SYM-AM-17-34

PROCEEDINGS
OF THE
FOURTEENTH ANNUAL ACQUISITION RESEARCH SYMPOSIUM

WEDNESDAY SESSIONS
VOLUME I

Acquisition Research:
Creating Synergy for Informed Change

April 26–27, 2017

Published March 31, 2017

Prepared for the Naval Postgraduate School, Monterey, CA 93943.

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

The research presented in this report was supported by the Acquisition Research Program of the Graduate School of Business & Public Policy at the Naval Postgraduate School.

Table of Contents

SYM-AM-17-034

PROCEEDINGS
OF THE
FOURTEENTH ANNUAL ACQUISITION RESEARCH SYMPOSIUM

WEDNESDAY SESSIONS
VOLUME I

Acquisition Research:
Creating Synergy for Informed Change

April 26–27, 2017

Published March 31, 2017

THIS PAGE INTENTIONALLY LEFT BLANK

Keynote: Vice Admiral David C. Johnson, USN—Principal Military Deputy, Assistant Secretary of the Navy for Research, Development and Acquisition

Vice Admiral David C. Johnson, USN—the son of a Navy captain and a Pensacola, FL, native, graduated from the U.S. Naval Academy in 1982 with a Bachelor of Science in Aerospace Engineering.

Upon commissioning, Johnson reported to Trident Refit Facility in Bangor, WA, where he served as docking officer, qualified as ship superintendent at Puget Sound Naval Shipyard, and earned his engineering duty dolphins. Johnson graduated from the Massachusetts Institute of Technology in 1989 with a naval engineer degree and a Master of Science in mechanical engineering. Subsequently, Johnson held submarine acquisition and repair positions at the Supervisor of Shipbuilding in Groton, CT; as a waterfront coordinator delivering Ohio-class submarines and later as the program manager's representative for the Virginia-class submarine; at Trident Refit Facility Bangor as the planning officer; and at Program Executive Officer (PEO) Submarines as the assistant program manager for USS *Jimmy Carter* (SSN 23).

Johnson became major program manager for the Virginia Program Office (PMS 450) in 2005. Under his guidance, the Virginia program reduced overall cost by $4 billion and delivered four submarines to the fleet. The program was awarded the 2007 DoD Value Engineering Award and the 2008 David A. Packard Award for Acquisition Excellence. Johnson also established and served as the first Undersea Enterprise chief technology officer.

Johnson's flag tours include PEO Submarines, deputy commander for Undersea Technology (SEA 073), deputy PEO Submarines for the Ohio SSBN Replacement Program, and commander for the Naval Undersea Warfare Center. In October 2015, he assumed responsibilities as principal military deputy for the Assistant Secretary of the Navy Research, Development, & Acquisition.

Johnson has received various personal and campaign awards, including the Defense Service Medal, Legion of Merit, and the Meritorious Service Medal with three gold stars.

THIS PAGE INTENTIONALLY LEFT BLANK

Plenary Panel: The 809 Panel. Reviewing the Regulatory and Legislative Framework to Improve the Efficiency and Effectiveness of Defense Acquisition

Wednesday, April 26, 2017	
9:30 a.m. – 11:00 a.m.	**Chair: Moshe Schwartz,** Executive Director, 809 Panel ***Current Research Agenda for the 809 Panel and Initial Findings*** William A. LaPlante, Vice President, Intelligence Portfolio, National Security Engineering Center ***Future Research Agenda*** Joseph W. Dyer, VADM, USN (Ret.), Commissioner, Chief Strategy Officer, National Spectrum Consortium

Moshe Schwartz—is a Specialist in Defense Acquisitions at the Congressional Research Service and an adjunct professor at National Defense University's Eisenhower Center. He has written numerous reports and testified before Congress on a variety of issues, including the acquisition of major defense acquisition programs, the use of contractors in military operations, DoD energy, and GAO bid protests. Temporary assignments include serving as a senior advisor to the Commission on Wartime Contracting in Iraq and Afghanistan, and as an advisor at ISAF headquarters in Afghanistan.

Before joining CRS, Schwartz served as a senior analyst at the Government Accountability Office, as an Assistant District Attorney in Brooklyn, NY, and as Vice President of the public relations firm KCSA Strategic Communications.

Schwartz received a BA from Yeshiva University, a JD from Yeshiva University's Benjamin N. Cardozo School of Law, an MBA from Carnegie Mellon's Tepper School of Business, and a master's in Public Policy Management from Carnegie Mellon's H. John Heinz III School of Public Policy and Management

William A. LaPlante—is Vice President of the Intelligence Portfolio in the National Security Engineering Center, a federally funded research and development center that MITRE operates on behalf of the U.S. DoD. In this role, Dr. LaPlante leads key initiatives in support of the nation's intelligence community.

Dr. LaPlante has more than 30 years of experience in defense technology, most recently as Assistant Secretary of the Air Force for acquisition. During his three years in that position, Dr. LaPlante led the $43 billion Air Force acquisition enterprise budget, bringing it into alignment with the greater Air Force vision and strategy. Under his leadership, the Air Force reaped nearly $6 billion in "should-cost" savings—the investment of these savings resulted in greater capability for our nation's warfighters. In recognition of his outstanding performance, the Air Force Association awarded Dr. LaPlante the W. Stuart Symington Award for the most significant contribution by a civilian in the field of national defense. In November 2015, the Air Force bestowed on him its Medal for Exceptional Civilian Service, the highest honor it bestows on a civilian employee. And in 2016, the Massachusetts Institute of Technology Security Studies Program presented him with the General James Doolittle Award in recognition of his contributions to U.S. air power.

Prior to entering public service in 2013, he was MITRE's Missile Defense portfolio director. During this time, Dr. LaPlante was appointed to the Defense Science Board (DSB), where he co-chaired a study on enhancing the adaptability of U.S. military forces. He has resumed his participation in the DSB, where he advises top DoD leadership on critical scientific and technological topics related to the effectiveness of the nation's military forces.

Before joining MITRE, he was the department head for Global Engagement at the Johns Hopkins University Applied Physics Laboratory (APL). In that role he was responsible for all of APL's work supporting offensive military capabilities. He was also a member of the APL Executive Council.

He holds a bachelor's degree in engineering physics from the University of Illinois, a master's degree in applied physics from Johns Hopkins University, and a doctorate in mechanical engineering from Catholic University of America.

Joseph W. Dyer, VADM, USN (Ret.)—Commissioner, Chief Strategy Officer, National Spectrum Consortium.

- Chief Strategist, The Strategy Group
- Chair, NASA Aerospace Safety Advisory Panel
- Former Chief Strategy Officer and Chief Operating Officer, iRobot Corporation
- Former Commander, Naval Air Systems Command

Vice Admiral (Ret.) Dyer is currently an independent consultant in the technology and defense markets.

From 2003 through late 2012, he was an executive at iRobot Corporation, serving consecutively as the President of the Government and Industrial Division, Corporate Chief Operating Officer, and as the company's Chief Strategy Officer.

Vice Admiral Dyer served as Commander of the Naval Air Systems Command, where he was responsible for research, development, test and evaluation, engineering and logistics for naval aircraft, UAVs, and air launched weapons and sensors, from June 2000 until his retirement in July 2003. He was assigned as Commander of the Naval Air Warfare Center Aircraft Division at Patuxent River in July 1997, and a month later he assumed additional responsibilities as the Assistant Commander for Research and Engineering of the Naval Air Systems Command.

From January 1994 to April 1997, Admiral Dyer served as F/A-18 Program Manager, leading engineering and manufacturing development efforts on the new F/A-18E/F, continued production and fleet support of the F/A-18C/D, and all F/A-18 foreign military sales. Under his leadership, the F/A-18 program won the DoD Acquisition Excellence Award and the Order of Daedalian. Earlier in his career, he served as the Navy's chief test pilot.

Vice Admiral Dyer graduated from North Carolina State University with a BS in chemical engineering and from the Naval Postgraduate School with an MS in financial management.

He is an elected Fellow in the National Academy of Public Administration and the Society of Experimental Test Pilots.

Panel 2. Maximizing Contracting Effectiveness

Wednesday, April 26, 2017	
11:15 a.m. – 12:45 p.m.	**Chair: Elliott Branch—**Deputy Assistant Secretary of the Navy (Acquisition and Procurement), ASN (RDA) ***Analyzing the Effects of Source Selection Method, Acquisition Type, and Service Component on Acquisition Outcomes*** Lieutenant Colonel Karen Landale, USAF, Naval Postgraduate School Rene Rendon, Naval Postgraduate School ***Data Consolidation of Disparate Procurement Data Sources for Correlated Performance-Based Acquisition Decision Support*** Samantha Nangia, ASN(RD&A) DASN(AP) Ryan Dickover, ASN(RD&A) DASN(AP) Thomas Wardwell, ASN(RD&A) DASN(AP) Randall Mora, AVUM, Inc. ***Customizing the Use of TINA (Truth in Negotiations Act) in the DoD*** Chong Wang, Naval Postgraduate School

Elliott Branch—is the Deputy Assistant Secretary of the Navy (Acquisition and Procurement) in the Office of the Assistant Secretary of the Navy (Research, Development, and Acquisition). He is the senior career civilian responsible for acquisition and contracting policy that governs the operation of the Navy's worldwide, multibillion-dollar acquisition system. Branch is the principal civilian advisor to the Navy Acquisition Executive for acquisition and procurement matters. He serves as the Department of the Navy's Competition Advocate General, and he is the leader of the Navy's contracting, purchasing, and government property communities.

Prior to joining the Navy Acquisition Executive's staff, Branch was the first civilian Director of Contracts at the Naval Sea Systems Command. In that role, he led one of the largest and most complex procurement organizations in the federal government. As the senior civilian for contracting at NAVSEA, Branch was responsible for the contractual oversight of the nation's most complex shipbuilding and weapons systems procurement programs. His duties involved the obligation and expenditure of approximately $25 billion annually.

He is a member of the Senior Executive Service (SES). Members of the SES serve in the key positions just below the top presidential appointees. They are the major link between these appointees and the rest of the federal work force. SES members operate and oversee nearly every government activity in approximately 75 agencies.

Branch spent time in the private sector, where he specialized in acquisition and project management education, training, and consulting for the federal workforce and its associated contractors. In this role, he was responsible for the design, development, delivery, and maintenance of a wide variety of course materials on subjects ranging from project management to contract law. Branch's clients included Computer Sciences Corporation, QSS Group, BAE Systems, the Pension Benefit Guaranty Corporation, and the Departments of Defense, Energy, Justice, and State.

Prior to that, he served as the Chief Procurement Officer for the government of the District of Columbia, where he was the agency head responsible for procurement operations and policy, and for formulating legislative proposals for local and congressional consideration. Branch led a staff of over

200 employees that supported over 40 city agencies, administered a $14 million annual operating budget, and oversaw the placement of $1.5 billion annually in city contracts.

Before joining the District of Columbia's government, Branch held various positions in the SES with the Department of the Navy (DoN). In 1993, he became a member of the SES as the Director of the Shipbuilding Contracts Division at NAVSEA. He next served as Executive Director of Acquisition and Business Management for the DoN, responsible for policy and oversight of contract operations throughout the entire Navy. While in this position, he also served as Project Executive Officer of Acquisition Related Business Systems. In this role, he was responsible for the formulation and execution of a multi-year effort transforming the Navy's acquisition system from a paper-based system into one that made use of electronic technologies and methods. In this role, Branch was directly responsible for a portfolio of projects worth more than $200 million.

Branch graduated with a Bachelor of Science degree in Economics from the University of Pennsylvania Wharton School and completed the Executive Program at the University of Virginia Darden School. He has received the Navy Distinguished Civilian Service Medal, the David Packard Excellence in Acquisition Award, two Presidential Rank Awards for Meritorious Executive, the Vice Presidential Hammer Award for Reinventing Government, and the 2012 Samuel J. Heyman Service to America Medal for Management Excellence.

Analyzing the Effects of Source Selection Method, Acquisition Type, and Service Component on Acquisition Outcomes

Lieutenant Colonel Karen A. F. Landale, USAF—is an Assistant Professor at the Graduate School of Business and Public Policy, Naval Postgraduate School, Monterey, CA. She teaches strategic sourcing and category management as part of the contracting curriculum. Lt Col Landale received her PhD from the Kenan-Flagler Business School at the University of North Carolina at Chapel Hill. Her research focuses on talent management, services marketing, and contracting in the public domain. She is married to LTC Gordon Landale, USA, and they have two daughters, Amelia and Caroline. [kalandal@nps.edu]

Rene G. Rendon—is an Associate Professor at the Graduate School of Business and Public Policy, NPS, where he teaches defense acquisition and contract management courses. He also serves as the Academic Associate for the MBA specialization in contract management. Prior to joining the NPS faculty, he served for over 20 years as an acquisition contracting officer in the United States Air Force. His career included assignments as a contracting officer for the Peacekeeper ICBM, Maverick Missile, and the F-22 Raptor. He was also a contracting squadron commander and the director of contracting for the Space Based Infrared Satellite program and the Evolved Expendable Launch Vehicle rocket program. Rene has published in the Journal of Public Procurement, the *Journal of Contract Management*, the *Journal of Purchasing and Supply Management*, and the *International Journal of Procurement Management.* [rgrendon@nps.edu]

Introduction

For years, one of the most hotly contested debates in contracting and acquisition has been the choice of source selection method and the contract-related consequences of that choice. While policy memos encourage contracting officers to “select the appropriate source selection process … to match the specific requirement, meet Warfighter needs, and deliver a contracted solution that will provide the required performance levels at the lowest cost” (Kendall, 2015, p. 3), stakeholders on both sides of the table have differing views about how the choice of source selection method affects contract outcomes.

Anecdotally, from the perspective of the government, lowest priced technically acceptable (LPTA) procedures offer a faster time-to-contract, as the technical acceptability criteria is binary and the evaluation of price—the most important factor in LPTA source selections—is objective. Hence, theoretically, the requirement can be put on contract faster, with less likelihood of protest. The sellers’ perspective, however, is that the LPTA source selection method stifles innovation, because price is more important than, say, an innovative approach that may ultimately better serve the government (Calisti, 2015). Critics argue that the LPTA method often results in the selection of a contractor that has undercut the cost of the requirement. They argue that the contractor has essentially achieved “buy-in” by proposing an unreasonably low price that will later have to be adjusted (i.e., increased) via modification in order to fulfill the terms and conditions of the contract. This sort of gamesmanship of the LPTA method has been the argument of federal contractors for many years. Further, opponents of the LPTA method believe the process represents a “race to the bottom” price-wise, and mockingly dub the outcomes achieved by LPTA contracts as “Lousy Project, Tragic Act” (Weckstein & Delgado, 2012). In other words, opponents feel LPTA source selections produce inferior products and services. Proponents suggest this is not the case, and that by providing clear technical acceptability criteria, the government can avoid receiving inferior products and services.

On the opposite spectrum of the best value continuum, the tradeoff (TO) source selection method is anecdotally believed to take more time because of the subjective nature of the evaluation and the increased likelihood of protest. Customers and contractors alike seem to prefer this approach, as it allows customers to feel a certain measure of control over selecting the contractor that represents the best value to the government—that by ranking the evaluation factors in terms of importance, they have the option of tailoring the evaluation to fully meet their needs. Contractors also seem to prefer this method, as it allows them to provide innovative solutions to government requirements, without the burden of competing mainly based on price. Proponents of the TO method argue that it results in higher quality products and services because contractors are not "squeezed" on price. Opponents argue that the method does not necessarily produce better contractual outcomes (i.e., better contract performance), particularly given the anecdotal belief that TO acquisitions take longer to put on contract.

