The

ITEA Journal

September 2009
Volume 30, Number 3

Published quarterly
by the International
Test and Evaluation
Association

Integrating Test & Evaluation

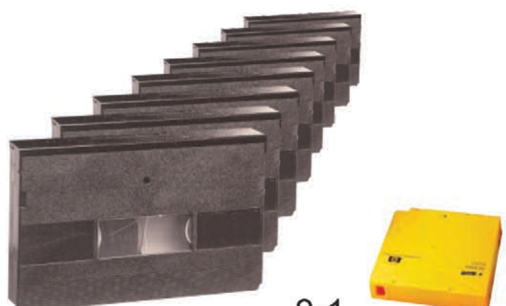
SYSTEMS ENGINEERING & INTEGRATION | ASSURANCE & RESILIENCE | INFORMATION TECHNOLOGY
Diverse stakeholder objectives.
Emerging technologies.
Tangible results.
Ready for what's next. A major transformation of our nation's digital infrastructure is underway. With this type of change comes complex cyber security challenges and you must be ready to meet these constantly evolving demands. Booz Allen Hamilton partners with government agencies, industry, and civil society to address their cyber challenges and help them achieve their goals. Our strategy and technology consultants build a common core of Test and Evaluation (T&E) competencies using a proprietary, strategy-focused T&E Boot Camp© program. We train a dedicated T&E workforce responsible for testing complex systems in air, land, sea, space, and cyberspace mission environments to help our clients transform organic T&E capabilities. Whether you're managing today's issues or looking beyond the horizon, count on us to help you be ready for what's next.
Ready for what's next. www.boozallen.com/rfwn
Booz | Allen | Hamilton
delivering results that endure

ec’trō·py, *n.*

a measure of a system to do useful work and grow more organized.

DESY Dub’ber, *n.*

a system to transfer 8 or more DCRsi® tapes onto a single LTO tape while preserving all data and search parameters of the original recording. Significantly lower maintenance costs, savings in storage space and extending life of valuable data.



8:1

CALCULEX

575-525-0131
Info@calculex.com
www.calculex.com

DESY™ Dubber

CONTENTS

TECHNICAL ARTICLES

DEPARTMENTS

ON THE COVER: The Global Power Bomber Integrated Developmental and Operational Test Team at Edwards Air Force Base, California, is in the final stages of integrated developmental testing and operational testing on the $1.35 billion B-2 Spirit Radar Modernization Program. Consisting of more than 193 personnel representing Air Force Materiel Command, the Air Force Operational Test and Evaluation Center, Air Combat Command, and Northrop Grumman, the B-2 integrated developmental test/operational test team provides seamless verification of B-2 upgrades. The cover image exemplifies the spirit of Integrated Testing. This issue addresses implementation and follows up the new Department of Defense policies to examine integrating contractor, developmental, operational, and live fire testing and evaluation throughout the system life cycle in a seamless continuum. (Cover graphic courtesy of Mr. Matt Guy of the Air Force Operational Test and Evaluation Center.)

■ ITEA Headquarters: 4400 Fair Lakes Court, Suite 104, Fairfax, Virginia 22033-3899; Tel: (703) 631-6220; Fax: (703) 631-6221, E-mail: *itea@itea.org;* Web site: *http://www.itea.org.*
■ ITEA is a not-for-profit international association founded in 1980 to further the development and exchange of technical information in the field of test and evaluation.
■ *The ITEA Journal* (ISSN 1054-0229) is published quarterly by the International Test and Evaluation Association at 4400 Lakes Court, Suite 104, Fairfax, Virginia 22033-3899. Single issue cover price for *The ITEA Journal* is $20. ITEA membership dues are $50 for individuals, $25 for full-time students, and $800 for corporations. Annual dues include a one-year subscription to *The ITEA Journal*. The annual subscription rate for libraries and other organizations providing timely reference material to groups is $60. All overseas mail (air mail or AOA) requires an additional $20. *The ITEA Journal* serves its readers as a forum for the presentation and discussion of issues related to test and evaluation. All articles reflect the individual views of the authors and not official points of view adopted by ITEA or the organizations with which the authors are affiliated.

POSTMASTER: Send address changes to: ITEA, 4400 Fair Lakes Court, Suite 104, Fairfax, Virginia 22033-3899.

ITEA Journal 2009; 30: 325

As I began to record my final input for the *ITEA Journal* as President, I reflect on the pleasure it has been to serve the organization in this capacity for the past two years. When I was first elected in 2007, the organization's primary goal was to move beyond Department of Defense (DoD) test and evaluation (T&E) as our primarily single constituent in membership and workshop/symposium participation. During my tenure, we continued to expand our themes and content to include increased participation from the international, non-DoD, and training communities. Our recent event in June, *Change Change Change for the 2020 Vision: The Test and Training Open Forum*, proved to be a testament to our expansion goals. I am pleased to report that this highly successful forum included a much higher ratio of speakers and participants from both the test and training communities than similar events in the past. A white paper outlining the forum's highly relevant results and recommendations is under development and will be published in an upcoming *ITEA Journal* article. I hope to see you all at our 2009 Annual Symposium, *Sharing Emerging Best Practices to Improve Test and Evaluation (T&E) to Meet Future Challenges* in Baltimore, September 28–October 1. The committee has put together an exciting program with many new format changes aligned to promote the newly approved mission, vision, and goals of ITEA.

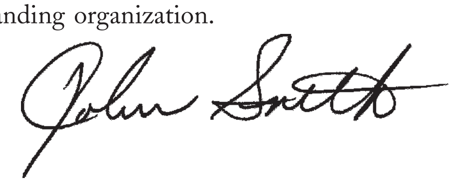
I am particularly excited about the direction we have taken to ensure that ITEA continues to meet the future needs of our expanding constituency. In June, I announced the approval of the ITEA Strategic Plan by the Board of Directors. The three-year plan is now available for download from the ITEA web site. The final step in this important endeavor was the development of a detailed Action Plan for implementing the following six goals outlined in the Strategic Plan:

- GOAL 1: Broaden participation within ITEA by the T&E community
- GOAL 2: Strengthen ITEA chapters
- GOAL 3: Improve the quality and relevance of educational courses
- GOAL 4: Increase opportunities for information exchange and networking through association events, publications, blogs, and other avenues
- GOAL 5: Increase opportunities to recognize accomplishments of T&E professionals
- GOAL 6: Ensure ITEA's governance is effective and efficient

The Action Plan includes three-year timelines and objectives, accountability measures and budget recommendations for ensuring accomplishment of the established mission and vision of the organization. The plan was approved at the June Board meeting and execution of the resulting 2009 actions and metrics are underway. Our way forward is an aggressive one and success relies solely on the hard work and determination of our volunteer members. I urge anyone interested in assisting in this important and rewarding process to contact the ITEA Executive Director, Ms. Lori Tremmel Freeman.

John Smith

I quoted Winston Churchill in my first *President's Corner* in December 2007, "We make a living by what we do, but we make a life by what we give." Nearly two years later, I remain in awe of what ITEA accomplishes by the sheer hard work and dedication of its volunteers. I want to take the time to thank those volunteers and the ITEA staff for their support during my tenure. I know that I called in a lot of favors from my colleagues over the past two years and am extremely grateful for the response and support that I ultimately received. I am proud to be a part of this important and viable community. I will remain on the Board of Directors in 2010 and look forward to furthering the ITEA vision and mission as I help in other capacities to implement the Strategic Plan. I look forward to working with all of you on this important endeavor as we all roll up our sleeves and collaborate to ensure that ITEA remains the premier professional association, meeting the needs of the international T&E community in the future. I thank you all for supporting me and allowing me the opportunity to serve as President of this outstanding organization.



ITEA Journal 2009; 30: 327–329

Test and Evaluation for the ''On-Demand'' World

Steven J. Hutchison, Ph.D.
Test and Evaluation Executive,
Defense Information Systems Agency, Arlington, Virginia

Defense Information Systems Agency (DISA), is a Combat Support Agency dedicated to providing our nation's leaders and warfighters the best information technology solutions needed to accomplish the mission. We provision, operate, and secure the Global Information Grid (GIG) in support of net-centric operations. DISA is also an acquisition organization with a large test and evaluation (T&E) workforce that specializes in T&E for information technologies (IT).

Steven J. Hutchison, Ph.D.

Our vision in DISA T&E is to be a center of excellence for advancing the art and science of T&E for IT. Toward that vision, we have established four goals:

- Goal 1: Support the warfighter—provide responsive, relevant, value-added test and evaluation services.
- Goal 2: Be the premier tester and evaluator of Joint net-centric warfighting capabilities in the Department of Defense (DoD).
- Goal 3: Team for maximum efficiency. Empower our teams to design and execute robust test plans and provide objective assessments to decision makers.
- Goal 4: Develop and retain a highly qualified workforce.

Bottom line—My job is to help our programs field enhanced IT capabilities to our warfighters. I do that by ensuring that our T&E organizations are engaged with the programs to help them find and fix problems early, stay objective in their assessments, and invest in the right technologies that will keep us relevant in the fast-paced, ever-changing, "on-demand" world of IT.

In DISA, we are trying to shift acquisition and testing of IT to be more responsive to warfighters needs—our first goal. The shift that I'm addressing is highlighted in the current strategic planning guidance (SPG): "Evolve planning, determination of needed capabilities, acquisition, test and evaluation, programming and training toward 'demand driven' processes focused increasingly on developing and testing joint capability portfolios to meet Combatant Commander preferred approaches." Demand-driven processes, Joint capability portfolios, and combatant commander approaches all signal change …and remind us of who we work for—the warfighter.

We are expending a lot of energy and intellectual capital on finding ways to improve our acquisition and T&E processes. Many of you are aware that DISA is the sponsor of a National Academies of Sciences study on improving processes for acquiring and testing IT in the DoD. We should have their recommendations in November. We've also engaged the Business Executives for National Security (BENS) and the National Defense Industrial Association (NDIA) to hear their recommendations. Recently, a Defense Science Board completed a study on IT acquisition and made recommendations to institute a new acquisition model geared specifically to IT. One common theme emerging from these studies is that the DoD should adopt agile development and test practices. Nowhere is the "on-demand" requirement more prevalent than in IT—whether it's bandwidth, spectrum, storage capacity, intrusion protection, or computing, we expect our IT capabilities to provide what we need, when we need them. To support the warfighter with relevant IT today requires us to shift how we do business to support test on demand.

DISA is an acquisition organization for Joint capabilities, likewise T&E of Joint capabilities is an essential part of our mission in DISA; hence, our second goal. Our test organizations provide full spectrum T&E services (developmental, operational, interoperability, and security T&E) across all of the Joint capability areas. It is imperative that we get this right. The SPG speaks to "developing and testing joint capability portfolios," and I would like to emphasize 'joint' in that statement because, simply put, the processes we have in place today are not designed to support Joint programs. To accomplish the guidance, we are going to need to formalize processes for obtaining Joint troops as test units, approved Joint mission threads, and relevant Joint scenarios. We are working with Joint Forces Command to improve these processes for Joint testers.

In the world of IT, agility is the name of the game. It frustrates me when we have capability improvements ready to go but can't get them fielded. I believe this has led our warfighting units to circumvent the acquisition process and buy commercial capabilities. Think about that, because it is its own form of acquisition process: the commanders in the field weigh the risks and make a decision—speed is important, costs low, risks manageable, capabilities improve. Our DoD acquisition process should do the same thing, and our T&E organizations will have to be better, faster, and more mission-focused than ever before. It is essential that we combine efforts and work as a team to streamline test processes and eliminate redundancy. Our third goal is to team for maximum efficiency, and empower the capability test teams to design, execute, and report results of T&E activities.

The commercial sector succeeds with agile methods due to teamwork and empowerment—the opposite of our DoD acquisition process today! Fortunately, our leaders have recognized the value of teamwork in achieving greater efficiency. Almost 2 years ago, in December 2007, the Director, Operational Test and Evaluation, the Honorable Dr. Charles McQueary, and then Undersecretary of Defense for Acquisition, Technology, and Logistics, the Honorable John J. Young, Jr., signed into policy that T&E "shall be integrated and seamless throughout the system lifecycle." Integrated testing, the theme of this issue of *The ITEA Journal*, is all about teamwork.

In the IT world, we have a lot of decision makers to satisfy to get a capability fielded; therefore, our test teams should be composed of members representing all of their interests. There's a milestone decision authority, an interoperability certifier, and a designated approving authority who certifies information assurance. A typical integrated master schedule will have developmental test events, a security test event, and an operational test event, all conducted by different organizations and spanning a period of several months. An integrated test model for IT could bring all test activities together and cut that time in half or more. I think this will yield considerable cost savings, but more important, if testers work as a team, we can produce a more complete evaluation to give the decision makers.

Integrated testing is essential to the transformation of T&E to the "on demand" world. For programs following agile processes, T&E has to be on demand. It means being able to assemble the right team of skilled testers, translate user stories into an effective test design, use test automation as much as possible, analyze the data, and share the results—all in a period of days, not weeks or months. It is essential that our testers understand the Joint environment and be able to assemble the right components quickly and at low cost. To test capabilities such as those DISA is developing, we need an on-demand Joint test environment with the means to find the right set of test assets, schedule their use, collect data, and then release the asset when completed. Those who oversee these agile programs have to adjust their processes accordingly, which means that governance bodies such as the overarching integrated process teams (OIPTs) have to be able to work within the agile framework.

Last, our fourth goal is the development and retention of the workforce. The SPG highlights training as part of the evolution toward demand-driven processes. We have a professional acquisition workforce and a quality acquisition university to support them. However, IT acquisition is different; we should not expect our program managers and testers to be proficient in IT acquisition and T&E when we train them through a program of instruction based for the most part on weapons platforms. As we evolve training, let's address the education of our IT acquisition and T&E workforce. This is one of many initiatives DISA is pursuing.

The challenges for the T&E community are great but not insurmountable. Our job is to set the conditions for success. At DISA we're taking on all of these challenges, but we know that we don't have all of the answers. We need your help to get the good ideas turned into action. Let's work together to develop the processes, methods, training, and infrastructure to ensure we "test like we fight." We look forward to working with all of you to transform acquisition and testing to the on-demand world. ❑

STEVEN J. HUTCHISON, Ph.D., a member of the Senior Executive Service, is the Test and Evaluation Executive for the Defense Information Systems Agency. He is responsible for developing and enforcing T&E policy and procedures, representing DISA to the DoD T&E community, and providing direct support to DISA programs for T&E of IT capabilities. Prior to his arrival at DISA, Dr. Hutchison served in the office of the Director, Operational Test and Evaluation (DOT&E), Office of the Secretary of Defense, as a net-centric warfare systems analyst. While in DOT&E, he had oversight responsibilities for several of the major warfighting

information systems, including the Global Command and Control System–Joint, the Service variants of the Distributed Common Ground/Surface System, and the Net Enabled Command Capability. Dr. Hutchison also served for 6 years in the Army Test and Evaluation Command as an evaluation officer and later as the Assistant Technical Director. E-mail: Steven.Hutchison@disa.mil

ITEA Journal 2009; 30: 331–332

Implementing Integrated Testing

Christopher DiPetto
Office of the Under Secretary of Defense, Acting Director, Developmental Test & Evaluation, Arlington, Virginia

Current Department of Defense (DoD) acquisition policy mandates the use of integrated testing. The policy not only makes economic sense but also has the potential to reduce risk, as early, integrated testing often involves more realistic operational scenarios than traditional developmental testing and therefore allows earlier discovery of operational failure modes. As more programs have attempted to implement the policy, however, they have encountered obstacles that have prevented them from fully realizing the benefits of integrated testing. Issues that present difficulty in integrated testing fall into three principal areas: sharing and access to data; shared control of test events; and overreaction by some observers to the test results. I believe the real obstacles to fully implementing integrated testing are mostly cultural and can be overcome with appropriate action by acquisition leaders.


Christopher DiPetto

DoD policy memos and guidance documents define what we mean by "integrated testing." The Defense Acquisition Guidebook, Test and Evaluation (T&E) chapter (chapter 9) provides the formal definition and additional detail. The definition focuses on collaborative planning and execution of tests to provide a shared or common data set for independent evaluations and reporting. It is important to note that the definition is not "*integrated test and evaluation*" but "*integrated testing*." Although the testing is planned and executed collaboratively by the contractor, government Developmental Test (DT) and Operational Test (OT) communities, the evaluations are performed independently to fulfill respective roles and missions.

The challenges regarding sharing and access to data seem to be associated largely with ensuring the pedigree of the data and proprietary issues with contractor data. As defined, integrated testing includes contractor testing and can result in claims of proprietary data rights. In order to share the data from contractor events, provisions for data access must be included in program contracts. The converse is also true if contractors are expected or allowed to use data collected during government test events. Discussing data access issues up front, before the contract is signed, can set the expectation and allow for an equitable arrangement for obtaining access to contractor data. In order to assist programs with these types of T&E contractual matters, my office has produced a guide, "Incorporating Test and Evaluation into Department of Defense Acquisition Contracts" (http://www.acq.osd.mil/sse/pg/guidance.html).

Ensuring the pedigree of the data refers to understanding the configuration of the test asset and the actual test conditions under which each piece of data was obtained. While the primary purpose of integrated testing is to increase the value and efficiency of test events, the practice of sharing data also could result in a reduction in the acquisition timeline if we use shared data to satisfy multiple objectives. By infusing operationally relevant profiles and a mission perspective during integrated testing, and establishing and maintaining the data pedigree, much of the data needed by the Operational Test Agency (OTA) could be obtained before Initial Operational Test and Evaluation (IOT&E). If the data collected during integrated testing is adequate, it could serve to shorten the dedicated OT&E phase.

The issue with shared control of test events appears to be acceptance and use of data captured from test events outside the evaluator's sole control. It is easy to have confidence in data from a test you controlled. If tests are planned and conducted collaboratively, though, all stakeholders (both DT and OT) have control of the event, and all should be able to accept the data from the tests. The independence of the separate evaluation is not compromised by the fact that the source of the data was an integrated event.

Another issue is the potential overreaction by observers to test results. Many program managers view problems discovered during testing as bad news. Most understand that T&E results are important in maturing the system design through the systems engineering process, but reports of problems discovered in early testing could be misinterpreted by outside observers. The potential for these misunderstandings creates a disincentive to stress the system early. It creates an incentive to perform DT in tightly scripted scenarios in order to demonstrate successful system operation. Unfortunately, this conduct merely postpones the discovery of operational failure modes, resulting in expensive rework and causing disruption to program schedules. We need the acquisition and user communities to realize that the product of T&E is knowledge about the system's capabilities and/or limitations, not problems. Testers and evaluators must develop knowledge that is relevant and timely for the decisions being made, and report results in a mission-oriented context. Ultimately, we need comprehensive knowledge from T&E results to assist in managing risks and better decision making.

Integrated testing holds a promise of greater testing efficiencies and improving the quality of the information provided to the decision makers. The challenge to the T&E community is to implement robust integrated testing and change the culture to fully realize the benefits to the acquisition process and ultimately to the warfighters. ❑

CHRISTOPHER DIPETTO currently serves as the acting director, Developmental Test and Evaluation (DT&E) within the Office of the Secretary of Defense (OSD) with responsibility for oversight of 180+ major weapons systems developmental test and evaluation (T&E) programs and as the functional advisor for the T&E acquisition career field. Mr. DiPetto is the principal advisor to the Secretary of Defense and the Under Secretary of Defense (AT&L) on DT&E. He has over 29 years of experience in the Department of Defense with extensive experience in congressional matters, energy policy, acquisition, and T&E. Previous jobs in OSD included assignments as staff specialist for naval weapons and for strategic and tactical airlift programs. Mr. DiPetto served with the Naval Air Systems Command as a flight test manager, planning and executing Tomahawk flight tests. He began his career at the Navy's Pacific Missile Test Center as flight test engineer for surface and subsurface launched missile systems.

He earned a bachelor of science from the New York Institute of Technology in 1979. A graduate of the U.S. Naval Test Pilot School, he logged hours in a variety of U.S. Navy fixed and rotary wing aircraft. He is a 1994 graduate of the Harvard Kennedy School of Government, Senior Executive Fellows program; a 1993 graduate of the Defense Systems Management College's Program Management Course; and a Level III certified acquisition and T&E professional. E-mail: gordon.tillery.ctr@osd.mil

ITEA Journal 2009; 30: 333–338

Hardware-in-the-Loop Testing of Wireless Sensor Networks

Kenneth G. LeSueur
Subsystem Test and Analysis Branch,
U.S. Army Redstone Technical Test Center, Huntsville, Alabama
Emil Jovanov, Ph.D.
Electrical and Computer Engineering Department,
The University of Alabama in Huntsville, Huntsville, Alabama

The addition of network interfaces and fusion algorithms to sensor systems results in an increase in the complexity of the required test and evaluation technologies and methods. Existing testing capabilities are not adequate for testing of tactical networked sensor hardware in complex battlefield configurations. We are developing a system for testing the wireless sensor networks that support tactical hardware interacting in real time with emulated network nodes in an augmented reality test scenario. The test bed provides a realistic simulated representation of a tactical network that allows faithful testing of networked systems focusing on hardware-in-the-loop testing of sensors and sensor fusion systems. Systems can be tested using this method in a controlled, repeatable environment not feasible in field testing. The system design combines dedicated high performance computing resources with a scalable, high fidelity network emulation and a computer generated forces model to virtually represent the tactical network, force movement, interactions, and communication loads to systems under test. This article presents the test bed design framework, preliminary performance results, and a concept for determining the requirement and performance envelope for test bed utilization.

Key words: Real-time network emulation; wireless tactical sensor test bed; high performance computing; hardware-in-the-loop.

Realistic testing of tactical wireless sensor networks requires enhanced technologies and techniques utilizing real-time hardware-in-the-loop (HWIL) test methods. A mixture of live and simulated network nodes operating in real time and immersed in an augmented virtual environment is the optimal approach to obtaining highly accurate test data with the ability to scale the network size to tactical force levels. Performing a fully live field test is impractical because of the size and variations of the terrain required and the sheer number of tactical network nodes needed to completely represent the full force structure and equipment (i.e., networks, radios, sensors, and weaponry). The HWIL test bed approach described in this article combines dedicated high performance computing (HPC) resources with a scalable, high fidelity network simulation and a computer generated forces (CGF) model to virtually represent the tactical network, force movement, interactions, and communication loads to systems under test. The network emulation and CGF models are required to interoperate and scale to the size of an Army brigade combat team that will have thousands of network nodes. The use of this test method allows testers to interface a small number of real hardware nodes with virtual components to produce an operationally realistic environment.

HWIL testing background

Sensor and missile systems have historically relied on HWIL testing methods to determine many aspects of networked system under test (NSUT) performance. The technique of HWIL testing is based on using actual tactical hardware and software interfaced to a suite of stimulus and measurement systems including modeling and simulation tools. Laboratory HWIL test methods are considered the highest fidelity alternative to live field testing because of the inclusion of the actual hardware and software in the test, as opposed to pure simulation or mathematical models (Almendinger and LeSueur, 2007). To realize the benefit of HWIL, we must ensure that the NSUT interacts with a virtual environment that completely replicates the real world if the NSUT is to perceive and respond as it would in an actual fielded situation.

In-band stimulation of sensor and missile systems with visible, infrared, acoustic, and seismic sensors has become common place in the test and evaluation (T&E) community. Other forms of stimulation of the NSUT to complete the virtual environment include physical motion (pitch, yaw, and roll of the platform), launcher or vehicle electric interfaces, and Global Positioning System (GPS) to name a few.

An example of this type of HWIL capability is the Advanced Multispectral Simulation Test Acceptance Resource (AMSTAR) (Almendinger and LeSueur 2007). The AMSTAR is unique in the ability to perform real-time multispectral scene generation and projection into a common missile seeker aperture mounted on a multiaxis flight motion simulator. The AMSTAR is equipped with a dynamic infrared scene projector, millimeter wave projection, and a semiactive laser return simulator. The three projected beams are combined to provide simultaneous in-band stimulation into a single sensor aperture *(Figure 1)*.

With the increase in complexity of sensors systems that include network interfaces, the HWIL test methods must be enhanced. The network addition to the test item is another aperture whose input must be properly stimulated and output accurately measured. Often other networked systems receive and respond to sensor traffic and changing conditions in the network environment. Therefore, simply playing back recorded network traffic cannot be used to comprehensively test the functionality of a NSUT.

The real-time sharing of information from multiple sensors across a wireless network coupled with the use of sensor fusion creates a "system of systems" where the combined performance has the potential to be greater than the sum of the individual system capabilities. The T&E of the system of systems must take into account the performance difference when sensors share information such as detections, identifications, moving object tracks, photos, and live video across the sensor network. The performance of the system of systems depends greatly on the performance of the network linking the individual nodes.

The software emulation approach to network testing is more cost effective, scalable, and adaptable than hardware emulation (Werner-Allen, Swieskowski, and Welsh 2009). The network simulation provides the background traffic present in tactical situations and transports sensor data to consuming applications, with

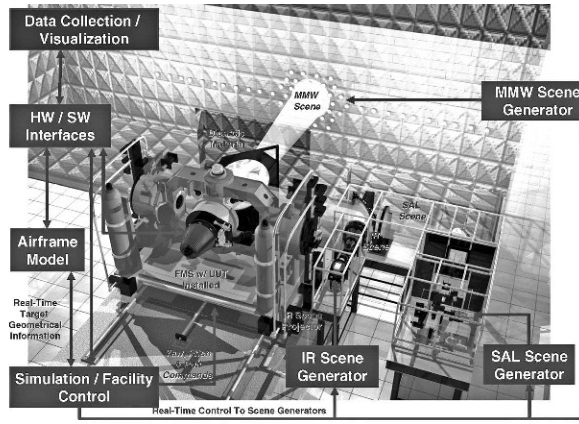


Figure 1. HWIL to simulation interface.

realistic representation of radio frequency propagation and terrain effects, delivery time, packet loss, collisions, and bandwidth availability. By providing interfaces to actual tactical hardware, the wireless sensors can be tested in a reliable and repeatable way not available through typical field level test methods.

Proposed testing method

The core of the wireless tactical sensor test bed is the high fidelity, real-time, network simulation made possible by a parallel computing platform upon which the simulation runs. The system is implemented using EXata network emulation and OneSAF CGF running on a parallel Linux blade system. The test system integrates sensors to the simulated virtual environment through wireless gateways and Ethernet connections.

EXata is a wireless network emulator that connects to live networks and supports real-time operations (SNT 2009). EXata creates a simulated network that interfaces with real networks allowing for software, hardware, and human-in-the-loop test applications to communicate over all layers of the network.

Network emulation with HWIL interfaces

The need for having real wireless sensor hardware connected with the simulated network drives the requirement for operating in real time. As the size and complexity of the modeled network increases, the test bed computing resources must continue to perform all required calculations in real time to allow system testing (Hamida, Chelius, and Gorce 2008). When real-time performance cannot be maintained, the test and simulation results are considered invalid and adjustments to the test bed, such as the size of the simulated environment or simulation fidelity, must be performed until real-time performance can be reliably maintained.

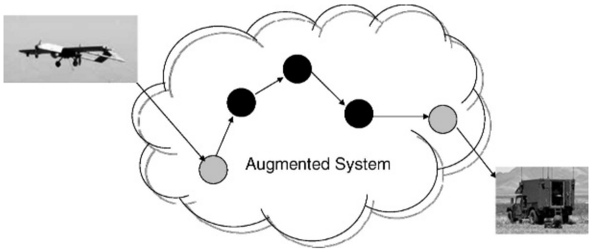


Figure 2. HWIL to simulation interface.

Each HWIL device under test will have a corresponding virtual node in the network simulation engine. This connection approach is illustrated in *Figure 2* where the example of a HWIL network node is an unmanned aerial system (UAS) supplying video to a network of simulated nodes, and a second HWIL command system is receiving the UAS video. The light-colored nodes correspond to HWIL items in the real environment and provide the interface to the emulated network represented by the black nodes. The video quality and latency are affected by operations of the emulated nodes just as they would be tactically.

Parallel processing of simulation environment

Network emulation allows interfacing and testing of a few samples of real hardware with virtual components to produce operationally realistic numbers of network nodes. The Army brigade combat team is the target size for the development of the tactical wireless sensor network test bed. The brigade combat team will have thousands of heterogeneous networked nodes with a wide range of processing power and network bandwidth requirements. Both the network simulation code and the CGF model are required to scale to this magnitude.

The system architecture with the major interfaces between the HPC and the HWIL interfaces is presented in *Figure 3*. The HWIL interface supports operation of different wireless network configurations including wireless sensors, network missile systems, vehicles, and UAS.

EXata can operate in a shared memory or message passing interface parallel processing environment. There are manual and automatic parallel workload distribution methods. The manual method allows the simulation operator to assign the workload associated with simulated nodes to computational cores as desired. The automatic method assigns network nodes to computational cores in sequential order as the emulated nodes are defined in the environment. After all cores have one node assigned, the process repeats starting from the first core until all of the emulated nodes have been assigned.

Test bed performance results

The test bed HPC system is scheduled to integrate at the Redstone Technical Test Center in early fall 2009. To establish performance expectations of the completed test bed, we executed initial test cases on an existing parallel computing platform with an early release of EXata 2.0. For these preliminary test cases, the

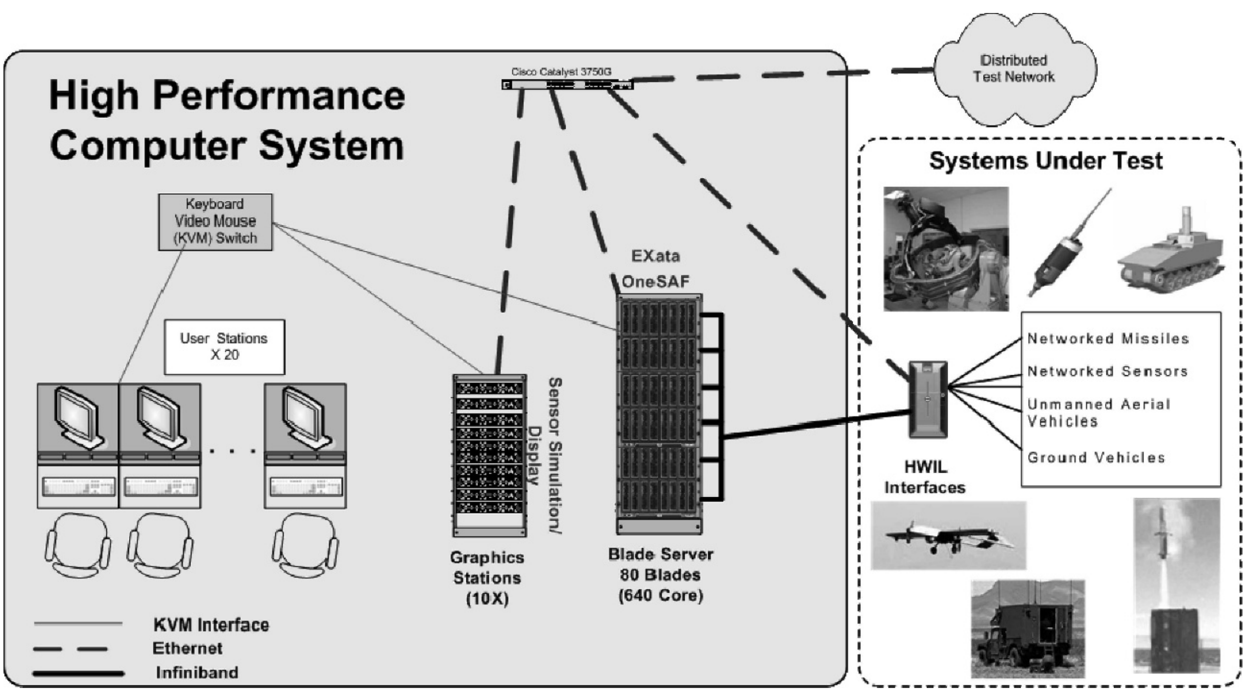


Figure 3. System interface architecture.

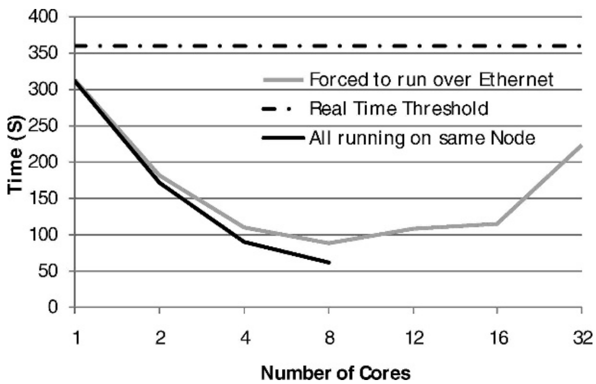


Figure 4. Parallel execution time for scenario 1.

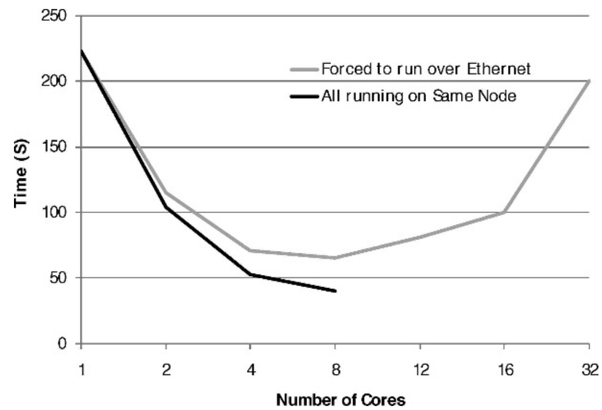


Figure 5. Parallel execution time for scenario 2.

simulation is executed on a Silicon Graphics Inc. (Freemont, California) Altix XE 320 cluster with gigabit Ethernet connections. One connection is designated for administrative function, and another one is for data transfers. The cluster nodes consisted of two Intel Xeon 2.5 GHz E5420 four-quad-core processors (total of eight cores/node) and 8 GB of RAM per node. The system has a total of 32 processors or 128 cores.

Experimentation scenario description

Two existing representative scenarios were chosen to be used in the test bed evaluation tests. The two scenarios were selected based on differences in the simulated node count and the complexity and fidelity of the environment model. The first scenario has a node count of 500 radios. Each radio is mobile and travels according to a group mobility model that limits the movement of nodes to an area around an established group location. The propagation path loss model used was the irregular terrain model, and an associated terrain file was loaded for use by the irregular terrain model.

The second test scenario has 800 radios simulated and 68% of the nodes are mobile with their movement defined by an input mobility file. A two-ray propagation path loss model is used in this scenario.

The two test scenarios were selected because of their size, variations in computational load, and tactical relevance. The second scenario has a larger node count but uses a more simplistic two-ray propagation model. Because only a subset of the nodes is moving, the number of path loss calculations that need to be performed during the test is limited.

Preliminary results

The first scenario was executed and the simulation run time was measured for a section of the scenario *(Figure 4)*. The real-time threshold for this scenario section was 360 seconds. The simulation run time is plotted with the number of computer cores in *Figure 3*. The gray line shows the measured performance from 1 to 32 processing cores when the interprocessor communications occur over the Ethernet interconnect. The black line shows the measured performance when the computational node is allowed to use a shared memory interface. Note that this data stops at eight cores, the number of cores on a node. There is slight improvement in performance when using the shared memory communication versus the Ethernet.

The test results from the second scenario are shown in *Figure 5* in a similar format. In both scenarios, the optimum performance occurs when operating on eight computer cores for the given size, scenario complexity, and modeled environment fidelity.

The parallel efficiency of the two scenarios using eight processing cores and communicating across the Ethernet interconnection is:

$$Scenario\_1\_Efficiency = \frac{\frac{313\ \text{s}}{88\ \text{s}}}{8\,Cores} * 100 = 44.5\%$$

$$Scenario\_2\_Efficiency = \frac{\frac{223\ \text{s}}{65\ \text{s}}}{8\,Cores} * 100 = 42.9\%$$

Concept of a performance and requirement envelope

As seen in the preliminary results section, the real-time performance of network emulation capabilities is a multidimensional problem that includes the size (number of nodes) of the scenario, the amount of computing resources available, and the fidelity or complexity of the simulation environment. There is a need to establish a wireless sensor network performance or requirement envelope to aid in NSUT test planning and resource allocation. This performance or

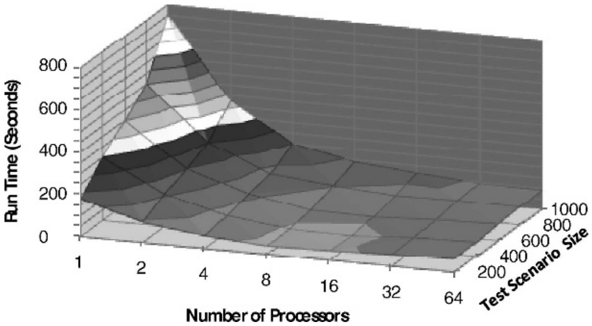


Figure 6. Test bed performance or requirement envelope.

requirement envelope can be established through the collection of a set of empirical test data and the development of a functional model of the test bed.

Figure 6 shows a three-dimensional plot from a notional test bed configuration and set of scenarios. The notional test scenario real-time threshold is 100 seconds. These data are used only to describe the utility in establishing the envelope.

For the notional case, the fidelity or complexity of the scenario is held constant while the number of radio nodes and number of test bed processors are changed. The resulting simulation runtime is ploted on the vertical axis. Once the envelope is developed, test configurations can be determined based on several limitations or requirements. For example, *Figure 7a* shows a notional case where the number of computation cores is limited to four (potentially a field test where access to an HPC is not available). A plane is drawn across the envelope corresponding to this test limitation.

Figure 7b shows an associated graph where the processor count is limited to four and the scenario size is plotted against the simulation run time. The point where the runtime crosses the 100-second real-time threshold establishes that a maximum of 500 nodes, at the given fidelity or complexity, can be emulated on a four-processor machine while maintaining real-time performance.

Another use of the performance or requirement envelope is shown in *Figure 8a*. In this case the test application demands a scenario that simulates 800 radios. A plane is drawn to show where this requirement intersects with the envelope.

Figure 8b shows the companion graph where the scenario size is set to 800 and the simulaiton run time is plotted against the number of processing cores. The point where the runtime crosses the real-time threshold (100 seconds) shows that 16 or more processors are necessary to maintain real-time performance for the 800-node scenario.

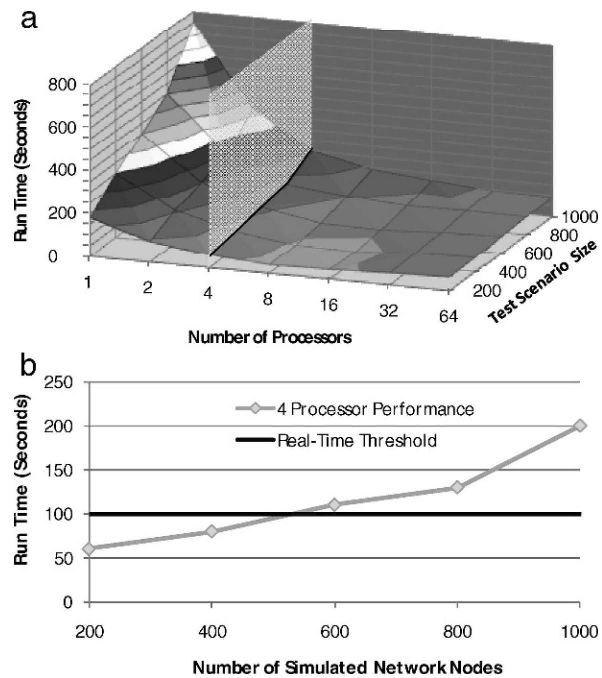


Figure 7. (a) Performance envelope with four-processor plane. (b) Performance plot using four cores.

A different three-dimensional graph is necessary for each of the various levels of simulation fidelity and scenario complexity desired for a set of test applications. The development of a test bed performance model validated with samples of empirical test data will

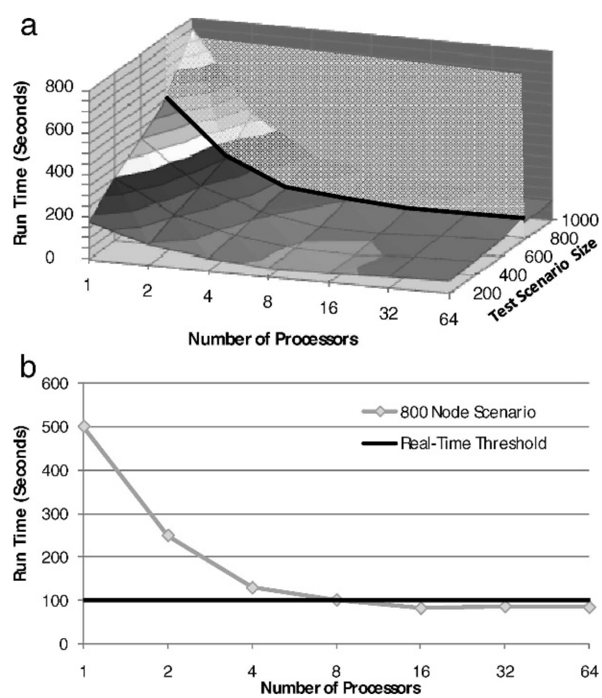


Figure 8. (a) Performance envelope with 800-node scenario plane. (b) Performance plot for 800-node scenario.

allow the complete suite of graphs to be generated. The performance or requirement envelope provides an estimation tool useful for many functions including test planning, experiment design, and resource allocation.

Conclusion

Current test methods are not adequate for testing tactical wireless network hardware in realistic battlefield environments (LeSueur and Jovanov 2009). The use of an augmented wireless sensor test bed operating in a real-time HWIL configuration is a viable solution to the T&E challenges associated with tactical networked sensors operating in complex battlefield configurations. It is established that real-time performance improvements can be realized when operating on parallel computer cores for a given test scenario. Through this evaluation, it is determined that general performance thresholds can be measured, but the results are highly scenario dependent. A performance or requirement envelope is needed to accurately predict wireless sensor test bed performance based on multidimensional setup parameters.

Follow-on research and testing are needed in several areas to fully realize the benefits of the test bed.

- Future performance enhancements will be realized when the simulation engine is transferred to the new HPC cluster with Infiniband interconnections. The new system will have increased computing performance, and the interconnect architecture will have lower latency.
- Interfaces to the CGF model must be completed to provide more realistic platform movements and tactically appropriate network loads.
- Development and validation of a test bed performance model is needed to aid in the generation of the performance or requirement envelope.

The implementation of the tactical wireless sensor network test bed enhances the test and analysis of system performance in a realistic real-time, high-fidelity simulated environment not achievable through standard test processes. The test bed allows the community to evaluate large tactical NSUT performance parameters such as throughput, latency, jitter, dropped packets, message completion rate, channel interference, jamming, bottlenecks, power consumption, and reliability just to name a few. The primary advantage of this architecture is the inclusion of live hardware in the test, which will be immersed in an augmented environment that allows the item under test to perceive and respond to stimulus just as it would in the real world. Each layer of the network can be tested because of the high fidelity simulation made possible by utilizing parallel processing to maintain real-time performance. ❑

Kenneth G. LeSueur serves as the chief technologist in the Subsystem Test and Analysis Branch at Redstone Technical Test Center. His work and research have been concentrated in HWIL testing, distributed testing, modeling and simulation, and high performance computing. He received his master's degree in computer engineering at the University of Alabama in Huntsville and is currently working on his doctoral dissertation. E-mail: ken.lesueur@us.army.mil

Emil Jovanov, Ph.D., is an associate professor in the Electrical and Computer Engineering Department at the University of Alabama in Huntsville, with more than 25 years of experience in the design of application specific hardware and software system. He is recognized as the originator of the concept of wireless body area networks for health monitoring and one of the leaders in the field of wearable health monitoring. Dr. Jovanov is a senior member of IEEE, associate editor of the IEEE Transactions on Information Technology in Biomedicine *and* IEEE Transactions on Biomedical Circuits and Systems, *member of IEEE EMBS Technical Committee on Wearable Biomedical Sensors and Systems, and serves on the editorial board of Applied Psychophysiology and Biofeedback. E-mail: Emil.jovanov@uah.edu*

References

Almendinger, F., and K. LeSueur. 2007. Design tradeoffs in the development of the advanced multi-spectral simulation test acceptance resource (AMSTAR) HWIL facilities. In *Proceedings of the SPIE* 6544, Orlando, FL: SPIE Conference Proceedings, April.

Hamida, B., G. Chelius, and J. Gorce. 2008. Scalable versus accurate physical layer modeling in wireless network simulations. In *22nd Workshop on Principles of Advanced and Distributed Simulation*. Rome, Italy: IEEE Computer Society (ACM Portal), pp. 127–134, June.

LeSueur, K. G., and E. Jovanov. 2009. Performance analysis of the augmented wireless sensor network testbed. *41st Southeastern Symposium on System Theory*. Tullahoma, Tennesee: University of Tennessee Space Institute, pp. 319–323.

SNT (Scalable Networks Technologies). March 16, 2009. SNT Knowledge Base. Exata. http://www.scalable-networks.com/pdf/exata.pdf (accessed June 2009).

Werner-Allen, G., P. Swieskowski, and M. Welsh. 2005. MoteLab: a wireless sensor network testbed. In *IPSN 2005 Fourth International Symposium on Information Processing in Sensor Networks*. Los Angeles, CA: IEEE, pp. 483–488, April.

ITEA Journal 2009; 30: 339–344

Turning in the Wind: Frank W. Caldwell and the Variable-Pitch Propeller

Jeremy R. Kinney
National Air and Space Museum,
Smithsonian Institution, Washington, D.C.

Editor's Note: Turning in the Wind *is the first of a two-part series on Frank Walker Caldwell, a pioneer of American propeller technology. This article outlines his overall achievements; the second will focus on Caldwell's development of propeller testing and evaluation facilities for the U.S. government during World War I.*

Key words: Propeller design; aircraft testing; aviation history.

Propeller design remains perhaps the most unappreciated of the important technological developments in early aeronautics. Not surprisingly, then, one of the main exponents of propeller research continues to be all but unknown. This man—Frank W. Caldwell, America's leading propeller engineer and designer during the aeronautical revolution of the 1920s and 1930s—oversaw most of the major improvements in propeller design and construction of that period *(Figure 1)*. He became the first to develop the metal, multipiece ground-adjustable pitch propeller and the hydraulically actuated variable-pitch propeller for practical purposes.[1] Caldwell's propellers, which hastened the advent of the modern airplane, coexisted with other big advances of the time in such areas as streamline design, cantilever monoplane wings, retractable landing gear, engines, fuels, and supercharging.

The aeronautical revolution of the 1920s and 1930s witnessed the transformation of the slow, fabric-covered, strut-and-wire-braced biplane existing at the end of World War I into the high-speed, streamline, cantilever monoplane seen just before World War II. A key technology in this transformation was the variable-pitch propeller, which enabled the angle, or pitch, at which each propeller blade rotated through the air to vary according to different flight conditions. Among many interwar milestones, the propeller innovations of Frank W. Caldwell came to be regarded as major technological achievements.

From 1918 to 1938, Caldwell effectively set the pace of discovery in the American propeller industry, yet despite his specialization, he recognized the importance of integrating the propeller advances with those of the larger technical systems of the airplane.

