Four Testing Types Core to Informed ICT Governance for Cyber-Resilient Systems

Keith F. Joiner
School of Engineering and IT
University of New South Wales
Canberra, Australia
k.joiner@adfa.edu.au

Amit Ghildyal
School of Business
University of New South Wales & Department of Defence
Canberra, Australia
Amit.Ghildyal@defence.gov.au

Narelle Devine
Department of Human Services
Australian Government
Canberra Australia
narelle.devine@humanservices.gov.au

Alan Laing
Chief Information Officer Group
Department of Defence
Canberra, Australia
alan.laing@defence.gov.au

Anne Coull
Information Security
Westpac Banking Group
Sydney, Australia
anne.coull@student.unsw.edu.au

Elena Sitnikova
Australian Centre for Cybersecurity
University of New South Wales
Canberra, Australia
e.sitnikova@adfa.edu.au

Abstract—**Research on ICT projects continues to report very high cost and schedule overruns, as well as many high-profile ICT projects experiencing high incidences of unexpected cyber-vulnerabilities. Consequently, there is renewed interest in ICT governance from diverse areas. Some of the proposed governance models considered have great complexity while others appeal to simplicity for success. Three diverse and practical research efforts in ICT governance in Australian Government, as well as observations in the Banking Sector, came to similar concerns about the importance and type of ICT testing and expertise critical for ICT project governance to build cyber-resilience. Today's ICT Governance critically depends on: (1) information coming from all four types of testing, (2) the management of the testing as a coherent whole, and (3) that such test capabilities must endure through the whole life-cycle, so as to provide a sufficient degree of commercial and architectural independence to enable hard and timely decisions. Further, cyber-resilience challenges ICT testing to cope with increasing system configurations, threat permutations, future upgrades and threat sequencing. Therefore, this research uniquely calls for all ICT test types to use new combinatorial test design techniques for efficient screening and cyber-threat rigor. These lessons were shared at a special conference panel on ICT governance for resilient systems [1]-[4], where for the first time authors called for ICT governance frameworks to directly include test-informed previews in all decisions so that ICT can be more innovative, competitive, and cyber-resilient. This paper outlines the four testing types and lists the test infrastructure and combinatorial test design skills necessary for each.**

Keywords- ICT governance; usability testing; cyber-resilience; penetration testing; integration testing; project success factors; stress testing.

I. INTRODUCTION

Difficulties with ICT projects abound in all parts of the World [1]. There are also reports of many high-profile ICT projects experiencing high incidences of unexpected defects and cyber-vulnerabilities despite apparently extensive testing [6]-[8]. Research by [9] into 1,471 IT projects showed that cost overrun averages were not unremarkable to other projects (27%) but that there was, what they describe as, a *fat tail* of risk. They summarize that '*Fully one in six of the projects in the sample was a Black Swan, with a cost overrun of 200%, on average, and a schedule overrun of almost 70%.*' Reference [10] cites a U.S. Department of Defense (DoD) finding that *'85 percent of software intensive projects finished over time or budget; half of projects doubled original cost estimates; projects slipped an average of 36 months; and one-third of projects were cancelled.*' He goes on to cite military standards that '*inadequate software reliability can double or triple field support and maintenance costs*,' meaning that even those software-intensive projects that eventually succeed can remain a sustainment burden through-life. These sobering findings are alongside ever-increasing software functionality in systems, systems interconnectivity and autonomy [11]-[12], as well as increasingly sophisticated and cost-effective cyber-threats [13]-[14]. Reference [8] proposes governance involving continuous system monitoring through-life and his assessment is one of a field continuously exploring the bounds of achievement:

'... there will be notable failures, some great successes, and a large number of projects that get delivered in a sub-optimal state. That represents the norm for large software projects ... it is critical to understand that SoS [systems of

systems] *generate emergent behavior that can't always be reliably triggered by normal test inputs.*'

Governing the complexity of the software systems and their functions is significantly affected by the increasing number and sophistication of cyber threats to both open and closed system architectures [7] [14]. Cybersecurity is increasingly moving from avoiding cyber-attack in the form of barriers, to being able to sustain and recover from cyber-attack, or 'fight through' [7]. Cyber-resilience has many definitions, such as *'the ability* [for] *cyber systems and cyber-dependent missions to anticipate, continue to operate correctly in the face of, recover from, and evolve to better adapt to advanced cyber threats*' [15], or more simply as *'the capacity of an enterprise to maintain its core purpose and integrity in the face of cyber attacks*' [16]. Current ICT governance is seriously challenged by this shift to be more adaptable, omnipresent and evolving in the design of systems, support to operations and test infrastructures.

Three separate and diverse Australian Government-based research efforts in ICT governance, as well as an assessment in the Banking Sector, were found to have similar concerns about the importance and type of ICT testing and test expertise critical to ICT governance and the ability to build cyber-resilience: namely, systems integration, usability testing, stress testing and security testing (vulnerability, penetration & high assurance testing - VPAT). Each of these research efforts will be briefly critiqued before covering the four test types. Finally, the paper draws the common threads of that research in order to recommend on the role of testing in supporting ICT governance to achieve cyber-resilient systems.

II. STATE OF THE ART

A good example of the growth in software-intensive functionality and the associated software and cybersecurity difficulties in a project is the Joint Strike Fighter (F35) aircraft program [17], where capability growth has been limited by uncontained software deficiencies in development and now in early operational testing. Reference [18] examines the testing and certification difficulties of combining such software functionality levels into advanced aircraft software that have intelligent autonomy as well. He supports the push towards a more continuous certification approach, one that combines both test-based verification and analysis-based verification.

Growth in ICT technology in society more broadly is rampant, leading some reviews to have predicted a so-called '*C generation*' of *'digital natives*' [19] and other researchers to predict a shift from the Information Age to a new Synthetical Age [20]. For example the prediction by [19] has come true that a '*highly connected generation will live "online" most of their waking hours, comfortably participate in social networks with several hundred or more contacts, generate and consume vast amounts of formerly private information, and carry with them a sophisticated "personal cloud" that identifies them in the converged online and offline worlds.*'

Consideration of current and emergent cybersecurity risks must also occur early in the software development lifecycle. Failure to understand the types of threats by designers and developers often leads to security flaws in software projects that are either costly to remediate or that place the owner at additional cybersecurity risk for the life of the product.

Both [9] and [21] attribute part of the difficulty with IT projects and software-intensive systems to limited understanding by engineers and managers of how to implement the emerging technology, too often leaving it entirely to software engineers and engaging these much too late in the process. Reference [1] extends this difficulty with software to the business and government leaders of such projects, while [22] extends that leadership concern to cybersecurity and [23] to cybersecurity in Australian Defence in particular.

Preview or pre-contractual test and evaluation of prototypes, where necessary using modelling and simulation, is key to de-risking projects [24]-[26]. Reference [27] outlines that software development has long been capable of rapid prototyping and they cite early research showing that user performance improves about 12 percent with each design iteration and that the average time to perform software-based tasks decreases about 35 percent from the first to the final iteration. While [27] is concerned for usability, [8] reinforces this same early approach for reliability, stating, '*The availability and continued development of tools for modelling SoS now provide a useful toolset for testing, evaluating and understanding the behavior of large complex systems in a virtual environment.*' For example, [11] explains how federated systems integration laboratories (SILs) connected by dedicated test networks and live, virtual and constructive (LVC) simulation capabilities have enabled the U.S. DoD to do early preview of modelled new systems in representative complex and interconnected operating systems-of-systems where they are intended to be used.

Adjusting ICT governance to these challenges has seen new standards, such as the ISO/IEC 38500:2015 that provide guiding principles for the members of governing bodies of both public and private enterprises in making decisions for their ICT use [28]. The ISO/IEC 38500 standard is limited in its guidance for developing cyber-resilient ICT through projects and through-life. There is an ICT governance support package called COBIT5 that provides a Performance, Compliance and Risk Control Framework for ICT project management [29]. This deeper and trademarked framework does not directly include benefits realization around cyber-resilience; at least not one that is 'test led' in the way proposed herein. That said, the quality framework of COBIT5 would likely adapt readily to provide such a test focus with appropriate regard to the other key governance factors.

Better governance frameworks of ICT projects need to under-pin readily available test capability for the necessary usability, preview de-risk and whole-life cyber-resilience monitoring to occur; however, research literature on such governance appears scarce. This scarcity is most likely driven in part by beliefs that extant project governance can be stretched or sped-up to cope.

III. AUSTRALIAN ICT GOVERNANCE EFFORTS

ICT project problems and cyber-vulnerabilities have not lessened the pace of advanced software functionality in all aspects of governments and society. Collectively these factors have seen renewed interest in ICT governance, from areas as diverse as program management offices (PMOs), departmental reform, and high-assurance security. Some of the proposed governance models considered have great complexity and isolation to ICT-only organizational structures in attempts to build prophetic and prescient oversight; while others, appeal to simplicity for success and leverage extant PMO reviews. Some governance models seek great rigor and acceptance before operational service, while others focus on the wherewithal for a life-long learning and development. Ironically, both these approaches of upfront rigor and through-life development, see the developing cyber-threat as reinforcing their approach.

