Chapter 1: Introduction

Not another book on mobile? Come on Wisely, how an earth can you justify yet another book with 4G in the title? I suppose the best answer is that I have never seen a book quite like this one! In truth it grew out of a short, two day, course on the future of mobile that myself and a few colleagues have given at Oxford for the last few years. The course was intended as an overview of the key up and coming technologies as well as commentaries from leading experts in each field as to why they were important and what the key issues were. In trying to find a book to support the course it proved impossible to find something that was both comprehensive – covering all the topics – but also having sufficient detail to avoid missing the key facts about the technologies. There are plenty of books dedicated to (say) the IMS or 3G but they are jam-packed with the nitty gritty detail of the protocols – required reading if you are working in the area, but if you just want a quick overview then they are very hard work indeed. Overview books seemed equally unsatisfactory; suffering from being very unevenly written – with each chapter often written by different authors of varying quality and consistency. Worse still was the “Spangliase effect” – an affectionate term by one of my colleagues for English prose written by native Spanish speakers that seems to be grammatically correct but just sounds odd and is hard to read fluently. In addition you really got only a very superficial level of detail – which often didn’t offer much insight into the technology and its shortcomings (such as why Wireless LANs (WLANs) have such high power consumption and why it is so hard to do anything about it). Both types of books were also weak on the commercials – lots about the technology but no commentary as to whether it would be a success or what the leading operators really thought about it. For that you really need to attend some of the expensive non-academic conferences that are run by the likes of IIR or Marcus Evans. If you are not a speaker then they can cost upwards of £2000 a go. You also probably need an insight from key operators and vendors – things they are very wary of talking about in public.
So, to summarise the case for the defence, this book attempts to cover all the key technologies in the mobile space at a decent level of detail – between the horror of describing all the protocols and a shallow summary. I have also tried to tackle the key commercial questions and offer an opinion on the main issues in mobile today. Lots of people should read this book – I need the money! – no this book is really for people in the industry who want an update into what’s happening, for people with an engineering background who want to know how mobile comms work at a deeper level and for all those with an interest in mobile comms over the next few years.

So what of your author? Well I have been in telecoms for 20 years at BT. I have worked in mobile for 10 years – looking at WLANs before they were popular, and working, for the last few years, on convergence between fixed and mobile. I was technical manager for the EU BRAIN/MIND projects and am now heading up BT’s convergence research. In relation to bringing you this book I have been to many conferences – boring audiences from Moscow to Melbourne – read many dull books and scoured the Internet for news and useful links. I have endured a dodgy shoulder (leaning over tables), vibration white finger (bad typing habits) and pushy salesmen from mobile vendors. In order to survive all this you need a very good sense of humour and you will find I have included a few jokes along the way to lighten things up.

If you have an engineering background then you can easily follow the book – you might need a primer on IP or Mobile radio and GSM – but there are a couple listed at the end of the chapter – although most of the points on mobile are included in the book. If you don’t have an engineering background then there is still plenty of commercial insight to be had and I have tried to isolate the more detailed sections and make them as self-contained as possible.

The book follows on a bit from *IP for 3G* (Wisely et al.) which was published in 2002 – which covered IP more thoroughly with chapters on IP Mobility, IP QoS and SIP (Session Initiation Protocol). Things have moved on a lot in mobile – in my view this is a very exciting time for mobile as this chapter explains – with IP QoS and mobility proving to be rather “red herrings” – in that mobile standards haven’t really used them and the IETF has been very slow at moving these on. It is rather the air interfaces – WLAN, WiMAX, HSPA, LTE – that are the buzzwords for the industry and that is how I have chosen to structure the chapters. They sort of go in an order of 3G, WLANs, 3.5G, WiMAX and then a couple of chapters on service creation (IMS) and mobile services themselves. I apologise for readers who already own *IP for 3G* because the chapter on 3G is lifted straight from that book – in order to make this book complete in its own right (if you feel hard done by then email me and I will send you a copy of my Oxford slides that accompany the course for free).

