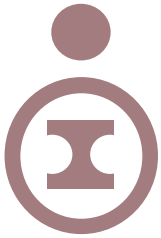
White Papers From IML

The Disappearance of Telecommunications

Why does an SMS message cost so much to send when other methods of transmitting data are so much cheaper? This paper looks at evolving network technologies and a possible impact their combination could have on the way we send data and make phone calls

Presented by Intercai Consultants
Mark Norris and Chris Angell



INTERCAI
MONDIALE LIMITED

1. Introduction

The huge popularity of text messaging is remarkable considering that the service was developed by mobile operators in the early 1990s as something of an afterthought. It was never expected to take off as it has and would be characterized by most people as a crude, if functional, service. Yet basic text messaging is now a major part of the telecommunications business.

In the United Kingdom alone, nearly 2 billion text messages are sent every month. With each message up to 160 characters long this equates to a constant data flow of just over 1Mbit/sec for the month.

But why is it so expensive to send a text message? After all, the cost of sending each 160 character message is around 10p. Considering the amount of data being conveyed this seems to be too expensive by a factor of thousands when compared with typical rates for data transmission.

By way of illustration, a domestic broadband connection can carry about half the monthly UK messaging traffic at a cost of around £20, yet the bill for the same amount of data delivered as text messages is £200M. Of course, this is something of a disingenuous comparison as point to point transfer is only one part of a messaging service. Nonetheless, the example does beg the question; why are some telecom services expensive?

One of the reasons often cited is that the equipment used to provide the services is specialised, so is expensive both to buy and to operate. But this is beginning to change. Many of the large operators are now replacing their traditional infrastructure with routers, media gateways and soft switches on the expectation that they will be cheaper, easier to manage and will enable greater flexibility in service provision.

This paper looks at the way in which technology developed for the IT industry can be used to deliver telecommunications services, sometimes with devastating cost implications.

2. Consumer Equipment

It is becoming increasingly difficult to distinguish one piece of consumer electronics from another. A computer with a wireless card and an application such as Skype™ does much the same as a 3G mobile phone and a PDA with communication facilities is not dissimilar to both of them. Yet they are quite different in the level of user customisation they allow. A PC can be extensively personalised, to suit individual tastes. The same may appear to be true of a mobile phone but strip

Skype allows calls to be made and messages sent to other Skype users over an Internet connection. It also allows calls to conventional telephones, typically for the cost of a local call. Although usually used with a PC on a fixed (broadband) network, Skype works equally well over a 3G wireless data service and would work on many other wireless data services.

away the ring tone and other superficial aspects and they all do much the same with no personalisation potential left.

There is no real tradition of providing users with the facilities to customise their phone. Indeed, the preference is to lock it into a specific service provider with a defined set of service options.

It is not unreasonable to expect mobile phones with USB connectors and Bluetooth connectivity that allow the user to configure the device to communicate as they see fit (see box). This sort of facility could spark a trend of swapping service profiles in much the same way that people now swap SIM cards – one day you may be a T-mobile customer, the next you may choose to be a customer of local wireless data provider. The latter may well use a different air interface and communication protocols but a standard PC can cope with a variety of communication modes, why not a common-or-garden mobile phone.

A mobile phone with Bluetooth connectivity could provide the gateway between a Bluetooth headset and a Bluetooth LAN access point opening up voice over the internet as an alternative to mobile technology. The user chooses a function profile (just as with a conventional mobile) which, in this case selects Bluetooth for outgoing calls and GSM for incoming (Bluetooth would not be used for incoming calls). Calls in this case include voice and SMS.

Indeed, there are already dual band mobile phones but these are expensive. They are, however, just another piece of consumer electronics. If they don't evolve to provide all of the flexibility and external connectivity that you get with other devices, it is only a matter of time before you get a telecom module on your camera, your watch or even your toaster.

