Security Test and Evaluation for Multilevel-Mode Accreditation: Lessons Learned

Deborah J. Bodeau, Nancy R. Brusil, Irene N. Chang, Mary Jo Reece

The MITRE Corporation
202 Burlington Road, Bedford, MA 01730

Abstract

In this paper, we describe MITRE's experience and lessons learned in performing Security Test and Evaluation (ST&E) for SPADOC 4B. ST&E drew on contractual testing but involved extensive government testing of the system. ST&E included testing in a representative stressed environment and in failure situations. ST&E also involved verification of safeguards from security disciplines other than COMPUSEC, including procedural and physical security. SPADOC 4B has been accredited to operate in multilevel mode.

1: Introduction

ST&E is required as part of the overall risk management program for a computer system by Air Force Regulation (AFR) 205-16 [1]; similar regulations apply for other services and agencies. As defined in this regulation, the purpose of ST&E is to determine whether security measures are in place which reduce the risk of operating the system to an acceptable level. Specific ST&E objectives are to identify physical, procedural, administrative, communications security (COMSEC), TEMPEST, and system security measures; to verify that those security measures are used properly; and to identify areas in which improvement is needed. ST&E thus includes not only functional testing of system security features but also verification that safeguards from other security disciplines have been properly implemented.

The ST&E Plan and the ST&E Report were major components of the SPADOC 4B accreditation package. Other components were (1) a description of the system, its operational environment, and security safeguards, (2) a risk analysis, and (3) a plan for identifying, tracking, and resolving security-related discrepancies. The accreditation package was presented in the format indicated by the North American Aerospace Defense Command (NORAD)/Space Command Supplement to AFR 205-16 [2].

This work was supported by the U.S. Air Force.

2: Background

In this section, we describe the Space Defense Operations Center (SPADOC) system mission and the government's strategy for acquiring that system. For more information on the acquisition strategy and on lessons-learned from the requirements definition, design, and development phases of the SPADOC 4B program, see [3].

2.1: System mission

SPADOC 4 is the name of a system currently being acquired by the Air Force for use by U.S. Space Command. It supports two missions: space surveillance and space defense. The space surveillance mission involves maintaining a catalog of orbital parameters (orbital element sets) for all the man-made objects orbiting the earth, performing astrodynamical calculations for various users (primarily military but also research, educational, and commercial), and predicting breakup and decay of aging satellites. The results of these calculations are then output in response to specific or standing requests by users to support their various space missions. The space defense mission operation involves assessing foreign activities, identifying threats to space assets, and warning of an attack on space assets. The conduct of space defense operations depends on precise knowledge of the position of all space assets, both friendly and hostile. As a result, the catalog maintenance functions are critical to the performance of space defense mission operations.
2.2 Development approach

The SPADOC 4 program is the fourth stage in the development of the Space Defense Operations Center. Prior phases involved heavy reliance on a mixture of manual operations and processing done on a number of unconnected systems. Acquisition of SPADOC 4 is being done in an evolutionary manner in three phases: SPADOC 4A, 4B and 4C. SPADOC 4A achieved initial operational capability in April 1989.

SPADOC 4B achieved initial operational capability in August 1991. It has been accredited to operate in multilevel mode. SPADOC 4C is being acquired in two versions. No new security requirements will apply to SPADOC 4C; however, the existing accreditation must be maintained as the system is upgraded to include new mission functionality and interfaces.

2.3: Security concept of operations

SPADOC 4A operated in the System High-Secret security mode of operations: all users were cleared for all data that SPADOC 4A handles, but not all users had need-to-know for all data (thus requiring discretionary access controls); all interfaces were to Secret systems, which handle products from SPADOC 4A as Secret. SPADOC 4A handled data classified no higher than Secret; no categories or markings applied to data handled by SPADOC 4A.

SPADOC 4B enhances SPADOC 4A functionality and performs catalog maintenance/catalog products functions on a subset of the space catalog. SPADOC 4B interfaces with a larger collection of systems than SPADOC 4A, including some systems at Secret facilities which will distribute hardcopy outputs to unclesared individuals based on the labels provided by SPADOC 4B. It handles data classified no higher than Secret; markings (known as Special Handling Designators, or SHDs) apply to some of the data it handles. Some users are Canadian military personnel not formally authorized for some data on the system. SPADOC 4B therefore operates in the multilevel security mode of operations, as defined in DODD 5200.28 [4].

