

## 8. Security and data management

Information stored on computers can be sensitive and needs to be looked after very carefully. This means restricting access to the information. Inappropriate and unauthorised access to this information is likely to have serious consequences and could result in legal penalties, identity theft, financial loss, fraud and invasion of privacy.

Other risks to information stored on computers include loss due to accidental deletion, or overwriting parts of files in error; mechanical damage to hard discs, which are the most fragile parts of a computer; power failure whilst work is in progress; accidental damage to hardware, such as fire or damage caused by spilling a drink.

Most of these risks can be managed by adopting efficient procedures for saving work and making regular backups.

### Network security

Security is of paramount importance to any network as the loss of data, in particular, personal or confidential data can have many serious consequences. Risks to data become greater as it is shared across a network.

### User access levels

It is not desirable that every user should be able to access all the data on a computer system. User access levels are one method used to allow certain users read and/or write access to data on a computer system. For example, in a program used within a company, an administrator, possibly the owner, will have read and write access to all data on the system. An assistant, however, will not have access to confidential data such as employees' salaries. User access levels will define which users can change and view, view but not change, or not view stored data.

### Suitable passwords

Passwords are commonly used to prove a person's identity to a computer system, thus allowing them access to relevant data.

Different programs may require a user to use different complexities of password, as well as different character lengths. An example of a simple password may be the user's town of birth, or the word 'password'. A more complex

#### INTERESTING FACT

The average Internet user has 25 online accounts, 6.5 passwords and waits an average of 3.1 months before changing passwords.

password may require the user to use a combination of upper and lower case alphanumeric characters, for example 'Pa55word1234'. Increasingly, computer programs require users to use a combination of upper and lower case alphanumeric characters as well as other non-alphanumeric characters such as @ ! ~ - / \ %, for example 'P@55word/1234!'.

Another user can guess short simple passwords, or a hacker may have access to programs that have the ability to try multiple guesses in quick succession. This is known as a *brute force attack*. Passwords that use a combination of upper and lower case alphanumeric characters as well as other non-alphanumeric character, will be much harder to guess and will take longer to 'brute force'.

As a rule of thumb, the following formula can be used to determine the number of attempts it would take to brute force a password.

$$\text{Attempts} = \text{Number of characters}^{\text{Password length}}$$

So a password, such as 'computer' (8 characters), which only contains lower case characters from the 26 letter English alphabet will take:

$$\text{Attempts} = 26^8 = 208,827,064,576$$

(on a typical 3.5GHz computer, this would take less than 6 seconds to brute force)\*

Whereas a password that contains upper and lower case alphanumeric characters, such as 'Computer1' (9 characters), has  $26 + 26 + 10 = 62$  possible characters. This will take:

$$\text{Attempts} = 62^9 = 13,537,086,546,263,552$$

(on a typical 3.5GHz computer, this would take just over 1 hour to brute force)\*

\*assuming one attempt per clock-tick

## Encryption techniques

Encryption is the conversion of data, using an algorithm, into a form, called cyphertext that cannot be easily understood by people without the decryption key.

When data is encrypted, a logical operator is sometimes used, called the XOR logical operator.

## XOR

The XOR logical operator has two inputs and one output. The output is 1 only if A and B are different.

Input (A)	Input (B)	Output (A XOR B)
0	0	<b>0</b>
0	1	<b>1</b>
1	0	<b>1</b>
1	1	<b>0</b>

When encrypting data, the XOR logical operator is performed on the original data and a *key*. The key is a secure binary number, known only to the sender and recipient.

In this example, we will encrypt the data 10101010, using the key 11110000.

<b>Original Data</b>	<b>10101010</b>	
<b>Key</b>	<b>11110000</b>	XOR
<hr/>		
<b>Cyphertext</b>	<b>01011010</b>	

The original data, 10101010, is now encrypted and can be transmitted as 01011010.

To recover the original data, the cyphertext is *XOR'ed* with the key.

<b>Cyphertext</b>	<b>01011010</b>	
<b>Key</b>	<b>11110000</b>	XOR
<hr/>		
<b>Original Data</b>	<b>10101010</b>	

Other, more complex techniques are also used to encrypt data e.g. SHA256 and Blowfish.

## Compression and compression types

Compression is the process of making a file size smaller. This may be advantageous as it allows more data to be stored on the disk and files may also be transferred more quickly. There are two primary methods that are used to compress files stored on a computer system; these are *lossy* and *lossless*.

## Lossless compression

Lossless compression uses an algorithm that compresses data into a form that may be decompressed at a later time without any loss of data, returning the file to its exact original form. It is preferred to lossy compression when the loss of any detail, for example in a computer program or a word-processed document, could have a detrimental effect.