3. Access to the Network

The last few years have seen a plethora of potential alternatives – Bluetooth, DECT, WiFi, wiMax and more – for wireless access to a network. Each of these offers a different approach to the delivery of telecommunications services. For example DECT gives cordless access to a fixed line service, Bluetooth, being very short range, offers hands-free functionality to mobile phones or PDAs, WiFi and wiMax provide an access point (usually at a hot-spot) to the Internet. A combination of these access technologies (such as a Bluetooth headset

Users of modern equipment generally do not want to be encumbered with a wide range of choices each time they switch on. In most cases it is preferable for the equipment to operate seamlessly and transparently so that the user can get on and, well, use it. At the same time users want the best value for money and so a system which detects alternative communication routes and chooses (perhaps generically) the cheapest without consulting the user might be the way to go. A basic GSM (or similar) handset would need the capability of downloading additional communication protocols (say Bluetooth or 802.11x) with which it could then connect to a local system. A one-off charge for the download would be added to the user's monthly telephone bill. The protocol could stay resident in the handset (incurring a high one-off charge) or be discarded when no-longer needed and reloaded when required (incurring a much lower one-off charge).

communicating with a wiMax access point) could be used to provide a VoIP (Voice over IP/Internet) service, thereby providing an alternative access mechanism to the established GSM mobile service. Alternatively a mobile phone equipped with a WiFi transceiver could transmit text messages (of any length) as an alternative to the GSM Short Message Service.

And when other service providers, with the ability to deliver long distance traffic, come on the scene the services supported by wiMax, WiFi etc al provide are easily available as an existing consumer base. A current example of this is the 'hot-spot' which gives wireless access over a limited area so that a user can connect to the Internet (and onward to their final destination). The way in which a more general, multi-service network (i.e. one to which a range of wireless technologies can work) is developing is considered in the next section.

Of course, there is an immediate concern that such piecemeal networking will result in mayhem. Indeed, there is a school of thought that believes that telecommunications must be built on the firm foundation of Homogeneity, Ubiquity and Stability (HUS) if they are to evolve. Without one or more of these foundations new technology will ultimately fail to deliver what users want. The alternative access methods discussed here will not be available everywhere thus ubiquity is missing.

However, because the various technologies discussed are here are complementing each other and the mobile networks will provide a wide coverage back-up, a different kind of ubiquity is provided together with a homogeneity through multiple service delivery methods.

In the early days of such harmonisation of communications, the stability aspect of HUS might be less than ideal but again good old mobile will be there as the (expensive) safety net. The user can set-up a profile within his CPE which chooses from a priority list of access methods, with cheapest at the top, to mobile at the bottom (assuming it remains the most expensive delivery method). Decisions are then made by the CPE and the user is left with a transparent communication system delivering services at the lowest available cost.

It appears that it is entirely feasible to serve a population with wireless. Some bands are licensed, some are not and each option has its specific characteristics, but the basic message is that there are lots of options for connecting a mobile device to a network point of presence.

4. The Multiservice Network

The traditional approach to providing telecommunication services has been to construct a hierarchical network that sets up an end to end link for the duration of a call. This model has evolved over many years and has been engineered to deliver a high level of performance and predictability. In terms of structure, the picture of a traditional network is fairly fixed with customers attached to local

exchanges (Class 5), transit and core switches (Class4) delivering bulk traffic. There is a similarly familiar picture for mobile network with base stations collecting wireless traffic to be delivered via a Mobile Switch Centre with various location registers allowing customers to be tracked.

