Technical Report
1179

Operational Cyber Testing Recommendations Version 1

P.J. Donovan
W.N. Herlands
T.R. Hobson

2 May 2014

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS

Prepared for ASD(R&E) under Air Force Contract FA8721-05-C-0002.

This report is based on studies performed at Lincoln Laboratory, a federally funded research and development center operated by Massachusetts Institute of Technology. This work was sponsored by the Assistant Secretary of Defense for Research and Engineering, ASD(R&E), under Air Force Contract FA8721-05-C-0002.

The 66th Air Base Group Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Gary Tatungian
Administrative Contracting Officer
Enterprise Acquisition Division

Massachusetts Institute of Technology
Lincoln Laboratory

Operational Cyber Testing Recommendations
Version 1

P.J. Donovan
W.N. Herlands
Group 59

T.R. Hobson
Group 58

Technical Report 1179

2 May 2014

Lexington Massachusetts

EXECUTIVE SUMMARY

This recommendation report is based on observations made by the MIT Lincoln Laboratory (MIT LL) Cyber Measurement Campaign (CMC) team during a major Spring 2013 Pacific Command (PACOM) exercise. The goals of this report are: (1) to provide high level insight into Operational testing of cyber defensive technology, (2) to provide recommendations on Operational testing formulated from our observations and experiences, along with insightful input from the experienced PACOM team, and (3) to present a proposal overview of integrated Science and Technology (S&T) and Operational testing approach.

Operational testing associated with a major military exercise is often very limited in scope, duration, and resources, and as such is constrained in repetition of tests for statistical significance, re-running of tests with missing, corrupt or anomalous results, deep dives into technology under test, and evaluation of a broad set or metrics or analytical results. Its strength is in connection with the Operational "warfighter" community, and overall realism in test cases and environment.

Science and Technology (S&T) testing also has limitations, often in Operational connection as well as realism. Its strengths draw from a reasonable timeline and inclusion of a number of technical subject matter experts, which allow for technical deep dives, an array of measurements and metric evaluations, re-running of tests and investigations into anomalous results, and more.

Knowledge in both Operational and S&T testing allows the formulation of a number of recommendations for Operational testing that optimize the insight gained from the tests within the bounds of the stated limitations. The recommendations include developing and reviewing in detail a test plan well in advance of the test event with all relevant parties, including the test team, the developers, the Red Teams, the IT teams and other stakeholders who can be valuable participants. The recommendations also include some risk reduction efforts, like developing configurations and scripts ahead of time, verifying a number of test components in advance, checking logs and data through out the test event, and performing some on-the-fly pre-analysis during the event to ensure data is being adequately collected and testing is not going awry.

Finally, an integrated testing approach is proposed which includes Operational testing and S&T testing in multiple stages of technology evaluation, from mature prototype through deployment which allows appropriate metrics and testing methodology to mature over time and which utilizes the strengths of each type of testing to enhance the other.

ACKNOWLEDGMENTS

The PACOM team, led by Dr. Matthew Goda, Mr. Kevin Jordan and Mr. Eamon Jordan, has been gracious and insightful throughout our collaborations with them. We thank them and their team for their time, insights and shared experiences and appreciate our overall inclusion in the testing event.

Mr. Douglas E. Stetson of MIT Lincoln Laboratory and Mr. George Jones of Software Engineering Institute (SEI) at Carnegie Mellon University whose expertise augmented their thoughtful input to this document.

TABLE OF CONTENTS

1. INTRODUCTION

This recommendation report is based on observations made by the MIT Lincoln Laboratory (MIT LL) Cyber Measurement Campaign (CMC) team during the preparation, testing event and analysis phases of a Pacific Command (PACOM) major Spring 2013 exercise, as well as limited previous participation in Operational testing events. These recommendations are shaped by the Science and Technology (S&T) testing experience of the MIT LL CMC team, as well as valuable input and experience of the PACOM test team, and are limited to Operational testing of cyber defensive technology.

The goals of this report are: (1) to provide high level insight into Operational testing of cyber defensive technology, (2) to provide recommendations on Operational testing formulated from our observations and experiences, along with insightful input from the experienced PACOM team, and (3) to present a proposal overview of integrated S&T and Operational testing approach. We expect to provide updates and revisions to this content over the course of the CMC program, as experience in Operational cyber defensive technology testing is gained.

