



OPERATIONAL TEST
AND EVALUATION

OFFICE OF THE SECRETARY OF DEFENSE
1700 DEFENSE PENTAGON
WASHINGTON, DC 20301-1700

AUG 01 2014

MEMORANDUM FOR COMMANDER, ARMY TEST AND EVALUATION COMMAND
COMMANDER, AIR FORCE OPERATIONAL TEST AND
EVALUATION CENTER
COMMANDER, OPERATIONAL TEST AND EVALUATION
FORCE
DIRECTOR, MARINE CORPS OPERATIONAL TEST AND
EVALUATION ACTIVITY
COMMANDER, JOINT INTEROPERABILITY TEST COMMAND

SUBJECT: Procedures for Operational Test and Evaluation of Cybersecurity in Acquisition Programs

The cyber threat has become as real a threat to U.S. military forces as the missile, artillery, aviation, and electronic warfare threats which have been represented in operational testing for decades. Any data exchange, however brief, provides an opportunity for a determined and skilled cyber threat to monitor, interrupt, or damage information and combat systems. Real-world cyber adversaries regularly demonstrate their ability to compromise systems and inflict damage. The Department of Defense (DOD) acquisition process must deliver systems that provide secure, resilient capabilities in the expected operational environment. Operational testing must examine system performance in the presence of a realistic cyber threat.

Operational Test Agencies (OTAs) will include cyber threats among the threats to be encountered in operational testing for DOT&E oversight systems with the same rigor as other threats. The purpose of cybersecurity operational test and evaluation is to evaluate the ability of a unit equipped with a system to support assigned missions in the expected operational environment. The system is considered to encompass hardware, software, user operators, maintainers, and the training and Tactics, Techniques, and Procedures used to carry out the Concept of Operations. The operational environment includes other systems that exchange information with the system under test (system-of-systems to include the network environment), end users, administrators and cyber defenders, as well as representative cyber threats. Early involvement of programs with the operational test community is required to ensure that system requirements are measurable and testable, and that the rationale behind the requirements and the intended operational environment are understood. An adequate operational test gathers sufficient data to identify all significant vulnerabilities of a system in the operational environment to capture their effect on mission accomplishment. I will use the results of the cybersecurity testing, in part, to determine the operational effectiveness, suitability, and survivability of the system.

This memorandum, which supersedes previously published guidance that described a "six-step" process, specifies a two-phase approach for operational cybersecurity testing in



support of operational test and evaluation for DOD acquisition programs.¹ These procedures apply to all oversight information systems, weapons systems, and systems with connections to information systems, including major defense acquisition programs (MDAP), major automated information systems (MAIS), and special access programs. The requirement for operational cybersecurity testing is independent of any requirements for certification and accreditation.²

DOT&E will determine adequacy of operational testing for cybersecurity based on adherence to these procedures, and will review and approve Test and Evaluation Master Plans (TEMPs) and Operational Test Plans accordingly. OTAs are encouraged to apply these procedures to all tested systems. OTAs shall conduct test and evaluation of systems with Sensitive Compartmented Information subject to Intelligence Community Directive 503 dated September 15, 2008, following these procedures to the extent possible and providing all required test data to DOT&E.

Procedures

All oversight systems capable of sending or receiving digital information are required to conduct cybersecurity testing. This includes uploading or downloading data by physical means such as Universal Serial Bus (USB) connections or removable data devices. Any system that has a two-way data connection with a network external to the system, whether direct or indirect, is required to conduct both phases of cybersecurity operational testing described below. The level of testing required for systems that do not have a two-way connection with an external network will be determined by the OTA and approved by DOT&E on a case-by-case basis based on an examination of system architecture and network protocols. For systems with incrementally fielded capabilities or frequent software upgrades, OTAs will assess the changes with consideration of previous testing results, known vulnerabilities, developmental test data, systems architectures, and other defensive mitigations in order to conduct a risk assessment at each delivery and propose an appropriate level of cybersecurity testing to DOT&E. Data and findings from both phases must be made available to DOT&E in a timely manner.