Caldwell was born on December 20, 1889, at Lookout Mountain near Chattanooga, Tennessee. His father was the president of the local iron works, and young Frank spent a lot of his time there learning about technology in the bustling industrial city. In 1908, Caldwell entered the mechanical engineering program at the Massachusetts Institute of Technology (MIT) where, in addition to his academic talents, he distinguished himself with his pleasing Tennessee drawl. Exhibiting a strong interest in aviation, Caldwell and another student, Hans Frank Lehmann, designed and flew a contest-winning glider. They also collaborated on a bachelor's thesis in 1912 entitled, "Investigation of Air Propellers," one of the earliest attempts at plotting out a comprehensive propeller-testing program.[2]

Caldwell graduated from MIT in 1912 and joined the Curtiss Aeroplane and Motor Company in Buffalo, New York. Curtiss sent him to Columbus, New Mexico, near the Mexican border, in 1916 to investigate the recent propeller failures on the Curtiss JN aircraft during the incursions of General John J. Pershing's forces into Mexico. Caldwell found that the wooden propellers, manufactured in the American northeast, dried and split in the heat of the southwestern desert. He studied local furniture-making practices and adopted native woods and improved glues as the solution, establishing a small factory that constructed 80 propellers for the expedition.[3]

When the United States entered World War I in 1917, it initiated an aviation production program

Figure 1. Frank W. Caldwell, 1889–1974.

funded by an unprecedented $640 million federal commitment. The program established an aeronautical research and development facility at McCook Field in Dayton, Ohio, to be operated (beginning in December 1917) by the newly formed Airplane Engineering Division of the U.S. Army Air Service. The division specialized in the modification of existing aircraft to increase performance, as well as the design, testing, and construction of new aircraft. The Propeller Department dealt specifically with enhancing the efficiency and durability of propellers, of which Caldwell became the civilian chief engineer, responsible for the research, design, and testing of all aircraft propellers used by the United States during the war.[4]

World War I served as a catalyst for aeronautical development in the United States, and it directly stimulated the development of the variable-pitch propeller. The wooden, fixed-pitch propeller, which was efficient for only one predetermined flight condition, gave satisfactory performance at speeds less than 100 miles per hour and at low altitudes. Until 1917, this limitation proved to be acceptable because the highest output of any American-made engine was the 90-horsepower Curtiss OX-5. The introduction of the 400-horsepower Liberty engine, however, represented a formidable challenge for Caldwell, who now had to conceive of a propeller capable of withstanding the forces of the new power plant.[5]

The broader American aeronautical community recognized the propeller dilemma from an early time. Indeed, William F. Durand, chairman of the civilian federal agency known as the National Advisory Committee for Aeronautics (NACA)—the predecessor of NASA—asserted in 1918 that the invention of the variable pitch propeller represented "one of the appliances for which the art of navigation is definitely in wanting."[6] The NACA leadership also knew that the realization of a practical metal propeller was one of the "very important problems now confronting the air services of the nation." In short, the NACA sought almost from its inception to foster a metal, variable pitch propeller.[7]

The NACA's objectives reflected the pursuit of two important and intertwined propeller design trends that began in the United States during World War I: the search for new materials used in the construction of propellers and the perfection of the mechanism for changing blade pitch. But before variable pitch could be a reality, engineers and designers had to develop a type of propeller that would enable controlled pitch variation. Caldwell led both developments.

Caldwell first addressed the issue of propeller composition. Wood reigned as the primary material from the earliest period to World War I because it was strong and easy to fabricate. But Caldwell also found that wood possessed several disadvantages. Variations in temperature and humidity, as well as water, sand, and stone, easily damaged the aerodynamic and structural integrity of wooden propellers, which led to an average service life of only 6 months. Additionally, wooden propellers deteriorated rapidly under the stress of high output engines, making them unsuitable for use on the new aircraft.[8]

It became clear to Caldwell that metal offered the needed durability and strength. Metal also offered a specific performance advantage because it could be fashioned into thinner blades toward the tip, thus increasing the propeller's efficiency. Moreover, by 1920, propeller engineers knew that thin blade sections were ideal for high-speed applications because they did not suffer from compressibility burble (with the accompanying sharp increase in drag at high speeds). Wooden propellers had thick airfoils for the purposes of structural integrity, but to withstand faster rotations, propellers needed to be thinner, which required the strength of metal.

The initial tests of steel propellers led Caldwell to reject the material. Wartime designs had exhibited low resistance to torsional stresses and were expensive to fabricate.[9] But Caldwell cleverly recognized that the construction method, not the material, had stymied development. The Propeller Department began work on a drop-forged steel propeller in 1918. Caldwell argued that drop-forging would produce a uniformly strong part.[10] The U.S. Army contracted with the Standard Steel Propeller Company of Pittsburgh, Pennsylvania, in 1920 to construct various forged-steel designs in 1920. Just as important, the contract initiated a long-term collaboration between Caldwell and Thomas A. Dicks, Standard Steel's chief engineer.[11]

The initial destructive whirl tests showed that the "steel" in Standard Steel lacked practical applications. Caldwell then proposed drop-forged duralumin blades. Composed of aluminum, copper, manganese, and lesser amounts of iron, magnesium, and silicon, duralumin enabled a closer adherence to wooden propeller design. Also, duralumin's lower density allowed a stiffer blade with a greater area than steel. Through a comparison of the blade deflections of wooden and steel propellers under load, Caldwell attempted to construct a duralumin propeller with a degree of flexibility somewhere between that of the old and new materials, and with a weight slightly above that of a wooden propeller.

Meanwhile, Dicks at Standard Steel developed a two-piece steel hub that used retaining shoulders to secure the blades at the blade root and a clamping ring to fasten the complete unit together. This assembly compensated for errors in the vertical balance of the blades, allowed interchangeability of parts, and permitted the aircraft operator to adjust pitch angle on the ground for anticipated performance regimes.[12] By 1925, the detachable drop-forged duralumin blades, split hub, retaining shoulders, and clamping ring represented a major milestone in propeller design, in effect ushering in the standardized metal ground-adjustable pitch propeller.

The appearance of the new propeller in the mid-1920s came at just the right time for military and commercial aircraft. During a routine takeoff exercise at the Norfolk Navy Yard, a new Curtiss SC torpedo-bomber scout lost its wooden propeller on takeoff because of structural failure. Fortunately, the crew survived the crash, landing at the end of the field. After an investigation, the U.S. Navy authorized procurement of 100 duralumin propellers, which marked the first use of the new propeller by the U.S. government.[13] A new generation of pilots, both military and civilian, adopted the new propeller during the 1920s as aviation expanded *(Figure 2)*.

The lessons Caldwell learned with the ground-adjustable propeller provided valuable design knowledge in realizing a practical variable-pitch propeller,

Figure 2. Charles A. Lindbergh specifically requested a Standard Steel ground-adjustable propeller in preparation for his solo transatlantic crossing in May 1927.

the next technological plateau for the American propeller community, which materialized during the early 1930s. Imbued with a sense of possibilities, Caldwell and other engineers even went so far as to predict that variable pitch would offer up to 40% more thrust at takeoff and a 60% faster rate of climb.[14]

Regardless of the rhetoric, nothing illustrated the need for variable pitch more than the lagging performance of the new aircraft—low-wing monoplanes with high wing loadings—rolling out of factories during the late 1920s, Indeed, with its ground-adjustable propeller set for cruise, a Boeing Monomail failed to even get off the ground during its first test flight in 1930. Test pilot Eddie Allen had to adopt a compromise propeller setting that did not generate full performance in either regime. Even then, the new mail plane needed every inch of runway to become airborne.[15]

By this time, Caldwell had been directing work on variable-pitch mechanisms for many years—since the opening of McCook Field in late 1917. One early design was a mechanically actuated, pilot-controlled propeller developed in 1917 by Seth Hart and Robert L. Eustis of Los Angeles. The Hart and Eustis propeller suffered from rapid wear of its control mechanism, and its wooden blades wore out quickly. Both Standard Steel (1920) and Caldwell (1922) designed early mechanically actuated, controllable propellers.[16]

Caldwell's long experience with these designs influenced his decision to concentrate on hydraulic actuation and to develop the new device privately. Consequently, he joined Standard Steel as chief engineer in June 1929.[17] Shortly after he got to Pittsburgh, a new aviation holding company, the United Aircraft and Transport Corporation (UATC), acquired Standard Steel expressly to get Caldwell into

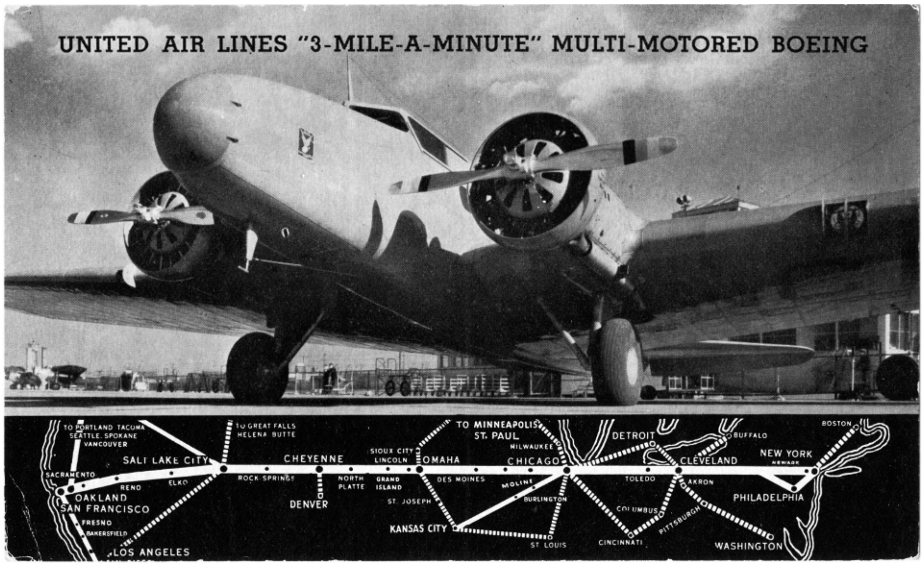


Figure 3. Boeing 247 transports equipped with Hamilton Standard two-position controllable-pitch propellers accelerated the United Air Lines route across the United States.

its ranks. UATC merged Standard Steel with Hamilton Aero Manufacturing of Milwaukee to create the Hamilton Standard Corporation in November 1929.[18]

At its founding, Hamilton Standard suffered from dwindling government contracts and the economic upheaval of the Great Depression. President Eugene E. Wilson asked Caldwell if he had any ideas for new products. Caldwell unveiled a drawing of a two-position controllable-pitch propeller he had patented in 1929. His design used the engine's oil supply and centrifugal force exerted by counterweights to keep the propeller blades at the desired pitch. The propeller was the aeronautical equivalent of a manual automobile transmission—the "gear shift of the air"—where "low" gear provided efficiency at takeoff and "high" gear provided efficiency at cruise. Wilson recognized the potential and quickly ordered a development program, at the time a heavy burden for the company. Caldwell's team fabricated the first propeller in 1930 and the new propeller went into production at the end of 1932.[19]

Yet, Hamilton Standard's new product ran into a solid phalanx of design conservatism. At first, the Army, Navy, and industry rejected the variable-pitch propeller.[20] The most notable example of this reluctance involved the development of the Boeing Model 247, often regarded as the first modern airliner. Boeing engineers rejected variable-pitch propellers because of anxieties about weight, cost, and mechanical complications. But without Caldwell's innovation, the new airliner performed poorly. During the United Air Lines acceptance flights at Cheyenne, Wyoming, in February 1933, the 247 could barely attain the 6,000-foot altitude needed to clear the Rocky Mountains.

Sensing an opportunity, Hamilton Standard sent Caldwell to Cheyenne to prove the value of its new innovation. His tests concluded that variable-pitch propellers reduced the 247's takeoff run by 20%, increased the rate of climb by 22%, and raised the cruising speed by 5.5%. Caldwell's invention maximized the performance of an already revolutionary airplane. Boeing placed the first production order for Hamilton Standard's controllable-pitch propellers for its transport aircraft *(Figure 3)*.[21]

The dramatic improvement in performance offered by the variable-pitch propeller won Caldwell and Hamilton Standard the prestigious 1933 Collier Trophy. President Franklin D. Roosevelt bestowed the honor personally, on behalf of the National Aeronautic Association, saying that the new device enabled "modern planes and engines to realize to the full the improvements in design" brought by the aeronautical revolution.[22] Roosevelt added that, "The success of his [Caldwell's] propeller has revealed a new horizon of aeronautics and taken the limits off speed. Henceforth, our pace through the air will be as fast as the daring and imagination of the engineers."[23] The aeronautical community got the point: The variable-pitch propeller had become a vital component of any new airplane design, and Hamilton Standard, expanding to meet demand, took orders for 1,000 controllable-pitch propellers by the spring of 1934.[24]

Caldwell next turned his attention (in 1932) to the ultimate variable-pitch mechanism, the constant-speed propeller, or the "automatic gear shift of the air." A constant-speed propeller changed blade pitch automatically according to varying flight conditions while

Figure 4. The full-feathering feature of the Hydromatic was crucial to the safety of American bomber crews over Germany and Japan.

the engine speed remained the same. After introducing an intermediate design that employed counterweights, Caldwell and his team developed a new propeller that relied on hydraulic pressure for all pitch actuation. Introduced in 1938, the Hydromatic propeller employed three major improvements. First, it incorporated a new safety feature called "quick-feathering," which prevented propeller windmilling after engine failure.[25] Second, the propeller's "quick-acting" cams provided better control of pitch variation, facilitated multiengine synchronization, and removed the risk of "overspeeding" the engine while diving. Finally, the sealed operating mechanism performed under constant engine oil pressure, making the Hydromatic easier to maintain and operate.[26]

The Hydromatic propeller played a pivotal role in the American air campaigns of World War II. Virtually the entire U.S. Air Force and Navy inventory, from multiengine bombers to fighter and transport aircraft, employed Hydromatic propellers, which shortened runway requirements, boosted climb, and added fuel efficiency *(Figure 4)*. D. Adam Dickey, the head of the Army's propeller program, traveled to Germany at the end of the war to evaluate aeronautical developments. Dickey concluded that while they were ahead in many areas, the Germans were dramatically behind the United States in propeller development, which he attributed to the "early vision" of Frank Caldwell.[27]

During the decisive years of World War II and the early Cold War, Frank Caldwell moved up the corporate ladder at United and served as director of research. He supervised the design and construction of one of the world's leading industrial wind tunnel facilities.[28] With his direct work in propellers finished, he supervised fundamental research until his retirement in 1955.

Frank Caldwell lived a long life. His other honors included the Sylvanus A. Reed Award (1935) and an honorary fellowship (1946) from the Institute of Aeronautical Sciences. In retirement, he served as a consultant to Hamilton Standard where he was friendly, open, and willing to discuss his experiences with younger engineers. Caldwell often appeared at corporate functions, always identifiable by his trademark crew cut. He died on December 23, 1974, at his home in West Hartford, Connecticut, at the age of 85.[29] ❑

Dr. Jeremy R. Kinney is a curator in the Aeronautics Division of the National Air and Space Museum, Smithsonian Institution, Washington, D.C. He curates the aircraft propulsion collection, which includes many of Caldwell's propellers. Kinney received his doctorate in the history of technology from Auburn University, Alabama, in 2003. E-mail: kinneyj@si.edu

Endnotes

[1]A ground-adjustable pitch propeller is fixed in pitch. A variable-pitch propeller can change its blade angle, whether it is on the ground or in the air.

[2]"Frank Walker Caldwell," *Collier's Encyclopedia* 4 (New York: Collier, 1957), 322–323; "F.W. Caldwell Dies at 85," *West Hartford News*, January 2, 1975, p. 4.

[3]"Frank Walker Caldwell," *Who's Who in American Aeronautics* (New York: Aviation Publishing Corporation, 1928), 17–18; "F.W. Caldwell Dies," *The Hartford Courant*, December 24, 1974, p. 4; George Rosen, *Thrusting Forward: A History of the Propeller* (Windsor Locks, CN: United Technologies Corporation, 1984), 31.

[4]"Frank Walker Caldwell," *Collier's*, 322–323; Lois E. Walker and Shelby E. Wickam, *From Huffman Prairie to the Moon: The History of Wright-Patterson Air Force Base* (Washington, D.C.: Government Printing Office, 1986), 178–181.

[5]D. Adam Dickey, interview by Lois E. Walker, September 2, 16, 30, 1983, interview K239.0512-1712, transcript, U.S. Air Force Oral History Program, U.S. Air Force Historical Research Agency, Maxwell Air Force Base, Alabama, 6, 163; Ronald Miller and David Sawers, *The Technical Development of Modern Aviation* (New York: Praeger, 1970), 71–72.

[6]William F. Durand, "Some Outstanding Problems in Aeronautics," Wilbur Wright Memorial Lecture, *Annual Report NACA, 1918* (Washington, D.C.: Government Printing Office, 1919), 40–41.

[7]Problems in Propeller Design" *Aviation and Aeronautical Engineering* 4 (June 1, 1918): 108.

[8]Frank W. Caldwell, "The Construction of Airplane Propellers," *Aviation* (May 1, 1917): 300; Dickey Oral Interview, 1; Rosen, *Thrusting Forward*, 30, 32, 36.

[9]Frank W. Caldwell, "Development of the Drop-forged Metal Propeller," *Aeronautical Engineering: Contributions of the ASME Aeronautics Division* (April–June 1930), 73; Dickey Oral Interview, 13, 16, 20–21, 36; Rosen, *Thrusting Forward*, 24, 37.

[10]Caldwell, "Drop-forged Metal Propeller," 73.

[11]William F. Trimble, *High Frontier: A History of Aeronautics in Pennsylvania* (Pittsburgh: University of Pittsburgh Press, 1982), 116.

[12]Caldwell, "Drop-forged Metal Propeller," 74–75.

[13]Eugene E. Wilson, *Slipstream: The Autobiography of an Air Craftsman* (Palm Beach, FL: Literary Investment Guild, 1967), 77–78, 155–156; Rosen, *Thrusting Forward*, 38, 40.

[14]Rosen, *Thrusting Forward*, 42.

[15]Harold Mansfield, *Vision: The Story of Boeing* (New York: Popular Library, 1966), 42, 44, 46.

[16]"New Type Adjustable Pitch Propeller," *Bulletin of the Airplane Engineering Department* 1 (June 1918): 78–88; Miller and Sawers, *Technical Development*, 72; Rosen, *Thrusting Forward*, 42.

[17]"Frank Walker Caldwell," *Collier's*, 322–323; Miller and Sawers, *Technical Development*, 73.

[18]"U.A.T. Propeller Firms Now Hamilton Standard," *Aviation* (November 30, 1929): 1080.

[19]"A Gear Shift For The Airplane," *New York Times*, November 25, 1935, sec. 8, p. 7; Wilson, *Slipstream*, 165–166; Rosen, *Thrusting Forward*, 42; Miller and Sawers, *Technical Development*, 74.

[20]Wilson, *Slipstream*, 166–168; Mansfield, *Vision*, 44.

[21]"Transport Cruises 171 M.P.H. Using New Type Props," *Boeing News* 4 (May 1933): 1; Charles H. Chatfield, "Controllable-Pitch Propellers in Transport Service," *Aviation* 32 (June 1933): 180; Wilson, *Slipstream*, 168–170; F. Robert van der Linden, *The Boeing 247: The First Modern Airliner* (Seattle, WA: University of Washington Press, 1991), 46.

[22]"Caldwell Wins Aviation Trophy," New York *Times*, May 29, 1934, pg. 7, col. 4; C.B. Allen, "Hamilton Standard Wins Collier Trophy for Controllable-Pitch Propeller," *The Bee-Hive* 8 (June 1934): 1.

[23]Franklin D. Roosevelt, quoted in *Technology Review* 36 (July 1933–1934): 1.

[24]Wilson, *Slipstream*, 170; Miller and Sawers, *Technical Development*, 75.

[25]When an engine fails, the airstream causes the blades to rotate like a windmill. Feathering, or turning the blades so they are parallel to the airstream prevents windmilling.

[26]"The Hamilton Standard Constant-Speed Propeller," *The Bee-Hive* 10 (September 1936): 1–3; Frank W. Caldwell, "A Review of the Hydromatic Propeller," *The Bee-Hive* 14 (July 1939): 1–3; Rosen, *Thrusting Forward*, 46, 49.

[27]Dickey Oral Interview, *x*.

[28]"Frank Walker Caldwell," *Collier's*, 322–323; Wilson, *Slipstream*, 256.

[29]"F.W. Caldwell Dies," p. 4; "F.W. Caldwell Dies at 85," p. 4.

ITEA Journal 2009; 30: 345–353

Spiral 1 Space Threat Assessment Testbed (STAT) Development

Lance Baxter
U.S. Air Force Arnold Engineering Development Center, Arnold Air Force Base, Tennessee
Randy Simpson, John Prebola, and Marc Smotherman
Aerospace Testing Alliance (ATA), Inc., Arnold Air Force Base, Tennessee
Douglas K. Moore, Rudolf Goldflam, and Dawn Battles
ATK – Mission Systems, Goleta, California
Dave Higham and Dave Perkes
ATK – Space Systems, Logan, Utah

The development of the first phase of the Space Threat Assessment Testbed (STAT) facility, known as Spiral 1, will provide a capability to test space hardware without the expense of space flight. Key features of STAT include the capability to replicate effects of complex natural space environments occurring at low earth and geosynchronous orbits. In addition, STAT will simulate key artificial threats and provide a near-real-time connection capability that enables the involvement of ground station hardware, software, and operators in the control, test, and evaluation process. STAT will also lay the foundation for near-real-time connectivity to a satellite operations center, enabling the U.S. Air Force to perform integrated system testing and training while also assisting in the development of tactics, techniques, and procedures for future space operations.

Key words: Capability recovery; defensive counterspace operations; hardware, software, human controllers; integrated testing; natural space environment; realistic orbital conditions; satellite assets; simulation; spacecraft systems; survivability; thermomechanical performance.

The Arnold Engineering Development Center (AEDC) mission is to provide high-quality, economical, and timely state-of-the-art test and evaluation (T&E) services in support of the U.S. Department of Defense, NASA, commercial, and international aerospace programs. In support of this T&E mission, technology research and development programs are conducted to further advance testing techniques and equipment, new facilities are designed and constructed, and the existing facilities are kept current through maintenance and modernization programs. The Space Threat Assessment Testbed (STAT) facility is supported by a substantial infrastructure, which includes liquid nitrogen, handling systems, computers, communication networks, and diagnostics. In addition, capabilities include laboratory support equipment, modeling and simulation tools, and applied technologies research.

Overview and purpose of STAT

Air Force Doctrine Document (AFDD)-2-2.1 (AFDC 2004) lists three primary missions for counterspace operations: space situational awareness (SSA), defensive counterspace (DCS), and offensive counterspace. STAT Spiral 1 supports the acquisition and utilization of systems to support the SSA and DCS missions. Spacecraft systems, in some cases subsystems and ground systems, work together to provide the capabilities to perform these functions. Ground testing these systems in the appropriate natural environments against various threats will provide the data necessary to understand how systems perform and to identify indicators of an attack. As systems and subsystems are designed to provide defensive counterspace capabilities, their performance must be characterized in a controlled manner and evaluated as an integrated system. This includes integrated T&E, ground support and crew involvement, understanding the effects on satellite operations, and participating in mitigation activities. This will enable evaluation of the ability of a subsystem or system to detect, discriminate, and attribute attacks and of the vulnerability of a subsystem or system to the natural environment and to enemy action.

SSA testing

SSA is the result of sufficient knowledge about space-related conditions, constraints, capabilities, and activities—both current and planned—in, from, toward, or through space. Achieving SSA supports all levels of planners, decision makers, and operators across the spectrum of terrestrial and space operations. SSA involves characterizing, as completely as possible, the space capabilities operating within the terrestrial and space environments. SSA information enables defensive and offensive counterspace operations and forms the foundation for all space activities. The primary mission of STAT Spiral 1 is to provide a capability that does not currently exist to enable characterizing our own space capabilities that are or will be in the operational space environment. That characterization includes the operation of subsystems and/or systems in the natural space environment and under the influence of certain enemy actions.

DCS testing

DCS operations preserve U.S.-friendly abilities to exploit space to its advantage via active and passive actions to protect friendly space-related capabilities from enemy attack or interference. Friendly space-related capabilities include space systems such as satellites, terrestrial systems such as ground stations, and communication links. DCS operations are key to enabling continued exploitation of space by the U.S. and its allies by protecting, preserving, recovering, and reconstituting friendly space-related capabilities before, during, and after an adversary attack. As DCS capabilities are developed, their ability to detect, discriminate, find, fix, track, target, and locate threats in a challenging environment must be independently evaluated under representative operational controlled conditions. Testing in an integrated environment including space segment hardware, ground control hardware, software, and human controllers provides the greatest level of understanding of the integrated system capabilities.

Natural space environment

There have been many descriptions of the natural space environment over the years, and with each new version comes new information. One of the latest space environment books released at the time of this writing is by Vincent Pisacane and describes each component of the space environment in detail (Pisacane 2008). The worldwide importance of operating in this environment is also highlighted by the November 2008 release of a new space environment standard by the European Cooperation for Space Standardization (ESA-ESTEC 2008). Laboratory investigations of effects of such environments on space materials abound (Barrie et al. 2002). If one considers replicating all aspects of the space environment in a ground-test facility, it becomes apparent that such a difficult undertaking is ultimately cost prohibitive, if not impossible because of chamber effects and limitations. The following space environment subset was chosen for STAT after considering tradeoffs of complexity, orbits of interest, and fidelity:

1. protons,
2. electrons,
3. atomic oxygen (AO),
4. solar radiation,
5. electric thruster ions,
6. surface and optical contamination, and
7. spacecraft charging and charge migration.

The first two environment components comprise the natural charged particle environment. This typically includes protons, electrons, and heavier ions trapped in the earth's magnetic field, streaming from the sun, or originating in deep space. The plasma environment consists of low-energy (few eV) electrons and ions that are associated with spacecraft charging. The next energy group of protons (less than 1,000 keV) and electrons (few keV) have energies too low to penetrate deep into the spacecraft and are primarily responsible for surface damage. STAT simulation of surface damage relies primarily on a low-energy proton source. More energetic particles in the radiation environment penetrate beyond a surface and generate secondary radiation. In addition to the surface effects, they affect solar cells, electronics, and electro-optics through total ionizing dose. STAT simulates total ionizing dose using a medium-energy electron source. The overlap of these groups is ill-defined and can cause confusion when discussing the damage potential to spacecraft hardware, especially when considering qualification of thin film coatings and micro electromechanical systems components. STAT will include hardware that spans much of the radiation environment and provides a more realistic charged particle test environment for many space assets. Displacement damage caused by high-energy protons, neutrons, and ions is not simulated in STAT in order to control facility cost associated with such sources.

AO found in low earth orbit is highly reactive with many materials. It is formed when ultraviolet radiation from the sun interacts with diatomic oxygen. The mean free path of AO is large at low earth orbit, and thus recombination is limited. This results in primarily AO atmosphere above 120 km altitude. Surfaces of orbiting spacecraft interact with these particles and become degraded to varying degrees. The most dramatic AO effects were seen on the Long Duration Exposure Facility experiment of the 1980s (Levine 1991). Ever since the Long Duration Exposure Facility experiment, material survivability to AO has been an ongoing research effort, and since 2001, the Materials International Space Station Experiment (MISSE) experiments have been conducted at the International Space Station (NASA 2008). New space-deployed materials and hardware must be capable of surviving the low earth orbit AO environment. For this reason, AO environment simulation is a critical part of STAT.

The sun is the most important object in the solar system, and solar radiation is continually emanating toward the planets. Whereas there is variability in the intensity, on average the exoatmospheric earth and near-earth spacecraft are illuminated with total solar irradiance of approximately 1,366 W/m^2 (ASTM 2000). Sunlight is useful for spacecraft as solar cells can convert solar photons into electrical power to operate the spacecraft. Unfortunately, the solar spectrum also contains the same ultraviolet light that is energetic enough to dissociate oxygen and can damage spacecraft materials and coatings. At longer wavelengths of the solar spectrum, the incoming radiation causes the spacecraft to heat up, requiring careful thermal management of any space object. If the spacecraft thermal balance is disrupted by emissivity changes of surface materials due to the effects of space environments, the incoming solar flux can cause excessive heating resulting in performance degradation or loss. STAT will have a broad spectrum solar simulator that spans the ultraviolet to infrared spectrum to properly evaluate spacecraft and component thermomechanical performance and survivability. Electric thrusters are becoming more common on a variety of NASA and commercial spacecraft (Pidgeon et al. 2006; Polk et al. 2001). Although there has been some testing and modeling of these devices with respect to how they interact with the spacecraft, there are still concerns because of limited operational experience. Of primary concern are how the charge exchange ions near the exit of the spacecraft interact with nearby surfaces and components. A full thruster is not planned for inclusion in STAT; instead, hardware will be added to simulate the low-energy ions in the charge exchange cloud.

When a low-pressure, elevated temperature environment is encountered, most hardware will give off some amount of material. Cold surfaces tend to condense these materials, depending on the surface temperature and outgassed material condensation temperature. This is known as outgassing and redeposition of volatile condensable materials and occurs in the space environment and in ground-test chambers (Prebola et al. 2009). If the condensable material impacts performance of a system it is known as contamination. Contamination control of spacecraft and hardware operating in these environments is essential because even small amounts of contamination can significantly degrade performance. STAT Spiral 1 will include hardware to reproduce the outgassing products of large spacecraft surfaces that will not fit within the test volume. This will allow the determination of how outgassed species may impact test article performance.

Some spacecraft in the orbital environment have large surfaces capable of collecting charged particles. Positive and negative particles may collect on different surfaces in different areas of the spacecraft because of potential buildup, material properties, or other factors. Sufficient charge buildup can result in sudden discharge, thereby damaging the spacecraft. A number of relevant references can be found in the recent spacecraft charging paper from the Air Force Research Laboratory (Lai 2007). STAT Spiral 1 is a medium-scale test bed and is not sized to accommodate large spacecraft. An induced charging system will be employed to properly replicate the portion of the spacecraft that is not in the chamber.

STAT mechanical and chamber systems

The STAT facility vacuum, mechanical and chamber systems are designed around support of the natural and threat sources to provide mechanical mounting, vibration isolation, a class 7.0 clean room, and a high-vacuum/cryogenic test environment. The proposed ATK STAT layout, shown in *Figure 1*, illustrates the main chamber, the antechamber for the test article, and all the sources that illuminate the test article through ports in the main chamber. The design meets requirements for 1×10^{-5} Torr vacuum level, with all sources operational and nominal 80 K Cryoliner temperature based on delivery of similar designs to multiple customers during the last 20 years. Modeling based on data from similar chambers and the baseline system design indicate the STAT system will reach steady-state test conditions in 3 days or less from the start of chamber evacuation, remain at test conditions for at least 500 hours, and return to ambient conditions in 2 days or less, thus meeting STAT threshold requirements.

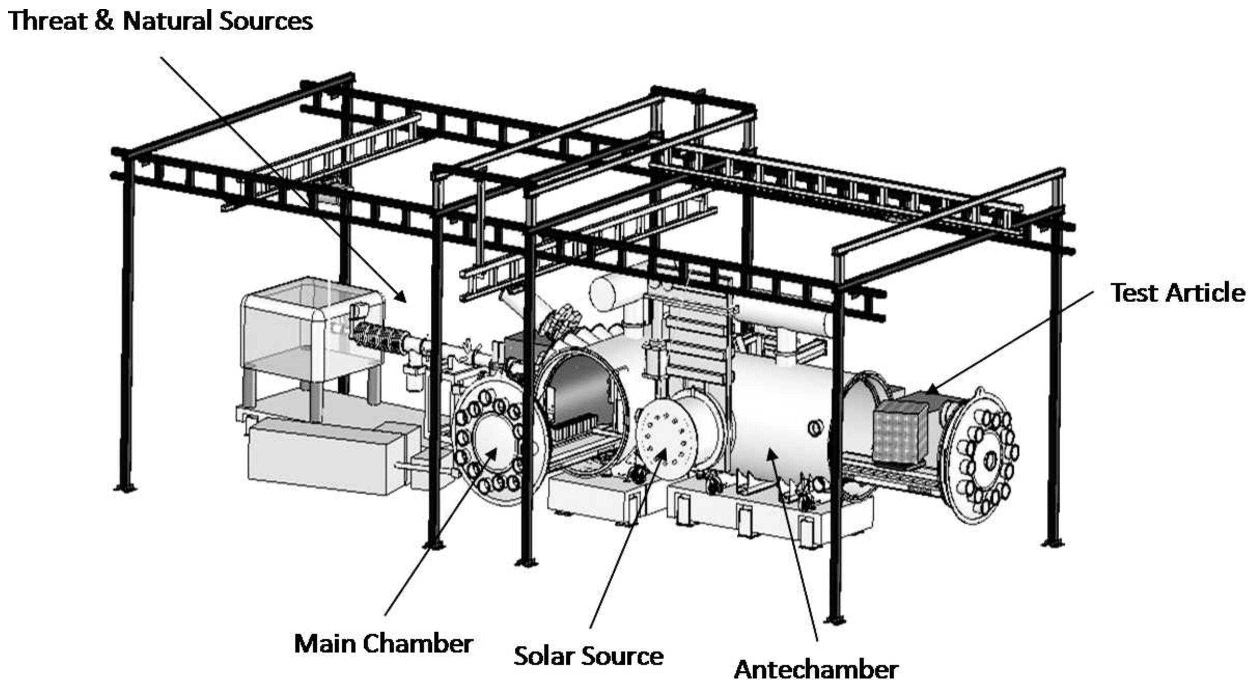


Figure 1. STAT vacuum chamber and sources.

The chamber design consists of two vacuum shells in a T-shaped configuration and includes isolation/gate valves that will enable each chamber to be valved off from each other and from the STAT test volume. The standard cryogenic vacuum cryoliner that has been employed for numerous applications is approximately 2 m internal diameter. At this diameter, it will give AEDC about 10 percent growth potential over the 75-cm, on-a-side cube test article expected for the initial capability. The main chamber uses the same design and enables source input on one side, expansion through the width, expansion through the intermediate gate valve, and then subsequent presentation at the test article. Small changes (less than 5 percent) in the diameter of the chamber and cryoliner do not significantly affect the manufacturing cost of the overall chamber and cryoliner. In addition, each simulation/threat source is separated from the STAT chamber by a valve and has independent controls, vacuum, and power supply that can be operated autonomously. This enables each source to be verified, repaired, or replaced independent of chamber operations. The source-beam expansion volume is further separated from the test article by a gate valve. This arrangement, shown in *Figure 2*, enables test article access, servicing, replacement, and diagnostics without having to open the complete STAT system volume. The antechamber test volume can be pumped and cooled independently of the main chamber. With both sides of the gate valve at cryovac conditions, either side can be warmed and recooled to cryovac without disturbing the other side of the chamber.

Utilizing a novel, hydraulically formed, toroidal shell tank, the shroud design maximizes the heat-transfer area by placing the saturated liquid nitrogen across a very thin metal boundary surrounding the entire inner shroud surface. This design approach has proven to keep cryoliner temperature to within a few degrees of 80 K even when it is coated with a highly emissive surface and exposed to a 300 K radiation. The chamber systems using liquid nitrogen cryogenics remain consistently below 1×10^{-6} Torr. The internal surface of the liner is painted with Aeroglaze Z-306 paint as the baseline on the cryogenic liner interior. ATK has successfully used this paint scheme on many chambers that have been in operation for more than 10 years and undergone hundreds of cryocycles without any need to rework the paint. Z-306 is known for its excellent infrared absorption properties, but its reflectance properties can vary dramatically across the solar spectrum. The reflectivity versus wavelength of Z-306 was measured on baffles to be in the four to 10 percent range over a measured solar spectrum from 250 to 2,500 nm. During the STAT design, work will continue to investigate optical blacks for their cryogenic and optical properties in the solar region, as well as resistance to AO erosion.

The Test Article Positioning System (TAPS) consists of a 1-m travel test article deployment stage that will run in the same direction as the cylindrical

Figure 2. Typical antechamber design enables easy access to a test article.

axis of the antechamber. This will give AEDC the ability to deploy and retract the test article along the projection axis of each environmental source to vary intensity. A ± 180-degree yaw cryogenic rotary stage will also be provided that will be mounted on the deployment stage to enable each side of the test article to be presented to each of the sources. Cabling and liquid nitrogen connections will run through the center of the rotary stage mechanism to minimize cable drag and wind up. The design will accommodate a 100-kg cube test article measuring up to 82 cm on a side without hitting the walls of the enclosure. The cryogenic TAPS positioning stage is capable of supporting 120 kg, a 20 percent margin above the 100 kg requirement. Umbilical connection providing power, data, and cryogenics is provided through the base of the motion table. This arrangement limits the motion to 180 degrees in either direction to prevent kinking of the connections. The TAPS baseline mechanical interface to the chamber will be via a rotary platen with a large cable pass-through in the center of the cryogenic rotary stage. The platen will be fabricated from 6061 series aluminum and have an embedded heat exchanger to enable the platen to be kept at the same temperature as the rotary stage base. As a baseline, the final electrical connections to the test article consist of resistance temperature detectors, Type T thermal couples, CAT 6 cables, four-conductor twisted shielded pair, and coax and triax cables. This bulkhead feed can be accomplished either above or below the test article, depending on its final configuration. For routing below the test article, routing cables under the table and then through another bulkhead in the side of the table support and then through the rotary platen to the test article is recommended to minimize cable drag.

The STAT chamber will be enclosed in a Class 7.0 clean room environment with penetrations provided for a high bay roll-up door, personnel entry doors, vacuum, and electrical components. Western Environmental Corporation is the lead contractor for the clean room built for the STAT system. Western Environmental Corporation clean room development addresses familiarity with AEDC constraints on such construction because they have developed other clean rooms for AEDC. The environmental sources and other STAT components that must be close to the chamber will be inside the clean room. These include the local control racks for vacuum, source control consoles, and TAPS.

The STAT Vacuum System consists of two identical pumping systems as shown in *Figure 3*, one for each chamber to enable high-vacuum evacuation independently of each other. The approach will use four turbomolecular pumps (TMPs) in each pumping subsystem backed by independent Leybold screw roughing/backing pumps to provide clean, dry vacuum pumping using standard components, which address safety concerns with pumping atomic oxygen. Based on history with similarly sized chambers, the pumping system described below will achieve less than 10^{-6} Torr vacuum when coupled to the chambers. This vacuum system is cost effective with ease of maintenance using common pumping components and contains complete redundancy to mitigate single-point failures. The proposed

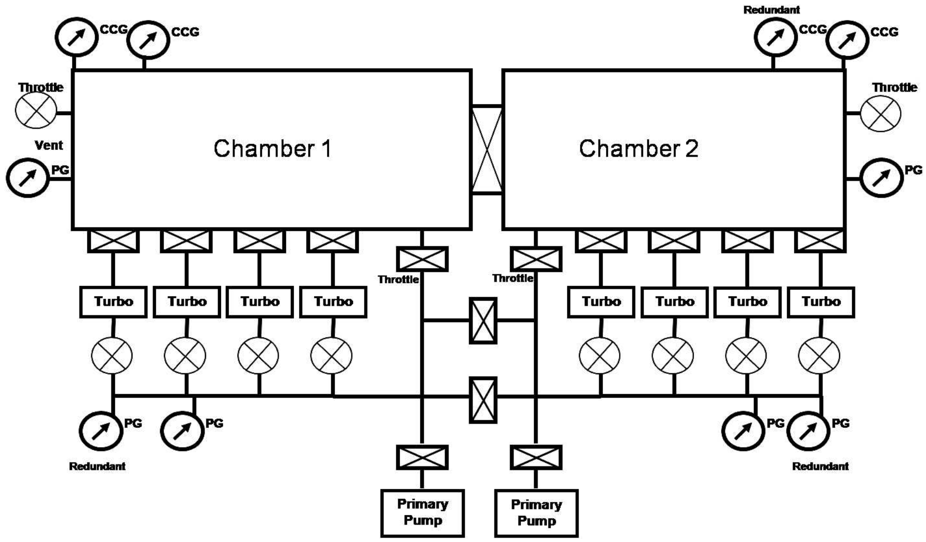


Figure 3. Vacuum system schematic.

vacuum system is a throughput type system where all gases are exhausted through the high-vacuum pumps and the primary backing pumps. It is proposed to use four Pfeiffer 1201 TMPs per chamber (a total of 8). Three of the 1201 TMPs will provide about 3,800 L/s pumping speed, whereas the fourth TMP provides additional pumping speed for backup if one TMP fails or if needed for extreme thermal loads. Three TMPs can handle the gas load during AO generation. A single smaller TMP is typically used on standard space chamber systems of similar size.

The liquid nitrogen system consists of a 5,000-gallon vacuum-jacketed supply tank located outside the STAT Facility and vacuum-jacketed transfer lines to the chamber and vented to the outside. The transfer lines will be factory manufactured and assembled onsite using bayonet connections. Each section will have a static vacuum insulation established at the factory through an evacuation/relief port and measured by a thermocouple vacuum gauge. Vacuum-jacketed pneumatically controlled isolation valves will be provided for flow control on the supply side. Overall heat load on the chamber from steady-state radiation and conduction heat and operational heat load from the sources indicate about a 400-gallon liquid nitrogen usage per day (2,800 gallon per week). The proposed design will have additional capacity to accommodate thermal emissions from the test article with margin.

Data Acquisition and Control System (DACS)

The STAT DACS comprises the STAT software architecture, processing, and display support for the operators and maintenance personnel. As shown in *Figure 4*, it comprises computer workstations, data storage redundant arrays of independent drives (RAIDs), intercom operator consoles, and projected displays. Computer software configuration item (CSCI) servers and clients provide STAT subsystem control as well as data acquisition and analysis, source scenario generation, data storage, and system health monitoring. The DACS is a distributed system that uses remote data access and control methods for collecting information and controlling system operations. This enables both independent and integrated operation of system elements, as well as supporting incremental and independent development, integration, and test with low risk. The DACS is composed of several systems: the chamber monitor and control system, source control and monitor system, vibration isolation and monitor system, contamination monitoring system, auxiliary power system, and video system.

The DACS is a distributed system that is interconnected via data network and the ethernet. The data network is the primary method to control and receive real-time status throughout the STAT system. The data network is defined and mapped out, creating a common interface to the complete STAT DACS. The ethernet will be used for file transfer and remote control of individual computers via Microsoft Remote Desktop. All data acquired by the DACS are time-tagged with Inter-Range Instrumentation Group-B (IRIG-B).

The chamber monitor and control system is based on a programmable logic controller with distributed I/O that controls and monitors all chamber systems, including monitoring of chamber temperatures and pressures, control and monitoring of the vacuum

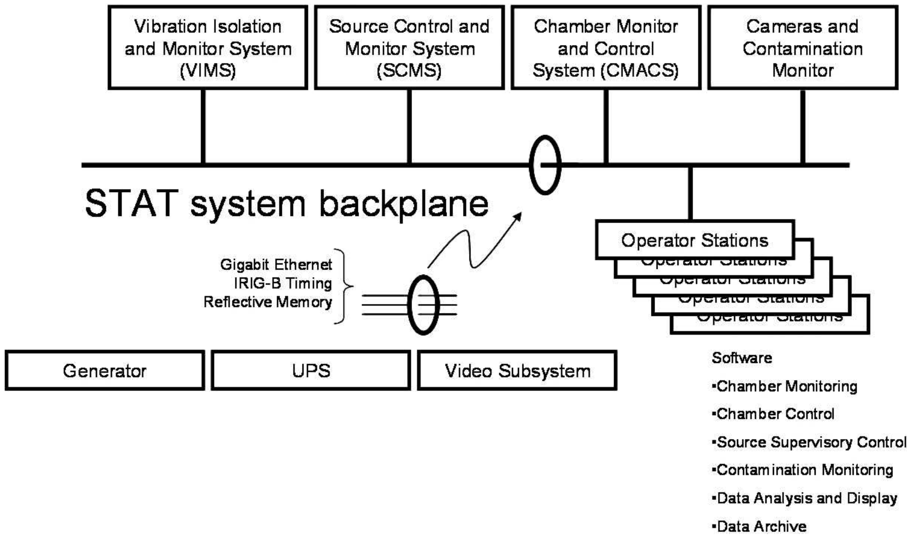


Figure 4. Data acquisition and control schematic.

pumping system, control and monitoring of the liquid nitrogen system, control of any chamber heaters, and monitoring of all utility systems. The chamber monitor and control system provides operators, via any of four facility operator stations, with system schematics overlaid with real-time data. From these schematics, operators control any function of the chamber and monitor resulting response. The chamber monitor and control system also provides real-time plotting of any STAT parameter over any time frame of the STAT test, with as many as 10 parameters per plot, and archives all STAT data acquired by all STAT DACS systems at a user-defined rate to a 10 TB RAID.

The source control and monitor system is a personal computer-based system executing Windows that monitors and controls all natural and threat sources and the test article positioning system. The source control and monitor system provides operators with schematics of the source systems overlaid with real-time data, which enable operators to remotely control any function of the sources or test article. The source control and monitor system provides automated calibration routines of source outputs, and when in test mode, controls all source outputs based on script files, developed manually or via AFGEOSPACE software. Source control up to 1,000 hours can be accommodated.

The vibration isolation and monitor system is a Windows-based personal computer that monitors and controls the STAT chamber vibration and level. The vibration isolation and monitor system analyzes data taken from accelerometers and places the results in a data network. It receives commands sent over the data network from an operator station and controls the chamber level and damping parameters.

The contamination monitoring system acquires data from a residual gas analyzer and quartz crystal microbalance and places these data on data network. Control of both devices is performed on a Windows-based personal computer at an Operator Station.

The auxiliary power system consists of an uninterruptible power supply with backup generator. The uninterruptible power supply provides 80 kVA of power, whereas the backup generator is rated to 65 kW. The uninterruptible power supply and generator provide backup power to the chamber vacuum system and critical DACS systems, enabling a controlled return to vacuum in the case of a total power loss.

The video system consists of three overhead projectors and three display screens located in the STAT control room. The system enables operators to display computer video output, operator station video, or closed circuit television video on any of the three display screens. Each screen may be subdivided into four separate screens, providing display of up to 16 different signals.

