CPNI

Centre for the Protection of National Infrastructure

CAE:

REVIEW AND CHALLENGE

CONTENTS

FIGURES

TABLES

01. INTRODUCTION

One of the challenges in exploring the justification for a system's security or safety is to establish when to stop developing the assurance case. Review and challenge are also fundamentals to the process of developing a case that there is justified confidence in. The purpose and focus of review can be varied as it can be part of:

- development and architecting of the case
- formal confidence building (e.g. independent review)
- formal decision-making (e.g. go/no go operational decisions)

The actual confidence needed in a case is of course dependent on the decision being made and the purpose of the case.

This guidance document brings together three pragmatic pieces of advice to assist in review and challenge:

1. reviewing a case and the underlying models it might depend on;
2. 'stopping rules' for representing the justification graphically in Claims, Argument and Evidence (CAE).
3. an overarching sentencing statement that poses the questions the judge of the case should consider when making a decision based on the case.

The sentencing statement is mapped to aspects of the Assurance 2.0 methodology that would support it and is more technical than the other guidance; it includes a number of developments that have been consolidated in 2022 (technical background to confidence in cases and Assurance 2.0 is available here [1]. In particular, the notion of indefeasibility that means the justification is well supported, all reasonable doubts and objections have been considered and countered, and that there is confidence no significant doubts remain that could change the decision. To achieve this, it is necessary to examine an assurance case from diverse perspectives that focuses on both 'positive' aspects of the case, such as the evidence and argument in support of its claims, as well as those that consider the 'negative' aspects. As with all aspects of assurance, the level of rigour has to be commensurate to the criticality and novelty of the decision being considered.

02.
SIGNPOSTING

This is the fourth CAE guide in the stack of resources for security-informed safety assurance. Figure 1 below shows its location in the set of guides (highlighted in red).

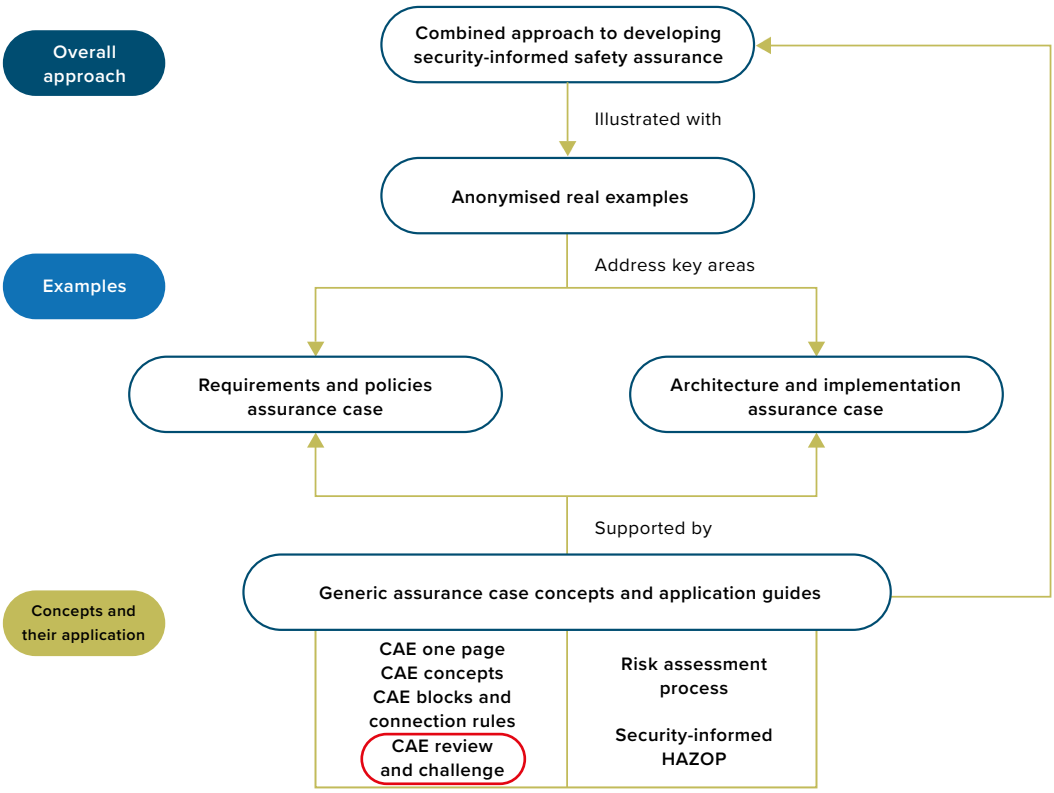

Figure 1: Location of this guide in the set of resources

03. GUIDANCE

3.1 REVIEW AND CHALLENGE

Review and challenge are fundamental to developing a case that there is justified confidence in. The purpose and focus of review can be varied as it can be part of:

- development and architecting of the case;
- formal confidence-building (e.g. independent review); and
- formal decision-making (e.g. go/no go operational decision).

3.1.1 INITIAL REVIEW

An initial review of the case should establish:

- an understanding of what it is for and what it concerns;
- the range of stakeholders involved;
- the systems it concerns;
- the status of the project in terms of its criticality and the decision making it supports; and
- the provenance of the case, how the case was produced (e.g. template, brainstorming, etc.).

The next step is to gain an overall impression in order to:

- understand the main claims;
- review the architecture of the case;
- establish what the key evidence is;
- assess the topology, whether it is in normal form, if there are any nodes with an excessive number of subclaims; and
- appraise the status of completion (e.g. evidence identified but lacking).

3.1.2 TECHNICAL REVIEWS

If the case passes an initial review then more technical reviews should be undertaken of the CAE structure. This should address the verification and validation of the CAE as well as consider its fitness for purpose and explore other design and assurance options. Questions to be addressed are defined in Table 1 and Table 2.

Question	Discussion
Are the CAE concepts properly applied and is the CAE properly formulated?	Are the claims, arguments and evidence actually claims, arguments and evidence?
Is the structure of the case sufficiently complete?	Does the evidence provide a link to the top-level claim? Are all leaves of the CAE tree subclaims that are either recognised as assumptions or evidence?
Does the case follow the connection rules?	See ‘CAE Blocks and Connection Rules’. Are deviations justified?
Are the CAE blocks applied correctly?	Do the CAE blocks comply with their definitions? (see ‘CAE Blocks and Connection Rules’). For instance, justify that dependability = {reliability, availability} or justify that if system model is A and B then response of system is response time of A + response time of B.

Does the case reflect the real world?	Perform validation – does the case actually reflect the real world? Have the CAE blocks side-claims been sufficiently justified? E.g. 'Is the system made of A and B (only A and B?)' or 'Does the property distribute?'
Is the case feasible?	Are the claims realistic? Is it possible to get the evidence at appropriate cost? Are the assumptions well-articulated? Are they valid?

Table 1: Stakeholder preferences for training topics

Question	Discussion
Are there better ways of achieving the same assurance?	Could the case use different types of evidence? Has maximum use been made of evidence? (e.g. operating experience).
Is the impact on other systems and processes acceptable?	Will the case have an unacceptable impact on system design or operation? (e.g. complex architecture from diversity requirements, complex operator actions from not using computer-based systems).

Table 2: Optioneering

In the review, the following should be considered:

- Using an explicit model-based approach to reasoning about the system behaviour.
- Knowledge of known vulnerabilities in systems and previous issues in assurance justifications. It is important to broaden the appreciation of what can go wrong. This could be captured in a variety of checklists.
- Applying checklists and prompts specific to the CAE blocks.
- Applying hazard analysis techniques to the case itself to assess impact of issues with it and focus the review.
- Developing a diverse case to explore the validity of the claim. For example, a "'counter-case' could be developed on why the system does not have the claimed property.
- Whether the case reflects the assurance principles set out in Table 3.

Effective understanding of the hazards and their control should be demonstrated.	Intended and unintended behaviour of the technology should be understood.	Multiple and complex interactions between the technical and human systems to create adverse consequences should be recognised.	Active challenge should be part of decision-making throughout the organisation.	Lessons learned from internal and external sources should be incorporated.	Justification should be logical, coherent, traceable, accessible and repeatable with a rigour commensurate with the degree of trust required of the system.

Table 3: Assurance principles

The IAEA guide [2] provides background to the safety and dependability assessment along with assurance principles.