---

1 Internet Engineering Task Force – the experts who make up the Internet protocols.
2 WLAN – Wireless LANs, HSPA (high speed packet access – next phase of 3G – sort of 3.5G), WiMAX – Worldwide Interoperability for Microwave Access (sort of wide area WLAN) and LTE – long term evolution, the mobile industry equivalent of WiMAX. Much more about these later.
The rest of the introduction is really setting the scene – it includes a prologue that looks at the history of mobile (very briefly – thank goodness I hear you say), introduces the props (the key technologies) as well as the main characters. After that it is up to you in what order you read the book, but it has been written to be read linearly and the later chapters assume you know some of what precedes them. I hope you enjoy reading the book and that at least something in it is useful or makes you smile.

1.1 Prologue – The Generation Game

Cellular mobile has been around for over 30 years – Figure 1.1 shows a 1970s scene with what might be called a zero generation mobile. Things have moved on a lot since then – not only in fashion and hairdressing – with the launch of analogue cellular systems in 1980 (1G). Such systems were characterised by their lack of roaming – many countries developed their own system – lack of security and poor battery life. Second generation (2G) digital systems were launched in 1990. Several systems (see Table 1.1) were launched but it has been GSM, 18 years later, that has come to dominate mobile communications around the world. As of March 2008 there were a little over 3 billion GSM phones in use around the world – representing about 85% of the total. 3G, by contrast had only 350 million or so subscribers (although rising fast) (Ref. 1). It is also reckoned that GSM will not be switched off until 2018 at the earliest in the UK alone. It could almost be argued that the whole success of mobile is based on GSM – standardisation, volumes and falling costs driving a virtuous circle of success.

Figure 1-1. Zero generation mobile. (Source: BT Heritage. Reproduced by permission of BT Heritage © British Telecommunications plc.)
2G was about voice – good quality mobile voice with roaming, secure encryption and improved talk and stand-by times. SMS – the now ubiquitous messaging service – happened almost by accident, since it was originally intended as an internal communication system for GSM engineers. Even today voice and SMS accounts for over 90% of the typical non voice revenue of UK mobile operators (Ref. 2). General packet radio system (GPRS) – often called 2.5G – was introduced to allow (nearly) always-on data connection and volume-based charging for data sessions. Previously, users of wireless application protocol (WAP) – a menu-based, cut-down, version of HTML specially designed for handsets – had to have a circuit connection (with time-based charging) open for the duration of the session. GPRS offers data users up to about 50 kbit/s – something like dial-up rates – which improves the WAP experience. 3G was launched in Europe on the Isle of Man by BT Cellnet in 2001 and was sold as offering users communication “any time, any place, anywhere” – the so called Martini effect. It was intended to offer high-speed data services, video telephony, lower costs and a host of novel services as well as the “mobile Internet”. In Europe at least, however, take-up has been slower than predicted with currently 8% of UK mobiles being 3G (some 5 million out of 65 million [Ref. 3]) with most used only for voice and SMS. Only now are we seeing a move to mass market mobile data as enhancements to 3G’s data capability have been rolled out and prices dramatically reduced.