A simplified version of lossless compression on a word-processed document may be to replace a common string, such as 'the', with a token such as the symbol @. One character takes 1 byte of memory; therefore, the string 'the' would take 3 bytes.

Original uncompressed text	The word <b>the</b> , is <b>the</b> most frequently used word in <b>the</b> English language.	71 characters (bytes)
Compressed text	@ word @, is @ most frequently used word in @ English language.	63 characters (bytes)




This is an 11% reduction in the file size!

## Lossy compression

Lossy compression is a technique that compresses the file size by discarding some of the data. The technique aims to reduce the amount of data that needs to be stored.

The following versions of the Eduqas logo show how much of the data can be discarded, and how the quality of the images deteriorate as the data that made up the original is discarded. Typically, a substantial amount of data can be discarded before the result is noticeable to the user. The compression ratio is calculated using the simple formula:

$$\text{Compression ratio} = \frac{\text{Original file size}}{\text{Compressed file size}}$$

		
Original image 100 kB	Compressed image 10 kB (compression ratio = $100/10 = 10$ or 10:1)	Compressed image 5 kB (compression ratio = $100/5 = 20$ or 20:1)
<b>File size</b>	<b>File size</b>	<b>File size</b>

Lossy compression is also used to compress multimedia data, such as sound and video, especially in applications that stream media over the Internet.

## Network policies

### Acceptable use

Network policies are documents written to outline the rules that users are required to follow while using a computer network. Each document is often several pages long, written and agreed by a committee. Following its publication, network users will be expected to adhere to the rules.

Typical rules set out in these policies include a list of unacceptable types of website that should not be visited and activities that are not allowed on the network, such as gambling and installation of unauthorised software.

## Disaster recovery

Given the amount of important data often stored on a computer network, it is essential that an effective disaster recovery policy be in place. Disasters include:

- fire, flood, lightning, terrorist attacks etc.
- hardware failure, e.g. power supply unit failing
- software failure, e.g. virus damage
- accidental and malicious damage, e.g. hacking

There are usually three parts to a disaster recovery policy:

- **before the disaster:** risk analysis, preventative measures and staff training
- **during the disaster:** staff response – implement contingency plans
- **after the disaster:** recovery measures, purchasing replacement hardware, reinstalling software, restoring data from backups

## Backup

A **backup** is a copy of data that can be used if the original data is lost.

Backups of all data should be made regularly as the older the backed up data becomes, the less likely it is to match any current data stored on a computer system.

A backup policy sets out how often and to what medium backups are made. The backup medium is generally different to the active storage medium. Historically, the medium used was magnetic tape backup.

A typical backup policy would require that three different backups be kept at any given time, with one of these being stored off-site. The oldest backup copy would be named the *grandfather*, the second oldest backup being named the *father* and the most recent backup being called the *son*. When a new backup is made, the oldest backup, the *grandfather* is overwritten and becomes the *son* backup, with the original son becoming the *father* and the father becoming the *grandfather*. This backup policy is called the *grandfather-father-son* method.

### INTERESTING FACT

Key causes of data loss are:

- 78% hardware failure
- 11% human error
- 7% software failure
- 2% computer viruses
- 1% other

## Archiving

Data held on computer systems is often **archived**. Archiving is the process of storing data that is no longer in current or frequent use. It is held for security, legal or historical reasons. The process of archiving data frees up resources allows on the main computer system and faster access to data that is in use.

eBay – Customer’s usernames, encrypted passwords, email addresses of 145 million users stolen. No payment information was taken

use. It is reasons. resources allows

## Cybersecurity

Online networks are vital to many business operations, but they are liable to attacks targeted to access confidential data, such as customers’ details or technical information about products etc. This level of data is very expensive to gather and its loss could result in loss of reputation and even business failure.

### INTERESTING FACT

October 2013. Email addresses and passwords for 150 million users and credit card data for 2.9million users stolen from Adobe’s network.

### INTERESTING FACT

eBay – Customer’s usernames, encrypted passwords, email addresses of 145 million users stolen. No payment information was taken.

Cybersecurity refers to the range of measures that can be taken to protect computer systems, networks and data from unauthorised access or cyberattack. Cyberattacks are carried out using various types of **malware** (malicious software), including;

**Viruses.** Viruses are programs that can replicate themselves and be spread from one system to another by attaching themselves to host files. They are used to modify or corrupt information on a targeted computer system.

**Worms.** Worms are self-replicating programs that identify vulnerabilities in operating systems and enable remote control of the infected computer.

**Spyware.** Installed by opening attachments or downloading infected software. Spyware can be used to collect stored data without the user’s knowledge.

