

Metrics-based Approach for Evaluating Air Traffic Control Automation of the Future

Mike M. Paglione
Federal Aviation Administration

Shurong Liu, Robert D. Oaks,
Dr. Hollis F. Ryan,
General Dynamics Network Systems

Mike M. Paglione is the Conflict Probe Assessment Team Lead in the Simulation and Analysis Group with the Federal Aviation Administration's Simulation and Analysis Group. Shurong Liu is a Senior Staff Member, Robert D. Oaks is a Principal Staff Member, and Dr. Hollis F. Ryan is a Senior Analyst with General Dynamics Network Systems. They all work at the William J. Hughes Technical Center, Atlantic City International Airport, NJ, 08405.

Background

In 1996 the Federal Aviation Administration (FAA) established the Conflict Probe Assessment Team (CPAT) at the William J. Hughes Technical Center (WJHTC) to evaluate the accuracy of the conflict probes in Decision Support Tools. Since its inception, CPAT has measured the conflict prediction accuracy of the User Request Evaluation Tool (URET),¹ measured the trajectory modeling accuracy of both the URET and the Center TRACON Automation System (CTAS),² assisted in the accuracy testing of the URET Current Capability Limited Deployment (CCLD),^{3,4} which is the operational implementation of the URET, and has been involved with the regression testing of the URET as it has been deployed throughout the National Airspace System (NAS). As the result of both the individual and team expertise and the suite of software tools created through these activities, CPAT is leading an effort to develop metrics for testing the accuracy of the En Route Automation Modernization (ERAM) system, which is the replacement system for modernization of the NAS. This paper discusses how the CPAT team members have been working with other test groups located at the WJHTC to develop metrics for ERAM testing.

En Route Automation Modernization

The development of ERAM will be especially challenging because it is a safety critical system replacing hardware and software that has been extensively enhanced and modified since it was originally developed in the 1970's. In addition, the air traffic controllers who use this system are very familiar with the system and exploit its strengths and weaknesses. This is not unique; the development of any complex computer system is a difficult effort, with each system having unique challenges that require the system developers to adhere to a software development methodology that organizes the many tasks.

Over the years a multitude of software development processes have been proposed and tried with varying degrees of success. But regardless of the specifics of the software development process there are four basic phases that must be addressed in any methodology: a design phase, a development phase, a testing phase, and operational deployment.

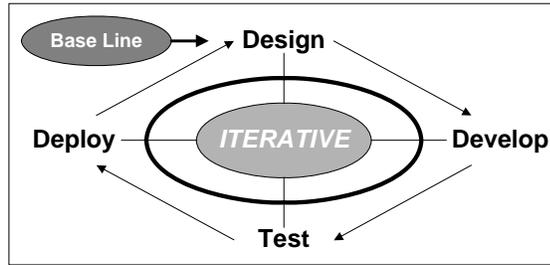


Figure 1. Recipe for project success*

In most modern software development methodologies, it is generally accepted that these phases must interact iteratively. Figure 1 depicts such an interaction, which the Standish Group describes as their recipe for project success.⁵

The scope of this paper is the testing phase of the ERAM development effort. Specifically, this paper presents an overview and examples of testing metrics defined to support the development of ERAM, many of which are an extension of CPAT's accomplishments.

Testing

As shown in Fig. 1, the testing of a complex computer system is an iterative process that is an integral part of all phases of software development, from initial design through deployment. But testing is an ill-defined task that can never be considered to be a completed. This is reflected in the following quotes that are often cited when discussing testing:

Testing is the process of comparing the invisible to the ambiguous, so as to avoid the unthinkable happening to the anonymous.

-- James Bach[†]

Program testing can be used to show the presence of bugs, but never to show their absence.-

-- Dr. Edsger Wybe Dijkstra[‡]

Often testing is decomposed into two components: verification testing and validation testing.

