Safety Standards – a New Approach

John Knight

University of Virginia

Charlottesville, VA USA

Abstract Safety standards provide great value, but despite their benefits, standards and the culture that goes with them have a variety of weaknesses. In this paper, I review these various weaknesses and propose a new approach that defines a technical structure for standards based on desired properties of conformant artifacts, a development and maintenance process designed to ensure technical quality, and a funding model designed to make standards both freely available and revenue generators for their developers.

1 Introduction

Underlying much of safety engineering is a collection of standards. Standards provide immense value. There is no question that the various standards that are used in the safety-engineering field have provided the safety community, and thereby the public at large, with safer systems at lower cost.

Despite their demonstrated value, standards have weaknesses, for example they tend to trail behind the technical state of the art. They can also do harm, for example in some cases, standards imply a level of quality in conforming artifacts that is not necessarily present, and standards can lead to excessive resource consumption in achieving conformance.

A careful examination of standards themselves, their development, and their uses reveals a number of technical and procedural issues, and I claim that the time has come for standards to be held to a higher standard, a much higher standard. In this paper, I argue that a new approach is needed to all aspects of the preparation, content, availability and maintenance of standards. I refer to this new approach as the Standard for Standards, or SfS.

The new approach to standards is centred on the need to structure standards so that they establish properties of the subject artifact. In this paper, I introduce the SfS and base many of the ideas on earlier work that developed the filter model of critical system certification (Steele and Knight 2014). Along with the core focus on artifact properties, I introduce a revised development process and a new financial model for standards. The SfS maintains the established benefits that standards

bring, adds additional benefits, and eliminates many of the difficulties that standards present.

2 The roles of standards

Standards play a special role in engineering, because standards are not themselves an engineering technology. In various ways, standards define the techniques and technologies that are used in safety engineering by pointing engineers in specific directions. Whether a standard is prescriptive, i.e., demands the use of certain techniques, is goal oriented, i.e., establishes goals that safety engineering has to meet, or is somewhere in between, the standard sets a direction that has to be followed if conformance is to be claimed.

Looking more closely at existing standards, we see that standards tend to play roles in three major areas of safety engineering:

Development. In many aspects of safety engineering, standards are written with the goal of ensuring that certain technical methods are used. By including expected technical methods, various properties are assumed in conforming system artifacts, at least implicitly.

Certification. By definition, if certification includes conformance with one or more standards, the associated standards help to define the certifying organization's meaning of the term certification and thereby the expectations of the organization.

Education. Standards educate those subject to the standard. Because a standard embodies the experience and opinions of its authors and, in some cases, its users, a standard plays a major role in education. Successful conformance with a standard requires an understanding of the technology, goals, and intent of the standard. Such depth and breadth are unlikely to be in the experience of all involved in safety-critical systems development, and standards itemize the many applicable techniques and technologies.

3 Issues with standards

Despite their immense value, there are issues with standards that need to be addressed. The issues are in the areas of preparation, content, meaning, maintenance and availability. Each of these areas is examined in the following subsections.

3.1 Preparation of standards

Clearly, standards influence device safety. Thus, their preparation has to be undertaken with great care. Presently, most standards-development processes strive to achieve the necessary levels of care by making public the discussions surrounding the creation of the standard. In essence, participation is voluntary and open to all.

This volunteer/open-to-all approach is attractive, but the approach is flawed in several ways:

- There is no guarantee than the group creating a safety standard will have the necessary technical skills. Safety expertise is highly specialized, diverse, and, in some essential areas, held by relatively few. Standards bodies could seek out specialists, but there is neither a requirement to do so nor any substantial evidence that they do.
- Participation is restricted to those who can volunteer where 'volunteer' actually means that a participant has the time and resources to participate. Inevitably, the effect is that industry representatives dominate many standards development groups. Since, by definition, an industry representative represents his or her employer, that representative is likely to be influenced by the interests of his or her employer.
- Standards are prepared by discussion and inspection but not assessed through practical application before being published. With a document as complex as a safety standard that is stated in natural language, what are the chances that there will be elements that are incomplete, ambiguous, inconsistent or otherwise inadequate? The problem is amply illustrated by the explanatory documents that have evolved for many standards.

