CITY UNIVERSITY
LONDON

CSR
Centre for Software Reliability

Evaluating the resilience and security of boundaryless, evolving socio-technical Systems of Systems

Robin Bloomfield and Ilir Gashi

September 2008

reb@csr.city.ac.uk, i.gashi@city.ac.uk

020 7040 8423, 020 7040 0273

Centre for Software Reliability
City University
London
EC1V 0HB
United Kingdom

We hope to keep this report up to date so if you are inserted in obtaining the latest version please contact the authors in the addresses above.

Table of Contents

Executive Summary

This report describes the research we have conducted on identifying the technological advancements envisaged in the period 2007-2027, the research challenges that will be faced for the assessment of dependability and security of these complex systems and systems of systems (SoSs), and the assessment framework envisaged for these systems. The research has been conducted as part of a project titled "EMR (Extra Mural Research) proposal SoSoS Development Methodologies for Secure System Evolution" which has been sponsored by UK Defence Science and Technology Laboratory (DSTL).

The report is structured as follows:

- Part 1: Define and characterise the 2027 research challenge by a review of existing road maps, technology watch papers and by brainstorming with DSTL (and NATO) contacts.
- Part 2: Propose and justify a Security Evaluation Framework based on existing approaches from other sectors and disciplines. This draws on present state of the art in assessment of SoS of COTS based components and in particular explain and understand why existing methods do not either scale or otherwise translate to the SoSoS context.
- Part 3: Define the characteristics of a security evaluation framework based on the future context explored in Part 1 with the proposed framework in Part 2. It also characterises the research challenge.

The research challenges from the context we and others envisage are as follows:

- *Critical societal role* – The technological advancements discussed in section 2.2 and 2.3 show that an unprecedented reliance on technology may be created. Due to the increased access to advanced technology the future generations will become more vulnerable to either *deliberate* or *unintentional* disruption of the system implemented with these technologies.
- Unprecedented *scale and complexity* coming from the ubiquity and pervasiveness of systems that are driven by adaptation and evolution, ambition and requirements, the blurring of boundaries and the increased tempo of threats and operations. It occurs through design and deliberate policy and accidentally, or as side effect, of other trends. It is shaped by the economic and political forces: systems were once seen as technical, then socio-technical and now we can see political-economical-socio-technical (p-e-s-t) systems
- Concerns regarding *privacy* are likely to become especially important. The widespread data sharing and communication that will be required to make the new systems function may lead to serious infringements of privacy. Various infringements are possible such as *identity theft*, *data laundering*, *disclosure of personal data*, *surveillance*, risks from *personalised profiling* etc.
- *Trust / Assurance* – The new technological advancements, especially those on ICT systems will require substantial level of *trust* to be placed on the new systems. Hence the new technology needs to be *trustworthy*. Building trustworthy systems (especially complex interconnected SoSs) will pose a significant challenge since current methods for both *building* and *assuring* a system that is sufficiently dependable or secure will most likely not scale well when applied to these complex systems.
- *Inter-dependence and dependence* – The future socio-technical ICT systems, as discussed in section 2.2 and 2.3, are likely to be complex inter-connected and inter-dependent systems in a much greater scale than they are now. Hence the assessment of the various security and reliability attributes cannot be done in isolation for each constituent component of the systems or SoS.

- *Interoperability* – There is likely to be an increased difficulty in architecting complex systems of systems. There might for example be inconsistencies between architectures (including inconsistencies in the interfaces) of the different systems of an SoS or inconsistencies in the timescales (timebands) in which the different systems interact.
- *Socio-technical aspects* – The future systems are expected to become increasingly socio-technical in nature, hence the role of the human users and operators within the system, and their fallibilities also need to be considered in the overall assessment of the system.
- *Blurring of boundaries* – The complexity of the future systems and the large inter-dependence and interconnectedness of the components will make it very difficult to define what the boundaries of a system or a SoS are, i.e. where does one system end and another one begin. This makes security and dependability assessment of a SoS very difficult. Assumptions may need to be made about the systems boundaries which may not hold in practice, hence leading to wrong conclusions being drawn from the assessment. We envisage future blurring of boundaries between human/machines and between classes of devices.
- *Information explosion* – The future systems will deal with radically increased volume of information due to advance in sensor and networks technologies. There will also be greater pressure on both systems and human decision makers to deal with information in shorter response times. This will challenge effective decision-making progressively at all levels.
- *Rapid obsolescence of technology* – The predicted pace of new technological innovations, listed in section 2.2 and 2.3, is likely to render existing technologies obsolescent more quickly than at present and also lead to heterogeneous systems composed on many generations of technology.
- *Tempo* – There will be increasing tempo to operations and systems with dynamic and ad hoc coalitions being formed. Reconciling timing issues as systems and organisations are brought together to form larger SoSs also becomes an issue. There might be inconsistencies in the timescales within which the different systems interact and there might also be differences in the time domain within which two organisations, that form part of the same SoS, interact, e.g. which tasks/process are considered more urgent will depend on the organisation.

These implications might seem very incremental in that they are just extending social and technological trends we can already see. However this hides the fact that they would apply to many different future scenarios. Even those in this report are quite disparate but one can imagine a wide range of futures of different economic wealth distribution, of relationships between the state and the individual, of levels of social cohesion and of conflict and threat. However we have not analysed extreme scenarios where we see anti-technology refusenik cultures, disenchantment with a technology (e.g. as a result of successful attack of key systems, their oppressive use by the state, breakdown of complex systems such as financial markets), or scenarios of extreme state control. Nor have we tested the robustness of the work to developments in quantum computing.

The challenges posed by current and future systems and threats leads us to propose an ambitious shift in perspective to an evaluation framework that attempts risk based, market sensitive, psychologically aware, evidence based approach to the assessment and communication of security (and dependability, resilience). The approach must be capable of addressing the scale and complexity of adaptive systems of systems (p-e-s-t systems) that have heterogeneous human and technical components and are deployed across a variety of organisational, political and legal boundaries. This report is on the *evaluation* of systems; however the actual *achievement* of trustworthy systems to time and budget is not a solved problem, as the current difficulties with the NHS National Programme for IT (NPfIT) system [1] testifies.

The proposed framework would be composed of:

- Evaluation and communication of risk-based resilience.
- The definition of the evaluation target and the assumed threat model.
- A claim-argument-evidence assurance case approach.
- Methods for addressing scale and tempo required of both events and operations and system evolution.

The framework is probabilistic and should explicitly address uncertainty, both aleatory uncertainty in the world as well as epistemic uncertainty arising from incompleteness of knowledge. The framework should also:

- Consider the economics of security and the role of the markets.
- Have methods for handling decisions and evidence of different levels of confidentiality and trust.
- Consider the (possibly conflicting) expectations of multiple stakeholders.
- Be justified and validated (unlike many current standards).
- Be repeatable and trustworthy but recognising the importance of human judgements
- Be adapted as threats adapt to how the evaluation works.
- It should be applicable in graduated manner commensurate with the importance and criticality of the system or service being considered.

The framework has been defined to address issues along the following themes:

- Confidence / Trust
- Diversity and Heterogeneity
- Complexity and Emergent properties which may arise from the novelty of the systems developed / integrated
- Structure
- Resilience
- Tempo and adaptation
- Markets

We then elaborated some of the research directions along these themes into two inter-related broad areas:

- Resilience and security cases as an overarching approach within which we address: resilience models, claims decomposition, arguments, scalability and tempo. The argumentation will be about a service or system in an environment. To give meaning to the claims and to understand them they should be based on a coherent set of underlying models.
- Models for giving meaning to and supporting the evaluation. These will be disparate, multi-formalism models. We need to understand the interrelationships, the required abstractions and levels of fidelity. They will include models of the system environment and particular economic and threat models.

We also point out that an evaluation framework is not a neutral, technical object. It will have an impact on society and as with any risk based approach there will be those that benefit and those that suffer the costs of the framework. There may be unintended side effects and the market may operate to deliver certain levels of resilience but not necessarily the required level of resilience required for society critical systems (and occasionally the market may even hinder / prevent the delivery of security/resilience). The need for supporting policy and regulation should be born in mind and the research on the evaluation framework should feed into and enable research underpinning any emerging political initiatives. For example, on the tradeoffs between and conflicts of interest that may arise between the government, society and individuals, such as issues regarding intellectual property rights and privacy, human rights issues, individual liberty etc.

1 Introduction

This report describes the research we have conducted on identifying the technological advancements envisaged in the period 2007-2027, the research challenges that will be faced for the assessment of dependability and security of these complex systems and systems of systems (SoSs), and the assessment framework envisaged for these systems. The research has been conducted as part of a project titled "EMR (Extra Mural Research) proposal SoSoS Development Methodologies for Secure System Evolution" which has been sponsored by UK Defence Science and Technology Laboratory (DSTL).

The project proposal stated that there would be three main tasks of the research conducted as part of this project:

- Task 1: Define and characterise the 2027 research challenge by a review of existing road maps, technology watch papers and by brainstorming with DSTL (and NATO) contacts.
- Task 2: Characterise the present state of the art in assessment of SoS of COTS-based components and in particular explain and understand why existing methods do neither scale nor otherwise translate to the SoSoS context.
- Task 3: Characterise the research challenge by addressing the INDEED themes of adaptation and diversity; confidence and uncertainty; time and structure and responsibility and trust. We would consider issues from the perspectives of:
    - probabilistic approaches to diversity and adaptation in socio-technical systems that incorporate changing perceptions of dependability, learning and adaptation with experience.
    - models of organisational responsibility and the underlying trust between agents.
    - argument modelling approaches that address confidence, uncertainty and diversity and by using trust models to support the presentation of assurance cases.
    - the notion of timebands as structuring mechanism.

The rest of the report will present the research conducted for each of these tasks. The report is structured as follows:

- Part 1: Define and characterise the 2027 research challenge by a review of existing road maps, technology watch papers and by brainstorming with DSTL (and NATO) contacts.
- Part 2: Propose and justify a Security Evaluation Framework based on existing approaches from other sectors and disciplines. This draws on present state of the art in assessment of SoS of COTS-based components and in particular explain and understand why existing methods do not either scale or otherwise translate to the SoSoS context.
- Part 3: Define the characteristics of a security evaluation framework based from the future context explored in Part 1 with the proposed framework in Part 2. It also characterises the research challenge.
- Summary and Conclusions: Summarises the report, presents the main conclusions and lists suggestions for possible extensions to the report.
- Appendix A: Details three scenarios developed by a NATO Research Task Group on the Dual Use of High Assurance Technologies.
- Appendix B: Details analysis of the three scenarios presented in Appendix A.
- Appendix C: Details research trends and challenges in the assessment of COTS components.

- Appendix D: Lists the assessment techniques and methods that are possible for COTS-based development.

2 Part 1: The future context

2.1 Introduction

This section characterises the systems and their dependability context predicted to occur up to 2027 and from this summarises the main research challenges faced for assessing their dependability and security. Of course it would be too unrealistic to try and predict the exact nature of the products and services that will emerge. Instead we hope to capture sufficient of the trends that we can provide some design criteria for a security evaluation framework. We have used a variety of sources to identify the predictions of technological advancements and trends and similar to [2] we followed a three-forked approach:

- *Top-down scenario driven approach*: in partnership with a NATO Research Task Group IST-048/RTG-020 we developed three Scenarios to identify primarily the security and dependability challenges that can be faced by large SoSs. We have provided details of each of these scenarios and the analysis that was performed on these scenarios in Appendices A and B respectively. Additionally we have studied Scenarios developed by others, for example [3], [4], [5].
- *Lateral approach*: we studied existing roadmaps and strategic trends documents to critically analyse the predictions made on those documents for the future technological advancements and the research challenges faced when and if these emerging technologies are deployed. Main sources for this section were [6], [7], [8], [2] and the references therein.
- *Bottom-up approach*: we studied and reviewed current state of the art in the dependability and security assessment of systems to identify the research challenges faced when the current approaches are attempted for assessment of emerging systems or SoSs. The main sources for this work have been the recently completed state-of-the-art reviews (in which Centre for Software Reliability (CSR) was also a contributor) conducted as part of the EU-funded ReSIST [9] and AMBER [10] projects and the references therein.

2.2 System characteristics and context

The AMSD EU sponsored *Dependability Roadmap for the Information Society in Europe*[1] [2] identified the key dependability related characteristics of future systems and these can be grouped as issues of:

- Scale and complexity.
- Adaptation and evolution.
- Blurred boundaries.
- Heterogeneity.
- Multiplicity of faults and threats.

We have taken these as a starting point and augmented with additional insights from a complementary EU Study on the dark side of ambient intelligence [5], and from the UK Foresight programme [3]. The defence and security context are provided by the *Global Strategic Trends Programme* [6], *Network Enabled Capability* [11], *Global Information Grid* [12] and by discussing with DSTL experts.

[1] Robin Bloomfield was one of the principal authors.

2.2.1 Scale, complexity and pervasiveness

In [2] we envisaged a growing scale and complexity of systems. As hardware capabilities improve and costs reduce, there is continuing pressure to attempt to build systems of ever greater scope and functional sophistication, especially for the software components: given Moore's law, Metclaf's law and Bell's Law this was hardly a prescient prediction.

The reasons why software is difficult have been covered elsewhere [13]. The scale and complexity comes from the ubiquity and pervasiveness driven by adaptation and evolution, ambition and requirements, the blurring of boundaries and the increased tempo of threats and operations. It occurs through both design and deliberate policy as well as accidentally, or as side effect, of other trends. It is shaped by the economic and political forces: systems were once seen as technical, then socio-technical and now we can see political-economical-socio-technical (p-e-s-t) systems e.g. in health service, financial markets etc. In the defence context there is the momentum given by the Global Information Grid [12], Network Enabled Capability [11] and Network Centric Operations as well as by the tempo and variety of military operations.

It is now commonplace to predict a growth in the use and pervasiveness of Information and Communications Technology (ICT) ([6], [2], [14]). In the DCDC report [6] it is stated that "*Wearable and implanted wireless ICT is likely to be accessible to all that can afford it in the second half of the period [2020 onwards], and users will be linked through sensors and networks that are enabled by computers that are significantly more capable than at present, possibly by 100bn times if quantum computing reaches its potential*". The aforementioned sources also comment on the security and privacy concerns that will develop from the widespread use of these pervasive applications (e.g. [6] states that "*…the majority of the global population will find it difficult to 'turn the outside world off'. ICT is likely to be so pervasive that people are permanently connected to a network or two-way data stream with inherent challenges to civil liberties;* ***being disconnected could be considered suspicious***"). For more concrete scenarios of the predicted technological advancements and their use the reader is advised to read the scenarios depicted in [2].

These trends are also reflected in the Calls for Proposals FP7-ICT-2007-1 and FP7-ICT-2007-3 of the Seventh Framework Programme (2007-2013) (FP7) of the European Commission on ICTs [15]. The principal stated objectives of the funding are:

- *"…to research, demonstrate and validate new computing architectures and algorithms that will allow designing, programming and managing future high-performance ICT components with up to one* ***Tera (10^{12}) devices integrated in a single chip.***"
- *"…to investigate an invisible, implicit, embodied or even implanted* ***interaction between humans and system components****, for natural interaction (including communication) in surrounding environments, themselves augmented with pervasive and ubiquitous infrastructures and services.*"
- *"…to overcome major scientific, technological and theoretical challenges for quantum technology to deliver on its promise to radically outperform its classical counterpart not only in terms of processing speed, capacity and communication security, but also, in the ability to solve classes of practical problems which currently cannot be solved*".

The pervasiveness and growth of the internet can also be expected to continue. Semantic web [16], for example, which is an evolving extension of the World Wide Web, may bring a new way in which the internet information is organised and used.

2.2.2 Adaptation, evolution and architecture

The increased scale and complexity of systems is partly explicitly intended (as in the NHS National Programme for IT (NPfIT) system [1]) but also an effect of federation and integration of systems of systems. Systems will be under continuous incremental development and deployment – they are never finished, evolution is incessant, upgrades, changes in functionality and new features are being added at a continuous pace [1]. Two main strands of development are:

- self-configuration and adaptation – systems are expected to be able to respond to the changing circumstances of the ambient where they are embedded;
- multiple innovative types of networking architectures and strategies for sharing resources – GRID-like, peer-to-peer, "on-the-fly" services, Trusted Web services etc.

This is emphasised in [6] that predicts that our reliance on **networks** and the **complex** nature of our **environment**, (with often poorly understood properties), will increase. They state that "*higher bandwidth, greater processing power, larger datasets, smaller sensors and greater understanding of the dynamics of physical and virtual network behaviour will converge to allow new types of network connection. However, the growth of many networks is not and cannot be governed by top-down planning and occurs in a decentralized manner, often analogous to naturally occurring systems*".

Web services technologies are expected to continue growing [17]: "*Over the next 15 years [from 2008], we'll see the rise of many trade-secret-based companies that do everything from image rendering to statistical calculations to heat-flow analysis ....No large corporate program will be able to run effectively without using of these services, and to do so, it will have to tell its secrets to service providers. Webs of trust will become webs of contracts, and control over security will be out of the hands of any individual organization*".

Continued growth is also expected in the development and adoption of virtualisation of computing and computer resources. For example, [17] states "*the current trend to use a single physical-logical PC for everything will reverse—for example, I'll eventually use my "office work" computer, my "Web browsing" computer, my "e-banking" computer, my "personal theatre" computer, and so forth. The fact that they may all exist in the same box is irrelevant because what matters is that they'll be self-contained and non-interfering, yet will also have different levels of security and dependability*".

Virtualisation hence promises a new way of organising computing resources and more flexibility in the way users interact with the systems, but security concerns will arise from the closer proximity of virtual machines and the location of resources in the same physical host(s).

2.2.3 Blurring of boundaries

The AMSD roadmap [2] discussed the boundary-less nature of the systems and interconnectedness – few systems have a clear-cut frontier, and they are in systems within systems, within larger systems; in addition connectivity might be achieved through common underpinning information infrastructures that become a critical factor. Nodes and services of the system might be reachable from everywhere and the overall system might exhibit unpredictable emergent behaviours.

According to predictions made in [6], the English language will most probably consolidate its position as the globally dominant language for data and global services. But there will be a proliferation of other trans-national languages (e.g. Spanish, Mandarin or Arabic) hence "*sophisticated **translation devices** are likely to become widely available"*. More generally "*interdisciplinary advances involving **cognitive science** are likely to enable us more effectively to map cognitive processes*" [6]. Amongst the advances predicted are self repairing networks (which are predict to be developed in the next 10 years) and the prediction that before the year 2035 the mapping of human brain functions and the replication of genuine intelligence is possible. Whether or not this is achieved the more near term implications are that greater understanding of cognition will allow us to develop new approaches to computation particularly to interfaces.

2.2.3.1 Synthetic environments and decision support

Interfaces are crucial to decision making. In [6] it is predicted that combined advances in social science, behavioural science and mathematical modelling will lead to more informed decision making. Hence "*advanced processing and computational power will permit a new level of pattern recognition (Combinatronics) enabling the decoding of previously unrecognised or undecipherable systems and allowing the modelling of a range of biological to social, political and economic processes".*

Stemming from these advancements, **simulation** is predicted to become an increasingly powerful tool to aid policy and decision makers (but it may also "*blur the line between illusion and reality*" [6]). Later in the same report [6] it is predicted that the advancements in the ICT and Cognitive sciences fields are likely to produce advanced decision-support tools which are likely to be revolutionary leading to opportunities for novel or decisive application. Even though they predict that most of the advancements are likely to have positive effects, some advances are also predicted to present potential negative effects (e.g. perverse applications, such as the use of genetic engineering to produce designer bio-weapons).

2.2.4 Heterogeneity

As is evident with current systems, there will also be continuing increase in heterogeneity of the future systems caused by the use of components of different generations and the need to support legacy services, applications and protocols along with more modern variants. There will be multiple COTS and SoSs that are brought together for particular missions. The components and systems will be of different provenances and trustworthiness. This will be reflected in the range of standards that they might comply with and there will be an increasing need to address interoperability issues.

The increase in heterogeneous and disparateness of systems is partly due to the blurring of boundaries between systems; partly due to the different levels of scale; partly due to the blurring of human/device boundary – wrist-held gadgets, wearable devices, implantable devices; and partly due to the adaptation and evolution discussed above.

Advancements in biotechnology are predicted and the software related impacts are "*development of artificial sensors capable of interfacing with the human mind and prosthetics capable of mimicking human actions precisely, improving human performance beyond current levels*" [6]. The same source also states that due to the high costs of the biotechnology research that the distribution and benefits will be unevenly distributed in the society.

Details of short-medium term funding on biotechnology and ICT convergence (2007-2013) can also be found in FP7 call ICT-2007.8.3 [15] of the European Commission.

Advancements are also predicted in the area of nanotechnology. In [6] it is claimed that "*advanced nanotechnology, at the interdisciplinary frontier where physics, chemistry and biology meet, will be a key enabler of technological advance and will underpin many breakthroughs, including materials and sensor development and their application in manufacturing, synthetic reproduction and health care. Nanotechnology will result in* ***more-capable systems*** *and artefacts that are* ***smaller***, ***lighter***, ***cheaper*** *and less energy hungry. Out to 2020, its application is likely to be predominantly in electronics and materials, including bacteria resistant agents, stain resistant materials and nanocomposite materials. After 2020, nanodevices are likely, such as nanobots*". Additionally in [17] it is claimed "*The devices we use to connect to information systems will become smaller and smarter, but also more specialized and more diverse. Most of the functionality we're used to on our desktops today will move "into the cloud" as connectivity gets more ubiquitous, more robust, and cheaper.*"

Most of the advancements above are clearly to do with hardware, but they will also impact the software applications that are developed to make use of the underlying layer of nanotechnology hardware.

Details of short-medium term funding on nano-technology (2007-2013) can also be found in FP7 call ICT-2007.8.1 [15] of the European Commission.

2.3 The security context

2.3.1 Multiplicity of faults and the threat space

The AMSD roadmap [2] rather succinctly characterised the threat space in terms of multiplicity of fault types – in particular the growing danger of malicious faults, both due to individual or organised external attackers, and due to deceitful insiders – and the need to model and understand the multiple interactions and interconnections (human, technical, social, market) among systems, when they all depend on each other.

More recent publications ([11], [12]) amplify this and focus these especially in a security and military context. For example in [11], regarding the UK MoD planned Networked Enabled Capability (NEC), it is stated that "*The growing threat from cyber attack, to which we will become more vulnerable as our reliance on the network increases, must also be contained through new measures*". Both [11] and the US DoD Global Information Grid [12] document envisage the tighter integration and interoperability of military and defence agencies, departments and systems over the network hence exposing these networks to an increased risk from cyber-attacks.

Some of the reasons for the increase in the threat space may be:

- *Wider availability of technology* – Technology will become more widely available and affordable [6]. This technology is likely to have both military and commercial use. It will benefit the less technologically capable, particularly through cheap, novel applications. But the availability of this technology to the less technologically capable and inexperienced users will also increase the dangers of system or application mis-configurations which will create vulnerabilities in systems, making them more prone to malicious attacks and machine hijacking.
- *Rapid Mass-Mobilization* – The pervasiveness of ICT will enable communities of common interest to be established very quickly and coordinate the mobilization of significant numbers of people. According to [6] "*Rapid mobilization – 'Flashmobs' - may be undertaken by states, terrorists and criminals, and may involve dispersed communities across international boundaries, challenging security forces to match this potential agility and ability to concentrate*". An example of rapid mass-mobilisation may be the cyber-attack against Estonia in 2007 [18]. Even though the exact perpetrators of the attack are not known for sure (Estonia initially blamed the Russian government for involvement; the Russian government denied any involvement) it seems like a large number of the attacks may have come from disparate sources influenced by a common cause [19].

Innovations in forensics and counter-forensics tools and applications are also predicted, especially in commercial settings where litigation proceedings to get remuneration from consequences of attacks or breaches may be launched more frequently. For example, [17] states "*Forensics and counter-forensics, a major trend in security to come within the social fabric. Military [computer security] failures don't allow appeal: if you didn't protect your transmissions, you won't get far complaining about being exploited. Commercial security failures are quite different, and if you collected the right evidence, you could recover by suing.*" The Estonia attack [19] highlights difficulties with attack attribution (and not only governmental but also commercial sites, such as banks, were targeted in this attack).

2.3.1.1 Unintended side effects

There might also be various unintended side effects (especially in the security context) from the advancements in technology. For example, the intelligence agencies' capacity to penetrate the cyber-space may be reduced due to the increasing use of obscure languages and commercially available 'strong' encryption tools and algorithms [6]. We may also see increasing pressure from nation states on developers of system to provide secret trap-doors and others surveillance mechanisms as a price to pay for access to markets.

In [6] it is stated that "*innovation, research and development will originate from more international and diffuse sources and will proliferate widely, making regulation and control of novel technologies more challenging*". Further [6] states that (the emphasis is ours) "*the exploitation of these may have catastrophic results, especially those associated with **nanotechnology**, **biotechnology** and **weapon systems**. These may be unintended, for example 'runaway' nanotechnology or biotechnology, or intended, such as the development and use of directed energy or electromagnetic-pulse weapons*".

Other side-effects include:

- *Information warfare* – [6] reports that some states are developing sophisticated "*information warfare capabilities*" by exploiting the pervasiveness and pliability of digital information to gain commercial or political advantage. They predict that the threat of opportunistic hacking and network manipulation will also continue at an increasing rate and intensity. They predict that both hostile nation states and opportunistic criminal or terrorist hackers will represent a significant threat to military ICT systems. Hence these ICT systems will require robust and comprehensive protection.
- *Technology leakage* – The interconnected nature of the networked systems will increase the risk of deliberate or accidental technology or information *leakage* [6]. This is likely to happen in spite of stringent regulations and security. This may lead to a "*widening number of state and non-state actors accessing advanced and sensitive technologies*". For military and other safety-critical systems or SoSs this may result in hostile states organisations obtaining access to novel weapons and devices. [6] also states the risk (from the defence and law-enforcements point of view) from "*ethical scientists*" who may reveal details of advanced programmes "*in the interests of ensuring a 'level playing field' and balance of risk.*" Risks of technology and information leakage are also discussed in the three scenarios (Appendix A) developed in partnership with a NATO Research Task Group.

2.3.2 Insights from the scenarios

A NATO Research Task Group on the Dual Use of High Assurance Technologies (IST-048/RTG-020) has written a set of three scenarios of *potential* attacks against critical infrastructures. The task group was headed by Dan Craigen (University of Toronto, Canada), Ann Miller (University of Missouri-Rolla) and Robin Bloomfield (CSR, City University). A final report detailing the scenarios and the analysis performed is in preparation. Details of the scenarios and initial analysis performed are provided in Appendices A and B of this report.

The "Terms of Reference" (ToR) for the Research Task Group (RTG), provides the historical background for the group as follows:

"As a result of the terrorist attacks on the United States of America on September 11, 2001, NATO and its member countries have been actively investigating means for combating terrorism.

High Assurance Technologies (such as formal methods) have normally been used to develop systems requiring high degrees of assurance as to functionality, safety and security. One of the key benefits of such technologies is their ability to ferret out subtle problems with system requirements, design and implementation."

The ToR continues by justifying the RTG to NATO as follows:

"An important aspect of combating terrorism is the protection of NATO network-enabled systems from exploitation by terrorists and other foes."

The three scenarios are entitled:

- Scenario 1: Medical and Financial Services Sector – (present time)
- Scenario 2: Oil and Gas Sector – (present time)
- Scenario 3: Electrical Power Sector – (Circa 2012)

In what follows we will provide a brief analysis of Scenario 3 (developed by Robin Bloomfield at CSR, City University) with details of threat targets and the goals and intentions of the attackers. Details of how the attack in this scenario may unfold as well as details on the other two scenarios are given in Appendices A and B.