Conclusion/summary

The new STAT facility will offer a unique test capability for satellite assets, exposing them to realistic orbital conditions and environment and artificial threats. In addition, STAT will provide a real-time connection capability that enables involvement of ground station hardware, software, and operators in the T&E process. The complexity of STAT will present numerous challenges throughout the development process; however, the importance to understanding integrated spacecraft performance in the natural and threat environment is a challenge worth facing. The new STAT facility establishes a new approach to integrated T&E of space systems. ❑

LANCE E. BAXTER is the director of the 649th Test Systems Squadron at Arnold Engineering Development Center (AEDC) Arnold Air Force Base, Tennessee. The Test Systems Squadron directs investment project planning and systems engineering to assure that Research, Development, Test and Evaluation (RDT&E) systems are readily available at the world's largest and most advanced aerospace ground test complex. The delivered ground test systems support aircraft, missile, space launch, and satellite development for the U.S. Department of Defense (DoD), NASA, and commercial customers. Mr. Baxter received a bachelor of science degree in electrical engineering at the U.S. Air Force Academy and received a master of science degree in electrical engineering from the University of New Mexico. He currently works in the federal civil service but held technical and management positions on active duty in the U.S. Air Force (USAF) for more than 8 years and continues to serve in the USAF Reserve. His military positions included engineer and laboratory manager, command protocol officer, manager of space testing at AEDC, IMA to the C-130 AMP deputy program manager, and analyst. He is the author/co-author of multiple published papers. E-mail: lance.baxter@arnold.af.mil

RANDY SIMPSON is a mechanical engineer at Arnold Engineering Development Center (AEDC) employed by Aerospace Testing Alliance with 35 years of experience. For 10 years he has been responsible for test buildup for various test installations in the Space Test Area. For the remainder of his career he has been specifically involved in the development of the 7V and 10V Aerospace Chambers and Space Threat Assessment Testbed. Mr. Simpson received his bachelor of science degree from Vanderbilt University and his master of science degree from the University of Tennessee, both in mechanical engineering. E-mail: Randy.simpson@arnold.af.mil

JOHN PREBOLA is currently the Aerospace Testing Alliance space environments lead for STAT and is a member of the ATA Technology and Analysis Branch. He has worked with space environment simulators and facility development since the Characterization of Combined Orbital Surface Effects project began in 2000. Prior to this, he worked to develop various plasma combustion techniques for high- altitude ignition of turbine engines. Mr. Prebola received his bachelor of science degree from North Carolina State University in 1997 and his master of science degree from Virginia Tech in 1998, both in aerospace engineering. E-mail: John.prebola@arnold.af.mil

MARC SMOTHERMAN is a controls engineer at Arnold Engineering Development Center (AEDC) employed by Aerospace Testing Alliance with 29 years of experience. For 10 of these years he has been responsible for the design, implementation, and maintenance of motion control systems for the 16-feet and 4-feet wind tunnels at AEDC. The past 19 years he has been technical project manager for the development of data acquisition and control systems for the 7V and 10V Aerospace Chambers, 16T Propulsion Wind Tunnel, and Space Threat Assessment Testbed. E-mail: Marc.smotherman@arnold.af.mil

DOUGLAS K. MOORE, is the director of system engineering at ATK Mission Research, Santa Barbara, California. He has more than 25 years experience as a research scientist/engineer, program manager, integration and test engineer, and systems engineer. He has been responsible for technical, schedule, and cost performance on contracts with total value more than $30 million. Mr. Moore has managed program tasks across 7 MRC locations and 35 full-time equivalent staff in technical areas of system engineering, structural design, hardened microelectronics, cryogenic special test equipment, focal plane array evaluation, composite hardware fabrication, and electromagnetic compatibility. E-mail: Douglas.Moore@atk.com

DR. RUDOLF GOLDFLAM is the chief engineer at ATK Mission Research, Santa Barbara, California. Dr. Goldflam provides technical leadership in the testing of space sensors, infrared systems and electronics, and provided engineering services in system design, electro-optics, and electronics for space and airborne applications. Dr. Goldflam co-founded a radiation test laboratory designed for operational and hardware-in-the-loop testing of space and strategic sensors. Primary areas of technical expertise are optical/IR sensor system design, development and testing; Focal Plane Array and electronics survivability and operability and radiation hardening; development of system specifications and requirements flow down to subsystem level; system engineering and subsystem evaluation, design, and testing; space environment definition; sensor signal processing; and chemical dynamics of atmospheric gases. Dr. Goldflam earned his bachelor of science degree in physics, nuclear engineering, a master of science in nuclear physics, and a doctor of philosophy degree in chemical physics. He has 40 years of professional experience. E-mail: Rudolf.Goldflam@atk.com

DAWN BATTLES is ATK Mission Research Santa Barbara's lead program manager and manages the successful execution of specific contract and/or program efforts, including the management of technical milestones, program teams, schedule, program budget and expenses, and contract deliverables. Ms. Battles has been with ATK Mission Research for more than 10 years in program management, contract administration, and subcontract administration within the defense industry, managing time and material,

firm-fixed price, and cost-type contracts for services, research and development, and hardware development. Her experience encompasses contractual compliance, pricing/estimating, proposal development, fact-finding and negotiations, change order management, contract close-outs, export/import compliance, and government contract requirements. E-mail: Dawn.Battles@atk.com

DAVE HIGHAM is the group leader of the ATK Systems Development Group located in Logan, Utah. He has more than 20 years experience in the design and development of the cryogenic vacuum chamber, cryogenic motion control, and cryogenic optical systems. Mr. Higham has led or participated in the development of complex low background (cryogenic) Opto mechanical Infrared calibration systems. He has been involved in the production of more than 20 cryogenic vacuum chamber systems currently in use at government test facilities, university research centers, and major aerospace corporations around the world. Many of the subsystem designs from these cryogenic calibration chambers have resulted in the production of devices including cryogenic stepping motors, vacuum pumping systems, cryogenic fluid control valves, cryogenic translation stages, cryogenic filter wheels, cryogenic blackbodies, ambient and cryogenic reflective and refractive collimators, and ambient and cryogenic steering mirrors systems. Mr. Higham has a bachelor of science degree in mechanical engineering from Utah State University. E-mail: Dave.higham@atk.com

DAVID PERKES is the engineering manager and senior engineer at ATK Mission Research in Logan, Utah. He has more than 25 years experience as a systems engineer, engineering manager, and electrical engineer. He has overseen and performed the engineering and design of several space simulation chambers for Raytheon, Applied Physics Lab, Lockheed, and Space Dynamics Lab, including many subassemblies for customer-owned chambers including components for the AEDC 7V and 10V chambers. Mr. Perkes specializes in electro-mechanical controls and software and is skilled in chamber temperature, vacuum, and data acquisition systems as well as controls for cryogenic electro-mechanical devices such as linear and rotary stages, gimbals, filter, and aperture wheels. He has a bachelor of science degree in electrical engineering from Utah State University. E-mail: Dave.perkes@atk.com

References

AFDC. 2004. Counterspace Operations. Airforce Doctrine Document (AFDD) 2-2.1. August 2, 2004. Washington, DC: HQ Air Force Doctrine Center.

ASTM International. 2000. *Solar constant and zero air mass solar spectral irradiance tables*. ASTM E490-00a. West Conshohocken, PA: Author.

Barrie, J. D., M. J. Meshishnek, P. D. Fuqua, and S. C. Rostel 2008. Simulated space environmental exposure of optical coatings for spacecraft solar rejection. *Appl. Opt*. 41: 3150–3155 (2002).

ECSS-E-ST-10-04C. 2008. *Space engineering: Space environment*. Noordwijk, The Netherlands: ESA-ESTEC, Requirements & Standards Division.

Lai, S. T. 2007. Spacecraft charging–present situation and some problems. AIAA 2007-4525, 38th AIAA Plasmadynamics and Lasers Conference, Miami, FL.

Levine, A. S. (editor). 1991. LDEF–69 Months in Space, First Post-Retrieval Symposium (Part 2). NASA Conference Publication 3134, NASA Scientific and Technical Information Program.

NASA. 2008. *Material International Space Station Experiment (MISSE) home page*: NASA. http://misse1.larc.nasa.gov/

Pidgeon, D. J., R. L. Corey, B. Sauer, and M. L. Day. Two years on-orbit performance of SPT-100 electric propulsion. AIAA 2006-5353, 24th AIAA International Communications Satellite Systems Conference (ICSSC), San Diego, CA, June 2006.

Pisacane, V. L. 2008. *The space environment and its effects on space systems*. AIAA Education Series. Reston, VA: American Institute of Aeronautics and Astronautics.

Polk, J. E., et al. (2001). Performance of the NSTAR ion propulsion system on the Deep Space One Mission. AIAA 2001-0965, 39th AIAA Aerospace Sciences Meeting & Exhibit, Reno, NV.

Prebola, J. L., C. R. Menako, D. S. Crews, and D. H. Crider 2009. Addition of ultraviolet radiation to the AEDC 1559 chamber. AIAA 2009-0558, 47th AIAA Aerospace Sciences Meeting and Exhibit, Orlando, FL.

Acknowledgments

The research reported herein was performed by the Arnold Engineering Development Center (AEDC), Air Force Materiel Command (AFMC), and ATK Space Systems, Inc. Work and analysis were performed by personnel of ATK and ATA Inc., AEDC Group, technical services contractor for AEDC. Further reproduction is authorized to satisfy needs of the U.S. Government.

2010 ITEA Journal *Themes*

The ITEA Publications Committee has established themes for the 2010 issues of *The ITEA Journal* and invites articles in the following areas:

The Role of T&E in Systems Engineering (March issue). Systems engineering is the engineering of complex systems and is intrinsically multi-disciplinary just as test and evaluation (T&E) are. Systems engineers, with their broad view of a program, are in a unique position to diagnose problems in the event of system failure. In T&E the operational requirements of a system must be decomposed to technical requirements and a strategy developed for measuring parameters that can lead back to assessment of mission performance. Systems engineering supplies the process and tools and along with integrated testing is the foundation for future T&E. Integrated testing is the collaborative planning and collaborative execution of test phases and events to provide shared data in support of independent analysis, evaluation, and reporting by all stakeholders. The purpose of integrated testing is to identify system deficiencies early, comply with accelerated schedules, and reduce cost. Design of experiments enables an efficient test design considering all key factors and conditions affecting performance. This issue examines all aspects of systems engineering as well as design of experiments; T&E workforce and training; verification, validation, and accreditation; standards, metrics, data, analysis, and more. ***(Manuscript deadline: December 1, 2009)***

User-Centric Systems (June issue). Systems of systems and network-centric systems are viewed as force multipliers deriving from the mutual connectedness of the elements and the perceived value of timely, critical information. Yet the right information provided to the right person at the right time does not guarantee success. The human is the key and in emerging complex systems of systems the human is more than an operator, and instead is part of the system, a node in the network. Testing a system includes objectively testing the user and requires characterizing human performance. Systems can become so complex that training the user to operate them is no longer possible; rather the systems must be designed with human limitations in mind. This issue examines cognitive performance and measures in addition to traditional form, fit, and function; instrumentation; personal protective equipment; human-machine integration; and situation awareness. ***(Manuscript deadline: March 1, 2010)***

Simulation – Where is T&E Today? (September issue). Simulation is not new and is known by terms such as modeling & simulation and live-virtual-constructive simulation. In one form or another it has been around T&E for more than 20 years and spawned simulation-based acquisition in the Department of Defense, simulation-based design in industry, and a host of other initiatives and hopes. Yet the predictions and expectations have not been realized and the capabilities are often oversold. Where is simulation today in the business of T&E? What has prevented realizing the full power of simulation, what needs to change, or have we arrived already? This issue looks at simulation past, present, and future in T&E and addresses technology, policy, history, success stories, and lessons learned; as well as simulation experience in operational testing, training, design, and other applications. ***(Manuscript deadline: June 1, 2010)***

Cyberspace Test and Evaluation (December issue). Cyberspace is the fifth combat domain, beyond air, land, sea, and space and is the realm of computers, networks, and software. The terrain of cyberspace is not physical but is virtual and ever in flux as network topology and system connectivity dynamically change. In the Department of Defense (DoD) the importance was recognized by creation of cyber commands. Beyond the DoD, cyberspace encompasses commercial networks, the communications industry, power distribution, commerce, transportation, and nearly everything that touches our lives and business today. Systems and networks are subject to continual attacks including spam, phishing, viruses, Trojan horses, worms, root kits, spyware and other malware, and distributed denial of service. Cyber-crime is multi-jurisdictional and spam is being replaced by scam. This issue looks at cyber-infrastructure, data-driven security, information assurance, information operations, electronic warfare, network electronic attack, and other cyber-threats and defenses. ***(Manuscript deadline: September 1, 2010)***

In addition: T&E articles of general interest to ITEA members and *ITEA Journal* readers are always welcome. Each Issue includes specialty features, each 2-3 pages long: **"Featured Capability"** describes unique, innovative capabilities and demonstrates how they support T&E; **"Historical Perspectives"** recall how T&E was performed in the past, or a significant test or achievement, often based on personal participation in the "old days" of T&E.; **"TechNotes"** discusses innovative technology that has potential payoff in T&E applications or could have an impact on how T&E is conducted in the future. **Interested authors**: should submit contributions to the **ITEA Publications Committee Chairman (itea@itea.org, attn.: Dr. J. Michael Barton)**. Detailed Manuscript Guidelines can be found at **www.itea.org** under the ITEA Publications tab.

ITEA Journal 2009; 30: 355–359

ATEC's Approach to Integrated Testing

James J. Streilein, Ph.D and N. Dianne Luna
U.S. Army Test and Evaluation Command, Alexandria, Virginia

Integrated testing is not new to anyone in the acquisition/testing field. When asking members of the Service testing communities if they are doing integrated testing, the reply is almost always going to be affirmative. With that said, what do they mean by integrated testing? Historically, different people have offered different definitions.

Key words: ATEC System Team; Ballistic Missile Defense System; combined DT/OT; coordination; integrated DT/OT; planning testing strategy.

In an effort to provide a common point of reference, the Office of the Secretary of Defense (OSD) provided a definition of integrated testing to the Services in 2008. They defined it as

> *"…the collaborative planning and collaborative execution of test phases and events to provide shared data in support of independent analysis, evaluation and reporting by all stakeholders particularly the developmental (both contractor and government) and operational test and evaluation communities."* (OSD, 2008)

U.S. Army Test and Evaluation Command (ATEC) has further refined this definition to obtain a better understanding of where we can go beyond integrated testing. When talking about integrating testing, ATEC uses the terms "Integrated" Developmental Test and Operational Test (DT/OT) as well as "Combined" DT/OT. Integrated DT/OT, a special case of a Combined DT/OT, is a single-phased event that generates data to address developmental and operational issues simultaneously under operational conditions. The execution strategy for this event is based on the requirements of the program. Combined DT/OT is a single event that produces data to answer developmental and operational system issues. A Combined DT/OT is usually conducted as a series of distinct DT and OT phases at a single location using the same test items. For the case in which a single phase can be used to simultaneously meet developmental and operational issues, this testing will be referred to as an Integrated DT/OT. Combined DT/OT and Integrated DT/OT are encouraged to achieve time, cost, and resource savings. However, they must not compromise DT and OT objectives in accordance with the Defense Acquisition Guidebook.

Roadmap

ATEC has the distinct advantage among the Service Operational Test Agencies (OTAs) of having the developmental testers, operational testers, and the system evaluators all organized under one command. ATEC utilizes a highly effective team structure for each evaluated system. This team structure is labeled the ATEC System Team (AST). The team is composed of personnel from each subordinate command activity (SCA) within ATEC. The core members include the developmental tester, operational tester, and the evaluator. The AST has a Chair who is the command's lead person on that system.

The AST plans, manages, and coordinates test and evaluation (T&E) for assigned systems. An AST is established upon notification of a new requirement from any source or receipt of justification for new requirements. The office of the ATEC Deputy Chief of Staff for Operations coordinates the formation of the ASTs across ATEC.

The AST identifies and resolves system T&E issues and, when necessary, elevates unresolved issues up through the chain of command. The AST presents a single coordinated ATEC position at the T&E Working-level Integrated Product Team (T&E WIPT) meetings. The AST has many responsibilities. For example, the AST is responsible for the synchronization and integration of all ATEC efforts for the assigned system; speaks with one voice when interacting with organizations outside ATEC; reviews all requirement documents, request for proposals, and statement of works. The AST Chair leads all AST activities.

The AST members (both testers and evaluators) are involved early in the T&E planning activities in order to offer their expertise and begin identifying resource requirements. The AST ensures the T&E strategy is aligned with and supports the approved acquisition strategy, so that adequate, risk-reducing, nonredundant gathering of T&E information is provided to support decision events.

The evaluation strategy typically utilizes ground test activities, where appropriate, to include hardware-in-the-loop simulation, prior to conducting full-up system-level testing in realistic environments. The required technical progress includes reliability, desired capabilities, and satisfaction of critical operational issues and criteria to mitigate technical and manufacturing risks. This will increase the likelihood of operational testing and evaluation (OT&E) success by testing in the most realistic environment possible. In addition, the AST will assess system-of-systems command, control, communications, computers, intelligence, surveillance, and reconnaissance prior to operational testing to ensure that interoperability under representative load conditions will represent stressed OT scenarios.

The AST must keep ATEC and the SCA leadership informed by conducting a series of management reviews of planning, progress, and results; continually updating program status and milestones in the ATEC Decision Support System.

Methods/processes

Planning for T&E begins at the earliest stages of the development of user needs, science and technology, system requirements, development, and acquisition processes. Evaluators on the AST participate in the Integrated Concept Development Team review of the Initial Capabilities Document when a new system or new technology is being considered for development.

By policy, every T&E strategy will collaboratively plan all testing and Modeling and Simulation (M&S) activities as an efficient continuum to include appropriate use of integrated/combined testing. Both developmental and operational testers, in concert with the system evaluator, assist the Materiel Developer (MATDEV) and Combat Developer in credible events as appropriate to the program. Early involvement can significantly reduce test time and cost through comparative analysis, data sharing, and use of all credible data sources.

The primary purpose of the T&E WIPT is to optimize the use of appropriate T&E expertise, test assets, targets, instrumentation, facilities, simulations, and models to achieve test integration, thereby reducing costs to the Army and decreasing acquisition cycle time. The T&E WIPT supports the collaborative T&E strategy, resolves issues, and assists the Program Manager (PM)/MATDEV in developing and coordinating the Test and Evaluation Master Plan.

Expectations

What is expected by the overall acquisition community with the increase of integrated testing? Program Executive Officer (PEO)/PMs expect to save time and money. Because of that, integrated testing is very attractive to them. However, with this type of integrated DT/OT, it is very unclear (from the tester's perspective) who is in charge of the planning, execution, and reporting.

Current OSD guidance is that developmental and operational test activities shall be integrated and seamless throughout the system's life cycle. This guidance is understood, but very difficult for test agencies to implement due to the lack of resourcing. OSD expects better results from the systems acquisition as demonstrated via higher pass rate in initial OT&Es.

Challenges

- Approval of plans—Director OT&E approves operational test plans but they only provide advice on developmental testing. With an integrated DT/OT, who would approve the plans?
- Changing cultures—Traditional ways of planning, executing, and reporting test events must change into a more cohesive and cooperative process. How do we decide who is in charge of each test event?
- Competition for services between the test centers—If both DT and OT assets are being used for test events, will there be enough assets to fulfill DT-specific or OT-specific events?
- Maintain the "independent" evaluator—Will the evaluator be able to remain independent if they are integrated into the developmental cycle early in the process?
- ATEC's early involvement in DT—Challenges to ATEC's early involvement (DT) need to be overcome (for example, negative OTA perception by the PMs and increased personnel and funding requirements). How does ATEC convince PMs that being involved early will help the process, not hinder it?

Benefits/lessons learned

The acquisition community needs to be aware that early involvement by ATEC leads to a high return on investment ROI for the PM. ATEC has proven this by having success after success with acquisition programs. Over the past few years, ATEC has many success stories where early involvement by the testing organization has saved time and money for the PM.

Figure 1. M915A5: In an effort to gain early operational feedback, two 88M Soldiers accompanied test drivers on the cross-country drive from Portland, Oregon, to testing grounds at Aberdeen Proving Ground (APG), Maryland. Each M915A5 (above) was instrumented in order to fully utilize all mileage in the final reliability evaluation.

Figure 2. Mine Resistant Ambush Protected (MRAP): Military Subject Matter Experts from the U.S. Army, Marine Corps, Air Force, and Navy assisted Army Test and Evaluation Command (ATEC) providing ground personnel at Aberdeen Test Center (ATC) in generating data for the early assessment of the MRAP vehicles.

Examples

The following examples of ATEC's approach to integrated testing provide a clear understanding of the above policy statements:

1. The M915A5 is a commercial purchase standard acquisition of a line haul semi-tractor with some add-on military equipment for fittings for up-armoring packages *(Figure 1)*. The M915A5 T&E program will start with a cross-country drive from the West Coast manufacturer to Aberdeen Proving Ground, Maryland, on the East Coast. DT instrumentation and data collection will be utilized before delivery to the DT location. During the DT, the reliability/durability testing will be accomplished to the operational mode summary/mission profile usage with a mix of civilian and military drivers. There will be a traditional pure OT, but reliability data from both the developmental and operational testing will be used in the evaluation of reliability and durability, thereby allowing a reduction of the OT miles to fewer on each truck than possibly needed to address reliability.
2. On rapid acquisition programs, in comparison to traditional acquisitions, ATEC ordinarily can only get a very limited single test in one location. For example, the initial Mine Resistant Ambush Protected (MRAP) vehicle programs had a very limited functional checkout as opposed to a full DT, a very limited user operational period fit in around the DT schedule and a parallel set of survivability sets *(Figure 2)*. Based on an integrated assessment by ATEC and U.S. Marine Corps Operational Test and Evaluation Agency, MRAP decision makers were able to decide which specific variants to buy and field in specific quantities. Based on continuing cycles of buying, testing in both DT- and OT-like environments, as well as Theater usage as reported by the ATEC Forward Operating Assessment team; MRAPs have been continually produced, improved, tested/evaluated, fielded, and had field feedback leading to what is viewed as a success story in quickly increasing Soldier survivability. Of course this quick reaction approach also has had negative consequences in which the field has experienced problems that would have been found and fixed via a more standard testing approach. These realized risks have had an even larger impact because of the many variants of models from different manufacturers that have been fielded.
3. Certainly the most integrated T&E program being worked by ATEC is for the Ballistic Missile Defense System (BMDS) *(Figures 3 and 4)*. The BMDS is a system-of-systems made up

Figure 3. Ballistic Missile Defense System (BMDS): U.S. Army Test and Evaluation Command (ATEC) is the lead of a multi-service Operational Test Agency (OTA) providing an independent operational report of the integrated BMDS capability to defend the United States, its deployed forces, friends, and allies against ballistic missiles of all ranges and in all phases of flight. (Approved for Public Release 2009 MDA Book).

of developmental programs, upgraded programs to existing programs, and integration of existing systems. A list of the BMDS currently fielded systems includes Ground-Based Mid-course Defense (GMD), Aegis BMD Surveillance and Track Destroyers, and Sea-Based X-Band (SBX) radar. Some BMDS systems fielded for defense against Regional/Theater ballistic missions are Aegis Engagement Cruisers and Destroyers with Standard Missile–3 (SM-3) interceptors and Patriot PAC-3. The Theater High Altitude Area Defense system is being developed and tested under BMDS. The BMDS has been removed from the normal acquisition rules by an act of Congress. However, as the lead OTA, ATEC, with support of the other involved Service OTAs, has developed a fully integrated T&E approach, which is making use of collaborative testing on multiple occasions. The BMDS will be fielded in capability sets and each capability set has a planned T&E campaign that is built from pure DT events, operationally realistic events, partial OT events with real users in the loop, pure OT ground and flight test periods, and real users in real operational periods leading to final user acceptance as a mission capable war fighting system.

Summary

An integrated or combined testing approach may offer an effective means of shortening the time required for testing and may achieve cost savings. However, when this approach is used, it is imperative that extensive coordination is done well in advance of the event in order to ensure all (DT and OT) requirements are addressed. The utilization of the AST and the T&E WIPT can help with this coordination by providing mutual support and sharing mutual beneficial data across the acquisition community. Every commodity area, every system, and every system phase needs a tailored T&E program. Every T&E program ought to be looked at to see if the best mix of separate, integrated and collaborative testing has been built into the program. ATEC early T&E planning reviews take a serious look at each program from an integrative approach to T&E. ATEC lessons-learned have shown the value of this approach. We expect to make further advances on this path in the

Approved for Public Release. 08-MDA-3982 (25 NOV 08)

Figure 4. Integrated Developmental Test and Operational Test (DT/OT).

future. As discussed, there are difficulties in fully achieving integrated testing. However, as a community of acquisition and T&E experts, there is no doubt we will be able to make substantial progress toward our ultimate goal of providing systems to our soldiers that are proven to meet requirements within established time and cost constraints. ❑

N. DIANNE LUNA has been a policy analyst at the U.S. Army Test and Evaluation Command (ATEC) since October 1996. She earned a bachelor of science degree in management and business from National-Louis University, Evanston, Illinois. Mrs. Luna has over 30 years of federal service with emphasis on organizational and technical analysis, planning, policy, compliance, and implementation of military issues and/or programs. E-mail: dianne.luna@us.army.mil

JAMES J. STREILEIN, Ph.D., has been the executive technical director/deputy commander, of the U.S. Army Test and Evaluation Command (ATEC) since April 2007. He earned a doctor of philosophy degree in mathematics from Pennsylvania State University, University Park, Pennsylvania, and a bachelor of science degree in mathematics from Carnegie Mellon University, Pittsburgh, Pennsylvania. Dr. Streilein provides oversight and technical direction to all command efforts to include testing, evaluation, modeling and simulation, and instrumentation. Dr. Streilein has over 35 years of federal service starting from 1974 to 1996 at the U.S. Army Materiel Systems Analysis Activity (AMSAA) in a variety of technical and managerial positions. He became a member of the Senior Executive Service in August 1991 upon selection as the Chief, Reliability, Availability, and Maintainability Division of AMSAA. In the 1996 reorganization of Army test and evaluation, Dr. Streilein was selected as the first director of the Evaluation Analysis Center (EAC) of the Operational Test and Evaluation Command (OPTEC). In September 1999, the Army reorganized test and evaluation and Dr. Streilein was selected as the first director of the newly formed Army Evaluation Center (AEC) of ATEC. E-mail: james.streilein@us.army.mil

2009 ITEA ANNUAL SYMPOSIUM

Aero & Defense TEST 2009
Air.Space.Land.Sea

Baltimore Convention Center
September 28 - October 1, 2009

Co-located with

INTERNATIONAL TEST & EVALUATION ASSOCIATION
2009 ANNUAL SYMPOSIUM

The 2009 International Test and Evaluation (ITEA) Annual Symposium
will be an absolutely unique event that brings together the significant technical and exhibit programs of ITEA and the newly expanded Aero & Test Defense 2009 Exposition. We look forward to seeing you in Baltimore
– **RADM Charles 'Bert' Johnston, Retired**, Symposium Chair

"Sharing Emerging Best Practices to Improve Test and Evaluation to Meet Future Challenges"

Keynote Speakers

Mr. Frank J. Anderson, Jr., President, Defense Acquisition University

Dr. Ashton Carter, Under Secretary of Defense, Acquisition Technology & Logistics (AT&L) (Invited)

Dr. Werner J.A. Dahm, Chief Scientist of the Air Force

Major General Tom Henry Knutsen, Defense Attaché

Maj Gen (Ret) Michael Kostelnik, Assistant Commissioner, U.S Customs and Border Protection Office of Air and Marine

Mr. Gordon M. Kranz, Director, Systems & Software Engineering, Deputy Under Secretary of Defense, AT&L (Invited)

Sue C. Payton, Assistant Secretary of the Air Force for Acquisition

Panel Discussions with Q&A

- T&E Visions: Investing in Future T&E Needs
- Operational Test Agency Commanders: Key Best Practices that Have Improved OT&E
- Visions for T&E Workforce
- International Collaboration on T&E Issues

Tutorials Being Offered
(4-hour sessions on Mon.)

- Systems Engineering and T&E
- Testing 1-2-3: A Fundamental Course
- T&E for Unmanned Autonomous Systems: An Overarching Perspective
- Introduction to Concepts for Analyzing System and System of Systems in Complex Environments
- Fundamentals in Leadership
- Cross Command Collaboration Effort; LVC M&S, Data Collection, Analysis, and Instrumentation – Cross Collaboration Effort
- Design of Experiments – The Bridge to Systematic Innovation

NEW Tutorials
(2-hour sessions on Tues. & Wed.)

- MIL STD 810F Changes and Challenges
- Generational Diversity Dynamics for Leaders
- Experimental Designs for Complex Test Systems
- DoD 5000 Update – Defense Acquisition University
- Fundamental Statistical Concepts and Methods Needed in a Test and Evaluation Toolkit
- Random Vibration and Shock Testing

Technical Tracks

New Expanded Selection to Meet Attendee Requests

- Future Technologies and Facilities Executing Network-Enabled Battlefield T&E Challenges Future T&E Challenges
- Testing Complex Systems
- Lessons Learned and Best Practices in T&E
- Flight Testing
- T&E Using Wind Tunnels
- Test Methods and Performance Evaluations for Intelligent Systems
- Systems Engineering, Systems Architecture, and Test Planning
- Chemical-Biological-Nuclear
- Use of Modeling and Simulation in T&E
- Interoperability Testing
- Testing Non-Kinetic Weapons
- Component Testing
- Testing New Warfare Systems: Cyber, Counter IED, and Information Assurance
- Design and Test of Complex Systems
- Knowledge Management within Test and Evaluation
- Improvements in Air Platform/Avionics T&E Capabilities
- Testing Autonomous and Tele-Operated Systems

Golf

The Annual Golf Tournament will be held at Mountain Branch Golf Course, 1827 Mountain Road in Joppa, Maryland, on Monday, September 28th with a Shot Gun start at 1 pm. Hole sponsorships are available for $200. Contact Rich Pace for more information at **rpace2@csc.com** or **301-737-8111**. Registration forms are available on the ITEA website.

Academia Day and Career Zone

On Thursday, October 1 from 9:00 am until 1:00 pm, ITEA and the organizers of the Aero & Defense Test 2009 exhibition and forum are hosting a significant academic program with the opportunity for students and educators to tour the exhibition hall, talk with exhibitors, and attend a seminar on Test and Evaluation disciplines, opportunities, and careers.

During Academia Day, organizations will have the opportunity to tap into the latest talent pool of aerospace/aeronautical engineering / mathematics and computer science majors at the Career Zone. These undergrad/post grad/doctoral candidates are all looking to gather information for upcoming positions in the industry. Send your Human Resources specialist to take advantage of this opportunity. Space is limited and a minimal fee of $250 is required for booth space. Contact Ms. Nancy Roediger at **nancy.roediger@leadingedgeem.com** to reserve your space today!

Industry and Government Leaders: The academic program will provide a rare venue for important introductions for our future workforce to some of the world's leading corporations and Government agencies toward establishing mutually beneficial networking relationships. Take advantage of participating in the Career Zone. These introductions could then lead to internships, training opportunities, and future employment.

Exhibits

The Exhibit hall is 80% full. Act now to book your booth at a special ITEA corporate member rate! Contact Leading Edge Media at **800-913-5022** or Mr. Bill Dallas, ITEA Exhibit Manager at **703-631-6226** or wdallas@itea.org to book or discuss options. Visit **www.aerodefensetest.com** for more details.

Sponsorships Available

ITEA would like to thank the sponsors who have already pledged for the 2009 program. If your company is interested in participating, please contact Mr. Bill Dallas at **wdallas@itea.org**.

PLATINUM: Northrop Grumman Corporation, Scientific Research Corporation – SRC, Wyle

GOLD: CALCULEX, Inc., EWA Government Systems, Inc., Science Applications International Corporation - SAIC

SILVER: Advanced Systems Development, Inc., JT3, LLC

BRONZE: Advanced Test Equipment Rentals, Cobham Analytic Solutions, DRS Training and Control Systems, Inc., Georgia Tech Research Institute - GTRI, Quadelta, Inc.

Accomodations

Sheraton Baltimore City Center Hotel
101 West Fayette Street, Baltimore, MD 21201
866-837-5182

ITEA has established a room block for our attendees. This hotel is in a prime location to the Convention Center and offers our attendees a special room rate of $159 *including a full breakfast* until Sept. 7. We encourage you to make your reservations early to take advantage of this rate.

On line registration available

WWW.ITEA.ORG

ITEA Journal 2009; 30: 361–368

ATEC Initiatives in Response to the Office of the Secretary of Defense Policy Guidelines for Test and Evaluation

Frank J. Apicella, Kerry W. Wyant, Ph.D. and Christopher M. Wilcox
U.S. Army Evaluation Center,
Aberdeen Proving Ground, Maryland

The Army Test and Evaluation Command (ATEC) is using mission-based test and evaluation (T&E), among other initiatives, to implement the Department of Defense T&E guidelines. Mission-based T&E focuses on the identification and alignment of system components and functions with the tactical missions and warfighting functions/tasks that the system supports. The approach facilitates testing in an "operationally realistic" environment and evaluating "in the mission context at the time of fielding." Further, it facilitates the assessment of system functionality, the assessment of the effect of system functionality on operational capability, and the assessment of the capability of the warfighter to accomplish mission tasks.

Key words: Capabilities and limitations assessment; data management & documentation; early test; experimental design; integrated DT and OT; mission-based T&E; operational realism; typical user.

Section 231 of the National Defense Authorization Act for Fiscal Year 2007 directed a review, and amendment, if necessary, of defense acquisition test and evaluation (T&E) policies and practices. In response to Section 231, the Office of the Deputy Under Secretary of Defense for Acquisition, Technology and Logistics (USD [AT&L]) produced a Department of Defense (DoD) Report to Congress on Policies and Practices for Test and Evaluation, dated July 17, 2007, in order to satisfy this legal mandate. This report discussed eight key T&E principles.

The intent of this article is to describe how the U.S. Army Test and Evaluation Command (ATEC) is implementing these eight principles. The principles along with ATEC initiatives for implementing the principles are provided in *Table 1*.

The first of these principles encourages the T&E community to broaden its focus away from pass/fail, final milestone decision-oriented assessments. It suggests providing periodic assessments to the materiel developer with emphasis on progress made versus requirements met/not met. This method had been successfully applied in support of the Ballistic Missile Defense System and Future Combat Systems programs. A similar approach is used for Rapid Acquisition Program Systems, where capabilities and limitations assessment reports are updated every 6 months with highlights of the ongoing T&E efforts.

The second principle attests to the value of very early test events to obtain initial insights into the strengths and weaknesses of early design concepts. It also implies that more emphasis on early experimental design for these test/experimentation events should result in testers being able to more efficiently gather information on system performance. The increased analytical rigor that design of experiments brings should result in better early estimates of system performance and potential.

The ultimate goal of integrated testing is to ensure that all stake holders (Program Manager [PM], developmental testers, operational testers, and evaluators) collaborate so that all can use the data from any test event to satisfy their needs. There are complicating factors, which make fully attaining this goal difficult; for example, developmental testing (DT) is characterized by the use of a "test-fix-test" process, which allows the system design to constantly be improved and refined; final operational testing (OT) is usually conducted with systems that are nearly fully mature and are "production representative." However, we have made great strides in approaching this goal over the last several months. Many more DT events are being conducted with what we refer to as more "operational flavor." Soldiers are participating in many

Table 1. How to achieve the Office of the Secretary of Defense policy guidelines.

Principle	How to Implement
Measure improvements to mission capability and operational support.	Conduct periodic or annual assessments of observed levels of improvement. Each assessment becomes the "rolling baseline" for the last.
Experiment to learn strengths, weaknesses, and the effect on operational capabilities.	Increased emphasis on DT/OT experimental design: test under controlled conditions to determine capabilities and limitations.
Integrate DT and OT.	Consider using a single test event to address both DT/OT T&E requirements. Use an OT experimental design during DT. Give DT an operational flavor with appropriate threats and Soldier operators.
Begin early, be operationally realistic, and continue throughout the life cycle.	Get involved very early in the development process. Liaison with the PM and contractors. Coordinate and assist the combat developer in developing requirements and rationale.
Evaluate in the mission context expected at the time of fielding.	Ensure an operational environment with realistic and representative threats that are expected at the time of fielding plus 5 years. Apply a "mission-based test and evaluation" (MBT&E) focus to the system evaluation.
Compare to current mission capabilities.	Whenever possible, include in the OT experimental design a comparison to the baseline system. Factors such as terrain, mission, operational mode summary and mission profile (OMS/MP), and tactics, techniques, and procedures (TTP) must be held constant between the current system and the system under development. A model-test-model approach may provide the best estimates of system comparison.
Use all available data and information.	Develop a data management strategy and define the data early. Use any credible data source: historical data as well as data from contractor testing, DT, Mod/Sim, etc. for system evaluation. Data must be properly managed and documented (using a data model) to facilitate understanding of differing sources of like data.
Exploit benefits of Modeling and Simulation	Supplement/complement live testing with virtual and constructive simulations. Plan for early verification, validation, and accreditation of models and resource the effort.

DT events either as observer subject matter experts (SMEs) or as system operators if safety constraints allow.

Early involvement of the T&E community in system development has many potential benefits. This allows the Army Evaluation Center (AEC), for example, to provide the data requirements to testers early on, so they can structure test events to satisfy OT as well as DT evaluation requirements. Some benefits that we have seen include the following:

- contributions in drafting the T&E portions of the system specifications,
- technical support for the source selection process,
- assistance in the development of more realistic requirements,
- improved contractor understanding of requirements,
- more informed and balanced assessment reports,
- problems identified earlier and fixed more economically, and
- a much more timely developmental process.

The Stryker program is an excellent example of the efficiencies gained through early involvement. Hundreds of improvements were identified and fixed during an extensive, early T&E program for the initial eight Stryker variants. The system was developed and fielded quickly and in record time.

Evaluating in a mission context has many benefits. In the past, evaluation products have frequently highlighted technical performance requirements that have not been achieved. The senior decision maker "so what" question that sometimes has proven difficult to answer is "What are the operational mission impacts of not meeting these requirements?" To facilitate being able to address this question and others, ATEC has adopted a new mission-based T&E (MBT&E) methodology for all its programs.

We do not always have the resources to conduct a live test in which the system under test is compared with the baseline system. Also, controlling some factors and conditions (level of unit training, different leadership styles, limited sample sizes, learning effects, etc.) for such events makes them difficult to execute and analyze. ATEC has adopted a "model-test-model" approach as a preferred evaluation strategy for many of these types of system tests.

Traditionally, ATEC has always endeavored to use all possible data sources in system evaluation. Utility of some of these data has been improved by closer, early coordination between the contractor and the T&E community. Adoption of common data definitions is a new initiative, which will greatly facilitate the utility of this data exchange.

The utility of modeling and simulation (Mod/Sim) in support of the T&E process has expanded markedly. Mod/Sim has been used effectively as a "wrap-around" to simulate large, joint forces for major tests. Mod/Sim tools have extended specialty analysis areas such as survivability; reliability, availability, and maintainability (RAM); and integrated logistic support (ILS). Force-on-force models have been used to augment and extend system assessments. This is a large growth area for T&E.

Acceptance of these principles has prompted areas of increased emphasis and several new T&E process innovations. The rest of this article focuses on four of the above-mentioned initiatives: MBT&E, experimental design, data management and documentation, and integrating developmental and operational testing. These best practices are essential to improving our strategy for system evaluation and to enabling the identification, definition, structuring, and prudent resourcing of the areas of study and the data needed to test and evaluate an Army system. ATEC System Teams (ASTs) responsible for the T&E of Army systems address these initiatives in their T&E plans.

MBT&E

MBT&E is a methodology that seeks to assess the "strengths and weaknesses" of a system and its components and "the effect on operational capabilities." Purposes and requirements for the methodology were derived from recent T&E policy-shaping initiatives such as the initiatives shown in *Table 1*. The overall development of MBT&E has been done using an interagency working group that assessed individual mission-based concepts, developed a combined methodology from these concepts and from T&E policy requirements, coordinated the methodology with organizations throughout the acquisition community, executed pilot projects, and incorporated comments and lessons-learned from the coordination and execution efforts. The result is a baseline MBT&E methodology that was published in January 2009 and is in use to develop ATEC T&E strategies today. Continuing methodology development is leading to improved efficiencies and synchronization of the efforts of the Services' Operational Test Agencies (OTAs) and the Director, Operational Test and Evaluation (DOT&E).

MBT&E consists of a framework and a procedure. The MBT&E framework is composed of four main elements: mission, system, evaluation, and test. The methodology links the information in these elements to create a T&E strategy that associates the functionalities a system must possess with the warfighting functionalities/capabilities it must support. Further, it identifies the associated T&E measures and standards, operational conditions, and the data required to support the evaluation. The MBT&E procedure is a step-by-step description of activities that guide the development and execution of the T&E strategy. It describes the activities for the following:

- analyzing unit missions and tasks,
- analyzing the materiel system of systems components and functionality,
- developing T&E measures and data sources,
- executing the evaluation,
- and reporting of unit and system effectiveness, suitability and survivability.

The framework and procedure provide a method to conduct the following:

- planning of integrated and synergistic T&E requirements,
- continuous evaluation from program inception through deployment, and
- evaluation simultaneously focused on system performance and warfighter capabilities in the mission context.

Table 2 shows an example of the MBT&E framework applied to an unmanned aerial system (UAS) unit conducting an attack by fire mission. The mission is broken down into three "phases": (a) conduct reconnaissance, (b) direct attack, and (c) indirect attack. Each phase is further broken down by the unit tasks necessary to accomplish the mission. The mission, phases, and associated unit tasks are shown in the column headings. The system components and functions are shown in the row headings. The system functions are linked to the tasks by the addition of an evaluation measure in the intersecting box in the table. The operational context of the task is applied to the system functions through these linkages and is used to

Table 2. Mission-based test and evaluation (MBT&E) framework example—unmanned aerial system.

UAS: Unmanned Aerial System
OM: Operational Measure
TM: Technical Measure

SYSTEM COMPONENT		SYSTEM FUNCTION	MISSION/TASKS					
			Attack by Fire (ART 7.5.1)					
			Conduct Reconnaissance			Direct Attack		Indirect Target
			Arrive in Area of Operations (AO)	Detect & Locate Surface Targets (ART 3.2)	Intel Support to Targeting (ART 2.4.1)	Lethal Direct Fire Against Surface Target (ART 2.4.1)	Provide Fire Support (ART 3.3)	2.3.2 Indirect Fire, Surface-to-Surface (ART 3.3.1.1)
UAS	Airframe	Navigate to AO	- Time to Arrival in AO - Navigation Accuracy					
		Execute search pattern	- Time on Station					
	Communication Equipment	Send call for fires					- % Fire Calls Sent/ Received	
	Sensor	Detect target		- % Targets Detected - Detection Range				
		Locate target		- Location Accuracy				
		Identify target			- % Targets Identified			
	Laser Designator	Designate Target				- % Targets Hit - Spot Jitter		
	Missile	Destroy Target				- % Targets Killed		
M109 Paladin	Communication Equipment	Receive Firing Order					- Time to Process Fire Order	
	Firing System	Fire Round						- % Targets Killed

develop the operational conditions addressed in the measure. This example illustrates the design of a system of systems evaluation by looking at the performance of the UAS to conduct a direct attack and to support an indirect attack using a M109 Paladin. It also illustrates the use of tasks in the Army Universal Task List to provide a doctrinal reference and help establish measures and standards used in the evaluation.

The achievement of an integrated MBT&E strategy comes with the analysis of data sources to support the evaluation of the measures. *Table 3* shows the measures from *Table 2* and the linkages to all possible data sources. Possible data sources include contractor tests, Mod/Sim, hardware-in-the-loop, and interoperability certification. Integrated use of the data is achieved by identifying the right type of data at the right time in the development program. Synergistic use of the data is achieved by combining the data from more than one source into an "accumulated" knowledge of system and task performance.

The execution of the MBT&E pilot projects and ongoing implementation of the MBT&E methodology within ATEC have yielded a long list of lessons learned and observations. Lessons learned are being applied to improve the development and robustness of MBT&E strategies. Some key observations include the following:

- Impact of system performance on mission task is being achieved.
- Task context is critical to designing integrated T&E and enabling more operationally relevant developmental test.
- Improved efficiencies and aligning of expectations are achieved by combat and materiel developer participation in the MBT&E process.
- MBT&E is similar to the processes used by the U.S. Navy's Operational Test and Evaluation Force, and the U.S. Air Force's Operational Test and Evaluation Center.

The last observation has led to an effort within the Services' OTAs and DOT&E to develop a common

Table 3. MBT&E framework example—measure to data source linkages.

MEASURE	DATA SOURCE: Contractor Tests	Modeling & Simulation	Interoperability Testing	Hardware-In-The-Loop	DT Flight Tests	Operational Tests
UAS Time to Arrive in Area of Operations					x	x
UAS Time on Station	x					x
UAS % Targets Detected					x	x
UAS % Targets Identified					x	x
UAS % Targets Hit		x			x	x
UAS % Targets Killed		x				
M109 Paladin % Targets Killed		x				
M109 Paladin Time to Process Fire Order			x			x
UAS Navigation Accuracy	x					
UAS Sensor Detection Range				x	x	
UAS Target Location Accuracy					x	
% Fire Calls Sent and Received			x			

process with the goals of synchronizing processes, products, and training; and enhancing Joint Services operational testing.

Experimental design

A statement of the problem facing our analysts is essentially this: Given ever increasing resource constraints, assess to the extent possible the effectiveness of an Army system under the full range of operational conditions within which it is intended to operate. In ATEC, we are dealing with this problem by placing greater emphasis on formal experimental designs versus reliance, sometimes on less structured or short "free-play" demonstrations. Experimental designs permit a rigorous structuring of test events such that the full range of operational conditions can be addressed while conserving resources. Moreover, the use of experimental designs to structure test events allows the evaluator to assess from "improvement to mission capability and operational support," to "learn strengths, weaknesses, and the effect on operational capabilities"; and to make comparisons "to current mission capabilities." Tools and procedures used to structure test events include factorial designs, fractional factorial designs, sample sizing, and statistical analysis.

Experimental designs are represented during early command reviews of the evaluation strategy using a matrix showing how the factors and conditions are combined. An important step in getting to the matrix is a factors and conditions table. In this table, each factor, the conditions the factor may assume, and the method of controlling the factor is given.

Table 4. Factors and conditions table template.

Factor (F)	Control	Conditions (C)
F_1	Control Type	$C_1 \ldots C_J$
F_2	Control Type	$C_1 \ldots C_K$
F_3	Control Type	$C_1 \ldots C_L$
.	.	.
.	.	.
.	.	.
F_N	Control Type	$C_1 \ldots C.$

Factors may be systematically varied, uncontrolled, "tactically" varied, or held constant. Tactically varied factors are changed in accordance with unit tactics, techniques, and procedures and have the same status as uncontrolled factors, unless incorporated into the test design.[1] An experimental design requires that at least one factor is systematically varied. If all of the factors in the table are tactically varied (and not incorporated into a test design) or are uncontrolled, it may be very difficult to explain why a system performed well or performed poorly.

Analysts prepare a factors and conditions table and a matrix early in the planning process. A factors and conditions table template is given in *Table 4*. An example is provided in *Table 5*. The system represented in *Table 5* is the Global Position System (GPS) Guided Fuze. Four factors (weapon, temperature, quadrant angle, and munitions) are systematically varied, two factors (range and mission-oriented protective posture [MOPP]) are held constant, and one factor (weather) is uncontrolled. Given the controlled factors, an experimenter could assess the effects of the varied factors on delivery accuracy. The factors and conditions given in *Table 5* are represented in a matrix as a full factorial design in *Table 6* and as a fractional factorial design in *Table 7*. Fractional factorial designs can enable savings in the total number of test executions without giving up any information about the effects of the factors and their first-order interactions. As shown in *Table 7*, there is a 33 percent (12/36) reduction in the number of rounds required to assess the effects of the four systematically varied factors on accuracy. In these examples, each combination of conditions is executed five times to help ensure test conditions are representative.

The use of experimental designs to structure test events allows the evaluator to assess the effects of the full range of operational conditions on evaluation measures, assess improvements to mission capability

Table 5. Global Position System (GPS) Guided Fuze system factors and conditions table.

Factor	Control	Conditions
Weapon	Systematically Varied	M109A6, M777A2*
Temperature	Systematically Varied	Hot, Ambient, Cold
Quadrant Angle	Systematically Varied	High, Low
Munitions	Systematically Varied	M795, M549, M107
Range	Held Constant	90% of maximum
MOPP	Held Constant	MOPP 0
Weather	Uncontrolled	

* The M109A6 is the 155mm self-propelled howitzer known as Paladin; the M777 is a towed lightweight 155mm howitzer.

and operational support, make comparisons to currently fielded equipment, and conserve limited resources.

Data documentation

AEC must be able to communicate their data requirements to the tester; design and create data sets that can be used to analyze the data received from the tester; and ultimately, archive that data that can be used for accrediting models and simulations and future comparative analysis. The role of the system evaluator within the AST is critical. They must ultimately decide what data are needed, how to get that data, and what the source of that data is.

ASTs must also be able to use data from all sources. They must be able to use data from past assessments in order to measure improvements in functionality and operational capabilities. They must also be able to use contractor data and DT data in order to identify deficiencies early in the development process. The crucial first step toward accomplishing these goals is to properly manage the data requirements identified during the requirements analysis.

Well-managed data are visible, accessible, and understandable. Data are made visible if they are properly tagged with descriptors. The DoD Discovery Metadata Specifications can be used to help accomplish making data visible. Data are accessible if they are held in a common area and can be searched and retrieved by the members of the interested community. Data are made understandable through precise, well-formed definitions documented in a data dictionary.

Currently, increased emphasis is being placed on making data understandable. Analysts in AEC are asked to follow naming and defining conventions when developing data dictionaries.[2–4] In the data dictionaries, the analysts name and define the objects and events

Table 6. Global Position System (GPS) Guided Fuze system full factorial design.

Temperature	Weapon Type	Quadrant Angle High			Quadrant Angle Low		
		M795	M549	M107	M795	M549	M107
Hot	M109A6	5	5	5	5	5	5
Hot	M777A2	5	5	5	5	5	5
Ambient	M109A6	5	5	5	5	5	5
Ambient	M777A2	5	5	5	5	5	5
Cold	M109A6	5	5	5	5	5	5
Cold	M777A2	5	5	5	5	5	5

Table 7. Global Position System (GPS) Guided Fuze system fractional factorial design.

Temperature	Weapon Type	Quadrant Angle High			Quadrant Angle Low		
		M795	M549	M107	M795	M549	M107
Hot	M109A6	5		5	5	5	
Hot	M777A2	5	5		5		5
Ambient	M109A6	5	5			5	5
Ambient	M777A2		5	5	5	5	
Cold	M109A6		5	5	5		5
Cold	M777A2	5		5		5	5

being studied (e.g., unit, mission, task, sensor type), and name and define the data elements (e.g., unit identifier, mission date-time, task outcome code, sensor type name) used to describe the objects and events under study.

AEC must be able to communicate data requirements to the tester, use the data received from the tester, and archive data for future use. ASTs must also be able to use data from diverse sources including historical data and contractor data. This is made possible if data are understandable and are documented in a data dictionary that is shared throughout the test and combat and materiel development communities.