Keyloggers are a type of spyware that can be used to track keystrokes and capture passwords, account numbers for fraudulent use.

Parents can use keylogger software to monitor their children’s online activity.

**Trojans.** A Trojan is a program that appears to perform a useful function, but also provides a ‘backdoor’ that enables data to be stolen.

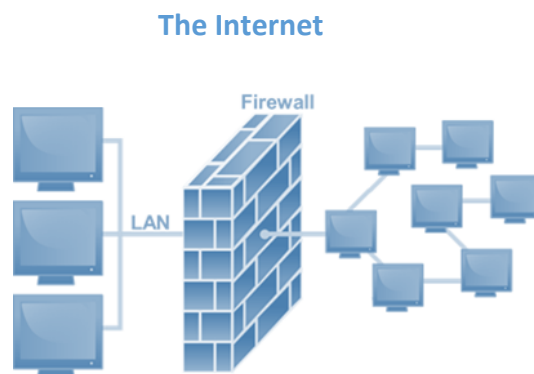
## Protection against malware

**Install Virus protection software**, also called anti-virus software, is a program that can be loaded into memory when the computer is running. It monitors activity on a computer system for the signs of virus infection. Each virus has its own unique 'signature' that is known

to virus protection software and stored in a database. Data stored on a computer system is scanned to see if any of the virus signatures within the database exist on the system.

There are many thousands of known viruses, and new viruses are created daily. Virus protection software therefore needs to be updated regularly to combat these.

**Use a firewall.** A firewall can be a software or hardware security system that controls the incoming and outgoing network traffic. Packets of data are analysed to determine whether they should be allowed through or not.



The basic function of a firewall is to monitor where data has come from and where it is going and to determine if this communication is allowed. It does this by checking a list of pre-defined rules.

**Keep your operating system up to date.** New ways to bypass the operating system's built-in security are often discovered and can be covered by installing the security patches issued by the operating system manufacturer.

**Use the latest versions of web browsers.** As for operating systems the manufacturers of web browsers seek to continually improve their products and remove possible security vulnerabilities. Most browsers will download updates automatically, but will need a restart for the update to be installed.

### INTERESTING FACT

Some advanced viruses attempt to evade the virus protection software by changing their own code so that they no longer match the "signature" in the virus signature database.

These are known as *polymorphic viruses*.



**Look out for phishing emails.** Emails that ask you to confirm personal details are usually fakes. They should be caught by the spam filter, but be suspicious and do not provide any sensitive information.

If you suspect you have malware on your computer you will need to download and run a **malicious software removal tool** that should detect and remove malware not blocked by the anti-virus software.

## **Forms of Cyberattack**

Internet protocols, operating systems and network equipment all present inherent technical weaknesses that must be recognised and protected against. User behaviour can also compromise security e.g. sending sensitive documents to unintended recipients, opening malicious attachments to scam emails, or using the same passwords for multiple systems. Specific forms of attack include;

### **Shoulder surfing**

Shoulder surfing is using direct observation to get information. It is relatively simple to stand next to someone and watch as they fill out a form, or enter a PIN number, but shoulder surfing can also be carried out long distance with the aid of binoculars or even CCTV.

### **SQL injection**

SQL injection is a technique where malicious users can inject SQL commands into an SQL statement, via web page input. Injected SQL commands can alter SQL statements and compromise the security of information held in a database.

### **DoS attack.**

Denial of service (DoS) attacks do not attempt to break system security, they attempt to make your website and servers unavailable to legitimate users, by swamping a system with fake requests—usually in an attempt to exhaust server resources.

A DoS attack will involve a single Internet connection. Distributed denial of service (DDoS) attacks are launched from multiple connected devices that are distributed across the Internet. These multi-person, multi-device attacks target the network infrastructure in an attempt to saturate it with huge volumes of traffic.

#### **Carphone Warehouse data breach DDoS – 2015**

In July 2015, major UK smartphone retailer Carphone Warehouse suffered a serious data breach, which it later transpired, might have been aided using a DDoS 'distraction' attack. Up to one in five DDoS incidents are later found to be part of a data theft snatch in which IT staff are occupied fending off the DDoS, giving attackers more opportunity to sneak in and out.

### Password-based attacks.

Passwords are a good starting point, but are they secure? Cyber criminals have ways of finding out your password.

<b>Dictionary attack</b>	This uses a simple file containing words found in a dictionary. This attack uses exactly the kind of words that many people use as their password.
<b>Brute force attack</b>	Similar to the dictionary attack but able to detect non-dictionary words by working through all possible alphanumeric combinations from aaa1 to zzz10. It's not quick, but it will uncover your password eventually.
<b>Guess ???</b>	A user-generated password is unlikely to be random. Passwords are likely to be based upon our interests, hobbies, pets, family etc. Educated guesses often work.

