CAMBRIDGE IGCSE COMPUTER SCIENCE 0478

UNIT 3

Data communications and networking

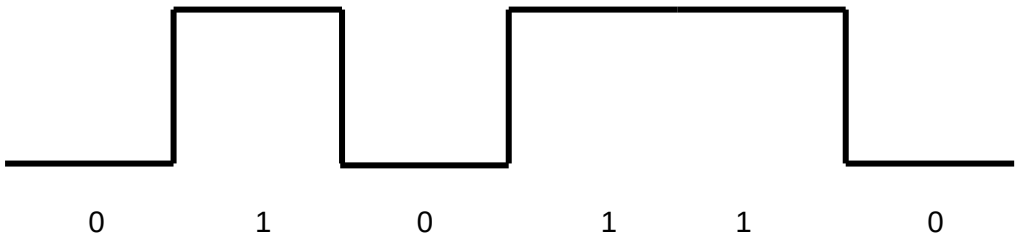
Contents

Data transmission between a processor and its peripherals, and between computers in a network, is a central element in everyday life; the internet has become an unconscious way of life for many people today. This unit looks at the principles underpinning data transmission in these contexts.

Part 1 Serial and parallel data transmission

Show an understanding of what is meant by transmission of data

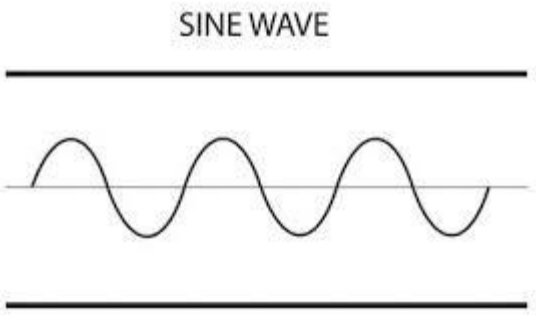
When data are transmitted from one computer to another they travel along some medium (copper wire, optical fibre, as radio waves, infra-red) in the form of bits. For example, when the data are transmitted using wires they are transmitted in the form of voltage changes. The sender will generate one voltage to represent 1 and a different voltage to represent 0. The sender has a clock and it transmits bits at regular intervals. At any point in time there will be a number of bits being carried along the wire so the voltage may be depicted as follows:



Baud Rate
The rate that voltage changes is called the baud. In the simple case described above, if the voltage changes 10 times every second the baud is said to be 10.

Bit Rate
The bit rate is the term given to the rate that bits are transmitted. In the simple case described above the bit rate is the same as the baud. If we could generate four voltages, instead of two, we could use each change in signal to represent two bits.

Bandwidth
When a signal is transmitted along a wire (or any other medium) it is reluctant to travel as a square wave. Waves are naturally in the form of sine waves



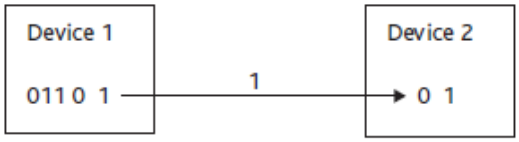
All media are capable of transmitting a set of waves that have a range of frequencies. The frequency of the wave is the rate at which the wave repeats itself. The range of frequencies that a medium can transmit is known as its bandwidth. The wider the bandwidth, the more

data that can be transmitted. An optical fibre has a very high bandwidth so it can transmit a very large amount of data. A normal telephone wire has a very low bandwidth so it is not possible to transmit many data.

Distinguish between serial and parallel data transmission

Serial transmission of Data

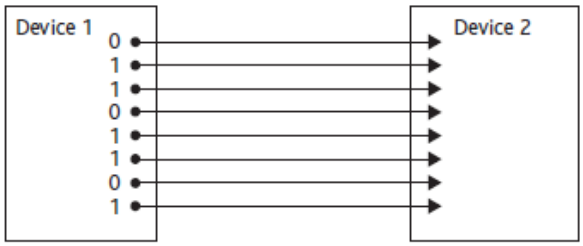
Data needs to be transmitted between devices in a computer system. The easy way is by using a single wire that links the two devices. Data are transmitted down the wire in the form of bits, so a byte that represents a single character is transmitted as eight bits in sequence, one signal for each bit.



The diagram shows the data byte 01101101 in the process of being transmitted from Device 1 to Device 2. As there is only one wire, only one bit can be transmitted at any time.
This is known as serial transmission of data.

Parallel transmission of Data

If the devices are connected by more than one wire, then more bits can be sent simultaneously. A sensible number of wires would be eight, because then a whole byte can be sent at the same time. This type of data transfer is called parallel data transmission.



Show understanding of the reasons for choosing serial or parallel data transmission

Serial transmission of Data
Serial transmission has the advantage of being simple and reliable because the next bit is not transmitted until the current one has been received. However, because only one bit can be transmitted at a time, the transmission is slow. All peripheral devices that connect through a Universal Serial Bus (USB) use serial data transmission.

