ARTICLE

Measure and Share Project—Channeling the Power of Data, Analytics, and Visualization

Dr. Michael K. Shields, Peter Firey

Abstract

The Test and Evaluation (T&E) Science and Technology (S&T) Program within the Test Resource Management Center (TRMC) is developing tools and technologies for advancing testing and other forms of assessments to improve the Department of Defense (DoD) ability to evaluate the technical and operational performance of DoD systems and services in contested all-domain environments. This article introduces one of the S&T program's initiatives within the Cyber Test Technology Area—the Measure and Share (Me&S) project. This article details the initiative's foundational presuppositions, objectives, concept of operations, and technical areas of the research.

Foundational Presuppositions and Objectives

The authors identified five presuppositions that guided this research. First, objective science begins with measurements. Qualitative observations, while sometimes useful, are often based upon unstated assumptions and biases. Second, information and knowledge are most effective when they are most broadly available to all personnel who have a need-to-know. Analyses and decisions made without relevant information, which is otherwise known but not made available, are often not optimal. Third, using machine readable artifacts and machine analyses can improve test efficiency and effectiveness. To this end, data must be collected, stored, and indexed in a way that modern tools such as artificial intelligence (AI) and machine learning (ML) can be applied. Fourth, the cyber testing community, the operational community, the acquisition community, and the intelligence community communicate using similar terms with different meanings unique to their area of expertise. In these cases, each party believes the communication was more effective than it actually was warranting an appropriate feedback mechanism. Ineffective communication often results in bad warfighting outcomes. Fifth, the use of ontologies (which include taxonomies and precise definitions of terms) can improve machine analysis and communication.

With these presuppositions in mind, we enumerated the following objectives for Measure and Share (Me&S):

- **Objective 1**: Measure and improve the efficacy and efficiency of cyber testing and assessments including but not limited to formal and informal T&E events, training, and exercises.
- **Objective 2**: Enable the analysis of the DoD data related to cyber-dependent capabilities, vulnerabilities, threats, testing and assessment results, and operations.
- **Objective 3**: Enable more comprehensive and appropriate dissemination (sharing) of the data.
- **Objective 4**: Provide communication and feedback channels based upon well-defined ontologies and terms of communication to include, for example, informing combatant and major commands, with mission threat analyses.

The first and second objectives of Me&S will be met by providing a flexible data fabric with appropriate security controls. The data fabric is an architecture that enables seamless data integration and access across a distributed environment. The data fabric implementation must enable the use of modern data analytics techniques to provide stakeholders needed information. It must provide mechanisms to query the data and create new analytics as necessary. It must support analysis of the data from individual tests and analysis of multiple events. For example, the analytics must enable the analysis of technical and operational performance and test resource efficiencies. The Me&S tool must enable access to all publicly available information to include licensed data as well as controlled access to classified and limited distribution information related to cyber vulnerabilities, threats, adversary tactics, operational mission requirements, and testing and assessment results.

The third objective is to enable appropriate sharing of the data. That is, the data should be accessible based on the classification of the data, the user's clearance level, valid need-to-know requirements, and as allowed by licenses or subscriptions. This objective includes the dissemination of raw data and the results of analytics.

The fourth objective is to enable better information sharing and communication by using modern data science tools such as ontologies and to provide feedback channels to ensure data that is disseminated to various communities is properly defined and thus understood by those communities.

Concept of Operations

The Concept of Operations (CONOPS) for Me&S is provided in Figure 1. The objects shown in yellow are being developed as part of this research project. The objects shown in blue and gray are tools and data sources that exist outside of Me&S. This CONOPS consists of four basic processes that occur asynchronously.

The first process in the CONOPS is represented by the vertical double arrow (#1) where open-source data from public sources, operational information in the form of mission snippets, and intelligence information is imported into the data fabric. For example, an automated process monitors the National Vulnerability Database (NVD); when the NVD is updated, the process automatically packages the data, indexes it, and saves it to the data fabric. Similar automated processes monitor other public sources. In addition, the DoD could subscribe to commercial sources of vulnerability and threat data, and that data could be automatically uploaded to the fabric where it would be available to all appropriate users. In addition, data sources for mission snippets and mission requirements are also imported as they are updated and published. Finally, a process must be developed to import intelligence information.

