EVOLVING BEYOND UMTS - THE MIND RESEARCH PROJECT

P L Eardley[1], J Eisl[2], R E Hancock[3], D J Higgins[1], J M J Manner[4], P M Ruiz[5]

[1] BTexact Technologies, UK, [2]Siemens AG, Germany, [3]Siemens/Roke Manor Research Limited, UK, [4]University of Helsinki, Finland, [5]Agora Systems SA, Spain

INTRODUCTION

With the first commercial 3G systems nearing deployment and the standardisation of future UMTS releases well advanced, research has now started to map out the longer term evolution of mobile networks to extend and improve UMTS in more fundamental ways. The EU IST project MIND (1) (and its predecessor BRAIN (2)) have taken a wide ranging view of this subject, considering user applications, network architectures, and future broadband radio interfaces with a timescale of 5-10 years beyond UMTS. The main network concepts considered here are unified mobility management and QoS support in an 'all-IP' infrastructure. Security is also a critical issue in future networks, but here MIND will take its lead from other research activities and current standardisation work. The main additional goals are to enable multi-radio integration and novel mechanisms which allow the cost effective deployment of a high density of basestations – a likely necessity for future high bandwidth wireless access, given the shortage of available spectrum.

THE MIND NETWORK CONCEPT

The basic components of the MIND network are shown in Figure 1, which shows it as a peer network to UMTS and other access technologies. Our usage scenario considers only IP-based applications and services, so the MIND network only handles the transport of IP data. Indeed, the basic goal of the MIND network can be summarised as making mobile access look the same as access through a fixed IP network (such as ADSL).

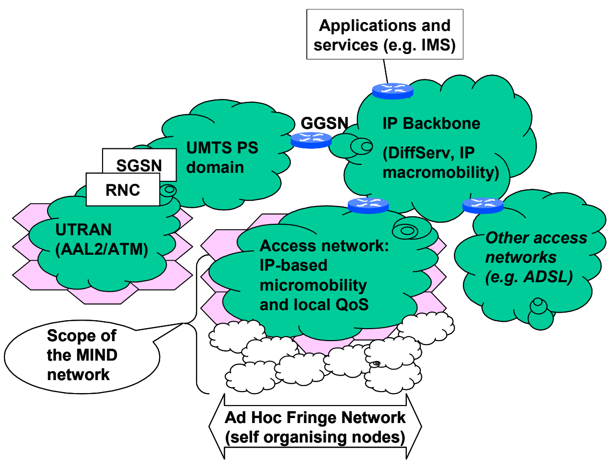

Figure 1: The MIND Network

In this sense, the MIND network is comparable to the combination of a radio access network and mobile core network packet domain. For the end-user, the MIND network can therefore provide UMTS services in the same way as is currently being considered for WLANs (e.g. Findlay et al (3)), either basic Internet access or services from the Internet Multimedia Subsystem (IMS).

As well as providing IP access, the MIND network is itself fully IP-based, both in the sense of using IP transport internally, and using IP mechanisms to support terminal mobility and QoS support all the way from the mobile device to the inner boundary of the access network. Indeed, we treat even the basestations as routers implementing mobile-optimised routing and QoS mechanisms (as compared to the UMTS case, where the first router for IP traffic is the GGSN). This architectural approach is sometimes referred to as 'all-IP' or 'native-IP' radio access. An effect of this is that the particular physical air interface being used is significant only at the very edge of the network, so a single access network naturally supports heterogeneous radio environments.

FUNDAMENTAL COMPONENTS

Implementation of the MIND network requires the development of the basic functions of terminal mobility and QoS support. These are considered for two regions:

- A traditional managed infrastructure, using wired (or at least static) links for routing from the basestations towards the core network, and

- At the outer edge, networks of self-organising, unmanaged routing nodes – either fixed nodes, or possibly even end user devices themselves. This part is referred to as a the 'ad hoc fringe'.

The terminal mobility solution includes both handover signalling and re-routing within the network to support seamless mobility between basestations and within the ad hoc fringe. QoS support includes protocols for signalling the QoS requirements of particular application flows and carrying out admission control, as well as traffic management mechanisms for the actual QoS provisioning itself. These two fundamental building blocks are described in the remainder of this section. The following section then considers the operation of these components to support the particular network features of multiradio support and multicast.

