Institution: City, University of London

Unit of Assessment: Computer Science

Title of case study: Reducing risk in computer-based systems by using a "claims, arguments evidence" framework to communicate and evaluate risks.

Period when the underpinning research was undertaken: 1st January 2000 – **31/12/2017**

Details of staff conducting the underpinning research from the submitting unit:

Name(s):	**Role(s) (e.g., job title):**	**Period(s) employed by submitting HEI:**
Professor Robin Bloomfield	Professor	Since 2001
Dr Peter Popov	Reader	Since 1997
Professor Bev Littlewood	Emeritus Professor	Since 2020, until then Professor (since 1983)
Professor Peter Bishop	Professor	Since 2001
Professor Lorenzo Strigini	Professor	Since 1995
Dr Andrey Povyakalo	Senior Lecturer	Since 2001
Dr Kizito Salako	Lecturer	Since 2019, until then Research Fellow (since 2004)
Dr Vladimir Stankovic	Senior Lecturer	Since 2009
Dr David Wright	Research Fellow	Since 1985
Dr Katerina Netkachova	Research Fellow	From 2013 to 2019

Period when the claimed impact occurred: August 2013 – present

Is this case study continued from a case study submitted in 2014? Y

1. Summary of the impact (indicative maximum 100 words)

A method for demonstrating that the risk in a computer-based system is acceptably low, structured through “claim-argument-evidence” ("CAE") links and supported by quantitative models, is widely and increasingly adopted by industry and regulators in the UK and worldwide.
This method has originated from research conducted in the Centre for Software Reliability (CSR). It requires explicit arguments linking evidence to the claims made about, e.g., safety and security; it encourages rigour and the use of analytical probabilistic models.
The impact through use in industry, listed in the REF2014 case, has continued and has increased with new adopters and adoption of an extension to the method. New beneficiaries include companies and regulators from railway, energy and autonomous vehicles.
The research has informed safety policies for complex critical infrastructures and contributed to new standard and guidance documents worldwide.

2. Underpinning research (indicative maximum 500 words)

The underpinning research spans several decades. “Assurance cases" extend the approach of safety cases, well-structured set of documents, to demonstrate that the risk posed by critical systems are acceptably low. Assurance cases have been widely adopted by industry and regulators, in the U.K. and worldwide.
Research conducted in CSR since 2014 continued the directions in research, which led to the previously reported impact in REF 2014, with several significant extensions of the method:

- support for users of assurance cases to better structure them and to make them more trustworthy by incorporating probabilistic models. The advances concern:

  - improving rigour by defining semantics for fragments of CAE arguments ("CAE Blocks"); improving usability of the CAE framework by introducing various guidance elements, e.g., the “helping hand” visual aid [3.6];
  - formulating explicitly the need for validation of system dependability in the presence of uncertainty [3.1] and demonstrating the benefits that a probabilistic model-based assessment can add to assurance cases. The research team reduced the difficulty of applying model-based assurance to very complex systems, such as interdependent critical infrastructures. Incremental refinement [3.3] allows assessors to progress from a very abstract model of the complex system to a high-fidelity model, as required by stakeholders.
- extensions to security and to security-safety “co-engineering”. In case studies in industrial automation, power grids and medical devices (in projects SeSaMo, AQUAS, RITICS-CEDRICS, I3S), the researchers developed models of the analysed systems which capture the essential aspects of an assessment captured in an assurance case. For instance, in assessing safety under cyber-attacks, it is essential to model credibly how successful attacks to the computer systems can degrade safety of the controlled engineered system [3.5].
- more recently the research team addressed the assurance gaps in critical applications of machine learning and artificial intelligence, with case studies from autonomous vehicles (studied in the recently completed TIGARS and the ongoing ICRI-SAVe projects);
- The team also provided more supportive tools for probabilistic modelling, especially of large, complex systems. Their PIA-FARA tool [3.3] supports "what if" analyses of accident/intrusion propagation scenarios in complex infrastructure, and integration of analysis results into the CAE framework (e.g., in RITICS-CEDRICS project);
- The research team demonstrated potential pitfalls in extending the use of "fault injection" (a well-established technique for probing resilience mechanisms, e.g., IEC 61508 and ISO 26262) as some do to *quantifying* a system's resilience against design faults. A widely recognised problem is making injected faults "realistic"; their modelling [3.4] demonstrated more serious issues. The LoS from Intel Labs [5.2] identifies the practical impact of this insight for their work on building dependable autonomous vehicles.
- Further work on probabilistic aspects of confidence (all projects above). An important, ongoing line of research concerns “conservative” Bayesian assessment, e.g. [3.2]. Bayesian methods bring advantages, recognised by some regulators, but their complexity invites shortcuts that undermine rigour and hence safety. The approach by the CSR team simplifies rigorous application while guaranteeing against over-optimism. This approach is now being applied to autonomous vehicles, to help translate experience of safe operation into the level of confidence that it supports in future safety.