Choosing which method is appropriate for a given acquisition is clearly established by policy and is not the focus of this research. Instead, we aim to use scientific methods to confirm or deny the anecdotal beliefs associated with each source selection method. We use multivariate analysis of variance (MANOVA) and multivariate analysis of covariance (MANCOVA) methods to determine if statistically significant differences in contract outcomes exist based on source selection method. This first-of-kind research uses actual contract file data from the Air Force and Navy to test hypotheses associated with the anecdotal beliefs. Specifically, we examine whether differences exist in Contractor Performance Assessment Reporting System (CPARS) scores and procurement administrative lead time (PALT) based on choice of source selection method (LPTA or TO), while taking into account several different covariates related to the acquisitions.

The remainder of this paper proceeds as follows: The Literature Review section provides a detailed review of the contract management process, the best value continuum, and the relationship between contract type and source selection method. Following that is a discussion of the data collection and analysis methodologies, results of the analysis, and finally, a review of practical and managerial implications, as well as limitations and areas for further research.

Literature Review

Contract Management Process

The contract management process consists of three main phases that encompass six basic steps (for a more thorough review, see Garrett, 2010). The pre-award phase consists of three steps: procurement planning, solicitation planning, and solicitation. The award phase consists of just one step: source selection. Finally, the post-award phase consists of two steps: contract administration and contract closeout.

The first step, procurement planning, involves determining whether the government should produce the requirement organically or outsource production. This is known as the "make or buy" decision. Procurement planning also involves scoping out the requirement, conducting market research, and discussing acquisition strategy in terms of the type of contract to use, the appropriate source selection method, and the appropriate procurement method (sealed bidding or contracting by negotiation). The results of market research will indicate the availability of commercial items or services that meet the requirement, the nature of the competitive environment, and the variability in the technology used in industry to develop the supplies or services. Based on the results of the market research, the solicitation document can be developed.

The second step, solicitation planning, occurs after the decision to outsource has been made. In solicitation planning, the acquisition team continues to refine the requirement and the procurement methods, and it establishes the evaluation criteria that will be used to select a contractor. Clearly, these first two steps—procurement planning and solicitation planning—have a significant impact on the resulting success or failure of the contract. Poor planning or an inadequate requirement definition in the procurement and solicitation planning steps can result in unclear solicitation documents or in the inability to properly evaluate and choose an offer that represents the best value for the government. It is critical that the acquisition team has a clear understanding of the requirement, of how it will be solicited, and of how proposals will be evaluated. Any confusion or uncertainty will be passed on to potential offerors, who may interpret the requirement differently. In terms of the research performed in this study, these first two steps are the most impactful.

The third step, solicitation, involves publicizing the requirement and instructing potential offerors how, where, and when to submit their proposals. Clarifying questions often arise, and the government buyer ensures all questions are answered and provided to all potential offerors. Lamoreux, Murrow, and Walls (2015) note that

> the fourth step, source selection, involves using the evaluation criteria established during the solicitation planning step and specified in the solicitation document to formally evaluate each offer. Depending on the size and complexity of the procurement, this may involve source selection boards, technical panels, and any other expert required to evaluate the offers received. Further, the source selection may involve directly negotiating with one or more vendors on price, technical factors, or personnel. Finally, the acquisition team selects the winner during this step; it is the most vulnerable to protests from unsuccessful vendors. (p. 15)

A successful source selection is a reflection of a successful planning process. Source selection is the execution of the evaluation strategy that was designed during solicitation planning, which highlights the importance of ensuring the acquisition team has adequate time to properly plan for the acquisition.

The fifth step, contract administration, is typically the longest step in terms of the overall life of the acquisition. In this step, the contractor produces the good or service, and the government monitors performance and provides feedback. Both parties play an active role in ensuring the terms and conditions of the contract are enforced.

Finally, contract closeout, the sixth step, involves confirming that all work has been accomplished and the contractor has been paid in full before finalizing contract details and closing the contract. This step also includes the important task of assessing the contractor's performance using the Contractor Performance Assessment Reporting System (CPARS).

Best Value Continuum

In government contracting, the best value continuum recognizes the fact that there are a variety of ways in which an organization can obtain the best value for their dollar. The Federal Acquisition Regulation (FAR) states,

> An agency can obtain best value in negotiated acquisitions by using any one or a combination of source selection approaches. In different types of acquisitions, the relative importance of cost or price may vary. For example, in acquisitions where the requirement is clearly definable and the risk of unsuccessful contract performance is minimal, cost or price may play a dominant role in source selection. The less definitive the requirement, the

> more development work required, or the greater the performance risk, the more technical or past performance considerations may play a dominant role in source selection. (FAR 15.101)

For practical purposes, we typically envision the best value continuum using its poles: on one end is LPTA, and on the other, TO.[1] Both strategies can result in the best value to the government, but selecting a proposal that represents the best value varies for each method.

In LPTA source selections, best value is obtained by choosing the lowest priced offer that still meets established minimum quality thresholds (i.e., technical acceptability). The government establishes minimum thresholds and conveys them via the solicitation document. LPTA works best when the requirement is well-defined and the risk of unsuccessful performance is minimal. It “should be used in situations where the DoD would not realize any value from a proposal exceeding its minimum technical or performance requirements” (Government Accountability Office [GAO], 2014, p. 6). The LPTA method is typically used in contracting commercially-available goods or services, as the market has already established reasonably acceptable quality levels, and, assuming an adequate number of offerors supply the market, competition is based on price alone. Source selection for an LPTA requirement is typically performed by ranking the proposals from lowest to highest price, then evaluating whether the lowest-priced proposal meets the minimum quality thresholds (i.e., whether the lowest-priced proposal is technically acceptable).[2] If it is, the evaluation stops, and the lowest-priced offeror is declared the winner. If the lowest-priced proposal is not technically acceptable, it is removed from the competition and the next lowest-priced proposal is evaluated for technical acceptability. The process continues until the evaluation team finds the lowest-priced, technically acceptable offer. In general,

> LPTA acquisitions tend to be simpler than tradeoffs, [as] contracting offices can move more quickly through the six-step contract management process, reducing administrative operating costs. [T]he generally inflexible nature of the LPTA source selection method does not grant contracting officers discretion, which serves as a guard against the appearance of favoritism, promoting the perception of integrity, fairness, and openness. (Lamoureux et al., 2015, p. 20)

TO source selections, on the other hand, acknowledge that best value may result from higher quality ratings, which might consist of a host of factors (e.g., technical capability, management practices, past performance, etc.), and that higher quality may cost more. The TO method allows the government to establish which evaluation factors are most important and which are less important, and the government is allowed to trade cost or price factors for non-cost or non-price factors. Using the TO method “is appropriate when it may be in the best interest of the Government to consider award to other than the lowest priced offeror or other than the highest technically rated offeror” (FAR 15.101-1(a)). In a memorandum

---

[1] In reality, LPTA is on one end of the continuum and highest technically rated offer (HTRO) is on the other. Because the FAR requires the evaluation of cost or price in each source selection, the federal government can never make an award based only on the HTRO.

[2] Ranking of non-price criteria is not permitted. Technical acceptability is binary: A proposal is technically acceptable or it is not.

detailing the appropriate use of source selection processes, Under Secretary of Defense for Acquisition, Technology, and Logistics, Frank Kendall, asserts that "whenever the Warfighter is willing to pay more for above threshold requirements or performance standards and may benefit from an innovative and technologically superior solution to meet their needs, a tradeoff source selection process between cost or price and non-cost factors is optimal" (Kendall, 2015, p. 2).

Offerors still have to meet minimum standards; however, they may be rewarded for surpassing minimum standards where advantageous to the government. The government must establish how they will assess each offeror's quality, cost, and past performance, as well as the relative importance of these factors and any subfactors. The government communicates the importance of each evaluation factor through numerical and/or textual ranking specified in the solicitation document. For instance, the government might state that technical capability is twice as important as cost, which is twice as important as past performance. This implies a sort of numerical ranking (e.g., technical capability is worth 40 points, cost is worth 20 points, and past performance is worth 10 points). Alternatively, a textual ranking might say something like "technical capability is significantly more important than cost, which is more important than past performance." Using this sort of language implies that the technical capability is the most important factor, and that it is much more important than either cost or past performance. Naturally, the evaluation of "significantly more important" or "more important" are left open to interpretation when comparing offerors. This sort of subjective assessment provides the government the flexibility to select the offeror that represents the best value to the government; however, it is also subject to potential pitfalls. One potential pitfall is that the evaluation of each offer may take more time and involve many rounds of internal discussions. A second potential pitfall is that the subjective nature of the assessment results in higher risk of the government failing to comply with the evaluation process as stated in the solicitation, which can result in a protest that delays the acquisition.

The TO method works best for complex acquisitions where requirements are not well-defined, and where increased contractor capability could make the acquisition less risky. Source selection for a TO requirement typically involves a source selection authority, a source selection advisory council, and a source selection evaluation board. Members of the source selection evaluation board evaluate each evaluation criteria independently, scoring proposals according to the source selection procedures established in the solicitation. The independent scores for each evaluation criterion are presented to the source selection advisory council, which then makes an award recommendation to the source selection authority. The source selection authority is the ultimate decision-maker—they can choose to accept the recommendation or choose a different offeror for the award.

Clearly, the TO source selection process is more bureaucratic than its LPTA counterpart. Further, because of the subjectivity involved in evaluating and rating proposals, TO source selections are often more susceptible to protests. However, "proponents of tradeoffs argue that the initial costs of a higher-priced vendor are ultimately more efficient, as the incentive structure encourages vendors to avoid cutting costs that could jeopardize the effort after award" (Lamoureux et al., 2015, p. 21).

In sum, the best value continuum balances the need to receive quality goods and services for the customer with the need to procure those goods and services in a way that is fiscally responsible for the taxpayer. Many articles and reports discuss the implications of choosing one source selection method over the other, see, for example, GAO (2014), Duncombe and Prentice (2013), and Nichols and Totman (2013). For most acquisitions, the choice of source selection method that best fits the requirement is clear. However, some acquisitions do fall into gray territory, and for those, the choice of source selection method ultimately comes down to a cost-benefit analysis. For a detailed discussion of the costs and benefits of each method, particularly the tradeoff method, see Lamoureux et al. (2015).

Contract Type and Source Selection Methodology

While contract type and source selection methodology are two distinct decisions, source selection method is influenced by contract type. Further, both decisions are influenced by the type of requirement being outsourced and the results of market research during the procurement planning step.

In federal government contracting, there are two overarching contract types: fixed-price and cost-type contracts. FAR 16.202-1 states,

> A firm-fixed-price contract provides for a price that is not subject to any adjustment on the basis of the contractor's cost experience in performing the contract. This contract type places upon the contractor maximum risk and full responsibility for all costs and resulting profit or loss. It provides maximum incentive for the contractor to control costs and perform effectively and imposes a minimum administrative burden upon the contracting parties. (FAR 16.202-1)

Understanding this risk, contractors often apply a buffer in their proposed pricing to account for uncertainty. The more complex the requirement, the larger the buffer. For this reason, fixed-price contracts are typically used for commercial products and services. Naturally, the more clearly-defined the requirement, the more it lends itself to the LPTA source selection method, where price is considered the most important factor. In other words, when the requirements are well-defined and technical acceptability is easy to describe and evaluate, the determining factor for award is price—hence the relationship between fixed-price contracts and the LPTA source selection method.

On the other hand,

> Cost-reimbursement types of contracts provide for payment of allowable incurred costs, to the extent prescribed in the contract. These contracts establish an estimate of total cost for the purpose of obligating funds and establishing a ceiling that the contractor may not exceed (except at its own risk) without the approval of the contracting officer. (FAR 16.301-1)

Unlike fixed-price contracts, which are recommended for use whenever practical, cost-reimbursement contracts should only be used when the requirement cannot be sufficiently defined or when uncertainties in contract performance do not allow costs to be estimated sufficiently for a fixed-price arrangement (see FAR 16.301-2). In fact, "acquisition teams are prohibited from using cost-reimbursement contracts to procure commercial items, limiting their use to complex, uniquely governmental efforts" (Lamoureux et al., 2015, p. 17). Given the unique nature of many defense-related needs, it is not always possible for the federal government to have a well-defined requirement. Many of the weapons systems it procures have no equivalent anywhere in the world—they are purposefully different and represent innovative capabilities to achieve competitive advantage over our adversaries. Because they are "new to the world" requirements, they are often less defined and more difficult to clearly articulate to potential offerors. Less defined, more complex requirements are better procured using cost-type contracts. Because of the need for innovative solutions, cost-type contracts typically lend themselves to the TO source selection method, where the cost/price factor can be traded off for more important factors, such as technical capability.

Hypotheses

The purpose of this research is to empirically analyze popular assumptions related to source selection method and subsequent contract outcomes. To do this, we test four hypotheses.

Given that LPTA source selections typically occur when requirements are well-defined and lower risk, and the fact that LPTA source selections generally lend themselves to greater objectivity than TO source selections, we posit that LPTA source selections are faster (i.e., take less time from requirement generation to contract award) than TO source selections:

- Hypothesis 1: LPTA acquisitions have a shorter PALT than TO acquisitions.

Further, given that TO source selections are more flexible in allowing the government to trade cost/price for non-cost/non-price factors, and that TO source selections allow the acquisition team to rank the evaluation factors to best meet the needs of the requirement, we posit that TO source selections result in better contract performance[3] than LPTA source selections:

- Hypothesis 2: TO acquisitions produce higher CPARS scores than LPTA acquisitions.

We also examine whether different types of acquisitions (product acquisitions versus service acquisitions) produce different PALTs or CPARS scores. Because the data we collected were from systems-level buying organizations and/or from high dollar value contracts, the products and services acquired are more complex than those typically purchased at the installation level. Thus, given the similarity in complexity, we find no reason why product acquisitions and service acquisitions, using the same general procedures, would produce different contract outcomes:

- Hypothesis 3a: There is no difference in PALT between product acquisitions and service acquisitions.
- Hypothesis 3b: There is no difference in CPARS scores between product acquisitions and service acquisitions.

Next, we examine whether or not the contract outcomes are different between the service components. Because all service components are subject to the Federal Acquisition Regulation (FAR) and its Defense supplement (DFARS), we find no reason why different service components using the same general procedures would produce different contract outcomes:

- Hypothesis 4a: There is no difference in PALT between service components.
- Hypothesis 4b: There is no difference in CPARS scores between service components.

With the hypotheses in place, we turn to the details regarding the data and the analyses.

---

[3] We use contractor performance (i.e., CPARS scores) as a surrogate measure for contract performance. The rationale is that if the contractor's performance is successful, the contract would also be considered successful.

Methodology

Data Collection

To collect the data required for this research, five teams of graduate students traveled to seven different Air Force and Navy contracting offices and pulled the data from actual contract files. Our goal was to choose contracts that were as similar in complexity as possible in order to better understand the effects that source selection method might have on contract outcomes. Thus, we purposely chose to collect data from systems-level buying organizations and/or high dollar value contracts. It is important to note that the contracting databases currently used in the Department of Defense (DoD) do not automatically collect these data. Thus, "scraping" the data from the physical contract files was required.[4]

Variables Examined

In group comparison statistical methods, like the ones used in this study, independent variables (IVs) serve as the grouping variables. They are categorical in nature (i.e., no single observation can belong to more than one group) and have at least two different categories, or groups. We have three IVs for this study: choice of source selection method (LPTA or TO), acquisition type (product or service), and service component (Air Force or Navy). Each IV is binary, where LPTA, product, and Air Force all equal zero, and TO, service, and Navy all equal one.

Dependent variables (DVs) are variables whose values depend on the IV. For this reason, they are often termed "outcome" or "response" variables. The DVs we chose for this study are meant to provide answers about how long the contracting process took (a process metric) and how well the contractor performed (a performance metric). Accordingly, we chose (1) PALT as the measure of time-to-contract and (2) CPARS scores as a measure of contractor performance. PALT is measured by the number of days from requirement identification to contract award. Consistent with FAR 42.15, CPARS data were collected and used for the following reporting categories: (1) cost control, (2) quality, (3) schedule, (4) business relationship, and (5) subcontracting. CPARS measures each category using the following Likert-style scale: 1 = unsatisfactory, 2 = marginal, 3 = satisfactory, 4 = very good, and 5 = excellent. These scores serve as a proxy for contractor performance, with higher numbers indicating better performance. Although we have CPARS data for each category, the average across the first four categories was used in this research, as the subcontracting category had relatively few cases, and the listwise deletion resulted in too few cases to run the analyses.