- Verification testing is the testing that ensures that the software meets the requirements specified by the customer. Verification testing is usually characterized by the question, "Are we building the product right?"
- Validation testing, on the other hand, is characterized by the question, "Are we building the right product?" Validation testing is the testing that ensures that the

* Adapted from "Extreme Chaos" by The Standish Group.⁵

[†] James Bach is a contemporary author and founder of Satisfice, Inc., a company dedicated to teaching and consulting in software testing and quality assurance.

[‡] Dr. Edsger Wybe Dijkstra (1930-2002) was a Dutch computer scientist and physicist.

software meets the needs of the user. This testing often includes performing a systematic comparison of the software's results to increasingly complex cases of real-world experimental data.

Both verification testing and validation testing assess the software's correctness and completeness; but they are concerned with different evaluation criteria. In other words, verification testing establishes whether a system corresponds to its specification, while validation testing is defined in terms of comparing the system to a baseline or target.

Another important aspect of testing is its affect on project cost. Barry Boehm in his book titled "Software Engineering Economics"⁶ evaluated a number of software projects and estimated how the relative cost of fixing an error significantly increases as the project progresses in phase. Figure 2 summarizes the results of this study. Errors are about 15-50 times more costly to fix when found in the testing phases as compared to errors that are detected when the requirements are defined; but after the system has been deployed, it is roughly a hundred times more costly to fix an error. This heavily cited reference, may even be underestimating the cost.

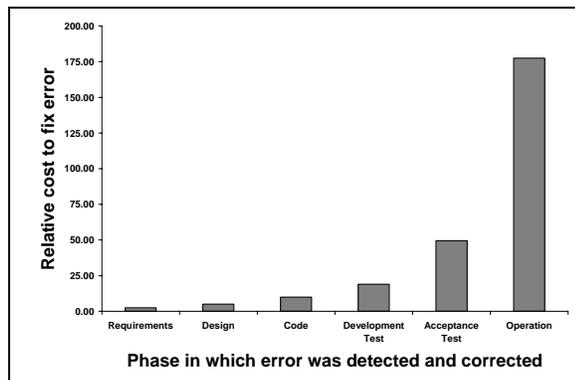


Figure 2. Increase in cost-to-fix through the software development cycle[§]

In addition, when dealing with safety critical systems such the NAS, safety may be a bigger issue. In a recent article in the Wall Street Journal, which cites examples of software errors found in numerous deployed onboard computer systems, Michaels and Pasztor state that software errors, "while extremely rare, are emerging as a top safety challenge in the air."⁷ Although this article focuses on aircrafts' onboard computer software, the authors feel that the safety challenge can also be appropriately directed toward the ground-based air traffic control systems.

Automation Metrics Test Working Group

To facilitate the ERAM testing, the FAA Solution Implementation Division ERAM Test Group, located at the WJHTC, formed the Automation Metrics Test Working Group (AMTWG) in 2004. This is a cross-functional team formed with members from a half dozen organizations located at the WJHTC. The team's charter is to support the developmental and operational testing of ERAM by developing a set of metrics that quantify the effectiveness of key system functions in ERAM. The targeted system functions are the Surveillance Data Processing (SDP), the Flight Data Processing (FDP), the Conflict Probe Tool (CPT), and the Display System (DS) modules. The focus of the

[§] Adapted from *Software Engineering Economics* by Barry W. Boehm.⁶

AMTWG is to go beyond verification testing (i.e., the requirement-based testing) and to emphasize validation testing by linking the metrics directly to the services provided by NAS. Whenever appropriate, the metrics are designed to measure not only the performance of ERAM, but also to measure the performance of the existing Host Computer System (HCS), which will enable comparison of the functionality in ERAM to the same functionality in the legacy HCS. For logistical purposes, the AMTWG categorized the metrics based on the targeted ERAM subsystems.

As illustrated in Fig. 3, the AMTWG divided the project into three key phases: a metrics identification phase, an implementation-planning phase, and a data collection and analysis phase.

