Chapter 6 Security, Privacy and Data Integrity

6.1 Data Security

6 Security, privacy and data integrity	
6.1 Data Security	
Candidates should be able to:	Notes and guidance
Explain the difference between the terms security, privacy and integrity of data	
Show appreciation of the need for both the security of data and the security of the computer system	
Describe security measures designed to protect computer systems, ranging from the stand-alone PC to a network of computers	Including user accounts, passwords, authentication techniques such as digital signatures, firewall, anti-virus software, anti-spyware, encryption
Show understanding of the threats to computer and data security posed by networks and the internet	Including malware (virus, spyware), hackers, phishing, pharming
Describe methods that can be used to restrict the risks posed by threats	
Describe security methods designed to protect the security of data	Including encryption, access rights

Data Security	Data Integrity	Data privacy
• Security is keeping data safe • Security is preventing data loss • e.g. usernames, passwords, encryption, firewall	• Integrity is making sure data is correct • Integrity is making sure data received is same as data sent • e.g. parity checks, double entry	Privacy is keeping data confidential, only accessible to authorized personnel

Why is it important to keep database secure?

e.g. why school database must be protected?

- Stop unauthorized data access, so that student's personal data is not accessible to unauthorized people
- Prevent loss of data, avoid student data getting deleted, e.g. student marks
- Prevent unauthorized changes to data, preventing students changing their grades

Preventing data loss and restricting data access

- Usernames and passwords
  - User has a username and password access to resources, can be limited to specific accounts
  - Cannot access them without valid username and password, prevents unauthorized access to system
  - To prevent unauthorized access
- Biometrics / Fingerprints / Retina Scan / Iris Scanner
- Password protect file
- Two step verification
- Token authentication / use of dongle / swipe cards
- Firewall
  - A firewall can be software or hardware. It sits between the user's computer and an external network (such as the internet) and filters information in and out of the computer.
  - Prevents unauthorized access to data
  - Monitors incoming and outgoing traffic
  - Blocks transmissions from unauthorized sources
  - Blocks signals that don't meet set of requirements
  - Keeps a log of signals
  - Applications can have network access denied
  - Maintains an allow list / deny list of IP addresses
  - Stops unauthorized access, by preventing hackers gaining access to system
- Anti-virus / Anti malware
  - Scans for malicious software
  - Quarantines or deleted any malicious software found
  - Scans at regular intervals
  - Should be kept up to date

  When should antivirus check for virus?
  - When a new file is downloaded
  - External storage device is connected

- Use of proxy server
- Auditing
  - Logging all changed to system
  - To identify anything suspicious, unauthorized use
- Application Security
  - Applying regular updates
  - Finding and fixing any security vulnerabilities in any application
- Encryption
  - Contents are scrambled so they cannot be understood without a decryption key
  - Cannot be understood by unauthorized personnel
- Physical Measures
  - Locked doors, keyboards
  - Secure methods of access

- Access Rights
  - Different access rights for individuals users
  - To stop users editing data they're not permitted to access
- Digital signatures
  - Digital signatures protect data by providing a way of identifying the sender of, for example, an email.
  - (email) message put through hashing algorithm to produce a digest
  - Digest encrypted with sender's private key (to create the digital signature)
  - the (digital) signature can only be decrypted with matching sender's public key

- Disk Mirroring
  - Data is written on two or more disks simultaneous

Threats to computer and data security

Malware

Malware is one of the biggest risks to the integrity and security of data on a computer system. Many software applications sold as antivirus are capable of identifying and removing most of the forms of malware described below.

Viruses

Programs or program code that can replicate and/or copy themselves with the intention of deleting or corrupting files or causing the computer to malfunction.