These flaws can easily detract from the value of a standard. Some standards are presented to the community in draft form in order to elicit comment from a wide audience. The generality of the elicitation inevitably means that the comment process is not necessarily complete and is certainly ad hoc.

3.2 Content of standards

There is an expectation by the community that standards will embody the best available technology and that their presentation will allow determination of conformance to be fairly straightforward. A criticism that is seldom heard is that some standards are, in fact, technically flawed and poorly presented. Here are some examples of the problem:

> **From IEC 61508:** 'This International Standard sets requirements for the avoidance and control of systematic faults, which are based on experience and judgment from practical experience gained in industry. Even though the probability of occurrence of systematic failures cannot in general be quantified the standard does, however, allow a claim to be

> made, for a specified safety function, that the target failure measure associated with the safety function can be considered to be achieved if all the requirements in the standard have been met.' (IEC 2010)

This statement says that required probabilities of failure, even in the ultra-dependable range (for example the region of 10^{-9} failures per hour of operation) in the presence of systematic (design) faults can be assumed provided the procedural elements of the standard are followed. There is no scientific basis for this assumption.

> **From IEC 61508:** 'The appropriateness of competence shall be considered in relation to the particular application, taking into account all relevant factors including safety engineering knowledge appropriate to the technology." (IEC 2010)

This statement appears in the part of the standard addressing the competence of personnel. There is no definition of 'appropriate', and so the meaning of 'appropriate' (and hence conformance with the standard in this area) is left to chance.

> **From RTCA DO-178B:** 'The basic principle is to choose requirements development and design methods, tools, and programming languages that limit the opportunity for introducing errors, and verification methods that ensure that errors introduced are detected.' (RTCA 1992)

This statement appears in the section of the standard that relates to software life-cycle planning. DO-178B is a software assurance standard not a safety standard, and so the focus of the standard is on software technology. The intent of the quoted statement is quite clear and seems to dictate the use of programming languages that can support the best possible software fault detection. The implication is that languages with weak type systems and confusing syntax, such as C, would be rejected in favour of languages more capable of supporting software assurance, such as Ada. In practice, the quoted statement is rarely followed in the development of systems with the result that safety-critical systems judged to be in conformance with DO-178B are often written in C.

> **From Defence Standard 00-56:** '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." (MoD 2007)

This statement is at the core of the standard and defines the expected material for submission of an application for conformance. Surprisingly, the standard does not define the terms 'compelling', 'comprehensible' and 'valid'. Thus, conformance is subject to local definitions of these terms, and, as a result, determination of conformance is likely to be inconsistent. An attempt to define these terms can be found in the work of Graydon et al. (Graydon et al. 2010).

3.3 The meaning of conformance to standards

At the heart of the technical difficulties with standards is the following question:

> If an entity is determined to be in conformance with a standard, what can be said about properties that the entity possesses as a result of conformance?

Answering this question is problematic, because standards, including safety standards, usually do not address the goal of establishing precise, rigorous properties of the subject artifact directly. Rather, standards require various process, procedure, or process goal activities that tend only to imply properties of the subject artifact.

As an example of the problem, consider the following example from RTCA DO-178B (RTCA 1992). Objective 2 in Table A3 of Annex A is defined as: 'High-level requirements are accurate and consistent'. The objective references the following from the body of the standard:

> '**6.3.1b. Accuracy and consistency:** The objective is to ensure that each high-level requirement is accurate, unambiguous and sufficiently detailed and that the requirements do not conflict with each other.'

Nothing further is presented about exactly what can be expected from conformance with this objective. In practice, experts with experience using the standard, both developers and experts licensed by the FAA, provide a great deal of interpretation of such statements.

A significant consequence of the issues raised by the question above arises with safety cases. The reason that safety cases have to distinguish between direct and indirect evidence is precisely because of this issue. Direct evidence is evidence about the artifact itself. Indirect evidence is evidence about the development process (likely in conformance with one or more standards). Indirect evidence provides less confidence in claims based on that evidence because the evidence is not directly about the artifact.

3.4 Maintenance of standards

A common criticism is that standards limit innovation. To the extent that standards reflect the state of the art when written, they actually encourage innovation. In addition, their role in educating the community that has access to them likely introduces newer technology.