Combined and integrated DT and OT

Combining or integrating DT and OT is becoming more and more necessary in order to address the greatest number of evaluation questions with the least use of limited T&E resources. Combined DT and OT describe a single test event that produces data to answer both developmental and operational system issues and is usually conducted as a series of distinct DT and OT phases. Integrated DT and OT describe a single-phased event that generates data to address developmental and operational issues simultaneously under operational conditions. Integrated DT and OT have the potential to answer both DT and OT issues more efficiently in terms of the time and resources normally required by separate tests. It is also the most difficult type of testing to execute as it requires maximum coordination and cooperation among all members of the test community. AEC analysts are required to consider integrating DT/OT in the evaluation strategy while considering the following guidelines:

- Fully consider all safety issues.
- Achieve maximum cooperation among test team(s).
- Consider use of highly experienced, soldier SMEs for early test events.
- Consider use of typical user soldiers for later test events.
- Enhance operational realism of DT events, whenever possible.
- Use a single experimental design to answer both DT and OT questions.

Currently, the most acceptable form of integrating DT and OT is to enhance the operational realism of the DT. This is usually done by considering realistic threats and observing operational conditions and scenarios specified in the Operational Mode Summary/Mission Profile. Sometimes imparting a realistic operational "flavor" to a test is relatively easy. One example would be the varying of message load when testing interoperability between two communication systems. Imparting a realistic operational "flavor" can also be difficult. For example, the required use of soldier operators requires a safety release that may not be available during DT.

A question will always arise as to whether the conditions under which testing occurs are sufficiently operational to permit the use of data generated at a DT event to answer operational issues. The AST will make this judgment. If a single-phased event cannot be used to generate data to address developmental and operational issues simultaneously under operational conditions, a combined event with distinct DT and OT phases can be implemented. Conducting DT with deference to operational realism followed by an operational phase may be much easier to execute.

An experimental design must be used if the goal of testing is to predict how a system will perform under different conditions. An experimental design that accommodates the needs of the developmental tester and operational tester must be used for integrated testing. In contrast, combined DT/OT can employ separate experimental designs for the different DT and OT phases.

Summary

The eight OSD guidelines introduced in the 2007 DoD Report to Congress on Policies and Practices for Test and Evaluation provide a basis to better fulfill the key objectives of Defense acquisition: to acquire quality products that satisfy user needs, to do so with measurable improvements to mission capability and operational support, to accomplish this in a timely manner, and to ensure purchase at a fair and reasonable price. AEC has undertaken to implement these principles by introducing the initiatives in *Table 1*. The initiatives include MBT&E, experimental design, data documentation, and integrating DT and OT. These best practices are essential to our evaluation strategy. They enable the identification, definition, structuring, and efficient resourcing of the areas of study and the data needed to test and evaluate an Army system and are addressed by ASTs early in the planning process. ❑

FRANK J. APICELLA is the technical director for the U.S. Army Evaluation Center. With over 35 years of T&E experience, he has held many key T&E positions. He received his bachelor of science degree from Norwich University and has a master of science degree in administrative science from the Johns Hopkins University.

He is a graduate of the U.S. Army War College, the Harvard University Senior Executive Fellowship Program, the Western Executive Development Institute, and the Federal Executive Institute. He is a recipient of the Meritorious Civilian Service Award. In both 1997 and 1998 he received the Dr. Wilbur B. Payne Memorial Award for Excellence in Analysis. E-mail: frank.apicella@us.army.mil

KERRY WYANT, Ph.D., is the chief of the Methodology and Analysis division of the U.S. Army Evaluation Center. He has over 30 years of experience in T&E and holds a doctor of philosophy degree in experimental psychology from the University of Oklahoma. E-Mail: kerry.wyant@us.army.mil

CHRISTOPHER WILCOX is a senior engineer with the Maneuver Air Evaluation Directorate of the U.S. Army Evaluation Center, Aberdeen Proving Ground, Maryland. He is directly responsible for the planning, execution, and reporting on the system evaluation of aviation-related weapons systems. He has had previous engineering assignments with the Missile Defense Agency, Pentagon, Washington, D.C., and with the U.S. Air Force. He graduated from the University of Tennessee in 1997 with a master of science degree in aerospace engineering. He also holds an associate of science degree from Ocean County Community College, New Jersey, and a bachelor of science degree from Embry Riddle Aeronautical University, Daytona Beach, Florida. E-mail: chris.wilcox1@us.army.mil

Endnotes

[1]The stimulus for a unit to tactically vary comes from some external change in the test unit's physical or tactical environment. If the conditions resulting in a tactical change are made part of the test design, there will be start and finish times for the conditions assumed for a tactically varied factor. Data collected between the start and finish time for a specific condition will have had that condition held constant.

[2]International Organization for Standardization and International Electrotechnical Commission (ISO/IEC) international standard 11179 - *Metadata registries (MDR)*.

[3]DOD Directive 8320.02 Data Sharing in a Net-Centric Department of Defense, April 23, 2007.

[4]DOD 8320.02-G Guidance for implementing Net-Centric Data Sharing, April 12, 2006.

ITEA Journal 2009; 30: 369–374

Theory and Practice of Integrated Test for Navy Programs

Michael O. Said
Office of the Assistant Secretary of the Navy, Research,
Development and Acquisition (ASN[RDA]),
Test and Evaluation, Washington, D.C.

The role of test and evaluation (T&E) is to provide acquisition program decision makers at all levels with information on a weapon system's effectiveness and suitability. T&E supports validation of the weapon system's capabilities and limitations and their associated concept of operations. T&E is an integral part of the systems engineering process, where contractor test, developmental test, live fire test, and operational test all contribute to design understanding, design iteration, and ultimately design confirmation. T&E results are needed to support the formal Department of Defense (DOD) acquisition decision making process from early phases through life cycle management, to include technology development phase, engineering and manufacturing development phase, production and deployment of system phase, and planned product improvements. T&E is most effective and efficient when it is conducted early in a program with direct assimilation to the systems engineering plan and integrated with related activities. In response to findings and recommendations from past T&E process studies and today's acquisition policy, the U.S. Navy has implemented integrated testing within its weapon systems programs and within its T&E enterprise. Results, although at times difficult to measure, indicate that the benefits in terms of cost avoidance and schedule reductions for Navy programs have been positive. Challenges abound, but lessons learned have been identified and best practices are being shared as discussed in this article.

Key words: Collaboration; combined DT/OT; contractor testing; Navy policy; operational testing.

In accordance with Department of Defense (DOD) and U.S. Navy acquisition instructions, integrated testing is defined as the collaborative planning and collaborative execution of test and evaluation (T&E) phases and events to provide shared data in support of independent analysis, evaluation, and reporting by all stakeholders. This includes the developmental (both contractor and government) and operational T&E communities. Integrated testing leverages early and continuous operational testing (OT) with contractor testing (CT) and developmental testing (DT) to form a cohesive testing continuum that supports an operationally realistic evaluation of the system in development. Navy policy emphasizes early and continuous T&E throughout the acquisition life cycle of a weapon system.

The purpose of T&E is to gain knowledge that can be used to (a) advance system development, (b) make programmatic acquisition decisions, and (c) inform users about the system's operational characteristics and performance. Preferably, integrated testing efforts need to start during material solutions analysis and requirements development, to allow for test community understanding of objectives, influence the evaluation of technology alternatives and ascertain the use of operationally realistic test environments. Early involvement also allows for early identification and resolution of deficiencies resulting from T&E in a cost effective and timely manner. Integration of efforts also breaks down stovepipe barriers and enhances efficiency in cost, schedule and performance.

Integrating the enterprise

In 2005, the Navy completed a critical study involving a bottom-up review of T&E domain processes, and major cost and schedule drivers that impacted Navy acquisition programs. The study was led by Commander, Operational Test and Evaluation Force (COMOPTEVFOR) and included broad participation from the Navy T&E and acquisition community. A consensus by study participants emerged that T&E cost drivers identified in earlier reports were still valid and that prior improvement recommendations were slow in implementation. A critical recommendation was to provide a strategic top-down approach in Navy T&E enterprise management to make change happen and to help coordinate efforts across the T&E domain. For the most part, programs do a good job of coordinating the testing of systems for platforms within their purview. However, the T&E process usually begins and ends within programmatic stovepipes.

The Navy study also found that the enterprise view of T&E was limited in regards to the coordination across the acquisition business enterprise and the diverse weapon system development efforts to address the availability of testing resources, T&E policy and improvement initiatives, development efforts and fleet training. The study identified a need for a single T&E process owner and chartered a T&E task force. That task force evolved into the current Navy Enterprise T&E Board of Directors (BoD) to provide synergy and continuous improvement across the T&E domain in support of acquisition programs.

The 2005 T&E study, amongst other findings and recommendations, concluded that integrated testing should be viewed as an order of magnitude increase in collaboration over traditional DT/OT. The study recommended the conduct of integrated CT/DT/OT, using the program's T&E Working Integrated Product Team (WIPT) to develop, coordinate, and execute DT and OT events through phase specific integrated test plans within the Test and Evaluation Master Plan (TEMP).

Navy T&E Executive

In January 2007, Assistant Secretary of the Navy, Research, Development and Acquisition (ASN[RDA]) in concert with Vice Chief of Naval Operations (VCNO) and the Assistant Commandant of the Marine Corps (ACMC), realigned functions that created the Department of the Navy (DON) T&E Executive. Specifically, CNO (N091), Director, Navy T&E and Technology Requirements, was designated and assigned collateral duties as the DON T&E Executive. For matters pertaining to T&E policy, requirements, and operational test resources, the DON T&E Executive reports to VCNO and ACMC as needed. For matters pertaining to development in acquisition programs, the DON T&E Executive reports to ASN(RDA). The DON T&E Executive works with the Principal Military Deputy RDA and all Program Executive Officers (PEOs) and System Commands (SYSCOMs) (i.e., Naval Sea Systems Command [NAVSEA], Naval Air Systems Command [NAVAIR], Space and Naval Warfare Systems Command [SPAWAR], and Marine Corp Systems Command [MARCOR]) to improve the efficiency and effectiveness of T&E enablers across the DON acquisition enterprise.

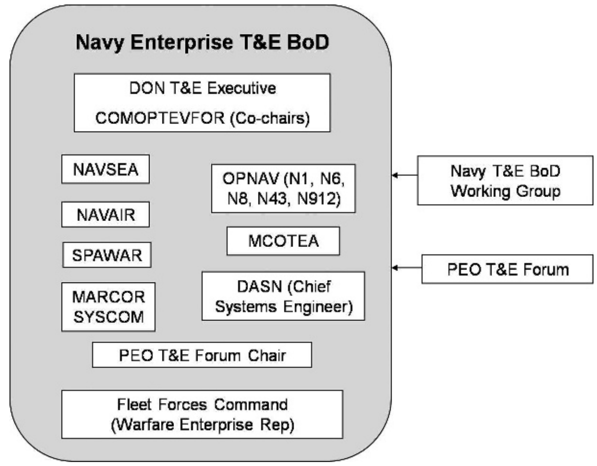


Figure 1. Navy enterprise Test and Evaluation (T&E) Board of Directors (BoD) and participating members.

Navy Enterprise T&E BoD

The Navy Enterprise T&E BoD is unique to the Navy as a Service component. The group was formally established in January 2008 by ASN(RDA) with the authority and responsibility to develop corporate priorities for T&E and to bridge program and domain enablers in support of the Navy warfighting and acquisition enterprises. The group's purpose is to identify and oversee continuous and integrated process improvement to more efficiently and effectively meet the T&E needs of weapon system programs. Because of the group's initial success, a Secretary of the Navy (SECNAV) instruction was issued in May 2009 making the Navy T&E BoD a permanent entity. The DON T&E Executive co-chairs the Navy Enterprise T&E BoD with COMOPTEVFOR. (See *Figure 1*, which shows the other participating members of the board and its support elements.)

Integrating the testing

As noted, integrated testing is being used in support of Navy acquisition programs. However, the specific approach for integrated testing is not mandated because each acquisition system under development is different. As a result, each program determines how best to implement the integrated testing approach for

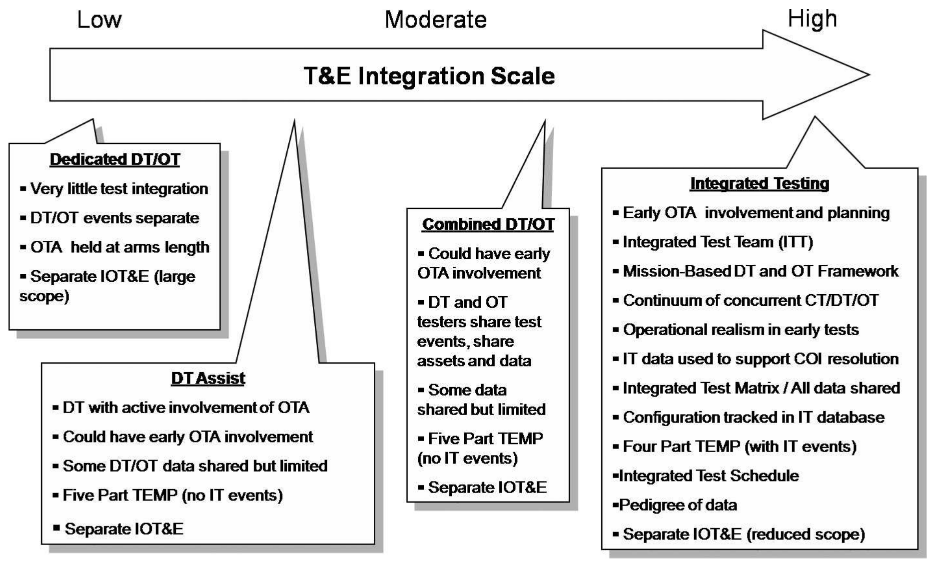


Figure 2. The spectrum of Test and Evaluation (T&E) integration.

its unique acquisition strategy and the technical parameters that need verification. The integrated testing spectrum can span from combined DT/OT events to "full" integrated testing, where specific "integrated test and evaluation" events are called out in the TEMP. *Figure 2* provides an overview of the possible spectrum of integrated testing approaches being used and their basic characteristics.

Furthermore, a summary of integrated testing approaches for Navy air and ship system programs is summarized below. Expeditionary warfare, combat systems and command, control, communications, computer, and intelligence (C4I) systems are using a combination of combined DT/OT and enterprise testing approaches, as outlined in their TEMPs and the integrated test program schedules.

Air programs

NAVAIR with affiliated PEOs is aggressively implementing integrated testing in close to 80 percent of its programs. Such programs include Super Hornet Naval Strike Fighter (FA-18E/F), Poseidon Maritime Multi-Mission Air (P-8A), Advanced Hawkeye Airborne Warning and Control System (E-2D), and Growler Electronic Attack Aircraft (EA-18G).

Based on inputs from T&E action officers and program managers, the specific examples of efficiencies and performance benefits from integrated testing as seen by the P-8A Poseidon program include the following:

- development and test schedule agility, allowing them to minimize the impact of Boeing's recent labor strike;
- potential reduction of flight test program from 3,500 hours at program Milestone B to the current 3,100 hours;
- reduction in the number of test aircraft from seven to six;
- volume and pedigree of flight and lab data generated by the integrated test process as well as comprehensive modeling and simulation to allow a program Milestone C without a dedicated flight phase operational assessment;
- potential reduction in initial operational T&E (IOT&E) by 5 months.

Full integration of OT throughout the program development cycle is the number one reason the Littoral Surveillance Radar System (LSRS) program has seen the following benefits:

- schedule shortened by 4.5 months as a result of integrated testing;
- efforts on time and under cost through initial operating capability and predicted the same for full operation capability;
- significant cost avoidance over the 5 years of integrated testing.

Benefits seen to date on both the Growler (EA-18G) program and the Multi-Mission Helicopter (MH-60) Pre-Planned Product Improvement (P3I) program have been reductions in schedule as early performance reviews yielded cost savings and improved performance as the programs track toward IOT&E. Close cooperation in defining and manning test flights has also reduced sortie requirements to date. Integrated testing allowed for compressed development schedules and provided some cost reductions. The early OT involvement contributed to earlier identification of deficiencies that allowed for design improvements with subsequent system performance improvements that avoided costly corrections late in the schedule.

Ship programs

NAVSEA with affiliated PEOs and their ship acquisition programs are implementing integrated testing but in a variety of approaches. Ship systems must plan for longer design/build/test timelines, complex systems integration, and smaller production quantities. This reflects the way ships are typically contracted, built, and tested (i.e., produced as one or two first ship quantities and tested in distinct stages since major subsystems are incorporated into the ship as separate program elements). At this time, most programs are addressing integrated T&E using combined DT/OT. Newer programs are approaching a fully integrated test concept. As program plans mature and TEMPs are revised, such as for Future Carrier (CVN 78), Zumwalt Class Land Attack Destroyer (DDG 1000), Future Cruiser (CG[X]), and Littoral Combat Ship (LCS), they will move more and more toward the full collaboration of integrated testing. Other programs, such as Virginia Class Submarine program (SSN774) and Ohio Class Conversion program (SSGN), are using combined DT/OT approaches to provide for T&E efficiencies. This is because developmental and production phases were too far along upon implementation of the full integrated testing methodology in DOD/ASN acquisition policy instructions.

Program Executive Office for Integrated Warfare Systems (PEO[IWS]) has historically implemented various levels of integrated testing. To conserve expensive weapons and target assets, PEO(IWS) has expanded collaborative efforts with Aegis combat systems and Ship Self-Defense Air Warfare enterprises by implementing enterprise testing. The enterprise testing approach is well suited for integrated warfare systems (i.e., combat systems and C4I systems) that are employed across multiple classes of ships. Well-defined test scenarios and data collection efforts maximize information needed for all evaluations while minimizing the number of test events that each class of ship must participate in to demonstrate effectiveness and suitability.

Integrated testing best practices

To date, some integrated testing best practices from Navy programs have been collected and are summarized below:

- At program initiation, the integrated testing culture needs the full support of all stakeholders.
  - The attitude starts at the program manager level and should filter down.
- The integrated test team (ITT) needs to be formed early to support test planning, modeling and simulation, and CT/DT/OT test execution for each acquisition phase.
  - ITT objectives, procedures, processes, and memorandums of agreement (MOAs) need to be established to build trust between participants.
  - Proprietary data agreements for the ITT should be established as needed and appropriate.
  - Early planning and meetings are needed to set up the ITT for CT/DT/OT collaboration, T&E framework development, and event execution.
- The Operational Test Agency (OTA) should be funded and brought on board early.
  - The number of billets needs to be identified early and funded for a permanent program presence.
- The OTA should be granted wide access to program data and meetings for formal and informal periods of test.
- OTA representatives should be given free access to the production facility, test sites, test articles, and program meetings, as appropriate.
- OTA needs to support requirements reviews and Concept of Operations development.
- To expedite the process, as appropriate, the prime contractor should be incentivized in the design/build contract to address and resolve operational issues and deficiencies identified by the OTA to improve systems design and operation.
- The program should plan and budget for resolution of deficiencies identified during T&E, since they will undoubtedly occur.

Integrated testing challenges

A summary of challenges and lessons learned identified to date from Navy programs implementing integrated testing are provided below:

- Increased T&E planning efforts are required upfront to plan and execute an integrated testing approach.
- Earlier funding for the T&E workforce (involves CT, DT, and OT workforce) will be needed.
- The T&E effort needs to be engaged for program life to support P3I and capability block upgrades.
- Data management and transparency of the system under test must be carefully managed. During

periods of integrated testing where the design/build contractor and the OTA are involved, contractor data rights or proprietary issues may come into play.
- Pedigree of the test data must be managed and maintained in order to fully understand its validity in resolution of critical operational issues by the OTA.
- The early involvement by T&E stakeholders will impact staffing at SYSCOM/PEOs, Naval Warfare Centers, and OTA and may require reexamination to ensure support can meet program requirements.

Integrated testing pathfinders

To further promote integrated testing and continuous process improvement, the Navy is implementing pathfinder(s) to further identify the lessons learned and best practices. An approach to implementing a pathfinder program is to

1. identify a suitable program (warfare domain area of interest);
2. solicit SYSCOM/PEO/Program Office/OTA concurrences to proceed;
3. implement integrated testing best practices and lessons learned to date;
4. implement a forward-fit test integration effort vice a back-fit effort;
5. start integrated testing efforts early in the program, for best results;
6. determine the extent that certain warfare area acquisition programs are, or are not, suited for integrated testing; and
7. determine the impediments, collect lessons learned, identify challenges and policy changes, if needed.

Summary

At the strategic level, to help improve the effectiveness and efficiency of T&E for acquisition programs, the Navy aligns and governs the T&E enterprise using the DON T&E Executive and Navy T&E enterprise BoD. This process is unique to the Navy and is working well. To improve effectiveness and efficiency in T&E for programs, the Navy has adopted integrated testing. The specific approach is not dictated and is achieved in a number of different ways that suit individual programs. Each program T&E WIPT defines the specific approach to be followed for that program. The basic principle of integration and enhanced communications between organizational elements sets the stage for improved planning and execution of test events. This provides for and has shown to provide testing efficiencies that result in cost and schedule reductions.

Integrated testing does not eliminate the requirement for an independent IOT&E event by an OTA. Independent activity test data are needed by statute to support a full rate production decision. However, the expectation (and results seen to date) is that the IOT&E period will be less in scope and time due to the early involvement of operational testers throughout the entire continuum of system development. Integrated testing entails a significant departure from the legacy DT and OT methodology and encompasses an additional planning paradigm. Early coordination and collaboration between both DT and OT teams in the integrated test planning process provide an earlier-than-normal sharing of data that continues throughout the development and test periods. This sharing will support the monitoring and assessment of system capabilities, attributes, performance parameters, and measures of effectiveness and suitability in order to support resolution of critical operational issues upon completion of IOT&E.

Conclusions

The anonymous quote "First a thing is impossible, then it's difficult, then it's done" applies to integrated testing for a complex weapon system program. Robust testing and early involvement by test activities allow discovery to take place at the front end of the program where it is far less expensive to implement design changes. IOT&E can then be used to confirm what is already known. The integrated testing culture needs to be implanted early in a program to provide the greatest benefit. Cost is reduced by sharing of resources, elimination of duplicative testing, and the early identification and correction of deficiencies. Schedule is shortened by combined versus serial events and the sharing of high demand test assets. ❑

MICHAEL SAID, Office of the Assistant Secretary of the Navy, Research, Development and Acquisition (ASN[RDA]), Test and Evaluation (T&E), is on staff to provide direct support to the Principal Military Deputy, Deputy ASNs (DASNs), DASN Action Officers, SYSCOMs, PEOs, and Navy acquisition programs in the area of T&E. His role is to coordinate efforts closely with CNO (N912). He also provides coordination with OSD(ATL), Director, Operational Test & Evaluation (DOT&E); COMOPTEVFOR; Marine Corps Operational Test & Evaluation Activity (MCOTEA); Defense Test Resource Management Center (DTRMC); other CNO codes; other Service components and industry on issues, policies, and acquisition *reform to achieve efficiency and effectiveness in T&E of warfare system programs. E-mail: michael.o.said@navy.mil*

References

ASN(RDA) Memo. *Realignment of DASN (RDT&E) Functions and Creation of the Navy T&E Executive*, 7 Jan 2008, Washington, D.C.

ASN(RDA). *Navy Enterprise T&E Board of Directors*, SECNNAVINST 3900.44, 28 May 2009, Washington, D.C.

COMOPTEVFOR. *COMOPTEVFOR Report for Streamlining Navy T&E Processes, CNO Guidance for 2004 Task 73, Volume 1*, July 2005, Norfolk, Virginia.

COMOPTEVFOR. *Operational Test Director Manual*, COMOPTEVFORINST 3980.1, 23 April 2008, Norfolk, Virginia.

DASN(RDA) ALM. *Implementation and Operation of the Defense Acquisition System and Joint Capabilities Integration and Development System*, SECNAV Instruction 5000.2D, 16 October 2008, Washington, D.C.

OSD(ATL) and DOT&E Memo. *Definition of Integrated Testing*, 25 April 2008, Washington, D.C.

OSD(ATL). *Operation of the Defense Acquisition System*, DOD Instruction 5000.02, 2 December 2008, Washington, D.C.

Acknowledgments

The author wishes to thank the following individuals for their expert advice, input, and feedback during the development of this article regarding integrated testing principles, best practices, and results achieved to date in Navy programs: CAPT Lee Schonenberg, CAPT Jim Hertlein, CDR Karl "Das" Glaeser, CDR James Fleming, and Mr. Jerry Manthei; all of the Navy T&E and Technology Requirements Office (OPNAV N912).

ITEA Journal 2009; 30: 375–380

Integrated Testing: A Necessity, Not Just an Option

Beth Wilson, Ph.D.
Raytheon Company, Sudbury, Massachusetts

Department of Defense policy states that developmental and operational test activities need to be integrated whenever possible to improve overall test and evaluation efficiency with increased emphasis on operational relevance. The National Defense Industrial Association Systems Engineering Division Developmental Test and Evaluation Committee has been evaluating existing integrated testing policies, methods, and practices to identify an implementation framework with best practices for sharing data, involving developmental test and operational test stakeholders in integrated test planning, and collaboratively executing an integrated test program. Barriers to integrated testing were identified looking at the cultural constraints placed on planning, people, and data. While the definitions and mandates are recent, the practice of integrated testing is not new. The framework described focuses on existing policy and captures best practices.

Key words: Collaborative planning; collaborative execution; shared data; DT&E; OT&E; Army; Navy; Air Force; contractor support; resource planning.

The motivation for integrated testing has long existed in the desire to identify system deficiencies early, meet schedule compression demands, and reduce costs. In the Secretary of Defense (2009) memo dated April 6, 2009, integrated testing was identified as one of the initiatives successfully institutionalized to reform acquisition and requirements processes. *Table 1* provides the Integrating Test and Evaluation section of that memo. The purpose of test and evaluation (T&E) is described as translating test data into knowledge to support management decisions associated with fielding system capabilities. The motivation for the integrated test policy is to improve the complementary relationship between development and operational testing throughout the system life cycle.

To that end, a formal definition of integrated testing appeared in the Office of the Secretary of Defense (OSD 2008) memo cosigned by the Director of Operational Test and Evaluation and the Deputy Undersecretary of Defense for Acquisition and Technology. This definition focused on the need for collaborative planning and collaborative execution of a test program to provide shared data for independent analysis. The Defense Acquisition Guidebook (DAG) chapter 9 (Test and Evaluation) (DoD, 2009) has long described the need for an integrated test and evaluation strategy and amplifies this by describing the concept as a single test program to address issues from developmental test and evaluation (DT&E) and operational test and evaluation (OT&E).

The National Defense Industrial Association (NDIA) Systems Engineering Division DT&E Committee was charged with focusing on integrated testing in view of the OSD memorandum mandating the use of integrated testing (OSD 2007). The Committee used the preceding definition to develop an implementation framework from the key points in the definition:

- collaborative planning,
- collaborative execution,
- shared data.

While the formal definition is recent, the concept is not new. The Department of Defense (DoD) has written policies regarding combined activities relating to DT&E and OT&E since the early 1970s. In soliciting best practices for integrated testing, many examples emerge showing successful implementation before the Department issued the recent policies and definitions. The Army, Navy, and Air Force had already defined integrated testing in their respective test and evaluation policies. *Table 1* lists the integrated testing definitions identified in the DoD and Services policy and procedure documents.

Table 1. Policy definitions of integrated testing.

OSD Memo April 6, 2009	The fundamental purpose of T&E is to provide knowledge to assist in managing the risks associated with developing, producing, operating, and sustaining system and capabilities. The goal is early identification of technical, operational, and system deficiencies to facilitate timely corrective action. To achieve that goal, DoD has implemented new policies designed to improve the integration of developmental and operational testing through the system lifecycle. The objective is to improve the complementary relationship between developmental and operational testing, eliminate costly redundancy, and ensure test results are a key element of the knowledge base that informs management decisions.
OSD Memo April 25, 2008	Integrated testing is the collaborative planning and collaborative execution of test phases and events to provide shared data in support of independent analysis, evaluation, and reporting by all stakeholders, particularly the development (both contractor and government) and operational test and evaluation communities.
DAG chapter 9 (Test & Evaluation)	The concept is to conduct a single, combined test program that produces credible qualitative and quantitative information that can be used to address development and operational issues.
Army DA PAM 73-1	A T&E strategy that reduces the multiple and redundant products and processes, and encompasses the development of a single integrated system evaluation plan and a single integrated test or simulation strategy, leading to a single system evaluation report for the customer. The process also increases the use of contractor data for evaluation and expands the use of M&S with the goal of reducing T&E costs. Integrated T&E strategies may include combined DT or OT events where appropriate.
	Integrated DT or OT, a special case of a combined DT and OT, is a single phased event that generates data to address developmental and operational issues simultaneously under operational conditions. The execution strategy for this event is based on the requirements of the program.
Navy OT&E Framework and IT Methodology	IT is a cooperative approach to T&E where CT, DT, and OT entities work to blend or integrate the T&E requirements throughout the defense acquisition process. Integration of CT, DT, and OT does not involve the analysis and reporting aspects of T&E, which remain solely under the purview of the respective CT, DT, or OT organization.
Air Force AFI 99-103	The collaborative planning and collaborative execution of test phases and events to provide shared data in support of independent analysis, evaluation, and reporting by all stakeholders, particularly the developmental (both contractor and government) and operational test and evaluation communities.

The Department of the Army (2003) describes the need for a single integrated test strategy and the use of shared data. The Navy describes the collaboration of contractor testing, development testing, and operational testing to blend the test and evaluation requirements (USN 2005). The Air Force (2009) describes the collaborative planning and collaborative execution of test phases and events as part of its definition of integrated testing.

Barriers to integrated testing

If the concept of integrated testing is so well-defined and has existed for so long, why was a joint DT&E and OT&E memorandum required to mandate it? During his keynote presentation to the NDIA Systems Engineering Conference, Dr. McQueary (Director, Operational Test and Evaluation) reported that in 2007, four of eight (50%) systems delivered for initial operational test and evaluation (IOT&E) were found to be "Not Suitable" and in 2008 two of six (33%) suffered the same fate (McQuery, 2008). This fact alone would indicate something in the process is broken. Even though the practice of integrated testing is accepted, why are the products being produced failing to capitalize on values of this concept? The barriers appear to be cultural and/or based in misunderstanding.

The most common barrier the NDIA DT&E Committee identified was the misrepresentation of the law. The specific language depicting independent operational test and evaluation is in US Code: Title 10 (Subtitle A, Part IV, Chapter 141 § 2399) and reflected in Services policies. *Table 2* lists the relevant excerpts from these documents related to contractor participation in integrated testing.

Section (d) "Impartiality of Contractor Testing Personnel" defines the restrictions on contractor involvement in the conduct of OT&E. Section (e) "Impartial Contracted Advisory and Assistance Services" further describes restrictions on contractor involvement in establishing criteria for operational testing and operational evaluation. The statute clearly states that the contractor cannot be involved in OT&E conduct, establishing OT&E criteria, and operational evaluations. The statute does not prohibit, however, contractor participation in OT&E to provide technical understanding of test incidents, logistic support and training, support to test failure analysis, and unique software and instrumentation support. For clarification, the contractor (sometimes referred to as the prime contractor) in this instance is the entity contracted to develop and design the system under test and not the Scientific Engineering and Technical Assistance (SETA) type contractor providing direct technical support to the government.

Existing policies describe allowed interaction of the contractor with OT&E activities. The DAG describes combining contractor testing with government DT and

Table 2. Statute and policy regarding contractor involvement in integrated testing.

Title 10 U.S. Code Section 2399 (d)	In the case of a major defense acquisition program (as defined in subsection (a)(2)), no person employed by the contractor for the system being tested may be involved in the conduct of the operational test and evaluation required under subsection (a). The limitation in the preceding sentence does not apply to the extent that the Secretary of Defense plans for persons employed by that contractor to be involved in the operation, maintenance, and support of the system being tested when the system is deployed in combat.
Title 10 U.S. Code Section 2399 (e)	(1) The Director may not contract with any person for advisory and assistance services with regard to the test and evaluation of a system if that person participated in (or is participating in) the development, production, or testing of such system for a military department or Defense Agency (or for another contractor of the Department of Defense). (2) The Director may waive the limitation under paragraph (1) in any case if the Director determines in writing that sufficient steps have been taken to ensure the impartiality of the contractor in providing the services. The Inspector General of the Department of Defense shall review each such waiver and shall include in the Inspector General's semi-annual report an assessment of those waivers made since the last such report. (3) (A) A contractor that has participated in (or is participating in) the development, production, or testing of a system for a military department or Defense Agency (or for another contractor of the Department of Defense) may not be involved (in any way) in the establishment of criteria for data collection, performance assessment, or evaluation activities for the operational test and evaluation. (3) (B) The limitation in subparagraph (A) does not apply to a contractor that has participated in such development, production, or testing solely in testing for the Federal Government.
DAG chapter 9 (Test & Evaluation)	Integrating T&E consists of many aspects, all designed to optimize test scope and minimize cost. For example, separate contractor developmental testing might be combined with governmental developmental test and evaluation, with control being exercised by a combined test organization.
Army DA PAM 73-1	Discussions with system contractor personnel may be necessary to ensure full technical understanding of test incidents observed during the IOT&E or related activities. All discussions will be held separately from any scoring or assessment activities.
Navy OT&E Framework and IT Methodology	"Integrated testing" blends or combines contractor, developmental, and OT to form a cohesive testing continuum. This integration cannot occur unless the participants (CT, DT, and OT) have determined their entering requirements for adequate testing of the system under evaluation. IT does not remove or combine any of OPTEVFOR's current or future requirements for reporting based on a separate (OPTEVFOR) analysis of the shared test information produced by the IT effort.
Air Force AFI 99-103	System contractors may be beneficial in providing logistic support and training, test failure analyses, test data, and unique software and instrumentation support that could increase the value of operational test data.

OT to optimize test scope and minimize cost. The Department of the Army (2003) describes contractor support to OT&E with separation from the evaluation activities. The USN (2005) document describes blending of contractor, development, and operational testing to form a cohesive testing continuum while maintaining independent evaluation of shared data. The USAF (2009) document also describes contractor support to operational testing.

The statute does allow contractor support to OT&E, and current policy reflects this. The statute also allows for sharing of data collected throughout the integrated testing continuum. Referring to the same Services policy and procedure documents, *Table 3* lists excerpts that describe the use of shared data.

To establish an integrated test program that allows for and encourages collaborative planning, collaborative execution, and shared data, the Services must give early consideration to resource planning and contract requirements. While the OT&E personnel can participate in contractor test planning and DT conduct, these resources need to be identified early in the program to implement effective collaborative planning. Collaborative execution means funding and staffing these integrated teams. Shared data may require contract language to support the data dictionary development and release of contractor collected data.

The NDIA DT&E Committee investigated the barriers to integrated testing. The most consistent cause for these perceived barriers was misrepresentation of the statute and misunderstanding of existing policy. Once you remove the cultural barriers, however, there still exist barriers that need to be addressed early to facilitate effective integrated testing strategies.

Implementation framework for effective integrated testing

Within existing policy, integrated testing can be (and has been) effectively implemented. As part of this evaluation, the NDIA DT&E Committee solicited best practices for integrating the people, integrating the planning, and integrating the data to achieve successful integrated testing. The DT&E Committee received many examples from multiple industry companies for various contracts awarded from different Services. These best practices were consolidated to

Table 3. Shared data policy for integrated testing.

DoD Policy DoDI 5000.02 Enclosure 6	Evaluations shall take into account all available and relevant data and information from contractor and government sources.
Army DA PAM 73-1	The T&E WIPT goals are to develop a mutually agreeable T&E program that will provide the necessary data for evaluations.
	Support the CE process by accomplishing early, more detailed, and continuing T&E documentation, planning, integration, and promote the sharing of data.
Navy OT&E Framework and IT Methodology	OT uses the shared data from the IT period to "answer" or achieve resolution on as many measures of effectiveness (MOE) and measures of suitability (MOS) as possible. The goal being to have sufficient data or test information at the end of the IT phase to resolve most COIs, pending successful completion of the final independent OT phase.
Air Force AFI 99-103	Operational testers may use data from sources such as DT&E, integrated testing, and OAs to augment or reduce the scope of dedicated operational testing if the data can be verified as accurate and applicable.
	Test teams and TIPTs should use as much contractor T&E data as possible if its accuracy can be verified. Contractor T&E data should be visible in the common T&E database.

depict amalgamation statements that collectively describe the inputs received.

Collaborative planning

Integrating the test planning requires early collaboration to make efficient use of test assets across the T&E continuum. Integrated test planning can improve test efficiency and streamline the T&E schedule. Collaborative planning reduces duplication of effort and ensures established test coverage without gaps.

The earlier testing is involved in the planning process, the more efficient integrated testing can be conducted. This has been demonstrated during the development of operational design requirements early in system design. By conducting concept of operations experiments with mock-ups of the operational layout, the developmental and operational community implemented real platform mission threads using preliminary displays. By conducting this integrated exercise during the initial requirement development program phase, the user community directly effected key system design requirements through a discussion of performance parameters. The result was a product baseline that better met customer expectations with fewer surprises in performance characteristics.

Early involvement in test planning is essential to integrated testing. Review and approval of contractor deliverable documentation by the integrated test community can greatly assist in the planning process. An OT&E and DT&E community review of the program contractor test plan prior to submittal for approval allowed examined planned ground, simulation, and flight tests to ensure that a realistic approach was being taken. The integrated test program planners and community worked with the integrated test team to ensure that an integrated operational environment was applied to the entire system test program. This ensured that realistic test scenarios were developed throughout the contractor's and integrated test team's program. It also improved communication and set expectations across the test program.

Combined review of documentation can assist with integrated test planning. Cases were presented where the Operational Test Agency (OTA) participated in a review of the developmental test cases. As part of that review, a matrix was developed showing which scenarios had been identified for DT and contrasted with those being identified for OT. Some of the DT scenarios selections were changed to OT scenarios so that the OTA could evaluate system behavior and displays during DT.

Collaborative execution

Integrating the people means chartering and implementing integrated test teams. These teams provide the coordination and cooperation for executing an integrated test strategy. The sharing of people across the test continuum provides an opportunity for early OT&E influence on test design and scenarios.

Early participation of test personnel yields exceptional benefits as was seen when end-users acted as test operators during contractor testing, DT preparation, dry runs, and conduct. The operator interacted with the system in the way it would be used when fielded. By having the operator use the system, the development team identified display issues early and made corrections during DT.

Sharing personnel resources and expertise allows for collaborative testing and early operational relevant information in reports. When testing a new capability for an added mission, the OTA observed the DT preparation effort. When the opportunity presented itself, the OTA executed one of their planned tests and noted a difference in behavior. The DT team was able to explain that this was an artifact of the upgraded performance that provided a graceful degradation feature. This eliminated the chance of the OTA noting this as a deficiency and requiring a subsequent investigation.

Sharing of personnel provides early system interaction, improved system understanding, and knowledge to develop superior tests. In one example, there was a contractor facility with actual equipment interfaced to simulated sensors and weapons, and a government facility with actual equipment interfaced to real sensors and weapons. The government provided personnel to assist with contractor testing. The contractor gained experienced technicians at no cost, and the government technicians gained experience with the system prior to government testing. In addition, the government technicians reported progress to the OTA providing visibility into and confidence in the contractor testing.

Shared data

Integrating the data means maximizing the data that are available and usable for all participants. Common data formats are necessary for those data items intended for sharing. Designing the test program to include shared data also means incorporating operational realism in DT&E.

Maximizing the utility of the test events and data are paramount to integrated testing. Examples were provided where test assets were shared during tests. In particular, DT events with scripted firings were supplemented by OT firings introducing "tactical surprise." Using more operationally realistic procedures for the DT firings allowed the OTA to include these data to increase the sample size to support operational effectiveness confidence computations. The strategy saved the program significant cost in reduced test assets. The cooperation also reduced time on the test range.

Sharing of data and tools reduces the burden of independent analysis, evaluation, and reporting by all stakeholders. This was the case for a system when in preparation for DT the OTA participated in meetings to identify data reduction tools. Where possible, the DT team used common data reduction tools so that the OTA had access to the DT data in the formats desired for OT&E. In addition, the OT&E data reduction tools were used during DT&E preparation. This allowed the DT&E team to detect a problem early where data formats not used by the DT&E team had been degraded and would have caused problems for OT&E. The problem was corrected before the DT event.

Summary

The NDIA Systems Engineering Division DT&E Committee focused its evaluation of integrated testing on identifying existing policies, methods, and practices. The DT&E Committee identified barriers to integrated testing and offers best practices that look at removing those barriers in the data, resources, and cultural constraints. From this perspective, the DT&E Committee identified potential collaborative approaches within current policies focusing on best practices that defined roles and positions for the people involved, interactions between stakeholders, when in the test continuum these resources were involved, and output products for these interactions.

The DT&E Committee also developed a list of attributes associated with effective integrated testing programs. *Table 4* uses the Jeff Foxworthy format to list attributes of integrated testing.

Table 4. Attributes of integrated testing.

If you find the contractor data augmenting the OT data, *you might be doing integrated testing*
If the DT and OT personnel recognize each other in the airport, *you might be doing integrated testing*
If the OT personnel influences DT scenarios, *you might be doing integrated testing*
If the DT system is operated by end users, *you might be doing integrated testing*
If the CT, DT, and OT teams are sharing data in a common format, *you might be doing integrated testing*
If the OT confirms DT results, *you might be doing integrated testing*

A need for integrated testing exists to facilitate early identification and correction of system deficiencies, to make OT&E a process of confirmation instead of discovery, to minimize surprises when the completed system is delivered to the war fighter, and to reduce cost and schedule with shared resources and reduced duplication. While the statute prohibits contractor involvement in OT&E conduct, it does allow for contractor support. There is nothing in the statute that prevents collaborative planning and execution of an integrated test program to provide shared data to support independent analysis.

Effective integrated testing requires an implementation framework that integrates the people, planning, and data. Integrated test teams need to introduce operational realism early in the program to accelerate the discovery of deficiencies. Integrated test teams need to implement early and collaborative planning efforts to streamline the test program. Integrated test teams need to share data to address development and operational issues as early as possible in the test and evaluation program.

If your program's operational testing confirms your developmental testing results, you may indeed have benefited from integrated testing. ❑

BETH WILSON is a senior engineering fellow at Raytheon Company, Sudbury, Massachusetts. She earned a doctor of philosophy degree in electrical engineering (EE) from the University of Rhode Island (Kingston, Rhode Island), a master's degree in electrical engineering from Brown University (Providence, RI), and a bachelor's degree in electrical engineering from the State University of New York at Buffalo (Buffalo, New York). She is the industry cochair for the National Defense Industrial Association Systems Engineering Division Developmental Test and Evaluation Committee and is the industry lead for the Integrated Contractor/Developmental/Operational Test focus area. E-mail: Beth_J_Wilson@raytheon.com

References

Department of Defense (DoD). 2009. *Interim Defense Acquisition Guidebook,* Chapter 9 (Test and Evaluation). Washington, D.C.: Department of Defense (DoD). Available online at https://acc.dau.mil/dag (Accessed July 29, 2009).

Department of the Army. 2003. Test and Evaluation in Support of Systems Acquisition. Pamphlet DA PAM 73-1, May 30. Washington, D.C.: Department of the Army.

McQueary, C. 2008. Keynote Presentation. In: *NDIA 11th Annual Systems Engineering Conference Proceedings,* October 20–23, 2008 in San Diego, California. Available online at http://www.dtic.mil/ndia/2008systems/2008systems.html (Accessed July 29, 2009).

OSD (Office of the Secretary of Defense). 2007. Memorandum, "Test and Evaluation Policy Revisions," December 22. Signed by Dr. Charles McQueary (Director, Operational Test and Evaluation) and John Young (Under Secretary of Defense for Acquisition, Technology, and Logistics). Washington, D.C.: Office of the Secretary of Defense.

OSD. 2008. Memorandum, "Definition of Integrated Testing," April 25, 2009, signed by Dr. Charles McQueary (Director, Operational Test and Evaluation) and James Finley (Deputy Undersecretary of Defense for Acquisition and Technology). Washington, D.C.: Office of the Secretary of Defense.

Secretary of Defense. 2009. Memorandum, "Acquisition and Requirements Reform," April 6, signed by Dr. Robert Gates (Secretary of Defense). Washington, D.C.: Department of Defense.

US Code: Title 10, Subtitle A, Part IV, Chapter 141, Section 2399 "Operational Test and Evaluation of Defense Acquisition Programs." Available online at http://straylight.law.cornell.edu/uscode/html/uscode10/usc_sec_10_00002399 (Accessed July 29, 2009).

USAF (United States Air Force). 2009. Department of the Air Force Instruction AFI 99-103, "Capabilities-Based Test and Evaluation." March 20. Washington, D.C.: U.S. Air Force.

USN (U.S. Navy). 2005. Commander Operational Test and Evaluation Force Policy and Information Notice 05-1. 2005. "Operational Test and Evaluation (OT&E) Framework and the Integrated Testing (IT) Methodology." June 28. Washington, D.C.: U.S. Navy.

Acknowledgments

This article documents the collective effort of the NDIA Systems Engineering Division DT&E Committee, which is comprised of over 100 representatives from industry and government. As the industry lead for the focus on integrated testing, I wish to thank the 35 members of the DT&E committee that contributed to this article.

ITEA Journal 2009; 30: 381–385

Chemical and Biological Defense Test and Evaluation—A Hallmark of Integrating DT and OT

Charlie Holman, Ph.D.
U.S. Army Test and Evaluation Command, Alexandria, Virginia
Robert L. Berkowitz
VRC Corporation, Alexandria, Virginia

The most operationally realistic testing of chemical and biological defense systems uses actual biological and chemical warfare agent; however, testing live or actual agents is restricted to laboratory containment chambers, which are operationally unrealistic environments. This state of affairs has driven the chemical and biological defense community to integrate developmental testing (DT) and operational testing (OT) to support the evaluation process. Three paradigms are commonly used to integrate chemical and biological defense DT and OT. They are (a) conducting DT with systems before and after OT, (b) developing agent simulant relationships in DT, which are then applied to OT data, and (c) modeling and simulation. This article supplies chemical and biological defense system examples for each paradigm.

Key words: Biological agent detection; chemical agent detection; developmental testing; integrated testing; modeling and simulation; operational testing; public safety; realistic testing.

The most operationally realistic testing of chemical and biological defense systems will use actual biological and chemical warfare agent. As a result of public law, treaty, and concern for public safety, testing with actual agent is restricted to laboratory containment chambers, which is the least operationally realistic environment. As a result, the chemical and biological defense test and evaluation (T&E) community has been forced to develop and use paradigms that combine operational testing and developmental testing. The operational test (OT) part brings the realism of actual warfighters executing missions in combat-like environments. The developmental test (DT) part brings the realism of actual biological or chemical agent. These paradigms are the following:

- conducting DT with systems before and after OT,
- developing agent simulant relationships, and
- modeling and simulation.

A simulant is a relatively harmless substance that has some of the properties of agents and can be released into the environment. These three basic paradigms are not mutually exclusive and may be used in combination on the same system under test.

This article describes how these three basic paradigms to integrate DT and OT have been used as the lynch pins for evaluation of chemical and biological defense systems.

Paradigm 1: DT with systems before and after OT

During OT, systems are used by warfighters while executing wartime missions. Use, care, and maintenance of the system under test are typical of what could be expected during actual operations. OT, like wartime, tends to be a strenuous environment. During OT, system performance is often degraded. The total effect of an OT on the performance of system under test can be determined by conducting DT performance testing on those systems both before and after the OT. The effect of less use than is measured by a whole OT on the performance of a system under test can be measured by periodically removing systems from OT for DT performance testing. This paradigm has been routinely used on protective garments since the Joint Service Lightweight Integrated Suit Technology (JSLIST) testing

Figure 1. Beach assault from Joint Service Lightweight Integrated Suit Technology operational testing.

in the early 1990s. It was also applied to the Joint Biological Point Detection System (JBPDS) as well as other detectors.

During OT, the JSLIST candidate protective garments were worn by U.S. Marines, soldiers, airmen, and sailors while conducting combat missions. These missions included offensive and defensive actions at 29 Palms Marine Corps Base, a beach assault at Camp Pendleton *(Figure 1)*, shipboard operations, and flight line operations. The missions were intended to put the same wear, stresses, and strains on JSLIST as would occur during combat. The garments were worn for up to 60 days. They were worn 23 hours a day. Warfighters not only conducted missions in JSLIST but also slept and ate in this protective garment. The suit was laundered every other week. A random sample of JSLIST protective garments and the base line protective garment were removed from the OT after 15, 30, 45, and 60 days of wear and sent to DT performance testing. It was known that mission types were not equivalent. Offensive actions result in more wear on garments than defensive operations. Hence, care was taken to randomly choose garments and to spread mission types somewhat uniformly over the 60 days of OT.

