University of Manchester
School of Computer Science

Comp60242: Mobile Computing 2010

B1. Overview of mobile computing & communications

BMG Cheetham

1 Introduction

This course is about computing hardware and software designed to be used in locations that are not necessarily fixed. The definition encompasses both mobile computing and telephony where wireless (radio) provides links to networked 'base stations' or 'access points' with provision for 'handover' from one to another. Wireless networks supporting mobility may be termed either 'Cellular' (as evolved from traditional cellular mobile phone networks) or Nomadic (which refers to wireless LANs, PANs, cordless phones & maybe WANs). Include satellite communication links as cellular (with large cells).

Until quite recently, a computer was a large and heavy piece of equipment that was installed in a room, connected to the mains, and left there to allow users to come along and use the software installed on it. When such computers were connected by wires to other computers to form networks, distributed computing became possible where the software and data being used in one location may have been sent from another computer some distance away. Alternatively, the data generated in one location could be sent to another computer for faster processing. The wired communication links were similar to the links that already existed for fixed telephones, which were also quite bulky and remained fixed on peoples' desks or kitchen tables. In many cases, wired links designed for telephony were used for data.

Mobile computing in the early days of computer networks meant that a user could access a given computer from different locations over wired links, which may have been telephone channels. Accessing a central computing facility from home or a hotel via a telephone modem gave a useful degree of mobility. Portable or 'luggable' computers equipped with telephone modems were prized and expensive commodities. In wired telephony, mobility meant that a user could divert his calls within an institution or access a fixed line from a limited distance by means of a cordless phone using radio.

Mobile telephony has these days come to mean 'cellular wireless' telephony, where wireless communication take place between a battery powered 'mobile phone' and a 'base-station' whose antenna generally sits on top of a tower or a tall building. With cellular mobile telephony, the ideal is to provide universal radio access to telephone & computer networks via base-stations each serving an area called a 'cell'. The link should be maintained even if we move from one cell to another. Each cell is allocated a frequency band which must be shared by many users within the cell. Neighbouring cells must have different frequency bands. Distant cells may re-use the same frequency bands.

The term 'cordless' is used for devices that communicate over short distances at home or within an institution.

The term 'cellular' refers to the way large numbers of base-stations are located to achieve universal coverage. Each base-station serves users who happen to be within an area referred to as its cell and arrangements are made for users to be 'handed over' to other base-stations if they move. Wireless telephony has been available for many decades but cellular mobile phones based on digital rather than analogue technology have been available only since about 1992. The growth of mobile telephony over the past decade has been phenomenal. Mobile phones are now quite powerful computers capable of communicating and processing data as well as speech. Therefore this course on 'Mobile Computing' must include material on 'mobile communications' and 'mobile telephony'.

Satellites are 'base stations in the sky'. A 'geo-stationary' satellite must be 22,300 miles high. Low orbit ones can be between about 100 to 1000 miles up, but we need lots of them, e.g. 77 or 66 for the 'Iridium' system. Planes go up to about 8 m and the Moon is 240,000 m away.

Question: If radio waves travel at 300,000 km/s (≈ 186,000 m/s) what would be the 'round trip delay' (go + return) between a transmission to and from a satellite, or the moon?

Answers: Geo-stationary: 0.24 s., Low orbit: 1/1000 to 1/100 seconds. Moon: 2.6 s

'Mobile computing' also refers to the use of lap-top computers, tablets and PDAs capable of being battery powered and communicating by radio to 'access points' which connect them to a (normally) wired computer network and ultimately the Internet. In some ways access-points are similar to base-stations as used in mobile telephony. An access-point communicating by radio to one or more mobile computing devices forms an 'infrastructure type' wireless 'local area network' (LAN) or WLAN. It is possible for a number of wireless enabled computers to come together and intercommunicate without an access-point and in this case they are said to form an 'ad-hoc' rather than an 'infrastructure type' WLAN.