The security concept of operations for SPADOC 4C is basically unchanged from that of SPADOC 4B. However, since SPADOC 4C will replace the system that currently performs space surveillance, it will interface to a larger number of systems than SPADOC 4B. All digital communications between SPADOC 4B and other computer systems are through a single system providing network services, the Communications Segment Subsystem (CSS).

2.4: System overview

The operational SPADOC 4B system is located at Cheyenne Mountain Air Force Base (CMAFB) outside Colorado Springs. It has as its primary processors two IBM 3090-200J mainframes, one prime and one backup. The backup is also used for exercise and training. The IBM mainframes are linked together by HYPERchannel buses, with the host bus on the CSS side and display buses on the user console side. MicroVAX 1Is provide an interface between the display bus and the graphic display devices (Megateks). A PDP-11/84 (the Communications Interface Adapter, or CIA) provides an interface between the host bus and the CSS. Storage media consist of disks and magnetic tapes.

Commercial off-the-shelf (COTS) software includes the IBM operating system and related utilities (MVS, RACF, JES2, VTAM, RMF, TSO-E, DFP, and RTX-E), NETEX software for communications over the HYPERchannel, DEC's RSX-11M+ operating system for the CIA, VMS for the MicroVAXes, and the ADABAS database management system. Developed software includes operating system augmentations, the operator-system interface for users, augmentations to ADABAS, and mission applications.

A copy of the SPADOC 4B system is located at the SPADOC Off-Site Test Facility (SOSTF) at Peterson Air Force Base. This system is in a slightly different configuration in that it consists of an IBM 3081 and a 3090 as opposed to two 3090s for the CMAFB system. The SOSTF system is used for training and maintenance.

3: ST&E process

In this section, we describe how the SPADOC 4B ST&E effort was organized. We first identify sources of guidance. We describe our approach to the overall ST&E, which included not only system testing but also verification of the presence and effectiveness of physical, administrative, and procedural safeguards. We then discuss how we structured our testing of the SPADOC 4B system security features.

3.1: Sources of guidance

For systems under the approval authority of AFSPACECOM/LK, the LKXS ST&E Guideline [5] applies. This document provides high-level guidance on how to perform ST&E. It is intended primarily for an operational system rather than for a system prior to initial operational capability. It identifies considerations to be addressed during ST&E planning (e.g., schedule, impacts on connected systems) and
makes recommendations for participants in the ST&E test team.

The LKXS ST&E Guideline includes twelve checklists in which specific types of safeguards are identified:

1. Procedural safeguards related to hardware utilization;
2. Physical, procedural, and system safeguards related to remote terminals;
3. Procedural safeguards to control access to systems and data files;
4. Procedural and system safeguards to protect the integrity of the operating system and the resources it controls;
5. Administrative and procedural safeguards related to personnel;
6. Administrative, procedural, and physical safeguards to restrict physical access to system resources;
7. Procedural and system safeguards to ensure the integrity of application programs and data;
8. Administrative and procedural safeguards to protect ADP products;
9. Administrative and procedural safeguards to protect magnetic media;
10. Procedural and technical safeguards related to erasure and declassification of magnetic media;
11. Procedural safeguards related to protection of data files; and
12. Procedural and physical safeguards related to the facility in which the system resides.

We were able to use these checklists to gather information from the Air Force Space Command (AFSPACECOM) and United States Space Command (USSPACECOM) organizations responsible for implementing nonsystem safeguards.

The Air Force Cryptologic Support Center (AFCS) has drafted an Air Force System Security Instruction on ST&E [6]. This document provides more detailed guidance than does the LKXS Guideline. It includes examples of report formats and schedules, specific duties of ST&E team members, and sample ST&E plans, worksheets, and reports. We found these helpful.

