

Safety Argumentation for a Nuclear Reactor Protection System – an Assessor’s View

Xueli Gao

Safety and Risk Department, Institute for Energy Technology, Norway. E-mail: xueli.gao@ife.no

Peter Karpati

Safety and Risk Department, Institute for Energy Technology, Norway. E-mail: peter.karpati@ife.no

Bjørn Axel Gran

Digital System, Institute for Energy Technology, Norway. E-mail: bjorn.axel.gran@ife.no

Alan Wassying

McMaster Centre for Software Certification, McMaster University, Canada. E-mail: wassying@mcmaster.ca

Structured safety argumentation has several advantages over safety demonstrations provided through a free text form. However, there are few publicly available examples of broadly accepted safety assurance cases with sufficient detail to demonstrate best practice. Furthermore, they usually reflect the system developers’ viewpoint. This paper presents simplified extracts of a safety assurance case from a case study that uses an assessor’s viewpoint to structure the argument. The case study is based on relevant sections of US Nuclear Regulatory Commission regulation. The argument is partial and focuses on the conceptual design level of the “trip” safety function allocated to the Reactor Protection System of a nuclear power plant. Reflections and general observations from the discussion with an expert assessor aim to support readers with practical considerations for similar safety assurance cases.

Keywords: safety demonstration, structured safety argumentation, safety assurance case.

1. Background

One activity of the Halden Human-Technology-Organisation project, OECD NEA (nd), is to perform a case study focusing on what contributes to a scientifically and logically sound structured safety argument, with illustrative examples from the nuclear power field. This is expected to be published in 2023. The first three authors of this paper are leading that case study, which has the working title “HTOR-027: Safety argumentation case study on APR1400’s Reactor Protection System interactions – regulator’s view”. The regulator’s view was supported by an expert assessor familiar with the relevant regulations.^a The motivation for the case study, extracts from it, and principles and observations made through discussion with the expert assessor are the focus of this paper.

Structured safety argumentation has several

advantages over safety demonstrations provided through free text forms, which often contain implicit, interpretation-dependent reasoning about safety. These advantages include greater flexibility through greater support for performance based approaches, and better identification of ambiguities and missing information. As agreed by safety assurance and licensing specialists in the nuclear field, applying structured, scientifically and logically sound reasoning in safety demonstration reduces regulatory uncertainty, see Hauge et al. (2014) and Karpati et al. (2017). Despite increased interest and effort in this endeavour, there are many open questions on the application of structured argumentation for safety demonstration. Different domains, organizations and disciplines seem to be at different levels of maturity of its application. Publicly available, sufficiently complex examples are extremely difficult to find, as is state-of-the-practice, step-by-step guidance. Surprisingly, the assessor’s view of safety assurance is hardly ever documented in the form of

^aThis paper contains personal opinions and viewpoints of the assessor and does not represent any official position of the U.S. Nuclear Regulatory Commission.

safety argumentation. A starting point is to consider what an assessor needs in order to conduct an effective safety evaluation of a design certification application. There are various options in this regard, and we present one of those – safety assurance cases (ACs) that can be used to evaluate ACs submitted for review. In this option we consider at least two ACs. The first AC (not shown in this paper) is a safety argument submitted to an assessor. The second AC is developed by the assessor to help evaluate the submitted AC. It facilitates and improves the assessor’s evaluation of the submitted ACs. In reality, the assessor may develop multiple ACs to evaluate submissions. We will demonstrate parts of one such AC.

Section 2 briefly introduces the case study. Section 3 presents extracts of the evaluation focused AC. The extracts are chosen to walk readers through the development of the AC. This is followed, in Section 4, by our observations regarding the approach. Finally, our conclusions and a look into the future are provided in Section 5.

