ITEA Journal 2008; 29: 63–66

Towards Better Control of Information Assurance Assessments in Exercise Settings

David J. Aland
Wyle, Arlington, Virginia

By the adoption of certain limited techniques, the assessment of Information Assurance in both acquisition and fielded systems can achieve a higher level of rigor than available using current methods. These techniques do not replace traditional Blue/Red team activities but are used to augment them and provide a means by which replicable data may be recorded and analyzed without raising the level of risk to the exercise planner.

Key words: Exercise planning; network assessment; network penetration & exploitation; network protection; network vulnerability; risk; training.

The testing of Information Assurance (IA) in Department of Defense (DoD) information systems is addressed at numerous points throughout the life cycle of these systems, for the most part in the development and acquisition process. In 2002, a Congressional mandate added a requirement for post-fielding assessments of DoD networks. These assessments were to be accomplished during major exercises, a shared environment often familiar to the operational testing community. But this additional venue also created a challenge for both assessment and exercise planners—how to best integrate network evaluations into highly complex training events that depend upon the network that is also being evaluated. This would necessitate the integration of both the training events and assessment events, and a deeper level of synchronization between the two.

There are three key goals to such a process: (a) make the best possible use of the existing IA assessment capabilities; (b) provide meaningful and nondisruptive training in a warfare area (Information Operations) that had previously received little attention; and (c) structure events to gather meaningful observations and data regarding effectiveness of IA systems, practices, and policies. In order to accomplish this, it is necessary to design exercise events that emphasize the various aspects of IA in a manner that adds value to the training exercises and is consistent with the skills and expertise of the teams from the agencies that normally conduct DoD network assessments. This also requires IA teams to adhere in some degree to scripted events and timelines. In addition, it requires exercise planners to place greater emphasis on IA events, an area which is only now growing in prominence in most exercise scenarios. For the IA teams, this means greater constraints, and for the exercise planners, greater risk. For the operational evaluator, this could only mean many more variables in the shared testing environment.

Assessment process

Inherent to the DoD IA assessment process is the use of traditional DoD IA teams: Blue teams (technical and nontechnical vulnerability audits) and Red teams (technical adversarial penetration and exploitation tests). The missions of these teams differ, despite the common focus. The Blue team assessment most frequently consists of a collaborative review of technical and administrative support to a system or network, often including the use of scanning tools, password crackers, and low-intensity penetration tests. The goal of a Blue team assessment is to identify and document vulnerabilities caused by configuration, process, or management shortfalls. Conversely, a Red team assessment is usually a limited-duration "attack"—a network-based adversary, operating within some preset limitations, which attempts to find and exploit at least one area of vulnerability to gain internal access to a network or system. In many cases, such an attack will be accompanied by modest exploitation of that access, usually in the form of data exfiltration or modification, in order to demonstrate the operational impact of the vulnerability exploited. For these reasons, as well as others (including technical limitations, operational considerations, and resources), the Blue team activities could be described as being "a mile wide, but an inch deep," whereas the Red team activities would be "a

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
MAR 2008		**00-00-2008 to 00-00-2008**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Towards Better Control of Information Assurance Assessments in Exercise Settings	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Wyle,241 18th Street S. Suite 701,Arlington,VA,22202-3419**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**4**	

mile deep, but an inch wide." The differing focus of each team provides very different products.

Information Assurance is normally described as consisting of four fundamental tasks or principles: protect, detect, react, and restore. Due to the fundamental character of the established DoD IA controls (DoD Instruction 8500.2), the focus of most DoD IA assessments (and pre-acquisition testing) is on network protection, with limited insight or investigation into network detection, reaction, and restoration capabilities. Most Blue team events similarly focus on protection, with some view to detection. Red team assessments also focus on protection (through penetration and exploitation events) but can allow greater assessment of the other three tasks, if structured to do so. However, because of the limitations most often imposed on Red team events (whether technical, operational, or resource), the detect, react, and restore functions are not often examined in any depth, nor in a reproducible fashion. The "traditional" modus operandi of most Red teams is to find and exploit a single vulnerability, making comparison of one event to another relatively difficult, with only a few common characteristics. Employment of wider testing can significantly expand the cost, in both time and resources, of any given Red team event, making such an expansion typically impractical. Furthermore, such an expansion may be contrary to the interests of the exercise planner, as they may increase risk to other training objectives.

