An Architectural Framework for a 'Carrier Class Mobile Internet'

Paul Reynolds[1]

Orange SA & University of Plymouth, United Kingdom

Abstract

This invited paper considers an architectural framework for a 'Carrier Class Mobile Internet'. This framework is characterised by a move away from 'mobile access' (as advocated by the 3rd Generation Partnership Project - 3GPP) or 'edge mobility' (as advocated by the Internet Engineering Task Force - IETF) to one where mobility can be found everywhere. In one sense it may be called 3.5G or 4G. This new architectural framework is based upon IETF protocols and is constructed in a way that separates the concerns of access, core and services; it is also underpinned by an effective quality management system necessary to meet carrier requirements. This architectural framework is based upon the author's extensive experience working on the development of the Internet including his technical leadership of the Mobile Wireless Internet Forum (MWIF) and a contributor to the IETF.

Background – The Internet and 3G

The Internet is simply a 'community of common interest' underpinned by a shared philosophy of personal empowerment and freedom of expression: transferring the 'power to be creative' from an anonymous corporate to the individual and granting them the freedom of choice. Outsiders often perceive this as anarchical; popular misconceptions of its demographics include, male dominated, ageing pony-tailed hippies; academics with no understanding of commercial reality, etc., such stereotypes are, of course, far from reality.

The Internet community believes in the concept of a 'bazaar' approach to building networks, services and applications; allowing the individual to walk metaphorically into an open market and select the building blocks that they need to develop a service. Of course, this means that the task of integrating such blocks into a whole becomes the responsibility of the individual. This is opposite to many carriers 'walled garden' approach where you can only buy, or use, what they have to offer.

The IETF is the technical powerhouse of the Internet; it is open to any individuals who come together to form an international community of network designers and researchers concerned with the evolution, and smooth operation, of the Internet. It is a loosely self-organised group of people who exists as a collection of happenings. It has a technical focus upon the transport of data where it needs to prescribe necessary protocols at layer 2 (e.g. MPLS), layer 3 (e.g. IPv4, IPv6, Mobile-IP), layer 4 (e.g. UDP, TCP) and layer 5 (e.g. Diameter, COPS, SIP) of the OSI model. The IETF does not prescribe an overall architecture; individuals are free to use IETF protocols in any combination they deem fit: in a sense it is the environment that selects a solution or 'protocol set' by a process of market survival. But it does have a set of architectural principles that guide the protocol developer; these include intelligence at the edges, separation of concerns, and no inter-dependencies of protocols.

When the Internet was conceived no account was taken either of mobility or nomadicy; it was fixed or wire-bound. As laptop terminals became commonplace a 'wouldn't it be wonderful if we could support, seamlessly, the terminal in a nomadic mode' drove the development of a new protocol called Mobile-IP. This protocol, despite its name, was never intended to support micro-mobility (hand-off) - rather nomadicy, or macro-mobility, to allow the movement of a terminal through space and continue the session at a latter time, for example, from office to home. In compliance with the architectural philosophy, the intelligence associated with Mobile-IP for support of nomadicy is maintained within the end-points, for example, the home agent and terminal. With the rise of mobile telephony device penetrations, the support of hand-off, that is true mobility, is currently being actively considered and as a result a number of proposals circulate within the IETF including Hierarchical Mobile-IP, Cellular-IP and Hawaii-IP; in addition, a number of proprietary solutions exist.

The IETF's interest in micro-mobility galvanised the mobile carrier community, who saw the Internet as having the potential to become a major threat to their market. To regain leadership they set up a rival organisation called the Mobile Wireless Internet Forum (MWIF) whose prime task was to design an architecture, using IETF protocols, to support mobility in a way that meets their network centric needs. Recently, many MWIF ideas have been taken up by the 3GPP and 3GPP2. Once again the status quo has been, or at least appears to be, maintained.