But standards are expensive to produce, and standards development organizations, such as RTCA, the IEC, and the IEEE, are understandably reluctant to change them. Once in place, standards tend to be static and so tend to drop behind the state of the art. For example, RTCA DO-178B was published in 1992, and it was in effect for 20 years before DO-178C replaced it in 2012. Although standards like DO-178B try to include provision for the introduction of new technology as time passes, the community tends to reject innovation because of the likelihood of certification difficulty.

3.5 Availability of standards

In principle, standards are freely available. In practice, the cost of obtaining a copy of many standards limits their availability substantially. Some standards documents are available for download for no charge (e.g., U.K. Defence Standard 00-56 (MoD 2007) and U.S. Department of Defense Standard 882E (DoD 2013)), but the majority of standards are available for download only after a purchase price has been paid. IEC 61508 Edition 2, for example, costs $2,743 from the International Electrotechnical Commission. A hardcopy of DO-178C costs $290 from RTCA Inc. Charging for copies of standards is a major source of revenue for standards organizations.

As has been noted by others (Jelliffe 2013), prices such as that for IEC 61508 Edition 2 limit access considerably. In practice, standards though nominally available are basically unavailable to many, e.g., educational institutions and those not required to use the standard. The price might even preclude consideration of use of the standard by development organizations that are not required to use the standard even though the developers expected to realize value from use of the standard.

The lack of availability of standards to educational institutions severely hampers the technical value of standards. Students are unlikely to be able to gain exposure to the whole field of development using safety standards under present circumstances.

Charging for individual copies of standards also has the effect of promoting the creation of pirate copies and of organizations restricting the number of copies available to their staff in order to keep the cost down.

4 A new approach

With the issues outlined in the previous section as a starting point, I propose a Standard for Standards (SfS), i.e., a new approach to all the different aspects of safety standards.

The SfS that I propose has three parts:

- a technical structure based on desired properties of conformant artifacts
- a development and maintenance process designed to ensure technical quality
- a funding model that generates revenue for standards-development organizations from those who gain significant value from standards.

Each of these parts is explored in detail in the following three subsections.

4.1 Technical structure of standards

4.1.1 Certification and standards

The role of certification is to protect the public interest. The challenge faced by any regulating agency is to discriminate between entities submitted for certification that would expose the public to an unacceptable level of risk from those that do not. This discrimination has to be as accurate as possible. The worst possible outcome would be a false positive, i.e., a decision to certify an entity that, in fact, would expose the public to an unacceptable level of risk.

The primary use of standards is within certification processes. Certifying agencies frequently require the use of standards, and establishing conformance is the most significant aspect of the certification process. Understanding the role of standards in certification is an important first step in understanding standards, and how best to develop and use them.

4.1.2 The filter model of certification

The filter model (Steele and Knight 2014) is a framework for certification that provides a structure for modelling and analysis of certification. The thesis of the filter model is that certification is fundamentally a mechanism for filtering the stream of applications received by the agency, i.e., for stopping deployment of entities that are unacceptable.

The notion of filtering leads to the question: 'What would cause an application to be filtered out?' Or, to put the question another way: 'What is going to get trapped in the filter?' The statement in the paragraph above -- *stopping deployment of entities that are unacceptable* -- is intuitively reasonable but not a statement upon which one could reasonably act.

The answer to these questions is that the filter has to trap applications with defects that would make the entity unacceptable. In the case of safety-critical software, for example, a filter needs to detect faults in the software that could lead to service failures at a rate or of a severity that is unacceptable.

In principle, this filtering could be effected by the regulating agency entirely internally. The agency could construct a filtering infrastructure that examines the entities for which certification is sought and makes an informed decision. Although possible, this approach is quite impractical for any entity of reasonable complexity. For example, a suitable filter for safety-critical software would entail comprehensive analysis of the software and all the development artifacts to establish a suitable level of freedom from faults. The resources required alone make the notion of filtering within the certifying agency quite untenable. Also significant is the fact that what the agency would be doing within its filtering infrastructure would duplicate much of what the developers did. The duplication would not be complete, but there would likely be extensive overlap.

4.1.3 Standards as filters

In practice, standards deployed throughout development organizations *are* the filter needed by certifying agencies. The filter is manifested as one or more standards and the filter, i.e., the standards, is located in the development organizations. The purpose of standards in this context is twofold:

- to shift the burden and therefore the cost of undertaking the filtering process to the development organization
- to align the filter concept desired by the certifying agency with the practices undertaken by the developers thereby ensuring that development activities match the necessary filter structure.