Parallel transmission of Data
Parallel transmission of data is obviously faster than serial transmission because all the bits are travelling at the same time. However, because of the fine tolerances in the transmission, it is less reliable as the bits can become muddled up. If one bit is delayed because of the resistance on its wire, for example, it may arrive in time to be counted as a bit in the next byte! This problem, where the bits become out of sequence, is called “skew”. Parallel transmission is only suitable for short distances.

Show understanding of the need to check for errors

When data, of whatever type, are transmitted from one device to another, they are transmitted as a series of binary digits. Any data that are transmitted are going to be made up of a very large number of bits.
Consequently, there are bound to be occasions on which the data are not transmitted correctly or on which they become corrupted during transmission. There are only two possible types of error that can occur; either a 1 is received as a 0 or a 0 is received as a 1. Mistakes rarely occur, but when they do occur they can be very serious, as the data are no longer correct. This makes it important that there should be methods for checking the data when they are transmitted.

Explain how parity bits are used for error detection

A parity check involves checking that the number of 1 bits in a byte totals to an even number (called “even parity”) or an odd number (called “odd parity”). If two devices that are communicating decide to use odd parity, there must always be an odd number of 1s. If a byte is received with an even number of 1s, an error must have occurred. For example, the byte 01011000 is sent. It has three 1 bits so it passes the odd parity check. When it is transmitted, the byte received is 11011000. This has four 1 bits, which is an even number, so there must have been an error in transmission. The receiving device would ask for it to be sent again. Although this example uses odd parity, even parity can equally well be used. The two devices have to agree which type of parity to use.

Identify current uses of serial and parallel data transmission such as Integrated Circuits (IC), Universal Serial Bus (USB)

Serial transmission of Data

Practically all long-distance communication transmits data one bit at a time, rather than in parallel, because it reduces the cost of the cable.

Keyboard and mouse cables and ports are almost invariably serial -- such as PS/2 port and Apple Desktop Bus and USB. The cables that carry digital video are almost invariably serial -- such as coax cable plugged into a HD-SDI port, a webcam plugged into a USB port or Firewire port, Ethernet cable connecting an IP camera to a Power over Ethernet port, FPD-Link, etc.

An integrated circuit or monolithic integrated circuit (also referred to as an IC, a chip, or a microchip) is a set of electronic circuits on one small plate ("chip") of semiconductor material, normally silicon. ICs can be made very compact, having up to several billion transistors and other electronic components in an area the size of a fingernail.

Integrated circuits are used in virtually all electronic equipment today and have revolutionized the world of electronics. Computers, mobile phones, and other digital home

appliances are now inextricable parts of the structure of modern societies, made possible by the low cost of producing integrated circuits. Solid State storage use integrated circuits, for example, flash drives.

Many communication systems were generally originally designed to connect two integrated circuits on the same printed circuit board, connected by signal traces on that board (rather than external cables).

Integrated circuits are more expensive when they have more pins. To reduce the number of pins in a package, many ICs use a serial bus to transfer data when speed is not important. Some examples of such low-cost serial buses include SPI, I²C, UNI/O, and 1-Wire.

Parallel transmission of Data

If you have a printer connected to your computer, there is a good chance that it uses the parallel port**.** While USB is becoming increasingly popular, the parallel port is still a commonly used interface for printers.

Parallel ports can be used to connect a host of popular computer peripherals:

- Printers
- Scanners
- CD burners
- External hard drives
- Iomega Zip removable drives
- Network adapters
- Tape backup drives

Part 2 Internet principles of operation

Show understanding of the role of the browser and Internet server

LANs and WANs
In schools, colleges and offices around the world, computers are connected together in some way to form a network. These vary from local area networks (LAN) as found in school, for example, to massive wide area networks (WAN) such as the Internet.

LANs essentially allow the sharing of resources both hardware and software, easier communication between users and the ability to control and monitor computer use. The LANs usually have some form of file server (where common files and software are stored) and print server. LANs are frequently connected together to become part of a wide area network and made use of, for example, routers and broadband modem.

The Internet
The internet is a world-wide system of computer networks. It is possible to access any computer connected to this network provided you are given the necessary permissions.

A protocol is a set of rules which is used by computers to communicate with each other across a network.

In reality, the internet took off in the 1990's with the introduction of HTML (Hypertext mark up language) and WWW (World Wide Web) which uses http (hypertext transfer protocols)

In order for a computer to operate on a network, there are a range of different components that are required:

A Router is a device that transfers data from one network to another in an intelligent way. It has the task of forwarding data packets to their destination by the most efficient route. In order to do this, the router has a micro computer inside it. This holds a table in memory that contains a list of all the networks it is connected to, along with the latest information on how busy each path in the network is, at that moment. This is called the 'routing table'.