The DT performance testing include various types of swatch testing with various chemical warfare agents, and whole system testing in which warfighters wore JSLIST in a simulant chamber performing a prescribed task list. In addition to the garments with 15, 30, 45, and 60 days of wear, new garments were also tested in DT performance testing.

The JSLIST evaluation compared the effect of mission wear time on the performance degradation of JSLIST candidates to the baseline. The fielded protective garment served as the baseline. This strategy of performing DT both before and after OT resulted in identifying the JSLIST candidate garment that best increased the amount of protection (Musgrave et al. 1997).

The JBPDS was tested in accordance with this paradigm in the Ambient Breeze Tunnel (ABT) before and after OT I. The ABT is a DT wind tunnel in which simulant concentration can be accurately controlled and measured. This testing revealed that after OT there was significant and substantial degradation in detection performance. The degradation in performance was traced to a LASER in the detection system. The LASER design was upgraded. Subsequent testing demonstrated that the new LASER did not degrade during OT (Chipman et al. 2001).

This paradigm combines or integrates DT and OT results to produce a powerful evaluation tool to ascertain if system performance is degraded during OT.

Paradigm 2: Agent-simulant relationships

A simulant is a relatively harmless substance that possesses, for the attributes being measured or tested, properties similar to agents of interest and that may be

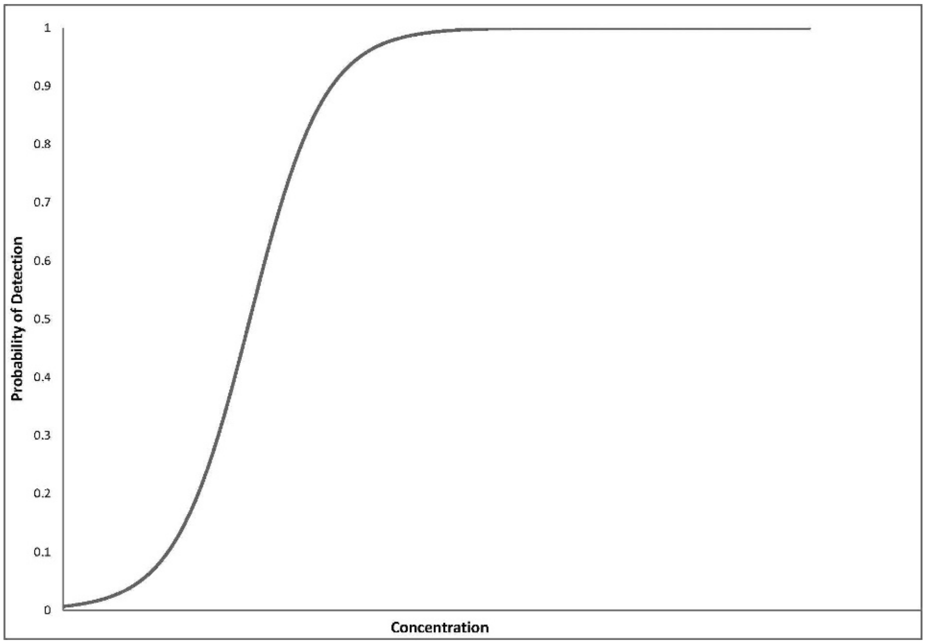


Figure 2. S-shaped or sigmoid curve depicting the relationship between agent detection and agent concentration.

released into the environment since the simulant is not dangerous. Much effort is invested in simulant selection and development. For testing detectors, it is desirable that the system under test "sees" the simulant the same way as it "sees" the agent that the simulant is simulating. Other properties of interest include similarities in dissemination and cloud dynamics. No simulant is an exact match in all properties of interest to the agent.

Traditional biological detection simulants include ovalbumin for toxins, the bacteriophage Male Stereotype 2 (MS2) for viral biological warfare agents, *Erwinia herbicola* for vegetative biological warfare agents, and *Bacillus subtilis* for spore-forming biological warfare agents. Chemical simulants are much more numerous than biological simulants; two commonly used simulants for chemical detectors are triethyl phosphate for nerve agent and acetic acid for blister agent.

The Biological Integrated Detection System (BIDS) and BIDS Pre-Planned Product Improvement (P3I) were tested in the early to mid 1990s. They used both biological warfare agent and the traditional biological detection simulants in DT laboratory containment chamber testing and traditional biological detection simulants in OT. Inferences focused on laboratory results with agent and OT results with simulant. It was noted in DT that the detection performance of BIDS for a group of agents lay between the values of two of the stimulants. This finding enabled us to predict that the field performance of BIDS against these agents would be bounded by the OT performance against these two simulants.

For both biological warfare agent detectors and chemical warfare agent detectors, the variable that has the most profound effect on detection performance is agent concentration. As concentration increases, detection performance tends to increase. At some high level of agent concentration, all releases are detected. As concentration decreases, detection performance tends to decrease. At some low level of agent concentration, no releases are detected. The plot formed by agent-concentration and detector-performance is s-shaped or sigmoid as is depicted in *Figure 2*. Logistic regression is a statistical framework based on a sigmoid relationship. Logistic regression has been useful in modeling detection performance and in developing the agent-simulant relationships in both the JBPDS and the Joint Chemical Agent Detector.

Three of the traditional biological detection simulants—ovalbumin, MS2, and *Erwinia herbicola*—have been considered inadequate representations of actual biological warfare agent (Fitch et al. 2004). In an effort to obtain simulants that better represent biological warfare agents of interest, closely related organisms or vaccine strains were used for viruses and bacteria. Toxoids were used for toxin. A toxoid is a toxin that has been denatured or broken into nonhazardous components. The toxoids, closely related organisms, and vaccine strains are referred to as agent-like organisms (ALO). To eliminate any chance of infection from the closely related organisms or vaccine strains, these were killed with ionizing radiation.

The DT portion of the Whole System Live Agent Test of JBPDS was conducted to test JBPDS detection performance when challenged with biological warfare agent and ALOs. Both live and killed agent and ALO were used. JBPDS performance testing was conducted with both living and killed ALOs, and living and killed agents, in laboratory containment chambers. JBPDS performance testing was conducted with only killed ALO simulants in the ABT and field. The end results of this DT testing were twofold:

1. JBPDS detection performance when challenged with live agent in a laboratory containment chamber was characterized.
2. Relationships were developed to relate JBPDS detection performance when challenged with killed ALO simulants to JBPDS detection performance when challenged with live agent.

The JBPDS OT used the new killed ALO simulants. The relationships developed in DT between JBPDS detection performance with killed ALO simulants and live agent were used to predict JBPDS performance as if it had been challenged with live agent (Holman et al. 2008).

The evaluation of JBPDS provided predictions of JBPDS performance against live agent based upon OT challenges with killed ALO simulants and relationships developed in DT.

Paradigm 2, the development agent-simulant relationships, combines or integrates relationships developed in DT with OT results to support an integrated evaluation of how the system performs against agent.

Paradigm 3: Modeling and simulation

The keystone in both the Joint Service Lightweight Standoff Chemical Agent Detector (JSLSCAD) and the Joint Biological Standoff Detection System (JBSDS) evaluations was the use of modeling and simulation to integrate DT and OT results to predict system detection performance.

JSLSCAD modeling and simulation was conducted by a team that included Johns Hopkins University Applied Physics Laboratory (JHU APL), the Joint Project Manager for Contamination Avoidance, and both Dugway Proving Ground and the U.S. Army Evaluation Center from the Army Test and Evaluation Command. The backbone of the JSLSCAD model was a model developed by Honeywell, which was used to support JSLSCAD development. Spectral backgrounds, meteorological data, and other data were collected by JHU APL in various locations of tactical interest and used in the JSLSCAD model. The vapor, liquid, and solid tracking transport and dispersion model, with chemical and biological analyzer, was used to provide cloud size, concentration, and cloud propagation information for chemical warfare agents. A scanning model was used to allow the JSLSCAD model to have multiple chances of detecting a chemical cloud as it moved downwind. OT and DT field tests and laboratory tests provided a base for validation, verification, and accreditation of the modeling and simulation.

The modeling and simulation results were the best source of information on the detection performance of JSLSCAD. The modeling and simulation provided a means to integrate DT and OT in an evaluation that provided information on the effectiveness in detecting chemical warfare agents in a threat realistic environment. Model results can be found in Holman et al. (2007).

The JBSDS modeling and simulation was conducted by a team that included JHU APL, the Joint Project Manager for Biological Defense, Air Force Operational Test and Evaluation Center, and the Army Test and Evaluation Command. The model was based on the capability of the system to re-analyze "playback" data from previous trials. Some of the parameters in the system "playback" are then changed from parameters that are indicative of simulant to parameters that are indicative of biological warfare agent. The basic playback uses field DT or OT data. The agent unique parameters are measured in a laboratory during a DT event.

These modeled results are the best source of performance data to evaluate the detection performance of JBSDS. The modeling and simulation provided a means to integrate DT and OT in an evaluation that provided information on the effectiveness in detecting and discriminating biological warfare agents in a threat realistic environment. This process is described in Shirakawa, Russell, and Holman (2008) and Przybylowicz et al. (2003).

Paradigm 3 modeling and simulation provides an opportunity and means to integrate DT and OT results to support an integrated evaluation of how a system performs when challenged with agent.

Conclusion

Biological and chemical warfare agents cannot be released into the environment during OT, and they can only be tested in DT laboratory containment chambers. Hence, the only way to obtain chemical and biological defense system operational relevant evaluations is to integrate DT and OT. ❑

Charlie E. Holman is chief of the Chemical and Biological Defense Evaluation Division at the U.S. Army Test and Evaluation Command in Alexandria, Virginia. He has 19 years of experience in evaluating chemical and

biological defense systems. He received a bachelor of science degree in biology from the Pennsylvania State University, University Park, Pennsylvania; a master of science degree in biology from Bloomsburg State University, Bloomsburg, Pennsylvania; a master of mathematics degree from the University of South Carolina, Columbia; and a doctor of philosophy degree in biodefense from George Mason University, Fairfax, Virginia. E-mail: charlie.e.holman@us.army.mil

ROBERT L. BERKOWITZ is currently a senior analyst with the VRC Corporation in Alexandria, Virginia, supporting the Chemical and Biological Defense Evaluation Division at the Army Test and Evaluation Command. He received his bachelor of arts degree in mathematics from Temple University in Philadelphia, Pennsylvania, in 1968, and his master of arts degree, also in mathematics, in 1972 from the University of Maryland, College Park. He has supported many different government agencies including the U.S. Army, Navy, and Federal Aviation Administration. E-mail: robert.berkowitz@us.army.mil

References

Chipman, M., C. E. Holman, K. Finanger, J. Fuller, R. Jernigan, L. Wollenberg, B. Rodriguez, et al. 2001. System assessment of the Joint Biological Point Detection System. U.S. Army ATEC Report, Alexandria, Virginia.

Fitch, J. P., B. P. Mark, W. Colclasure, M. Coleman, P. E. Coyle, H. H. Hill, N. B. Jackson, et al. 2004. *Review of test and evaluation methodology for biological point detectors*. Washington, D.C.: The National Academies Press.

Holman, C., I. Kuria, N. Dunn, J. Timmerman, S. Colegrove, P. Serguievski, J. Kleimeyer, et al. 2007. Report of the modeling and simulation of the Joint Service Lightweight Standoff Chemical Agent Detector (JSLSCAD). U.S. Army Test and Evaluation publication, March 2007. (Classified at the secret level.)

Holman, C. E., M. Moe, C. T. Russell, C. Jennings, R. Bartholomew, and D. Anderson. 2008. Technical Note: Whole System Live Agent Test (WSLAT) methodology and performance assessment. U.S. Army Test and Evaluation publication, September 2008.

Musgrave, D., N. Dunn, F. Gasiorowski, J. Winters, S. Tackett, and C. Holman. 1997. System evaluation report of the Joint Service Lightweight Integrated Suit Technology I (JSLIST I) Developmental and Operational Tests (DTs and OTs). U.S. Army Test and Evaluation publication, April 1997.

Przybylowicz, E., E. Browell, D. Chase, J. De Haseth, R. Flagan, P. Griffiths, D. Layton, T. Reichardt, and L. Rhomberg. 2003. Testing and evaluation of standoff chemical agent detectors. Washington, D.C.: The National Academies Press.

Shirakawa, K., C. T. Russell, and C. E. Holman. 2008. Estimating performance of a standoff biological detector system agents actual biological warfare agent. MORSS, June 10–12, 2008. New London, Connecticut: U.S. Coast Guard Academy.

ITEA Journal 2009; 30: 386–394

The System Engineering and Test (TSET) Approach for Unprecedented Systems

Lora G. Weiss, Ph.D., Rusty Roberts, and Stephen E. Cross, Ph.D.
Georgia Tech Research Institute,
Georgia Institute of Technology, Atlanta, Georgia

The rapid pace of development of new systems coupled with a strong desire from warfighters to quickly field systems with advanced technologies and innovation poses new Test and Evaluation (T&E) challenges. These challenges start with the realization that most T&E procedures are derived from a historical, requirements-based approach to acquisition, which inherently is a sequential process. For innovative and unprecedented systems, i.e., the kind of system for which there is no experience in building similar systems or in their test or use, T&E cannot follow a sequential approach. Throughout military history, development of unprecedented systems has occurred when there has been a simultaneous advance in technology and operational need such as is occurring now in the domain of unmanned systems. T&E needs to evolve to be integrated with the development process. Waiting for the results of developmental and operational testing will only exacerbate the delay in rapidly fielding advanced capabilities. This article presents the tenets of using the system engineering and test approach for evaluating unprecedented systems and moving testing to the forefront of the system development process.

Key words: Adaptability, collaborative development, CONOPS discovery, early involvement, emergent capabilities, evolving environments, experimentation, innovation, system alternatives, systems engineering and testing, unmanned autonomous systems.

The current Test and Evaluation (T&E) paradigm is founded in decades of experience, yet the world has changed significantly, particularly over the last decade. There is the realization that new test methods are needed to get our systems to the warfighter sooner; some strong sentiments have been recently expressed, which reflect the need for unconventional approaches to achieving this goal. U.S. Secretary of Defense Robert M. Gates, at his speech at the Air War College at Maxwell Air Force Base, Montgomery, Alabama, on April 21, 2008, stated:

> "For the kinds of challenges America faces and will face, the armed forces will need principled, creative, reform-minded leaders, men and women who…want to do something, not be somebody. An unconventional era of warfare requires unconventional thinkers." (Gates 2008)

Chris Dipetto, Deputy Director for the Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics, stated in his presentation "A New Vector for Developmental Test and Evaluation (DT&E)" to the International Test and Evaluation Association, Tidewater Chapter, 2007 (DiPetto and Stuckey 2007) that we need to change "… the mind-set of all persons involved to focus on trading greater capability for earlier fielding." And, the 2008 report of the Developmental T&E Committee of the National Defense Industrial Association (NDIA) stated that we need to "…specifically address T&E policy recommendations for incorporating T&E expertise early in the acquisition cycle, integrating developmental and operational testing, and improving suitability of weapon systems during development."

Numerous reports, findings, presentations, and committees are voicing a need for supporting a new direction in T&E, which often happens too late in a system's life cycle. This is especially true for unprecedented systems.

An unprecedented system is a system that has never been developed before. In a 1989 report from the Air Force Studies Board, an unprecedented system was defined as one for which development is based on a new technology, a new architecture, and/or a new acquirer/development team (i.e., they have never built anything like this before, or they have never worked together as a team), as opposed to systems with precedence, which include those for which the requirements are "consistent and well understood" (Beam 1989). This report detailed development methodologies that could be best used to reduce risk associated with the attributes of unprecedented systems.

It is interesting that at various times in military history, there has been a rapid advance in technology simultaneous with doctrine that has led to unprecedented new capabilities. These capabilities have often resulted from discovery-based experiments in exercises. One historical description of this comes from a fascinating article published in 1999 (Perry 1999). Perry reviewed the rapid development of air doctrine during the 1930s and the simultaneous advances in technology. The development centers of that day were co-existent with training and test facilities, which in turn led to simultaneous consideration of system definition, system test, and system use. It is in this spirit that we pursue a new approach—one that applies historical lessons—called The System Engineering and Test (TSET). TSET is enabled by approaches such as the Joint Mission Environment Test Capability (Lockhart and Ferguson 2008) and is coupled with evolving tools to support systems engineering to bring test considerations forward in the system engineering process. TSET also helps to consistently and effectively address test issues throughout all phases of the system life cycle. Further, the approach discussed will enhance the likelihood that different stakeholder constituencies (e.g., user, developer, acquirer, tester, trainer, sustainer) will be more likely to have a shared understanding of key issues related to the system during the different life cycle phases (Saunders 2005). This work is timely as current federal legislation is addressing the issue of earlier use of system engineering methods and test consideration in major weapons systems acquisition programs (U.S. Senate 2009)—points that are consistent with a recent National Research Council study (Kaminsky and Lyles 2008).

This article presents the tenets of using TSET for evaluating unprecedented systems and moving testing to the forefront of the development process. The overarching approach emphasizes three major aspects; specifically, there is a strong need for

- earlier discovery and experimental testing of systems and engaging the evaluators in the T&E process sooner rather than later;
- testing across a larger breadth of representative and realistic operational environments, both live and virtual, while stressing evaluation of unprecedented, innovative, and interacting systems operating in unpredictable environments; and
- engaging in campaigns of experiments to explore system alternatives through the co-evolution of systems, technologies, and concepts of operations (CONOPS).

The challenges

When systems are tested, there is usually one or a few variations of the system operating in the test environment, yet today's warfighting operations require collections of disparate platforms developed by different contractors to be interoperable, support joint and coalition missions, and be effective in environments that are unknown or were unforeseen when the contract specification was written. The U.S. Air Force Science Advisory Board recently noted "… it has become increasingly apparent that although the United States Air Force (UASF) *buys* systems in isolation, it does not *use* systems in isolation" (Saunders 2005).

Today's testing has a strong focus on system requirements and in assessing whether each requirement is met. New T&E methods are needed to allow system capabilities to be discovered and evaluated without the strict confines of a pass/fail test and to enable an understanding of the system relative to potential and actual missions. Based on the mission or operational environment, unprecedented systems and systems-of-systems may execute behaviors that cannot be precisely predicted. Discovery of capabilities and assessments of systems need to support evaluation of actions and judge whether the actions are reasonable and acceptable. Testing of these systems needs to focus on capabilities and potential missions. These notions are straining the current methods of T&E. New approaches to T&E must be adopted to support system discovery, innovation, and advances, where the T&E methods evolve and adapt, just as the systems do.

The challenges that seem to plague many programs include

- an incomplete understanding of the operational need, the potential operational need, and the related performance requirements;
- an often inadequate and ineffective tradeoff analysis between performance, schedule, and cost requirements, where performance is broader than component performance;
- an inadequate approach for defining key performance parameters (KPPs) that links them back to

needs, requirements, and potential needs and capabilities;
- testing considerations, both operational and developmental, that start too late;
- the lack of a shared understanding by stakeholder groups, including the developers, the testers, and the end users; and
- acknowledging that while the Department of Defense (DoD) has many facilities to support T&E, these resources are mostly devoted to formal systems testing (i.e., there is an inadequate ability to *simultaneously* address technical, operational, and production gaps in capabilities).

Many of today's systems are evolving to be much more capable of supporting collaborative and unscripted operations. Systems are moving away from being point solutions to instead being used in unorthodox ways. Based on the mission or operational environment, these systems have the potential to be used in ways that were never precisely predicted. Advances in sensors and systems are supporting operations in unstructured and hostile conditions, while providing the opportunity to identify new capabilities in situ. Systems are operating with other systems that have differing capabilities. Their operations are becoming more and more characterized by nonlinear responses and responses based on incomplete information. Yet, these characteristics do not preclude robust operations. What is unique is that new approaches are needed for discovering the capabilities and then testing and measuring this robustness, especially in nondeterministic and evolving environments. In today's wartime environment, where we need a rapid turnaround to get systems to the warfighter, adaptability is crucial.

Unprecedented systems will be used to their fullest extent only when the end user is confident in their operation. The best way to accomplish this is to increase the interactions between developers, testers, and users. Early involvement with operational test agencies will result in early identification of operation T&E expectations and needs. T&E needs to provide methods to explore system alternatives through innovative interactions of technologies, CONOPS, and experimentation and to explore the interplay between technology and CONOPS, just as was done in the 1930s.

The big "E" word

It has been said many times that T&E activities do not fund Experimentation, the *other big* "E" word. Somehow, the view that experimentation must be kept separate from T&E has infiltrated our systems thinking at a time when T&E and Experimentation truly need to be integrated. This does not mean that T&E must fund Experimentation or that program managers must pay for extra T&E out of their program development budget. It means the two must be coordinated and work together. It includes concepts as simple as providing technology developers with validated T&E tools to use when they are doing their own testing, and it includes concepts as expansive as the double helix, system-level experimentation approach described in the Air Force's 2006 Scientific Advisory Board report (Cross and Fouse 2006). The double helix approach supports discovery-level experiments by the testers as well as the developers *(Figure 1)*.

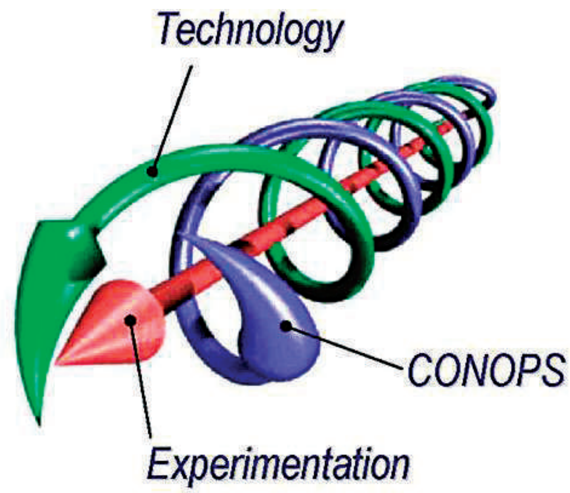


Figure 1. Co-evolution of systems. The double helix. Adapted from Air Forces 2006 Scientific Advisory Board report (Cross and Fouse 2006).

For unprecedented systems, where discovery experiments are needed to explore new CONOPS and new means to satisfy the CONOPS, co-evolution is required. Co-evolution is the idea that innovative technologies and CONOPS evolve and are developed together and evaluated through experimentation. This is key to creating a shared understanding of how the systems work, how they will be used, and how they could be used. T&E needs to be part of the experimentation process, and experimentation needs to be part of the T&E process, because a significant amount of testing, albeit system development testing, occurs early in the development process and is effected through experimentation.

The Air Force Scientific Advisory Board (Cross and Fouse 2006) noted that system level experimentation "increases the ability to discover game-changing ways to 'fly and fight' BEFORE the fight." Experimentally derived systems development has proven that higher payoffs are attained over requirements-focused devel-

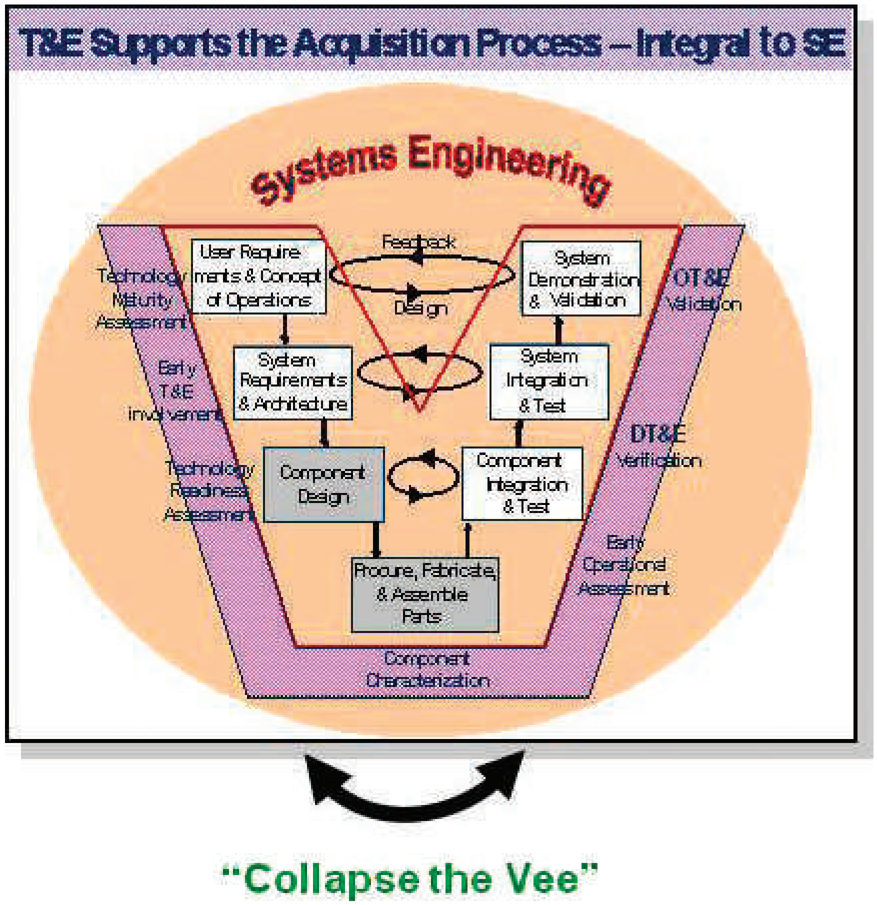


Figure 2. Collapsing the ''Vee'' so that testing and evaluation becomes an integral part of the Systems Engineering (SE) process.

opment. The Predator unmanned air vehicle was declared "not operationally effective or suitable" (Pogo 2002), yet it was deemed an immediate success since it could transmit live video feeds (Newman 2002). The Predator went from concept to deployment in less than 30 months. If experimental testing had been included before production, many initial glitches and cost increases may have been averted, and the Predator may have passed operational evaluation (Pogo 2002). The Predator evolved from an Advanced Concept Technology Demonstration (ACTD), and although the Predator is in high demand, even ACTD experimentation is not enough. Experimentation may be conducted through ACTDs or Joint Capability Technology Demonstrations (JCTDs). However, these programs are largely demonstration experiments, not the discovery experiments that are needed to explore and understand development of unprecedented systems.

One of the main tenets of system-level experimentation is to conduct iterative campaigns of discovery experiments to create a deeper understanding of future environments. Experimentation must be with the system, not just the technology, and it must stress the system via unconstrained adversaries. Is this experimentation the responsibility of T&E or is it a program manager's responsibility? The answer is "yes" and "yes"; everyone must be involved. Yes, it is experimentation, but tools to assess the system should come from T&E; the stakeholders (end users, warfighters, developers, and testers) must partake in early discovery experimentation. The result will be a better warfighting capability.

Collapsing the "Vee"

The best way to begin addressing these challenges is to develop new test technologies and approaches that increase interactions between developers, testers, and users. This amounts to collapsing the "Vee," so that T&E becomes an integral part of the system engineering process *(Figure 2)* (Buede et al. 2005). Constant, interactive feedback from the testers to the developers is crucial. To save both time and money, as test methods are developed and validated, they need to be promulgated among program managers for use in their programs. Newly developed test methods pro-

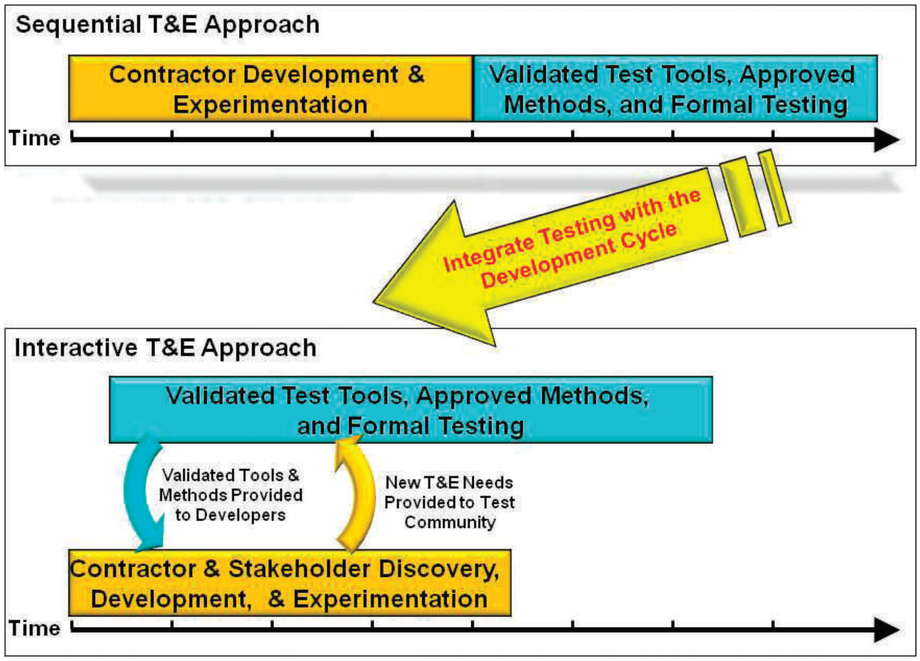


Figure 3. Integrating testing and evaluation into the development cycle.

vided to developers will allow system capabilities to be evaluated and enable an early understanding of the systems relative to mission success. As discovery, experimentation, development, and testing become aligned, the turnaround time to get systems to the warfighters will be dramatically reduced, resulting in T&E becoming part of a quick reaction capability to get systems in operation.

New approaches to T&E can be adopted that support advances in system capabilities, where the T&E methods evolve and adapt, just as the systems do. *Figure 3* shows how the T&E process can be adapted to support developers by providing them validated tools for experimentation, while the developers provide feedback to the T&E community on identifying new test needs. The relationship between testers and developers becomes interactive and integrated, where testers working closely with developers will discover new system CONOPS, and where developers attain an understanding of test needs. This relationship is integral and will accelerate introduction of innovative systems.

The TSET concept

The TSET concept is based on the double helix, system-level experimentation concept. The concept emphasizes that exercise-based experiments are needed to explore new CONOPS and that they are also needed to satisfy the CONOPS. Co-evolution is expected and is key to capturing these capabilities as part of the program record and to help create a shared understanding. By using the same environment for both developing the system performance requirements and testing the system, one can replay any state of the system so that a shared understanding is facilitated. Methods to support dynamic representations of test plans are linked to the experimental test environment, which then generate discussions on how KPPs will be measured and how the requirements will be tested. A collection of T&E tools can be used to facilitate these interactions, so that experimentation is not biased based on a contractor's suite of tools.

Validated and exploratory modeling and simulation tools can enable analysis of emergent systems in environments that are difficult to physically replicate or that present unforeseen circumstances. They allow numerous scenarios to unfold in shorter amounts of time, and they support new approaches to introspection assessments of the systems. T&E tools to support the experimental environment can assist in testing physical aspects of the systems at all levels, from component testing to bench testing and ground testing through collaborative operations. They can support testing payloads, sensors, weapon management, and data collection for large amounts of disparate test data. These would be the same tools that are used for actual T&E measurements, but they are provided to the experimental sites to support early testing and assessments. They also enable communication of shared visions and expectations, as well as the development of new CONOPS and the discovery of new uses for the systems. Tools for identifying new capabilities or analyzing new CONOPS are also needed to enable increased operator interactions with new families of systems, to provide metrics for mission effectiveness, to assess new types of "-ilities," and to determine new approaches for discovery and evaluation of unprecedented systems. These tools cannot be provided after the system is completed but rather need to be developed in concert with the technologies, the system, and the CONOPS. Everyone must be engaged from the start.

As complexity and capabilities increase, additional test scenarios will be needed to analyze and assess the operational envelope of unprecedented systems. As the number of tests increases, the testing time will also increase, and a larger burden will be placed on T&E. New testing approaches are needed that relieve this burden while ensuring the operational envelope is covered. System developers and testers need to interact early on in the development cycle to explore the test needs for these systems and to identify how to best test capabilities.

While it is difficult to reduce the level of complexity, system transparency can be increased. By employing experimental testing through TSET, transparency can be increased, and the states of subcomponents in the system can be exposed to allow evaluators to inspect the intent of these subsystems and discover new uses for the systems.

TSET applied to unmanned systems

As an example TSET domain of application, consider unmanned systems. Today, unmanned systems are operating in-theater with untested collaborative capabilities. The vehicles are heterogeneous, in that they are developed by different contractors they have different levels of autonomy, they have different sensors and capabilities, and they are physically disparate. Unmanned air vehicles built by one contractor have never autonomously collaborated with unmanned surface vehicles built by another contractor, and no one knows how they would perform if deployed together. Their integrated use, however, is rapidly growing in the military. As improvements in autonomy, sensing, and reasoning advance; collaborating, multi-vendor unmanned systems will be increasingly employed to support challenging, tactical operations. The anticipated increase in sophistication drives the need to collaboratively design, develop, test, and evaluate heterogeneous unmanned vehicles for full-spectrum dominance and joint operations (Robinson, 2008). We need a paradigm shift in T&E of unmanned systems that enables rapid discovery and flexible assessment of force-on-force capabilities of the effectiveness of disparate unmanned systems collaborating in theater-wide scenarios, while simultaneously ensuring safety of operations and stability to programs of record.

Addressing the test complexity of interacting, heterogeneous, intelligent, and autonomous unmanned systems requires a flexible experimental test environment for coupled hardware and software capability discovery and validated assessment tools for quantifiable analysis. TSET operations in experimental test environments support a less-formal, but physically meaningful way for unmanned systems developers to bring their systems for early-on exploration and testing, and make it possible for the end users and testers to become involved in the product sooner rather than later.

To enable this objective, TSET has several components: test tools and methodologies for component evaluation, modeling and simulation for high-level discovery, CONOPS exploration, analysis, and a flexible experimental test environment for coupled hardware and software experiments. These components are linked by analysis tools for assessments prior to fielding the systems and for quantifiable on-range evaluations of the systems. *Figure 4* presents the concept.

The modeling and simulation enables discovery and evaluation of potential emergent capabilities in environments that are difficult to physically replicate or that present unforeseen circumstances. It provides methods and tools to assess the impact of collaborative unmanned vehicle decision making, and it helps capture the amount of human awareness needed by the testers of the systems. As unmanned systems become more autonomous, the capability to communicate knowledge and information to the testers may not exist. For example, as systems become more autonomous and physically smaller (e.g., micro-unmanned vehicles), less data may be transmitted to a tester (because of power and size constraints) while, simultaneously, the systems will be making decisions autonomously and in unpredictable environments. It is not clear what information is necessary for the tester to be able to evaluate the autonomous aspects of these systems. Test tools need to evolve in parallel with the systems to determine what is really needed to test these systems and to discover what the systems are inherently capable of doing. The testers can stress the systems in unforeseen situations within a controlled digital environment with tools that are flexible and adaptable. They gain an understanding of the systems and their capabilities, which allows them to explore new CONOPS.

Such tools do not currently exist, and since technologies are still in development, it is difficult to create specifications and requirements for the *assessment* tools (let alone the systems). However, baseline test capabilities can be initiated and adapted as the systems evolve.

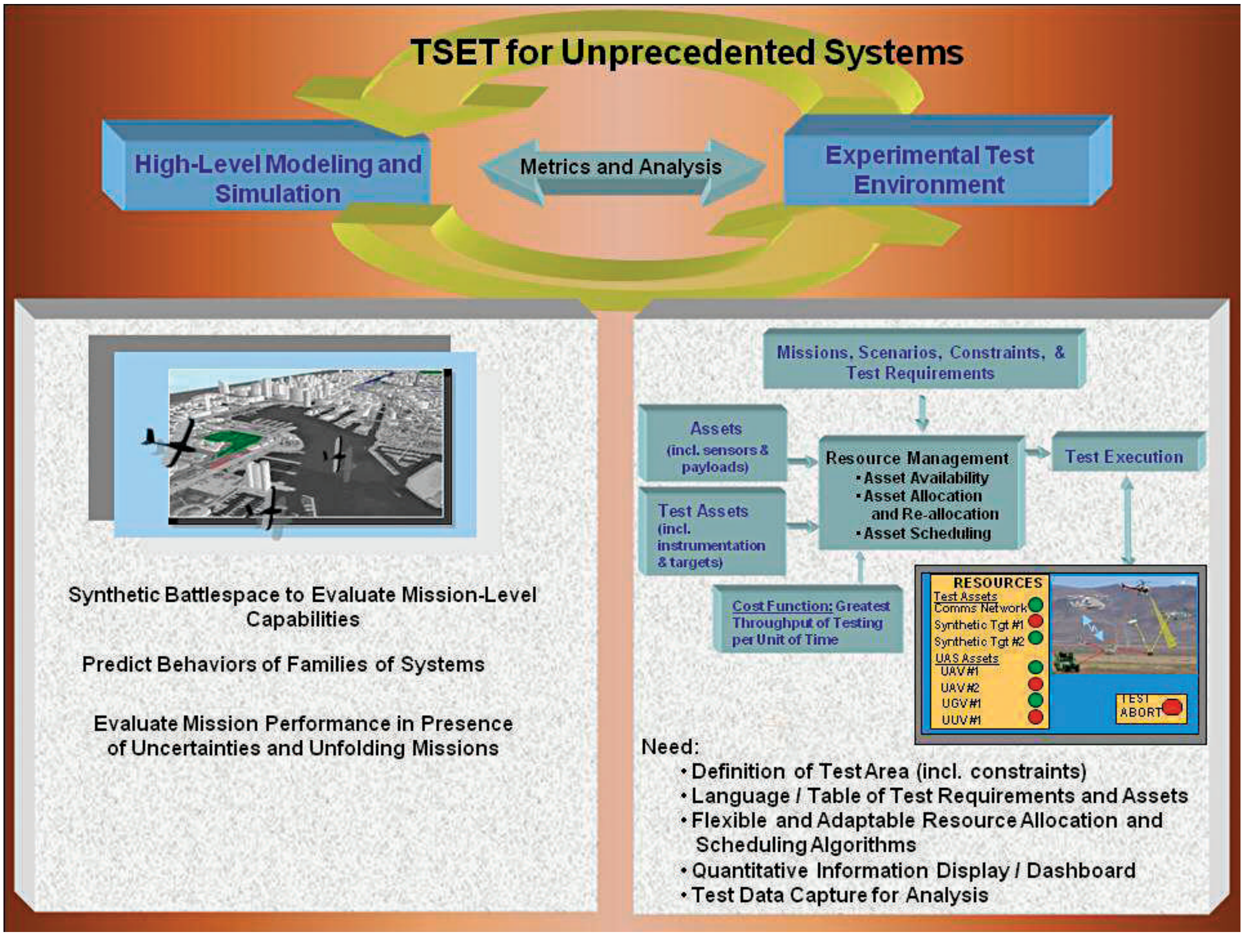


Figure 4. The System Engineering and Test (TSET) for unmanned systems.

For example, an unmanned system may display behaviors that emerge from actions and interactions with other manned and unmanned systems. The behaviors may be nondeterministic and unpredictable. System engineers may observe an emergent behavior and integrate technologies to reinforce or suppress what is observed, but it is difficult to know the performance gains attained from the emergent behaviors. Unmanned systems T&E requires capabilities to predict how modifications in the external environment may alter a behavior or possibly trigger an emergent behavior and what effect it has on mission performance. Tools are needed for exploration and assessment of unmanned systems that leverage advances in knowledge discovery and predictive analysis and that support an ability to identify potential outcomes of unmanned systems operations in nondeterministic environments. TSET methods allow unmanned systems capabilities to be discovered and evaluated without the strict confines of a requirements test and, instead, enable an understanding of unmanned vehicle autonomy relative to mission success.

The experimental test environment supports a less-formal, but physically realistic area for unmanned systems concepts to be tested early. It enables aspects of systems to be discovered and evaluated while there is still time to make adjustments. It enables end users and testers to become involved with the product sooner rather than later, and it enables exploration of physical interactions among multi-vehicle missions, where systems may have common goals, but decentralized control.

The modeling and simulation and the experimental test environment are connected by analysis tools. As the shift from systems-based development to capabilities-based development continues, new metrics and methods are needed to assess systems. The emphasis on system measures of performance will shift to measures of effectiveness, and those measures of effectiveness will evolve based on the systems, scenarios, and missions that are identified by testers and developers alike during early experimentation. For example, testing long-duration, persistent unmanned system operations will require long-duration analyses. New metrics will be needed for testing a different type of information overload in these long-duration scenarios, but we will not know what information is needed until we begin system-level experimentation.

Conclusions

As more unprecedented systems are used for multimission or collaborative operations, new challenges are placed on testing. There is a strong need to conduct systems engineering testing for coupled exploration of technologies, CONOPS, and capabilities. This is driven by increased connectivity, increased capabilities, and operations in complex adaptive environments. The TSET concept enables testing unprecedented systems and operations, where new information and knowledge are gained, and where increased interaction and integration results from early involvement from the end user. The TSET approach allows T&E to evolve as the systems evolve. It enables earlier testing of systems and engages the evaluators in the T&E process sooner. It supports testing across a larger breadth of representative and realistic operational environments, and it can be attained without causing major disruptions to current programs, but recognizing changes are needed. ❑

Dr. Lora Weiss is a lab chief scientist at the Georgia Tech Research Institute (GTRI), where she conducts research in all aspects of behavioral systems, from behavior-based unmanned and autonomous systems, to behavioral analysis of individuals and groups. She has supported research in intelligent autonomy for manned and unmanned systems, robotics, and control. Dr. Weiss chairs the ASTM standards development committee for Autonomy and Control of Unmanned Maritime Vehicles. Dr. Weiss is on the Board of Directors for AUVSI, the world's largest nonprofit unmanned systems organization. She has provided over 150 technical briefs to high-ranking DoD officers and DoD technology offices. E-mail: lora.weiss@gtri.gatech.edu

Rusty Roberts is the director of the Aerospace, Transportation and Advanced Systems (ATAS) Laboratory at the Georgia Tech Research Institute. ATAS develops advanced systems concepts, builds system prototypes, and performs research on technologies related to aerospace, transportation, power and energy, threat systems, and food processing. A nationally recognized expert in test and evaluation, Roberts is currently vice president and executive committee member of the International Test and Evaluation Association. He also started and presently leads a GTRI-wide T&E initiative that brought together the resources to provide science and technology support to the U.S. Joint Forces Command and the Office of the Secretary of Defense Test Resource Management Center. E-mail: Rusty.Roberts@gtri.gatech.edu

Dr. Stephen E. Cross is a vice president of the Georgia Institute of Technology, director of the Georgia Tech Research Institute, and a professor of industrial and systems engineering. He is retired USAF officer, a graduate of the Flight Test Engineer's Course at the USAF Test Pilot School, and a current International Test and Evaluation Association (ITEA) member. Before joining Georgia Tech in 2003, he was the director and CEO of the Carnegie Mellon Software Engineering Institute (SEI). Dr. Cross has published over 60 papers on information technology and technology transition. He is a fellow of the Institute of Electrical and Electronic Engineers (IEEE), the associate editor for the Journal of Information, Knowledge, and Systems Management, and a former editor-in-chief of IEEE Intelligent Systems. A former member of the USAF Scientific Advisory Board, Dr. Cross has supported many recent National Research Council and Defense Science Board studies on systems engineering. E-mail: Stephen.Cross@gtri.gatech.edu

References

Beam, W. 1989. *Adapting software development policies to modern technology.* Air Force Studies Board. Washington, D.C.: National Academy Press.

Buede, D., K. Forsberg, H. Mooz, C. Plowman and B. Tufts 2005. Systems engineering processes, *A Guide to the Systems Engineering Body of Knowledge (SEBoK)*, INCOSE, http://www.paper-review.com/g2sebok/seboksection2.htm (accessed July 15, 2009).

Cross, S. and S. Fouse 2006. *System Level Experimentation*. USAF Scientific Advisory Board. Washington, D.C.: Storming Media. http://www.stormingmedia.us/05/0593/A059364.html (accessed July 15, 2009).

DiPetto, C. and R. Stuckey 2007. A new vector for Developmental Test and Evaluation (DT&E). *The ITEA Journal of Test and Evaluation*. 28(1), pp. 160–166.

Gates, R. M. 2008. Secretary Gates remarks at Maxwell-Gunter Air Force Base, Montgomery, Alabama, *U.S. Department of Defense News Transcript.* Washington, D.C.: Department of Defense. Available online at http://www.defenslink.mil/transcripts/transcript.aspx?/transcriptid=4214 (Accessed July 15, 2009).

Kaminsky, P. and L. Lyles 2008. *Pre-Milestone A and Early-Phase Systems Engineering*. National Research Council. Washington, D.C.: National Academy Press. http://books.nap.edu/openbook.php?record_id=12065 (accessed July 15, 2009).

Lockhart, R. and C. Ferguson 2008. Joint Mission Environment Test Capability (JMETC). *The ITEA Journal of Test and Evaluation*. 29(2), pp. 160–166.

Newman, R. 2002. The little Predator that could. *Air Force Magazine* Online, vol. 85, No. 03. Available at www.airforce-magazine.com/MagazineArchive/Pages/2002/March%202002/0302 predator.aspx (Accessed July 15, 2009).

Perry, J. 1999. Air Corps Experimentation in the Interwar years—A Case Study. *Joint Force Quarterly*. Available online at http://www.dtic.mil/doctrine/jel/jfq_pubs/0922.pdf (accessed July 15, 2009).

Pogo. 2002. Fighting with failures series: Case studies of how the Pentagon buys weapons, Predator UnManned Aerial Vehicle. Washington, D.C.: Pogo. http://www.pogo.org/pogo-files/alerts/national-security/ns-puav-20020325.html (accessed July 15, 2009).

Robinson, R. 2008. Unmanned & autonomous: GTRI wins contract to develop technology roadmap for test and evaluation of new systems. *Research News*. http://gtresearchnews.gatech.edu/newsrelease/uav-test.htm (accessed July 15, 2009).

Saunders, T. 2005. *System-of-Systems Engineering for Air Force Capability Development*. USAF Scientific Advisory Board, SAB-TR-05-04. Washington, D.C.: DTIC. http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA442612 (accessed July 15, 2009).

Senate Bill 454. 2009. A bill to improve the organization and procedures of the Department of Defense for the acquisition of major weapon systems, and for other purposes. Washington, D.C.: Library of Congress. http://thomas.loc.gov/cgi-bin/bdquery/D?d111:2:./temp/~bdK3Ke::|/bss/111search.html| (accessed July 15, 2009).

ITEA Journal 2009; 30: 395–400

The Future of Spectrum in Weapon Systems Testing

Darrell E. Ernst
The MITRE Corporation, McLean, Virginia

Until recently, the test community had almost exclusive use of the bands they use for testing. The only users they had to share with, with a few trivial exceptions, were other testers. This is changing. The 1980s saw the advent of wireless consumer products, and the recent growth of WiFi, and WiMAX, and the expansion of the 5-gigahertz instrumentation radar band to enable non-radar usage of the band are examples of the continued growth in the use of the radio spectrum that represents a threat to the frequency bands used for testing. There are two ways in which the test community loses access to the spectrum. The first way is loss of allocation. The second way that access to the spectrum can be restricted is through the addition of users to the bands in use. There are some within the spectrum business who advocate commercializing all the radio spectrum, with the possible exception of that needed for public safety. Acquisition program managers increasingly will have to address spectrum issues early in their programs, including issue such as loss of access, increased schedule competition due to increased sharing and increased bandwidth demand, changes in technology that afford new opportunities, and risks of increased costs due to the imposition of access fees.

Key words: Radio spectrum access; bandwidth expansion; electromagnetic spectrum; frequency bands; spectrum trading companies; commercialization; satellite digital audio broadcast; wireless devices.

The physical entity commonly referred to as the radio spectrum is arguably the single most important resource for open air testing for at least the past 50 years. It is used for transporting measurement data, measuring position, velocity and radio reflectivity, command and control of test instrumentation and systems under test, range safety, and a host of other applications. The use of the radio spectrum has increased steadily with time. This increase is a function of data rate growth, decreased technology cost, and increasing test requirements. The test community is not the only sector expanding its use of the radio spectrum. Radio usage and bandwidth expansion are growing rapidly in many sectors, including the military, civil aviation, medical community, emergency response and management, commercial telecommunications, consumer wireless applications, and the sciences. Indeed, the demand for spectrum has made it a commercially valuable asset, with auctions bringing in billions of dollars in revenue. Beginning this November, spectrum will become a commodity when the first private spectrum trading company opens its doors (Cox 2008). Because there can only be a finite number of spectrum users in any location at a given time, the test community will be challenged to maintain and increase its access to the radio spectrum as the pressures of market demand place the frequency bands we use at ever greater risk.