Cooperative Vulnerability and Penetration Assessment

The purpose of this phase is to provide a comprehensive characterization of the cybersecurity status of a system in a fully operational context, and to substitute for reconnaissance activities in support of adversarial testing when necessary. This phase consists of an overt and cooperative examination of the system to identify all significant vulnerabilities and the risk of exploitation of those vulnerabilities. This operational test shall be conducted by a vulnerability assessment and penetration testing team through document reviews, physical inspection, personnel interviews, and the use of automated scanning, password tests, and

¹ DOT&E Memos, "Procedures for Operational Test and Evaluation of Information Assurance in Acquisition Programs," dated 21 January 2009, "Clarification of Procedures for Operational Test and Evaluation of Information Assurance in Acquisition Programs," dated 4 November 2010, and "Test and Evaluation of Information Assurance in Acquisition Programs," dated 1 February 2013.

² DODI 8500.01, "Cybersecurity," dated 12 March 2014, and DODI 8510.01, "Risk Management Framework (RMF) For DOD Information Technology," dated 12 March 2014.

applicable exploitation tools. The assessment should be conducted in the intended operational environment with representative operators including system/network administrators.

The minimum (core) data to be collected in this phase are identified in Attachments A and B, and include the evaluation of selected cybersecurity compliance metrics, cybersecurity vulnerabilities discovered, intrusion, privilege escalation and exploitation techniques used in penetration testing, and metrics for password strength. The assessment should consider operational implications of vulnerabilities as they affect the capability to protect system data, detect unauthorized activity, react to system compromise, and restore system capabilities. This testing may be integrated with Developmental Test and Evaluation (DT&E) activities if conducted in a realistic operational environment, and approved in advance by DOT&E. It may use data from earlier operational testing or operational testing of related systems as appropriate. OTAs should share the results from this assessment to permit the correction of deficiencies or when necessary to support a comprehensive adversarial assessment.

Adversarial Assessment

This phase assesses the ability of a unit equipped with a system to support its missions while withstanding validated and representative cyber threat activity. In addition to assessing the effect on mission execution, the OTAs shall evaluate the ability to protect the system/data, detect threat activity, react to threat activity, and restore mission capability degraded or lost due to threat activity. This test phase should be conducted by an operational test agency employing a National Security Agency certified adversarial team to act as a cyber aggressor presenting multiple cyber intrusion vectors consistent with the validated threat. The assessment should be designed to characterize the systems vulnerability as a function of an adversary's cyber experience level, relevant threat vectors, and other pertinent factors. The adversarial team should attempt to induce mission effects by fully exploiting vulnerabilities to support evaluation of operational mission risks. The adversarial assessment should include representative operators and users, local and non-local cyber network defenders (including upper tier computer network defense providers), an operational network configuration, and a representative mission with expected network traffic.

When necessary due to operational limits or security, tests may use simulations, closed environments, cyber ranges, or other validated and operationally representative tools approved by DOT&E to host cyber threat activity and demonstrate mission effects. The aggressor team may use as much data from the vulnerability and penetration assessment phase as necessary to develop and execute this assessment when insufficient opportunity exists for the adversarial team to conduct independent reconnaissance or to ensure that all critical vulnerabilities are assessed during this phase.

The minimum (core) data to be collected are specified in Attachment C, and include metrics characterizing the system protect, detect, react, and restore capabilities, as well as the mission effects induced by the cyber threat activity. A meaningful evaluation of mission effects will be system-specific, and should be expressed in terms of performance parameters already being used to assess operational effectiveness, suitability, and survivability. Mission effects could include shortfalls in the confidentiality, integrity and availability of critical mission data.

In cases where direct measurement of mission effects in the operational setting or in a simulated environment is not feasible, due to safety or operational concerns, the OTA shall propose an alternative assessment method, involving subject matter experts, by which they ascertain the effect of the vulnerabilities discovered on system performance. For enterprise systems, the evaluation must consider continuity of operations, and for systems primarily concerned with financial data, financial fraud must be evaluated alongside other mission effects.