When a data packet arrives, the router does the following:-

- Reads the data packet's destination address
- Looks up all the paths it has available to get to that address
- Checks on how busy each path is at the moment
- Sends the packet along the least congested (fastest) path

Other tasks the Router can perform:

- Exchange protocol information across networks
- Filter traffic - helps prevent unathorised intrusion by malware

Routers are also needed to enable a computer to connect to the internet, after all, the internet is just one vast external network.

If the data packet has a destination address outside the local networks, then the router may send it to the internet modem and then on to the ISP's router at the other end of the line. Their router will then pass forward the data packet towards its destination. For a computer say in the UK to connect to a web server in the USA the data packet will pass through many routers around the world.

Network cards are needed if the computer does not have a built-in network chips on the motherboard. They allow the signal from the network to be transmitted to the machine – this could be via a fixed cable, infra red or radio waves.

A modem converts the digital data from the computer into a continuous analogue wave-form that the telephone system is designed to deal with (MODulation). The reason for this is that the telephone system was originally designed for the human voice i.e. continuous signals. The modem also converts the analogue signal from the telephone network back into digital data that the computer can understand. (DEModulation). Hence the word MODEM.

Standard ADSL modems come in two forms: An external box that links to your computer through an USB port or network cable, or an internal modem that is plugged directly to the motherboard inside the computer.

In addition to telephone modems, radio has now become very popular as a means of connecting to the internet. The device that allows you to do this is called the Wi-Fi modem

Web browser

Web browsers are software that allow a user to display and interact with web pages and files from the internet.

The software interprets the coding language of the websites and displays the translation instead of showing the actual coding. Consequently, a user can simply launch a web browser by clicking on the appropriate icon from the desktop and there is no need to know

the commands which are required to interpret the website coding once it has been accessed.

Server

A server is software that responds to the requests of other programs, known as clients. In a web server the software handles web pages so that remote internet users can communicate with the website. The server accepts and processes these requests and supplies the resources required from the website.

Web servers are also embedded in many devices such as webcams, routers and printers.

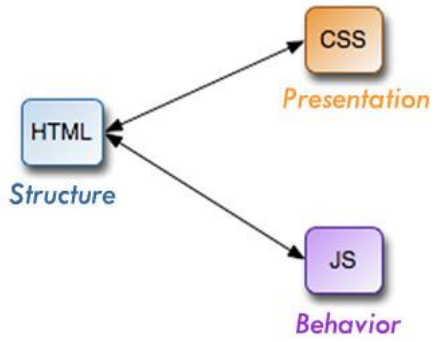
Distinguish between HTML structure and presentation

The makeup of a webpage could be viewed as a combination of the following four elements:

- **Content** is the collective term for all the browser-displayable information elements such as text, audio, still images, animation, video, multimedia, and files (e.g., Word, PowerPoint, PDF, etc.) of web pages. Content does not require any additional presentational markups or styles in order to fully convey its message.
- **Structure** refers to the practice of using HTML on content to convey meaning and to describe how blocks of information are structured to one another. Examples: "this is a list" (ol, ul, li), "this is headings and subheadings" (<h1>, <h2>, ..., <h6>), "this section is related to" (<a>), etc..
- **Presentation** (or **Style**) refers to anything related to how the content and structure is presented. Examples: size, colour, margins, borders, layout, location, etc.
- **Behaviour** (or **Interactivity**) refers to the employment of client-side script (e.g., JavaScript) to create interactivity between the webpage and its users.

Often, a clear distinction between content and structure is difficult, because content could also be viewed as the information with its semantic coding as well as its structure. In practice, the makeup of a webpage can simply be viewed as a combination of three elements: Structure, Presentation, and Behaviour.

The term essentially refers to the "separation of the content made meaningful by structure and the presentation, ” or simply the "separation of the structure (HTML) and presentation (CSS)".



Example	Structure or Presentation?	HTML or CSS?
A heading	Structure	XHTML
Size of a heading	Presentation	CSS
A paragraph	Structure	XHTML
Color of the text in a paragraph	Presentation	CSS
A table of figures	Structure	XHTML
A border around table cells	Presentation	CSS
An image, such as a portrait photo	Structure	XHTML
An image, such as a tessellating background	Presentation	CSS
A group of navigation links	Structure	XHTML
The placement of a group of navigation links on a page	Presentation	CSS

(source: P. Griffiths, "HTML Dog: The Best-Practice Guide to XHTML&CSS, " New Raider, 2007)

The Practice

- Ideally, the separation of structure and presentation will produce an HTML document which contains the content and structure and a separate CSS file which contains everything that controls presentation.
- The perfect website separation system will store content in a database for a complete isolation and management of content information. Structure, on the other hand, is dealt with through a collection of template package built by a server-side scripting language like PHP or ASP.net.