3.1.3 IMPLICIT AND EXPLICIT MODELS

The CAE will address a property of a system or organisation and as such relies on explicit or implicit models to give it meaning, e.g. our ideas of what a 'X-ray scanner' is.

A model represents the system in a way that is relevant to the claim or property being justified. If the model justifies the claim being assessed, attention can be focused on the residual doubt surrounding the evidence used and the validity of the model.

As part of deciding whether enough has been done, consideration might be given to whether the model has been applied correctly, and whether it is valid. The stopping rules then concern the following questions:

- Does the model capture the required behaviour?
- Is it based on sound principles?
- Has it been applied correctly?
- Does it provide adequate results?
- Is it valid?
- Does it capture all credible fault types?
- Does it contradict other evidence (and vice versa)?

So the flow of the claim process is as follows:

1. Develop an appropriate model.
2. Increase the detail and rigour of the model until a firm judgement can be made about the claim.
3. Check that the model has been applied correctly.
4. Check the trustworthiness of the evidence used.

The stopping rule is illustrated by exploring how a case for the time response of a smart sensor might be developed. At the system level, it is supposed that a temperature has to be measured and transmitted to a controller. As mentioned above, claim decomposition can be driven by a number of partitioning approaches by architecture, attributes or activities. In this case, it might be decided to 'concretise' the attribute timeliness as a response time for these abstract signals, and consider a separate claim for accuracy. The response time for the signal would then be apportioned to different components – an architectural decomposition – and a specification for the device in terms of its concrete inputs and outputs would be arrived at. So:

- the abstract attribute timeliness would be used to prompt the definition of temperature response time;
- the system response time would be refined to produce a smart device response time; and
- the system temperature would be related to the measured signal.

The time response would then be apportioned to different parts of the smart device (the A/D conversion, the output D/A and the main processing) and arrive at a software response time requirement.

In order to justify the response time, a model of resource usage for the software would be needed. The first attempt might be a simple yet conservative model that could be used to try to show that the response time is deterministic by design and is within the bounds. The inadequacies with this model could then be analysed and a more detailed justification developed.

In this example, a focus on justifying the claim from the device requirements and design might miss possible failure modes and sources of timing problems. For example, in examining a real device it might be found that part of the lookup table code uses loops with different numbers of iterations in a binary search – not strictly deterministic but expected to be upper bounded (so accuracy and timing become related, because a bigger lookup table will provide more accurate results). Demonstrating that this is satisfactory from the design point of view requires access to the code or a very detailed pseudo-code like description of the algorithm used and therefore raises the related issue of how much can be done 'black-box'.

Table 4 shows how the stopping rule questions have been interpreted in this example.

Question	Comment
Does the model capture the required behaviour?	The model is about timing and not some other aspect. It provides an upper bound on the execution time.
Is it based on sound principles?	Can examples be found in the literature? Yes for 'worst case execution time' in general but not for this specific simplified application of it.
Has it been applied correctly?	The results have been reviewed and checked with independent diverse calculations.
Does it provide adequate results?	Is the bound calculated within the required response time? If not, understand why. Remove some approximations, detail model or abandon approach and accept negative result. Does the model increase understanding and insights?
Is it valid? Does it capture all credible fault types?	Are the assumptions credible? Look for any credible mechanisms that would lead to significant time delays and stop when it can be shown that these are not present or have a quantified impact.
Does it contradict other evidence (and vice versa)?	Is the model consistent with test results? Can the degree of pessimism be explained? Does it explain the fastest response time?

Table 4: Example review questions for smart sensor devices

3.1.4 DEFEATERS

In reviewing and challenging the case, it is being looked at from both a positive and negative aspect, searching for any doubts in the reasoning behind each argument step of the CAE. These doubts or counter-beliefs are captured as part of the CAE approach in the form of defeaters.

The systematic search for defeaters as part of the case development builds confidence that the overall claim is true or, in other words, that the case is indefeasible. Confidence in the case can be increased by identifying, and then eliminating, potential defeaters through more detailed analysis. This approach supports efforts to reduce confirmation bias in security cases and allows the challenge and review procedure to be formally recorded and auditable.

3.2 CAE STOPPING RULES

How do we know that what we have designed is enough? Having a CAE diagram that is large will be difficult to maintain and communicate, while having a CAE diagram that is too small means that it may fail in expressing the argument in an effective way and omit important detail.