2. OVERVIEW TESTING CYBER TECHNOLOGY FOR THE WARFIGHTER

2.1 OPERATIONAL TESTING

The focus of the observed test event was on demonstrating to the warfighter impact and usefulness of proposed technologies, with a scope limited by a moderate size test team and an objective of testing three government-funded, pre-deployment technologies over a few weeks' time. Thus, our observations and recommendations are primarily bounded to this instantiation of Operational testing. For this, there is little to no screening of the technologies prior to the test event. If a technology seems a viable option to address current warfighter needs, it is selected for the event. There is no small scale assessment prior to the test event to ensure the technology actually functions, functions as advertised, and is compatible with the planned testing environment. This lack of screening is unfortunate, as the test event itself is resource intensive-including a sizable team of people with a variety of disciplines, potential participation by one or more Red Teams, significant network and computer reservation time, and more. Additionally, once a technology is inserted into a prominent testing event, it may become a sensitive issue to report poor results, for both the developers and funding sources who are looking to transition the technology as well as to the Operational test team who has spent valuable resources in attempting to demonstrate its utility.

The testing itself is all done within a set, contiguous period of time and expected to be fast paced and realistic. This leads to a number of strengths, and some limitations. A key strength of the testing is a substantial tie-in with the technical needs and expectations of the Operational community. Thus, the tests benefit from realistic test cases and threat models that are provided and vetted by the Operational community. The tests also benefit from usually running in near to real or actual Operational environments. This helps determine if the technology is compatible and robust in the intended deployment environment and the results obtained are expected to represent the effectiveness and performance of the technology well if deployed. The limitations, on the other hand, are largely a result of the fast-paced testing over a short, single period of time. These limitations include limited repetition yielding low statistical significance in results, little to no investigation of edge cases which may arise in deployment or which may provide deeper understanding of the technology, and little agility in technology or testing configuration due to not having access to particular subject matter experts during the specific set testing period when something unexpected inevitably arises.

With good planning and preparation, the strengths of operational testing can be amplified and the limitations constrained. The recommendations within this report are focused on doing exactly that by identifying critical steps that can be taken to optimize the execution of all stages of a test event.

2.2 S&T TESTING

For the purposes of this report, we will focus the S&T testing overview to a scope that is complementary to the Operational testing previously described. For S&T testing, considerable effort is put into establishing a relevant threat model and meaningful measures of effectiveness (MoEs)

(e.g. how well the technology defends against a particular threat), measures of performance (MoPs) (e.g., resource overhead associated with the implementation of the technique) and measures of safety (MoSs) (e.g., compatibility with the host system or environment). The test or experimentation period may be spread over many months, may be repetitive and may be sometimes unrealistic when diving down into thorough understanding of the technology under test. The team may use experimentation in combination with analysis and modeling and simulation, as each type of assessment has various strengths and the output of each can yield valuable input to better shape the others.

S&T testing also has strength and limitations, which contrast well with those of Operational testing. In S&T testing, there is almost always much more time available to the test team. The testing can be done in phases, and there is an opportunity to test or experiment in phases: screening of technology is possible; investigation of stress tests, outliers and edge cases is possible; repetition for statistical significance is considered necessary. Between phases, there is an opportunity to perform further research as needed and to seek out the advice or participation of additional subject matter experts. Some of the challenges of S&T testing are the lack of direct tie-in with the Operational community, which presents a risk of technology under investigation being irrelevant to the Operational user, becoming irrelevant over the lifetime of the testing and experimentation, or being evaluated against metrics developed for a poor threat model. Further, the environment is often a inferior (i.e., low-medium fidelity) representation of the Operational environment, as even a simulated environment may be lacking the nuances that are found to significantly impact the deployment of the technology.

As the Operational testing and S&T testing strengths and limitations are complimentary in their contrast, the evaluation of a technology could certainly benefit from an integrated approach. Thus, a proposal for integrated Operational and S&T testing is presented later in this report.

