terminal, access network and air interface were derived. Already available technical solutions were compared with these requirements and, when necessary, new approaches were developed.

C. *Refining Business models*

Understanding the roles of the players involved in the scenarios has been fundamental for refining and extending the BRAIN business models. In the case presented above, two novel roles can be defined:

- Auxiliary Network Provider: An individual that uses her device to provide access services to end users (e.g., Stephanie offering access to her colleagues).
- Extended Network Provider: Provides access services through a hotspot high performance wireless network (e.g., the train company offering access to Stephanie).

The business relationships between the traditional and these new roles has been studied in detail and corresponding business models have been evolved. The next figure depicts the resulting value chain for the scenario presented in the previous section.

Fig. 11 shows the flow of services and money between involved parties. The values proposed for the revenues are only representational. Interesting relationships and business prospects for the different players can be derived from a similar analysis to the one presented here. One major conclusion is that the customer ownership in future systems seems to be one of the key aspects.

VIII. SERVICE CREATION

A flexible and dynamic service architecture must be capable of supporting the BRAIN/MIND business models and, in particular, allowing the full range of value chain players, from content providers to users themselves, to participate in the service creation process. In traditional telecom networks the role of service creation has fallen, by and large, onto the network operator. Services such as divert-when-busy, voicemail and free-phone services, involved deep changes, often at the exchange level, to the

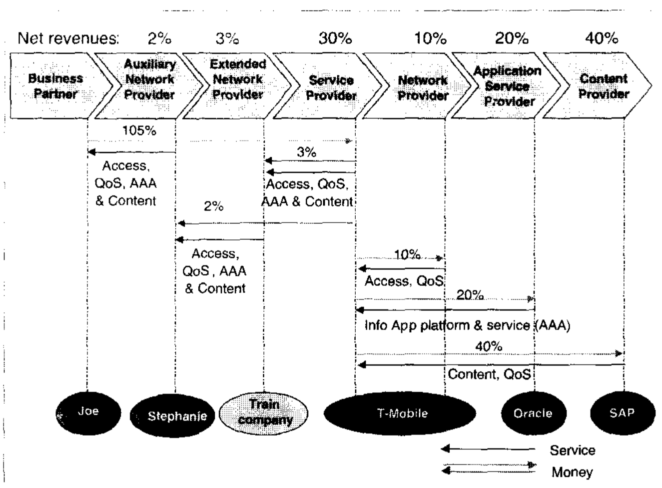


Fig. 11. Value chain for the "Nomadic Worker" scenario.

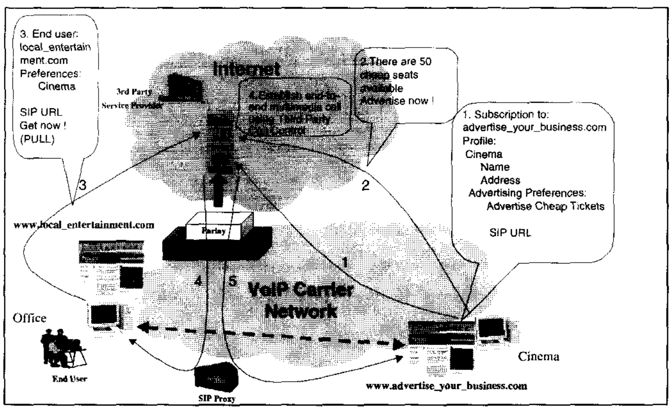


Fig. 12. SIP used for service creation.

network. Operators were not prepared to open up the necessary interfaces or yield up the potential revenue to third parties, restricting service creation opportunities. With the IP revolution users can have a copper wire provided by a telco, Web access by an ISP and buy goods in an online bookshop with a credit card. But online shopping is just one service and is typical of the client-server type services that are available today. If we are to move beyond these, to services that include presence and real-time multimedia as well as offering users integrated services then a new architecture for dynamic and flexible service provision is required. As an example we might have a user register that he is looking for entertainment tonight and his preferences include war movies and historical films. A cinema might be having trouble filling the screen for its latest blockbuster war movie and so advertises 50 cheap seats. A really useful service would put the two in touch and initiate (perhaps) a transaction.