- In the metrics identification phase the AMTWG generated a list of approximately one hundred metrics that map to the services and capabilities found in the “Blueprint for the National Airspace System Modernization 2002 Update.”⁸ These initial metrics were published in the “ERAM Automation Metrics: Progress Report of the Automation Metrics Test Working Group.”⁹
- Next, in the implementation-planning phase, the initial metrics were prioritized for more detailed refinement. This was documented in the “En Route Automation Modernization: Automation Metrics and Preliminary Test Implementation Plan.”¹⁰ It lists the metrics, gives the rationale for selecting them, and provides a high level description on how the highest priority metrics will be measured. The paper provides each metric’s traceability to the basic controller decisions, ERAM Critical Operational Issues (COIs), and the development contractor’s Technical Performance Measurements (TPMs). The categories of high priority metrics are: (1) SDP radar tracking metrics, (2) SDP tactical alert processing metrics, (3) FDP flight plan route expansion metrics, (4) FDP aircraft trajectory generation metrics, (5) CPT strategic aircraft-to-aircraft conflict prediction metrics, (6) CPT aircraft-to-airspace conflict prediction metrics, (7) additional system level metrics, and (8) DS human interface metrics.
- The final phase is the data collection and analysis phase, during which the AMTWG will document the further refinement and application of these metrics on the current legacy systems in a series of Metric Reports. The AMTWG is planning the delivery of four Metric Reports in fiscal years 2005 and 2006 with one covering each of the ERAM modules: SDP, FDP, CPT, and DS. The AMTWG is publishing these reports in multiple drops to provide the ERAM Test Team with timely information. The drops coincide with the approaches used to implement the metrics. This phase

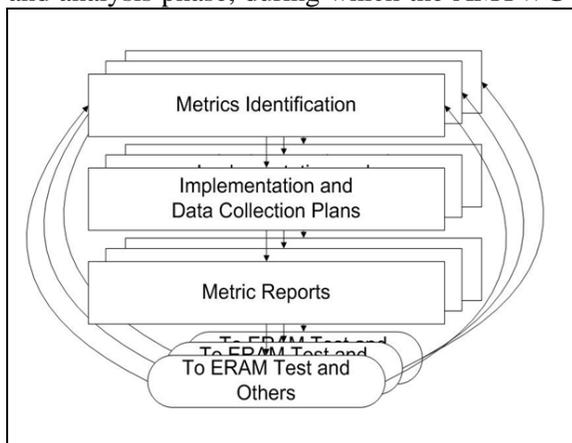


Figure 3. Iterative Process of Data Collection/Analysis Phase

will continue in fiscal year 2007 and beyond by examining more metrics from the list. The data collection and analysis phase is depicted in Fig. 3, which illustrates how the Metric Reports generate feedback from the ERAM Test Team and other sources (Program Office, development contractor) to identify more metrics or expand upon existing ones. The AMTWG will revise the implementation planning and generate new Metric Reports.

The metrics defined by the AMTWG are either absolute or comparative in nature, with the comparative metrics being applied first to the current NAS automation systems and then later to ERAM. The use of the metrics is also iterative with frequent output. The metrics are being applied currently on the legacy NAS to flush out and establish their credibility. The metrics focus on validation during development and operational testing and will later support verification by providing valid test cases, expertise, and tools when applicable.