3. References to the research (indicative maximum of six references)

The research outputs on which this impact case is based have been published in selective peer-reviewed forums - high impact factor technical journals [3.1-3.3], in the proceedings of the prestigious International Symposium on Software Reliability Engineering [3.4, 3.5], one of the few conferences using both double-blind reviews and review moderation by senior members of the PC. Article [3.6] is published in IEEE Software, which reaches a very wide audience of practitioners.
A more complete list of publications related to the impact case can be seen at:
https://researchcentres.city.ac.uk/software-reliability/research/REF2021/_nocache.

3.1 Bishop P.G., et al., *Towards a Formalism for Conservative Claims about the Dependability of Software-Based Systems.* IEEE Transactions on Software Engineering, 2011. **37**(5): p.708-717.

3.2 B. Littlewood and A. A. Povyakalo, *Conservative reasoning about epistemic uncertainty for the probability of failure on demand of a 1-out-of-2 software-based system in which one channel is “possibly perfect* IEEE Transactions on Software Engineering, 2013. **39**(11): p.1521-1530.

3.3 R. E. Bloomfield, et al., *Preliminary interdependency analysis: An approach to support critical-infrastructure risk-assessment.* Reliability Engineering & System Safety, 2017. **167**: p.198-217.
3.4 P. Popov and L. Strigini, *Assessing Asymmetric Fault-Tolerant Software*, in *IEEE 21st International Symposium on Software Reliability Engineering*. 2010, IEEE: San Jose, CA, USA, p.41-50.
3.5 P. Popov, *Models of reliability of fault-tolerant software under cyber-attacks* in *The 28th IEEE International Symposium on Software Reliability Engineering (ISSRE'2017)*. 2017, IEEE: Toulouse, France. p.228-239.
3.6 R. Bloomfield and K. Netkachova, *Building Blocks for Assurance Cases*, in *IEEE International Symposium on Software Reliability Engineering* 2014, IEEE: Naples, Italy. p.186-191.
(also: K. Netkachova and R. E. Bloomfield, *Security-Informed Safety.* IEEE Computer, 2016. **49** (6): p.98-102.)

4. Details of the impact (indicative maximum 750 words)

Failure of critical computer systems could result in death, injury, financial loss and damage to the environment. "Assurance cases", well-structured set of documents, to demonstrate that the risk posed by a critical system is acceptably low, have been widely adopted by industry and regulators, in the U.K. and worldwide. Assurance cases require explicit arguments linking evidence to the claims made (or goals pursued) about e.g., safety and security. The approach to assurance via assurance cases was developed over many years with essential contributions from City staff, Prof. Robin Bloomfield and Prof. Peter Bishop, both part-time professors at City, University of London and leading personnel at Adelard LLP[1].

This impact case is about a *specific form* of assurance cases with the following distinct characteristics developed at City, University of London:

- Assurance cases are built using CAE, recently extended with CAE blocks [5.4];
- Assurance cases rely not only on informal reasoning, e.g., based on expert judgement, but also on the rigour of models suitable for quantitative risk assessment.