2. Case Study

The scope of the case study covers safety functions allocated to the Reactor Protection System (RPS) of a nuclear power plant (see KEPCO and KHNP (2014, 2013a,b)). Basically, the RPS is responsible for initiating an emergency shutdown of the reactor when required. The case study AC is based on application documents submitted to the US Nuclear Regulatory Commission (NRC), KEPCO and KHNP (nd). Although this material was submitted to the US NRC, the case study is not constrained by US NRC regulations, NRC Library (nd). The purpose of the case study is to broaden the repository of examples and share lessons learned about building and evaluating safety ACs. The main components implementing the “trip” safety function, are illustrated in Fig. 1.

3. An Assessor’s Safety Assurance Case

This section explains the main structure of the assessor’s argument. We used the Claim-Argument-Evidence (CAE) notation Bishop and Bloomfield (2000); Adelaar (nd) to document the AC. CAE is a graphical notation that describes the argu-

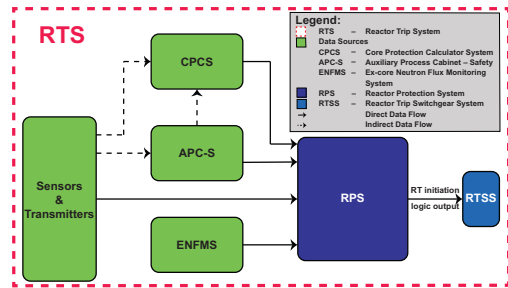


Fig. 1. Components of the Reactor Trip System

ment in a tree structure, starting with a top-claim, supported by sub-claims and eventually evidence from system artifacts. The argument fragments shown illustrate only a fraction of the actual argument created in the case study. Further, nodes containing contextual information for the individual argument nodes are not presented.

3.1. A top-claim in an assessor’s assurance case

The top-claim in the assessor’s AC may be part of a more complete safety argument. It is chosen by the assessors to represent an aspect that they feel deserves particular attention (e.g., risk-informed and/or safety focused). It also helps an assessor evaluate the soundness of the argument in a submitted AC. The top-claim, *Claim-1*, “Interactions of the RPS with its operational environment will not degrade the performance of the ‘trip’ safety function”, is an example of what an experienced assessor will consider while reviewing a submitted AC for a nuclear protection system. This claim will not be the top-claim in any submission. For reasons discussed in Section 4, it is possible that it is not even one of the sub-claims that supports the top-claim in an application. This particular top-claim was defined by the authors and reflects the way assessors think and what they will want to check in their reviews. This was confirmed by the expert assessor. Note that identified interactions between the RPS and its operational environment are analyzed in the case study but not included in this paper.

The scope of *Claim-1* was actually too broad for the case study. It was therefore refined through

additional constraints, resulting in *Claim-2* and then *Claim-3*, the text of which is “No credible failures of the systems which provide direct signal inputs to RPS will lead to a ‘missed trip’ caused by the RPS’s system design features”. This is the top-claim actually used in the case study. It focuses on “failure”, which is central to one regulatory requirement.

In the rest of the paper the wording of sub-claims is somewhat simplified for readability by repeating only the relevant parts of their parent claim. This is not intended to change the meaning of the sub-claim.

3.2. Supporting Claim-3

When a claim is an obvious combination of *n* components, it is usual to decompose the claim into *n* sub-claims. Thus, *Claim-3* is decomposed into the conjunction of *Claim-4* and *Claim-5* as shown in Fig. 2. We further develop the sub-argument under *Claim-4* to demonstrate that the system design and involved equipment protects the health and safety of the public. The sub-argument under *Claim-5* (not developed here) demonstrates that the equipment is operated and maintained such that it is always capable of performing its design function.

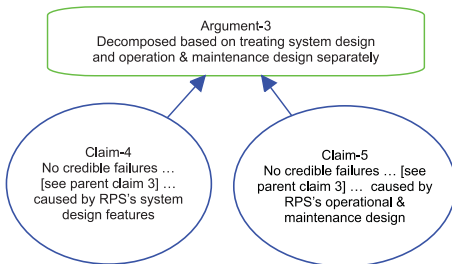


Fig. 2. Decomposition based on separating system design from operational & maintenance design

3.3. Using regulatory defined system failure types to support Claim-4

Claim-4 is decomposed through *Argument-4* (see Fig. 3) based on the failure types described in the relevant regulation.