3.2: Overall approach

Our overall test approach was initially described in the SPADOCS 4B Security Certification and Accreditation Plan (SCAP). The SCAP is a living document. Initially, it defined the goals of certification and accreditation (C&A) activities, the relation of these activities to programmatic milestones, organizational responsibilities, and the basic C&A strategy. This strategy addressed the extent to which COTS features would be tested and the techniques for verifying compliance with system security requirements (testing, inspection, and analysis). The SCAP was updated periodically to add further details, particularly with respect to ST&E. Eventually, we extracted the information about ST&E and put it in a separate ST&E Plan. In the ST&E Plan, we identified the scope of the ST&E effort, the goals of system ST&E, the constraints that applied to system ST&E, and the phases of system ST&E.

3.2.1: Scope. Frequently, there is confusion about the scope of an ST&E effort: how much of the operational environment is addressed? We found it useful to distinguish between ST&E as a whole and system ST&E. The scope of SPADOCS 4B ST&E as a whole included safeguards from all security disciplines. However, our efforts focused on system ST&E.

For nonsystem safeguards, we identified the responsible organizations in the accreditation plan. The accreditation plan was coordinated with those organizations so that they had accepted responsibility for participating in ST&E to the extent of completing the LKXS checklists. Physical, procedural, administrative, and TEMPEST safeguards for CMAFB as a whole were certified by the organizations responsible for those safeguards. Since ST&E supported interim accreditation for Initial Operational Test and Evaluation (IOT&E), we also asked the test organization to certify procedural safeguards specific to the test environment.

3.2.2: Goals of System ST&E. The goals of system ST&E were as follows:

1. Certify the system against its contractual requirements.

Because the contractual requirements had been validated by the Designated Approving Authority (DAA), determining how well the system met those requirements was a fundamental part of ST&E. As we discuss below, this goal was met primarily by contractual testing (Developmental Test and Evaluation, or DT&E), but some independent testing was also needed. For DT&E, we reviewed the contractor's test procedures, resulting in numerous revisions and expansions, and we witnessed the security-relevant parts of DT&E (i.e., those test cases to which security requirements were allocated). In witnessing the tests, our goal was to ensure that the actual system behavior was as expected and to look for anomalies; we analyzed the test results and the audit trail. Contractual verification involved significant analysis of design specification and verification documentation: a formal model, a Descriptive Top-Level Specification (DTLS), and
Determine the extent to which the system complies with the Trusted Computer System Evaluation Criteria (TCSEC).

Since the TCSEC [7] is a DOD standard, the DAA wanted to know how well the system met its criteria. We mapped TCSEC criteria statements to the functional and assurance requirements in the contract. This mapping was simplified by the fact that the contractual requirements were tailored from the B2 criteria with the concurrence of the DAA. Some TCSEC requirements were interpreted; some requirements were stated at the B1 level; some were stated at the B3 level; and some requirements specific to the mission and/or the target operational environment were added to the TCSEC-based requirements.

Identify potential vulnerabilities and assess how easy they are to exploit.

We initially focused on potential vulnerabilities that could be exploited by ordinary users or maintenance personnel. To do this, we tested the system, using a broader range of inputs to security-critical functions than were tested during DT&E, in a range of representative stressed environments and by inducing failures of hardware and software components (and the resulting restart/recovery activities) to determine whether such failures could result in security problems. We later expanded our scope to vulnerabilities that could be exploited by trusted users; see below.

3.2.3: Constraints. The constraints on system ST&E were due primarily to how the test system differed from the operational system:

1. External Connectivity. During ST&E, we used a test driver system to inject test messages into SPADOC 4B. There was no connectivity to operational systems.

2. Location. The terminals were located in a test area separate from the operational environment. This prevented testing from interfering with normal operations.

3. User Profiles. Special user profiles were set up to facilitate testing.

4. Use of Batch Jobs. Batch jobs were used to change user security levels and to perform other functions that otherwise would require a large number of operations from interactive panels. Operationally, the use of batch jobs is expected to be strictly limited; the activities which required their use during testing are not expected to be part of normal operations.

SPADOC 4B testing at CMAFB did not allow the use of test drivers (software specially written to exercise specific system features) or utility software which could dump locations in memory or on disk, since such software could potentially bypass system security and will not be allowed on the operational system. As a result, testing was at the user interface.