Overcoming obstacles

The agencies that sponsor Blue and Red teams are experiencing a growing demand for their services, as the number of critical mission functions migrating into automated information systems grows. Working within limited budgets, and facing a long lead-time for the development, training, and employment of skilled operators, the Blue and Red teams cannot practically expand the scope of their assessments without having to reduce the quantity of assessments they can perform. Given the limits in funding and manpower, one possible solution would be to establish means by which these assessments can provide greater depth of assessment without requiring additional time, personnel, or other resources.

For the "customer"—that is, the unit being assessed or sponsoring the assessment—a very robust IA assessment can potentially derail other testing or training objectives, and for that reason, most Blue and Red teams must operate within a series of constraints or written ground rules established in advance of the event. These ground rules serve to protect critical training events from disruption and yet create de facto limits on the scope and quality of the IA assessment. Most unit commanders would be reluctant to expand the scope of IA events without some form of assurance that critical functions or events would not be impaired.

For both reasons, the assessment planner is faced with limitations that all too often render the assessment findings for any one event essentially unique—a product of the variable selection of limitations imposed by both the assessment agency and the one being assessed. In order to widen the available data for analysis, trending, and long-term issue identification, the evaluator working in this shared environment requires a better form of controlled metrics and conditions but often has the least influence over the environment itself. From an operational test and evaluation standpoint, this is a considerable obstacle: conducting an assessment in an environment that is not controlled by the assessor, using resources that are, to a greater or lesser extent, also not controlled by the assessor.

A better way

The needs of all three stakeholders—the Blue/Red teams, the assessed unit, and the operational assessor—can be met by the application of a common solution: establishment of a set of core events that are more closely controlled but do not raise the cost of conducting an assessment, and that do not increase risk but do improve the consistency of the data gathered.

In order to do so, these events must: (a) leverage tasks already being performed (or that can be performed) by the Blue and Red teams; (b) maintain or decrease the level of risk currently available through existing limitations; and (c) be sufficiently consistent that they can be performed repeatedly, and in the same manner, during a variety of assessments of systems, networks, and locations. This may require all three parties to make adjustments to their current processes, but these adjustments are relatively small, particularly in view of the gains to be realized.

The implementation of more controlled test events must make use of the highly developed skills of Blue and Red teams in achieving system or network penetration, and exploiting those penetrations; demonstrate the operational/training risk such penetrations and exploitations produce without actually incurring any significant risk; and provide a consistent set of tests that can be repeated and compared in subsequent evaluations and assessments. The main attribute in achieving all three goals is control.

Such control can be achieved in a number of ways: (a) by establishing alternative, but equally fixed, boundaries for test events; (b) by conducting tests against non-operational entities; (c) by applying precise

amounts of force/stimulation during tests; (d) by segregating tests into discrete events or phases; or (e) by limited automation.

Examples of the kinds of controlled test events that might meet these conditions include

