Sandia National Laboratories

The Digital Assurance for High Consequence Systems (DAHCS) Mission Campaign Whitepaper

Sandia National Labs DAHCS Mission Campaign

September 2024

U.S. DEPARTMENT OF ENERGY
National Nuclear Security Administration

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND2024-12664R

Developing the scientific foundation needed to create **rigorous**, **rapid**, **cost-effective**, **generalizable** digital assurance **across high consequence systems' lifecycles**

The Digital Assurance for High Consequence Systems (DAHCS) Mission Campaign (MC)

The DAHCS (pronounced "Dax") MC is a 7-year, $45 million research portfolio within Sandia's Laboratory Directed Research and Development program. The DAHCS MC arose in response to a great need: to ensure that the use of digital technologies does not weaken our nation's high consequence systems.

Digital technologies offer many benefits in speed, cost, and flexibility, and we seek to reap those benefits without introducing new system failures. However, digital technologies cannot be evaluated the same way as analog technologies. Initiatives across the nation highlight the capability gap that prevents efficient, effective digital assurance[1].

The Challenge Today's *digital assurance*[2] tools, techniques, and methods are inadequate to confidently characterize, assess, and manage digital risk; they are ad hoc, slow, costly, and rarely scalable to increasingly complex digital technologies. The rapidly evolving cyber threat landscape exacerbates this problem because digital assurance **now** must secure against digital risks **now and in the future**, including those introduced by rapidly evolving technologies, adversaries, and systems[3].

[1] For example, https://www.whitehouse.gov/wp-content/uploads/2024/02/Final-ONCD-Technical-Report.pdf, https://www.whitehouse.gov/briefing-room/statements-releases/2023/03/02/fact-sheet-biden-harris-administration-announces-national-cybersecurity-strategy/, https://www.cisa.gov/resources-tools/resources/secure-by-design, https://www.congress.gov/bill/117th-congress/house-bill/4346

[2] For our purposes, *digital assurance* includes processes, measures, and/or controls applied to digital technologies to ensure that a given system fulfills its intended purpose, even given current and future digital threats [NNSA SD 452.4-1 Nuclear Enterprise Assurance (NEA), 1/27/2022]. We include digital technologies both **within** and **influencing** HCS, and we include threats such as active adversaries, cyber attacks, supply-chain issues for components and tools, an insider, natural environmental hazards (both digital and physical), and both unintended behaviors (e.g., from errors) and emergent behaviors.

[3] Herkert J, Borenstein J, Miller K. The Boeing 737 MAX: Lessons for Engineering Ethics. Sci Eng Ethics. 2020 Dec;26(6):2957-2974. doi: 10.1007/s11948-020-00252-y. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7351545/

General-purpose digital assurance is a *wicked problem*. To address rampant discontinuity[4] across a vast discrete state space (which is a key characteristic of digital technologies), digital assurance capabilities must be situation- and system-specific. Today, we lack the scientific foundations to efficiently evaluate digital technologies with rigor and confidence.

For some systems, though, the cost of failure is catastrophic, e.g., death or existential threat to a nation. These *high consequence systems* (HCS) are created to serve very specific missions. In the DAHCS MC, we focus on four HCS types: nuclear deterrence, hypersonics, satellite, and individual critical infrastructure (e.g., nuclear power generators) systems. However, digital assurance capabilities for these HCS are insufficient for efficient, effective evaluation of risks from digital technologies.

DAHCS MC Strategy We aim to address this gap by bridging from foundational science research to engineering that supports Sandia missions. We invest in research needed to characterize, assess, and manage digital risk across HCS writ large. We focus on aspects of this digital assurance problem that are unique to HCS but applicable across many types of HCS – and specifically on discovering metrics and *principles* (abstractions, approaches, and assumptions) unique to assuring that embedded cyber-physical controllers do not *fail to function* (i.e., their *availability* and *reliability* is assured). With this focus, we seek to develop the scientific foundation needed to create **rigorous (including repeatable)**, **rapid**, **cost-effective**, **generalizable** digital assurance **across HCS lifecycles**, including in design, qualification, and sustainment.

DAHCS MC research should close the digital assurance capability gap for HCS by (1) identifying the boundaries within which we can confidently build and maintain HCS and (2) advancing state-of-the-art digital assurance tools, techniques, and methods across digital abstraction levels to support informed decision making about risk. By creating an integrated community of funded researchers, we aim to create and deliver a process and ecosystem that enables rigorous digital assurance at any point in a system's lifecycle. Ultimately, we aim to support decision makers, including systems designers, software developers, and program managers, in understanding systems-level implications of trade-offs against digital risk to HCS

[4] Billions of interconnected transistors within a device lead to more discrete digital states than particles in the observable universe and thus to overwhelming numbers of behaviors, and tiny perturbations (physical or electrical) can dramatically change *behaviors* in digital systems. *Behaviors* map inputs to outputs / effects over time.

missions. And, by shaping culture and building community (e.g., through our Multi-Institution Community of Practice, or MiCoP[5]), we aim to transform this domain from one driven by expert-dependent pockets of excellence — in techniques like red teaming, trusted hardware, risk assessment, secure system design, formal methods, human systems, vulnerability research, optimization, emulation, and modeling — into a sustainable, scalable, and rigorous discipline with enduring research communities and communities of practice that crosscut traditionally independent "cyber research" and "systems engineering" communities.

