

Certification and Safety Cases

Patrick Graydon, M.S., University of Virginia, Charlottesville, VA

John Knight, Ph.D., University of Virginia, Charlottesville, VA

Mitchell Green, Ph.D., University of Virginia, Charlottesville, VA

Keywords: military standards, safety case, system safety management

Abstract

Certifying agencies have begun to require a safety case as part of the product certification process. The standards upon which such certification regimes are built define properties that the safety case must have, e.g., “compelling,” and “valid”, yet leave these terms undefined. Unaided judgment of these properties leaves doubt about how approval will proceed. We introduce an operational definition of these terms in the form of a comprehensive certification process for certification based upon the submission of a safety case. The process defines how certification could be conducted. The process also defines the properties that an acceptable safety case must have since successful certification with this process implies that the safety case has the desired properties. We illustrate our approach to certification using hypothetical argument fragments.

Introduction

Safety cases have become an established component of the safety-engineering field. In this paper, we discuss their role in certification. A safety case is a:

“...comprehensive and defensible argument that a system is acceptably safe to operate in a particular context.” (Kelly, 1998)

A safety case must include both evidence of the safety of a system and a safety argument that explains how this evidence supports the conclusion that the system is adequately safe to operate in the intended operating environment. In Europe, certifying agencies have begun to require the construction of a safety case as part of the product certification process. Usually, this requirement is documented as a standard, and the requirement has become policy for several agencies in the U.K. (Civil Aviation Authority, 2010, Ministry of Defence, 2007a, 2007b).

In a safety-case-based certification regime, a certifying agency (or its agent) must examine the submitted safety case and either accept it — thereby clearing the system for deployment and use — or reject it. Typically, neither the standard that requires the safety case nor any associated guidance provides a precise definition of the mechanism by which the certification decision will be made. The standards established by certifying agencies demand that the safety case developed by the applicant have certain properties, for example that the included safety argument be *compelling*. Checking of such properties is a crucial aspect of the certification process, yet these properties are not precisely defined thereby leaving compliance difficult to ascertain.

In this paper, we present a comprehensive process for certification based upon the submission by an applicant of a safety case to a certifying agency. By definition, our process informs both the applicant and the certifying agency of how certification will be conducted. More importantly, our process *operationally defines* the properties that an acceptable safety case must have. Successful certification implies that the safety case has the desired properties that are otherwise undefined.

We illustrate our approach to certification using various items from a safety case for a hypothetical autopilot system. Our illustration is not an evaluation but indicates the possibilities that the approach offers.

Existing Certification Regimes

The standards upon which existing certification regimes are built require that a safety case have specific properties. Defence Standard 00-56 Part 1 (Ministry of Defence, 2007a), for example, demands that the safety argument be “compelling,” “comprehensible,” and “valid”:

“9.1 The Contractor shall produce a Safety Case for the system on behalf of the Duty Holder. The Safety Case shall consist of a structured argument, supported by a body of evidence, that provides a compelling, comprehensible and valid case that a system is safe for a given application in a given environment.”

Defence Standard 00-56 is a leading example of a certification standard based on safety cases, and, in general, we find the standard to be of the highest quality. However, neither the standard nor the associated guidance defines “compelling”, “comprehensible,” or “valid”. Such properties come from the fundamental definition of a safety case (Kelly, 1998) and are indeed crucial properties.

Without a precise statement of the properties that an acceptable safety case must have, certification becomes unpredictable and unrepeatable. Without precision in the definition of certification, developers have little guidance on what will prove acceptable, and certifying agencies have little guidance on what constitutes an appropriate decision-making procedure. As Haddon-Cave observed in his report concerning the loss of RAF Nimrod XV230 in Afghanistan in 2006 (Haddon-Cave 2009):

“The tenets ... that Safety Cases must be ‘Credible, Consistent, Complete, Comprehensible and Changeable’ are, in my view, too amorphous to inject real rigour and focus into the process”.

Developers and certifying agencies would, we argue, be better served by a testable definition for each of these terms rather than by relying upon the personal judgment.

Related Work

Defence Standard 00-56 Part 2 (Ministry of Defence, 2007b) provides guidance on compliance with Part 1. Part 2 calls for “scrutiny” of the evidence and of the argument itself. This scrutiny is to be provided by an independent team and, for the most critical systems and components, must

cover all of the evidence and the evidence generating processes. “Scrutiny”, however, is not defined.