3. TEST PREPARATION PHASE

3.1 TEST PLAN DEVELOPMENT

Ideally, the test plan development is led by the core test team with significant participation from the technology developers, Red Teams (if applicable), representative Operational users, IT team members responsible for configuring the environment, a selection of relevant technology area subject matter experts (SME), and other applicable stakeholders to ensure testing objectives and metrics are accurate and well founded given a detailed understanding of the technology under test. This test plan development, with all of its participants, should ensure that a strong, appropriate baseline is established, testing methodology is sound, scope is sufficiently limited given the testing duration and team capacity, expectations are set well in advance, possible challenges and other complexities can be brainstormed and discussed, and so on.

A good test plan must include a realistic, viable threat model, as measuring the technology's effectiveness against this is a critical component of a successful defensive cyber technology evaluation. Well-defined threat models should be agreed upon by the test team and technology experts prior to designing test plans and should drive the planning of the testing environment and test scenarios such that the key security properties of the technology are accurately evaluated throughout the tests.

All aspects of the developing test plan need to be read and thoroughly vetted by the relevant parties beforehand, to ensure the right parties have paid particular attention to specific details. How this is done should be deliberately planned out given the specifics of the testing objectives and plan, as well as the roles of the participants. For example, the objectives of the technology, metrics and test cases should be reviewed in great detail by the technology development team to ensure the focus accurately demonstrates the key capabilities of the technology. The specific details of the planned configuration of the technology and the overall environment should be examined by the IT team as well as the developers to ensure details such as IT resources can be configured in the planned manner, and any nuances that may affect implementation and integration of the technology itself can be discussed.

The test team should aim to have the test plan in as final a draft as possible two to four weeks beforehand, in order to allow time to discuss in detail the implementation plan, the schedule and delegation of testing milestones, what additional participants should be on call for the event, and other important details that only come up when describing vocally what is intended for each of the plan's specifics. Thus, ideally the test team and other key technical participants will reserve some appropriate number of days two to three weeks ahead of the event to spend in person talking through all aspects of the plan in much detail, and identifying specific actions for all participants to accomplish leading up to and during the test event. This will also help identify where additional resources are needed, and where implementation assumptions fail.

3.2 IMPLEMENTATION

This section discusses considerations for constructing a suitable environment for executing the test plan. The test environment consists of the hardware, operating systems, software, and networking components required for testing a technology. The environment should not only be capable of accurately executing and measuring the tests defined in the plan but should also represent the meaningful aspects of the deployment environment. Further, it should be verified prior to any time-critical testing events. We offer several recommendations for realizing such an environment.

3.2.1 Test Environment Design and Configuration

Operational testing provides insight into how a technology will perform in the intended deployment environment. While a test environment will never completely match a deployment environment, it should closely approximate the environment and be capable of adequately testing key features of the technology. The test plan should include a system diagram that is vetted thoroughly by the operators, technology experts, IT team, and Red Team(s) (if applicable) to ensure the environment is capable of testing the unique aspects of the technology against the unique operational components.

Once the representative environment components have been agreed upon and noted in the diagram, testers should confirm that these components are available and suitable for testing. If no suitable components are available, the test may need to be reconsidered with respect to the objectives specified in the test plan, potentially resulting in the removal of the test. For example, if the technology provides a defense against buffer overflow attacks, one must ensure that versions of the applications selected from the deployment environment indeed possess buffer overflow vulnerabilities.

In some cases, there may be suitable alternatives for testing the key aspects of the technology. The applications may be swapped out with other applications from the deployment environment that do indeed possess buffer overflow vulnerabilities or the applications may be artificially injected with vulnerabilities. However, note it is much less meaningful to substitute non-operationally relevant or generic applications, especially if these generic applications provide an oversimplified or irrelevant test case. In this case, the results may already be known or be insignificant to the warfighter, and the overall test may not provide useful insight.

The configurations of the system should also be representative of the deployment environment and targeted on testing the key features offered by the technology. Returning to the buffer overflow example, it may be necessary to ensure that auxiliary buffer overflow defenses such as stack canaries are disabled in order to effectively isolate the actual buffer overflow defense in the technology under consideration.