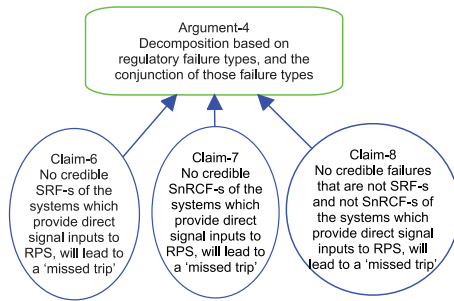


Fig. 3. Decomposition of the system design claim

US NRC’s GDC (General Design Criteria) Appendix A considers two main categories of failures: (1) Single Random Failures (SRF) and (2) Systematic, non-Random, Concurrent Failures (SnRCF). SRF-s are considered in *Claim-6*, SnRCF-s in *Claim-7*, and all other types of failures in *Claim-8*. GDC Appendix A states: “the development of these General Design Criteria is not yet complete”, and so this decomposition is not complete either, but considered to be sufficiently complete on a state-of-the-practice basis.

3.3.1. Reflections on characterizing failures

As mentioned, *Argument-3* was easy to construct. It simply separated system design from system operation & maintenance. Systems are always designed for a particular manner of operation and maintenance. The system has to be safe at time of deployment, but also for (possibly) many years of use. Safety of any system is not absolute. Its safety in the future depends on maintenance plans and how the system design prepared for changes in the future. Its safety also depends on operating procedures and how likely it is that operators comply with those procedures. Therefore, an application must explicitly include how the system design relates to planned operation and maintenance. Representing this argument is often difficult in a tree-based visualization, because the links between the design-argument and the operation-related argument are difficult to document without making the argument overly complex. Thus, *Claim-4* and *Claim-5* look simple at this level, but have cross-cutting concerns lower down in the arguments.

US NRC’s regulations include certain specific prescriptions for design techniques; one such prescription is the ‘single failure criterion’. Generally, different design techniques implicitly imply different claims and thus they require different arguments and evidence. Random failures like a pipe break or a valve failure cannot be prevented; they are considered credible. However, different parts of equipment will break or wear out independently. Redundancy and independence are included in the protection system design to mitigate, as much as possible, against single random failures causing loss of protection. Redundancy and independence do not protect the system against all systematic failures including earthquake, fire, hot, wet environments of LOCA (Loss of Coolant Accident), and design errors. For example, all the containment transmitters are subject to the same hot, wet environment during a LOCA. That is a potentially systematic cause of failure, therefore all the equipment is qualified to work in that environment. The equipment qualification criteria should exceed the projected worst case environmental limits in which the equipment is operated (i.e., margin to address uncertainties), as one way of addressing that systematic failure.

For both failure categories (SRF and SnRCF), the question is why a particular set of techniques is adequate to address them. The set of techniques deemed adequate depends on the specifics of the situation. What is good enough in a certain situation is typically an engineering judgment based on industry best practices. However, this complicates the assessment process, especially for applicants. With regard to assessment there are three possible outcomes: (1) the reviewer disagrees with the applicant, (2) the reviewer needs more explanation or information for further consideration, and (3) the reviewer agrees with the applicant. Prior agreement on common critical situations facilitates efficient and effective reviews. Rigorous engineering analysis of these situations can be used to guide development of AC patterns/templates acceptable to both assessor and applicant.

Although human errors are out of scope for this part of the argument, it is good to remember that they could also fall into both categories

of single and systematic failures. For example, a mistake made once leading to a failure versus a mistake made regularly, (e.g., the same technician incorrectly calibrates all the equipment). A systematic mistake may affect redundant components and lead to a degraded performance of the related safety functions.

We noted that the development of types of failures in the GDC is not complete. This does not relieve an applicant from additional considerations that were omitted in the current GDC. The argument in a safety AC must deal with many forms of incompleteness, and this is just one of them. That is why a claim related to “other failures” was included in *Argument-4*.