Kelly has proposed a four-step process for argument review (Kelly 2007). According to Kelly, the sufficiency of an inductive argument step depends upon: (1) the extent to which the argument / evidence presented “covers” the conclusion; (2) the independence of multiple grounds given for a single conclusion; (3) the range of situations under which evidence or a conclusion is thought to hold; (4) the degree to which the argument or evidence “directly” addresses the conclusion; (5) the degree to which evidence / argument is relevant to the conclusion; and (6) the degree to which the argument is sensitive to changes in the evidence. Kelly tasks the reviewer with auditing the evidence presented to verify that the evidence exists and has the claimed properties. Reviewers are asked to assess: (a) the degree of “buggy-ness” of the evidence; (b) the level of peer review to which the evidence has been subjected; (c) the experience and competence of the personnel hand-generating the evidence; and (d) the qualification and assurance of tools used to produce the evidence. Finally, Kelly calls for the reviewers to challenge the argument at hand. Reviewers are to use their domain knowledge to guide a search for evidence that would rebut or undercut an intermediate or final conclusion.

The certification process that we define is organized as a *phased inspection* (Knight 1993), technology that was developed for software inspections. A phased inspection is a rigorous inspection process carried out by human inspectors that is structured as a sequence of phases. Each phase is designed to establish a specific property of the inspection target, and the ordering of the phases allows later phases to assume properties established in earlier phases. Each phase is conducted by an inspector or inspectors who have skills suitable for assessing the specific property that is the subject of the phase. Our process confirms a set of properties that include those described by Kelly. An argument-based certification process has value as an operational definition of argument adequacy to the degree that that process represents the best that can be achieved. Accordingly, our contribution is to propose enhancements, extensions, and additional rigor for the process that Kelly proposes.

Certification Based Upon Safety Cases

The approach to certification defined in this paper is *operational* and involves two participants: (a) the *applicant* who acts for the developers and who is seeking the certification of the subject system; and (b) the *certifier* who is charged with protecting the public interest by ensuring to the extent possible that the subject system is adequately safe. The certifier might be a government organization, an industry group or a professional association, or an agent acting for the primary certifying body. It is especially important in the latter case that the certifier’s responsibility to act independently in the public interest is well understood. For example, after investigating the Nimrod Safety Case effort, Haddon-Cave observed, “the outcome might have been different” had QinetiQ’s role as an ‘Independent Safety Auditor’ been made clear (Haddon-Cave 2009).

Informally, the certification approach that we have defined is for the two participants to engage in a *structured dialog* so as: (a) to establish that the safety case has certain *essential qualities*; and (b) to establish the truth of the *safety claim* being made about the subject system.

The structured dialog between the participants might begin at any point in the development of the subject system. At one extreme, the dialog begins when the subject system is being planned and at the other, the dialog begins when the subject system is completed. Certification is defined to be completed satisfactorily when the certifying agent has determined that the safety case complies with the operational definition of the required properties. Examination of the safety case by the certifying agent at any point is acceptable. We echo the suggestion made by Kelly (Kelly 2004) that safety-case development should begin early and that the content should be reviewed frequently. We do not discuss the temporal aspects of this issue further in this paper.