Show understanding of the concept of Media Access Control (MAC) address, Internet Protocol (IP) address and cookies

TCP/IP
Protocols are the rules used for devices to communicate. TCP/IP stands for Transmission Control Protocol/Internet Protocol. TCP/IP is a set of rules that looks after data transmission on the internet. TCP/IP allows networks of different types to communicate with each other.

Internet Protocol (IP) Address
Each computer on a TCP/IP network has an IP address. This is a 32-bit numerical identifier, for example, 10000011 01101011 00010000 11001000. Each group of eight bits is called an octet. Each octet can store numbers ranging from 0 to 255. Binary digits are hard for a human to work with, so the address is normally quoted in four groups for decimal numbers, for example, 131.107.32.200. IP addresses can be permanently allocated to a device (static addressing), but they can alternatively be allocated as needed (dynamic addressing), so a particular computer will not always have the same IP address each time it connects. This saves the network administrator effort and makes efficient use of the addresses available. These can locate a resource on a network by its IP address.
Because of the huge growth of the internet, in order to make sure that there are enough addresses, there has been a move towards 128-bit addresses.
When you access a web page on the internet, you have to tell your browser the address to go to. Long numbers are hard to remember, so we normally use domain names liked chesterhouse.co.za instead. There are special servers called DNS servers that convert domain names such as chesterhouse.co.za into numerical IP addresses.

MAC Address
MAC stands for Media Access Control. Each physical interface connected to a network, such as a network card, has a unique number written to the device by the manufacturer. The MAC address is used to identify a device on a network. Some switches, known as level 2 switches use MAC addresses to identify devices on a network.
A MAC address is usually given as six pairs of hexadecimal numbers, for example, 01:1F:33:69:BC:14

Cookies
A cookie, also known as an HTTP cookie, web cookie, or browser cookie, is a small piece of data sent from a website and stored in a user's web browser while the user is browsing that website. Every time the user loads the website, the browser sends the cookie back to the server to notify the website of the user's previous activity.[1] Cookies were designed to be a reliable mechanism for websites to remember stateful information (such as items in a shopping cart) or to record the user's browsing activity (including clicking particular buttons, logging in, or recording which pages were visited by the user as far back as months or years ago).

Show understanding of what is meant by hypertext transfer protocol (http) and Hypertext Markup Language (HTML)

HTML (Hypertext Mark up Language)
This is a type of computer language; all the text, graphics and design elements of a webpage are tagged with codes that instruct a web browser to display the files on a computer screen. It is easy to see this when looking at the extension name of the file i.e. *.html* or *.htm* which will be shown at the end; for example:
http://www.programs.com/definitions.htm

http (hypertext transfer protocol)
This is a set of rules for transferring files on the web; https was later developed to allow secured transactions on the web to take place. In theory, https works by transmitting the usual information by using some form of encryption; this should mean that the information cannot be accessed by anybody other than the client (user) and end server.

Part 3 Data storage

Show understanding that sound (music), pictures, video, text and numbers are stored in different formats

Each type of application software has its own set of standard file types. The following list illustrates examples of each format but is not an exhaustive list.

Sound
MP3
WAV
AIFF (Audio Interchange File Format)

Pictures
GIF (Graphics Interchange Format)
JPEG
TIFF (Tab Image File Format)
BMP
EPS (Encapsulated Postscript)
PICT (PICTure)
PCD (Kodak Photo CD)

Video
MPEG
MOVIE (also .QT)
AVI

Text
ASCII, TXT (text only), text with break lines
RTF (Rich text format)
PS (Postscript file)
DOC (Microsoft word file format)

Numbers
TSV (Tab separated variable)
CSV (Comma Separated variable)

SYLK (Symbolic Link)
DIF (Data interchange format)
DBF (dbase format)

Identify and describe methods of error detection and correction such as parity checks, check digits, checksums, Automatic Repeat reQuests (ARQ)

Parity Checks
This is used to check data following potential transmission errors; an extra binary digit is added to each binary number before transmission. Systems that use EVEN parity have an even number of 1s; systems that use ODD parity have an odd number of 1's.

For example, if a system uses EVEN parity and the number being transmitted is: 1101110 then an extra 1 is added to give the number even parity i.e. 1101110**1**; but if the number being transmitted was: 1101100 then an extra 0 is added since the number already has an even number of 1s i.e. 1101100**0**. The parity is checked at the receiving end to make sure none of the binary bits have been transmitted incorrectly.

Check digits
This is an extra digit added to a number which is calculated from the other digits; the computer recalculates the check digit after the number has been input. Check digits are used on bar code numbers and ISBN's.

There are a number of ways that check digits are generated; in the example that follows, we will consider the **ISBN-10 method** which makes use of the *modulo 11 system.*

Example
We will consider the number 0-221-43256-?