3.4. Single random failures (Claim-6)

Claim-6 is refined into *Claim-9* (see Fig. 4) through generic reasoning where mitigating such a failure requires the failures be identified, characterized, and addressed (*Argument-5*). Then *Claim-9* is decomposed through *Argument-6* into the three sub-claims, *Claim-10* (identification), *Claim-11* (characterization) and *Claim-12* (addressing).

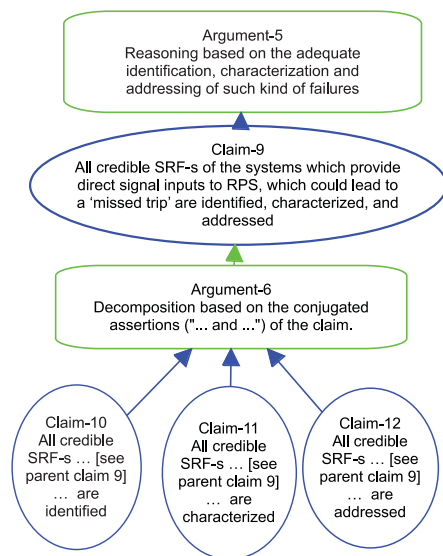


Fig. 4. Sub-argument of the SRF related claim

Each of those terminal claims are underpinned by two kinds of evidence. One kind refers to the selected approach, (method, technique), and explains why it is adequate (e.g., generally accepted state-of-practice), why it fits the purpose and shows a sufficiently complete description of how the approach can be applied correctly. The other kind of evidence refers to the documentation of how the selected approach was applied, also showing that the approach was used as expected, and that the outcome is also as expected. Checking the completeness of the outcome and validating the results can be done e.g., through independent expert review, or by applying an alternative, relevant state-of-the-practice approach with an independent expert team. In this case, a Failure Mode and Effects Analysis (FMEA) was used KEPCO and KHNP (2014), and the FMEA table's columns served as the documentation of its application (*identification*: “No.”, “Names” and “Failure Mode”; *characterization*: “Cause”, “Symptoms and Local Effects Including Dependent Failures”, and “Method of Detection”; *addressing*: “Inherent Compensating Provision”, “Effect on PPS”, and “Remarks and Other Effects”). FMEA is a generally accepted approach for identifying single random failures, when executed appropriately with qualified engineers.

3.4.1. Reflections on dealing with single random failures

When *Claim-9* is decomposed, a hidden assumption is that the state-of-the-practice is mature enough and is an effective means to achieve the purpose of the claim. A common situation when any of the three steps (*Claim-10*, -11 or -12) might fail, is when the envelope of state-of-the-practice is pushed, i.e., there is even a small change in a relevant feature of a new AC compared to an earlier AC in which best practice was used/developed. An example of failed characterization is the Fukushima nuclear disaster, NRC (nda). The hazard of a tsunami was known, and a seawall was prepared to address it. However, the characterization of the worst-case tsunami (how high it can be) failed since Fukushima experienced the most powerful earthquake ever recorded

in Japan which triggered an unexpectedly high tsunami with 13–14-meter waves.

Consideration about the propagation of the effects of a SRF from one redundancy to another belongs to the systematic failures related sub-argument. A usual way to analyze such a situation is a conserving bounding analysis of limiting events (in contrast to detailed analysis of all events). This approach was historically applied for simple, redundant, independent silo systems. However, it was wrong in some cases, such as in the safety analysis of the Three Mile Island Unit 2 reactor: the small break LOCA was thought to be less limiting than a large break LOCA, but turned out to produce worse results, NRC (ndb). Implicit assumptions and human errors also played a role in this accident.

