EVOLVING THE VV&A PROCESS FOR THE ALSP JOINT TRAINING CONFEDERATION

John A. Tufarolo
Ernest H. Page

The MITRE Corporation
1820 Dolley Madison Boulevard
McLean, Virginia 22102-3481

ABSTRACT

The Aggregate Level Simulation Protocol (ALSP) Joint Training Confederation (JTC) provides the environment for several large-scale computer-assisted command post exercises (CPXs) each year. The JTC consists of a collection of Service simulations and evolves according to an annual development cycle. The JTC environment poses several challenges to simulation model verification, validation and accreditation (VV&A). Some of these challenges are of a technical or fundamental nature, others are more programmatic. We briefly describe several of these challenges and the efforts currently underway to address them.

1. INTRODUCTION

The Aggregate Level Simulation Protocol (ALSP) began as a Defense Advanced Research Project Agency (DARPA) funded effort in the spring 1992 to investigate the feasibility of interconnecting existing constructive simulations to form a cohesive exercise environment (DARPA 1990). Subsequent to a successful demonstration of these capabilities in the fall of 1992, ALSP was fielded as the infrastructure for a premier joint exercise environment. Since 1992 this environment, the Joint Training Confederation (JTC), has grown from two primary simulations to eight and has supported several major exercises each year. For 1996 these include: RSOI and UFL, hosted by the Korean Battle Simulation Center, Seoul, Korea; Prairie Warrior, hosted by the National Simulation Center, Ft. Leavenworth, Kansas; and United Endeavor, hosted by the Joint Training and Analysis Simulation Center, Suffolk, Virginia. For a complete list of JTC-supported exercises refer to (Page et al. 1996).

Many aspects of ALSP and the JTC have been necessarily experimental in their nature. Although many of the technologies leveraged are mature ones, the JTC arena provides a novel context for their application. As a result, the processes and methodologies guiding the JTC have followed an evolutionary path. Many problems encountered have admitted solution, but many challenges remain.

The techniques and methodologies for verification, validation and accreditation (VV&A) are an example of this phenomenon. The techniques for VV&A of discrete event simulations are well studied and widely utilized (e.g., see (Balci 1994b)). Likewise, software testing methodologies are well known (e.g., see (Hetzel 1988)) and are the basis of a variety of standards (e.g., see (Department of Defense 1988)). However, the application of these techniques and methodologies within the JTC is not straightforward — a variety of technical and programmatic barriers to their application exist. The size and scope of the JTC is increasing to keep pace with the ever-expanding training requirements, yet the time and dollars allocated to testing promises to remain fixed. In order to maintain JTC acceptability (a function of sufficient validity), the process of engendering and quantifying JTC credibility must strive to be both flexible and optimal.

This paper describes the latest efforts in refining the VV&A processes within the JTC. Section 2 provides a review of the JTC development process. Evaluation of the impact on this process on VV&A is given in Section 3. Section 4 identifies some current initiatives to improve cost-effectiveness of VV&A in the JTC. Conclusions are given in Section 5.

2. THE JTC DEVELOPMENT PROCESS

Page et al. (1996) describe the VV&A practices for the ALSP JTC focusing on the testing process and the technical and operational challenges introduced through the Advanced Distributed Simulation (ADS) arena. A more general description of the JTC development process is given in (Miller and Zabek 1996).