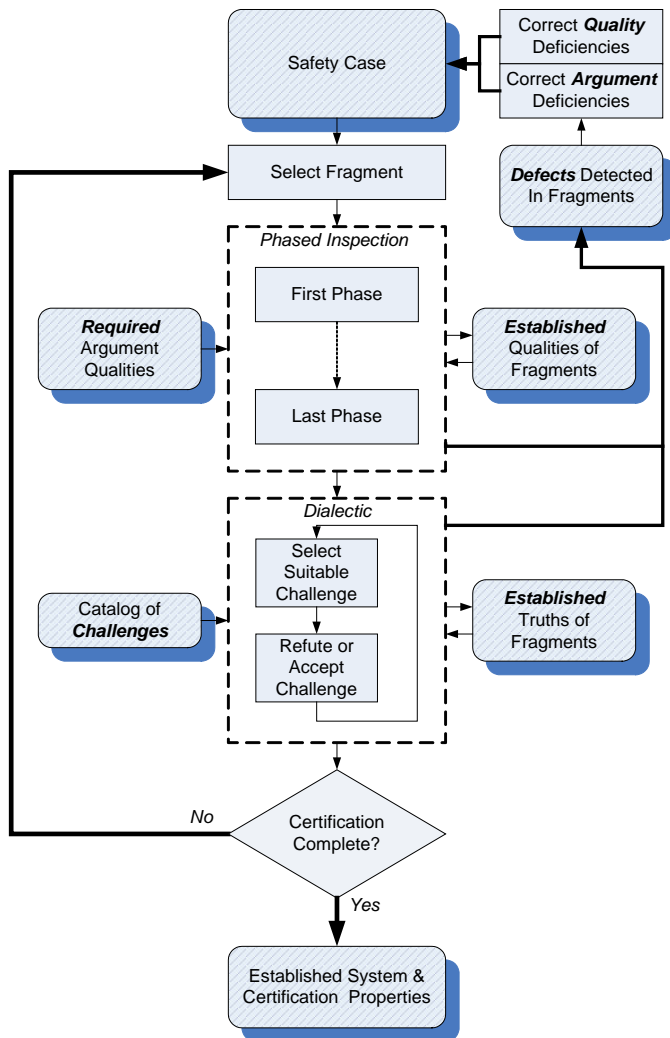


Figure 1. Certification based on a safety case.

parts of the argument as certification proceeds.

Establishing Essential Qualities

The essential qualities that we require of a safety case are established by phased inspection with a separate phase for each quality. The specific qualities and the order of the associated phases are as follows:

Figure 1 shows an outline of the certification process. The safety argument is treated as a set of *fragments*. Each fragment comprises a small number of related argument elements that together support a single sub-claim. Each fragment is examined in two stages. The first establishes the essential qualities of the fragment, and the second establishes the truth of the top-level claim in the fragment.

The certification process requires that two parts of the safety argument, the top-level claim and the argument context, be treated as a fragment that is examined first. The remainder of the argument is then partitioned into fragments at the discretion of the participants. Fragments are examined until the entire argument has been examined, and conclusions about the qualities of fragments and the truth of their claims are maintained as the argument is traversed. The conclusions about one fragment will be input to the analysis of others, and conclusions about multiple fragments are combined in order to establish conclusions about increasingly larger

Terminology Understood Identically By All Readers: The certifier cannot evaluate a safety case adequately if the certifier and the applicant do not share a common understanding of the terms in which the argument is written. Serious safety consequences might result if the certifier and applicant understand a term differently *but are unaware of the difference*. Techniques have been developed to deal with this problem in other contexts (e.g., Wasson 2006); these can be applied to a greater or lesser extent. The certifier should require a definition for any term that is not common knowledge or that he or she knows has been used to convey different meanings in similar contexts. When the meaning of a term appears to be central to the meaning of a critical portion of the safety argument, the certifier should insist upon its definition and test this definition by inquiring about hypothesized examples. Completion of this phase yields both the greatest practical confidence that the certifier and applicant share a vocabulary and, secondarily, confidence that a later third-party reader will also understand the terms used in the safety case.

Absence of Vagueness: As with terminology, the certifier cannot evaluate a safety case adequately if the natural-language text used within the various documents is vague. Elimination of vagueness can be based upon the removal of sentence forms that are known to be problematic and upon the certifier's own experience.

Document Syntactic Validity: A safety case comprises a number of documents. Each of these documents must follow a prescribed format and style. Of particular concern is the safety argument, which must conform to the syntactic rules of the language in which the argument is encoded. In the Goal Structuring Notation (GSN) (Kelly 2004), for example, the arrows must point in the correct direction, the DAG must be acyclic, the text in a strategy parallelogram must describe an argument strategy, etc.

Evidence Availability and Sufficiency: Evidence citations in a safety case take the form of an identifier for an artifact and a claim about that artifact. The certifier should ascertain whether the given identifier is sufficient to uniquely identify a particular version of a particular artifact from among the project's many artifacts, whether the identified artifact is available in the safety case, and whether it has the claimed property. Determining whether an artifact has the claimed property might be difficult for many reasons. The process of verifying the claim might be time-consuming. Judging claims might require expertise that a certifier cannot be expected to have. If the certifier cannot reasonably ascertain the truth or falsity of an evidence claim, he or she should instead challenge the applicant to replace the evidence with a sub-argument supported by simpler evidence.