The second process is represented by the left horizontal double arrow (#2) where the system under test (SUT) information is imported into the data fabric. In Me&S, a SUT may be an entity as large as a joint task force; a single platform such as a ship, aircraft, or ground vehicle; a component such as a Line

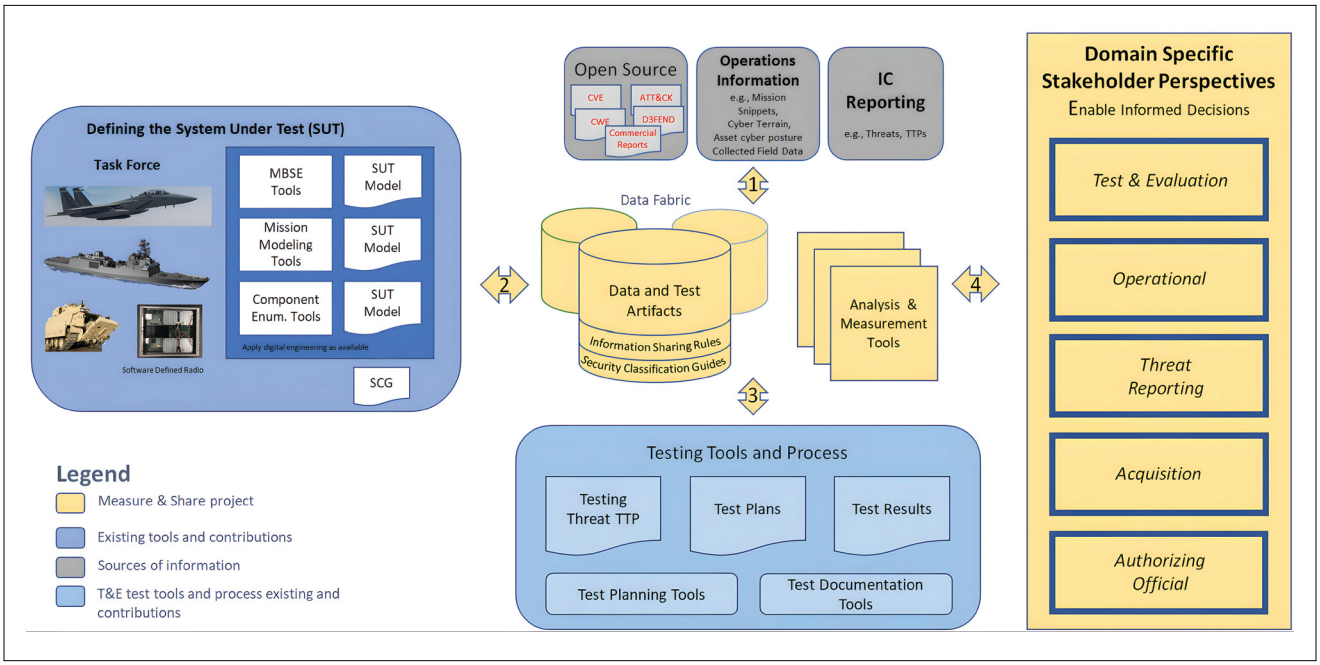

FIGURE 1. The Measure and Share (Me&S) Concept of Operations (CONOPS).

Replaceable Unit from an aircraft; or a subcomponent such as the software for a software defined radio. The goal is for the SUT to be defined by Model Based System Engineering (MBSE) tools where the models and descriptions of the SUT are in a machine-readable format (e.g., Bill of Materials) and can be ingested into the data fabric. This process uses an ontology to map the various elements in the SUT description so that they can be queried based upon multiple levels of abstraction. For example, a user could query for test results based on the specific version of a SUT (e.g., X-10 CICU v. 8.1.2), based on the SUT for all versions (e.g., X-10 CICU), or based upon a generic description (e.g., aircraft information management systems). It is recognized that several different MBSE tools are being considered by various program offices within the DoD, and TRMC intends to support the major tools. Each of these tools has its own ontology. Further, within the DoD and across commercial industries, different ontologies are employed. The Me&S program intends to implement support for using these ontologies and provide the ability to infer knowledge across the varied ontologies. In addition, there will be instances where the SUT is defined by text alone (e.g., a document). In these cases, Natural Language Processing (NLP) will be used to create a machine readable SUT definition.

