Security, privacy and data integrity

A-Level Computer Science

Security, privacy and integrity —three different ideas

These sound alike but mean different things:

- **security** 安全—protecting data from **unauthorised** 未授权 access, change or destruction.
- **privacy** 隐私—an individual's right to **control who sees their personal data**, with consent and a clear purpose.
- **integrity** 完整性—the data being **accurate and complete** —not corrupted or accidentally changed.

A file can be secure (only the right people can open it) but lack integrity (a typo corrupted it); or accurate but not private (anyone can read it). All three are needed.

Why security matters

Two things to protect: the **data** itself (keep it confidential, intact and available) and the **computer system** (a compromised system can attack others, steal credentials, or be held to ransom).

Threats from networks and the internet

- **malware** 恶意软件 (malicious software):
  - **virus** 病毒—self-copying code that attaches to other programs and spreads when they run.
  - **worm** 蠕虫—self-copying code that spreads over **networks** 网络 with no user action.
  - **Trojan horse** 木马—looks useful but hides malicious code.
  - **spyware** 间谍软件—secretly collects information (keystrokes, passwords).
  - **ransomware** 勒索软件—encrypts your files and demands payment.
  - **adware** 广告软件—pushes unwanted adverts.
- **phishing** 网络钓鱼—fake emails/sites that trick users into giving credentials.
- **hacking** 黑客入侵—unauthorised access, often via weak passwords or software flaws.
- **denial of service** 拒绝服务 (DoS/DDoS) —floods a server so real users cannot reach it.
- **eavesdropping** 窃听—capturing data in transit (a risk on open Wi-Fi).
- **man-in-the-middle** 中间人攻击—an attacker secretly relays or alters messages between two parties.
- **social engineering** 社会工程—tricking people into giving up information.

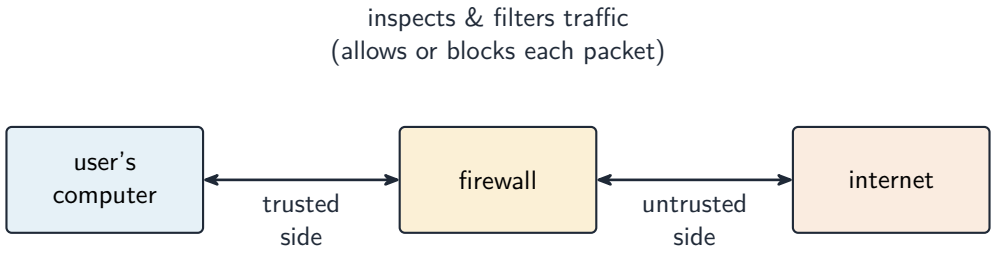
Security measures

A standalone PC

- a strong password; **antivirus** kept up to date; prompt **software updates**; **backup** 备份 to separate media; full-disk **encryption** 加密; a locked screen.

A networked PC

All the above, plus a **firewall** 防火墙, per-user **permissions** (admin rights only for admins), central account management, and **audit logs** 审计日志 (who logged in, what they touched).



A firewall sits between the user's computer and the internet

Across the internet

- **VPN** 虚拟专用网—encrypts traffic between the user and the corporate gateway.
- **HTTPS / TLS** —encrypt web traffic.
- intrusion detection —watches traffic for known attack patterns.

Matching measures to threats

- **interception in transit** → **encrypt** the data (HTTPS, VPN). Intercepted ciphertext is useless without the key.
- **unauthorised access** → strong **authentication** 身份验证 (long passwords; **two-factor authentication** 双因素认证 with a phone code or key); user **authorisation** 授权; lock-out after failed logins.
- **malware** → antivirus and real-time scanning; patching; avoid untrusted downloads.
- **phishing** → user training; email filtering; check the URL before entering credentials.
- **internal threats** → the **least-privilege** 最小权限 principle (give each user only what they need); auditing.
- **DDoS** → rate limiting and traffic filtering.

Protecting the data itself

- **encryption** —turn **plaintext** 明文 into **ciphertext** 密文 with a key. **Symmetric encryption** 对称加密 (AES) uses one shared key; **asymmetric encryption** 非对称加密 (RSA) uses a **public key** 公钥 and a **private key** 私钥. Protects data at rest and in transit.

- **access control** 访问控制—file permissions (read/write/execute), enforced by the OS.
- **authentication** —verify the user: something you know (password), have (token, phone), or are (fingerprint); strongest combined.
- **backups** —keep copies (some **off-site**) so loss or corruption is recoverable.
- **physical security** —locked server rooms, cable locks.

A security token shows a changing code for two-factor authentication ("something you have")

Image: RSA Security, Product image (www.tokenguard.com)

A fingerprint reader checks "something you are" —a feature of the person, not a password

Image: HID DigitalPersona, Product image (store.fulcrumbiometrics.com)

Data integrity

Data has integrity when it is accurate and complete. Two techniques: **validation** (catch bad data before storing) and **verification** (confirm data was entered or transferred correctly).

Validation —does the data make sense?

