Rise of the Robots

EXPLORING RISK & ASSURANCE CHALLENGES FOR RAS

08 JUNE 21

NIKITA JOHNSON

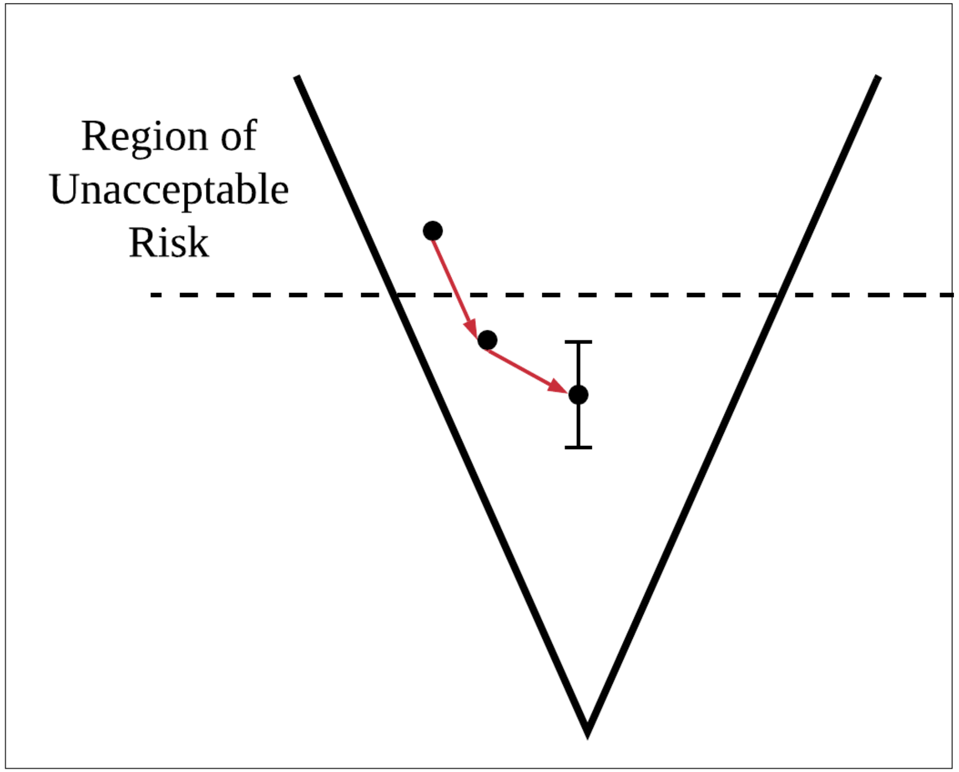
Agenda

There are fundamental differences between safety and security that have significant implications for co-assurance

- critically survey the current state-of-the-art techniques and standards
- technical and socio-technical challenges
- SSAF - a candidate solution
- discussion about ways forward

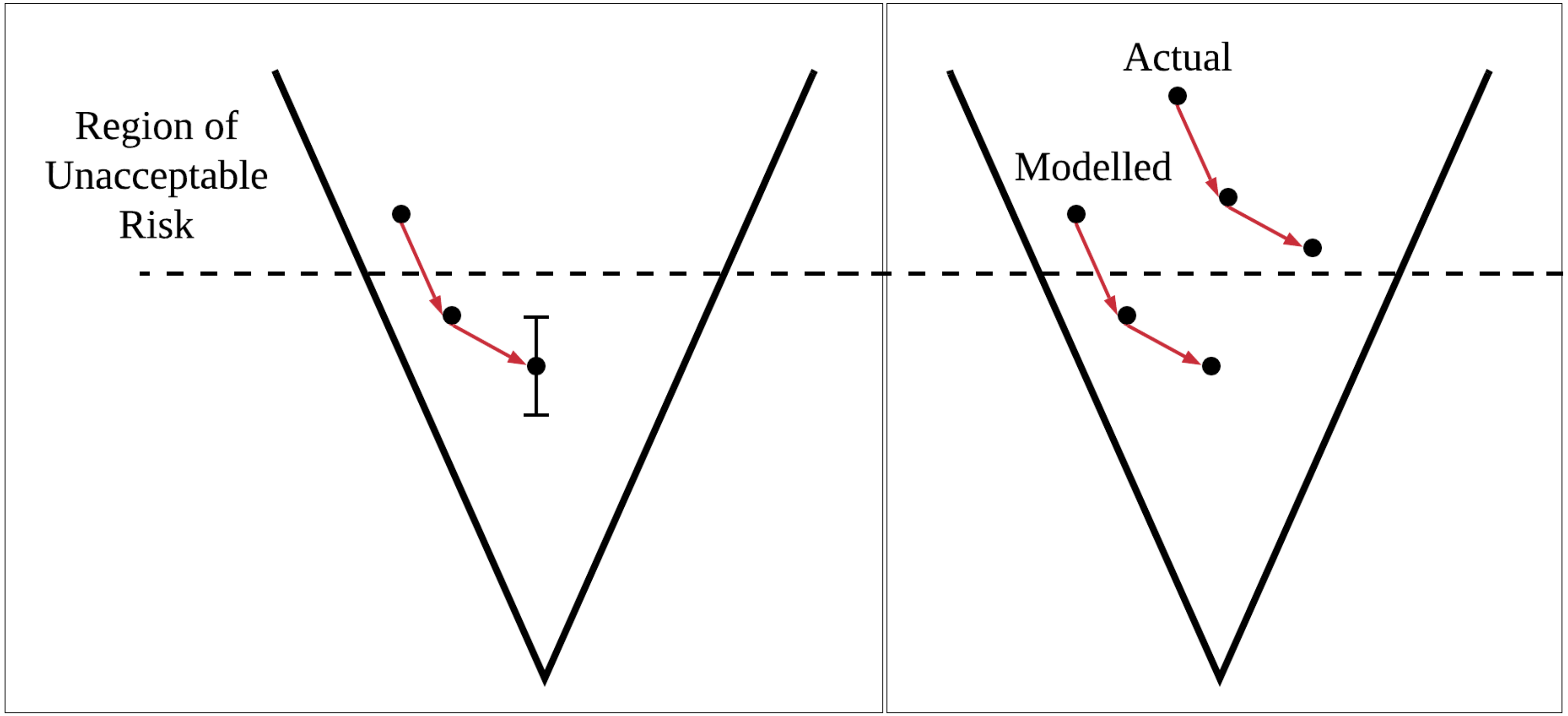
1. Risk Challenge

SAFETY-SECURITY CO-ASSURANCE



(a) Representation of Risk Reduction

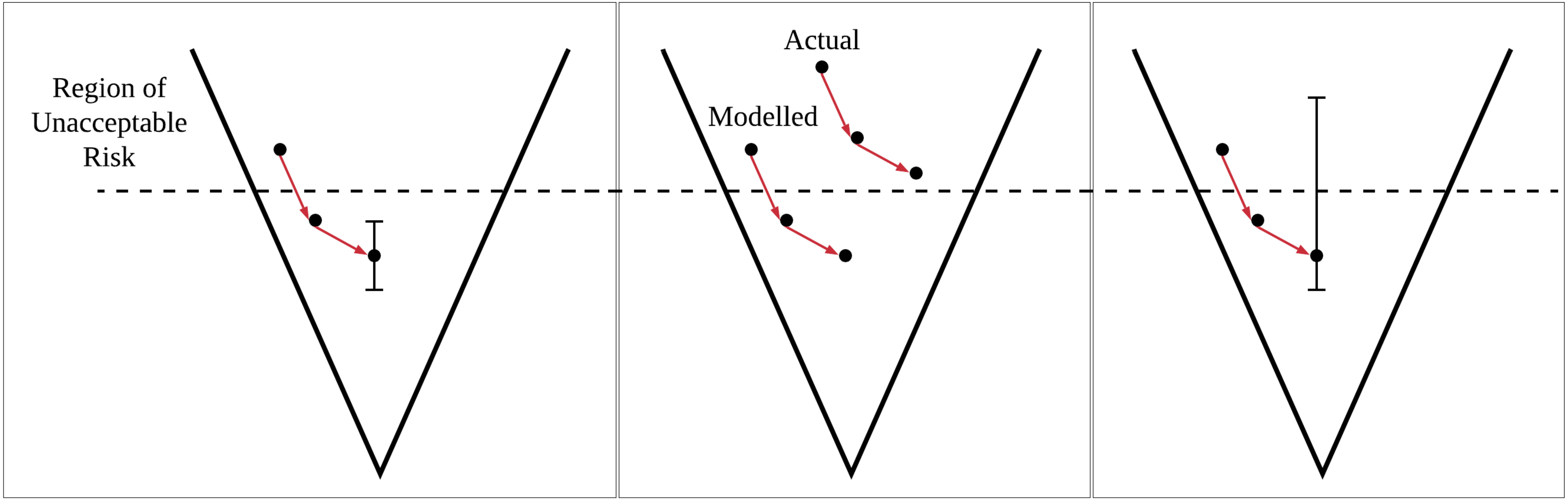
Risk Challenge



(a) Representation of Risk Reduction

(b) Problem 1: Incorrect Risk Estimation

Risk Challenge



(a) Representation of Risk Reduction (b) Problem 1: Incorrect Risk Estimation (c) Problem 2: Low Confidence

Risk Challenge

2. Existing Approaches

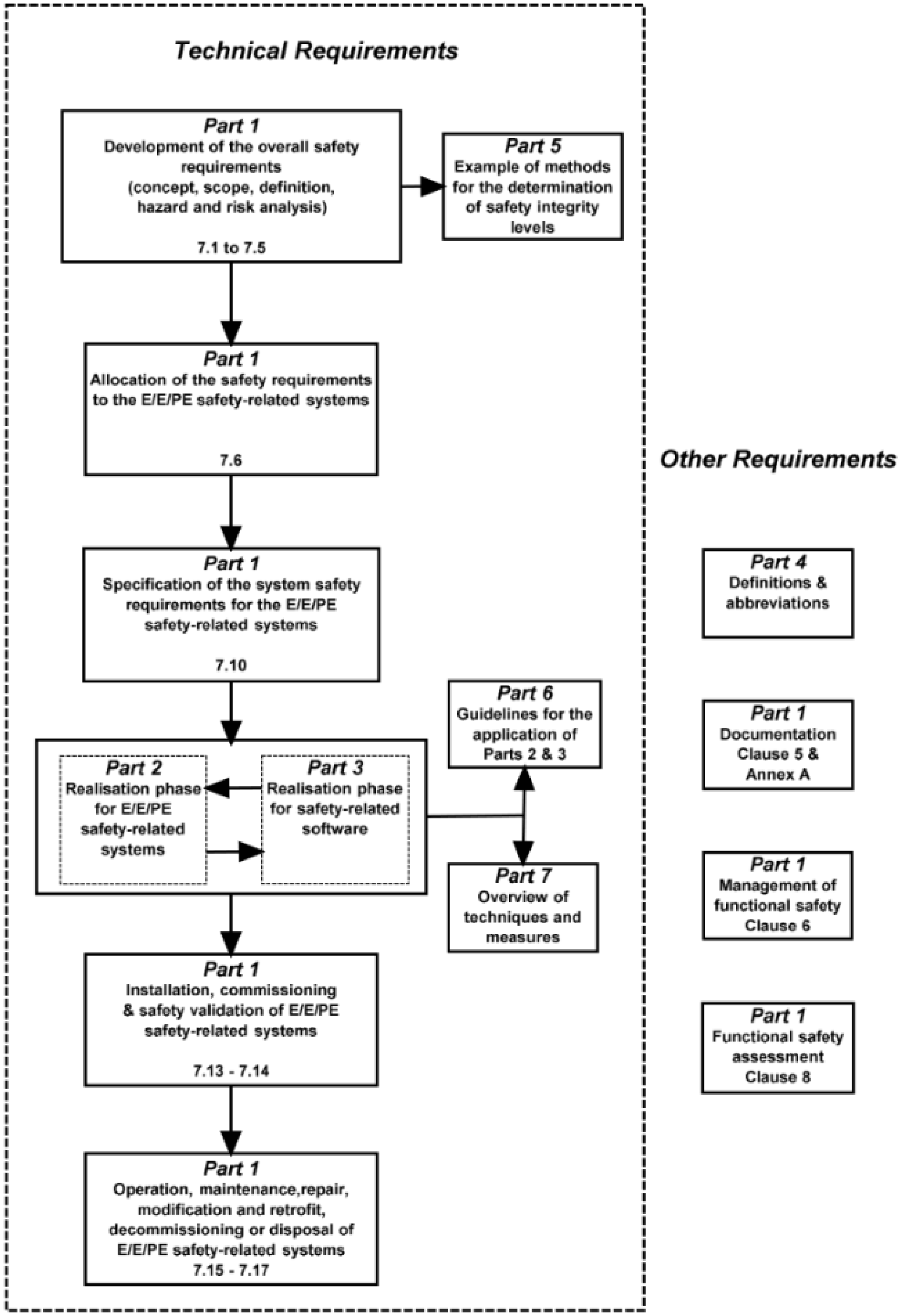
SAFETY-SECURITY CO-ASSURANCE

Safety-Security Standards

Domain	Safety	Security	Both
General		Common Criteria ISO 27K-Series NIST 800-Series NIST Framework NCSC Guidance	IET Code of Practice
Aerospace	ARP 4754A DO-178C	DO-326A	
Automotive	ISO 26262		PAS 11281
Defence	Def Stan 00-56	JSP 440	
Healthcare (Medical Devices)	ISO 14971 FDA Safety Guidance	AAMI TIR 57 FDA Security Guidance	
Industrial Control	IEC 61508	IEC 63443 HSE IACS NIST 800-82	IET TR 63069
Nuclear	ONR Safety Principles	ONR Security Principles	
Rail	CENELEC EN 51028	CENELECT TS 50701	CPNI Rail Guidance

IEC 61508

General Safety Standard



Technical Requirements

Part 1
Development of the overall safety requirements
(concept, scope, definition, hazard and risk analysis)
7.1 to 7.5

Part 5
Example of methods for the determination of safety integrity levels

Part 1
Allocation of the safety requirements to the E/E/PE safety-related systems
7.6

Part 1
Specification of the system safety requirements for the E/E/PE safety-related systems
7.10

Part 6
Guidelines for the application of Parts 2 & 3

Part 2
Realisation phase for E/E/PE safety-related systems

Part 3
Realisation phase for safety-related software

Part 7
Overview of techniques and measures

Part 1
Installation, commissioning & safety validation of E/E/PE safety-related systems
7.13 - 7.14

Part 1
Operation, maintenance,repair, modification and retrofit, decommissioning or disposal of E/E/PE safety-related systems
7.15 - 7.17