Unnecessary Argument Elements: A safety argument might contain unnecessary elements. For example, elements that are not necessary in the final argument might be left behind as the applicant extends and revises an initial argument. Unnecessary elements should be discarded.

Assumption Necessity and Reasonableness: Assumptions are critical elements of a safety case. All assumptions need to be checked for necessity and reasonableness. If it is practical to replace an assumption with evidence and argument, the assumption is not necessary and the certifier should ask the applicant to replace it. An assumption is reasonable if it is plausible and if the certifier can construct no equally plausible argument that the assumption is false. Certifiers

should consider accident and incident experience when attempting to rebut assumptions: an assumption that proved false in practice should not be accepted in a similar context.

Freedom From Well-Known Fallacies: A fallacy is an instance of unsound or invalid reasoning. Because fallacious reasoning in a safety argument can lead to belief in a false proposition, fallacies must be removed from safety arguments. A taxonomy of well-known fallacies has been developed for safety arguments (Greenwell 2006). Guided by such a list, the certifier should systematically examine the selected fragment for well-known forms of fallacious reasoning.

Establishing Argument Validity

The truth of the safety claim of a fragment is established in a carefully planned interaction called a *dialectic*. A dialectic is defined as:

“The art of critical examination into the truth of an opinion; the investigation of truth by discussion.” (Oxford English Dictionary, 2010)

The principle that drives the dialectic is the development of *challenges* by the certifier in a systematic and comprehensive examination of the safety argument. Each challenge that the certifier develops is submitted to the applicant, and the certifier and applicant then engage in a debate to resolve the disagreement. The goal of the interchange is for the applicant and the certifier to arrive at an agreed upon truth about the claim that the safety argument makes.

The structure of the dialectic is that, for each fragment, the certifier poses a sequence of one or more challenges to the applicant. The applicant then either successfully dismisses the challenge or accepts that the argument is insufficient. In the latter case, the argument has to be modified.

The certifier’s ability to challenge different aspects of the argument will be limited by his or her understanding of what is “reasonable” given the domain, the potential consequence of failure, and the engineering technologies involved. We describe the organization and use of a collection of sources of challenge that support the certifier’s examination by spotlighting common problems and misconceptions and offering evidence in support of common challenges. Thus, the challenges that the certifier poses come from the certifier’s own experience and from a predefined set of challenges where the elements of the set are in following six different categories:

Omission of expected practice. In many areas of engineering, there are standard practices that are used routinely. Their omission might indicate either: (a) a defect in the argument; or (b) an opportunity to include additional evidence that would strengthen the argument. This category lists standard practices in all expected areas to support challenges about standard practices that might have been omitted.

Common dependence of subarguments. Independent subarguments are used in circumstances where a single argument does not provide adequate confidence. This category lists known or

expected circumstances were the possibility exists of a chain of events that could lead to both subarguments being false.

Negative experience. Events such as incidents, accidents, and development failures frequently provide new engineering insights that are not distributed widely and uniformly. This category lists the accumulated failure experience with systems of the subject type.

Unrealistic assumptions. System safety relies upon the ability of a system to cope with a variety of component failures. Failure rates of components must be modeled correctly. This category lists typical failure rates of components enabling challenges to be generated based upon the realism of the rates being used in the argument.

Inapplicability. The basis of a safety argument frequently relies upon details of circumstances, and so the circumstances must hold in order for the argument to be valid. This category lists guidance on locating such occurrences.

Inadequate strategy. A strategy being used in the fragment might not provide acceptable assurance that the conclusion follows from the premises. The strategy might be fundamentally flawed or one or more premises might be missing. This category lists the details of known, acceptable strategies and provides guidance on generating challenges to strategies.

Improper use of patterns. Patterns have emerged as a valuable asset in the creation of safety cases, but they must be used correctly in order to be effective. This category lists all approved patterns and includes details of their syntax, parameters, applicability and limitations.