The CAE blocks make the construction of assurance cases easier for practitioners, leading to a wider adoption, and the assurance cases themselves become more expressive and clearer.
Quantitative models help an assessor to decide on the claims or serve as evidence supporting or refuting the claim, especially in those cases where direct empirical evidence is difficult to obtain. Some claims may be ruled out based on results obtained with models. An example of such a claim is "failures of the versions in multi-version software are independent", which City academics' probabilistic models have demonstrated not to be credible.
The quantitative models the research team has used range in complexity: from simplified probabilistic models, suitable as a risk-communication tool to high fidelity - typically a hybrid of probabilistic and deterministic - models of complex cyber-physical system such as models of critical infrastructures. The simplified models are useful to make the stakeholders' engagement easier by hiding the overwhelming system complexity. High fidelity models, instead, enhance the ability of experts to make well-informed decisions in cases where the expert judgements about how good the system is are hard due to system complexity.
Impact includes:

- Reduced risk of harm from malfunctioning or intentional subversion of critical systems (e.g., nuclear, transportation, power supply, defence, medical) through application of a well-structured evidence-based argument aided by models.
- Improved confidence in assurance: rather than depending on expert judgement or informal reasoning, with attendant significant uncertainty, by using modelling we can narrow this uncertainty or articulate its causes and implications.
- Better understanding of future widely deployed systems, e.g., connected and autonomous systems. We discover that in this era of ubiquitous use of machine learning and artificial intelligence, our approach typically leads to significant savings not only in research and

[1] Robin Bloomfield is a founding partner at Adelard LLP. Peter Bishop is the Chief Scientist at Adelard LLP.

development but also in setting directions for the development of such systems. This advantage comes from rigour in modelling new technical systems, and thus ability to clearly articulate doubts about unsubstantiated claims, (due to lack of awareness by newcomers dealing with assurance).

4.1. Impact via Adelard LLP.

Major impact is achieved through the long-term collaboration with Adelard LLP.

- The CAE Blocks framework is a way of structuring arguments:
  - It is a core part of the "IAEA Software Dependability Assessment guideline" [5.3], released in 2018, affecting nuclear safety worldwide.
  - The UK CPNI is expected to publish later in 2021 examples of security informed safety cases based on CAE Blocks.
  - The CAE approach is supported by Adelard's commercial tool ASCE. According to the vendor, over 300 organisations are using ASCE worldwide, at least 50% of them use the CAE approach.
- CAE has been used in Adelard LLP on their projects on assessing security informed safety of industrial systems and in the development of codes of practice for security informed safety:
  - For the rail industry and for connected autonomous vehicles [5.4].
  - In developing a regulatory cyber-maturity model for air traffic management for the Civil Aviation Authority (CAA).
  - In research conducted by Adelard funded by Assuring Autonomy International Programme (AAIP), a partnership between the Lloyd's Register Foundation and the University of York, and the UK Department for Transport (DfT) on projects on autonomous systems and in the TIGARS project. This has led to "Safety case Templates for Autonomous systems" [5.6], and a new approach to assurance dubbed "Assurance 2.0 Manifesto" [5.7]. Assurance 2.0 is the basis for a project within the DARPA ARCOS program on automated certification (https://www.darpa.mil/program/automated-rapid-certification-of-software).
  - City's stochastic modelling approach and tool, PIA-FARA, supporting the application of CAE to complex systems, such as critical infrastructures, has informed the work of Adelard with the National Cyber-Security Centre (NCSC) on software tools.
- From 2019 Adelard have been training a multi-disciplinary team of managers, engineers from chemical process Control and Instrumentation (C&I) from a major hazards site in using CAE Blocks and elements of Assurance 2.0, with over 100 completing the course to date. Adelard have a long-term project to follow up and support.

