1.1 Data representation

Chapter at a glance:

The basic building block in all computers is the binary number system.

A binary digit is commonly referred to as a BIT; 8 bits are usually referred to as a BYTE.

The byte is the smallest unit of memory in a computer.

It should be pointed out here that there is some confusion in the naming of memory sizes.

The unit was established by the International Electrotechnical Commission (IEC) in 1998, has

been accepted for use by all major standards organizations, and is part of the International System

of Quantities. The kibibyte was designed to replace the kilobyte in those computer science contexts

in which the term kilobyte is used to mean 1024 bytes. The interpretation of the kilobyte to denote

1024 bytes, conflicting with the SI definition of the prefix kilo (1000), is still common, mostly in

informal computer science contexts.

The IEC convention is now adopted by some organisations. Manufacturers of storage devices often

use the denary system to measure storage size. For example:

1 kilobyte = 1000 byte

1 megabyte = 1000000 bytes

1 gigabyte = 1000000000 bytes

1 terabyte = 1000000000000 bytes and so on.

The IEC convention for computer internal memories (including RAM) becomes:

1 kibibyte (1 KiB) = 1024 bytes

1 mebibyte (1 MiB) = 1048576 bytes

1 gibibyte (1 GiB) = 1073741824 bytes

1 tebibyte (1 TiB) = 1099511627776 bytes and so on

**Example Question: A company advertises its backup memory device as having 500 GB of storage. A customer wishes to know how many 8 MB files could be stored on the device. The company claimed that up to 62 500 files (assuming each file is 8 MB) could be stored. The customer calculated that 64 000 files could be stored. Explain the difference between these two storage values. Show any calculations you use in your explanation.**

**......................................................................................................................................................**

**....................................................................................................................**

**–company calculation is based on 1 GByte = 1000 Mbyte – so (500 × 1000)/8 = 62 500 files**

**– customer calculation based on 1 GByte = 1024 Mbyte – so (500 × 1024)/8 = 64000 files**

**– giving the difference of 1500 files [3]**

***Binary-to-Decimal & Denary-to-Binary Conversion*** –use binary notation (place values) i.e. 128, 64, 32, 16, 8, 4, 2, 1.

For Binary-to-Hexadecimal conversion firstly groups of 4 bits are made from right to left and each

group is converted separately using 8, 4, 2, 1 notation.

For **Hexadecimal-to-Binary conversion** each hex digit is separated by other and then each hex digit is converted separately using 8 4 2 1 notation.

For **Denary-to-Hexadecimal conversion** LCM of the denary number is taken.

For **Hexadecimal-to-Denary conversion** hexadecimal notation (place value) is used eg. 4096 256 16 1

Memory Dump is display of memory contents and address in hexadecimal on screen or printed on

paper. It is powerful fault-tracing tool for expert programmers.

Hexadecimal are used in HTML to represent colour codes (RGB Model). For example: # ff0000 for

bright red and #980000 for darker red.

MAC Addresses are unique number of NIC (Wi-Fi, Bluetooth. or wired connection i.e. Ethernet).

They are 48 bit long, but converted into 12 hexadecimal digits (in 6 pairs) making them short and

easier to understand. For 00-1C-2A-FF-01. 1st 3 pairs represent manufacturer while the other

represent serial number of product.

48 bit long address means there are 281,474,976,710,656 possible MAC addresses in the world.

UAA (Universally Administered MAC Address) are most common. These are the MAC addresses

set by manufacturer

LAA (Locally Administered MAC Addresses) are changed locally to bypass firewall, or to assign

MAC address of specific format.

URL encoding:

Web addresses can be written using hexadecimal rather than denary. Hexadecimal codes are

preceded by a % sign. For example, the word “www.ruknuddin.com” is written as:

r u k n u d d i n

in hex %72 %75 %6B %6E %75 %64 %64 %69 %6E

w w w . r u k n u d d i n . c o m

%77 %77 %77 %2E %72 %75 %6B %6E %75 %64 %64 %69 %6E %2E %63 %6F %6D

**Machine code and Assembly code** are examples of low-level languages and are used by software developers when producing, for example, computer games. They look difficult but they have many advantages at the development stage of software writing (especially when trying to locate errors in the code). Using hexadecimal makes it much easier, faster and less error prone to write code compared to binary. Character: Any text, number or symbol.