Covariates are secondary variables that can also affect the relationship of primary interest: the relationship between the IV and the DV. For this study, our goal is to parcel out the effects of covariates in order to more clearly see the relationship between the IVs and the DVs. We identified six potential covariates: (1) contract dollar value (VALUE), (2) number of reviews the solicitation and contract were subject to prior to award (NUMREVIEWS), (3) number of evaluation factors in the source selection plan

[4] While not the focus of this study, we found during the course of our research that a more comprehensive database is needed that captures many metrics the DoD should be capturing in order to quickly and continuously monitor performance of our contracts and contracting processes. See the Areas for Further Research section for more details.

(NUMEVALFACT), (4) number of offers received (NUMOFFERS), (5) number of contract line items in the contract (NUMCLINS), and (6) number of people on the source selection team (NUMPEOPLE). Each of these covariates could potentially affect PALT and/or CPARS scores, thus our goal was to parcel out their effect(s) in order to more clearly understand the effect of the IVs on the DVs.

Data Description

Our sample consists of 139 cases, which is sufficient for accurate analysis. The distribution of cases is unbalanced for each IV. There are 61 LPTA cases and 78 TO cases; 40 product acquisition cases and 99 service acquisition cases; and 52 Air Force cases and 87 Navy cases. This unbalanced design can cause ambiguity about the mean as the intercept and make assignment of sums of squares more difficult. There are, however, solutions to these issues. A weighted mean can be used in place of the grand mean and the STATA software (v12) we used for these analyses automatically handles the assignment of the sums of squares. Thus, we proceed with our analysis despite these limitations.

Analysis

Because our intent is to analyze differences in contract outcomes based on source selection methodology, acquisition type, and service component, a group comparison statistical methodology is necessary. We seek to find if there are differences in contract outcomes by group, both excluding and including the effect(s) of covariates.

MANOVA/MANCOVA

We use both multivariate analysis of variance (MANOVA) and multivariate analysis of covariance (MANCOVA) to assess group differences. Both methods create a new dependent variable using the information from the given dependent variables (PALT and CPARS scores). This new dependent variable is created in a way that maximizes differences between the grouping variable (the IVs).[5] Clearly, the only difference between the two methods is the use of covariates: MANOVA looks for differences in the DVs using the IVs only, while MANCOVA takes into account the effects of covariates when looking for group differences in the DVs. We use both methods to gain a better understanding of the effects of the covariates on the DVs. We describe the more complex method (MANCOVA) in detail.

MANCOVA addresses the following questions: Are mean differences among the groups on a combination of DVs (after adjusting for covariate effects) likely to have occurred by chance? Taken from another angle, is there a significant difference between the mean value for PALT and CPARS scores in LPTA source selections versus the mean value for PALT and CPARS scores in TO source selections, once the effects of the covariates have been parceled out?

If differences in outcomes are found using MANCOVA, we dig deeper to better understand the differences using univariate analysis of covariance (ANCOVA). ANCOVA

[5] There is much debate as to whether Likert-type items like those used in our DV CPARS are considered interval or ordinal (see Carifio and Perla, 2007, for a review). MANCOVA requires the items be considered interval; however, we also used the Kruskal-Wallis H test, which considers CPARS an ordinal variable, to confirm our results. Due to space limitations, the results of this analysis are available from the first author.

also assesses group differences, however because it is univariate in nature, this method assesses one DV at a time (PALT *or* CPARS individually, rather than PALT *and* CPARS simultaneously). This method helps isolate where the difference(s) is (are) occurring.

Assumption Testing

Like all statistical methods, MANCOVA requires that certain assumptions about the data be tested to ensure accurate results. We tested six assumptions before proceeding. Beginning with 147 raw observations, we first searched for multivariate outliers using Mahalanobis' Distance. We found four outliers and chose to drop those observations from subsequent analyses (n = 143), as outliers are known to significantly affect MANCOVA (Tabachnick & Fidell, 2007). Further, we searched for univariate outliers in each cell of our design (see Table 1). We found four univariate outliers, which were subsequently deleted (n = 139).

Table 1. Cell Design

Cell Design		
	Air Force	**Navy**
Product Acquisition	LPTA (6)	LPTA (18)
	TO (2)	TO (14)
Service Acquisition	LPTA (13)	LPTA (24)
	TO (31)	TO (31)

Second, we tested multivariate normality among the DVs by examining density graphs, determining multivariate skewness and kurtosis in order to identify variables that might require transformation. For the DVs, PALT was deemed to be non-normal and was normalized via a logarithmic transformation. We also considered the normality of the covariates, as covariates are useful in reducing error, but not if they are non-normal and thus reduce power (Tabachnick & Fidell, 2007). Several covariates required transformation. Specifically, VALUE, NUMREVIEWS, NUMOFFERS, and NUMCLINS all received a logarithmic transformation, and to normalize NUMPEOPLE, the square root was taken. After these transformations, all variables were deemed to be multivariate normal. All further analyses and statistical output use the transformed variables, however the written results back-transform the variables into their original form for a better understanding of the effects. We use the untransformed variable nomenclature in the text for ease of reading.

Third, we assessed linearity by examining scatter plots of (1) the paired DVs, (2) all pairs of covariates, and (3) all pairs of DV-covariate combinations for each grouping variable (a total of 168 plots). The plots revealed that NUMCLINS and NUMPEOPLE were consistently not linear, thus those covariates were removed from further analyses. Other variables failed linearity sporadically, and we were careful to remove offending pairings.

Fourth, we assessed homogeneity of regression for each DV and grouping variable (a total of 24 assessments). This test was performed using an analysis of covariance (ANCOVA) that included the independent variables, each of the remaining covariates (VALUE, NUMREVIEWS, NUMEVALFACT, and NUMOFFERS), and the interaction between the independent variables and the covariates. When the interaction terms are not significant, the relationship between the dependent variables and each of the remaining covariates is the same at both levels of the independent variables, and the assumption of homogeneity of regression is upheld. There were three violations of homogeneity of regression. First, when PALT is the DV, the interactions between service component and number of reviews is significant. This means that the number of reviews the contracts we examined went through differed significantly between the Air Force (mean = 6.5 reviews) and the Navy (mean = 5.5 reviews). Second, the interaction between acquisition type and number of offers is significant when PALT is the DV, meaning that product acquisitions (mean = 4.23 offers) receive significantly different number of offers than service acquisitions (mean = 4.43 offers). Finally, when CPARS scores is the DV, the interaction between service component and value is significant. Again, the value of the contracts differs significantly between the Air Force (mean = $52,000,000) and the Navy (mean = $32,300,000). We were careful to remove the offending covariates, where appropriate, from our analyses.

Fifth, we checked for multicollinearity by assessing the pooled within cell tolerance for each DV. The DVs are not highly correlated in any cell, thus multicollinearity is not an issue.

Finally, we checked for homogeneity of covariance matrices between groups using the multivariate test of means provided in STATA (v12). This test checks whether or not population variances and covariances of both dependent variables are equal for each of the IV groups. The results showed that all grouping cells are homogenous (source selection method: Box's M $\chi^2(3) = 1.88$, $p = .5967$; acquisition type: Box's M $\chi^2(3) = 4.12$, $p = .2484$; and service component: Box's M $\chi^2(3) = 6.08$, $p = .1078$).

Table 2 provides the remaining covariates available for each MANCOVA and subsequent ANCOVA.

Table 2. Covariates Available for MANCOVA/ANCOVAs

Covariates Available for MANCOVA/ANCOVAs			
Grouping Variable	**MANCOVA: PALT & CPARS Scores**	**ANCOVA: PALT**	**ANCOVA: CPARS Scores**
Source Selection Method	VALUE* NUMEVALFACT NUMOFFERS	VALUE NUMREVIEWS NUMEVALFACT NUMOFFERS	NUMEVALFACT NUMOFFERS
Acquisition Type	VALUE NUMEVALFACT	VALUE NUMREVIEWS NUMEVALFACT	VALUE NUMEVALFACT NUMOFFERS
Service Component	NUMEVALFACT*	VALUE NUMEVALFACT	NUMOFFERS

*Although these variables are not fully linear with both DVs, their departure from linearity was minor. We tested the MANCOVAs with and without these variables, and the results were similar. We chose to include them in our analyses.

With all assumptions tested, we performed the MANOVAs and MANCOVAs. The results are provided in the next section.

Results

Descriptive Statistics

Basic descriptive statistics for each variable are shown in Table 3. The table presents results for each grouping variable.

Table 3. Descriptive Statistics

Descriptive Statistics						
Variable	Obs	Mean	StdDev	Min	Max	Grouping Variable
PALT (days)	**133**	**303.02**	**271.71**	**3**	**1019**	-
	60	143.38	110.02	3	482	LPTA SS
	73	434.22	294.52	21	1019	Tradeoff SS
	38	228.79	198.03	3	953	Product Acq
	95	332.71	291.75	8	1019	Service Acq
	51	329.10	294.40	21	1019	Air Force
	82	286.79	257.13	3	990	Navy
CPARS (average rating)	**69**	**4.00**	**.78**	**2.5**	**5**	-
	20	3.63	.67	3	5	LPTA SS
	49	4.15	.79	2.5	5	Tradeoff SS
	14	3.50	.64	2.5	5	Product Acq
	55	4.13	.77	3	5	Service Acq
	35	4.07	.78	3	5	Air Force
	34	3.93	.80	2.5	5	Navy
Contract Dollar Value	**139**	**$39,700,000**	**$85,800,000**	**$27,819**	**$450,000,000**	-
	61	$9,846,556	$57,400,000	$27,819	$450,000,000	LPTA SS
	78	$63,000,000	$96,800,000	$36,000	$432,000,000	Tradeoff SS
	40	$32,100,000	$84,900,000	$145,481	$450,000,000	Product Acq
	99	$42,700,000	$86,300,000	$27,819	$432,000,000	Service Acq
	52	$52,000,000	$105,000,000	$36,000	$432,000,000	Air Force
	87	$32,300,000	$71,300,000	$27,819	$450,000,000	Navy
Number of Reviews	**118**	**5.89**	**5.83**	**1**	**28**	-
	56	5.77	5.46	1	25	LPTA SS
	62	6.00	6.19	1	28	Tradeoff SS
	35	4.11	4.12	1	22	Product Acq
	83	6.65	6.28	1	28	Service Acq
	44	6.52	6.05	1	28	Air Force
	74	5.53	5.69	1	25	Navy
Number of Evaluation Factors	**129**	**2.67**	**.86**	**1**	**5**	-
	55	2.13	.55	1	3	LPTA SS
	74	3.07	.83	2	5	Tradeoff SS
	35	2.40	.77	1	4	Product Acq
	94	2.77	.87	1	5	Service Acq
	48	2.42	.61	1	4	Air Force
	81	2.81	.95	1	5	Navy
Number of Offers	**139**	**4.37**	**4.33**	**1**	**23**	-
	61	3.85	4.39	1	23	LPTA SS
	78	4.78	4.27	1	22	Tradeoff SS
	40	4.22	3.39	1	12	Product Acq
	99	4.43	4.67	1	23	Service Acq
	52	6.40	5.70	2	23	Air Force
	87	3.16	2.63	1	12	Navy
Data presented is in its original form, before transformation.						

MANOVA

We begin with a series of MANOVAs to determine the primary effects the grouping variables have on contract outcomes. We examine all three grouping variables together, and then each individually. Where significant effects are found, ANOVA is used to identify which outcome variable(s) is(are) affected.

Using all three grouping variables, we find the overall model is significant (*Wilks' Λ* = .7141, $F(6, 118) = 3.61$, $p < .01$). *Wilks' Λ* is high, suggesting that 71% of the variance in the outcome variables is not explained by the three grouping variables. The results suggest that the source selection method is driving significance.

When source selection method is the only grouping variable, the model is again significant (*Wilks' Λ* = .7818, *F*(2, 61) = 8.51, $p < .01$). Follow-up ANOVAs show that both PALT and CPARS scores are significantly affected by source selection method (PALT, *F* (1, 131) = 45.34, $p < .01$, partial $\eta^2 = .25$ and CPARS scores, *F* (1, 67) = 6.50, $p < .05$, partial $\eta^2 = .09$). Although source selection method significantly affects both PALT and CPARS scores, it has a much more profound impact on PALT than on CPARS scores. The mean PALT for TO acquisitions is 67% longer than the mean PALT for LPTA acquisitions, whereas the mean CPARS rating for TO acquisitions is 13% higher than the mean CPARS rating for LPTA acquisitions. These results support Hypotheses 1 and 2.

Using only acquisition type as the grouping variable, the model is not significant (*Wilks' Λ* = .9228, *F*(2, 61) = 2.55, ns). However, post-hoc ANOVAs indicated that there is a significant difference in CPARS scores between product and service acquisitions (*F* (1, 67) = 7.85, $p < .05$, partial $\eta^2 = .10$). The mean CPARS score for service acquisitions is 15% higher than the mean CPARS score for product acquisitions. These results support Hypothesis 3a, but not 3b.

When service component was used as the grouping variable, the model was not significant (*Wilks' Λ* = .9876, *F*(2, 61) = .38, ns). Post-hoc ANOVAs found the same—there were no statistically significant differences in PALT or CPARS ratings between Air Force acquisitions and Navy acquisitions, supporting Hypotheses 4a and 4b.

MANCOVA

With the primary tests of the IVs on the DVs complete, we add covariates to our model to determine the impact that related aspects of the acquisition process have on PALT and CPARS scores.

Using source selection method as the grouping variable, we find the model is significant (*Wilks' Λ* = .5110, *F*(8, 106) = 5.28, $p < .01$). The substantially lower *Wilks' Λ* shows that an additional 27% of the variance in the DVs is captured when the covariates are included. Further highlighting the importance of the covariates, the univariate ANCOVAs show that when the covariates are included, source selection method is no longer significant. Instead, it is the value of the acquisition (*F* (1, 106) = 9.53, $p < .01$, partial $\eta^2 = .08$), the number of evaluation factors (F (1, 106) = 6.27, p < .05, partial $\eta^2 = .06$), and the number of offers (*F* (1, 106) = 7.02, $p < .01$, partial $\eta^2 = .06$) that significantly affect PALT, while no variables significantly affected CPARS scores. The mean number of evaluation factors and number of offers for LPTA acquisitions is 2.13 and 3.85, respectively, while the mean for TO acquisitions is 3.07 and 4.78, respectively. These results do not support Hypotheses 1 or 2; seeming to disprove popular assumptions about the effects of source selection method on contract outcomes. Because MANCOVA represents a more realistic view of the acquisition process, we deem these results to be more robust than the MANOVA results.

Using acquisition type as the grouping variable, the model is significant (*Wilks' Λ* = .5156, *F*(6, 108) = 7.07, $p < .01$). With the covariates included, we were able to capture an additional 41% of the variance in the DVs. The univariate ANCOVAs showed that it is again the value of the acquisition (*F* (1, 107) = 13.20, $p < .01$, partial $\eta^2 = .11$) and the number of evaluation factors (*F* (1, 107) = 7.60, $p < .01$, partial $\eta^2 = .07$) that significantly affect the PALT. In this analysis, CPARS scores are affected by acquisition type, with services acquisitions receiving statistically significantly higher CPARS scores than product acquisitions (*F* (1, 58) = 6.59, $p < .05$, partial $\eta^2 = .10$). CPARS scores were 15% higher for service acquisitions than for product acquisitions. These results support Hypotheses 3a, but not 3b—with the covariates included, there is no difference in PALT between product acquisitions and service acquisitions; but there is a difference in CPARS scores between product acquisitions and service acquisitions.