### 1.2 The Props – WLAN, WiMAX and All That

Figure 1.2 shows (admittedly not the best version) of the most famous diagram in mobile. However, it is a very useful starting point for understanding why
there are so many mobile technologies. The fundamental thing about radio is that there is always a trade-off between range, bitrate and mobility. Imagine you have been given a chunk of spectrum (maybe you bought it in a government auction) and you want to get busy in the shed and build a small mobile radio system that you and your community can use. There will always be some power constraint on how much power you can transmit – either because of restrictions on the spectrum (someone else shares it) or the battery power on the mobile terminals or just the cost of the power amplifiers on the base-stations. You then need to decide what bandwidth the users need – this will be much higher for video than for voice say. Let’s, for the sake of argument, say you want video – 1Mbit/s for good quality on a PDA maybe. This will set the range of your system – in conjunction with maximum transmit power and the frequency of the spectrum (the lower the frequency the better the range). If you had 10MHz of spectrum you could give your 10 neighbours 1MHz of dedicated spectrum. But then more people want to join and you soon realise that most of the time the spectrum is unused – so you decide to have a pool of spectrum resources and allocate it as users actually want to use it – 100 people can join as, statistically, only a few of them use it at any one time. Then more people want to use it – a bit of research shows that fancy codecs and compression of the video means that users don’t really need a 1Mbit/s video stream. They can send only the parts of the scene that change – suitably compressed. This has a bursty traffic profile – when the scene is slowing changing there is little traffic – so you can multiplex many more users – say 1000. But now Quality of Service is an issue – you need complicated protocols to signal, admit and police requests because the system
capacity is near to maximum sometimes and the bandwidth required fluctuates. Then more people want to join – so you need to upgrade the system but this is going to cost and so you want to start charging. However, some people pay but others are freeloading – so you need a security system for authenticating and authorising users. Then, with all this cash, you want to extend the system to people on their bikes and in their cars. You soon notice that your coverage range is much reduced as the mobility of the users rises (basically because of Doppler and other radio effects associated with movement). In order to keep the coverage area – after much debate with the vicar about mammon and greed in society – you pay him a hefty whack to put up another base-station on his church and pay BT to connect it to your servers. However, users moving from the coverage area of one base-station to another want seamless operation – so you have to engineer a handover solution and make sure the cells overlap but not on the same frequency otherwise they interfere. Congratulations – you are now a mobile operator.

What this fanciful romp shows is that mobile systems have a number of attributes that are trade-offs:

- Data rate;
- Capacity;
- Coverage;
- Mobility;
- Frequency planning.

And some that are optional in that they are not essential but desirable:

- Quality of service support;
- Mobility support;
- Security;
- Battery life;
- Handover support.

Overall the laws of physics don’t change – which is a pity as there are quite a few “interesting things” I’d like to do that contravene them. What really happens, as technology evolves, is that Moore’s law and electronics/software advances and gets cheaper to the point where mobile systems are better able to approach the fundamental limits of capacity that are set by the laws of physics.³

³ Such as Shannon’s law which basically says the capacity of a comms channel is proportional to the bandwidth of the channel and decreases as the noise increases. In mobile systems this noise mostly comes from other users – within the same or neighbouring cells. Modern systems – like WiMAX and LTE are said to approach the Shannon limit.
You can see this in Figure 1.2 where I have drawn a hyperbola to show that for most systems bandwidth range is constant. The constant goes up in time – as electronics and software advances and becomes cheap enough for mobile systems – but slowly. I reckon the equivalent Moore’s law constant for mobile (i.e. how long it takes mobile efficiency to double, all other things being equal) is about five years – i.e. the capacity of 1MHz of spectrum doubles every five years and quadruples every 10 years or one mobile generation.

So the different technologies you will read about in the next four chapters (3G, HSPA, LTE, WLAN and WiMAX) offer different solutions to the above functions and different trade-offs. Apart from WLANs they offer complete solutions – with a fully integrated architecture and tightly coupled functions. WLANs are the odd man out in that they provide only a minimal set of functions but very high data rates over a short range. The WLAN chapter will explore the consequences of this and explain how this came about.

### 1.3 The Players – The Jester, the Harlequin and the Guy with the Beard and Sandals

The players – well they are a mixed assortment. Firstly, in order of importance I suppose, come the mobile operators (MNOs they are often called – Mobile Network Operators). The MNOs are the control freaks of the industry. They sell most of the phones – often in their own shops. They control authentication through the SIM card. They also are the guardians of the services that you are or are not allowed to have on your handset and/or use on their network. This comes about because most handsets are subsidised (in Western markets particularly – even Apple is now allowing the iPhone to be subsidised). When you get one from a MNO the deal is that the handset is cheap but it is loaded with applications specific to that operator and the whole phone is then “locked down” to prevent alteration of the settings. Worse still the handset is locked to the network so that you can’t just buy a SIM from another operator and drop it in your existing handset. Mobile operators hate that sort of thing. They also have battled long and hard against the Internet business model (Figure 1.3) – launching their own

![Figure 1-3. Comparing the Internet (top) and Mobile (bottom) value chains. (Source: Author. Reproduced by permission of © Dave Wisely, British Telecommunications plc.)](Image)
“Mobile Internets” – which were in fact walled gardens of sites that the MNO has selected (i.e. mostly those that generated a rake-off). The MNOs are still in land grab phase in the developing world – signing up armies of new customers for voice. In the developed world the game is to offset static user numbers and falling voice revenue with new data services.