The JTC development cycle comprises the transformation of user requirements and existing training simulations into a user-validated system (a "confederation" of simulations). A portrayal of the overall ALSP system
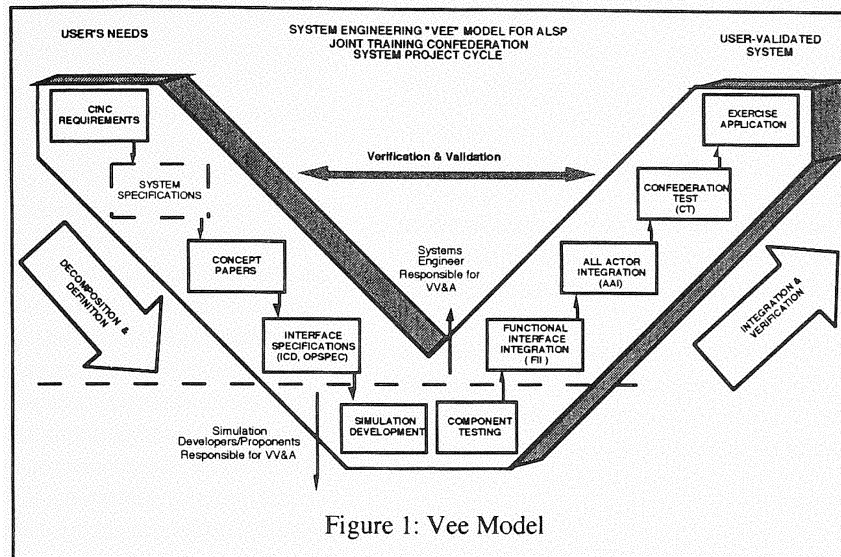
Figure 1: Vee Model

project cycle is presented in Figure 1, as viewed using a system engineering "vee" model adapted from (Forsberg and Mooz 1992).

A single iteration of the system life cycle begins with user needs (requirements) at the upper left and concludes with a user-validated system at the upper right. Descending the left side of the "vee" reflects decomposition and definition of the system. Ascending the right side of the "vee" represents system integration and verification. Each level of the model provides a correspondence between: (1) defined requirements and specifications, and (2) integration and verification events. In order to promote prudent system development, requirements and specifications should not be defined in a way that cannot be tested or verified. Similarly, integration and test events should have corresponding requirements and/or specifications they propound to verify and/or validate.

2.1 Overview

For the JTC, user needs are captured in the form of CINC/Service requirements. These high-level requirements are defined, agreed upon and prioritized for implementation. Based upon available funding (as well as the perceived availability of future funding) a collection of high-level designs are constructed in the form of concept papers. These concepts are evaluated against a variety of criteria and, if accepted, are refined into a set of model specifications and implemented within the various simulations comprising the JTC. Subsequently, the system of simulations is integrated, tested and evaluated. The "final" system consists of an accredited confederation (new versions of individual simulation software along with the ALSP Infrastructure Software) and associated documentation. It is this final

system that is intended for deployment and subsequent use at exercise locations.

2.2 Design and Specification

From CINC/Service requirements, sets of *concept papers* are written. A concept paper describes the new or modified confederation functionality being proposed, identifies the relevant requirements, and suggests methods for achieving the desired capability. Concept papers outline the applicability of the concept to the JTC, impact on existing simulations, and estimated level of effort for affected simulations. These papers also provide a forum for bringing new simulations into the JTC.

Concepts are evaluated and approved through the Interface Working Group (IWG). The interested reader should refer to (Fischer 1994) for a description of the organization and management of the Joint Training Confederation.

Subsequent to concept approval, detailed designs are formulated. These designs span several documents: Interface Control Documents (ICDs) describe message syntax, content, and sequencing requirements. Actor Implementation Documents (AIDs) detail the ICD implementation within a specific simulation ("actor" in the parlance of ALSP). The Operational Specification (OPSPEC) provides the model definition in terms of objects and attributes as well as simulation-object ownership and behavioral "scenario descriptions" and serves as the primary communicative model representation.

Currently the dashed item in Figure 1 labeled "System Specification" does not exist explicitly for the JTC. Individual simulations maintain system specifications describing private objects and behaviors within their

respective simulations. Ideally, these specifications should be consolidated to form a single system description describing objects and behaviors — both private and public — for the JTC *as one system.*

2.3 Development and Implementation

The simulation developers use the design documents to generate modified versions of their simulation software. By organizational agreement, each Service (or model proponent) has the responsibility of VV&A for its simulation (Fischer 1994). Similarly, each Service *reserves* responsibility for evaluating any code changes made with respect to either Service or JTC requirements. This important point is elaborated in Section 3.

2.4 Integration and Test

Following simulation development, integration and testing of the confederation takes place. Three test stages are identified: Functional Interface Integration (FII), All Actor Integration (AAI), and Confederation Test (CT). These tests are supported by participants in the JTC community.