2.3.2.1 Implications of Scenario 3: Electrical Power Sector

The scenario in Appendix C provides a number of insights into the nature of potential incidents for SoS and implications for as security evaluation framework. The scenario describes an incident, initially from the point of view of a controller of the electricity network as he attempts to control and recover the electricity grid in the face of doubt and confusion caused by bad weather, possible attacks, heightened threat levels and malfunctioning information and control systems

One immediate insight from the scenario is that in a complex SoS, such as the electrical power supply grid and supporting infrastructure, it is over simplistic to think of an attack being designed and executed by a single, identifiable agent. While this is of course possible, the scenario shows how a number of malicious, uncoordinated events together with stresses due to accidental events, environmental effects and opportunistic escalation challenge the resilience of the complex system: using a disease metaphor, there is no need for opportunistic infections to co-ordinate. Not only had the adversary absorbed the vocabulary of systems of systems and "boundary-less" systems but had developed an approach that broadened effects-based planning to a more ecological or bio-inspired view. The incident described is just one of the many routes through to failures: it is not just a "security" event.

The scenario uses a mixture of "attacks" that are un-coordinated in the normal sense of having short term communication between the actors but rather through a long term, decentralised understanding and intent. The use of markets and ownership may require large nation-state levels of resource to influence and to hide their intentions.

The threat agents assumed that challenges to the power system were bound to occur so pursued tactics that would increase the magnitude of these effects, make their management more difficult and especially hamper recovery. The adversary made significant use of system engineering and risk analysis skills to understand complex systems. They were aware of "normal accident" theory and had knowledge of complex systems and the relationship between topology and cascading effects. They had the ability to run simulations, picking up on published work and the availability of topological information. They had designed their approach so it did not rely on a single type of attack or a single vulnerability as these could be found. They had understood that consequences in interdependent infrastructures increase non-linearly with time to repair[2].

The difficulty in operation caused by bad weather had been amplified by:

- Behaviour of the protection devices;

[2] So damage something like = geographic impact*time^n. Would be interesting to run some models to investigate further.

- Incorrect risk information and Management Information System (MIS) problems had made recovery more difficult.

However hampering recovery and causing losses would be possible without the sophisticated compromise of the protection devices. The adversary might have been satisfied with just prolonging the disruption or could initiate it with more conventional attacks (e.g. graphite on sub stations, fires etc).

The adversaries had sown an easily detectable "bug" by using a friend in the contracting company to miss-set the trip levels in some of the protection devices. The malicious code they had inserted by intercepting the devices after manufacture had not been detected, despite the industry approved audits and assessments that had taken place, and some had been activated and tested in this incident. However once ran the code also self destructed[3]. They had made use of computer science tools for reverse engineering and mathematical modelling of the sensor code. The adversary had access to the code, unlike the industry assessors and auditors.

They had also learnt about the forensic capability of the system, the times to restore, had cast doubt on the risk information system, learnt how to stimulate consequential social unrest and exploit the legal system.

Another theme was the desire to escape detection, to have the incident seen as a conjunction of several unfortunate events. Part of the scenario was the doubt that it was hard to decide whether it was orchestrated or not. This was in part to avoid the massive retaliation that might follow an overt attack but also as a learning exercise[4] so as to retain the potential for future escalation and pave the way to a "shock and awe" variant.

2.3.2.2 Threat Targets

The scenario concerns a system of systems – the electricity generation and supply systems and its supporting control and maintenance systems and the people, organisations and institutions that are stakeholders. The "system" included the

- Technical system of electricity supply and control;
- The organisation and management of that system;
- The legal and insurance system;
- The supply chain;
- The maintenance system.

Also some abstract elements such as:

- Situational awareness;
- Doubt and confusion;
- Confidence in the system.

It should be noted that the consideration of threat targets as the more concrete concerns may result in missing some important points (e.g. an effective attack on an eVoting machine might be in the voters' confidence in the accuracy and robustness of the system).

[3] Might have inserted code to be date specific, or randomly start. Could be wiped by rebooting the OK version then reload it but would then need to control type of reboots. Would need to compromise checksum and fault detection measures. Could be inserted before shipments, as in this version, but the connectivity of the devices, albeit over a partially Government approved network, might lead to credible scenario with external attack.

[4] The adversaries were familiar with the High Reliability Organisation [HRO] literature.

2.3.2.3 Goals/intentions

A variety of goals and intentions have been discussed for the immediate incident:

- Challenge the state and corporations and show their vulnerability and fallibility;
- Decrease confidence in the state's competence;
- Increase feelings of insecurity and under attack (prompting further measures);
- Economic damage via loss of supply (secondary goal);
- Learn how to mount a larger "shock and awe".

There were also goals with respect to their capabilities

- To develop and prove stealth based disruption.
- Insert vulnerabilities and sources of instability that then promote and allow instabilities to grow.

2.4 Summary of evaluation challenges

The technological developments mentioned so far have various cultural, societal, political and environmental implications. They will be embedded in a world that is coping with the impact of climate change, rapid modernisation of China and India, competition for food and natural resources and increased sophistication, strategy and multiplicity of adversaries.

The discussion of the research challenges of assessing the dependability and security of these systems will take place later in the report. We will first summarise the main implications that arise from these technological advancements. We have touched upon some of these issues already in the report. Some of the main implications are in the fields of / due to:

- *Critical societal role* – The technological advancements discussed in section 2.2 and 2.3 show that an unprecedented reliance on technology may be created. Due to the increased access to advanced technology the future generations will become more vulnerable to either *deliberate* or *unintentional* disruption of the system implemented with these technologies. As mentioned in [6] countries with developed economies (where the adoption of these advanced technologies is likely to be more widespread) are likely to be more vulnerable to such disruption compared with less technologically advanced societies.
- Concerns regarding *privacy* are likely to become especially important. The widespread data sharing and communication that will be required to make the new systems function may lead to serious infringements of privacy. Various infringements are possible such as *identity theft*, *data laundering*, *disclosure of personal data*, *surveillance*, risks from *personalised profiling* etc.
- *Trust / Assurance* – The new technological advancements, especially those on ICT systems will require substantial level of *trust* to be placed on the new systems. Hence the new technology needs to be *trustworthy*. Building trustworthy systems (especially complex interconnected SoSs) will pose a significant challenge since current methods for both *building* and *assuring* a system that is sufficiently dependable or secure will most likely not scale well when applied to these complex systems.
- *Dependence and inter-dependence* – The future socio-technical ICT systems, as discussed in section 2.2 and 2.3, are likely to be complex inter-connected and inter-dependent systems in a much greater scale than they are now. Hence, the assessment of the various security and reliability attributes cannot be done in isolation for each constituent component of the systems or SoS.

- *Interoperability* – There is likely to be an increased difficulty in building complex systems of systems. There might for example be: inconsistencies between architectures (including inconsistencies in the interfaces) of the different systems of an SoS; but also inconsistencies in the timescales (timebands) in which the different systems interact.
- *Socio-technical aspects* – The future systems are expected to become increasingly socio-technical in nature, hence the role of the human users and operators within the system, and their fallibilities also need to be considered in the overall assessment of the system.
- *Complexity and System boundaries* – The complexity of the future systems and the large inter-dependence and interconnectedness of the components will make it very difficult to define what the boundaries of a system or a SoS are, i.e. where one system ends and another one begins. This makes security and dependability assessment of a SoS very difficult. Assumptions may need to be made about the systems boundaries which may not hold in practice hence leading to wrong conclusions being drawn from the assessment. We envisage future blurring of boundaries between human/machines and between classes of devices.
- *Information explosion* – The future systems, will deal with radically increased volume of information due to advance in sensor and networks technologies. There will also be greater pressure on both systems and human decision makers to deal with information in shorter response times. This will challenge effective decision-making progressively at all levels. As is stated in [6] "*Greater personal, corporate and military dependence on ICT and commercial interconnectedness and applications will create greater vulnerabilities and fragility, magnifying the impact of information denial, failure or manipulation.*" Continuing leakage and diffusion of sensitive information (which can have potentially dangerous national security implications if it concerns for example weapons systems) will continue to happen at potentially a greater rate.
- *Rapid obsolescence of technology* – The predicted pace of new technological innovations, listed in section 2.2 and 2.3, is likely to render existing technologies obsolescent more quickly than at present [6] and also lead to heterogeneous systems composed on many generations of technology.
- *Tempo* – There will be increasing tempo to operations and systems with dynamic and ad hoc coalitions being formed. Reconciling timing issues as systems and organisations are brought together to form larger SoSs also becomes an issue. There might be inconsistencies in the timescales in which the different systems interact and there might also be differences in the time domain in which two organisations, that form part of the same SoS, interact, e.g. which tasks/process are considered more urgent will depend on the organisation.

These implications might seem very incremental in that they are just extending social and technological trends we can already see. However this hides the fact that they would apply to many different future scenarios. Even those in this report are quite disparate but one can imagine a wide range of futures of different economic wealth distribution, of relationships between the state and the individual, of levels of social cohesion and of conflict and threat. However we have not analysed extreme scenarios where we see anti-technology refusenik cultures, disenchantment with a technology (e.g. as a result of successful attack of key systems, their oppressive use by the state, breakdown of complex systems such as financial markets), or scenarios of extreme state control. Nor have we tested the robustness of the work to developments in quantum computing.

3 Part 2: Towards a security/resilience evaluation framework

3.1 Trends in evaluation – the evaluation context

The achievement of security is mediated by a complex interaction of politics, regulation, culture and markets. The products that are available, the trust and threats, the methodologies used, the incentives to co-operate or compete are all shaped by markets. If we ignore or misjudge markets, information assurance and security evaluation initiatives can be blunted or have widespread unintended consequences.

For markets in security to operate effectively users and suppliers need to be able to evaluate the risks, costs and benefits of alternative courses of action. Even if we are not concerned about markets as a whole, an individual, an SME or large corporate as well as vendors and system integrators and security policy makers all need to evaluate alternative courses of action[5]. There is value not only in being secure, but also in knowing that you are secure, i.e. value of confidence. If we could assess and communicate security risk in the same manner that other risks are handled then there is the potential for:

- Awareness and discussion of appropriate risk based policies;
- The ability to make trade offs between different security risks and between different types of risks;
- A more refined approach to the insurance of digital risks.

Our current inability to make such risk assessments can be seen as contributing to a market and policy failure. We frame this failure as an inability to make informed risk-based decisions on the costs and benefits of security. In terms of the interaction of economics and security, there is a strong interdisciplinary academic research community and we provide a brief review in section 3.2.2.

However security is just one aspect of overall dependability that users and the market might be concerned with: typically users want an overall and comparable view of other identified risks. We therefore propose to consider this in terms of a risk based approach to resilience and introduce the concept in section 3.1.1.

We also propose to draw on the work of other sectors in evaluation of communication of risks of computer based systems. While it would be false to present the work in other attributes (such as safety) as having solved the problems of achieving and evaluating risk there are both frameworks and detailed technical aspects that could be readily adapted or deployed in security. For example, in the use of "assurance cases" [20], barrier models, worst case reliability bounds etc. We therefore briefly review these approaches in section 4.2.

It is not that risk assessment and security are strangers. The BS1779 2001 standard provides explicit normative guidance on risk assessment that is also extremely generic and could apply to any number of risks. The problems arise because there are particular uncertainties in these risk assessments and any IA policy should be constructed with a view to how to deal with these uncertainties.[6] One obvious candidate is some form of probabilistic risk assessment.

One oft cited reason for the inapplicability of risk and probabilistic approaches is the very nature of security. The most obvious uncertainty concerns the behaviour of socio-technical systems: we cannot say with certainty when and how they will be attacked and whether they will fail and in what manner. The problem is not only predicting the behaviour of an

[5] This is rather a simplification. Market failure and lack of information can benefit some in the market. Fear, Uncertainty and Doubt (FUD) is a recognised commercial strategy.

[6] We have in some sectors the problems caused by adopting goal, risk based, regulation without any technical basis for how this will be implemented.

intelligent, malicious adversary but also the overall complex dynamic system that they are attacking. There might also be some cultural aspects concerning the IT security community favouring deterministic rather than probabilistic reasoning - for example in the way risk and uncertainty is dealt with within security standards. The concept of risk underpins all of this, in both security and safety publications. The safety community seems more fluent in expressing this risk. Although risk is the combination of consequence and probability the security standards are very coy about mentioning probability. For example: BS 1779-200 has 114 entries for risk and non for probability. HMG1 has 36 entries for risk and non for probability, 1 for likelihood. HMG19 has 69 entries for risk and non for probability, 3 for likelihood. There are 133 "risks" in BS 7779-2:2002 and no entries for probability but 2 for likelihood

It is now generally accepted that this uncertainty is best represented using a probability calculus. Whilst there are alternatives to probability as means of expressing this uncertainty, such as fuzzy/possibility theory, these do not have the power of probability, nor do they easily fit into a wider engineering framework, in particular for enabling quantitative risk assessment. The necessity for probabilistic formulation of security claims is elaborated in [21]. It is certainly true that predicting when a particular attack will occur is very difficult, but by analogy with insuring against crime, aggregation of behaviour can lead to the ability to predict probabilistically and can be seen as an emergent property.

This kind of uncertainty has been called "*aleatory*" (or, sometimes, stochastic), and concerns uncertainty in the world. What is less well accepted is the uncertainty involved in the assessment of dependability. The uncertainty here e.g. uncertainty about the accuracy of claims for a system's dependability arises from incompleteness and inaccuracy of evidence, doubt about the truth of assumptions, and other non-physical causes. This second kind of uncertainty has been called "*epistemic*" (or, sometimes, subjective, or state of knowledge) as it arises from incompleteness of knowledge [22]. This uncertainty impacts upon the confidence with which we can make information assurance claims. Although there is a clear need to deal with uncertainty formally, there is a lack of theoretical and empirical work on fundamentals (although see some earlier work [23]). Relevant work is work on economics and security (section 3.2.1) and the interdisciplinary work on confidence (section 3.2.2).

While advocating a probabilistic approach this needs to be done with a mature appreciation of the benefits and pitfalls of such a policy. As with any policy there can be unintended consequences: in probabilistic risk assessment any quantification of the risk can be used and abused. In policy terms numbers can easily travel long social distances but the caveats and understandings surrounding them get left behind, so that bare figures of 10^{-n} failure can be misleadingly quoted. There is already considerable Government guidance in [24] on risk communication in general, on risk management [25] and in the Information Assurance Governance Framework [26].

Additional policy risks of probabilistic approaches are the unintended redistribution or shifting of risks. A market might operate to reduce easily addressed risks at the expense of higher end risks that can not be so readily communicated and might even be increased by reducing the low end risk. For example, if the lifetime cost of security products is low and the perceived threat high there might be the temptation to install myriad firewalls, virus checkers, IDS that might introduce some new but very unlikely routes of serious compromise.

There are a variety of other issues that would have to be considered:

- ➢ The result of risk assessment of socio-technical systems is very dependent on where the boundary is drawn (e.g. for e-voting [27]).
- ➢ There are interesting balances between trustworthy and trusted: an intended consequence of making an e-voting system so secure that voters do not comprehend and so do not trust the system as much as a less secure one [28].

- Risk communication is a well trodden path and any security work would need to take best practice into account (as we expect it does, see [24]).
- Risk assessment relies heavily on expert judgment.

These are general issues for risk assessment. Peculiar to security is the intelligent and well resourced adversary and any policy appraisal would have to consider the implications of the adversaries knowing the basis of policy and how this might manifest itself (e.g. manipulation of markets, introducing a new type of security attack on the confidence in the risk assessment). Hence the evaluation needs to be two-pronged: evaluation of the adversaries (their strategies etc.) and the evaluation of the system defences.

3.1.1 Resilience

One interesting trend in evaluation and policy is to consider an "all hazards" approach that addresses both malicious and accidental attacks on systems (e.g. in EU CIP Directives [29]). In addition the notion of dependability, or dependability and security, as an umbrella term to capture the need to address all attributes (safety, security, availability etc) rather than just a single one and the use of the term "resilience".

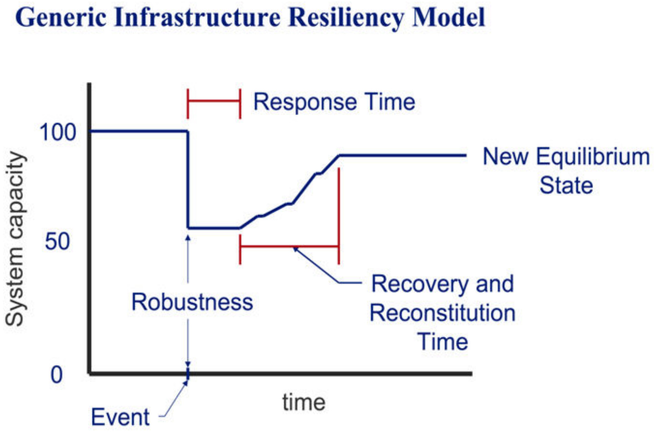

Figure 1 - Resilience

Resilience has its common sense meaning but is used in a variety of ways. The US Department for Homeland Security (DHS) and UK resilience viewpoints consider the loss due to an incident as an indication of how resilient a system is. This is shown in Fig 1. In [30] the emphasis is on the ability of a system to adapt and respond to changes in the environment. We propose to distinguish two types of resilience:

- *Type1: Resilience to design basis threats*. This could be expressed in the usual terms of availability, robustness, etc.
- *Type 2: Resilience to beyond design basis threats*. This might be split into those known threats that are considered incredible or ignored for some reason and other threats that are unknowns. (The Rumsfeld part of resilience)

A security evaluation could then be seen as evaluation of resilience for certain threats (e.g. malicious ones) and for certain attributes (confidentiality, integrity, availability). This evaluation of the security part of resilience would then address the different stages of Figure 1. See Table 1 for some initial details for critical infrastructures.

Table 1 – Security assessment of threats to Critical Infrastructure Assurance

Impact on resilience
Security assessment of threats to Critical Infrastructure Assurance (CIA) during the following phases: ➢ Reduce frequency of events ♦ Warning, operator support ➢ Increased robustness ♦ Network design, topology, redundancy, diversity ♦ Understanding of events and scenarios ➢ Detection time ♦ Communication between services ♦ Variety of forecasting approaches ♦ Detection of compromises ➢ Decision time ♦ Situational awareness ♦ Planning and training (Scenarios) and use of synthetic environments ➢ Recovery time ♦ Resource deployment; dependent assets identified ♦ Awareness state of other networks ♦ Communication and co-ordination ➢ Learning from experience

3.1.2 Evaluation and cases

For critical systems it is important to know whether the system is trustworthy and to be able to communicate, review and debate the level of trust achieved. In the safety domain, explicit Safety Cases are increasingly required by law, regulations and standards. Increasingly, the case is made using a goal-based approach, where claims (or goals) are made about the system and arguments and evidence are presented to support those claims. The need to understand risks is not just a safety issue: more and more organisations need to know their risks and to be able to communicate and address them to multiple stakeholders from the boardroom to back office and beyond. The type of argumentation used for safety cases is not specific to safety alone, but it can be used to justify the adequacy of systems in different applications, including security critical, business critical or service critical. An international community has begun to form around this issue of generalised assurance cases and the challenge of moving from the rhetoric to the reality of being able to implement convincing and valid cases [20], [31].

The "case" and associated supporting tools can be seen as having a number of roles:

- *Reasoning and argumentation*: as an over-arching argumentation framework that allows us to reason as formally as necessary about all the claims being made. Here there are two very different viewpoints: the one that sees argumentation as primarily a narrative and the other where we seek to model judgements in a formal framework. There are some hybrid approaches where the case can be seen to integrate and communicate a selection of formal analyses and evidence, e.g. it would not seek to reason formally about the timing of a component but leave that to a separate analysis. The balance between these two approaches should be part of on-going research.

- *Negotiation, communication, trust*: as a boundary objective between the different stakeholders who have to agree (or not) the claims being made about the system. To this end it has to be detailed and rigorous enough to communicate the case effectively and allow challenges and the subsequent deepening of the case.[7]

The assurance case contains a number of strategies that can be explained in terms of a "triangle" (see Fig. 2 below) of:

- The use of accepted standards and guidelines.
- Justification via a set of claims/goals about the system's safety behaviour.
- An investigation of known potential vulnerabilities of the system.

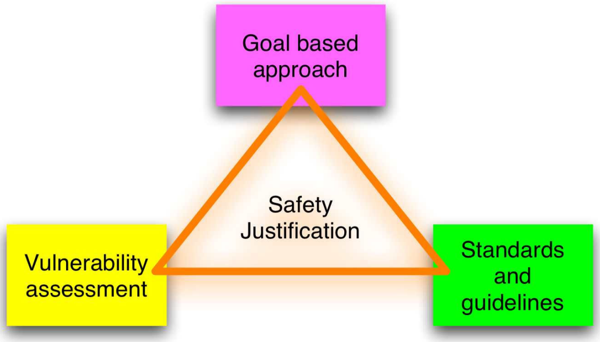

Figure 2: Safety Case triangle

The first approach is based on demonstrating compliance to a known standard—which is the approach that is normally used. The second approach is goal-based—where specific goals for the systems are supported by arguments and evidence at progressively more detailed levels. The final approach is a vulnerability-based argument, where it is demonstrated that potential vulnerabilities within a system do not constitute a problem—essentially a "bottom-up" approach as opposed to the "top-down" approach used in goal-based methods.

These approaches are not mutually exclusive, and a combination can be used to support a safety justification, especially where the system consists of both off-the-shelf (OTS) components and application-specific elements. While there is considerable experience in other sectors with the claims-argument-evidence (CAE) structure for safety justification, this type of structuring is novel to the nuclear industry.

There are perceived large benefits in developing and generalising the goal-based Assurance Case approach to security and critical infrastructure.

3.1.3 COTS components

Software systems are increasingly created through the utilisation of commercial-off-the-shelf (COTS) components. COTS components are used due to financial pressures on system developers to reduce costs and shorten the systems' development and delivery times. In this section we will briefly summarise the main characteristics of COTS components, problems with COTS components assessment, state-of-the-art in COTS assessment and the research challenges. We have provided a more detailed review of COTS assessment in Appendix C.

[7] An engineering analogy is to see the case as part of a "signal processing system" – the licensing and certification process – that seeks to reject false claims and accept good ones. For critical systems the probability of false positives has to be very low (i.e., there must be a very low chance of accepting a flawed system).

COTS components come in a variety of forms [32], from components that form part of a program (e.g. various graphical, statistical or mathematical libraries of functions) to complete systems of integrated software and hardware components (alarm systems, Programmable Logic Controllers (PLCs), medical devices, Enterprise Resource Planning (ERP) systems, air-traffic management systems etc).

The main characteristics of COTS components are:

- They already exist and cannot be re-engineered (exceptions are some open-source off-the-shelf (OTS) with less restrictive license agreements for which the code is available for changing).
- Due to their general-purpose use, most components may contain functions that are not necessary for a specific application.
- COTS with a substantial user-base are subject to continuous change and evolution to meet users' evolving requirements.

Contrasted with bespoke systems, there are several challenges that assessors are faced with when dealing with COTS components (more details in Appendix C):

- Non-compliance with standards
- Establishing COTS provenance
- Problems with unwanted/unneeded functionality in the component
- Problems stemming from patches and updates to component
- Difficulties with optimal COTS components selection

Several assessment approaches are reported in literature for dealing with some of these problems. There is for example a vast literature on COTS component selection (see Appendix C for a review) which ranges from selection and assessment of components for use in commercial settings to safety-critical settings such as medical, nuclear power and military.

Research challenges for COTS components closely mirror those for bespoke systems and SoSs - which will be presented in the next sub-section and in Section 4, even though there are some differences for COTS component assessment which we have highlighted in Appendix C.

3.2 Current research trends

3.2.1 Economics of security

An important interdisciplinary research area is in the field of *Economics of Security*. Two recent reports [33], [34] provide extensive details of this field and references to other important works. As stated in [34], "*security failure is caused by bad incentives at least as often as by bad design. Systems are particularly prone to failure when the person guarding them does not suffer the full cost of failure.*" Hence approaches such as microeconomics and Game theory for example, become important for engineers of computer security. Security mechanisms are used for such purposes as digital rights management and accessory control which introduces strategic issues. As stated in [34] "*where the system owner's interests conflict with those of her machine's designer, economic analysis can shine light on policy options*". In [33], the authors provide a set of 15 recommendations about what information security issues should be handled at the EU Member State level and what issues should require harmonisation (or at least coordination) amongst the member states (see the Executive Summary of [33] for a listing of these recommendations). Some of the research challenges in Economics of Security, listed in [34], [35] and [36] are:

- *Network topology and Information Security* – [34] cites several recent works investigating the network security and the nature of spread of viruses and other security risks, as well as strategies for defending from their spread. The authors state that it remains a research challenge to reconcile the generated network models (which have been done using various simulations) and real computer networks to get more accurate measures of the various costs/gains (e.g. from the defenders point of view, the gain in security that can be gained from a malicious node (or a group of them) being brought down/removed).
- *Psychology and Security* – Several issues are discussed in [34] and [36] as being at the crossroads of Security, Economics and Psychology:
  - *Inappropriate obedience* – e.g. card thieves calling up cardholders and pretending to be from a bank and demanding the PIN number from customers – how many give the PIN numbers away and why does this happen?
  - *Security usability* [36] – how to design easy to use and easy to configure security devices and security systems (especially those that are to be used by the general public)?
  - *Study of deception* –why is phishing successful?

Economics and security have been addressed at a series of academic workshops on the Economics of Information Security (WEIS) since 2002 and for CIIP there is a more recent Workshop on the Economics of Securing the Information Infrastructure [37]. Of particular relevance to security risk assessment are those related to cyber insurance [20], covert conflict [38], costs of infrastructure failure [39] and empirical papers on justifying security investments [40], [41], [42]. In addition there are studies of the security market, and economic models of vulnerability research [43], the vulnerability black market [44] and forensics [45]. There is work that makes an analogy with the statistical value of life used in safety policy and proposes a similar measure for the Statistical Value of Information and there is other empirical work on the consequences of privacy breaches [46]. In the area of CIP a number of studies have adopted the Leontieff approach to model interdependencies and also economic impact [47] and this is related to military effect based operations where economic impacts are an important part of planning certain interventions short of war [48].

The active multi-disciplinary research community in economics and security and the work on security and risk should, on the one hand, be encouraging but also underline the research challenged that is faced.

3.2.2 Interdisciplinary approach to assessment of SoSs

[6] states that "*the breadth and depth of the application of innovation will generate an unprecedented reliance on technology. Increased access to and the rapid cultural assimilation of technology will render future generations increasingly vulnerable to the deliberate or unintentional disruption of technology-based utilities. Sophisticated societies are likely to be more vulnerable to such disruption as they increasingly exist in a virtual environment in contrast to less technologically advanced societies*". Hence the need to *build* dependable and secure systems will remain but the difficulty of *assuring* that deployed systems are sufficiently dependable will also increase. Due to the complexity of the envisaged systems discussed in section 2.2 and 2.3, which are required to work in highly interoperable, interconnected and pervasive manner, it will remain a challenge to research how well the current approaches to security and dependability assessment can scale to these new systems. Clearly the assessment will need to be *interdisciplinary*, bringing in research from computer science, statistics, psychology, sociology and ethnography amongst others to assess the multi-faceted nature of these emerging systems. At CSR, City University we have been involved in a large-scale long-term project called DIRC (Dependable Interdisciplinary Research

Collaboration) sponsored by the UK EPSRC which has investigated these issues for current technologies. This research is continuing in the successor project to DIRC called INDEED (INterdisciplinary DEsign and Evaluation of Dependability) which is also sponsored by EPSRC. We will discuss the inter-disciplinary approach to assessment in section 4 of this report according to the themes of:

- Adaptation and diversity
- Confidence in dependability cases
- Responsibility and trust
- Time and structure

3.2.3 Emergent properties in complex and adaptive systems

The research challenges outlined in sections 2.2 and 2.3 highlight the need to deal with assessment of *complex* and *adaptive* systems. As stated in [49] "*As our world becomes a more interconnected place, so called "systems" ideas and perspectives become increasingly important. A central issue is the* ***emergent behaviour*** *of complex systems. In complex systems, non-linear interactions between component parts give rise to high-level "****emergent****" organisation that is not straightforward to explain*." In the same report [49], it is explained that a large class of natural complex systems (e.g. neural systems, ant colonies, animal matting habits, stock markets etc.) exhibit aggregate (emergent) properties that allow them to *adapt* to changing circumstances in an efficient and effective manner despite lacking central authority or control responsible for this ability. Therefore such systems can be very *robust* to perturbation and also *behaviourally agile*; ICT engineers would like to design both of these properties into their technical systems. A cognitive sciences perspective [50] outlines four approaches to pin down the notion of emergence, namely:

- Collective self-organisation
- Un-programmed functionality
- Interactive complexity
- Uncompressible unfolding (i.e. behaviour of a macrostate requires simulation of the microstate)

Research challenges remain however to better define, understand and explain these notions of *emergence* in complex ICT systems. Often *emergence* is used loosely, and incorrectly, to describe behaviour that has been surprising.