A. Department of Human Services (DHS)

The DHS is Australia's administrator of all forms of social security and health payments and due to the high costs involved, works closely with the Australian Taxation Office. Reform efforts in these departments have been focused on automation right through to the customer (public) and exploiting the benefits of the paperless '*enter once, use many times*' approach to improve efficiency and effectiveness. DHS places a high priority on both governance and research. This is not just limited to ICT governance, cybersecurity also has a focused effort on governance and there are many policies that form an Information Security Management Framework. The Technology Innovation Centre is an example of the priority DHS places on research. Innovation across ICT and cyber-operations is key to delivering solutions that utilize the most contemporary and beneficial technology and processes. This includes investigating how new market technologies can be adapted to assist both customers and staff.

Understandably the DHS projects deal with large numbers of users (public) and require high privacy and security, so as to avoid exploitation at every level from individuals, through organized crime to state-based disturbances. Governance reform was led in these departments from around 2011 by the adoption of portfolio, program and project management offices (P3O) [30]-[32]. The P3O focus is evident from their slogan '*Right Projects, Right Way, Right Results*' [33]. According to [34] their P3O encountered resistance by individual projects such that a symbolic large-scale model of the process was built in the foyer with a funnel shape to reinforce projects would be culled or reset if necessary for excessive risk or poor reviews [35]. What emerged from this P3O is a governance model focused on delivering successful ICT projects through informed decision-making; which in turn, is based on evidence-based testing of four types as shown in Fig. 1 [36]. This elegant solution has limits that derive from its deliberately simple project-level portrayal, such as it ends when the project achieves business use without any through-life expression coming from a business handover. Also, this model does not deal with the other operational or legacy systems in the business ICT architecture, except insofar as the integration, cybersecurity and user testing discloses. There has been a focused effort to increase the test capability at DHS and in recent years a dedicated test director has been established as well as the opening of an advanced Cybersecurity Operations Centre. Future effort is on improving the representativeness of the operating test environment (i.e., SIL), particularly to model more realistic cyber-attack surfaces, both for in-service systems and developing systems as much as possible within their intended in-service architecture.

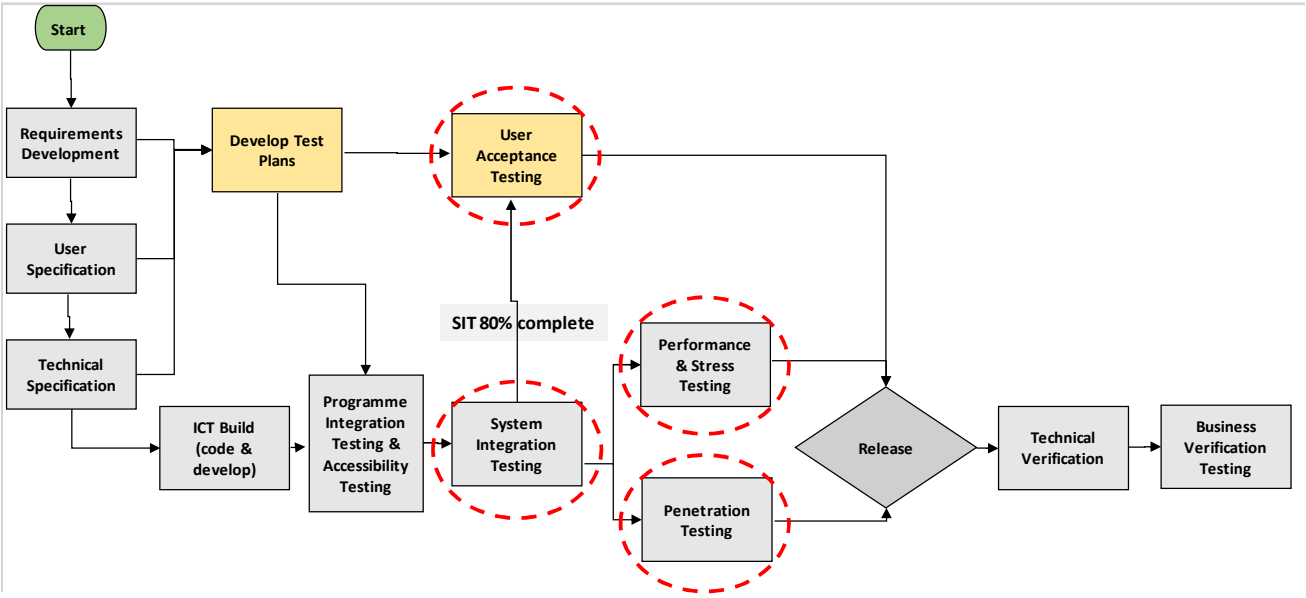

Fig. 1. Example of ICT Project Governance model (adapted [36])

B. *Department of Defence - generally*

The Australian DoD has also adopted a P3O governance model to its acquisitions following a first-principles review in 2014 [37] somewhat along the lines envisioned by [35], but so far without his proposed P3O accreditation or charter. The DoD differs from the DHS department in having a much smaller percentage of purely ICT projects, a greater complexity of interconnectivity between systems (i.e., families-of-systems-of-systems) [11], and many other competing acquisition domains for complex platforms like ships, submarines and aircraft. As such, ICT sits in acquisition and through-life operations as one of many disciplines in a matrix model, led by a Chief Information Officer (CIO) Group. The First-Principles Review sought to simplify acquisition policies and realign several different investment lines like estate, ICT and warfare platforms [37]. Unlike DHS, whose CIO is primarily responsible for the ICT projects, the current DoD governance structure sees the CIO have a significant role in managing a few ICT projects but as a specialist adviser to some 140 projects, 40 programs and five portfolios as required. The CIO's specialist role advising all acquisition projects and in-service portfolios is seriously challenged by rising demand and a paucity of complex ICT acquisition skills, especially in cybersecurity [38]-[39]. The demand is driven by the DoD's significant rise in software-intensive systems, its increasing cyber-threats ([14], [23]) and the increasing internet (/intranet) connectivity of its platforms.

Governance efficacy in such a CIO model is in the CIO primarily advising at scheduled project milestone approvals. Hence, a governance framework can be more about the decision-making approach that will pervade decisions no matter where they occur in the lifecycle or the program and project that is under review.

C. *DoD Research into Improved Governance*

A framework under development for the DoD is shown in Fig. 2 [40]. This model supports strategic alignment between business and IT for the creation of organizational value [41]. It provides an agile and benefits-driven approach to the governance of current capabilities and rapidly emerging and converging future technologies. Such technologies are not necessarily understood nor envisaged, especially with the advent of a new Synthetical Age [20] (or 4th Industrial Revolution [42]). The proposed framework is designed to support decision-making on investments on technological innovations that, while being disruptive, would be required in the organization's technology stack to generate benefits in the future [43].

Key to this decision framework is to understand that information systems investment does not provide any sustained advantage by itself, nor does it have any inherent value. Value is created by the organization's ability to convert and use the IT resource. Researchers call this *benefits realization*. Firms develop information systems to realize benefits after the implementation of the system [44].

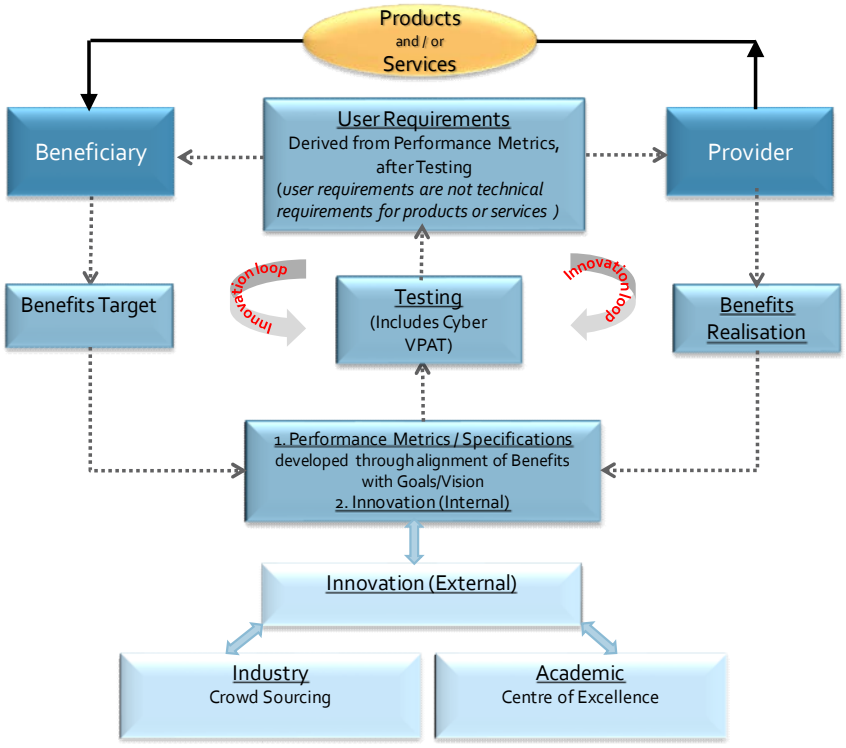

Fig. 2. Conceptual ICT Agile Governance Framework for innovation-led approach to benefits-realization (adapted [40])

The realization of benefits also comes from interventions; that is, changes to the way the business is conducted and how people work. There are two types of interventions – problem-based and innovation-based. In problem-based interventions, improvement targets such as return-on-investment form the basis of business case. The well-known approach of Enterprise Resource Planning is an example of problem-based intervention. In the case of innovation-based interventions, it is difficult to specify the end targets because there is uncertainty about the implementation success. This uncertainty implies that objectives and scope may well change during implementation, as the organization learns more about its environment and the evolving technology [46].