In our case, a generally accepted approach, FMEA, was applied. However, this may not always happen. If the applied approach is relatively new without general acceptance yet, then its adequacy and reliability should be demonstrated in the safety argument in addition to the already mentioned features. Beside the focus on SRFs, FMEA determines the required surveillance requirements of the failures and identifies whether they are self-revealing or not. For the non-self-revealing failures, surveillance test and checks are determined. Even if the FMEA results are imperfect, the fallback position of the reasoning about the safety I&C systems is that any remaining SRFs (non-self-revealing and not addressed by surveillance requirements) can be tolerated concurrently with a single random failure because of the redundant architecture with independent divisions (as required by US NRC regulations). Whatever hazard analysis is used, the goal always is to reduce the space of potential remaining failures to be as small as practical.

3.5. Systematic, non-random, concurrent failures (*Claim-7*)

Claim-7 (see Fig. 3) is decomposed through *Argument-7* into *Claim-14* and *Claim-15* (see Fig. 5 based on the characterizability of anticipatable SnRCFs). The non-anticipatable failures belong to the sub-argument under *Claim-8*. Char-

acterizability of a failure is defined as the potential to describe how a system behaves under some anticipated failure conditions so that the related feature(s) can be directly addressed by engineering methods. For example, a software bug is anticipatable but often not characterizable, i.e., one cannot say how the system running the software will behave.

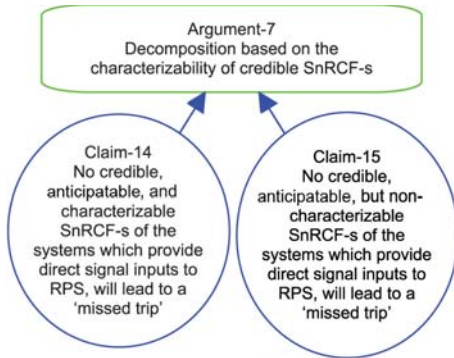


Fig. 5. Decomposition of the SnRCFs related claim into characterizable and non-characterizable SnRCFs

The sub-argument under *Claim-14* develops similarly to the sub-argument under *Claim-6*: (1) first *Claim-14* is refined based on adequate identification, analysis and addressing of such failures, then (2) that complex claim is decomposed into its conjugated three atomic claims, and finally (3) each atomic claim should be underpinned by two kinds of evidence referring to the documentation of the selected approach and how it was applied. The sub-argument under *Claim-15* also develops similarly, with the exception that characterization is not possible and therefore the means used to address the identified non-characterizable SnRCFs are indirect. This and the previous sub-arguments are not presented here.

3.5.1. Reflections on dealing with systematic non-random concurrent failures

The argument represents the SnRCF space as composed of two discrete parts where the line is drawn based on whether such a failure can be addressed directly or indirectly. However, deciding where the line is might not be simple in practice, because the failure space behaves as a con-

tinuum rather than as an aggregation of discrete parts. Characterizable failures can be addressed directly (e.g., mitigated or eliminated) by engineering methods, like fire barriers, fire suppression systems, etc. Examples of such failures are earthquake, fire, tsunami, and environmental conditions like temperature. In such cases, equipment can be designed to withstand the anticipated characteristics, while for the anticipatable but non-characterizable cases, it cannot be done. Examples of non-characterizable failures are software bugs and human and design errors. Mitigations include the use of conservative design practices and assumptions, high quality equipment, diversity, and defence-in-depth.

3.6. Remaining failures

Claim-8 covers the remaining failures which are not included in the other two parallel branches of the argument shown in Fig. 3. The reason for not including them is usually their novelty, likelihood, or underestimated severity. Since such failures are not anticipated, they cannot be addressed by direct means. Therefore, they are addressed by indirect safety measures which comply with state-of-the-practice which serve to reduce the likelihood and/or severity of consequences. Such indirect means again include conservative design practices and assumptions, high quality equipment, diversity and defence-in-depth.

3.6.1. Reflections on coverage of failures

The safety engineering team, with adequate qualifications and experience, is expected to reduce the space of the *Claim-8* failures as low as practically possible. Indirect safety measures are usually more comprehensive and not limited to specific design considerations. To avoid non-anticipatable failures, the following considerations are applied in nuclear safety: (1) do not do things which are too dangerous to do, and (2) conservatism. Regarding (2), digitization in the nuclear field is slow, and thus it lags behind the state-of-the-practice in computer science and software engineering. Therefore, nuclear practitioners end up using proven technologies and not cutting edge techniques. By the time a technique or technol-

ogy is used in nuclear power plants, it is usually proven in use and reliable. In addition, the level of complexity posed in digital systems for nuclear protection systems is (or should be) orders of magnitude less than the complexity in many other application domains.