---

[1] The author works in Orange Innovation, he is responsible, among other things, for the advocating the use Internet technologies and the development of innovative solutions based upon such technologies. He is also a Visiting Professor of Mobile Communications at the University of Plymouth. The views expressed in this paper are his own and do not represent the policy of either institution. He can be contacted at <paul.reynolds@orange.co.uk>

The Emergence of a Mobile Internet

Numerous market indicators point to the emergence of a vast mobile Internet market over the coming years. Many forecasts predict that this emerging market will reach many millions of customers as early as 2008. Along with these market forecasts, various live indicators support these predictions. The NTT DoCoMo iMode service, deservedly referenced in virtually every mobile industry article has demonstrated phenomenal growth of customers, network use and applications development. Similar experiences have occurred in Europe using short message service (SMS) and in the USA with a plethora of mobile wireless initiatives from user devices to applications. Early successes, like Aether Systems, OmniSky, and Orange SPV, point to future expansion of the market; this early burst of activity is a precursor to the more sustainable, mass-market availability of mobile Internet services. There have been disappointments too, for example, the slow take-up of WAP has lead many mobile carriers to be wary of implementing a true mobile Internet.

A distinguishing feature of the mobile Internet market is the considerable commitment already made by mobile carriers, applications providers and equipment manufacturers. Mobile carriers top this list with the signing of over one hundred contracts for 2.5G (GPRS) technologies and their commitment of over 100 billion Euro to the acquisition of radio spectrum to support 3G services. Whilst global deployment of 3G is not expected until 2005, the financial commitment to license acquisition indicates that mobile carriers will be forced to rapidly push toward service implementations. Equipment manufacturers have followed these commitments by investing heavily in ongoing technology and standards development activity. Probably, even more important is the appearance of a large community of applications developers from the World Wide Web (WWW) and Wireless Application Protocol (WAP) industries. For example, the Open Mobile Architecture (OMA) currently has over 250 members, many of whom are applications developers.

Another interesting feature of the mobile Internet market reflects the entry of various new players in the start-up mode as well as crossover from the Internet to the mobile domain. In addition, several equipment ventures, such as Megisto and Watercove, have begun around the notion of a new generation of Internet equipment focused on the mobile customer. The strong participation of new ventures in this market suggests the widespread belief that the demands and opportunities of the mobile Internet will create a vigorous and varied market consistent with the earlier generations of the Internet, Voice over IP (VoIP) services and WWW applications.

Users also show a willingness to change: early adopters of PC-based video conferencing services are willing to trade-off the inconveniences and lower quality of the calling experience for the tremendous cost savings that can be enjoyed over traditional circuit-switched video telephony. Most of the traffic generated over these early networks traverse international routes, where the greatest cost savings can be realised. Current technology also empowers users, for example, Microsoft's operating systems, with their built in SIP client, are able to drive the adoption and market size of Internet telephony, thus presenting a unique near-term revenue opportunity for new entrants to a once closed market. They will also tend to drive some mobile carriers into lower value bit transporters, as the terminal, with its intelligence and SIP client, is able to decide which carrier to use.

Numerous technical and cultural hurdles stand between the forecast and reality. Whilst there is evidence of the necessary technologies emerging in the desired timeframe, the blurring of the edge between the mobile and Internet domains poses difficult technical and cultural challenges. Technically, the Internet was never designed to support mobility whereas existing mobile networks were designed to support voice and circuit switched data. A new generation of Internet technologies, largely unproven in operation, will be required to support the mobile Internet.

Challenges of Building a Carrier Class Mobile Internet

Over time, the Internet has focused on the delivery of services to a stationary but increasingly robust customer device over constantly improving access networks. Personal computers and laptops have increased in processing power, screen quality, memory and battery power, driving an insatiable demand for web-based applications and bandwidth. The most common of access technologies, telephone access, has increased data transfer rates through improving technology, including DSL. The common attributes of these latest access techniques include high reliability and stable quality; once connected, today's Internet customer is rarely dropped and enjoys consistent network quality from the access connection.