CSR has conducted research on evaluating the structure of COTS and applying some ideas from complexity theory to understand what might be general topological properties of certain classes of COTS [51]. A promising line of research would be to expand this to different classes of evolving COTS and to develop evaluation models based on it. There is also interesting work in the EC-funded IRRIIS (Integrated Risk Reduction of Information-based Infrastructure Systems) project (see next section), of which CSR is a contributing partner, on applying complex systems style modelling to critical infrastructures [52].

3.2.4 Interdependencies

The challenge of understanding the behaviour of critical infrastructures and particularly information infrastructures is well known and a variety of national, European and US research initiatives and workshops have been actively considering the research agenda posed by this challenge. In the UK, the emerging Government Information Assurance Strategy [53] and within Europe FP7 [54] have both identified the need for research on interdependencies. One important aspect of infrastructures is their interaction and interdependency. Unforeseen interdependency can be a source of threat to systems and a dominant factor in our ability to understand risk. However interdependency is also central to providing tolerance to attack and failure, a means for adaptation and overall resilience.

Interdependency analysis can be undertaken at a variety of phases: from planning and feasibility through to emergency or situational management. Interdependencies are sometimes considered according to the different perceived layers (e.g. of physical, control and supervisory, management) and also in terms of abstraction such as effects, services and implementation. For each of these abstractions there are a wide range of possible modelling approaches and theories that can be deployed ranging from qualitative models, stochastic activity networks, complexity science style models and high-fidelity simulation. These can be deployed at a variety of abstraction levels e.g. to model the detailed implementation topology or to model the service topology and cascading effects.

The interdependency analysis needs a sufficiently rich model for the analysis to discover and assess the risks:

- Societal aspects need assessment as they provide possible hidden sources of commonality.
- Modes of operation have to be rich enough. Degraded modes of operation can amplify risks as levels of redundancy assumed at design time become defeated.
- Non-linearities in failure models (e.g. increased failure rates due to stress from nodes in the same locality) can lead to escalation and cascading effects.

Some of the IRRIS results can be found at [52]. Of particular interest are the models of cascade and common mode failure arising from Task 2.1 of IRRIIS. Currently CSR is leading a study on interdependencies between critical infrastructures [55].

3.2.5 Formal methods and static analyses

The contemporary static analysis and formal methods landscape is a rich one covering a range of approaches to the analysis of software code, designs and specifications, and it encompasses:

- Theorem proving of proof obligations generated from comparing a safety or security property or specification with an implementation
- Proof that unwanted behaviour such as deadlocks and divide by zero does not occur (this is independent of specification)
- Tools to extract and understand program structure
- Approaches to guide generation of test cases and techniques that link static and dynamic analysis

There have been tremendous advances in the capability of theorem provers and model checkers since the early industrial experiences (see for example Rushby papers [56], [57], [58]). There are a numerous tools for program slicing and flow analysis (Codesurfer and path inspector), model checkers (nuSMV, Spin, SAL), abstract interpreters (Polyspace, Astree), and integrity analysis tools (Coverity, ESC) and tools for verifying C programs (Spin, Why) (a list of useful links can be found in [59]). Microsoft has developed focused tools for attacking particular problems such as buffer overflow and in the embedded area there are tools such as SCADE which have plug-ins for model checking and other techniques [58]. Although there are many tools available, their use is often not off-the-shelf and will require additional engineering to cope with language variants and the particular processors used in embedded systems. In addition they may not address all the attributes that are of interest e.g. some will address only integrity issues but not correctness. There is continuing work on correctness by constructions such as that within the RODIN and now DEPLOY projects.

The rich landscape of static analysis means that many approaches may now be cost effectively used at lower integrity levels and that recommendations in standards, if based on cost/benefit considerations, could be out of date and not accurate. Examples of such shifts in the costs are give by the US Department of Homeland security sponsored work on integrity checking of many open source programs[8] [60] and the work done running proofs of the Mondex purse that showed that mechanical proof could now be added for an extra 10% of the overall effort on the less formal manual proof [61]. However, in part because of the rich landscape, we need to be clear how the techniques contribute to the safety argument being made e.g. removal of vulnerabilities, removal of all faults of a certain class, demonstration of absence of certain behaviour, proof of correctness properties.

Progress in the development and deployment of these techniques will have impacts on various aspects of evaluations:

- Existing security standards recommendations will be superseded as the techniques become cost-effective and their deployment appropriate and proportionate.
- The claims made in the security cases might change (e.g. to all vulnerabilities of class X have been found, when tools and techniques for finding X have been deployed) and the evidence with it.
- Threat models and likelihood need updating e.g. finding X now easier given the advance in tools.
- More generally the "arms race" between deploying these tools on products and their use by adversaries needs assessing.

3.2.6 Benchmarking and fault-injection related testing

There has been a lot of previous work on fault injection, robustness testing and more recently on dependability benchmarking. A review of state of the art and current research trends in these fields are given in [62].

Benchmarks for performance of computers systems (especially database management systems [63]) have existed for over twenty years. More recent attempts in the last decade have been made to define dependability benchmarks (see [64] for dependability benchmarking of operating systems, and [9] and [62] for a review of the dependability benchmarking field). Dependability benchmarking of a system involves the evaluation of dependability and/or performance attributes of a system either experimentally or with a combination of experimentation and modelling [64]. Dependability benchmarking combines the *workload* defined by existing performance benchmarks (e.g. TPC-C [63]) with a *faultload*. The faultload defines the types of faults that are used with the workload to derive dependability measures for the system.

Fault injection also involves the definition of a faultload and can be used for several purposes, for example to:

- Assess the fault coverage of implemented (either software or hardware) fault-handling mechanisms.
- Assess the error propagation and error latency of a system.
- Assess the response time of system recovery following a failure.
- Assess and verify failure mode assumptions of systems, sub-systems and components.

[8] Coverity found an average of 0.434 bugs per 1,000 lines of code in 17.5 million lines of C code from open source projects

The *representativeness* of both the workload and especially the faultload remains a key issue for both dependability benchmarking and fault injection as well as robustness testing, i.e. how well the faultload and workload represents the typical use of the target system (also referred to as "operational profile" [65] in reliability growth modelling). The definition and *representativeness* of the *faultload* is considered the most difficult part with these techniques.

3.2.7 Modelling and model-based assessment

[66] defines a model as "*an abstraction of a system that highlights the important features of the system organisation and provides ways of quantifying its properties neglecting all those details that are relevant for the actual implementation, but that are marginal for the objective of the study.*" Various modelling approaches to system dependability exist (*non-state-space* models such as *reliability block diagrams*, *fault trees* and *reliability graphs*; and *state-space* models such as *homogeneous continuous time Markov chains*) and the choice of a model for given system and a given context depends on many factors, such as the complexity of the system, the dependability attributes to be evaluated, the accuracy required, as well as the resources available for the study.

As stated in [62] "*the modelling and analysis of complex (large, dynamic, heterogeneous, ubiquitous) systems still needs continued research, both in model construction and in model solution. A crucial point in this context is also to assess the approximations introduced in the modelling and solution process to manage the system complexity, as well as their impact on the final results*."

Regarding the largeness, dynamicity, heterogeneity and ubiquity that we have discussed in section 2 [62] states: "*The role of modelling in a more comprehensive assessment process is, on the contrary, not well addressed in the literature. The largeness, dynamicity, heterogeneity and ubiquity of current computing systems actually calls for the development of a composite and trustable assessment framework including complementary evaluation techniques, covering modelling and experimental measurements. Mechanisms are needed to ensure the cooperation and the integration of these techniques, in order to provide realistic assessments of architectural solutions and of systems in their operational environments".*

A thorough overview of modelling approaches, strategies for building them as well as tools supporting model-based assessment is given in [62].

3.2.7.1 Reliability and availability evaluation and modelling

There is a wide variety of techniques for modelling the reliability and availability of systems. Guidance and lists can be found in IEC standards, NATO reports and elsewhere (see also Appendix B of this report). The handbook of Software Reliability Engineering [67] gives a comprehensive guide.

The reliability of the system or product (both during its development and use) can be evaluated using various reliability modelling techniques, such as:

- *Reliability growth modelling* – modelling the reliability of the system dynamically during the system development or use where faults are being fixed/removed from the system, hence the trend is usually that reliability will *grow* in the long term (but not always in the short term as a "fix" may introduce a new potentially more harmful bug in the system leading to reliability decay).
- *Statistical testing* - which involves the creation of a test harness to perform the testing; an 'oracle' against which the results obtained from the system under test are compared; and an accurate definition of the 'operation profile', i.e. the profile under which the system is expected to be used.
- *Evaluation of field experience* – using field data (tracking and recording faults and incidents in a product or system) to gain insights into the reliability of the product.

Difficulties with getting accurate measures stem from the difficulty in measuring the operational profile of the products in their given installation (however, some companies may offer incentives to users (such as cheaper products) if they are willing to allow the vending company to collect detailed data about the product usage, which can improve the quality of the data collected).

- *Reliability block diagrams* – modelling, in a diagrammatic form, the chain of events that are necessary for the successful operation of a system.

Reliability growth modelling is the most useful of the techniques listed above when arguing about software reliability as it allows an organisation to monitor and argue about the reliability of the system as it is being developed (or, if it is in operation, as it is used). *Statistical testing* is also a very useful technique to estimate the reliability before a system or component is deployed in operation (but the operational profile is crucial to the accuracy of the results obtained from statistical testing). A plethora of Reliability models exists (some better than others) and there are techniques (developed at CSR, City University [68]) which can help with *recalibrating* the results of the reliability growth models.

The issue is how and whether these can be deployed in the security area. There is some research reported in the literature of trying to apply these to security vulnerabilities. Research challenges are in the following areas:

- The conceptual aspects: definitions of vulnerability/defect; consideration of time and effort or other surrogate measures, consideration of appropriate measures and the need to partition the threat or risk space and then combine into an overall measure.
- The applicability and scope of different models and how they might be developed to be applied to security vulnerabilities.
- The problems of data and application of the models.

There is some recent work reported in [69-71], [72], [73].

4 Part 3: A Future Resilience Evaluation Framework

In this section we first summarise the scope and design of the proposed security framework and then group the problems and research challenges based on the themes of: *confidence/trust*, *diversity and heterogeneity*, *complexity and emergence*, *structure*, *resilience*, *adaptation and tempo*, *communication* and *markets*.

4.1 Introduction

The challenges of current and future systems and the anticipated threats that they face leads us to propose an ambitious shift in perspective to a framework that attempts risk based, market sensitive, psychologically aware, evidence based approach to evaluation and communication of security as part of an overall resilience case. The approach must be capable of addressing the scale and complexity of adaptive systems of systems (p-e-s-t-systems) that have heterogeneous human and technical components and are deployed across a variety of organisational, political and legal boundaries. The framework should address the scale and tempo imposed by threats, operations and system evolution.

4.2 Resilience and security cases

We propose that resilience and security cases should form the overarching approach within which we would address: *resilience requirements*, *claims decomposition*, *arguments*, *scalability* and *tempo*. The argumentation will be about a service or system in a particular environment and should adopt a claim-argument-evidence assurance approach.

The claims should be based on a coherent a set of underlying models and theories. These models would give meaning to and support the evaluation; these will be disparate, multi-formalism models. We need to understand their interrelationships, the required abstractions and levels of fidelity. They will also include models of the system environment particularly economic and threat model - this is discussed in more detail in the next sub-section.

A resilience case would be a structured argument based on *assumptions* and *evidence*, which supports a *claim* that a system meets its specified resilience requirements at a particular level of *confidence*. We propose that a flexible assurance approach should be adopted based on strategies that justify claims of behaviour, the absence or adequate mitigation of vulnerabilities and sufficient compliance with standards. This should build on current research emerging in safety cases, especially from the nuclear industry.

As noted above, the concept of *safety cases* has existed as a core requirement of safety standards for several years. Consolidation and applications of “cases” to security are still in their infancy and there are even research challenges to apply them to present systems and even more so to the systems envisaged in this report. Some of these research challenges are:

- *Claim definition and decomposition*. The first problem faced by someone building a dependability case is to decide exactly what is to be claimed, and the detail of how such claims should be expressed. There is a need to develop a core claims language for expressing dependability, security and resilience claims and a method for structuring the assurance case. Mechanisms are need in the claims language to support scalability and modularity of viewpoints and security policies. Claims need to address both technical and non-technical attributes of the system.
- *Resilience and security policies.* The evaluation will also require developments in dynamic, adaptive security and resilience policies that can form the basis of the top-level claims about the system.

- *Cascade failures*. There is a need to extend claims to complex systems properties and be able to express claims about cascade, common mode failures and interdependencies and define and evaluate credible supporting technical arguments usable for SoSs.
- *Confidence*. There is inherent uncertainty that needs to be handled in resilience cases: we cannot claim that dependability claims are true with certainty. Such uncertainty arises from many sources: doubts about assumptions upon which a dependability case will be founded (e.g. correctness of an oracle for a case based upon testing); strength of evidence (e.g. number of test cases, and number of these that are correct) etc. Central to this is the uncertainty of "knowledge about the world". Typically this involves an inevitable subjectivity and is harder to handle than "natural" uncertainty or randomness (such as that involved, for example, in a dependability claim itself). There is a need for a formal, rigorous (claim, confidence) calculus that supports quantitative and qualitative dependability arguments and supports the description and propagation of confidence in claims.
- *Robust and diverse arguments*. Similar to the concept of design diversity for fault tolerance, diverse arguments [74] are an approach to gain confidence in dependability claims by using multiple diverse 'legs' to reason about the dependability of a system (e.g. *formal verification* and *testing*). Further research is needed on the applicability of these tools and the complex dependencies that will exist between the uncertainties in the argument legs: for example, a dependability claim supported by two arguments legs, each of which singly incurs 10% doubt in the truth of the claim, does not directly translate into a 99% confidence from the two legged argument. The applicability of diverse arguments to complex systems of the future needs to be further investigated.
- *Dynamic cases.* Current approaches to cases are static, can take a long time to develop and could not cope with the tempo envisaged. Research is needed to how this will impact the cases approach, the drivers for more automated reasoning and the balance between the analyst and the advisory systems. As stated in [7], the assessment should most probably move from *off-line* and *pre-deployment*, to *continuous* and *automated operational assessment*. The assessment should include *validation* and *verification* approaches as well as *quantitative* and *probabilistic* approaches. This is related to benchmarking and obtaining good metrics which are discussed in the next two subsections.
- *Scalable cases.* While there are examples of safety cases for very large systems it remains a challenge to develop convincing rigorous cases with a surrounding process of challenge and evaluation of complex systems. Progress is required in the models used to support the evaluation and in methods of composition of system cases into a SoS case.
- *Communicable cases*. Security should be evaluated in user or stakeholder terms and this could be either with respect to some expected service (e.g. a service is available but respects confidentiality of assets) or with respect to the integrity and confidentiality of assets. The framework will operate with different level of security attached to different parts of it: the reasoning of one part to another needs to be sufficiently exposed to be communicated and used but so that security itself is not compromised. This will require advances in security policies for cases themselves.
- *Standards.* The framework should recognise the importance and limitations of standards. It should build on existing standards and frameworks that cover the conduct of evaluations and the gathering and trustworthiness of evidence.
- *Information and meta-data* will be important assets to be addressed by an assurance case and sources of threats and vulnerabilities.

Explicit in the idea of resilience is the need to allow security to be evaluated and communicated for the following phases of the resilience lifecycle:

- Pre-event
- Gestation
- Initiation
- Detection
- Recovery
- Learning and adaptation

There need to be appropriate measures and methods for describing the risk requirements during these phases. There is a need to understand how to formulate the security and resilience requirements and how they might change with threat levels (societal state), and how to address the multiple tradeoffs (e.g. between attributes, between phases of the resilience curve) and uncertainties.

The framework must provide timely and valid judgements to inform decision makers in terms of confidence in security claims. Timeliness is especially important in the recovery phase (for example, in the power grid scenario presented in Appendix A of this report, uncertainties about responsibilities within the organisation caused confusion and delays in the recovery phase following the breach).

Risk communication should allow informed discussion of security requirements between the myriad of stakeholders and the trade offs between different aspects of the resilience curve in both evaluation and design activities. We envisage a need for pubic debate, as there has been for tolerability of safety risks, to determine and articulate acceptability of risks from the emerging p-e-s-t systems. There may be a role for concepts such as ALARP (i.e. risk has been reduced to "As Low As Reasonably Practicable) to be applied to resilience (for instance ARARP - As Resilient As Reasonably Practicable).

The design and deployment of such a framework should be aware of how economics of security can be made to support the approach, and how markets may inhibit it. There is a need to understand how markets do/don't deliver resilience (e.g. how they might optimise the ability to defend against mild attacks), how the common good might be served by maintaining more resources for recovery but is not optimal for a single organisation, and how insurance might help or hinder resilience. This should feed into both the design of the evaluation scheme itself and into analyses that should support regulatory or resilience policies.

4.3 System and environment models and theories

A coherent set of underlying models and theories is needed to add meaning to and support the evaluation: these will be disparate, multi-formalism models. Security is a property of the wider socio-technical system and the supporting modelling will need to take into account:

- The boundaryless and fluid nature of the socio-technical systems.
- That while a services-based view provides a good abstraction for describing requirements, faults and vulnerabilities break that abstraction.
- The need to address a variety of interfaces – technical, organisational, contractual, political, legal – and provide convincing stopping rules that somehow limit the analysis yet recognises the scale of the system.
- The heterogeneous nature of the socio-technical systems (systems of systems, nano, bio, human systems, different organisations, legal systems, provenance).

- The inevitable epistemic and aleatory uncertainties in structure and behaviour. The need to cope with an uncertain supply chain of uncertain provenance.

These are challenging and hard requirements to satisfy for systems of today. The evaluation approach has implications for architectures and middleware and for design for assurance (which is outside the scope of the present discussion).

4.3.1 Models for evaluation - complexity and fidelity

Applying model-based dependability and security evaluation to increasingly complex systems will lead to proportionally complex and large models. Model complexity is a general term which indicates that the model construction process and/or the model solution process are "difficult" to perform [7]. These difficulties come from the characteristics of a system, such as *heterogeneity* and *largeness,* which result in problems like *state-space explosion*, *stiffness* for the analytical model solution, *rare event* problem for simulation and *intrusiveness of the monitoring system* for the experimental evaluation methods. Hence continued research is needed in methods for both *model construction* and *model validation*. The models need to be able to assess the impact of *accidental* as well as *deliberate* threats to a system.

The strategy for handling the scale of the systems envisaged is quite straightforward:

1. Build larger models and run them for longer, enabled by advances in processing and software engineering
2. Use abstraction to capture the essential properties
3. Use composition and federation to combine models together
4. Use real-time data for calibration and validation

All parts of this strategy face challenges for the multi-domain, multi-attribute nature of the models. The response to scale and uncertainty in boundaries and structure can not just be to build bigger models but needs research to address how stopping rules and the factoring of the problem can be determined and justified.

The evaluation of resilience should be supported by a range of disparate models. The next few sub-sections discuss some of these modelling issues.

4.3.1.1 Selecting the most appropriate model and fidelity level for a given assessment task

Dealing with very large scale SoSs would make "observability" of the entire system problematic for any assessor. Simplified models of the system (e.g. in the form "my part of the system is modelled in detail, the rest of the system as a monolithic whole") are likely to be used. We already touched on the problems of dependence between the sub-systems earlier and the model complexity. Additionally, when several simplified models are applicable, selecting the best one is not easy. E.g. in software reliability growth modelling (see section 3.2.7.1 for a summary) it is known [68] that the best predictive model is impossible to know with certainty in advance: which of the available models is best (i.e. which one will give the most accurate predictions of the system's reliability) depends on the model itself and on the data to which the prediction is applied. Objectively assessing the model's predictive quality in the context of large systems operating in a changing environment is a particularly important and hard problem to solve.

There is a need to address a variety of interfaces – technical, organisational, human, contractual, political, legal – and provide convincing stopping rules that somehow limits the analysis yet recognises the scale of the system.

4.3.1.2 Modelling adaptation

Human behaviour is an inherent part of socio-technical systems. We need to develop approaches that allow us to factor what is the socio-technical part form the purely socio and so bound and limit the analyses required. We need to understand the myriad of psychological aspects of human adaptive behaviour. The malicious aspects of this are addressed in the threat modelling, the macro-behaviour in terms of economics and markets but we also need to address the user or stakeholder adaptation to systems.

As we mentioned before, there is a growing awareness that (almost) all systems are socio-technical in nature. The role of humans in the overall dependability and security of systems needs much more sophisticated modelling than it has received so far, especially in the area of *confidence* in *claims* that are made (see *Resilience and security cases* section 4.2 above). Early results [75] suggest that there may be interesting aspects of diversity between humans and "computers" – e.g. they may differ usefully in what they find "difficult" in certain problems.

4.3.1.3 Inter-organisation boundary models

Many failures, especially for complex socio-technical systems are caused by inappropriate responses to communications across the boundaries between organisations (or between different departments within an organisation). Messages may be misunderstood and cause behaviours (through either action or inaction) that may be correct in the view of the recipient of the message but different from that which the message was meant to elicit.

Some of these failures fall in the category of "*responsibility failures*" (the term was coined in the U.K. DIRC project [76], of which CSR, City University was a partner). Responsibility failures are bound to get even more prevalent in the complex systems of the future where boundaries between the systems may be very blurred. Hence it remains a research challenge to study responsibility modelling for the future systems with the aim of reducing responsibility failures.

Inter-organisational models and theories are needed to assess recovery in multi-stakeholder, multi-jurisprudence situations.

4.3.1.4 Interdependencies and cascade failures

The resilience case should be able to express claims about cascade, common mode failures and interdependencies and define and evaluate credible supporting technical arguments usable for SoSs. There is a need for supporting theories and models that address:

- the nature, mitigation and recovery from cascade failures and interdependencies and how “coincidence” and large variations can occur (c.f. the debate from Mandelbrot and others on markets and “fat tailed” distributions[9])
- the balance between the key role that redundancy and diversity has in achieving resilience and the side effect of introducing unwanted vulnerabilities through interdependencies between systems and components

4.3.1.5 Interaction of markets, policy and security evaluation

Any evaluation framework that is deployed will have an impact on, and can exploit the interaction of markets, policy and individual behaviour. This needs to be understood to prevent unintended side effects of the policy and to identify effective methods for deployment (see section 4.2 for some details of what is required.)

The interactions could be:

- Impact on individual behaviours (intersection of psychology and security).
- Market behaviour (aggregated impact of the risk communication).

[9] Not to be confused with markets and fat cats ☺

- Implications for distribution of trust and security within a population for different scenarios (e.g. is there an analogy between wealth and trust distribution in a population? Can we adapt these economic models to trust and confidence?).

4.3.1.6 Threat models

The evaluation will only be valid with respect to certain threat models. The so called Design Basis Threats (DBT) should address:

- long term, informed, capable and patient adversaries throughout the supply chain.
- co-ordinated attacks and simultaneous attacks that may be spontaneous, opportunistic and loosely coupled through ideology.
- multiple events; recognising that events are not necessarily independent and beware of "fat tails" i.e. ignoring low frequency large events as judge erroneously improbable.
- the importance of non-physical assets and associated attacks, e.g. on confidence, situational awareness etc.
- the continuing role of insiders.

The threat model is closely related to the assurance strategy. For example an assurance strategy might make pessimistic assumptions that allow a relatively coarse threat model to be used. It may also argue levels of protection that are not dependent on any particular threat but rely on detection and recovery.

The threat models need to be consistent with and interwork with other environment models that are part of an all hazards approach. For example the credible environment challenges and resulting degraded modes and stresses on the system will impact the threats and attacker behaviour.

4.3.2 Defining the evidence base and obtaining good metrics

It is not that we are short of security metrics ([77] has some 900) but we are short of those that allow us to measure and predict operational measures of risk with a known confidence that the stakeholder can use to make informed decisions. Obtaining good metrics which would allow for a more dynamic real time view and assessment of security and dependability of the emerging and future systems or SoSs remains a challenge. Examples of the type of metrics which would be useful include: the lag between the release of a patch and its installation; the lag between new viruses and matching detection capabilities etc.

Automatic tools for data collection would help with providing high quality data which would allow for more *dynamic online* assessment of the dependability or security of a running component or sub-system. A difficulty will be to detect and collect "non-self-evident" (i.e. non-crash) failures. Recent work we have done with SQL database servers [78] suggest that a large proportion of known bugs of these servers cause non-self-evident failures (more than 50% for most servers in our study). For these types of failures our study showed that *design diversity* (using more than one diverse database server) or *data diversity* (sending syntactically diverse but semantically equivalent requests to the servers) has very good potential for improving fault tolerance (both failure detection and diagnosis of the failed component).

Obtaining good quality data may also support the development of efficient early warning intrusion detection systems, which will help the administrators to react efficiently to attacks.

4.4 Evaluation of the approach

The framework should have properties, such as:

- It should be justified and validated (unlike many current standards).

- It should be repeatable and trustworthy but recognising the importance of human judgements.
- It should be adapted as threats adapt to how the evaluation works.
- It should be applicable in a graduated manner.

In fact the assurance case approach can be applied reflectively to justify the framework itself.

4.4.1.1 Experimentation and benchmarking

Measurement and benchmarking of dependability and security of computer systems for their comparative evaluation is also an important research area. Performance benchmarks, such as TPC-C [63] benchmark for transactional database systems, are well established. Definition of dependability and security benchmarks on the other hand is still in its infancy (some initial work on dependability benchmarking is reported in [64]) and more research will be required to establish: to what extent are dependability and security benchmarks of SoS usable and what are their limitations; what are the limitations on applying them to large scale systems of the future. CSR, City University is a partner in a EU-funded Assessment Measurement and Benchmarking of Resilience (AMBER) [10] project which, amongst other aims, is researching the issues of benchmarking.

Work is required on the interaction between field experience, experimental evaluation, synthetic environments and simulation. It is closely related to the need for model validation and justification for the compositions that will be required to justify SoS.

4.5 Issues not addressed

There are two important issues that we have not addressed and their implications for the framework should be considered further. They are: *extreme scenarios* and *quantum computing*.

4.5.1 Extreme Scenarios

We have not analysed in this report extreme scenarios where we see anti-technology refusenik cultures, disenchantment with a technology (e.g. as a result of successful attack of key systems, their oppressive use by the state, breakdown of complex systems such as financial markets), or scenarios of extreme state control. We propose that the impact of these on the design and implementation of the framework should be assessed and the role such a framework may have in preventing or mitigating such scenarios should be considered.

4.5.2 Quantum computing

As noted above we have not tested the robustness of the work to developments in quantum computing. The possible impact of quantum computing should be assessed in terms of the new threats it might pose, the disruption to design bases assumptions, the new forms of evidence and argument that might be used in assurance and the new forms of attack from credible claims of an adversary in having such a capability.

4.6 A thematic view

4.6.1 Problems/challenges per INDEED theme

The table below groups the problems and research challenges based on an adapted set of main themes from the UK EPSRC-funded DIRC and INDEED projects, discussed previously. The themes are: *confidence/trust*, *diversity and heterogeneity*, *complexity and emergence*, *structure*, *resilience*, *adaptation and tempo*, *communication*, *markets* and *longer term adaptations*.