The MIND Mobility Solution

The fundamental requirement of any mobile network is that it should support the mobility of users and their terminals as they move around and between the access networks of different operators. We distinguish *personal* mobility – enabling a user to access any network and receive consistent services – and *terminal* mobility –maintenance of a continuous service even as the user's device changes basestation or indeed access technology. This is a very broad problem, and UMTS uses several cooperating protocols to address it. UTRAN procedures support terminal mobility, while at the other extreme, protocols from the IETF such as SIP relate mainly to user mobility. GPRS protocols relate to both functions.

MIND has adopted a slightly different functional split in its approach to the mobility problem. Consistent with the project's focus on IP service support, the fundamental starting point is that the user's terminal is allocated an IP address on attachment to the access network which can be used unchanged throughout a single access network domain; however, the allocation mechanism, and the utilisation of the address in user mobility procedures are almost fully independent of the rest of the operation of the MIND network, which focuses on terminal mobility. This provides for a very flexible integration of MIND with other networks, compared for example to an approach which assumed use solely of Mobile IP signalling and transport format for all aspects of mobility management. MIND devices can use MIP if they wish – to achieve mobility between networks of different types and different operators, but other options are also allowed. For instance, mobility between a UTRAN and a MIND access network could be enabled by integrating the latter with GPRS macromobility protocols, as a special network under a customised SGSN. Several other (e.g. SIP-based) approaches are also possible, and all are equally decoupled from the MIND-specific terminal mobility support.

In practice, micromobility is assumed to cover the majority of all handovers that take place. Our solution developed first in BRAIN focuses especially on seamless handover for the users, while proper idle mode support and the scalability and resilience required for a 'carrier class' network represent major issues as well, especially as they have been ignored until recently in IP terminal mobility research. A particular design goal has been the independence between procedures involving the terminal (e.g. handover signalling over the air) and routing within the network, with the aim of allowing freedom of network implementation in the second area depending on operator scenario, while maintaining roaming capability for terminals regardless of the details of the network they attach to.

Existing mobility protocols have been investigated for these requirements and design goals. The outcome of this research activities is the definition and implementation of a new micromobility management protocol suite, with universal components for handover signalling and idle mode support, along with a particular network-internal routing protocol called the BRAIN candidate mobility protocol (BCMP). BCMP has similarities to Hierarchical MIP, and like GTP within UMTS it uses tunnelling for delivering datagrams to the different access routers. The main improvements are the independence of procedures involving the terminal, and intrinsic features for resilience and high scalability within a single access domain, up to tens of millions of users. Simulations have been carried out to verify the practical benefit of the newly defined protocol and to demonstrate that it overcomes deficiencies imposed by other mobility management solutions, as reported in Keszei et al (4).

A key benefit of the MIND network is that it provides mobility support for terminals with heterogeneous radio access technology. Mobility management in the access network is purely based on IP network layer mechanisms – therefore the MIND access network handles all mobility issues, which are specifically related to the radio access technology. With this access network architecture a single mobility management solution can accommodate diverse radio access technologies, like UMTS, WLAN and Bluetooth access. Intertechnology handover between different radio networks are supported as well. To the user of the MIND network mobility management should be transparent – i.e. there should be no considerable difference compared to a fixed network access.

The MIND project aims to extend the BRAIN mobility solution by additional features like support for multihomed terminals and wireless ad hoc network fringes. Ad hoc networks provide the facility to extend a wireless access network infrastructure on demand with minimum administration effort and fast deployment. An intertechnology handover between an ad hoc network and a 3G based access ideally is triggered by the multihomed station. The decision which interface(s) to use should be based on the availability of the radio bearer signal – but additional criteria could be taken into account as well. Available bandwidth or charging schemes may be considered, when taking a decision for performing a handover between different access points. Though WLAN access could be more preferential, a mobile terminal could trigger a handover to the UMTS standby access after the ad hoc access is disrupted by some external event. Though the ad hoc network extension provides new promising business models for mobile users, the volatility of the network topology imposes some challenges for the mobility management solution. Interworking between a routing protocol in an ad hoc network stub and BCMP protocol in the fixed part of the access network is just one aspect to mention here.

The MIND QoS Architecture

The support for QoS in UMTS is based on four traffic classes: conversational and streaming classes for bi- and uni-directional streaming applications, respectively, an interactive class for delay sensitive data applications, and the background class. Currently there are 11 traffic parameters that are used to define the service in each class including peak and guaranteed bit rates, transfer delay and various error ratios.