The radio spectrum is considered by spectrum regulators to span the electromagnetic frequency range from 3 kilohertz to 300 gigahertz (3×10^{11} Hertz or cycles per second) (U.S. Department of Commerce 2007). There are many factors that dictate the use of the radio spectrum, but two factors dominate: (1) the amount of information that can be conveyed via the spectrum is directly proportional to the bandwidth that is used; and (2) the higher the frequency used, the more difficult it becomes to use the spectrum for information communications. The first factor drives our efforts to use ever-greater frequencies to meet the demands for more information throughput, whereas the second factor increasingly constrains our ability to use the upper portions of the spectrum. By the 1950s the test community was able to efficiently communicate information in the range between 100s of kilohertz up to about 300 megahertz. However, pressures on the spectrum by other users drove the test community to seek access to higher frequency bands (Jeffries 1970). In the very early 1960s a technology breakthrough in a classified military space program gave the community the capability to use the spectrum efficiently and inexpensively up to around 5 gigahertz. During the 1960s and first part of the 1970s the test community developed the radars, telemetry systems, and other communications systems that are still the backbone of our open air test infrastructure.

The threat

The test community, from a spectrum access standpoint, lived in a blissful time of abundance from the 1960s until the late 1990s. But the same technologies that gave them the capabilities they enjoyed fostered a little noticed revolution in the commercial telecommunications industry. The 1980s saw the advent of wireless consumer products that ultimately led to the establishment in 1985 of the International Mobile Telecommunications 2000 (IMT-2000) initiative by the International Telecommunications Union (UMTS 2008). The purpose of IMT-2000 was to adjust the international radio regulations to enable access to the spectrum by the many new wireless products waiting in the wings. The first shot across the bow for the U.S. test community was the Omnibus Budget Reconciliation Act of 1993, or OBRA-93, which was the U.S. response to IMT-2000 (U.S. Department of Commerce 1995). In this legislation, Congress directed that 200 megahertz of radio spectrum previously reserved for Federal government use be reallocated to the Federal Communications Commission (FCC) for auction to commercial telecommunications enterprises. The Executive Branch, in fact, released 220 megahertz to the FCC, more than 80 percent of which had been used by the Federal elements of the test community. Even then, the community paid little notice to the taking, mainly because the actual losses would not be felt at the user level for many years. It was not until Congress, enamored by the amount of money brought to the Federal coffers by the OBRA-93 auctions, passed the Balanced Budget Act of 1997 (BBA-97), demanding another 20 megahertz, that the test community realized they were under attack. Thirteen of the 20 megahertz for BBA-97 came from the spectrum used for flight test telemetry (U.S. Department of Commerce 1998). The Department of Defense test organizations partnered with the National Air and Space Administration (NASA) and the aircraft manufacturers to confront the threat to their access to the spectrum. The partners devised a plan that continues to operate to this day. There is good reason for the continuance of the coalition—the threat to their spectrum access not only continues, but in fact, continues to increase. The growth of WiFi, and now WiMAX, and the recent expansion of the 5-gigahertz instrumentation radar band to enable non-radar usage of the band are examples of the continued growth of the threat to the bands used for testing.

There are two ways in which the test community loses access to the spectrum. The first way is loss of allocation. This means that the international and/or national radio regulations are changed to reallocate a particular band from one kind of radio service to an alternative service. This is what happened with the bands reallocated under OBRA-93 and BBA-97. For example, parts of the bands previously allocated for aeronautical telemetry for Federal users were reallocated for non-Federal Mobile service.

The second way that access to the spectrum can be restricted is through addition of users to the bands in use. Until recently, the test community had almost exclusive use of the bands they use for testing. The only users they had to share with, with a few trivial exceptions, were other testers. This is changing. The same technological factors that are driving the commercial sector are also driving the Federal users. Since Federal users must use Federal spectrum where possible, new user systems must be assigned to these bands or compete with non-Federal users for access to the shared Federal/non-Federal bands. As this is often not feasible for a number of reasons, these new Federal users must be assigned to the already-crowded Federal bands. So, for example, the military's new data link for exchanging targeting information operates in a number of bands, including one band used for aeronautical telemetry and one band used for test article position measurement. One critical band shared with the civil flight test community is at risk of being shared with a commercial medical telemetry service, and a band used by range instrumentation radars must now share the band with WiFi users.

Reallocation and band sharing are not the only risks on the ascendant. Many of the new wireless services operate in bands adjacent to bands used by the test community. Many of these devices are built as inexpensively as possible. This means that they do not necessarily constrain their emissions to the band in which they are supposed to operate. Many of these devices splatter over into adjacent bands, despite FCC regulations. If one of these devices is operating near a sensitive test system, it can interfere with the system to the extent of rendering it unusable. There is a documented case of a wireless Internet device interfering with a sensitive tracking radar (Fleischman 2008) to such an extent that the unmanned vehicle that was being tracked was lost. Numerous other instances of interference have been reported, and test range spectrum managers report that instances of both in-band and adjacent band interference are on the rise.

These threats to the spectrum critical to the test mission do not bode well. The test community is receiving ever-more-sophisticated systems for testing that require ever-increasing-amounts of spectrum to obtain the necessary test data. One study shows that the demand for telemetry spectrum has continued to grow exponentially for more than 30 years, and the end is not in sight (Ernst, Hoh, and Portigal 2004). In 1972 a high data rate test was around 100 kilobits per second and needed about 300 kilohertz of spectrum to transmit the data. By 2022 a high data rate test will transmit 500 megabits per second and, even assuming the use of the most advanced modulation techniques, will need on the order of 400 megahertz of spectrum. Because the U.S. only has about 215 megahertz of spectrum allocated for aeronautical telemetry, such a high data rate test would not be possible unless something changes. Fortunately, such a change has occurred.

Responding to the threat

In 1997 a manager in the Pentagon office responsible for test range resources support recognized the threats that recent spectrum actions posed to the U.S. test ranges. He established a spectrum initiative to address the threats. One part of the initiative led to the establishment of a program to develop a modern network-based data communication system that would use the spectrum more efficiently. A second part of the initiative was the establishment of a science and technology program to investigate technologies that would increase the efficiency with which the spectrum is used for data transmission. A third part of the initiative resulted in a decision at the 2007 World Radiocommunication Conference to allow aeronautical telemetering in several radio sub-bands with an aggregate total bandwidth in the Americas of approximately 1.4 gigahertz of spectrum. The combination of these undertakings has substantially improved the outlook for one of the most important communication applications used by the test community. The aggregate economic impact of the initiative was estimated in one study to be on the order of $10 billion ($10,000,000,000) in cost avoidance (Ernst, Kahn, and Portigal 2007).

The Pentagon's telemetry spectrum initiative is illustrative from several standpoints. It demonstrates that changes to the way spectrum is acquired and used take years to see results. But it also shows that a consistent, focused, well planned and well managed program can be effective in addressing spectrum access challenges. Important to this narrative is that the work on the initiative gave hints to the future of radio spectrum usage and the challenges that we may face in adjusting to the revolution occurring in the world of radio.

Radio technology revolutionized the global human culture in the early part of the 20th century. Broadcast radio and then television became a part of people's daily lives. The invention of radar led to greatly improved air travel and weather forecasting. And then radio technology became passé as the personal computer and the Internet exploded onto the scene. Very rapidly, people wanted to be connected to the Web wherever they were, and they wanted access to entertainment and to their friends. This has led directly to the second revolution in radio technology. There are many different kinds of radio applications entering the marketplace, from simple cell phones to personal wireless devices, to broadcast radio from space, very wideband network services, automotive concierge services, personal medical telemetry, NASCAR telemetry, automotive anti-theft services, and many more. All of these radio applications have one thing in common—they all need to use the electromagnetic spectrum. And it is a law of physics that at any given point in space at any given instant of time there can only be one value of field strength for a given frequency of electromagnetic energy. This fact makes it very difficult to have more than one user of that space–time–frequency coordinate. Technology is beginning to make some inroads on this limitation, but there is a long way to go before everyone can use the entire radio spectrum simultaneously. That means we are stuck with the current way of sharing of the spectrum: one user per space–time–frequency coordinate. During the more than 100 years of sharing the spectrum, the management of its use evolved into a structure whereby the spectrum was divided up into about 490 bands, with the 30-plus types of "radio service" assigned to different bands to separate incompatible services from each other. For example, communications between airplanes and air traffic controllers do not operate in the same frequency bands that commercial broadcast radio operates because the airplanes would interfere with the home and automotive radio receivers. The implications of this basic radio management structure for the new wireless services are significant. In order for them to be operated, they must fit into the existing spectrum structure. But since the bands in which they can operate are already licensed to other users, ways must be found to accommodate both the new services and the incumbents. One of the user groups most affected by this new and growing push on the spectrum is the test community, so it is important to understand what is happening and what it bodes for the future.

One way that new radio services are admitted to the spectrum is to move the incumbent users from a particular band or set of bands and allocate those bands to the new service. This is the method that was used to accommodate the so-called "3G" or "third generation" wireless services and the satellite digital audio broadcast services. Both of these services moved into bands that had been used previously by the test community. At the time of the reallocations, no new spectrum bands were opened for the displaced test users. Instead, the test community was required to absorb the loss, with the concomitant impact to test operations.

Another way to accommodate new services is to place limits on the services. So, for example, a new service that is restricted in the amount of power it may radiate may be assigned to the same bands as other power-limited devices such as garage door openers and wireless key entry systems. A more complex approach is to place power, geographic limits, and a regulatory status of "secondary" on a service and allocate it to a band where the incumbent users operate at higher power levels, have primary status, but operate in geographically restricted areas. This is precisely the solution being proposed for a new medical telemetry system that would share a band with the test community (FCC 2008).

In general, any method of introducing new categories of spectrum usage necessarily places restrictions on either the incumbents or both the new services and the incumbents. This admitting of new radio services is not new, but in recent years it has been steadily increasing. This has put pressure on all user sectors to find new ways of improving the efficiency of the way the spectrum is used. One promising technology is the so-called *software defined radio*. The basic idea is that the radio is embodied mostly in software rather than in physical hardware components. This has introduced the possibility of using software coupled with advanced digital signal processors to use complex waveforms that enable radios to work in a more crowded spectrum. The so-called *cognitive radio* carries the software defined radio concept further by giving the software defined radio intelligence. Two promising techniques being explored with cognitive radios are *white space mining* and *policy-based radio*. Radios that mine radio white space listen for quiet parts of the spectrum and send out data in these quiet spaces, i.e., narrow frequency bands that are momentarily not being used by someone else. Policy-based radios have some means of knowing their location (e.g., Global Positioning System) and have a database listing all users in a given region. By combining its knowledge of where it is located, what authorized users are in its locale, and the rules by which it is permitted to operate, the radio attempts to adapt its operating frequency and bandwidth to unused parts of the spectrum at the operating location. Both of these technologies are still very immature for general usage, but they may have some applicability to test applications.

Spectrum as a commodity

Another trend in the spectrum management community is the increased interest by national administrations to use the radio spectrum as part of their economic infrastructure. Spectrum auctions worldwide since the mid-1990s generally have been very successful, and many nations are openly advocating the commercialization of the spectrum to the maximum extent possible. As a corollary to that, some nations, such as the United Kingdom, are charging government users for use of the spectrum. This concept is being watched with interest here in the U.S. and is being studied closely. How such an approach would be implemented in the test community is unknown. Most of the open air testing conducted within the U.S. involves two entities: a test range that provides the test infrastructure, including numerous communications services; and a program office that pays the test range for the use of its resources and services. The costs are generally based on some formula involving the cost to the range to operate and maintain each of the resources. But such an approach does not easily lend itself to valuing the radio spectrum resource. It takes no more labor to operate and maintain a radio using 2 megahertz of spectrum than it does for a radio using 20 megahertz. If the labor-based costing method were used, narrow bandwidth users would object to paying the same rates as high bandwidth users, and there would be no incentive to conserve bandwidth. One approach that is often proposed is based on the cost to obtain the data. If the total cost of a particular test is $10 million, then the cost to replace the data from that test is $10 million, discounting inflation. The cost per bit of data is then easily calculated. Some simple calculations convert the number to the cost per Hertz per hour of operation of spectrum usage. The next step is the difficult part: how does one determine what percentage of this cost should be rental fee paid to the government? Some propose valuing the spectrum based on FCC auction income. Others suggest using methods by private industry to bill for their wireless services. Regardless of what method of valuing the spectrum is used, a user fee based on the number of Hertz used per hour would almost certainly result in a decrease in the amount of data transmitted via radio.

The repercussions of such an approach need very careful study.

There are some within the spectrum business who advocate commercializing all of the radio spectrum with the possible exception of that needed for public safety. The implications for the test community of this approach are very difficult to assess without knowing the regulatory details, but there are some hints that can be gleaned from some past examples. One DOD training range rented a band for years from a cellular service provider, but had to give it up when urban expansion led to the cellular provider reclaiming the spectrum to bring service to the area surrounding the range. Another illustrative example occurred in Europe. After the flight test community in Germany lost most of its spectrum to the "3G" auction in the late 1990s, one test organization rented spectrum from the post office to conduct testing on a fighter aircraft.

The test community is not watching the spectrum drama idly. Given that there are certain categories of test in which data can only be obtained wirelessly, and recognizing that the radio spectrum is under great pressure, some elements of the test community are exploring alternatives to the current ways of obtaining test data. Led primarily by the Test Resource Management Center, but involving partnerships with others such as NASA and the Aerospace and Flight Test Radio Coordinating Council, various initiatives are underway to ensure the future. In the near term, spectrum efficient radios have been developed and are now available commercially. In development are a spectrally efficient, network-capable, common range instrumentation system and a spectrally-efficient, multi-band, wideband, long-range wireless telemetry network. A new initiative just getting underway is the Test Resource Management Center's *15 Gigahertz and Above* initiative, which will commence with a preliminary "state of the technology" survey that will address the potential for electromagnetic communication between 15 gigahertz and the ultraviolet band. Free space optics is very high on the list of candidate research areas.

Summary

The genie is out of the bottle. Revolutionary changes are occurring in the way humans use the electromagnetic spectrum, especially the radio spectrum. The test community is a major user of the spectrum and is being swept along in the revolution. The days are rapidly approaching when we can no longer just call the frequency management office and schedule a test. Acquisition program managers increasingly will have to address spectrum issues early in their programs. Test resource managers will have to look beyond their 5-year budget window and adapt to the changing spectrum world. They will have to participate in international spectrum forums and technology programs that enable continued access to the spectrum. Range commanders and range directors will have to incorporate spectrum planning, budgeting, and administration into their business enterprise processes at the strategic level.

All testers should be conversant with the basic spectrum issues that could affect them. This includes loss of access, increased schedule competition due to increased sharing and increased bandwidth demand, changes in technology that afford new opportunities, and risks of increased costs due to the imposition of access fees.

It is incumbent upon all of the test community, including range commanders, range directors, spectrum managers, resource managers, program managers, test planners, and test directors to become informed about the spectrum requirements and spectrum challenges that will affect their enterprise, and to do so in sufficient time to mitigate adverse impacts to their missions. That worthy philosopher, Francis Bacon, admonished us to "…leave the future to divine Providence" (Knowles 1959, p. 24). Such a philosophy may be appropriate for an outing in Las Vegas, but it is poor guidance for a tester. Much better are the words of Richard Hengist Horne: "Ye rigid Ploughmen! Bear in mind your labor is for the future hours" (Knowles 1959, p. 262). And just as it is for the ploughmen, so it is for the tester, especially with regard to radio spectrum. ❑

DARRELL ERNST is a principal space systems engineer with the MITRE Corporation. He has been an advisor on spectrum issues to the test ranges in the United States for more than 10 years. In that capacity, he has been involved in all of the major test range spectrum encroachment issues resulting from government regulatory actions. He was a leader in the initiative to gain access to more spectrum for flight testing through the World Radiocommunication Conference process. He was a member of the U.S. delegation to the World Radiocommunication Conference and to Working Party 8B of the Radio Bureau of the International Telecommunications Union. In that capacity he was a technical advisor to the delegation on issues associated with the use of spectrum for aeronautical telemetering. He is one of the founders of the International Consortium for Telemetry Spectrum (ICTS) and speaks internationally for the ICTS on current issues. Mr. Ernst began his career in the U.S. Air Force in 1961 as an airborne instrumentation technician, recording telemetry

data from rockets and spacecraft while flying in remote regions of the world. He participated in all the major U.S. strategic missile development programs and the Mercury, Gemini, and Apollo manned space programs. After completion of his graduate work he was assigned to the Air Force Communications Command in which he was responsible for the assessment of the performance of the Air Force's worldwide microwave backbone communications system. His last three tours in the Air Force were with the Air Force Test & Evaluation Center. After retirement from the Air Force he worked for TRW on the Space Shuttle program, and on what eventually became Schriever AFB. Mr. Ernst is co-inventor on several communications-related patents. He is a member of Ruritan National, the International Test & Evaluation Association, and the ICTS. He teaches an International Test & Evaluation Association course on radio spectrum management for test and evaluation. Mr. Ernst has a bachelor of science degree in mathematics from Auburn University and a master of science degree in operations research and statistics from Rensselaer Polytechnic Institute. He has done post-graduate work in industrial processes and is the author of numerous papers, most of which are not in the public domain. E-mail: dernst@mitre.org

References

Cox, J. 2008. Meet the eBay of wireless spectrum. *Network World*. 12 September 2008. http://www.networkworld.com/news/2008/091208-wireless-spectrum.html (accessed September 30, 2008).

Ernst, D. E., Y. S. Hoh, and D. Portigal. 2004. *Projected growth of spectrum requirements for aeronautical telemetering*. McLean, VA: The MITRE Corporation.

Ernst, D. E., C. A. Kahn, and D. L. Portigal. 2007. *The economic importance of adequate aeronautical telemetry spectrum*. McLean, VA: The MITRE Corporation.

Federal Communications Commission (FCC). 2008. *Petition for rulemaking*. ET Docket No. 08-59, Aerospace and Flight Test Radio Coordinating Council, Federal Communications Commission filing, May 2008. Washington, DC: FCC.

Fleischman, G. 2008. *Radar violators in Florida*. Seattle: WA: Wi-Fi Net News. http://wifinetnews.com/archives/004757.html. January 28, 2005 (accessed September 25, 2008).

Jeffries, C. E. 1979. "Past History and Present Status of Aeronautical Telemetry Bands," Minutes of the Inter Range Instrumentation Group Telemetry Working Group, Range Commanders Council Secretariat, White Sands Missile Range, NM, March 1979.

Knowles, E. 1959. *The Oxford dictionary of quotations* (2nd edition). New York: Oxford University Press, 24, 362.

UMTS. 2008. *UMTS/3G History and Future Milestones*. Pirkanmaa, Finland: UMTSWorld.com. http://www.umtsworld.com/umts/history.htm (accessed September 25, 2008).

U.S. Department of Commerce. 1995. *Spectrum Reallocation Final Report–Response to Title VI – Omnibus Budget Reconciliation Act of 1993*. February 1995. Washington, DC: U.S. Department of Commerce National Telecommunications and Information Administration.

U.S. Department of Commerce. 1998. *Spectrum Reallocation Report–Response to Title III of the Balanced Budget Act of 1997*. February 1998. Washington, DC: U.S. Department of Commerce National Telecommunications and Information Administration.

U.S. Department of Commerce. 2007. *Manual of Regulations and Procedures for Federal Radio Frequency Management*. May 2003 Edition, January 2007 Revision. Washington, DC: U.S. Department of Commerce National Telecommunications and Information Administration.

ITEA Journal 2009; 30: 401–408

Paths for Peer Behavior Monitoring Among Unmanned Autonomous Systems

Rick Dove
Kennen Technologies, Questa, New Mexico

Aberrant behavior in unmanned weapon systems can be caused by design flaws, system malfunction, malevolent control penetration, and human error. As systems manifest more autonomous operation, unpredictable emergent behaviors add to this list. The near-term expected future calls for swarms of heterogeneous unmanned autonomous systems (UAS) to be employed on a mission. The emergent group behaviors will add new dimensions to testing, posing potentially explosive centralized monitoring and evaluation tasks with large groups. The impossibility of specifying and testing all potential situational conditions is recognized, and the safety of the testing environment itself is of concern. Lessons from social animal life show that peer behavior monitoring has evolved to detect and mitigate aberrant behavior among members, and mitigating that behavior if it is evaluated as intolerable. This article explores a foundation for peer behavior monitoring among UAS under both test and mission conditions.

Key words: Aberrant behavior; emergent social behavior; ethics; parallel pattern recognition; peer monitoring; robots; self-organizing systems-of-systems; soldiers.

Aberrant behavior in mobile unmanned autonomous weapons is likely. Regardless of the degree of autonomous control, aberrant behavior can be caused by design flaws, system malfunction, malevolent control penetration, and human error. In fully autonomous operation, unanticipated emergent behaviors are both likely and desirable in dealing with the infinite possibilities of situational reality. Simulation and test of individual units with these autonomous capabilities have their own sets of challenges and cannot predict how these units will behave in group operations. Individual behavior cannot be ignored as simulation or testing advances to group behavior and poses an explosive centralized monitoring and evaluation task with large groups.

Social animal life exhibits built-in systemic mechanisms for detecting aberrant behavior among its members, and mitigating that behavior if it is evaluated as intolerable. This article identifies a foundation for employing socially attentive monitoring in Unmanned Autonomous System (UAS) predeployment testing, and in perpetual peer evaluation after deployment. The suggested approach was instigated by studies of self-organizing systems-of-systems in a graduate systems-engineering course at Stevens Institute of Technology (Dove 2007) and deemed possible by capabilities of a new pattern detection-engine technology (Dove 2009). The foundation explored in this research was shaped by this targeted technology.

This article reports on part one of a two-part study, identifying a promising behavior detection approach that might benefit from a massively parallel pattern recognition capability. Part two of the study investigates the potential of massively parallel classification technology for leveraging the detection approaches outlined in this article.

In *The Principles of Ethics*, Herbert Spencer reaches into the animal kingdom to support his theories on the origins and enforcements of natural laws within social groups:

> *"There arises such general consciousness of the need for maintaining the limits, that punishments are inflicted on transgressors—not only by aggrieved members of the group, but by the group as a whole. A 'rogue' elephant (always distinguished as unusually malicious) is one which has been expelled from the herd: doubtless because of conduct obnoxious to the rest—probably aggressive. It is said that from a colony of beavers an idler is banished, and thus prevented from profiting by labours in which he does not join:*

a statement made credible by the fact that drones, when no longer needed, are killed by worker-bees. The testimonies of observers in different countries show that a flock of crows, after prolonged noise of consultation, will summarily execute an offending member. And an eye-witness affirms that among rooks, a pair which steals the sticks from neighbouring nests has its own nest pulled to pieces by the rest." (Spencer 1893, 12–13)

Though stories of beaver and rook justice, and anecdotal witness to crow judgment and execution exist, scientific evidence is illusive; nevertheless, the values and varieties of peer judgment constraining and enforcing societal behavior are well known among humans and are studied and observed in animal (Flack et al. 2006) and insect societies (Heinze 2003, Monnin et al. 2002, Ratnieks, Foster, and Wenseleers 2006).

This article suggests that peer evaluation of behavior is necessary and valuable in the domain of autonomous unmanned systems when they are working together as a team on a warfighting mission, and perhaps even more so when these systems are being tested, as they are less likely to be well behaved. The suggestion is prompted by the positioning and planning for an Unmanned Autonomous System Test (UAST) focus area described in a 2008 Broad Area Announcement:

"Due to the mobility inherent in all UAS, their close proximity to humans (e.g., soldiers, testers, population centers, etc.) and their capability for unpredictable behavior; a reliable fail-safe system is needed. This effort seeks technologies for all aspects of system safeties as they pertain to UAS, Systems of Systems, and Complex Systems. This includes safe test conduct, testing for top level mishaps, safety fail-safes, truth data assessment for safety, and safeties associated with large numbers of collaborating UAS." (Office of the Secretary of Defense 2008, 21)

It is also recognized that testing outcomes can have "an almost infinite number of possibilities, depending on UAS cognitive information processing, external stimuli, operational environment, and even possible random events (hardware/software failures, false stimuli, emergent behavior, etc.)." (Office of the Secretary of Defense 2008, 23)

Emergent behavior is later recognized as something less than random:

"UAS formation control, swarming, and aggregate intelligent agent behavior are an emergent characteristic of this technology arena. … System behavior, in a multi-agent system, can be difficult to predict and often unexpected system behaviors occur which lead to poor system performance. These unexpected system behaviors result from unforeseen group actions of agent groups and agent-group behavior that is not directly coded by the agent designers." (Office of the Secretary of Defense 2008, 54–55)

Such unexpected system behaviors can be good as well as bad. In fact, the goal of fully autonomous intelligent behavior is creative problem-solving in situations without precedence. It is unlikely that unleashing a swarm of UASs that are only capable of dealing with well-defined cataloged situations will be effective.

We cannot know the situations that will arise, nor can we directly control how things should play out. Instead, we must recognize and embrace uncertainty within a framework of governance principles that will bound the outcomes within an acceptable space. The principle described in this article classifies behavior as unacceptable based on absolute boundary-infraction recognition, rather than attempts at imperfect reasoning or restrictions to specifically approved behaviors.

UAS will necessarily be tested and fielded in situations that have no precedence in cataloged responses. How will we constrain the outcomes to those we can live with? More to the point, how will we detect unacceptable behavior in time to intervene if unacceptable consequences are the likely outcome?

Dove and Turkington (2009) characterize agile systems as class 1 if they are (operator) reconfigurable and class 2 if they are (systemically) reconfiguring. It can be useful to think of a class-1 UAST system testing class-2 UAS systems. This agile-system class distinction arose from a graduate course in the School of Systems and Enterprises at Stevens Institute of Technology (Dove 2007). The course reviews the literature in various bodies of knowledge relevant to self-organizing systems, and challenges collaborative student analysis to identify recurring and necessary patterns across bodies of knowledge. Five cycles through to date, this investigation is beginning to yield some promising fundamental patterns. One in particular is the genesis of this article's focus: successful social systems often exhibit a pattern of peer behavior-enforcement arising when the stability of the social system is put at risk.

A related body of work led by Ronald Arkin (2007) at Georgia Institute of Technology is concerned with

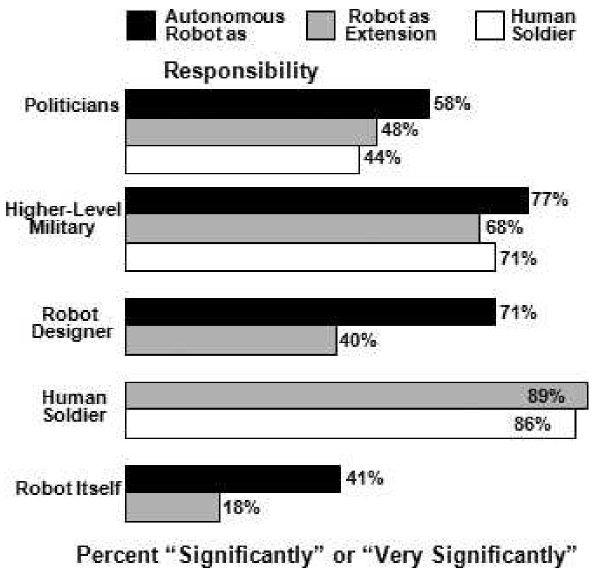


Figure 1. Responsibility for lethal errors by responsible party based on a survey of 430 respondents from demographic groups. The soldier was found to be the most responsible party, and robots the least. (Reproduced with permission from the survey reported in Moshkina and Arkin 2007.)

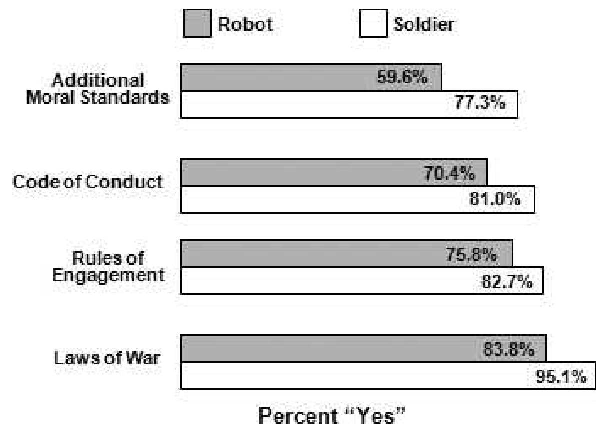


Figure 2. Ethical behavior for soldiers and robots. Applicability of ethical categories is ranked from more concrete and specific to more general and subjective. (Reproduced with permission, from the survey reported in Moshkina and Arkin 2007.)

ethical behavior of UAS used in military operations and recognizes the potential for peer monitoring: "When working in a team of combined human soldiers and autonomous systems, they have the potential capability of independently and objectively monitoring ethical behavior in the battlefield by all parties and reporting infractions that might be observed."

A recent survey investigated opinions about the use of, and responsibilities for, lethal autonomous systems among four demographic groups (Moshkina and Arkin 2007). A total of 430 respondents were distributed demographically as 54% robotics researchers, 30% military, 16% policymakers, and 27% general public. *Figure 1* depicts who the respondents felt was responsible when behavior was unacceptable. Our interest here is in the autonomous devices, not the "robot as extension" case, in which a human directs the unmanned system. Interesting to note: lethal mistakes made by a UAS are blamed on higher-level military, UAS designers, and politicians, in that order.

The survey showed that all four demographic groups want ethical standards for UAS to be considerably higher than those for soldiers. *Figure 2* shows how the groups felt about specific constraints that should be enforced. Monitoring for ethical behavior infractions is a subset of what must be monitoring for safe and secure behavior overall.

Two sections follow that discuss social peer-behavior monitoring; first in terms of temporal relationships and then in terms of spatial relationships. Temporal behavior monitoring compares the temporal event sequence of the intended tactical plan against the actual sequence of events, on an agent-by-agent basis. Spatial behavior monitoring compares the spatial aspects of the intended plan against actual trajectories on an agent-by-agent basis.

Temporal behavior leverage

Our fundamental interest is in the ability to detect and evaluate certain aspects of the behavior of team members as they work toward a common goal. This common goal may encompass a set of tasks that are not necessarily shared by all team members but are nevertheless a part of the activities pursuing common goal achievement. Task plans for achieving this common goal will have constraints. For instance, the end may not justify any possible means. Achievement may also have constraints on team member behavior (e.g., team members are expected to work toward team goals according to an established coordination plan).

A team is defined as a collection of members (agents) working toward a common goal. Working together implies some form of activity coordination. Coordination comes in a range of forms from centralized planning and micro-direction of the agents at the one extreme, to mindless local-reaction agent-behaviors resulting in emergent swarm effects at the other extreme. Our interest is in neither extreme, but rather with autonomous agents that possess and employ both self-awareness and social awareness of other team members and their behaviors as they jointly pursue a mission. Agents will have a sense of team and a sense of mission, and use this information to detect when another agent is clearly not behaving in the team's interest.

In a broader socially aware sense, the team's interest includes the team's image among outsiders—a weapons-toting team member gone rogue can impair the team's long-term existence likelihood. This is a new behavior focus not seen in prior research of Multi-Agent Systems (MAS).

Malone and Crowston (1994) in a broad interdisciplinary survey of the "*emerging research area, sometimes called coordination theory,*" define coordination as "*managing dependencies between activities.*" Noting the onslaught of the electronically connected world, they proposed that new forms of organizing and new forms of coordination structures would emerge. They also observed that different disciplines were already exploring domain-specific coordination concepts, and that there was now value to be gained in finding domain-independent underlying principles of coordination. Their stated intent was to help kick-start the development of a *theory of coordination* by illuminating these cross-discipline similarities, noting that "*It is not enough just to believe that different systems are similar, we also need an intellectual framework for 'transporting' concepts and results back and forth between the different kinds of systems.*" This idea is germane to the present discussion as this article suggests that coordination concepts of social systems inform how we deal with aberrant behavior in UAS.

About the same time as Malone and Crowston pulled together their survey, Jennings (1993) modeled coordinated agent communities on a foundation of commitments and conventions. Jennings defines commitments as mutually agreed upon plans of action, and conventions as the means for monitoring commitments under changing circumstances. He goes on to suggest that all coordination mechanisms can be seen as joint commitments and their related social monitoring conventions. Jennings acknowledges that it is infeasible in any community of reasonable complexity for total monitoring to occur, due to communication bandwidth and processing time. In our own human experience we see this to be true in teamwork, where some awareness of other team-member activity is naturally maintained, but any attempt at totally detailed and continuous monitoring knowledge is impossibly overloading and counterproductive.

Important to our monitoring interests, Jennings shows why the behavior (alone) of a collection of agents as seen by an outside observer is insufficient to determine if coordination is present, and concludes that "*coordination is best studied by examining the mental state of the individual agents.*" He then goes on to say: "*The exact make up of this mental state is still the subject of much debate, however there is an emerging consensus on the fact that it contains beliefs, desires, goals and commitments (intentions).*"

Jennings raised issues that are addressed in Gal Kaminka's Ph.D. thesis (2000) and related publications (Kaminka and Tambe 1997, 1998, 2000). Kaminka pursued what Jennings dubbed the *social conventions* aspect and developed a "mental state" representation based on goal hierarchies presumably shared by a team of agents—recognizing that some agents may have tasks different than others and some may choose to achieve a common task differently than others. Notably his work features primary examples of unmanned autonomous (aerial) systems, where individual UAS (agents) monitor and recognize when a member of the team doesn't behave as mutually agreed to in the plan. Kaminka's approach also enables an agent to detect self-failure often but not always.

Kaminka credits inspiration to Leon Festinger's seminal work on *social comparison theory* (Festinger 1954), which is founded on the hypothesis that humans have a drive to evaluate their own opinions and abilities and will employ a social comparison process if this evaluation cannot be tested directly.

Kaminka (1997) kicks off this path of work by proposing an approach to failure detection which he called unique to a multi-agent setting: the key idea being that agents observe each other and use that information to inform themselves about the situation and about the appropriateness of the behaviors of self and others. Basically each agent evaluates its own behavior by observing that of others, and comparing those observations with its own behavior, beliefs, goals, and plans.

Kaminka's early tack had Festinger's self-centered focus: agents used cues from others to evaluate their own fitness. Subsequently, his investigations broadened to both a centralized agent that could monitor team and other-agent behavior, and multiple agents monitoring team and other-agent behaviors.

Kaminka makes the case in his thesis for distributed monitoring and detection, showing that a centralized monitor using his algorithms does as well as can be done, whereas multiple monitor/detectors among socially aware agents do best, as they can exploit their own local state as part of the information. He shows that the centralized approach can provide either sound (no false positives) or complete (no false negatives) results, whereas the decentralized approach provides both sound and complete results—meaning no incorrect detections and no missed detections. He also shows that this can be accomplished without any one agent monitoring all the agents, and without all the agents having this monitoring capability.

Recent work is getting even closer to the detection of threatening aberrant behavior. Avrahami-Zilberbrand and Kaminka (2007) extend the social comparison concept into the detection of suspicious behavior by an agent. The general approach is to monitor a large group of agents and note that one or some agents are behaving decidedly different than expected. Two types of suspicious behavior recognition are employed: explicit and implicit. Explicit recognition classifies behavior as suspicious if it reflects a reference pattern known to be suspicious. Implicit recognition classifies behavior as suspicious if it does not conform to cataloged reference patterns of "normal" behavior. This work is part of a more general interest in dynamic tracking of multi-agent teams.

Sviatoslav Braynov (Braynov 2004, Braynov and Jadliwala 2004) has investigated the use of coordination graphs built from filtered action data to recognize coordinated behaviors among multiple agents maliciously working toward an undesirable goal. This is done by an aberrant behavior detector examining forensic data, with suggestions that real-time log-data examination might recognize a coordinated attack in early stages of setup and initiate counteraction. This approach may be useful for identifying the agents, actions, and situational conditions that participate in the manifestation of an emergent behavior. Proactively, such emergent behaviors that are determined to be undesirable could thereafter become recognizable patterns that generate increasing states of concern as manifestation of the conditions increases.

In summary, the cited works in this article bring the concepts of social awareness into play with good effect for detecting behaviors not in keeping with team goals, agent tasks, and coordination plans. Separate paths by Kaminka and Braynov are beginning, respectively, to attack computational scaling issues and the detection of coordinated alien activity within groups.

Spatial behavior leverage

Stephan Intille (1999) opened an interesting path that explored *visual recognition of multi-agent action*. His work analyzed films of American football games and identified the plays being made according to visual analysis of the trajectories of the players, matching the offense player trajectory's against the team's playbook patterns. There is a considerable difference between idealized chalk-board play patterns *(Figure 3)* and actual game-time trajectories given the unpredictability of the 11 defensive-team players, as well as the infinite variety of trajectories the 11 offensive-team players may take in reaction to defensive play. Yet he built a system that could recognize appropriate single-agent and multi-agent actions in this domain under "noisy" trajectory data of player and ball movements.

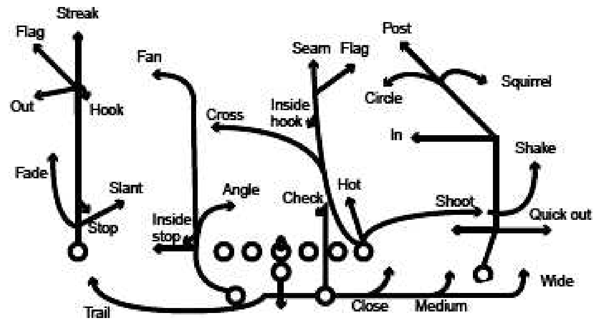


Figure 3. The different types of pass patterns that receivers can run constrained by the rules and nature of the game. (Reproduced with permission, from Intille 1999.)

"Recognizing Planned, Multiperson Action" (Intille 2001) is a mature digestible presentation of his work covering approximately 6 years. He has focused on plan recognition, attempting to identify the play by classifying the observed actions, movements, and spatial relationships of the players. Intille notes certain aspects of American football and the nature of its team interaction that shape the recognition approach:

- Infrequent collaborative replanning—though agents adjust their playing to fit the real-time situation, the intended play coordinates the general action.
- Agent based goal description—each agent has a goal (e.g., catch pass, block, receive handoff, etc.) for any given play. The system attempts to identify the goals the agents are pursuing based on spatial and temporal relationships of agents and their trajectories.
- Existing taxonomies—a common, fairly universal terminology exists among coaches and fans for describing all low-level agent actions (e.g., blocking, running through a hole) and higher level aggregated actions (e.g., executing a specific play)—within the boundaries of game constraints and experience. This forms a succinct and closed domain language. *Figure 3* shows some of that common terminology and the nature of its reference.
- Purposeful behavior—every agent is expected to contribute to the play's objective, nothing happens without a reason.
- No statistical databases—large statistical databases describing much of what has transpired in football action do not exist. A recognition algorithm cannot be based on that type of resource. Instead, a linguistic description of plays is provided by a domain expert (coach).

- Structured but uncertain—each offensive play begins as a highly structured coordination plan. The defensive agents rarely cooperate, so a great deal of variation exists in individual agent movements, individual agent goal achievement, and overall trajectory maps.

This model has potential for describing joint UAS maneuver patterns and detecting when an agent is not contributing as planned. Though a great deal of latitude is expected in the execution of a maneuver pattern, general characteristics should prevail and indicate an individual UAS not working on team behalf. Judgment of cause and severity for out-of-scope behavior is a separate issue not dealt with here.

Concluding remarks

This article reported on the first part of a two-part study, instigated by a technology for massively parallel pattern recognition and studies in self-organizing systems-of-systems. This first part identified a research base for aberrant behavior detection in multi-agent systems that might benefit from a massively parallel pattern recognition capability. The second part of the study, to be published shortly, will indicate how the detection approaches outlined in this article might be implemented advantageously in a massively parallel classification technology.

The temporal and spatial discussions in prior sections are complementary, each having a potential role in a socially attentive solution platform. It is not suggested that the research referenced here is the only way to approach the problem effectively, or that what was presented is completely sufficient, but rather that this basis appears promising as a foundation for a solution path worth exploring.

It is likely that the future of UAS is pervasive employment in human society, regardless of purpose, warfighting or otherwise. Such "things" will need to be socialized, as do the children of all species. Simple behavior safeguards will not be sufficient. Right or wrong, ready or not, we will expect these things to exhibit respect for life and property, ethics, self-control, and peer-policing capabilities approaching our own. To the extent that they don't, we will object to their presence.

In a test environment, especially in early years, as well as later with the presence of legacy units, such detection mechanisms are not likely to be present on board. The physical location of these mechanisms is not necessarily important, provided suitable sensor data are available. Under testing conditions, the testing arena is likely bounded and populated with various observer installations and mobile facilities. Suitable sensors located in these facilities, and perhaps sent from transmitters on board UAS, can provide the raw data feeds. As was shown by Kaminka (2000), it is not necessary to have a one-to-one ratio of monitors to agents in order to ensure high detection accuracy. Thus, multiple such mechanisms might be located in testing and observation facilities in suitable proximity to the testing arena. Alternatively, special units could be deployed among the UAS under test much as many field sports employ referees on the field.

In the end, such mechanisms also belong on board as an integral part of every UAS, as UAS will operate outside of ready observation and are subject to attrition by enemy destruction. In such live cases, aberrant behavior must be detected and evaluated to sense control penetration by the enemy as well as malfunction that threatens the mission or might provide a disabled UAS to the enemy for later recovery.

This concept of a socially aware team of autonomous agents has application well outside the UAS and UAST focus of this discussion. For instance, socially aware security agents can be employed as a community watch among networked groups of computers or sensors, keeping watch on each other. For another instance, Braynov (2004) is investigating ways in which coordination graphs can be employed in the recognition of coordinated attacks by groups of autonomous agents working toward a common goal. The platform suggested here for UAST can merge with Braynov's work to pursue application in intrusion detection areas. ❑

Rick Dove is a founding partner of Kennen Technologies, an adjunct professor in the School of Systems and Enterprises at Stevens Institute of Technology, and Chairman of Paradigm Shift International. He is author of Response Ability—The Language, Structure, and Culture of the Agile Enterprise*; and* Value Propositioning—Perception and Misperception in Decision Making. *He is an active researcher and consultant in agile and self-organizing systems. He has 40 years experience at start-up, turn-around, and interim executive management. He holds a bachelor's degree in electrical engineering from Carnegie Mellon University and has done graduate work in Computer Science at the University of California, Berkeley. E-mail: rick.dove@kennentech.com*

References

Arkin, Ronald C. 2007. Governing lethal behavior: Embedding ethics in a hybrid deliberative/reactive robot architecture. Technical Report GIT-GVU-07-

11. Mobile Robot Laboratory, College of Computing, Georgia Institute of Technology, Atlanta, Georgia. http://www.cc.gatech.edu/ai/robot-lab/online-publications/formalizationv35.pdf (accessed May 15, 2009).

Avrahami-Zilberbrand, Dorit, and Gal A. Kaminka. 2007. Towards dynamic tracking of multi-agent teams: An initial report. In *Proceedings of the AAAI Workshop on Plan, Activity, and Intent Recognition*, July 23, Vancouver, Canada, 17–22. Menlo Park, California: AAAI Press. www.cs.biu.ac.il/~galk/Publications/Papers/pair07.pdf (accessed May 15, 2009).

Braynov, Sviatoslav. 2004. A model for detecting malicious groups. *3rd European Conference on Information Warfare and Security*, June 28–29, London, UK, 21–30. Reading, UK: Academic Conferences International.

Braynov, Sviatoslav, and Murtuza Jadliwala. 2004. Detecting malicious groups of agents. In *Proceedings the First IEEE Symposium on Multi-Agent Security and Survivability*, August 29–31, Philadelphia, Pennsylvania, 90–99. Piscataway, New Jersey: IEEE.

Dove, Rick. 2007. *Self-organizing agile systems and enterprises: Architectural patterns enabling self-organizing systems-of-systems* (course binder for SDOE-683). Hoboken, New Jersey: Stevens Institute of Technology, School of Systems and Enterprises.

Dove, Rick. 2009. Pattern recognition without tradeoffs: Low-cost scalable accuracy independent of speed. In *Proceedings Cybersecurity Applications and Technologies Conference for Homeland Security (CATCH)*, March 3–4, Washington, D.C., 255–260. Piscataway, New Jersey: IEEE. www.kennentech.com/Pubs/2009-PatternRecognitionWithoutTradeoffs.pdf (accessed May 15, 2009).

Dove, Rick, and Garry Turkington. 2009. On how agile systems gracefully migrate across next-generation life cycle boundaries. *Journal of Flexible Systems Management* 10 (1): 17–26.

Festinger, Leon. 1954. A theory of social comparison processes. *Journal of Human Relations* 7 (2): 117–140.

Flack, Jessica C., Michelle Girvan, Frans B. M. de Waal, and David C. Krakauer. 2006. Policing stabilizes construction of social niches in primates. *Nature* 439 (7075): 426–429.

Heinze, Jürgen. 2003. Reproductive conflict in insect societies. In *Advances in the Study of Behavior*, ed. P. Slater and J. Rosenblatt 34: 1–57. New York: Academic Press.

Intille, Stephen S. 1999. Visual recognition of multi-agent action. Ph.D. dissertation, School of Architecture and Planning, Massachusetts Institute of Technology. http://web.media.mit.edu/~intille/papers-files/thesis.pdf (accessed May 15, 2009).

Intille, Stephen S. 2001. Recognizing planned, multiperson action. *Computer Vision and Image Understanding* 81 (3): 414–445. http://web.media.mit.edu/~intille/papers-files/cviu01.pdf (accessed May 15, 2009).

Jennings, Nick. 1993. Commitments and conventions: The foundation of coordination in multi-agent systems. *The Knowledge Engineering Review* 8 (3): 223–250. https://eprints.kfupm.edu.sa/30491/ (accessed May 15, 2009).

Kaminka, Gal A. 2000. Execution monitoring in multi-agent environments. Ph.D. dissertation, University of Southern California. www.isi.edu/soar/galk/Publications/diss-final.ps.gz (accessed May 15, 2009).

Kaminka, G. A., and M. Tambe. 1997. Towards social comparison for failure detection. In *Proceedings Fall Symposium on Socially Intelligent Agents*, November 8–10, Cambridge, Massachusetts, 64–66. Menlo Park, California: American Association for Artificial Intelligence.

Kaminka, G. A., and M. Tambe. 1998. What is wrong with us? Improving robustness through social diagnosis. In *Proceedings of the 15th National Conference on Artificial Intelligence*, July 26–30, Madison, Wisconsin, 97–104. Menlo Park, California: American Association for Artificial Intelligence.

Kaminka, G. A., and M. Tambe. 2000. Robust multi-agent teams via socially-attentive monitoring. *Journal of Artificial Intelligence Research* 12: 105–147. http://teamcore.usc.edu/papers/2000/jair-socmon.pdf (accessed May 15, 2009).

Kaminka, G. A., D. V. Pynadath, and M. Tambe. 2001. Monitoring deployed agent teams. In *Proceedings of the Fifth International Conference on Autonomous Agents*, May 28–June 1, Montreal, Canada, 308–315. New York: ACM. www.cs.biu.ac.il/~galk/Publications/Papers/agents01.pdf (accessed May 15, 2009).

Malone, Thomas, and Keven Crowston. 1994. The interdisciplinary study of coordination. *ACM Computing Surveys* 26 (1): 87–119. www.cs.unicam.it/merelli/Calcolo/malone.pdf (accessed May 15, 2009).

Monnin, T., F. L. W. Ratnieks, G. R. Jones, and R. Beard. 2002. Pretender punishment induced by chemical signaling in a queenless ant. *Nature* 419 (6902): 61–65.

Moshkina, Lilia, and Ronald C. Arkin. 2007. Lethality and autonomous systems: Survey design and results. Technical Report GIT-GVU-07-16. Mobile Robot Laboratory, College of Computing, Georgia Institute of Technology, Atlanta, Georgia. www.cc.gatech.edu/ai/robot-lab/online-publications/MoshkinaArkinTechReport2008.pdf (accessed May 15, 2009).