Table 2 – A thematic view of research problems and challenges

Themes	Problems / Challenges (section number where discussed in more detail in this document)	Research directions and potential solutions
Confidence / Trust	➢ Difficulties with defining nature of claims to be made. ➢ Difficulties with expressing uncertainty in the values of the dependability and security attributes. ➢ Difficulties with handling the dependence *between the sub-system* dependability and security values. ➢ Difficulties with handling the dependence *between various dependability and security attribute values*. ➢ Difficulties with decomposing claims of dependability, i.e. what is to be claimed, and the detail of how such claims should be expressed (4.2). ➢ Difficulties with expressing *confidence* in the claims that are made for dependability or security of a system or SoS (4.2). ➢ Difficulties with establishing system or sub-system provenance, especially when SoSs are build with COTS components (Appendix C). ➢ Difficulties with ensuring effective *communication* between the myriad of stakeholders and of risks and the trade offs between different aspects of resilience curve in evaluation and design (see section 4.2). ➢ Difficulties with ensuring timeliness properties in risk and responsibility communications in the recovery phase (for example, in the power grid scenario (depicted in Appendix A of this report), uncertainties about responsibilities within the	➢ See *resilience* theme below. ➢ Assurance case approach to assessing dependability and security. As discussed in section 4.2, claim-argument-evidence assurance approach should be adopted based on strategies that justify claims of behaviour, the absence or adequate mitigation of vulnerabilities and sufficient compliance with standards. This should build on current research emerging in safety cases, especially from the nuclear industry. ➢ Probabilistic approach to allow assessors to express their doubts in the values of the dependability and security attributes and hence provide a mechanism for quantifying the uncertainty in the values of the dependability and security attributes. Some solutions which can be used with smaller systems are presented in Appendix C. But further research is required for their applicability to larger systems. ➢ The approaches and models of the previous bullet points will clearly depend on the quality of the data and the metrics that are collected. Hence there is a need for good data to be obtained to allow for more objective evidence-driven assessment (see discussion in section 4.3.2). ➢ The elements of the framework defined in

	organisation caused confusion and delays in the recovery phase following the breach).	section 4.2 should provide for both the communication and evaluation of risk: ♦ Security cases by their nature provide a boundary object between stakeholders and promote communication ♦ These will be augmented with synthetic environments (via improvements in simulation, reasoning and information visualisation) to provide the tempo required.
Diversity and heterogeneity	➢ Difficulties with handling the dependence *between the sub-system* dependability and security values, unless strong independence assumptions for coincident failures of dependent components or sub-systems are made. ➢ Difficulty with dealing with heterogeneous SoSs made up of different constituent components of different levels of provenance (see section 2.2.4). ➢ Need to assess the balance between the key role that redundancy and diversity has in achieving resilience and the side effect of introducing unwanted vulnerabilities through interdependencies between systems and components	➢ Various models for modelling diversity have been proposed (with CSR, City University being a leading contributor). A review of these approaches can be found in [79]. ➢ The evaluation framework will need to deal with issues of heterogeneity as stated in section 4.3.
Complexity and Emergent properties	➢ Difficulties with assessing complex *inter-connected* and *inter-dependent* systems (3.2.3, 3.2.4). As mentioned under the Trust / Confidence heading, the assessment of the various security and reliability attributes cannot be done in isolation for	➢ Studying models and approaches developed for natural complex system in Biology and Physical sciences might help to gain a better understanding of complex ICT systems and the emergent behaviour that

	each constituent component of the systems or SoS. ➢ Difficulties with establishing system boundaries due to the complexity of the future systems and the large inter-dependence and interconnectedness of the components, i.e. where does one system end and another one begin (section 3.2.4). This makes security and dependability assessment of a SoS very difficult as assumptions may need to be made about the systems boundaries which may not hold in practice hence leading to wrong conclusions being drawn from the assessment.	is derived from this complexity. Some work has been done in this area and there is a need to challenge this to see whether one can move from a useful metaphor to applicable results. This will also help the assessor chose the right level of complexity of a given model to assess the dependability and security of these systems (also discussed in section 4.3.1.1). ➢ One part of this work might be to expand the work done on the structure of COTS outlined in [51] which we summarised in section 3.2.3.
Structure	➢ Similar to the issues discussed under complexity above, the structure and organisation of SoSs of the future are highly likely to become a major cause for concern for the developers and assessors alike (see also section 2.2). ➢ There are difficulties stemming from the lack of clear definitions of sub-system boundaries in a complex SoS (2.2.3). ➢ The need to address a variety of interfaces – technical, organisational, contractual, political, legal – and provide convincing stopping rules that somehow limits the analysis yet recognises the scale of the system (4.3.1.1). ➢ Difficulties stemming from the increased pervasiveness and ubiquity of future SoSs (see section 2.2.1). ➢ Difficulties with establishing what the roles of human operators and users of the complex ICT system systems	➢ As we discussed in section 3.2.2 the assessment of future SoS will need to be *interdisciplinary*, bringing in research from computer science, statistics, psychology, sociology and ethnography amongst others to assess the multi-faceted nature of these emerging systems. ➢ Develop stopping rules to prevent cases escalating to become the assurance of everything (4.3.1.1). ➢ Inter-organisational failures and modelling the roles of humans in the future systems will require *responsibility modelling* to be incorporated in the interdisciplinary approach (see sections 4.3.1.2 and 4.3.1.3). ➢ Extend complex systems approaches to investigate how to deal with uncertainty in structure (section 4.2 (*scalable cases* part) and section 4.3).

	are (4.3.1.2). ➢ New challenges for both dependability and security assessment may be posed from the growing use of *virtualisation* to organise the computing resources of ICT SoSs (see section 2.2.2) as well as *nanotechnology* and *biotechnology* (see section 2.2.4)	➢ Investigate scalable approaches to composing non-functional properties of systems and dealing with lack of independence. Related to calculus of confidence (see section 4.2 (*confidence* part) for a discussion).
Resilience	➢ Difficulty with definition of top level claims and sub-claims to be made for a system (section 4.2 *Claim definition and decomposition* part). ➢ Difficulty with considering "all hazards" in system assessment that addresses both malicious and accidental attacks on systems (see section 3.1.1).	➢ Security evaluation should be part of a holistic assessment of resilience addressing type I and type II resilience (see section 3.1.1).
Adaptation and tempo	➢ Difficulties with assessing systems in changing environments, i.e. as systems evolve due to patches, upgrades and new releases as well as change in the operational use of the system (see section 2.2.2). ➢ Difficulties with selecting the most appropriate model for assessing a given dependability or security attribute at runtime (see section 4.3.1.1)	➢ As stated in section 4.2: The framework should address the scale and tempo required of both events and operations and system evolution. There will be a need for: ♦ Approaches that can compose security of components or mitigate the impact of components or subsystems via wrappers or middleware. ♦ Approaches that can quickly deploy strategies e.g. worst case assumptions or some bounding assumptions ♦ Dynamic, real-time, assurance cases that provide decision-making support, or even autonomy, for parts of living or dynamic cases. This will require an understanding of human/computer trade offs in decision making and defined delegation

		policies. ➢ Related to the last bullet point above, online assessment is one way of getting accurate and up to date measures of dependability and security (see section 4.3.2 for more details). At CSR, City University we are involved in an EU funded project ReSIST [7], as part of which we will develop an online assessment engine for reliability of a small application. It remains research challenge to study how well online assessment scales in larger systems and SoSs.
Markets and longer term adaptations	➢ Difficulties with assessing security failures which may be caused by bad incentives (see section 3.2.1 on Economics of security). ➢ Difficulties with assessing issues which lie at the cross roads of psychology, economics and security (such as deception, inappropriate obedience, security usability etc., see section 3.2.1 for more details).	➢ The design and deployment of framework should be aware of how economics of security can be made to support the approach, and how markets may inhibit it. There is a need to understand how markets do/don't deliver resilience e.g. how markets might optimise ability to defend against mild attacks; how the common good might be served by maintaining more resources for recovery but is not optimal for a single organisation and how insurance might help or hinder resilience (see section 4.3.1.5).

5 Summary and further work

This report described the research we have conducted on identifying the technological advancements envisaged in the period 2007-2027, the research challenges that will be faced for the assessment of dependability and security of these complex systems and systems of systems (SoSs), and the assessment framework envisaged for these systems.

The research challenges, from the context we and others envisage, were summarised as:

- *Critical societal role* – The technological advancements predicted in the reviewed literature show that an unprecedented reliance on technology may be created.
- There is likely to be unprecedented increase in *scale and complexity* coming from the ubiquity and pervasiveness of systems that are driven by adaptation and evolution, ambition and requirements, the blurring of boundaries and the increased tempo of threats and operations.
- Concerns regarding *privacy* are likely to increase due to the widespread data sharing and communication that will be required to make the new systems function. Various privacy infringements are possible such as *identity theft*, *data laundering*, *disclosure of personal data*, *surveillance*, risks from *personalised profiling* etc.
- The new technological advancements, especially those on ICT systems will require substantial level of *trust* to be placed on the new systems. Hence there will be increased pressure on building *trustworthy* systems and SoSs.
- The future socio-technical ICT systems, are likely to be complex *inter-connected* and *inter-dependent* systems in a much greater scale than they are now, increasing the difficulty in performing holistic assessment of the various security and reliability attributes of complex SoSs and their constituent parts.
- There is likely to be an increased difficulty in building complex SoSs due to problems with *interoperability* of the constituent systems and components (e.g. inconsistencies in the interfaces of the different systems of an SoS; inconsistencies in the timescales (timebands) in which the different systems interact etc.).
- The future systems are expected to become increasingly *socio-technical* in nature, hence the role of the human users and operators within the system, and their fallibilities also need to be considered in the overall assessment of the system.
- The complexity of the future systems and the large inter-dependence and interconnectedness of the components will make it very difficult to define what the *boundaries* of a system or a SoS are, i.e. where does one system end and another one begin.
- The future systems will deal with radically increased *volume of information* due to advance in sensor and networks technologies. There will also be greater pressure on both systems and human decision makers to deal with information in shorter response times.
- The predicted pace of new technological innovations is likely to render existing technologies *obsolescent* more quickly than at present and also lead to heterogeneous systems composed on many generations of technology.
- There will be increasing tempo to operations and systems with dynamic and ad hoc coalitions being formed. Reconciling timing issues as systems and organisations are brought together to form larger SoSs also becomes an issue.

The challenges posed by current and future systems and threats lead us to propose an ambitious shift in perspective to an evaluation framework that attempts risk based, market sensitive, psychologically aware, evidence based approach to the assessment and communication of security (and dependability, resilience). The approach must be capable of addressing the scale and complexity of adaptive systems of systems (p-e-s-t systems) that have heterogeneous human and technical components and are deployed across a variety of organisational, political and legal boundaries.

The proposed framework would be composed of:

- Evaluation and communication of risk-based resilience.
- The definition of the evaluation target and the assumed threat model.
- A claim-argument-evidence assurance case approach.
- Methods for addressing scale and tempo required of both events and operations and system evolution.

We then elaborated some of the research directions along these themes into two inter-related broad areas:

- Resilience and security cases as an overarching approach within which we address: resilience models, claims decomposition, arguments, scalability and tempo. The argumentation will be about a service or system in an environment. To give meaning to the claims and to understand them they should be based on a coherent set of underlying models.
- Models for giving meaning to and supporting the evaluation. These will be disparate, multi-formalism models. We need to understand the interrelationships, the required abstractions and levels of fidelity. They will include models of the system environment and particular economic and threat models.

We also point out that an evaluation framework is not a neutral, technical object. It will have an impact on society and as with any risk based approach there will be those that benefit and those that suffer the costs of the framework. There may be unintended side effects and the market may operate to deliver certain levels of resilience but not necessarily the required level of resilience required for society critical systems (and occasionally the market may even hinder / prevent the delivery of security/resilience). The need for supporting policy and regulation should be born in mind and the research on the evaluation framework should feed into and enable research underpinning any emerging political initiatives. For example, on the tradeoffs between and conflicts of interest that may arise between the government, society and individuals, such as issues regarding intellectual property rights and privacy, human rights issues, individual liberty etc.

There are number areas to be considered for extending the analysis in this report:

- Consideration of extreme scenarios where we see anti-technology refusenik cultures, disenchantment with a technology (e.g. as a result of successful attack of key systems, their oppressive use by the state, breakdown of complex systems such as financial markets), or scenarios of extreme state control.
- The impacts of quantum computing should be assessed in terms of the new threats it might pose, the disruption to design bases assumptions, the new forms of evidence and argument that might be used in assurance and the new forms of attack from credible claims of an adversary in having such a capability.
- More detailed analysis on the assessment issues that arise from potentially conflicting expectations of the different stakeholders (e.g. curtailment of some individual privacy for the wider good of the society).

- Periodic revisions of this document to update the research questions that may have been solved and addition of new ones, especially in the light of new reports and roadmaps (e.g. the upcoming NATO NEC Security Research Strategy report [80]).

Overall further work is envisaged to address and prioritise the research challenges summarised in this report and develop the evaluation framework presented in Part 3 of this report. This remains to be discussed further with DSTL and other interested stakeholders.

6 Acknowledgements

This work has been sponsored by the UK Defence Science and Technology Laboratory (DSTL) via the "EMR (Extra Mural Research) proposal SoSoS Development Methodologies for Secure System Evolution". The work has also been partially supported by Engineering and Physical Sciences Research Council (EPSRC) under the INterdisciplinary DEsign and Evaluation of Dependability (INDEED) project. We would like to thank Victoria Page, Helen Philips, John Yesberg and Lisa Hamilton from DSTL for useful discussions and suggestions. We would also like to thank Dan Craigen and Ann Miller for allowing us to use the scenarios developed as part of the NATO Research Task Group RTG-020 on the Dual Use of High Assurance Technologies (IST-048/RTG-020). We would also like to Thanks Nick Chozos of Adelard for reviewing an earlier version of this report. And finally we would like to thank our colleagues at CSR for discussions and useful suggestions on improving the report and especially Professor Bev Littlewood who reviewed an earlier version of this report.

7 References

1. House-of-Commons-Committee-of-Public-Accounts, *"Department of Health: The National Programme for IT in the NHS",* 2007; Available from: http://www.parliament.the-stationery-office.co.uk/pa/cm200607/cmselect/cmpubacc/390/390.pdf.
2. AMSD, *"A Dependability Roadmap for the Information Society in Europe: Part 1 - An Insight into the Future",* 2003
3. Curry, A., T. Hodgson, et al., *"Intelligent Infrastructure Futures The Scenarios – Towards 2055",* UK Office of Science and Technology, 2005.
4. Antonio, D., H. Seppo, et al., *"D1.2.1 Scenario analysis",* IRRIIS(Integrated Risk Reduction of Informationbased Infrastructure Systems), 2006.
5. Alahuhta, P., P.D. Hert, et al., *"D2: Dark scenarios in ambient intelligence: Highlighting risks and vulnerabilities",* SWAMI (Safeguards in a World of Ambient Intelligence), 2006.
6. DCDC, *"Global Strategic Trends Programme 2007-2036",* 2007.
7. ReSIST, *"D13: From Resilience-Building to Resilience-Scaling Technologies: Directions ",* 2007.
8. US-Department-Of-Defense, *"Joint Vision 2020",* 2005.
9. ReSIST, *"D12: Resilience-Building Technologies: State of Knowledge ",* 2006.
10. AMBER, *"AMBER - Assessing, Measuring, and Benchmarking Resilience",* 2008; Available from: http://amber.dei.uc.pt/.
11. MoD (2008), *"Joint Service Publication 777 - Network Enabled Capability",*
12. DoD, *"Global Information Grid Architectural Vision",* 2007; Available from: http://www.defenselink.mil/cio-nii/docs/GIGArchVision.pdf.
13. Littlewood, B., *"The Use of Computers in Safety-Critical Applications, Final Report of the Study Group on the Safety of Operational Computer Systems constituted by the Advisory Committee on the Safety of Nuclear Installations, HSE Books London",* 1998; Available from: http://www.hse.gov.uk/nuclear/computers.pdf.
14. Gligor, V.D., T. Haigh, et al., *"Information Assurance Technology Forecast 2005",* IEEE Security and Privacy, 2006. **4**(1): p. 62-69.
15. European-Commission, *"ICT - INFORMATION AND COMMUNICATION TECHNOLOGIES: Work Programme 2007-08"*. 2007.
16. W3C, *"W3C Semantic Web Activity",* 2008; Available from: http://www.w3.org/2001/sw/.
17. Bellovin, S.M., T.V. Benzel, et al., *"Information Assurance Technology Forecast 2008",* IEEE Security and Privacy, 2008. **6**(1): p. 16-23.
18. Aaviksoo, J., *"Statesmen's Forum: Jaak Aaviksoo, Minister of Defense, Republic of Estonia ",* 2007; Available from: http://www.csis.org/component/option,com_csis_events/task,view/id,1440/.
19. Wikipedia, *"2007 cyberattacks on Estonia",* 2008; Available from: http://en.wikipedia.org/wiki/Cyberattacks_on_Estonia_2007.
20. Bloomfield, R.E., S. Guerra, et al., *"International Working Group on Assurance Cases (for Security)",* IEEE Security & Privacy, 2006. **4**(3): p. 66-68.
21. Littlewood, B. and L. Strigini, *"Redundancy and diversity in security",* in *ESORICS (European Symposium on Research in Computer Security)*. 2004. Sophia Antipolis, France: Springer.
22. Helton, J.C. and W.L. Oberkampf, *"Special Issue: Alternative representations of epistemic uncertainty",* Reliability Engineering and System Safety, 2004. **85**(1-3).
23. Littlewood, B., S. Brocklehurst, et al., *"Towards operational measures of computer security",* Journal of Computer Security, 1993. **2**(3): p. 211-229.
24. UK-Resilience, *"Communicating Risk",* 2008; Available from: http://www.ukresilience.info/publications.aspx.

25. *"Risk Management Assessment Framework",* 2008; Available from: http://www.hm-treasury.gov.uk/media/6/6/17A8166B-BCDC-D4B3-16668DC702198931.pdf.
26. Cabinet-Office, *"Information Assurance Governance Framework",* 2008; Available from: http://www.cabinetoffice.gov.uk/csia/ia_governance/content/~/media/assets/www.cabinetoffice.gov.uk/csia/ia_governance_app1%20doc.ashx.
27. Bryans, J., B. Littlewood, et al., *"E-voting: Dependability Requirements and Design for Dependability",* in *Workshop on Dependability and Security in e-Government (DeSeGov 2006), at at First International Conference on Availability, Reliability and Security (ARES 2006)*. 2006. Vienna, Austria: IEEE Computer Society Press, p. 988-995.
28. Randell, B. and P. Ryan, *"Voting Technologies and Trust",* IEEE Security & Privacy, 2006. **4**(5): p. 50-56.
29. EPCIP, *"The European Programme for Critical Infrastructure Protection (EPCIP)",* 2006; Available from: http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/06/477&format=DOC&aged=1&language=EN&guiLanguage=en.
30. Hollnagel, E., D.D. Woods, and N. Leveson, eds, *"Resilience engineering : concepts and precepts"*. 2006, Ashgate Pub Co.
31. DHS-Build-Security-In, *"Assurance Cases",* 2008; Available from: https://buildsecurityin.us-cert.gov/daisy/bsi/articles/knowledge/assurance.html.
32. Jones, C., R. Bloomfield, et al., *"Methods for assessing the safety integrity of safety-related software of unceratin pedigree (SOUP)",* 2001; Available from: http://www.adelard.com/web/hnav/resources/reports/hse_soup.html.
33. Anderson, R., R. Böhme, et al., *"Security Economics and the Internal Market"*. 2008.
34. Anderson, R. and T. Moore, *"Information Security Economics – and Beyond",* Information Security Summit 2008, 2008.
35. Anderson, R., *"Security Engineering - : A Guide to Building Dependable Distributed Systems",* 2nd ed. 2008, Indianapolis: John Wiley & Sons.
36. Cranor, L., *"Security Usability"*. 2005: O'Reilly
37. *"The Workshop on the Economics of Securing the Information Infrastructure"*; Available from: http://wesii.econinfosec.org/workshop/.
38. Nagaraja, S. and R. Anderson, *"The Topology of Covert Conflict",* in *Workshop on the Economics of Information Security (WEIS)*. 2006. Cambridge, UK.
39. Dynes, S., E. Andrijicic, and M.E. Johnson, *"Costs to the U.S. Economy of Information Infrastructure Failures: Estimates from Field Studies and Economic Data",* in *Workshop on the Economics of Information Security (WEIS)*. 2006. Cambridge, UK.
40. Rowe, B.R. and M.P. Gallaher, *"Private Sector Cyber Security Investment: An Empirical Analysis ",* in *Workshop on the Economics of Information Security (WEIS)*. 2006. Cambridge, UK.
41. Liu, W., H. Tanaka, and K. Matsuura, *"An Empirical Analysis of Security Investment in Countermeasures Based on an Enterprise Survey in Japan",* in *Workshop on the Economics of Information Security (WEIS)*. 2006. Cambridge, UK.
42. Herath, H. and T. Herath, *"Justifying Spam and E-mail Virus Security Investments: A Case Study ",* in *Workshop on the Economics of Information Security (WEIS)*. 2006. Cambridge, UK.
43. Sutton, M. and F. Nagle, *"Emerging Economic Models for Vulnerability Research ",* in *Workshop on the Economics of Information Security (WEIS)*. 2006. Cambridge, UK.
44. Radianti, J. and J.J. Gonzalez, *"Toward A Dynamic Modeling Of The Vulnerability Black Market",* in *The Workshop on the Economics of Securing the Information Infrastructure*. 2006.

45. Moore, T., *"The Economics of Digital Forensics"*, in *Workshop on the Economics of Information Security (WEIS)*. 2006. Cambridge, UK.
46. Hasan, R. and W. Yurcik, *"Beyond Media Hype: Empirical Analysis of Disclosed Privacy Breaches 2005-2006 and a DataSet/Database Foundation for Future Work"*, in *The Workshop on the Economics of Securing the Information Infrastructure*. 2006.
47. Santos, J.R., Y.Y. Haimes, and C. Lian, *"A Framework for linking oil and gas cybersecurity metrics to the inoperability input-output model"*, in *Western Economics Association (WEA) Annual Conference*. 2006. San Diego, CA, USA.
48. Gallagher, M.A., A.W. Snodgrass, and G.J. Ehlers, *"Input-Output Modeling for Assessing Cascading Effects"*, Military Operations Research Society 2005. **10**(2).
49. Bullock, S. and D. Cliff, *"Complexity and Emergent Behaviour in ICT systems"*, 2004; Available from: http://eprints.ecs.soton.ac.uk/11478/1/HPL-2004-187.pdf.
50. Clark, A., *"Mindware: an introduction to the philosophy of cognitive science"*. 2001: Oxford University Press.
51. Bloomfield, R., *"A descriptive model of failures in complex systems"*, in *Supplemental Volume of Dependable Systems and Networks (DSN'03)*. 2003. San Francisco, CA, USA: IEEE Computer Society, p. B76-77.
52. IRRIIS, *"Integrated Risk Reduction of Information-based Infrastructure Systems"*, 2008; Available from: http://www.irriis.org/.
53. Cabinet-Office, *"A National Information Assurance Strategy"*, 2008; Available from: http://www.cabinetoffice.gov.uk/csia/~/media/assets/www.cabinetoffice.gov.uk/csia/nia_strategy%20pdf.ashx.
54. EU, *"Seventh Research Framework Programme"*, 2008; Available from: http://cordis.europa.eu/fp7/home_en.html.
55. CETIFS, *"Interdependency Feasibility Study"*, 2008; Available from: http://www.csr.city.ac.uk/projects/cetifs.html.
56. Rushby, J., *"Just-in-Time Certification, Auckland, New Zealand."* in *IEEE International Conference on the Engineering of Complex Computer Systems (ICECCS)*. 2007. Auckland, New Zealand: IEEE, p. 15-24.
57. Rushby, J., *"What Use Is Verified Software - Invited paper presented in a special session on the Verified Software Initiative"*, in *IEEE International Conference on the Engineering of Complex Computer Systems (ICECCS)*. 2007. Auckland, New Zealand: IEEE.
58. Rushby, J., *"Automated Formal Methods Enter the Mainstream"*, Communications of the Computer Society of India: Formal Methods Special Theme Issue, 2007. **31**(2): p. 28-32.
59. *"Static Source Code Analysis Tools for C"*, 2008; Available from: http://www.spinroot.com/static/.
60. ScanCoverity, *"265 PROJECTS ON THE LADDER"*, 2008; Available from: http://scan.coverity.com/rungAll.html.
61. Woodcock, J. and L. Freitas, *"Z/Eves and the Mondex electronic purse"*, in *invited talk to ITCAC* 2006.
62. Moorsel, A.v., A. Bondavalli, et al., *"D2.1 - State of the Art"*, 2008; Available from: http://www.amber-project.eu/documents/md_$HTTP_POST_VARS[editbin]_amber_d2.1_stateoftheart_v1.0.pdf.
63. TPC, *"TPC Benchmark C, Standard Specification, Version 5.0."* 2002; Available from: http://www.tpc.org/tpcc/.
64. Kanoun, K. and Y. Crouzet, *"Dependability Benchmarks for Operating Systems"*, International Journal of Performability Engineering, 2006. **2**(3): p. 277 - 289.
65. Musa, J.D., *"Operational Profiles in Software-Reliability Engineering"*, IEEE Software, 1993. **(March)**: p. 14-32.

66. Balbo, G., *"Introduction to stochastic petri nets",* in *Lectures on Formal Methods and Performance Analysis, Lecture Notes in Computer Science*. 2001, Springer Verlag. p. 84-155.
67. Lyu, M.R., ed. *"Handbook of Software Reliability Engineering"*. 1996, McGraw-Hill and IEEE Computer Society Press.
68. Brocklehurst, S., P.Y. Chan, et al., *"Recalibrating software reliability models",* IEEE Transactions on Software Engineering, 1990. **16**: p. 458-470.
69. Alhazmi, O.H., Y.K. Malaiy, and I. Ray, *"Security vulnerabilities in software systems: A quantitative perspective",* in *IFIP WG 11.3 Working Conference on Data and Applications Security*. 2005, p. 281-294.
70. Alhazmi, O.H. and Y.K. Malaiya, *"Modeling the vulnerability discovery process",* in *16th IEEE International Symposium on Software Reliability Engineering (ISSRE'05)*. 2005. Washington, DC, USA: IEEE Computer Society Press, p. 129-138.
71. Alhazmi, O.H. and Y.K. Malaiya, *"Quantitative vulnerability assessment of systems software",* in *IEEE Reliability and Maintainability Symposium (RAMS'05)*. 2005. USA: IEEE Computer Society, p. 615.
72. Woo, S.-W., O.H. Alhazmi, and Y.K. Malaiya, *"Assessing vulnerabilities in Apache and IIS HTTP servers",* in *IEEE International Symposium on Dependable, Autonomic and Secure Computing*. 2006: IEEE Computer Society, p. 103-110.
73. Ozment, A., *"Vulnerability Discovery & Software Security",* 2007; Available from: PhD Thesis, University of Cambridge.
74. Bloomfield, R. and B. Littlewodd, *"Multi-legged arguments: the impact of diversity upon confidence in dependability arguments",* in *Dependable Systems and Networks (DSN-03)*. 2003. San Francisco, USA: IEEE Computer Society, p. 25-34.
75. Alberdi, E., A.A. Povyakalo, et al., *"Effects of incorrect CAD output on human decision making in mammography",* Academic Radiology, 2004. **11**(8): p. 909-918.
76. DIRC, *"Dependability Interdiciplinary Research Collaboration project",* 2004; Available from: http://www.dirc.org.uk.
77. Herrmann, D.S., *"Complete Guide to Security and Privacy Metrics"*. 2007: Auerbach.
78. Gashi, I., P. Popov, and L. Strigini, *"Fault tolerance via diversity for off-the-shelf products: a study with SQL database servers",* IEEE Transactions on Dependable and Secure Computing, 2007. **4**(4): p. 280-294.
79. Littlewood, B., P. Popov, and L. Strigini, *"Modelling software design diversity - a review",* ACM Computing Surveys, 2001. **33**(2): p. 177-208.
80. *"NEC Security Research Strategy",* Pre-released RTO Technical report - NATO, 2008.
81. Francis, R., *"B.C. plans province-wide electronic medical record system",* 2007; Available from: http://www.itbusiness.ca/it/client/en/home/News.asp?id=42429&cid=3.
82. Gonsalves, A., *"McKesson Offers Health Care Apps On Red Hat Linux ",* 2007; Available from: http://www.informationweek.com/news/management/showArticle.jhtml?articleID=197008898.
83. Tippu, S., *"Indian offshorers move offshore ",* 2006; Available from: http://www.itwire.com/content/view/5870/945/.
84. Trent, W., *"Offshorers to Offshore?"* 2006.
85. Associated-Press, *"LexisNexis theft much worse than thought",* 2005.
86. Givens, B., *"The ChoicePoint Data Security Breach (Feb. '05): What It Means for You",* 2005; Available from: http://www.privacyrights.org/ar/CPResponse.htm.
87. Privacy-Rights-ClearingHouse, *"A Chronology of Data Breaches ",* 2008; Available from: http://www.privacyrights.org/ar/ChronDataBreaches.htm.
88. Bächer, P., T. Holz, et al., *"Know your Enemy: Tracking Botnets",* 2005; Available from: http://www.honeynet.org/papers/bots/.