The third process is represented by the bottom vertical double arrow (#3) where the personnel responsible for designing, conducting, and reporting cyber tests and assessments use the Me&S analytics to inform and enable all steps in the testing process. The test engineer can use the SUT description to identify the system boundaries for the test. For example, using the notional aircraft introduced by the Government Accountability Office (GAO) and shown in Figure 2, the test engineer could specify the boundaries of the test as the flight software system. Using the software bill of materials, the test engineer could determine that the flight software system is composed of a primary and backup flight management unit together with several sensors and user interface modules connected by three data buses as the test boundary.

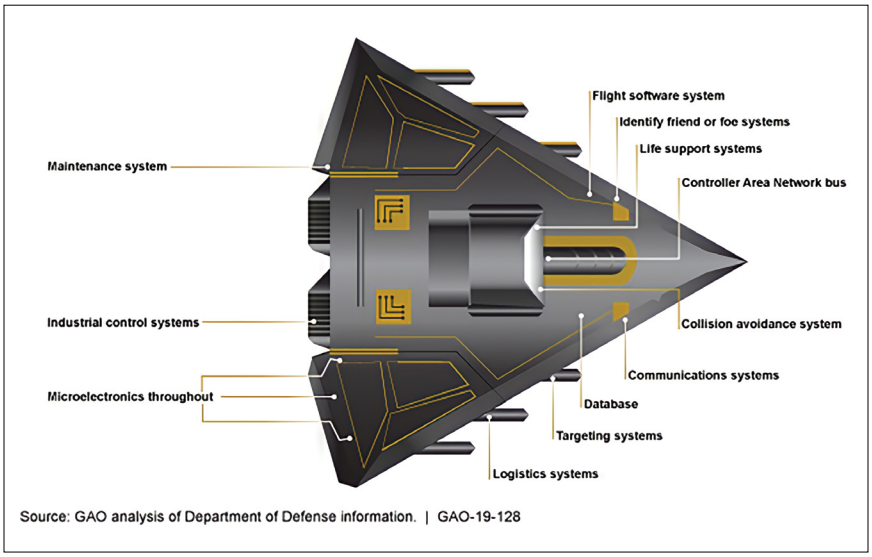

FIGURE 2. GAO Notional Aircraft

The test engineer could also use the description of the components within those boundaries to query the data fabric for open-source threat reporting and prior test artifacts using the hardware and software bill of materials for the components. Analytic techniques could use information from those queries to recommend specific tests to perform and provide priority ranking on those based upon analytics as described below. The test engineer could then use that information to develop and upload the test plan. Analytics would then provide efficacy scores for that test plan. When the test plan is executed, test results and artifacts would be uploaded to the data fabric. The intention is to provide tools to upload the machine-readable artifacts and have an analytic produce the initial draft of the test report. Artifacts could include raw data, logs, and test scripts and test vectors that could be modified and reused by others. After the test is completed, analytics could use the information uploaded to provide efficacy scores for the test that was actually executed.

The fourth process in the CONOPS is represented by the right horizontal double arrow (#4) where various stakeholders use different perspectives (dashboards, visualizations, etc.) designed for them to query the data and run analytics to provide specific information tailored to them. For example, the test engineer could use one of the *Test and Evaluation* perspectives designed for test engineers to accomplish the steps in the third process outlined above. During the Me&S research, TRMC intends to implement several example perspectives and corresponding analytics for T&E leadership, operational users, threat reporting users, program management office users, acquisition and sustainment users, and authorizing officials. The Me&S system also allows for stakeholders to design their own analytics and perspectives.

Technical Areas ofResearch

To enable this CONOPS, TRMC has identified several technical areas to research and develop tools. The first technical area is the data fabric where data are stored in a secure, distributed manner, and includes the core query process. This research will be described in detail in a future article. The remaining technical areas are summarized below.

Data Import Tools

The second technical area is data import tools—broadly divided into tools to import threat information (both open-source information and intelligence information), tools to import SUT descriptions, and tools to import test plans, test results, and test artifacts. The objective of this area is to enable analysis. The tools must be developed to import data into the data fabric where data can be analyzed. These tools must be designed to operate in existing frameworks where the data are created.