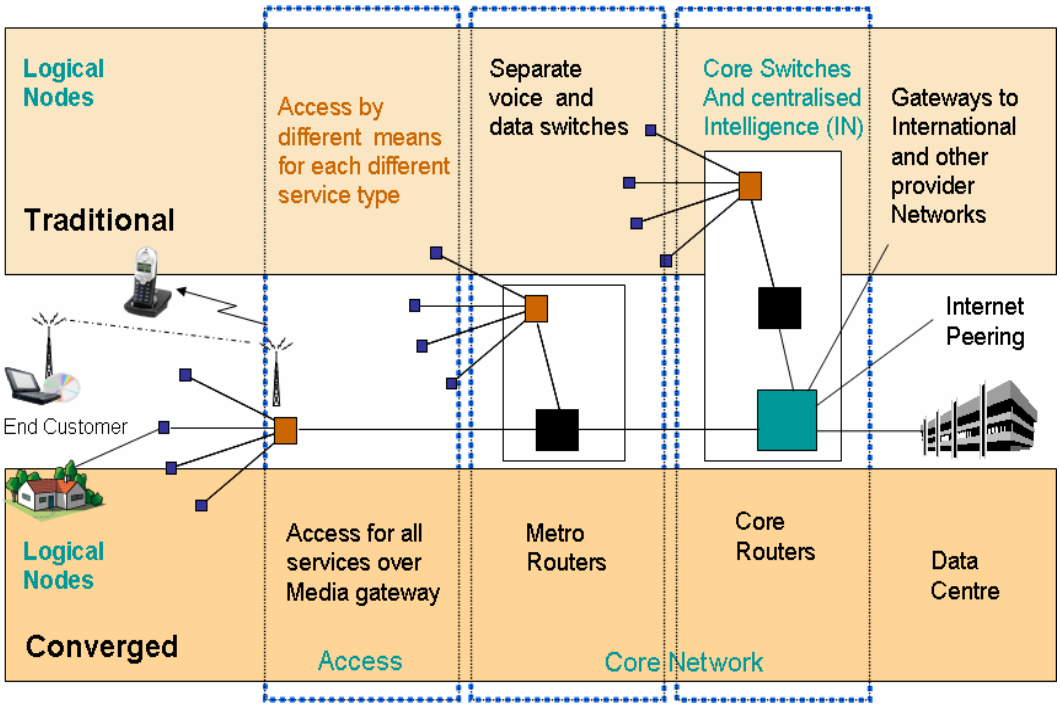
This basic picture has evolved very little over the last few years. Facilities such as Intelligent Network (IN) and Messaging services have been added and interfaces, such as Parlay (which allows services to be delivered independent of a specific network operator) have been implemented but the network, in the main, is fairly stable.

In many ways, this slow evolution is a good thing. It has allowed operators to focus on availability, quality of service, traffic engineering and all of the other operational issues that characterise a stable service.

On the other side of the coin is the fact that the established hierarchical telecom network is not well suited to the carriage of packet based data traffic. This problem is manifest in the number of overlay networks implemented to provide data services. Hence there has been growing pressure to change as the demand for data services has started to outstrip that for voice. And the advent of relatively inexpensive equipment, capable of dealing with the huge volumes of traffic carried by the public telecom network, has triggered that change.

The hierarchical network is rapidly becoming a historical view. Soft switches and Media gateways offer a cheaper and more flexible alternative. And, being packet based, all traffic types are equal – if it can be connected to the network, it can be delivered over the same equipment.

This multi-service, or converged, network is illustrated below. It is the central pasts of the diagram that are of most interest to the majority of operators as this is where they see a cheaper and simpler alternative to the status quo.



In considering how the market as a whole might develop, it is probably more interesting to look at the two ends of the diagram. To the right, there are the many mechanisms that can be used to connect to a converged network, in much the same way that an ISP connects to the Internet. Some of the more promising candidates for this were introduced in the previous section. To the left are the all of the other networks and services that can be reached once the converged network to which a user has connected has been traversed.

It would be far too simplistic to suggest that a converged network allows any sort of device to access any sort of network or service. There is a lot more to communication than simply having a common infrastructure for delivering the packets used to carry all the different service types. Names and numbers have to be resolved to ensure accurate addressing, routes have to be set up to ensure efficient delivery and priorities have to be set so that the quality requirements of each particular service are met.

These are all important issues that demand careful design. That said, there is reasonable expectation that adequate standards will be in place in all of the key areas. The way in which the various facilities outlined below combine to support converged services is a complex and evolving subject, not one that we address in any depth in this paper. But there is a sound basis for building a full range of reliable and effective services.