Next, using service component as the grouping variable, the model was significant (*Wilks' Λ* = .6839, *F*(4, 110) = 5.75, *p* < .01). We were able to capture an additional 30% of the variance in the DVs by including the covariates. The univariate ANCOVAs once again showed the importance that value (F (1, 120) = 22.54, $p < .01$, partial $\eta^2 = .16$) and the number of evaluation factors (F (1, 120) = 4.81, $p < .05$, partial $\eta^2 = .04$) has on PALT. No variables significantly affected CPARS scores. These results support Hypotheses 4a and 4b—with the covariates included, there are no differences in PALT or CPARS scores between the Air Force acquisitions and Navy acquisitions.

Discussion & Conclusion

We broke new ground in this research by scientifically testing popular assumptions related to source selection methods and their subsequent contract outcomes. Further, we empirically showed that contract outcomes are the same between service components, but not necessarily between acquisition types. The following practical and managerial implications are provided for this research.

Using a simple MANOVA, we found there are significant differences in contract outcomes based on source selection method. The subsequent ANOVAs showed that PALT was 67% longer for TO source selections than for LPTA source selections. Further, the CPARS scores were, on average, 13% higher for TO source selections when compared to LPTA source selections. While these results appear to lend credence to popular anecdotes, when the details of the acquisition (i.e., the covariates) were included in the analysis, source selection method did not affect PALT or CPARS scores. Instead, it was the value of the acquisition, the number of evaluation factors, and the number of offers that affected PALT; and no variables affected CPARS scores.

This is a very interesting finding, as it suggests that it is the *details and processes of the acquisition itself*—some of which are controllable by the acquisition team—that affect time-to-contract. Specifically, the more evaluation factors included in the solicitation, the more time it takes to evaluate them and award a contract. This is common sense, of course, but the finding generates the question of the optimal number of evaluation factors. Is there a tipping point at which the number of evaluation factors included in the solicitation significantly affects PALT? Finding the answer to this question would help procurement teams plan their solicitations accordingly—either reducing the number of evaluation factors, or planning for extra time to assess many evaluation factors.

Further, the number of offers a requirement receives significantly affects PALT. Those in the field understand this finding, as more offers require more time to properly evaluate before awarding a contract. The number of offers a requirement receives is related to how wide the procurement team "cast the net"—how many offerors in the market were eligible to receive the contract. A procurement team can cast a wide net by using full and open competition solicitation methods, or they can cast a narrower net by limiting eligible respondents to small businesses, or even a sole source, when justified. Clearly, there is a balance to achieve between inspiring maximum competition and awarding the contract in the desired amount of time. Maximum competition often results in lower prices and increased quality, but comes with the cost of extended evaluation time, and, thus, a longer PALT. Minimizing competition might allow the contract to be awarded faster, but the procurement team may not achieve the best business deal. Here, again, the question of the optimal number of offers is raised.

Finding that there were significant differences in CPARS scores based on the type of acquisition is also interesting. We found that service acquisitions receive 15% higher CPARS scores than product acquisitions. These results may highlight criticism the DoD has received in recent years concerning the department's failure to properly evaluate/score service contract performance (e.g., lack of proper scoring metrics, failure to properly oversee and measure service performance, etc.). Higher CPARS scores for service acquisitions might reflect these failures. Without adequate justification to downgrade performance, scores may be artificially high.

Finally, we found no significant differences in PALT or CPARS scores between service components. This suggests that federal regulations, policies, and practices are being applied in a uniform manner across service components.

In summary, we feel the most important finding of this research is that the covariates matter. In other words, it is the details of the acquisition, solicitation document, and source selection processes that affect the time-to-contract. Each service component should ensure their processes are expedient and supportive, always aimed at producing optimal contract outcomes for the customer in a way that is least burdensome for the acquisition team.

Areas for Further Research

More data are needed to substantiate the results found in this research. Access to more data might also permit more covariates to pass assumptions, allowing for more comprehensive analyses.

References

Calisti, S. R. (2015, March–April). Lowest price technically acceptable: Why all the debate? Defense AT&L, 17–21.

Carifio, J., & Perla, R. (2007). Ten common misunderstandings, misconceptions, persistent myths, and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences, 3*(3), 106–116.

Duncombe, B. A., & Prentice, J. D. (2013). Lowest price technically acceptable procurements: The good, the bad, and the reality. *Dayton Business Journal*. Retrieved from http://www.bizjournals.com/dayton/news/2013/10/09/lowest-price-technically-acceptable.html

Federal Acquisition Regulation (FAR), 48 C.F.R. ch. 1 (2016).

Garrett, G. (2010). *World class contracting*. Riverwoods, IL: CCH Incorporated.

Government Accountability Office (GAO). (2014). *Defense contracting: Factors DOD considers when choosing best value processes are consistent with guidance for selected acquisitions* (GAO-14-584). Washington, DC: U.S. Government Printing Office.

Kendall, F. (2015). *Appropriate use of lowest priced technically acceptable source selection process and associated contract type*. Retrieved from http://bbp.dau.mil/docs/Appropriate_Use_of_Lowest_Priced_Technically_Acceptable_Source_Selec_Process_Assoc_Con_Type.pdf

Lamoureux, J., Murrow, M., & Walls, C. (2015). *Relationship of source selection methods to contract outcomes: An analysis of Air Force source selection* (MBA project). Retrieved from Calhoun Database: http://hdl.handle.net/10945/47987

Nichols, R., & Totman, J. C. (2013). Feature comment: Myth-busting the LPTA conundrum. *The Government Contractor, 55*(46). Retrieved from https://www.cov.com/~/media/files/corporate/publications/2013/12/myth-busting_the_lpta_conundrum.pdf

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Boston, MA: Pearson Education.

Weckstein, K., & Delgado, A. (2012). *The growing backlash against LPTA source selection*. Law 360. Retrieved from http://www.brownrudnick.com/uploads/117/doc/The_Growing_Backlash_Against_LPTA_Source_Selection.pdf

Data Consolidation of Disparate Procurement Data Sources for Correlated Performance-Based Acquisition Decision Support

Samantha Nangia—is a PMP and DAWIA-certified Project Manager and Business Technology Subject Matter Expert (SME). In her current role, she helps manage and oversee the suite of Acquisition Technology Solutions for the Department of the Navy (DoN) on the staff of the ASN(RD&A) DASN (AP) eBusiness Policy and Oversight (eBPO) Division. She has been dubbed the Navy's Procurement Data SME and is working to transform the current set of legacy Navy contract reporting and BI capabilities into a single data-rich enterprise solution. Nangia received her bachelor's degree in Computer Science from Rollins College in Winter Park, FL, in 2002 and her master's degree in Business Administration from Georgetown University in 2009.

Ryan Dickover—PhD, currently serves on the ASN(RD&A) staff as Director of DASN (AP)'s eBPO division. This division is responsible for the integration and governing of the suite of information systems used to acquire over $150 billion in DoN assets annually. Dr. Dickover is a member of the Defense Acquisition Corps and is DAWIA Level III certified in both Contracting and Information Technology. He earned his bachelor's degree with honors and a master's degree in Business Administration with an Information Systems Management concentration from Old Dominion University and a PhD from Capella University in 2009. His study put forth one of the first quantitatively derived mathematical models for forecasting Enterprise Resource Planning System ROI in the public sector, and was based on data collected from Naval and DLA field activities.

Thomas Wardwell—is a 1986 graduate of the United States Naval Academy and spent 20 years as a Supply Corps Officer honing his skills in logistics, contracting, and information systems. He volunteered as the Head of Contracts supporting contingency and counterterrorism operations in the Horn of Africa, where he envisioned and implemented an electronic contract writing solution to create a data-centric procurement environment with improved data visibility and a central reporting capability to facilitate enhanced decision making. He currently serves as a Navy civil servant improving electronic contract writing capabilities and acquisition business systems and defining the capabilities for a future department-wide procurement system while overseeing the DoN's portfolio of 120 systems and applications supporting the Acquisition Business Mission Area.

Randall Mora—is Founder, President, and CEO of Avum, Inc. In Avum, he has built, and continues to build, a professional services and computer sciences research and development firm that specializes in large-scale systems and software development. With more than 35 years of experience in systems architecture, development, and deployment, Mora is currently focusing on cognitive computing approaches and enterprise architectures to solve complex mission critical solutions for achieving interoperability between cross-domain disparate systems. Mora is currently applying this technology across multiple contracts, including major government and commercial business applications.

Introduction

Frank Kendall, then Under Secretary of Defense for Acquisition, Technology and Logistics, released the first defense acquisition system performance report in June 2013. This report focused primarily on performance related to the collective outcomes of Major Defense Acquisition Programs (MDAPs), but additionally explored various descriptive dimensions and acquisition approaches of the same (Kendall, 2013). Each annual report builds on the work previously conducted, and focuses on data-driven analysis relying on statistical techniques to identify trends that improve the defense acquisition community's insights into how contract incentives are motivating better contractor/vendor performance (Kendal, 2016).

Nevertheless, large amounts of data (in modern jargon, “Big Data”) are now available for research in the area of defense acquisition. Over the past several years, changes in electronic commerce have increased the amounts of both structured and unstructured data available—both in runtime and archived environments. This electronic data, from a variety of different acquisition agencies, can be obtained by a variety of means and used for a multitude of purposes (Snider et al., 2014).

Traditional statistical and trend analysis methods thus far have been primarily relied upon to explore trends and test metrics in the sets of acquisition data at hand. Sometimes, spreadsheets of linear regression correlation are employed, or, in some more modern applications, multivariate structural equation models via scientific applications such as SPSS and AMOS are leveraged for their ability to evaluate complex variable relationships, such as nested or recursive if-then patterns (Byrne, 2016).

However, not only are today’s modern datasets large in magnitude, they are also large in variety and complexity (Gartner, 2013). Furthermore, to address this state of data, new statistical modeling techniques, more powerful than before, have had to be created. This is due to the older methods finding difficulty with some of the size problems Big Data represents, such as privacy and security concerns (Parms, 2017). Thankfully, computer power necessary to employ the modern techniques is less expensive today, the software near free, and the storage capacities available now yield bewildering capacities at a fingertip, and with amazingly fast access speed. In fact, these performance parameters appear to continue along a Moore’s trend line against critical opposition (Magee, Basnet, Funk, & Benson, 2015). Presently, one of the more interesting of the new statistical modeling techniques is *neural network algorithm machine learning*.

Neural network modeling involves utilizing a “powerful computational data model that is able to capture and represent input/output relationships.” This model was developed out of the desire to create artificial intelligence systems capable of completing functions that were previously executed solely by the human brain. One benefit of using neural network modeling lies with its capacity to display and comprehend both linear and non-linear relationships from the data to which it is supplied (NeuroSolutions, 2015).

Research Question

Because “Big Data” is present in the Defense Acquisition Business space, and, because the demand to critically understand real cause-and-effect relationships between variables within that data is persistent from the Acquisition community, this paper’s research question is, *Can a neural network modeling technique be confidently relied upon to meaningfully explore variable relationships within acquisition business datasets?* Because, if it is, then any question may be reasonably asked by anyone of such a dataset; and, via the neural network-enabled tool, the answers they receive will come with scientific statistical confidence as to whether they can be trusted as interesting or useful answers.[1] In order to explore this research question, the study opted to use business data on contractor performance and attempted to isolate predictive variables from past performance information predictive of good performance.

[1] The role of human judgement of course, notwithstanding.

Methodology

This research uses the *Simple Action Research Mode*l (MacIsaac, 1995). Direct participation by the researchers went into answering the research question. In accordance with the steps of direct action: (1) the problem was defined (i.e., Can neural network modeling be applied to Big Data sets in acquisition?); (2) an Action Plan was developed (described in detail below, but generally it was to obtain a subset of Big Data, cleanse it for use, program a neural network tool, write hypotheses postulating expected correlative relationships between variables or variable sets, and execute testing of the hypotheses via the neural network for validation); (3) Execution of the Plan (which was a success: the data was obtained and cleaned, the hypotheses generated, and the neural network tool coded, tested and exercised over several cycles); and (4) Learning and Evaluation, which was completed via the documentation of results in this paper.

It is important to note that the Simple Action methodology employed here is evaluating the paper's research question regarding the applicability of the neural networking modeling technique to big data in the acquisition environment; as such, the actual statistical correlative output of the hypothesis are of a secondary value only (i.e., they are for the purpose of experimenting with the neural network environment itself, as opposed to for discovery in their own right).

Creation of the Cognitive Learning Environment

To build the data environment for evaluation, multiple sources of acquisition data were imported and fused together with iterative slices of multiple groups taken out for analysis. Further breakdown of the environment is described in the Study Plan section. Multiple open-source data analysis and machine learning tools were used to iteratively create models and generate graphs of the data slices. Human evaluation was involved in looking for patterns in the data which may have explained best performance across programs and portfolio groups in the past in order to produce a testable hypothesis. The underlying goal was to find patterns with the best chance of explaining *contractor/vendor performance improvement.*

The cognitive environment is generated in two phases: a simulation phase and a predictive phase. Simulation phase models generated from simple datasets perform predictive analytics. The predictive phase models were generated via Predictive Model Markup Language (PMML)[2] and integrated into a simple prototype for the proof of concept. PMML was used as a standard to integrate defined and tested models into the decision support toolsets. Once those models were iteratively perfected (i.e., acceptable levels of false positives were observed based on training and testing the datasets), we exported the PMML from the models and integrated the new capability into our decision support components.

[2] The Predictive Model Markup Language (PMML) is an XML-based predictive model interchange format conceived by Dr. Robert Lee Grossman, then the director of the National Center for Data Mining at the University of Illinois at Chicago. PMML provides a way for analytic applications to describe and exchange predictive models produced by data mining and machine learning algorithms. It supports common models such as logistic regression and feedforward neural networks (Wikipedia, n.d.).

Building the Cognitive Learning Application Framework depended on a predictive model being *able to learn from past experiences and make significantly intelligent decisions.* Thus, bringing together the archived acquisition data and building a model exportable to PMML was the main concern of this research. Figure 1 outlines the researchers' process of creating the Cognitive Learning Environment: mining, fusing, and modeling the datasets. It is important to note the methodology is iterative in nature, requiring the team to return to previous steps during model development. For example, during the Patterns/Analytics step, the need to slice the data differently was identified, which necessitated new data, cleaning, transformation, and so forth. Also, through the development of reusable components, time required for iterations was significantly reduced. Normally, Data Selection and Preprocessing are the most time-consuming steps; usually taking around 80% of the total effort required to build an analytical model (Baesens, 2014).

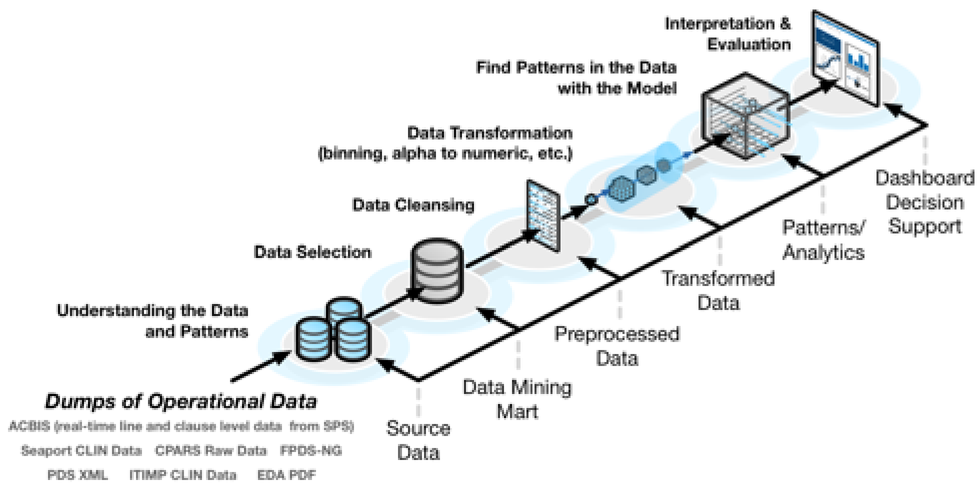


Figure 1. Project Methodology

Hypotheses

This research observes it is commonly claimed: contracts structured with incentivized performance line items (i.e., Cost Plus Incentive Fee, Cost Plus Award Fee, Fixed Price Incentive Fee, Fixed Price Award Fee, Cost Plus Fixed Fee, etc.) are associated (or expected) to enjoy good or better performance than otherwise structured; shorter duration contracts perform better than longer duration ones; competed contracts perform better than sole-source awards; negotiated clauses have an impact on performance, either for good or ill.