The first category of tools is for importing threat information. Most of the open-source threat information is provided in a well-organized machine-readable format. For these sources, the import tools are trivial to implement. In some cases, the information is presented in a format designed for human consumption, and NLP tools will have to be developed to process and organize the information before importing.

The second category of tools is for importing SUT descriptions. In this category, three types of tools are envisioned. In the case where the SUT is well defined with MBSE tools, this information is available in a machine-readable format and the tools simply import the data with the proper ontology to describe the data. In the case were the SUT can be described using tools that generate a bill of materials (e.g., a software bill of materials), the bill of materials can be converted to the correct format and imported. The final case for SUT import is the situation where the SUT is defined in a collection of documents in human readable formats, and tools will be implemented that use NLP to import the data.

The final category of import tools is those for importing test plans, test results, and test artifacts. Several programs exist in the DoD for creating test plans and documenting test results. Initially, the intention is to implement modules for these existing tools that export the data in a format that can be directly imported. In addition, tools will have to be designed and implemented to use NLP to extract test plan and test results from documents. Finally, a simple tool will be created to directly import and annotate test artifacts.

Analysis and Analytics

The next technical area covers the framework for performing analysis and the specific analytics. The most important feature of the Me&S architecture is the ability to analyze the saved data. Me&S is designed so that analytic queries will work with the data made available via a need-to-know process from all the connected distributed stores in the Me&S data fabric. The intention of the Me&S project is to use an existing

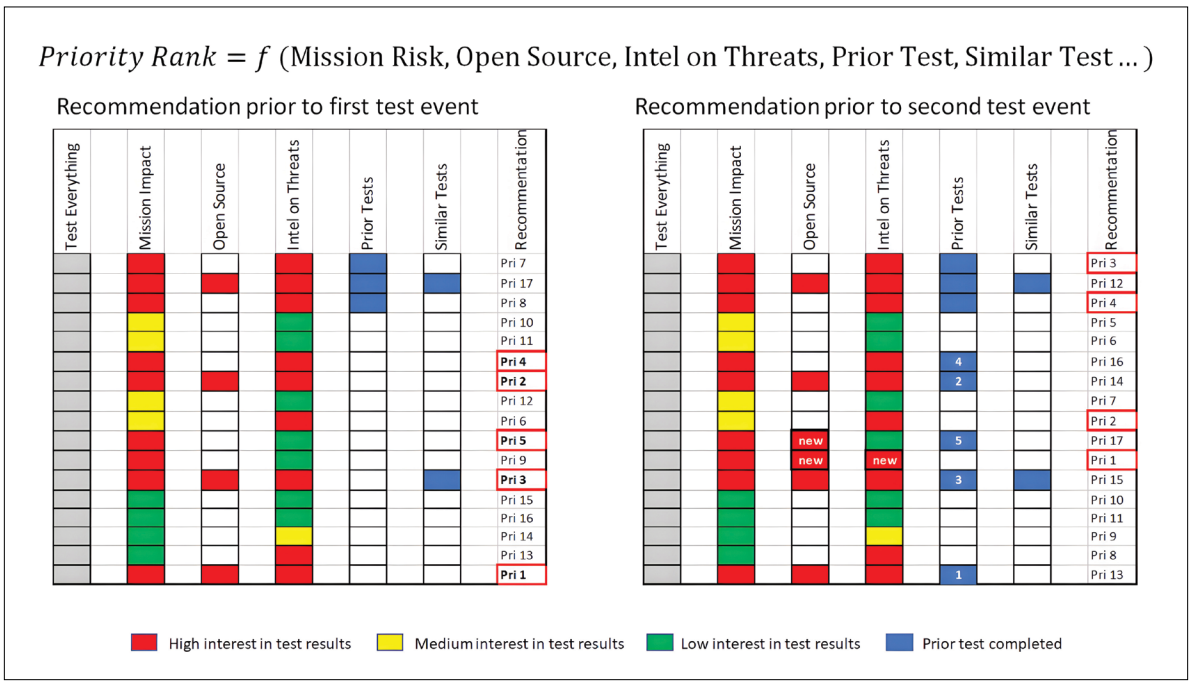

FIGURE 3. Hypothetical Example of a Test Priority Recommendation Analytic.

framework for processing data and implementing AI/ML algorithms. Our initative is currently studying these options and one or more will be selected to integrate into Me&S. The intention is for TRMC to fund the development of an initial set of analytics to process the data and provide recommendations and efficacy metrics to various stakeholders. One specific area of research will be the development of efficacy analytics. The Me&S data processing framework will use an open architecture so that these initial analytics can be used as examples and various users can implement their own analytics.