Current market initiatives to create mobile Internet services employ the existing mobile networks for mobile access. However, a true mobile Internet requires a business model that is more than simply 'mobile access' towards one that includes such players as content aggregators, content providers, services providers, transport providers, etc. It also requires engineering, deployment and operation of a new generation of Internet-based technologies within a domain previously dominated by voice technologies.

The introduction of Internet technologies to the mobile network challenges many assumptions made in creating the current Internet. For example, access connections employing radio technologies exhibit frequent anomalies including varying bit error rates, connection bandwidth and availability. The need to support mobile Internet services will create the following demands upon emerging mobile Internet technologies:

- Dynamic IP address assignment to millions of devices that are, virtually, 'always on';
- Strong service level agreements and quality of service capabilities aimed at minimising the use of precious RF spectrum;
- Protocols that manage changing customer attachment points;
- Triple A services[2] that include support for roaming;
- Session management that maintains customer data transactions when radio connection fails;
- Extension of security technologies to allow mobile access to secured domains such as corporate Intranets;
- Application awareness that manages the relationship between ongoing applications and the current state of the radio access including Quality of Service (QoS) and bandwidth available.

These demands present the industry with significant technology development and deployment challenges. However, few of the current participants are well suited to handling the engineering, deployment and operational issues of building networks capable of delivering mobile Internet services on a 'carrier-class' scale. Most mobile carriers have little experience in Internet technologies or Internet network operation; they face a stiff challenge in developing this expertise. Indeed, most mobile carriers are unable to differentiate between 'mobile access to the Internet' and an 'intrinsic mobile Internet'. Conversely, ISPs, whilst having much of the IP expertise to handle these issues, lack mobile awareness and, due to their size and focus, have difficulty redirecting the resources necessary to address this complex and emerging market. They face organisational and technical hurdles in tackling the disruptive change of mobile services. Manufacturers often build and operate networks for mobile carriers, however, there is no historical precedent for deployment and operation of Internet based networks by such manufacturers.

Cultural compatibility is at least as important as the technical challenges of a mobile Internet; currently the two industries are virtually orthogonal. For example, the expectation of users of the Internet that services are provided free means that a business case developed that shows metered, per packet charging, for Web services is unlikely to be successful. In addition, successful mobile Internet deployment is going to require cross industry teams working together to meld Internet and mobile concepts and technologies. The status quo, where there is little co-operation among mobile and Internet service providers, will be a significant barrier to success. While numerous examples exist of the differences between the mobile and Internet industries, perhaps the best example involves the development and use of standards. The mobile industry is a standards-driven culture. Mobile industry player's focus on the development of a standard first, followed serially by the development of technology and finally leading to deployment. This approach results in deployment of mature, reliable carrier class services from an early stage but slows the time to market for many new services. In contrast, the Internet community tends to deploy technology quickly to gain perspective on market demand and uses this experience to refine the technology. The market experience, if successful, leads to establishment of a pseudo-standard, by mobile industry convention, referred to as a Request for Comment (RFC) approved by the IETF. The bottom line is that the Internet service provider industry experiences a technology evolution cycle of months, while the mobile industry's lasts years.

The mobile Internet will not be incarnated without a sense of purpose or drive from the user community. These users need to demand applications that require 'extreme Gbit/s' rates and are 'mobile aware'; the services need to be user-centric allowing individuals to tailor services to their own needs. Technologies will need to bring to bear a number of innovative solutions including low cost embedded, even disposable, radio microchips and software defined radio systems supporting both dynamic configuration of the air interface and dynamic spectrum allocation, thus eliminating the need for ownership of unique spectrum.

Mobile software agents, supporting personal profiles and personalisation of data, and, intelligent agents learning, adapting and acting on behalf of a user even pre-empting their needs and wishes will need to be deployed. These agents require minimal user involvement, so that the technology complexity is hidden from the user; these will be supported by middleware which will be used to support homogeny and autonomous adaptive networks that self manage their structure to meet ever changing demands; allow ad-hoc networking and distributed mobility management.