Test Planning Requirements

Test and Evaluation Master Plan (TEMP)

Minimum standards for cybersecurity test planning in the system's TEMP are contained in Attachment D. The TEMP must define a cybersecurity test and evaluation strategy that uses relevant data from all available sources, including information security assessments, inspections, component-and subsystem-level tests, system-of-system tests, and testing in an operational environment with systems and networks operated by representative end users and network service providers. DOT&E must approve the use of test data from system development activities to support resolution of OT&E issues and measures. The TEMP should contain or provide references to enough information on network architecture, external network connections, intended operational environment, and the anticipated cyber threat to assess testing adequacy. Cybersecurity must also be integrated into the evaluation structure. The TEMP must identify resources required to execute the cybersecurity test and evaluation, including funding, responsible organizations, and threat documentation. The TEMP must identify the operational testing events where the two phases will be performed.

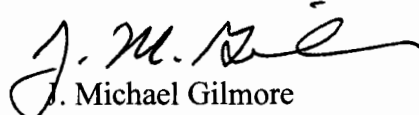
Operational Test Plan

Specific requirements for cybersecurity test planning in an operational test plan are contained in Attachment E. The test plan must contain details of how the operational test agency will execute the vulnerability and adversarial assessments, including resources, schedule, expected tools, and data to be collected. The plan must identify the environment to be used for both phases of testing, and known test limitations due to anticipated deviations from the intended operational environment. The test plan must identify the specific cyber threat(s) that the adversarial team will portray, the data to be collected during both phases of assessment, and the plan for assessing mission effects.

Test Reports

Cybersecurity test reports must adequately identify all significant system vulnerabilities and evaluate their operational impacts. Reports and data (including at a minimum the data described in Attachments A, B, and C) from the teams executing these assessments shall be made available to DOT&E and contain sufficient detail to support independent analysis of the test results, conduct, and adequacy. Distribution of test results from either assessment phase

shall also include program offices, and should include accreditation authorities when significant issues are identified which may require system modifications or retesting.


J. Michael Gilmore
Director

Attachments:

- A – Core Cybersecurity Compliance Metrics
- B – Core System Protection Data and Metrics
- C – Core Cyber Defense Performance Data and Metrics
- D – TEMP Cybersecurity Content
- E – Operational Test Plan Cybersecurity Content

cc:

Director, NSA
Director, DISA
DoD CIO
Director, Army TEO
Director, Navy Test and Evaluation and Technology Requirements (N912)
Director, T&E, HQ, USAF
Commander, U.S. CYBERCOM
Director, JCS

Attachment A: Core Cybersecurity Compliance Metrics

The metrics listed here are the minimum compliance baseline to be verified during the cooperative vulnerability assessment and penetration testing phase.

Title	Measurement	Notes
Account Management (Protect)	Accounts are established only after screening users for membership, need-to-know, and functional tasks, and disestablished promptly when they are no longer required.	NIST Special Publication 800-53 Revision 4: Control AC-2
Least Privilege (Protect)	Accesses are granted to users following the principle of least privilege.	NIST Special Publication 800-53 Revision 4: Control AC-6
Identification and Authentication (Protect)	Organizational users are uniquely identified and authenticated when accessing the system, including when using group accounts.	NIST Special Publication 800-53 Revision 4: Control IA-2
Content of Audit Records (Detect)	Audit records contain sufficient information to establish the nature, time, location, source and outcome of malicious events, as well as the identity of any individuals associated with such events.	NIST Special Publication 800-53 Revision 4: Control AU-3
Audit Review, Analysis and Reporting (Detect, React)	Audit records are reviewed and analyzed promptly for indications of inappropriate activity, and any findings are reported to the appropriate cyber defenders.	NIST Special Publication 800-53 Revision 4: Control AU-6
Continuous Monitoring (Protect, Detect)	The system is continuously monitored for vulnerabilities, to include regular assessments by cybersecurity test teams.	NIST Special Publication 800-53 Revision 4: Control CA-7
Configuration Settings (Protect)	The system is installed in accordance with an established baseline configuration following the principle of least functionality, and any deviations from this baseline are recorded.	NIST Special Publication 800-53 Revision 4: Control CM-6
Backup, Recovery and Restoration (Restore)	System data is routinely backed up and preserved, and a recovery and restoration plan for the system is provided.	NIST Special Publication 800-53 Revision 4: Controls CP-9, CP-10
Device Identification and Authentication (Protect)	The information system uniquely identifies and authenticates devices before establishing a connection.	NIST Special Publication 800-53 Revision 4: Control IA-3