You have to feel sorry for the mobile equipment manufacturers. It has been said that since the .com bust the total number of Telecoms jobs has shrunk by 500 000. It is the mobile vendors who have taken the biggest hit with major mergers – such as Alcatel-Lucent and Nokia-Seimens – showing the consolidation going on in the industry. Mobile vendors are under threat from newer rivals from China and the Far East – such as Huawei and many new handset vendors – who are pushing down prices. Vendors are very keen that MNOs upgrade their networks with the next version that has been carefully worked on by both of them in standards. They are not keen on “left field” initiatives from the computer industry vendors to offer different systems in the form of WiMAX and WLANs. The computer industry – led by companies such as Intel – however, has been busy adding functionality to WLANs and developing WiMAX – which has been called “WLAN on steroids”. If all the laptops and PDAs have WLAN and WiMAX in them then that is a huge boost for these technologies over the more traditional systems from mainstream mobile vendors.

Spare a thought for the customers who, ultimately, have to pay for all this. Customers really like mobile phones and they really like chatting on them. If you live in a developed country and haven’t got at least one mobile then most people assume you are either a hippie or living on the streets. People have been paying a “mobile premium” of 10–100 times for voice (over and above what they would have paid on a fixed phone) for the privilege of being mobile. Even today this is still 2–10 \times \text{C} for voice (in-country) and much higher when roaming. Most users say they would talk for longer if tariffs were cheaper – which they have been getting for the past few years. Users have also taken to texting (luv it or h8 it) which has also been a massive hit. Users have not taken to video-telephony or picture messaging and are just showing an interest in the Mobile Internet.

Then we have the standards guys (I say guys because 95% of them are men). Standards are very dull – very long tedious meetings in dreary locations arguing about minutiae and eating too much. However, without standards there is no way the mobile phone revolution could possibly have happened. It has lowered prices by an order of magnitude compared to the fragmentary diverse standards of first generation that were different for pretty much every country. GSM now has 85% of the world market with handset prices falling year on year as volumes have ramped up. There are two standards “camps” that we will be mainly concerned with. The mobile camp is dominated by 3GPP and 3GPP2 with various industry for an ancillary to this. Then there is the computer camp with the IEEE responsible for WiMAX and WLAN standards. The IETF (Internet Engineering Task Force) – otherwise known for their resemblance to 1970s heavy metal bands (all beards and sandals) – are responsible for IP standards.
Finally the industry police (regulators) have a major role in the play. They control the spectrum – the “fuel of mobile”. They also have the power to set prices and to order additional competition. This is not the place for a debate about the merits of regulation but its effect can be seen in the price of DSL across Europe where, as we shall see, it has a major effect on the possible future of mobile.

A mention of the bit-players perhaps? The fixed operators are busy trying to deliver something called convergence – where the fixed and mobile worlds become part of the same telecom experience. They are mostly just trying to avoid fixed to mobile substitution which is gaining traction – in Austria 20% of people don’t have a fixed line. Then there are the Internet giants – Google, Amazon, Microsoft, Wisely.com – who are looking at the mobile space as a way to break out of the fixed Internet. And what about the researchers? – underpaid, locked away in dingy labs, no appreciation .

1.4 The Plot – The Elephant in the Room

The elephant in the room is the Internet. The Internet is now everything in communications – despite the fact that there are four mobile phones for every PC (a fact the mobile vendors are very keen to tell you) – this does not reflect where the economic power lives. Ninety-five per cent of those handsets are used for voice and text and less than 1% have been used to book tickets or watch videos. The power of the Internet giants (Amazon, Google . . . ) is so large, as we noted above, that the mobile operators have started to give up trying to build a separate mobile Internet and are partnering with them.