### IP address spoofing.

A spoof is a hoax, or a trick. IP address spoofing involves an attacker changing the IP address of a legitimate host so that a visitor who types in the URL of a legitimate site is taken to a fraudulent or spoofed web page. The attacker can then use the hoax page to steal sensitive data, such as a credit card number, or install malware.

### Social engineering

Internet users frequently receive messages that request password or credit card information to "set up their account". Social engineering involves tricking a user into giving out sensitive information such as a password, by posing as a legitimate system administrator.

Examples of social engineering attacks carried out by deception include **phishing**, which is an attempt to acquire users' details using fake emails and websites, and **pharming**, where users are unknowingly re-directed to a fake website, again with the intention of identity theft.

## **Identifying vulnerabilities**

### **Footprinting.**

Footprinting is the first step in the evaluation of the security of any computer system. It involves gathering all available information about the computer system or network and the devices that are attached to it. Footprinting should enable a penetration tester to discover how much detail a potential attacker could find out about a system and allow an organisation to limit the technical information about its systems that is publicly available.

### **Ethical hacking**

Ethical hacking is carried out with the permission of the system owner to cover all computer attack techniques. An ethical hacker attempts to bypass system security and search for any weak points that could be exploited by malicious hackers. This information is then used by the system owner to improve system security.

### **Penetration testing**

Penetration testing is a sub set of ethical hacking that deals with the process of testing a computer system, or network to find vulnerabilities that an attacker could exploit. The tests can be automated with software applications or they can be performed manually.

Penetration test strategies include;

- Targeted testing, testing carried out by the organization's IT team and the penetration testing team working together.
- External testing, to find out if an outside attacker can get in and how far they can get in once they have gained access.
- Internal testing, to estimate how much damage a dissatisfied employee could cause.
- Blind testing, to simulate the actions and procedures of a real attacker by severely limiting the information given to the team performing the test.

## **Protecting software systems**

### **Secure by design**

Secure by design is an approach that seeks to make software systems as free of vulnerabilities as possible through such measures as continuous testing and adherence to best programming practices. At the design stage malicious practices are taken for granted and it is assumed that the new system will have invalid data entered or will be the subject of hacking attempts. These issues are taken into account and corresponding security measures

are considered to ensure security is not an afterthought and so reduce the need for addressing vulnerabilities and patching security holes as they are discovered in use.

Some examples of attacks that should be prevented during design and testing include;

### **Buffer overflow attacks**

A buffer overflow occurs when a program tries to store more data in a buffer (temporary data storage area) than it was intended to hold. This may occur accidentally through programming error, or it may be caused intentionally in a buffer overflow attack, where the overflow data may contain codes designed to change data, or disclose confidential information. Thorough testing, particularly of any library routines used, will help to prevent this type of attack.

### **Permissions**

Every time you want to install an app you are asked to give permission for the software to access certain settings and features of your device e.g. Facebook's Messenger App, which boasts over 1,000,000,000 downloads, requires permission to access a large amount of personal data and requires direct control over your mobile device. It is unlikely that many of those who downloaded this app read the full 'Terms of Service' before accepting them. It is not always easy to understand what you are permitting an app to do. Should you uninstall an app because its permissions are suspicious?

App developers are keen to develop interactive products that are useful, but they need to consider the scope of access and limit the number of permissions required at the design stage. There are malicious apps, but you can avoid them by using common sense.

### **Scripting restrictions**

Same Origin Policy (SOP) is a security measure that prevents a web site's scripts from accessing and interacting with scripts used on other sites. Running scripts from other sites would be dangerous because a malicious script from a compromised site could interact with a script from a legitimate site without restriction, potentially leading to malware infections or sensitive data being compromised.

A programmer can control the range and type of scripts allowed by setting the restrictions in, e.g. an HTML page header, or by using standard script execution settings such as unrestricted, trusted, restricted etc.

### **Accepting parameter without validation**

Dynamically generated HTML pages can introduce security risks if inputs are not validated on the way in. Malicious script can be embedded within input that is submitted to web pages and this could then appear to browsers as originating from a trusted source.

Approaches to prevent this type of cross-site scripting attack rely on the design of validation rules that will check and filter input parameters.

## **The role of cookies**

**Cookies** are data stored on a computer system. They allow websites to store a small amount of uniquely identifying data on your computer system while you are visiting. This may be useful as the website can then identify you in future without requesting that you identify yourself each time, i.e. by entering a username and password. Another use of a cookie would be when adding items to a shopping basket over a period of time. The cookie allows you to store this information between separate browsing sessions.