3.2.4: Test phases. System ST&E was initially planned for two phases:

1. Basic Functionality. In Phase 1, we tested system security features without an operational load on the system. This phase was intended to provide government verification of the results of DT&E and to provide a broader range of expected and unexpected inputs. This phase lasted three days.

2. Stressed Environment. In Phase 2, we tested system security features in representative stressed environments. We also induced a variety of hardware and software failures and tested the restart/recovery mechanisms. Phase 2 lasted four days.

Both these phases were based on the assumption that system security administration personnel were trusted. After Phase 2 testing, the DAA added a requirement to identify vulnerabilities associated with security administrators: the System Security Officer (SSO) and Database Administrator (DBA). We therefore added a third phase of system ST&E; Phase 3 lasted three days.

A day of testing consisted of an eight-hour shift plus some additional time for data reduction. Additional time was required for off-line data analysis and report generation.

3.3: ST&E experience

In the following paragraphs, we describe our experience with SPADOC 4B ST&E. We discuss how the ST&E schedule related to the program schedule, the relation between ST&E and DT&E, the makeup of the ST&E team, sources of information for developing test procedures, and test results.

3.3.1: Schedule. SPADOC 4B in CMAFB received interim approval to operate in restricted
multilevel mode in December 1989 to allow DT&E and dry runs for IOT&E. Restrictions included precluding connections to external systems and allowing only U.S. personnel on the system. The DAA required the Program Office to request reapproval prior to IOT&E, since IOT&E involved connecting to external systems, and again prior to trials period, since Canadian personnel participated in that period of operational verification.

System Test took place from August through October 1990. When the contractor identified a discrepancy between the specified and actual system behavior during software integration or DT&E dry runs, a contractor-internal discrepancy report (DR) was generated. DRs were tracked and closed by the contractor. When the government identified a discrepancy during DT&E, a discrepancy notice (DN) was generated. DNvs were tracked jointly by the contractor and the government and were closed only by the government, based on regression testing or analysis.

Concurrently with DN closure, we performed the first phase of system ST&E (30 October-1 November 1990). We scheduled ST&E for this time to ensure that the contractor would have time to fix any discrepancies we identified in the nonstressed environment. Several security-related DNs were written as a result.

A major upgrade to the mission functionality, integrating the functions added as part of SPADOC 4A maintenance, was made after most DNs were closed. We deferred the second phase of system ST&E until after the "A-to-B roll" so that the stressed environments in our tests could be more representative of the operational and IOT&E environments. During Phase 2 (14-17 January 1991), we tested the security features while running scenarios developed to verify quantitative performance requirements (QPRs) and while performing failovers from the prime to the backup processor. Our test schedule was tightly constrained; ST&E had to be completed and a draft of the ST&E report had to be available to the DAA forty-five days prior to the accreditation briefing at which the Program Office requested interim approval in multilevel mode for IOT&E. We performed a third phase of system ST&E in response to a new requirement which the DAA imposed at the first of the accreditation background briefings on 17 January. During Phase 3 (29-31 January 1991), we used information about potential vulnerabilities to try to modify security-relevant software and tables, using the privileges associated with system security administrators.

3.3.2: Relation to DT&E. For some security features, a large number of tests could potentially be performed to verify correct behavior. DT&E included enough tests to provide contractual verification of the requirements these features were intended to meet. However, our ST&E included more exhaustive testing to detect possible inconsistencies in the implementation. For example, a subset of all possible security labels was used during DT&E to verify labeling capabilities; during Phase 1, we used all possible security labels. We uncovered a minor coding error that caused a user's directory listing to show a file labeled with two categories as having only one category; this error was not revealed during DT&E.

Similarly, DT&E included a number of induced hardware and software component failures and verification that the security features worked properly when a failure occurs and after restart/recovery. However, this was a subset of all possible failures, and we identified a number of other types of component failures which could be of concern. In particular, we wanted to verify that the display devices were properly sanitized when disconnected or when a component failure occurs. We included restart/recovery testing in Phase 2.

In addition, DT&E was organized according to the types of requirements being tested. Thus, for example, security requirements were tested primarily in Volume II of the System Test Procedures, while tests of exercise/training functionality were allocated to Volume VIII and QPRs were tested in Volume XXII. This allowed the DT&E test teams to focus on specialized areas but meant that many requirements were not tested together to investigate interactions between different functional areas.