IP is the key to the Internet and to understanding where mobile is going – in my view anyway. Hence the title IP44G even though developments in IP itself are only a minor part of the book (in fact IP hasn’t changed much in the last five years – IPv6 was just over the horizon then and is just over the horizon now). It is worth remembering what IP really is and why it has been so successful because this is what is driving the likes of LTE and WiMAX as well as the IMS (Internet Multimedia Subsystem – IP service creation for mobile networks as explained in Chapter 6). Firstly, let’s think what IP is – Figure 1.4 shows the standard IP stack – which hopefully brings some recognition? The key point about IP is that you can put it over anything – any layer 2 technology from ATM to Ethernet to carrier pigeons (there was an Internet draft about IP over carrier pigeons (Ref. 4)). The other key point about IP was that you could put anything in the IP packets – music, voice, gaming data – and the packets were treated just the same by all the routers. This is called the transparency/layering principle and is very powerful and meant that all the services and control (old and new) was de-coupled from basic IP operation. “IP over everything and everything over IP”. IP is the bottleneck and the stack might be more appropriately drawn as in Figure 1.5.

The other big design principle of IP is the “end-to-end” principle – meaning that end points are the best place to put functionality whilst the network should be dumb – just transporting IP packets around but never opening them or performing functions on them other than routing them to their final destination.
Figure 1-4. The IP stack (Source: Author. Reproduced by permission of © Dave Wisely, British Telecommunications plc.)

Figure 1-5. IP as the bottleneck (Source: BT. Reproduced by permission of © British Telecommunications plc.)
It is important because it allows IP systems to be very flexible, to do many different tasks simultaneously (from serving web pages to voice flows to downloading files and so on) and to do new things without changes being required in the network. The IP architecture is very loose – with one protocol per function and no coupling between them.

2G mobile has none of these characteristics! It is what is known as a stovepipe solution – meaning every part of the stack is interdependent on every other part. This made sense when the service set was very limited – in the case of GSM it was just voice (SMS was originally for engineers to signal to each other). Each part of the GSM stack was optimised for voice and the result was far more efficient than a voice over IP system could ever have achieved. But, and this is the big but, when it came to IP services GSM was next to useless. It was also very hard to upgrade or change even for voice – offering higher voice quality (so that you could listen to CD quality music say) was impossible because that would have necessitated changing all the base-stations, handsets and so on. Even trivial new services – such as short code dialling (e.g. 901 to get my messages) took a long time to appear and needed a new standard and major upgrades to the network just to make it work for roaming customers. Figure 1.6 shows the general idea.

3G was born before the Internet. As we shall see in the next chapter 3G designers opted for a revolutionary air interface and an evolutionary network. The result: 3G is very good for voice but very bad for IP. It is bad for IP firstly because the air interface doesn’t cope well with IP applications – the throughput and capacity is limited in relation to current IP applications. The latency (end to end delay) is also high (300–500 ms) compared to DSL as is the time taken to start a browsing session from idle of a couple of seconds. Secondly, there is nothing in the standards that allows the MNOs to create and control Internet services – they are reduced to a bit carrier with the “value-add” escaping to the likes of Google – whilst this is perfectly in line with the Internet business model it is not well aligned with the existing mobile business model – that covers services and content as well as connectivity.

![Figure 1-6. The IP way and the Telco way. (Source: BT. Reproduced by permission of © British Telecommunications plc.)](image-url)
So 3G was never going to be the solution for the mobile Internet. Two solutions to carrying IP packets at DSL-like data rates came forward. The computer industry came up with WLANs – for connection to portable computers – PDAs and laptops. The mobile industry came up with 3.5G – enhancing the existing infrastructure and the 3G air interface in particular to better cope with IP. The two camps have also developed WiMAX – once called “WLAN on steroids” – and LTE (3.9G) – see Figure 1.7. Both are very similar in having a very efficient air interface that is optimised for IP and takes advantage of electronic progress. Both technologies feature “all-IP” networks that are “flatter” (i.e. have fewer layers) and more like routed IP networks.