Test plans created to support the AAI are intended to test the inter-simulation functionality. Formal reports from the AAI are not generated. However, results are informally promulgated to the simulation developers to permit refinement of their interfaces.

The CT is the pinnacle of the JTC development process, and serves as the user acceptance test. Testing is conducted by the user community, and formalized test reports are generated. Stress tests are added to provide a means to operate the JTC under larger, "exercise-like" loads. The CT culminates with a written test report presented to the Review Panel for approval. This report describes the tests performed at the CT and the results obtained. The Review Panel promulgates the official accreditation of the ALSP JTC.

2.5 Deployment

Confederation deployment includes delivery and employment of new simulation and ALSP Infrastructure Software at user sites. The JTC Systems Engineer provides support during "first-use" — the first major exercise to operate a newly accredited JTC. Associated with the JTC deployment, exercise centers typically conduct additional on-site testing of the JTC using their specific environment and data. This supplementary testing, to include Systems Engineer support if requested, is funded by the exercise sites.

3. EVALUATION

An evaluation of the VV&A process within the Joint

Training Confederation must begin with a recognition that a singular "VV&A Agent" role is not defined for the JTC (Fischer 1994). Such an agent, if singularly defined, would require the omnipotence to establish and oversee the VV&A within the individual simulations as well as the VV&A of the confederation as a whole. In the current JTC management structure each Service (or model proponent) has the responsibility of VV&A for its simulation. The ALSP Systems Engineer assumes responsibility for VV&A of the system as a whole (see the demarcation of VV&A responsibilities in Figure 1). As Balci (1994a) indicates, a collection of valid sub-components does not necessarily result in a valid system when these components are integrated. Therefore, validation of the whole requires validation beyond that of the parts. However, the limited authority of the Systems Engineer in this role implies that most of the testing undertaken to support the system-level validation activity is necessarily functional or "black-box".

3.1 The Vanilla Approach to Systems Testing

Adopting the term used by (Harel 1992) we describe the vanilla approach to systems testing (testing "in the large" as delineated in (Hetzel 1988)) as being comprised of four stages: (1) component test, (2) integration test, (3) system test, and (4) acceptance test (assuming the existence of a customer for the system). Figure 2 depicts the JTC test process superposed on the vanilla approach. Although the mapping is not as clean as the figure indicates, the absence of a component phase of system testing is evident. While lack of a component test may not directly affect validation, it potentially results in inefficiencies in the development cycle — since faults may go undetected until later in the cycle thus tending to increase the cost of their correction.
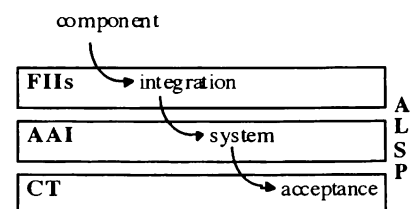

Figure 2: JTC Test Process Superposed on the Vanilla Approach

3.2 Problems Associated with a Traditional Approach

The addition of a component test phase to JTC development would provide the general vanilla framework, and as such should tend to increase cost-effectiveness. However, a number of complications (with respect to

model validation and accreditation) are associated with the application of such a traditional testing approach within the context of the JTC. A few such complications are noted below.

- *The test system and fielded system differ in scale.*

The systems and personnel assembled to support and participate in a JTC exercise are extensive, typically involving thousands of people and hundreds of workstations over a month long period. In contrast, the 1996 ALSP JTC confederation test (user acceptance test) involved roughly 100 people for two weeks, and approximately 40 simulation workstations.

Both practicality and cost limit the size and scale of JTC testing. A test cannot cost a much as an exercise. In fact, testing costs arguably must be many orders of magnitude less than an exercise. As a result, the system that is used to support training is much larger in scale than the system tested.

- *The test system and fielded system differ by platforms.*

Exercises are usually held at military "simulation centers" (for examples, see section 1). Each center has a significantly different hardware, software, and network infrastructure. Therefore, test results are only applicable (for certain measures, e.g. performance) with respect to a particular infrastructure (the one used for the test) and, of course, subject to the caveat of the aforementioned scale problem.