3.2.2 Verification

Operational testing events are often not repeatable and are constrained by time and resources. Therefore it is critical that the environment and the collected data be accurate from the start of the event. To meet this objective, the environment and test cases should be configured and verified prior to the event. Test and measurement scripts developed in advance can be used to verify

representative pieces of the environment via sample test runs. Also, ideally, an early version of each test case can also be stood up and verified in advance (and can be developed further into the actual test cases for the event itself, rather than this being a duplicate or extraneous effort). Comparing the sample results collected from the scripts with expected results can reveal deficiencies in the data collection techniques, the reporting mechanisms, and the system components themselves. The sample results can also be used for confirming that data is collected with high enough fidelity to compute metrics but low enough fidelity to meet storage requirements. The size requirements for a full test run can be extrapolated from representative sample runs and compared against available storage space.

3.3 RISK REDUCTION

The uncertainties of new technologies present additional risk to any testing process. Unmitigated risks can result in increased costs, lost opportunities, and misguided technology selection. Increased costs can result from the need to rerun tests due to improper configuration or data collection. Lost opportunities can occur when resources for testing are constrained and the technology under test was unable to properly utilize these resources (e.g., an underutilized Red Team that is available for a limited time period). Misguided technology selection can result from improper design or execution of tests that lead to erroneous analysis and technology adoption decisions. In test plan development and test execution, it is important to continuously think of what might (with some reasonable probability) go wrong or what complexity may arise, and whether safeguards are employable within some reasonable cost. Of the many things one could consider, below is a representative list of a number of commonly seen pitfalls that should be safeguarded against or provisioned for:

- Technology does not work as expected and cannot be updated in the available timeframe
- Understanding of the benefits and weaknesses of technology is insufficient
- Test components, test cases, and test scenarios are not representative of the deployment environment or do not exercise the key features of technology
- Data collected is unsuitable for computing metrics
- Data collection is lost due to insufficient storage, hard drive failure, etc.

3.3.1 Early Preparations

There are many tasks that can be done prior to the event, rather than during the event itself. Doing so will not only free up resource time during the event, but will allow more time for debugging, handling unexpected complexities, etc. As mentioned previously, test cases can be developed mostly or entirely before the event. Determine which parts of each test case, from configuration through analysis can be configured, written or developed and do so in advance of the event. If the network or systems are available before the event, have them configured. In some cases, the test team may have access but not complete "ownership" of networks and systems before the event. In most

cases, the network devices and other systems can be configured and verified to run correctly in this configuration, and then the configuration of each device is exported. These exports can be loaded again later or during the event. This is both time saving during the event, and often more reliable as issues can be worked out without the pressure and constraints of the test.

The software that are a planned part of each test case can also be loaded and installed. Developing the test cases early allows the test team to utilize the software in ways similar or same to the planned usage for the test case, which again helps identity and correct issues or complexities beforehand.

Finally, usually test events and test cases require many scripts - to automate data collection, program and application control, data processing, logging, etc. In our experience, the vast majority of scripts can be written and tested well in advance of the test event. Although it seems data processing and analysis scripts seem that they are unneeded until after the event, we recommend these scripts also be written and tested on realistic (e.g., same format) dummy data sets. This allows for processing and running quick on-the-fly analysis on critical data sets through out the event to ensure non-corrupt data is collected during each test, and at a couple of set intervals throughout the testing days.

3.3.2 Graduated Testing Approach

In an effort to avoid these pitfalls and support better overall risk management we suggest a graduated testing approach that begins with smaller, internal tests and progresses towards larger, higher risk tests involving external parties. Such a graduated approach is designed to minimize the impact of many common pitfalls by exposing and addressing them earlier in the testing process.

Consider a new technology requiring both performance testing as well as a defensive effectiveness assessment by an external Red Team with tight time and cost constraints. The internal Operational testing team can often conduct the performance tests during the earlier stages of testing, prior to any Red Team participation. This provides the testers with the opportunity to familiarize themselves with the use and configuration of the technology as well as the expected output prior to introducing external pressures. During the Red Team assessment the testers will be better equipped to direct the Red Team to the particularly relevant areas, answer questions on the fly, and detect abnormalities in output.

The initial performance tests may also uncover previously unknown issues with the technology that need to be fixed by the technology developers prior to additional testing. It is important to ensure developer availability in order to address minor issues that may be uncovered. Depending upon the time constraints specified in the timeline, major issues with the technology would likely prevent testing from proceeding further.

