39th International T & E Symposium
2022 Technical Program

Thursday, September 29, 2:00-4:00pm

Session 1 **Cyberspace Test Technology**
Chair Min Kim, Deputy Executing Agent for the TRMC T&E/S&T Cyberspace Test Technology (CTT)

2:00 p.m. **"Activity and Content Enhancement – Next Gen Traffic Generation Toolkit"**
Steve Durst, Terry Champion, and Eric Renouf, Skaion

On-host and network traffic generation are essential aspects to conducting effective cyber operations testing and training, but are often minimized. Current approaches to traffic generation for cyber testing are unrealistic, don't scale well, reveal artifices inherent in the test platform, and inhibit true knowledge capture that makes test results meaningful. We present the Advanced Cyber Environment, a next-generation traffic generation toolkit for implementing and controlling highly realistic on-host user behavior, network traffic, modern social interactions between users, and live network services, all suitable for use in a closed or "clean room" environment.

2:30 p.m. **"Measure and Share: TRMC T&E/S&T Cyberspace Test Technology's Project to Improve Cyber T&E Impacts Across DoD"**
Dr. Mike Shields, TRMC T&E/S&T and Pete Firey, MITRE

The Measure and Share Initiative is to Measure the Efficacy of Cyber Test and Evaluation, and Share the results at an appropriate classification level by providing a relevant perspective to the DoD stakeholders.

The Goals of this initiative are:

- *Improve T&E Results*
- *Improve JTF and Service Commanders Cyber Knowledge*
- *Enable Better Acquisition Outcomes*
- *Improve Intelligence Community Reporting Impacts*

This presentation will describe the Measure and Share Initiative in depth including the Concepts of Operations, current status and a way forward.

3:00 p.m. **"Advanced Automated Machine Learning System"**
Dr. Himanshu Upadhyay, Dr. Leonel Lagos, Santosh Joshi, Jayesh Soni, and Michael Perez, Florida International University

Florida International University has developed the Advanced Automated Machine Learning (AAML) system in collaboration with Test Resource Management Center. AAML enables machine learning model development, prediction, advanced analysis and visualization using the data available from different test technology

domains using traditional machine learning / deep learning and hybrid machine learning approaches.

The team has developed an automated machine learning (ML) system using an Artificial Intelligence (AI) based advanced analytics platform. AAML is a standalone system which allows analysis of data collected from different test technology domains by using traditional machine learning / deep learning and ensemble learning approaches to generate machine learning models, make predictions, then apply advanced analytics and visualization to perform analysis. This system enables automated machine learning using AAML platforms like AI based Advanced Analytics and the Analytics Control Center. This capability can be applied to multiple Data Sources.

Artificial Intelligence based Advanced Analytics Platform:
This platform is the analytics engine of AAML which provides pre-processing, feature engineering, model building and predictions. Primary components of this platform include:

- ***Machine Learning Server:*** *This module is deployed to build ML/DL models using the training data from the data sources and perform predictions/analysis of associated test data based on the AAML-generated ML/DL models.*
- ***Traditional Machine Learning Algorithms:*** *ML algorithms like Logistic Regression, Linear regression, Decision tree, Random Forest, One Class Support Vector Machine, Jaccard Similarity etc. are available for model building.*
- ***Deep learning Algorithms****: Deep learning algorithms such as Deep Neural Networks and Recurrent Neural Networks are available to perform classification & anomaly detection using the TensorFlow framework and the Keras API.*

Analytics Control Center:
This platform is a centralized application to manage the AAML system. It consists of following main modules.

- ***Data Source:*** *This module allows the user to connect to the existing data to the AAML system to perform analytics. These data sources may reside in a Network File Share, Database or Big Data Cluster.*
- ***Model Development:*** *This module provides the functionality to build ML/DL models with various AI algorithms. This is performed by engaging specific ML algorithms for five types of analysis: Classification, Regression, Time-Series, Anomaly Detection and Clustering*
- ***Predictions:*** *This module provides the functionality to predict the outcome of an analysis of an associated data set based on model built during Model Development.*
- ***Manage Models and Predictions:*** *These module allows the user to manage the ML models that have been generated and resulting predictions of associated data sets.*

Data Sources:
This platform allows user to connect to three different types of data sources to be analyzed:

- ***Network File Share****: A pre-defined network shared drive to analyze datasets for model building / prediction. Data will be stored in the form of .csv files.*
- ***Database****: A pre-defined database source to analyze datasets for model building / prediction. Data will be stored in training and test database tables.*
- ***Big Data Cluster:*** *A pre-defined Hadoop Distributed File System (HDFS) source to analyze selected datasets for model building/prediction. Data will be stored in csv format in the cluster.*

