Interpreting ALARP

Catherine Menon*, Robin E Bloomfield [†], Tim Clement [††]

**[††]Adelard LLP, UK, [cm ¦ reb ¦ tpc]@adelard.com*

Keywords: ALARP, SoS, risk reduction

Abstract

This paper explores some of the common difficulties in interpreting the ALARP principle, and traces the potential effects of these difficulties on system risk. We introduce two categories of risk reduction approach which permit us to characterise the risk profile of a system in more detail and discuss their application to Systems of Systems (SoS).

1 Introduction

Reducing system risk So Far As Is Reasonably Practicable (resulting in risks which are As Low As Reasonably Practicable (ALARP)) is a legal requirement for all safety-related systems. However, the concept of ALARP is often interpreted in a way which does not fully satisfy this legal requirement. Furthermore, it may be desirable to specify the desired risk profile of a system in more detail than simply requiring that the system risk be reduced ALARP. In this paper we present an analysis of these issues and discuss ways in which the individual risks in a system may be balanced to achieve a particular risk profile.

Section 2 of this paper discusses existing guidance on ALARP. Section 3 illustrates some of the difficulties, and presents a mathematical formulation of ALARP to aid conceptualisation. Section 4 identifies different types of risk reduction efforts which can be used to balance individual risks in a system, while Section 5 discusses some confounding factors. Section 6 summarises our discussions on ALARP and risk transfer, and Section 7 presents conclusions and future work.

2 Existing guidance on ALARP

As the body responsible for the enforcement of health and safety legislation, the Health and Safety Executive (HSE) includes the ALARP concept as a legal requirement and provides guidance in [1] which addresses both the use of good practice and arguments for ALARP from first principles.

2.1 Health and Safety Executive

Good practice is defined as compliance with those standards for controlling risk that HSE has judged and recognised as satisfying the law in this area. Where relevant good practice of this form exists, a duty holder is *required* to base his or her ALARP argument on compliance with this. However, adequate good practice may not exist for the use of new technologies or complex systems. In this situation, the HSE provides guidance based on first principles. This approach requires the duty holder to make a qualitative or quantitative judgement to support the ALARP argument.

We note that the HSE also provide guidance on other aspects of risk management, including hazard identification, risk aggregation and societal risk assessment. However, there is no significant discussion of the interactions between different risks in a system.

2.2 SAPS and ONR guidance

The nuclear Safety Assessment Principles (SAPs) [2] and Office for Nuclear Regulation (ONR) Guidance on the Demonstration of ALARP [3] present similar guidance to that above. However, they extend this by explicitly considering the need to balance the risks within a single system. Risk transfers are discussed, with the motivating requirement being that the decrease in one risk should be measurably greater than the associated increase in other risks. They also discuss the need to balance risks over time, with short-term high risks being justified if they result in a long-term decrease in risk. Small increases in risk which are balanced by factors which make an overall improvement to health and safety issues are also permitted by this guidance provided the good practice requirement for ALARP is met. We discuss these points further in Sections 5 and 6 and provide examples of situations where their use might be advised.

3 Interpretations of ALARP

One approach to constructing ALARP arguments makes use of argument templates, particularly GSN patterns. In [4] and [5] we introduce a more formal basis for structured arguments (called Fog and sponsored by the UK and Swedish nuclear industry). This approach is based on the construction of an argument from a number of basic building blocks.

In the course of the empirical work in the project to identify requirements for the Fog “blocks” we have analysed a wide range of real cases from a number of different industries. This research has identified a common argument pattern in which the claim is made that the system risk has been reduced ALARP because the risk from each individual hazard has been reduced ALARP.

This is true if and only if ALARP is shown to be a property that distributes across the system, i.e. semi-formally

$$ALARP(X+Y+Z\ldots) = ALARP(X)+ALARP(Y)... \quad (1)$$

There is typically no justification presented for this assumption when this argument pattern is used. However, in order to make use of the basic building blocks, the Fog approach does require that justifications of assumptions are presented. That is, Fog requires us to examine the nature of the composition operator, "+", of the function denoted by "ALARP", and of the enumeration indicated by "…. ". In the examination of these, we have identified a number of circumstances in which Equation (1) does not hold. For example, situations in which:

- The identified set of hazards is incomplete, and credible vulnerabilities are ignored
- There are interdependencies between hazards which are not accounted for, leading to double counting of these sources of risk
- There are interdependencies between systems, covert channels and common mode failures which are not adequately accounted for. All of these can increase the complexity of combining risks
- Risk mitigations are not independent. For example, an identified mitigation for one risk may increase other risks, e.g. a mitigation for the risk of fire on a submarine may increase the risk of drowning.
- Costs are amortized over several hazards and cannot be justified when assessed against each of these individually. For example, the cost of static analysis tools may not be justified in terms of the resulting risk reduction for a single system, but may be justified if used over several systems
- There are limited resources subject to a threshold effect of aggregation. For example, operator attention may be a mitigation against several hazards. When these hazards present themselves simultaneously, this resource is depleted and may be inadequate as a mitigation against any single hazard.

This list of circumstances in which the ALARP property does not distribute (as shown in Equation 1) is not intended to be exhaustive. However, it is sufficiently comprehensive to identify a need to examine the validity of any assumption that ALARP decomposes in this way. Compositionality in assurance is an active research area [6], and decomposition in its various forms – modes of operation, structure, sets of hazards – is often a source of problems and imprecision in cases.

3.3 Mathematical presentation of ALARP

In this section we present an informal mathematical formulation of ALARP, intended to aid conceptualisation.

Let $R(R_1(D), ..., R_n(D))$ be the risk function of a system *S* for which the chosen design, implementation and operation methodologies are represented by *D*, and for which there are *n* identified sources of risk to a defined exposed group. The risk associated with the *i*-th source, in the presence of all other sources of risk, is therefore represented by a function $R_i(D)$.

We note that where all risks are independent, the system risk would be represented as:

$$R = R_1(D) + \ldots + R_n(D) \quad (2)$$

Let *M* be the set of all possible design, implementation and operational choices and methodologies for the system *S*. Each finite subset M_q of *M* then represents a potential set of choices we can make about how to design, implement and operate the system. In practice a number of stakeholders typically have input into these choices, and their input is guided by their experience and competency in the relevant areas. They will use this experience to identify a finite set $\{M_a, \ldots M_m\}$ of possible ways of designing, implementing and operating the system. (We note that this finite set may not include an ALARP design, but for the purposes of this formulation will assume that it does.)

Each of the implementation sets M_i in $\{M_a, \ldots M_m\}$ has an associated cost, C_i. Here, C_i represents the cost – taken from some defined starting point – of implementing the system using the set of techniques represented by M_i. It should be noted that different subsets M_q, M_p may have the same cost $C_q = C_p$.

For a given choice of implementation, design and operational techniques M_i we can then discuss the value of the risk function $R(R_1(M_i), ..., R_n(M_i))$; that is, the system risk associated with these particular implementation, design and operational choices. Similarly, we can also do this with risk $R_j(M_i)$ associated with the *j*-th source of risk. By varying the implementation choices over the set $\{M_a, \ldots M_m\}$ we can graph the system risk against the varying costs $\{C_a, \ldots C_m\}$ of implementation choices.

Figure 1 shows how the system risk and individual risks might vary with the cost of different implementation choices.

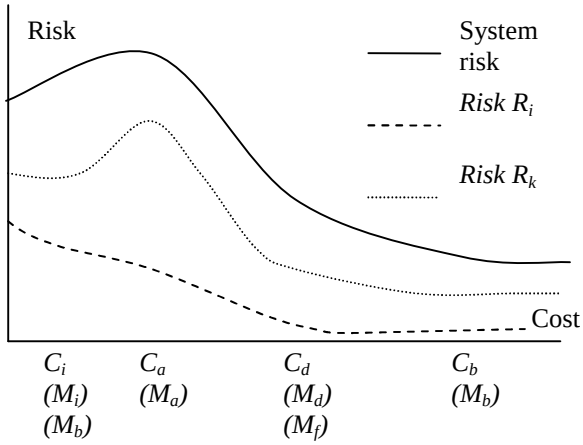

Figure 1: Risks and costs

A risk *R* (resulting from a set of design, implementation and operation choices M_p) is ALARP if the cost of any further risk reduction is grossly disproportionate to the associated

reduction in risk [7]. Using the vocabulary and assumptions above, we can rewrite this as:
A set M_p with associated system risk R will be judged ALARP if there is no other set M_q with associated system risk R' for which

$$R' < R \text{ and either}$$
$$C_q < C_p \text{ or } (R - R') / (C_p - C_q) < F \qquad (3)$$

for some constant $F<0$ which has been agreed to be the disproportionality factor for ALARP.