In the new governance decision-making framework (Fig. 2) the *Beneficiary* initiates the requirements for the new ICT by identifying the benefits that the new technology would generate. While the Beneficiary has a limited idea of benefits, perhaps gained from previous projects, the expertise on identifying, proposing and supporting the realization of benefits rests with the ICT *Provider*. Therefore, the important consideration here is to align the *Beneficiary's* expertise in the organizational goals or problems with the *Provider's* expertise in identifying, proposing and supporting the realization. When this alignment occurs, it should synergistically and iteratively achieve ongoing innovative solutions underpinned by persistent performance metrics, specifications, test and evaluation. The alignment occurs in a closed continuous "*innovation loop*," providing the goods and services along different phases of acquisition through sustainment and until the Disposal Phase. In this continuous innovation loop, the *Beneficiary* and the *Provider* benchmark innovations external to their organizations with the support of specialists, such as academia and the wider industry, shown in the bottom two boxes (Fig. 2).

This framework initiates the following best-practice arrangements:

a) Access to best practices and innovation through external agencies. The internal innovation loops are coupled to external innovation programs of research centres such as academic centres of excellence, industry R&D, and crowd sourcing. The external research centres should trigger further innovative solutions; as an example, using forecasting techniques like *reference class forecasting* [45]. These forecasting techniques also address the risks arising out of optimism bias and strategic bias situations where many benefits in the business case are overstated so as to get the project approved, leading to the promised benefits not being completely realised [46]. Use of the business principle of *incremental enlargement* [42] coupled with *reference class forecasting* would assist with identifying realistic benefits targets prior to each investment review.

b) Best practice contracting arrangement. According to [47] governance structures in a commercial environment will benefit from being of an "ongoing kind" where parties preserve cooperation during contract execution. He suggests a flexible approach with the "*contract as framework*" in contrast to the more familiar concept of the "*contract as legal rules*". The contract in Figure 2 can be viewed as a flexible framework and not a rigid one which often serves as a legal weapon, protective device, or heirarchy. The flexible framework allows collaboration and sharing of information that hopefully leads to reduced contractual overheads.

c) Collaboration between three groups. This framework, should bring about a partnership of three groups – the organisation that desires ICT-led change, ICT industry (includes the provider), and an ICT academic research organisation and/or other expertise such as crowd-sourcing.

This DoD-funded research into ICT governance has found the need to focus projects on demonstrating compliance to the benefits approach through the four key ICT testing areas outlined later in this paper. This is because benefits inherently involve the same areas of usability, integration with the in-service operating environment, network performance, security and cyber-resilience, as well as important trade-offs between these benefit characteristics. This research is now focusing on characterizing the governance approach across the ICT capability lifecycle and the necessary tailoring for capabilities with differing levels of software-intensity in the systems.

D. DoD – High Assurance Review

Concurrent to the broad ICT governance framework research just outlined, the Australian DoD has been reviewing its governance of high-assurance ICT capabilities in support of many other government departments [48]. Such capabilities must be based on products that have undergone a high assurance (HA) evaluation, characterized by a rigorous investigation, analysis, verification and validation of the products or systems against a stringent information security standard, in this case the DoD's Information Security Manual (ISM), in order to protect highly classified information. Such capabilities have historically been assured through High Grade cryptography — the processes and standards that evolved from the experiences of World War 2. Over the years, these ICT security evaluation processes and standards have evolved, divided and come back together. In 1985, the so-called Orange Book [49] contained the U.S. DoD's Trusted Computer Systems Evaluation Criteria, which was the first widely released systematic set of standards for securing computer information systems. It was influential among U.S. allies as the basis of national standards. By December 2000, the Orange Book was retired being effectively subsumed into the so-called Common Criteria published by the U.S. National Institute of Standards and Technology (NIST) [50]. A parallel set of processes and standards have developed in the U.S. [51], U.K. [52] and Australia [53]. All of the approaches to HA have two aspects in common:

- *Compelling evidence.* HA is a property of the evidence, not the system. It also makes assumptions about the independence and expertise of the entity evaluating the evidence.
- *Specified requirements*. HA needs requirements that are simple enough to be analyzed in a reasonable

time and are refutable. This makes it possible to evaluate whether the design satisfies the requirements, in other word is effective, and whether the implementation matches the design.

To satisfy these approaches the governance used to manage systems protecting highly classified information evolved into a set of prescriptive policies applying to discrete security compartments, where isolation was the main security enforcing mechanism. The growth in the demand for real-time collection, processing, exploitation and dissemination of intelligence, targeting and geospatial information from increasing numbers of capable collection assets, has seen much of the HA edge, if not eroded, certainly outsourced and at greater risk of compromise. Such risk also derives from the growth and reach in sophisticated cyber-threats that contest the Western pursuit of information dominance [11] [14] [54]; or put another way, a joint and networked force [55]. The underlying cause of this growth is from the fact that all new Government capabilities have a strong ICT component. For the Australian DoD, this has meant a significant increase in the number of systems designed to secure highly classified information or connect to other systems that protect highly classified information.

Currently the approach used to assess the security of systems protecting highly classified information has not been able to keep up with the demand [48]. The increasing HA demand and the changing nature of ICT led the U.S. DoD 15 years ago to develop an improved HA evaluation methodology [56]. Other allied nations have generally not followed suit and this has arguably led to a general weakening of their comparative ability to evaluate and certify the security of systems protecting highly classified information. This is despite a number of research and policy efforts over the years to improve HA efficiency and effectiveness, such as policy initiatives like approving public domain cryptographic algorithms for protecting highly classified information and the ongoing research into high-trust techniques like formal methods.

To address the demand issue the recent Australian DoD HA governance review [48] found that the HA responsibility needed to be spread across the P3O reviews and be more clearly focused on benefits realization through informed ICT testing. The necessary test areas were found to be the four areas outlined later in the paper, albeit some being more specialized, in-house and secure. Specifically the HA review found that in order for P3Os to deal effectively with HA, HA must scaffold more into the whole ICT life cycle.

To inculcate HA and security more broadly into all aspects of ICT, two sets of processes are proposed: one set to influence the behavior of the ICT life cycle and the other set to measure, test and evaluate the security performance throughout the ICT life cycle. The first process set is known variously as supply chain management or Information Security Industry Engagement (ISIE). The second set of processes is generally known simply as test and evaluation, though in our case we should specify the purpose as conducting an Information Security Evaluation (ISE). These two process sets interact and feedback upon each other, with the ISE providing the compelling evidence and the ISIE

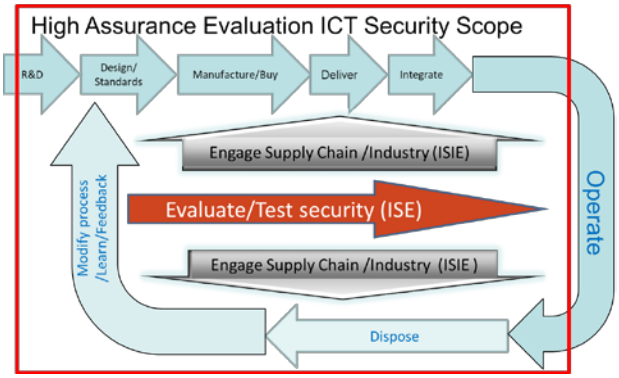

Fig. 3. ICT Security Cycle & HA Evaluation Scope (Red Box).

managing the specified requirements. Feedback and interplay between ISE and ISIE processes can be complex, where overlapping boundaries abound. For example, with ISIE the broad aim is to manage ICT Supply Chain Compromise per the concerns outlined by [57]; in other words, to manage an occurrence within the ICT supply chain whereby an adversary jeopardizes the confidentiality, integrity, or availability of a system or the information the system processes, stores, or transmits. One way to manage this is to use Formal Methods [58] as a tool to specify in a mathematically rigorous way the security requirements. The process of proving that these specified requirements meet the security objects also provides the evidence needed for ISE [18] while also managing the supply chain issues [57]. The HA Evaluation process can be summarized as the functions within the red box of Fig. 3.

IV. FOUR TESTING TYPES

This paper will now clarify the four ICT testing types assessed as crucial to informed project governance for evidence-based benefits realization, sound integration, consistent cybersecurity and thus ultimately more cyber-resilient operating environments (families-of-systems-of-systems [11]). Each type of testing will briefly examine the unique test design and analysis skills that are needed.

A. Usability Testing

Software performs functions for systems replacing both mechanical systems and human operators alike in a continuous frontier of increasingly complex heuristics that also includes new language development, new processing and proprietary boundaries. As such, it is rare that software in systems technology ever repeats functions in precisely the same way to the same purpose and for the same user. Human-machine interfaces have been well researched since computers evolved [27] and this research has clearly shown the efficiency and effectiveness benefits of usability testing that were cited earlier, including standard usability test metrics. Yet, ICT projects abound with poor performance stemming from under-researched user requirements [1] and from the authors' experiences they rarely use structured and iterative usability testing as shown in the software development cycle at Fig. 4.