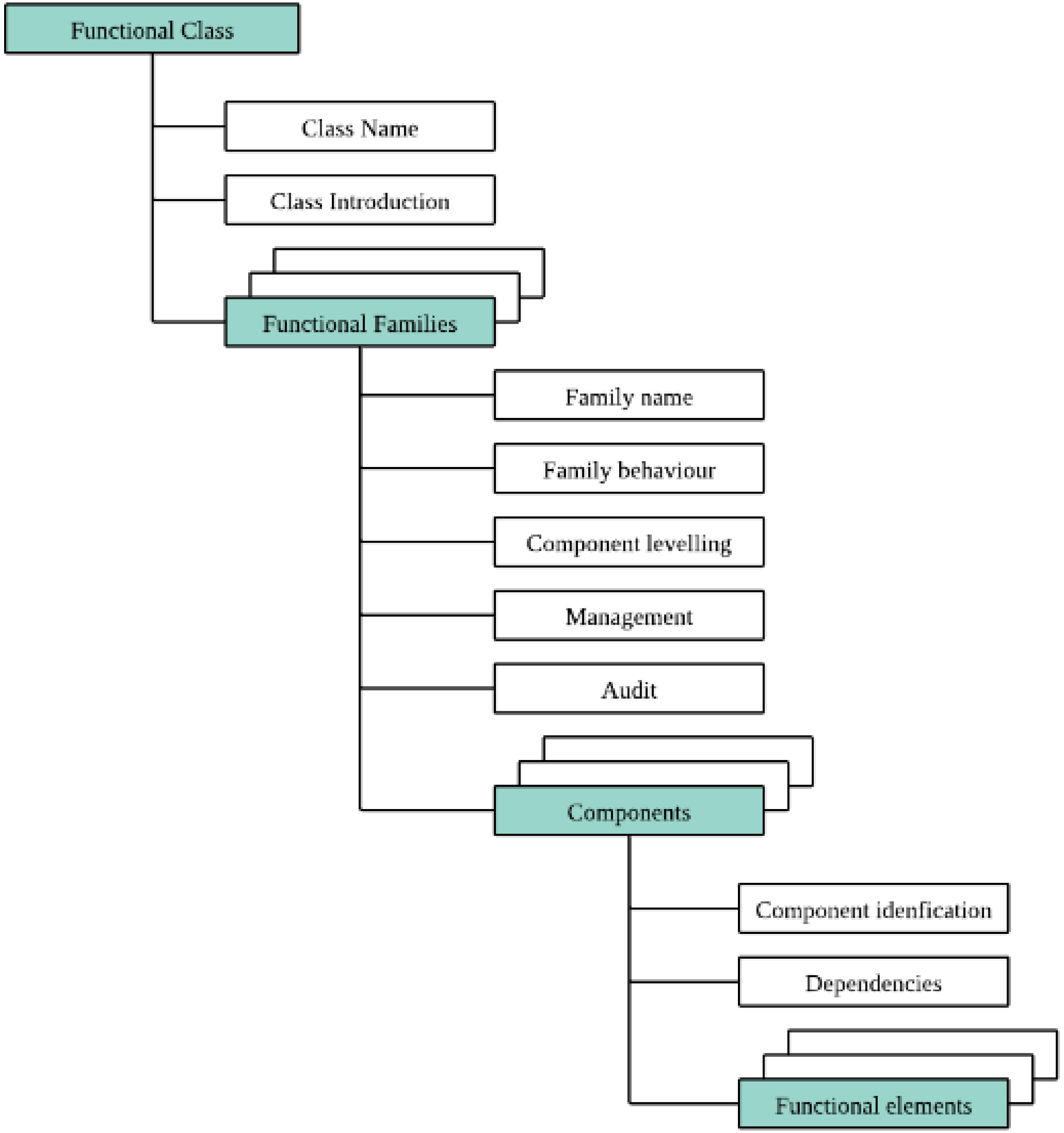
Other Requirements

Part 4
Definitions & abbreviations

Part 1
Documentation
Clause 5 & Annex A

Part 1
Management of functional safety
Clause 6

Part 1
Functional safety assessment
Clause 8

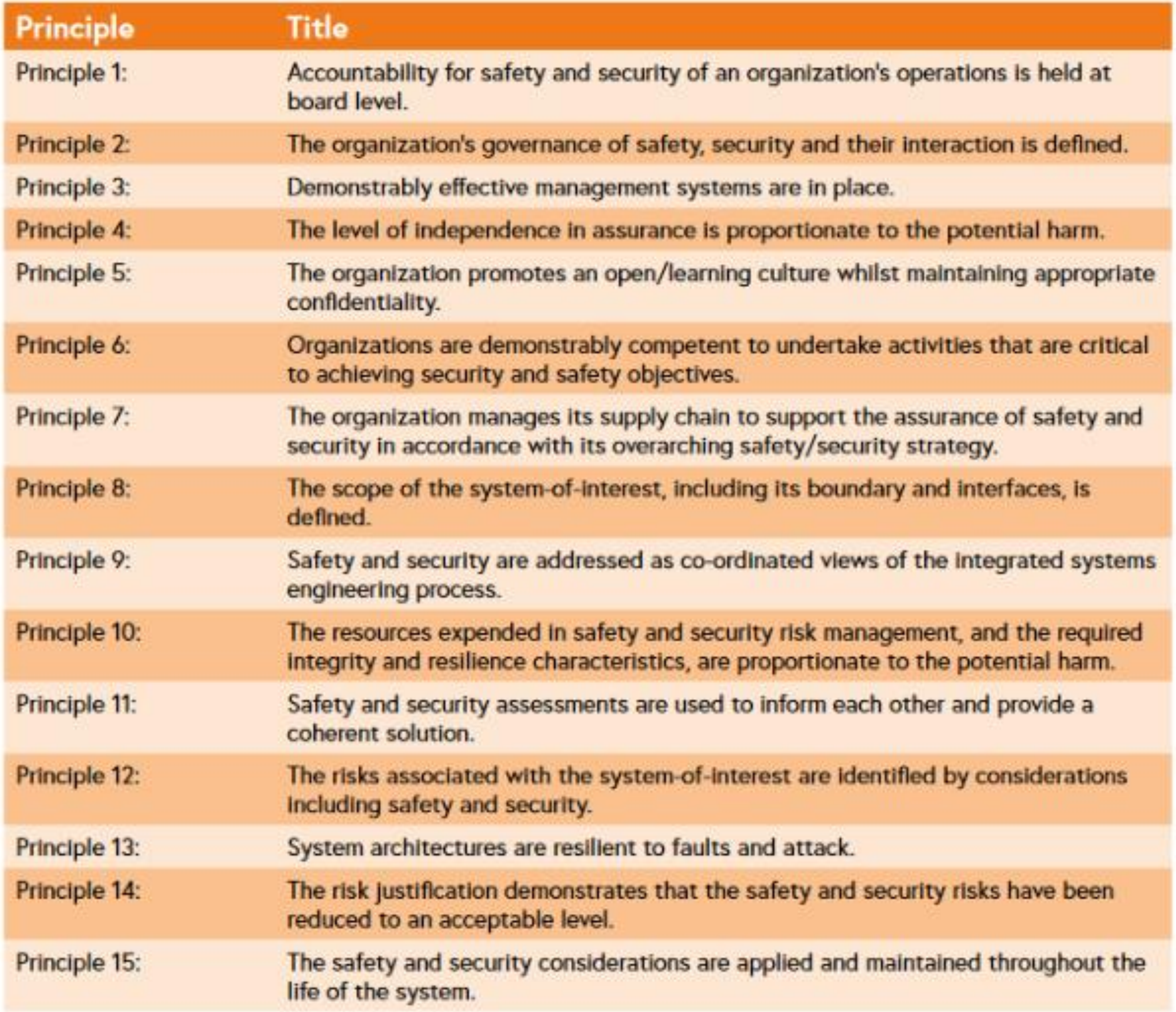


Common Criteria

ISO 15408

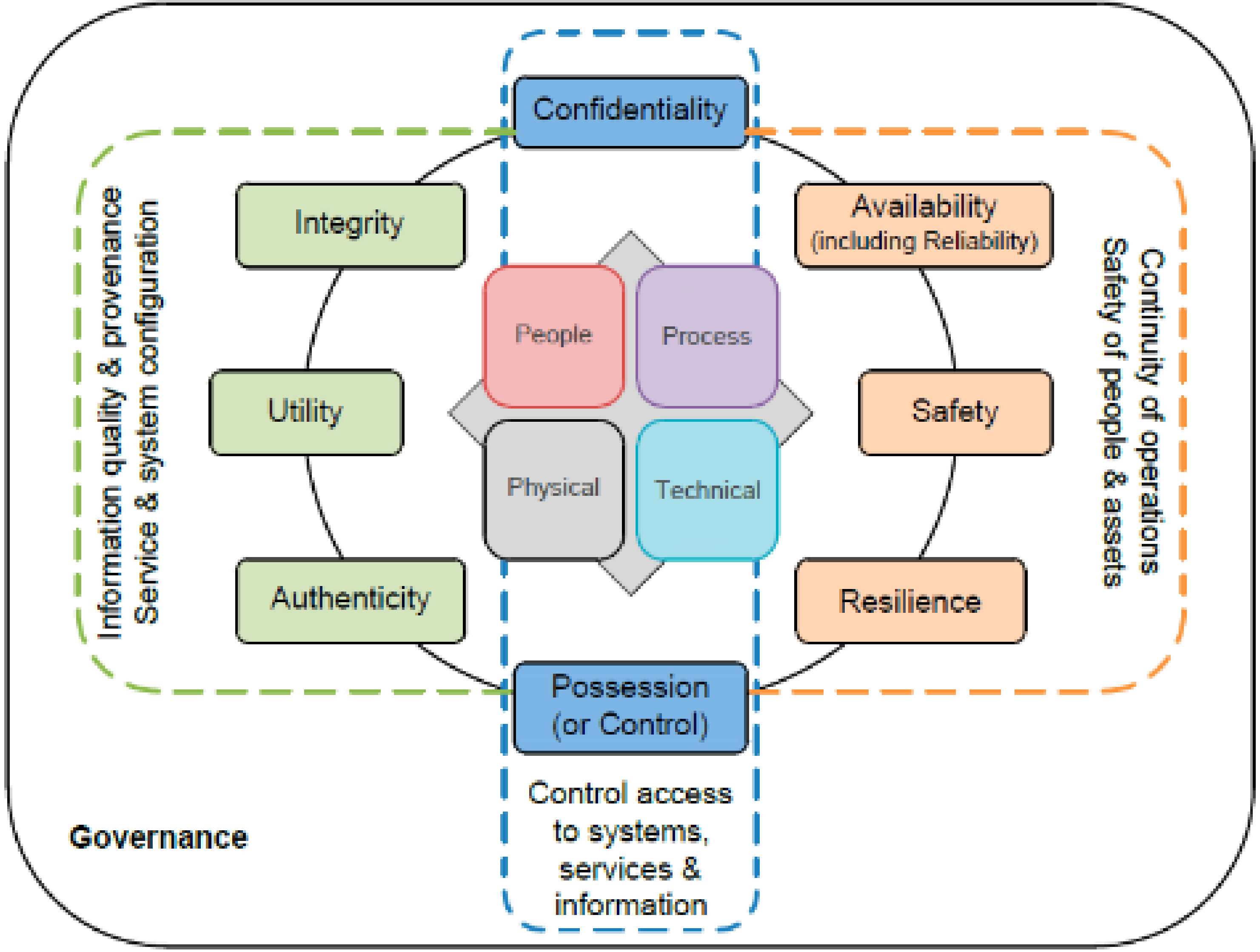
General Security Standard

IET Code of Practice – Cyber Security and Safety

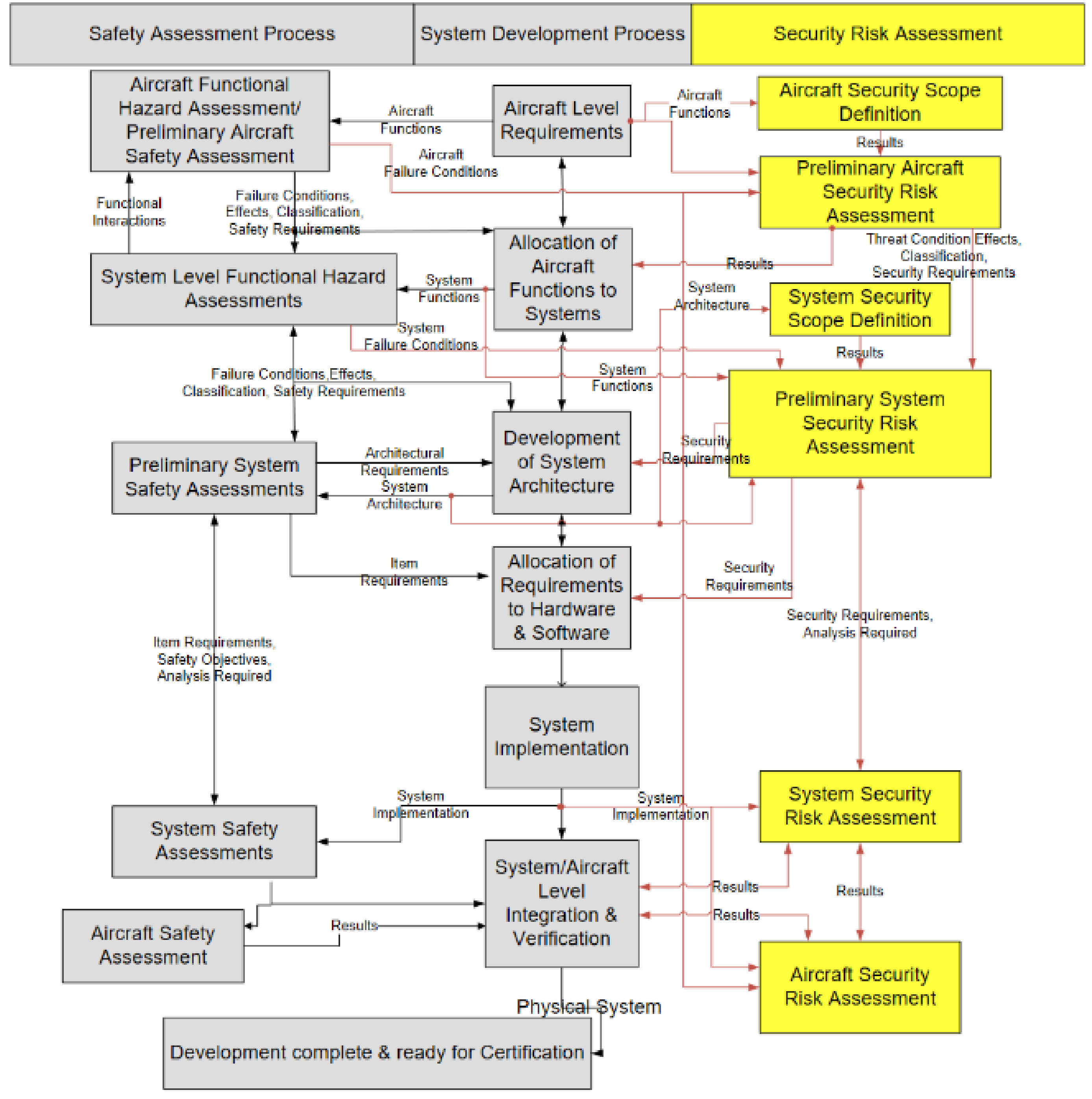


Principle	Title
Principle 1:	Accountability for safety and security of an organization's operations is held at board level.
Principle 2:	The organization's governance of safety, security and their interaction is defined.
Principle 3:	Demonstrably effective management systems are in place.
Principle 4:	The level of independence in assurance is proportionate to the potential harm.
Principle 5:	The organization promotes an open/learning culture whilst maintaining appropriate confidentiality.
Principle 6:	Organizations are demonstrably competent to undertake activities that are critical to achieving security and safety objectives.
Principle 7:	The organization manages its supply chain to support the assurance of safety and security in accordance with its overarching safety/security strategy.
Principle 8:	The scope of the system-of-interest, including its boundary and interfaces, is defined.
Principle 9:	Safety and security are addressed as co-ordinated views of the integrated systems engineering process.
Principle 10:	The resources expended in safety and security risk management, and the required integrity and resilience characteristics, are proportionate to the potential harm.
Principle 11:	Safety and security assessments are used to inform each other and provide a coherent solution.
Principle 12:	The risks associated with the system-of-interest are identified by considerations including safety and security.
Principle 13:	System architectures are resilient to faults and attack.
Principle 14:	The risk justification demonstrates that the safety and security risks have been reduced to an acceptable level.
Principle 15:	The safety and security considerations are applied and maintained throughout the life of the system.

PAS 1885 – Automotive Cyber Principles



Aerospace – DO-326A



Safety Assessment Process

System Development Process

Security Risk Assessment

Aircraft Functional Hazard Assessment/ Preliminary Aircraft Safety Assessment

Aircraft Functions

Aircraft Level Requirements

Aircraft Functions

Aircraft Security Scope Definition

Results

Aircraft Failure Conditions

Preliminary Aircraft Security Risk Assessment

Functional Interactions

Failure Conditions, Effects, Classification, Safety Requirements

Allocation of Aircraft Functions to Systems

Results

Threat Condition Effects, Classification, Security Requirements

System Level Functional Hazard Assessments

System Functions

System Architecture

System Security Scope Definition

System Failure Conditions

Results

Failure Conditions,Effects, Classification, Safety Requirements

System Functions

Preliminary System Security Risk Assessment

Development of System Architecture

Security Requirements

Preliminary System Safety Assessments

Architectural Requirements

System Architecture

Item Requirements

Allocation of Requirements to Hardware & Software

Security Requirements

Security Requirements, Analysis Required

Item Requirements, Safety Objectives, Analysis Required

System Implementation

System Implementation

System Implementation

System Security Risk Assessment

System Safety Assessments

System/Aircraft Level Integration & Verification

Results

Results

Results

Aircraft Safety Assessment

Results

Aircraft Security Risk Assessment

Physical System

Development complete & ready for Certification

Safety-Security Standards

Domain	Safety	Security	Both
General		Common Criteria ISO 27K-Series NIST 800-Series NIST Framework NCSC Guidance	IET Code of Practice
Aerospace	ARP 4754A DO-178C	DO-326A	
Automotive	ISO 26262		PAS 11281
Defence	Def Stan 00-56	JSP 440	
Healthcare (Medical Devices)	ISO 14971 FDA Safety Guidance	AAMI TIR 57 FDA Security Guidance	
Industrial Control	IEC 61508	IEC 63443 HSE IACS NIST 800-82	IET TR 63069
Nuclear	ONR Safety Principles	ONR Security Principles	
Rail	CENELEC EN 51028	CENELECT TS 50701	CPNI Rail Guidance

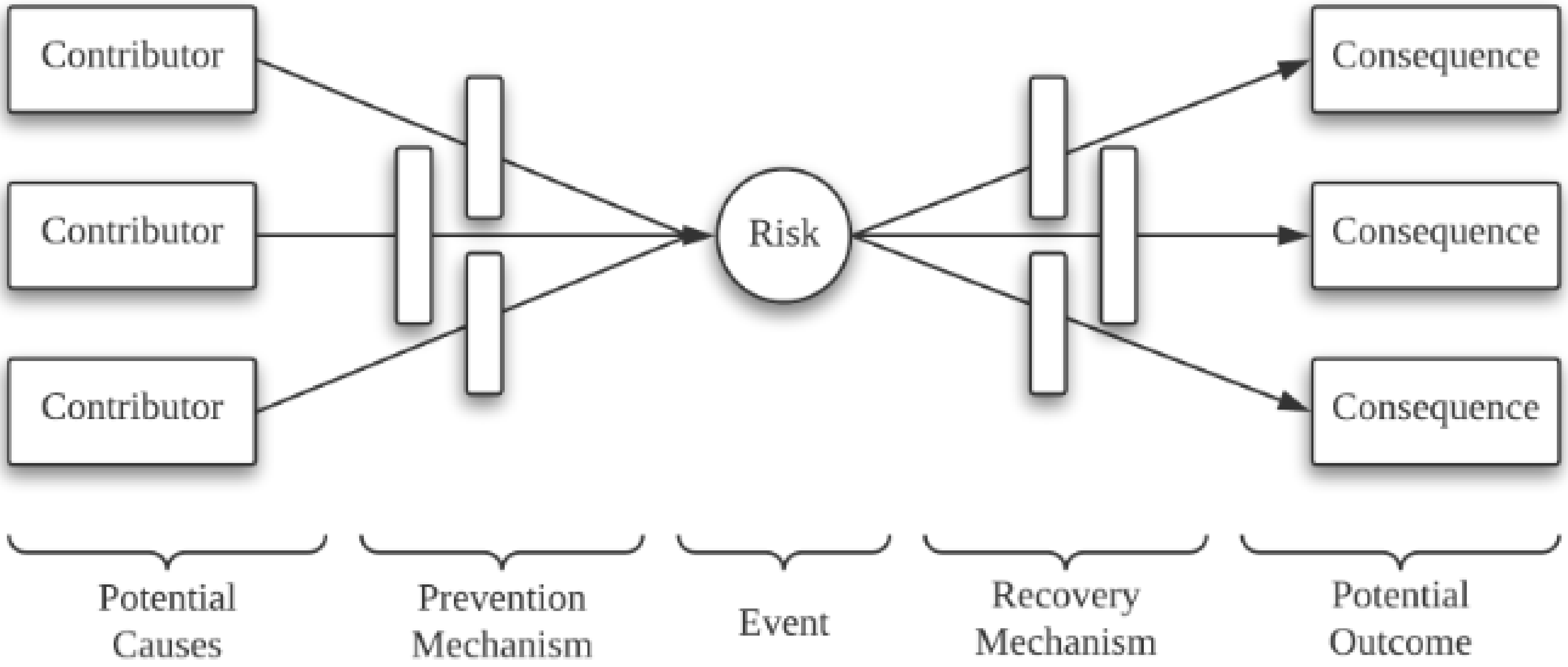
Safety-Security Approaches

1. Hazard Analysis
   - Security-Aware Bowtie
   - Security-Aware STPA: STPA-Sec and STPA-SafeSec
   - Security-Aware Guidewords: FMEVA, FMVEA
2. Mitigations and Control
   - Security-Integrated Fault Trees: Attack-Defence Trees
3. Architectural and System Analysis
   - Architecture Trade-off Analysis Method (ATAM)
   - Dependability Deviation Analysis (DDA)
4. Assurance
   - Static analysis and testing for security (*e.g.* category theory applied to cryptography)
   - Argument structures for security