One example of a type of analytic is a recommendation for test planning. To illustrate how this analytic might work, we have a hypothetical SUT composed of several components and subcomponents. For this hypothetical SUT, a very large set of potential tests exist that could be planned and executed to test for the presence of a particular cyber vulnerability in the various components. Each row in Figure 3 is for one of those potential tests. Given unlimited resources and time, the "Test Everything" column shows all the tests one would perform. Because resources are limited, it is beneficial to rank the potential tests in a priority order. We propose that this ranking should be a function of the mission risk if an exploitable vulnerability exists, the open-source intelligence indicating the potential that the vulnerability does exist, and the intelligence regarding a potential adversary's ability to exploit the vulnerability. In addition, an assessment of what similar tests have already been performed should factor into the priority. In our hypothetical example, the particular test indicated by the first row is for a component that, if compromised, would have a high impact on the mission of the SUT (as indicated by the red cell in the second column). In addition, there is no open-source information indicating that a vulnerability (e.g., a Cyber Vulnerability and Exposures report) exists (as indicated by the white cell in that column). The presence of relevant threat intelligence is shown by the red cell, and the fact that a prior test event included a test for that vulnerability is indicated by the blue cell in the fifth column. The data analytic processes these data and provides the priority for each potential test to the engineer who is planning the test event.

In this hypothetical example, we assume that the resources are available to perform the top five priority tests. When a second test is being planned, the process is repeated as shown in the right table of the figure. Now the five tests performed from the first recommendation show up as prior tests, and there is new open source and intelligence information. Thus, a new set of priorities are recommended. The benefit of using an analytic is to prioritize those tests that may have the most impact on advancing the survivability of the SUT in contested cyberspace.

Perspectives

The final technical area of the Me&S research and development project is the perspectives. Each of the stakeholders requires a mechanism enabling them to interact with the data, including running distributed queries, running analytics on the results of those queries, and visualizing the results of the queries and analytics in a format that is designed for the needs of the user. We broadly divide the perspectives into categories based upon the broad roles of the users. These categories include *Test and Evaluation*, *Operational*, *Threat Reporting*, *Acquisition*, and *Authorizing Officials*. Within each of these categories, we anticipate developing multiple perspectives for the various roles within that category. For example, in the *Test and Evaluation* category there will be perspectives for users who create tests plans, those who execute test plans, and for the various levels of leadership from test team leads to senior DoD organizational leaders. In addition to the initial developed perspectives, templates and tools will be provided to enable users to create their own perspectives.

Conclusions

The Me&S initative is in the second year of a multi-year research and development program. The intent of this program is to significantly improve the DoD's ability to operate in contested cyberspace by improving cyber testing and assessments and making the results of those test events available to the various stakeholders who operate, manage, and sustain the DoD systems and services that our warfighters depend upon for mission success. NEJ

Acknowledgements

The authors wish to thank the many reviewers of this article for their suggestions and improvements.

AUTHOR BIOGRAPHIES

MICHAEL K. SHIELDS, PHD, is the President and CTO of Vigilant Cyber Systems and currently supports TRMC as the Chief Scientist of the T&E S&T program in the Cyber Test Technology area. He has 30 years of research and development experience, most of that in Cyber Testing. He is a Naval Academy and Naval Postgraduate School graduate and retired U. S. Navy Nuclear Submarine Officer.

PETER S. FIREY is a Senior Principal at The MITRE Corporation and currently supports TRMC's T&E S&T Cyber Test Technology area as its senior cybersecurity T&E advisor and systems engineer. He has 40 years research and development experience advancing a broad spectrum of DoD and U.S. Government information technologies to include collaboration and command and control capabilities. He has a B.S. in Computer Science from the Virginia Tech University.