Spyware

software that gathers information by monitoring, for example, key presses on the user's keyboard. The information is then sent back to the person who sent the software – sometimes referred to as key logging software

Similarities between virus and spyware

- Both are pieces of malicious software
- Both are downloaded / installed/run without the user's knowledge
- Both can pretend to be / are embedded in other legitimate software when
- downloaded // both try to avoid the firewall
- Both run in the background

Difference between virus and spyware

- Virus can damage computer data; spyware only records / accesses data
- Virus does not send data out of the computer; spyware sends recorded data to third party
- Virus replicates itself; spyware does not replicate itself

Phishing

Phishing is when someone sends legitimate-looking emails to users. They may contain links or attachments which, when clicked, take the user to a fake website, or they may trick the user into responding with personal data such as bank account details or credit card numbers. The email often appears to come from a trusted source such as a bank or service provider. The key is that the recipient has to carry out a task (click a link, for example) before the phishing scam causes harm

Protection against phishing

- Users need to be aware of new phishing scams. Those people in industry or commerce should undergo frequent security awareness training to become aware of how to identify phishing (and pharming) scams.
- Do not click on links unless certain that it is safe to do so; fake emails can often be identified by greetings such as 'Dear Customer' or 'Dear emailperson@gmail.com', and so on.
- It is important to run anti-phishing toolbars on web browsers (this includes tablets and mobile phones) since these will alert the user to malicious websites contained in an email.
- Look out for https and/or the green padlock symbol in the address bar (both suggest that traffic to and from the website is encrypted).
- Regularly check online accounts and frequently change passwords.
    - Ensure an up-to-date browser, with all of the latest security upgrades, is running, and run a good firewall in the background at all times. A combination of a desktop firewall (usually software) and a network firewall (usually hardware) considerably reduces risk.

Pharming

Pharming is malicious code installed on a user's computer or on a web server. The code re-directs the user to a fake website without their knowledge (the user does not have to take any action, unlike phishing). The creator of the malicious code can gain personal data such as bank details from users. Often, the website appears to belong to a trusted company and can lead to fraud or identity theft
illegal access to a computer system without the owner's permission.

Protection against pharming

- using antivirus software, which can detect unauthorized alterations to a website address and warn the user
- using modern web browsers that alert users to pharming and phishing attacks
- checking the spelling of websites
- checking for https and/or the green padlock symbol in the address bar. It is more difficult to mitigate risk if the DNS server itself has been infected (rather than the user's computer)

Hacking – illegal access to a computer system without the owner's permission.
Protection – Firewall

6.2 Data Integrity

6.2 Data Integrity	
Candidates should be able to:	Notes and guidance
Describe how data validation and data verification help protect the integrity of data	
Describe and use methods of data validation	Including range check, format check, length check, presence check, existence check, limit check, check digit
Describe and use methods of data verification during data entry and data transfer	During data entry including visual check, double entry During data transfer including parity check (byte and block), checksum

Validation

Validation checks that data entered is reasonable. One example is range check, present check, type check.

Check digit

An example of a check digit calculation is **modulo-11**. The following algorithm is used to generate the check digit for a number with seven digits:

1. Each digit in the number is given a weighting of 7, 6, 5, 4, 3, 2 or 1, starting from the left.
2. The digit is multiplied by its weighting and then each value is added to make a total.
3. The total is divided by 11 and the remainder subtracted from 11.
4. The check digit is the value generated; note if the check digit is 10 then X is used.

For example:

Seven digit number:	4 1 5 6 7 1 0
Weighting values:	7 6 5 4 3 2 1
Sum:	$(7 \times 4) + (6 \times 1) + (5 \times 5) + (4 \times 6) + (3 \times 7) + (2 \times 1) + (1 \times 0)$
	$= 28 + 6 + 25 + 24 + 21 + 2 + 0$
Total:	= 106
Divide total by 11:	9 remainder 7
subtract remainder from 11:	11 − 7 = 4 (check digit)
final number:	4 1 5 6 7 1 0 4