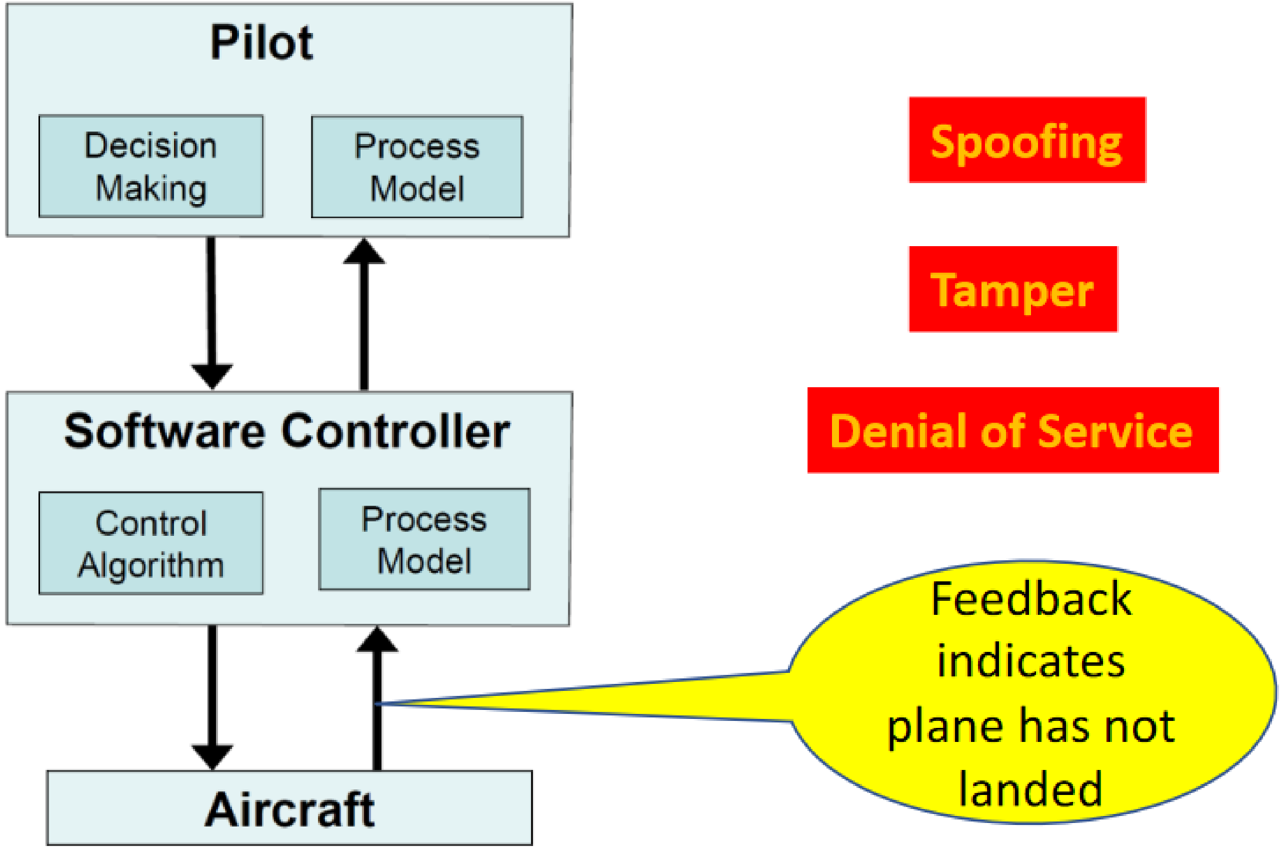
Safety-Security Approaches

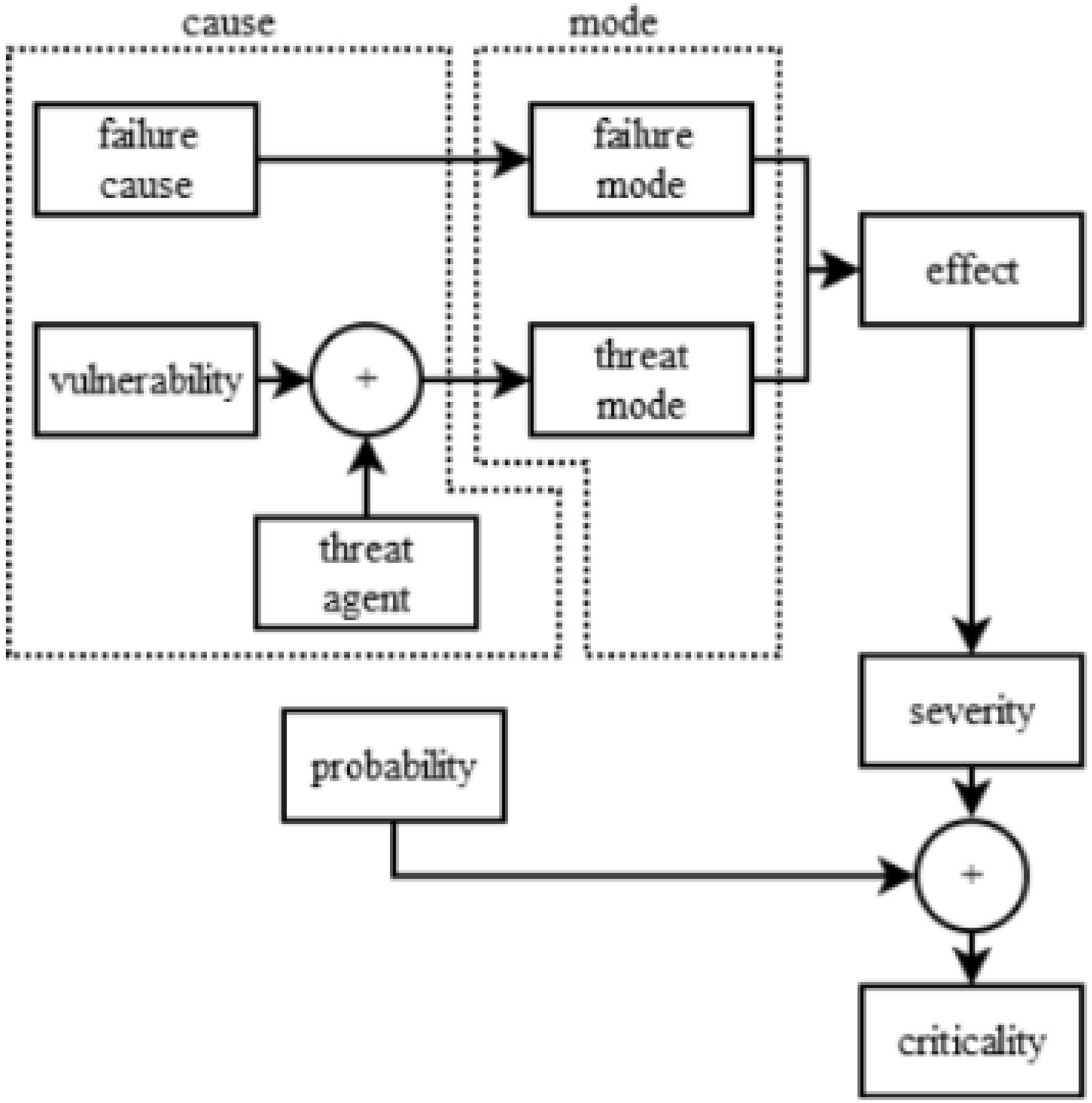
1. Hazard Analysis
   - Security-Aware Bowtie
   - Security-Aware STPA: STPA-Sec and STPA-SafeSec
   - Security-Aware Guidewords: FMEVA, FMVEA
2. Mitigations and Control
   - Security-Integrated Fault Trees: Attack-Defence Trees
3. Architectural and System Analysis
   - Architecture Trade-off Analysis Method (ATAM)
   - Dependability Deviation Analysis (DDA)
4. Assurance
   - Static analysis and testing for security (*e.g.* category theory applied to cryptography)
   - Argument structures for security

Bow-tie analysis

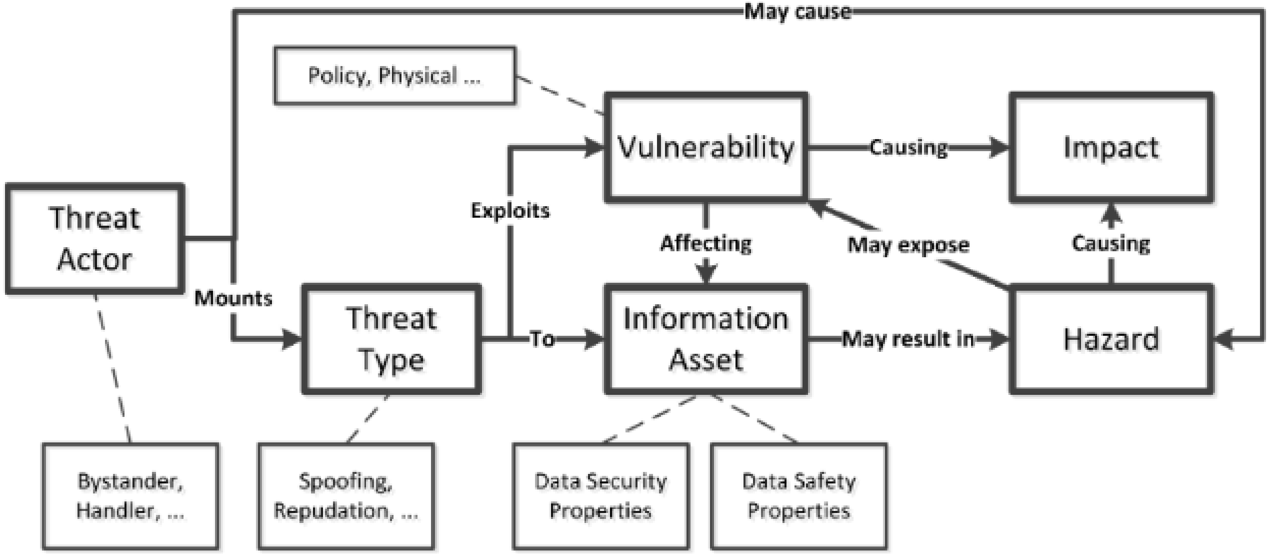


STPA-Sec


Pilot
Decision Making
Process Model
Software Controller
Control Algorithm
Process Model
Aircraft
Spoofing
Tamper
Denial of Service
Feedback indicates plane has not landed


cause
mode
failure cause
failure mode
effect
vulnerability
+
threat mode
threat agent
severity
probability
+
criticality

FMEVA



CRAF – Cyber Risk Assessment Framework (Guideword)

Safety-Security Approaches

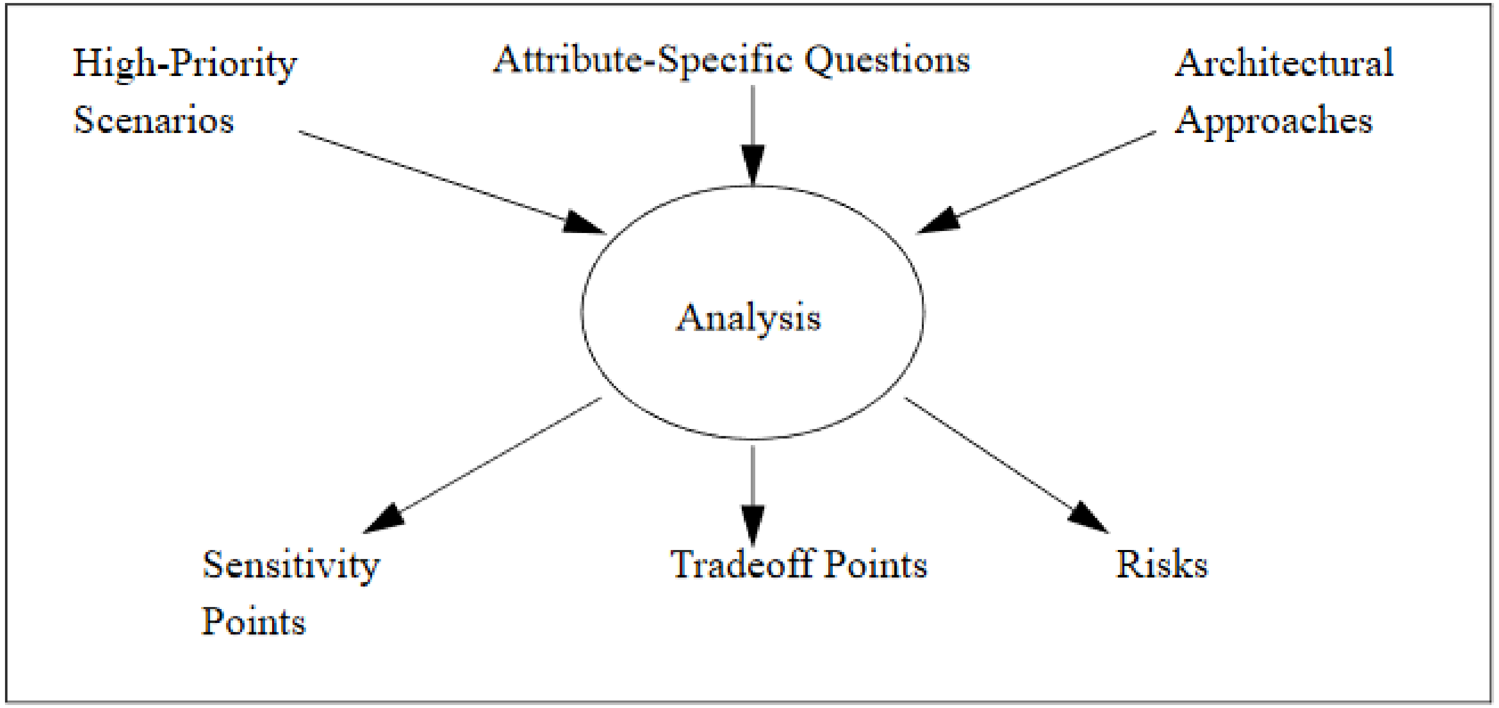
1. Hazard Analysis
   - Security-Aware Bowtie
   - Security-Aware STPA: STPA-Sec and STPA-SafeSec
   - Security-Aware Guidewords: FMEVA, FMVEA

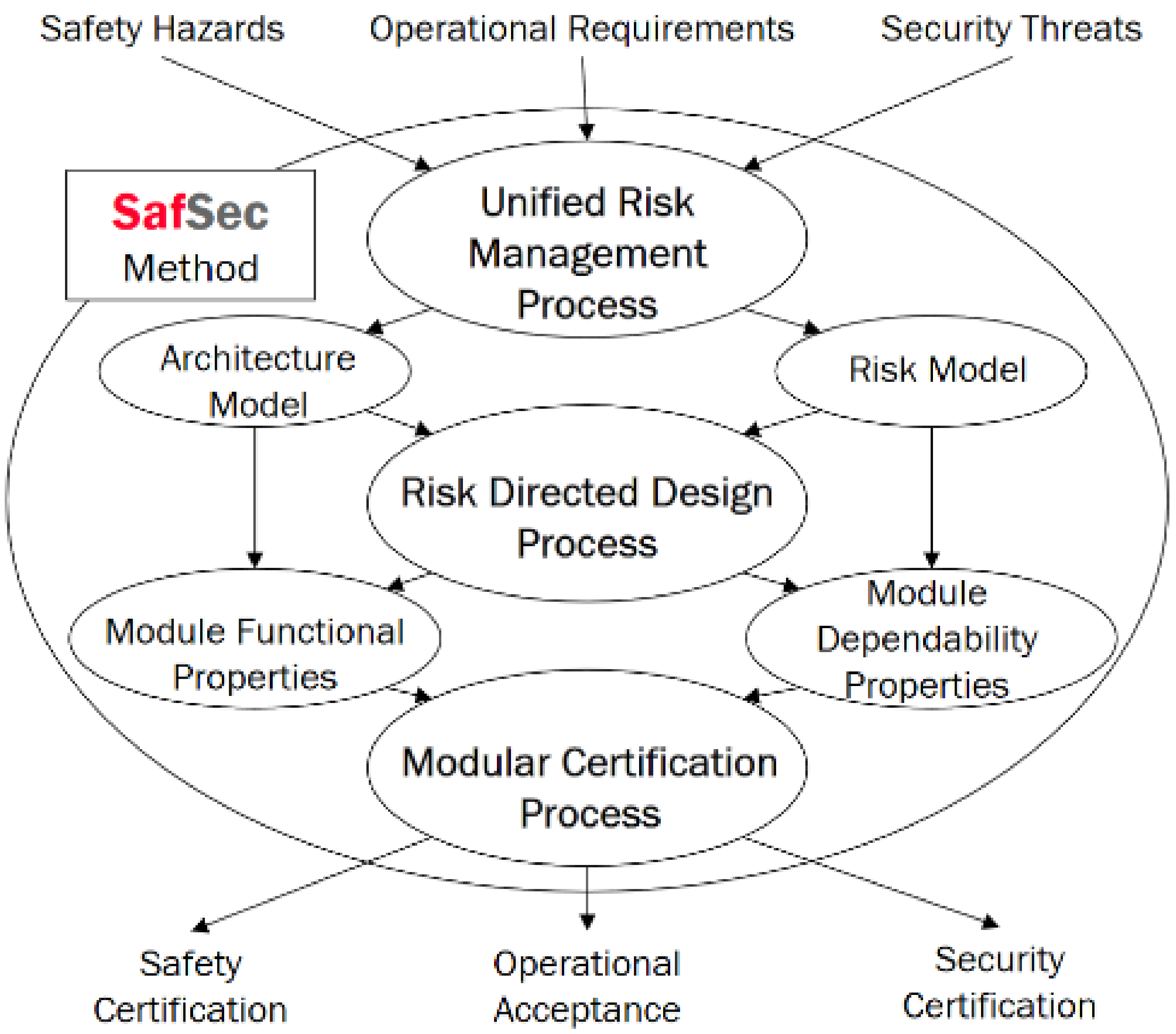
2. Mitigations and Control
   - Security-Integrated Fault Trees: Attack-Defence Trees

3. Architectural and System Analysis
   - Architecture Trade-off Analysis Method (ATAM)
   - Dependability Deviation Analysis (DDA)

4. Assurance
   - Static analysis and testing for security (*e.g.* category theory applied to cryptography)
   - Argument structures for security

ATAM – Architecture Trade-off Analysis Method





SafSec Method & DDA

However!
Uncertainties & Challenges Remain:

- Technical Uncertainties
  - Lack of unifying language leads to ambiguity in expression of models
  - Model complexity and interactions; timing and incomplete information
  - Intent of the attacker currently not well considered for systems and safety
  - How to incorporate different risk? Comparing apples and oranges
  - Model divergence and change over time
  - Completeness of the methodology

- Socio-Technical Uncertainties
  - Lack of unifying underly philosophy leads to misunderstandings and miscommunication
    - *e.g.* openness vs. security-through-obscurity
  - No standard practices means that integration varies between project or people
  - Differences in proportionality and resources
    - *e.g.* Industry shortage of Suitably Qualified and Experience People (SQEP) for security