Authenticator Management (Protect)	The cryptographic strength, maximum lifetime and storage methods for system authenticators (e.g., password, tokens) are compliant with organizational policy.	NIST Special Publication 800-53 Revision 4: Control IA-5
Default Authenticators (Protect)	System authenticators (e.g., password, tokens) are changed from their default settings.	NIST Special Publication 800-53 Revision 4: Control IA-5
Physical Access Control (Protect)	The information system, including data ports, is physically protected from unauthorized access appropriate to the level of classification.	NIST Special Publication 800-53 Revision 4: Controls MP-7, PE-3
Boundary Protection (Protect, Detect)	The system monitors and controls data exchanges at the external boundary and at key internal boundaries, including: <ul style="list-style-type: none"> • Firewalls or guard • IPS/IDS/HBSS¹ 	NIST Special Publication 800-53 Revision 4: Control SC-7
Secure Network Communications (Protect)	Network communications are secure and remote sessions require a secure form of authentication.	NIST Special Publication 800-53 Revision 4: Controls SC-8, SC-23
Update Management (Protect)	Security-related software and firmware updates (e.g. patches) are centrally managed and applied to all instances of the system in accordance with the relevant direction and timeliness.	NIST Special Publication 800-53 Revision 4: Control SI-2
Malicious Code Protection (Protect)	Mechanisms for preventing the deployment of malicious code (e.g., viruses, malware) are installed, configured and kept up-to-date.	NIST Special Publication 800-53 Revision 4: Control SI-3

¹ Intrusion Protection System/Intrusion Detection System/Host-Based Security System

Attachment B: Core System Protection Data and Metrics

The data and metrics listed are the minimum to be collected during the cooperative vulnerability assessment and penetration testing phase.

Title	Measurement	Notes
Vulnerabilities	Cyber vulnerabilities with descriptions and DISA severity codes ¹	<p>Descriptions shall include the nature of the vulnerability, affected subsystem(s) and implications for system protect, detect, react and restore capabilities.</p> <p>Include description of tools used.</p>
Intrusion/Privilege Escalation/Exploitation Techniques	<p>Intrusion/privilege escalation/exploitation techniques</p> <ul style="list-style-type: none"> • Specific technique employed • Starting point • Success/failure result • Time to execute • Level of difficulty (low/medium/high) <p>Starting point is the point internal or external to the system under test from which a scan or penetration attempt is initiated.</p>	<p>If technique successful, state affected system(s).</p> <p>Level of difficulty grades:</p> <ul style="list-style-type: none"> • <u>Low</u>: Technique can be executed by an actor without formal training or material support, e.g. a "script kiddie" • <u>Medium</u>: Technique can only be executed by an actor with some formal training and material support, but does not require a high level actor • <u>High</u>: Technique can only be executed by an actor with state-of-the-art training and ample material support, e.g. a nation state
Password Strength	<p>Number of passwords attempted to crack</p> <p>Number of passwords cracked</p> <p>For each cracked password:</p> <ul style="list-style-type: none"> • Privilege level • Level of difficulty required • Reason for password weakness (e.g., default password, low complexity) 	<p>Can consider the use of tokens where appropriate.</p> <p>Include description of tools used.</p>

¹ Defense Information Systems Agency (DISA) *Application Security and Development Security Technical Implementation Guide (STIG) Version 3, Release 6 (24 January 2014)*.

Attachment C: Core Cyber Defense Performance Data and Metrics

The data and metrics listed here are the minimum to be collected during the adversarial assessment phase.