The mobile industry also came up with the IMS – a platform for service creation and control that would allow them to create services and capture value from Internet services (such as Voice over IP, video calling and Instant Messaging).

So, you see, everything in mobile is being driven by the need to carry IP and IP services. The Internet is now so big and so valuable that if the mobile industry is to move on from voice and messaging then it will have to be through IP. It doesn’t matter that the MNOs may or may not save money with IP or that it is or is not technically better – it is simply the power and size of the Internet that is driving mobile forward. People will pay a mobile premium for a service they use in the fixed world – they paid a huge premium for mobile voice. If mobile operators follow the same paradigm of taking a successful service and making it mobile then they may expect at least some premium for the mobile Internet. Whether they capture any further value is a moot point and one of my top 10 questions posed below.

1.5 Epilogue – The Road to 4G

I think this is a very exciting time to be involved in the mobile industry. With HSPA finally offering some of the promise of 3G as a technology capable of
delivering mobile data and Wireless Cities – whole centres covered by WLANs. Then there are the new radio systems – WiMAX and LTE – with “magic bullet” technologies like OFDM (Orthogonal Frequency Division Multiplexing) and smart antennas – systems that promise to be all-IP from the start.

With so much hype and conflicting messages from different parts of the industry I would like to set 10 key questions that we can return to in the conclusion to see if the book has shed any light on:

1. Why has 3G been so slow to take-off?
2. What really is the Mobile Internet?
3. How good is 3.5G?
4. Why are WLANs so limited (range, power consumption etc.) and is that going to change?
5. Is all the hype around WiMAX justified?
6. Is LTE just the same as WiMAX?
7. Will the fixed Internet business model be carried over into the mobile space?
8. How will mobile operators create attractive services of their own? And will they succeed?
9. What services (if any) – over and above voice and messaging and the Internet– do users really want?
10. How much are users willing to pay for new services?

All of these questions are linked to the issue of “what is 4G?” In some sense that is the main theme of the book as we follow the attempts of the mobile and computer worlds to deliver mobile Internet. Of course it is quite hard to define a mobile generation (as there have been only three) but what has distinguished them so far has been:

- A gap of about 10 years;
- A completely new air interface and network;
- New terminals;
- About 4 times the capacity of the previous generation;
- New spectrum.

On this metric only LTE and, possibly WiMAX qualify as new generations. But not everybody buys this argument (Ref. 5). Some people believe that converged solution – with many different access technologies such as: WLAN, WiMAX, 2G and 3G is what 4G is really about. Others say it is defined by a
service set. All the chapters in the book build towards an answer – which is revealed in the final chapter after the above questions have been revisited.

The ITU (International Telecommunications Union – see the who’s who of standards in the next chapter) have defined 4G as 100Mbit/s – and on that scheme neither LTE nor WiMAX qualify. Indeed LTE has been officially registered with the ITU as a 3G technology (insiders call it 3.9G) with something called “LTE-advanced” destined as the 4G version. Much of this can be put down to the industry trying not to let 3G seem dated before it has achieved a significant market share (around the world 3G has less than 10% of all connections) and an attempt to talk down spectrum prices in auctions currently taking place. 4G is still a confusing term in the industry and one of the topics we will return to when we have seen what some of these technologies have to offer.

If you don’t know about the history of 3G and how it works then I suggest you shuffle off to Chapter 2 and find out. If you know all about mobile then move directly to Chapter 3 on WLANs – if you know all about those then you are a smart Alec who should be on Mastermind and not reading this book!

References

[4] IP over Avian Carriers (IPoAC) – RFC 1149. On 28 April 2001, IPoAC was implemented by the Bergen Linux user group who sent nine packets over a distance of three miles each carried by a pigeon – they received four responses.

More to Explore

Radio Tutorial


Overview of GSM

– try http://www.iec.org/online/tutorials/gsm/
Or http://www.palowireless.com/gsm/tutorials.asp which has a handy list of web resources on GSM tutorials.

IP design