Examples of Results to Date

The following provides examples of activities that have been completed to date:

- Several measurement techniques have been developed for the SDP radar tracking metrics using both recorded and simulated air traffic data.
- A strategy using descriptive and inferential statistical techniques to compare the converted route processes of ERAM to the HCS has been developed and tested.¹¹
- A study was conducted to measure the accuracy of the radar tracking function of the HCS.¹²
- A study was conducted measuring HCS tracker error.¹³
- A mapping of air traffic controller commands to the Air Traffic Control capabilities listed in the NAS Modernization Blueprint 2002⁸ was extended to the mapping of each ERAM View by operational position.¹⁴
- The frequency of use of air traffic controller commands using the legacy system was measured.¹⁵
- A study was conducted of how controllers use the legacy system during special situations such as weather, traffic management initiatives, emergencies, and outages.¹⁶
- A study was conducted that provided an analysis of major areas where new ERAM features may affect how controllers do their jobs.¹⁷

Future Plans

It is planned that the metrics developed by the AMTWG will be used for the ERAM Release 1 development and operational testing program in the following ways:

- To help to identify issues as early as possible.
- To support the plans and procedures used for Developmental and Operational Testing.
- To address a core COI ensuring ERAM supports ATC operations with at least the same "effectiveness" as the current NAS.
- To establish benchmarks on legacy systems.

- To provide data driven scenarios, methods, and tools to apply to ERAM to compare to current NAS.
- To leverage information obtained by subject matter experts to support ERAM testing questions.
- To provide valid baseline methods and measurements for future FAA Air Traffic Organization (ATO) Test Programs.

Broader, more system level metrics (e.g. aircraft time in hold) will be measured in collaboration with the Human Factors Group, who are planning other simulations and analysis with human-in-the-loop studies to evaluate the ERAM DS subsystem. These studies will use the Target Generation Facility and Pseudo-pilot systems in the I²F, both located at the WJHTC.

Later, with the experience gained using the ERAM Release 1 metrics-based approach, the metrics will be applied to future ERAM releases. The metrics can also be used in Next Generation Air Transportation (NGATS) initiatives, where systems will be proposed for new air traffic control concepts. In fact, these metrics will support the development of future NGATS Requirements by defining NAS capabilities based on measurable performance data.

Benefits

Overall, metrics-based testing provides several significant benefits to the ERAM program:

- In support to the ERAM Test Team, it will provide additional data points to help develop test cases and measurements that will supplement or enhance the requirements verification testing.
- The use of these metrics is a risk reduction activity for the entire program because they evaluate the effectiveness of key subsystems and functions.
- The metrics provide the ERAM Program Office with supporting data assessing the benefits of ERAM.
- The metrics-based testing activities allow key ERAM Test Team personnel to receive in-depth experience on the ERAM subsystems and similar existing functions in the legacy automation system. This, in turn, will increase their effectiveness in reviewing contractor test plans and procedures.
- Finally, the tools, metrics, and traffic scenarios may be incorporated, where appropriate, in the formal ERAM Test Program.

References

¹Paglione, Mike M., Robert D. Oaks, Dr. Hollis F. Ryan, J. Scott Summerill, “User Request Evaluation Tool Daily Use Time-shifting Trajectory Prediction Accuracy Degradation Study,” Federal Aviation Administration, ACT-250, William J. Hughes Technical Center, Atlantic City International Airport, NJ, December 15, 1999.

²Cale, Mary Lee, Michael Paglione, Dr. Hollis Ryan, Dominic Timoteo, Robert Oaks, “URET Conflict Prediction Accuracy Report,” DOT/FAA/CT-TN98/8, Federal Aviation Administration, ACT-250, William J. Hughes Technical Center, Atlantic City International Airport, NJ, April, 1998.

³Paglione, Mike, Dr. Hollis F. Ryan, Robert D. Oaks, J. Scott Summerill, and Mary Lee Cale, “Trajectory Prediction Accuracy Report User Request Evaluation Tool (URET)/Center-TRACON Automation System (CTAS),” DOT/FAA/CT-TN99/10, Federal Aviation Administration, ACT-250, William J. Hughes Technical Center, Atlantic City International Airport, NJ, May, 1999.

⁴Conflict Probe Assessment Team, “URET CCLD Final Accuracy Scenario Delivery Refresh Data, Revision 1,” available on CD-ROM, Federal Aviation Administration, Engineering and Integration Branch, ACT-250, William J. Hughes Technical Center, Atlantic City International Airport, NJ, November, 2000.