However, there are shortcomings to the UMTS QoS architecture if an IP-based mobile node wants support for delay sensitive flows. First of all, the UMTS QoS management is mainly effective within the UTRAN. In addition, there is no standardised scheme to map IP-layer QoS mechanisms to the UMTS QoS architecture. Thus, Integrated Services (IntServ) parameters sent with RSVP or Differentiated Services (DiffServ) Code Points do not directly trigger QoS within the UMTS network. Moreover, the complexity of the current UMTS QoS solution makes it difficult to handle multiple service classes efficiently. The exploitation of the IETF Differentiated Services architecture in UMTS has been discussed in 3GPP and in Venken et al (5) and Yang and Karias (6). Support for DiffServ in the UTRAN may be included in a future release of the specification. Still, support for IntServ and IP QoS in general is missing from the UMTS core network. Fundamentally, the many layers of mapping from application layer, to IP QoS, to UMTS bearer classes, to specific air interface characteristics, make it hard to develop applications which have a consistent QoS behaviour across different networks and implementations.

For MIND, the basic QoS architecture within the managed part of the access network is based on IETF IntServ, RSVP, and DiffServ architectures. Applications use IntServ and RSVP to signal their QoS requests to the access network and the correspondent node. The core of the access network is based on DiffServ, while at the edges of the access network the IntServ parameters are mapped to DiffServ Per-Hop Behaviours. Per-flow states are only kept at the network edges, so the core of the network can scale better with the number of flows. The architecture also allows mobiles to use pure DiffServ with less sensitive flows. Furthermore, if end-to-end QoS is not available, for example, due to a QoS unaware correspondent node, the architecture still includes a localized QoS signalling mechanism to trigger QoS within the access network.

Support for QoS in the 'ad hoc fringe' can be addressed as an application of QoS routing: routing mechanisms evaluate the routing options and use alternative paths in order to support the QoS requested by the sender. This approach places the burden of QoS management mainly on the nodes at the edges of the ad hoc network, minimising the dependency on resource negotiation with the rest of the nodes in the interior. This is appropriate since there is no single administration which can negotiate on behalf of those nodes. QoS routing is fundamentally the same problem in fixed and mobile ad hoc networks, although the changes in mobile networks can be more frequent and the properties of wireless links creates a more unpredictable environment. The issue of QoS routing in the Internet has been elaborated in Crawley et al (7).

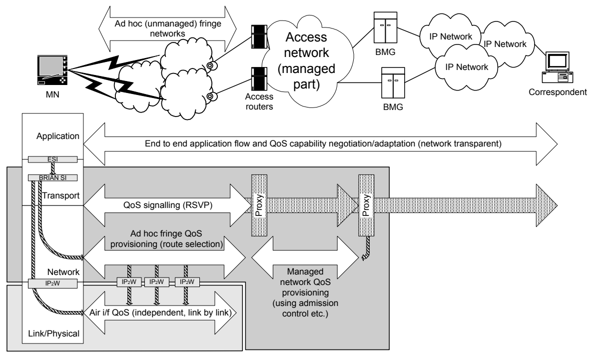

Figure 2: QoS Provisioning Architecture

To continue to provide consistent QoS support in this more complex network environment containing both managed and unmanaged parts, we aim that the QoS support within the 'ad hoc fringe' integrates seamlessly with the fixed network QoS mechanisms. This simplifies the implementation of mobiles and the way QoS is supported in a heterogeneous network environment. Since the common nominator in all these different networks is IP, we aim to extend the IntServ over DiffServ architecture from the basic managed network. The key functionality is the mapping of the IP layer QoS information to the underlying link layers at the edge of different sub-networks. The relative service levels of DiffServ would be well suited to ad hoc networks and IntServ and RSVP would enable communicating the resource request end-to-end and enable controlling the amount of traffic in the networks and protect the resources from congestion.

The QoS negotiation and admission control in a complex network as envisaged in MIND depends on the usage scenario. In the basic case, the operation uses AAA servers for QoS authorisation and deploys IntServ and DiffServ for QoS support. If a mobile is connected via the ad hoc fringe, the scenario is somewhat more complex, since the key nodes are probably those providing the fringe itself, and the nodes offering the connectivity may have their own mechanisms for admission control and resource sharing. It is then the responsibility of the routing functions implicitly to adapt the possible paths to the QoS available via each, so as to continue to enable at least some level of negotiated QoS for the end to end application flows.