89. Lemos, R., *""Data storm" blamed for nuclear-plant shutdown",* 2007; Available from: http://www.securityfocus.com/news/11465.

90. Ballard, M., *"UK Treasury knew of US hunt through British bank data",* 2007; Available from: http://www.theregister.co.uk/2007/02/16/swift_hm_treasury/.

91. Ballard, M., *"Europe demands say on US data trawling",* 2007; Available from: http://www.theregister.co.uk/2007/02/15/eu_grabon_us/.

92. Gaudin, S., *"Homeland Security Creates National Computer Forensics Institute ",* 2007; Available from: http://www.informationweek.com/news/management/showArticle.jhtml?articleID=198000260.

93. Reed, T., *"At the Abyss: An Insider's History of the Cold War "*. 2004: Presidio Press.

94. Prevelakis, V. and D. Spinellis, *"The Athens Affair",* 2007; Available from: http://www.spectrum.ieee.org/jul07/5280.

95. Ramalingam, A., A. Miller, and K.T. Erickson, *"SCADA System Vulnerability Analysis",* in *Working together: R&D Partnerships in Homeland Security Conference*. 2005. Boston, MA.

96. Strom, D., *"Six steps to a wireless site survey",* 2006; Available from: http://www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9004641&source=NLT_PM&nlid=8.

97. Poulsen, K., *"Slammer worm crashed Ohio nuke plant network",* 2003; Available from: http://www.securityfocus.com/news/6767.

98. Gellman, B., *"U.S. Fears Al Qaeda Cyber Attacks ",* 2002; Available from: http://www.securityfocus.com/news/502.

99. McMillan, R., *"Hackers break into water system network",* 2006; Available from: http://www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9004659.

100. Goodin, D., *"Electrical supe charged with damaging California canal system",* 2007; Available from: http://www.theregister.co.uk/2007/11/30/canal_system_hack/.

101. *"Ems powerline crossing",* 2008; Available from: http://en.wikipedia.org/wiki/380kV-Ems-Overhead_Powerline_Crossing.

102. *"Power outage",* 2008; Available from: http://en.wikipedia.org/wiki/Power_outage.

103. Ivey, M., A. Akhil, et al., *"Grid of the Future White Paper on Accommodating Uncertainty in Planning and Operations",* 1999; Available from: http://eetd.lbl.gov/certs/pdf/certs-uncertainty.pdf.

104. OMG, *"Unified Modeling Language (UML), version 2.1.1",* 2007; Available from: http://www.omg.org/technology/documents/formal/uml.htm.

105. Yang, Y., Jesal Bhuta, et al., *"Value-based processes for COTS-based applications",* IEEE Software, 2005. **22**(4): p. 54-62.

106. Dean, J., *"An Evaluation Method for COTS Software Products"*. 2000.

107. Lewis, P., P. Hyle, et al., *"Lessons Learned in Developing Commercial Off-The-Shelf (COTS) Intensive Software Systems",* 2000; Available from: http://www.cebase.org/www/researchActivities/COTS/LessonsLearned.pdf.

108. Kontio, J., S.Y. Chen, et al., *"A COTS Selection Method and Experiences of Its Use",* in *Twentieth Annual Software Engineering Workshop,NASA Goddard Space Flight Center*. 1995. Greenbelt, Maryland.

109. Phillips, B.C. and S.M. Polen, *"Add Decision Analysis to Your COTS Selection Process",* in *CroosTalk -The Journal of Defence Software Engineering*. 2002.

110. Boehm, B., D. Port, et al., *"Composable Process Elements for Developing COTS-Based Applications",* in *Symposium on Empirical Software Engineering. (ISESE'03)*. 2003: ACM-IEEE, p. 8-17.

111. Ochs, M., D. Pfahl, et al., *"A Method for Efficient Measurement-based COTS Assessment and Selection -Method Description and Evaluation Results",* in *7th Symposium on Software Metrics*. 2001. London, England: IEEE Computer Society, p. 285-294.

112. Comella-Dorda, S., J. Dean, et al., *"A Process for COTS Software Product Evaluation",* in *International Conference on COTS-Based Software Systems (ICCBSS'02)*. 2002. Florida, USA: Springer-Verlag, p. 86-92.

113. Ncube, C. and N. Maiden, *"PORE:Procurement Oriented Requirements Engineering Method for the Component-Based Systems Engineering Development Paradigm",* in *International Workshop on Component-Based Software Engineering*. 1999.

114. Alves, C. and J. Castro, *"CRE: A Systematic Method for COTS Components Selection",* in *XV Brazilian Symposium on Software Engineering (SBES)*. 2001. Rio de Janeiro, Brazil.

115. Gregor, S., J. Hutson, and C. Oresky, *"Storyboard Process to Assist in Requirements Verification and Adaptation to Capabilities Inherent in COTS",* in *International Conference on COTS-Based Software Systems (ICCBSS'02)*. 2002. Florida, USA: Springer-Verlag, p. 132-141.

116. Burgués, X., C. Estay, et al., *"Combined Selection of COTS Components",* in *International Conference on COTS-Based Software Systems (ICCBSS'02)*. 2002. Florida, USA: Springer-Verlag, p. 54-64.

117. Ruhe, G., *"Intelligent Support for Selection of COTS Products",* in *Web, Web-Services, and Database Systems*. 2003: Springer, p. 34-45.

118. Tran, V. and D.-B. Liu, *"A Risk Mitigating Model for the Development of Reliable and Maintainable Large-Scale Commercial-Off-The-Shelf Integrated Software Systems",* in *Reliability and Maintainability Symposium (RAMS'97)*. 1997: IEEE Print, p. 361-367.

119. Jeanrenaud, J. and P. Romanazzi, *"Software Product Evaluation: A Methodological Approach",* in *Software Quality Management II: Building Software into Quality*. 1994, p. 55-69.

120. Kunda, D. and L. Brooks, *"Applying Social-Technical Approach for COTS Selection",* in *UK Academy for Information Systems (UKAIS'99)*. 1999. University of York, England.

121. Likert, R., *"A Technique for the Measurement of Attitudes"*. 1932, New York: McGraw-Hill.

122. Gashi, I., P. Popov, and V. Stankovic, *"Uncertainty Conscious Assessment of Off-The-Shelf Software: a Baysian Approach",* Elsevier Information ad Software Technology Journal, 2008: p. Accepted for publication.

123. Littlewood, B., P. Popov, and L. Strigini, *"Assessment of the Reliability of Fault-Tolerant Software: a Bayesian Approach",* in *SAFECOMP-2000*. 2000. Rotterdam, the Netherlands: Springer, p. 294-308.

124. FDA, *"Guidance for Industry, FDA Reviewers and Compliance on Off-The-Shelf Software Use in Medical Devices",* 1999; Available from: http://www.fda.gov/cdrh/ode/guidance/585.pdf.

125. IEC, *"IEC 60880 - Nuclear power plants - Instrumentation and control systems important to safety - Software aspects for computer-based systems performing category A functions ",* 2006; Available from: http://webstore.iec.ch/webstore/webstore.nsf/artnum/036058.

126. DS, *"Requirments for safety related software in defence equipment",* 1997; Available from: http://www.dstan.mod.uk/data/00/055/01000200.pdf.

127. Bloomfield, R., J. Cazin, et al., *"Validation, verification and certification of embedded systems"*. 2004, National Aerospace Laboratory NLR.

128. Bloomfield, R., S. Guerra, and D. Sheridan, *"Evaluation of Integrity and Availability in Safety-Related Applications"*. 2006, Adelard LLP.

129. Avizienis, A., J.-C. Laprie, et al., *"Basic Concepts and Taxonomy of Dependable and Secure Computing",* IEEE Transactions on Dependable And Secure Computing, 2004. **1**(1): p. 11-33.

130. Dewsbury, G. and J. Dobson, eds, *"Responsibility and Dependable Systems (Hardcover)"*. 2007, Springer-Verlag

131. Bhuta, J. and B. Boehm, *"Attribute-Based COTS Product Interoperability Assessment",* in *International Conference in COTS-Based Software Systems (ICCBSS'07)*. 2007. Banff, Alberta, Canada: IEEE Computer Society, p. 163 - 171.

Appendix A

This appendix contains details of three scenarios which have been developed by a NATO Research Task Group on the Dual Use of High Assurance Technologies (IST-048/RTG-020). The task group was headed by Dan Craigen (University of Toronto, Canada), Ann Miller (University of Missouri-Rolla) and Robin Bloomfield (CSR, City University). A final report detailing the scenarios and the analysis performed is in preparation.

The "Terms of Reference" (ToR) for the Research Task Group (RTG), provides the historical background for the group as follows:

"As a result of the terrorist attacks on the United States of America on September 11, 2001, NATO and its member countries have been actively investigating means for combating terrorism.

High Assurance Technologies (such as formal methods) have normally been used to develop systems requiring high degrees of assurance as to functionality, safety and security. One of the key benefits of such technologies is their ability to ferret out subtle problems with system requirements, design and implementation.

However, experience has also been obtained in which such technologies can be used to identify effectively weaknesses in NATO network-enabled systems, thereby allowing for enhanced defence. From a formal methods perspective, the phrase "formal methods-based tiger teaming" is meant to capture the idea of using formal methods to help drive the identification of weaknesses in such systems. The interaction of mathematical modelling with experimental testing of such systems has been shown to be very effective, but not widely known, in identifying weaknesses and thereby being a first step in defending such systems."

The ToR continues by justifying the RTG to NATO as follows:

"An important aspect of combating terrorism is the protection of NATO network-enabled systems from exploitation by terrorists and other foes."

The three scenarios, discussed in turn, are entitled:

- Scenario 1: Medical and Financial Services Sector – (present time)
- Scenario 2: Oil and Gas Sector – (present time)
- Scenario 3: Electrical Power Sector – (Circa 2012)

Scenario 1: Medical and Financial Services Sector – (present time)

Author: Dan Craigen (University of Toronto)

1. Isabelle had felt unwell for some days and that fateful morning she fainted. An ambulance was dispatched and she was taken to the nearby HMO Acme Hospital. There she was given an RFID identity bracelet. Her medical records were "securely" downloaded from the medical IT and records data collecting/storing company MedRecsOnly.[10] The doctors quickly identified the problem, admitted her, and put her on a course of medication that was expected to take 3-4 days to have full effect. MedRecsOnly had obtained licenses in a number of countries to provide such records to properly authenticated medical institutions; they were also able to provide data analysis to pharmaceutical and marketing companies – at least in the aggregate.
2. HMO Acme Hospital was one of a handful of hospitals world-wide on the vanguard of adopting an enterprise-wide IT services infrastructure that covered everything from classic inventory control through to the new PDAs all medical staff carried to access a patient's medical records.[11] The system had been developed by Medirixx, one of a number of new (and established) companies that entered the burgeoning health field in which the value propositions for enhanced sharing of information and enhanced automation were compelling. For example, it was now much easier to share the results of tests, MRIs and X-Rays between the various medical institutions (including palliative care facilities, emergency hospitals, walk-in clinics and regular doctors' offices); this greatly reduced the need for repetitive testing and examination. It also helped substantially when patients were unconscious or otherwise unable to speak for themselves. Medirixx interoperated with the MedRecsOnly databases.
3. The PDAs were able to link to the RFID identity bracelets and in conjunction with a secure wireless connection to the hospital routers were able to alert staff as to when patients' medication was due (amongst other services). Significant design effort had been directed at the security architecture with specific considerations for confidentiality, integrity and authentication. Medication was now routinely mixed in an automated pharmacy that accepted downloaded instructions from the Medirixx hospital enterprise system and "securely" entered into the dispensation units, which were labelled with a unique patient identifier (equal to that on the RFID identify bracelet – thereby allowing for confirmation).
4. On the evening of Isabelle's admission the new shift arrived. Hillary had been in the nursing profession for years and was well regarded by her colleagues. On activating her PDA and putting on her RFID identity bracelet (so her whereabouts in the hospital was known) she received a notice that Isabelle was due her next intake of medicine. Hillary, as was her common practice, visually inspected the medications and the accompanying documentation and found that everything seemed to be in order – there were no observable abnormalities. So, after talking briefly with Isabelle, gave her the prescribed medication.

[10] See http://www.fraserinstitute.ca/admin/books/chapterfiles/Feb07ffAtherley.pdf for a discussion. One particular observation made in the Fraser Institute paper is as follows: "*The vulnerabilities of information technology beset the financial services industry. The industry can insure consumers and reimburse them for losses that it is unable to protect them against. But in health care, the same vulnerabilities can result in irretrievable losses of the health and even the life of patients.*"

[11] See, for example, [81] and also "[82].

5. Two hours later Isabelle went into Anaphylactic shock and died. Just as troubling was that another ten patients all died that evening – some due to Anaphylaxis; some due to cardiac arrest; others, the cause was not immediately clear. Given that HMO Acme was one of the smaller hospitals, this was clearly anomalous, especially given that all the patients were not in the Extensive/Emergency Care wing.
6. The next day turned out to be just as bad, with another fifteen patients succumbing. The hospital was locked down and autopsies fast tracked. Early results showed that some patients had been given medication to which they were lethally allergic. Others had, in effect, been poisoned, by unexpected combinations of medication. Yet others, had died since the applied medications were given at inappropriate doses.
7. Disturbingly, though not fully recognized for another day, numerous other hospitals were faced with unexpected mortalities. Over all, thousands died – a number comparable to those who died during the 9/11 attacks.
8. Investigators hypothesized that the patients were not receiving their proper medication, yet an initial review of the security logs indicated that the patients were being treated properly. Not knowing of the events in the other hospitals, HMO Acme investigators immediately shifted their attention to the possibility that one of the medical caregivers was responsible; this avenue, however, was rapidly rejected. Gradually, the global reach of the problem became known.
9. Fortuitously, it was noted that a non-Medirixx audit log of the chemicals being used by the automated pharmacy was not consistent with the Medirixx database. Forensics experts finally proved that the instructions being sent to the automated pharmacy by the Medirixx system was inconsistent with the Medirixx logs. Somehow, the integrity of the system had been compromised.
10. Identification of the compromise meant that the HMO Acme had to stop using Medirixx and revert to their emergency paper backup procedures; thereby severely impacting efficiencies. All elective surgeries were cancelled and even some critical patients were moved to other nearby hospitals. HMO Acme had to bring in extra staff and, at least initially, shut down the automated pharmacy. Though the hospital did its best, patients were ill-served.
11. It became apparent that the heightened mortalities in the other Medirixx supported hospitals were also due to an integrity failure with the Medirixx system.
12. Medirixx commenced a full review of their systems, but corporate disaster loomed. They were one of the leading IT firms on NASDAQ and they saw a substantial decline in what had been a NASDAQ darling for some years. Other firms in the IT sector also saw their stock prices depressed. NASDAQ had its worst few days since the Telecom/IT collapse of the early 2000s.
13. Medirixx was a well regarded firm, which took seriously the concerns arising from their life-critical software. Substantial design effort went into developing a highly trusted system. Medirixx, however, like other firms, felt the pressures of the competitive market place and decided to offshore the coding. SafelyC was hired after a due diligence review by Medirixx and some early trials.
14. SafelyC, however, also felt commercial pressures, and unbeknownst to Medirixx, subcontracted out portions of the work (to OnwardsC) and, to make sure that full interoperability was maintained, included overall design information of the Medirixx system.[12]

[12] See [83] and [84] as indications of the offshorers offshoring trend.

15. OnwardsC, however, was a front company for a terrorist/crime gang (TC), which evaporated on completion of the attacks. TC was known to western intelligence and policing agencies, but many aspects of their organization remained a mystery; though it seemed they ran profitable criminal enterprises and had been making some headway with their terrorist objectives. Increasing resources were being directed by the intelligence and policing agencies to roll back the advances and to depress their criminal enterprises. TC was beginning to feel the pressure.

16. A core group of OnwardsC employees also belonged to TC and were extremely well trained; mostly in western universities. These employees were experts at information technology, including data mining, secure technologies, and other cutting edge technologies. OnwardsC also had access to medical expertise. While being extremely careful to meet their contractual obligations to SafelyC, including satisfying an exhaustive set of test cases, they also inserted Trojan horse software that included analytical techniques and means for working around the Medirixx designed system. (The simple analytical techniques were inserted so that the medical records could be analyzed for allergies or for patients taking medication where slight changes of dosage could result in mortality.)

17. Like much software, the Medirixx software was patchable over the Internet. OnwardsC designed their Crimeware so that it could be activated by a carefully selected communication. Since all the firewalls would allow for Medirixx updates to pass through, OnwardsC had no problem having their carefully crafted message sent out by a compromised Botnet to the thirty-one hospitals.

18. Other members of the TC gang shorted Medirixx on the NASDAQ, along with some other related medical IT stocks and waited. Medirixx trading was stopped for a few days, but once trading commenced, it continued its free fall. TC apparently made millions.

19. But, this was only part of the story. TC had greater plans than an apparent financial motivation. Through their expertise and contacts with the black hat ecosystem, TC had successfully penetrated Lemangtrix – a huge player in the burgeoning space of world-wide personnel data acquisition and analysis, with significant contacts into government and intelligence agencies.[13] Lemangtrix had information on over a billion individuals, with terabytes of data. Lemangtrix made use of state-of-the art hardware, software and networking technology. They were fully aware that penetration of their database, similar say to the problems that TJX had in 2007 and previous years,[14] could cripple their business.

20. TC and their black hat associates had carefully acquired tens of millions individual records from Lemangtrix and had made use of their own Botnets (and peer-to-peer technologies) to distribute the data "securely" on compromised PCs globally. (The data was acquired through various means including social engineering, improper backup procedures, and penetration of the Lemangtrix databases.)

21. Three days after they activated their Meditrixx Crimeware, TC activated their Botnets.[15] Emails containing data from the compromised Lemangtrix data base were sent out by the PCs – essentially as SPAM. The refinement was that where ever they had the email address for a specific individual, the private information was sent to that individual and to thousands of others. In addition, the email noted that the information had been sent by

[13] See [85] and [86]. Visit [87] for a chronology of data breaches since 2005. Over 100 million records containing sensitive personal information have been compromised.

[14] "The TJX Breach: It's even worse than they thought," Briony Smith, itbusiness.ca, 21 February 2007.

[15] "[88] provides an excellent synopsis of Botnets and how a "honeynet" was used to acquire information.

MegaBank for confirmation and that MegaBank on receiving confirmation would open accounts for the individual. Furthermore, a “distributed denial of service” attack was launched on both MegaBank’s web presence and on their VOIP network; essentially isolating the bank.[16]

22. Again, TC had shorted stock – this time in MegaBank and other finance related stocks. As news of the apparent MegaBank penetration hit, there was, to say the least, a public uproar. MegaBank was in a state of crisis. After collapsing by 30% the stock was frozen, but other financial stocks continued to trade, to a detriment. Worldwide, stock markets were in significant decline.

23. It was days before investigators determined that Lemangtrix had been the source of information (though the experts did not identify all of the channels) and that MegaBank had nothing to do with the event.

24. The political and economic costs were profound. MegaBank never fully recovered. Lemangtrix was essentially put out of business. Though governments had made use of the intelligence that firms such as Lemangtrix provided, the reality of the public revolution and revulsion was that all the major data acquisition and analysis firms came under substantial scrutiny and the public became fully aware of the information being acquired. Substantial regulation was in the offing.

25. The combination of the Lemangtrix-MegaBank fiasco and the Medirixx software failure severely undercut the public’s perception of the benefits of Information Technology and Electronic Commerce. A neo-Luddite movement formed throughout the West and high technology, science and rationalism was under attack. It might not have been societal collapse but the events were disruptive to the existing technological foundations. Compelling value propositions were now primarily related with privacy and security rather than with gee-whiz innovations. Firms with complex legacy systems, which formed much of the world’s network and IT infrastructure, were under particular legislative and public pressure.

26. Internationally, Intelligence and Police agencies spent substantial effort at analyzing what actually happened. Following the money,[17] investigators determined that there had been substantial short ordering of Lemangtrix, MegaBank and Medirixx. Fortunately, international agreements regarding financial activities, simplified the process of determining the cash flow. It was determined and certified that the profits (in the hundreds of millions of dollars) had been sent to four specific banks. Each of these four banks was associated with the TCTarget crime organization, which had a global reach and was, amongst other activities, involved with both the drug trade and arms smuggling. Under severe political and public pressure, the intelligence and policing agencies, along with substantial military support, was directed towards the crippling of the TCTarget organization.

27. How was it that TCTarget became the target? The TCTarget arms smuggling had “irritated” TC since TCTarget had been sending arms to a splinter group opposed to TC; this splinter group were of the opinion that TC was too moderate. Further, TC, while merging terror and crime, was led by a certain group of fanatics. They had enough money to fund their efforts, but they were coming under increasing pressure from Western intelligence, police and military. By directing the profits of their short selling to TCTarget they significantly removed that pressure, allowing them much more room to manoeuvre

[16] For a recent possible DDOS on a nuclear plant see [89].

[17] [90] and [91]. Note also that US Homeland Security has created a national computer forensics institute as reported in [92].

and speed up attaining their objectives, while significantly weakening the economic and technological strength of their adversaries.

Scenario 2: Oil and Gas Sector – (present time)

Author: Ann Miller (University of Missouri-Rolla)

1. The oil slick was miles long when the tourist fishing boat "Keys" came across it. The Keys, centred out of Key West, catered to wealthy tourists who went trophy fishing for tuna. It was early morning and there had been no notification overnight. The Keys captain called in the spill to the US Coast Guard and an investigation was launched as to the extent and reason for the occurrence.
2. The spill was clearly massive and it became immediately apparent, especially with the choppy waters and the expectation of sub-hurricane winds and rain in the next few days, that it would be extremely difficult to contain and, in fact, there would need to be a massive cleanup along the Gulf States shores. Massive loss of fowl and marine life accrued. This would be expensive and difficult.
3. Millions of barrels of oil were estimated to be involved, but there had been no reported major breaches of oil tankers in the area.
4. Investigations led to the source of the oil leak as one of the feeder lines of a deep-sea well which had ruptured, spilling millions of gallons of oil and natural gas into the gulf.
5. The well is owned and operated by Oil4U, a mid-size oil and natural gas provider, with a typical network-centric system for their operations. Their network consisted of corporate intranet for accounting, billing, payroll, yield rate calculations, etc, and a SCADA system for Supervisory, Control, and Data Acquisition of their deep-sea wells located in the Gulf of Mexico.
6. Inter-agency and Oil4U investigators noted that the rupture on the feeder line was consistent with an internally generated rupture. It was concluded that the pressure on the feeder line was too great for the joints and welds at the rupture location.
7. As the investigations proceeded, three alternative scenarios were prioritized.
8. **Alternative Scenario 1**: Was there a malicious modification of the SCADA control software, which was programmed to slowly increase the maximum allowable pressure of oil and gas which flowed through the feeder line? If true, such an attack would have been pre-meditated and performed either by the original vendor or an intermediary third party who had access to the original software and made the malicious modifications. This was realistic given that spyware has been found embedded in COTS software. It was also possible that there had been third party intervention on the part of an adversary intelligence agency, terrorist group, or organized crime.[18]
9. **Alternative Scenario 2**: Oil4U acknowledged that their SCADA system failed for a relatively short time due to an accidental error when a maintenance technician brought a laptop infected with a computer virus into the on-shore control room. The virus infected the PCs in the system, turning them into "zombies," which continually flooded the controllers with nuisance messages. While Oil4U operators responded rapidly, rebooting the controllers within minutes, the DoS attack continued, shutting down the controllers every time they re-booted. Unfortunately, during the attack, one of the controllers could

[18] This allegedly occurred above ground in June 1982 when "*the pipeline software that ran the pumps, turbines, and valve settings was programmed to produce pressures far beyond those acceptable to the pipeline joints and welds.*" The changes are alleged to have been implanted in the host software by a foreign intelligence service (see [93]). Further, a recent digital forensics examination detailed an actual case of modified telephone switch software [94].

not read data from a pressure sensor which indicated increasing pressures in one of the feeder lines. Though the operators realized that there was a problem in the SCADA network, there was still a time lag until a manual override was executed and while there was indication of a lower than expected pressure on one of the feeder lines this was not viewed as being particularly abnormal, since some operators ignore or under-set Pressure Sensitive Low (PSL) alarms. Ultimately, the virus was removed, but only after extensive environmental and economic damage.[19]

10. **Alternative Scenario 3**: Oil4U had always claimed that their SCADA network was securely separated from their enterprise network. Communication between the two networks was only allowed through a properly configured, state-of-the-art firewall and routinely followed regular auditing of their SCADA network. Thus, they were surprised when it seemed that the Human-Machine Interface (HMI) of their SCADA network seemed to have been breached. Within ten minutes, OIL4U's automated auditing and analysis tools confirmed their suspicion that an attacker had invaded the SCADA HMI. The operators reacted professionally and diligently worked to mitigate the risk. Within a further twenty minutes the system was restored to normal. However, the operators were unaware that within the first ten minutes the attacker had sent a command to one of the controllers to modify the maximum allowable pressure on its feeder line – to a value in excess of its rated limits. The operators were unaware of the spill due to the PSL settings and only later confirmed the rupture. Their initial focus was to identify how the penetration occurred. Oil4U's control network is only connected to the corporate network and requires passing through a state-of-the-art firewall. The firewall is configured to only allow the billing server to communicate with the control network so as to permit the gathering of billing data in near real time. The billing server had been compromised. To compromise the billing server, there were three entry points: the Wireless LAN, the Modem Bank and the Internet firewall. However, the billing server had no communication with the wireless LAN or modem bank, implying that access had occurred through the Internet firewall. The only traffic allowed through the Internet firewall is the web server and this server is located in the DMZ. The web server passes through the firewall to retrieve billing information allowing Oil4U's corporate customers and contractors access to their account status. The web server had also been compromised and the IP address of the attacker was identified. Oil4U determined that the attacker had compromised a corporate customer's account and that the customer's computer had been turned into a zombie.[20]

11. What actually happened? While all of the above scenarios were all valid, what in fact happened was that a disgruntled employee decided to cause serious financial loss to Oil4U as well as damage to their reputation. She used the wireless hub to log in from home and change the parameters on the feeder line.

[19] That a scenario along these lines is feasible consider the Harrisburg, PA water treatment plant event [95] and, in a self-contained process control lab, see [96].

[20] In January 2003, the SQL Slammer worm began attacking computer networks. Users of the business network at Ohio's Davis-Besse nuclear power plant began to notice a network slowdown. Investigation revealed the worm had spread from the plant's business network to its operations network, causing enough congestion to crash the computerized panel used to monitor the plant's most crucial safety indicators. Minutes later, the Plant Process Computer, another monitoring system, crashed as well. The plant's firewall had initially blocked Slammer, but the worm still managed to reach the plant through a high-speed connection from a contractor's network. Had the plant's operations network been properly protected from either the contractor's network or the plant's own business network, the infiltration would not have happened. (See [97].) It has also been demonstrated in a self-contained process control lab. (Paper submitted for publication by Trent and Miller.)

12. That this form of attack is possible, consider the following, extracted from a European Union briefing paper:[21] “In Queensland, Australia, on April 23, 2000, police stopped a car on the road and found a stolen computer and radio transmitter inside. Using commercially available technology, Vitek Boden, had turned his vehicle into a pirate command centre for sewage treatment along Australia's Sunshine Coast. Boden's arrest solved a mystery that had troubled the Maroochy Shire wastewater system for two months. Somehow the system was leaking hundreds of thousands of gallons of putrid sludge into parks, rivers and the manicured grounds of a Hyatt Regency hotel. Janelle Bryant of the Australian Environmental Protection Agency said "marine life died, the creek water turned black and the stench was unbearable for residents." Until Boden's capture - during his 46th successful intrusion - the utility's managers did not know why. Details of Boden's intrusion, not disclosed before, show how easily Boden broke in - and how restrained he was with his power. Boden had quit his job at Hunter Watertech, the supplier of Maroochy Shire's remote control and telemetry equipment. Evidence at his trial suggested that he was angling for a consulting contract to solve the problems he had caused. To sabotage the system, he set the software on his laptop to identify itself as "pumping station 4," then suppressed all alarms. Hunter Watertech's chief executive admitted that Boden "*was* the central control system" during his intrusions, with unlimited command of 300 control nodes governing sewage and drinking water alike. "He could have done anything he liked to the fresh water", the chief executive said. Like thousands of utilities around the world, Maroochy Shire allowed technicians operating remotely to manipulate its digital controls. Boden learned how to use those controls as an insider, but the software he used conforms to international standards and the manuals are available on the Web. He faced virtually no obstacles to breaking in.”[22]

[21] Text extracted from [98].