Recently, a vendor measure of performance has become available for calculation: the Superior Supplier Incentive Program (SSIP) composite score.

Therefore, this paper takes as its set of testing hypotheses, for the purpose of direct action experimentation of a neural network environment toolset, the following:

- H_1—Contract structures incentivizing performance (i.e., CPIF, CPFF, CPAF, FPIF, FPAF) result in higher Superior Supplier Incentive Program (SSIP) composite scores.
- H_2—Shorter contract duration results in better performance outcomes.
- H_3—Contracts that are competed result in better performance outcomes than sole-source.

- H_4—The mixture of negotiated clause inclusions impacts vendor performance outcomes.

Data Collection

This research began with utilizing an SSIP sample set of contracts from the FY16 review (i.e., FY13, FY14, and FY15 contracts). This data represents the study source boundary. The SSIP master data was derived from the Tri-Service SSIP Selection Methodology inclusive of subsets of CPARS[3] master data by suppliers[4], thus, initially limiting the dataset for analysis (Wardwell et al., 2016).

Analysis proceeded grouping data by similar Product Service Code (PSC) portfolios and similar dollar ranges. Performance was indicated by composite SSIP score (ranging from 0 to 4). Varying contract structures within the SSIP sample set were investigated for potential correlations between contract-type, CLIN mix, contract length, extent-competed, and clause inclusions. Subsequently, these relationships were analyzed within vendors or specific programs and contrasted against available program metrics aside from CPARS metrics.

Core data inputs were derived or pulled from acquisition data sources to which DASN (AP) had access. The following is the comprehensive list of data sources used throughout this research:

- SSIP sample set of contracts from the FY16 Review
- Army Contracting Business Intelligence System (ACBIS)
- Standard Procurement System (SPS)
- FPDS-NG[5]
- PDS XML[6]
- ITIMP CLIN Data[7]
- EDA PDF[8]

Initially, the SSIP sample set of contracts from the FY16 review matched corresponding ACBIS contracts. This allowed CLIN-level information from SPS and DFARS clause inclusions to be included for analysis. Table 1 outlines each attribute and the data source from which they were sourced. Some or all of the attributes were used during different phases of the research. As an example, during the beginning of the project, datasets were limited to the SSIP FY16 Review boundary. This limited research to 1,762

---

[3] Contractor Performance Assessment Reports System
[4] To determine which companies and business segments will be rated in a given FY, the Services use USASpending.gov to aggregate Systems contracts' obligations for the last three FYs by supplier and business segment. For each agency, the funding agency can be found by using the funding codes DoN (1700), AF (5700), and ARMY (2100). The obligations are maintained for all companies in the Air Force Industrial Liaison Office data warehouse. However, only the Top 100 or so (by obligation amount) are pulled for SSIP consideration.
[5] Federal Procurement Data System-Next Generation
[6] DoD Procurement Data Standard, Extensible Markup Language
[7] Navy's Integrated Technical Item Management and Procurement system, Contract Line Item
[8] DoD Electronic Document Access system, Portable Document Format file

records. After cleansing and transformation (i.e., matching/fusion) 972 contracts remained. Analysis of this dataset led to a determination that using source attributes directly from CPARS and including the full set of CPARS data (e.g., 174,138 records), which included contracts from the Navy, the Army, and the Air Force, would be a preferable set for testing the models.

Later iterations of the dataset focused on contract-level details and CPARS ratings. SPS line item details from ACBIS were suspended.

Table 1. Data Attribute Sources

Attribute	Sourced From
Contract-Type (CLIN mix)	ACBIS (SPS line item details)
Contract-Type (contract-level)	FPDS-NG
Extent-Competed	FPDS-NG
Contract Length	FPDS-NG/CPARS
DFARS Clause Inclusions	ACBIS
CLIN Count	ACBIS (SPS line item details)
CPARS Award Value	CPARS
Quality, Schedule, Cost Control, Management, Small Business	CPARS
SSIP	SSIP from the FY16 review
PSC & Portfolio Group	Defense Procurement and Acquisition Policy (DPAP) office

Data Selection, Cleansing, and Transformation

Data selection, cleansing, and transformation of this research was continuous throughout, and encompassed a significant amount of the effort in bringing operational data together correctly. Initial assumptions changed based on accuracy of data, and cleaning what became required. For instance, CPARS data is notoriously inconsistent when it comes to referencing IDIQ contracts. Sometimes, the CPARS references the base IDIQ, which is meaningless for analysis purposes. In other instances, the task order number is referenced.

The following outlines tasks performed to conduct preliminary data selection, grouping, and fusion between sets with the intent of matching with fusion while mitigating issues preventing a dataset:

1. Load of individual data into a data repository for selection and cleansing.
2. Determination of which award-value (CPARS) or obligation-amount (FPDS-NG) would better serve the research objective.
3. Analysis of the FPDS-NG contract numbers in preparation for matching to CPARS data. Contract number formatting varies across different source data repositories, thus, care must be taken to ensure data across those sources match properly. For instance, non-alphanumeric characters (dashes, spaces, etc.) must be removed, and the first 13 characters must be selected, unless the contract number starts with “GS,” “HHS,” “LC,” “NN,” or “V.” In those cases, the last 13 characters are selected. This method yields a 94% match

with FPDS-NG PIID[9] or Referenced IDV[10] PIID; however, matching algorithms are modified and tuned until the highest percentage match across the data emerges.

4. Contract-type, extent-competed, contract-length and PSC were selected from the base contract or mod with the highest obligated amount change.
5. For indefinite delivery/indefinite quantity (IDIQ) contracts with "D" in the 9th position of the contract number, the algorithm took into account both the "contract-num" and "order-num" values for successful match. Order number was parsed from the CPARS contract number and used both contract number and order number to match with FPDS-NG data using the highest obligated amount change of the mods (28% of the CPARS contracts). IDIQ contracts without order numbers were removed from the CPARS data (27%), as these would not produce meaningful results.
6. Portfolio and Portfolio Groups for each contract were identified by matching the FPDS-NG PSC code with data from the DPAP Product and Service Code (PSC) Selection Tool.

Analytics & Modeling

To perform the analytics and modeling of the data itself, different representations were built of results (e.g., Excel pivots, scatter plots, graphs, charts, etc.). This first step was conducted to steer further neural modeling efforts. Working with defense acquisition subject matter experts, iterative evolutions of the dataset were visualized to attempt discovery of patterns representing algorithms executable at high levels of accuracy. The iterations generated numerous views into the dataset, and participants were successful in coming to an understanding of the best way to begin modeling. Visual data exploration proved important in supplying initial insights into the data, which researchers then adopted throughout the modeling (Baesens, 2014).

To be useful to the community at large, end users should not have to acquire and learn complex analytic software to obtain predictions from these models. To facilitate this goal, once models are trained and tested, they're anticipated to be exported as PMML. The PMML can then be fed into a wide variety of systems and programming languages, which can be used to run the model(s) against incoming data. A Gartner 2017 Magic Quadrant for Data Science Platforms (called in 2016 "Advanced Analytics Platforms") evaluated a new set of 16 analytic and data science firms over 15 criteria and placed them in four quadrants, based on completeness of vision and ability to execute (Piatetsky, 2017). KNIME[11] was in the top quadrant with SAS, IBM, and RapidMiner. This study found KNIME to exhibit a flexible and extensible design, display an ease of use and verified capability to export PMML (essential to support of developing this research's data modeling and Cognitive Learning Application Framework).

The KNIME platform was selected for this study's use as the neural network tool. The KNIME advanced analytics platform is an open-source analytics platform, and was used

[9] Procurement Instrument Identifier
[10] Indefinite Delivery Vehicle
[11] KNIME Analytics platform: https://www.knime.org/products

to run computational analysis against the historical dataset to generate various sets of predictive models. Numerous models, such as polynomial and logistical regressions, were used to predict numerical values for variables, whereas *neural networks* and *decision trees* were used to predict categorical values for variables. Operational data was loaded into local databases (i.e., PostgreSQL) and quickly iterated regardless of it being functional in Excel for preliminary analysis or in KNIME for modeling.

Figure 2 illustrates the modeling process for the paper's problem set. Significant time was spent segmenting the dataset and running the data through models until patterns emerged that are relative to the expected results sought (i.e., contract structures performance).

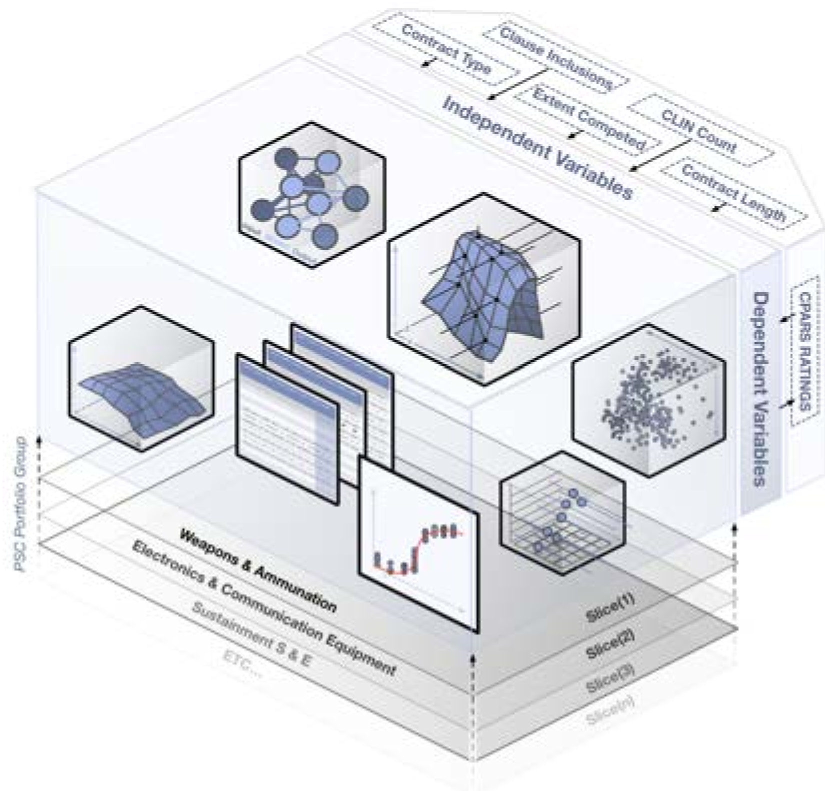


Figure 2. Analytics and Modeling Paradigm

Figure 2's illustration reveals how data in the study was segmented by PSC Portfolio Groups and used CPARS ratings (both by composite SSIP scores and individual ratings values) as dependent variables. In contrast, numerous independent variables were inputs to the modeling process, only some of which are depicted.

Data Analysis and Results

Modeling used contract-type and extent-competed as independent variables with composite CPARS ratings as dependent variables. The research goal at this stage was to find patterns in the data revealing correlations between contract-type and extent-competed to be neural network modeled, and, thus, exercise the paper's hypotheses via statistical correlation testing and address the paper's research question. Contract-length became another independent variable added to the set for modeling and, from that point forward, any attribute available and deemed pertinent to the research hypothesis objective was incorporated into operational data or transformed and sanitized. As a result, multiple back and forth passages through previous steps of data selection, cleansing, and transformation were avoided. Excel pivots were created against the full CPARS dataset, which supplied 174,138 records for final analysis in the study. No obvious significant pattern emerged from observing the pivot tables. Table 2 summarizes initial descriptive data for contract-type in the sample.

Table 2. Pivot of Contract-Type and SSIP[12]

Contract Type Description	Average of ssip	Count of ssip
COMBO	3.17	3,617
COST NO FEE	3.08	9,047
COST PLUS AWARD FEE	3.22	4,801
COST PLUS FIXED FEE	3.05	31,491
COST PLUS INCENTIVE FEE	2.76	3,380
COST SHARING	2.42	17
FIRM FIXED PRICE	2.82	116,660
FIXED PRICE AWARD FEE	3.12	366
FIXED PRICE INCENTIVE FEE	2.76	1,902
FIXED PRICE LEVEL OF EFFORT	2.86	522
FIXED PRICE REDETERMINATION	2.87	175
FIXED PRICE WITH ECONOMIC PRICE ADJUSTMENT	2.84	714
ORDER DEPENDENT	2.43	209
Grand Total	**2.89**	**172,901**

Table 3 summarizes initial descriptive data for extent competed in the sample.

Table 3. Pivot of Extent-Competed and SSIP[13]

Extent_Competed	Average of ssip	Count of ssip
COMPETED UNDER SAP	2.71	4,979
FOLLOW ON TO COMPETED ACTION	3.03	260
FULL AND OPEN COMPETITION	2.93	78,917
FULL AND OPEN COMPETITION AFTER EXCLUSION OF SOURCES	2.88	35,975
NOT AVAILABLE FOR COMPETITION	2.98	20,036
NOT COMPETED	2.80	31,516
NOT COMPETED UNDER SAP	2.87	2,150
Grand Total	**2.90**	**173,833**

Working With the Predictive Analytic Environment

The pivot table input yielded prima facie evidence against H_1 (that incentivized contract-types yield greater contractor performance) and H_3 (that competed contracts yield greater contract performance). With this starting point in hand, the study moved into the regression analysis phase and began running polynomial and linear regression models

[12] 1,237 unmatched records from Table 2 were removed
[13] 305 unmatched records from Table 3 were removed.

against the data with KNIME software to evaluate how close target measures of interest were with hypothesized outcomes. We ran both linear and polynomial regression algorithms against the data to attempt to find a mathematical formula that would take the extent-competed and contract-type as inputs and give a predicted value of the SSIP score.

Notes on Inputs: The regressions require numeric values for the dependent and independent variables. The SSIP score (the dependent variable) is a numeric value; however, the independent variables (contract-type and extent-competed) are strings. The type was converted to numbers by assigning a unique integer value to each distinct value (i.e., firmed fixed price was set to 10).

Figure 3 illustrates the KNIME's workspace used for these regressions. Node 1 utilizes an Excel file as input, and the data is passed into a partitioning node (Node 6) to give us a set of data on which to run the regressions, and a small set of data used to test the regressions. The data is passed to three different nodes. Node 2 runs the polynomial regression (with a max polynomial degree of 2), Node 10 runs a linear regression and Node 11 generates a 3D scatter plot to assist in visualizing the data.