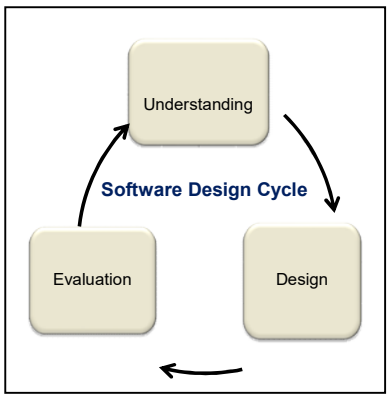

Fig. 4. Software Design Cycle (adapted [27])

Where user interface requirements are documented and the software tailored to those, they are often then tested to those functions for correct coding by other software, as if they were now axiomatic. Functional testing is usually the unfortunate substitute of better metrics of user efficiency, effectiveness and satisfaction that should be tested iteratively for improvement. Hence, software functionality is often operationalized with a high degree of expectation mismatch. According to [27]:

'*The fact that computer software is sometimes poorly designed and therefore difficult to use causes a variety of negative consequences. First user performance suffers; researchers have found the magnitude of errors to be as high as 46 percent for commands, tasks and transactions in some applications. Other consequences follow, such as confusion, panic, boredom, frustration, incomplete use of the system, system abandonment altogether, modification of the task, compensatory actions, and misuse of the system … The trend toward a greater number of functions, called creeping featurism, is an important problem because the additional functions make the interface more complex and increase the number of choices a user must take.*'

The authors' experiences are that most project managers are not software specialists and that when usability testing is described to them, their greatest fear is that the requirements will increase (creep) and that even unwanted features will cost to be taken out. Most project managers also do not realize software can be virtually modelled and thus it can undergo the first usability testing even before the code is written, ideally even before full developmental contract [24] [25] [60].

One example of the benefits of the P3O-led usability testing can be seen in the successful digital linking of taxation and social security in Australia. By contrast, a large software project in the Australian DoD for a command support environment debuted 20 major applications in an operational evaluation where only one had received something approximating usability testing (one iteration) and only half were suitable enough to continue use, albeit commencing with usability upgrades. In the case of one application, the live evaluation found just three people whose function would use the application and none had been consulted previously in the development. Similarly, a major battle-management system in the Australian DoD debuted operationally using an off-the-shelf foreign application that did not adequately match the military culture or organizational structures, and while that rapid evaluation did ratify the broad benefits of digitizing battlefield plans, a redevelopment using three iterations of representative usability testing has since fundamentally improved acceptance and effectiveness. Furthermore, the redevelopment has enabled an assessment of, and improvement in, cybersecurity, which was not envisioned in the original off-the-shelf project.

Good usability testing fundamentally contributes to cyber-resilience. Modern context-dependent security-monitoring algorithms (i.e., beyond 2-factor) only work if you can accurately capture what normal use looks like. Unusual and unnecessary features simply provide avenues for malicious exploitation by increasing the attack surface and in high-assurance applications, will require repeated cybersecurity checking against each wave of new threats. Additionally, security controls that are not user friendly often result in the users finding ways to avoid having to comply with security policies; an overall degradation of cybersecurity in a system that in its pure design was more cyber-resilient.

Usability testing must also continue to be undertaken in applications where "human-in-the-loop" functions become "human-on-the-loop" and then "human-out-of-the-loop." New standards are being developed to specifically cover intelligent and autonomous systems and how they are to be ethically designed [61], where the operational commander needs the usability testing to accurately ensure and record that the systems use always accords with human intent. Such ethical design will require the structured experimentation with a representative sample of commanders that comes with usability testing. While that sounds very military, such autonomous intelligent systems are set to rapidly pervade health care and other domains via the internet-of-things (IoT) [62]. Within Australia's public sector areas there have been too many examples where policy initiatives have not had sufficient preview or trial [63]-[64] — early usability preview of software functionality prior to developmental contract is key to such de-risking strategies being evidence-based. For example, DHS has developed dedicated user testing suites that bring members of the public in to examine and participate in such aspects as application development, online design and user interfaces. This enables early engagement with the very people that will use the final systems. Understanding their needs and constraints and integrating that knowledge with in-house developers assists in ensuring the products are fit-for-purpose upon initial release. As a delivery agency to almost every Australian there is a balance between placing cybersecurity measures (multi-factor authentication, password complexity rules etc) embedded in the user interface to attempt to demonstrate and guide the cyber-behaviors of the individuals and moving cybersecurity features behind the firewall to prevent the user from being inconvenienced by them.

Usability testing metrics and constructs are not difficult test design skills and methods to manage and an introductory text is [27]. It is advisable to have contract coverage of the usability testing and at least three to five iterations, as well as some method to ensure representative sampling of the users. As mentioned earlier, at least the first de-risk usability testing should be pre-contractual (i.e., offer definition

activity in tendering) based on virtual modelling (pre-coding), especially if some applications and operating systems are claimed to be off-the-shelf (i.e., mixed maturity architecture). Preferably at least one person in the test team needs to have had university-level education in introductory human factors and another in human-computer interfaces (i.e., graphical user interfaces). For efficiency of determining significant factors and identifying confidence levels, at least one of the test design team needs to be experienced in design of experiments (DOE) test design techniques (i.e., [65]), especially efficient screening of main test factors with high throughput testing (HTT) based on two-way combinatorial test designs [66]. Ideally, that test designer should be the person qualified in human factors, as usability testing should not be diverted to center on requirements verification or be overlaid with more technical testing. If original user requirements or functionality are verified in usability testing it should not be disclosed to, or constrain in anyway, the representative users. Subsequent iterations of usability testing could move to more classical DOE test methods as the non-significant factors are removed from analysis and test performances converge on the acceptable user performance metrics.

B. Systems Integration Testing

The greatest difficulties in system integration testing are: first, efficiently dealing with the multiple configurations of hardware and software configurations present in distributed ICT architectures [67] and second, having representative operating systems from the system-of-systems or families-of-systems where the new system will debut [10]. The ideal first step is having a P3O culture to deliver virtual models of new software prior to full developmental contract that can then be tested in a software integration laboratory (SIL), software system support center (SSSC) or distributed live-virtual-constructive (LVC) test integration network [11].

Organizing and sustaining a representative test architecture for evolutionary improvement faces many proprietary and funding challenges, as well as the challenge of staying representative when portions of the network exist in the public domain, such as the internet and self-funded public users. Given the existential and also evolving cyber-threat to the public sector [22] and financial industry, proprietary outsourcing of parts of their ICT architectures and ad hoc approaches to other parts are resulting in increased vulnerabilities being introduced during ICT build and integration that will become too risky at some stage in the future. Proprietary support will need renegotiation so as to be representatively maintained in a SIL / SSSC / LVC test network, and to be independently and regular tested for vulnerabilities against cyber-threats without disclosure of the test methods used. Such arrangements fundamentally challenge corporate and public sector outsourcing models for ICT from the last few decades and associated fixed-price and intellectually-protected contracting arrangements, in favor of more cooperative security arrangements and fee-for-service. Reference [68] assesses that the contractual and project methods used for system safety can be leveraged to achieve this greater flexibility. Certainly, more mature process instantiations for cybersecurity testing like the U.S. DoD [69] and industry [70] are forcing cooperative flexibility like that hitherto seen in safety. This becomes particularly difficult at scale and DHS is an excellent example of the constant effort required to maintain test and development environments that are representative of the production system.

The Australian DoD is proposing battle-laboratories of mixed SoS for greater integration [71] and there are options to cost-effectively leverage the U.S. DoD LVC test networks [11]. While the main forces returning departments and major corporations to such in-house testing are the risks in cyber-threats, for system integration testers this is a welcome reprieve from trends towards disparate sub-system testing and limited opportunities for SoS testing with high operational risks [8]. According to [72], U.S. DoD project managers have often been focused at narrow capabilities, lacked program or portfolio support to consider a wider good, facing cost and schedule pressures, and of very limited tenures compared to the capabilities they deliver. The advent of P3Os, the cyber-risks, and the need for representative integration centers will provide much needed focus and support for such ICT project managers to be fully informed in the way envisioned by governance models at Figures 1 and 2.

System integration test skills demand high levels of ICT integration knowledge around fusing of applications, operating systems and datalinks, as well as management of regular and emergency demand, cyber-defense; and for DoDs, how to expand and contract services based on military priorities and deployment. Such advanced technical skills and experience often comes at the expense of formal education in test design and analysis, such that tests run by integration experts alone are usually successful but sub-optimal. At least one of the integration test team should be educated in test design methods so that integration tests are efficiently screened early to focus subsequent testing on the significant factors. The HTT combinatorial methods for all two-way combinations are ideal for functional testing in an integration environment [7] [78], followed by a focused modelling on interactions involving significant integration factors with a more orthogonal test design. Combinatorial test design packages with algorithms to optimize orthogonality offer combined efficiency and diagnostics to the software integration industry with genealogies in both Japan [73] and the U.S. [74], but they require higher education of the testers than random (fuzz) test methods, and understanding by the beneficiaries, who associate fewer test runs with greater product risk. Significant dedicated test experience is required if testers are to use the efficiency of combinatorial methods at the meso-level with the investigative advantages of the fuzz methods at a micro-test level.

C. Performance and Stress Testing

Performance and stress testing of the integrated and usable ICT or software-intensive architecture is necessary to ensure manageable demands during the full range of usage cycles. User satisfaction, system stability and effectiveness

are inherently linked to timely performance. Not only is maintaining the networks challenging but also producing replica test load. For banks with 3.5 million online user authentications making 2.5 million payments per day, and departments such as DHS with over 26 million users and approximately 600 000 authentications daily, it is difficult to reproduce that level of load in virtual environments. Some system errors can only be replicated under load and testing for these errors is complex and expensive. The only other alternative is to run scripts in production environments whilst under the assessed load, which significantly increases the risk. If there are errors in the code and they do then appear during load testing the associated error is then potentially exposed to all those using the system at that point in time. For large systems, this risk is often too high to accept.