3. Candidate Solution

THE SAFETY-SECURITY ASSURANCE FRAMEWORK

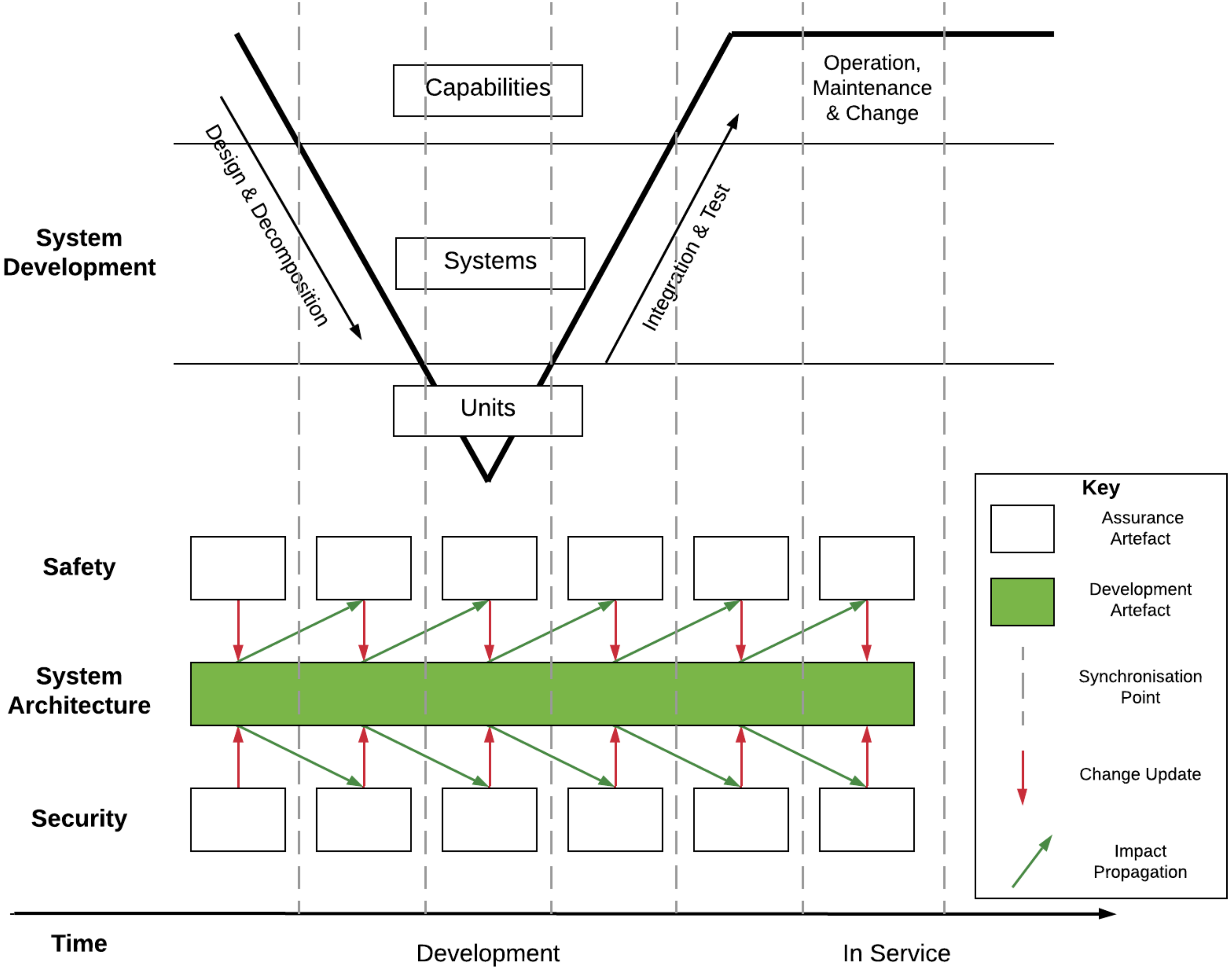
The Safety-Security Assurance Framework SSAF

- Independent Co-Assurance

- Synchronisation Points

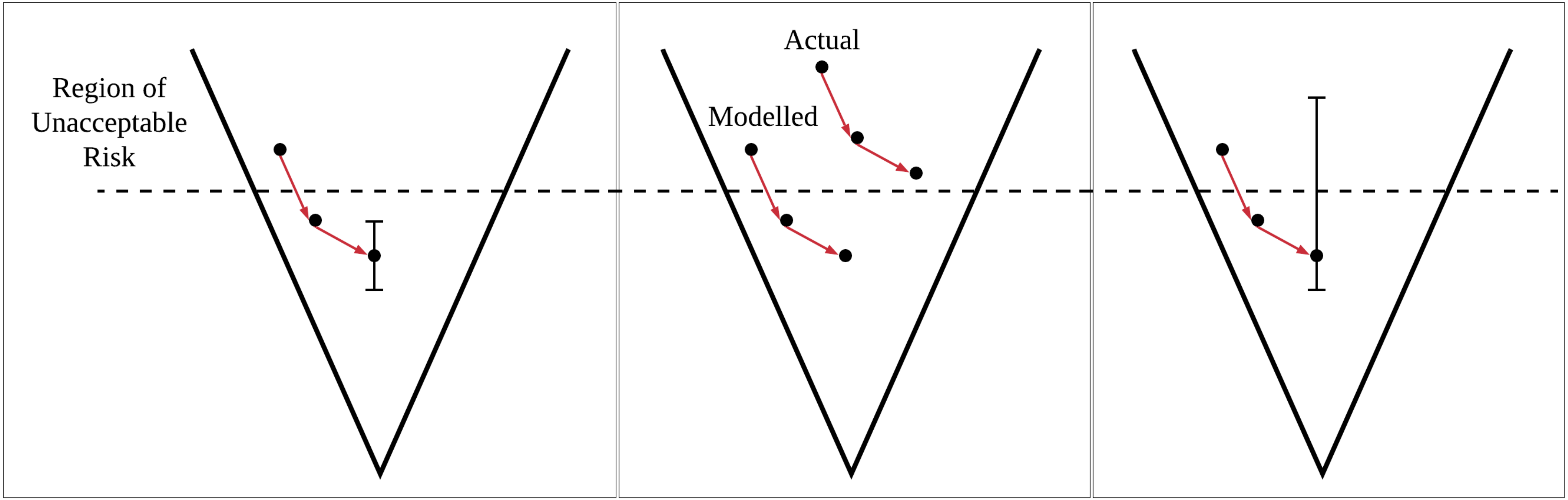
- Information Needs

- Trade-off



4. Causal Model & Patterns

SAFETY-SECURITY CO-ASSURANCE



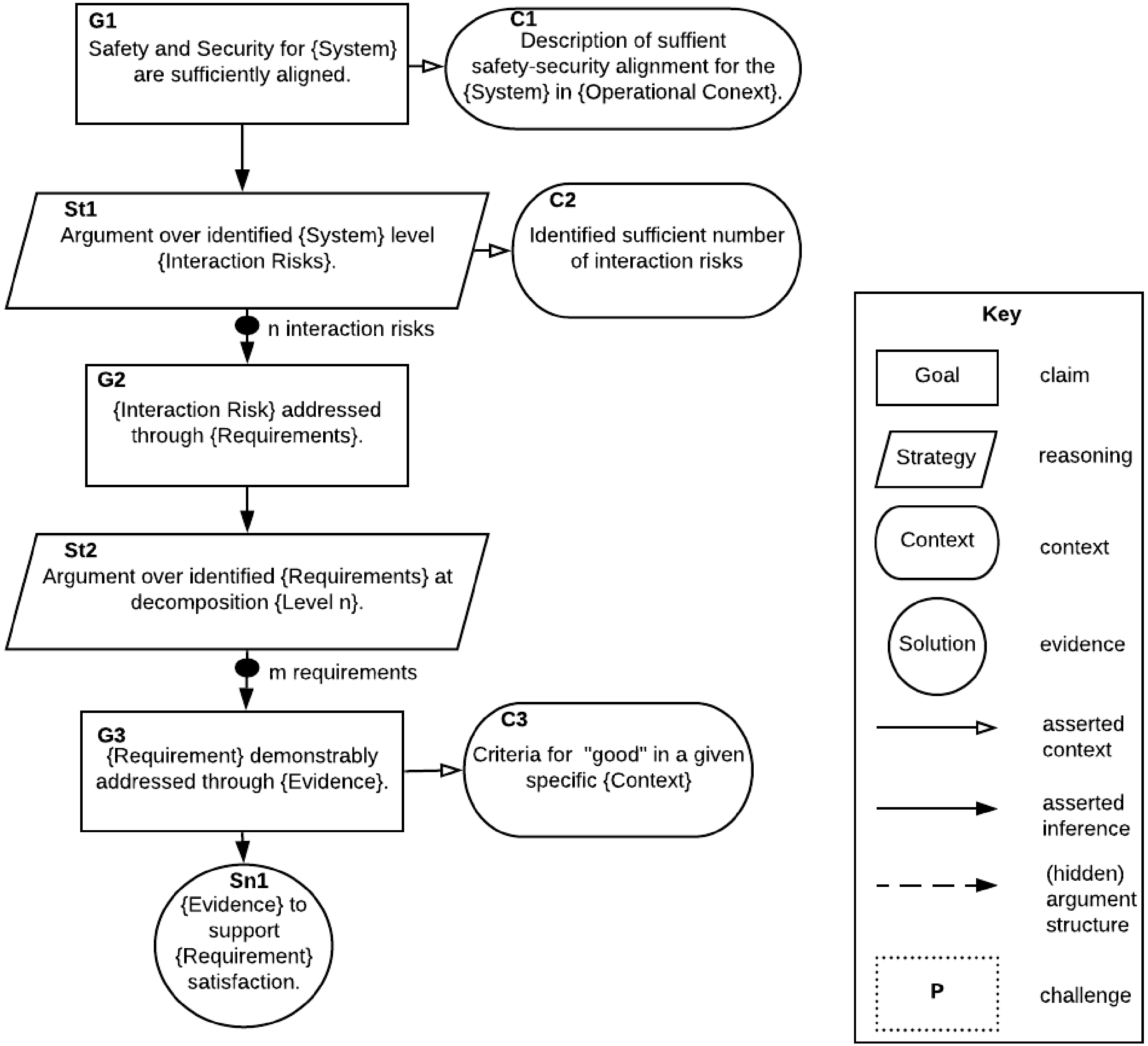
(a) Representation of Risk Reduction

(b) Problem 1: Incorrect Risk Estimation

(c) Problem 2: Low Confidence

How to Represent Risk Reduction?

Technical Risk Argument



G1
Safety and Security for {System} are sufficiently aligned.

C1
Description of suffient safety-security alignment for the {System} in {Operational Conext}.

St1
Argument over identified {System} level {Interaction Risks}.

C2
Identified sufficient number of interaction risks

n interaction risks

G2
{Interaction Risk} addressed through {Requirements}.

St2
Argument over identified {Requirements} at decomposition {Level n}.

m requirements

G3
{Requirement} demonstrably addressed through {Evidence}.

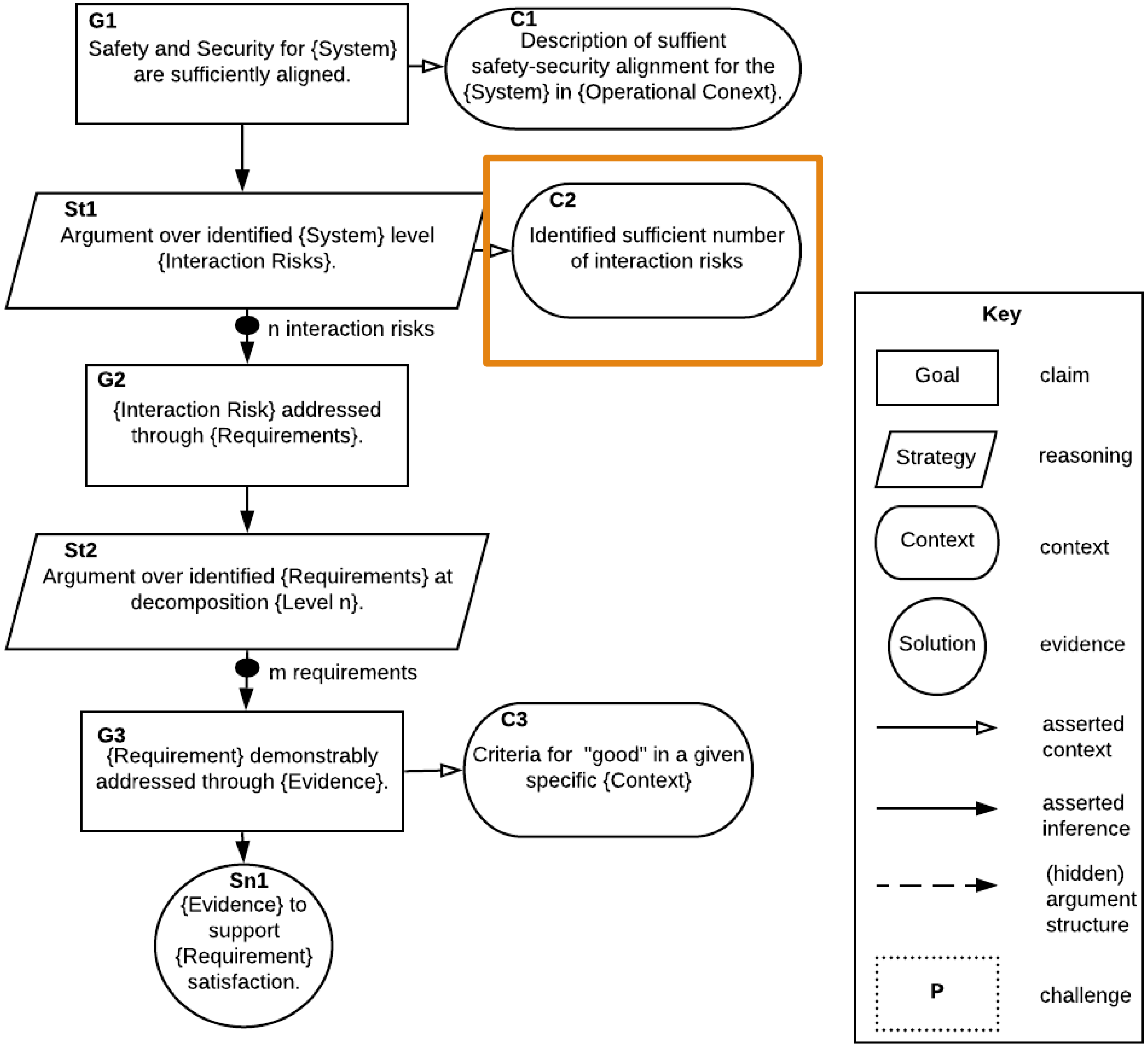
C3
Criteria for "good" in a given specific {Context}

Sn1
{Evidence} to support {Requirement} satisfaction.

Key

Goal	claim
Strategy	reasoning
Context	context
Solution	evidence
——▷	asserted context
——▶	asserted inference
– – –▶	(hidden) argument structure
P	challenge

Technical Risk Argument



G1
Safety and Security for {System} are sufficiently aligned.

C1
Description of suffient safety-security alignment for the {System} in {Operational Conext}.

St1
Argument over identified {System} level {Interaction Risks}.

C2
Identified sufficient number of interaction risks

n interaction risks

G2
{Interaction Risk} addressed through {Requirements}.

St2
Argument over identified {Requirements} at decomposition {Level n}.

m requirements

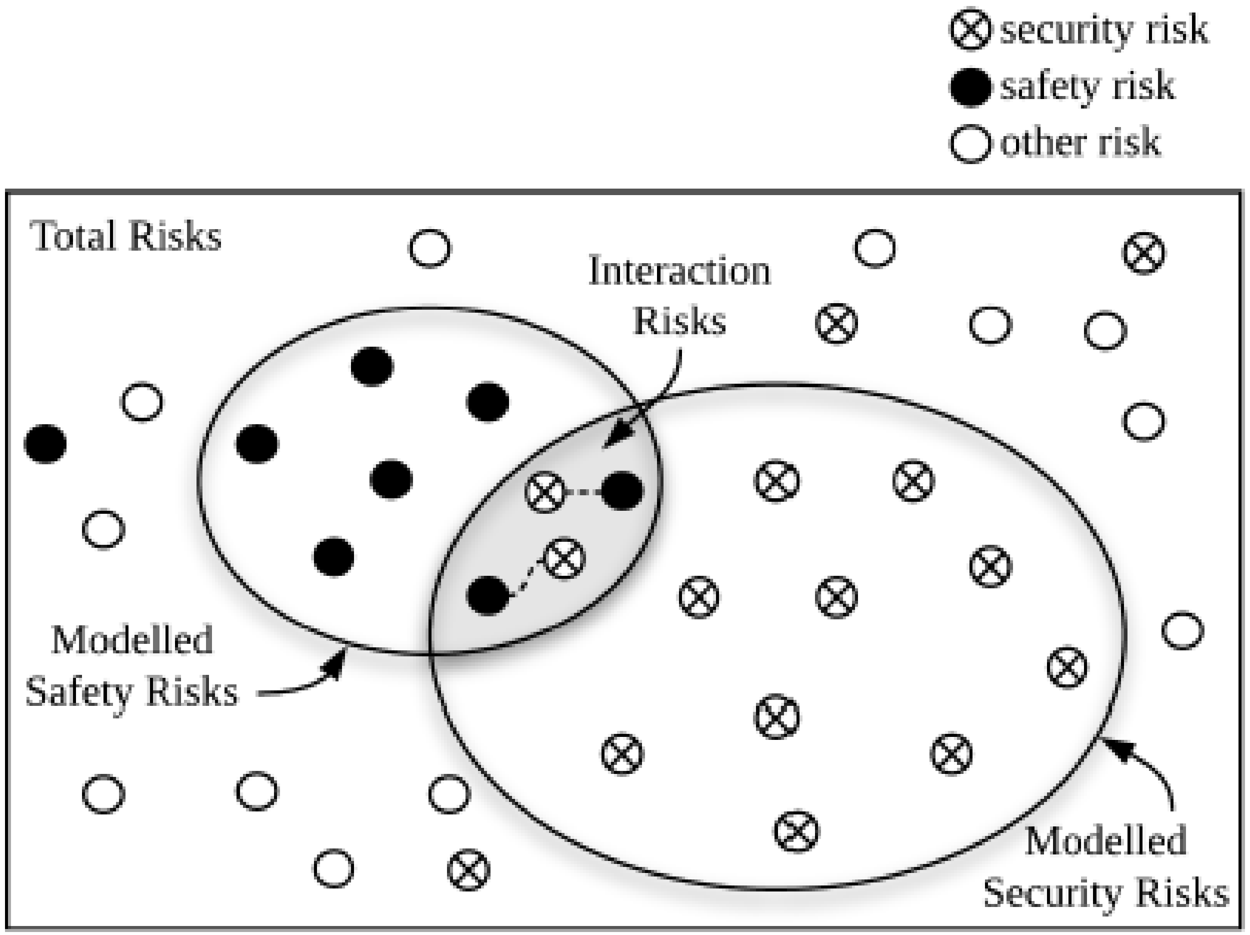
G3
{Requirement} demonstrably addressed through {Evidence}.

C3
Criteria for "good" in a given specific {Context}

Sn1
{Evidence} to support {Requirement} satisfaction.

