Introduction to Security

AP Cybersecurity

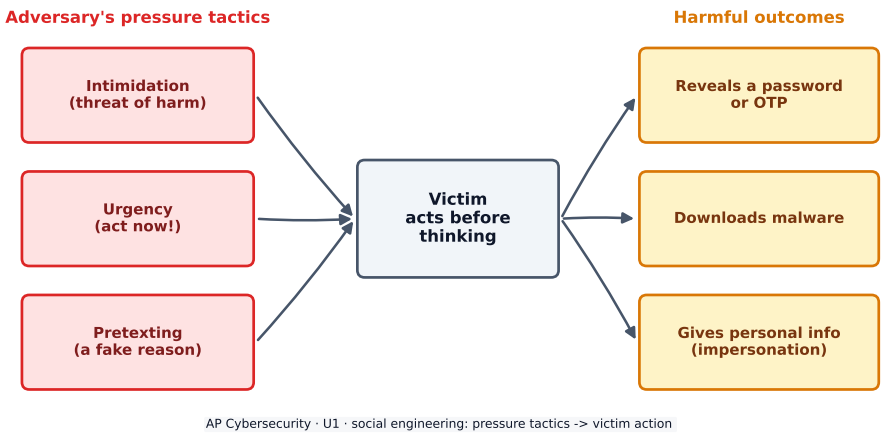
Understanding Social Engineering

The weakest part of any computer system is often the **human** using it. **Social engineering** 社会工程学 is the art of tricking people into breaking security - giving away a password, opening a bad file, or clicking a bad link. The attacker (we call them an **adversary** 对手) does not need to break the code; they only need to fool a person.

Most social engineering happens by **email**, text message, or social media, though it can also happen in person or by phone. The goal is **elicitation** 套取信息 - getting sensitive information out of someone without them realising.

Adversaries lean on two powerful feelings:

- **Intimidation** 恐吓 - the adversary threatens a bad result if you do not obey. Fear pushes you to act.
- **Urgency** 紧迫感 - the adversary invents a deadline ("reply in the next hour or your account closes"). When we feel rushed, we stop thinking carefully about whether an action is safe.



Social engineering uses psychological pressure to make a victim act before they think

The **impact** 影响 on a victim can be serious. They might reveal personal details (name, address, pet's name, birthday) that are later used to answer security **challenge questions** 安全问题 and **impersonate** 冒充 them. They might hand over a **one-time password (OTP)** 一次性密码, letting the adversary log in as them. Or they might download **malware** 恶意软件 that steals data from their browser.

Worked example. A phishing email reads: *"Over 90% of staff have already verified their account - confirm yours in the next hour or lose payroll access."* Two tactics are stacked here. "In the next hour" is **urgency** (a deadline that rushes you), and "over 90% of staff have already" is **consensus** (social pressure to follow the crowd). Naming each tactic - not just calling the email "suspicious" - is exactly what an exam answer needs.

Suspicious Website Logins

A hardware security key: strong authentication reduces damage when a password is phished

A **password attack** 密码攻击 is any attempt to log in using guessed or stolen passwords. In an **online** password attack the adversary tries passwords against a real login page. The warning signs are visible in the logs:

- many failed logins in a short time,
- login attempts at unusual hours,
- login attempts from unknown devices.

Adversaries succeed because people choose **weak** 弱 passwords. Common patterns include a word plus a two-digit year plus a special character (like `Summer24!`), or a pet's or family member's name. Because these patterns are so common, an adversary can build a **dictionary** 字典 of likely passwords from information gathered about you and let an automated tool try each one.

To make **authentication** 身份验证 stronger:

- Create passwords that are **long, random, and unique** - a **password manager** 密码管理器 can generate and store them for you.
- Avoid names, dates, and meaningful words.
- Turn on **multifactor authentication (MFA)** 多因素身份验证, which asks for extra proof (like a texted code) on top of the password.

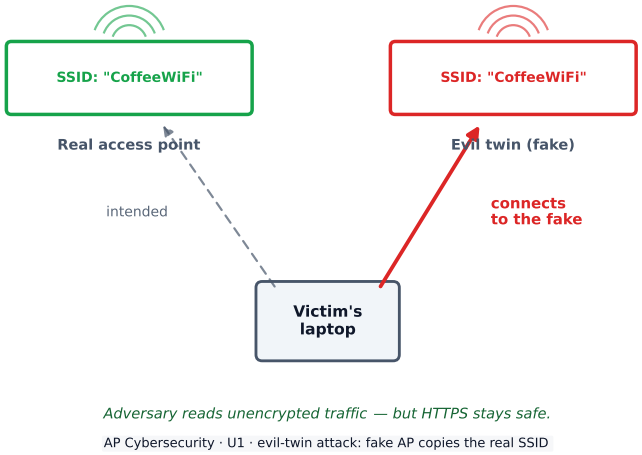
Best Practices for Public Networks

A security token: one-time codes and tokens stop password-only logins from being enough

Not all adversaries are the same. We classify them by **skill**: low-skilled attackers buy ready-made tools online and reuse known **exploits** 漏洞利用, while high-skilled attackers write their own tools and can discover brand-new holes called **zero days** 零日漏洞. Their **motivation** 动机 varies too - greed, revenge, politics, or belief.