Although access-points may be considered the WLAN equivalents of base-stations, and lap-top computers, or PDAs, the equivalents of mobile phones, there are some fundamental differences between WLANs and cellular mobile telephone networks. Cellular mobile telephony is strongly connection oriented and based on traditional 'time-division' and 'code division' multiplexed access techniques (TDMA & CDMA) in highly expensive licensed radio frequency bands. In Europe, these bands, referred to as the '900' and '1800' MHz bands extend from 890 to 960 MHz and from 1710 to 1880 MHz respectively. In contrast, WLANs are packet-orientated with packet-based multiplexing in freely available unlicensed radio frequency bands referred to as the '2.4 GHz' and the '5 GHz' Industrial Scientific & Medical(ISM) bands. These unlicensed radio bands, in the frequency range 2.4 to 2.48 GHz and 5.15 to 5.8 GHz are currently unused except by microwave ovens, garage door openers and a small number of similar devices. The range of cellular mobile phones, up to many miles, is considerably greater than the 100 meters or so offered by WLANs.

With WLANs, there is currently no need for service providers or users to pay for use of radio frequency bands. A company can charge for the use of an access-point, in a 'hot-spot', for example, but this can be very inexpensive since the equipment is relatively simple and there is no need for the 'hot-spot provider to pay for and then charge the user for the right to use a certain radio frequency band. . The user can thus have Internet access via 'hot-spot' connections to high capacity wired networks, and can buy his own access-point to allow wireless access to his broadband connection at home. An institution such as a university can buy and install WLAN access points and incur no further costs apart from maintenance.

Wireless LANs can replace or augment wired LANs for more convenient data communication and access to the Internet. With the increasing use of computer networks, including the Internet, for speech (VoIP) and real time multimedia, it is logical to consider the use of WLAN technology for providing convenient access to telephone services as well as data networks. Once WLANs are used for telephony, they are in direct competition with cellular mobile telephony. The outcome of this competition is still unresolved, though it is likely that some form of convergence of telephony and data services will emerge to the benefit of all users. Already it is quite feasible for users at home and in small institutions to use their Skype or MSN VoIP communications over a WLAN link. Large institutions very often have high quality VoIP telephony over voice-dedicated networks or reserved network capacity and there is much interest in providing wireless access to this VoIP telephony over wireless LANs.

The advantages of WLANs and the use of license-free radio frequency bands have been stated, but there are some possible disadvantages especially if convergence with telephony is envisaged. Firstly, the use of unlicensed frequency bands means that, in principle, anyone can transmit in these bands and thus inadvertently or perhaps maliciously interfere with their efficient use. It may be that WLANs will become ‘victims of their own success’ if lots of users make use of them and thus cause failure due to congestion. Such failure is especially likely with voice and multimedia transmissions and it has been shown that only about eight VoIP users can be accommodated by a single IEEE802.11b WLAN access point without incurring failure due to congestion. This restriction for VoIP users is due to the ‘contention based’ access mode of WLANs which was designed for data transmissions before the idea of using WLANs for telephony was ever thought of. Now that the problem has been realised several solutions have been proposed and will be presented later in this course.

2. Generations of mobile telecommunications standards

The word of telecommunications and computer networks is governed by standards published by organisations such as the International Telecommuncations Union (ITU) which has, for many years, defined standards for fixed and mobile telephony. This body was largely responsible for defining the four (or five) generations of mobile telephony standards as summarised in the table below:

- 0G Radio telephones
- 1G (1983) Cellular analogue for voice – e.g. AMPS
- 2G (1991) Cellular digital for voice & slow data – e.g. GSM, IS95
- 2.5G(≈1998) Introduce GPRS (56-114 kb/s)
- 2.75G(≈2003) Add EDGE(E-GPRS) (up to 384 kb/s)
- 3G (≈2001) IMT2000 for speech & faster data - UMTS etc
- 3.5G(≈2007) HSPDA (1.8-7.2 Mb/s downlink); UL: 384 kb/s
- 3.75G (≈2010) HSPA+ (DL: 56, UL: 22 Mb/s) etc.
- 3.95G (?) 3GPP-LTE, mobile WIMAX, etc.
- 4G (?) ITU-‘IMT Advanced’

The meanings of some of the acronyms used in the table are listed below:

- AMPS – Advanced (analogue) mobile phone system
- GSM – (European) Global system for mobile comms
- IS95 – USA equivalent of GSM
- GPRS –General packet radio system for 2G (56-114 kb/s)
- EDGE – Enhanced GPRS (≈ 384 kb/s)
- IMT2000 – International mobile telecomms (3G standard)
- UMTS – Universal mobile telecoms system
- HSDPA – High speed downlink packet access
- HSPA+ – High speed packet access
- LTE – Long term evolution (from 3G to 4G)
- WiMAX- Worldwide Interop for Microwave Access

- ITU – International telecomms Union
- 3GPP – 3G Partnership Project (ex GSM)
- 3GPP2 – 3G Partnership Proj 2 (ex IS-95 & CDMA2000 in USA)

'3G-IMT 2000' is a family of standards for 'third generation' mobile telecomms which includes:
+ GSM & EDGE legacy from 2G
+ UMTS – Cellular access to voice & high speed data
+ DECT - for cordless phones
+ WiMAX – for point-to-point & wide area networks.
UMTS is gradually being evolved towards the ideals of the next generation (4G) as defined below.

'4G-IMT Advanced' is the ITU definition of what is required for the 4th generation of cellular wireless standards. It is not a standard, but a set of goals which are, in summary:

(1) To fuse cellular mobile & nomadic access into a seamless layered architecture that is transparent to user
(2) By approximately 2010, to achieve 100 Mb/s for mobile access & 1000 Mb/s (1GB/s) for nomadic access.
(3) To pursue world-wide common spectrum & open global standardisation.

Only 2 technologies had been proposed, by Sept 2009, for realising these goals:
'3GPP-LTE-Advanced' (a standard due to be published in 2010) and
IEEE 802.16m (enhanced mobile WiMAX).

3GPP-LTE (Long Term Evolution towards 4G) is a project to evolve from 3G-UMTS towards 4G. The new standard: '3GPP-LTE-advanced' is expected by the end of 2010. With parallel developments in nomadic technology (wireless LANs etc.) this is intended to achieve the 4G goals. The new developments include:
+ IP for all voice (VoIP) & data (with seamless handover)
+ Enhanced precoding & forward error correction (FEC)
+ New radio transmission techniques (OFDMA & SC-FDMA)
+ Multiple antennas (MIMO)
+ Flexible spectrum usage,

3. Some background on wired telephone &computer networks

To understand the mobile computing issues in this course, it is useful to have some background information on traditional wired telephone and computer networks. This section presents some interesting and important issues.

3.1. Wired telephone networks

An interesting difference has traditionally existed in the ways that computers and telephones use their wired network links. Telephone networks were always considered to be 'circuit switched' in that the caller and the called user were given links that appeared to be a direct pair of wires as may have been used at the invention of telephony more than 100 years ago. In those very old days, a telephone operator would have interconnected users by physically connecting the wires from one telephone to those from another. Local telephone exchanges serving a few thousand users within a one or two mile radius in a city continued to do this, albeit mechanically, until fairly recently, say until the 1970s. However, exchange-to-exchange links and international links have long been multiplexed through high capacity channels each carrying hundreds and thousands of calls. Although there was not necessarily a direct wired connection, the concept of 'circuit switched' links remained in telephony, the philosophy being that making a call would reserve the necessary links to the local exchanges and a share of the resources of high capacity multiplexed channels for the duration of the call.

Originally, all speech was conveyed as an analogue electrical signal, i.e. as a continuous variation of voltage, and the sharing of high capacity channels was achieved by 'frequency division multiplexing' (fdm). The 'circuit switching' concept remained when exchange-to-exchange traffic was digitised and influenced the way the hundreds and thousands of digitised speech channels were multiplexed along single high capacity channels. Digitised speech channels were and still are multiplexed by 'time-division-multiplexing' (tdm) in such a way that they experience only a very small delay.

If there are 32 speech channels each with 8000 digitised samples per second, 8 bits per sample, there are many possible ways of time-division multiplexing these channels into a single bit-stream. For example we could send 800 samples of the first channel followed by 800 samples of the second, and so until the 32nd channel. Then we start again with the first channel. This would introduce the need for digital storage to collect each 800 sample block before it is transmitted, and more seriously would introduce a delay of 0.1 second into each channel. This would be in addition to the unavoidable 'propagation delay' delay of about 0.001 seconds per 100 miles (or about 200 km) due to the time it takes electrical signals to travel along wires. The propagation delay is about 30 ms for a 3000 mile trans-Atlantic wired link between Manchester and New York. If we transmit one sample (8 bits) from the first channel followed by one from the second channel, one from the third and so on, the need for storage is eliminated and the 'framing' delay becomes 1/8000 second i.e. 0.125×10^{-3} seconds. The intention is to convey each speech channel with imperceptible delay and this 'minimal delay' multiplexing approach has been adopted in wired telephony links for many years.