In fact, as we discuss in Section 5, there are confounding factors – including variations in risk over the life of the system – which mean that this characterisation of ALARP requires further study.

4 Characterising the risk profile

From the discussions in Section 3 we can infer that there may be a number of different implementations which all reduce system risk ALARP, but which do it in different ways. ALARP guidance - such as that discussed in Section 2 - typically does not offer much insight into how to make the choice between these different implementations; indeed this is arguably beyond the scope of such guidance. In this section we present a number of different risk reduction approaches such as "fairness in improvement" and "fairness in outcome" which provide alternative ways of balancing individual risks (permitting an increase in one if it corresponds to an equivalent or greater decrease in another) in order to achieve an ALARP system risk.

Without loss of generality we will use only two risks, A and B, to illustrate each of these approaches. We note that although we present two approaches in detail here, there are a potential number of ways in which to combine these approaches to obtain a desired risk profile.

4.1 Fairness in improvement

The aim of this approach is to achieve a similar absolute risk reduction for all individual risks.

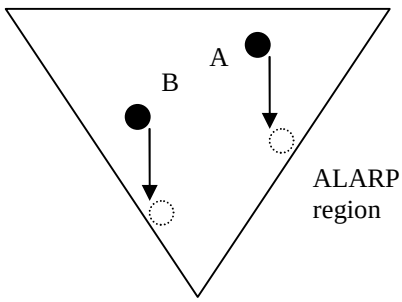

Figure 2: Fairness in improvement

Fairness in improvement means that our risk reduction attempts prioritise reducing *both* risks A and B, regardless of the relative cost of these reductions compared to each other (as long as the reductions themselves are reasonably practicable), and regardless of whether making these reductions to B means that for technical reasons further reductions cannot then be made to A. That is, where two risk reduction efforts both reduce risk to ALARP, and one of these is to implement moderate reductions for both risk B and risk A, and the other is to implement significant risk reductions for A and none for B, then a fairness in improvement approach would recommend the first choice.

Using a fairness in improvement approach can mean that no individual risk is as low as possible; that is, as low as we would have achieved by reducing that risk alone. However, this approach ensures that the risk reduction effort confers a certain minimum benefit on all system risks. Where A and B refer to the risks encountered by two different exposed groups, a fairness in improvement approach would mean that both groups affected by the system will benefit from a certain minimum risk reduction, regardless of the degree of risk they faced initially.

Fairness in improvement may be a good approach where there are justifiable reasons why it is acceptable for one risk to be higher than another; for example where it is justifiable that one group of people encounter a higher risk from the system than others. This can result from the societal acceptability of different risks. For a given system such as a nuclear plant, this is usually higher for workers than for the general public.

4.2 Fairness in outcome

The aim of this approach is to achieve a similar level of risk for all individual risks.

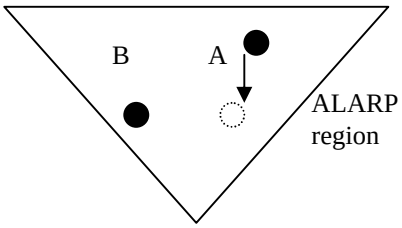

Figure 3: Fairness in outcome

Fairness in outcome means that our risk reduction attempts prioritise the more severe risk A, regardless of the relative cost of reducing risks A and B compared to each other, and regardless of whether making these reductions to A means that for technical reasons further reductions cannot be made to B. That is, where two risk reduction efforts both reduce risk to ALARP, and one of these is to implement significant reductions to the higher risk A, and the other is to implement moderate reductions to both risk A and risk B, and the other is to implement significant risk reductions for A and none for B, then a fairness in outcome approach would recommend the first.

Using a fairness in outcome approach can mean that the risk reduction efforts are concentrated on only a few risks, with no benefit for the other risks. However, this approach ensures that the areas of greatest risk are targeted by reduction efforts.

Where A and B are two different exposed groups, a fairness in outcome approach would mean that the group at highest risk is guaranteed to benefit from risk reduction, regardless of the degree of risk faced by other affected groups.

Fairness in outcome may be a good approach where one affected group takes on a disproportionate amount of the entire system risk, without any justification. It may also be useful where a few risks are particularly high.