ICT reports are, like many military systems, often replete with the use of averages, sometimes inappropriately aggregated across diverse mission scenarios without regard for the underpinning statistical distributions and appropriate confidence limits. Like stress loads, it is difficult for banks and DHS, due to the scale, to produce accurate reports on performance during the testing phase due to the inability for all components to be integrated, combined with the absence of load. The U.S. DoD has worked hard to improve test plans and test reports to deliver better reporting of performance metrics [75]-[76], including in the presence of cyber-threats and varying cyber-defense postures [11] [77]. Better education of ICT testers in test design and analysis techniques offers not only efficiency gains in the testing, but better rigor in reporting results and managing with operational models the cycles of demand. Operational models should not simply be based on deterministic predictions, but backed by probabilistic performance test modelling. The skills necessary to do this are readily available in most six-sigma industry accreditations [65] [66].

D. Security Testing – Vulnerability and Penetration

The pervasive cyber-threat to DoDs, public sector, finance and industry means cooperative vulnerability and penetration assessment (CVPA) is no longer an option but rather about managing an acceptable risk of how much testing is enough [7]. Not testing, simply means not knowing, and thus an unassessed risk, while not re-testing fielded systems at some interval means an atrophy of security confidence at an unknown rate [77]. Additionally, over time, systems acquire aggregated cyber-risk as different elements of complex systems are deployed. In critical systems, operators are mounting continuous defensive cyber-operations, sometimes extending to supply-chain monitoring through-life [57] [68], but in Australia these precautions are largely only on live networked systems [23]. Outside the U.S. and particularly the U.S. DoD, there is still limited understanding of the risk of cyber-threats to software-intensive systems that are only occasionally updated or networked. However, recent DoD testimony to the Australian Senate announced a program of what is being termed "*cyber-worthiness*" of capabilities [79], hopefully following research recommendations like [80].

Public sector and financial systems are vulnerable to more sophisticated probing and logic disruptions that can now be electromagnetic lodged at low power with no connectivity [23]. Without CVPA and some defensive posturing even for fielded legacy systems, significant risk exists that at a time of a potential enemy or criminal entity's choosing, systems will be denied or interfered without detection for an unknown period of malicious intrusion [78].

The cost of mounting expensive CVPA and defensive capabilities will be borne by either a slower pace of computer-based services to the public and DoD capability, or increasing market differentiation. Reference [11] documents a widening difference in systems integration and cyber-resilience between DoDs of even close allies like Australia and the U.S.. While the cybersecurity of two militaries might seem irrelevant to much of the public sector or critical industries, the reality is that such differentiation as that described therein can soon be expected in public sectors and markets. Regular CVPA on representative operational test architectures (i.e., federated SILs) is needed for as much known threat as possible. The capability to do so needs to be introduced and funded at a portfolio-level, so as to enable informed decision on each new system release and collectively how to strategically posture resilience of the operating systems. It may be that for some services, risks are low and public or consumer risks can be tolerated, even deliberately targeting cheaper or efficient services with perhaps a greater explanation of consumer and public risk. Whereas for other services, capabilities may be compromised and costs raised to enable greater cyber-protections. As always, not testing is not knowing and that means no informed choice.

Militaries have a unique advantage in that many of their combat systems can be isolated to a certain degree from the internet. This is not the case of other Government services such as the ATO, DHS and banks, where their core business is linking Australian citizens to payments and services through the internet. The use of intelligence provided by military counterparts in persistent threats is of great benefit. Ultimately, the threats that only a few years ago were aligned to largely espionage or a criminal intent have now converged.

Forming a CVPA test capability is dramatically easier if the other ICT test capabilities (i.e., usability, integration & performance) are robust and appropriately part of project governance and a benefits-realization decision-making culture. Inevitably in capped schedules and budgets, increased cyber-resilience involves trade-offs between:

- user requirements, such as determining through structured test what users value more;
- ICT build, such as limiting use of code libraries to those known to be cyber-secure;
- integration, such as limiting connectivity to limit cyber-threat exposure; or
- performance, such as increasing the threat detection algorithms and reducing system processing for main functions.

Having the other ICT testing well run and iterative, as shown in governance models at Figures 1 and 2, enables

Governance to make these cyber-resilience trade-offs in an informed way. The next level of maturity for large organizations is to combine them and ensure CVPA is integrated into the formal testing cycle. However, to be robust, it should also be conducted to systems at regular intervals post-delivery, with appropriate levels of funding set aside to address the vulnerabilities that are then identified.

The test design, test analysis skills and test infrastructure required to manage CVPA testing are, with only a few key additions, supported by the test skills and test infrastructure of the other ICT test types. For example, industries, public sector and government departments that have invested in SILs, SSSCs or LVC test networks can adapt these to allow for multi-security CVPA testing — in essence extending integration and capability upgrade infrastructure to be cyber-ranges that can concomitantly manage evolving cyber-threats and deliver greater cyber-resilience. If such infrastructure has been outsourced and is proprietary, then contractual changes will be needed to safeguard connection to government-managed representative cyber-threats. For example, the Australian National Audit Office (ANAO) assessed that DHS had security controls in place to provide protection from external attacks, internal breaches and unauthorized information disclosures [81]. This was achieved by prioritizing activities that were required to implement the top four Australian Government mitigation strategies and by strengthening supporting governance arrangements. This prioritization was largely enabled by the in-house capability that DHS possesses and the lack of reliance of contracts and service providers. Similarly, the challenge for Australian banks is to be compliant with the Australian Prudential Regulation Authority (APRA) and Sarbanes Oxley (SOX) by implementing the top eight mitigation strategies and establishing a cyber-resilience culture.

Similar to test infrastructure, additional test design skills can be added to integration and performance testers to manage the additional rigor necessary for testing cyber-resilience. Again combinatorial test design has been instrumental in achieving greater cyber-resilience with three-way through to six-way combinatorial test rigor being achieved, often while deriving new efficiencies [6] [74] and other defect-protection rigor. An example of this approach is the industry six-sigma software testing award overview by [82]. The University of NSW [66] has adapted test design education to give early awareness of these additional cybersecurity test techniques using the freeware by [74] as a reasonable simplification of the test design packages used by big software industries [73].

Industry and departments have been slow to adopt another protective process layer, which has led industry bodies to develop minimum additional cyber-planning and testing checks to overlay standard systems engineering [70] [83]. These process links and explanations offer the greatest promise to normalize cybersecurity in industry, albeit that industry using system engineering practice.

Probably the last and most difficult extension for CVPA testing from hitherto ICT testing, is the skill of defensive (blue) and penetration (red) teams war-gaming the cyber-threat as described well by [78]. These are military skills applied in a new domain and unfortunately necessary for public sector and critical industries to adopt if they are to be reasonably defensive to malicious threats. Legal protections in cyber are a long way from being instituted [14] [22] [84] and deterrence critical depends on timely attribution, which unfortunately remains difficult. Even if legal recourses become viable, public sector and industry war-gaming is necessary at some level for the defensive capability to exist to collect evidence for legal recourse.

DHS has proven this applicability outside of DoD. In 2017 they ran the first government cyber war-games on a cyber-range built in-house and representing a fictional city. Ten departments and agencies combined to form five teams that conducted both defensive and offensive play and were assessed on skills outside the technical, such as teamwork, communication, leadership and critical thinking. To be able to defend, understanding how to attack is critical. Ultimately, it is another human behind the opposing keyboard and being able to understand how they may manipulate the systems to maliciously achieve their aim will ultimately direct a diligent defender to monitor, protect and defend the right elements of the system.

The other skill that is difficult to build, and also tested during the DHS cyber war-games, is the ability to translate the technical nature of cyber-operations to both the boardroom and the media. In times of cyber-disruption the ability to deal with the media in what is inevitably an uncertain time, where the nature of network problem (network outage or cyber-attack) is unknown, is another complex skill. In large organizations and government departments having a technical team to conduct CVPA integrated with an engagement team capable of doing that media translation is key.

Building CVPA test skills and infrastructure requires education to be improved to merge the necessary knowledge and skills into industry-accredited packages [38]. Having teams of testers that are able to conduct the required testing is an initial start. Having testers that are able to schedule tests to match the development schedule is the next step. Moving testing to the left in the software development lifecycle and conducting CVPA throughout development is even better [77]. But ultimately, designing and building cyber-resilience in, by having cyber-operations staff embedded with both the design and development teams from the start, is key. Ensuring that at the first conceptual design any ideas that will invoke cyber-vulnerabilities are discussed and the risks are clearly articulated to cyber-aware business owners early [78], such that testing throughout both design and development is combined with training the developers. Outsourcing, offshoring, and high turnover of developers all magnify this challenge. Often similar mistakes are seen multiple times because developers aren't made aware of new and emerging cyber-threats and how the way that they code allows cyber-criminals and state-based actors alike to exploit those flaws.

Having centralized code libraries sees any vulnerability that has been introduced exponentially deploy through the network as code with security flaws is drawn from a central library. Automated code scanning, during ICT develop and

test, is an effective way to assure absence of known code vulnerabilities. There is constant tension between cost and benefit in testing and in particular CVPA. All organizations and departments have differing risk appetites and what may be acceptable to one will not be at all palatable to another. There is a significant cost in increasing the cyber-resilience of any organization or department, however, the reputational cost if information or systems are lost or exploited, in most cases, far outweighs the required investment to secure it. Technology is only one element however, without equal investment in the people and an understanding by the beneficiaries of the business implications, the CVPA system will never achieve full maturity. A layered security approach has to be designed to be both complex and obscure [70] [83].