Validation 验证 checks data against sensible rules, automatically:

- **range check** —within limits (a month is 1–12).
- **length check** —the right number of characters.
- **type / character check** —the right kind of data (a phone field allows only digits).
- **format check** —matches a pattern (an email must contain `@`).
- **presence check** —required fields are not empty.
- **check digit** 校验位—an extra digit computed from the others (ISBN, card numbers) that spots transcription errors.
- **lookup check** and **consistency check** (e.g. delivery date order date).

Validation catches data that is wrongly **formatted**, but not data that is the right format yet factually wrong ("Bob" for "Bib").

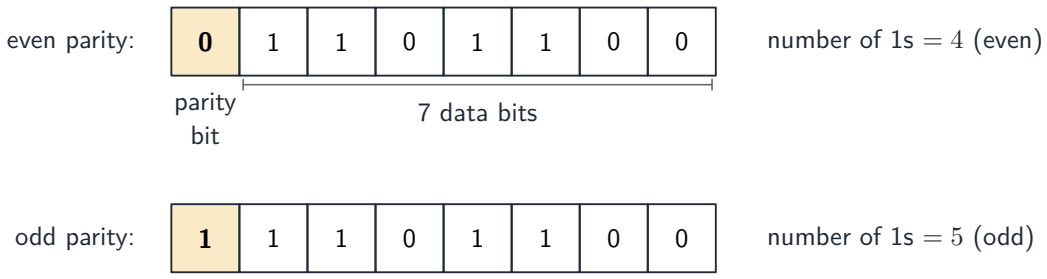
Verification —was the data entered or transferred correctly?

Verification 核对 checks the data was not changed in moving from one place to another.

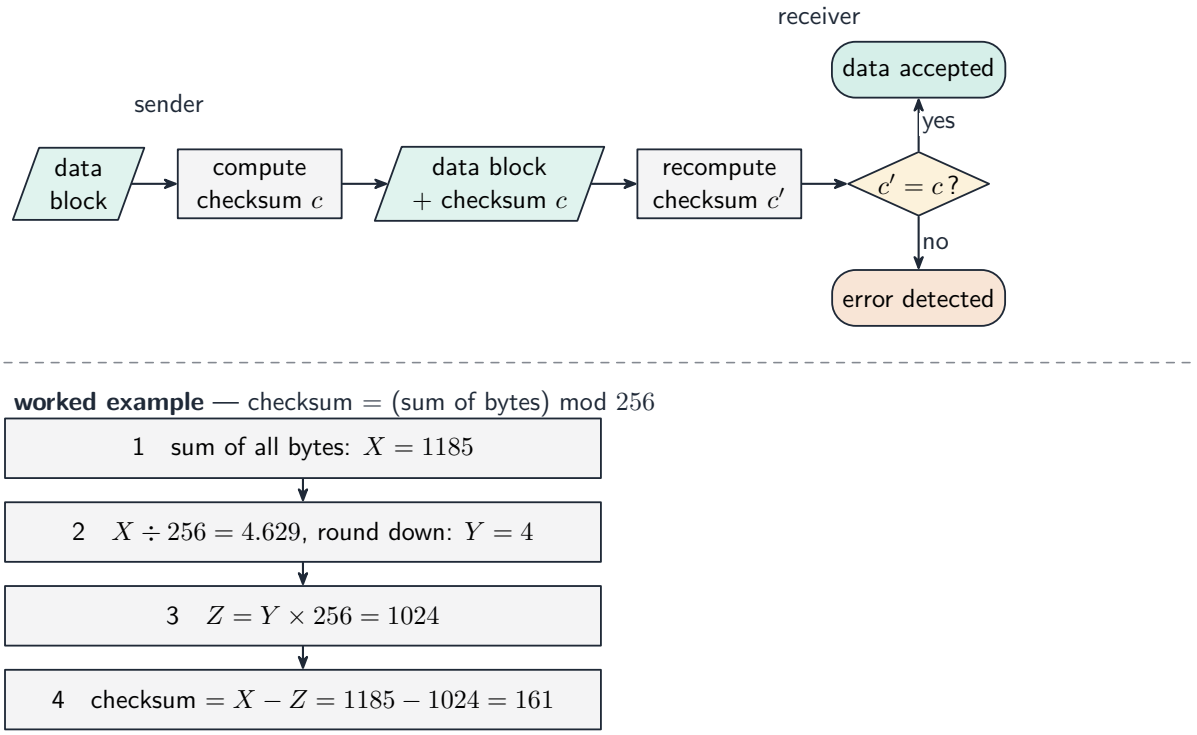
During entry: **double entry** (type it twice and compare, as for a new password) or **visual** check.

During transfer (bits can flip):

- **parity check** 奇偶校验—an extra bit makes the number of 1s even (even parity) or odd. The receiver re-counts. Catches single-bit errors.
- **checksum** 校验和—the sender sends a summary value of the data; the receiver re-computes it and compares.
- **cyclic redundancy check** 循环冗余校验 (CRC) —a stronger checksum using polynomial division, catching many more error types.



The parity bit is set to make the number of 1s even or odd



Working out a checksum for a block of data

Verification only proves what arrived matches what was sent —not that the data is correct, and not against deliberate tampering. **Validation** asks "is this sensible?"; **verification** asks "was this copied correctly?" —use both.