In the past service creation has been anything but dynamic with provision of services taking up to several days. In the Internet, multimedia future services will be created almost in real-time. If I am at my local team's football match I may have a camera in my PDA and could use WLAN hotspot to relay the game to the local network. This would probably involve a multicast, various rate adaptors and QoS requests to the network, as

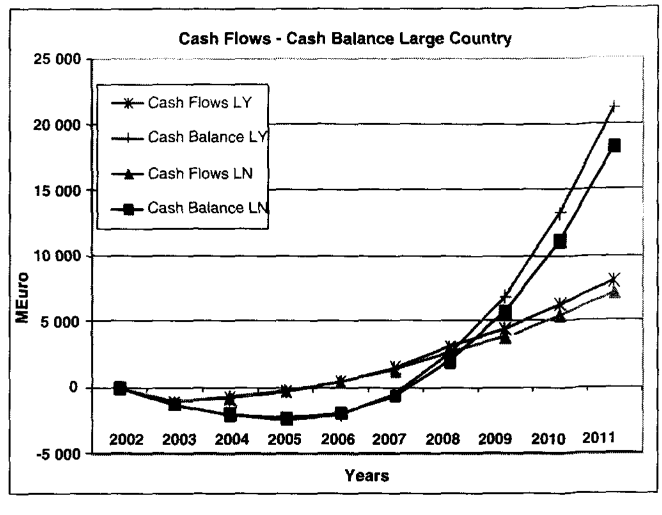


Fig. 13. Example cash flows of an operator deploying UMTS with (LY) and without (LN) a Wireless LAN complementary system.

well as endless renegotiations to deal with changed loads, etc. All this requires a very dynamic architecture that can respond in real-time to service requests and changes. Fig. 12 shows an example studied in the projects of how SIP and Parlay (a generic network interface for third party applications) could be used to create a dynamic and flexible service platform in the case of the cinema example cited above.

IX. TECHNO – ECONOMIC ANALYSIS

In the frame of MIND and in cooperation with the IST project TONIC, a techno-economic analysis on the provision of IP services over wireless networks has been performed [4].

The analysis assumed UMTS technology for 3G and an enhanced HIPERLAN/2 for WLAN. To this avail, the BRAIN architecture was taken as reference. Both 3G and WLAN were considered to be owned by a single network provider. Futher inputs were service sets including demand and average revenue per user (ARPU) estimates. Two deployment scenarios having demographic characteristics similar to large and small European countries were studied.

Fig. 13 shows the cash flow and cash balance curves in the large country case with (LY) and without (LN) WLAN deployment.

The comparison of cash balances in that figure shows that the revenues gained from offering broadband services over WLAN more than offset the additional costs. This accounts for the improvement of both the Net Present Value and the Internal Rate of Return. Note that the global discounted investment for WLAN turned out to represent only 2% of the total, while global discounted revenues resulting from it accounted for 8% of the total.

The techno-economic study concluded that adding WLANs to the roll out of UMTS, even where licences were not paid for, can be done at a relatively low cost. However clear benefits, as described in section VI, can be obtained from it. Recent moves from established network providers towards complementing their 2.5 and 3G offerings with WLAN in Europe (Telenor, Sonera), USA (VoiceStream, Nextel) and Japan (NTT

DoCoMo) seem to corroborate the attractiveness of this combination.

X. CONCLUSIONS

The concept of a "system beyond 3G" is based on the combination of several access technologies like 3G for cellular coverage, WLAN for hotspot coverage and DSL for broadband fixed access. These are unified by running IP at the network layer and are interconnected by an IP backbone. By offering users a choice of bandwidth and price, as well as common services, it is expected that take up and use of multimedia services will be greater than if the various access networks exist only in isolation. There are many technical challenges that need to be solved before advanced multmedia services can be created and delivered seamless across such a system. In particular:

1. An IP-based access network supporting QoS and mobility is required for WLAN deployment.
2. WLAN efficiency, support for QoS and integration with IP networks needs to be enhanced.
3. Ad hoc extensions to networks must be considered for greater flexibility.
4. A dynamic and flexible service provision framework is needed.
5. Middleware for the rapid adaptation of applications to network QoS must be developed.
6. Scenarios and business models must be developed to support the business case for systems beyond 3G.

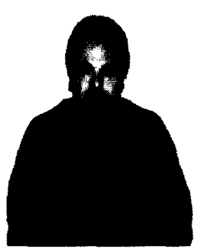
Within this paper we have tried to outline the progress that the IST projects BRAIN and MIND have made towards solving some parts of these challenges. We believe that with the roll out of 3G, WLANs and DSL, the world over there will have powerful economic reasons, for either operators or third parties, to unify the customer experience (e.g., billing, services, personalization) through the commonality of IP. We believe it will be systems beyond 3G that will come to dominate the telecommunications arena over the next decade.

ACKNOWLEDGEMENTS

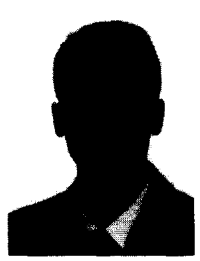
This work has been performed in the framework of the IST project IST-2000-28584 MIND, which is partly funded by the European Union. The authors would like to acknowledge the contributions of their colleagues from Siemens AG, British Telecommunications PLC, Agora Systems S.A., Ericsson AB, France Telecom S.A., King's College London, Nokia Corporation, NTT DoCoMo Inc, Sony International (Europe) GmbH, T-Systems Nova GmbH, University of Madrid, and Infineon Technologies AG.