- Mission-focused assessments (alternative limits). Assessment plans are designed around one or more specific mission areas and are limited to impacting those missions and the network components supporting the missions designated. For the purposes of an IA assessment, risk would be limited largely to the system or systems targeted, and the assessment focuses on determining the impact to the designated mission supported by the targeted systems. This method would also allow extrapolation from prior acquisition testing into the broader testing of systems in their intended operational environments while limiting "spillover" effects into other systems or portions of the network.
- Repetitive vector assessments (alternative limits/precise force/segregation). Assessment team activities are organized as a series of repeated events, with each event specifically focused on testing a discrete segment of a system/network, or functional attribute. Such events can be conducted as multiple attacks along a limited set of identified attack vectors (authentication, known vulnerabilities, etc.) to statistically determine the rate of success and/or failure, as well as root causes. They can also be conducted as a series of events constructed to be increasingly detectable over time to statistically determine thresholds of sensitivity.
- Automated test events (alternative limits/automation). These events would be a controlled series of indicators (which may not necessarily require the services of either a Blue or Red team) that replicate the symptoms of abnormal network activity, internal traffic loading, or data-exfiltration. These would be used to evaluate network team responses and detection capability. Such automated events would be useful in accomplishing repetitive vector assessments as well as proxy target events.
- Proxy target events (alternative limits/non-operational). Assessment teams focus on locating and exfiltrating target files specifically placed at critical network locations as a means of determining depth of penetration, potential mission impact (without actually disrupting operations), attack pathways, and effectiveness of specific defense and detection devices ("Capture the Flag"). Alternatively, essentially harmless target files (or limited purpose macros constructed to replicate unauthorized activities) can be planted at critical network locations as a means of determining the ability of the network management and defense systems/personnel in detecting and reacting to these activities ("Scavenger Hunt").
- Adversary Level-of-Effort Metrics (alternative limits/precise force). If the level of effort expended by a Red team is one de facto measurement of the level of network protection, detection, and reaction (just as the level of force applied in kinetic testing is a de facto measurement of material strength), then the need to more precisely measure and express the level of effort brought to bear against the network or system is essential to scoping an assessment and analyzing the results. These metrics would include observation of success/failure along selected Red team attack vectors, time expended, manpower/tool levels, and possibly time-sensitivity factors (i.e., Was a successful attack achieved within a critical time-span?).
- Test Range events (non-operational/segregation). While the best method for observing risk to operational networks is to conduct tests on the operating network, one method for reducing actual risk to those networks is to conduct discrete or high-intensity tests on a simulacrum—a similarly configured test network that does not convey risk to actual network components or systems. While this type of test is more akin to laboratory testing than to live system testing, the use of a test network (and, potentially, simulations or models) allows the assessment of specific issues that would otherwise induce unacceptable degrees of risk to operating and operational systems and networks.
- Casualty testing (non-operational/precise force/segregation). One of the most critical IA precepts is the ability to reconfigure or restore a system following a casualty, system attack, or other debilitating event. The very nature of such events causes most network owners to shun such testing. The risk incurred in "bringing down" any portion of the network, however, can be ameliorated by inducing the casualties in a very limited scope (specific systems, specific durations, specific network segments) and observing the subsequent actions.

Conclusion

Implementation of some, or all, of these types of assessment/test events can meet the goals of all three stakeholders in the IA assessment process: (a) they are intended to provide a baseline for Blue and Red team activities, but only a baseline—they do not replace the existing skills and techniques employed by these teams, nor do they represent any significant expansion to their

tasks; (b) they serve to increase the degree of control and decrease the risk present in conducting such assessments in operational environments, while preserving the most critical attributes of those environments in the scope of the assessment; and (c) they provide a standardized basis by which multiple assessments can be compared, either of the same system, or of same/similar networks and environments.

Each of the three major stakeholders must accept some change to the way they currently conduct these assessments. For the Blue and Red teams, it means incorporating a more scripted structure to the often more freely executed penetration and exploitation efforts, but it does not replace the element of "free-play" in the assessment. All of the tasks described above are within the current scope of skills and expertise for these teams and should not require additional personnel, time, or significant resources. For the exercise planner, it means incorporating more aggressive events into the exercise structure, but it also means a significant reduction in the risk represented by those events. For the operational evaluator, it means developing more specific assessment plans, but it also means a greater return in terms of observations and replicable data.

For each of the stakeholders, the greatest obstacle to implementing such an approach may be essentially cultural. It will require IA teams to think like exercise planners, assessment planners to think like IA teams, and exercise planners to think like operational testers. In the end, however, all three are likely to find that the final product of the assessment/exercise event is a better view to how well DoD networks are performing. ❑

DAVID J. ALAND is an employee of Wyle, supporting the Office of the Secretary of Defense, Operational Test and Evaluation Directorate (OSD DOT&E) in the assessment of Information Assurance and Interoperability during major DoD exercises. He is a graduate of the U.S. Naval Academy and U.S. Naval War College, and a retired Naval officer with prior experience as Sixth Fleet Communications and Information Systems Officer (N6) and as deputy to the Navy Chief Information Officer. E-mail: david.aland@wylelabs.com