[22] Another incident was reported by Computer World Security in [99]. Yet another incident in November 2007 [100].

Scenario 3: Electrical Power Sector – (Circa 2012)

Author: Robin Bloomfield (CSR, City University)

Background

1. Karl Wilhelm stretched his arms over his head, logged-off and got out of his operator's chair. It had been a difficult 24 hours.
2. Over the past few years erratic weather had been causing insecurity (lack of robustness) in the electricity supply. In the winter a combination of extreme cold snaps increasing demands and high winds often caused some local damage to distribution and transmission lines that challenged the European network. In the summer, the increasing use of air conditioning and seasonal variation in alternative generation methods presented their own challenges.
3. The market in electricity was also showing some turbulence. The prices had been widely varying over the past 5 years in part due to changing perceptions in the security of gas and oil supplies. This had lead to a number of takeovers in the deregulated markets covering both the transmission companies and the generating companies. There was a still a proliferation of companies (and also individuals thanks to success of micro-generation is some parts of Europe) that owned generation and transmission network. In the past few years there had been consolidation in the market with considerable investment and ownership from outside Europe.
4. The electricity sector was seen as an example of successful deregulation. Market forces and the political threat of a unified regulatory and control regime across Europe had motivated the industry to improve its investment in control and management systems and to improve overall risk assessment and co-ordination.
5. There had been significant investment in SCADA and control systems following awareness in the mid2000s of the vulnerability to cyber attacks of the previous generation of systems. The new generation of PLCs and smart protection devices had been engineered to remove the egregious problems that were at first suspected in incidents in the Gulf of Mexico (see Scenario1), Australia, etc. Long standing European engineering companies had supplied the equipment but many of the components had been developed and manufactured in the Far East allowing much cheaper and more extensive refurbishment.
6. The new generation of systems relied on one or two Government certified operating systems, and although these were an improvement over previous designs they had gained features and connectivity and lost any remaining security by obscurity that the old propriety systems used to have. However the design was thought to be a great improvement over previous systems particularly in the way it allowed suppliers to build a number of levels of access control to network and remote devices.
7. Not the entire power network was controlled by this next generation of equipment but a sizable part of Europe was controlled PwrProm[23] who had invested in a modernised core business. Their refurbished control (and simulation) centres had allowed for improved training but also a reduction in staffing levels. To address the workload in adverse conditions special co-ordination and evaluation teams had been set up. In part building on the nuclear industry practice of separate teams (because of the different knowledge sets) to improve incident diagnosis and also the nuclear practice to rely on automatic controls

[23] A fictitious French/Russian joint venture.

for 25 minutes following an incident. However this level of reliance on automation had not been possible to achieve due to the coupling and fast dynamics of the grid and the many possible network configurations that could be presented to the operators.

8. This change in ownership had led to the "rationalisation" of many of the corporate networks and information systems and the migration towards more modern international corporate Enterprise standards. This has enabled a more integrated approach but careful best practices had been followed to separate the control and management systems using diverse and multiple firewalls. This was in part a result of a number of CI penetration scares in 2005-2010 (e.g., Oil and Gas Services scenario).

9. Although the control and stability of the network had improved in recent years, not all physical interactions were understood or predicted beforehand. The growth of the network and the success of micro-generation had led to some surprises. Although the 2007 Polish- Portugal oscillation in the frequency of the supply was now well understood there were still transient flows that surprised the power experts.[24] The increased in decentralised generation (wind farms, micro-generation) remained a modelling and control challenge and this was compounded by the more recent Service Level Agreements (SLAs) not requiring such quality of control e.g. during a frequency excursion[25] the wind generators often disconnected more rapidly then fossil fired plant. This protected the generators' investment at the expense of the overall network.

10. Risk analysis and sharing of information made for a more resilient and aware operation. While there was no centralised control of the European grid, these proposals having faltered in the success of the distributed market driven approach, long gone were the days when each distribution controller could only phone or fax the controllers of management centres around Europe. More modern information sharing had been put in place following the European problems of early 2000s and there was some real-time risk assessment of security supply and external threats.[26] Control centres had displays – risk "dash-boards" - that included a summarised traffic light status of interdependent networks (e.g. so the telecoms operators could see some limited operational information on the energy supply status).

The incident develops

11. Although the winter had seen some severe challenges to the interconnected grid and to the local distribution to supplies, these had been resolved quickly. There had been some rural areas where damage to mobile phone transmission systems and a shortage of skilled maintainers had increased the outage beyond what was anticipated most people had been reconnected within days rather than weeks. Power engineers were still studying the performance of the grid over the winter as there had been some unusual oscillations in the power that were not understood, (nor were they serious).

[24] For example, in trying to match analysis with the Nov 2006 German incident. See [101] and links thereof.

[25] The frequency and phase of the electricity supply has to be controlled. The frequency is held around 50Hz in Europe and as more power is demanded it will dip and then be restored as generation increases. Power stations and the control systems, governors, on their turbines try and maintain the frequency but if this fails there is a frequency collapse, disconnection and subsequent blackout. More details on power systems can be found at [102]

[26] See for example the goals of IRRIIS [52].

12. It was now March a cold, but not an unseasonably, cold day. The operating margin for the central sector of the grid was lower than usual but the single failure[27] criterion was still nominally satisfied and the contingency analysis tools did not indicate any unsatisfactory risks. Karl Wilhelm surveyed his computer screen as saw the usual minor alarms and status messages. Across the control room the relatively new display that showed the risk status of neighbouring electricity operators and the national Telco was at their normally constant green/ok. His experienced colleague, Helmut Boch, was behind him in the shift office filling in paperwork, or perhaps taking a break.
13. Further south, severe weather was affecting the Alps. Although there had been a scarcity of snow this winter the winds and storms had been quite intense. He had heard on the news the possible problems that this was causing and some orange alert message popped up on his screen advising him of a possible request for greater transmission across his area. He acknowledged this and began to watch carefully as the market requests went in for increased power from northern Europe and the plants slowly began to respond.
14. He soon began to see an escalating demand for power and began to urgently consider the routing options before him. Two of his small links were repeatedly tripping out and the main X-Y line was operating above its normal conditions. This was perfectly allowed from time to time but only gave him about 40 minutes to find alternatives.
15. The phone rang and Helmut came across to say that the management information system had just gone down. It was not “patch Tuesday” but the system had been under maintenance and there had been some difficulty bringing it back up. They were having to role back to a previous configuration, an activity they often rehearsed, but this would take 30 minutes to reboot. The screen across the room that provided some overall management information began to grey out indicating that their data was stale. The “risk status” of the other infrastructure went out at the same time and both wondered out loud how they were connected to the management IT but they started to flick again soon afterwards indicating they were coming back up[28].
16. They soon begun to notice abnormal flows in the network, however much of the control was now automated and the protection could be seen working causing transmission lines to trip out and then reconnect. The system configuration and dynamics was complex and fast moving and Karl had no way of checking that the protection was working as specified. A few of the disturbances propagated and began to challenge the anti-cascading technology. This was playing havoc with the system dynamics and the frequency was fluctuating close to the limits. The operators attempted to stabilise the situation but the problems developed further leading to isolation of part of the transmission network.
17. While Karl was attempting to reconnect part of the network he noticed that the “risk status” of Belgium had turned red and was flashing indicating as serious security incident was predicted imminently or in progress. This was confirmed by a barrage of messages they were getting alerting them to unspecified but malicious disruption. They were also seeing news feeds from CNN where a previously unknown organisation, x’HFS, was posting claimed exploits on the web confirming the threats. Following the operating procedures, and as a precaution, they put a call out for extra staff to come in from their holiday. They shifted their attention to following the procedure to reduce their dependence on part of the network under attack by reducing the north-south flows. Unfortunately the unusual operating configuration caused by the bad weather in the south and the already difficult situation led to some clumsy disconnections. They worked hard

[27] Industry currently plans for loss of any single element (N-1 contingencies) or, less frequently, for the loss of multiple elements (N-X). In practice this contingency risk analysis is quite complex. See [103].

[28] A recent UK financial infrastructure system had (unrelated but confusing) problems and at the same the MIS went down.

to stabilise the remaining network but much was in the hands of the automation that seemed to do what it was designed to do, shedding load, isolating the network and shutting regions. They of course made some mistakes as they working under extreme pressure. The networks stabilised and a series of islands but significant urban centres were left disconnected.

18. Some long range indicators worked giving operators in Spain and Italy enough information to isolate their systems in time and some of the long range anti-cascading technology worked but the indication from the control centre was that 50% of the lights had gone out in Europe.
19. The control room team now paused their control actions to assess the situation. Training in incident recovery had identified special roles for the recovery team and they sought to stabilise the worst effected areas and establish the priority for reconnection based on the contingency plans and models. It soon became clear to then there was a serious security incident taking place. The security threat level had increased and the control and protection systems demanded the highest levels of authentication through trusted connections before they would obey any restart or reconfiguration commands. This security lock down also affected the black start capability of some of the power stations further prolonging the blackout.
20. These problems were compounded by the fact that many companies used the same outsourced IT security expertise that was in demand elsewhere. Chaos in transport and the need to help friends and family put further pressure on staffing levels. This was exacerbated by increasing rumour and an external claim, that Europe faced an escalating cyber attack. Having lost trust in their own information systems the operators improvised using ad hoc mobile networks and informal internet connections to verify the system data and to maintain contact with their families. Luckily operators also had personal and professional ties having trained with each other and the control centres and stations begun to work out a co-ordinated picture of what the situation was and worked steadily to restore supply. The weather began to moderate and after a difficult 24 hours most urban centres had been reconnected to the main grid but the there were continuing problem at a local level.
21. Some of the local centres required manual intervention to reset and replace failed protection devices. In a few urban areas there had been some what appears to be spontaneous, locally organised, unrest hampering restitution by burning cars in the street and threatening repair crews and causing confusion and overload for the emergency services.
22. Restitution was also being hampered from an unexpected quarter.[29] The control rooms soon had the CEO and company lawyers on the phone asking strangely detailed questions about what had been effected, how much was due to their action and how much was automated. They began to feel uneasy about the attitudes and being treated as if they were responsible. Of course they had made some a few mistakes while operating under pressure but they felt the management interest was misplaced. What was happening was the supply chain had sought legal redress for the damage caused and there were legal actions to secure evidence even at the expense of restoring operation. The need for evidence collection and uncertainty of the legal situation had paralysed some of the senior management.[30] There were suggestions that operators were personally liable for the priorities of the decisions made in reconnection and the subsequent, often implicit, trade-offs that (e.g. that company X lost more than Y, distress to person X increased but not to Y as couldn't get to hospital in time). [31]

[29] Might also like to add that MIS problems lead to leakage of billing and bank account data causing yet more management distractions.

[30] We worked with some insurers who alerted us to these issues of restitution.

[31] Responsibility modelling might have been deployed. This looks for vulnerabilities in certain alignments and patterns of responsibilities.

23. Luckily the telecoms were largely unaffected although there was some loss of local loops due to power problems. The mobile operator has secured supplies so that for the first responders. They continued to operate but as the power remained down they began to reduce their service to those with the "essential service" phones in the few urban centres affected. There had been some rural areas where damage to mobile phone transmission systems and a shortage of skilled maintainers had increased the outage beyond what was planned most people had been reconnected within days rather than weeks.
24. The official reports into the incident found that the claims of exploits made by x'HFS were not credible and they were just seeking publicity. The root cause of the incident was the difficulties in controlling the power system following the bad weather and some of the malfunctioning protection devices. This had been attributed to some mis-calibration of trip thresholds by outsourced contractors and this was being remedied[32]. Some protection devices had been found to have failed with memory problems and this was attributed to the problems in manufacture and test (a consequence of cheap components). QA procedures and acceptance testing had been tightened and there had been checks on the protection devices concerned.[33]
25. The MIS had in fact been attacked and it seemed some "script-kiddies" – not exactly "kiddies" but recreational and hobbyist hackers – had got lucky replaying some of the tactics used in the Oil and Gas Sector scenario to access and disrupt the risk status warning system. This unexpected interconnectivity that the "kiddies" stumbled across was somewhat surprising as the system design came from a research project focused on interdependencies but the product had rather quickly transitioned from research into deployment on the basis that it was not directly operational and could not worsen the situation as there would be external confirmation of its messages.
26. The lack of MIS and the security lock down meant that only limited local forensic information had been captured. The stock market movements had not been investigated as this was not considered a large security incident.
27. However the most likely explanation offered by the incident investigators is not necessarily the most accurate one. Someone somewhere was considering their next step.

Postscript

Adversary strategy

In this complex system it is over simplistic to think of an attack being designed and executed by a single, identifiable agent. While this is of course possible, the scenario shows how a number of malicious, uncoordinated events together with stresses due to accidental events, environmental effects and opportunistic escalation challenge the resilience of the complex system: using a disease metaphor, there is no need for opportunistic infections to co-ordinate. Not only had the adversary absorbed the vocabulary of systems of systems and "boundaryless" systems but had developed an approach that broadened effects-based planning to a more ecological or bio-inspired view. The incident described is just one of the many routes through to failures: it is not just a "security" event.

The scenario uses a mixture of "attacks" that are uncoordinated in the normal sense of having short term communication between the actors but rather through a long term, decentralised understanding and intent. The use of markets and ownership may require large nation-state levels of resource to influence and to hide their intentions.

[32] Know of an incident was caused by a misconceived modification of smart device and was propagated by modification in a piece of CI

[33] Restoration under pressure, mixed teams from everywhere could spread bad component, put things in place for next time.

The threat agents assumed that challenges to the power system were bound to occur so pursued tactics that would increase the magnitude of these effects, make their management more difficult and especially hamper recovery. The adversary made significant use of system engineering and risk analysis skills to understand complex systems. They were aware of “normal accident” theory and had knowledge of complex systems and the relationship between topology and cascading effects. They had the ability to run simulations, picking up on published work and the availability of topological information. They had designed their approach so it did not rely on a single type of attack or a single vulnerability as these could be found: their approach is *n-1* secure. They had understood that consequences in interdependent infrastructures increase non-linear with time to repair. [34]

The difficulty in operation caused by bad weather had been amplified by:

- Behaviour of the protection devices
- Incorrect risk information and MIS problems had made recovery more difficult

However hampering recovery and causing losses would be possible without the sophisticated compromise of the protection devices. The adversary might have been satisfied with just prolonging the disruption or could initiate it with more conventional attacks (e.g graphite on sub stations, fires etc).

They had sown an easily detectable “bug” by using a friend in the contracting company to improperly set the trip levels in some of the protection devices. The malicious code they had inserted by intercepting the devices after manufacture had not been detected, despite the industry approved audits and assessments that had taken place and some had been activated and tested in this incident. However once run it also self destructed.[35] They had made use of computer science tools for reverse engineering and mathematical modelling of the sensor code. The adversary had access to the code, unlike the industry assessors and auditors.

They had also learnt about the forensic capability of the system, the times to restore, had cast doubt on the risk information system, learnt how to stimulate consequential social unrest and exploit the legal system.

Another theme was the desire to escape detection, to have the incident seen as a conjunction of several unfortunate events. Part of the scenario was the doubt that it was hard to decide whether it was orchestrated or not. This was in part to avoid the massive retaliation that might follow an overt attack but also as a learning exercise[36] so retain the potential for future escalation and pave the way to a “shock and awe” variant.

Threat Targets

The scenario concerns a system of systems – the electricity generation and supply systems and its supporting control and maintenance systems and the people, organisations and institutions that are stakeholders. The “system” included the

- Technical system of electricity supply and control
- The organisation and management of that system
- The legal and insurance system

[34] So damage something like = geographic impact*time^n. Would be interesting to run some models to investigate further.

[35] Might have inserted code to be date specific, or randomly start. Could be wiped by rebooting where OK version then reloaded but would then need to control type of reboots. Would need to compromise checksum and fault detection measures. Could be inserted before shipments, as in this version, but the connectivity of the devices, albeit over a partially Government approved network, might lead to credible scenario with external attack.

[36] The adversaries were familiar with the High Reliability Organisation [HRO] literature.

- The supply chain
- The maintenance system

Also some abstract things such as

- Situational awareness
- Doubt and confusion
- Confidence in the system

In fact thinking of threat targets as the more concrete concerns misses important points (e.g. a good attack on an eVoting machine might be in the voters' confidence in the accuracy and robustness of the system).

Appendix B

The purpose of this appendix is to show how the various evaluation, validation and verification techniques, listed in Appendix C of this report, may be used to assess a complex ICT system. We will use three scenarios developed by a NATO Research Task Group (RTG) to illustrate how the different techniques outlined in the report could be used to reduce the probability of occurrence of incidents and shortcomings outlined in the scenarios.

The next three subsections will present a bottom-up analysis of each of the scenarios. The structure of the table for each sub-section is the same. The columns

- The first (left-most) column contains a sequential numbering of the analysis for easier cross-referencing
- The second column lists main parts of the scenario from the Security point of view
- The third column lists and explains which existing or emergent validation and verification techniques could be used to best address the scenario's security problems identified in the first column

Analysis of Scenario 1: Medical and Financial Services Sector – (present time)

No.	Specific issue in the Scenario	How we might address it with existing or emergent V&V techniques
1	In points 13 and 14 of the scenario it is stated "13. … Substantial design effort went into developing a highly trusted system. Medirixx, however, like other firms, felt the pressures of the competitive market place and decided to *offshore* the coding. SafelyC was hired after a due diligence review by Medirixx and some early trials. 14. SafelyC, however, also felt commercial pressures, and unbeknownst to Medirixx, *subcontracted* out portions of the work (to OnwardsC) and, to make sure that full interoperability was maintained, included overall design information of the Medirixx system."	In this scenario the outsourcing company further outsourced their development and it was OnwardC which was run by a criminal/terrorist group. It is very difficult to defend against these types of deliberate vulnerabilities by outsourced companies, but both SafelyC and Medirixx could have reduced the probability of the vulnerability being present on the system by using the following techniques: ➢ Checking and vetting procedures for components developed by contractors (these may include static techniques such as: ♦ Inspections ♦ Walkthroughs ♦ Code Reviews ♦ Compliance with standards ➢) as well as ♦ Review of Quality of supply ♦ Supplier competency ♦ Supplier process assessment ➢ Security cases Validation and verification techniques (both at SafelyC and at Medirixx) would complement the above techniques to improve the product security against hacking attacks in general are: ➢ Use of *Static analysis* and *formal methods* for the developed code to minimise the risk of security vulnerabilities left in. ➢ *Testing* (including hiring special "tiger" teams of hackers) to do *penetration testing* on the delivered system and check for vulnerabilities. ➢ *Review of compliance with best practice documents* for configuration of defence devices such as Firewalls, Intrusion Detection Systems (IDS), Intrusion Prevention Systems (IPS), Access Control Rights etc. The last bullet point, with respect to configuration of the firewall, is much more difficult in this scenario as

		the incorrect dosage levels were downloaded during 'legitimate' patching of the system and hence these connections would have been allowed to proceed in the firewall. Hence in these cases the most likely defence would have been some from of architectural *diversity* in the system: ➢ *Redundancy* and /or *Diversity* in the system deployed (i.e. the deployment of more than one version of drug dosage control developed from a contractor other than SafelyC). However the presence of diversity complicates the architecture and may lead to performance penalties (hence a more thorough assessment is required to assess the tradeoffs between functionality, reliability, security and performance) ➢ *Wrappers* (could plausibility checks have been written for acceptable limits of dosage given to patients) ➢ Watchdogs
2	In point 19 of the scenario it is stated: "19. But, this was only part of the story. TC had greater plans than an apparent financial motivation. Through their expertise and contacts with the black hat ecosystem, TC had successfully penetrated Lemangtrix – a huge player in the burgeoning space of world-wide personnel data acquisition and analysis, with significant contacts into government and intelligence agencies. Lemangtrix had information on over a billion individuals, with terabytes of data. Lemangtrix made use of state-of-the art hardware, software and networking technology. They were fully aware that penetration of their database, similar say to the problems that TJX had in 2007 and previous years, could cripple their business."	TC had managed to penetrate the Lemangtrix personnel database via social engineering and insiders in the company. Very difficult to protect against these types of attacks. Possibilities for reducing the probability of this type of attack include: ➢ *Review* of *security policy* to ensure that rules for higher privilege access rights are tightened ➢ *Responsibility modelling* so that roles and responsibilities within an organisation are properly understood and any gaps are identified. This may also help with the definition of a sound security policy. As a defence against penetration attacks from outside then the following review should also be done: ➢ *Review of compliance with best practice documents* for configuration of defence devices such as Firewalls, Intrusion Detection Systems (IDS), Intrusion Prevention Systems (IPS), Access Control Rights etc. Data backup procedures also should be checked to ensure that backup files are securely stored and proper access rights are in place (which should be defined in the *Security Policy*).
3	*Security economics* issues pertinent in the scenario.	Even though the end aims of TC were proved to be the elimination of the rival organisation TCTarget and

		creation of panic and confusion in the pubic of the enemy governments, the intermediate steps that TC used to get to its final aims are also related to the area of *Security economics* (see a recent report on this field [33]). Several recommendations have been made in the aforementioned report about changes in legislation to improve the security of systems. Even though most of the recommendations within them contain the assessment techniques we have listed in Appendix C, the architects of systems and solutions ought to be aware of the complex economic and financial incentives behind the attackers' actions. This insight may help the architects to better understand the threats and the motivations and therefore lead them to the building of more effective defence mechanisms in their systems.
4	In point 20 of the scenario it is stated: "20. TC and their black hat associates had carefully acquired tens of millions individual records from Lemangtrix and had made use of their own Botnets (and peer-to-peer technologies) to distribute the data "securely" on compromised PCs globally. (The data was acquired through various means including social engineering, improper backup procedures, and penetration of the Lemangtrix databases.)".	The compromise on the system and the techniques available to reduce the probability of the compromise occurring were covered in 2 above. Point 20 of the scenario details the methods used by the company to distribute the compromised data and therefore reduce the public trust (and hence the share price) of Lemangtrix and MegaBank. The response by MegaBank and Lemagtrix following the divulgence of the data will have to be largely defensive aiming to reduce the uncertainty and panic of their customers.
5	Forensic aspects to do with "following the money" issues raised in point 26: "Internationally, Intelligence and Police agencies spent substantial effort at analyzing what actually happened. Following the money, investigators determined that there had been substantial short ordering of Lemangtrix, MegaBank and Meditrixx. Fortunately, international agreements regarding financial activities, simplified the process of determining the cash flow. It was determined and certified	This point of the scenario raises important issues with respect to the forensic investigation aspects used by law enforcement and high-tech crime investigation units. TC in this case have been quite successful in covering their tracks and shifting the responsibility to the rival organisation TCTarget which suffered the brunt of the retaliation.

	that the profits (in the hundreds of millions of dollars) had been sent to four specific banks. Each of these four banks was associated with the TCTarget crime organization, which had a global reach and was, amongst other activities, involved with both the drug trade and arms smuggling. Under severe political and public pressure, the intelligence and policing agencies, along with substantial military support, was directed towards the crippling of the TCTarget organization”	

Analysis of Scenario 2: Oil and Gas Sector – (present time)

No.	Specific issue in the Scenario	How we might address it with existing or emergent V&V techniques
1	The scenario first describes that an oil spill has happened at an offshore Oil extraction platform in the Gulf of Mexico. It then proposes three alternative explanations of what could have caused the oil spill. Alternative explanation 1: "Was there a malicious modification of the SCADA control software, which was programmed to slowly increase the maximum allowable pressure of oil and gas which flowed through the feeder line." *Description continues in the next row ...*	Regardless of whether the modification was malicious or not, better testing of the product (either by the original vendor of the COTS component or in-house once the system was delivered) would have increased the probability of the fault being detected and fixed. Therefore some of the testing techniques, listed in Appendix C of this report, would have helped with increasing the probability of preventing the fault from remaining undetected in the COTS component before it was made operational. A recommendation would be to use the *Boundary value analysis* technique to check how the system reacts to extreme values. In more detail the following evaluation would help here: ➢ *Black box and white box testing* (the latter, if the COTS code is available) ➢ Static analysis ➢ Formal method techniques ➢ Checking and vetting procedures for components developed by off-the-shelf component suppliers
2	*The scenario continues from the sentence in 1 above:* "If true, such an attack would have been pre-meditated and performed either by the original vendor or an intermediary third party who had access to the original software and made the malicious modifications. This was realistic given that spyware has been found embedded in COTS software. It was also possible that there had been third party intervention on the part of an adversary intelligence agency, terrorist group, or organized crime."	If the product vendor has left a backdoor on the COTS software then it is very difficult to protect against them. Continued monitoring of the system from human administrators is a possibility but this may be expensive and not clear whether it would have had an effect in this case. When COTS components are used in safety critical environments more thorough testing is required as mentioned in 1 above. But high assurance may be difficult to obtain without using of some form of diversity in the system architecture: ➢ *Redundancy* and /or *Diversity* in the system deployed (i.e. the deployment of more than one version the COTS component). As mentioned in the earlier scenarios, there are tradeoffs between functionality, reliability, security and performance when diverse components are employed as complexity of the system architecture is increased. ➢ *Wrappers* (could plausibility checks have been written to check for maximum allowable pressure of oil and gas which flowed through the

		feeder line?) ➢ Watchdogs
3	The 2nd alternative explanation for the cause of the oil spill: "Oil4U acknowledged that their SCADA system failed for a relatively short time due to an accidental error when a maintenance technician brought a laptop infected with a computer virus into the on-shore control room. The virus infected the PCs in the system, turning them into "zombies," which continually flooded the controllers with nuisance messages." *Continues in the next row…*	Possibilities for reducing the probability of this type of attack include: ➢ *Review* of *security policy* to ensure that a sandboxed environment is created for new machines (or even known machines with an uncertain security status since last connection in the system) until they are deemed safe to be reconnected to the network. This should be done automatically for all machines, and circumventing this restriction should only be allowed to a restricted number of individuals and only for a well-defined set of tasks. ➢ *Responsibility modelling* so that roles and responsibilities within an organisation are properly understood and any gaps are identified. This may also help with the definition of a sound security policy. As a defence against penetration attacks from outside then the following review should also be done: ➢ *Review of compliance with best practice documents* for configuration of defence devices such as Firewalls, Intrusion Detection Systems (IDS), Intrusion Prevention Systems (IPS), Access Control Rights etc.
4	*Continues from previous row*: "While Oil4U operators responded rapidly, rebooting the controllers within minutes, the DoS attack continued, shutting down the controllers every time they re-booted. Unfortunately, during the attack, one of the controllers could not read data from a pressure sensor which indicated increasing pressures in one of the feeder lines. Though the operators realized that there was a problem in the SCADA network, there was still a time lag until a manual override was executed and while there was indication of	The scenario illustrates the importance of *human factor analysis* during system design especially the consideration of: ➢ the level of human intervention required by the system ➢ the complexity of operator actions required ➢ the potential for operator error to directly lead to a system failure ➢ the level of training available to the operator *Risk assessment* would also help for assessing and quantifying the risk of failure both from safety as well as reliability or security point of view: ➢ Hazard and Operability Studies (HAZOPS). ➢ Failure Modes and Effects Analysis (FMEA) ➢ Event Tree analysis

	a lower than expected pressure on one of the feeder lines this was not viewed as being particularly abnormal, since some operators ignore or under-set Pressure Sensitive Low (PSL) alarms. Ultimately, the virus was removed, but only after extensive environmental and economic damage."	➢ Fault tree analysis and dependence diagrams
5	The 3rd alternative explanation for the cause of the oil spill: "Oil4U had always claimed that their SCADA network was securely separated from their enterprise network. Communication between the two networks was only allowed through a properly configured, state-of-the-art firewall and routinely followed regular auditing of their SCADA network. Thus, they were surprised when it seemed that the Human-Machine Interface (HMI) of their SCADA network seemed to have been breached. Within ten minutes, OIL4U's automated auditing and analysis tools confirmed their suspicion that an attacker had invaded the SCADA HMI. The operators reacted professionally and diligently worked to mitigate the risk. Within a further twenty minutes the system was restored to normal. However, the operators were unaware that within the first ten minutes the attacker had sent a command to one of the controllers to modify the maximum allowable pressure on its feeder line – to a value in excess of its rated limits. The operators were unaware of the spill due to the PSL settings and	This is yet another "access via the backdoor due to a compromised trusted source" case (in this instance a trusted corporate customer account is the trusted source). Similar to recommendation in row 3 above, the Security policy needs to be reviewed especially with regard to: ➢ *Review of compliance with best practice documents* for configuration of defence devices such as Firewalls, Intrusion Detection Systems (IDS), Intrusion Prevention Systems (IPS), Access Control Rights etc. This would ensure a better separation of the safety critical SCADA parts of the system from the billing and management parts. If sufficient assurance cannot be obtained about the effectiveness of the protection devices (firewalls IDS etc) then a review of the *architectural design* of the protection systems out to be done. *Diversity* in the protection devices deployed (such as diverse Firewalls, diverse IDSs) may bring higher assurances about the secure separation of SCADA sub-system from the billing and management sub-system

	only later confirmed the rupture. Their initial focus was to identify how the penetration occurred. Oil4U's control network is only connected to the corporate network and requires passing through a state-of-the-art firewall. The firewall is configured to only allow the billing server to communicate with the control network so as to permit the gathering of billing data in near real time. The billing server had been compromised. To compromise the billing server, there were three entry points: the Wireless LAN, the Modem Bank and the Internet firewall. However, the billing server had no communication with the wireless LAN or modem bank, implying that access had occurred through the Internet firewall. The only traffic allowed through the Internet firewall is the web server and this server is located in the DMZ. The web server passes through the firewall to retrieve billing information allowing Oil4U's corporate customers and contractors access to their account status. The web server had also been compromised and the IP address of the attacker was identified. Oil4U determined that the attacker had compromised a corporate customer's account and that the customer's computer had been turned into a zombie."	
6	The scenario description then divulges that the actual cause of the oil spill was "… a disgruntled employee decided to cause serious financial loss to Oil4U as	The assessment, verification and validation techniques to reduce the probability of this event are similar to those described in 2, 3 and 5.

	well as damage to their reputation. She used the wireless hub to log in from home and change the parameters on the feeder line.”	