Stopping rules give advice on the question “How do I know I have done enough?”, “When should I stop adding detail to the CAE case?” It is not possible to give a simple set of rules to answer these questions as this will require expert judgement on the purpose of the case, who is using it and for what purpose. Table 5 presents a useful set of questions to help in this judgement.

Question	Comment
Is it appropriate to make the case more detailed? **Are you dealing with properties that do not make sense at a finer level of detail?**	If you are working top down to develop the CAE, there may be many issues that can be best addressed by ‘divide and conquer’ - applying the CAE decomposition block to grow and detail the structure. At some level of detail this tactic will cease being useful as reductionism is not appropriate (e.g. is it a system level or emergent property that cannot be assessed in terms of components).
Is there a good balance between the graphical and narrative?	Sometimes, graphical is not helpful. There are some aspects where the graphical approach might not be appropriate and that linking or including tables is a better approach. For example, representing a hazard log graphically may provide an unhelpfully large cauliflower-like figure. Is there a good balance between graphical and narrative?
Is there unnecessary clutter in the CAE? **What are the essential aspects to the case?** **Would some aspects be better presented in a narrative or as a set of assumptions?**	As the case develops claims and information that were thought to be relevant may not be needed. It may be helpful to prune the CAE structure. Following the guidance on using CAE blocks helps restrict the case to the essential items. It is best to use the CAE to capture the relationship between established arguments rather than reproduce arguments for which there is already a good notation. A proof of security property should be done in the appropriate mathematics.
Has the need for different viewpoints by different stakeholders with different levels of detail been addressed?	It might be useful to provide summaries of the case for different stakeholders. Some might require a vertical slice (e.g. interested in a particular property or subsystem), while others might want a more strategic view of the high-level claims and the key evidence. Guidance on summarising cases can be found in ‘CAE Blocks and Connection Rules’.

Table 5: Reviewing graphical CAE

3.3 SENTENCING STATEMENT

In Assurance 2.0 a so-called 'sentencing statement' has been developed that helps guide the assessor or case developer in the case that is being presented. (Note: the term sentencing as it is used in some industries to indicate the evaluation of the findings in safety justifications and the judgement of their significance. It is similar to how the term is used to describe a judge's deliberations.)

An idealised view is taken of how someone might express the judgement they are making on the acceptability of a system documented in an engineering justification or assurance case. It addresses how an informed user might make this judgement.

They might make the following statement:

> "*On the basis of this assurance case and an examination of other relevant documentation, I judge the proposed {product, service, system} to be {adequately safe and effective, unsafe, secure and effective, insecure or ineffective, insufficient to make a judgement}. I believe my judgement is sound and valid because …*"

The details of their reasoning and how the framework would support it would be captured in a sentencing statement in table below. The table shows how elements of Assurance 2.0 support the engineering reason. For a specific assessment, the methodology commentary would be replaced with answers to the issues raised in the 'engineering reasoning' column.

Engineering Reason	Role of CAE Assurance 2.0-based methodology
Understand the context and criticality of the decision. **I understand how the top-level claim is appropriate for the decision being taken. I understand this decision and the impact it will have and I understand the confidence needed in my judgement.**	Foundational core of framework to support formulation and analysis of top-level claims and understanding of the criticality of system. Supporting methodology will support exploration of top-level claims: sometimes claims can be changed to provide clearer case. Undertake sensitivity analysis of changing system boundary, e.g. ensure socio-technical aspects properly captured. Use of chain of confidence to explore alternative arguments if top claim not valid. Explore confidence/claim trade-offs.
Understand the system. **My judgement is based on an understanding of the role of the product/ system/service and how it contributes to the overall system, its complex interactions, its failure modes and emergent properties.**	The importance of the system context is key, supported by links to system models (outside of the framework).
A golden thread of reasoning from evidence to claim. **I have identified, and where missing developed myself, a clear thread of reasoning linking evidence to direct intermediary subclaims and these subclaims to the top-level claim.**	Foundational core of framework to provide clear thread. Explicit use of different types of reasoning, CAE concepts and CAE blocks to provide a clear thread. Identification of deductive and inductive steps.