A graduated testing approach can also aid in the management of timelines. Time estimates gleaned from early tests can be extrapolated to more realistically assess the timeframes defined in the test plan. As time expectations evolve, scheduling adjustments can be made earlier in the process. Allocating sufficient time between early stage, low risk tests allows for additional fixes to be made to the testing environment or the technology itself prior to the high risk tests.

3.3.3 Additional Considerations

Unexpected results uncovered during the analysis stage may demand revisiting certain tests. To facilitate revisiting these tests, the test infrastructure should be kept available as long as possible through the analysis phase after the event. In the case of unrepeatable tests, contingency tests can be considered.

Criteria could be defined in the test plan that specifies the requirements for proceeding to later stages. In addition to improving the chances of success for later stages, this also provides an opportunity to evaluate any insight that may have changed expectations about security benefits and costs such that the technology should no longer be considered and further testing should be halted.

Version control of configurations, scripts, etc., is also recommended, as it can greatly reduce duplication of efforts, reduce code triage upon introduction of a bug, allow multiple parties to make updates at once, and provide a backup of all key software components.

4. TESTING EVENT

Developing automated tests and procedures in the preceeding weeks should help the testing event itself proceed smoothly. This early preparation will ensure that minor bugs and unexpected roadblocks do not snowball into fundamental testing failures. To help mitigate problems during testing, testers must maintain constant vigilance of testing results and be prepared to dynamically modify their plans.

4.1 VIGILANCE

Throughout the test it is crucial to maintain proper awareness of the state of all running tests. The test team should routinely consult a checklist to ensure that all systems are properly functioning, especially towards the start of testing in order to catch any bugs early in the process. Logs and data, described in more detail below, should be checked at the end of every test and at least every half day to ensure data is not lost or corrupt.

4.1.1 Log Information

Logging and debugging information are ideal methods for comprehensive situational awareness. If properly designed, these logs should provide the tester with an understanding of whether the program under test is functioning correctly and the correct data is being collected. Logging information is particularly useful because it can be tagged with a timestamp. Using this information, a tester can diagnose the source of a problem and determine when the problems began.

4.1.2 Data

While logging information provides a window into the data collection, testers can also sample the data directly. While it is impractical to look at all the data being collected, a small set of data can serve as a litmus test for proper collection. Additionally, observing the data directly can help inform preemptive changes in the analysis and reporting phase.

4.1.3 Early Analysis

As described previously, data processing and analysis scripts should be written in advance of the test event. Processing a representative subset of the test data after each test twill facilitate finding data loss or corruption before they ruin an entire testing event.

4.1.4 Schedule

Comparing the current testing status to the schedule of tests allows testers to determine if they are slipping critically behind schedule. Testing schedules should have built-in buffer room for each test to ensure that minor set backs do not endanger the entire process. However, if significant delays do occur, parts of the testing should be replanned.

4.2 REPLANNING

If parts of the testing are severely behind schedule or certain tests appear not to be technically feasible, then elements of the testing procedure will have to be rescheduled. It is essential that all aspects of the testing procedure be prioritized according to their importance. Low priority tests can then be factored out during a time crunch while testers focus on conducting high priority tasks. Alternatively, replanning may be necessary if initial results warrant further investigation. In this case, new tests may be added according to their priority relative to previously scheduled tests.

4.3 DYNAMIC CHANGES

Testing invariably requires some amount of online hacking and debugging. From scripting a new test to troubleshooting faulty hardware, it is essential that testing teams have qualified personnel who are both intimately familiar with the testing procedure and experienced with the technology at hand. Remember that problems can arise at any stage in the technology, not only in the narrow functionality under test. Depending on the scope of the test, it is often advantageous to have a few engineers specifically tasked with understanding the underlying systems and prepared to triage problems quickly and effectively.

5. ANALYSIS AND REPORTING

While analysis of test results is the final step in the testing process, it is an important opportunity to verify the accuracy of the results and evaluate whether further testing is necessary. Only after such evaluation is it proper to finalize a testing report.

5.1 DATA EVALUATION

As data from the tests is processed, evaluators should look for aspects of the data which warrant further investigation.