Specific CVPA test design and analysis skills of the types outlined by [66], [78] and [83] need to be available in country and *en masse*, but this requires industry and public sector to commit to their staff undertaking the education and placements. Furthermore, this requires governance structures and awareness regarding the necessity for cyber-resilience and the wherewithal to achieve that through CVPA testing.

E. Security Testing - High Assurance Evaluation

The aim of information security evaluations (ISE) is to make sure that the effort required to defeat exceeds the value of the material being defended. An evaluation aims to measure that effort and compare it with the value of the protected material and the resources available to a likely threat. In general both active and passive exploitation requires all three of the following factors [48]:

- a vulnerability to gain the initial access;
- an implant or processing system to retain access; and
- a communications system to manage the command and control (inward) and an export means (outward).

The aim of both policy and technology is to block at least one of these factors. Essentially all threats exploit failures of either policy or technology in these three areas.

The aim of ISE is to show that no security failure state exists by demonstrating the nonexistence of known or likely failure states. Proving the non-existence of something is generally not possible and so in most cases ISEs measure the likelihood of the non-existence of something by searching for it and not finding it over a period of time. The longer spent looking and not finding, the more likely the non-existents' case is. As such, the level of trust or assurance one has in the system is proportionate to the effort expended in trying to defeat a system and failing.

The most common approach at lower security levels is to use process and procedures to systematically search for failures in policy and the design, rather like a check list. Most schemes and standards, such as [50] have this property. They have a list of items or controls that are needed to be enacted and checked systematically to determine if the system matches the policy. This has the advantage of making the effort required to secure a system easier to measure and manage, albeit through the rigor of compliance. Active searches for security failures, such as CVPA, also use lists of known failures and threats to see if they exist in built systems, but in most instances with a degree of war-gaming above that of systematic compliance. Such active approaches are harder to cost and outsource, due to the complexity of the failures being searched for, but it has the advantage of identifying new failures due to combinations of known failures and human ingenuity. Active approaches can therefore be difficult to justify and maintain for highly complex systems and those at scale.

Improving the governance of ISEs is the key to being able to have visibility of total risk across all systems. For example, being non-compliant with any of the set ISE controls does not explicitly lower the cyber-resilience of the system, however aggregated across many systems it may pose risk in areas not considered in isolation.

In higher security levels, such as the one used to evaluate HA equipment, the approach is different [51]-[53]. An unfettered search for a failure is conducted, and then for all the ones found, a theoretical attack is developed and then costed, using a rigorous well-tested method based on the HA standard. If the cost of the attack exceeds the value of the material being defended then the system is said to be secure. The critical part of this approach is the costing model. This model, developed over many years, determines over time: 1) the value of the material being protected, 2) the resources of an adversary and 3) the resources required to run an attack.

The unfettered search for a failure examines two aspects of security being the design and the implementation. The two metrics used to measure the effectiveness of these security aspects are: 1) the cost to defeat security, and 2) the effort undertaken looking for a new defeat and not finding one. The cost is measured in terms of resources, such as effort, money, knowledge etc. and the time required to defeat the security. The effort is measured by the number of people months spent examining the system and not finding a new defeat.

The current Australian HA standard [53] contains a number of built in parameters around the investigation effort, the resources an adversary has and the length of time required protecting the system or information. For example, it assumes that highly classified information needs to be protected for 30 years at least, from all possible organizations and to spend from 6 to 24 person months, depending on the complexity of the system, not finding a new defeat. For less highly classified systems, it assumes that 3 to 12 person months have been spent showing that no organization will have the resources to defeat the security for 10 years.

A key difference between the HA and general ISE evaluation methods is that, the HA one focuses on the resources required and available to exploit security failures and defeat a security system, while the general ISE one focuses on going through a list of possible security failures and removing any present. The HA one is hard to plan, requires skilled staff, but has proven to be quicker and very effective. The general ISE one is easier to plan and requires less skilled staff but is less effective, takes longer for higher security levels and can be difficult at scale.

Where the logic is not based solely on classification it is possible to combine the two methods. This can be done by evaluating the key cyber-terrain of the organization or department. In order to know how much to invest in cyber-

resilience it is important to know where to channel that funding and effort. By understanding what parts of the network may be attractive to state-based actors or cyber-criminals it is possible to conduct those methods for HA against a small subset of the larger network. An indicative ISE method is the Attack Cost Method summarized as follows from [48].

a) Method. The Attack Cost Method is a search for the best attack that will defeat the security, then that attack is costed to determine when that attack becomes possible. There are two basic approaches to searching for the best attacks or vulnerabilities. One starts from first principles assuming nothing known about the system or device. The other approach uses security assessments from multiple sources, combining the result across the whole system. Both of these approaches are used. The method assumes that an adversary has a finite number of resources measured as money (R). It also measures the payoff in terms of plain-text documents equivalents (p) where one highly classified document is 1.0 p and one less highly classified document is 0.1 p and so on, and there is an estimated expected payoff (Pe). Cost of the attack (C) is also measured as money. So the basic idea is that for a secure system the cost (C) of all possible attacks exceeds the resources (R) available to an adversary, or the cost per plain text documents equivalents (C/P) or the cost per plain text documents equivalents (C/P) exceeds the expected cost per plain text documents equivalents (R/Pe).

b) Attack-Cost Method steps:

- Develop an adversary model and determine the adversary resources (R) and expected payoff (Pe).
- Develop a usage model or scenario and determine the value or payoff of the user's data (P) and how it will be used.
- Launch an unfettered search for vulnerabilities.
- Develop attacks from the vulnerabilities and detail the attack proving that it exists.
- Rigorously cost the attacks using a costing model developed with respect to context, calculate the cost (C) of the attack over time.
- Using the cost (C) calculate the payoff (p) and resource limits (R) over time, noting that over time R goes up and the value of the payoff goes down.
- Repeat until required assurance level is reached.
- Write up report and recommendations, put comments to the manufacture and have a trusted third-party review the report and evidence, certify the results and note any improvement and maintenance plans.
- Acceptance by the Accreditation Authority.

c) Indicative set of assurance levels are:

- 6 person months of not finding an attack for a highly classified level of assurance.
- 3 person months of not finding an attack for a less highly classified level of assurance.
- 1.5 person month of not finding an attack for a moderately classified level of assurance.
- 3 person weeks of not finding an attack for a classified level of assurance.
- 1.5 person weeks of not finding an attack for all other levels of assurance.

V. FUTURE RESEARCH AND WORK

All three of the research avenues described here are still ongoing and the collaboration to compare findings will continue under the auspices of the University of NSW Australian Centre for Cyber-Security (ACCS). Each of the authors is passionate about improving cybersecurity education along the industry-accreditation lines outlined by [38]. As such, the collaboration will hopefully have feedback from test practitioners in each of the four ICT test areas based on their experiences undertaking closely mentored and industry-placed research assignments. Australia's efforts on cyber-testing is seminal and so early industry-based feedback will be crucial to build the experience base around the ICT governance frameworks, so as to confirm what works well and what does not, especially for cyber-resilient systems. Countries with similar challenges to Australia in ICT governance are welcome to leverage the research collaboration.

VI. CONCLUSION

Difficulties with ICT projects abound in all parts of the World, with research reporting as many as one in six such projects exhibiting cost and schedule overruns in excess of 200 percent. There are also reports of many high-profile ICT projects experiencing high incidences of unexpected cyber-vulnerabilities. These project problems and cyber-vulnerabilities have not lessened the pace of advanced software functionality in all aspects of governments and society. Collectively these factors have seen renewed interest in ICT governance, from areas as diverse as program management offices, departmental reform, and high-assurance security. Some of the proposed governance models considered have great complexity and isolation to ICT-only organizational structures in attempts to build prophetic and prescient oversight from only brief project reviews, while others appeal to simplicity for success.

Three separate and diverse Australian Government research efforts in ICT governance, as well an assessment in the Banking Sector, have found similar concerns about the importance and type of ICT testing and test expertise critical to ICT governance and the ability to build cyber-resilience; namely, usability testing, systems integration testing, performance testing and cyber-security testing. These research efforts all found that ICT Governance critically depends on: (1) information coming from all four types of testing, (2) some test understanding in management to appreciate fully the outputs, and (3) that such test capabilities must be enduring (i.e., through-life, however short) so as to provide a sufficient degree of commercial and architectural independence to make hard and timely decisions.

These lessons on the importance of testing to ICT governance seem almost to have been forgotten in a rush to be technologically and managerially adroit, yet if done as outlined from these research efforts, could see a resurgence in test-informed project reviews that are: (1) innovative, (2) give lower risk competitiveness, and (3) greater cyber-resilience.