Public **Wi-Fi** is a favourite hunting ground. Three wireless attacks you must know:

- **Evil twin** 双胞胎恶意热点 - the adversary sets up a fake **access point** 接入点 with a name (**SSID** 服务集标识符) copied from the real network. Victims connect to the fake one, and the adversary reads their traffic (though **encrypted** 加密的 sites like HTTPS stay safe).
- **Jamming** 干扰攻击 - the adversary floods the air with a strong radio signal so no one can connect. This is one kind of **denial of service (DoS)** 拒绝服务 attack.
- **War driving** 战争驾驶 - the adversary drives around detecting wireless networks and where their signal leaks outside a building.



An evil-twin access point copies the real network's name so victims connect to the attacker

To protect yourself on public networks: check that the network name **exactly** matches the one you intend to join, prefer encrypted sites, and consider a **virtual private network (VPN)** 虚拟专用网络, which encrypts all of your traffic to the VPN operator.

AI-Based Cybersecurity Attacks

Artificial intelligence gives adversaries powerful new tools. With enough voice and image samples, an adversary can build a **deepfake** 深度伪造 avatar to impersonate someone on a call. **Large language models (LLMs)** 大语言模型 let them write convincing **phishing** 钓鱼 emails in perfect, native-sounding language - removing the clumsy wording that once gave scams away.

AI also helps adversaries **on the back end**: crafting prompts that pull secret data out of an LLM, planting false information on websites so it poisons an LLM's training data, scanning the internet to gather facts about a target, and even writing new malware.

You can defend against many AI-augmented attacks: agree on a **shared secret** 共享秘密 word with close contacts to verify identity, enable MFA (so a cloned voice alone cannot log in), never type sensitive data into a chatbot, and always double-check AI output against reliable, non-AI sources.

AI writes code, and that cuts both ways. Adversaries use **AI-enhanced coding tools** 人工智能辅助编程工具 to write new **malware** faster than they could by hand, to modify existing application code so that it performs malicious activity, and to scan a codebase for **vulnerabilities** 漏洞 to attack. The skill barrier falls: someone who

could not previously write an exploit can now ask for one, so the number of capable attackers rises even when no new technique is invented.

Leveraging AI in Cyber Defense

The same technology defends us. AI tools can analyse an application's own source code, identify **vulnerabilities** in it and **recommend mitigations**; they can also review firewall rules and access settings and **recommend** safer options - though a human expert must always check the advice before applying it. AI can scan application code for weaknesses and suggest detection rules.

△ **A recommendation is not a fix.** The CED is explicit that the advice must be reviewed and implemented by a **knowledgeable programmer**: an AI tool can be confidently wrong about whether a flaw is exploitable, and applying a suggested patch without understanding it can introduce a new fault of its own.

Its biggest advantage is **scale**. A medium network produces millions of events every day - far too many for people to read. AI can quickly sort the harmless events from the likely-malicious ones, **alert** human staff, or take an automatic action. This lets defenders catch an attack and respond in seconds instead of days, preventing loss and damage.

That scale is what makes **threat detection and response** 威胁检测与响应 possible in practice: an AI system flags malicious activity as it happens, so the response team can **intervene quickly** enough to prevent loss, harm, or destruction of digital infrastructure — rather than reading the logs days later and finding out what was taken.

Exam tips

- When a question asks you to rank risks, remember **high risk = high impact AND easy to exploit**. A parking-lot Wi-Fi leak matters less than an **open internal port that lets an adversary spoof a device**.
- Learn the social-engineering tactics by name - **intimidation, urgency, pretexting, authority, consensus, scarcity, familiarity** - and be ready to spot which one an email is using.
- **Encryption still protects you on an evil twin**: the adversary sees your traffic but cannot read HTTPS. Say *what* is exposed, not just "it's unsafe".
- For "how to make authentication stronger", **MFA** is almost always part of the answer, plus long/unique passwords from a manager.
- AI is **dual-use**: the same tool (LLMs, code analysis) appears on both the attack and the defense side. Read the question carefully to see which side it asks about.