5 Confounding factors

In this section we discuss some of the reasons why a particular risk reduction approach might be used even where the resultant risk balance does not fully satisfy the required risk profile. One fundamental reason for this may be that the required risk profile is not specified; there is no further information provided besides the legal requirement that the system risk be reduced ALARP. This can be due to a lack of vocabulary to discuss risk balancing.

More generally, it may be the case that once the system risk has been reduced ALARP, no further effort is made to determine whether there is another way of achieving this. That is, the ALARP claim is justified because there are no further reasonably practicable measures – including a complete design change – to reduce the risk further, but no effort has been made to examine and assess alternative reasonably practicable measures which reduce the risk to the *same* level. This may be due to the cost involved in assessing these further solutions (and we note that, unlike the requirement for ALARP, this search for alternatives which present the same risk is not legally required) or it may be because the alternative solutions have been searched for and not found. Some more specific confounding factors are discussed below.

5.1 Areas of influence

Where there are multiple components or subsystems within a system, the stakeholders responsible for each subsystem might be required to furnish their own safety analysis and justify their own individual ALARP claims for this subsystem. Indeed, this is a relatively common requirement when procuring complex military systems, despite the observation that we have made in Section 3 that the decomposition of ALARP in this way is not universally valid. In this situation the scope of the individual safety analyses and ALARP claims are determined by the *area of influence* of each stakeholder.

One example of this is a military warfare SoS containing two discrete subsystems: training and operating. Here, A represents the "operational" subsystem, and B the "training" subsystem. There are two possible outcomes in terms of stakeholder influence. First, the stakeholders responsible for A and B may be able to influence safety beyond the boundaries of their individual systems. That is, there may be a mechanism for them to engage in mutually-agreed risk transfer, such as agreeing to transfer part of the operational risk to the training system. This might be done by increasing the fidelity of the training and extending it to cover all foreseeable conditions. This reduces the operational risk associated with undertaking (potentially unfamiliar) manoeuvres, but clearly increases the risk associated with undergoing this training.

This risk transfer means that the system risk associated with training is no longer ALARP (as it could be reduced further by decreasing the fidelity of training, e.g. performing this in daylight hours only).

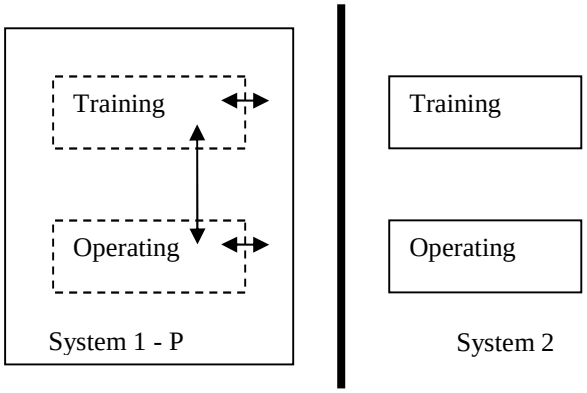

Figure 4: Two contrasting systems

The left-hand side of Figure 4 shows a situation where stakeholders can influence safety beyond the boundaries of their individual systems. The dashed lines indicate a "porous" information boundary. That is, information and risk can be transferred explicitly between these two subsystems and the wider system context, as indicated by the arrows. In this situation the stakeholders for the Training and Operating subsystems are aware of the overall personnel risk, are aware of the risks associated with the other subsystem, and have a mechanism for transferring risk between them. As indicated by the solid line around the entire system, there is also a defined wider system boundary which includes all subsystems which contribute to the personnel risk. This means that the overall personnel risk can be calculated by aggregating the (known) risks from the interactions of the subsystems.

In contrast, it is also possible that the stakeholders responsible for A and B might have no influence beyond the boundaries of their subsystems. They may have each been tasked with the requirement to justify an ALARP claim for their subsystem in isolation, which precludes the possibility of transferring risk between subsystems. Furthermore, these stakeholders may not be aware of the existence of each other, nor of the boundaries of any wider system context for which risks would also need to be assessed.

This is shown in the right-hand side of Figure 4. The solid lines around subsystems indicate that information cannot pass between them, nor can any information about the internal subsystem risks be used within a wider system context. Furthermore, there is no defined wider system boundary (no solid line around the system as in the left hand side). This

means that the overall personnel risk is not explicitly considered at any point.

In the first situation (where risks and information are visible), the risks associated with the subsystems (Training and Operating) are known and can be transferred provided the explicit system risk remains ALARP. Fairness in outcome is therefore a feasible and possibly recommended approach to use. This would allow the risk reduction efforts to be focused on the higher risk subsystem (Operating), while leaving the Training risk as is. Here, the "porous" information boundaries mean that an ALARP argument can still be made which takes the total aggregate (including transferred) risk into account.