Key conclusions of this research are:

- A benefits-approach to ICT governance as shown in Fig. 2 should give more cyber-resilient operations through informed ICT capability life-cycle decisions.
- Usability testing is crucial to user satisfaction and needed even when software-intensive systems seek to replace an operator or commander.
- Development contracts should cover three to five usability test iterations, with the first iteration ideally being on a virtual software model prior to the development contract, so as to de-risk project scoping.
- Test teams need human factor engineering expertise to successfully conduct proper iterative usability testing as well as a governance culture of refining user requirements.
- Integration testing critically depends on a representative operational test environment such as a SIL, SSSC or LVC test network to be effective with significant parallel benefit to then extend such infrastructure cost effectively to do proper full cyber-attack surfaces in CVPA testing.
- The high number of permutations in integration and later performance testing requires test design skills in combinatorial HTT to be efficient. Six sigma test courses with practical competency assessments in industry are key to realizing such efficiency benefits.
- There is a balance between embedding good cyber-culture in the user interface to teach good cyber-behavior and moving cybersecurity rearward so as not to inconvenience the user.
- Sound ICT test infrastructure and test skills in usability, integration and performance testing, backed by project governance and benefits realization in the ICT test types, are crucial determinants in the preparedness and ease for CVPA testing to be incorporated and evolve for cyber-resilient systems.
- A CVPA test capability needs some additional combinatorial test design and analysis skills to deliver the necessary rigor or high-assurance against malicious intent.
- Cybersecurity processes have now been efficiently mapped to industry systems engineering so as to adequately enable CVPA testing in newly developed systems.
- The most difficult of CVPA skills and experience to acquire, particularly outside DoDs, is the defensive and penetration posturing of teams for war-gaming, but the reward for these efforts should be sound cyber-risk profiling and value-adding to public confidence and commercial marketing.

While these findings and guidelines come from Government reviews, commercially-based authors have assessed where these are universal for industry to follow; albeit sometimes to a lesser extent.

These common research threads show the somewhat unique finding that preview testing should be required directly in all ICT governance frameworks; if not for the many *a priori* reasons such testing already should exist, then certainly now for cyber-resilient systems. Furthermore, increasing system configurations, threat permutations and possible future upgrade and threat sequencing mean that ICT testing needs to use new combinatorial test design techniques for efficient screening and cyber-threat rigor.

REFERENCES

[1] N. Devine, “Department of Human Services ICT Governance for cyber-resilience,” presentation at special track on Critical Test Capabilities for Informed ICT Governance of Cyber-Resilient Systems (CTC-Gov-CRS), CYBER 2018 IARIA conference, Athens, 18-22 Nov., 2018

[2] A. Ghildyal, “An Agile Innovation-led Benefits Realisation Approach for ICT Governance,” track CTC-Gov-CRS, CYBER 2018 IARIA conf., Athens, 18-22 Nov., 2018

[3] A. Coull, “Cyber Security Transformations in dynamic and disruptive environments,” track CTC-Gov-CRS, CYBER 2018 IARIA conf., Athens, 18-22 Nov., 2018

[4] A. Laing, “High Assurance Evaluation,” track CTC-Gov-CRS, CYBER 2018 IARIA conf., Athens, 18-22 Nov., 2018

[5] S. Jenner, “Why Do Projects Fail and More to the Point What Can We Do About It? The Case for Disciplined, 'Fast and Frugal' Decision-making,” PM World J., pp. 1-18, 2015.

[6] D. R. Kuhn, R. N. Kacker, L. Feldman , and G. White, “Combinatorial Testing for Cybersecurity and Reliability,” Information Technology Bulletin, Comp. Sec. Div., Inf. Tech. Lab., NIST, 2016

[7] P. Christensen, “Introduction to Cyberspace T&E,” tutorial at 32nd Annual International Test and Evaluation Symposium, 18-21 August 2016, Director, National Cyber Range

[8] B. Normann, “Continuous system monitoring as a test tool for complex systems of systems,” ITEA J., vol 36, pp. 298-303, 2015

[9] B. Flyvbjerg and A. Budzier, “Why Your IT Project Might be Riskier than You Think,” Harvard Business Review, pp. 24-27, 2011

[10] M. Hecht, “Verification of software intensive system reliability and availability through testing and modeling,” ITEA J., vol. 36, pp. 304-312, 2015

[11] K. F. Joiner and M. G. Tutty, “A tale of two Allied Defence Departments: New assurance initiatives for managing increasing system complexity, interconnectedness, and vulnerability,” Aust. J. Multi. Eng., pp. 1-22, 2018

[12] U.S. DoD Defense Science Board (DSB), “Summer Study on Autonomy,” pp. 28-30, 2016

[13] K. Geers, D. Kindlund, N. Moran, and R. Rachwald, “World War C: Understanding Nation-State Motives Behind Today’s Advanced Cyber Attacks,” Fireeye Corp., 2017, https://www.fireeye.com/content/dam/fireeye-www/global/en/current-threats/pdfs/fireeye-wwc-report.pdf.

[14] C. H. Heinl, “The Potential Military Impact of Emerging Technologies in the Asia-Pacific Region: A focus on cyber capabilities” in “Emerging Critical Technologies and Security

in the Asia-Pacific," R.A. Bitzinger, Ed., Palgrave Macmillan: Hampshire, U.K., 2016

[15] Bodeau, Graubery, Heinbockel, and Laderman, "Cyber Resilience Engineering Aid – the updated Cyber Resilience Engineering Framework and Guidance on Applying Cyber Resilient Techniques," MITRE Corp., Bedford, MA, 2015

[16] Information Technology Sector Resilience Working Group, "Cyber Resilence White Paper: An IT Sector Perspective," IT Gov. Coordination Centre, Washington D.C., 2017

[17] U.S. DoD, Directorate of Operational Test and Evaluation, "Annual Reports to Congress on DoD Programs - F35 Joint Strike Fighter," at www.dote.osd.mil, 2015-2018

[18] D. Cofer "Taming the complexity beast," ITEA J., vol. 36, pp. 313-318, 2015

[19] R. Friedrich, M. Peterson, A. Koster, and S. Blum, "The rise of Generation C & Implications for the world of 2020," Booz & Company, now Strategy&, Price Waterhouse & Cooper (PWC) report, at https://www.strategyand.pwc.com/media/file/Strategyand_Rise-of-Generation-C.pdf, 2010

[20] S. Reay Atkinson, T. Tavakoli, A. Goodger, N. Caldwell, and L. Hossain, "The Need for Synthetic Standards in Managing Cyber Relationships," 3rd Int. Conf. on Soc. Eco-Informatics, Nov. 18-20. Lisbon: IARIA, 2013

[21] J. O. Grady, "Systems Requirements Analysis," London: Academic Press Elsevier, pp. 252-253, 2006

[22] G. Austin, "Australia rearmed! Future needs for cyber-enabled warfare," Discussion Paper No. 1, Aust. Centre for Cyber Sec., Uni. NSW, Canberra, at https://www.unsw.adfa.edu.au/australiancentre-for-cyber-security/news/australia-rearmed, 2016

[23] K. F. Joiner, "How Australia can catch up to U.S. cyber resilience by understanding that cyber survivability test and evaluation drives defense investment," Inf. Sec. J., vol. 26, pp. 74 - 84, 2017

[24] E. Copeland, T. Holzer, T. Eveleigh, and S.Sarkani, "The Effects of System Prototype Demonstrations on Weapon Systems," DefenseAR J., vol. 22(1), pp. 106–134, 2015

[25] K. F. Joiner, "How New Test and Evaluation Policy is Being Used to De-risk Project Approvals through Preview T&E," ITEA J., vol. 36, pp. 288-297, 2015

[26] Australian Senate. "Senate Inquiry into Defence Procurement," Canberra: Australian Parliament House, Ch. 2 & 12, 2012

[27] C. Wickens, J. Lee; Y. Liu, and S. Becker, "An Introduction to Human Factors Engineering," 2nd Ed. New York: Pearson Prentice Hall, 2014

[28] ISO/IEC, "ISO/IEC 38500:2015 Information technology -- Governance of IT for the organization,", available at https://www.iso.org/standard/62816.html , last accessed 8 Nov, 2018

[29] ISACA, "COBIT 5: A Business Framework for the Governance and Management of Enterprise IT," available from www.isaca.org, last accessed 7 Nov, 2018.

[30] L. Tjahjana, P. Dwyer, and M. Habib, "The Program Management Office Advantage: A powerful and centralised way for organisations to manage projects," American Management Assoc., New York, 2009

[31] K. Sandler and S. Gorman, "PMOs – Results from 4th Global PPM Survey of Sep. 14," Presentation to ProjectCHAT industry symposium, March, Sydney, at www.pwc.com/gx/en/consulting-services/portfolio-programme-management/assets/global-ppm-survey.pdf, 2015

[32] S. Dixon, "Everything You Wanted To Know About PMOs (in one presentation)," Association for Proj. Management, accessed from www.apm.org.uk on 27 Oct. 2015

[33] B. Robertson, "Right Projects, Right Way, Right Results: Building portfolio, program and project capability – The Australian Taxation Office journey," presentation at Proj. Gov. Controls Symp., Uni. NSW, Canberra, 7 May 2015

[34] B. Grey and P. Harrison, "Right Projects, Right Way, Right Results: Building portfolio, program and project capability – The Australian Taxation Office journey," presentation at ProjectCHAT industry conf., Sydney, 17 Mar. 2015

[35] K. F. Joiner, "Implementing the Defence First Principles Review: Two Key Opportunities to Achieve Best Practice in Capability Development," Canberra: Australian Strategic Policy Inst., Strategic Insights No. 102, at www.aspi.org.au, 2015

[36] K. Terrell, "Going the Extra Mile," keynote Proj. Gov. Controls Symp. 11th May, Uni. NSW Canberra, 2016