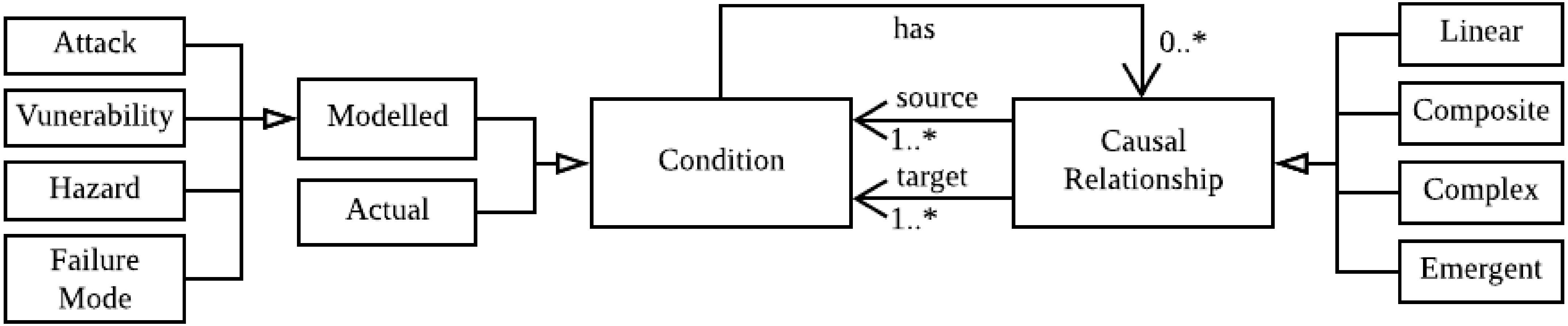
Key

Goal	claim
Strategy	reasoning
Context	context
Solution	evidence
——▷	asserted context
——▶	asserted inference
– – –▶	(hidden) argument structure
P	challenge

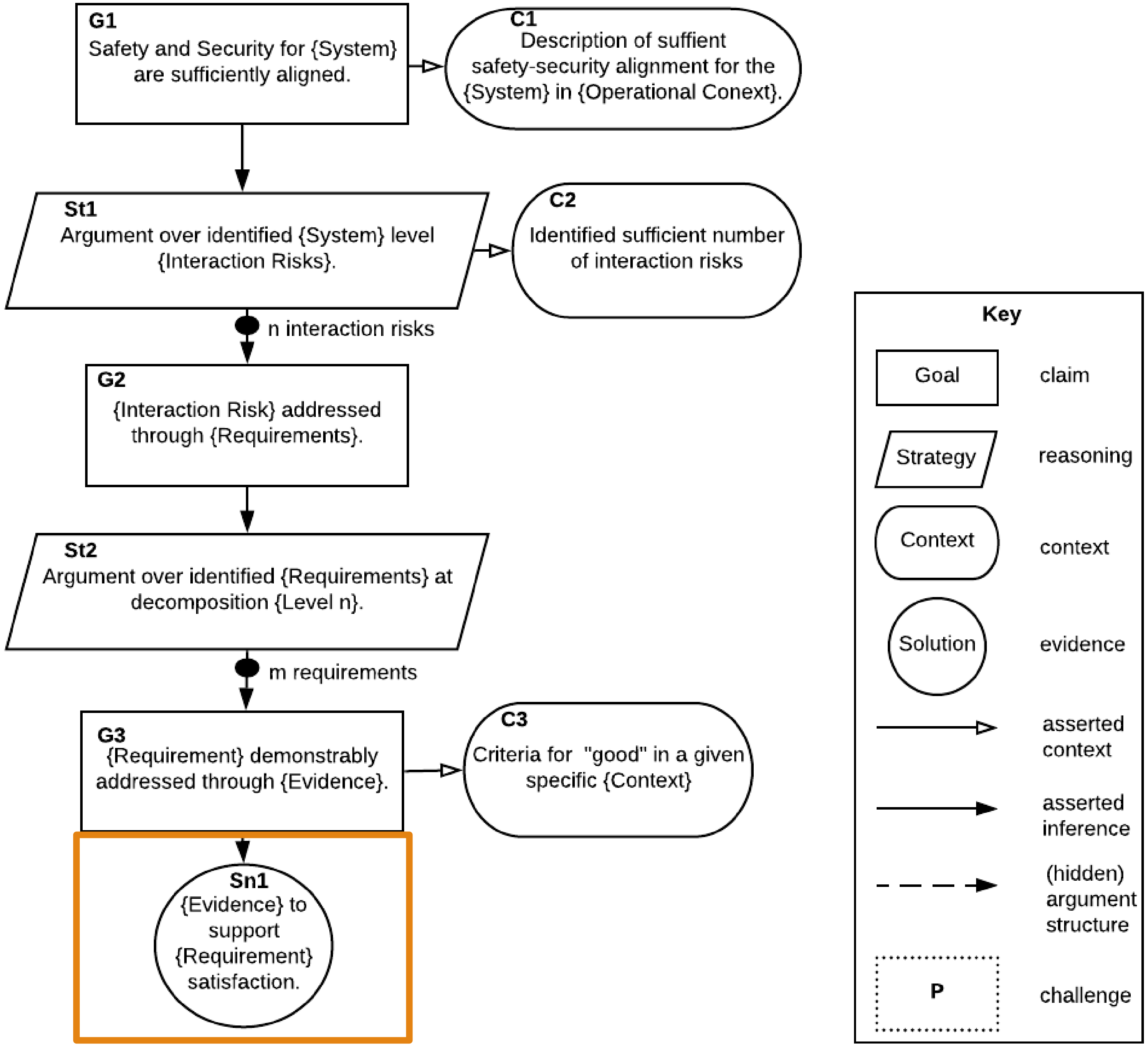


Interaction Risks

SSAF Causal Model



Technical Risk Argument



G1
Safety and Security for {System} are sufficiently aligned.

C1
Description of suffient safety-security alignment for the {System} in {Operational Conext}.

St1
Argument over identified {System} level {Interaction Risks}.

C2
Identified sufficient number of interaction risks

n interaction risks

G2
{Interaction Risk} addressed through {Requirements}.

St2
Argument over identified {Requirements} at decomposition {Level n}.

m requirements

G3
{Requirement} demonstrably addressed through {Evidence}.

C3
Criteria for "good" in a given specific {Context}

Sn1
{Evidence} to support {Requirement} satisfaction.

Key

Goal claim

Strategy reasoning

Context context

Solution evidence

asserted context

asserted inference

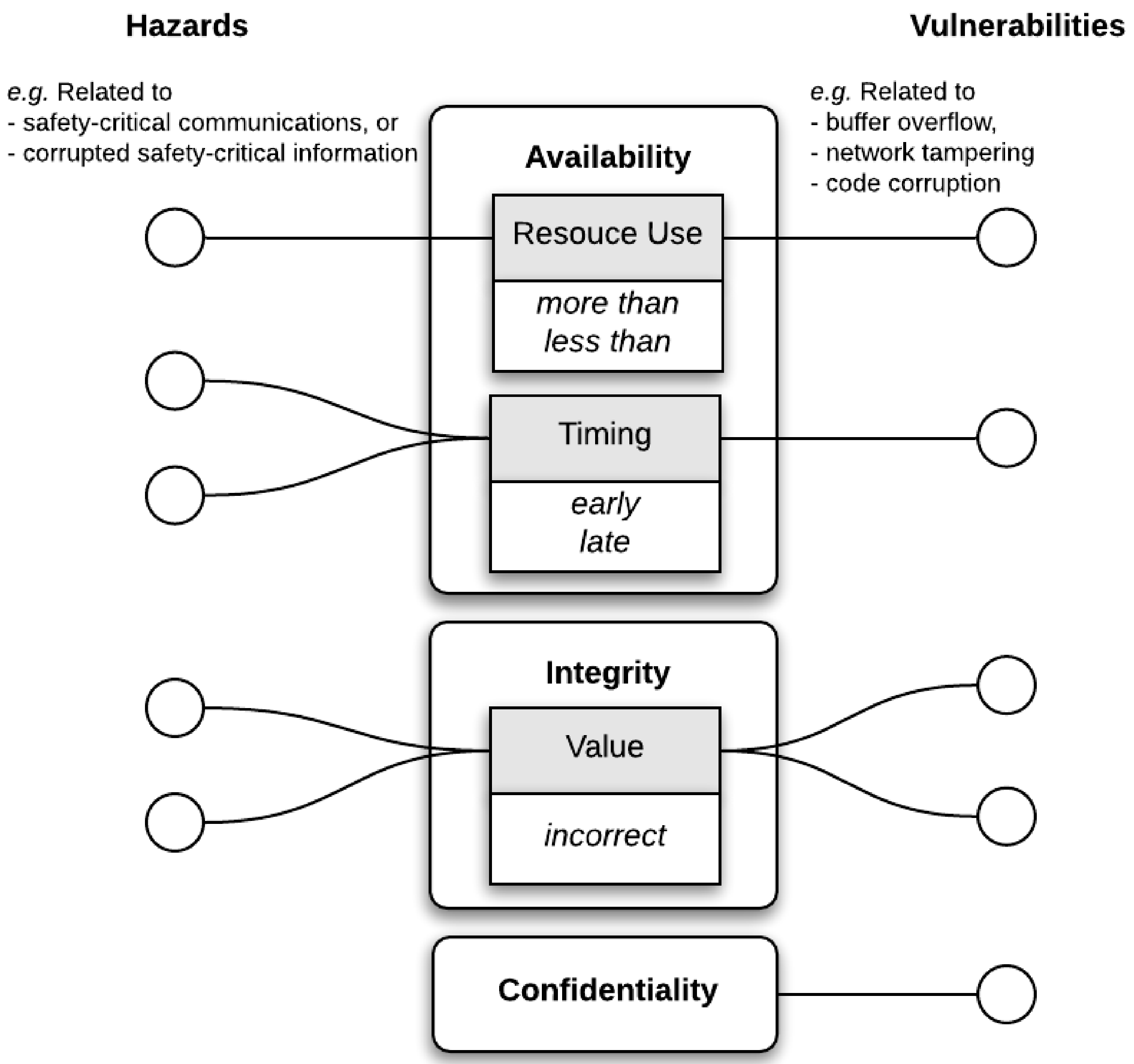
(hidden) argument structure

P challenge

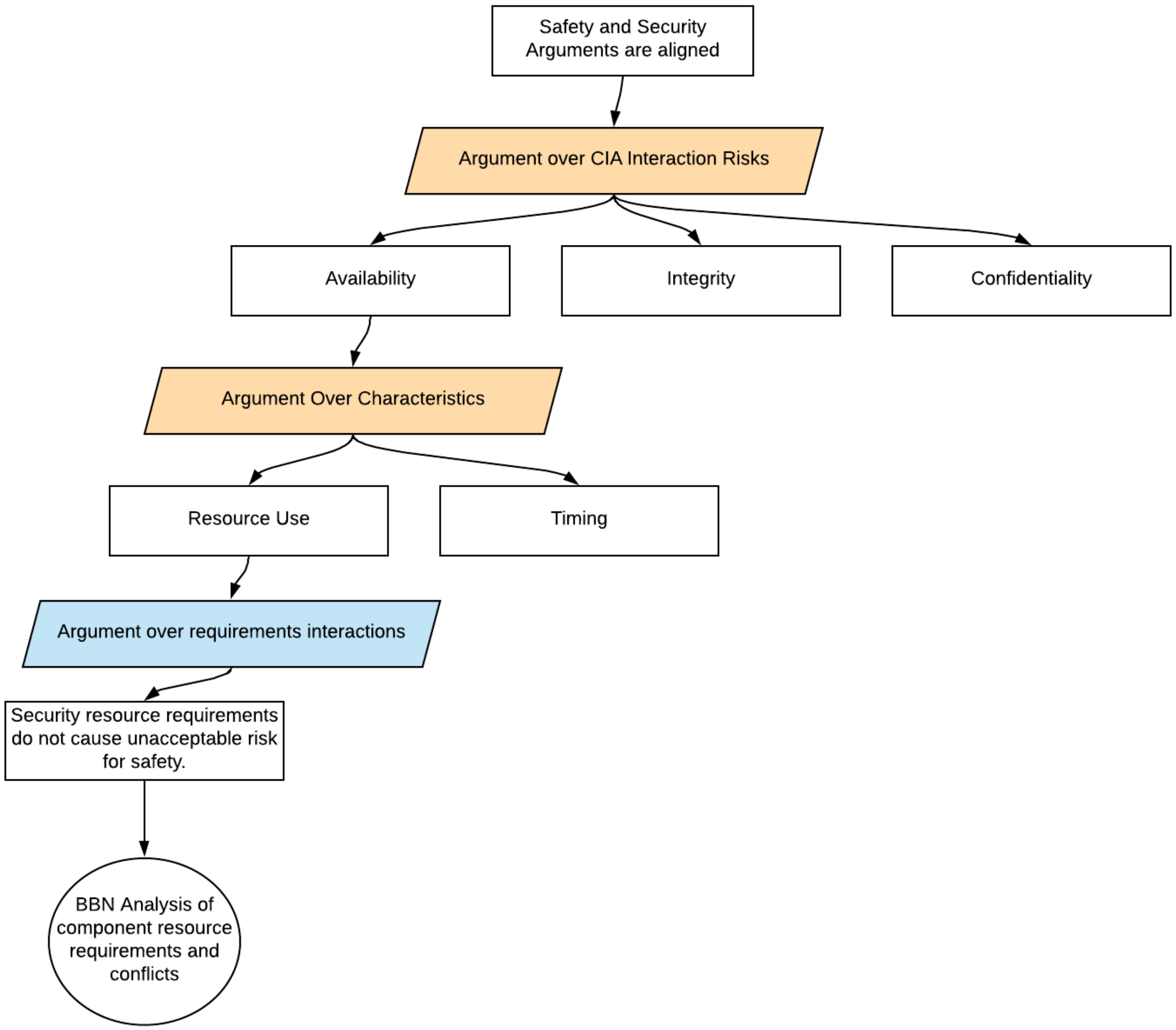
Examples: Links for safety-security

CR.ID	Condition		Causal Relationship	
	Source	Target	Label	Method
	Safety Requirements	Security Requirements	trade-off	ATAM
	Security Requirements	Safety Requirements	trade-off	ATAM
	Threat Condition	Safety Requirements	influence	STPA-Sec
	Threat Condition	Safety Requirements	influence	STPA-SafeSec
	Vulnerabilities	Failure	cause	FFA
	Vulnerabilities	Hazards	contribute to	SAHARA, DDA, UML, FTA
	Safety Consequence	Attack	motivates	ADT
	Threat Condition	Hazard	safety impact	Standard
	Security Controls	Safety Requirements	conflict with	ad-hoc

Interactions using sub-attributes



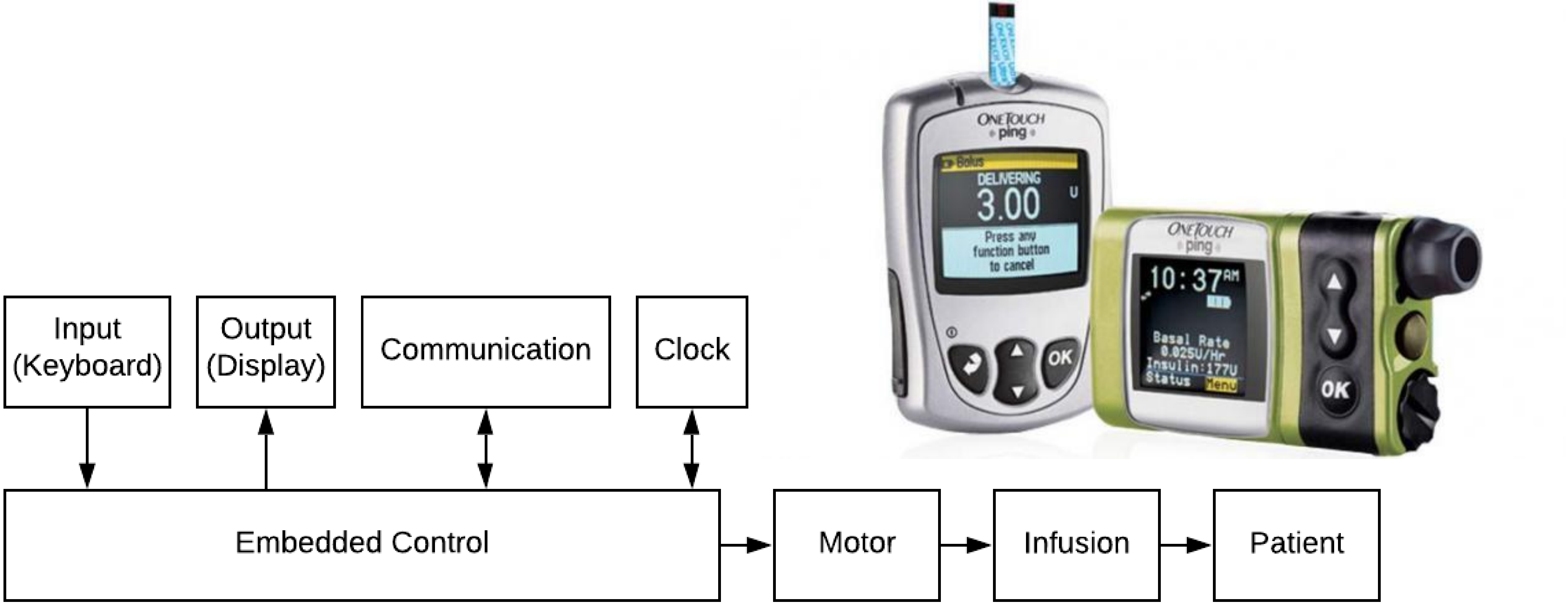
Technical Risk Argument


Safety and Security Arguments are aligned
Argument over CIA Interaction Risks
Availability
Integrity
Confidentiality
Argument Over Characteristics
Resource Use
Timing
Argument over requirements interactions
Security resource requirements do not cause unacceptable risk for safety.
BBN Analysis of component resource requirements and conflicts