Office of the Secretary of Defense. 2008. *Unmanned and Autonomous System Testing (UAST) Broad Agency Announcement <BAAUAST0002>*. U.S. Department of Defense, Issue Date April 4, 2008.

Ratnieks, F. L., K. R. Foster, and T. Wenseleers. 2006. Conflict resolution in insect societies. *Annual Review of Entomology* 51: 581–608. www.people.fas.harvard.edu/~kfoster/RatnieksetalAnnreventomol2006.pdf (accessed May 15, 2009).

Spencer, Herbert. 1893. Sub-human justice. In *The Principles of Ethics*, Volume 2, 8–16. New York: D. Appleton and Company.

Acknowledgments

This work was supported in part by the U.S. Department of Homeland Security award NBCHC070016. *Figures 1 and 2* are included with permission from Ronald Arkin. *Figure 3* is included with permission from Stephen Intille.

ITEA Journal 2009; 30: 409–416

Challenges in Software Safety for Army Test and Evaluation

Frank Fratrik
U.S. Army Aberdeen Test Center,
Aberdeen Proving Ground, Maryland

As the capabilities of software intensive systems grow so does the complexity of functions controlled via software. Similarly, software test and evaluation (T&E) efforts have become increasingly difficult to quantify and scope appropriately. T&E efforts for traditional programs have hinged on system level testing in realistic or simulated environments to verify and validate the systems. Application of these traditional methods to software intensive systems continues to hold value, but it no longer provides exhaustive data. A number of T&E deficiencies are surfacing in test programs for these software intensive systems as controllability and visibility related to software functionality decreases. The result is testing that can fail to uncover critical problems, potentially with catastrophic results. This article describes successes and shortcomings with current test and analysis methodologies for software intensive systems. As a part of the Army Test and Evaluation Command (ATEC), the author looks from an unbiased viewpoint at relevant current practices and the outlook for future T&E in regards to software safety. Recent examples of Army software test and analysis efforts, current Army T&E guidance for software safety, and a path forward for increasing confidence in software safety will be discussed.

Key words: Residual risk; risk; software intensive systems; software safety; T&E current practices; test and analysis methodologies.

Developmental test (DT) programs for many military systems over the past several decades have successfully verified safety to a high degree of confidence. However, with the rise of complexity and criticality allotted to software control, software test methodologies must be expanded to achieve the same high confidence required by those responsible for system safety.

The goal of this discussion is to increase awareness of software safety and identify practices and methods, based on existing guidance, which will reduce hazards associated with military systems that use software to achieve system safety objectives. The target audience includes developers, program managers (PMs), acquirers, and others involved in test and evaluation (T&E) as it relates to system safety.

Because of the data gaps when testing software in a completed system, analysis of both safety engineering and software engineering efforts must be accepted as data points in building a case for software safety. Only after earlier development efforts have been assessed can the right system level DT scope and environments be identified.

Current Department of Defense (DoD) software safety guidance includes a series of analyses in combination with testing to provide the best confidence for systems with safety critical software. Many current programs still lack the suggested development processes and rigor. In recent years this has led to systems with safety critical software displaying incomplete hazard analysis, lack of requirements traceability, poor design, and insufficient testing. Deficiencies in these key analysis areas resulted in only a partial data set being available for safety decision makers.

To fill these data gaps several methods are considered. Test results ranging from unit level up to system level will be assessed for their value in software safety. Relevant analyses, including hazard analysis, requirements analysis, architectural and detailed design analysis, and software code analysis, used to compliment test data will also be discussed.

DoD guidance for software safety

> Safety Critical *(from MIL-STD 882C). A term applied to a condition, event, operation, process,*

or item of whose proper recognition, control, performance or tolerance is essential to safe system operation or use, e.g., safety critical function, safety critical path, safety critical component.

As previously stated, the goal in this discussion is not to introduce novel concepts relating to safety critical software. Instead, shortcomings with adherence to existing guidance are highlighted and a path forward to concurrence with those guidelines is identified. Multiple current resources exist for obtaining software safety guidance in a DoD context; however, little is regulatory. In 1999, the Joint Software System Safety Committee (JSSSC) Software System Safety Handbook (JSSSC, 1999) was released, which provides significant content in regards to identification, development, and verification of safety critical software. This is the most comprehensive single source for DoD software safety guidance.

Multiple other sources are also available, including an International Test Operating Procedure (ITOP) for Safety Critical Software Analysis and Testing (U.S. Army, 1999) the U.S. Army Communications and Electronics Command (CECOM) Software System Safety Guide (U.S. Army, 1992), and the U.S. Naval Sea Systems Command (NAVSEA) Weapon System Safety Guidelines Handbook (NAVSEA, 2006). A draft Software System Safety Policy is also in circulation for the U.S. Army Aviation and Missiles Command (AMCOM) (AMCOM, 2008).

These are by no means the only sources of guidance. It is critical to be aware that guidance for software safety exists and to find the most relevant information that applies to your system and its goals.

Software testing versus hardware testing

For those who have been in the T&E domain for a number of years, it may seem unclear why software information cannot be obtained as part of a traditional system level testing for verification of safety. Software testing limitations are characterized by the JSSSH E.13.1, General Testing Guidelines as follows:

"Systematic and thorough testing is clearly required as evidence for critical software assurance; however, testing is 'necessary but not sufficient.' Testing is the chief way that evidence is provided about the actual behavior of the software produced, but the evidence it provides is always incomplete since testing for non-trivial systems is always a sampling of input states and not an exhaustive exercise of all possible system states." (JSSSC, 1999)

It is because of these limitations that a gap exists when testing systems with safety critical computer software components.

Safety Critical Computer Software Components *(from MIL-STD 882C). Those computer software components and units whose errors can result in a potential hazard, or loss of predictability or control of a system.*

Generally, traditional system level DT is able to characterize the system down to the smallest hardware components. Environments can be simulated for whole systems or parts of systems. Failures can be isolated and characterized by looking at failure points. As an example, system level vibration testing will submit all hardware elements to the test environment. Elements such as vehicle chassis, weapon mounts, engine components, and bolts are all stressed in this test. Failure points can be visually inspected.

On the contrary, software is not intended to be characterized by exposure to traditional system level testing. Complex software will commonly use only a subset of the total software code to achieve common mission functionality. This leaves rarely used (but possibly safety critical) code untested. Specific software functions can be difficult to stimulate or difficult to observe for correct output. Consider the example of a built-in-test (BIT) function achieved via software. The BIT may require internal subsystem failures to occur to exercise software code responsible for detection and reaction. A system level test might require causing damage to some internal subsystems or it may not be possible even via destructive testing due to constraints with observing outputs in a black box test.

Safety critical computer software component identification

If additional methods are to be applied to gain confidence in software safety, it is wise to focus the increased scope on safety critical computer software components. In order to do this, we must identify what software in a given system has a safety impact. Software safety must be considered a subset of system safety and flow system level hazards down to subsystems that are relevant. System level hazards should be allocated to hardware, software, or some combination of both.

Software can play a role in safety in many systems where it is a cause of a hazard. Examples could be causing hardware to perform unsafe actions or guiding an operator to make unsafe decisions. Software can also be a control of a hazard. It may be used to prevent hazards or limit severity after a mishap occurs. In either case, the software role in system safety must be defined based on top level hazards related to the entire system.

Software by itself does not impact safety; however, when coupled with critical hardware the software can become safety critical. Software that performs functions like these is likely to be safety critical:

- arm, enable, release, launch, fire, or detonate a weapon system;
- control movement of gun mounts, launchers, and other equipment;
- control movement of munitions or hazardous materials;
- monitor the state of the system for purposes of ensuring its safety;
- sense hazards and/or display information concerning the protection of the system;
- control or regulate energy sources in the system;
- exercise autonomous control over safety critical hardware;
- generate outputs that display the status of safety critical hardware systems;
- compute safety critical data.

These are merely a sample of potential examples derived from the JSSSH, section E.1.1.3 (JSSSC, 1999). For every system a detailed analysis of the software role in system safety must be conducted.

Increasing confidence in safety critical software

Moving forward requires the following two assumptions:

1. Correct safety critical software functionality is required for safe system operation.
2. Software is typically not sufficiently characterized by system-level testing alone.

To bridge the implied gap resulting from these assumptions, the recommendation is a disciplined approach to system, safety, and software engineering that manages software's role in system safety.

This approach results in evidence that safety is a prominent part of specifying, designing, building, and testing safety critical software. Analysis of this evidence can in itself result in increased confidence in the safety of associated software. Also, this analysis can be used to better scope and execute system level safety testing. JSSSH guidance covers a broad spectrum of analyses and artifacts; however, the minimum set of information required for the software safety analyses in this discussion includes these elements:

1. Hazard Tracking System (HTS) data,
2. software requirements,
3. software design,
4. software problem reports (SPRs),
5. software test plans and results,
6. safety assessment report (SAR).

Software safety analyses

The analyses described in this section are not exhaustive but are considered in this discussion to be the *minimum* required to support safety decision makers. In all cases, the number and rigor of software safety analysis methods should be proportional to the software hazard criticality, as defined in the JSSSC Handbook (JSSSC, 1999) and MIL-STD 882C (DoD, 1993). In some cases, the analyses will be minimal and aim to prove that software is not safety critical. For each area, there are typically multiple different methodologies that can be applied. They are ideally to be conducted by a safety agency independent of developers and acquirers, such as the Developmental Test Command (DTC) in the Army domain. The safety agent does not have the primary responsibility for developing these artifacts but is instead a consumer who uses and assesses them. Nominal responsible developing entities for Army systems are listed for each artifact; however, this may vary between organizations or programs.

Generally, these six elements flow chronologically in their development, but all will typically mature and develop iteratively. The value of any one of these elements is limited in and of itself. The largest benefits can be realized if all six are analyzed and compared in the context of each other.

HTS data analysis

Data from a HTS, sometimes called a Hazard Tracking Database, is typically a PM responsibility in an Army setting. Often generation of this artifact is delegated to a developer. This is considered a starting point for software safety analysis, as it is a primary source for identification of system level hazards and subsystems that could contribute to or prevent mishap occurrence. From a software safety perspective it is key that hazards allocated to software are clearly identified. Chapter 12 in Section II of the NAVSEA Weapon System Safety Guidelines Handbook is a helpful resource for planning, implementing, and updating an HTS (NAVSEA, 2006).

Software requirements analysis

Software requirements are normally a developer responsibility. Software requirements must be derived to specify the software role in both system functionality and system safety.

In addition to general requirements best practices for all requirements, safety requirements should be identified explicitly. Tracing up to system requirements and system hazards provides validity. Tracing down to software test cases, and possibly software design elements, is needed to allow for verification. All requirements need to be written in a way that they are testable.

Software requirements analysis can identify any gaps in traceability or other shortcomings that could impact the software artifacts to be derived from the requirements specification. More detailed expectations for derivation and tracing of safety critical software requirements is detailed in the JSSSH, section 4.3.4 Derive System Safety Critical Software Requirements and section 4.3.5.3 Traceability Analysis (JSSSC, 1999).

Software design analysis

Software design is typically a developer responsibility and normally flows out of software requirements. This can include both architectural design and detailed design. Software design best practices, such as maximizing cohesion and minimizing coupling, translate well to the safety domain. Isolation of software with safety impact in design can decrease the risk of a safety impact from other outside software modules.

Identification of critical software modules will alert software maintainers to potential impacts of changing software code in a safety critical area. Another benefit is that the impact of changes can be assessed using design information. This aids in identifying proper regression test scoping.

There are many design elements that can aid in coding with safety in mind, such as using interlocks, safety flags, watchdogs, or other techniques. Each of these can be considered during software design analysis. More detailed expectations for design of safety critical software can be found in the JSSSH, section 4.3.6 Detailed Software Design, Subsystem Hazard Analysis (JSSSC, 1999).

Developer SPR analysis

SPR generation is typically a developer responsibility and begins as early as practical after software implementation has begun. Formal SPR development ensures that all discovered problems are addressed and tracked to closure. SPRs that are not closed in a given software version can then be compiled and assessed for their impact.

It is important that SPRs be categorized for criticality. Those with safety impact should be made explicit and given highest priority for fix implementation. SPRs intended for closing should show traceability to test cases for verification of a successful fix. SPR tracking as a function of time can also be a useful metric for considering software maturity or completeness of test. SPR tracking is supported by the JSSSH, section E.13.1 General Testing Guidelines (JSSSC, 1999).

Developer software test planning and execution

SPRs are typically generated as a result of testing, which is the next focus area. To be clear, this is developer-run software focused testing, scoped to fully verify software requirements. Often this is called software formal qualification test (FQT). This may or may not use the full, complete system. Many times parts of the system are simulated or decomposed to increase the ability to stimulate inputs or observe outputs of the software.

This is the first activity where it is highly desirable to have a participant attend as a representative of the independent safety agency. The results need to be considered a primary data source for software safety. On-site involvement increases confidence when leaning heavily on a developer-run event.

Since this is a software requirements–based test, traceability from requirements should be evident. Specifically, identification of test cases aimed to verify safety requirements must be identified in test plan and test description documents. More rigorous testing should be executed on safety elements to ensure that off nominal or abnormal execution cannot result in a hazard. Examples could be fault injection, boundary condition testing, zero value testing, or input rate variation.

Software testing in this environment should measure the amount of software code that is exercised by testing, and justify any shortcomings, especially for untested safety critical code. More detailed expectations for developer software test planning and execution can be found in the JSSSH, section 4.4.1 Software Safety Test Planning and section E.13.3 Formal Test Coverage (JSSSC, 1999).

SAR analysis

The SAR is typically a PM responsibility in an Army setting. It can be considered a snapshot in time of the current system and its safety implications. A current description of system hazard analyses is mapped against the intended use, which could be a test event, demonstration, or field use. Required mitigations should be evident for those who will control the environment where the system is to be used.

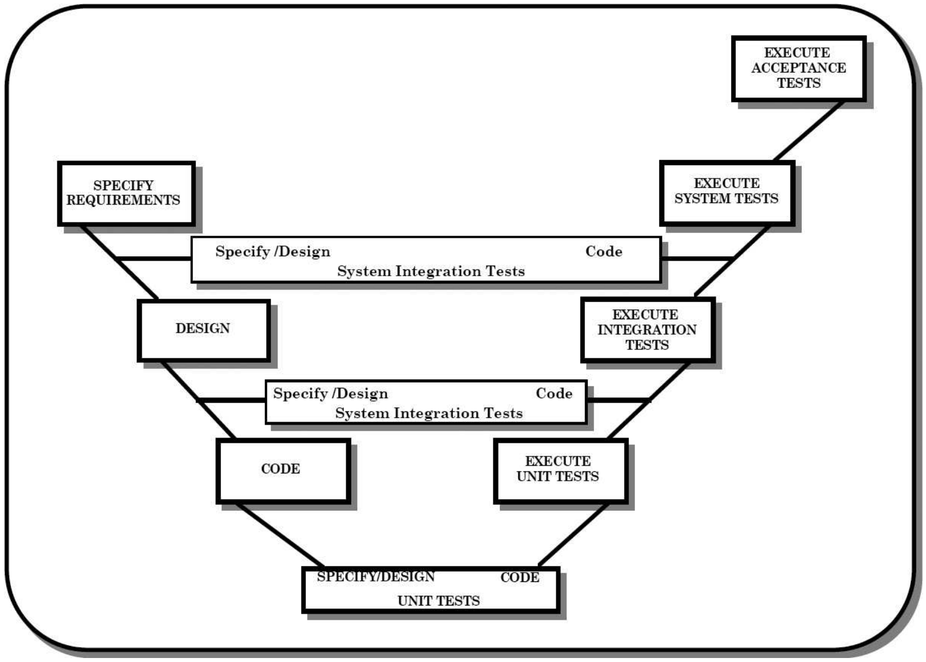


Figure 1. Levels of software testing (from Joint Software System Safety Committee (JSSSC) Software System Safety Handbook, section 2.6.4.2.2).

For software, there should be a description of the software safety approach, to include design constraints, coding standards, or other applicable software safety methodology. There must be a description of the safety characteristics of the specific hardware and software configuration intended for use. Open and residual hazards with causes allocated to software should be identified. Chapter 27 in Section II of the NAVSEA Weapon System Safety Guidelines Handbook is a potential resource when generating and updating a SAR (NAVSEA, 2006).

Software safety analyses summary

Without delivery of the artifacts in this section, the resulting analyses cannot be completed. The result of this information gap is that hazard contributions allocated to software should be considered to have elevated probability.

A second key point is that this information must be developed and delivered in order to lower software-related hazard probabilities even for “off-the-shelf” or prototype systems. The choice to buy products to perform safety critical functions or resource limitations does not preclude them from being assessed for their correct and safe performance.

Last, this is an iterative set of analyses. Any resulting outputs apply only to the configuration and environment for which the analyses were completed. As changes are made, each element should be revisited and considered for its impact.

Levels of software safety testing

Traditionally, upon delivery of the SAR, independent system level DT can begin. Verification of safety requirements should occur at the highest level of system integration possible; however, some safety critical software requirements will not be verifiable at the system test level.

As discussed previously, software FQTs are often a primary source for verification of software. Even at this FQT level, it will sometimes not be possible to fully verify all requirements of the software.

It is less desirable, but in some cases software integration tests and software unit tests can be used as data sources to achieve complete requirements or SPR fix verification. In this lower-level testing it is critical to identify assumptions and limitations of the test environment.

Although not preferred, static software code inspection can be used as a verification method for requirements not able to be verified via dynamic testing. Normally this method is not used as a primary verification source for a safety requirement but can be a good secondary source for additional verification confidence. This flow and scope of sources for safety testing is shown in *Figure 1*.

In addition to full verification of software requirements during developer testing, independent DT should include system level software-focused safety testing. This testing should be designed to compliment developer testing and address capabilities as well as

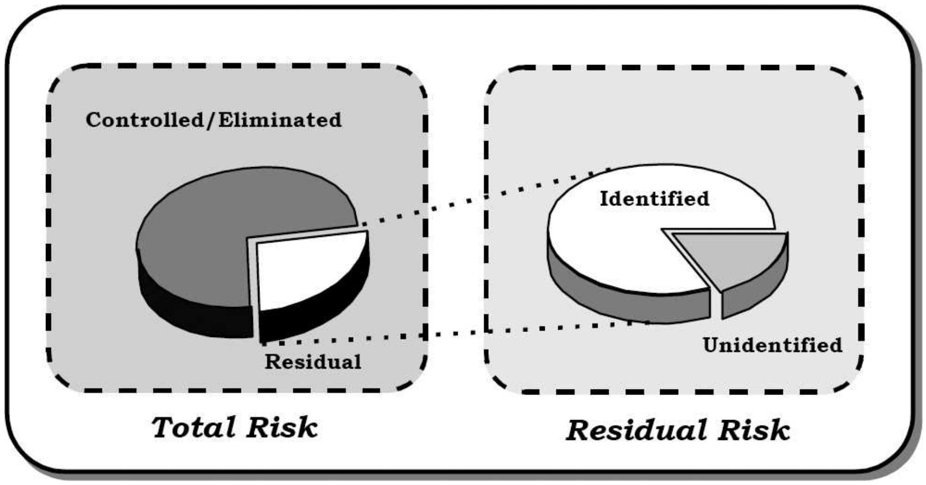


Figure 2. Total risk versus residual risk (from Joint Software System Safety Committee (JSSSC) Software System Safety Handbook, section 3.3).

requirements. System level software test design can focus time spent testing on safety-related software functionality. Training is key to allowing independent testers to identify expected test results. At this system level, exploratory testing is encouraged—often the defects not found during requirements-based testing are found during exploratory testing.

It is imperative to be aware that the primary method for showing safe software is verification of safety critical software requirements via test.

Residual risk identification

At the completion of Army DT, there is often a need to identify residual risks prior to soldier interaction with the system. This is done through DTC-issued safety releases and safety confirmations. Following analyses and test completion, residual risk can be derived from the total risk subset, as visualized in *Figure 2*.

For systems with safety critical computer software components, it is essential that any residual risk documents are limited to well-defined, tested configurations. Residual risk outputs should trace observed test results and safety analysis to identify residual system hazards. If the software is modified, the residual risk outputs must be updated or amended. Having requirements and design analyses completed will help properly scope any required regression testing without excessive budget or schedule impact.

A key difference in assessing residual risks relating to software is awareness of expected reliability. Software reliability considers errors resulting from specification, design, or implementation that are unfound during testing. Historically, software can be shown to meet a failure rate of no better than the 10^{-3} to 10^{-4} range according to the American National Standards Institute (ANSI) R-013-1992 and T/AST/046 (Nuclear Safety Directorate [UK], 2003). Goals for system safety often include showing 10^{-6} or better residual probability of mishap occurrence. Because of this, hazard mitigation relying on software should include other mitigation methods in hardware and/or procedural use to reduce probabilities to an acceptable level. Even with a most rigorous software development and test effort, the lower ends of mishap probability may be unreachable via software alone. Systems relying on software as the sole mitigation to a hazard should consider the residual risk if probability falls in the 10^{-3} to 10^{-4} range, which is labeled *remote mishap probability* in MIL-STD 882 (DoD, 1993). Initial design efforts can greatly reduce the amount and impact of safety critical software.

Application

The following are recent examples that show limitations associated with relying solely on system level testing to achieve confidence in system safety. Also, expected results of applying the existing DoD software safety guidance detailed in the previous sections are described.

Case 1—weaponized Unmanned Ground Vehicle (UGV)

In this example, a small, weaponized UGV system entered DT with weak system hazard analysis. Also, it had no hazard allocation to software, software requirement specification, software design documentation, or software requirements–based testing.

Actual results during DT included occurrence of multiple uncommanded motion events that were not prevented or halted by system software. Also, system-level software safety testing discovered that the system was capable of weapon fire via a switch other than the trigger.

The result via proper application of existing software safety guidance would have been earlier identification of elevated probability for hazards during test and field use associated with both uncommanded platform motion and uncommanded weapon firing.

Case 2—large, fast UGV

In this example, a 5,000-lb UGV, capable of tele-operation up to 50 mph, entered government DT with no hazard analysis, software requirement specification, software design documentation, or software requirements–based testing.

Actual results included system-level software safety testing, discovering that remote vehicle power loss left actuators in their previous state. The system was driving when power loss was induced. The result was continued UGV motion at the original speed with no method to emergency stop the UGV.

The result via proper application of existing software safety guidance would have been earlier identification of elevated probability for hazard during DT and field use associated with uncommanded motion.

Case 3—remote fire control system

In this example, a remote fire control system entered DT with incomplete system hazard analysis. Also software requirement specification, software design documentation, and evidence of software requirements–based testing were not delivered.

Safety findings were initially prepared solely using system level test results. Software safety artifacts were requested from the developer and delivered for analysis. Analytical results showed five additional system hazards, three of which had catastrophic severity, to have residual risk.

These risks did not manifest during testing and were not apparent without consideration of analytical outputs. Proper residual risk identification then allowed mitigations to be put in place to address the risks. Following iterations were able to further reduce the risks via design changes.

Application summary

Each of these systems required significant external mitigations to allow safe system testing. In all cases, significant safety issues were discovered late in the development cycle and resulted in costly rework to correct the problems. Fortunately, the problems in the first two cases were discovered through exploratory testing; however, consistency of problem discovery would increase significantly assuming a structured integration and test series leading up to system level DT. Each of the UGV systems had only system level test results as available inputs for safety decision makers, resulting in higher probabilities of residual hazard occurrence.

True value can be seen in case three, where hazards were identified analytically, mitigated via design when feasible, and accurately elevated for acceptance when needed.

Benefits to system safety

The primary safety benefits of consistently applying existing software safety guidance would be seen in several ways. Test safety would be increased by raising awareness of potential hazards during test. Those executing DT would not be exposed to unknown hazards.

Another critical benefit is that safety releases and safety confirmations could most accurately present residual hazards allocated to software. Appropriate mitigation techniques could be identified to safely maximize system functionality during testing, demonstrations, or field use.

In addition, the disciplines identified for generating software safety outputs are in line with current best practices for software engineering identified in the Institute of Electrical and Electronics Engineers (IEEE) 12207 "Standard for Information Technology–Software Life Cycle Processes" (IEEE, 1998) or the Software Engineering Institute (SEI) Capability Maturity Model® Integration (CMMI) (Carnegie Mellon, 2006). The added benefit would be promotion of software quality, software maintainability, software testability, and the discovery of software problems earlier in the life cycle when they are cheaper to fix.

Conclusions

System functionality executed by complex software has been increasing and will continue to grow in the future. System hazards related to software will be more important to quantify, while at the same time they cannot be exhaustively characterized through traditional system level safety testing. Those involved in T&E must be aware that DoD guidance is in place for software safety. Only through awareness and application of these practices to software specification, design, implementation, and testing can we ensure safe and appropriate tests resulting in credible residual hazard identification. ❑

FRANK FRATRIK is an electrical engineer at the US Army Aberdeen Test Center in Aberdeen Proving Ground, MD.

He has been performing test and analysis on software intensive systems for the past 5 years. Fratrik has been the system software test lead for unmanned ground vehicle systems with varied size and capability. His test work has also included software controlled systems such as: landmines, artillery fuzes, fire control systems, and other ground vehicle systems. Fratrik holds a B.S. in Electrical Engineering from Pennsylvania State University, University Park, PA and is also nearing completion of a M.S. in Systems Engineering from Johns Hopkins University, Baltimore, MD.

References

American National Standards Institute (ANSI). 1992. *ANSI R-013-1992, Software Reliability, 1992*. http://webstore.ansi.org/RecordDetail.aspx?sku=R-013-1992 (March 5, 2009).

Carnegie Mellon. 2006. *Capability Maturity Model (CMMI) CMMI® for Development, Version 1.2*. Carnegie Mellon Software Engineering Institute, August 2006. Pittsburgh, Pennsylvania: Carnegie Mellon Software Engineering Institute.

Department of Defense (DoD). 1993. *MIL-STD 882C, System Safety Program Requirements, 19 January 1993*. Wright Patterson Air Force Base, Ohio: HQ Air Force Materiel Command (SES). http://crc.army.mil/guidance/system_safety/882C.pdf (accessed March 5, 2009).

IEEE. 1998. *IEEE 12207, Standard for Information Technology-Software Life Cycle Processes*. New York, New York: Institute of Electrical and Electronics Engineers.

Joint Software System Safety Committee (JSSSC). 1999. *Joint Software System Safety Committee (JSSSC) Software System Safety Handbook, December 1999*. Washington, D.C.: Joint Services Computer Resources Management Group, U.S. Navy, U.S. Army, and the U.S. Air Force. http://www.system-safety.org/Document/Software_System_Safety_Handbook.pdf (accessed March 5, 2009).

NAVSEA. 2006. *NAVSEA Weapon System Safety Guidelines Handbook, SW020-AH-SAF-010, 1 February 2006*. Indian Head, Maryland: Naval Sea Systems Command Warfare Center Enterprise, Naval Support Facility, Indian Head.

Nuclear Safety Directorate [UK]. 2003. *T/AST/046, Technical Assessment Guide: Computer Based Safety Systems, (United Kingdom) Nuclear Safety Directorate—Business Management System, 10 January 2003*. http://www.hse.gov.uk/foi/internalops/nsd/tech_asst_guides/tast046.pdf (accessed March 5, 2009).

U.S. Army. 1992. *CECOM Software System Safety Guide, TR 92-02, May 1992*. Washington, D.C.: U.S. Department of the Army Headquarters.

U.S. Army. 1999. *FR/GE/UK/US International Test Operations Procedure (ITOP) 1-1-057 Safety Critical Software Analysis and Testing, 4 June 1999*. Washington, D.C.: U.S. Department of the Army Headquarters.

U.S. Army Aviation and Missiles Command (AMCOM). 1998. Software System Safety Policy, AMCOM Reg 385-17. Draft, March 15, 2008.

ITEA Journal 2009; 30: 417–420

Tech Refresh—It's More Than Plug-N-Play

Mary-Ann M. Boyce
L-3 Communications,
Atlantic City International Airport, New Jersey
Derek Kubont
Federal Aviation Administration,
Atlantic City International Airport, New Jersey

A Technical Refresh—or Tech Refresh—is the migration of an existing system to a new platform. There are several scenarios such as the following that may drive the need for a system's tech refresh:

- *hardware and/or software that is at, or approaching the end of, its expected life;*
- *hardware that may no longer be manufactured;*
- *Commercial Off-The-Shelf (COTS) software that is no longer supported by the developer;*
- *current system software made obsolete by advances in technology or in other interrelated systems;*
- *changes in the way the end user utilizes data from the existing system.*

There are five major factors—or drivers—that cause a tech refresh. In this article, we will examine the roles of these drivers: (a) hardware, (b) software, (c) technological advances, (d) economics, and (e) the system end user. Any one of the drivers may initiate the tech refresh. During a tech refresh, all five drivers are represented.

Key words: Economics; hardware; software; system end user; technological advances.

The existing system serves a need. The replacement system must conform to the functionality provided by the existing system. This is where the role of the tester is exposed. In order to verify the functionality, it is necessary to return to the original system design, specifications, and test documents to locate and utilize the requirements verification matrix. The replacement system must be able to meet all the requirements and capabilities of the original system as well as correct any problems detected in the original system. This verification is accomplished through regression testing.

The tech refresh system may also provide new capabilities that are beyond the scope of the original system. This is referred to by the term "requirements creep." It will be necessary to develop a matrix of these new capabilities and then determine which ones will be included in the upgrade and which will remain dormant for the current tech refresh effort.

Finally, the tech refresh system must pass a stringent battery of testing to "shake out" any new problems encountered owing to the testing of new hardware and/or software.

The results of thorough testing will ensure the tech refresh is successful and has produced a system that is reliable, sustainable, and maintainable for the foreseeable future.

Typical system life cycle

A typical life cycle for today's hardware and software lasts from 3 to 5 years. The serviceability actually depends on the original hardware decisions in the main memory and system (clock) speed of the processor. A hardware system purchased with optimal clock speed and fully populated memory will tend to remain in service significantly longer than a system that when purchased only met (then) current software requirements. Whether looking at a hardware system or a software system, the System Life Cycle (*Figure 1*)

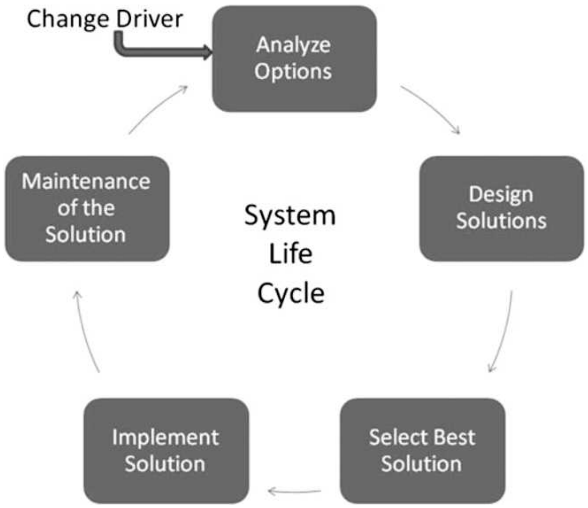


Figure 1. Typical system life cycle.

will follow a predictable path. It begins with input from one or more of the most common Change Drivers (*Figure 2*) and progresses through the phases shown.

The drivers

The drivers listed in *Figure 2* represent the most common causes for change; however, this is by no means a comprehensive list. Occasionally one driver will cause additional drivers to become "active."

Hardware

Hardware life cycles are most closely allied to the inherent capabilities (clock speed, memory capacity, and throughput); vendor support (maintenance contracts); equipment failures (Are failures due to poor engineering or manufacture?); and the availability of parts supplies necessary to repair failures or perform upgrades. If the existing hardware fails and vendor support or parts are not available, a new hardware platform (including the operating system) initiates change.

Software

Software life cycles are determined by the software developer. New versions of operating systems and applications are not synchronized to occur together, nor are they synchronized with new hardware released to the market. New software can be required (a) by an

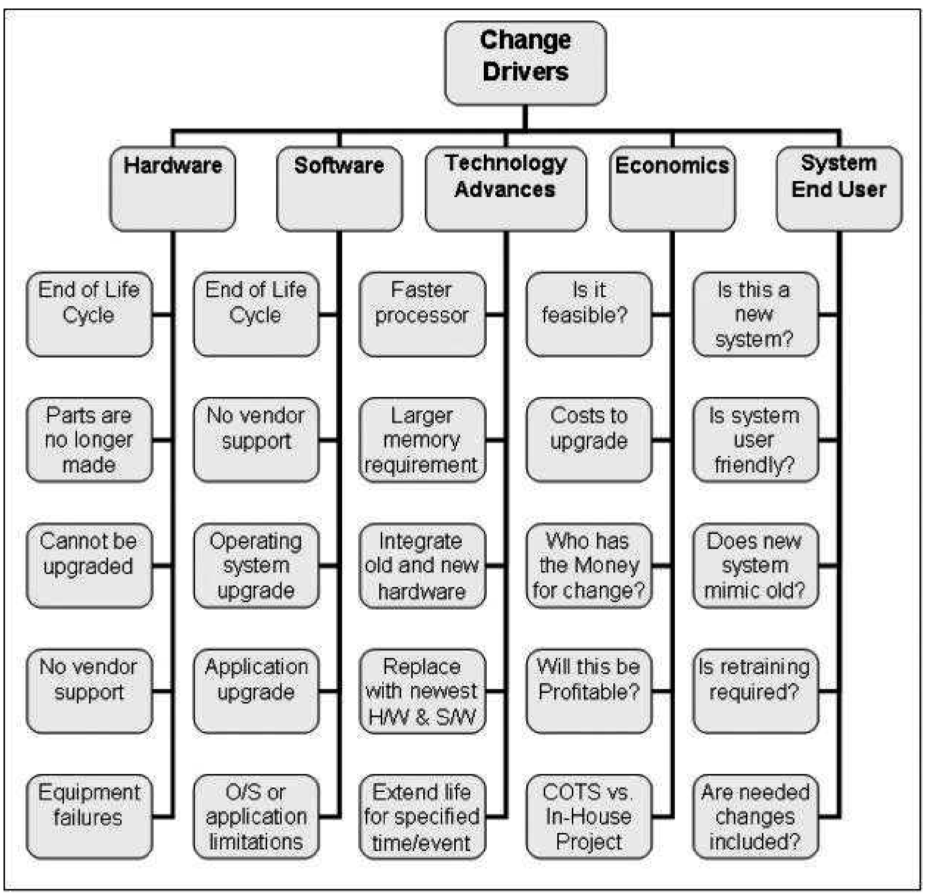


Figure 2. Most common change drivers.

operating system upgrade, (b) because the operating systems and/or Commercial Off-The-Shelf (COTS) applications currently in use have known limitations and vulnerabilities, and (c) the existing products are no longer supported by the vendor In the case of COTS versus homegrown (custom applications), the "homegrown" application has become more cumbersome and expensive to maintain than the currently marketed COTS product with a one-time purchase price.

Technological advances

Technological advances introduce new requirements including (a) integrating some new and more efficient hardware and/or software into the old system, (b) replacing all existing equipment with state-of-the-art commodity hardware and software, and (c) extending the life of the current system for a specified driver/event to occur. Technical advances are published on an almost weekly basis. It becomes necessary to determine how much the current system can be upgraded with replacement hardware and software before expending more resources into a near-to-obsolete system.

Economics

Economics is a look at the "bottom line" by the managers and financial experts. They must be provided with proof that a replacement system is feasible, cost-effective, and can be accomplished within budget and time constraints. The economics will determine where the money is coming from (end user, department, or corporate expense account). In all instances, the economic change driver will be looking for a change that will be a profitable venture and increase the "bottom line."

System end user

System end user has a high stake in the change process. They could be the ones that are requesting a new system. The end user is most concerned with the questions: Is it user friendly? Does it mimic the old system? Does it require retraining? These folks will also provide feedback about any necessary enhancements that must be made in the new system that were lacking in the existing system.

Requirements creep

The new system may also provide new capabilities that are beyond the scope of the original system. It will be necessary to develop a matrix of these new capabilities then determine which ones will be included in the upgrade and which will be left dormant for the present.

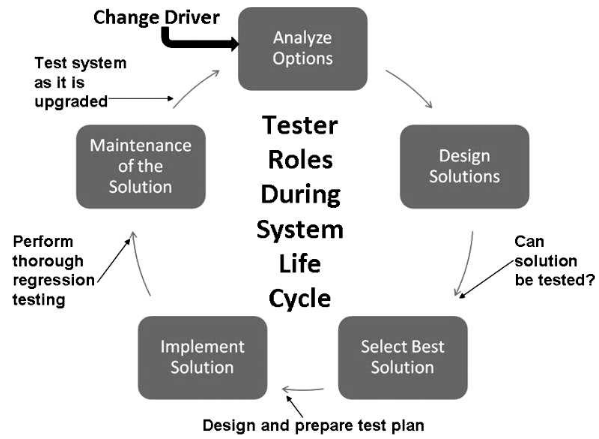


Figure 3. Tester roles during the system life cycle.

Tester roles

Testers should begin their tasks as soon as possible during the system life cycle (*Figure 3*). This process starts with a review of the proposed designs to make sure proposed solutions are appropriate, testable, and able to satisfy current and future system upgrade efforts.

The selected replacement system must be able to meet or exceed all the requirements and capabilities of the original system as well as correct any problems detected in the original system. This is accomplished through aggressive regression testing. Finally, the replacement system must pass a stringent battery of testing to "shake out" any new problems encountered as a result of the testing of new hardware and/or software.

Testers must also be prepared to perform both the regression testing and new tests to verify and validate any modifications to the system during the maintenance phase of the system life cycle.

Hardware

Hardware enters the tester's domain when it is received from the vendor or supplier. It will be subjected to a pre-installation hardware checkout, the installation and assembly testing, the installation and integration of the operating system (O/S) and application software, and a checkout test to make sure the operating system and application software are operable.

Software

Software has two possible paths to follow. The first path includes a new hardware platform. The tester must shake out the new O/S; integrate any programming language for software development from the existing system onto the new platform; test the existing application software for portability and reliability with the new O/S; test all communications connectivity by exercising all communications paths used by the system; revise the existing software system to function properly on the new platform (if needed), and finally perform a full set of regression tests using the old and new software application(s). The second path includes the existing hardware platform. Here the tester must integrate the upgraded O/S with the programming language and application for the existing system; test the existing application software for portability and reliability; and test all communications connectivity by exercising all communications.

Technological advances

Technological advances have made today's personal computer more powerful (as a processor) than the minicomputer of 1985. This has led to the localization of processing and the centralization of storage (use of networks and servers). In addition, equipment physical size has become a smaller footprint as new hardware designs are perfected and marketed. Technology is continually expanding the memory and processing capacities and increasing the throughput capabilities with each new product cycle in the marketplace. Buying a "ready made solution" from a vendor can save some expenses, but make sure the COTS products truly fit the needs better than a specifically designed hardware or software solution.

Economics

Economics dictates you need to get the "biggest bang for the buck" and this means that for any change there will be budget constraints. There is another concept sometimes called "leading edge versus bleeding edge." Most new products today undergo stringent prerelease testing, but even so, not all problems encountered are fully resolved and/or corrected when the product first hits the market. Folks who rush to buy the newest hardware or software are on the "bleeding edge" because they are the final testers for the product, which may or may not crash an existing system. Those on the leading edge wait for 6 months, get a product that is fully market-tested, and it is still considered a state-of-the-art product. Another economic trend is that costs for hardware tend to go down after 6 months to 1 year on the market and costs for COTS software usually have an annual price increase. Costs for custom software must be carefully negotiated and include milestones with both incentives and penalties based on meeting the contracted timeline.

Summary

Drivers are the behind the scenes initiators of change. There may be a single driver operating, or there may be several drivers that necessitate the end of one system's life cycle and the beginning of the new system's life cycle. In this article, we have reviewed the most common drivers including hardware, software, technological advances, economics, and requests from the system end users. The results of thorough testing of the chosen new system solution will ensure that the tech refresh is successful and has produced a system that will be reliable, sustainable, and maintainable for the foreseeable future. ❑

MARY-ANN M. BOYCE is a senior software engineer for L-3 Communications. She has over 30 years experience analyzing, designing, testing, and implementing system solutions for both aviation and business systems. She is also a senior adjunct Instructor in Computer Science at Atlantic Cape Community College for over 25 years. Ms. Boyce holds a history degree from University of Massachusetts, Dartmouth, and a computer science degree from State University of New York Institute of Technology. She is a professional member of the American Society for Engineering Education (ASEE), a senior member of the Institute of Electrical and Electronics Engineers (IEEE), and an associate fellow (Lifetime) of the American Institute of Aeronautics and Astronautics (AIAA). E-mail: m.boyce@ieee.org.

DEREK KUBONT is a Federal Aviation Administration (FAA) computer specialist at the FAA William J. Hughes Technical Center. He has over 6 years experience with aviation and networking systems. Currently Mr. Kubont is the lead of multiple En Route projects based within the National Airspace System (NAS). Mr. Kubont holds a computer science degree from Richard Stockton College of New Jersey. E-mail: derek.kubont@faa.gov

ITEA Journal 2009; 30: 421–433

Examining the Role of Metadata in Testing IED Detection Systems

Paul J. Fortier, Ph.D. and Kiran Dasari
University of Massachusetts Dartmouth, North Dartmouth, Massachusetts

A sensor provides capabilities to extract unique spatial and temporal instances of a measurement specific to a system under test. The measurement is only useful when the interrogation agent fully knows the measurement's context. After a measurement event, if insufficient descriptive, contextual information is available, the measurement is lost for use in ascertaining future value. This article defines generic concepts for metadata and sensors such that their design, selection, application, and use are fully described for users, allowing for seamless present or future uses. Within this context, we examine a definition of metadata, metadata types, metadata uses, concepts for sensor ontology, and multilevel sensor metadata. Metadata's role in detecting events; acquiring measurements; converting measurements to information, information fusion, and aggregation into complex structures; developing actionable knowledge and persistent storage; and retrieving derived knowledge are addressed through presentation of an illustrative example application, that of improvised explosive device detection system testing.

Key words: Metadata; sensors; domain context; measurement; information; actionable knowledge; knowledge storage and retrieval.

Localization, selection, extraction, processing, and movement of measurement data from a system under test (SUT) to a spatial and temporal data archive, as well as query selection and reuse for new analysis purposes, represents a challenging set of tasks. Significant effort is spent in translating raw measurements to useful information through the application of time and spatial associations, calibration details, and device pedigree rankings to name a few. Additionally, time and effort is spent adding further meaning to extracted information fragments to include data usage tracking, determining environmental and situational conditions relevant to the extracted measurement, as well as relevant contextual configurations (such as sensor settings, sensor base definitions, sensor placement, sensor pedigree, sensor orientation, etc.) to support formal assessment and interpretation of the raw measurement data. Information often is collected without regard for sensor health and is assumed correct and accurate. Such ad hoc methods are unacceptable if data are used in decision making or within human safety applications.

Data measurements represent raw atomic analog and or digital values derived from a discrete sampling device. Data alone provide little utility; consider a simple data set: *10 08 09 12 10*. The values could be interpreted in many ways; they could represent five distinct raw measurements for a single sensor (e.g., voltage levels), or they could represent a complex collection of grouped and organized items representing something totally different, such as a date (10-09-2008), time (1200 hours), and a temperature (10°C). To determine meaning requires that we add a relatively small amount of descriptive information to provide an accurate interpretation of the raw data. Without this descriptive information (metadata), there is little if any added utility to sampled and collected measurements.

The focus of this article is on the development of end-to-end metadata concepts aimed at derivation of actionable situational awareness for a target domain using information added to, extracted from, or derived from a collection of sensors. The sensors physically can be mobile or fixed, remote or local, appliqué or embedded, onboard or off board, dumb or smart. The primary driving requirement is the need to measure, collect, preserve, communicate, and share sensor-derived information for a variety of present and as yet unforeseen future test and evaluation applications. To accomplish this goal, we must make metadata

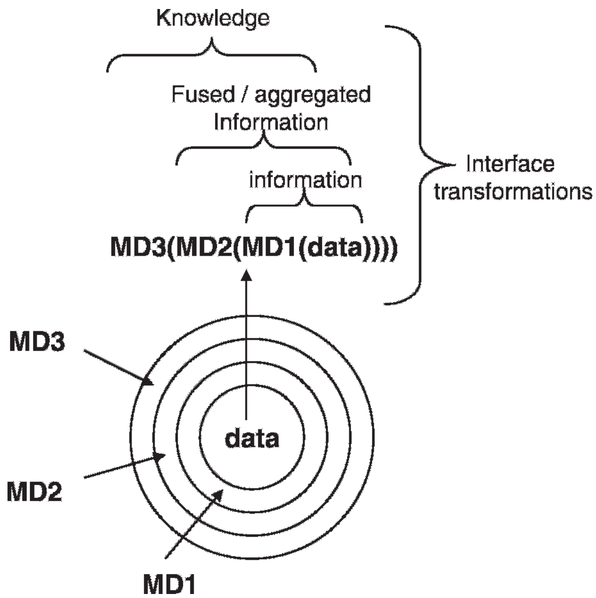


Figure 1. Metadata's role in data transformation.

available to aid in placing domain context and state to collected information to support information reuse or to support new initially unforeseen applications.

Metadata are the means through which raw data are transformed into information and are aggregated and fused into new synthetic representations, and finally through ontology into knowledge *(Figure 1)*. Metadata should aid in defining what data are present, what aspect of reality do the data represent, where the data are located, how the data got located there, how we acquire the data, how the data may be used, who may use the data, what available transformations or services are available to act upon this data, to name a few. The metadata management components for a test and evaluation (T&E) system should be able to clearly and concisely provide answers and direction to address each of these issues for all data and informational items under its management and control.

Metadata can be highly organized and formal, such as would be demonstrated by domain ontology (Borst 1997) or more ad hoc as found in an application program's data type specification. An ontology or context specification (Gauvin, Boury-Brisset, and Auger 2004) is the result of a study to categorize and organize items (Gruber 1993), that exist within some domain (Macintyre 1972). Such domain terminologies, optimized for human processing, are characterized by a significant amount of implicit knowledge. Deeper meaning for T&E can be achieved by constructing a T&E domain ontology. The benefits from such an exercise include: the ability to build more powerful and more interoperable information systems; support for the requirement to transmit, reuse, and share system and test data in real time and for future uses; semantic-based criteria to support different statistical aggregations; and possibly the most significant benefit an ontology brings to T&E systems is the ability to support the integration of knowledge and information (CCDA 2000; NISO 2004).

Metadata

Metadata has many uses, not simply to define information structure or domain transformations (Baca 1998; Duval et al. 2002). Metadata can describe how to process a collection of diverse information to represent some abstract construct (Bose 2002; Tambouris, Manouselis, and Costopoulou 2007) such as a synthetic or virtual sensor specification (Sethunadh, Athuladevi, and Iyer 2002). Metadata can describe system interaction, or how to convert information into knowledge (Ladner and Pe 2005). A requirement for the embedded instrumentation systems architecture (Michel and Fortier 2006, Visnevski 2008) was to generically define informational, control, and behavioral models to support embedded nonintrusive sensors in a T&E environment using metadata. Metadata can describe informational data flow, data aggregations, and data fusion supporting real-time synthetic sensor construction and operations (Sethunadh, Athuladevi, and Iyer, 2002; Visnevski and Johnson 2007).

Metadata classification

There is not one format or definition for metadata within the context of T&E systems applied to native, embedded, appliqué, or noncontact nonintrusive instrumentation (NII). Typical research and trade literature defines metadata according to structure and semantics (Qin and Prado 2006), or functions (services) supported (Tan 2004; Tannenbaum 1998). Within these two broad categories lie additional refinements for metadata classification. Upon review of the present research and trade literature, nine distinct classifications or types of metadata become apparent; these are as follows: descriptive, structural, administrative, preservation, usage, interface, transport, context, and process metadata. Each distinct type has a place within the T&E community and NII, and is briefly described.

Descriptive metadata are used to define, identify, and describe a measurement resource. The transducer electronic data sheets (TEDs) (Lee 2006b) and SensorML (Botts et al. 2004) represent a standardized digital means for providing specifications for sensor components and systems. SensorML's metadata includes identifiers, classifiers, constraints (time, legal, and security), capabilities, characteristics, contacts, and references in addition to inputs, outputs, parameters, and system location, which can be mined and used for discovery of sensor systems and observation processes. TEDs provide metadata to aid in the definition of the sensor measurement device often referred to as the transducer, as well as definitions for basic elements and functional elements of a sensor using IEEE 1451's concept of a functional block.