[2] Authentication, authorisation and accounting (AAA)

Carrier Class Mobile Internet Architectural Framework

In order to support a Carrier Class Mobile Internet that provides a means for management, control and billing, it is necessary to devise a new architectural framework, based upon IETF protocols, but containing elements and structures that enable the provision of consistent and measurable quality of service. In order to achieve this it is first necessary to split the framework into three separate architectures: for access, core and services.

'Access Network' Architecture

The mobile Internet will have a logical access network characterised by an improved capacity from the available spectrum and a move away from a cellular-only system to one that integrates separate, parallel access networks, for example, broadcast, cellular, cordless, WiFi[3], point-to-multipoint, and fixed access technologies, to an access technology independent interface within the core network supported by a mobility manager. This generic interface will use IETF protocols for mobility and hand-off management with consequence of openness towards radio access networks that can be adapted to several radio interfaces.

There is a need to support more elaborate radio access technologies and allow, for example, adaptive signal processing at both the base station and terminal. The access network will:

- Have a mixed mode radio interface with rapid adaptation to bursty traffic, for example, to support a digital terrestrial television return channel and its adaptation to asymmetrical bit rates;
- Use smart antennas at both the base station and terminal. The radio interface will be designed to take advantage of them, for example, include signalling to allow enhanced antenna processing;
- Support dynamic allocation of spectrum as loading demands;
- Use spread spectrum techniques;

There will be a need to relieve the regulated 2G and 3G spectrum whenever possible, and, to offer higher bit rates to nomadic customers when they are in a static situation by way of unregulated spectrum, for example, WiFi.

The access network will integrate broadcasting services, notably 'IP push' services. Different degrees of integration will be possible with an 'intelligent' server hiding complexity from the user. This approach raises considerable technical and regulatory challenges because it implies co-operation of heterogeneous networks that may be operated by different enterprise entities. The use of broadband coverage of specific areas or 'hot spots', for example, hotels, airports, etc., is already being provided albeit at the cost of a restricted mobility data access with techniques derived from wireless local area networks.

'Core Network' Architecture

A single universal logical core network will support the mobile Internet. Carriers will find that their core networks will evolve towards more open and universal solutions capable of supporting fixed, mobile and broadcast, allowing them to inter-work more deeply than they do today.

The core network will be populated with a number of functional elements. One key function is a Mobility Manager; based upon a Home Agent but enhanced with additional functionality in order to permit mobility control from within the network. Acting upon information reported by the access network, and in accordance with carrier policy, it will be capable of generating triggering messages telling the terminal the optimal target access technology to hand-off too. In addition, the mobility manager can store location information for each active terminal and provide the binding between the permanent home address and care of address (CoA) as required by Mobile-IP. It may also be able to analyse the Network Address Identifier (NAI) allowing the identification of the relevant Access Gateway and as such the access technology.

A second key function is a Subscriber Context Register: which is in effect a database that stores the current contexts for each subscriber, it could perform some similar functions to the HSS in 3GPP systems. 'Triple A' services are necessary to grant access each time a user associates itself to the network and performs inter-technology or administrative domain hand-off. This is associated with a Policy Enforcer that makes hand-off decisions, based on information made available by the Mobility Manager.

An Access Gateway function is also envisaged which is a logical entity that represents one access technology. It permits the allocation of dynamic CoA to the terminal roaming in the given Access Network; it is a standard IP-router that provides interconnection between the IP core and external IP networks such as the Internet or 'GPRS Roaming Exchange'[4]. It could implement complementary functions to manage QoS or security, for example through a firewall.

[3] Also known as Wireless LAN

[4] The GPRS Roaming Exchange (GRx) is the carrier intranet that interconnects the individual GPRS networks

'Service Network' Architecture

Key to the service network is a service agent whose role is to provide session processing functions, routing functions; telephony based services such as call forwarding and interfaces to external service logic. The service agent will not manage mobility, rather, it allows for the optimised relationships between the functional elements during mobility. This will result in improved end-to-end service that will be affected as a consequence of mobility. The service agent is agnostic to the bearer being used (lower layers) as well as to the applications (upper layers). Nevertheless, and as far as the mobility is concerned, the service agent must be aware of terminal mobility as it can have an impact on existing sessions under its control, for example, close or open media streams, re-negotiated codecs, etc.