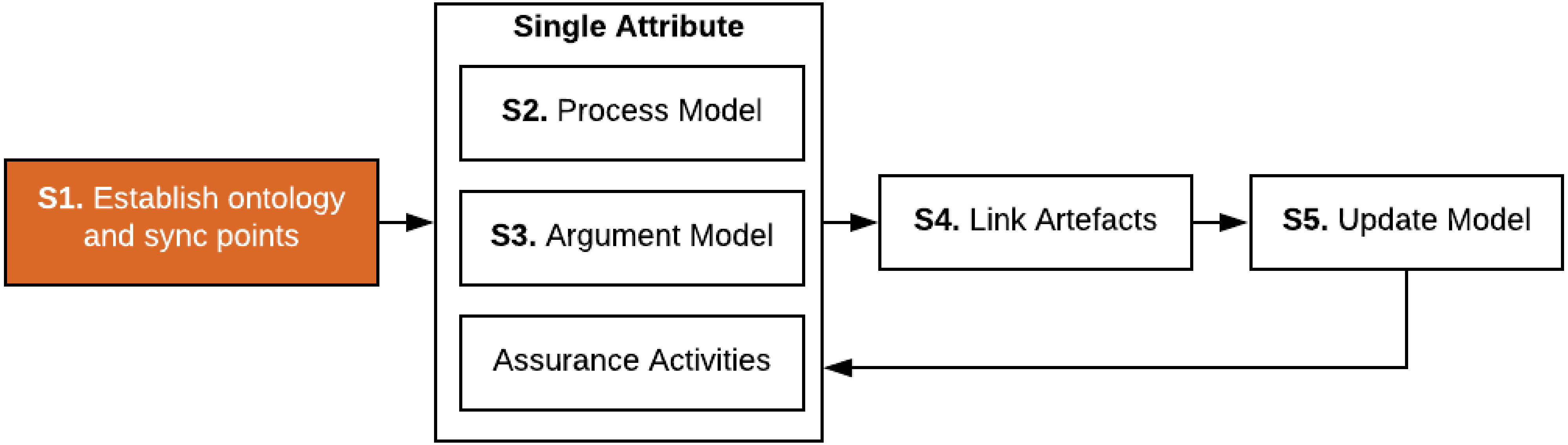
5. SSAF TRM Example

THE SAFETY-SECURITY ASSURANCE FRAMEWORK

Insulin Pump Case Study

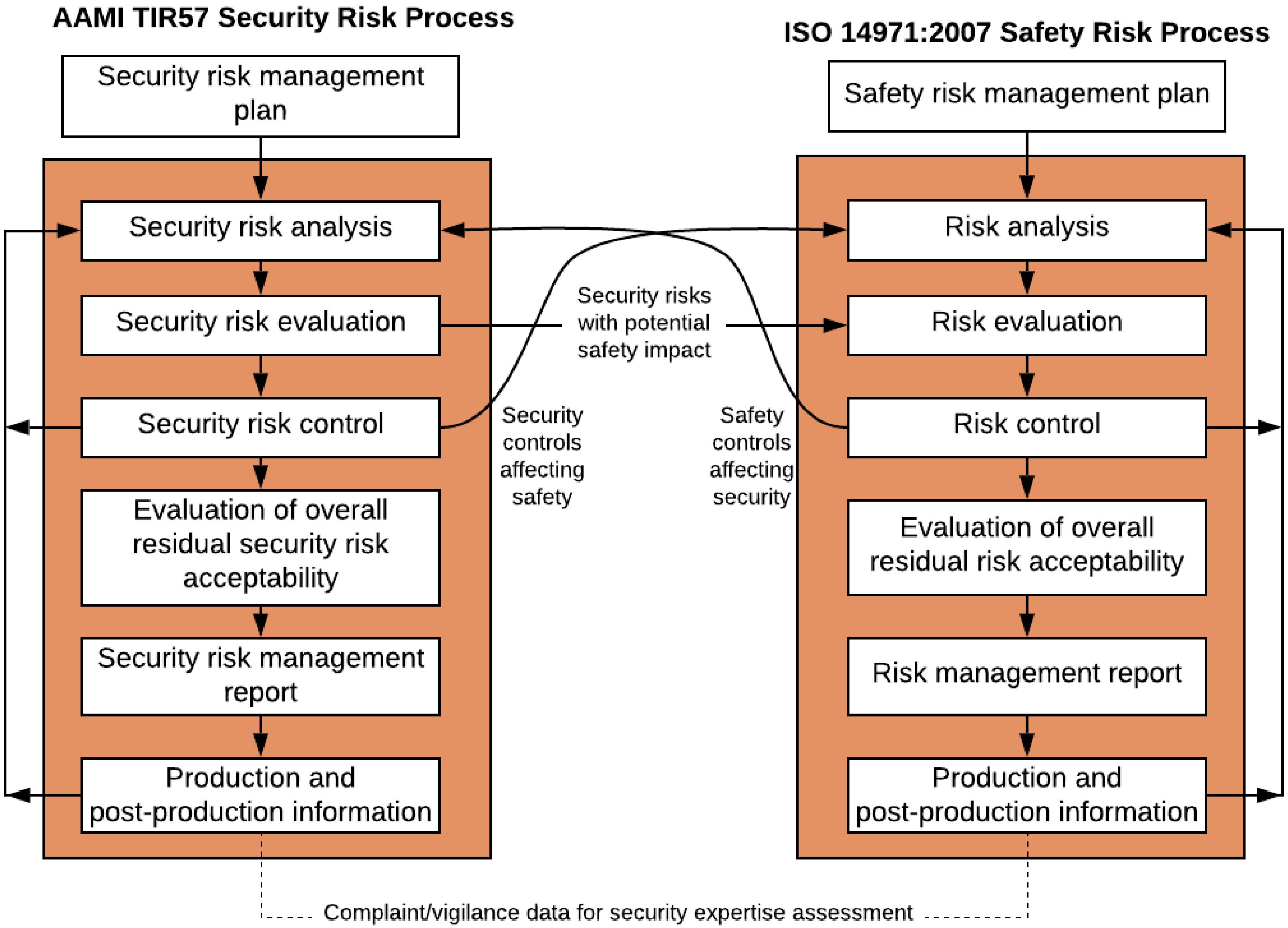


SSAF Technical Risk Process

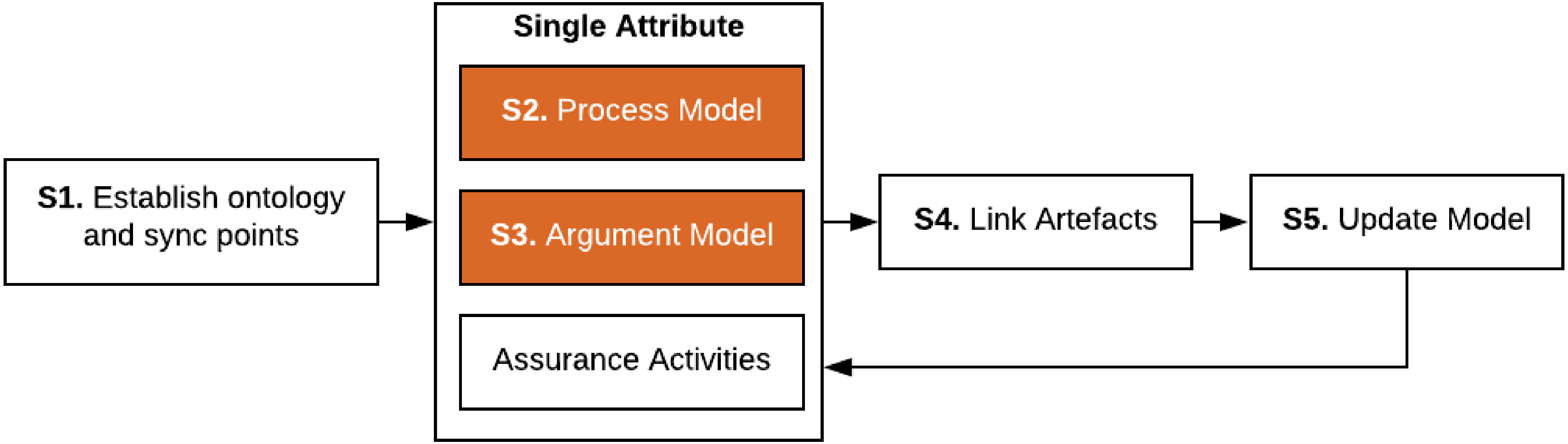


SSAF Technical Risk Process Step 1

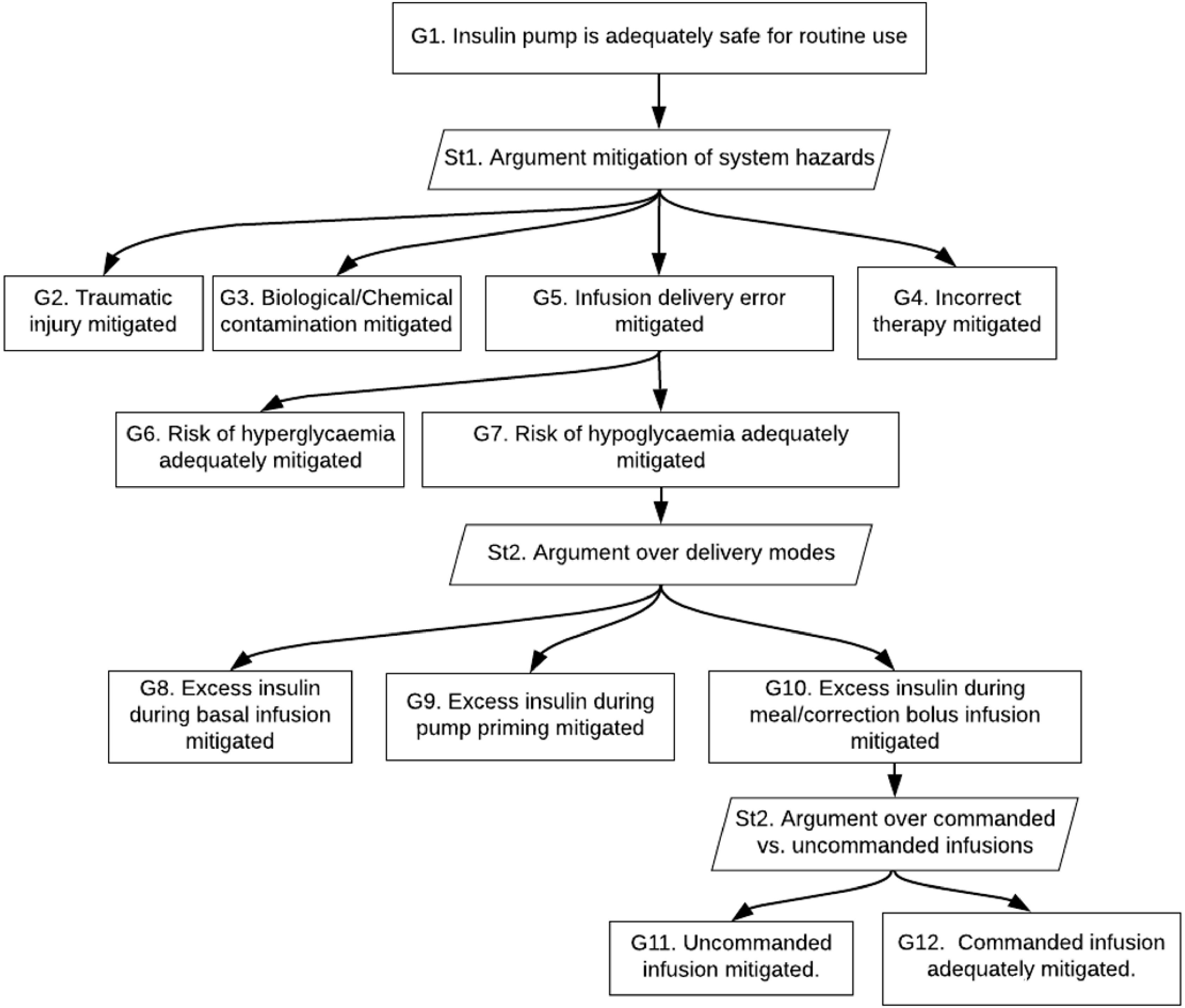
- ○ Ontology
- ○ Sync Points


AAMI TIR57 Security Risk Process
Security risk management plan
Security risk analysis
Security risk evaluation
Security risk control
Evaluation of overall residual security risk acceptability
Security risk management report
Production and post-production information
ISO 14971:2007 Safety Risk Process
Safety risk management plan
Risk analysis
Risk evaluation
Risk control
Evaluation of overall residual risk acceptability
Risk management report
Production and post-production information
Security risks with potential safety impact
Security controls affecting safety
Safety controls affecting security
Complaint/vigilance data for security expertise assessment

SSAF Technical Risk Process



SSAF Technical Risk Process Step 2 & 3



G1. Insulin pump is adequately safe for routine use

St1. Argument mitigation of system hazards

G2. Traumatic injury mitigated

G3. Biological/Chemical contamination mitigated

G5. Infusion delivery error mitigated

G4. Incorrect therapy mitigated

G6. Risk of hyperglycaemia adequately mitigated

G7. Risk of hypoglycaemia adequately mitigated

St2. Argument over delivery modes

G8. Excess insulin during basal infusion mitigated

G9. Excess insulin during pump priming mitigated

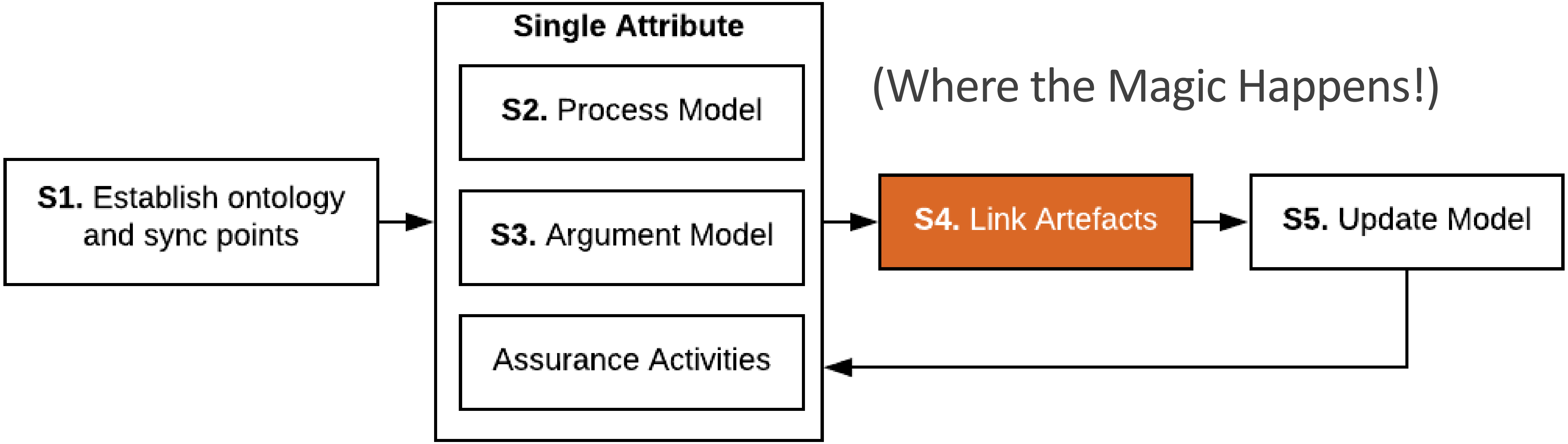
G10. Excess insulin during meal/correction bolus infusion mitigated

St2. Argument over commanded vs. uncommanded infusions

G11. Uncommanded infusion mitigated.

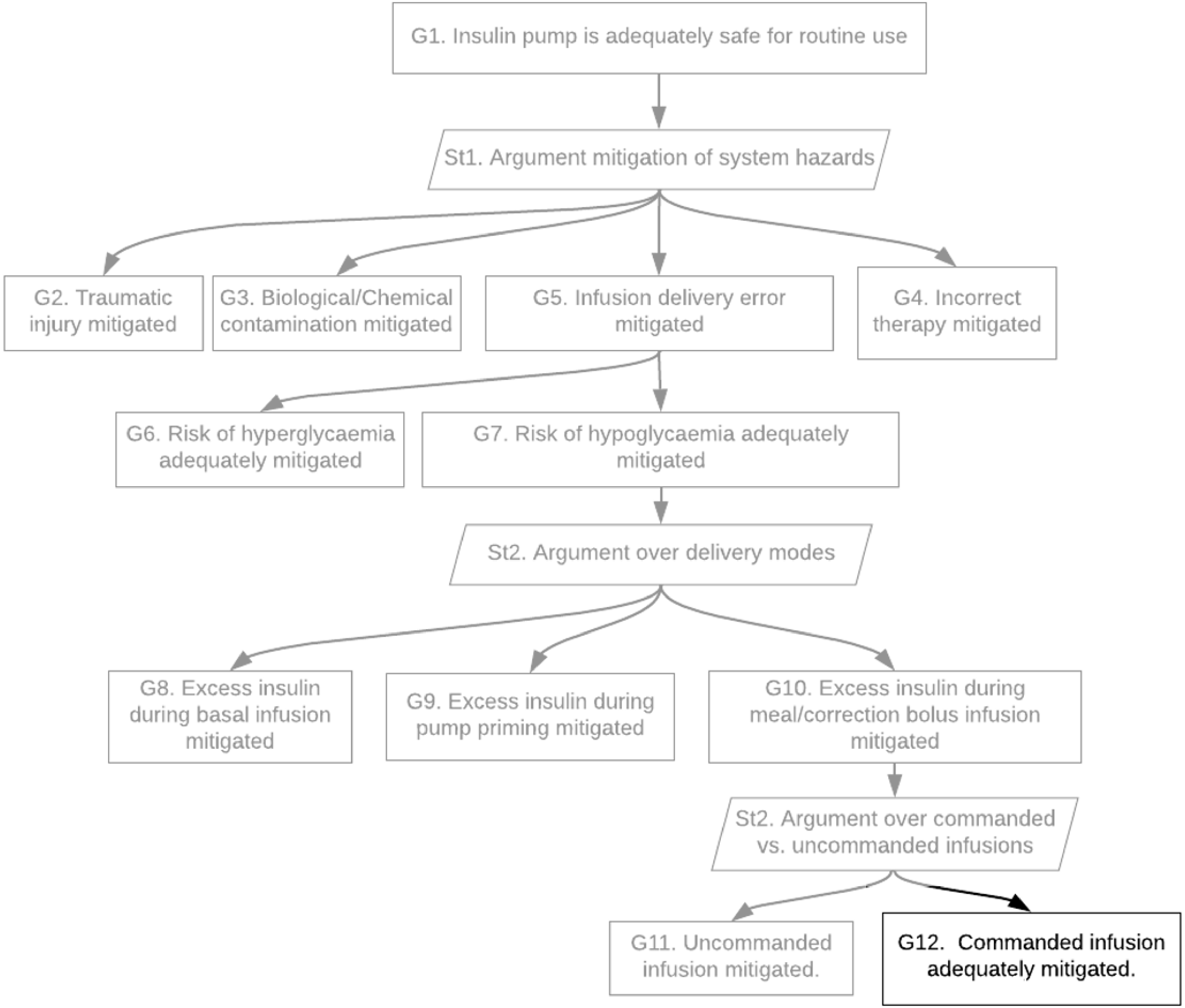
G12. Commanded infusion adequately mitigated.

SSAF Technical Risk Process



(Where the Magic Happens!)