**Why compress on the Internet?**

**The Internet can be slow at times, especially in it’s early days relying on early 56Kbps modems as opposed to the 3-4 Mbps average. Therefore if you could compress the files that were being sent on the Internet, then you sent smaller files, and smaller file sizes meant faster downloads. It also frees up the network and avoids clogging up the bandwidth.**

**Compression:** The method of reducing file size.

**Lossy Compression:** The file is reduced in size for transmission and storage; by permanently removing some redundant information from the file

**Lossless Compression:** The file is reduced in size for transmission and storage; it is then put back together again later producing a file identical to the original

**MIDI:** A MIDI file consists of a list of commands that instruct a device like an electronic organ, how to produce a particular sound or musical note.

Examples of MIDI commands include:

• **note on/off:** this indicates that a key has been pressed/released to produce/stop producing a musical note

•  **key pressure**: this indicates how hard the key has been pressed (this could indicate loudness of the music note or whether any vibrato has been used, and so on).

The whole piece of music will have been stored as a series of commands but no actual musical notes. Their size, compared with an MP3 file, is considerably smaller. MIDI is essentially a communications protocol that allows electronic musical instruments to interact with each other. The MIDI protocol uses 8-bit serial transmission with one start bit and one stop bit, and is therefore asynchronous. MIDI is essentially a communications protocol that allows electronic musical instruments to interact with each other.

**MP3:** File compression system for music which does not noticeably affect the quality of the sound. This is done using file compression algorithms which use PERCEPTUAL MUSIC SHAPING; this essentially removes sounds that the human ear can’t hear properly.

For example, an 50 megabyte music CD can be reduced to 5 megabytes.

50 MB x 90/100= 45 MB reduced. 50 MB x 10/100=5 MB new file size.

MP3 technology reduces the size of a normal music file by about 90 per cent.

**MP4:** MPEG-4 (MP4) format allows the storage of multimedia files rather than just sound. Music, videos, photos and animation can all be stored in the MP4 format. Videos, for example, could be streamed over the internet using the MP4 format. It uses lossy compression.

**JPEG:** File compression format designed to make photo files smaller in size for storage and for transmission. It uses lossy compression and compresses a file between factor of 5 to 15.

For example 2000 pixels wide and 2000 pixels high image will have 2000 × 2000= 4,000,000 pixels. This is often referred to as a 4-megapixel image. A raw bitmap can often be referred to as a TIFF or BMP image (file extension .TIF or .BMP). The file size of this image is determined by the number of pixels. In the previous example, a 4-megapixel image would be 4 megapixels × 3 colours(RGB) =12 megabytes. This image will be compressed at factor of 5 (12/5=2.5 mb) to factor of 15 (12/15= 0.8mb). Text and number file formats: Text and numbers are usually stored in an ASCII format. Text files are also compressed. Lossless compression method is used for text and numbers. These use complex algorithms that work on redundancy or repeated sections of words (e.g. OU in yOUr, cOUntry or mOUntain).

THIS Repeated words, such as 1 SECTION

SECTION ‘THIS’ could be put into a data S2S Y3 2 1

SHOWS YOU dictionary and be replaced by W3LD WORK

HOW THIS ‘ 1’. Repeated word sections,

WOULD such as ‘HOW’ and ‘OU’ could

WORK’ be replaced by the numbers

‘2’ and ‘3’. Our phrase then

becomes

The following section shows, in very simple terms, how this could work:

**Algorithm:** step-by-step set of instruction to solve a problem.

**Register:** Immediate access store in the processor. It can store small piece of data.

**1.2 Communication and Internet technologies**

Chapter at a glance:

A **network** is defined as a collection of computers and peripheral devices (such as printers) connected together. Generally, a network over short distances is called a local area network (LAN) while those over great distances are wide area networks (WAN) . Network adapters: These adapters (also called network interface cards or NICs) connect computers to a network so that they can communicate.