The service agent will represent the user allowing the support of:

- Access to services via the best access route, in terms of cost and bit rate, without the need to manage several environments;
- Access to much higher bit rates in 'static' situations and when the required information is not personalised, for example, broadcast;
- 'Capacity-aware' access that adapts to, and makes use of, the available bit rate, for example, to synchronise files in relation to a distant computer, to download attached files, etc., when the availability of a broadband network has been detected;
- Access to the same services via variable data rates by way of content adaptation or application transcoding. This will entail the adaptation of service presentation before being transmitted so that its size can be matched to the bit rate available. This implies variable rate coding for voice and video, but also variable content formats or even service adaptation, for example, deleting or reducing images, and deleting access to certain applications.

CAN THE 3GPPS DELIVER AN ARCHITECTURE TO SUPPORT A CARRIER CLASS MOBILE INTERNET?

Standardisation within the 3GPP is focused upon a series of releases. Currently technology is being manufactured to Release 99. The architecture and much of the control functions such as session management, QoS management, mobility management are very much designed to be GPRS specific, and it relies upon an integration of 2G and 2.5G technologies onto a GSM and WCDMA based access network.

Release 4, gives focus to the Core Network and allows for IP based transport allowing CS voice to be transported across the IP core (VoIP packets). It does not affect the access networks or the terminal so is fully backward compatible: as far as the user is aware there is little to gain; perhaps it could be argued that the cost saving from using cheaper packet technology to provide a single core for both circuit and packet switched domains should be passed on to the end user.

Release 5 introduces two new functions into the core network - an IP Multimedia Subsystem (IMS) and Call Session Control Function (CSCF). In addition, the access network is upgraded to support real time hand-off of packet traffic. The IMS uses IETF IPv6 and SIP protocols, but in contrast to the Internet philosophy, it maintains control and only allows users to access services that it provides, that is, it is still is a walled garden concept[5]. The CSCF (unwittingly with reference to call, the 3GPP has exposed its telephony legacy), is a hierarchical mechanism that relies upon a series of entities including a Proxy, a Serving and an Interrogating CSCF. The CSCF too uses SIP plus a further IETF protocol called Common Open Policy Service – COPS. One further feature of Release 5 is its support of service roaming: in the traditional Internet model, such roaming is not supported, rather a user has to access their ISP by way of international dial-up toll free number or virtual private network.

3GPP2 currently has a better alignment to a carrier class mobile Internet architecture; without the constraints of legacy 2.5G systems they are able to consider the space at least compatible with Release 5. For example, their architecture uses standard IETF protocols including the support of macro-mobility with Mobile-IPv6, and significantly relies upon standard 'triple A' functionality. Thus it can be said that a carrier class mobile Internet is not supported in the 3GPPs Release 4 and 5 architectures. User-level IP[6] is not used as IP headers are not read or used for routing, being encapsulated in a 3GPP defined protocol called 'GPRS Tunnelling Protocol' (GTP), although classical IP routing is used between network elements, for example, RNC, SGSN and GGSN.

However, the opportunity exists in Releases 6 & 7 to take a revolutionary approach and adopt an architectural framework as advocated above.

SUMMARY

The mobile Internet can be defined as a simplified access for the user to all of their services across multiple radio technologies and networks. Their services adapting, notably to the available bit rate, without forcing them to manage the consequential complexities. It will be delivered by a diverse set of

[5] The Release 5 IMS, or more accurately the R5 QoS in IMS, allows for carriers' control over the access to IMS as well as to its GPRS/UMTS bearers and the associated QoS according to its own defined policies (called service based local policies). In the meantime, it allows for interconnection/inter-working with other IP multimedia systems

[6] IP Bearer Service Level

access networks controlled by a common core, with services handled by negotiation between mobile devices and a set of software agents brokering access to the applications.