APPLICATIONS OF MIND

Multi-Radio Access

There is broad consensus that post-3G networks will require the integration of different air interface types of widely different characteristics, to cover a wide range of application scenarios within a common framework. To optimise the integration it is desirable to maximise the network parts that are common between the different physical and link layers, which leads to the 'integration at the edge' paradigm introduced earlier. At the same time, adoption of a 'native IP' approach while making best use of air interface capabilities means that we must consider IP-link layer interactions in more detail.

Traditionally, IP protocols do not assume any specific functionality from the underlying link layer. For example, IP QoS architectures are based purely on IP-layer decision-making, packet buffering and scheduling through a single link-layer Service Access Point. The recent evolution of wireless communication has driven the design of wireless link layers including more functionality than what is required for simple FIFO queuing. For example, the Hiperlan/2 link layer provides priority-based packet scheduling and support for guaranteed bandwidth reservations, as described in Kadelka and Masella (8).

Unfortunately, these new functionalities are likely to introduce duplications and layering violations when the IP layer functions are taken into account. BRAIN studied a generic service interface called 'IP-to-Wireless' (IP_2W) between the IP layer and the link layer. This service interface is supported by an access technology specific convergence layer, which is meant to co-ordinate the IP QoS management and packet scheduling with the link layer and also to enable better support for address management and handovers with different link layers, for example. The resulting protocol stack functional model is shown in Figure 3. This will be extended to include support for ad-hoc behaviour, specifically issues such as address handling and routing support, and also support for unified resource management mechanisms.

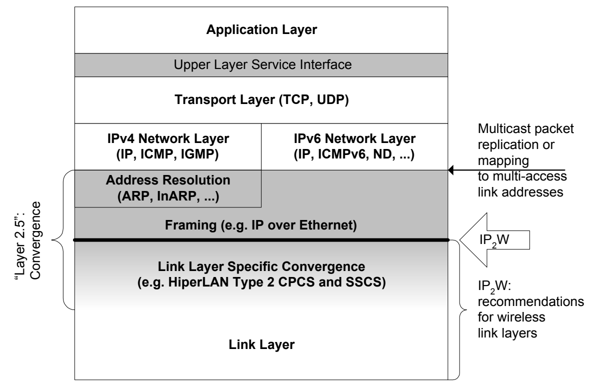

Figure 3: The IP_2W Protocol Model

Effectively the integration of access technologies at the edge of the access network allows for the normal mobility support to be equally applicable to handover between different access technologies. This creates the ability to have a seamless handover between access technologies: changes in access technology do not dictate a change in access network because this is common. Therefore, the network performance is comparable to a handover in a traditional RAN (e.g. only local routing updates). In the same way, a common QoS framework can be applied across the access technologies. This means that the QoS information from the lower layers can be used in the process of handover to ensure the application is aware of the QoS being offered on the link and then an adaptive application can adjust. Equally an application can provide QoS requirements which request the use of the best access technology for the level of QoS desired.

As has been mentioned the MIND project is looking at the area of RRM, and extensions to the IP_2W will be investigated for the support of this. RRM functionalities will need to be both above and below the IP_2W as routing occurs above but resource specific functionality occurs below. By extending the approach of the IP_2W enabling the provision of generic approaches above the IP_2W means that the possibility of a deployed RRM system working across a number of access technologies forming a complex communications network will be greatly increased. This would then provide significant input to the ability for the network to adjust and allow for access technology switching depending on demands on the network etc and allow for the more efficient use of all resources available within the system. Efficient management and use of Radio Resources is import to a bandwidth dependant industry.

This discussion of multi-radio access networks leads into the consideration of multi-homing, which is being considered within the MIND project. By the extension of IP through the RAN to the edge we can start to see multi-homed terminals being a realistic possibility. A terminal may for example have two or more co-existing connections via a singular access technology or a number of different access technologies. This raises a number of new issues which are being investigated, an example of this would be whether a terminal with two connections should advertise link availability across the two, this could cause a hole to be produced in the routing table as routes on the b-link are advertised on the a-link. This could be advantageous in terms of connectability but equally could mean that careful bounds on routing tables may have to be considered. Another issue would be that of addressing: does a multi-homed terminal have a separate address for each link or does it retain a single address advertised on both link. These are just some of the issues currently being considered in the MIND project towards the support of multi-homed terminals.

