Security and Safety Fundamentals

Dr Ayman El Hajjar

CyBOK © Crown Copyright, The National Cyber Security Centre 2026, licensed under the Open Government Licence

http://www.nationalarchives.gov.uk/doc/open-government-licence/

CyBOK Mapping

The lecture maps to the following CyBOK Knowledge Areas:

Full Coverage:

- **Introductory Concepts → Introduction to CyBOK**
- **Infrastructure Security → Cyber Physical Systems**

Partial Coverage:

- **Human, Organisational & Regulatory Aspects → Risk Management & Governance**
- **Attacks & Defences → Adversarial Behaviours**

Outline

What is Security?

What is Safety

The Traditional Separation

The Convergence Problem

Security-Informed Safety

What is Security?

What is Security?

Security

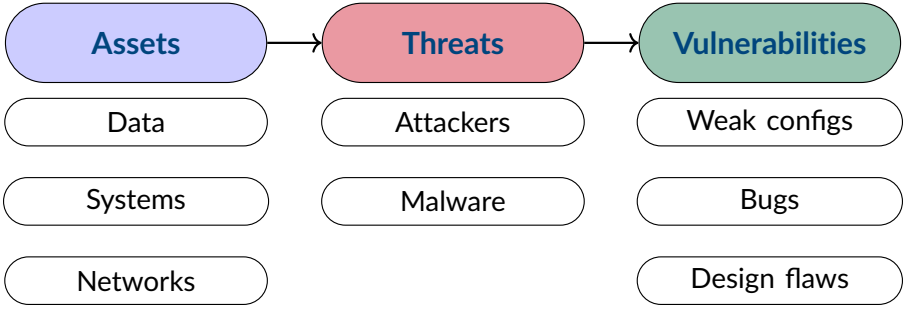
Security protects **information systems**, the **data** on them, and the **services** they provide from unauthorised access, harm, or misuse.

- It focuses on preventing unauthorised access, harm, misuse, disruption or compromise.
- It is commonly expressed through **confidentiality**, **integrity**, and **availability**.
- In connected systems, security failures can also affect physical or operational behaviour.

What is Security?

- **Security principles:**
  - **Confidentiality**: information is not disclosed to unauthorised parties
  - **Integrity**: data and systems are not tampered with or modified
  - **Availability**: systems and services remain accessible and functional
- **What security protects:**
  - Data, including personal, financial and intellectual property data
  - Systems and infrastructure
  - Networks and communications
- **What security protects against:**
  - Attackers and adversaries with malicious intent
  - Unauthorised users
  - Software vulnerabilities, exploits and misconfigurations

Security: Assets, Threats, Vulnerabilities



The Security Problem: Threats exploit vulnerabilities to compromise assets.

Security: A Practical Example

Example

A database with weak credentials is accessed by an attacker to steal customer data.

- **Vulnerability**: default or weak database credentials left unchanged after deployment
- **Threat**: attacker scans for exposed databases and attempts common credentials
- **Compromise**: attacker gains read access to customer records
- **Impact**: data breach, regulatory penalties, reputational damage and financial loss

Breach Path: Remove the vulnerability, reduce the threat opportunity, or protect the asset, and the breach path changes.

What is Safety

What is Safety?

Safety

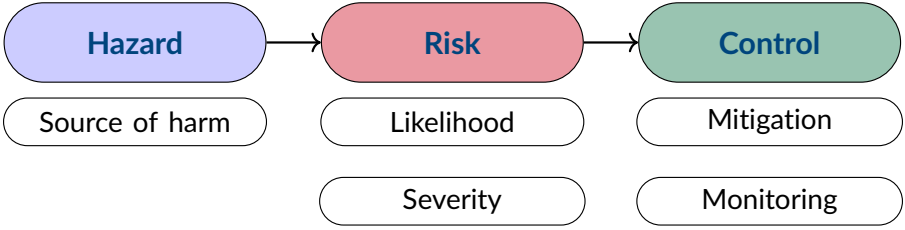
Safety is the state of being protected from danger, injury, or loss. In engineering, it means systems operate without causing **harm** to people, the environment, or critical operations.

- Safety focuses on preventing **hazards** from leading to unacceptable harm.
- It considers failures, misuse, environmental conditions and operational constraints.
- In cyber-physical systems, unsafe outcomes may result from both accidental faults and security failures.

What is Safety?

- **Safety principles:**
  - **Hazard control**: identifying and mitigating things that could cause harm
  - **Risk management**: assessing likelihood and impact of adverse events
  - **Assurance**: building confidence that systems are safe
- **What safety protects:**
  - Human health and life
  - Environmental integrity
  - Critical infrastructure function
- **What safety protects against:**
  - Equipment failures
  - Environmental conditions
  - Human error
  - **Malicious interference** is increasingly important

Safety: Hazards, Risks, Controls



The Safety Problem: Hazards create risks that must be managed and controlled.

Example

A medical ventilator failure can create a patient safety hazard. Controls may include alarms, backup power, maintenance checks and clinical escalation procedures.

Security and Safety as Separate Domains

Security and Safety

Why are **security** and **safety** traditionally treated as separate domains?

- Security has traditionally focused on malicious actors, data, systems and services.
- Safety has traditionally focused on hazards, failures, accidents and physical harm.
- Connected and cyber-physical systems make this separation harder to maintain.