By contrast, in the right-hand system there is no potential for achieving fairness in outcome, as neither stakeholder is aware of the risks associated with the other. Because of this lack of transparency, risk cannot be traded off. Any risk reduction efforts made in this system are not likely to include balancing. Stakeholders for each system are trying to reduce their (subsystem) risk ALARP, which may in turn be increasing the risks associated with the other subsystem. That is, risk transfers are likely to be performed without being explicitly recognised as such by either stakeholder. An equilibrium eventually reached between these (unwanted) risk transfers is not guaranteed to result in a system risk which is ALARP.

5.2 Risk over time

So far we have discussed risks which are present simultaneously in the system, and have not factored in variation in risk over a system's life. This is in accordance with the definitions of ALARP in [1], [2] and [3], which also do not explicitly address this issue. However, risks can emerge over time or can be present during part of the system lifecycle only. It may therefore be reasonable to balance risks over the system life, i.e. to allow a temporary increase in risk for a consequent decrease in risk at a different time. For example, we may allow a system to present a risk which is not ALARP – when considered in isolation – for a short time (e.g. during upgrades) if this results in a decreased system risk over the rest of the system life. Similarly we may be willing to accept a risk which is not ALARP over a short time in order to prevent an unacceptable increase in risk later (e.g. decommissioning a deteriorating nuclear system). Another situation is where a risk is likely to exist for a short time only, it may not be reasonably practicable to implement a design mitigation to reduce this risk, because the costs over the system lifetime are not judged reasonably practicable (although the costs considered over the duration of this risk may be relatively insignificant).

Societal acceptability is another important consideration when assessing the balance of risk over time. This is problematic because social acceptability may change over time, while an assessment will continue to reflect the values of society at the time it was undertaken. It may therefore be the case that further risk balancing is required in future if risks are deemed at that point to be outside the acceptable system risk profile.

5.3 ALARP and ACARP

In the above discussions, we have assumed that the individual risks are known quantities which can be precisely defined and reduced as required. However, in practice this is not the case. Instead, we have to deal with epistemic uncertainty – uncertainty about our knowledge of the world – and we *estimate* the degree of risk, with a certain amount of justified confidence in this estimate. Although in theory we may represent the risk by some continuous probability distribution of our beliefs (Bayesian), in practice the research in this area seeks to find a minimum set of questions to determine confidence in an upper and lower bound for the risks [9].

Often safety engineering is about undertaking activities that increase our confidence in a product or system without necessarily changing. For example, the nuclear industry has a formalised approach to Independent Confidence Building Measures [2].

If we only increase our confidence in the accuracy of the estimate, it is clear that this does not change the risk itself. However, increasing confidence in a risk estimate can alter our *perception* of the risk by defining it more precisely. This situation is shown in Figure 5, where we have expended effort in obtaining confidence in, and refining, the initial risk estimate. The new risk estimate is lower than the original estimate; that is, we have both increased our confidence in our risk assessment, and identified that the risk estimate is lower than originally thought.

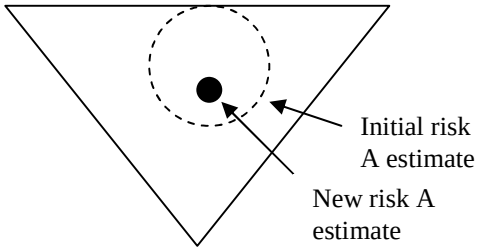

Figure 5: Increasing confidence in risk assessment

This shows that in some situations it may be reasonable to balance the cost involved in reducing risks with the cost involved in increasing our confidence in the assessment of those risks.

There is no consensus on how issues of confidence should be incorporated into the ALARP discussion. For example, it may be valid to use either the "worst case" estimate of the risk, or it may be valid to use the expected risk (which will be identified by a larger envelope); this is a potential area of future research.

6 Decision procedures

We have discussed some of the difficulties in interpreting the ALARP principle in accordance with legal requirements, and

in balancing individual risks in a system. While there is no prescriptive methodology for determining how risks are to be balanced in a system, consideration of the following issues can provide a way to approach risk reduction.