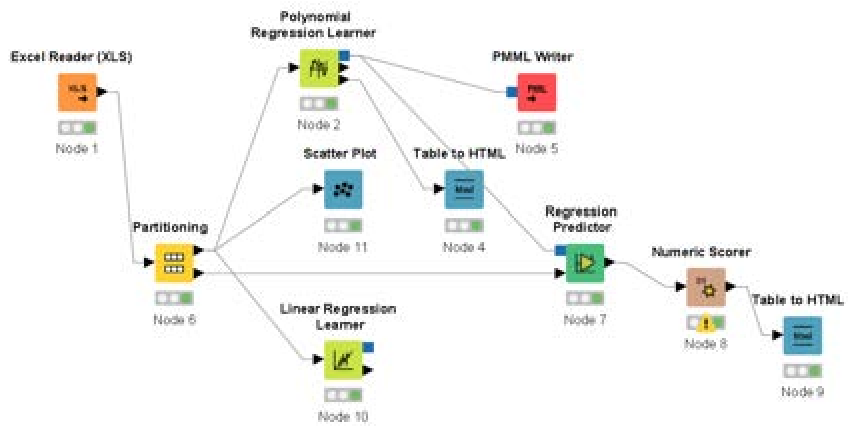


Figure 3. Initial Polynomial and Linear Regression

The results of the linear regression analysis showed no linear correlation existing in the dataset between independent variable types of "extent-competed" or "contract-type" and dependent variable types of contractor SSIP score. The results of the Polynomial Regression were pushed into a Regression Predictor and Numeric Scorer to output the model's statistics. Figure 4 shows all learned coefficients for the contract-type and extent-competed.

Regression involves numerous variables. The R^2 statistic, listed at the bottom of the figure, measures the proportion of variability in the SSIP score and CPARS rating (dependent variables) that can be explained by the contract-type and extent-competed (independent variables). This value is between 0 and 1, and can be negative, with a value close to 1 indicating that a large proportion of the variability in the response can be explained by the regression. The standard error is the average amount that the response will deviate from the true regression line (James et al., 2013).

Statistics on Polynomial Regression

Variable	Coeff.	Std. Err.	t-value	P>|t|
Contract_type_numeric	0.0528	0.0501	1.0537	0.2922
Extent_competed_numeric	0.0682	0.0806	0.8459	0.3977
Contract_type_numeric^2	-0.0114	0.0074	-1.5413	0.1235
Extent_competed_numeric^2	-0.0216	0.0177	-1.2241	0.2211
Intercept	2.4667	0.1129	21.8573	0.0

Multiple R-Squared: 0.4437
Adjusted R-Squared: 0.0047

Figure 4. Statistics on Polynomial Regression

The Numeric Score (Node 8) in the flow computes certain statistics between the numeric column's values (r_i) and predicted (p_i) values. It computes $R^2=1-SS_{res}/SS_{tot}=1-\Sigma(p_i-r_i)^2/\Sigma(r_i-1/n*\Sigma r_i)^2$ (can be negative!) ("Coefficient of Determination," n.d.); mean absolute error $(1/n*\Sigma|p_i-r_i|)$ ("Mean Absolute Error," n.d.); mean squared error $(1/n*\Sigma(p_i-r_i)^2)$ ("Residual Sum of Squares," n.d.); root mean squared error $(sqrt(1/n*\Sigma(p_i-r_i)^2))$ ("Root-Mean-Square Deviation," n.d.); and mean signed difference $(1/n*\Sigma(p_i-r_i))$ ("Mean Signed Deviation," n.d.). The computed values can be inspected in the node's view and/or further processed using the output table. Table 4 contracts the results from the Numeric Scorer.

Table 4. Numeric Score Results

Prediction (ssip_raw)	
R²	0.019
Mean absolute error	0.55
Mean squared error	0.458
Root mean squared error	0.676
Mean signed difference	-0.018

Next, we built more complicated predictive analysis algorithms in an attempt to find the correlations/predictive capabilities we were expecting (i.e., model contract structures that incentivize performance correlated to better performance outcomes). Figure 5 shows the training and testing of two different types of neural networks. It initially splits the data into a training set, and a validation set: we train the neural network on the training set and measure the performance on the validation set. Node 5 partitions the data (i.e., 80% Training and 20% Validation) and pushes the sets into both an RProp MLP Learner and a PNN Learner (DDA).

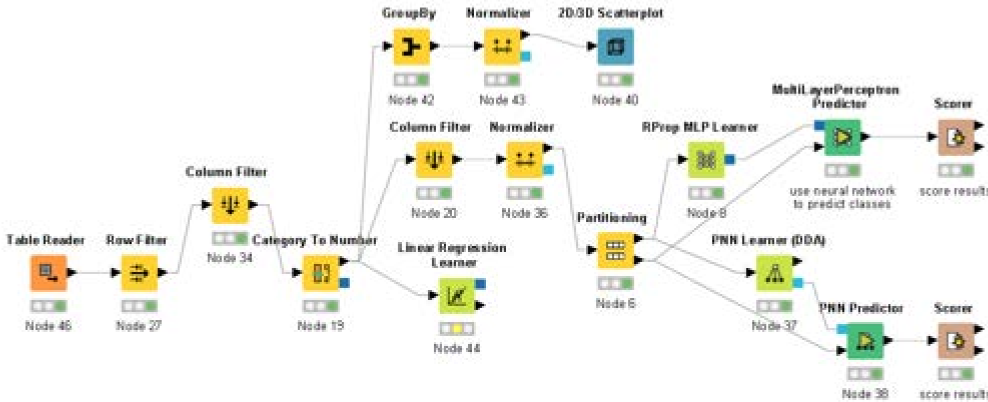


Figure 5. Neural Network Model

The RProp algorithm is used for multilayer feedforward networks. RProp performs a local adaptation of the weight-updates according to the behavior of the error function (Riedmiller & Braun, 1993).

The PNN Learner (DDA) trains a Probabilistic Neural Network (PNN) based on the DDA (Dynamic Decay Adjustment) method on labeled data using Constructive Training of Probabilistic Neural Networks (Berthold & Diamond, 1998) as the underlying algorithm. This algorithm generates rules based on numeric data. Each rule is defined as a high-dimensional Gaussian function that is adjusted by two thresholds, theta minus and theta plus, to avoid conflicts with rules of different classes. Each Gaussian function is defined by a center vector (from the first covered instance) and a standard deviation, which is adjusted during training to cover only non-conflicting instances. The selected numeric columns of the input data are used as input data for training, and additional columns are used as classification targets. Either one column holding the class information, or a number of numeric columns with class degrees between 0 and 1, can be selected. The data output contains the rules after execution, along with a number of rule measurements. The model output port contains the PNN model, which can be used for prediction in the PNN Predictor node.

The KNIME tool was able to create an abstraction layer around the algorithms that were being run in the neural network nodes. A basic understanding of what the nodes are doing and how to configure them for optimum performance facilitates quicker iterations and an overall feasibility study that supports rapid prototyping.

The MultiLayerPerceptron and the PNN Predictor nodes are used to validate the resulting trained model against the test dataset. For the MultiLayerPerceptron Predictor, if the output variable is nominal, the output of each neuron and the class of the winner neuron are produced. The PNN Predictor is doing a similar test (i.e., using the trained model to validate with the test data). In this case, it also outputs predicted data with an additional classification column (e.g., CPARS quality attribute).

Figure 6 and Figure 7 outline the confusion matrix and accuracy statistics for the model. An example of what it's showing is as follows: For the MultiLayerPerceptron model/predictor, using 20% of the data to score/validate based on contract-type and extent-competed, we successfully predicted 271 times Technical/Quality of Product or Service would be rated as Satisfactory. As well, 166 times Technical/Quality of Product or Service was rated as Satisfactory when it should have been Very Good. We only trained the model to predict Quality at this point. The accuracy for the MultiLayerPerceptron was 45.427% and

48.018% for the PNN Predictor. Evaluation ratings are defined as: E: Exceptional, V: Very Good, S: Satisfactory, M: Marginal, U: Unsatisfactory, and N: N/A.

Confusion Matrix - 0:16 - Scorer (score results)

File Hilite

quality \ Pr...	V	S	E	U	N	M
V	25	166	0	0	0	0
S	10	271	0	0	0	0
E	22	106	2	0	0	0
U	0	1	0	0	0	0
N	2	27	0	0	0	0
M	1	23	0	0	0	0

Correct classified: 298 Wrong classified: 358

Accuracy: 45.427 % Error: 54.573 %

Cohen's kappa (κ) 0.067

Figure 6. MultiLayerPerceptron Predictor

Confusion Matrix - 0:39 - Scorer (score results)

File Hilite

quality \ Pr...	U	V	S	E	N	M
U	0	0	1	0	0	0
V	0	14	164	13	0	0
S	0	8	267	6	0	0
E	0	9	87	34	0	0
N	0	2	27	0	0	0
M	0	1	22	1	0	0

Correct classified: 315 Wrong classified: 341

Accuracy: 48.018 % Error: 51.982 %

Cohen's kappa (κ) 0.13

Figure 7. PNN Predictor

Work on finding correlations within the data continued throughout the research. Through additional tuning and new independent variables, the study found the model accuracy statistics showed positive results. New problem sets with higher accuracy statistics should find themselves into a defense procurement toolset soon and/or a sequel to this research.

PMML Integrated Decision Support Tool Kit

The Java PMML API (GitHub, 2017) is an open-source project that provides a PMML producer and consumer libraries for the Java/JVM platform. Using these libraries, a Java wrapper is created around each model. These wrappers each take a pre-defined set of inputs and provide a predicted output value. These wrappers then become pluggable components that can be added to any Java-based tool, such as websites, web services and stand-alone applications. For example, a tool that runs Pre-Validations on PDS XML[14] (Defense Procurement and Acquisition Policy, 2008) files can quickly be updated to take the information it parses from the PDS file, and feed it to these various models to augment the validation results to include predictive analytics.

Early in the research project, a simple model prototype was built to export the PMML from this model. This PMML was used in a Java application as a Proof of Concept (POC) for the ability to use a PMML-defined model and integrated its predictive capabilities into the Cognitive Learning Application Framework, thus, having an end-to-end solution for decision support capabilities in the application toolsets.

Future work in this space involves making the models used by these wrappers dynamic. The current iteration utilizes static PMML that was generated using a historical dataset. It would be desirable to have the ability to update the models in real time as new data comes in.

Conclusion

This study's research question was, *Can a neural network modeling technique be confidently relied upon to meaningfully explore variable relationships within acquisition business datasets?* This paper's result was positive to the research question.

The study's open architecture framework (i.e., the Cognitive Learning Application Framework [CLAF]) for Acquisition Decision Support and Business Intelligence successfully integrated and prototyped a neural network model using a PMML standard and explored variable relationships using four test hypotheses addressing contract performance data. Regarding the study's test hypotheses, results were inconclusive. Only H_1 (incentivized contract types correlate with higher vendor performance scores) and H_3 (competed contracts correlate with higher vendor performance scores) were thoroughly evaluated, and proved to be inconclusive via initial standard regression technique. Due to datasets being too small for substantive use in big data network evaluation, or, because of time limitations preventing necessary dataset concatenation, H_2 (shorter duration contracts correlate with higher vendor performance scores) and H_4 (contract clauses have impact on vendor performance score) could not be evaluated.

The study's main function went on to explore H_1 and H_3 as a means of addressing the research question via a direct action methodology of research. Experimenting with the neural network mode of analysis, the study attempted accurate prediction of vendor performance scores given an input of one of the hypothesized independent variables. The study's neural network obtained a maximum accuracy score of 49%. Obtaining this level of accuracy required careful, and sometimes tedious, assembly of statistical and logical

[14] The Procurement Data Standard (PDS) is a system-agnostic data standard that is adopted and implemented DoD-wide for creation, translation, processing, and sharing of procurement actions. It defines the minimum requirements for contract writing system output to improve visibility and accuracy of contract-related data, to support interoperability of DoD acquisition systems and to standardize and streamline the procure-to-pay business process. Further, the PDS will improve visibility of contract-related data, enabling senior DoD leadership to make better informed business decisions. And finally, this data standard will support future migration to enterprise and federal systems and processes where appropriate.

components. Although work has stopped, it is anticipated to resume post-publication in order for more potential latent variable relationships to be discovered, or, presumed relationships tested for the potential to be dispelled.

The value of discovering through this study's experience that there's evidence that the neural network modeling technique is applicable to big data sets in acquisition is that, now, any question for which there is a discrete data element available, or derivable, within those sets, can expect a trusted answer (if interesting and useful) with scientific statistical confidence.

Findings

The KNIME advanced analytics platform can write to PMML simply by dragging the PMML writer to the model. Figure 3 illustrates this capability in Node 5 (PMML Writer). This and the Application Frameworks created during this research project will enable future researchers to quickly bring new data into the modeling process, as well as integrate exported PMML models into Defense Acquisition Decision Support tools.

Cognitive computing (neural network modeling) solutions promise better-informed buying and increased compliance, and may make this faster and easier to accomplish once generated. Also, from this experience, the study found that loading more than just enough data helps significantly during the modeling phase.

Given the success of this research, it is recommended that government and industry oversight entities build a cognitive learning component for acquisition support that uses archived acquisition data from known repositories. The component can then be used as a stand-alone tool for the acquisition community and/or integrate into existing acquisition community toolsets and contract writing applications.

Afterwards, the models could be leveraged in decision support or Business Intelligence (BI) dashboards by the acquisition community.

In fact, a simulation tool could even be envisioned that would allow contracting officers (KOs) to perform scenario testing surrounding new agreements—one that would, given any variables, project performance indices based on purchase type or agreement structure, or other discovered latent relationships.

Areas for Further Research

The combination of the Cognitive Learning Acquisition Framework and a Big Data archive together form a methodology for an Application Framework, enabling a dynamic information analysis space to build intelligence into acquisition decision support tools. Future practical research into the feasibility and capability of applications leveraging such a framework could include the following: decision support for contracting officers and program managers surrounding contract structure tied to true historic performance and delivery outcomes; modeling of contract incentives, structures, and policies, and their impact on performance, delivery, costs, and schedule across major programs; volume of modifications and manual and/or late payments tied to contract, clause, and line item structures, as well as the overall quality of the contract data and compliance with contracting rules and regulations; EVM[15] outcomes correlated to initial negotiated contract terms; award and incentive fee payouts tied to EVM and CPARS metrics; vendor past performance within specific product service codes correlated to historic contract structure; and Q&A support for initial acquisition planning (i.e., "What vendors typically support this product or service?", "What type of contract is most widely used?", "Is this work typically competed?", and "What clauses above and beyond the typical prescriptions accompany this type of buy?"). Finally, the incorporation of additional public and Defense datasets within the financial, logistics, and commercial spaces into the Big Data archive is warranted to provide opportunities for further exploration of data relationships for use in acquisition decision making.

References

Baesens, B. (2014). *Analytics in a big data world*. New York, NY: Wiley.

Berthold, M., & Diamond, J. (1998). Constructive training of probabilistic neural networks. *Neurocomputing*. Retrieved from http://www.sciencedirect.com/science/article/pii/S0925231297000635.

Byrne, B. M. (2016). *Structural equation modeling with Amos: Basic concepts, applications, and programming*. New York, NY: Routledge, Taylor & Francis Group.

Coefficient of determination. (n.d.). In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Coefficient_of_determination

Defense Procurement and Acquisition Policy. (2008). Procurement Data Standard and other enterprise initiatives. Retrieved from http://www.acq.osd.mil/dpap/pdi/eb/procurement_data_standard.html

Gartner. (2013). Gartner identifies top technology trends impacting information infrastructure in 2013. Retrieved from http://www.gartner.com/newsroom/id/2359715

GitHub. (2017). Java PMML API. Retrieved from https://github.com/jpmml

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). Unsupervised learning. In G. James, D. Witten, T. Hastie, & R. Tibshirani, *An introduction to statistical learning with applications in R* (pp. 373–374). New York, NY: Springer Science+Business Media.

Kendall, F. (2013). *Performance of the defense acquisition system, 2013*. Washington, DC: Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics.

Kendall, F. (2016). *Performance of the Defense Acquisition System: 2016 annual report*. Washington, DC: DoD.