(i) The position of each digit is first considered:

10	9	8	7	6	5	4	3	2	1 ← Digit Position
0	2	2	1	4	3	2	5	6	? ← Number

(ii) Each digit in the number is then multiplied by it's digit position and the totals are added together:
i.e. (0x10) + (2x9) + (2x8) + (1x7) + (4x6) + (3x5) + (2x4) + (5x3) + (6x2)
= 0+18+16+7+24+15+8+15+12
= 115 total

(iii) The total is then divided by 11 (modulo 11) and the remainder, if any, is subtracted from 11. The answer then gives the check digit.
i.e. 115/11 = 10 remainder 5
i.e. 11 – 5 = 6 (check digit)
hence, the final number is: 0-221-43256-6

Checksums

A checksum is a way of summarising a block of data such as a USB or a network data packet. At its simplest, it consists of the arithmetical sum of all the numerical values of all the elements of the block. The sum is reduced to a standard number of digits and transmitted with the block. When the block of data gets to its destination, the same mathematical calculation is performed on the data by the receiving device and the results is compared with the received checksum. If the two checksums match, the integrity of the data has been maintained. If the two checksums do not match then an error has been made in transmitting the data and the receiving device requests the sending device to re-transmit the

data. Even if one binary digit has changed in the data, the recalculated checksum does not match the received checksum and the data are rejected.

Checksum is similar in function to a parity bit for a byte or a check digit for a code number. More complex implementations of checksum involve more complex arithmetic to try to detect a wider range of errors.

Automatic Repeat reQuest (ARQ)

Automatic Repeat reQuest (ARQ), also known as Automatic Repeat Query, is an error-control method for data transmission that uses acknowledgements (messages sent by the receiver indicating that it has correctly received a data frame or packet)
and timeouts (specified periods of time allowed to elapse before an acknowledgment is to be received) to achieve reliable data transmission over an unreliable service. If the sender does not receive an acknowledgment before the timeout, it usually re-transmits the frame/packet until the sender receives an acknowledgment or exceeds a predefined number of re-transmissions.

The types of ARQ protocols include

- Stop-and-wait ARQ
- Go-Back-N ARQ
- Selective Repeat ARQ

Show understanding of the concept of Musical Instrument Digital Interface (MIDI) files, jpeg files, MP3 and MP4 files

Musical Instrument Digital Interface (MIDI)

MIDI is a technical standard that describes a protocol, digital interface and connectors and allows a wide variety of electronic musical instruments, computers and other related devices to connect and communicate with one another. A single MIDI link can carry up to sixteen channels of information, each of which can be routed to a separate device.

MIDI carries event messages that specify notation, pitch and velocity, control signals for parameters such as volume, vibrato, audio panning, cues, and clock signals that set and synchronize tempo between multiple devices. These messages are sent to other devices where they control sound generation and other features. This data can also be recorded into a hardware or software device called a sequencer, which can be used to edit the data and to play it back at a later time.

MIDI technology was standardized in 1983 by a panel of music industry representatives, and is maintained by the MIDI Manufacturers Association (MMA). All official MIDI standards are jointly developed and published by the MMA in Los Angeles, California, US, and for Japan, the MIDI Committee of the Association of Musical Electronics Industry (AMEI) in Tokyo.

Advantages of MIDI include compactness (an entire song can be coded in a few hundred lines, i.e. in a few kilobytes), ease of modification and manipulation and choice of instruments.

JPEG

JPEG (seen most often with the .jpg extension) is a commonly used method of lossy compression for digital images, particularly for those images produced by digital photography. The degree of compression can be adjusted, allowing a selectable trade-off

between storage size and image quality. JPEG typically achieves 10:1 compression with little perceptible loss in image quality.

JPEG compression is used in a number of image file formats. JPEG/Exif is the most common image format used by digital cameras and other photographic image capture devices; along with JPEG/JFIF, it is the most common format for storing and transmitting photographic images on the World Wide Web. These format variations are often not distinguished, and are simply called JPEG.

The term "JPEG" is an acronym for the Joint Photographic Experts Group, which created the standard.

JPEG/JFIF supports a maximum image size of 65535×65535 pixels.

MP3

MPEG-1 or MPEG-2 Audio Layer III, more commonly referred to as MP3, is an encoding format for digital audio which uses a form of lossy data compression. It is a common audio format for consumer audio streaming or storage, as well as a de facto standard of digital audio compression for the transfer and playback of music on most digital audio players.

MP3 is an audio-specific format that was designed by the Moving Picture Experts Group (MPEG) as part of its MPEG-1 standard and later extended in MPEG-2 standard.

The use in MP3 of a lossy compression algorithm is designed to greatly reduce the amount of data required to represent the audio recording and still sound like a faithful reproduction of the original uncompressed audio for most listeners. An MP3 file that is created using the setting of 128 kbit/s will result in a file that is about 1/11 the size of the CD file created from the original audio source. An MP3 file can also be constructed at higher or lower bit rates, with higher or lower resulting quality.