3:30 p.m. **"Vader Modular Fuzzer: What, Why and How"**
Arch Owen, Draper

The State of the Art in testing software for correctness, security and reliability includes the use of fuzzing, however, fuzzing is not widely adopted by the DoD test community. This is partially due to DoD Cyber testing policies and a lack of knowledge in fuzzing and tools suited to DoD needs. TRMC (Test Resource Management Center) T&E/S&T (Test & Evaluation/Science & Technology) CTT (Cyberspace Test Technology) has initiated an effort to expand DoD-wide awareness and experience in fuzzing, and to provide tools suited to DoD needs - specifically tools that are affordable, usable in closed spaces, suited to unique testing needs (e.g. real time embedded systems), easily adapted, incorporate the latest fuzzing techniques, and can be quickly learned by non-fuzzing experts. In order to support the DoD Fuzzing initiative, the DoD Fuzzing Working Group is being stood up. This presentation will describe the DoD Fuzzing Framework, an open source plan, current status, DoD Fuzzing Working Group and a way ahead.

Thursday 29 September, 2:00-4:00pm

Session 2 **Young T&E Professionals**
Chair Andy Novario, Test Area Manager, Department of Homeland Security (DHS)

2:00 p.m. **"T&E: It's all About Solution Assurance, Right?"**
Tara Francis, Systems Engineer, QinetiQ

Typically, on Defence programmes in the UK, the conduct of T&E activity is focused on the design and manufacturing phase, occurring towards the end of the V cycle, in an effort to verify that the end solution meets the Requirement set. Therefore, T&E is often wrongly perceived culturally as a hurdle or tick box exercise that is delaying entry to service.

Acquisition of large-scale capabilities currently takes many years but given the changing pace of technology, understandably there is a growing appetite to reduce the length and cost of programmes, to remain operationally effective. To achieve this, there is a general push across MoD and Industry to make T&E more agile and to begin using AGILE methodologies more, as a perceived way of reducing overall programme length and obtaining a viable product sooner. The attitude across the enterprise seems to be just do less T&E rather than the right T&E at the right time (which could in fact mean more T&E activity overall!).

Adopting new methodologies and attempting to increase agility within T&E activity brings an interesting cultural change challenge both within MoD and between MoD and Industry. Questions we now face are:

- *How do we expedite the implementation of new processes in an organisation that is historically reluctant to change and overburdened by extant processes?*
- *How can we update the language and definition of T&E such that, it also includes Trials and Experimentation, rather than just Test and Evaluation, promoting activity earlier in the lifecycle to support requirement definition and de-risk complex aspects of design?*
- *How do we balance risk between the Customer and Supplier, such that there is contractual flexibility to design and deliver capabilities without a fully defined requirement set but within constrained budgets?*
- *How do we make T&E a truly through-life capability, by exploiting digital tools alongside modelling and simulation techniques and improved data management from concept to disposal?*

In summary, to implement time and cost efficiencies in our acquisition programmes, through improved more efficient T&E, we need to consider more than just solution assurance and conquer the cultural challenge of both left and right shifting our ITEA processes for complex SoS.

2:30 p.m. **"Operational Testing: How MCOTEA and Marines Support the Acquisition Process"**
Major John Spahlinger

The Marine Corps has entered a new era with specific tasks set forth in the Commandants Planning Guidance and Force Design 2030. Innovation is at the forefront of this change to increase the speed at which systems are developed, tested, and fielded to maintain a technological advantage over our adversaries. Decision makers are in need of rapid information from experimentation and test events to make investment decisions that will have significant impacts on the Marine Corps' ability to fight and win. MCOTEA fills a critical role by executing meticulous planning, testing, and reporting with the sole purpose of providing independent and credible information to key stakeholders. Armed with a six-step process that combines the Marine Corps Planning Process with the Scientific Method, MCOTEA deploys civilian/military teams to partner with Fleet Marine Forces to conduct operational testing. Through operational testing, MCOTEA provides a defensible determination of a systems effectiveness, suitability, and survivability.