Title	Measurement	Notes
Protect	Adversarial activities <ul style="list-style-type: none"> • Description • Level of difficulty (low/medium/high) • Time to execute • Success/failure 	Include starting position, nature of the technique(s) used, target system, and cyber objective (e.g. exfiltration)
Detect	Time for defenders to detect each intrusion/escalation of privilege/exploitation	For each detected event, include the means of detection (e.g., IDS alert).
React	Defense activities <ul style="list-style-type: none"> • Description • Time elapsed • Success/failure Time for defenders to mitigate each detected intrusion/escalation of privilege/exploitation White cards used ¹ <ul style="list-style-type: none"> • Description • Time issued 	Include origin of response (e.g., user, system administrator, cyber defender) and nature of response (e.g., containment, quarantine, reporting).
Restore/Continuity of Operations	Time taken to restore mission capabilities after each degradation White cards used <ul style="list-style-type: none"> • Description • Time issued 	Includes assessment of ability of typical user operators to execute procedures. Should describe restoration activities undertaken (e.g., restore from backup, failover to alternate site)
Mission Effects	Reduction in quantitative measures of mission effectiveness Where direct measurement not feasible, independent assessment of mission effects (minor, major, severe) using Subject Matter Experts (SMEs)	Should include performance parameters already being used to assess system effectiveness. Adverse effects could include specific mission-critical tasks or functions impaired and any resulting shortfalls in the confidentiality, integrity, and availability of critical mission data.

¹ A white card is a simulated event in an operational test. White cards are used when a system is too fragile or operationally critical for the adversarial team to pursue an exploitation, or when the adversarial team is unable to penetrate the system, but there is still a desire to evaluate the ability of the system to react to a penetration. White cards should be used only when necessary.

Attachment D: TEMP Cybersecurity Content

Architecture	<p>Is the architecture of the system or system-of-systems under test clearly described or is a reference provided? Description should include:</p> <ul style="list-style-type: none"> • Major subsystems • Interconnections between major subsystems (e.g., Ethernet links), external connections (e.g., NIPRNet, SIPRNet), and any physical access points (e.g., USB ports) • System and test boundaries
Operational Environment	<p>Is the operational environment of the system described? Description should include:</p> <ul style="list-style-type: none"> • End users and system/network administrators • Supported missions • Cyber defenders (local and non-local) • Cyber adversaries
Evaluation Structure	<p>Is cybersecurity integrated into the evaluation structure?</p> <ul style="list-style-type: none"> • Should encompass protect, detect, react and restore cyber defense functions • Should be in support of mission accomplishment • Should require evaluation in the presence of a realistic cyber threat
Authority to Operate	<p>Will the system have accreditation to operate prior to operational testing? If not, why not?</p>
Time and Resources	<p>Is the schedule of test events and resources described? Description should:</p> <ul style="list-style-type: none"> • Show both phases of cybersecurity testing occurring in the context of planned test events. • Identify operational users and cyber defense resources, and adequate funding for test team resources. • Identify test resources such as cyber ranges or specific tools required to conduct cyber testing.
Cooperative Vulnerability and Penetration Assessment	<p>Is a cooperative vulnerability and penetration assessment planned prior to any adversarial assessment?</p> <p>Will testing include the collection of data and metrics in accordance with Attachments A and B?</p> <p>Are the data collection methods specified? These shall include:</p> <ul style="list-style-type: none"> • Automated scanning/exploitation tools • Physical inspection • Personnel interviews • Document reviews <p>Are deviations from the operational configuration anticipated? If so, what are the implications for test adequacy?</p> <p>Will the cyber team issue a separate report and provide data before the adversarial assessment?</p>
Adversarial Assessment	<p>Is an assessment planned using an NSA-certified adversarial team?</p> <p>Is the cyber threat validated by the intelligence community?</p> <p>Will the adversarial team portray the validated threat?</p> <p>Are any restrictions or test limitations anticipated? If so, how will these be resolved (e.g., white cards, validated simulated environment)?</p> <p>Are the operational cyber defenders specified?</p> <p>Will the test plan include the collection of data and metrics in accordance with Attachment C?</p> <p>Will the test agency observe system users, cyber defenders and the adversarial team?</p> <p>Will mission effects be determined by direct measurement or by independent assessment using Subject Matter Experts (SMEs)?</p> <p>Will the adversarial team issue a separate report and provide data?</p>