The compression works by reducing accuracy of certain parts of sound that are considered to be beyond the auditory resolution ability of most people. This method is commonly referred to as perceptual coding. It uses psychoacoustic models to discard or reduce precision of components less audible to human hearing, and then records the remaining information in an efficient manner.

MP4

MPEG-4 Part 14 or MP4 is a digital multimedia format most commonly used to store video and audio, but can also be used to store other data such as subtitles and still images. Like most modern container formats, it allows streaming over the Internet. The only official filename extension for MPEG-4 Part 14 files is .mp4. MP4 files work in a similar way to MP3, but they store audio, video, photo and animation files.

Show understanding of the principles of data compression (lossless and lossy compression algorithms) applied to music/video, photos and text files

When data other than text is being transmitted, e.g. on the Internet, it is important to limit the amount of data that needs to be sent to stop the time taken to download the data being unreasonably long. The amount of data can be limited by reducing the file size of pictures so that they take up only a small part of the screen or restricting them to a few colours. Speeding up the transmission of the data is achieved by reducing the amount of data that is sent. This is known as **file compression**.

Compression can be either lossy or lossless. Lossless compression means that no data is lost.

Lossy Compression
Lossy compression involved sacrificing some of the data in order to reduce the file size. Lossy compression techniques reduce the quantity of data in two ways. First by using complex mathematical encoding and secondly by deliberately losing some types of visual information that our eyes and brain usually ignore (this is called *quantization*).

For example, the video frame rate may be reduced from the normal 25 frames per second down to around 15 frames per second before there is a perceptible loss in quality. The frames themselves may be treated as separate still images, and compressed individually using JPEG compression. Different areas of a frame may be compressed by different degrees – an area of blue sky which lacks detail might be compressed by 25:1, whereas a person's face might only be compressed by 5:1. Depending on the amount of action in the video, only some areas of each frame may change from one frame to the next and only the changed data need be stored. The size of the picture may also be reduced, reducing the overall quantity of pixels to be stored.

If lossy compression is taken to an extreme, it can result in a significant loss of picture quality. The higher the compression ratio, the worse the resulting image. For instance, colour fidelity fades and the edges of objects become very obvious, until eventually the results is unwatchable.

JPEG image compression works in part by rounding off nonessential bits of information. There is a corresponding trade-off between preserving information and reducing size. A number of popular compression formats exploit these perceptual differences, including those used in music files, images, and video.
Lossy image compression can be used in digital cameras, to increase storage capacities with minimal degradation of picture quality. Similarly, DVDs use the lossy MPEG-2 Video codec for video compression.

Lossless Compression
Lossless compression uses mathematical techniques such as Huffman coding or Discrete Cosine Transformation (DCT), to reduce the quantity of information to be stored, while still being capable of reproducing the original image without any loss in quality.

Lossless data compression algorithms usually exploit statistical redundancy to represent data more concisely without losing information, so that the process is reversible. Lossless compression is possible because most real-world data has statistical redundancy. For example, an image may have areas of colour that do not change over several pixels; instead of coding "red pixel, red pixel, ..." the data may be encoded as "279 red pixels". This is a basic example of run-length encoding; there are many schemes to reduce file size by eliminating redundancy.

Compression standards include MPEG, M-JPEG, Cinepak, Intel's Indeo Video Interactive (IVI), Apple's Quicktime, Microsofts Directshow

Part 4 Security

Show understanding of the internet risks associated with viruses, spy-ware and hacking

Show understanding of security aspects of using the internet and understand what methods are available to help minimise the risks

Hacking

Hacking is the act of gaining illegal access to a computer system; many hackers do this to cause harm to a computer system or just to gain personal data (such as credit card numbers) for their own.

Impact of hacking on computer systems:

- Gaining access to personal data can lead to fraud, identity theft, or even disclosing personal information to harm the individual.
- Hacking could be used to gain sensitive or key data and even change or delete data once accessed.

Safeguards available to remove or minimise the risk:

- Use of firewalls
- Use robust passwords and user ids
- Use of encryption
- Anti-hacking software is also available to help prevent hacking.

Viruses

Viruses are programs that replicate themselves and are designed to disrupt a computer system; they often work by attaching themselves to computer files.

Impact of viruses on computer systems:

- Viruses can cause a computer to crash (i.e. stop functioning normally, lock up or stop it responding to software).
- Viruses can cause loss of files (if system files are lost this can lead to a system crash.)
- Viruses can cause corruption of files.

Safeguards available to remove or minimise the risk:

- Install and regularly use up to date anti-virus software.
- Do not use software from unknown sources (example, from the internet, CDs, etc)
- Take care when opening attachments from ‘unknown’ email addresses.

Spyware

Spyware is software that gathers information by monitoring key presses on the user’s keyboard and relays this information back to the person who sent the spyware. Spywaare can also install other spyware programs, read cookies and can even change the user’s web browser.