DAHCS MC Research Framework Our research roadmap (below) calls out eleven Research Challenges. Addressing these Research Challenges requires collaboration across community pockets of excellence, and it requires building towards a functional, generalizable ecosystem **with appropriate metrics** to characterize digital risk.

Scenario-Based Test & Evaluation (T&E) To measure MC progress and focus research, the DAHCS MC T&E team will use *assurance cases*[6] for identified HCS *proxy controllers* to address three scenarios (note: details are still slightly in flux):

A. *Rapid Reassessment*, providing, within two weeks, an updated assurance determination and proposed actions given a technical surprise (e.g., a new threat, a failed test)
B. *Rapid Build*, building, within six months, a new controller with requirements altered from a prior design but with as much digital assurance as possible within the timeframe
C. *100% Solution*, aiming to build, at whatever cost, an entirely cyber-secure, digitally assured controller (we assume this is impossible, but we aim for it)

Proxy Controllers: The T&E team will work with researchers to test developed tools, techniques, and methods on proxy controllers (the first proxy announced October 2024), and the T&E team may additionally test on hidden *validation proxies* to assess the progress of the DAHCS MC overall. Current proxies consist of a state machine application running natively on a microprocessor core on a simple system-on-chip (SoC); external communications include sensors, actuators, discrete input/output (I/O), and serial communications; and software and hardware are built using standard (stated) toolchains.

[5] dahcs-micop@sandia.gov
[6] D. J. Rinehart, J. C. Knight and J. Rowanhill, "Current Practices in Constructing and Evaluating Assurance Cases With Applications to Aviation," NASA, 2015.

Threat Models: The threat models focus on *failure to function*, or *availability/reliability*, and they include both a baseline case (i.e., no intelligent adversary, but unintended and emergent behaviors are of concern) and an advanced persistent threat (APT) case (i.e., an adversary with nation state resources and access of up to one *insider*, where an insider could be **human** or **digital**, e.g., a corrupted compiler, manufacturing defect, or overstressed part).

Research Roadmap We call for research to improve creating, evaluating, and using evidence in these assurance cases (which make *claims* about digital technologies both **within** and **directly influencing** the proxy controllers) within the following three Research Thrusts:

I. **Scalable Analysis**, to scale end-to-end DAHCS by at least two orders of magnitude in time/cost or complexity of handled technologies[7]. This includes **Assuring Physical Hardware** (i.e., claiming that the physical hardware presents the expected digital abstraction – hardware logic is covered in the next Research Challenge), **Behavior Coverage** (i.e., claiming that hardware logic, software, and component behaviors meet requirements), and **Force-multiplying Experts** (i.e., scaling the expertise and human judgment needed for DAHCS).
II. **Impact Analysis Amid Uncertainty**, to measure and increase confidence in an assurance case and its evidence, e.g., by identifying what additional information is needed to increase confidence by how much. This includes **Intelligent Adversary and Hazard Modeling** (e.g., explicitly accounting for adversary goals, choices, and capabilities), **Model Inference Given Partial Information** (i.e., overcoming obstacles to reasoning about a controller's implementation when relevant design or environment details are incomplete or unreliable), and **Failure Consequence Modeling** (i.e., enabling end-to-end reasoning about consequences of failures, including understanding *direct impacts* such as the impact of a single timing delay, understanding *aggregate failures* like bit flips caused by radiation in conjunction with a minor timing delay, and understanding *indirect impacts* such as the follow-on failures that arise from a single upstream failure).
III. **Integrating with Systems Engineering**, to support systems-level decisions about digital assurance, including comparing and making trade-offs between options. This includes **Digital Composition** (i.e., combining evidence across digital technologies as well as analysis techniques, abstraction levels, and

[7] This includes, e.g., designing for analysis.

processing contexts), **System Assurability Tradeoff Analysis** (i.e., directly comparing the impacts of implementation choices on digital assurance **as well as** other important characteristics like safety, reliability, resilience, size, weight, power, cost, or schedule), and **Evidence Communication for Decision Support** (i.e., supporting decision-makers with credible evidence about security and reliability, including characterizing factors that influence decision making).

Our other two Research Challenges call for **Revolutionary DAHCS** (i.e., approaches that provide end-to-end digital assurance of HCS and fall outside this roadmap) and **Targeted Evaluation** (focusing on rapid, proof-of-feasibility, or baselining of processes for our test controllers specifically).

In the DAHCS MC, vulnerability detection, IT systems, systems of systems, and existing algorithmic scaling research are out of scope unless pertaining to DAHCS principles.

DAHCS MC Outcomes We aim to build a foundation for digital assurance by delivering proof-of-concept capabilities that include hardware, software, mathematical frameworks, technologies, tools, techniques, theories, workflows, methodologies, metrics, processes, approaches, and methods. By striving for **scalability**, **generalizability**, **interoperability**, and **rigor**, research under the DAHCS MC will enable us to efficiently and effectively characterize, assess, and manage digital risk across many types of HCS.

In the DAHCS MC vision, digital technologies are designed and evaluated as easily as other important components of HCS; a robust research community continues to push the state of the art in DAHCS efforts; and decision makers can make confident, evidence-based statements about the digital assurance of high consequence systems in a timely manner because we can:

- Characterize the digital technologies within our systems at any point in their lifecycles
- Assess the risks to our systems from digital technologies and adversaries, moving well beyond vulnerability-focused security
- Select among design and implementation options that appropriately manage and accept digital risks while balancing against other trade-offs (e.g., resilience, safety, size, weight, power, cost)