Magee, Basnet, Funk, & Benson. (2015). Quantitative empirical trends in technical performance. *Technological Forecasting and Social Change.*

Mean absolute error. (n.d.). In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Mean_absolute_error

Mean signed deviation. (n.d.). In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Mean_signed_difference

NeuroSolutions. (2015). What are neural networks & predictive data analytics? Retrieved from http://www.neurosolutions.com/products/ns/whatisNN.html

[15] Earned Value Management

Parms, J. (2017). Privacy and security issues in the age of big data. Retrieved from https://www.business.com/articles/privacy-and-security-issues-in-the-age-of-big-data/

Piatetsky, G. (2017). Gartner 2017 magic quadrant for data science platforms: Gainers and losers. Retrieved from http://www.kdnuggets.com/2017/02/gartner-2017-mq-data-science-platforms-gainers-losers.html

Residual sum of squares. (n.d.). In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Residual_sum_of_squares

Riedmiller, M., & Braun, H. (1993). A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In *Proceedings of IEEE International Conference on Neural Networks* (pp. 586–591).

Root-mean-square deviation. (n.d.). In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Root-mean-square_deviation

Snider, K. F., Rendon, R. G., Soeters, J., Shields, P. M., & Rietjens, S. J. H. (2014). Retrieving what's already there—Archival data for research in defense acquisition. In *Routledge handbook of research methods in military studies* (pp. 78–91). London, England: Routledge/Taylor & Francis Group.

Wardwell, T., Nangia, S., Simeoni, C. J., Brittenham, R., & Campbell, T. (2016). *Tri-service SSIP selection methodology.*

Acknowledgments

We gratefully acknowledge the contributions of numerous individuals who provided assistance and support in this research project. Larry Wright, David Harris, Kevin Chiu, and Anita Doyle each provided significant time in addition to performing their primary work in order to provide assistance in conducting the data loading, cleansing, transformation, and analysis for this project. David Harris's previous training and research with cognitive engineering proved invaluable in bringing everything together, from building the models to integrating the PMML into a Java application.

Special thanks go to David Michalczuk for the outstanding 3-D diagrams he produced for this project. We all love your diagrams.

Finally, the close partnership with DASN (AP) contributed in many diverse ways, including accessing cleansed data that normally is not available to researchers. This made the normally difficult and time-consuming task of acquiring the operational acquisition data analyzed in this project conclude significantly faster than otherwise possible. Their subject matter expertise (SMEs) in acquisition systems, policies, and procedures were immeasurable in our success.

The responsibility for the contents of this study rests solely with the authors. Any errors or omissions are their responsibility and should in no way reflect on any of the individuals named above.

Customizing the Use of TINA (Truth in Negotiations Act) in the DoD

Chong Wang—Associate Professor, Financial Management, Graduate School of Business and Public Policy, Naval Postgraduate School

Introduction

Wang, Rendon, Champion, Ellen, & Walk (hereafter, Wang et al., 2016) identify the incentive problem that is characterized as a "moral hazard" in the DoD's current use of the Truth in Negotiations Act (TINA). One of the examples they concentrated on was the ineffective use of TINA in the context of firm-fixed-price (FFP) contracts. Specifically, a contractor under an FFP contract that is subject to TINA has the following ill incentive: The fear of being held accountable for any significant unfavorable cost discrepancy (i.e., the actual incurred cost is significantly below the ex-ante cost estimate negotiated with the DoD as the basis for contract fixed-price) would strongly motivate the contractor to shirk (i.e., reduce cost-saving effort) or even engage in cost padding (e.g., by opportunistically incurring or allocating more costs to the government contracts). Such behavior leads to deadweight welfare loss that is ultimately borne by taxpayers.

This study extends Wang et al. (2016) to a broader scope and greater depth. In particular, we propose to customize the use (or disuse) of TINA in the DoD for various contracting scenarios involving specific acquisition category (ACAT I through III), stage of the cycle (Milestones A, B, and C), and contract type. The bottom line is: we don't believe the TINA policy should be prescribed via a "one-size-fits-all" approach; rather, the use or disuse of TINA should be customized to various situations.

We continue to employ an economics-based, incentive-centric approach that focuses on investigation of agents' (i.e., DoD contractors') incentives under various settings. Then we generate our policy recommendation through a "with" and "without" TINA comparison.

TINA is a *federal* acquisition regulation, which goes beyond the DoD and DoN. We expect that significant cost savings can be generated for the DoD and DoN, *as well as other federal government agencies*, by providing such a framework described above.

The remainder of the report is organized as follows. The next section describes the DoD acquisition process. Following that is a section that describes how TINA is implemented in DoD acquisition via a "one-size-fits-all" approach. Building on those two sections, the following one (Customizing the Use [or Disuse] of TINA in the DoD Acquisition Process) tailors the use or disuse of TINA (i.e., TINA waiver) to various circumstances. We offer a conclusion in the final section.

DoD Acquisition Process: Category, Cycle, and Contract Type

The DoD procures goods and services through contracts. Schwartz (2014) interprets "acquisition" as "a broad term that applies to more than just the purchase of an item or service"; rather, "the acquisition encompasses the design, engineering, construction, testing, deployment, sustainment, and disposal of weapons or related items purchased from a contractor."

DoD acquisition is governed by the Federal Acquisition Regulation (FAR) along with the Defense Federal Acquisition Regulation Supplement (DFARS). Additional regulations such as TINA also provide rules.

Acquisition Category (ACAT)

Depending on program costs, DoD acquisition is divided into three categories. The biggest ticket purchase is Category I (ACAT I), also called Major Defense Acquisition Programs (MDAPs). The Under Secretary of Defense for Acquisition, Technology, and Logistics (USD[AT&L]) defines MDAPs as programs with more than $480 million (fiscal year 2014 dollars) in research, development, test, and evaluation (RDT&E) expenditures; or at least $2.79 billion (fiscal year 2014 dollars) in procurement funding; or as designated as a major defense acquisition program by the milestone decision authority (10 U.S.C., § 2430, Major Defense Acquisition Program Defined). A similar Category I definition, namely *ACAT IA*, also called Major Automated Information System (MAIS), with different dollar thresholds, exists for DoD acquisition of Automated Information Systems (AIS).

The next tier procurement is Category II, which is a major system defined as 10 U.S.C. 2302d (Reference (h)), yet which does not meet criteria for ACAT I or IA. Finally, Category III (ACAT III) includes any program that does not meet criteria for ACAT II or above, or any AIS program that is not a MAIS.

The following Table 1, reproduced from DoD Instruction (DoDI) 5000.02, *Operation of the Defense Acquisition System* (USD[AT&L], 2013), details the definitions of each acquisition category.

It is worth noting that as the acquisition category decreases from I to III, so does the level of oversight from the DoD and Congress. One should expect that the closest and most supervision being applied to MDAPs. Another difference between ACAT I (MDAPs) and non-MDAPs is the degree of information asymmetry between the DoD and the contractor. MDAPs are inherently more technologically complex than ACAT II and III programs and hence information asymmetry is more serious to start with for MDAPs.

A GAO (2015) report indicates that

> DOD requested $168 billion in fiscal year 2014 to develop, test, and acquire weapon systems and other products and equipment. About 40 percent of that total is for major defense acquisition programs or ACAT I programs. DOD also invests in other, non-major ACAT II and III programs that are generally less costly at the individual program level. These programs typically have fewer reporting requirements and are overseen at lower organizational levels than ACAT I programs, although they may have annual funding needs that are just as significant.

Table 1. Description and Decision Authority for ACAT I–III Programs

ACAT	Reason for ACAT Designation	Decision Authority
ACAT I	• MDAP (10 U.S.C. 2430 (Reference (n))) ○ Dollar value for all increments of the program: estimated by the DAE to require an eventual total expenditure for research, development, and test and evaluation (RDT&E) of more than $480 million in Fiscal Year (FY) 2014 constant dollars or, for procurement, of more than $2.79 billion in FY 2014 constant dollars ○ MDA designation • MDA designation as special interest[1]	ACAT ID: DAE or as delegated ACAT IC: Head of the DoD Component or, if delegated, the CAE (not further delegable)
ACAT IA[2, 3]	• MAIS (10 U.S.C. 2445a (Reference(n))): A DoD acquisition program for an Automated Information System[4] (AIS) (either as a product or a service[5]) that is either: ○ Designated by the MDA as a MAIS program; or ○ Estimated to exceed: ▪ $40 million in FY 2014 constant dollars for all expenditures, for all increments, regardless of the appropriation or fund source, directly related to the AIS definition, design, development, and deployment, and incurred in any single fiscal year; or ▪ $165 million in FY 2014 constant dollars for all expenditures, for all increments, regardless of the appropriation or fund source, directly related to the AIS definition, design, development, and deployment, and incurred from the beginning of the Materiel Solution Analysis Phase through deployment at all sites; or ▪ $520 million in FY 2014 constant dollars for all expenditures, for all increments, regardless of the appropriation or fund source, directly related to the AIS definition, design, development, deployment, operations and maintenance, and incurred from the beginning of the Materiel Solution Analysis Phase through sustainment for the estimated useful life of the system. • MDA designation as special interest[1]	ACAT IAM: DAE or as delegated ACAT IAC: Head of the DoD Component or, if delegated, the CAE (not further delegable)
ACAT II	• Does not meet criteria for ACAT I or IA • Major system (10 U.S.C. 2302d (Reference (n))) ○ Dollar value: estimated by the DoD Component Head to require an eventual total expenditure for RDT&E of more than $185 million in FY 2014 constant dollars, or for procurement of more than $835 million in FY 2014 constant dollars ○ MDA designation[5] (10 U.S.C. 2302 (Reference (n)))	CAE or the individual designated by the CAE[6]
ACAT III	• Does not meet criteria for ACAT II or above • An AIS program that is not a MAIS program	Designated by the CAE[6]

1. The Special Interest designation is typically based on one or more of the following factors: technological complexity; congressional interest; a large commitment of resources; or the program is critical to the achievement of a capability or set of capabilities, part of a system of systems, or a joint program. Programs that already meet the MDAP and MAIS thresholds cannot be designated as Special Interest.
2. When a MAIS program also meets the definition of an MDAP, the DAE will be the MDA unless delegated to a DoD Component or other official. The DAE will designate the program as either a MAIS or an MDAP, and the Program Manager will manage the program consistent with the designation.
3. The MDA (either the DAE or, if delegated, the DoD Chief Information Officer (CIO) or another designee) will designate MAIS programs as ACAT IAM or ACAT IAC. MAIS programs will not be designated as ACAT II.
4. AIS: A system of computer hardware, computer software, data or telecommunications that performs functions such as collecting, processing, storing, transmitting, and displaying information. Excluded are computer resources, both hardware and software, that are an integral part of a weapon or weapon system; used for highly sensitive classified programs (as determined by the Secretary of Defense); used for other highly sensitive information technology (IT) programs (as determined by the DoD CIO); or determined by the DAE or designee to be better overseen as a non-AIS program (e.g., a program with a low ratio of RDT&E funding to total program acquisition costs or that requires significant hardware development).
5. Acquisitions of services that satisfy or are expected to satisfy the definition of a MAIS in 10 U.S.C. 2445c, Reference (n), will comply with this instruction. All other acquisitions of services will comply with Enclosure 9 of DoD Instruction 5000.02 (Reference (b)).
6. As delegated by the Secretary of Defense or Secretary of the Military Department.

Acquisition Cycle

Schwartz (2014) identifies

> a three-step process of identifying the required weapon system, establishing a budget, and acquiring the system. These three steps are organized as follows:
>
> 1. The Joint Capabilities Integration and Development System—for identifying requirements.
> 2. The Planning, Programming, Budgeting, and Execution System—for allocating resources and budgeting.
> 3. The Defense Acquisition System—for developing and/or buying the item. (Schwartz, 2014, Summary)

These three steps (each of which is a system unto itself), taken together, are often referred to as “Big ‘A’” acquisition, in contrast to the Defense Acquisition System, which is referred to as “little ‘a’” acquisition.”

Figure 1, reproduced from Schwartz (2014), depicts the three-step process.

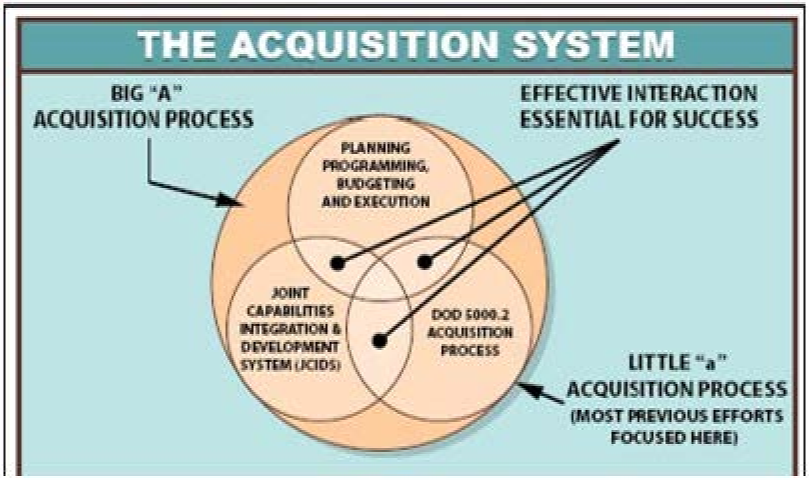


Figure 1. The DoD’s Defense Acquisition Structure

The cycle of the defense acquisition process contains three stages, namely, pre-acquisition, acquisition, and sustainment, with critical reviews identified by Milestones A, B, and C. The following Figure 2, reproduced from Schwartz (2014), describes the acquisition cycles.

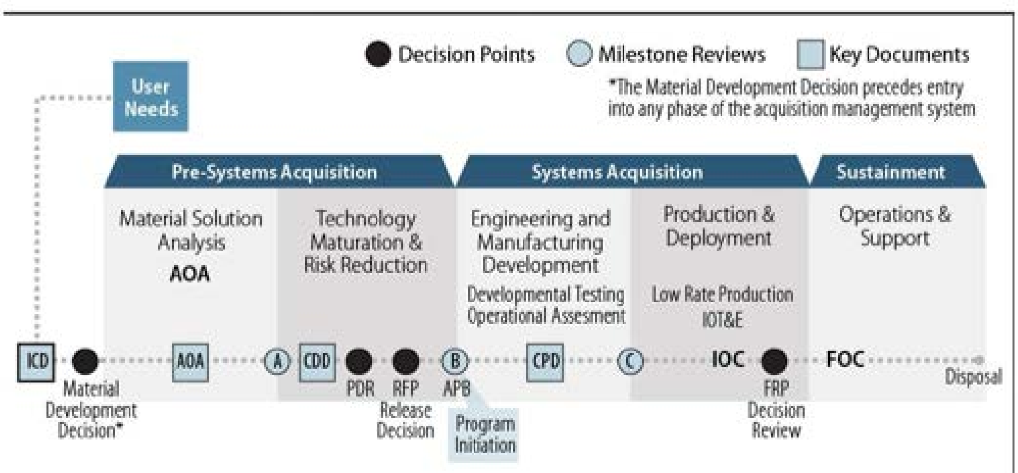


Figure 2. Defense Acquisition Milestones

As illustrated by Figure 2, each milestone needs to be passed in order to reach the next stage. In particular, three key phases—Technology Maturation & Risk Reduction, Engineering and Manufacturing Development, and Production and Deployment—immediately follow the passage of Milestones A, B, and C, respectively.

Schwartz (2014) points out,

> The official responsible for deciding whether a program meets the milestone criteria and proceeds to the next phase of the acquisition process is referred to as the Milestone Decision Authority (MDA). Depending on the program, the MDA can be the Under Secretary of Defense (Acquisition, Technology, & Logistics), the head of the relevant DOD component, or the component acquisition executive. (p. 7)

Contract Type

The contract types are broadly classified into two categories: fixed-price contracts and cost-plus contracts. One can imagine a spectrum with the firm-fixed-price (FFP) contract on one end, under which the contractor assumes all the risks and has the highest incentive to save costs. At the opposite end of the spectrum is the cost-plus-fixed-fee (CPFF) contract,[1] where the government pays the contractor its realized cost and sets a fixed fee (profit). The fixed fee is supposed to be independent of actual cost, although its level is implicitly related to the size of the project.[2] Under CPFF, the government bears all the cost risk and hence leaves the contractor little incentive to minimize cost. In between the two extremes, FFP and CPFF, are the various forms of incentive contracts including fixed-price-incentive-fee (FPIF) contracts, cost-plus-incentive-fee (CPIF) contracts, and cost-plus-award-fee (CPAF) contracts. The following descriptions of each contract type are based on the FAR, except for the "budget-based-cost-plus" scheme, which is not defined by the FAR and has no application thus far in the DoD.