4. Observations

The case study is at a stage where we can present relevant observations, supported by experience.

There is more than one valid structure for an AC. Different teams will probably generate different ACs even when using the same input. Therefore, principles or example templates that guide the structuring of an AC would be useful for standardization, which is need for improvements in efficiency and effectiveness. One of those principles could be to ensure robustness of the tree-structure of the ACs by pushing the more changeable argument elements lower in the argument tree. The higher level structure of the AC should be robust against future (likely) changes. This requires experience as well as effort expended in identifying likely changes.

This work has two primary focuses: 1) investigate how to develop structured safety assurance arguments in the nuclear power domain; and 2) investigate how to develop safety arguments structured based on what an assessor looks for when evaluating the safety of an application. The two focuses are complementary. We have discussed the first in some detail. The second, the assessor's view, is embedded in the discussion, but it can be made clearer with a simple example.

It is clear that there is more to showing that the RPS is adequately safe (top-claim labelled RPS-1) than is included in *Claim-1*, which is why (Section 3.1) it should never be the top-claim in an application. Generally, one cannot predict what the next level of sub-claims would be, since there are several options. For example: RPS-1.1 – show that the requirements lead to a safe RPS; RPS-1.2 – show the manufactured RPS complies with its requirements; RPS-1.3 – justify any behaviour not in the requirements. *Claim-1* combines elements of the argument that supports RPS-1.1 and the argument that supports RPS-1.2. This is useful for

assessors evaluating applications. It gives them a way of checking whether decomposed arguments with cross-cutting concerns are adequately connected. Inadequate connections are harder to find in free form text. Finally, developing a large structured AC according to an assessors' preferences is not clearly superior to multiple smaller ACs where each highlights a specific concern – much as we have done in our example.

The tripod of Process-Product-People is also of importance in safety argumentation and must be explicitly visible in the AC. This was included in the case study AC, where the selected method (“process, people”) and how it was applied (“process, product, people”) is documented.

Completeness issues are constant contributors to unsound argumentation. Dealing with completeness when refining or decomposing a claim into sub-claim(s) is not fully addressed in the presented AC due to space limitations. Efforts to mitigate incompleteness should be explicitly included in the AC. For instance, if a claim states that “all the hazards are identified”, there should be sub-claims that (1) “The hazard analysis (HA) used is such that it is likely to find all the hazards”; (2) “The HA team was qualified to do the HA”; and (3) “There was an appropriate effort expended to identify additional hazards and no additional hazards were found”.

Building an AC often uses a top-down process. The challenge is that when decomposing a claim into its sub-claim(s), we need to develop reasoning that shows the sub-claim(s) will fully support the parent claim. This must be performed at every step to justify each decomposition. This local reasoning makes many assumptions about what will be included in lower decompositions that are not yet determined. The logical argument in these tree-structured ACs is actually bottom-up. It starts with the evidence that supports terminal claims, and then progresses from sub-claims to parent claims, which in turn become sub-claims. The fact that the decomposition is performed top-down, but the logical reasoning is bottom-up, often results in confirmation bias – especially with regard to determining that specific evidence actually supports a specific terminal claim. An approach using

a bottom-up development process is outlined in Annable et al. (2022). This method generates the AC bottom-up based on specific evidence in the modelled system. The advantages/disadvantages of top-down versus bottom-up AC development is still the subject of research.

The nuclear domain could learn from other domains about how to apply the AC approach. To migrate such knowledge in an efficient and effective way, templates or frameworks could be used. These templates could have regulation specific appendices to allow mapping to particular regulatory context, (see Wassying et al. (2016)). Our proposal of an assessor focused AC specifically developed to facilitate review of submitted ACs is one step in the development of such templates.