SSAF Technical Risk Process Step 2 & 3



G1. Insulin pump is adequately safe for routine use

St1. Argument mitigation of system hazards

G2. Traumatic injury mitigated

G3. Biological/Chemical contamination mitigated

G5. Infusion delivery error mitigated

G4. Incorrect therapy mitigated

G6. Risk of hyperglycaemia adequately mitigated

G7. Risk of hypoglycaemia adequately mitigated

St2. Argument over delivery modes

G8. Excess insulin during basal infusion mitigated

G9. Excess insulin during pump priming mitigated

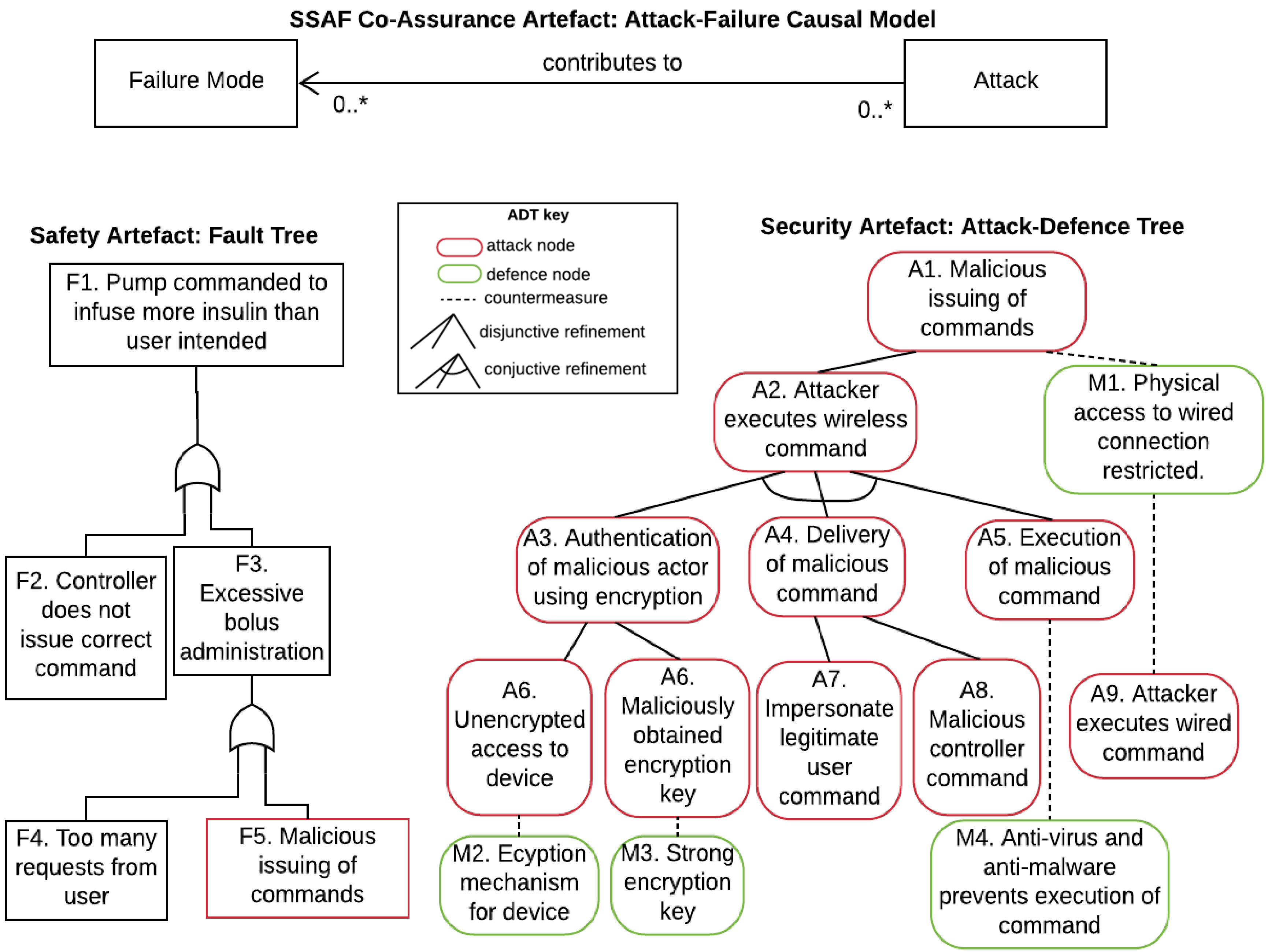
G10. Excess insulin during meal/correction bolus infusion mitigated

St2. Argument over commanded vs. uncommanded infusions

G11. Uncommanded infusion mitigated.

G12. Commanded infusion adequately mitigated.

SSAF Technical Risk Process Step 4



SSAF Co-Assurance Artefact: Attack-Failure Causal Model

Failure Mode 0..* ← contributes to — 0..* Attack

Safety Artefact: Fault Tree

F1. Pump commanded to infuse more insulin than user intended

F2. Controller does not issue correct command

F3. Excessive bolus administration

F4. Too many requests from user

F5. Malicious issuing of commands

ADT key

attack node

defence node

----- countermeasure

disjunctive refinement

conjuctive refinement

Security Artefact: Attack-Defence Tree

A1. Malicious issuing of commands

A2. Attacker executes wireless command

M1. Physical access to wired connection restricted.

A3. Authentication of malicious actor using encryption

A4. Delivery of malicious command

A5. Execution of malicious command

A6. Unencrypted access to device

A6. Maliciously obtained encryption key

A7. Impersonate legitimate user command

A8. Malicious controller command

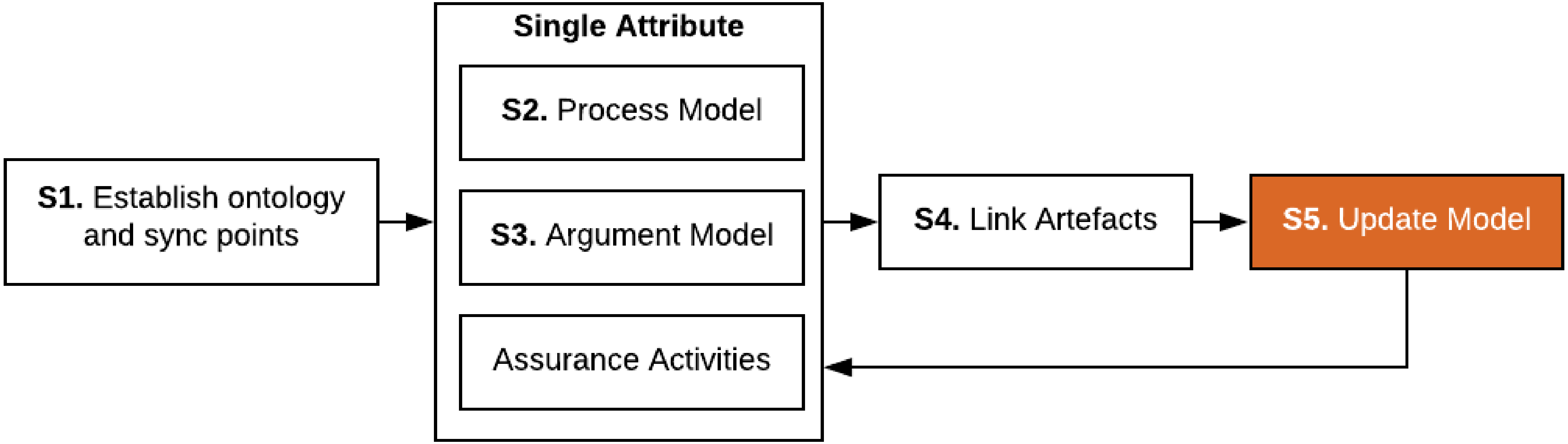
A9. Attacker executes wired command

M2. Ecyption mechanism for device

M3. Strong encryption key

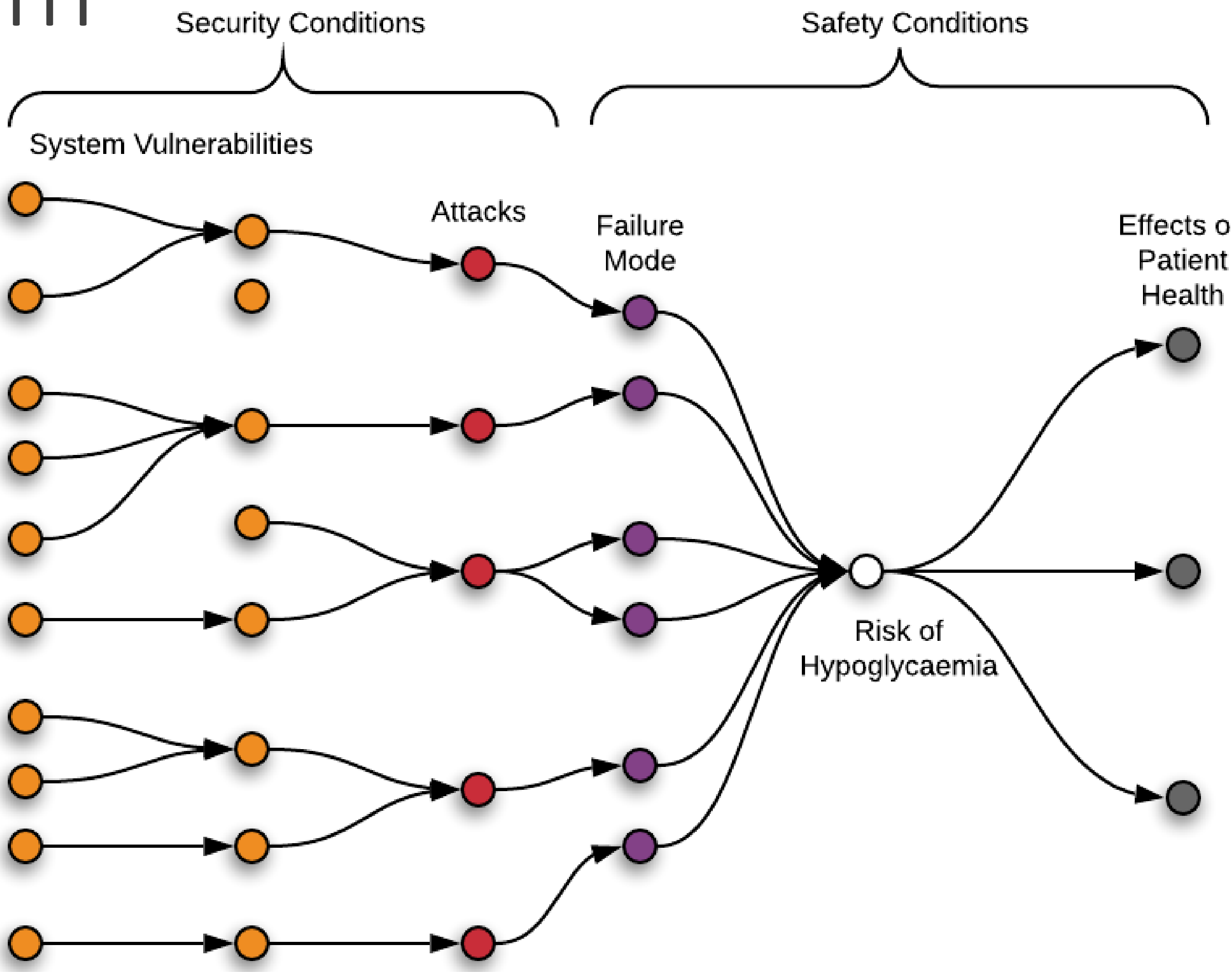
M4. Anti-virus and anti-malware prevents execution of command

SSAF Technical Risk Process



Co-Assurance Claim

Claim: All identified **attack vectors** that lead to **hypoglycaemia** (caused by excess insulin) have been mitigated.



Insulin Pump New Vulnerabilities

New Vulnerabilities

- R7-2016-07.1: Communications transmitted in **cleartext** (CVE-2016-5084)
- R7-2016-07.2: **Weak pairing** between remote and pump (CVE-2016-5085)
- R7-2016-07.3: **Lack of replay attack prevention** or transmission assurance (CVE-2016-5086)

Autonomous Infusion Pump: AAIP SAM Demonstrator Project

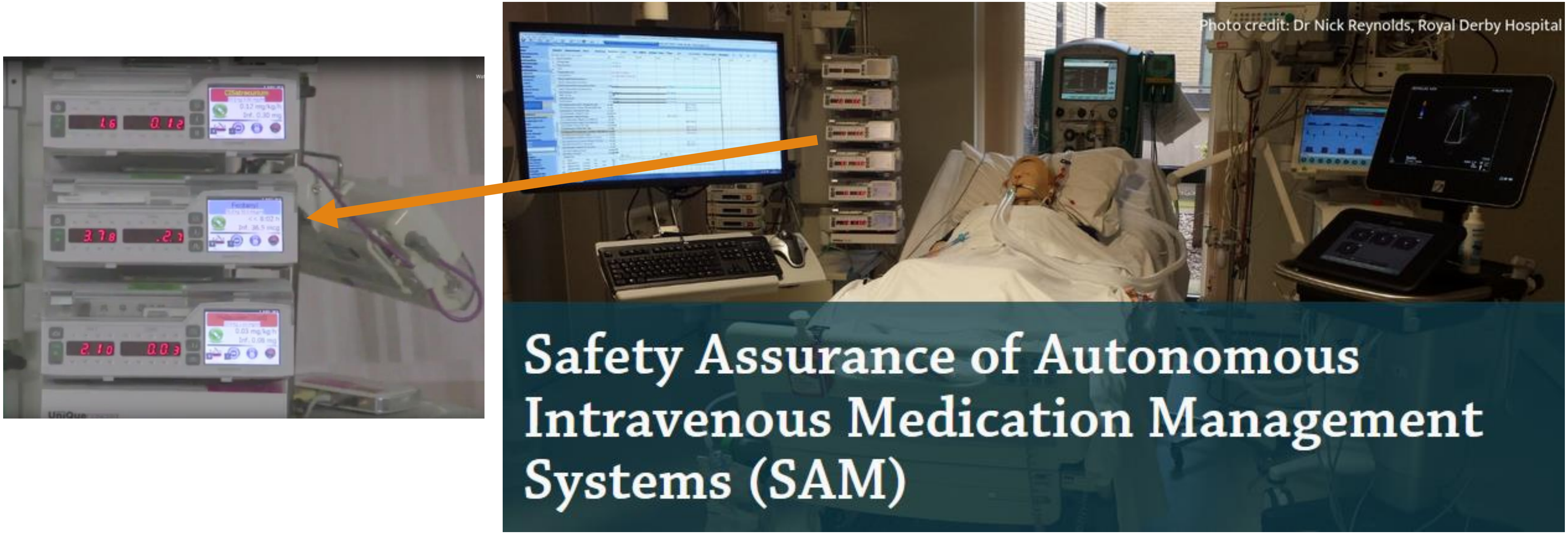
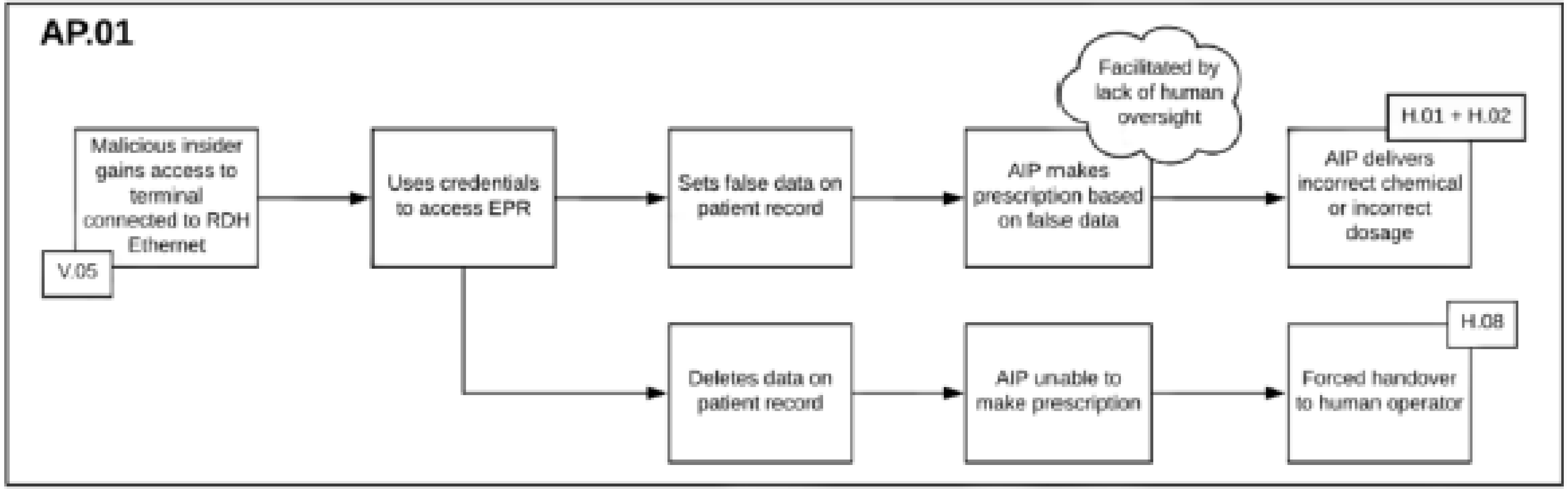


Photo credit: Dr Nick Reynolds, Royal Derby Hospital

Safety Assurance of Autonomous Intravenous Medication Management Systems (SAM)

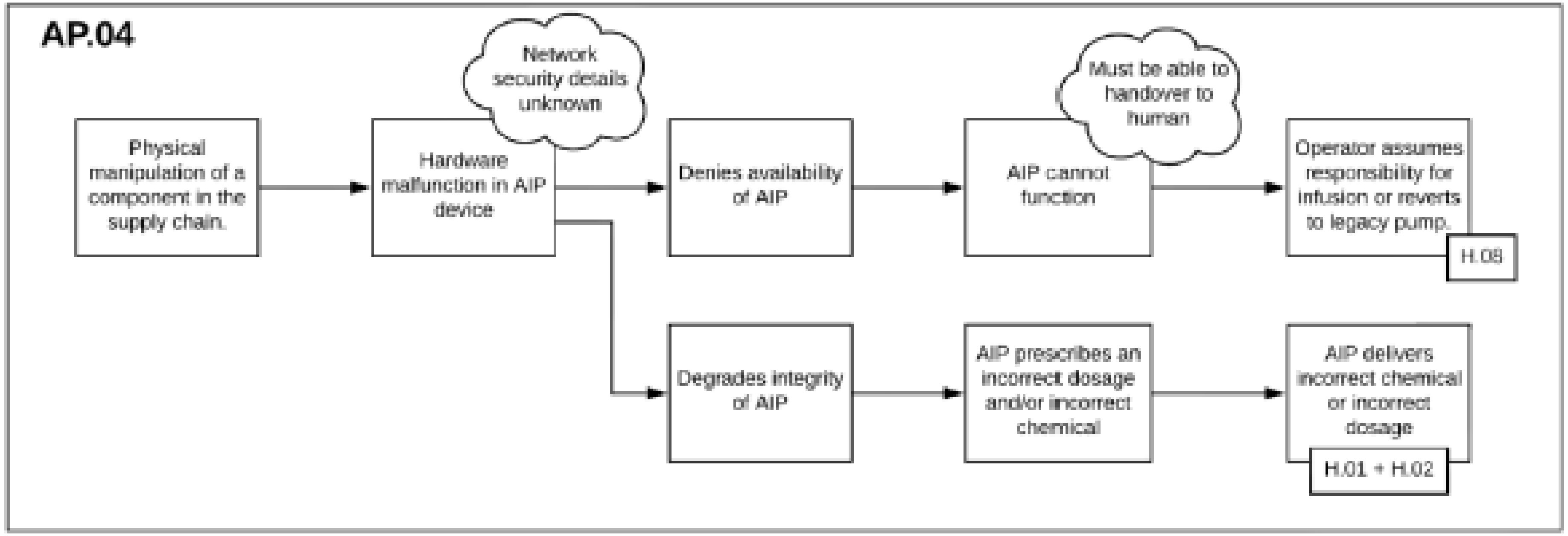
Autonomous Infusion SSAF Links


AP.01
Malicious insider gains access to terminal connected to RDH Ethernet
V.05
Uses credentials to access EPR
Sets false data on patient record
AIP makes prescription based on false data
Facilitated by lack of human oversight
AIP delivers incorrect chemical or incorrect dosage
H.01 + H.02
Deletes data on patient record
AIP unable to make prescription
Forced handover to human operator
H.08