Because loading could potentially have an impact on the correctness of a system's security features, we felt it was important to test the security features in a stressed environment by running one of the QPR scenarios as we tested security functionality. In addition, SPADOC 4B has the capability to perform live mission processing on one mainframe and exercise/training on the other. We felt it was important to verify that the actions of users running in exercise/training mode, who as trainees might make more security-related errors than users on the live processor, could not result in disclosure or confusion between live and exercise data.

Similarly, SPADOC 4B has the capability to be partitioned into two systems (sharing disk storage and the communications media) so that normal operations can be performed on one system (the "live side") while a new software or database release is tested on the other (the "test side"). The security implications of "partitioned mode" operations were not addressed during DT&E. We felt it was important to verify that the contractor's concept of operations for partitioned mode
would preclude risks that changes on the test side could cause disclosure, denial-of-service, or resource corruption on the live side. (For example, we wanted to verify that a tester would not have access to the live database.) As a result of our testing, inadequacies in the contractor's concept of operations were identified, and the plans for partitioning were improved.

3.3.3: Staffing. The system ST&E team consisted of the MITRE staff who planned and executed the tests, additional MITRE staff to support the QPR scenarios during Phase 2, contractor support, the Program Office Test Director, and AFSPACECOM observers. Only a small subset of the overall test team (the MITRE ST&E team and part of the contractor support team) was there for all testing; others provided part-time support on an as-needed basis.

The MITRE ST&E team provided a mixture of technical and other specialties. It included an overall security coordinator who had worked on the SPADOC project for several years, a former database administrator knowledgeable about the complex SPADOC 4B database, a staff member expert in the COTS operating system and utilities and in restart/recovery, and a project manager with a strong security background who ensured that the necessary resources were available when needed. All members had extensive knowledge of the SPADOC 4B design and of key components of the implementation, all had witnessed part of DT&E, and all had had some interactions with users and/or system administrators to clarify portions of the concept of operations. This background was key to the effectiveness of our test program. Our database and COTS specialists did not have security backgrounds prior to DT&E; their participation in dry runs of DT&E test cases to which security requirements were allocated and in test planning discussions provided the training needed to get an overall security perspective.

During Phase 2 testing, we were joined by MITRE staff with a deep knowledge of the mission applications and target operational environment, who executed the QPR scenarios. It is worth noting that MITRE identified several nonsecurity-related discrepancies (in the database, COTS, user interface, and mission applications areas) during ST&E.

The core contractor support included a tester capable of exercising the SSO functions, a specialist in the COTS operating system and utilities (in particular, RACF), and a staff member knowledgeable about the database and about message handling. The tester and database specialist were also familiar with the mission applications. As needed, they were joined by specialists in restart/recovery, mission applications (to run the QPR scenario), and exercise processing.

The contractor support was a key component of our test program. SPADOC 4B is a large and complex system with a number of specialized subsystems, including the SSO interface and the DBMS. While we could construct the command sequences we needed to implement our test procedures from the positional handbooks and COTS documentation, considerable time was saved by having contractor staff who were experienced in using those commands to help us in executing our procedures. Similarly, while we could investigate details of the implementation from the deliverable documentation, the contractor personnel were able to provide us with those details based on personal knowledge of and experience with the system.

3.3.4: Sources of information. In developing our system ST&E plans and procedures, we drew on a wide variety of sources. We started with the DT&E procedures and our experiences from witnessing DT&E and DN closure. We identified areas in which we felt more exhaustive testing would be useful.

We also consulted design documentation. The SPADOC 4B DTLS and the DTLS-to-Implementation Correspondence provided insight into the security design and implementation and into processing strings; some of our tests were intended to verify that the DTLS documentation accurately described the Trusted Computing Base (TCB).

We consulted minutes of the SPADOC Security Working Group (SSWG) and the Database Working Group (DBWG) to identify design decisions which, if not properly carried out, could result in incorrect or incomplete security functions. We developed procedures to test specific hypotheses about the implementation. Technical interchange meetings and teleconferences with the contractor allowed us to clarify our hypotheses and to obtain additional information.