Evidence-based decision-making. **The evidence provided is sufficient/ insufficient to provide confidence in the claims. I have identified the key confirmatory evidence.**	Review of evidence and tooling to integrate wide variety of evidence artefacts. Could also have subclaims reviewed using confirmation theory. The use of verified CAE patterns to add rigour and confidence.
Evidence-based decision-making. **I have also identified what evidence would be capable of disproving the claims and this has been sought and where available analysed.**	Explicit use of arguments and separation of inductive/deductive reasoning. Identification of rebutting defeaters. Methodology and tool support for seeking and addressing defeaters, guided by confirmation theory.
Actively seeking doubts. **In developing my confidence in my judgement, I have systematically identified sources of doubts (and attack) addressing both aleatory and epistemic issues, and judge that these have been addressed adequately.**	Systematically used a defeater identification and management system. Discuss major sources of doubt and the impact of undercutting defeaters and their aggregation. Reviewed any possible tipping points undermining confidence. Applying modified hazard analysis to the case itself. Developed own version of CAE case to support this.
Address biases and fallacies. **I am aware of potential biases in my judgement and have addressed these by e.g. seeking independent opinions, reviewing other similar cases, reflecting on past cases where misjudgements have been made by myself or my peers.**	Methodology and tool support for seeking and addressing defeaters. Incorporation of experience of flaws in critical systems to inform methodology (i.e. as prompts in defeater management system, use of templates to capture experience). Use of symmetrical Kemeny / Oppenheim formulation of likelihood confirmation measure to judge claim and counter claim for evidence integration. Development of counter-case and comparison. Deployed guidance on avoidance of common fallacies (outside of the framework). Psychological and human factors informed methodology (outside of the framework, part of specific deployment).
Support evolving case. **My judgement has been documented to enable other evidence to be incorporated into it as and when it becomes available.**	Use of graphical CAE notation and supporting narrative. Lack of evidence or weak evidence identified e.g. by traffic lighting (outside of the framework). Tooling support.
Support communication and challenge. **My judgement has been documented so that my understanding and rationale can be transferred to others and challenged and revised as necessary.**	Use of defined set of core concepts of CAE. Use of templates and common ontologies (outside of the framework). Provision of training material (outside of the framework).

The practitioner's statement would be supported by other aspects not covered in the CAE conceptual framework: the set of tools for managing the detail and scale of the case, the need for engineering processes and quality management system.

04. ACKNOWLEDGEMENTS

This document is based on material developed in earlier projects partially funded by the UK Control and Instrumentation Nuclear Industry Forum (CINIF) and guidance from previous CPNI projects and published research by Adelard.

05. BIBLIOGRAPHY

[1] B John Rushby, Robin Bloomfield, Assessing Confidence with Assurance 2.0, arXiv:2205.04522, https://doi.org/10.48550/arXiv.2205.04522

[2] Dependability Assessment of Software for Safety Instrumentation and Control Systems at Nuclear Power Plants" (NP-T-3.27),

Disclaimer

This guide has been prepared by CPNI and is intended to support the implementation of security-informed safety assurance and the Claims, Arguments and Evidence (CAE) methodology. This document is provided on an information basis only, and whilst CPNI has used all reasonable care in producing it, CPNI provides no warranty as to its accuracy or completeness.

To the fullest extent permitted by law, CPNI accepts no liability whatsoever for any expense, liability, loss, damage, claim or proceedings incurred or arising as a result of any error or omission in the report or arising from any person acting, refraining from acting, relying upon or otherwise using the [report]. You should make your own judgment with regard to the use of this document and seek independent professional advice on your particular circumstances.

No Endorsement

Reference to any specific commercial product, process or service by trade name, trademark, manufacturer or otherwise, does not constitute or imply its endorsement, recommendation or favour by CPNI. The views and opinions of authors expressed within this document shall not be used for advertising or product endorsement purposes.

© Crown Copyright 2022.

You may use or reuse this content without prior permission but must adhere to and accept the terms of the Open Government Licence for public sector information. You must acknowledge CPNI the source of the content and include a link to the Open Government Licence wherever possible. Authorisation to reproduce a third party's copyright material must be obtained from the copyright holders concerned.

CPNI
Centre for the Protection of National Infrastructure