Impact of spyware on computer systems:

- Monitoring key presses gives the originator access to characters in typed passwords, credit card numbers, sensitive data, etc. and can lead to serious fraud and identity theft implications.

Safeguards available to remove or minimise the risk:

- Certain anti-spyware software can identify and remove this type of code/software.
- The user should always be alert and look out for 'clues' that they are being monitored in this way.

Malware

Malicious, that is harmfully intended, software of all sorts, including but not restricted to, viruses.

Explain how anti-virus and other protection software helps to protect the user from security risks

The most reliable way of detecting and removing malware once it has arrived at a computer is anti-virus software, also known as a virus scanner. This software scans files as they are opened, copied and saved. If malware is detected, it prevents the malware from running. The software by be configured to automatically or manually remove malware. This consists of remove the malicious code from a file infected with a virus or deleting the whole of any other sort of malware file.

Some anti-virus software can also monitor a computer for malware activity that could lead to identity cloning, block a website know to be the source of malware or check links provided by a search engine for threats. Although anti-virus software is not usually included in an operating system, every computer should be protected by it and free editions are available for personal use.

Show understanding of how data is kept safe when stored and transmitted including:

a) Use of passwords both entered at a keyboard and biometric

User IDs and Passwords

When you log onto your network at school, you have to type in your User ID and Password. This identifies you to the network as an authorised user.

Any sensible company will ensure that staff need a User ID and Password to gain access to the system. This should reduce the risk of outsiders being able to get onto the system and damage data.

People should follow rules when choosing their password:

- Passwords should be kept secret at all times
- Passwords should not be something that is easy to guess such as pet's name or favourite football team.
- Passwords should include text and numbers or symbols
- Passwords should be a reasonable length e.g. over 6 characters
- Passwords should be changed regularly

Biometric Passwords

Biometrics use physical characteristics, like your face, fingerprints, irises or veins, or behavioural characteristics like your voice, handwriting or typing rhythm to identify you. Unlike passwords, your personal traits are extremely difficult to lose or forget. They can also be very difficult to copy. For this reason, many people consider them to be safer and more secure than passwords.

Biometrics uses unique features, like the iris of your eye, to identify you.

Biometric systems can seem complicated, but they all use the same three steps:

- Enrollment: The first time you use a biometric system, it records basic information about you, like your name or an identification number. It then captures an image or recording of your specific trait.
- Storage: They analyze your trait and translate it into a code or graph. Some systems also record this data onto a smart card that you carry with you.
- Comparison: The next time you use the system, it compares the trait you present to the information on file. Then, it either accepts or rejects that you are who you claim to be.

Systems use the same three components:

- A sensor that detects the characteristic being used for identification
- A computer that reads and stores the information
- Software that analyzes the characteristic, translates it into a graph or code and performs the actual comparisons

b) Use of firewalls both software and hardware, including proxy servers

Firewall

A firewall is a program or hardware device that filters the information coming through the Internet connection into your personal computer or into a company's network.

It is set up to allow mainly one way access, i.e. you can go out onto the Internet and access pages, but it checks everything coming back against a set of rules. If the data coming back is from an unauthorised source, then it is blocked.

You may have heard people saying, 'I can't get on that site at school because it's been blocked'; that is the firewall in action.

Proxy Server

A proxy server is used to protect a network from problems that can be associated with using the internet. A proxy server will send packets of data from a host on its own network to a server on the internet. Before sending, however, the proxy server will take the packet of data apart, check it and reassemble it. It can check for corrupted data, for viruses and for its destination, which can be blocked if necessary.

The proxy server also inspects packets of data being sent to hosts on its network. Again, it breaks the packets up, inspects them and reassembles them. Unwanted data, data from unauthorised sites and corrupted or infected data can be removed.

Sometimes, a proxy server will hide the IP addresses of individual hosts on the network, exchanging their IP addresses for its own. This adds a level of security from hackers.

Sometimes, proxy servers are set up to cache web pages. As a web page is received, it saves a copy of it. If the same page is requested again, even by a different host, it can send the page itself rather than relying on getting it from the internet. This makes access to pages much faster. Proxy servers can handle FTP transfers and Internet email using SMTP. For both, the proxy server can check all files and emails that pass through it and can remove any that might have viruses or have been received from unauthorised sites.

c) Use of Secure Socket Layer (SSL)

Websites need to have a system that protects sensitive data, such as credit card numbers and other account information, from being accessed by others. The standard method chosen to achieve this security is to employ a secure socket layer. This is another protocol and it has two distinct features. The first is to encrypt any data that is being sent and the second is to allow the computer to identify any server to which data is being sent.

Use of symmetric encryption (plain text, cypher text, use of a key)

Encryption

Encryption is the scrambling of data so that it becomes very difficult to unscramble and interpret. Scrambled data is called cypher text. Unscrambled data is called plain text. Unscrambling cypher text back to the original plain text is called decryption.