SSAF Link Attack-to-Hazard

- H.02 – Delivering Incorrect Treatment
- H.08 – Forced Operator Handover

Autonomous Infusion SSAF Links



AP.04
Physical manipulation of a component in the supply chain.
Hardware malfunction in AIP device
Network security details unknown
Denies availability of AIP
AIP cannot function
Must be able to handover to human
Operator assumes responsibility for infusion or reverts to legacy pump.
H.08
Degrades integrity of AIP
AIP prescribes an incorrect dosage and/or incorrect chemical
AIP delivers incorrect chemical or incorrect dosage
H.01 + H.02

SSAF Link Attack-to-Hazard

- H.02 – Delivering Incorrect Treatment
- H.08 – Forced Operator Handover

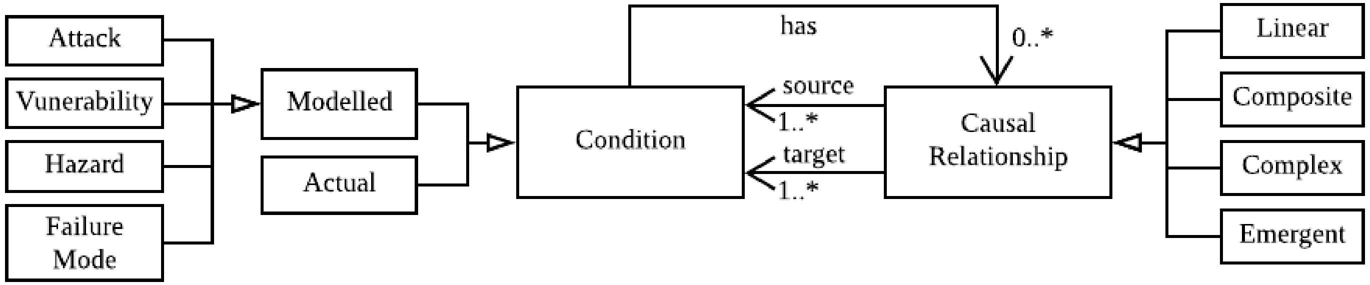
Autonomous Infusion Pump Co-Assurance

- New security risks to impact safety
  - Poisoning attacks, new types of spoofing specific to ML, oracle queries
- Greater uncertainty
  - Trained network deterministic, however unknown connections
- Greater demands on human operator competence
  - Handover
  - Explainability/understandability

Last two points beyond the scope of Technical Risk Argument

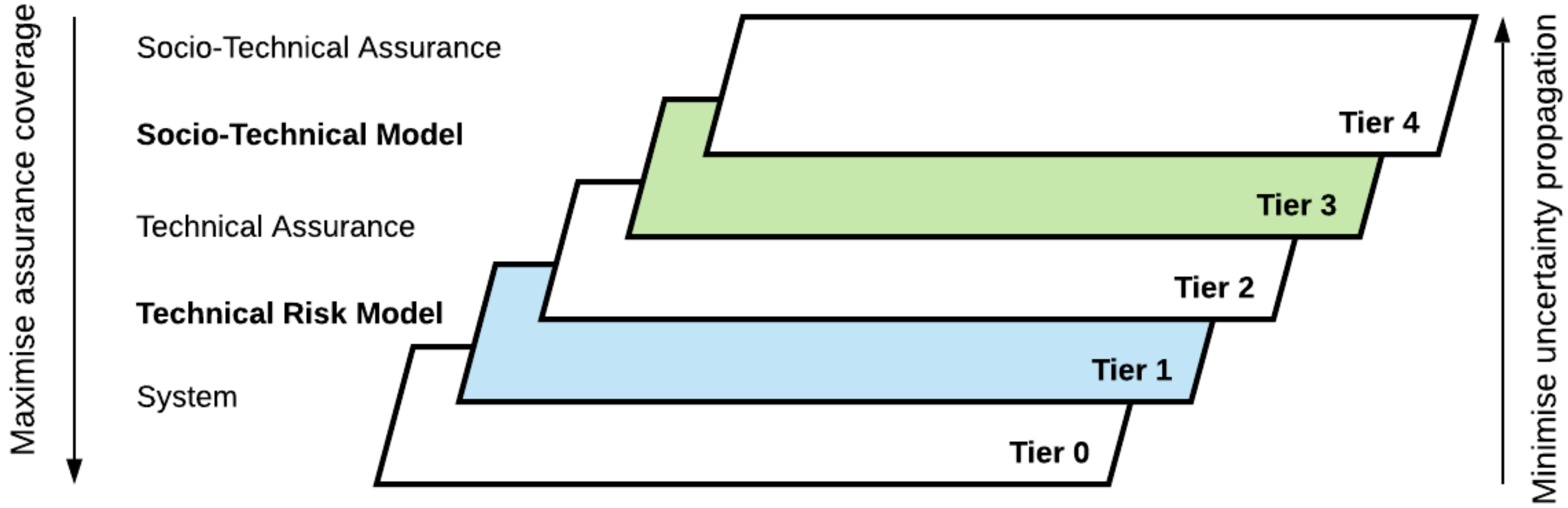
6. SSAF Socio-Technical Model (STM)

SSAF Causal Model (Tier 1)

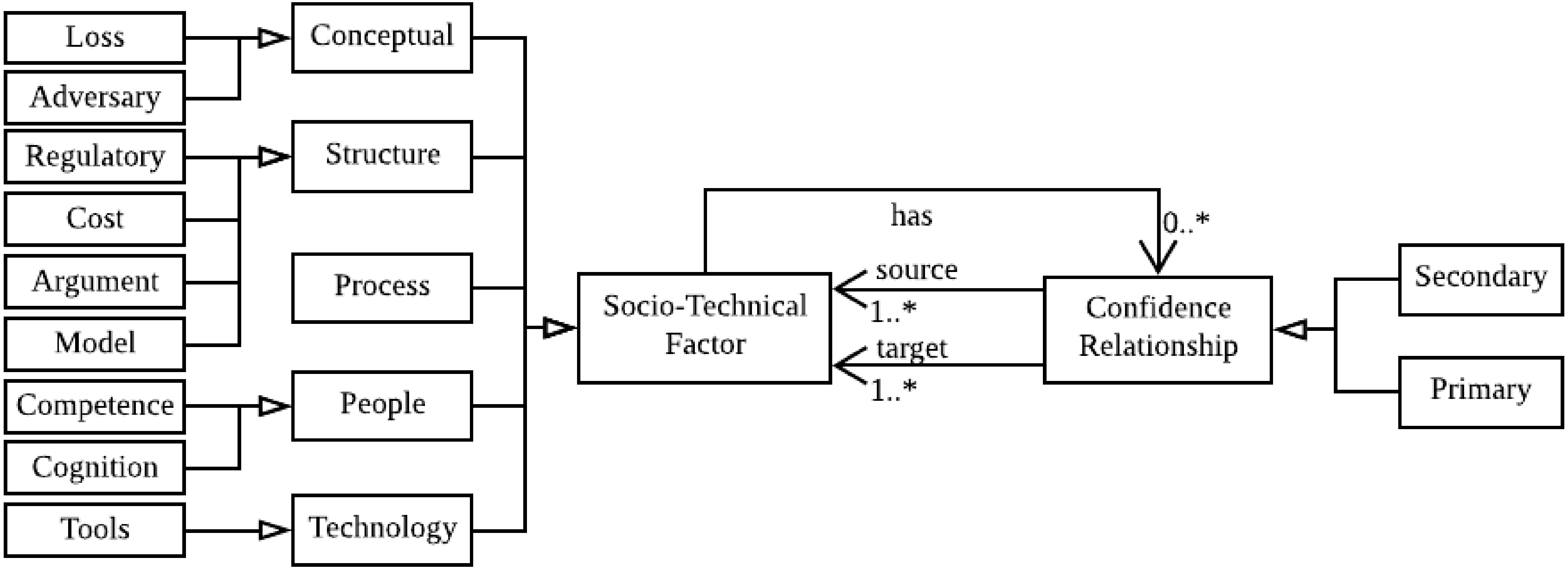


What if the model is wrong? ..

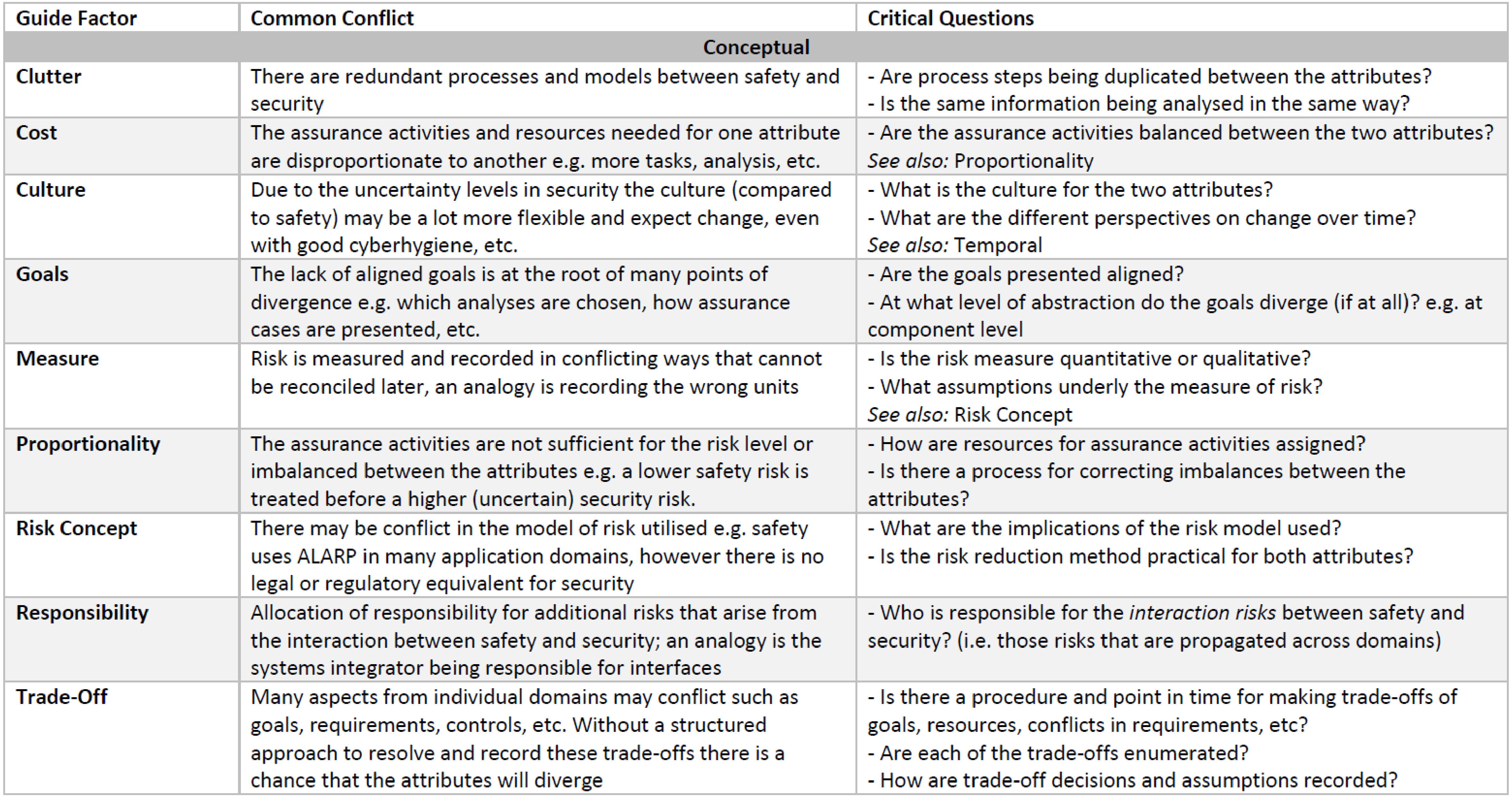
Assurance Surface


Maximise assurance coverage
Socio-Technical Assurance
Socio-Technical Model
Technical Assurance
Technical Risk Model
System
Tier 4
Tier 3
Tier 2
Tier 1
Tier 0
Minimise uncertainty propagation

SSAF Influence Model for Socio-Tech Factors

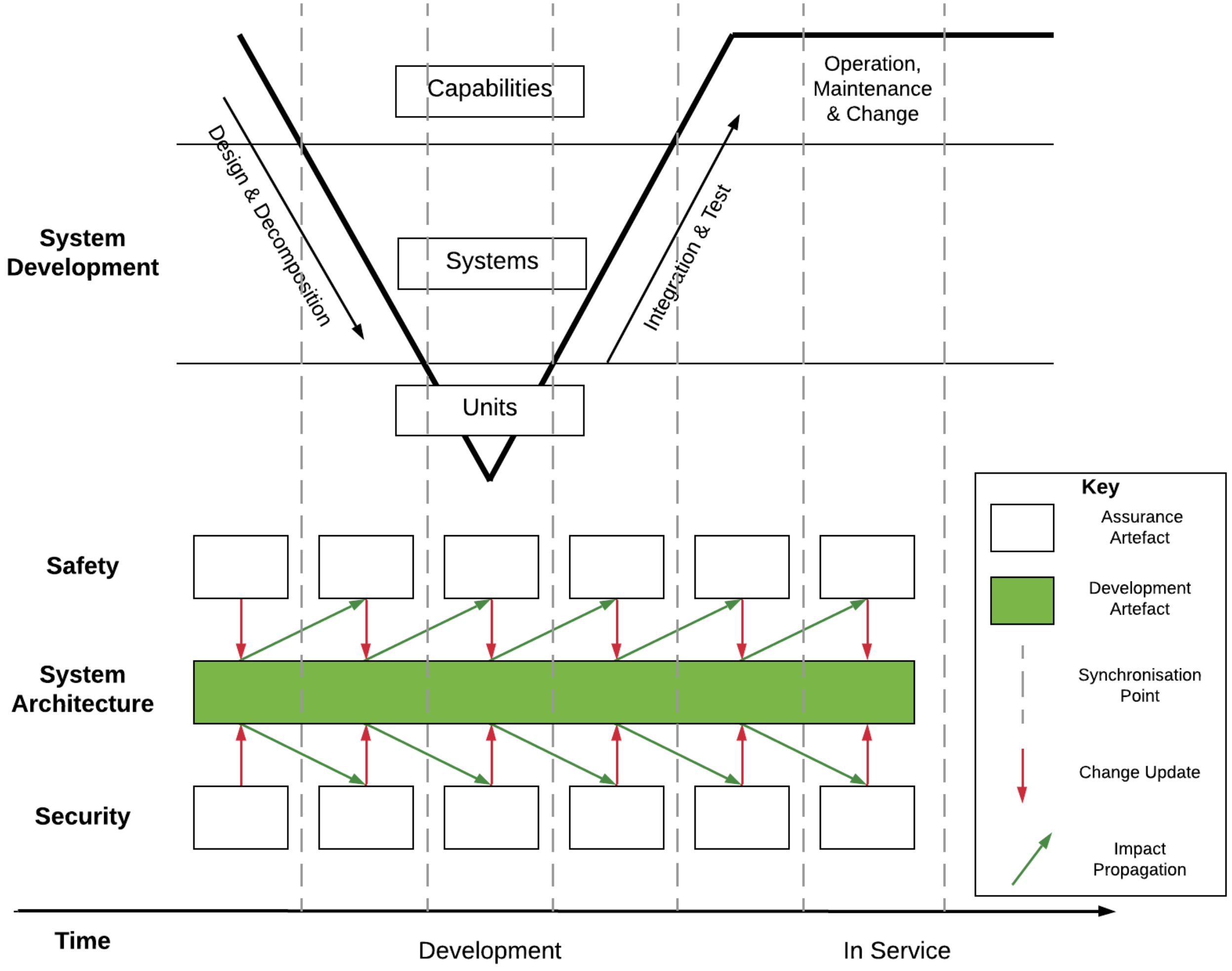


Schemes and Critical Questions



Guide Factor	Common Conflict	Critical Questions
Conceptual		
Clutter	There are redundant processes and models between safety and security	- Are process steps being duplicated between the attributes? - Is the same information being analysed in the same way?
Cost	The assurance activities and resources needed for one attribute are disproportionate to another e.g. more tasks, analysis, etc.	- Are the assurance activities balanced between the two attributes? *See also:* Proportionality
Culture	Due to the uncertainty levels in security the culture (compared to safety) may be a lot more flexible and expect change, even with good cyberhygiene, etc.	- What is the culture for the two attributes? - What are the different perspectives on change over time? *See also:* Temporal
Goals	The lack of aligned goals is at the root of many points of divergence e.g. which analyses are chosen, how assurance cases are presented, etc.	- Are the goals presented aligned? - At what level of abstraction do the goals diverge (if at all)? e.g. at component level
Measure	Risk is measured and recorded in conflicting ways that cannot be reconciled later, an analogy is recording the wrong units	- Is the risk measure quantitative or qualitative? - What assumptions underly the measure of risk? *See also:* Risk Concept
Proportionality	The assurance activities are not sufficient for the risk level or imbalanced between the attributes e.g. a lower safety risk is treated before a higher (uncertain) security risk.	- How are resources for assurance activities assigned? - Is there a process for correcting imbalances between the attributes?
Risk Concept	There may be conflict in the model of risk utilised e.g. safety uses ALARP in many application domains, however there is no legal or regulatory equivalent for security	- What are the implications of the risk model used? - Is the risk reduction method practical for both attributes?
Responsibility	Allocation of responsibility for additional risks that arise from the interaction between safety and security; an analogy is the systems integrator being responsible for interfaces	- Who is responsible for the *interaction risks* between safety and security? (i.e. those risks that are propagated across domains)
Trade-Off	Many aspects from individual domains may conflict such as goals, requirements, controls, etc. Without a structured approach to resolve and record these trade-offs there is a chance that the attributes will diverge	- Is there a procedure and point in time for making trade-offs of goals, resources, conflicts in requirements, etc? - Are each of the trade-offs enumerated? - How are trade-off decisions and assumptions recorded?

7. Conclusion



Implications

- it does not matter which analyses, methods or information as long as it is justified and delivered in a timely manner
- we can start to form patterns for interactions with safety
- make safety and security arguments explicit

Ongoing Work

- Safety-security co-assurance for manufacturing cobots

Further Open Questions

- Proportionality and stopping criteria for co-analysis?
- When to trigger synchronisation?
- Approaches to establishing shared understanding
- Identifying implications of change in assurance cases
- Guidance on making trade-offs
- Forensic activities after an incident
- Establishing a responsible person and accountability

…

Conclusion

- there is a lot of overlap between safety and security
- but! we need to understand the differences to avoid our arguments being undermined
- the adversarial nature of security adds a new level of complexity and uncertainty, but it becomes even *more* important to capture our reasoning and have structured processes.

Thank you! Any Questions?

Contact: nlj500 <at> york.ac.uk