Experiments have established that delays in excess of about 0.3 seconds make two-way telephone conversations difficult as may be demonstrated by the use of 'high-orbit' satellite links. A single link to a satellite and back introduces about 0.3 seconds propagation delay and if such a link is also used for the return path the overall 'round trip' delay of 0.6 seconds is well beyond the limit of acceptability.

3.2 Wired computer networks

The connection requirements of wired computer networks were considerably different from those of telephone networks to begin with. They were said to be 'packet switched' rather than 'circuit switched' since data consisting of text or sets of numbers were placed in frames, given an address, and launched into a network of linked computers with little consideration of how they would find their way to their destination. It was said that computer networks were analogous to a postal system, which conveyed letters to the destination written on the envelope often by circuitous routes. Packets may not arrive in the order they were sent in, and may not arrive at all. Packetisation delay is introduced by the need to collect data into reasonable sized frames before they are sent and in addition to the unavoidable propagation delay mentioned above, there is further delay with packetised transmission due to the need for specialised computers called 'routers' along the way to collect and read the addresses and convey each packet in the right direction towards its destination.

Packetised communication between computers was originally designed to convey data and not telephone conversations. Delays of a few extra seconds in data transmissions, in email for example, are not generally a problem though the arrival of packets out of order and the complete loss of packets is a problem that must be addressed. The great advantage of packetised transmission is its simplicity, robustness and flexibility. If the direct link between two cities malfunctions or becomes overloaded, routers can convey the addressed packets by alternative links, and are designed to do this automatically without human intervention and without necessarily losing any data at all. By contrast, the re-routing of telephone calls is not so straightforward.

3.3 Convergence of voice & data networks

The existence of two fundamentally different types of wired network, albeit sharing some communication links, may appear at first sight to be not such a bad thing. In principle the 'quality of service' (QoS) requirements of each type of network are different if we restrict the telephone service to speech and computer networks to data. However such restriction would have been impossible and undesirable for many reasons. Telephone service providers could see many advantages in offering data services as well as speech, and network providers were able to improve computer networks to such an extent that they could convey speech (VoIP) at much lower cost than is possible with traditional telephone networks. The convergence of computer and telephone networks seems inevitable and the adoption of packetised speech transmission for telephone exchange to exchange signalling, with the abandonment of tdm, will be a huge step, and perhaps the final step in ensuring that only one type of wired network will exist in the future. Telephone networks had the advantage of many decades start and the disadvantage of having maintain a huge existing infrastructure, originally based on analogue technology, while adopting the later digital technology.

4. Addressing

This section presents some background information on the concept of an 'IP address' which will be necessary for the coursework assignments.

Fixed telephone numbers and computer 'IP addresses' were, and still are, localised in that a telephone number beginning '0161'is always routed to Manchester and an IP address beginning 130.88 is always routed to a computer in Manchester university. IP addresses are 32 bit numbers expressed as four bytes with each byte converted to decimal. They are in many ways the equivalent of telephone numbers. If you work in an institution such as Manchester University, this institution will normally give your computer an IP address. When you connect to your broadband system at home, your computer must obtain an IP address from your router which itself will have an IP address as specified by your Internet Service Provider (ISP). If your portable (lap-top) computer has a fixed Manchester University IP address and you visit another institution, say Liverpool University, and connect it to a wired LAN there, it will not be recognised until the IP number is changed.

When using your computer to access the Internet, you will want it to be accessible to the whole world. However, when performing experiments in a laboratory this may be dangerous as you may make your computer, and others in your institution, vulnerable to viruses and hackers especially when your experiments involve wireless access with minimal or no security. Therefore we will be adopting the special IP addresses 192.168.0.XXX (with masking 255.255.0.0) in our experiments and running them with no wired connection to the university's wired networks.