⁵The Standish Group, “Extreme Chaos,” URL: <http://www.standishgroup.com>, Copyright © 2001.

⁶Boehm, Barry W., *Software Engineering Economics*, ISBN 0-13-822122-7, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1981.

⁷Michaels, Daniel, and Andy Pasztor, “Incidents Prompt New Scrutiny of Airplane Software Glitches; As Programs Grow Complex, Bugs Are Hard to Detect; A Jet’s Roller-coaster Ride. Teaching Pilots to Get Control,” *The Wall Street Journal*, Page A1, May 30, 2006.

⁸Federal Aviation Administration, “Blueprint for NAS Modernization: 2002 Update,” URL: <http://www.faa.gov/nasarchitecture/Blueprint2002.htm>, [cited October, 2002].

⁹Automation Metrics Test Working Group, “ERAM Automation Metrics: Progress Report of the Automation Metrics Test Working Group,” Federal Aviation Administration, ERAM & ECG Group (ACB-550), Solution Implementation Division (ACB-500), Simulation and Analysis Group (ACB-330), System Analysis Division (ACB-300), William J. Hughes Technical Center, Atlantic City International Airport, NJ, June 16, 2004.

¹⁰Automation Metrics Test Working Group, "En Route Automation Modernization: Automation Metrics and Preliminary Test Implementation Plan, Version 2.6," Federal Aviation Administration, Simulation and Analysis Group (ACB-330), System Analysis Division (ACB-300), William J. Hughes Technical Center, Atlantic City International Airport, NJ, June 16, 2005.

¹¹Baldwin, W. Clifton, "Comparison of Converted Route Processing by Existing Versus Future En Route Automation," DOT/FAA/CT-TN05/29, Federal Aviation Administration, William J. Hughes Technical Center, Atlantic City International Airport, NJ, August, 2005.

¹²Ryan, Dr. Hollis F. and Mike M. Paglione, "Comparison of Host Radar Tracks to Aircraft Positions from the Global Positioning Satellite System," DOT/FAA/CT-TN05/30, Federal Aviation Administration, William J. Hughes Technical Center, Atlantic City International Airport, NJ, August, 2005.

¹³Paglione, Mike M., W. Clifton Baldwin, and Seth Putney, "Host Radar Tracking Simulation and Performance Analysis," DOT/FAA/CT-TN05/31, Federal Aviation Administration, William J. Hughes Technical Center, Atlantic City International Airport, NJ, August, 2005.

¹⁴Dolka, Linda, Christopher Reilly, Lawrence Rovani, and Wayne Young, "Display System Air Traffic Function Mapping to Air Traffic Control Capabilities, Version 2," Federal Aviation Administration, William J. Hughes Technical Center, Atlantic City International Airport, NJ, January, 2006.

¹⁵Allendoerfer, Kenneth, Carolina Zingale, Shantanu Pai, and Ben Willems, "En Route Air Traffic Controller Commands: Frequency of Use During Routine Operations," DOT/FAA/TC-TN06/04, Federal Aviation Administration, William J. Hughes Technical Center, Atlantic City International Airport, NJ, May 2006.

¹⁶Allendoerfer, Kenneth, Carolina Zingale, and Shantanu Pai, "An Analysis of En Route Air Traffic Control System Usage During Special Situations," DOT/FAA/TC-TN6/11, Federal Aviation Administration, William J. Hughes Technical Center, Atlantic City International Airport, NJ, May 2006.

¹⁷Allendoerfer, K. R., B. Willems, C. Zingale, and S. Pai, "Methods for Examining Possible Effects of En Route Automation (ERAM) on Controller Performance," DOT/FAA/TC-TN6/14, Federal Aviation Administration, William J. Hughes Technical Center, Atlantic City International Airport, NJ, May 2006.