**Network hubs and switches**: Hubs and switches connect two or more computers to an Ethernet network.

**Routers** connect computers and networks to each other (for example, a router can connect your home network to the Internet).

**Modem**: Hardware device that converts signals from analogue to digital and vice versa; typically used to convert signals sent over the public service telephone network.

A **WEB BROWSER** is software which allows a user to display a web page on their computer screen. Web browsers interpret or translate the HTML code from websites and show the result of the translation.

**SIMPLEX DATA TRANSMISSION**: sending data in one direction only (i.e. from sender to receiver) Example: data being sent from a computer to a printer, from keyboard to processor etc.

**HALF-DUPLEX DATA TRANSMISSION**: Sending data in both directions but only one at a time (i.e. data can be sent from ‘A’ to ‘B’ or from ‘B’ to ‘A’ along the same line, but not at the same time). Example: a walkie-talkie, fax machine, reading or burning on cds, dvds.

**FULL-DUPLEX DATA TRANSMISSION:** Sending data in both directions simultaneously (i.e. data can be sent from ‘A’ to ‘B’ and from ‘B’ to ‘A’ along the same line, both at the same time). Example: a phone line, video recording and playing at the same time from DVD-RAM.

**SERIAL DATA TRANSMISSION** is when data is sent, one bit at a time, over a single wire or channel

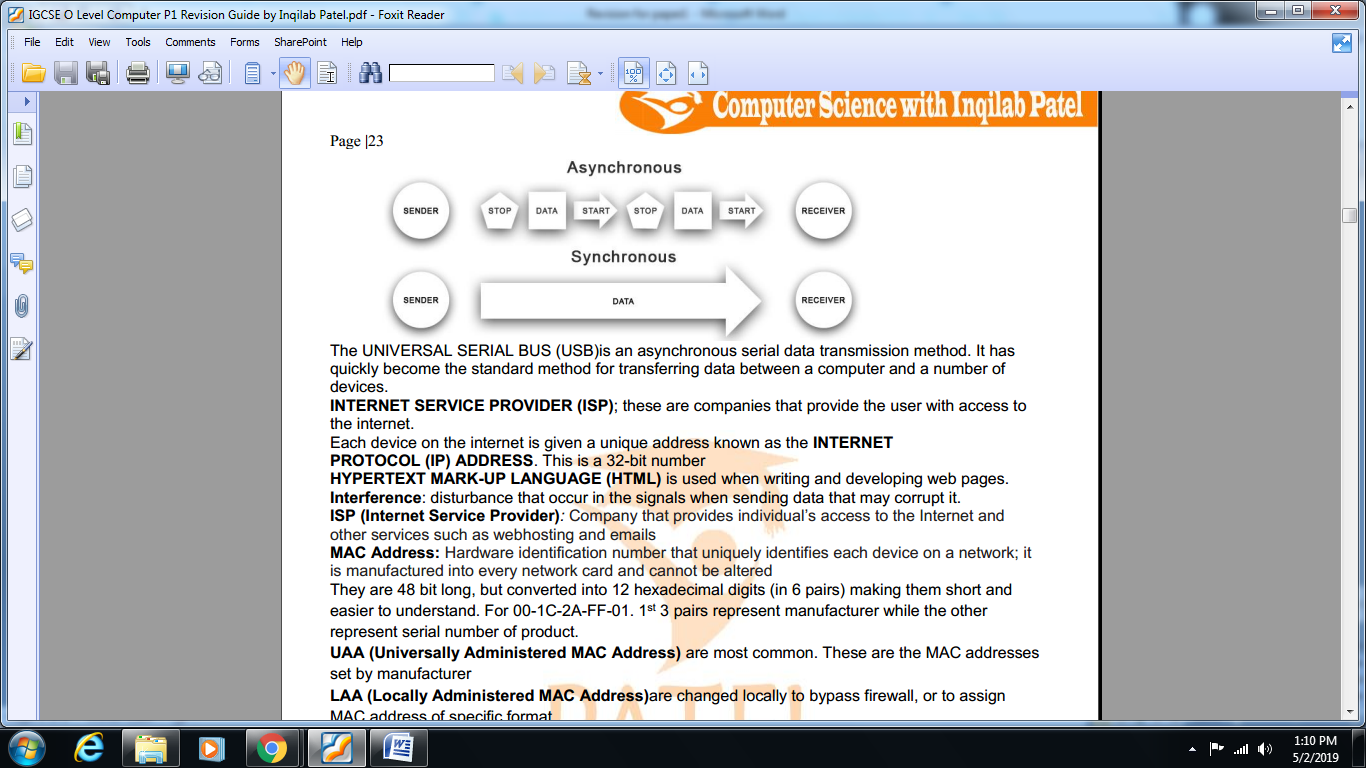
**Bit**: It is short of binary digit. It is the smallest unit of data in computer. It consists of a 0 or an 1.