In practice an IP address is assigned to the 'network interface' in a computer rather than to the computer itself. In older computers the network interface was an 'Ethernet' network card but these days it is more likely to be a chip-set on the motherboard. Each network interface has a unique hard-wired 48 bit 'physical address', also called its 'MAC address'. So once an institution has assigned an IP address to a computer's network interface, its networks can route packets with a matching IP address to the computer whose network interface has the appropriate physical or MAC address.

The localised way that IP addresses are defined simplifies the routing, but introduces an immediate problem with respect to mobility. This problem will be addressed in a later lecture.

Instead of giving a fixed IP address to each computer, it is often more convenient for an institution to allocate IP addresses dynamically using a 'dynamic host configuration protocol, 'DHCP server'. The DHCP server can identify a valid computer by the MAC (or Phy) address of its network interface as supplied by the owner of the computer. The DHCP server will have a table of valid MAC (Phy) addresses will and allocate IP addresses to these MAC addresses when the computers are connected and switched on. In principle, a computer can be given a different IP address every time it is switched on, but in practice DHCP servers will try to give the same IP address to a given computer each time it is switched on.

On switch-on, the DHCP server will also supply the IP address of a 'Domain Name Server (DNS)' which provides access to a universal world wide distributed data-base containing IP addresses of all the computers, servers and other devices you may ever want to contact together with their names. The IP address of your computer with its host-name XXX.cs.man.ac.uk will be there. The IP address of the host-name www.google.com, 216.239.39.99 will also be there, so that typing http://www.google.com into a web-browser is equivalent to typing http://216.239.39.99 .

5. Conclusions & learning outcomes from this lecture:

Meanings of 'mobility' have been discussed and some applications and issues were mentioned.
Similarities and differences between wireless LANs and mobile telephony have been explored.
Latency (delay) is an important consideration in telephony and is very low in traditional wired telephone networks based on the concept of 'circuit switching'. It is generally higher in computer networks.
The advantages and disadvantages of having wireless LANs in licence-free spectra have been outlined.
The goals of 4th generation mobile communication technology have been defined by ITU and others.
These goals integrate the provisions of cellular with 'nomadic' access to wi-fi & maybe WiMax.
Two 4G standards, based on 3GPP-LTE & WiMAX are imminent.
The trend towards convergence of voice & data networks is continuing.
The localised nature of IP addresses causes problems with mobility as will be discussed later in the course.

6. Problems & discussion points

1.What is the frequency range of the speech conveyed by 'plain old fashioned telephones' (POTs) as still used on wired networks and why is the lowest 300Hz bandwidth normally lost
2. How does the quality of a GSM call differ from that of POTS ?
3. In what year & in which country did it become possible to make phone call on a train?
4. When were car-phones first introduced?
5. When was the 'cellular' concept first introduced?
6. What is meant by 1st, 2nd, 2.5th, 3rd & 4th generation mobile telecommunications technology?
7. Why is the digital mobile phone industry considered more successful in Europe that in the USA.

8. With respect to 2G technology, what contributed to the greater success of the European GSM standards in comparison to IS95 in the USA?
9. Explain the terms: WLAN, WPAN, WMAN.
10. What is GPRS and EDGE and how does these compare with WLAN technology.
11. If a speech transmission system uses 130 ms speech frames, how far could we hope to communicate with acceptable delay for 2-way telephony.
12. For interactive music making, (e.g. Internet Choir) we believe that a ‘round trip’ delay of about 60 ms or less must be achieved. If we restrict the application to Western Europe, could this be achieved and how?
13. What advantages came from the change to digital (2G from 1G)
14. Why was it easier to update IS95 (& later versions) to 3G UMTS tech?
15. If IEEE802.11 is ‘wi-fi’ what is IEEE802.3 called?
16. What are the ‘ISM’ bands & what are they used for?
17. How would you define ‘nomadic’?
18. How do cells ‘breath’ (see refs)
19. What is cellular ‘hand-over’ & why should it be seamless?
20. If a speech transmission system uses 0.12 s speech frames, how far could we hope to communicate with acceptable ‘round trip’ delay of 0.3 s.
21. Why are geostationary satellites good for broadcasting but not so convenient for mobile computing & telephony.