- *The test system and fielded system differ by composition.*

The ALSP JTC may be viewed as a layered architecture. At the hub are a set of simulations (e.g. CBS, AWSIM, RESA, JECEWSI, TACSIM, CSSTSS, MTWS, and PSM for 1996 — see (Page et al. 1996) for details) that exchange information and coordinate the advance of simulation time over ALSP. Many of these simulations are interactive, i.e. they have user input terminals and battlefield situational displays. These devices are generally referred to as "controller workstations" and provide a layer around the simulation hub.

An additional layer of systems (referred to in the JTC community as peripheral software systems or PSS) are routinely attached to the JTC when employed at an exercise. These systems are influenced by actions and changes made within the JTC, and facilitate information exchange with the training audience. The training audience generally does not interact directly with workstations or PSS, but is (ideally) immersed in a wartime-equivalent environment. In some instances, the

PSS provide information in a facsimile of real-world formats. A very small number of the PSS support linking real-world command and control (C^2) systems and command, control, communications, computers and intelligence (C^4I) systems to the JTC. More often the linkage requires a manual interface (so called "man-in-the-loop") to convey information between the JTC and the training audience. The mélange of workstations, PSS, C^2/C^4I systems, and manual interface mechanisms provides the outer "external systems" layer of the JTC architecture.

During a CT, a system consisting of the JTC and a collection of external systems anticipated for the coming year's training exercises is tested. Unfortunately, the tested system is never identical to the one used at any exercise. Subsets of the simulations and external systems are selected, often based on political or cost rationale that supersede technical considerations. New external systems are occasionally used in exercises (to satisfy exercise requirements) that were not tested during a CT.

- *The test system and fielded system differ by code.*

The JTC development can be viewed as multi-level development life cycle. Each simulation component comprising the JTC has a separate life cycle (necessary to satisfy primary Service requirements) in addition to that of the JTC. These individual development life cycles must be coordinated to coincide with the JTC life cycle. Currently no formal mechanism exists to facilitate such coordination. Modifications to software can occur throughout the year (as dictated by the primary Service development cycle). By organizational agreement, configuration management authority falls to the Service.

- *The test system and fielded system differ by data.*

Many military simulations, including those in the JTC, are essentially "data driven," i.e. the outcome of any particular event in the model is sensitive (at varying levels of sensitivity) to values in the model database (see (Davis 1992) for design rationale). For the JTC simulations these databases include information regarding the types of battlefield equipment, order-of-battle (hierarchical) information, and details of the geographical region being played.

These databases differ greatly between exercises. One exemplary exercise includes next-generation (2010) weapons systems in a fictionalized European scenario. Another uses current weapon systems in a Korean defense scenario.

The results of any test are tightly coupled to the collection of databases underlying the simulations. As a

result, database tests must be conducted prior to each exercise. However these tests also fail to be definitive. Database parameters may be changed during an exercise, typically to correct an inappropriate action observed by exercise controllers, or to ensure the exercise scenario evolves as required to meet training objectives.

- *The test system and fielded system differ by personnel.*

Several interfaces between the training audience and the JTC simulations have been automated (i.e. a direct link between C^2/C^4I systems and simulations). Many of them have not. This is the primary role of an exercise controller: to communicate with the training audience via real world mechanisms, and interact with the simulations using controller workstations.

There is also a level of critical personnel generally referred to as "technical control". These personnel operate the simulations, the ALSP infrastructure software, and the attendant systems and networks.

Both the exercise controllers and technical personnel vary from exercise to exercise. Each can introduce mistakes that unintentionally disrupt the exercise. Novice workstation operators often input unexpected values, producing equally unexpected events in the simulations. System and network personnel will adjust systems and networks differently, resulting in varying behavioral characteristics.

- *System behaviors are subject to evaluation/modification.*

A mechanism for fault tolerance is typically used in training exercises under the nomenclature "white-cells". White-cells are a group of personnel that behave as referees and are used to adjudicate questions regarding the state of the simulation. When an undesired outcome (in terms of training objectives) is detected in the state of the simulation, an exercise white-cell can intercede and realign the simulated conditions to better meet the objectives.