**Bit rate**: the rate of transmitting data Serial Transmission: transfer of data bit by bit using single wire (bits are sent one after the other in a single stream).

**Parallel Transmission**: transfer of data in groups of bits using multiple wires.

**ASYNCHRONOUS DATA TRANSMISSION** refers to data being transmitted in an agreed bit pattern. Data bits (1s and 0s) are grouped together and sent with CONTROL BITS means START bit and STOP bit. Discussion forums and email are two examples of how asynchronous communication

**SYNCHRONOUS DATA TRANSMISSION** is a continuous stream of data (unlike asynchronous data which is sent in discrete groups). The data is accompanied by timing signals generated by an internal clock. This ensures that the sender and receiver are synchronized with each other. Chat rooms and online conferences are good examples of synchronous communication.



The **UNIVERSAL SERIAL BUS (USB)**is an asynchronous serial data transmission method. It has quickly become the standard method for transferring data between a computer and a number of devices.

**INTERNET SERVICE PROVIDER (ISP);** these are companies that provide the user with access to the internet. Each device on the internet is given a unique address known as the **INTERNET PROTOCOL (IP)** ADDRESS. This is a 32-bit number

**HYPERTEXT MARK-UP LANGUAGE (HTML)** is used when writing and developing web pages. Interference: disturbance that occur in the signals when sending data that may corrupt it.

**ISP (Internet Service Provider):** Company that provides individual’s access to the Internet and other services such as webhosting and emails

**MAC Address:** Hardware identification number that uniquely identifies each device on a network; it is manufactured into every network card and cannot be altered. They are 48 bit long, but converted into 12 hexadecimal digits (in 6 pairs) making them short and easier to understand. For 00-1C-2A-FF-01. 1st 3 pairs represent manufacturer while the other represent serial number of product.

**UAA (Universally Administered MAC Address)** are most common. These are the MAC addresses set by manufacturer **LAA (Locally Administered MAC Address)** are changed locally to bypass firewall, or to assign MAC address of specific format.

**Cascade style sheet:**

**IP Address:** Location of a given computer/device on a network; can be a static or dynamic value. IP addresses are 32 bit long converted into 4 groups of denary numbers. IP address starts from 0.0.0.0 and ends at 255.255.255.255.

**URL (Uniform Resource Locator):** The standard format for referring to is source on the Internet; also called Uniform Resource Indicator (URI); made up of:

●the protocol, e.g. http

●the domain name, e.g. ruknuddin.com

●the filename e.g. computer.html

**URL encoding:** Web addresses can be written using hexadecimal rather than denary. Hexadecimal codes are preceded by a % sign. For example, the word “www.ruknuddin.com” is written as:

r u k n u d d i n

in hex %72 %75 %6B %6E %75 %64 %64 %69 %6E

Page 24