Standards define combinations of processes, procedures, techniques, technologies and goals. Each element of a standard is present to help filter defects that otherwise might be present in a safety-critical entity. By combining elements in various ways, standards endeavour to provide a series of filter planes, each of which filters one or more types of defect. A standard defines a filter designed to 'catch' entities that do not have the properties desired for the entities of interest.

Separately, the results of conformance assessments are the decision processes that certifying agencies require about an artifact that is submitted for certification. The existence of a standard does not ensure that the filter provided by the standard has been used properly. If the filter has been used properly as determined by an assessment of conformance, then the effects of the filter in detecting defects and thereby potentially identifying defective applications can be assumed.

An immediate consequence of the filter model is that it provides a structure for modelling and analysis of standards. Since a standard is the manifestation of at least a major part of the filter required by a certifying agency, the goals of a certification filter can be used to create the associated standard.

4.1.4 Requirements for conformance to standards

Existing standards are influenced by the filtering concept. That is why safety standards, for example, attempt to encompass best practices. The associated practices are 'best' because their application tends to limit defects. But the content of most standards is developed in what amounts to an ad hoc and informal manner with no explicit, testable technical goal. The result is that there is no assurance of completeness of the resulting filter.

By contrast, a critical part of the SfS is to define precisely the properties required for entities conforming to a standard. For the most part a property will be the absence of a class of defects. The properties will be organized as hierarchies of abstraction to facilitate analysis.

With the properties defined, the filter elements within the standard that will be used to detect instances of the defects within the properties can be elaborated. The exact form of the filter will depend on the degree of assurance that is required for

the various properties. So, for example in a software standard, a property that requires absence of real-time frame overruns might use a filter element based on testing if the software is not safety critical or a filter based on detailed worst-case-execution-time analysis combined with proof of the scheduling algorithm for safety-critical software.

As an example of this concept, again consider a standard for safety-critical software. Rather than dictate procedures or processes, the standard might include the need to establish properties of a conformant entity such as the following:

Absence of faults in the software requirements. This high level of abstraction might be refined to a list of fault classes that are known to occur in the determination of software requirements. This property might be shown by a rigorous assurance argument that the software requirements are complete, consistent and unambiguous.

Absence of faults in the software specification. This property might be shown by a proof or rigorous assurance argument that the software specification is a solution to the stated requirements.

Absence of faults in the implementation of safety requirements. This property might be shown by a proof of compliance with formally stated safety requirements.

Absence of faults that raise exceptions during execution. This high level of abstraction might be refined to a list of types of exception that are known to be possible. This property might be shown by proofs of freedom from execution-time exceptions for each of the types of exceptions listed.

The union of a set of properties such as these provides a clear statement of what can be expected of an entity that conforms to the associated standard. The filter elements designed to ensure that the entity possesses the properties are clear and the degrees of confidence warranted in the filter elements are also clear.

4.1.5 Filters as safety-critical systems

A false positive in certification, i.e., a failure of the filter, is a serious outcome. The potential severity of a filter failure leads to the idea of treating the certification process itself as a safety-critical system, i.e., the activities associated with reaching a decision about certification are treated as a safety-critical activity. With this view, incorrectly certifying a system that might subject the public to an unacceptable level of risk is an accident, i.e., a certification accident.

With that view, certification itself can be analyzed with all of the technology of safety engineering. In particular, techniques such as fault-tree analysis and FMECA can be applied directly and immediately to certification and used to reveal deficiencies in the filter. Fault-tree analysis, for example, can be applied by treating a certification false positive as a hazard and developing the associated

fault tree. The faults identified in that process are actually weaknesses in the certification process.

Since standards are the realization of certification filters, the analysis of certification as a safety-critical activity can be applied to standards. This process leads to an assessment approach for standards. Fault-tree analysis applied to a standard, for example, provides an assessment not possible by any other means.