Analysis of Scenario 3: Electrical Power Sector – (Circa 2012)

No.	Specific issue in the Scenario	How we might address it with existing or emergent V&V techniques
1	Malfunctioning of the protection devices due to mis-calibration of trip thresholds by *out-sourced* contractors	Regardless of whether mis-calibration was deliberate or not, better testing of the product (either by the out-sourced contractors or in house once the system was delivered) would have increased the probability of the fault being detected and fixed. Therefore the testing techniques listed above would have helped with increasing the probability of preventing the fault from remaining undetected. A recommendation would be to use the *Boundary value analysis* technique to check how the system reacts to extreme values. In more detail the following evaluation would help here: ➢ Black box and white box testing ➢ Static analysis ➢ Formal method techniques ➢ Checking and vetting procedures for components developed by contractors (such as static analysis techniques as well as review of quality of suppliers): ♦ Inspections ♦ Walkthroughs ♦ Code Reviews ♦ Compliance with standards ♦ Review of Quality of supply ♦ Supplier competency ♦ Supplier process assessment ➢ Security cases
2	Malfunctioning of the protection devices due to memory problems caused by poor testing and quality assurance after the use of cheap *third party components*	Same recommendation as for 1 above.
3	In the running commentary of the scenario we have the following claim: “The attackers had sown an easily detectable ‘bug’ by using a friend in the contracting company to miss-set the trip	The attackers had used an insider in the company to insert malicious code. Very difficult to protect against these types of attacks. Possibilities for reducing the probability of this type of attack include: ➢ *Review* of *security policy* to ensure that rules for higher privilege access rights are tightened

	levels in some of the protection devices. The malicious code they had inserted by intercepting the devices after manufacture had not been detected, despite the industry approved audits and assessments that had taken place, and some had been activated and tested in this incident. However once run it also self destructed."	➢ *Responsibility modelling* so that roles and responsibilities within an organisation are properly understood and any gaps are identified. This may also help with the definition of a sound security policy.
4	Script kiddy programs were used to access and disrupt the risk status warning system (part of the management system)	Even though the scenario does not specify which script kiddy programs in particular were used, there is a danger from the growing number of automated malicious tools which allow even inexperienced users to launch sophisticated attacks with little training. Validation and verification techniques that will help to improve the product security against hacking attacks in general are: ➢ Use of *Static analysis* and *formal methods* for the developed code to minimise the risk of security vulnerabilities left in. ➢ *Testing* (including hiring special "tiger" teams of hackers) to do *penetration testing* on the system and check for vulnerabilities. ➢ *Review of compliance with best practice documents* for configuration of defence devices such as Firewalls, Intrusion Detection Systems (IDS), Intrusion Prevention Systems (IPS), Access Control Rights etc. ➢ Review of the *architectural design* of the system. The presence of architectural controls: ♦ *Redundancy* and /or *Diversity* in the protection devices deployed (such as diverse Firewalls, diverse IDSs), ♦ Wrappers ♦ Watchdogs ➢ After the product is released, then *regular patching* of the product will be the needed to maintain continued security.
5	The following claim is made on the scenario: "The attack on the management system may be a coordinated attack with the attack on the control system or it may be a separate attack from another	From the validation and verification point of view techniques to improve the system security are similar to those mentioned in 4 above. Additionally, for coordinated attacks, extensive *data collection* and use of real time alerting tools (such as the firewalls, IDS and IPS tools mentioned in 4) may allow a system administrator to anticipate attacks against one sub-

	terrorist / criminal group.”	system (in this case the management sub-system) based on knowledge of attacks against another sub-system (in this case the control sub-system). Tradeoffs in configuration will exist though between increasing the detection (reducing the false negative rate) and reducing the false alarms (reducing the false positive rate).
6	Operation and ownership of the market had reduced safety margins. Market mechanisms were efficient at aggregating some properties (efficiency) but not others (resilience, redundancy).	Several layers of operation and ownership may be involved: ➢ those that regulate the market, ➢ ownership of the generator and ➢ transmission companies. There is a need for proper regulations both in the industry and imposed by national governments and/or EU commission to ensure that those that regulate and evaluate the market are independent from the contractors, owners and transmission companies, to guarantee impartiality and reduce the conflicts of interest. This strategy is practiced in the British Nuclear industry.
7	A lack of security from obscurity, diversity of technology, operation and ownership.	The availability of open-standards has many benefits from the viewpoint of reliability and ease of development (due to the possibility of tapping talents from many countries and across different cultures) but this brings difficulties with the loss of security from obscurity. System designers and critical infrastructure regulators need to be aware of the open standards and incorporate this in the security policy. One strategy would be to have extensive security evaluation by the open standards community to identify the potential flaws in the system architecture. In this the lack of obscurity can be turned into an advantage rather than a drawback.
8	High risk of cascading failures to other critical infrastructures (telecoms). The scenario depicts an incident in which the telecom service in badly affected areas was reduced to only emergency/essential communications.	In complex systems of systems it may become difficult to place the boundaries of the system. The power grid has implication on the telecom network and breakdown in the power supply may lead to social unrests as was seen in the scenario. Various *Risk Analysis* techniques should be used to assess the risks of major power failures and ensure that sufficient contingency plans are put in place. This may also require long term public policy initiatives from the national governments to *diversify* the source of electricity supplies and ensure that enough redundancy and diversity is present in the network infrastructure to reduce the risk of major failures that propagate and cascade through the network.
9	The scenario contains the	One of the reasons reported for the mid-air collision of

	following text: "The phone rang and Helmut came across to say that the management information system had just gone down. It was not "patch Tuesday" but the system had been under maintenance and there had been some difficulty bringing it back up. They were having to role back to a previous configuration, an activity they often rehearsed, but this would take 30 mins to reboot."	two aircrafts over Uberlingen in 2002 was that the main radar system was being backed-up and was unavailable to the air-traffic controller who (having to manage two workstations at the same time) gave the wrong/delayed advice to Flight 2937 leading to the collision. The issues of having a critical part of the system down for a period of time (even if very short) needs to be carefully studied and ensure that sufficient *redundancy* and *diversity* (either in the form of redundant/diverse systems or additional human personnel) are in place to mitigate and shorten the at-risk-time.
10	The scenario contains the following text regarding cascading failures: "The system configuration and dynamics was complex and fast moving and Karl had no way of checking that the protection was working as specified. A few of the disturbances propagated and began to challenge the anti-cascading technology. This was playing havoc with the system dynamics and the frequency was fluctuating close to the limits. The operators attempted to stabilise the situation but the problems developed further leading to isolation of part of the transmission network."	The cascading failures problem depicted here is a serious issue in complex and seemingly boundaryless systems as the Power grid depicted in the scenario. *Risk analysis* (such as *Event-tree-analysis*) may help to flesh out the sequence of events that may happen if there is a failure and hence identify in advance to which parts of the system will the failure propagate to.
11	The scenario contains the following text: "Unfortunately the unusual operating configuration caused by the bad weather in the south and the already difficult situation led to some clumsy disconnections. They worked hard to stabilise the remaining network but much was in the hands of the automation that' seemed to do what it was designed to do, shedding load, isolating the network and shutting regions. They of course made some mistakes as they	The text highlights the important role the humans play in these complex socio-technical systems. Hence *Human factor analysis* and *Responsibility modelling* are very important both prior to the deployment as well as during the operational use of the system. All parts of *Human factor analysis* listed in Appendix C of the report (e.g.availability and quality of operation and maintenance instructions; the level of human intervention required by the system; the complexity of operator actions required; etc) would have helped mitigate the risk depicted in the scenario.

	working under extreme pressure. The networks stabilised and a series of islands but significant urban centres were left disconnected.”	
12	The following text exists in the scenario: “The control rooms soon had the CEO and company lawyers on the phone asking strangely detailed questions about what had been effected, how much was due to their action and how much was automated. They began to feel uneasy about the attitudes and being treated as if they were responsible. Of course they had made some a few mistakes while operating under pressure but they felt the management interest was misplaced. What was happening was the supply chain had sought legal redress for the damage caused and there were legal actions to secure evidence even at the expense of restoring operation. The need for evidence collection and uncertainty of the legal situation had paralysed some of the senior management. There were suggestions that operators were personally liable for the priorities of the decisions made in reconnection and the subsequent, often implicit, trade-offs that (e.g. that company X lost more than Y, distress to person X increased but not to Y as couldn’t get to hospital in time).	The problems depicted here stems from unclear *Responsibility modelling*. As listed in section Appendix C of the report, responsibility modelling makes a clear distinction between *Causal responsibility* - the obligation to ensure that some state of affairs comes about or is/is not maintained; and *Consequential responsibility* - the obligation to take the blame if some state of affairs does not come about or is/is not maintained. A clear definition of the responsibilities would reduce the need of interfering with operators during peak hours, which undoubtedly affects morale.

Applicability of techniques to general SoSs and their constituent components (including COTS)

This section lists list the techniques (described in Appendix C of the report) in a table and discusses whether they apply for the assessment of complex Systems of Systems, or their constituent parts (including COTS components) in general.

The table is structured as follows:

- The first column numbers the technique for easier cross referencing.
- The second column specifies the type of the assessment technique.
- The third column lists the assessment technique (in cases where the technique could b e further sub-grouped then we give the technique name as, for example, Sub-group: sub-sub-group: technique name.
- The fourth column asks, in simpler terms whether the technique can be applied for assessment of SoSs.

No.	Technique family	Technique name	Can the technique be used for assessment of SoSs?
1	*System development Life-cycle methods*	*Compliance with processes defined in standards*	Yes: compliance with standards (either for the whole SoS or for parts of the SoS) will most probably be required for most future SoSs
2		*Audit of specific aspects of a lifecycle*	Yes: however the lifecycle of the SoS is likely to be much more complicated as different parts of the SoS may be developed by different suppliers before they are integrated in the main system.
3		*Use of iterative development processes*	Yes: but due to the complex and compartmentalised nature of the SoSs the iterative process will need to be done at lower levels of granularity (at component or sub-system level)
4		*Competence management*	Yes: this will be required at many levels especially in the SoSs integration and the subsequent running.
5		*Review of the requirements process*	Yes: for the global set of requirements that the SoS is required to fulfil as well as the requirements of each individual component or Sub-system.
6	*Review of Quality of supply*	*Supplier competency*	Yes: if the different parts of the SoS are to be developed by a third or eternal party there needs to be a proper review of the competency of the suppliers (especially for safety-, security-, or mission-critical SoSs.
7		*Supplier process assessment*	Yes: same reason as for 6 above.
8	*Architecture*	*Redundancy*	Yes: redundancy in the different parts of the SoS (especially the safety critical ones) will be required to reduce the likelihood of complete systems failures (especially those due to hardware).
9		*Diversity: Design diversity*	Yes: diversity of suppliers may also lead to a lot of heterogeneity in the system (i.e. different parts of the system are constructed with different programming languages/development

			methodologies). Additionally more than one diverse instance of a component or sub-system may need to be developed to increase the probability of a safe/secure/reliable/available SoS (especially for safety-critical SoSs). However the introduction of diversity for these, already very complex, SoSs could have a detrimental effect on the manageability, hence possibly reducing the dependability of the SoS.
10		*Diversity: Data diversity*	Yes: SoSs are likely to be constructed to enable multiple modes of interaction with the system (or for the interaction of the different sub-systems between them).
11		*Diversity: Functional diversity*	*Yes:* the SOS could have functionally diverse components or sub-systems
12		*Wrappers*	*Note for the whole SoS:* probably wrapping will only be possible at the lower level of granularity (of components or sub-systems)
13		*Watchdogs*	*Yes:* same as for 12 above.
14	*Validation and verification techniques*	T*esting: Black Box (functional)Testing: Equivalence Classes* and *Input partitioning Testing*	*Not for the whole SoS:* testing the whole SoS may be very difficult if not impossible (hence some form of simulation may to be done). Therefore only sub-systems can be black-box tested.
15		T*esting: Black Box (functional)Testing: Boundary value analysis*	*Not for the whole SoS:* same as for 14 above.
16		T*esting: Black Box (functional)Testing: Error Guessing*	*Not for the whole SoS*: same as for 14 above.
17		T*esting: Black Box (functional)Testing: Stress testing*	*Not for the whole SoS:* same as for 14 above.
18		T*esting: Black Box (functional)Testing: Statistical testing* (see the *Reliability evaluation* section below)	*Not for the whole SoS*: same as for 14 above.

19		T*esting:* White Box (Structure-based) Testing: *Control Flow Analysis*	*Not for the whole SoS:* control flow analysis may have to be done at even lower levels of granularity than for Black box testing above. Hence the feasibility of this technique for the large SoSs may have to be questioned
20		T*esting:* White Box (Structure-based) Testing: *Data Flow Analysis*	*Not for the whole SoS*: same as for 19 above.
21		T*esting:* White Box (Structure-based) Testing: *Cause Consequence Diagrams*	*Yes:* in theory it is possible, but in practice may be infeasible due to the extremely high number of cause-consequence combinations and trees.
22		T*esting:* White Box (Structure-based) Testing: *Unit testing*	*Yes*.
23		T*esting:* White Box (Structure-based) Testing: *Integration testing*	*Yes*: this will be very important for SoSs as, by definition, an SoS is made up of many components or sub-systems which need to be integrated together in one large SoS.
24		*Static Analysis*	*Yes*: even though it is possible to do static analysis for the whole system, likely budgetary, time and complexity constraints may limit their use to a smaller number of safety-, mission-critical components within the SoS.
25		*Common-mistake Analysis*	*Yes.*
26		*Formal methods and semantic analysis*	*Yes*: similar to 24 above.
27		*Human factors analysis: the consideration placed on human factors in each phase of the lifecycle*	*Yes:* especially important for SoSs to have a socio-technical viewpoint of the system during its construction and assessment.
28		*Human factors analysis: availability and quality of operation and maintenance instructions*	*Yes*: of very high importance for SoSs. Because the SoSs are complex by definition the location if instructions and documents for each component and sub-system part need to be made available.
29		*Human factors analysis: the level of human intervention required by the system*	*Yes*: similar to 28 above.

30		*Human factors analysis: the complexity of operator actions required*	*Yes*: similar to 28 above.
31		*Human factors analysis: the potential for operator error to directly lead to a system failure*	*Yes*: similar to 28 above.
32		*Human factors analysis: the simplicity and intuitiveness of the user interface*	*Yes*: similar to 28 above.
33		*Human factors analysis: the stress placed on the operator*	*Yes*: similar to 28 above.
34		*Human factors analysis: the level of training available to the operator*	*Yes*: similar to 28 above.
35	*Risk assessment*	*Hazard and Operability Studies (HAZOPS)*	*Yes*.
36		*Failure Modes and Effects Analysis (FMEA)*	*Yes*: but due to the complexity of the system enumerating the complete failures of the system will be infeasible. Likely to be more useful at the component or sub-system level.
37		*Event Tree analysis*	*Yes:* similar to 36 above.
38		*Fault tree analysis and dependence diagrams*	*Yes:* similar to 36 above.
39	*Reliability* evaluation	*Reliability block diagrams*	*Yes*: likely to be useful for modelling the reliability of the hardware parts of the SoS. The 'blocks' could be constructed at several layers of abstraction, e.g. at a coarser level of granularity the components or subsystems of the SoS could form a block.
40		*Reliability growth modelling*	Yes: possible if the execution time of the SoS can be estimated well. But due to the complexity of the system and the many types of failures that can happen (correctness, security, safety etc.) strong and maybe unrealistic assumptions will need to be made in order to use a reliability growth model for an SoS
41		*Evaluation of field experience*	*Yes*: for a large SoS it is likely that a lot of dependability data will be collected (fault and failure reports, vulnerabilities, security breach

			attempts etc). The large amount of data is likely to overwhelm the administrators though and techniques are needed to utilise this data properly for dependability modelling and predictions.
42	*Dependability cases*	*Security cases*	*Unknown*: the development of this evaluation mechanism is still in its infancy and not clear yet how well it can be used in practice for large SoSs
43		*Safety cases*	*Yes*: likely to be mandatory for safety critical SoSs. Most likely to be useful at the component and sub-system level. Likely to get very complex at the SoS level unless simplifying assumptions are made.
44		*Diverse arguments*	*Unknown:* similar to 42 above.
45	*Responsibility modelling*	*Causal responsibility*	*Yes*: in large SoS it is very important to be clear about the responsibilities of humans in the SoS. For the Causal responsibility: important to make clear throughout the SoS who is responsible for carrying out a task or action.
46		*Consequential responsibility*	*Yes*: similar to 45 above. For consequential responsibility, important to make clear who has the responsibility of dealing with the consequences of a failure?

Appendix C - Task 2: State of the Art in the Assessment of SoS of COTS Components

Introduction

This section presents a review of frameworks and processes for assessment of COTS components, as well as evaluation, validation and verification techniques that can be used to assess the dependability and security of complex SoSs, including those built out of COTS components. It also presents research challenges in the assessment of SoSs made up of COTS components stemming from the limitations of the current state of the art. Hence this section addresses Task 2 of the proposal which states "*characterise the present state of the art in assessment of SoS of COTS based components and in particular explain and understand why existing methods do not either scale or otherwise translate to the SoSoS context*".

In the rest of this section we will first summarise the nature and characteristics of COTS components. We then provide a review of the assessment processes, frameworks and methods proposed in the literature for COTS components, including in safety critical domains such as the medical, nuclear and defence sectors. We also provide a summary of evaluation, validation and verification techniques for both systems and SoSs in general as well as for constituent components of SoSs (including COTS components). In Appendices A and B of this report we provide guidelines on how these techniques can be applied for the assessment of SoSs and their components by using as examples the SoSs depicted in the three scenarios that have been developed by a NATO research task group (of which CSR, City University is a contributing partner). And finally we summarise the research challenges of COTS assessment in large SoS context.

Nature of COTS components

Commercial Off-The-Shelf (COTS) software components (or simply Off-The-Shelf (OTS) software components) come in a variety of forms [32], such as:

- Components that form part of a program (e.g. various graphical, statistical or mathematical libraries of functions used to perform specific tasks; device drivers such as, for example, a JDBC (Java DataBase Connectivity) which provide implementations of Java interfaces for connecting to a database).
- Standalone programs and utilities (e.g. compilers)
- High level service programs (operating system kernels, database management systems programs, web servers, applications servers, office applications etc.)
- Complete systems of integrated software and hardware components (alarm systems, Programmable Logic Controllers (PLCs), medical devices, Enterprise Resource Planning (ERP) systems, air-traffic management systems etc.)

The main characteristics of COTS are:

- The components already exist and cannot be re-engineered (exceptions are some open-source OTS with less restrictive license agreements for which the code is available for changing)
- Due to their general-purpose use, most components may contain functions that are not necessary for a specific application.
- COTS with a substantial user-base are subject to continuous change and evolution to meet users' evolving requirements.

COTS-based development

Development of COTS-based systems (i.e. systems which will primarily be constructed from COTS components, “glue-code” and/or component “wrapping” (the wrapping may especially be required in defence and other safety–critical environments)) differs from the conventional system development. With COTS-based systems the emphasis shifts from conventional *specification*, *design* and *implementation* to COTS component *selection*, *configuration* (or tailoring) and *integration*. There will need to be an initial statement of requirements which clearly states what the system should do, and a *high level design* (especially if the COTS-based system requires the integration of several COTS components) to get an overall picture of the new system (UML Component Diagrams [104] could be one mechanism of showing the high level design). The COTS *selection* stage can then begin. During this stage conventional assessment techniques can be used and/or adapted to assess the most optimal COTS components to be used in a given system. The selection will need to be done from various functional as well as non-functional requirements and dependability attributes (reliability, performance, security etc). During the *configuration* stage the COTS component(s) are then tailored and configured to work optimally and consistently in the wider system in which they will be integrated (it is not uncommon for some complex COTS components such as Enterprise Resource Planning systems to take up to 18 months to configure). The final stage will be the system *integration*. The integration stage may require development of “glue-code”/middleware as well as component “wrapping”. This whole process will most likely need to be *iterative* (see Fig. 1 for details). For example the Configuration phase may tell us that some of the COTS components cannot be configured to run consistently with other components in the wider system hence the requirements may have to be revised and the selection process needs to be redone using the for the revised requirements. A similar process for the last three stages of development (selection, configuration and integration) is proposed in [105], which also gives finer level process elements for each of these three stages.

In this report we will be mainly concerned with COTS assessment issues. However in Appendix D we present a table which offers an outline of the various assessment and architectural techniques for these three main COTS-based system development phases.

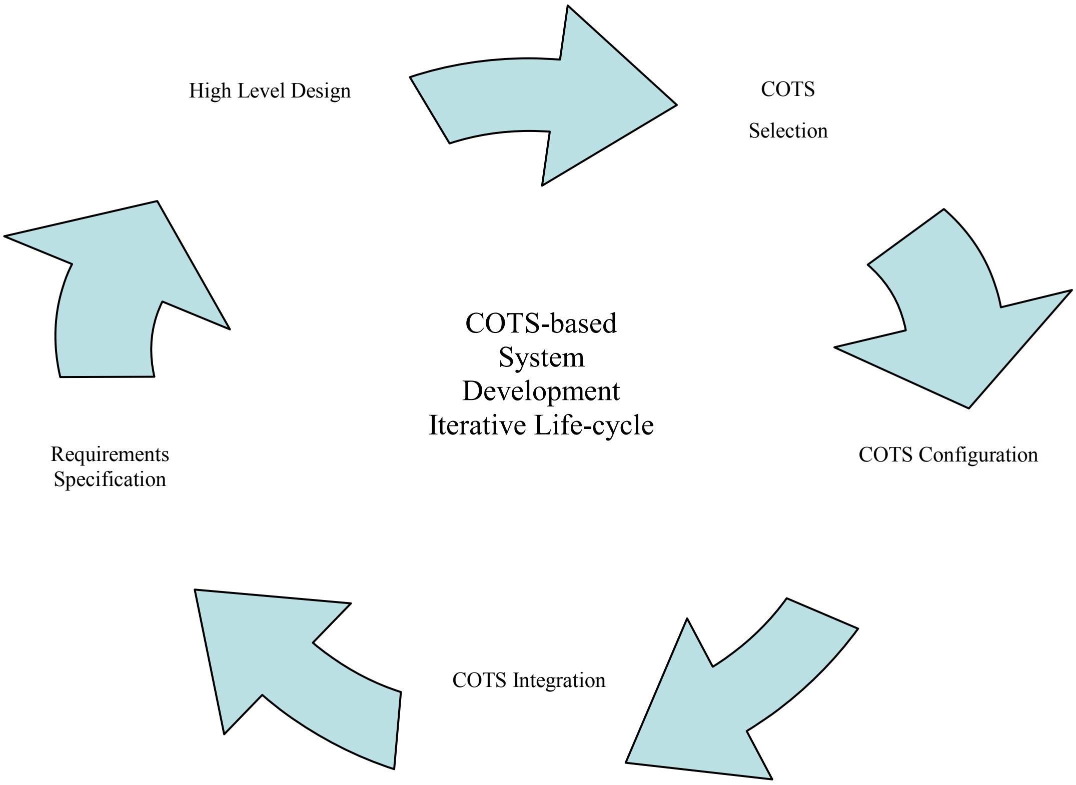
High Level Design
COTS Selection
COTS-based System Development Iterative Life-cycle
Requirements Specification
COTS Configuration
COTS Integration

Figure 3 – COTS-based development iterative system life cycle.

Problems with COTS component assessment

The main characteristics of COTS components highlighted in the previous sections lead to novel assessment problems compared with assessment of conventional bespoke systems, or similar assessment problems which may be more highlighted when assessing COTS components. In this section we will discuss some of these problems.

Non-compliance to standards and best practice documents

For safety- or business-critical applications, in particular, purpose-built products traditionally would come with extensive documentation, evidence of good development practice, compliance to standards and of extensive verification and validation. With mass-distributed COTS products on the other hand, users (system designers or end users) invariably find not only a lack of this documentation, but anecdotal evidence of serious failures and/or bugs that undermines trust in the product. As mentioned before, COTS components are usually developed for general-purpose use hence they may not comply with standards of a specific application.

Establishing COTS provenance

Since evidence about how the COTS component was developed may not be available, it is difficult for assessors to establish what the COTS component *contains* (e.g. whether any accidental / or malicious backdoor exists in the system). Additionally, evidence from testing prior to the release of the COTS component is rarely made available by the vendors. Hence, without extensive black-box testing it is difficult for assessors to assess key functionality, reliability and performance requirements that may be dictated by the target application, system or SoS, e.g.:

- whether the COTS component fulfils its functional requirements;
- whether the COTS component performs its functions within a sufficient level of reliability required by the target application (especially important when the COTS component is to be integrated in larger system or SoS)
- whether the COTS component responds within an acceptable time-delay as required by the target application.

Problems from unwanted/unneeded functionality

As mentioned before, COTS components often come with functionality that is unneeded for a given application. In some cases this unneeded functionality may also be dangerous in the new system or SoS, e.g. allowing unencrypted remote connections to the COTS component by default. In most cases the COTS component can be configured to tailor the need of the target application, but some unwanted behaviour may not have been documented well enough (or at all) by the vendor hence leaving the target application vulnerable.

Problems stemming from patches and updates

With bespoke software, updates and patches come from one source and hence their effects on the system are easier to control and predict. In large systems or SoSs built with large number of COTS components from many different vendors, patches and updates may cause serious system stability problems. The updates and patches for one COTS component may be inconsistent with those of other COTS components in the system or the application as a whole and hence lead to new failures. Since COTS components may themselves be composed of other COTS components, some of these changes and updates may not be self-evident (i.e. the higher level COTS may contain the same release version number even though it is made up of updated/patched COTS components).