5. Conclusion

We have presented the main ideas involved in extracts of a structured safety argument (AC) from the assessor's viewpoint about a specific safety concern in the reactor protection system of a nuclear power plant. The work is based on a detailed case study undertaken in the Halden Human-Technology-Organisation project, OECD NEA (nd). The structure of the simplified argument is explained step-by-step with additional concerns outlined at each step to help readers with practical considerations. Finally, generic observations are presented supporting further research related to this topic. There is ongoing work on extending and evaluating options in these argument(s), and in generalizing the results.

Acknowledgement

This work was funded and supported by the OECD Nuclear Energy Agency (NEA) Halden Human-Technology-Organisation (HTO) Project & Halden Reactor Project (HRP). Thanks to the members of the Halden HTO & HRP Projects, and Norbert Carte (US NRC) for review and constructive comments.

References

- Adelard (n.d.). Claims, Arguments and Evidence (CAE). <https://www.adelard.com/asce/cae>. Accessed: 2023-02-23.
- Annable, N., T. Chiang, M. Lawford, R. F. Paige, and A. Wassying (2022). Generating assurance cases using workflow+ models. In *Computer Safety, Reliability, and Security: 41st International Conference, SAFECOMP 2022, Munich, Germany, September 6–9, 2022, Proceedings*, pp. 97–110. Springer.
- Bishop, P. and R. Bloomfield (2000). A methodology for safety case development. In *Safety and Reliability*, Volume 20, pp. 34–42. Taylor & Francis.
- Hauge, A. A., P. Karpati, and V. Katta (2014, August). Summary of the 2014 Expert Workshop on Safety Demonstration and Justification of Digital Instrumentation and Control Systems in Nuclear Power Plants. Technical Report HWR-1113, OECD HRP.
- Karpati, P., V. Katta, and C. Raspotnig (2017, August). Expert Workshop on DI&C Safety Assurance with Special Focus on Experiences with Assurance Cases. Technical Report HWR-1220, OECD HRP.
- KEPCO and KHNP (2013a, September). Design Control Document TIER 2, Chapter 15, Transient and Accident Analysis, Revision 0. Technical Report APR1400-K-X-FS-13002, KEPCO.
- KEPCO and KHNP (2013b, September). Design Control Document TIER 2, Chapter 16, Technical Specification, Revision 0. Technical Report APR1400-K-X-FS-13002, KEPCO.
- KEPCO and KHNP (2014, December). Design Control Document TIER 2, Chapter 7, Instrumentation and Controls, Revision 0. Technical Report APR1400-K-X-FS-14002-NP, KEPCO.
- KEPCO and KHNP (n.d.). APR1400 Design Control Document and Environmental Report. <https://www.nrc.gov/reactors/new-reactors/large-lwr/design-cert/apr1400/dcd.html>. Accessed: 2023-03-27.
- NRC (n.d.a). Fukushima nuclear disaster. <https://www.nrc.gov/reading-rm/doc-collections/fact-sheets/japan-events.html#accident>. Accessed: 2023-05-05.
- NRC (n.d.b). Three Mile Island Accident. <https://www.nrc.gov/reading-rm/doc-collections/fact-sheets/3mile-isle.html>. Accessed: 2023-05-05.
- NRC Library (n.d.). Appendix A to Part 50—General Design Criteria for Nuclear Power Plants. <https://www.nrc.gov/reading-rm/doc-collections/cfr/part050/part050-appa.html>. Accessed: 2023-02-23.
- OECD NEA (n.d.). Halden Human Technology Organisation (HTO) Project. https://www.oecd-nea.org/jcms/pl_61937/halden-human-technology-organisation-hto-project. Accessed: 2023-03-27.
- Wassying, A., P. Joannou, M. Lawford, T. Maibaum, and N. Singh (2016). New standards for trustworthy cyber-physical systems. In *Trustworthy Cyber-Physical Systems Eng.*, pp. 337–367. CRC Press.