4.2 Development of standards

Irrespective of the technical structure of standards, the key requirements in the development of standards are to achieve technical excellence and to maintain technical excellence. To meet these requirements, the SfS includes the following three elements:

- Prior to coming into effect, a standard will be subjected to review by a panel of experts acting anonymously who were not involved in the development of the standard and whose participation in the panel is funded. The role of the panel is to comment on the completeness, accuracy, presentation, and appropriateness of the standard. Clearly, the deficiencies identified by the panel should be corrected before the standard is published.
- Prior to coming into effect, a standard will be subjected to an empirical study where the standard is used in the development of a typical system and the use of the standard is monitored. This empirical study could be based on a commercial development that would produce a product that would ultimately be considered conformant with the new standard. Again, the deficiencies identified by the empirical study should be corrected before the standard is published.
- After coming into effect, a standard will be subject to review and update after a specified interval. A typical interval might be five years. Some existing standards are subject to periodic review but by no means all. A standard might not need review after several years of use, but the most likely circumstances are that: (a) technological advances will necessitate change, and (b) experience with the standard will lead to insights that could improve the standard substantially.

4.3 Financing standards

To finance development, distribution, and maintenance of standards, the SfS includes a new funding model. This funding model would apply to any standard that is: (a) produced by volunteer group for public use; and (b) integral to a certification process that protects the public interest. The funding model is:

- The standard will be made available for the cost of reproduction no matter whom the publisher and purchaser are. For paper copies of standards that are delivered by regular mail, the cost would be the printing and mailing costs only. For electronic copies that are delivered via the Internet, the cost would be zero.
- A fee would be paid to the organization that developed the standard by any organization, public or private, making a claim of conformance with the standard for any artifact. Such a claim would arise with an application for certification or when offering an artifact for sale. Certifying authorities would not accept an application that claimed conformance with the standard without documented evidence that the requisite fee had been paid. The fee would be made up of two parts: (a) an initial fixed fee that gave the right to claim conformance; and (b) a fee for each instance of an artifact deployed that was established as a result of the claim of conformance.

The funding model is perhaps the most important and certainly the most radical aspect of the SfS. Nevertheless, the concept is quite simple – the value returned to the publisher of the standard is directly related to the commercial value of the standard to the users of the standard.

An electronic copy of a safety standard downloaded from a publisher's web site has a value that lies mainly in the reduction in risk that the copy yields. In many cases this is zero, because the person using the standard is not directly involved in system development. Despite the fact that the value is highly variable, difficult (actually impossible) to ascertain, and often zero, at present an electronic copy of almost all standards has a fixed cost that has to be paid to the publisher.

However, if a device is created using one or more standards, then part of that device's value results directly from being able to claim conformance with those standards. The value of a standard to the developer and to the community at large in that case is explicit and clear. Charging for claiming conformance as proposed in the SfS is a rational way to associate value with a standard where the value of the standard is realized when the standard supports a commercial activity.

To see the merit of this funding model, one only has to consider the 'value' of a copy of a standard that is never actually used to develop safety-critical systems, such as in an educational context. Charging for copies of the standard would be the normal practice despite the fact that the standard actually has value mostly in areas that are not immediately commercial, such as training.

5 Conclusion

Standards in the field of safety engineering are both essential and valuable. But the present approach to their creation and management leaves a lot to be desired.

I have proposed a new approach to the preparation, content, meaning, maintenance and availability of standards designed to enhance the value of standards and

to make their value explicit. Finally, I repeat the claim that the time has come for standards to be held to a higher standard, and I propose the SfS as a way to achieve that.

References

Graydon P, Knight J, Green M (2010) Certification and safety cases. International System Safety Conference, Minneapolis, MN

IEC (2010) IEC 61508 Functional safety of electrical/electronic/programmable electronic safety-related systems. International Electrotechnical Commission

Jelliffe R (2007) Where to get ISO standards on the internet free. http://www.oreillynet.com/xml/blog/2007/08/where_to_get_iso_standards_on.html. Accessed 24 October 2013

RTCA (1992) RTCA/DO-178B/ED-12B Software considerations in airborne systems and equipment. Federal Aviation Administration software standard, RTCA Inc.

Steele P, Knight J (2014) Analysis of critical system certification. HASE 2014: 15th IEEE International Symposium on High Assurance Systems Engineering, Miami FL

DoD (2013) MIL STD-882E Standard practice, system safety. U.S. Department of Defense

MoD (2007) Defence Standard 00-56 Safety management requirements for defence systems. U.K. Ministry of Defence