Positional handbooks and the Trusted Facility Manual provided us with details about specific commands and command sequences. The positional handbooks were still in draft form when we developed our test procedures. (The handbooks were refined and expanded during training.) We found a number of areas in which the handbooks were unclear or incomplete and made recommendations to improve them.

3.3.5: Results. We verified that the system security features work correctly in low-load and stressed environments, including not only the usual live/backup configuration but also the exercise and partitioned mode processing environments. We verified that hardware and software component failures do not degrade the correct functioning of the security features. We uncovered problems not identified during DT&E, not all of which were security-related.
We identified a number of quirks and subtleties about the system's behavior. Our recommendations for documenting this behavior have been incorporated into the SSO Positional Handbook and the user's operating instructions (OUs).

During Phase 3, we identified vulnerabilities related to the capabilities of system security administrators. We also recommended additional safeguards, some procedural or administrative (e.g., separation of duties) and some in the form of system changes. In one case, this involved changing one of the COTS products to remove the capability to observe and change the contents of memory.

4: Lessons learned

In this section, we summarize the lessons we learned from the SPADOC 4B ST&E effort that we feel could be of use to other programs.

1. **ST&E is a team effort.**

   ST&E requires close cooperation between the contractor and the government test team. The fact that our ST&E was not part of contractual verification helped to foster a cooperative atmosphere as did the fact that the contractor's ST&E support was partially funded by level-of-effort (LOE) funds. In contrast to DT&E, the contractor test staff were not pressured to stick to a prepared script. All team members were encouraged to take a "what happens if we try this?" attitude.

   The team effort is vital not only during the actual testing but also as the test procedures are written. We provided drafts of our test procedures to the contractor support team prior to testing; their comments and suggestions improved the procedures.

2. **The test team must include a broad range of expertise.**

   For a complex system, a mix of specialized backgrounds on both the contractor and government sides is essential. We needed team members with expertise not only in security but also in the database, the COTS OS and utilities, the mission applications, and restart/recovery. We also needed team members who knew how to get the resources (e.g., meetings or teleconferences with contractor and AFSPACECOM personnel, time on the system, documentation) we needed.

   The government should identify the specific contractor personnel wanted for the test team, based on experience during DT&E. This ensures that staff with the needed expertise will be present. It also creates a better chemistry in the test team, since the contractor test team members know that their personal contributions are valued.

   To be effective in performing ST&E, team members need not only technical backgrounds based on participation in design reviews and working groups and on documentation review but also some hands-on experience with the system. A government team member who had not witnessed part of DT&E participated in Phase 1 and was at a distinct disadvantage.

   The team should include members with detailed knowledge of the system security architecture. While functional requirements can be tested effectively at the user interface, such testing can fail to identify security vulnerabilities inherent in the design. The DAA expects ST&E to include a careful search for vulnerabilities. We found that the DTLS and the DTLS-to-implementation correspondence mapping provided the necessary information.

   Team members with an operational background are also needed. We were hampered by not having an experienced SPADOC 4A SSO on the team.

3. **Careful preparation is vital.**

   The contractor and government test teams had hands-on experience as well as a good theoretical knowledge of the system design so that we could have started testing with only a rough outline and developed our specific test procedures "on the fly." However, this would have wasted resources while we figured out what exactly to do next. Our time on the system was limited as was the availability of expert staff. Without a clear structure, we could not have been as exhaustive in our tests.

   The written test plans/procedures should include a "road map" of the overall test effort as well as the detailed procedures. Otherwise, it is easy to get bogged down in the details and to lose sight of what the procedures are intended to demonstrate or investigate. The procedures should be written to allow for flexibility. When we observed some interesting system behavior, we wanted the opportunity to pursue it further rather than sticking to a canned script. We did not emphasize this enough in our written procedures; as a result, one of the observers expressed concern whenever we deviated from them.

   The written procedures should provide enough detail that they can be repeated easily. It may be necessary to repeat specific procedures if something
goes wrong the first time. In addition, it is important to be able to perform regression testing after the system has been modified. For example, we repeated many of the tests we performed in Phase 1 during Phase 2 to demonstrate not only that the security features worked properly in a stressed operational environment but also that DN closure and the "A-to-B roll" had not changed the system's security-relevant behavior.