For example, during a recent exercise, an initial engagement involved a massive cruise missile attack by Opposing Forces (OPFOR) against friendly forces. As a result of operator errors during exercise set-up, many friendly forces did not have fire control for their air defenses. These units were virtually wiped out during the opening attack. Such an imbalance of forces threatened to seriously impact the training objectives for the exercise, so the referees interceded and the friendly force structure was replenished to the expected post-attack level.

3.3 Why Do Bad Things Happen at Exercises?

In "traditional" settings (i.e. simulation used for analysis) failure of the simulation is strictly a function of simulation invalidity. This isn't customarily the case in the interactive training simulation world. Numerous problems can be traced to causes outside the simulations themselves, including computer failure, network failure, and operator error. To a lesser extent, general software errors occur. Consequently, the items that a substantive portion of our testing efforts focus on, and the corresponding results that lead us to consider the simulations as "sufficiently valid" are frequently *not* contributing to exercise problems.

Although the testing efforts are valid and successful test results desirable, the downside is we end up with a confederation of valid simulations which can fail to support an exercise.

For example, in the first days of one exercise the network performance became significantly degraded such that the simulations were running so slow (at a ratio of 0.4:1 with real time) that the training audience was becoming affected. After several hours of investigation the problem was tracked to the fact that a workstation had been added to the network that morning and given the same IP address as one of the mainframes hosting the simulations. How VV&A can (or should) address problems like these is an open issue.

3.4 Is a Traditional Approach Infeasible?

The accumulation of factors noted above begs the question, "to what extent can model validity be determined during acceptance testing?" The answer seems to be a little, but only partially. The differences in scale, composition, data, and personnel demand that much of the validation activity may only be undertaken within the immediate context of an exercise — using the existent training system, data and personnel as the basis for testing. If this is true, then is there any value in conducting the traditional (component, integration, system, acceptance) tests? Should these tests be abandoned, and all test expenditures redistributed to pre-exercise efforts? Such course of action would seem unwise. The benefits gleaned from a disciplined, bottom-up testing approach extend beyond validation. They contribute to early error detection as well as verification, and provide feedback regarding the overall reliability of the system.

A cost-effective solution would seem to lie somewhere in the middle, as it often does; its precise location only identifiable by trial and error.

4. NEW DIRECTIONS

Although the challenges to the conduct of VV&A are numerous — some technical and others programmatic — affecting change within the process is necessarily slow. In addition to the natural resistance to change, the role of the JTC as a proven and reliable environment supporting training exercises demands that modifications to a process that works sufficiently well should be made with caution. There is wisdom to the philosophy "if it ain't broke don't fix it".

Several initiatives are underway for the 1997 development cycle.

- A test harness is being designed to support component testing at the system level. The harness will be "programmed" to stimulate an actor with respect to the interfaces it participates in (as described in the ICDs and AIDs).

- To address the sensitivity of model validation on the underlying model databases, a mechanism is being provided to allow each actor to *register* the values of key database elements with the ALSP infrastructure during the confederation initialization process. The database elements are enumerated attribute values defined at the confederation level — so-called *enumerations mapping*. A tool has been developed to collect this information and provide cross-referencing to facilitate the detection of incomplete entries or errant mappings. This facility could significantly improve what has been an entirely manual, labor-intensive and error-prone process. However, not all of the JTC actors can easily register their enumerations mapping. In some cases this mapping is resident in code rather than a separate database, thus making the creation of a routine that extracts this information a difficult (if not impossible) task.

- Another tool has been developed to monitor the message traffic and examine messages for invalid enumerations. Note that should each actor have the capability to register its enumerations mapping (as noted above) *and* if no database changes occur during an exercise, this tool would provide no additional benefit. However, in the absence of such capabilities and guarantees, this tool, while providing less information than an enumerations mapping cross-reference, can provide a valuable mechanism for quickly identifying certain common errors that occur during confederation operations.