Validation Checks

Validation test	Description	Example of data failing validation test	Example of data passing validation test
type	checks whether non-numeric data has been input into a numeric-only field	typing sk.34 in a field which should contain the price of an item	typing 34.50 in a field which should contain the price of an item
range	checks whether data entered is between a lower and an upper limit	typing in somebody's age as –120	typing in somebody's age as 48
format	checks whether data has been entered in the agreed format	typing in the date as 12-12-20 where the format is dd/mm/yyyy	typing in the date as 12/12/2020 where the format is dd/mm/yyyy
length	checks whether data has the required number of characters or numbers	typing in a telephone number as 012 345 678 when it should contain 11 digits	typing in a telephone number as 012 345 678 90 when it should contain 11 digits
presence	checks to make sure a field is not left empty when it should contain data	please enter passport number:.....	please enter passport number: AB 1234567 CD
existence	checks if data in a file or a file name actually exists	data look up for car registration plate A123 BCD which does not exist	data look up for a file called books_in_stock which exists in a database
limit check	Checks only one of the limits (such as the upper limit OR the lower limit)	typing in age as –25 where the data entered should not be negative	typing in somebody's age as 72 where the upper limit is 140
consistency check	checks whether data in two or more fields match up correctly	typing in Mr in the title field and then choosing female in the sex field	typing in Ms in the title field and then choosing female in the sex field
uniqueness check	checks that each entered value is unique	choosing the user name MAXIMUS222 in a social networking site but the user name already exists	choosing the website name Aristooo.com which is not already used

Verification

- Verification checks that the data entered is same as original. One example is double entry
- Checking that data entered is consistent to that of source
- Comparison of two versions of data
- e.g. double entry. Proof reading, visual check
- In event of mismatch user is forced to reenter data
- Does not check that whether data is sensible or not

Verification Checks from paper-based source during data entry

- Double Entry // The data from the form is entered twice by two different people from the paper form and automatically compared
- Visual check // The data from the compared by two different people after entry from to the paper form

Verification Checks

Parity Check

- Parity can be odd or even
- Parity check uses the number of 1s in a binary pattern
- If there is an even / odd number of 1s then parity is even / odd
- Following the transmission parity of each byte checked
- A parity bit is used to ensure parity is correct
- Automatically checks for errors on receipt of data
- Alerts if data has been received incorrectly, requests data to be resent
- Provides a verification check on data

How a parity block check can identify a bit that has been corrupted?

- Each byte has a parity bit
- An additional parity byte is sent with vertical and horizontal parity
- Each row and column must have odd/even number of 1s
- Identify the incorrect row and column
- Their intersection is the error

Give a situation where a parity block cannot identify corrupted bits?

- Errors in an even number of bits could cancel each other out
- Prevents error being identified
- Could appear to be correct

	parity bit	bit 2	bit 3	bit 4	bit 5	bit 6	bit 7	bit 8
byte 1	1	1	1	1	0	1	1	0
byte 2	1	0	0	1	0	1	0	1
byte 3	0	1	1	1	1	1	1	0
byte 4	1	0	0	0	0	0	1	0
byte 5	0	1	1	0	1	0	0	1
byte 6	1	0	0	0	1	0	0	0
byte 7	1	0	1	0	1	1	1	1
byte 8	0	0	0	1	1	0	1	0
byte 9	0	0	0	1	0	0	1	0
parity byte	1	1	0	1	0	0	0	1

How does a computer use parity byte to perform further check on received data?

- Parity bit of each column is worked out
- Computer generates a parity byte and compares
- If incorrect parity, there's an error occurred
- Incorrect bit can be identified

Checksum

- A calculation is done on block of data
- The result is transmitted with the data
- Calculation repeated at receiving end
- Results compared
- If different results an error has occurred

Automatic repeat request (ARQ)

Automatic repeat request (ARQ) is another method to check data following data transmission.
This method can be summarized as follows:

- ARQ uses acknowledgement (a message sent to the receiver indicating that data has been received correctly) and timeout (the time interval allowed to elapse before an acknowledgement is received).
- When the receiving device detects an error following data transmission, it asks for the data packet to be re-sent. » If no error is detected, a positive acknowledgement is sent to the sender.
- The sending device will re-send the data package if – it receives a request to re-send the data, or – a timeout has occurred.
- The whole process is continuous until the data packet received is correct or until the ARQ time limit (timeout) is reached.
- ARQ is often used by mobile phone networks to guarantee data integrity.