On the other hand, simply refusing to allow patches to COTS components is dangerous as the patches and updates may contain important security problem fixes, hence leaving the COTS component unfixed may leave the whole system vulnerable to attacks. Patching and update issues are also discussed in [33] (a recommendation given in [33] is to separate the security patches from functionality patches and updates: the first kind should be mandatory: the vulnerabilities posse a security risk to the application and others (if the application is online); whereas the second kind should be optional and it should be left to the users to decided whether applying the update will lead to system stability problems and hence may not be worth updating, at least in the short term).

Difficulties with optimal COTS component selection

For any given family of COTS components there may be a plethora of available solutions available from various vendors. For example, there are a multitude of available database management system (DBMS) products available in the market (both commercial closed-development and free open-source). For any given application, it is important to choose the most suitable component(s). The suitability of the component needs to be assessed from several different dependability attributes (safety, security, reliability, performance etc.). Since any assessment is conducted with limited resources and under various assumptions, which may not hold true in real operation, the outcome of the assessment is subject to uncertainty – the assessor may never be 100% sure that what they concluded during the assessment (both about the values of the attributes as well as the choice of a COTS component) will be confirmed when the COTS component is used in operation. Most of the existing approaches for COTS selection do not explicitly deal with uncertainty in the assessment, which may lead to sub-optimal COTS products being chosen.

Dependence Among Attributes

COTS component assessment requires dealing with multiple attributes of the COTS components being compared. For example reliability, performance, security, safety etc., as mentioned previously. The selection of a particular COTS component, thus, is a multi-criteria decision which taken under uncertain values of the attributes naturally leads to the question about the dependence between the uncertainties associated with the individual attributes. Ignoring the possible dependence between the attributes represents a particular form of belief: that assessing attribute X one can learn *nothing* about another attribute, Y. For example, performance of a COTS component will hardly tell anything about the quality of its documentation and vice versa. It is quite obvious, however, that not all COTS component attributes are like that. In many cases while assessing an attribute X the assessor may infer something about another set of attributes. For instance if we devise a prototype in order to assess the functionality of a COTS component in the process we will learn something about the performance (how quickly this COTS component responds to requests) and how reliable the COTS component is. A more subtle, but very useful concept, as we will see later, is that the uncertainties associated with the assessed attributes may be *dependent*. Informally, assume that we want to assess the reliability and performance of a COTS component. We may assume that the uncertainties associated with these two attributes are independent, in the statistical sense. Under this assumption learning something about reliability will tell us nothing about performance and vice versa. Now suppose that we have run a very long testing campaign and have repeatedly observed that whenever the response was late it was also incorrect and no other incorrect response has been observed. With such evidence of a strong positive correlation between the failures (incorrect responses) and the responses being late, we may accept that any change of our belief about the rate of failure should also be translated into a change in our belief about the rate of late responses. Current approaches for COTS assessment and selection invariably assume that the attributes are independent and do not allow for dependencies between their uncertain values to be captured adequately.

Current approaches to assessing COTS components

There exists a plethora of approaches for assessment of COTS components. The approaches differ amongst themselves based on which dependability attributes they consider (safety, reliability, security, performance etc.) as well as for which domains they were developed (defence, medical, nuclear, non-safety critical commercial, etc.).

In the next few sections we will provide: a brief survey of COTS component *selection* approaches; a summary of standards for COTS assessment in three safety-critical domains, namely health-care, nuclear power sector and the defence domain; and finally we summarise a list of *evaluation*, *validation* and *verification* techniques available for assessment of COTS components.

Overview of COTS selection approaches

Summary of main approaches

There is a wide variety of COTS component assessment approaches available. All of them start with an initial statement of requirements, which defines what is being sought. It has been proposed that the requirements initially should not be too stringent, since this would discard potentially appropriate COTS component candidates at a very early stage [106], [107]. It has even been suggested [107] that if the requirements are not flexible then the COTS-based development may not be appropriate and bespoke development could be more cost-effective. So initially [107] suggests distinguishing between essential requirements and those that are negotiable. The selection criteria are then based on the essential requirements.

Off-the-shelf-option (OTSO) [108] is a multi-phase approach to COTS component selection. The phases are: the search phase, the screening and evaluation phase and the analysis phase.

In the first phase COTS components are identified. In the screening and evaluation phase the components are further filtered using a set of evaluation criteria (established from a number of sources, including the requirements specification, the high level design specification etc.). In the analysis phase results of the evaluation are analyzed, which lead to the final selection of COTS components for inclusion in the system. Other similar multiphase process approaches for COTS component evaluation that have been proposed include CEP (Comparative Evaluation Process Activities) [109], CBA Process Decision Framework [110] which in addition to defining a process for COTS component assessment also defines two other processes: COTS integration ("gluing") and COTS configuration ("tailoring"); CAP-COTS Acquisition Process method [111] and PECA Process [112].

Procurement-oriented requirements engineering (PORE) [113] is a process in which requirements are defined in parallel with COTS component evaluation and selection. [113] propose using prototypes to develop knowledge concerning COTS components and their use within the wider system. Other methods that are centred on the requirements to assist with the COTS component selection process are CRE-COTS-Based Requirements Engineering Method [114], Storyboard Process [115], Combined Selection of COTS Components [116] and COTS-DSS [117].

CISD (COTS-based Integrated System Development) [118] and CDSEM (Checklist Driven Software Evaluation Methodology) [119] are both checklist-based evaluation methodologies. STACE (Socio Technical Approach to COTS Evaluation) [120] is a socio-technical approach to evaluation which builds on work of [113]and [108] and emphasizes the organizational issues related to COTS selection.

Most of the approaches above that we have surveyed are concerned with "process", i.e. which phases should be followed through the COTS assessment so that a single COTS component can be chosen for a given application. When it comes to measurements, assessment and testing of the COTS components, most of these approaches use check-list based approaches to assess the COTS components and scores in the Likert scale [121] are given for each attribute under assessment. The results of each attribute rae then usually combined with *multi-attribute utility theory* or *analytical hierarchical process* approaches so that the competing COTS components are ranked and the one with the highest score is chosen.

Two common deficiencies of all of these approaches is their failure to handle explicitly the *uncertainty* that is inherent during the assessment process and their failure to adequately capture and deal with the *dependence* between the values of the functional and non-functional attributes.

Uncertainty conscious approach to COTS component assessment and selection

At CSR, City University we have recently developed a method for assessment of COTS components when *reliability* (measured as *probability of failure on demand*) and *performance* (measured as *probability of late response on demand*, given a predefined *timeout* value) attributes of a COTS component are of main concern. Full details of the method and the underlying assessment model are given in [122]. The method is based on an extension and adaptation of a previous assessment model developed at CSR, City University [123]. Our proposed method tried to improve on existing methods of COTS component assessment which, as mentioned previously, invariably assumed that values of the dependability attributes of COTS components are known with certainty and that the uncertainties in the values between the different attributes are independent of each other. Our approach attempts to provide the assessors with the capability of expressing (using probability distribution) their doubt (uncertainty) in the values of reliability and performance attributes of a COTS component and the dependencies that may exist between the values of these two attributes. We illustrated how the assessment can be done with off-the-shelf database management system products [122]. In the same paper we also provide a discussion of how the method can be extended to assessments where more dependability attributes need to be considered, and the difficulties associated with assessments where more than two attributes need to be considered.

Domain specific assessment solutions

A report [32] prepared by Adelard for the UK Health and Safety Executive gives a very good overview of methods and techniques for assessing *safety* integrity of COTS components (or "Software Of Uncertain Pedigree" (SOUP) which is the term used for COTS components in the aforementioned report). The same report also reviews the recommendations for use and assessment of safety properties of off-the-shelf components in the medical, nuclear and defence domains.

COTS component use in the medical domain

The US Food and Drug Administration (FDA) of the Department of Health and Human Services have issued a guidance document for use of COTS components in medical devices [124]. The guidance document is risk based and hazard analysis is recommended to reduce the risk of safety-critical failures. The main purpose of the guide is to make recommendations on the needed documentation for all COTS software used in medical devices. It also provides recommendations of additional (special) needs and responsibilities of the COTS software vendor when the severities of the hazards from COTS software failure become more significant.

COTS component use in the nuclear power-plant domain

The first supplement of standard IEC 60880 of the International Electrotechnical Commission (IEC)[125] provides requirements for software for computer-based safety systems in nuclear power plants. It also contains a section on COTS software (the standard uses the term "pre-developed software" (PDS)). The evaluation process for COTS components recommended by the standard is as follows:

- An evaluation of the functional and performance features of COTS components and existing qualification documentation. In this stage the COTS component is treated as a "black box".
- A quality evaluation of the software design and development process of COTS components. In this stage a "white box" approach is taken.
- An evaluation of operating experience. This is done if there were weaknesses in the demonstration of the COTS quality in the previous two stages. At this stage the evidence required is:
  - The method for collection of data.
  - The operational history of defects, error report and other findings.
  - The operational history of modifications (patches/upgrades) made as a reponse to defects or for other reasons.
- A comprehensive documented assessment of the evidence from all steps of evaluation.

COTS component use in the defence domain

The UK Defence Standard (DS) 00-55 [126] addresses safety critical (SIL 4) software. This standard also contains a section (clause) on COTS software (the term used in the standard is "previously developed software" (PDS)). DS 055 is used in the context of DS 00-56 which addresses safety management.

DS 00-55 is targeted at software of very rigid safety-critical requirements. It requires that:

- All COTS software should be identified, and justified in the software safety case. The justification should include a safety analysis of the COTS software.

- COTS software to be used in the final delivered equipment should conform to the requirements of the standard for new software.
- Or, reverse engineering and validation and verification activities should be carried out on any COTS software that has not been produced to the requirements of the standard. Reverse engineering in this case should cover all stages of development including specification, design, verification and validation, and hence access to the source code, design and test documentation of the COTS software is required.
- All changes to COTS software made as part of its incorporation in the safety-related software should be to the requirements of the standard.
- "*Unreachable code should only be allowed to remain in the final application where it can be shown that the risks of leaving it in are less than the risks of modifying the code to remove it*"

Summary of evaluation, validation and verification techniques for COTS components

This sub-section outlines the various evaluation, validation and verification techniques that may be used to assess a complex ICT system, including those built out of COTS components. The main documents that have been consulted in deriving the list below are [127], [128] and [129] and the references therein. In Appendix B we provide details on how these techniques can be used for assessment of SoSs depicted in Appendix A. The techniques have been categorised depending on their objectives and purpose of use in assessment:

- *System development Life-cycle methods*: these methods are aimed at assessing how a given system or COTS product was developed, and they can be split into:
  - *Compliance with processes defined in standards*: quality (such as ISO 9001:2000), safety product development (IEC 61508)
  - *Audit of specific aspects of a lifecycle* (e.g. the requirements process, testing and test coverage etc.)
  - *Use of iterative development processes*
  - *Competence management* (e.g. assessment of the suitable training for personnel developing the product or system)
  - *Review of the requirements process* (since they are often identified as the most common source of problems in software development)
- *Review of Quality of supply*: similar to the review of the system development life-cycle methods above. It is performed to gain confidence in the development of the system but may be considered separately "*so as to emphasise its relationship with general organisational procedures and culture that affect any development made by the same supplier and supply chain, while lifecycle review focuses on the development of a particular product*" [128]. Techniques include:
  - *Supplier competency* - can be assessed directly or through evidence of a competency management system used by a supplier.
  - *Supplier process assessment – "assessment of the processes used by the supplier and the supply chain is based on an evaluation of the quality management systems and safety management systems in use*" [128].
- *Architecture:* the architecture of a system plays a significant role in its integrity and availability. Architectural controls exist which can improve the integrity and availability by monitoring, duplicating or diversifying equipment:

- *Redundancy* – using more than one redundant component or subsystem. Useful for protection against most hardware faults (and transient software faults) but not against design faults (hardware or software).
- *Diversity* – using more than one redundant, but diverse, component or subsystem (such as diverse Firewalls, diverse IDSs). Depending on how effective diversity is, this technique may be the most useful protection against design faults (however it does bring other architectural difficulties from creating middleware to make the diverse components work consistently). Various forms of diversity are possible:
    - *Design diversity* – more than one diverse component are used
    - *Data diversity* – the requests sent to the system may be syntactically diverse but logically equivalent (observed to be a useful mechanism with SQL database servers [78])
    - *Functional diversity* – the system is designed from functionally diverse components (for example a monitoring system made up of a pressure monitor and a temperature monitor)
- *Wrappers* - a wrapper will do additional plausibility checks on the input or output of the systems and protect the system from unreasonable or malicious demands.
- *Watchdogs* - similar to *wrappers*; the watchdog will constantly monitor the system and may take predefined corrective actions such as raising additional alarms or restarting the system (system 'rejuvenation' which may protect the system from certain (malicious or non-malicious) software ageing faults)

➢ *Validation and verification techniques:* "*verification is the assessment of code or a subsystem against its requirements specification; validation is the assessment of the final product against the user demands. V&V incorporates final product testing"* [128]. The main V&V techniques commonly used in assessment are:

- *Testing* - Various forms of testing techniques exist and have been reported in the literature (a review of techniques is given in [127]):
    - *Black Box* (Functional) testing - testing performed without visibility of the implementation of the component. Main techniques of black box testing are:
        - *Equivalence Classes and Input Partitioning Testing* – input space during testing is divided into a set of equivalence classes (input values which are treated the same way by the software); testing is done over as many different equivalence classes as possible to increase the test coverage.
        - *Boundary value analysis* – input space during testing consists of values that lie at the edge of an equivalence partition;
        - *Error Guessing* – testing for common errors (e.g. division by zero, an empty file, record, or field, negative numbers, alphabetic character for numeric field etc.)
        - *Stress testing* – testing under, for example, a much higher demand rate, a greater temperature or pressure etc., so that the COTS component is tested outside the limits of its normal operating range.
        - Statistical testing (see the Reliability evaluation section below)
    - *White Box* (Structure-based) Testing - testing performed with full visibility of the implementation of the component. Main techniques of black box testing are:
        - *Control Flow Analysis* – testing is performed to analyse the internal flow of control in a program or a component.

        - *Data Flow Analysis* - testing is performed to analyse the internal flow of data in a program or a component.
        - *Cause Consequence Diagrams* – diagrams are drawn to organise in a systematic way the combinations of inputs (causes) and outcomes (effects) into test cases.
        - *Unit testing* – testing is performed with individual software modules against their individual functional specifications.
        - *Integration testing* - testing performed when modules are integrated. At a higher level of granularity integration testing in COTS-based development would be performed once the COTS components have been integrated into the larger system.
        - *Static Analysis* - analysis of software code that is performed without actually executing the code. The analysis usually performed by an automated tool.
    - *Common-mistake Analysis* – similar to *error guessing*. For example, buffer overflows are a common problem with programs written in the C language, hence stress test C programs for buffer overflow problems. Depending on language and development methodology used a database of common-mistakes about these systems can be compiled.
    - *Formal methods and semantic analysis* – formal approaches to program correctness provide means for modelling the specification, design, or implementation of a system in a strict, precise mathematical formalism.
    - *Human factors analysis* – this involves considering the following:
        - *the consideration placed on human factors in each phase of the lifecycle*
        - *availability and quality of operation and maintenance instructions*
        - *the level of human intervention required by the system*
        - *the complexity of operator actions required*
        - *the potential for operator error to directly lead to a system failure*
        - *the simplicity and intuitiveness of the user interface*
        - *the stress placed on the operator*
        - *the level of training available to the operator*
- *Risk assessment*: there exists a range of techniques for assessing and quantifying the risk of failure both from safety as well as reliability or security point of view (even though the techniques below have been traditionally used in the safety and reliability domains):
    - *Hazard and Operability Studies (HAZOPS)* – a structured discussion involving the various stakeholders of the system which is driven by the models of the system and its components.
    - *Failure Modes and Effects Analysis (FMEA)* – a technique for assessing and prioritising the possible failures in a system.
    - *Event Tree analysis* – a diagrammatic representation of chains of events that can occur in the system. The root of the tree is a possible initiating event
    - *Fault tree analysis and dependence diagrams* – a similar modelling principle to Event tree analysis but constructed in the reverse direction: the root of the tree is a potential consequence.

- *Reliability* evaluation: The reliability of the system or product (both during its development and use) can be evaluated using various Reliability modelling techniques:
  - *Reliability block diagrams* – modelling, in a diagrammatic form, the chain of events that are necessary for the successful operation of a system.
  - *Reliability growth modelling* – modelling the reliability of the system dynamically during the system development or use where faults are being fixed/removed from the system, hence the trend is usually that reliability will *grow* (but not always as a "fix" may introduce a new potentially more harmful bug in the system leading to reliability decay).
  - *Statistical testing* - which involves the creation of a test harness to perform the testing; an 'oracle' against which the results obtained from the system under test are compared; and an accurate definition of the 'operation profile', i.e. the profile under which the system is expected to be used.
  - *Evaluation of field experience* – using field data (tracking and recording faults and incidents in a product or system) to gain insights into the reliability of the product. Difficulties with getting accurate measures stem from the difficulty in measuring the operational profile of the products in their given installation (however, some companies may offer incentives to users (such as cheaper products) if they are willing to allow the vending company to collect detailed data about the product usage, which can improve the quality of the data collected).

  The handbook of Software Reliability Engineering [67] gives a comprehensive guide. *Reliability Growth modelling* is the most useful of the techniques listed above when arguing about software reliability as it allows an organisation to monitor and argue about the reliability of the system as it is being developed (or, if it is in operation, as it is used). *Statistical testing* is also a very useful technique to estimate the reliability before a system or component is deployed in operation (but the operational profile is crucial to the accuracy of the results obtained from statistical testing). A plethora of Reliability models exists (some better than others) and there are techniques (developed at CSR, City University [68]) which can help with *recalibrating* the results of the reliability growth models.

- *Security cases*: this is an emergent technique based on the more widely used concept of a "safety case". The elements of a Security case are:
  - *Claims* about a property of the system or some sub-system.
  - *Evidence* (which can be facts (based on scientific principles and / or prior research), assumptions, or sub-claims, derived from lower level sub-arguments) used as the basis of the trust argument.
  - *Argument* which links the *evidence* to the *claims*.
- *Responsibility modelling*: *"To usefully reason about responsibilities in a complex socio-technical system, we must have some way of modelling the responsibility itself (as distinct from modelling the assignment of responsibilities). If we have some model of the responsibility, we can make better decisions about the appropriateness of responsibility assignment and so reduce vulnerabilities and consequent failures.*" The above paragraph is from a chapter in the book "Responsibility and Dependable Systems" [130]. The book is an output of the DIRC project [76], of which CSR, City University was a partner. The responsibility modelling makes a clear distinction between:
  - *Causal responsibility* - the obligation to ensure that some state of affairs comes about or is/is not maintained

- *Consequential responsibility* - the obligation to take the blame if some state of affairs does not come about or is/is not maintained.

Research challenges

In sections 3 and 4 of this report we reviewed the research challenges for assessment of future systems including complex SoSs in general. The research challenges for assessment of complex SoSs made up of COTS components are similar. The main issues facing the assessors of complex SoSs of COTS components are predicted to be:

- *Dealing with uncertainty and statistical independence in the assessment process*: dealing with inherent uncertainty in the assessment process and the dependence between values of different components of the system, as well as dependence between the values of various dependability attributes of even the same component, will become even more important in complex SoSs of the future, which may be constructed from a multitude of COTS components (both software and hardware). The assessment method we outlined in this Appendix C can only deal with a limited number of dependability attributes in the assessment process unless strong statistical independence assumptions are made. Hence further research is needed to investigate how the assessment methods that deal with uncertainty and dependence in the values of the COTS attributes, can scale in the SoS context where the number of dependability attributes that need to be considered may be higher.
- *Inter-dependence*: we already discussed the interdependency issues of SoSs in section 3.2.4. For SoSs constructed from COTS components interdependence between the constituent COTS components becomes very important both during system construction (to ensure overall system stability) as well as during assessment (to demonstrate that the various COTS components work reliably, securely, safely etc., together). Hence further research is needed in modelling and assessing the effects that these inter-dependencies between COTS components will have on the SoSs.
- *Patching and upgrading*: we already discussed the problems that can stem from patching and upgrades in this Appendix. Further research is needed on establishing the optimal patch and upgrade frequencies and methods. Clearly refusing or postponing patches of COTS components may help with maintaining overall system or SoS stability, but may have a detrimental effect on security if vulnerabilities are left un-patched.
- *Obtaining good metrics*: we already discussed the importance of good metrics for assessment of systems and SoSs in section 4.3.2. This issue remains very important for SoSs composed of COTS components. Data in the form of *bug* or *fault reports* as well as *vulnerability reports* are often available, but with absence of detailed failure data (i.e. counts of occurrence of these failures caused by these faults in operation) it is difficult to use the data in *reliability growth modelling* for instance (in the case of accidental faults) or security assessment (for deliberate faults).
- *Emergent properties*: as we discussed in section 3.2.3 unforeseen and /or unplanned emergent properties may arise as system complexity increases. When COTS components are used in SoSs (especially for those COTS components that are closed-development and hence the code is not publically available) the problems from emergent properties may be more pronounced due to difficulties with establishing COTS provenance and dependencies that may exist between the COTS components.
- *Online assessment*: we already discussed online assessment in section 4.3.2. It is clearly very much applicable for COTS components (for example, use the measurements during online operational assessment to decide when to upgrade a COTS component while in operation as part of a larger SoS). However further research and

experimentation is needed to learn more about the performance implications of online assessment and the difficulties with building the decision framework which will be required with any online assessment.

Appendix D

This appendix lists the assessment techniques and methods that are possible for the three stages of COTS-based development which were highlighted in Appendix C of this report.

Assessment techniques for COTS-based development phases – selection, configuration and Integration

The following table contains listings of assessment techniques and methods that can be used to assess COTS components during the three main phases of COTS-based systems development, namely *Selection*, *Configuration* and *Integration*. We have listed the techniques per dependability attribute and functional requirements.

Functional and Dependability attribute	Selection	Configuration	Integration
Functional properties	➢ Checklist approach against the requirements specification. ➢ Prototyping and testing. ➢ Review of product documentation to check whether required functionality is supported in the component.	➢ Review of documentation to ensure product is sufficiently tailorable and configurable for the new context. ➢ Prototyping and testing.	➢ Review of documentation to ensure interfaces are well documented to allow integration in the wider system. ➢ Prototyping and testing. ➢ Are architectural solutions listed in Appendix C possible/suitable for the component?
Interoperability	➢ Review of product documentation to check whether required interfaces for interoperating with other components in the wider systems are supported in the component. ➢ Prototyping and testing of the product interfaces for interoperability. ➢ [131] gives a list of 38 characteristics that "*define COTS product interoperability characteristics*" [131]. ➢ Review of	➢ Prototyping and testing to check the COTS component is sufficiently configurable to interoperate with other COTS components in the system (the attributes from [131] can be used to drive the testing)	➢ Interface testing to ensure interoperability with other components during system integration (again the characteristics of this interoperability can be derived from [131] and will also be driven by the specific requirements for a given system).

	interoperability field data available for the component.		
Reliability	➢ Reliability evaluation methods described in Appendix C. ➢ The Uncertainty conscious approach to assessment described in Appendix C ➢ Validation and verification techniques described in Appendix C (especially black box testing if the code of the COTS component is not available) ➢ Risk assessment techniques described in Appendix C ➢ Dependability cases. ➢ Review of reliability field data available for the component.	➢ Validation and verification techniques described in Appendix C can be used to verify the reliable configurability of the COTS component (e.g. are there any side effects that stem from a given COTS configuration). ➢ Risk assessment techniques described inAppendix C	➢ Reliability evaluation methods described in Appendix C can be done at the system level once the COTS component has been integrated in the wider system. ➢ Validation and verification techniques described in Appendix C can also be done at the system level. ➢ Risk assessment techniques described in Appendix C performed at the system level
Timing properties (Performance)	➢ The Uncertainty conscious approach to assessment described in Appendix C ➢ Experimentation with performance benchmarks (e.g. TPC-C for COTS database servers) to obtain measures of COTS response time and/or throughput rates. ➢ Review of performance field data available for the component.	➢ Experimentation and testing to obtain measures of time it takes for a COTS component to be configured. ➢ Experimentation and testing to obtain measures of COTS response time and/or throughput rates once configuration of the COTS component is changed.	➢ The Uncertainty conscious approach to assessment described in Appendix C performed at the system level. ➢ Experimentation with performance benchmarks to obtain measures of system response time and/or throughput rates following integration of the COTS into the wider system.

Availability	➢ Reliability evaluation methods described in Appendix C with emphasises now being placed on the proportion of *downtime* of the component during the testing period rather than probability of failure on demand.	➢ Reliability evaluation methods described in Appendix C with emphasises now being placed on the proportion of *downtime* the component will be in during the configuration process. ➢ Same method as above but with emphasis on the proportion of downtime the system will be in following configuration or tailoring.	➢ Reliability evaluation methods described in Appendix C with emphasises now being placed on the proportion of downtime of the component during the testing period rather than probability of failure on demand – but done at the system level following the COTS component integration into the wider system.
Provenance	➢ System development Life-cycle methods described in Appendix C ➢ Review of Quality of Supply methods described in Appendix C ➢ Testing and prototyping.	N/A	N/A
Resource usage	➢ Experimentation and stress-testing techniques to measure the response time, CPU-usage, memory-usage, file input /output, deadlocks, priority process allocation etc of a component.	➢ Following configuration, further experimentation and stress-testing techniques to measure resource usage of a component.	➢ Experimentation and stress-testing techniques to measure the response time, CPU-usage, memory-usage, file input /output, deadlocks, priority process allocation etc of the system following the integration of a component.
Robustness	➢ Experimentation and stress-testing techniques to test the component robustness. ➢ Review of the components documentation to check the level of	➢ Experimentation and stress-testing techniques to test the component robustness following configuration and tailoring. ➢	➢ Experimentation and stress-testing techniques to test the system robustness following the integration of the component in the wider system.

	architectural redundancy and diversity in the component which would allow it to tolerate certain kinds of failures. ➢ Review of reliability field data available for the component to check for the robustness of the component.		
Maintainability	➢ Experimentation and measurement of mean-time-to-repair of a component following failure. ➢ Review of the components documentation to check the level of architectural redundancy and diversity in the component which would allow it to tolerate certain kinds of failures and hence reduce the cost of maintenance for the component. ➢ Review of reliability field data available for the component to check for the mean-time-to-repair times for the component in operational use.	➢ Experimentation and measurement of mean-time-to-repair of a component following failure, once the component has been configured.	➢ Experimentation and measurement of mean-time-to-repair of the component once it has been integrated in the wider system.
Usability	➢ Task analysis ➢ Human factors analysis techniques listed in Appendix C ➢ Usability in safety-critical settings (such as air-traffic control) may also be measured in terms of the response-time and	➢ Task analysis – how easy the systems is to configure	➢ Usability assessment of the whole systems where the COTS components is integrated using the approach summarised in Appendix C ➢ Human factor analysis for the whole system following COTS

	correctness of the task performed hence the approach summarised in Appendix C of this report can be used for usability assessment.		integration. ➢ Responsibility modelling in organisation where the integrated system will be used.
Safety	➢ Safety cases ➢ Review of quality of supply as described in Appendix C ➢ Review of system development life-cycle methods for the component as described in Appendix C. ➢ Risk assessment techniques described in Appendix C.	➢ Safety cases need to be developed for safe configuration of components. ➢ Risk assessment performed on the component following configuration.	➢ Safety case for the whole system where the COTS component is integrated. ➢ Risk assessment for the whole system where the COTS component is integrated.
Security	➢ Security cases. ➢ Evaluation methods such as Common Criteria. ➢ “Red team” penetration testing. ➢ Review of quality of supply. ➢ Validation and verification listed in Appendix C in the context of malicious failures.	➢ “Red team” penetration testing following configurations. ➢ Validation and verification listed in Appendix C in the context of malicious failures, following a COTS component configuration.	➢ Security cases at the system level. ➢ Evaluation methods such as Common Criteria applied at the system level where the COTS component is integrated. ➢ “Red team” penetration testing applied at the system level. ➢ Validation and verification listed in Appendix C in the context of malicious failures applied at the system level.