Inadequate argument strength. The strength of an argument fragment is basically the confidence that the certifier has in the fragment. The degree of confidence required depends on the associated consequences of failure. This category lists: (a) a set of argument fragments known to be inadequate; (b) a set of argument fragments obtained from previous systems that were deemed inadequate in given circumstances; and (c) a systematic local review process for strength assessment based upon the likely consequences of failure should the argument fragment turn out to be false.

The overall confidence in the truth of a claim assessed by a dialectic process derives from confidence in the performance of the participants. By providing the set of challenges listed above, the argument assessment process becomes a largely guided process. All of the argument assessment process is aimed at leading the certifier and the applicant to be “As confident as reasonably practicable.”

Illustrative Examples

In order to illustrate the mechanism of certification that we have defined, we present a set of simple examples derived from a completely *hypothetical* autopilot system. We assume that the autopilot: (a) provides typical basic functionalities such as altitude hold and pitch hold; (b) is implemented as a software application running on top of a minimal real-time kernel; and (c)

operates on a redundant target computer platform. We further assume that the safety case for the autopilot has been prepared according to specific published guidelines (Eurocontrol 2006).

Property: **Terminology Understood Identically By All Readers**

Text: *“Flight angles are with reference to the primary coordinate system.”*

Issue: The notion of “primary” is not defined and likely to be interpreted differently.

Property: **Absence of Vagueness**

Text: *“The timing margin on communications delays is sufficient.”*

Issue: The statement is vague. “Sufficient” is subject to individual interpretation.

Property: **Document Syntactic Validity**

Syntax: The autopilot safety case contains an argument that is recorded in the variant of GSN defined by the Eurocontrol safety-case standard.

Issue: The Eurocontrol standard utilizes elements that are no longer legal GSN, and GSN is specified as the required syntactic standard.

Property: **Freedom From Well-Known Fallacies**

Text: *“Software will not raise exceptions during operation because no exceptions were raised by software during testing.”*

Issue: This is an argument from ignorance. Failure to observe a problem does not mean that it does not exist; it means that the problem does not arise *in the tested cases*.

Property: **Omission of expected practice**

Text: *“The autopilot software configuration was managed manually using a text-based change log.”*

Issue: Change logs are typically part of a comprehensive, tool-supported configuration management plan. Such an approach is expected in order to provide confidence that configurations are managed properly.

Property: **Negative experience**

Experience: *“An in-service autopilot that was previously certified failed and jeopardized the safety of the aircraft when the altitude exceeded flight level 340 following a sensor failure on the inertial measurement unit.”*

Issue: This experience seems to contradict the safety argument’s autopilot performance claims. While the incident might be a statistical anomaly (i.e. the notional 1 in 10^x), it should prompt especial scrutiny of the support for those claims. Do the incident investigation’s findings rebut any portion of that supporting argument?

Property: **Unrealistic assumptions**

Text: *“The target computers upon which the autopilot will operate have a failure rate of less than 10^{-8} per hour, and so failure of the target hardware need not be considered in the safety case.”*

Issue: This claimed failure rate is implausible. While implausibility is not impossibility, the assumption should be rejected and evidence demanded in its place.

Property: **Inapplicability**
Text: *“The autopilot will not be operated below flight level 100.”*
Issue: The details of this element of the context need to be documented including: (a) a check that the statement is correct; (b) a detailed and explicit version of the statement must be included in the context; and (c) a comprehensive check of the safety argument needs to be conducted to make sure that restriction to operation above flight level 100 is included systematically in the argument’s claims.

Property: **Inadequate strategy**
Text: *“The real-time deadline for computation of the control surface updates will be met because the software timing has been assessed by comprehensive testing.”*
Issue: Reliance on a limited form of testing is not an adequate strategy for concluding that crucial timing conditions are met. Minimally, a crucial claim such as this needs a strategy based upon a variety of forms of evidence possibly including analysis of the underlying schedule, model checking of the temporal structure of the design, and WCET analysis of the generated code.

Conclusion

Without testable definitions of terms such as “compelling”, terms that are required attributes of safety cases, neither applicants nor certifiers know exactly what constitutes an acceptable safety case for purposes of certification. In informal everyday use, such terms are highly subjective; an argument that is compelling to one observer might not be to another. This subjectivity has little effect in areas such as politics but needs to be eliminated to the extent possible in system safety engineering.