[37] D. Peever, R. Hill, P. Leahy, J. McDowell, and L. Tanner, "First Principles Review: Creating One Defence," DoD, Canberra. At https:// www.defence.gov.au/publications/reviews/firstprinciples/, 2015

[38] A. P. Henry, "Mastering the Cyber Security Skills Crisis: Realigning Educational Outcomes to Industry Requirements," discussion paper, Uni. NSW at http://dx.doi.orghttps://www.unsw.adfa.edu.au/ australian-centre-for-cyber-security/sites/accs/files/uploads/ACCS-Discussion-Paper-4-Web.pdf, 2017

[39] D. Lewis, "Cybersecurity skills shortage putting public, private sectors at risk, experts say," Australian Broadcasting Corporation news article, 13 June, at http://www.abc.net.au/news/2017-06-09/cybersecurity-skills-shortage-putting-australia-at-risk-expert/8601426G, 2017

[40] A. Ghildyal, "Realising Value through IT Governance: Issues and Solutions," Proj. Gov. Controls Symp., Uni. NSW, Canberra, 2-3 May 2017

[41] S. Wu, D. Straub, and T. Liang, "How information technology governance mechanisms and strategic alignment influence organizational performance: Insights from a matched survey of business and IT managers," Mis Quarterly, 2015

[42] J. Moavenzadeh, "The 4th Industrial Revolution: Reshaping the Future of Production," DHL Glob. Eng. & Manuf. Summit, Amsterdam, The Netherlands, 2015

[43] Porter and Heppelmann, "How smart, connected products are transforming companies," Harvard Bus. Review, 2015. vol. 93(10): pp. 96-114.

[44] K. Mohan, F. Ahlemann, and J. Braun, "Exploring the Constituents of Benefits Management: Identifying Factors Necessary for the Successful Realization of Value of Inf. Tech. in System Sciences," 47th Hawaii Int. IEEE Conf., 2014

[45] B. Flyvbjerg, "From Nobel prize to project management: getting risks right," Proj. Man. J., vol. 37(3), pp. 5-15, 2006

[46] J. Peppard, J. Ward, and E. Daniel, "Managing the Realization of Business Benefits from IT Investments," MIS Quarterly Executive, vol. 6(1), 2007

[47] O. Williamson, "Outsourcing: Transaction cost economics and supply chain management," J. Supply Chain Man., vol. 44(2), pp. 5-16, 2008

[48] A. Laing, (unpublished) "Review of High Assurance Testing," Chief Inf. Officers Group, Dep. of Defence, Canberra, 2018

[49] U.S. DoD Standard 5200.28, 1985.

[50] National Institute of Standards and Technology, "Common Criteria for Information Technology Security Evaluation," Version 2.0 / ISO IS 15408 (May 1998); Version 3.1 (Sep 2006–Apr 2017) ISO/IEC 15408:2005 and ISO/IEC 18045:2005 at https://www.commoncriteriaportal.org/

[51] U.S. DoD 8500.01E, October 24, 2002 at http://dodcio.defense.gov/Portals/0/Documents/DIEA/850001p.pdf

[52] U.K. MoD CESG CS3 - Cryptographic Standard - Implementation Standard For High Grade Products, Iss. 1.0, Aug 2012 (Unclassified Controlled unpublished)

[53] Aust. DoD, Defence Signals Directorate (DSD) High Assurance Standards - Cyber and Inf. Sec. Div., Nov. 2014 Version 1.0 (Unclassified Controlled unpublished)

[54] T. Vaidya, "2001-2013: Survey and Analysis of Major Cyberattacks," Comp. Sci. at arXiv:1507.06673v2, 2015

[55] Australian DoD, "Defence White Paper," esp. p. 50, pp. 81-82, available at www.defence.gov.au, 2016

[56] U.S. National Security Agency (NSA). Commercial Solutions for Classified Program (CSfC). https://www.nsa.gov/resources/everyone/csfc/ ,. accessed May 2018

[57] C. Alberts, J. Haller, C. Wallen, and C. Woody, "Assessing DoD System Acquisition Supply Chain Risk Management," CrossTalk, vol. 30(3), pp. 4-8, 2017

[58] S Chong, et al., "Report on the NSF Workshop on Formal Methods for Security," Cryptography & Sec. (cs.CR); Logic in Comp. Sci. (cs.LO) at https://arxiv.org/abs/1608.00678, 2016

[59] U.S. NIST, "Special Publication 800-161 Supply Chain Risk Management Practices for Federal Information Systems and Organizations," Apr. 2015, at https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-161.pdf

[60] W. Kramer, M. Sahinoglu, and D. Ang, " Increase Return on Investment of Software Development Life Cycle by Managing the Risk — A Case Study," Defense AR J., vol. 22(2), pp. 174–191, 2015

[61] IEEE Standards Association P7009 (under development) - Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems, available at http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html, 2018

[62] F. Nawaz, O. Hussain, N. Janjua, and E. Chang, "A proactive event-driven approach for dynamic QoS compliance in cloud of things" Proc. Int. Conf. Web Intel., pp. 971-975, ACM, 2017

[63] G. Banks, "Restoring Trust in Public Policy: What Role for the Public Service?," Australian J. of Public Admin., vol. 73(1), pp. 1–13, 2014

[64] P. Shergold, "Learning from Failure: Why large Government policy initiatives have gone so badly wrong in the past and how the chances of success in the future can be improved," Australian Public Service Commission, Canberra, 2015

[65] J. Antony, "Design of Experiments for Engineers and Scientists," London: Elsevier Ltd, 2014

[66] K. F. Joiner, "Six-Sigma Reform and Education in Australian Defence: Lessons-Learned Give Rigour and Efficiency to Ordnance, Aircraft and Ship Testing," Proc. 7th Int. Conf. Lean Six Sigma, Dubai, UAE, 7th – 8th May, 2018

[67] Troester, "National Cyber Range Overview," Pres. ITEA Cybersecurity Workshop, Belcamp MD, February, 2015

[68] S. Fowler, C. Sweetman, S. Ravindran, K. F. Joiner, and E. Sitnikova, "Developing cyber-security policies that penetrate Australian defence acquisitions," Australian Def. Force J., vol 202, July, 2017

[69] U.S. DoD, "Cybersecurity T&E Guidebook," Version 1.0, 1 July, 2015, available online in numerous locations.

[70] P. Nejib, D. Beyer, and E. Yakabovicz, "Systems Security Engineering: What Every System Engineer Needs to Know," 27th Annual INCOSE Int. Symp., Adelaide, July, 2017

[71] D. Scheul, "Force Integration – Integrated Capability Realisation for the ADF," Syst. Eng. Test & Eval. Conf., Sydney, 2 May 2018

[72] N. Smith, E. White, J. Ritschel, and A. Thal, "Counteracting Harmful Incentives in DoD Acquisition through Test and Evaluation and Oversight," ITEA J., vol. 37, pp. 218-226, 2016

[73] Tatsumi, K. 2013. "Combinatorial Testing in Japan," ICECCS 2013, 16 July, Singapore, Association of Software Test Engineering (ASTER) & Fujitsu Ltd

[74] D. Kuhn, R. Kacker, and Y. Lei, "Practical Combinatorial Testing," NIST Spec. Pub. 800-142, Oct., 2010

[75] D. Ahner, "Better buying power, developmental testing, and scientific test and analysis techniques," ITEA J., vol. 37, pp. 286-290, 2016

[76] D. Chu, "Statistics in Defense: A guardian at the gate," ITEA J., vol 37, pp. 284-285, 2016

[77] C. Brown, P. Christensen, J. McNeil, and L. Messerschmidt, "Using the developmental evaluation framework to right size cyber T&E test data and infrastructure requirements," ITEA J., vol. 36, pp. 26–34, 2015

[78] P. Christensen, "Cybersecurity Test and Evaluation: A Look Back, Some Lessons Learned, and a Look Forward!," ITEA J., vol 38(3), pp. 221–228, 2017

[79] Australian Senate, "Budget Hearings on Foreign Affairs Defence and Trade," Testimony by Vice Admiral Griggs, Major General Thompson and Minister of Defence, at http://parlview.aph.gov.au/mediaPlayer.php?videoID=399539&operation_mode=parlviews, circa 1900 hours, 29 May, 2018

[80] K. Joiner, E. Sitnikova, and M. Tutty, "Structuring defence cyber-survivability T&E to research best practice in cyber-resilient systems," Syst. Eng. Test Eval. Conf., Melbourne, May 2016

[81] Australian National Audit Office (ANAO), "Audit Report No 42 2016-17 – Cybersecurity Follow-up Audit," at https://www.anao.gov.au/sites/g/files/net4816/f/ANAO_Report_2016-2017_42.pdf, Mar. 2017

[82] N. Mackertich, P. Kraus, K. Mittlestaedt, B. Foley, D. Bardsley, K. Grimes, and M. Nolan, "IEEE/SEI Software Process Achievement Award 2016 Technical Report," Raytheon Integrated Defense Systems, Design for Six Sigma Team, March, 2017

[83] N. Mead and C. Woody, "Cyber Security Engineering: A Practitional Approach for Systems and Software Assurance," Pearson Education, 2017

[84] S. Reay-Atkinson, G. Tolhurst, and L. Hossain, "The Dichotomy of Decision Sciences in Information Assurance, Privacy, and Security Applications in Law and Joint Ventures," Int. J. Advances in Sec., vol 8(3-4), 2015