5. CONCLUSIONS

Within the JTC development process, cost-effectiveness serves as the overriding objective. While organizational and bureaucratic factors tend to hinder early error detection and strict configuration management, the VV&A process within the JTC seems to overcome these difficulties. The track record of the JTC as a training vehicle serves as witness to this phenomena. While the size and scope of the JTC continues to expand to keep pace with ever-expanding training requirements, the time and money allocated to testing promises to remain fixed (or even diminish). Given that model validity is a function of cost versus benefit, the challenge to the Systems Engineer for the JTC is to maximize the benefits (i.e. the validity) in the face of fixed costs.

REFERENCES

Balci, O. (1994a). Principles of Simulation Model Validation, Verification, and Testing. Technical Report. TR 94-24. Department of Computer Science. Blacksburg, VA, June 10. (To appear in: *International Journal in Computer Simulation*).

Balci, O. (1994b). "Validation, Verification, and Testing Techniques Throughout the Life Cycle of a Simulation Study." *Annals of Operations Research*(53): 121-173.

Davis, P. K. (1992). Generalizing Concepts and Methods of Verification, Validation, and Accreditation (VV&A) for Military Simulations. R-4249-ACQ. The RAND Corporation. Santa Monica, CA.

Defense Advanced Research Project Agency (DARPA) (1990). Advanced Distributed Simulation Program. Project Report. Advanced Research Project Agency, March.

Department of Defense (1988). MIL-STD-498: Software Development and Documentation, United States Department of Defense.

Fischer, M. C. (1994). "Aggregate Level Simulation Protocol (ALSP) Managing Confederation Development." *Proceedings of the 1994 Winter Simulation Conference.*

Forsberg, K. and H. Mooz (1992). "The Relationship of Systems Engineering to the Project Cycle." *Engineering Management Journal* 4(3): 36-38.

Harel, D. (1992). "Biting the Silver Bullet:Toward a Brighter Future for System Development." *IEEE Computer* 25(1): 8-20.

Hetzel, B. (1988*). The Complete Guide to Software Testing.* Wellesley, MA, QED Information Sciences, Inc.

Miller, G. and A. Zabek (1996). The Joint Training Confederation and the Aggregate Level Simulation Protocol. *Phalanx.* **29**: 24-27.

Page, E., J. Tufarolo and B. Canova. (1996). "A Case Study of Verfication, Validation and Accreditation for Advanced Distributed Simulation." *ACM Transactions on Modeling and Computer Simulation* (Submitted for publication).

AUTHOR BIOGRAPHIES

JOHN A. TUFAROLO is a Senior Simulation Engineer for the MITRE Corporation in Reston, Virginia, where he is currently the MITRE project manager for the Synthetic Theater of War (STOW) Command Forces (CFOR) program. He has over 10 years experience in modeling and simulation, supporting development, testing, and application of military training and analysis simulations. Mr. Tufarolo has worked as a systems engineer and lead test planner for the Aggregate Level Simulation Protocol (ALSP) program, and currently serves as a member of the Defense Modeling and Simulation Office (DMSO) High Level Architecture (HLA) simulation management group. He received a B.S. degree in electrical engineering from Drexel University in 1985 and is currently an M.S. candidate in Systems Engineering at George Mason University. He is a member of the ACM, IEEE, and SIGSIM.

ERNEST H. PAGE is a member of the technical staff of the MITRE Corporation where he is currently working on the Aggregate Level Simulation Protocol (ALSP) program. He received the Ph.D., M.S. and B.S. degrees in Computer Science from the Virginia Polytechnic Institute and State University (VPI&SU) in 1994, 1990 and 1988 respectively. Dr. Page serves as a member of the Defense Modeling and Simulation Office (DMSO) Verification, Validation and Accreditation Technology Working Group (VV&A TWG) and VV&A Technology Support Team. He has been active with the High Level Architecture (HLA) initiative, working with both the Test and Interface Specification Working Groups. Dr. Page is the Secretary/Treasurer for the Association for Computing Machinery (ACM) Special Interest Group on Simulation (SIGSIM). His research interests include discrete event simulation, parallel and distributed systems, and software engineering. He is a member of ACM, SIGSIM, IEEE CS and Upsilon Pi Epsilon.