The basic set of protocols used to control services on a converged network

Network Routing – MPLS

Multiprotocol Label Switching, (MPLS) provides a mechanism for setting up paths though an IP network. It gives network operators a great deal of flexibility to set up preferred routes and to divert traffic around link failures, congestion, and bottlenecks. Critically, it allows an operator to manage quality of service –different kinds of data streams can be based on different priority and service plan. For instance, those who subscribe to a premium service plan, or those who receive a lot of streaming media or high-bandwidth content can see minimal latency and packet loss.

When packets enter a MPLS-based network, Label Edge Routers (LERs) give them a label (identifier). These labels not only contain information based on the routing table entry (i.e., destination, bandwidth, delay, and other metrics), but also refer to the IP header field (source IP address), socket number information, and differentiated service. Once this classification is complete and mapped, different packets are assigned to corresponding Labeled Switch Paths (LSPs), where Label Switch Routers (LSRs) place outgoing labels on the packets.

Call control – SIP

The Session Initiation Protocol (SIP) is a signalling protocol used to establish a session (e.g. a telephone call) over an IP network. It is a request-response protocol that closely resembles two established Internet protocols, HTTP and SMTP. As such, it sits comfortably alongside Internet applications making SIP controlled telephony another web application that integrates easily into other Internet services. SIP allows for the establishment of user location (i.e. translating from a user's name to their current network address) and provides for feature negotiation so that all of the participants in a session can agree on the features to be supported among them. Hence SIP is a mechanism for call management - for example adding, dropping, or transferring participants.

SIP is not a session description protocol, or a resource reservation protocol and it has nothing to do with quality of service (QoS). Other protocols are needed for these functions and SIP is designed to function with SOAP, HTTP, XML, UDDI etc as a modular component of a larger IP telephony solution – a collection of telecom related session control protocols is commonly referred to as SIP-T. In terms of function, SIP overlaps with other protocols such as H.323, MGCP, and MEGACO but can co-exist with all of them. Virtually all equipment suppliers support both SIP and these others.

Numbering – ENUM

The system of telephone numbers is well defined and understood. There are billions of telephones, and billions of telephone numbers, each one no longer than 15 digits and globally unique (to the ITU E164 standard). Alongside this system many people will be using SIP phones and SIP clients on their PCs and these will have SIP URLs associated with them. Fortunately, SIP can carry phone numbers using the new telephone URL (e.g. tel: 441628478470).

The problem arises when one needs to associate a traditional phone with a resource on the Internet (this would be a requirement of any converged presence-type service). ENUM provides a way of entering telephone numbers onto the Internet Domain Name Service so that any application, including SIP, can discover resources available to a globally unique phone number. A SIP phone or proxy server would do the number domain translation and through classic DNS resolution discover a DNS resource that would give a SIP address at which the dialled number could be reached.

Gateway routing – TRIP

When a telephone number does not have an associated SIP resource, the call is sent to a telephone gateway, which connects to the PSTN. In an interconnect environment with many peering relationships between service providers, resources in the IP network need to be able to discover which telephone numbers are associated with which gateways.

Telephony Routing over IP is a policy driven inter-administrative domain protocol for advertising the reachability of telephony destinations between location servers, and for advertising attributes of the routes to those destinations. TRIP is designed to allow service providers to exchange routing information in order to avoid the over-provisioning or duplication of gateways. It uses established Internet routing protocols such as BGP and OSPF.

The above is no more than a sample of what could be a long list of specialised protocols – to it could be included the likes of Mobile IP, as a means of supporting roaming users on an IP network, Real Time Protocol (RTP) to support voice services, IPSec to provide security and many others. Each part of the IP network jigsaw is of considerable intrinsic interest and would be worthy of more than passing explanation and analysis. However, our focus here is much more on the ease with which new operators can enter the market, so we make the gross (but not altogether unwarranted) assumption that there is enough technology to support a wide range of new applications. The challenge is more one of imagination and integration than components.

A key point in this section is that, with a hierarchical network, control is centralised , the cost of equipment is high and this tends to limit the operator base to fewer 'heavy duty' players. With a multi-service network there is a homogeneous infrastructure with distributed intelligence built on low cost computing components and a raft of standard protocols.