Attachment E: Operational Test Plan Cybersecurity Content

TEMP Linkage	Is the proposed evaluation consistent with the approved Test and Evaluation Master Plan? If not, is the difference explained?
Architecture	Is the architecture of the system or system-of-systems under test clearly described or is a reference provided? Description should include: <ul style="list-style-type: none"> • Major subsystems • Interconnections between major subsystems (e.g., Ethernet links), external connections (e.g., NIPRNet, SIPRNet), and any physical access points (e.g., USB ports) • System and test boundaries
Operational Environment	Is the operational environment of the system described? Description should include: <ul style="list-style-type: none"> • End users and system/network administrators • Supported missions • Cyber defenders (local and non-local) • Cyber adversaries
Evaluation Structure	Is cybersecurity integrated into the evaluation structure? <ul style="list-style-type: none"> • Should encompass protect, detect, react and restore cyber defense functions • Should be in support of mission accomplishment • Should require evaluation in the presence of a realistic cyber threat
Time and Resources	Is the schedule of test events and resources described? Description should: <ul style="list-style-type: none"> • Include the dates and location for both phases of cybersecurity testing. • Identify the operational users, cyber defense resources, test articles, and test team personnel and equipment.
Cooperative Vulnerability and Penetration Assessment	Is a cooperative vulnerability and penetration assessment planned prior to any adversarial assessment? Will the cyber team review the system architecture, concept of operations, configuration, policies, and prior known vulnerabilities? Will cybersecurity compliance metrics be collected in accordance with Attachment A? Will the following data and metrics be collected in accordance with Attachment B? <ul style="list-style-type: none"> • <i>Cyber vulnerabilities with descriptions and DISA severity codes¹</i> • <i>Intrusion/privilege escalation/exploitation techniques</i> <ul style="list-style-type: none"> ○ <i>Starting point</i> ○ <i>Success/failure</i> ○ <i>Time to execute</i> ○ <i>Level of effort (novice/skilled/expert)</i> • <i>Password strength</i> Are the data collection methods specified? These shall include: <ul style="list-style-type: none"> • Automated scanning/exploitation tools • Physical inspection • Personnel interviews • Document reviews Will there be deviations from the operational configuration? If so, what are the implications for test adequacy? Will the cyber team issue a separate report and provide data before the adversarial assessment?

¹ Defense Information Systems Agency (DISA) *Application Security and Development Security Technical Implementation Guide (STIG) Version 3, Release 6 (24 January 2014)*.

<p>Adversarial Assessment</p>	<p>Is an assessment planned using an NSA-certified adversarial team?</p> <p>Is the cyber threat validated by the intelligence community?</p> <p>Will the adversarial team portray the validated threat? Consider:</p> <ul style="list-style-type: none"> • Potential attack vectors • Intent <p>Are the restrictions and test limitations specified? Will white cards or validated simulated environments be used where adversarial activity is not allowed?</p> <p>Are the operational cyber defenders specified? These should include any defenders appropriate to the scope of the test, such as:</p> <ul style="list-style-type: none"> • Local users and administrators • Command-level administrators and defenders • Tier 2 Computer Network Defense Service Providers <p>Will the following data and metrics be collected in accordance with Attachment C?</p> <ul style="list-style-type: none"> • <u>Protect:</u> <ul style="list-style-type: none"> - Adversarial activities (description, level of effort, timespan, success/failure) • <u>Detect:</u> <ul style="list-style-type: none"> - Time for defenders to detect each intrusion/escalation of privilege/exploitation • <u>React:</u> <ul style="list-style-type: none"> - White cards used (description, time issued) - Defense activities (description, time elapsed, success/failure) - Time for defenders to mitigate each detected intrusion/escalation of privilege/exploitation • <u>Restore/COOP:</u> <ul style="list-style-type: none"> - White cards used (description, time issued) - Time taken to restore mission capabilities after each degradation • <u>Mission Effects:</u> <ul style="list-style-type: none"> - Reduction in quantitative measures of mission effectiveness - Where direct measurement not feasible, independent assessment of mission effects (minor, major, severe) using Subject Matter Experts (SMEs) <p>Will the test agency observe system users, cyber defenders and the adversarial team?</p> <p>Will mission effects be determined by direct measurement or by independent assessment using Subject Matter Experts (SMEs)?</p> <p>Will the adversarial team issue a separate report and provide data?</p>
-------------------------------	---