The mobile Internet will be delivered over time. It has started with the introduction of nomadic services, for example, broadband access in 'hot spots' using WiFi connected to a fixed network; and access to mobile networks via a home radio interface which 'tunnels' through the fixed network. Later, it will evolve towards a more generic solution making the support for services by different networks easier, including:

– The implementation of QoS management protocols and parameters recognisable by all types of networks; each network being able to map the QoS protocols and parameters to its own requirements, for example, radio resource allocation[7];
– Generalisation of IPv6 and the emergence of consistent solutions for IPv4 & IPv6 interworking to hide complexity from the user;
– Convergence of security policies, for example, based on a triple-A architecture in the network and the SIM card in the terminal.

It will be possible to use differing technologies to support services in a way to optimise performance and cost trade-offs including:

– Auto-adaptation to the available bit rate and terminal capabilities;
– Flexible access networks capable of easily integrating new radio interfaces, allowing the implementation of higher performance radio solutions, for example, increased user peak bit rate, increased physical layer spectral efficiency, and more efficient resource allocation;
– Continuous connection to the mobile network via radio access technologies optimised for different contexts, for example, indoor-outdoor, new frequency bands, etc.;
– Applications taking advantage of the changes in communication conditions, for example, to carry out automatically certain heavy tasks when the available bit rate is increased;
– Universal core network;
– Multi-interface terminals, programmable at the radio level.

Along with this evolution, the mobile Internet architecture will need to deal with:

– The degree to which carriers can control services and service access as terminals become more powerful;
– Different business models between mobile carriers, ISPs and WiFi 'hot spot' operators;
– Any remaining regulatory hurdles, including, using broadcasting frequencies for combining telecommunications and broadcast services.

The challenge for mobile carriers in developing a carrier class mobile Internet is that they will need not only to embrace IETF protocols and technology but also adapt their philosophy to be a closer match to that of the Internet.

BIBLIOGRAPHY

'The Conflict Shaping the Mobile Internet', Next Generation Telecommunications, Reynolds, Issue 3, 2002

'IP Mobility Support', Perkins (Editor), RFC 2002, October 1996

'Mobility Support in IPv6', Johnson, Perkins, and Arkko, draft-ietf-mobileip-ipv6-21.txt, work in progress, February 2003

'Low Latency Handovers in Mobile IPv4', MIPv4 Handover Design Team, draft-ietf-mobileip-lowlatency-handovers-v4-00.txt, work in progress, June 2002

'Fast Handovers for Mobile IPv6', MIPv6 Handover Design Team, draft-ietf-mobileip-fast-mipv6-06.txt, work in progress, March 2003

'Hierarchical Mobile IPv6 Mobility Management', Soliman, Castellucia, El-Maki, Bellier, draft-ietf-mobileip-hmipv6-07.txt, work in progress, October 2002

'Motivation for Network Controlled Hand-off using IP mobility between Heterogeneous Wireless Access Networks', Reynolds, Njedjou, Bertin, draft-NC-inter-AN-handoffs-00.txt, work in progress, June 2003

'Mobility Related Terminology', Manner, Kojo, draft-ietf-seamoby-mobility-terminology-01.txt, work in progress, November 2002

DEDICATION

With over 35 years working in the telecommunications sector it was not surprising that the author is considered to be a thoroughbred 'bell-head'. In recent years he has had the intellectual pleasure of working closely with an equally thoroughbred 'net-head' - Tim Clifford. Tim worked with Paul in MWIF and joint ventures at UUNet & Lacuna Networks. They are living proof that despite their orthogonal philosophies they were able, not only to work together, but also come to incarnate jointly a new vision of a mobile Internet. This paper is dedicated to him.

Acknowledgement

I would like to acknowledge Lynn Johnson, Martin Harris, Xiaobao Chen and David Mulvey for reviews of the draft paper

[7] ITU-T has adopted 3GPP proposals for QoS classification and it is expected that 3GPP2 will follow.