The contractor team needed advance knowledge of some of our planned tests which involved situations they had not encountered in their own testing. For example, they were unsure what outcome to expect from many of the restart/recovery tests. (In fact, we discovered robustness in the HYPERchannel which they had not expected.) They also had to clarify their concept of operations for partitioned mode and to consider how security concerns for partitioned operations should be addressed.

Part of the necessary preparation is identification of which manuals and design documents are needed. We did not do this for Phase 1 and came across several situations in which we needed the SSO Handbook. For Phase 2, we brought the documentation we thought we might need; this resulted in fewer dead periods while questions were being investigated.

It is also necessary to identify what auditing should be turned on. We needed to test a broader range of auditing functions than are expected to be used in normal operations.

It is useful to have enough people (on both the government and the contractor sides) to allow running tests in parallel. However, these should be tests which are not expected to interfere with each other. We sometimes lost time figuring out that an unexpected result was due to another test using the same resources, which we could have predicted.

4. Preserve the test structure during testing.

While it is important to allow for flexibility ("what happens if...?") during ST&E, such investigations can result in failure to perform planned tests if time runs out or in unexpected results in planned tests if the what-if testing uses the same resources. Our testing ran more smoothly when we waited until after the planned tests were done to go back and investigate apparent anomalies further. A good mechanism for preserving the test structure is to take notes on the test procedures with pointers to a notebook where needed.

5. Keep detailed notes.

Detailed notes are useful not only when writing the test reports but also when something unexpected occurs. Software errors in a complex system are notoriously hard to replicate. Similarly, correct behavior may depend on enough different variables that it is hard to describe the exact conditions under which a given response will occur.

It is especially important to keep detailed notes of free-form tests, since the DAA could require regression testing of the entire set of ST&E procedures after a change is made to the system. However, care must be exercised in making notes about problem areas, since descriptions of vulnerabilities should be classified.

6. Allow extra time in the schedule.

We found that system initialization at the start of testing and audit data reduction at the end took a significant part of the time we had on the system. System initialization sometimes involved cleanup from use of the system during the previous shift. In addition, restart and recovery operations (which also involve cleanup) took a lot of time.

7. Ensure that a vehicle exists for resolving problems.

The DAA will expect the ST&E reports to provide recommendations for improving system security and, specifically, for resolving any problems that are found. The contractor may be unwilling to treat such problems as seriously as those that could affect contractual acceptance. We ensured that any discrepancies or errors we found would be treated contractually in the same way that problems found during the IOT&E dry runs were.

8. Test in a variety of representative environments.

Unusual conditions such as a heavy processing load or a component failure should not degrade security. However, the only way to verify this is to test the security features under those conditions. The fact that we had tested in a stressed environment and in all the operating modes (live/exercise and live/backup) gave our test results far more credibility to the DAA.

9. Use ST&E as an opportunity for visibility into other functional areas.

While the focus of ST&E is on the system security features, it provides the government test team with a unique opportunity for hands-on experience. Such experience is not available during more tightly
scripted periods of testing. In addition, testing in a variety of environments allows the government to observe interactions between system functions that may not have been visible during DT&E.

5: Conclusion

SPADOC 4B received an interim accreditation in multilevel mode in February 1991 to allow IOT&E with connectivity to external systems. This accreditation was based on the evidence developed through design analysis and ST&E and on the risk analysis that incorporated that evidence. SPADOC 4B received a second interim accreditation in multilevel mode in May 1991 to allow operational verification of the system; this interim accreditation covered a period up to ninety days after initial operational capability. It was based on the evidence presented for the February accreditation and on the tracking and resolution of security-related discrepancies identified during IOT&E. Final accreditation was granted in November 1991 based on prior evidence, on an ongoing program of security discrepancy tracking, and on implementation of additional safeguards identified in the ST&E Report and/or in the Risk Analysis Report.

ST&E was a fundamental component of the SPADOC 4B C&A process. It provided evidence to validate the assumptions used in the risk analysis. It required extensive effort to plan, execute, analyze results, and report, but this effort resulted in a credible and useful accreditation package.

References