And this sounds much more like the Internet than a traditional telecom network, so can we expect the same things to happen? Is it likely that a move to converged networks will see a plethora of small and specialist providers, rather than just a few, big full-service operators? Will converged networks as a whole experience the same sort of organic growth as we have seen with the Internet or will the tightly controlled and planned evolution of the established telephone network prevail?

In the next section we consider some of the possibilities and consequences of converged networks. And in particular, we go back to our introductory remark to ask whether this new infrastructure holds the prospect of shipping text messages at a price commensurate with the amount of data being sent.

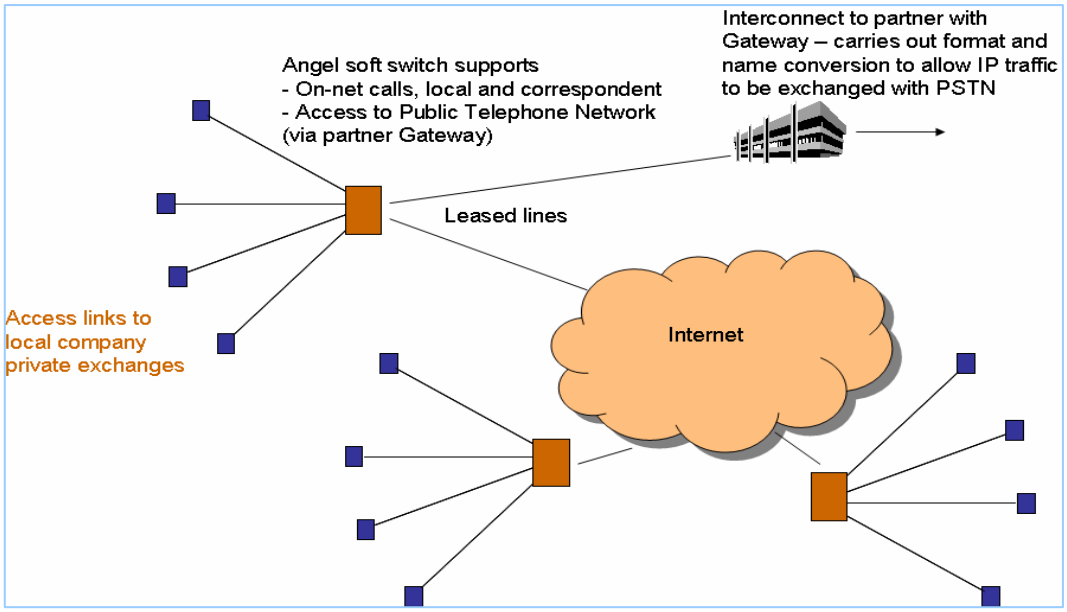
5. Scenario

It seems inevitable that a communications network comprised of inexpensive components that operate to open standards will invite competition. So, it is increasingly likely that new operators will be able to take advantage of the new direction in telecoms to offer competitive services, and probably at much lower rates in some instances.

Given the sheer variety of possibilities, there is no one path that they will take – the differences in customer need, geographical constraints and technical options means that each new age telecom provider will have to identify its role and the most appropriate selection of hardware, software and operational components needed to fulfil that role. That said, let's look at a couple of examples of what could readily happen in the years ahead in a brief history of the near future