In this paper, we have presented a comprehensive approach to certification that provides an operational definition of the various required attributes which are otherwise undefined. Satisfactory completion of the certification process implies that the associated safety case has those attributes. This operational definition is testable and provides both certifier and applicant with practical engineering goals.

References

1. Civil Aviation Authority. 2010. CAP670: Air Traffic Services Safety Requirements. February.
2. Eurocontrol. 2006. DAP/SSH/091: Safety Case Development Manual. October. <http://www.eurocontrol.int/cascade/gallery/content/public/documents/safetycasedevmanual.pdf>.
3. Greenwell, W. S., J. C. Knight, C. M. Holloway, and J. Pease. 2006. A Taxonomy of Fallacies in System Safety Arguments. Paper presented at the 24th International System Safety Conference, August, in Albuquerque, NM, USA.
4. Haddon-Cave, C., Q.C. 2009. *The Nimrod Review: An Independent Review Into The Broader Issues Surrounding The Loss Of The RAF Nimrod MR2 Aircraft XV230 In Afghanistan In 2006*. London: The Stationery Office.
5. Kelly, T. 1998. *Arguing Safety — A Systematic Approach to Managing Safety Cases*. PhD diss., University of York.

6. Kelly, T. 2004. A systematic approach to safety case management. Paper presented at the Society for Automotive Engineers 2004 World Congress in Detroit, Michigan, USA.
7. Kelly, T. 2007. Reviewing Assurance Arguments — A Step-By-Step Approach. Paper presented at the Workshop on Assurance Cases for Security — The Metrics Challenge at the International Conference on Dependable Systems and Networks, Edinburgh, U.K.
8. Knight, J. and E. Meyers. 1993. An improved inspection technique. *Communications of the ACM* 36, no. 11: 51–61.
9. U.K. Ministry of Defence. 2007. Defence Standard 00-56: Safety Management Requirements for Defence Systems, Part 1: Requirements.
10. U.K. Ministry of Defence. 2007. Defence Standard 00-56: Safety Management Requirements for Defence Systems, Part 2: Guidance on Establishing a Means of Complying with Part 1.
11. Oxford English Dictionary. 2010. <http://www.oed.com>.
12. RTCA Inc. 1992. DO-178B, Software Considerations in Airborne Systems and Equipment Certification, Washington DC.
13. Wasson, K. 2006. CLEAR Requirements: Improving Validity Using Cognitive Linguistic Elicitation and Representation. PhD diss., University of Virginia.

Biography

Patrick Graydon, M.S., University of Virginia, Department of Computer Science, 151 Engineer's Way, Box 400740, Charlottesville, VA 22904-4740, USA, telephone – (434) 249-2379, facsimile – (434) 982-2214, e-mail – graydon@cs.virginia.edu.

Patrick Graydon is a Ph.D. candidate in the Computer Science Department at the University of Virginia. His research interests include software dependability and the rigorous use of arguments in software development and system safety.

John Knight, Ph.D., University of Virginia, Department of Computer Science, 151 Engineer's Way, Box 400740, Charlottesville, VA 22904-4740, USA, telephone – (434) 982-2216, facsimile – (434) 982-2214, e-mail – knight@cs.virginia.edu.

John Knight is a professor of computer science at the University of Virginia. He holds a B.Sc. (Hons) in Mathematics from the Imperial College of Science and Technology (London) and a Ph.D. in Computer Science from the University of Newcastle upon Tyne. Prior to joining the University of Virginia in 1981, he was with NASA's Langley Research Center.

Mitchell Green, Ph.D., University of Virginia, Department of Philosophy, 120 Cocke Hall, P.O. Box 400780, Charlottesville, VA 22904-4780, USA, telephone – (434) 924-6922, facsimile – (434) 924-6927, e-mail – msg2m@virginia.edu.

Mitchell Green is the Horace W. Goldsmith Distinguished Teaching Professor In Humanities in the Department of Philosophy at the University of Virginia. Mitch Green specializes in philosophy of language, philosophy of mind, metaphysics, and aesthetics. He is also interested in decision theory and the theory of action. His current research interests include the relation between semantics and pragmatics, speech acts and their role in conversation, self-expression, self-knowledge, and attitude ascription.