3:00 p.m. **"Environmental Testing: Mechanical & Electronic Subsystems"**
Kalvin Krompetz, Test Engineer, General Dynamics Land Systems

The Cybersecurity Vulnerability and Assessment Test Environment (CVATE) Other Environmental testing simulates the environmental conditions that a product may be exposed to during its lifecycle. Environmental conditions include temperature extremes, humidity, corrosion, mechanical shock, and vibration. It is critical to know details about the environments a product will experience so that the correct profiles can be selected. By testing to those profiles we can show that the product will meet or exceed its performance requirements.
At General Dynamics Land Systems environmental testing typically falls into four categories: Design Verification Testing (DVT), Qualification Testing (QT), First Article Testing (FAT), and Control Testing (CT). Where a product is in its lifecycle and risk level will determine which phases of testing will be conducted. DVT is used in order to identify design improvement areas early in the lifecycle. This is done by exposing a product to the environments that are most likely to cause issues: high temperature, low temperature, shock and vibration. Once a design is ready for production, QT is performed. Qualification testing includes all environments and electrical inputs with the intention of stressing the design as much as possible to make sure it is ready to enter into production. FAT is conducted on a few of the first units built during production in order to prove out the manufacturing process. As a quality control test, CT is completed on units at specific intervals to ensure no unintended changes have occurred in the manufacturing process.

3:30 p.m. **"Discussion and questions"**
Andy Novario, Test Area Manager, Department of Homeland Security (DHS)

Session chair and presenters open discussion.

Thursday 29 September, 2:00-4:00pm

Session 3 **How Do Know My M&S is Good Enough?**
Chair Dr. Ade Britton, Lead Solution Architect, Global T&E Campaign, QinetiQ UK

2:00 p.m.. **"Modeling and Simulation in Todays Complex world; A Testers Perspective"**
Dr. Ade Britton, Lead Solution Architect, Global T&E Campaign, QinetiQ UK

Modeling and Simulation (M&S) has been used for decades to support the design and development of defense equipment. Typically this has been designed b the manufacturer for system under development, with the testing of the final system relying on live events in major test ranges and extensive operational T&E to confirm performance. However, drivers around pace of development, system complexity and increasingly persistent foreign oversight are challenging the reliance on live test events. A more coherent through-life approach to the use of the M&S is needed that is integrated with focused live events. Such as approach is potentially enabled through the use of digital engineering and increasing digital connectivity. Although such approaches have been used for many years for synthetic training, the requirements for test raise some challenging issues around model fidelity. In this paper, we explore several areas where M&S can be used to augment or reduce the reliance on live test and discuss the implications on model fidelity and the resulting implementation strategy.

2:30 p.m. **"Model Validation Levels in Digital Engineering"**
Corinne Weeks, STAT Expert, Scientific Test and Analysis Techniques Center of Excellence (STAT COE)

Digital engineering uses an integrated, model-based approach in order to speed up the acquisition process and provide capabilities to the warfighter as quickly as possible. As the Department of Defense shifts toward the digital engineering approach, it is critical that modeling and simulation results can be trusted in order to minimize the risk introduced by using models in place of physical articles. Trust is assigned to a model through validation, which determines the degree to which a model is an accurate representation of the real world from the perspective of the intended use. However, validation is often a one-time, subjective process resulting in a binary indicator of whether or not a model is valid. Digital engineering requires a new paradigm of model validation, where model validity can be continually reassessed as models change and improve over the course of the system lifecycle. Model Validation Levels (MVLs) aim to meet this need by redefining validation in terms of fidelity, referent authority, and scope, and providing an objective, rigorous validation metric which can be automated to continually assess model validity.

3:00 p.m. **"Software Validation in Cloud/DEVSECOPS/Lab Environment"**
Chad Cummings, NAVWAR Director of T&E for PEO-C4I & Project Overmatch T&E Pillar Lead

In support of getting software capability to the fleet faster, utilization of a DEVSECOPS virtualized environment is necessary. Virtualized environment

provide the necessary hosting capability to support early development and integration prior to functional hardware in the Loop (HWIL) lab based testing. Virtualized environment is required to be representative enough to support software hosting capabilities. This is a cost effective way to support development and early integration without the need to expensive HWIL test time. With the use of Test Automation and Automated analysis, development and early integration execution time can be severely reduced and manual HWIL testing will be limited to functionality testing that requires a Systems of Systems Enterprise architecture.

3:30 p.m. **"T&E Across the Virtuality/Reality Spectrum to Improve System Confidence"**

Dr. Jim Leathrum, Associate Professor and Chief Departmental Advisor in the Department of Modeling, Simulation and Visualization Engineering at Old Dominion University

Simulation-based test and evaluation (T&E) presents great opportunity to test autonomous vehicle software in the absence of a hardware platform. Testing in the virtuality/reality spectrum allows T&E to begin at a behavioral level given the development of a simulated representation of the hardware platform and to progress through a structural to physical representation. While early software testing is well known to reduce the development lifecycle, the testing of autonomous vehicle software requires a realistic representation of the environment, both the physical vehicle and the environment the vehicle operates within. This requires the development of the simulated environment to stay ahead of the physical development to be beneficial. Rapid simulation development generally requires reduced fidelity, sufficient to be useful. Insufficient fidelity can result in invalid software test results, while too much fidelity slows the development process. This talk discusses how the development of a simulation of the physical platform and the virtual environment in which it operates can be developed in parallel with system design/implementation, increasing fidelity as the lifecycle progresses. It highlights the importance of sufficient verification and validation (V&V) to ensure the simulation provides confidence in design decisions while understanding that V&V is time intensive, potentially making the simulation's relevance untimely.