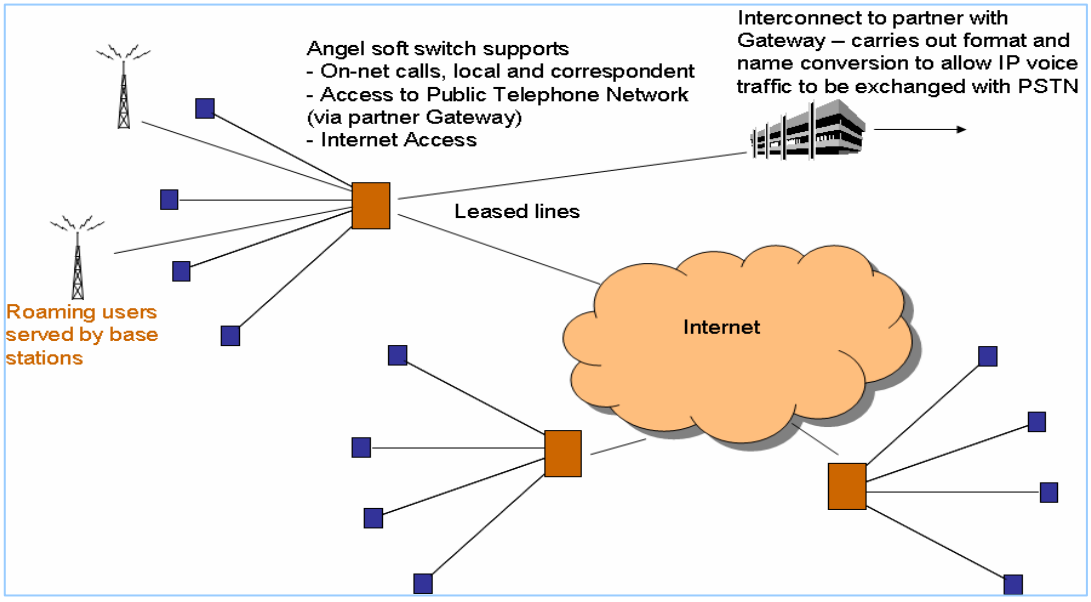
Angel Communication

This is a small operator that was set up in 2006 to serve the business community in and around the Salisbury area. The company is centred on a small but highly capable Metaswitch soft switch (which was initially used to delivers voice traffic to



and from the private branch exchanges of several co-operating organisations). Most of the traffic was on-net but Angell also managed connections to public networks.

In early 2007, Angel realised that the addition of two base stations to their network would allow them to expand their service offerings. In addition to supporting mobile telephony, they also move to provide a data service to their customers. They update their soft switch with Mobile IP, which allows them to manage roaming connections and make the necessary changes to their interconnect arrangement.



Although limited in terms of the equipment they deploy and the extent of their cover, Angel offer all of the services that their customers have come to expect. Furthermore, their closeness to their customers enables them to be responsive to their need. Both parties are happy – Angel has a viable business and their customers have a flexible supplier and reasonable rates.

Norwest communications

With the appearance of a new range of mobile phones, they decide to go in to the messaging business. They set up radio kit on business parks which are linked to their Cisco softswitch. Users can elect to use the service from their existing mobile phones – the message destination is translated for carriage over an IP network that has differential quality of service set up.

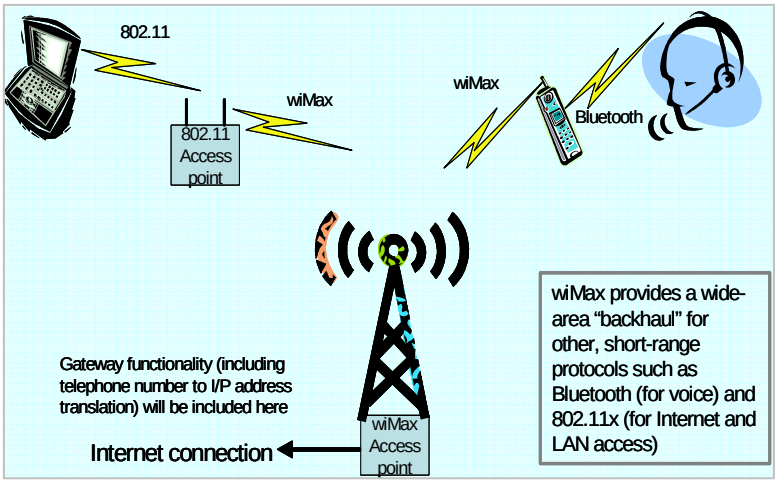
So what are the future prospects for the habitual text message user? If all of the advances mentioned above come to fruition, it is entirely possible that smart new operators will take advantage of the new order in telecoms and that texting from mobile devices will become a 'free' service, similar to the instant messaging facility that comes with most on-line services such as AOL's ICQ or MSN.