Structural metadata provides information to define or organize complex collections of information items or to define a composite measurement composed of these basic information items (Hall and Llinas 1997). The definitions may include algorithms (possibly even a chain of algorithms needed to produce a desired derivative measurement) to use in the selection, extraction, fusion, aggregation, and combination of signal streams into a synthetic measurement or intermediate computation for use in a further informational refinement.

Administrative metadata represents information needed to manage and supervise all system metadata. Administrative metadata may include access rights information to metadata and data items, version control information for an item, data creator information, location information, and data production information supporting maintenance and supervision of data resources (Park et al. 2006).

Preservation metadata provides information needed to store and maintain information in a persistent, recoverable form within an archive. Preservation metadata may include physical assessment of data, the media they are housed in, storage formats, refresh rates, history of refresh, rebuild or recovery history, redundant copy location, repository or media status, pedigree, provenance, and other information related to long-term storage management (Ledlie et al. 2005).

Usage metadata maintains a record of how a data item and related metadata were utilized. One could look at this form of metadata as state maintenance or log history of operations concerning a data item. The metadata may include data inputs, outputs, intermediate values, and process chain descriptions along with configuration information to allow for restoration or reconstruction of a usage thread (Groth, Luc, and Moreau 2004).

Interface metadata define inputs to a stage or phase of operations for a system, along with functionality or possibly protocols applied to the input to transform them to the appropriate output format. Typical definitions may include which additional metadata are added to inputs so outputs can be transformed and formatted correctly to pass useful information to the next requesting level or phase of systems operations.

Transport metadata are an essential type of metadata when a communications media of any kind is associated with a domain. Transport metadata are used to define the payload (packet, stream, etc.) format for a transmission protocol and the transmission protocol (e.g., Transmission Control Protocol (TCP), User Datagram Protocol (UDP), etc.) steps or traces that should be maintained to reconstruct access patterns. Transport metadata may also maintain information concerning quality of service and other network parameters (e.g., addressing formats, integrity of payload, correctness, etc.) to allow transported data to be extracted correctly and made available for use (Faulstich and Grace 2007). The Transducer ML standard defines a self-describing data exchange protocol and common metadata data format standard based on XML supporting data streaming between any sensor and a processing sink (TML 2008) (Havens 2007).

Context metadata refers to complex relationship-oriented information (Borst 1997; Bose 2002), concerning how data relate to each other and under what conditions these relationships hold, such as found in an ontological representation of a domain's knowledge. Context may include fundamental classification of items, the provenance of complex items, pedigree relationships for items, and possibly even definitions for use cases for instances of measurement and event classes in the domain.

Process metadata describe information concerning the behavior and interface of processes and workloads (Nle et al. 2006). Metadata describing processes, algorithms, and methods could be stored in an object code or interface library for reuse by different process model instances in another domain. The TENA object model (Noseworthy 2005) and LINUX binaries represent such forms of process metadata.

Uses of metadata

Just as there are multiple types of metadata, there are also many differing uses. Research literature reveals nine general categories defining metadata uses: archival and preservation, digital identification, resource discovery, e-resource organization and management, observational retrieval, operational logging, interoperability management, knowledge discovery, and application development.

Archival and preservation metadata provides instructional information to aid in storing, recovering, restoring, locating, describing media, lifecycle management, error handling, and assessing physical status of SUT data measurements.

Digital identification services aid in appropriately identifying a physical device. Metadata services allow for new device type determination, localization, categorization (e.g., provenance and pedigree), and description of basic features of a device in relation to existing devices.

Resource discovery metadata services provide for the request and discovery of embedded, appliqué, and test devices, including those over a network if applicable. Services to maintain directories of active and known sensors as well as interface services to provide for plug-and-play access of sensors support resource discovery.

E-resource organization and management metadata services keep track of sensors once discovered or inserted into a SUT. This could include specification and build of a synthetic sensor from existing sensor and processing inventories or through selection and linking of sensors and possibly external processing to form a new as-yet-unimagined synthetic sensor. Of interest to E-resource metadata are management oriented tasks such as sensor configuration, reconfiguration, availability assessment, calibration, and coordination, such as sensor subscription and automeasurement retrieval.

Observation retrieval is a basic service of sensors and requires metadata to support registered sensors and users (sources and sinks), automatic pushing of observed measurements, as well as ad hoc pulling of measurements based on user requests. Translations may require specialized translation algorithms and accuracy parameter (or lineage) metadata.

Operational logging is a primary user of metadata supporting pedigree and provenance associations for measurements and requires capabilities to trace all forms of actions within a system.

Interoperability has many meanings, and in general provides procedures, processes, policies, and mechanisms to support the use of an item developed and applied in one domain to another possibly unanticipated domain. Dependency on the degree of seamless operations required will dictate the amount of metadata needed to provide for data or functional interoperability. The basic idea is to provide metamodels allowing for code to run independent of a platform (hardware, operating system, language, etc.) or data requirements.

Knowledge discovery utilizes all forms of domain and context information provided through domain ontology and a resultant use case bases describing instances of the domain knowledge. Metadata linking these case bases and ontological metadata stores are needed to enhance knowledge discovery.

Applications development is typically not thought of as a fundamental user of metadata. However, in sensor networks for T&E there will be a need for a variety of services for application development to maintain, use, or generate metadata concerning programming interface (application programming interfaces, graphical user interfaces) system configuration, sensor status (including fault assessment, management, reconfiguration, correction, etc.), and workflow configurations.

Example systems use of metadata

Sensors and sensor networks were originally driven by military and security concerns, and are now being developed and fielded into previously unenvisioned applications (e.g., habitat and environmental monitoring, pollution assessment, renewable energy management, home energy management, crop assessment, weather forecasting, disaster alerts, and endangered species assessment [Biagioni and Bridges 2002; Mainwaring et al. 2004;Wang 2003]).

Recent efforts in sensor networks focus on the need to refine and standardize on interoperable services (Lee 2006a; Lee and Percivall 2008). Services can roughly be broken into three categories: sensor management, operational development services, and applications development services. Sensor management services define policies and mechanisms for using sensor operational services to localize, configure, control access, and manage operations of sensor networks. Sensor application development services focus on services and development tools aiding sensor definition, integration, configuration management, fault management, and system assessment. Services for sensor operations support sensor specification, interface specification, directory management, collection services, observational services, notification services, planning services, coordination services, transactional services, information aggregation and fusion services, persistent storage and archival services, operational logging services, and configuration services (Lee 2006b).

One integration effort generalizing sensor networks architecture and services is Embedded Instrumentation System Architecture (EISA) (Michel and Fortier 2006; Visnevski 2008). EISA is an initiative funded under the Test Resource Management Center (TRMC) T&E/Science and Technology (S&T) NII focus area and has as a fundamental objective to develop a common, comprehensive methodology for nonintrusively collecting massive amounts of T&E data supporting war fighter systems testing. EISA offers a metadata driven methodology and common architecture for heterogeneous data collection, aggregation, and fusion in a real-time synchronized and correlated fashion. These methods support the real-time instrumentation and sensor management supporting nonintrusive synthetic (virtual) and real test instrumentation.

Initial sensor services mappings to the EISA architectural framework *(Figure 2)* indicates metadata related to defining resources, managing sensor operational conditions, calibrating sources, and configuring collection and transport services are found in the lowest three levels of the Information Technical Reference Model (ITRM) pyramid (Joshi and Michel 2007).

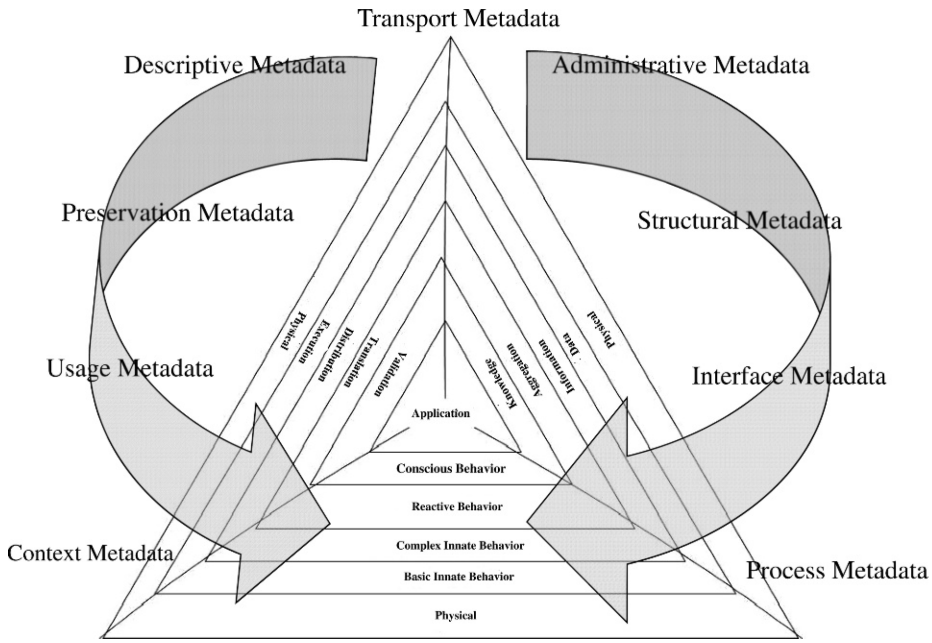


Figure 2. EISA and metadata mapping.

Metadata related to information aggregation and fusion, context refinement, scenario mapping, knowledge representation, and user administration are found in the upper layers. Metadata services such as test planning, measurement observation acquisition, alert configuration, and management utilizing myriad forms of metadata likewise map to numerous layers. EISA and the ITRM provide evolving templates to develop a mapping of metadata types and services for sensors applied to disparate domains such as medical informatics and environmental sustainability monitoring (Dasari 2008) as well as military and homeland security systems.

IED detection system T&E

For purposes of this article, the example focuses on metadata generated, extracted, or derived to construct and perform improvised explosive device (IED) detection systems testing within a generic test range *(Figure 3)*. Testing of IED detection products will be ongoing and evolving as the enemy's tactics and technology evolve. Each service has a number of products, projects, and proposed products in some stage of research and development. Such programs include Buckeye, developed jointly with the Army Corps of Engineers (Kauchak 2006) using imagery analysis; Shadow (Harpel 2007), an Army unmanned

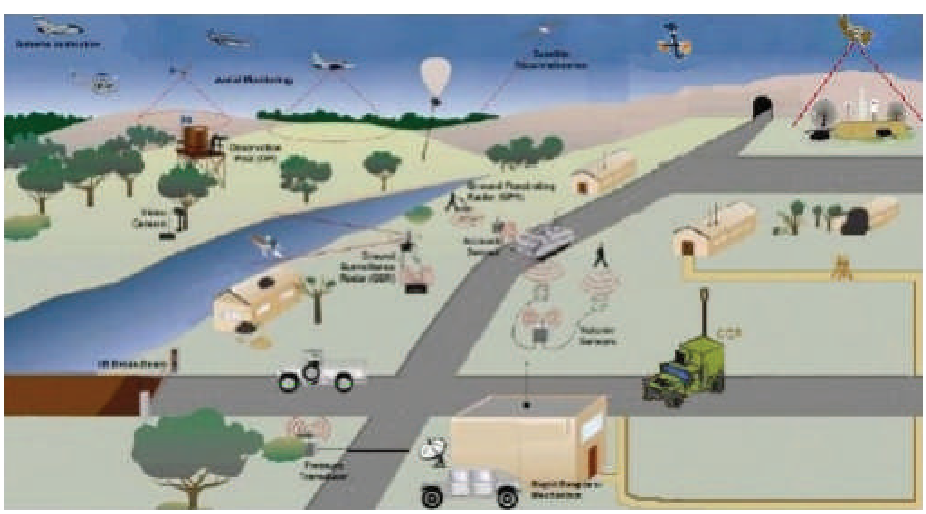


Figure 3. Test range for IED detection and defeat systems evaluation.

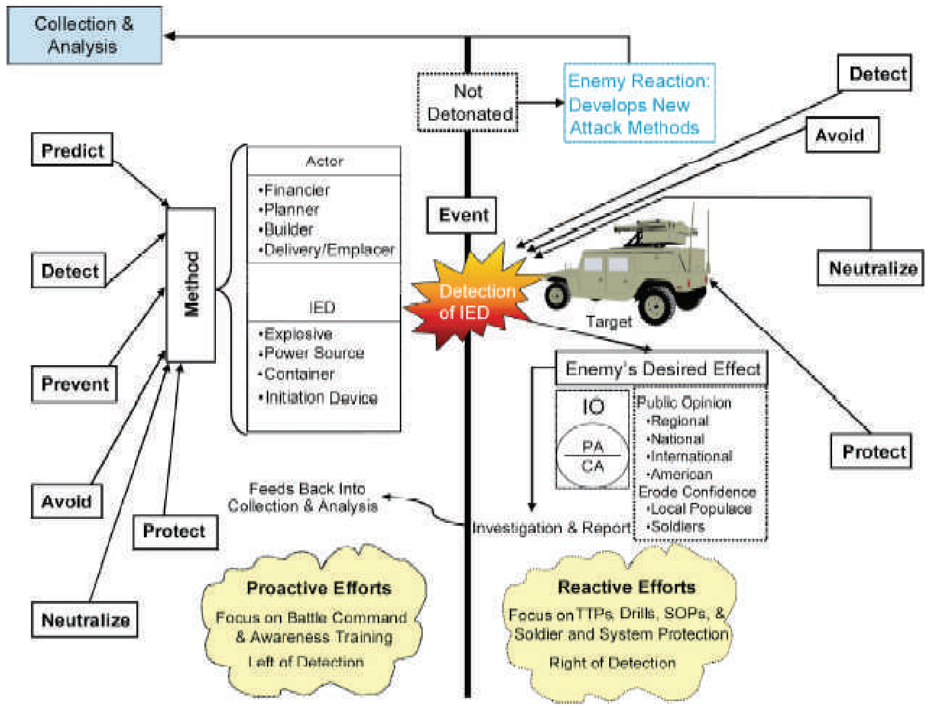


Figure 4. IED defeat system components.

autonomous vehicle (UAV) system using change detection and hyper spectral sensing; British Aerospace (BAE) systems' Talon Radiance II hyper spectral sensor system using both change detection and data mining technology to detect possible IED sites; EDO's Joint Counter Radio-Controlled Electronic Warfare (JCREW IED) jammer system; Northrop Grumman's Vehicle and Dismount Exploitation Radar (VADER) system using radar mounted on an airborne UAV system; and the Israeli Trophy Active Defense System (ADS) examined by the Army for fielding in Iraq.

The IED as a weapon has been known for decades in various forms (land mines, booby traps, suicide bomber, Kamikaze, etc.). It is only recently because of the conflicts in Iraq and Afghanistan that the Military has begun a rigorous examination of IEDs as a serious and coordinated threat and weapon. The Department of Defense (DOD) has begun development of technology for detecting a wide variety of these devices as part of a coordinated Joint IED detection and defeat organization. Typically IED devices are fielded in one of three primary forms: vehicle borne IED, suicide bomber (person-borne IED), and ad hoc munitions or Leave behind IED (e.g,. a C4 charge implanted in an animal carcass with cell phone detonator thrown on the side of a road, or as a pipe bomb taped to a target).

A complete anti-IED system requires an IED detection component, an IED assessment component, and an IED defeat component *(Figure 4)*. The IED detection system test example developed is limited to the testing of anti-IED detection hardware, software, procedures, metadata, and information during the arming, detonation, and assessment periods within the lifecycle of an IED *(Figure 5)*. Not included is the testing of anti-IED support platforms (such as a mine resistant ambush protection (MRAP) or joint EOD rapid response vehicle (JERRV), human operators, and countermeasures, though testing that includes such platforms as part of an environment for scoring and assessment of sensors and or procedures performance are considered. IEDs are constructed from a variety of elements *(Figure 6)*, all of which must be represented in a testing environment (typically not in their active form) for use in testing IED detection components.

Example scenario

In the IED detection system test scenario, we examine three phases of system T&E: the test planning stage, the test preparation phase, and the test execution phase. In each of these test phases, we examine the metadata requirements, existing tools, and techniques that exist to provide the needed metadata and services as well as define shortfalls and issues lacking.

In the scenario, a test engineer wishes to test the war fighting worthiness of a new IED distributed multi-sensor detection system. The test engineer needs to design a global situational assessment Data Acquisition System (DAS) using legacy, manufacturer, and newly

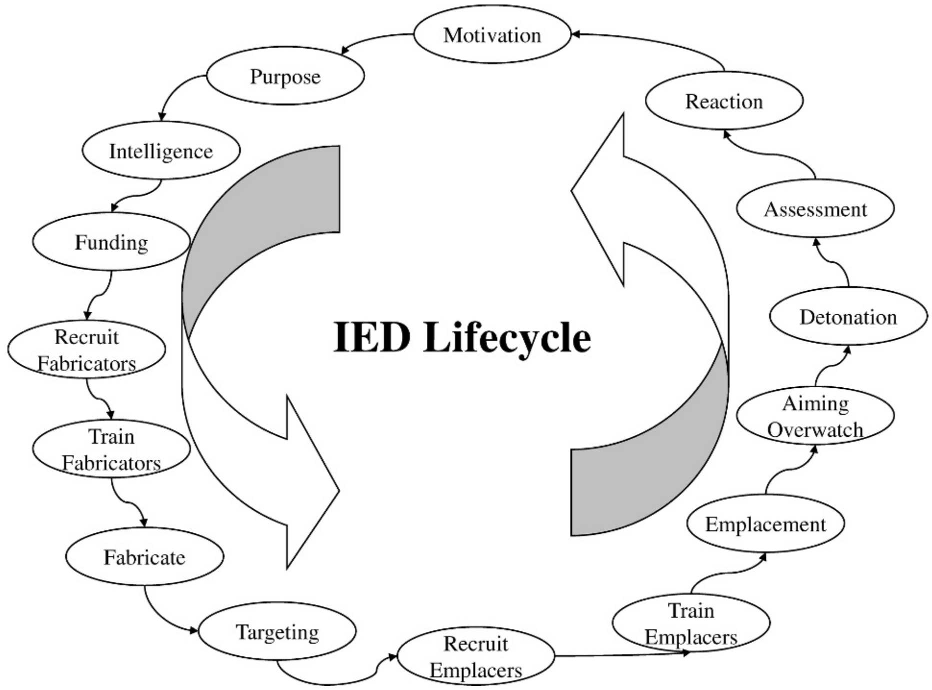


Figure 5. IED lifecycle.

developed NII components *(Figure 3)*. Installation of a variety of physical NII sensors; intermediate data collection and processing units; time, space, positioning information (TSPI) devices, and communications (telemetry) units must be performed on targets (IEDs), and detection platforms (e.g., explosive ordinance disposal [EOD] robots, fixed and maneuverable autonomous surveillance platforms, combat engineers, and EOD vehicles). All newly developed NII components installed are "smart" and can self-identify their capabilities using metadata to the test data acquisition unit. Legacy components must be defined manually and configured using master–slave metadata wrapping concepts. The DAS reconfigures all discovered and configured legacy elements to form a synthetic situational assessment sensor. The synthetic sensor uses data from all component sensors augmented with workflow, algorithmic, spatial, temporal, and contextual metadata to construct the virtual measurement in real time.

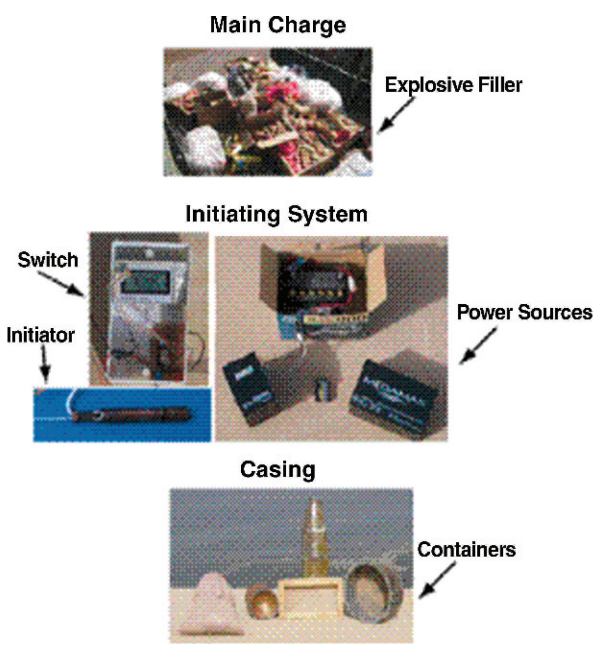


Figure 6. Components of an IED.

The configuration requires the use of sensor and test planning tools and configuration metadata available through vendors' metadata (e.g., TEDs, TransducerML, and SensorML) using a common set of standards for both hardware and software. Once configured and the test commences, sensor measurements are collected and tagged with appropriate metadata (e.g., time tag, location, pedigree, provenance, etc.) indicating all relevant test data for the DAS to use and for storage for future reference and use. It is assumed that none of the sensors is physically interfaced with a SUT devices' data bus, though through observational services, test engineers can have native generated measurement data made available for consumption outside of the SUT operational envelope. The DAS network composed of all relevant source and sink data collection sites is self-configuring. The DAS network provides additional services to ensure self-

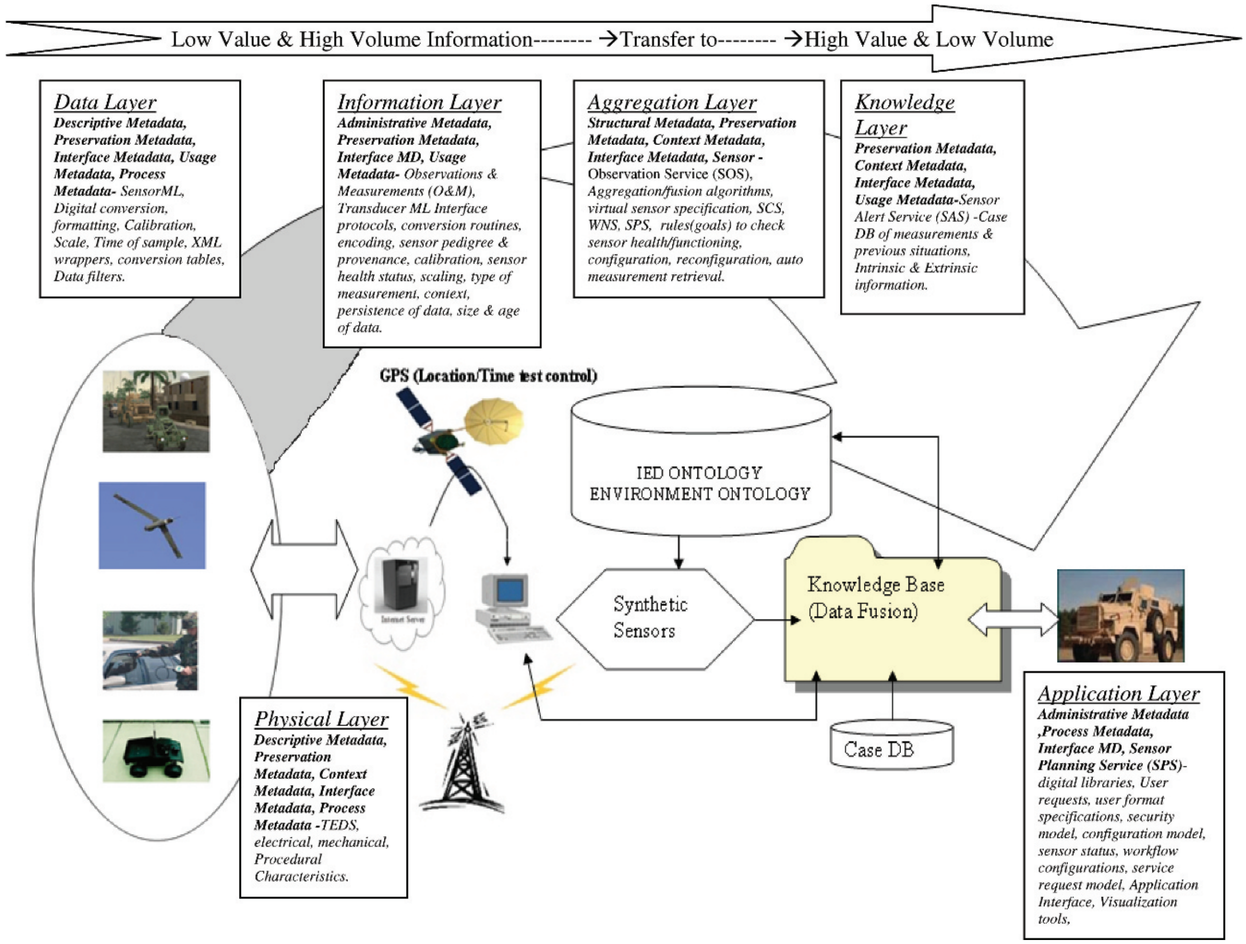


Figure 7. Control flow model and metadata use.

reliability through reconfiguration services and redundancy, which are specified and configured into the devices at initial DAS configuration time (using metadata). All collected data are stored redundantly on board each collection site SUT and in a composite repository off board.

Planning and test preparation phase

During the planning phase, the test engineer uses test planning services that include standardized tools for sensor planning (e.g., sensor web enablement [SWE] sensor planning service (SPS) services) *(Figure 7)*. The tools are used to manage requests by applications to plan events (e.g., set fixed optical sensors to take images at x frames per second, beginning at time t_i and ending at t_{i+j} to synchronize with scheduled IED firing event), configure SUT sensor resources, reconfigure resources (e.g., reset target IED simulators and emulators to provide desired test signatures), calibrate sensors (e.g., test sensitivity, resynchronize time clocks), detect or find sensors (e.g., request inventory of all available sensors, services available, and present status), support addition of new sensors, and initiate the collection and dissemination of measured data.

In the IED detection system test, a test engineer desires to construct a synthetic measurement from available measurements found on the detection mobile vehicle, an EOD talon robot, an AUV, and multiple fixed sensors of varying type and pedigree. The system requires context metadata (e.g., environmental conditions, ontological description of the test domain including GPS signatures from all in situ fixed sensors, derivable through open web services [OWS] and SWE services augmented with Open Grid Forum [OGF] and Open Ontology specification tools), administrative metadata (e.g., security enforcement for application construction and data use, access, and integration services for legacy wrapper access and builds, localization, and authorization for real and virtual sensor configuration historical files for use in configuration builds), descriptive metadata (e.g., specific sensor data used to define real and virtual sensor state such as TEDs, TransducerML, SensorML, as well as work flow specifications for building and operating synthetic instruments), and process metadata to locate sensors, algorithms, and distribute the tasks, based on priority, needed to build the user defined synthetic instruments from basic measurements and known sensors. Once synthetic sensor specification is completed, available resources must verify their capabilities to accomplish the task using structural metadata (e.g., configuration checks, status checks, and calibration checks) and preservation metadata (e.g., location and configuration of sources, sinks, and paths) and descriptive metadata elements.

Any required transformations will also be defined at this time using interface metadata (e.g., correlating map data with geographic information systems [GIS] data and TSPI data, along with sensor streams to facilitate real-time streaming synthetic sensor operations) and placed in the structural specification for the synthetic sensor. Many system level services and service oriented applications must be developed to allow for the actual control (e.g., error detection, error correction) of the system during the test.

Test execution phase

In the IED defeat system test example, a trace for a unique sensor measurement through two faces of the EISA model, the information and the control models, is performed. Within the examination is a definition for raw sensor measurements flow from a sensor to the monitoring applications, illustrating metadata extracted and used for interpreting measurements from the source through the application sink.

Raw measurements are detected and verified by a sensor (described through descriptive metadata such as SensorML), stored (using preservation metadata, e.g., Structured Query Language, Open Web Language), extracted and translated into information (using interface metadata, TransducerML, SensorML, sensor observation services (SOS)). The extracted information is also marked with usage tags and contextual tags (using extracted usage and context metadata, including provenance and pedigree information to support replay, query, or restoration). These added metadata initiate data provenance and pedigree chain formations for future data preservation. In the IED detection example, the extracted information is combined with additional measurement, spatial, and contextual information items using structural and process metadata to build a synthetic sensor measurement supporting the multisensor detection scenario. To perform data fusion and aggregation operations using multiple heterogeneous data fragments as inputs, developers must extract process metadata describing the fusion algorithm(s) to utilize (possibly using Sensor ML) or the dataflow processes to use a structural metadata item defining a synthetic sensors specification (also possibly using Sensor ML). Aggregated metadata are used along with context metadata to place the appropriate domain specific parameters to aid the goal directed synthetic computation (using TransducerML and Ontological specification). Context metadata, in the form of domain ontology, are used to place the synthetic data in a place and time for the IED defeat system and component under test.

The derived composite synthetic, aggregated measurement is then transferred to the knowledge layer to be used for additional actionable knowledge development. This may include incorporation into the knowledge base as a new case instance or an ontological instance. Transport metadata are used to aid in payload specification, packaging, and transmission. Along the synthetic sensors' data flow path (Figure 8), each metadata item extracted and used in the transformation of the raw measurement into actionable knowledge is tagged and stored using usage metadata, preservation, and administrative metadata supporting provenance storage, long term management, and postretrieval of the measurements.

The stored ontological and instance information is used to maintain lineage and pedigree of measurements. Using preservation metadata, the derived synthetic situational measurement is persistently stored, maintaining information such as how, what, where, why, and by whom was this measurement stored. Preservation information can later be used to retrieve and restore measurements for future uses. Along the dataflow path *(Figure 8)*, administrative metadata are used to determine if the application requesting the measurement has appropriate authorizations and to orchestrate the performance of predefined workflows (process metadata) performing desired actions.

Another important service provided is support for the configuration and logging services to aid in the collection of build information into usage metadata for future pedigree and provenance determination. The example illustrates a small fragment of the services needed to support a user's construction of an IED defeat system test scenario, from the initial definition of all sensor and service resources available, selection of desired resources, to configuration into desired analysis displays.

Standards for interoperability and reuse

No single standard seems to capture all the elements needed within the IED defeat system example test scenario. To capture requirements to support placement of items in a map coordinate system and within a

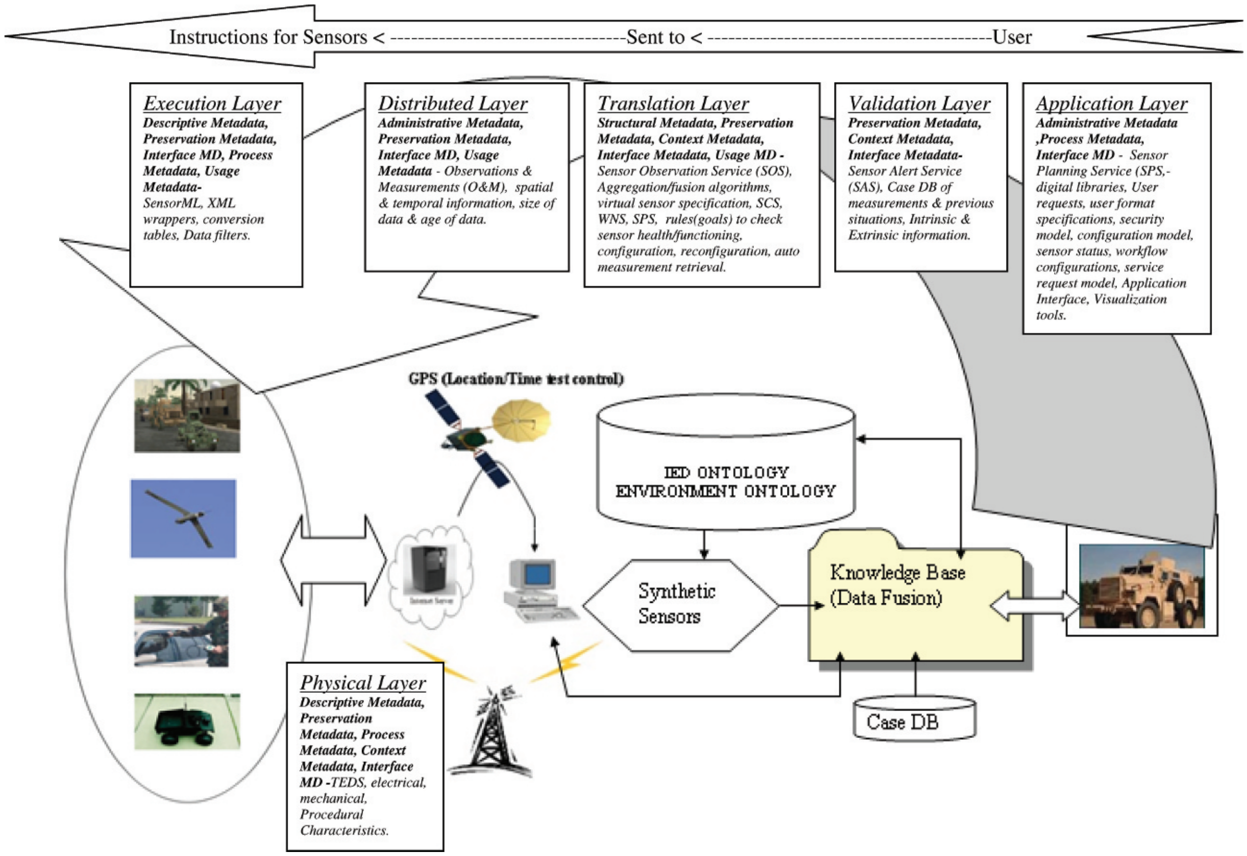


Figure 8. Information flow and metadata use.

space and time context, we require one set of services. Another set of services are needed to use these specifications to locate resources (sensors and services) to configure these services for a specific planned use, to control their operation to effectively extract the correct observations, to configure and control alerts allowing for real-time control of boundaries or events of interest, and to correctly capture data and provenance and pedigree metadata for use in postanalysis or modeling.

Figure 9 depicts a plausible configuration of available standards and services that would be needed to plan, build, configure, operate, and analyze the scenario we have postulated. The Open Geospatial Consortium (OGC) and the open grid forum are collaborating on collections of open standards that address many of the distributed computing and geospatial issues required by the testers in building distributed tests for systems of systems testing, such as is found in the IED detection system described within this article.

The OGC (2004) has specified a set of web services (OGC Web Services or OWS) standards that can be layered on top of sensor specific services to provide distributed geospatial services. One of the standards within this collection is the Web Feature Service, which provides standards for retrieval and update of digital representations of real-world entities tied to the earth's surface. The Web Mapping Service standardizes the integration and display of superimposed map

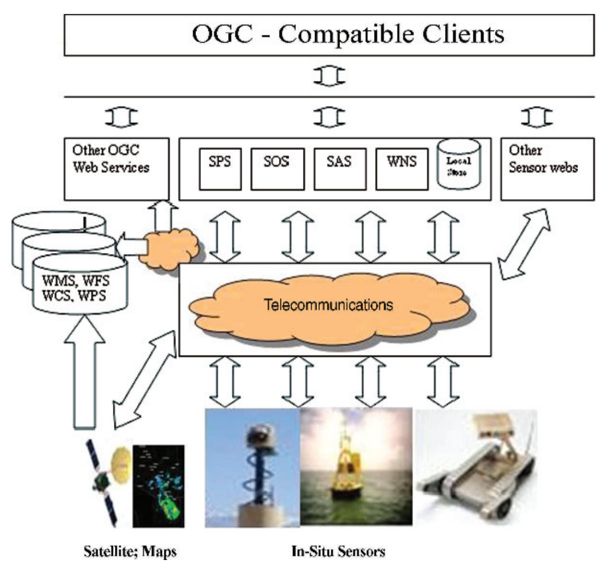


Figure 9. Available and evolving standards for example test use.

entities from multiple heterogeneous sources, the Web Coverage Services standardize access to spatially extended coverages (details), and the Web Processing Services provide standardized basic request and response interaction protocols and metadata formats for remote execution of any algorithm, calculation, or model operating on spatial data. The Catalogue service provides standardized services and metadata specifications to publish, discover, browse, and query metadata about information and services available.

The OGC's SWE standards provide distributed and local services for accessing, controlling and using sensors, instruments and imaging devices. The SWE standards consist of four primary standards: Sensor Planning Service, Sensor Observation Services, Sensor Alert Services, and Web Notification Service. The Sensor Planning Service standardizes the tasking of sensors or models.

Tasks include reprogramming, calibrating, starting, altering a sensor mission, and controlling simulation models. The Sensor Observation Services are used to standardize the methodology and metadata used in retrieval of measurement observations from a sensor or model. Included in the retrieved information is sensor system configuration and status metadata information. Sensor Alert Services provide standard protocols, interfaces, and metadata specifications for subscribing to or publishing alerts configured for sensors. The Web Notification Service defines a set of standard specifications that control and configure the way Web Services interact using a predefined notification collection or pattern.

The low level sensors also require additional standardized services for specifying individual instances of a sensor and models including workflows defining composite sensors and synthetic sensors, spatial locations, contextual information, and other relevant information that will make them available to the upper level services. The basic standards available for such descriptions include, but are not limited to, the IEEE 1451 standards, TEDS, TransducerML, and SensorML standards. These standards are further augmented with ontological standards for OWL (McGuiness 2004) and database standards (e.g., SQL database language [Melton 2003] or OMG—object database or metaobject facility [Pope 1998]) to provide tools to organize these resources into persistent collections of device specifications, measurement instances, and knowledge repositories.

Conclusion

In this article, a definition of metadata and how it is used in the context of sensor systems applied to T&E is developed. The term "metadata" is found to be more than simply "data about data." "Metadata" has many meanings and many applications dependent on the context. The type of metadata is not uniform, but instead is defined based on a few fundamental questions, what is it, how is it used, where is it used, who generated it, how is it related to other data. In this article, we developed these concepts and applied them to an example configuration, control, and execution thread within a heterogeneous distributed multisensor IED defeat system T&E scenario. The natural conclusion from this effort is to look toward metadata and sensor network service standards to realize the true value and strength of metadata to enhance seamless communication and interoperability and reuse of sensor and NII collected information for the T&E community. ❑

PAUL J. FORTIER, PH.D., is a professor of electrical and computer engineering at the University of Massachusetts Dartmouth located in North Dartmouth, Massachusetts. He is a member of the IEEE and ACM as well as numerous technical and working groups of these societies. He is a past member of the SQL Standards Committee, the Predictable Real-time Systems Task Group, the ANSI POSIX Working Group, and present member of the American Medical Informatics Associations HL7 EHR Standards Committee and the iNet Metadata, Test Article and Systems Management Standards Working Groups. Prior to his arrival at UMass Dartmouth, he served as a senior systems engineer at the Naval Undersea Warfare Center for 16 years. His research interest are in real-time and temporal databases, data mining and knowledge discovery, medical informatics, real-time and semantic concurrency control and transaction processing models, performance modeling, database language standards, and embedded sensor systems and sensor networks. E-mail: pfortier@umassd.edu

KIRAN DASARI is a graduate student in the Electrical and Computer Engineering Department at the University of Massachusetts at Dartmouth. He will complete requirements for his master of science in computer engineering in December of 2008. Mr. Dasari's interests are in business process models and in metadata models for ontological data representation. He will be joining the Ross School of Business at the University of Michigan, Ann Arbor in January of 2009. He has plans to set up his own consulting firm in emerging economies upon completion of his advanced degree.

References

Baca and Murtha, ed. 1998. *Introduction to metadata: Pathways to digital information*. Washington, DC: Getty Information Institute. Version 2.1. http://

www.getty.edu/research/conducting_research/standards/intrometadata/index.html (accessed June 10, 2008).

Biagioni, E. and K. Bridges. 2002. The application of remote sensor technology to assist the recovery of rare endangered species. *International Journal of High Performance Computing Applications*. 16 (3): 315–324.

Borst, W. N. 1997. Construction of engineering ontologies, Doctoral thesis, University of Twente, Enschede.

Bose, R. 2002. A conceptual framework for composing and managing scientific data lineage. In *14th international conference on scientific and statistical database management*, ed. J. Kennedy. 15–19. Edinburgh, Scotland: IEEE Computer Society.

Botts, M., ed. November 2004. *Sensor Model Language (SensorML) for in-situ and remote sensors*. Location: Publisher. http://vast.nsstc.uah.edu/SensorML/Sensor ML_04-019_1.0_beta.pdf (accessed October 20, 2007)

Committee on Catalog Description and Access (CCDA). 2000. *Task force on metadata: final report* (CC:DA/TF/Metadata/5), June 16. http://www.libraries.psu.edu/tas/jca/ccda/tf-meta6.html (accessed June 16, 2008).

Dasari, K. 2008. Development of sensor ontology for environmental monitoring and situational assessment. Master's thesis, University of Massachusetts Dartmouth.

Duval, Erik, Wayne Hodgins, Stuart Sutton, and Stuart L. Weibel. 2002. Metadata principles and practicalities. *D-Lib Magazine*. 8 (4): 16 pp.

Faulstich, R. and T. Grace. 2007. Integrated Network Enhanced Telemetry (iNet) overview. In *Proceeding of the 38th Annual International Symposium of the Society of Flight Engineers,* October 22–25, 2007/ Las Vegas, NV. Washington, DC: ITC.

Fortier, P. and K. Dasari. 2008. The role of metadata for improving the utility and life of sensor derived measurement. In *ITEA Annual Technology Conference,* July, Colorado Springs, CO, 10 pp. Washington DC: ITEA.

Gauvin, M., A. Boury-Brisset. and A. Auger. 2004. Context, ontology and portfolio: key concepts for a situational awareness knowledge portal. In *Proceedings of the 37th Hawaii International Conference on Systems Sciences,* 2004 January 5–8, 2004, Hawaii, HI, 10 pp. New York: IEEE.

Groth, P., Michael Luck, and Luc Moreau. 2004. A protocol for recording provenance in service-oriented grids. In *Proceedings of the 8th International Conference on Principles of Distributed Systems* (OPODIS'04), December 2004, Grenoble, France, 124–139. New York: Springer Verlag.

Gruber. and Thomas, R. 1993. A translation approach to portable ontology specifications. *Knowledge Acquisition*. 5 (2): 199–220.

Hall, D. and J. Llinas. 1997. An introduction to multisensor data fusion. *Proceedings of the IEEE*. 85 (1): 6–23.

Harpel, D. 2007. Shadow 200 TUAV proves mettle in heavy Iraq use. *Defense Systems Daily*. November 6.

Havens, S. (ED), 2007. *Transducer Markup Language Implementation Specification*. Wayland, MA: Open GIS Consortium. Available at http://www.opengeospatial.org/ standards (accessed April 15, 2009).

Joshi, H. and H. Michel. 2007. Integrating information-centric, control-centric and behavior-centric technical reference models for autonomous sensor networks *ICWN*. 2007: 319–324.

Kauchak, M. July 2006. New eye in the sky *Journal of Military Geospatial Technology*. 4 (3): 5 pp.

Ladner, R. and Frederick E. Pe. 2005. Metadata concepts to support a net-centric data environment. In *Net-centric approaches to intelligence and national security*. 29–54. New York: Springer Verlag.

Ledlie, J., Chaki, Ng, David A. Holland, Kiran-Kumar Muniswamy-Reddy, Uri Braun, and Margo Seltzer. 2005. Provenance-aware sensor data storage. In *Proceedings of First IEEE International Workshop on Networking Meets Databases*. April 8–9, 2005, Tokyo, Japan, 10 pp, New York: IEEE.

Lee, C. and Percivall, G. 2008, November. Standards based computing capabilities for distributed geospatial applications *IEEE Computer*. 41 (11): 50–66.

Lee, K. 2006a. Sensor networks pilots for DRM 2.0: Sensor standards harmonization working group meeting, September 12, 2006. http://colab.cim3.net/file/work/SIC0p/2006-04-12 (accessed April 14, 2009).

Lee, K. 2006b. An implementation of the proposed IEEE 1451.0 and 1451.5 standards. In *Proceedings of the Sensors Applications Symposium,* February 5–7, 2006, Houston, TX. 72–77. New York: IEEE.

MacIntyre, A. 1972. Ontology. In *The encyclopedia of philosophy*, ed. Paul Edwards. volume 5, 542–543. New York: Macmillan Publishing and The Free Press.

Mainwaring, A., et al. 2002. Wireless sensor networks for habitat monitoring. In *ACM International Workshop on Wireless Sensor Networks and Applications,* September 2002. September 28, 2002. Atlanta, GA, 8 pp. New York: ACM.

McGuiness and Van Harmelen (editors). 2004. Wayland, MA: Open GIS Consortium. http://www.omg.org/ standards (accessed October 20, 2008).

Melton, J. (editor). 2003. SQL Framework. New York: IEEE Society. http://www.jcc.com/sql.htm (accessed June 14, 2008).

Michel, H. and P. Fortier. May 2006. Development of an embedded instrumentation system architecture and its comparison to the test and training enabling architecture. In *Proceedings of the SPIE Symposium on Defense and Security*. December. Las Vegas, NV: SPIE.org. 7p.

NISO. 2004. *Understanding Metadata*. Bethesda, MD: NISO Press. http://www.niso.org/standards/resources/UnderstandingMetadata.pdf (accessed June 10, 2008).

Nle, N., U. Ppeler, D. Nicklas, T. Schwarz, and M. Grossmann. 2005. Benefits of integrating meta data into a context model. In *Proceedings of the 3rd International Conference on Pervasive Computing and Communications Workshop*. March 8–11, 2005, Kavai, HI. 12 pp. New York: IEEE Publishing.

Noseworthy, J. R. 2005. Developing distributed applications rapidly and reliably using the TENA middleware. In *Proceedings of the IEEE Military Communications Conference (MILCOM),* 7p. October 17–20, 2005, Atlantic City, NJ. 7pp. New York: IEEE Publishing.

OGC (Open Geospatial Consortium) 2008. Open GIS Specifications. Wayland, MA: Open GIS Consortium. www.opengeospatial.org (accessed June 5, 2008).

Park, S., J. Kim, K. Lee, K. Shin, and D. Kim. 2006. Embedded sensor networked operating system. In *Ninth IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing.* April 24–26, 2006, Gyeongju, South Korea, pp. 117–124. Washington DC: IEEE.

Pope. 1998. The COBRA Reference Guide. Waltham, MA: Addison Wesley Publisher. http://www.w3.org/2004/OWL/ (accessed June 14, 2008).

Qin, J. and J. C. Prado. 2006. The semantic and syntactic model of metadata. In *Alfabetização Digital e Accesso ao Conhecimento (Série Comunicação da Informacao Digital 2006)*. 4: 143–156.

Sethunadh, R., S. Athuladevi, and S. Iyer. October. 2002. Virtual instrumentation techniques in test and evaluation of launch vehicle *Avionics, Defense Science Journal*. 52 (4): 357–362.

Tambouris, E., N. Manouselis. and C. Costopoulou. 2007. Metadata for digital collections of e-government resources *The Electronic Library (TEL), Special Issue on Metadata and Semantics for Digital Libraries and Information Centres*. 25 (2): 176–192.

Tan, V. H. K. 2004. *Interaction tracing for mobile agent security.* Doctoral thesis, University of Southampton: England.

Tannenbaum, A. 1998. The metadata mystique. *The Journal of Data Warehousing*. 3 (4): winter.

Visnevski, N. 2008. Embedded instrumentation systems architecture. In *Proceedings of the IEEE Instrumentation and Measurement Technology Conference. IMTC 2008.* May 12–15, 2008, Lancaster, CA. pp. 1134–1139. Washington, DC: IEEE Publishing.

Visnevski, N. and T. Johnson. 2007. Embedded instrumentation systems architecture for mobile wireless sensing platforms. *ICWN*. 2007: 335–342.

Wang, H., J. Elson, L. Girod, D. Estrin, and K. Yao. 2003. Target classification and localization in habitat monitoring. In *Proceedings of the IEEE ICASSP,* April 6–8, 2003, Los Angeles, CA. pp. 844–847. Washington, DC: IEEE Publishing.

Acknowledgment

The authors thank the Test Resource Management Center (TRMC) Test and Evaluation/Science and Technology (T&E/S&T) program for their support. This work was partially supported by the T&E/S&T Program through the Naval Undersea Warfare Center, Newport, Rhode Island, under NUWCDIVNPT SEAPORTE contract N00178-04-D-4109-N407. We also wish to thank Purvis Systems Incorporated of Middletown, Rhode Island, for their support under ATMC work order number 080825-1 through the Advanced Technology and Manufacturing Center of the University of Massachusetts Dartmouth.