Firstly, any ALARP claim based on compliance with good practice should be examined to determine whether sufficient good practice exists across all applicable areas to support the claim. This is particularly true for SoS, for which there may be insufficient existing good practice and which are particularly vulnerable to many of the confounding factors (including a lack of information transparency) discussed in Section 5. Secondly, any claim that the system risk is ALARP should also include a justification for the factorisation of system risk into individual risks. Both double-counting and omission of risks should be avoided. Furthermore, as we have seen, it is important to determine whether this factorisation leads to independent risks or to risks which interact in some way.

Thirdly, an ALARP claim is legally required to be supported with a justification of the search space; it must be demonstrated that there is no reasonably practicable way to reduce the risk further. However, as stated in Section 5 it is also beneficial to consider if there are alternative ways to reduce the risk to the same level (instead of further). Should an alternative way to reduce the risk ALARP be identified, it is then important to consider the desired risk profile of the system. Explicit consideration of different risk reduction methods, such as fairness in improvement or fairness in outcome, may be of benefit here. Each of these risk reduction methods will result in a system which presents a different profile, and each of these may be vulnerable to a number of confounding factors, as discussed in Section 5.

Another fundamental issue to be discussed is that of confidence. Section 5.3 identifies situations in which our perception of risk can be altered by increasing our confidence in the risk estimate i.e. the expected value of the risk can be altered by reducing the uncertainty in the risk. In situations where we have little confidence in the accuracy of estimates, it is important to consider whether the cost of increasing confidence in our risk assessments - and therefore potentially changing our perception of these risks - is justifiably balanced against the cost of implementing risk reduction efforts.

Finally, any system-specific confounding factors must be assessed. This may include consideration of different risks over time. Nuclear systems are particularly vulnerable to this, as both upgrading and decommissioning a working nuclear plant involve procedures which differ significantly from those involved in standard operation. Systems with a long service life may also be vulnerable to a change in societal perceptions of acceptable risk, while SoS are particularly vulnerable to issues around information transparency.

6.1 Conclusions

In this paper we have discussed some of the difficulties in interpreting the ALARP principle in accordance with legal requirements. In the course of this, we have identified some issues with one of the more common ways of justifying an ALARP claim. This is based on the assumption that the ALARP property decomposes over subsystems, and we have identified several scenarios in which this assumption is invalid. We have also introduced a number of different risk reduction methodologies which focus on the desired balance of individual risks within the system, and identified confounding factors which can increase the complexity of this assessment

There is the potential for significant further work in this area, particularly in extending the mathematical formulation of Section 3 to provide a more rigorous foundation for the work. In particular, it would be useful to be able to express some of the confounding factors (such as risk examined over time) in a more formal setting. Finally, there is scope for further research in the area of confidence. In particular, we propose to examine in more detail the ramifications of using risk estimates for ALARP and of trading off the costs of increased confidence and risk reduction. More generally ALARP needs to be developed to deal explicitly with epistemic uncertainties.

References

[1] Health and Safety Executive, "Reducing Risks, Protecting People", (2001).

[2] Health and Safety Executive, "Safety Assessment Principles for Nuclear Facilities", (2006).

[3] Office for Nuclear Regulation, "Guidance on the Demonstration of ALARP", (2013).

[4] J. Valkonen, I. Karanta, M. Koskimies, K. Heljanko, I. Niemal, D. Sheridan, R.E. Bloomfield. "NPP Safety Automation Systems Analysis: State of the Art", *ISBN 978-951-38-7158-1*, (2008).

[5] R.E.Bloomfield, P. Bishop, S. Guerra, N. Thuy, "Safety Justification Frameworks: Integrating Rule-Based, Goal-Based, and Risk- Informed Approaches", *Proceedings of the 8th International Topical Meeting on Nuclear Plant Instrumentation, Control and Human Machine Interface Technologies*, (2012)

[6] See papers at The 7th Layered Assurance Workshop (LAW) http://www.acsac.org/2013/workshops/law/

[7] Ministry of Defence, “Defence Standard 00-56 Issue 4: Safety Management Requirements for Defence Systems”, (2007).

[8] IEC 15026-2:2011, "ISO/IEC 15026-2:2011. Systems and software engineering — Systems and software assurance, Part 2: Assurance case", (2011)

[9] B. Littlewood, P. Bishop, R.E. Bloomfield, A. Povyakalo, D. Wright, "Towards a formalism for conservative claims about the dependability of software-based systems", *IEEE Transactions on Software Engineering*, (2011)