Thursday 29 September, 2:00-4:00pm

Session 4 **Applying Agile Approaches to Autonomous Vehicle T&E**
Chair Dr. Robin Poston, CTEP, Dean of the Moody School of Graduate and Advanced Studies, SMU and Research Fellow STEP FedEx Institute of Technology, The University of Memphis

2:00 p.m. **"Actuating Thru Agile – An Incremental Test Approach"**
Joshua Strain & Javier Lujan, Raytheon

The Raytheon EO/IR products have been integrated on manned and unmanned aerial systems, fixed and rotary wing aircraft. Testing the electro-optical and infrared sensors, along with its stabilization system and complex mechanisms involves a variety of build and test events that are inherently risky to perform at the end of a traditional waterfall product development lifecycle. This presentation describes a novel approach to successfully applying the agile framework to implement early incremental test activities on complex hardware development to burn down risk, supply continuous feedback on the design, and ultimately lead to a smooth transition into final product verification and validation activities.

An example of a new actuating control mechanism is utilized to demonstrate the application of this agile approach to testing, showing how to develop a progressive test and demonstration plan up front to support the incremental framework. This approach enables demonstrations of product capabilities very early in the design stage to detect issues and reduce life cycle risk of rework, while also improving customer and stakeholder satisfaction. Cross-functional sprint cadences across the program are utilized as additional opportunities for cross-team learning and re-emphasizing design decisions and tradeoffs available. Finally, the presentation focuses on lessons learned around the importance of streamlining supply chain processes to support test equipment and prototype material procurement in support of the agile test plan.

2:30 p.m. **"Test Resource Management Center - Autonomy and AI Test Technologies"**
Catherine Tadlock, TRMC T&E/S&T AAIT Chief Engineer

Testers of autonomous and AI-enabled systems are consistently running into T&E S&T gaps. Their tools and methods require an intractable number of resources as these systems become more common and require continuous validation after the traditional acquisition cycle ends. Test Resource Management Center, Office of the Undersecretary of Defense, funds the development of technologies to address testing gaps. This paper will discuss the technologies in current and past development, the gaps they address, and what gaps the organization foresees testers contending with in the near future.

3:00 p.m. **"Software-Driven Test & Evaluation for Autonomous Vehicles"**
Greg Granito, Product Manager, Applied Intuition

Autonomous vehicles are rapidly moving from test labs to the real world. The most successful autonomy companies–the ones whose vehicles are driving the streets of

California and Texas–employ a similar approach to development and deployment: they use an agile, software-driven methodology that rapidly curates vehicle fleet data, runs simulated scenarios to test and tweak perception and control software algorithms, and provides real-time feedback on progress toward development, testing, and safety requirements.

By building a pipeline around this process, autonomy developers and test engineers can quickly set requirements based on safety, effectiveness and suitability of autonomous systems; understand current capabilities; and refine autonomy algorithms for more reliable performance across potential scenarios.

3:30 p.m. **"Towards Continuous Cyber Testing with Reinforcement Learning for Whole Campaign Emulation"**
Dr. Tyler Cody, Virginia Tech

Modern automated penetration testing uses rule-based procedures and model-checking concepts to search through all possible attacks on network models and identify those that violate some correctness or security property by generating an attack graph. By generating all possible attacks, modern, top-down approaches inherently do not isolate the few attacks that matter the most. This weakness is exacerbated in future network settings like 5G and Internet of Things (IoT) settings where networks are expected to have thousands of hosts (or more) and evolve over time. This has created a perception that the attack graph concept itself is inadequate, in turn hindering the automation of cyber testing. Recent research re-positions automated attack graph generation as a best practice in cyber defense by applying deep reinforcement learning (RL). While recent research into penetration testing with RL has seen a rapid growth in interest, a clear concept of operational use has not been defined. We define and provide formalism for the concept of whole campaign emulation (WCE). We present WCE as both a challenge problem and a framework for automating cyber T&E with RL. This manuscript captures an RL-oriented perspective on the past, present, and future of attack graph generation, and serves as a primer from researchers and practitioners alike. With WCE, organizations from small businesses to nation-states can feasibly institute continuous cyber T&E with low test costs and low disruption to operations.