Multicast

The support of IP multicast has been defined as optional in the UMTS Packet Domain (see (9)). The GGSN will support the Internet Group Management Protocol (IGMP) towards mobile terminals and multicast routing protocols towards the Internet. Downlink multicast packets are sent point-to-point to each of the terminals that has joined that multicast group. The corresponding protocol stack is shown in Figure 4.

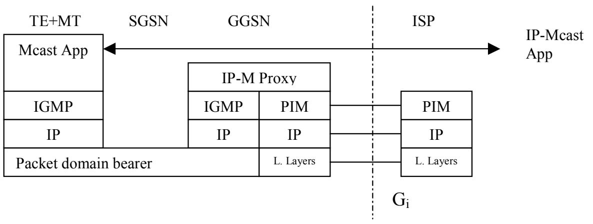

Figure 4: UMTS configuration for IP Multicast

The main goal of the multicast work performed in MIND has been to offer resource-efficient multicast support in the ad hoc fringe while retaining compatibility with native IP multicast especially in the end nodes. Several different configurations and mechanisms have been studied, and the conclusion is that the best choice is to have two different multicast routing protocols for the ad hoc extension and the access network as depicted in Figure 5.

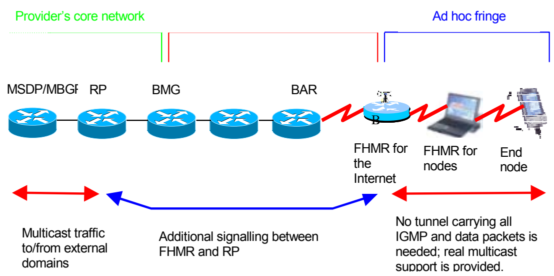

Figure 5: MIND Multicast Architecture

Here, all IGMP packets coming from the end-nodes towards the First Hop Multicast Router (FHMR) and vice-versa, instead of being tunneled or unicast, are forwarded using an ad hoc multicast routing protocol. This approach allow us to support many nodes with the benefits of using native IP multicast. The single point of failure of the GGSN is avoided, and the amount stored state towards the core of the network is much reduced (only groups joined, rather than the terminals in each group). All of this is as well as the reduction in air interface utilization towards the edge of the network where air interface resource are most constrained.

CONCLUSIONS

In this paper, we have presented a very brief overview of the architecture and major components of the MIND network, as an example of a system that can represent a direction for the next stage of mobile network evolution beyond 3G systems. Clearly, many aspects of MIND relate to the long term, for example the integration of self-configuring networks and the seamless integration of multi-radio access are still major research challenges. On the other hand, these will have to be met to achieve the user requirements for mobile networking in ten years' time. Meanwhile, we have also related the MIND technical approach to the corresponding elements of the UMTS architecture, in order – we hope – to provide some pointers for the next stages of 3G evolution.

Acknowledgement

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 Radio Systems AB, France Télécom 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. www.ist-mind.org

2. www.ist-brain.org

3. Findlay D A, Flygare H, Hancock R E, Haslestad T, Hepworth E M, Higgins D J, McCann S,2002, "3G Interworking with Wireless LANs", Proceedings of IEE 3G 2002 (to appear)

4. Keszei C, Georganopoulos N, Turanyi N, Valko A, 2001, "Evaluation of the BRAIN Candidate Mobility Management Protocol", Proceedings of the IST Mobile Summit 2001, 889-895

5. Venken V, De Vleeschauwer D, De Vriendt J, 2001, "Designing a DiffServ-capable IP-backbone for the UTRAN". Proc. of the 2nd Intl. Conference on 3G Mobile Communication Tech., March 2001, pp. 47-52.

6. Yang J, Kriaras I, 2000, "Migration to all-IP based UMTS networks", 1st IEEE Intl. Conference on 3G Mobile Communication Technologies, 2000, pp. 19-23.

7. Crawley E, Nair R, Rajagopalan B, Sandick H, 1998, "A Framework for QoS-based Routing in the Internet", IETF, RFC2386

8. Kadelka A, Masella A, "Serving IP Quality of Service with Hiperlan/2". Computer Networks, 37, September 2001, 17-24.

9. "Internetworking between the Public Land Mobile Network (PLMN) supporting Packet Based Services and Packet Data Networks (PDN)", 3GPP Release 1999 Technical Specification 29.061