This is helped, in no small measure, by the current developments that specifically target such services. The Session Initiation Protocol (SIP) standard, central to the operation of the converged network is being extended (as defined in RFC 3428) to provide basic Instant Messaging facilities. So the issue of providing a service over a converged network (once connected with an appropriately configured device) is largely covered.

There are still some open issues, such as the availability of the presence management data needed to establish whether a recipient is available to accept a call/message, but the basics are all in place. There is no compelling reason (at least, not a technical one) why a mere 160 characters of text should impact a teenagers bank balance. .

6. A Customer-Focussed Scenario

Aine Marks, a savvy communications user of Suffolk, is called to a conference in Portsmouth. At the conference centre facilities for wireless access to the Internet are provided via a local gateway to a wiMax radio access point connected to the local area network. Aine generally uses GSM with GPRS as well as SMS, but she notices (through advertising around the conference and in her hotel) that a local ISP is offering a "pay-as-you-go" wiMax and internet access bundle which would allow her to send messages and even use voice over the Internet via her usual mobile device.



Accessing the service is simple requiring a single menu selection to allow her handset to download the required protocols and register with the ISP. The ISP would bill Aine through her usual mobile service provider. The cost of the service for the short time she will be in Portsmouth, is very low, and will allow her to keep in touch with business colleagues and family without running up a huge bill. She still receives incoming calls and texts via GSM, however all outgoing communications will be, for the duration of her stay in Portsmouth (at the conference centre and her hotel) via the locally provided wiMax service.

7. Conclusion

It is too early to pass any sort of verdict on the impact of converged networks on the telecommunications market. However, with many of the barriers to entering the telecoms business falling, there is considerable expectation that services will be inherently cheaper to provide than is currently the case.

Taking a pessimistic view, it is possible that the market will fragment and that service provision will, in many cases sink to the lowest common denominator, with consumers confused by the range of service providers and disappointed in the quality of service provision.

A more positive view would be that the installation of a uniform IP-based infrastructure, supported by a raft of facilities for service delivery will lead to more consumer choice (and attendant confusion/responsibility) with prospect of much lower cost.

If the trend of regulators pushing incumbents to open their network to, for example, offer wholesale broadband continues it is reasonable to expect small operators to take advantage of a ready made delivery mechanism for their wares.

Several consequences of this have been explored in this paper. Prime among these is the way that the move to IP will have some impact messaging services. Doubtless there will be an impact on the voice market but with prices already low it seems more likely that the real impact will be in messaging as there is prospect for dramatic cost reduction.

And the key to this is provision of more commoditised cell phones plus wide area wireless access. The core network elements and mechanisms for their control will soon be in place. It seems likely that a converged network will trigger converged user devices and that a multi-service capability will admit multiple access technologies.

References

1. A good example of the application of wireless technology can be found in the white paper 'Bluetooth…' by Chris Angell. This paper can be found at www.intercai.co.uk along with a white paper by Steve Hodson that fully explains the relevance of Homogeneity, Ubiquity and Stability on the provision of telecom services

2. Detailed information on the way in which IP can be applied to provide converged voice and data network service can be following

- *Internet Communications Using SIP: Delivering VoIP and Multimedia Services with Session Initiation Protocol*, Henry Sinnreich and Alan B. Johnston, John Wiley and sons, 2001

- *Voice over IP: Systems and Solutions*, Richard Swale, Institution of Electrical Engineers, 2001

- *Mobile IP technology for m-Business*, Mark Norris, Artech House, 2002

- *A professional's guide to data communication in a TCP/IP world*, Bryan Carne, Artech House, 2004

3. Many of the protocols referred to in this paper are available from the Internet Engineering Task Force (www.ietf.org) as 'Requests for Comment' (RFC). For instance, the Session Initiation Protocol (SIP) is RFC 3261.

4. There are several books on the various wireless technologies, notably, *Advances in Mobile Radio Access Networks*, Jay Guo, Artech House, 2004

5. A good commentary on the evolving state of the telecoms market can be found in *The Great Telecom Meltdown*, Fred Goldstein, Artech House, 2004