1. Missing data. Even rigorously instrumented testing can go awry. Verifying the completeness of the final data is crucial to ensuring that no results are missing.
2. Anomalies. Unexpected deviations in the data may be indicative of testing errors or problems with the technology being evaluated.
3. Unexpected factors. The analysis stage is a time for reflection and evaluation. As the technology is better understood through the test results, unexpected relationships between seemingly benign factors may be discovered. Given these revelations, the results may have to be deemed invalid. If possible, re-tests can be executed or at the very least some simple, small scale tests can be set up and run by a small subset of the test team.

As the results of the data evaluation may create a list of tests to run (or re-run), it would be most beneficial if this process began before the testing resources are relinquished. However, due to the limited time frame and fast pace, it is likely that very little or no additional testing will be possible. Any unexplained phenomenon should be detailed in the reporting documents for future review.

5.2 REPORT

The structure and content of reports vary significantly depending on the intended audience and scope of the analysis. However, certain key features are found in all good testing reports.

1. Defined scope. Just as proper testing requires an understanding of the scope of the technology and the evaluation, proper reporting should clearly explain that scope, ensuring that the reader understands the purpose and limits of the testing.
2. Logical organization. The report should logically flow between sections providing a narrative of the technology and results which can be readily understood. It is often ideal to follow a similar organization to the (also well-organized) test plan, as the two documents are closely related and should contain similar content.
3. Relevance. A report is not only about organizing the test results, but also serves to educate others about the technology. The document should be useful to someone with no prior

knowledge of the test configuration and execution details, and with less technical knowledge than the test team. Where the focus of the testing in this report is on demonstrating the effectiveness and utility of a cyber technology to the warfighter, the warfighter is likely a key component of the intended audience.

6. PROPOSED INTEGRATED CYBER TECHNOLOGY TESTING

Operational and S&T testing are most often performed independently, resulting in duplicated effort and the absence of a valuable feedback loop between two complimentarily contrasting processes. Integration of the Operational and S&T communities throughout the entire lifecycle enables more meaningful, warfighter-relevant, and thorough testing of new cyber technologies. The following sections highlight some key input Operational test teams and S&T test teams can provide to each other, and a high level overview of the integrated approach.

6.1 OPERATIONAL INPUT TO S&T

The primary strength of this approach over disjointed testing lies in the opportunity to leverage a feedback loop across Operational and S&T testers. Some of the most meaningful input that Operational testers can provide to the S&T community includes:

Methods, Metrics, Tools, and Techniques (MMTTs) Requirements

A challenge of testing emerging cyber technology is the lack of established MMTTs. Rather than avoiding tests where appropriate MMTTs are not yet established, or limiting testing (e.g., to solely performance and not effectiveness), the Operational test team can specify MMTT requirements to the S&T test team.

Operational Properties

Operational knowledge about realistic use cases or expected mission applications for a cyber technology can guide the S&T community in developing more accurate metrics and models. Information about typical software, systems, networks and configurations can help guide the S&T team in designing their tests. For example, many S&T tests can be much more targeted by knowing whether or not source code is available in the Operational environment.

Additional Insightful Tests

In the course of testing, Operational test teams frequently discover additional areas or focuses for testing that would provide substantial insight into the technology but may be too focused to support these in an Operational capacity. Fortunately, such tests often fit well within S&T test environments.

Military Utility and Impact

While S&T testing may provide some measure of security achieved by technology, the measurement gains meaning when evaluated in terms of impact of compromising specific components or cost to overall mission. Operational testers possess domain knowledge that puts them in a better position to assess this impact. This information can be used by S&T to develop more accurate simulations, models and metrics.

6.2 S&T INPUT TO OPERATIONAL

Cyber Technology Performance and Effectiveness Limitations

S&T tests commonly assess the limits of a technology and can provide upper and lower bounds on its effectiveness and performance. Operational testers can use these bounds in structuring and calibrating their own tests.

MMTTs

MMTTs developed within the S&T community can be purposed for Operational testing. Additionally, MMTT requirements supplied by the Operational testers in early phases will drive the development of MMTTs by S&T for use in later phases of Operational testing.

General Results

Many of the tests in the S&T community are intended to be consumed by broad audiences and are often conducted using common systems, applications, and exploits. Components under test in the Operational community often have some overlap with these common components. Rather than duplicating effort, the Operational testers can build off of the S&T results and focus more on the components unique to the Operational environment.