4.2. Other impact

Impact was also achieved via other partnerships, e.g.:

- Our case is endorsed by Radiy, a major supplier of C&I for the nuclear industry, with more than 70 installations worldwide including safety protection systems for nuclear plants. The LoS, by senior executives of the company, acknowledges the impact of City's work, especially the modelling work, on Radiy's operation, including on the strategic decision to adopt design diversity in their portfolio of FPGA based products [5.1].
- The approach to rigorous assurance supported by quantitative models has been adopted by new actors (companies/regulators): Railway, Energy, Autonomous Vehicles. The City team was invited to join the Intel Collaborative Research Institute on Safety Assurance of Autonomous Cars (ICRI – SAVe), a recognition of the impact of the prior research conducted by the City team. A statement from the Co-Director of ICRI - SAVe, from Intel-Labs, Germany, acknowledges the impact of our work on assurance cases enhanced with rigorous modelling work on the current global effort on safety assurance of autonomous vehicles [5.2].
- Impact has been achieved in informing policies, e.g., of how models can be used to increase confidence in assuring resilience of complex interconnected critical infrastructures. The work we published [3.3] has informed to some extent the recent report by the Royal Academy of Engineering [5.5], especially in the part related to interdependencies. Bloomfield was a reviewer of [5.5].

- Finally, impact has been achieved via inclusion of outputs from our research in new standard/guidance documents, both international and national:
  - Dependability Assessment of Software for Safety Instrumentation and Control Systems at Nuclear Power Plant, IAEA Nuclear Energy Series [5.3]. This document impacts the nuclear industry worldwide setting guidelines for risk limitation based on the CAE blocks;
  - PAS 11281: 2018 ("Connected automotive ecosystems – Impact of security on safety – Code of practice", sponsored by CPNI, 2018) [5.4]. This guidance document impacts the UK industry working on connected automotive systems.
  - Code of Practice: Cyber Security and Safety, IET No. 211014, sponsored by the National Cyber Security Centre (NCSC), [5.8]. This cross-sector code of practice primarily targets the UK industry, but the actual impact may be broader.

5. Sources to corroborate the impact (indicative maximum of 10 references)

5.1 A letter of support from Director of RPC Radiy, & Vyacheslav Kharchenko, Head of Centre for Safety Infrastructure Oriented Research & Analysis at Radiy, Research & Production Corporation, Ukraine.
5.2 A letter of support from Head of Dependability Research Lab at Intel Labs, Germany, and Co-Director of Inter Collaborative Research Institute (ICRI-SAV).
5.3 Dependability Assessment of Software for Safety Instrumentation and Control Systems at Nuclear Power Plants, IAEA Nuclear Energy Series, No. NP-T-3.27, available at: https://www.iaea.org/publications/12232/dependability-assessment-of-software-for-safety-instrumentation-and-control-systems-at-nuclear-power-plants Accessed 14.12.2020.
5.4 PAS 11281: 2018 ("Connected automotive ecosystems – Impact of security on safety – Code of practice", sponsored by CPNI, 2018), ISBN 978 0 539 02394 7, 60 p., BSI 2018.
5.5 Royal Academy of Engineering, Cyber safety and resilience: strengthening the systems that support the modern economy, N. Jennings, Editor. 2018, Royal Academy of Engineering. p. 52.
5.6 Safety Case Template for Autonomous Systems, available at: http://arxiv.org/abs/2102.02625. Released early in 2021, but developed and shared with various stakeholders in 2020, hence included here. Accessed 04.12.2020.
5.7 Assurance 2.0 Manifesto, available at: https://arxiv.org/abs/2004.10474 Accessed 14.12.2020.
5.8 Code of Practice: Cyber Security and Safety, IET No. 211014, sponsored by the National Cyber Security Centre (NCSC), 93 p., available at: https://electrical.theiet.org/media/2516/cop_cyber-security-and-safety_linkable_secure.pdf. Accessed 14.12.2020.