The Traditional Separation

The Traditional Separation

- **Different contexts:** safety emerged in industrial and mechanical systems; security in IT and military domains
- **Different threat models:** safety assumes **accidental** failures; security assumes **intentional** adversaries
- **Different engineering approaches:** safety relies on redundancy and fail-safe design; security on access control, encryption, and threat modelling
- **Different regulatory frameworks:** safety: ISO 26262, IEC 61508, HSWA 1974; security: GDPR, Data Protection Act 2018, NIS Regulations

Security and Safety: The Traditional View

Primary Concern	Security	Safety
Threat Model	Malicious attackers	Failures and accidents
Focus	Systems, data, services	People and environment
Risk Approach	Vulnerability-based	Hazard-based
Typical Domain	IT, networks, software	Safety-critical systems
Regulatory View	Data protection	Safety standards

Why the Separation No Longer Works

The Traditional Separation

Why does the traditional separation between **security** and **safety** no longer work?

- A hospital insulin infusion pump is compromised and instructed to deliver an unsafe dose.
- This is a **security breach**: unauthorised access changes device behaviour.
- It is also a **safety failure**: the changed behaviour can cause patient harm.

Security-Safety Link: A security breach can now create a physical safety consequence.

The Convergence Problem

The Convergence Problem

- **Connected systems**
  - Industrial systems are now **networked** through IoT and cloud services
  - Medical devices communicate wirelessly
  - Autonomous vehicles rely on connected sensors
- **Security breaches create safety failures**
  - Compromised sensors $\Rightarrow$ incorrect control decisions
  - Malware in medical devices $\Rightarrow$ treatment errors
  - Attacks on power grids $\Rightarrow$ cascading failures
- **Malicious attackers are now a safety threat**
  - Traditional safety assumes **accidental** failures
  - Adversaries can **deliberately** cause failures
  - Safety standards do not adequately address this

Real-World Examples of Security-Safety Convergence

- **Medical Devices and Wearables**
  - Pacemakers, insulin pumps, glucose sensors, heart monitors
  - **Security flaw**: remote reprogramming, data interception, device spoofing
  - **Safety impact**: false readings, incorrect medical decisions
- **Autonomous Vehicles**
  - Rely on GPS, sensors, and V2X (vehicle-to-everything) communication
  - **Security flaw**: GPS spoofing or sensor manipulation
  - **Safety impact**: vehicle steers into oncoming traffic, collision, fatality
- **Industrial Control Systems (ICS) / SCADA**
  - Monitor and control critical infrastructure (e.g. power grids)
  - **Security flaw**: weak authentication, unencrypted protocols
  - **Safety impact**: blackouts, industrial accidents

Security-Informed Safety

Security-Informed Safety: A New Approach

Security-Safety Principle

"If It's Not Secure, It's Not Safe"

Bloomfield, Netkachova & Stroud, 2013

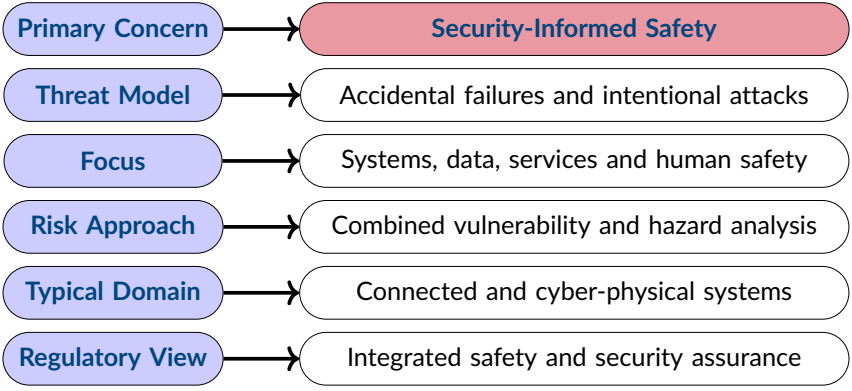
Security-Informed Safety: A New Approach

Security-Informed Safety

Security-informed safety integrates **security thinking** into traditional **safety engineering**.

- **Core idea:**
  - A system cannot be safe if it is not secure
  - Security breaches are a new class of **hazards** that safety engineering must address
  - Both accidental failures and intentional attacks must be considered
- **Practical implication:**
  - Safety engineers must understand cyber security threats
  - Security engineers must understand safety implications of breaches
  - Teams must work together on risk assessment and mitigation

Security-Informed Safety: The Integrated View



Integrated View

Security-informed safety considers how **security failures** can change **system behaviour** and create safety-relevant consequences.

Real security-safety convergence incidents

Stuxnet (2010)

- Targeted Iran's Natanz nuclear enrichment facility
- Compromised PLCs to manipulate centrifuge speeds
- Reported normal operation while causing damage
- **Impact**: centrifuges destroyed, programme delayed

Ukraine Power Grid Attacks (2015-2016)

- Remote access to SCADA systems in power distribution
- 2015: manual malicious commands; 2016: automated via Industroyer malware
- **Impact**: blackouts affecting hundreds of thousands

Summary

- **Security** = protecting assets from unauthorised access, harm, or misuse
- **Safety** = protecting people from harm caused by system failures
- **Traditional separation** made sense when systems were isolated
- **Connected systems** blur the boundary between security and safety
- **Security breaches are now safety hazards** in critical infrastructure
- **Neither discipline is sufficient alone**: security without safety ignores physical impact, safety without security ignores intentional threats
- **Security-Informed Safety** integrates both to engineer trustworthy systems

Key References

- Bloomfield, R., Netkachova, K., & Stroud, R. (2013). Security-Informed Safety in Cyber-Physical Systems. SAFECOMP **[Link]**
- Cardenas, A., Amin, S., & Sastry, S. (2008). Secure control: Towards survivable cyber-physical systems. **[Link]**
- Langner, R. (2013). To Kill a Centrifuge: A Technical Analysis of What Stuxnet's Creators Tried to Achieve. **[Link]**
- Case, D.U. (2016). Analysis of the cyber attack on the Ukrainian power grid. **[Link]**