REFERENCES

[1] J. Pereira, "Fourth generation: Now, it is personal," in *Proc. 11th Int. Symp. Pers. Indoor and Mobile Radio Commun.*, Sept. 2000, vol. 2, pp. 1009–1016.
[2] Available at http://www.ist-brian.org.
[3] Available at http://www.ist-mind.org.
[4] E. Mitjana *et al.*, "Seamless IP service provision: Techno-economic study by the MIND and TONIC projects," in *Proc. IST Mobile & Wireless Telecommun. Summit 2002*, Thessaloniki, June 2002.
[5] RFC 3234, available at http://www.ietf.org and for example the discussions, available at http://www.postel.org/mailman/listinfo/end2end-interest.
[6] J. Manner *et al.*, "Exploitation of link layer QoS mechanisms in IP QoS architectures," in *Proc. ITCom 2001*, Denver, Colorado, Aug., 2001., Published in "Quality of Service over Next-Generation Data Networks," in *Proc. SPIE*, vol. 4524, pp. 273–283.
[7] S. Bonjour *et al.*, "IP convergence layer with QoS support for HIPERLAN/2," in *Proc. IST Global Summit 2001*, Barcelona, Sept. 2001, available at http://www.ist-mind.org. See also "Functional specification of the IP service specific convergence sublayer for HIPERLAN/2," *IST MIND contribution to ETSI BRAN#28*, Apr. 16, 2002.
[8] D. Wisely *et al*, *IP for 3G*, Chap. 5, John Wiley 2002, 0-471-48697-3.
[9] D. R. Wisely, "SIP and conversational internet applications," *BT Technology J.*, vol. 19, no. 2, Apr. 2001.
[10] P. Eardley, A. Mihailovic, and T. Suihko, "A framework for the evaluation of IP mobility protocols," in *Proc. PIMRC 2000*, London, Sept. 2000.
[11] C. Keszei *et al.*, "Evaluation of the BRAIN candidate mobility management protocol," in *Proc. IST Global Summit 2001*, Barcelona, Sept. 2001, available at http://www.ist-mind.org.
[12] C. R. Lin and C. C. Liu, "An on-demand QoS routing protocol for mobile ad hoc networks," in *Proc. IEEE Int. Conf. Networks (ICON 2000)*, Singapore, Sept. 2000, pp. 160–164.
[13] A. Kassler *et al.*, "BRENTA - supporting mobility and quality of service for adaptable multimedia communication," in *Proc. IST Mobile Communications Summit 2000*, Galway, Ireland, Oct. 2000, pp. 403–408, available at http://www.ist-brain.org.
[14] A. Kassler *et al.*, "Enabling mobile heterogeneous networking environments with end-to-end user perceived QoS - the BRAIN vision and the MIND approach," , in *Proc. EWC*, Feb. 2002.
[15] M. Dohler *et al.*,"HIPERLAN/2 for vehicle to vehicle communication," in *Proc. PIMRC 2002*, Lisbon, Sept. 2002.
[16] ETSI BRAN - TR 101 957.
[17] T. Robles *et al.*, "Usage scenarios and business opportunities for systems beyond 3G," in *Proc. IST Mobile & Wireless Telecommun. Summit 2002*, Thessaloniki, June 2002.



Dave Wisely has worked for BT for 15 years in the fields of networks and mobility research. He pioneered optical wireless link in the early 1990s constructing a 4km, 1500nm system using optical amplifiers. He has worked in the field of mobility for the past six years, looking firstly at wireless ATM and HIPERLAN type 1 systems and more latterly into cellular mobile systems. He was one of the pioneers of an all-IP solution for future developments of 3G. He heads up BT's UMTS and 4G research unit and is technical manger for the influential EU IST BRAIN/MIND project. He has also just published a book entitled "IP for 3G."



Enric Mitjana has been with Siemens' Information and Communication Mobile Group since 1997. He was involved in the UMTS standardisation in the ETSI SMG2 and 3GPP. In 1999, he was delegated to Beijing, China, to set up an R&D centre concentrating on the TDD mode of IMT-2000. On returning to Munich, he was Project Manager in the Mobile Internet Solutions Division working on UMTS applications. Since June 2001, he is Project Manager of the IST MIND research project looking at systems beyond 3G. He is author and co-author of several papers on physical, network and service aspects of mobile communications systems. He holds several patents on radio communications and wireless networking.