Data encryption is performed by the use of a cryptographic algorithm and a key. The algorithm uses the key to scramble and unscramble the data. Ideally, the algorithm should be made public (so that it can be scrutinised and analysed by the cryptographic community), while the key remains private.

In symmetric encryption the same key is used to encrypt and decrypt the message. In asymmetric encryption, one key is used to encrypt a message and another is used to decrypt the message. This is known as public-key encryption.

Showing understanding that increasing the length of a binary key increases the strength of the encryption

The key is fundamental to the strength of the encryption. You need the one correct key before you can decrypt the cypher text. It follows, then, that the longer the key, the greater the range of possible values it could have. The range of possible values is called the keyspace. The greater the keyspace, the more difficult it is for an unauthorised person to discover the correct key.

Encryption cannot make unauthorised decryption impossible; it can merely make it improbable. With unlimited processing capacity and unlimited time available, all cryptosystems could be broken. The purpose of encryption is to make it as unlikely as possible that a cypher text could be broken within the period of time during which the contents should remain secret.

Show understanding of the need to keep online systems safe from attacks including denial of service attacks, phishing, pharming

Denial of Service attacks

In a denial-of-service (DoS) attack, an attacker attempts to prevent legitimate users from accessing information or services. By targeting your computer and its network connection, or the computers and network of the sites you are trying to use, an attacker may be able to prevent you from accessing email, websites, online accounts (banking, etc.), or other services that rely on the affected computer.

Impact of DoS on computer systems:

- The most common and obvious type of DoS attack occurs when an attacker "floods" a network with information. When you type a URL for a particular website into your browser, you are sending a request to that site's computer server to view the page. The server can only process a certain number of requests at once, so if an attacker overloads the server with requests, it can't process your request. This is a "denial of service" because you can't access that site.
- An attacker can use spam email messages to launch a similar attack on your email account. Whether you have an email account supplied by your employer or one available through a free service such as Yahoo or Hotmail, you are assigned a specific quota, which limits the amount of data you can have in your account at any given time. By sending many, or large, email messages to the account, an attacker can consume your quota, preventing you from receiving legitimate messages.

Safeguards available to remove or minimise the risk:

Unfortunately, there are no effective ways to prevent being the victim of a DoS attack, but there are steps you can take to reduce the likelihood that an attacker will use your computer to attack other computers:

- Install and maintain anti-virus software
- Install a firewall, and configure it to restrict traffic coming into and leaving your computer
- Follow good security practices for distributing your email address. Applying email filters may help you manage unwanted traffic

Phishing

Phishing is where the creator sends out a legitimate-looking email in the hope of gathering personal and financial information from the recipient. As soon as the recipient clicks on the link in the email/attachment they are sent to a bogus (fake) website where they will be asked for personal information.

Impact of phishing on computer systems:

- Access to personal information, such as credit card numbers, PINs, etc., can lead to fraud and other illegal use of whatever personal and financial information the creator can get.

Safeguards available to remove or minimise the risk:

- Many ISPs can filter out these phishing emails, but users should always be aware of the risk and exercise caution when opening emails.

Pharming

Pharming is where malicious code software is installed on the hard drive of the user's computer or on the actual web server. This code will re-direct users to bogus (fake) websites without the user's consent or knowledge. Unlike phishing, which requires the user to open an email, in pharming there is no need for the user to consciously take any action.

Impact of pharming on computer systems:

- Access to personal information, such as credit card numbers, PINs, etc., can lead to fraud and other illegal use of whatever personal and financial information the creator can get.

Safeguards available to remove or minimise the risk:

- Certain anti-spyware software can identify and remove pharming code.
- The user should always be alert and look out for 'clues' that they are being re-directed to another site.

Bibliography

- New Higher Computing by John Walsh (Hodder & Stoughton)
- Cambridge International AS and A Level Computing coursebook by Leadbetter, Blackford and Piper (Cambridge University Press)
- A Level Computing 5th Edition by Heathcote & Langfield (Payne-Gallway Publishers Ltd)
- A2 Level for OCR Applied ICT by K. Mary Reid (Heinemann)
- Cambridge IGCSE Computer Studies Coursebook by Leadbetter, Wainwright and Stinchcombe (Cambridge University Press)
- Cambridge IGCSE Computer Studies Revision Guide by Watson and Williams (Cambridge University Press)
- Numerous sections referenced throughout these notes www.wikipedia.org
- http://www.wisegeek.com/
- http://hwang.cisdept.csupomona.edu/cis311/design.aspx?m=sp
- http://www.igcseict.info/theory/6/internet/index.html
- http://www.teach-ict.com/gcse_computing/ocr/GCSE_A451_topics.html
- http://science.howstuffworks.com/biometrics.htm
- http://www.us-cert.gov/ncas/tips/ST04-015