Technology Expertise

The S&T community is typically afforded the opportunity to become familiar with new technologies early in the technology's lifecycle. This expertise can be utilized by the Operational test team in support of their learning the new technology and helping to ensure appropriate testing structure before the technology has become well understood throughout the Operational community. Additionally, the S&T team can be called upon for technology screening before great strides are made toward planning an Operational test event. Finally, the S&T team may provide on-call SME support throughout all phases of the test event.

6.3 INTEGRATED APPROACH

The proposed integrated approach is multi-phased, combining Operational and S&T aspects. The approach encompasses the period beginning with a mature prototype and ending with operational deployment, which could span several years, as conceptually illustrated in Figure 1. The MMTTs for assessing the technology will evolve over the course of testing, initially focusing on measuring the technology's fundamental effectiveness and eventually also focusing on the overall security posture impact the technology has on an operational system or network in real time. The process may begin with small scale technology screening by the S&T team to determine that the technology appears to "function as advertised." The first phase may focus on warfighter benefit, with S&T focused on assessing metrics of effectiveness based on warfighter-based threat model(s) and Operations focused on gaining warfighter buy-in and insight for the demonstrated technology. Mid-phases could be employed that lead to a final stage of testing that focuses on transitioning to deployment, such that S&T develops a mature set of metrics for real-time assessment of network or

host protection given the newly deployed technology, and Operations performs a final round of tests to ensure compatibility with the deployment environment. These latter phases may also a involve a training component so that the warfighters or end-users receiving the technology learn how to correctly utilize it. Throughout these phases, feedback can be given to the development team of ways to improve the effectiveness, performance or usability of the technology from the test teams and warfighters, and the S&T and Operational teams can collaborate with all of the valuable input and output described above.

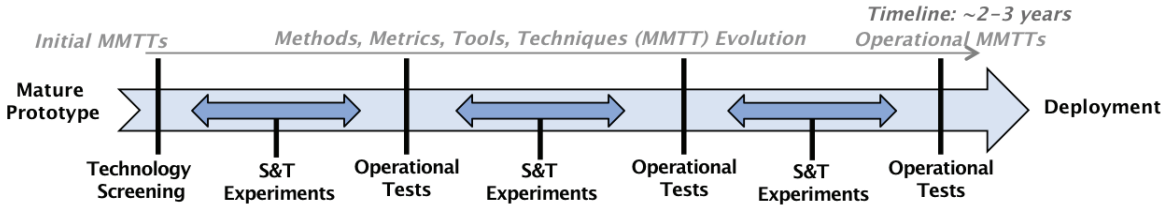

Figure 1. Conceptual timeline of proposed integrated Operational and S&T testing timeline.

7. SUMMARY

This report presents a high level overview of Operational and S&T testing and recommendations based on observations made during multiple phases of a major Spring 2013 Pacific Command (PACOM) exercise. Given the complimentary strengths of each type of testing, this report concluded with an integrated testing approach proposal which includes Operational testing and S&T testing in multiple stages of technology evaluation, from mature prototype through deployment, which allows appropriate metrics and testing methodology to mature over time and which utilizes the strengths of each type of testing to enhance the other.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
05–02–2013	Technical Report	

4. TITLE AND SUBTITLE

Operational Cyber Testing Recommendations (Version 1)

5a. CONTRACT NUMBER
FA8721-05-C-0002

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Paula J. Donovan, William N. Herlands, and Thomas R. Hobson

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

MIT Lincoln Laboratory
244 Wood Street
Lexington, MA 02420-9108

8. PERFORMING ORGANIZATION REPORT NUMBER

TR-1179

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Assistant Secretary of Defense, Research and Technology
4800 Mark Center Drive
Suite 16F09-02
Alexandria, VA 22350-3600

10. SPONSOR/MONITOR'S ACRONYM(S)
ASD(R&E)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This recommendation report is based on observations made by the MIT Lincoln Laboratory (MIT LL) Cyber Measurement Campaign (CMC) team during a major Spring 2013 Pacific Command (PACOM) exercise. The goals of this report are: (1) to provide high level insight into Operational testing of cyber defensive technology, (2) to provide recommendations on Operational testing formulated from our observations and experiences, along with insightful input from the experienced PACOM team, and (3) to present a proposal overview of integrated Science and Technology (S&T) and Operational testing approach.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified	Same as report	23	