Firm-Fixed-Price (FFP) Contracts

A firm-fixed-price contract provides for a price that is not subject to any adjustment on the basis of the contractor's cost experience in performing the contract. This contract type places maximum risk and full responsibility for all costs and resulting profit or loss on the contractor. It provides maximum incentive for the contractor to control costs and perform effectively and imposes a minimum administrative burden upon the contracting parties.

Fixed-Price-Incentive-Fee (FPIF) Contracts

A fixed-price incentive-fee contract is a fixed-price contract that provides for adjusting profit and establishing the final contract price by a formula based on the relationship of final negotiated total cost to total target cost. A fixed-price incentive contract specifies a target cost, a target profit, a price ceiling (but not a profit ceiling or floor), and a profit adjustment formula. These elements are all negotiated at the outset. The price ceiling is the maximum that may be paid to the contractor, except for any adjustment under other contract clauses. When the contractor completes performance, the parties negotiate the final cost, and the final price is established by applying the formula. When the final cost is less than the target cost, application of the formula results in a final profit greater than the target profit.

---

[1] The CPFF contract is the benchmark case for the cost-plus contract. Put another way, a "cost-plus" contract without mentioning whether it is "cost-plus-fixed-fee" or "cost-plus-incentive-fee" or "cost-plus-award-fee" would refer to a CPFF contract. However, throughout this paper, we reserve the use of "cost-plus" contract as a general category including all variations of cost-plus contracts.
[2] The cost-plus-a-percentage-of-cost contract type, which was used sometimes in U.S. DoD acquisition practice before the 1960s, is prohibited by FAR 16.102. This particular type of cost-plus contract rewards rather than penalizes a firm's cost inefficiency.

Conversely, when final cost is more than target cost, application of the formula results in a final profit that is less than the target profit, or possibly a net loss. If the final negotiated cost exceeds the price ceiling, the contractor absorbs the difference as a loss. Because the profit varies inversely with the cost, this contract type provides a positive, calculable profit incentive for the contractor to control costs.

Cost-Plus-Fixed-Fee (CPFF) Contracts

A cost-plus-fixed-fee contract is a cost-reimbursement contract that provides for payment to the contractor of a negotiated fee that is fixed at the inception of the contract. The fixed fee does not vary with actual cost, but may be adjusted as a result of changes in the work to be performed under the contract. This contract type permits contracting for efforts that might otherwise present too great a risk to contractors, but it provides the contractor only a minimum incentive to control costs.

Cost-Plus-Incentive-Fee (CPIF) Contracts

A cost-plus-incentive-fee contract is a cost-reimbursement contract that provides for an initially negotiated fee to be adjusted later by a formula based on the relationship of total allowable costs to total target costs. This contract type specifies a target cost, a target fee, minimum and maximum fees, and fee adjustment formula. After contract performance, the fee payable to the contractor is determined in accordance with the formula. The formula provides, within limits, for increases in the fee above the target fee when total allowable costs are less than target costs, and decreases in the fee below the target fee when total allowable costs exceed target costs. This increase or decrease is intended to provide an incentive for the contractor to manage the contract effectively. When total allowable cost is greater than (or less than) the range of costs within which the fee-adjustment formula operates, the contractor is paid total allowable costs, plus the minimum (or maximum) fee.

Cost-Plus-Award-Fee (CPAF) Contracts

A cost-plus-award-fee contract is a cost-reimbursement contract that provides for a fee consisting of (a) a base amount (which may be zero) fixed at inception of the contract and (b) an award amount, based upon a judgmental evaluation by the government, sufficient to provide motivation for excellence in contract performance. Since the award fee determination is made unilaterally by the government, this contract type is only appropriate when achievement is measurable by subjective evaluation rather than objective data, which is unlikely to be true under significant information asymmetry.

Budget-Based-Cost-Plus-Scheme (BBCPS) Contracts

A budget-based-cost-plus-scheme contract is a refinement of CPIF in the following sense: (a) Under BBCPS, the job of estimating target cost is shifted from the government to the contractor, and (b) moreover, both target fee and cost share coefficient vary with the estimated target cost rather than being constants under CPIF. A carefully designed BBCPS contract will desirably induce the contractor’s “truth-telling” behavior and hence effectively mitigates the agency problem and reduces information asymmetry.

BCPS belongs to the larger topic of “menu of contracts” discussed in the principal-agent literature. This body of literature has broad applications in executive compensation contracts, regulation, and government procurement contracts (Laffont & Tirole, 1986, 1993; McAfee & McMillan, 1987; Melumad & Reichelstein, 1989; Reichelstein, 1992).

Selecting contract type along with price requires sound judgment. The contracting officer also has to consider the implications of the contracting method. For example, FAR 16.102 (a) states that “contracts resulting from sealed bidding shall be firm-fixed-price contracts or fixed-price contracts with economic price adjustment.” Most often a decision on contract type and price is a negotiation process that hopefully will lead to a fair risk sharing and price that motivates the contractor to minimize cost and deliver a quality product.

The Use of TINA in DoD Acquisition

TINA Defined

TINA was first enacted in 1962 and has been amended many times since then. Wang et al. (2016) states,

> In a nutshell, TINA requires contractors (often sole-source) to submit "cost or pricing data" when they negotiate the price of a contract with the federal government. The contractors must certify that the information they provide is "current, complete, and accurate." Failing to disclose truthful information could lead to civil or criminal investigation. The intention of TINA is to protect the government and taxpayers from being ripped off by better informed contractors.

TINA Applicability

TINA applies to a wide range of procurements that include both fixed-price and cost-plus contracts. Any negotiated prime contracts or prime contract modifications that exceed $750,000 are subject to TINA. In a similar fashion, for any negotiated subcontracts or subcontract modifications greater than $750,000, certified cost or pricing data is required.

"Cost or Pricing Data" Defined

TINA defines *cost or pricing data* as "all facts that, as of the date of price agreement, or, if applicable, an earlier date agreed upon by the parties that is as close as practicable to the date of price agreement, prudent buyers and sellers would reasonably expect to affect price negotiations significantly."

In general, pure judgments are not deemed to be "facts" and hence are not cost or pricing data. However, Calhoon and Sybert (2012) point out,

> Cost or pricing data includes more than just historical accounting data; they are all the 'facts' reasonably relevant to evaluate estimates of future costs and to the validity of costs already incurred. This may include, but is not limited to:
>
> 1. Vendor quotes;
> 2. Nonrecurring costs;
> 3. Information on changes in production methods and in production or purchasing volume;
> 4. Data supporting projections of business prospects and objectives and related operations costs;
> 5. Unit-cost trends such yield rates and labor efficiency;
> 6. Make-or-buy decisions;
> 7. Estimated resources to attain business goals; and
> 8. Some information on significant management decisions (Calhoon & Sybert, 2012, p. 13)

Although some of the above information is hard facts, estimates and projections also can be used as "cost or pricing" data. It is worth noting that for most major weapon programs where technology is unbelievably complex, a big component of a cost estimate is based on faithful estimates and educated projections.

TINA Exemptions

According to Calhoon & Sybert (2012), TINA can be exempted if one or more following situations applies:

1. Adequate Price Competition
2. Prices Set by Law or Regulation
3. Commercial Items
4. Pricing Actions Less Than $750,000
5. Exceptional Cases—Waiver by Head of Contracting Activity (p. 7)

Note that TINA waivers are rarely given; consequently, TINA governs most major DoD contracts.

TINA Is a "One-Size-Fits-All" Approach

From what is described above, one can see that TINA is a "one-size-fits-all" approach. TINA is essentially a blanket application with very limited exception. In particular, TINA application does not (at least not directly) vary with acquisition category, cycle, and contract type. Intuitively, this approach does not make sense. In the subsequent chapter, we detail our arguments against the one-size-fits-all approach and accordingly propose to tailor the use of TINA to various combinations of acquisition category, cycle, and contract type.

Customizing the Use (or Disuse) of TINA in the DoD Acquisition Process

In this section, we continue our investigation of the role of TINA in the context of DoD procurement. The objective is to provide a guideline for the use or disuse of TINA for various combinations of acquisition category, cycle, and contract type.

We employ an economics-based, incentive-centric approach that focuses on investigation of agents' (i.e., DoD contractors') incentives under various settings. We generate our policy recommendation through a "with" and "without" TINA comparison.

Two key decisions need to be made to answer our research question. Namely, what is the right contract type for each combination of category and cycle, and further, given the selected optimal contract type, shall we impose or waive the TINA?

Table 2. Graphical Illustration of the Research Question

Technological uncertainty and cost vagueness descending →

Information asymmetry descending ↓

Acquisition Category (Product)	Pre-Milestone A	Pre-Milestone B	Pre-Milestone C	Production and Deployment	Operations
MDAP					
ACAT II					
ACAT III					

Table 2 illustrates the task graphically. On the vertical dimension, as the acquisition category descends from I to III, so does the information asymmetry between the government and the contractor. On the horizontal dimension, as the life cycle matures, the technological uncertainty gets resolved progressively and the cost vagueness runs down.

According to the *Defense Acquisition Guidebook* (DAG) provided by Defense Acquisition University, various contracts ranging from CPFF to FFP represent different risk allocations between the buyer (i.e., the DoD) and the seller (i.e., the contractor). Figure 3, reproduced from the DAG, illustrates this.

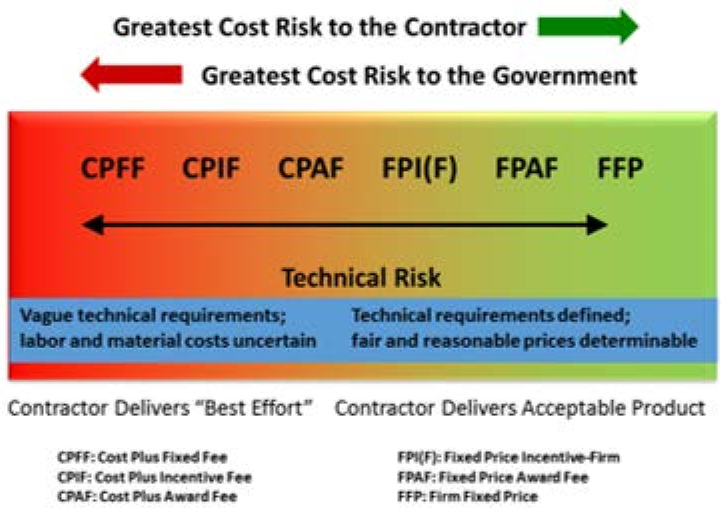


Figure 3. Risk to Contract Types

Moreover, the DAG also provides guidelines for the typical contract type that is used at different stages of the acquisition life cycle. Figure 4 is replicated from DAG.

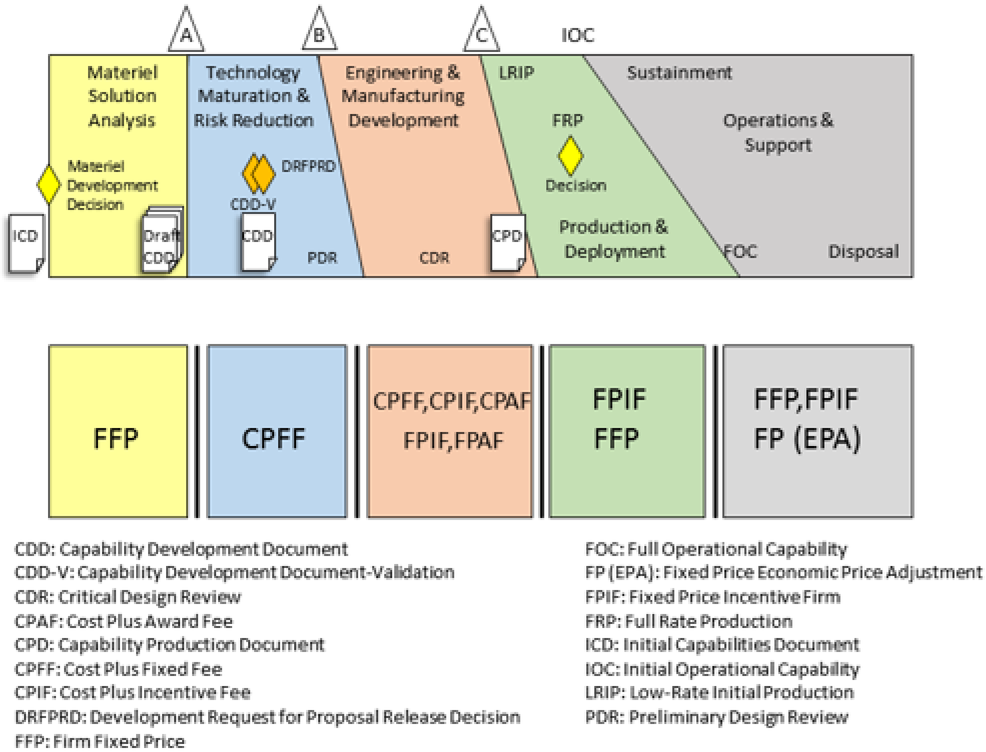


Figure 4. Typical Contract Types by Acquisition Phase

Applying the aforementioned framework of DAG to our Table 2 setting, which has one more dimension, acquisition category, we propose the following use of contract types for each cell of our Table 2.

The bottom line is: As the acquisition category descends from I to III, within the same life-cycle stage (with the exception of the first and last stage), we gradually shift toward the contract type that allocates more risk to the contractor and, in the meantime, takes away the risk from the DoD's shoulders. This induces contractors' better efforts.

Table 3. Contract Types

	Pre-Milestone A	Pre-Milestone B	Pre-Milestone C	Production and Deployment	Operations
MDAP	FFP	CPFF	CPIF	FPIF	FFP
ACAT II	FFP	CPIF	CPAF	FPIF/FFP	FFP
ACAT III	FFP	CPAF	FPIF	FFP	FFP

Now our task is to suggest either the use or disuse of TINA for each of the cells in Table 3. Let's first tabulate our recommendations in the following Table 4, followed by detailed explanations.

Table 4. Customizing the Use (Disuse) of TINA

	Pre-Milestone A (Material Solution Analysis)	Pre-Milestone B (Technology Maturity & Risk Reduction)	Pre-Milestone C (Engineering & Manufacturing Development)	Production and Deployment	Operations and Support
MDAP	FFP (TINA)	CPFF (TINA)	CPIF (No TINA)	FPIF (No TINA)	FFP (No TINA)
ACAT II	FFP (TINA)	CPIF (TINA)	CPAF (TINA)	FPIF/FFP (Maybe/MaybeNot TINA)	FFP (Maybe/Maybe Not TINA)
ACAT III	FFP (TINA)	CPAF (TINA)	FPIF (TINA)	FFP (TINA)	FFP (TINA)

Detailed Discussions/Justifications for Table 4:

1. The most notable part of this table is that it proposes a deviation from the current practice of TINA, which is essentially a one-size-fits-all prescription. Namely, we recommend varying the use or disuse of TINA with respect to acquisition category, acquisition life-cycle stage, and the corresponding preferred contract type.
2. For the red-colored cells, that is, ACAT I (MDAP) starting from Pre-Milestone C and continuing through the rest of the acquisition cycle, we propose to do away with the use of TINA. The polar case here, that is, the use of FFP in the context of MDAP, is thoroughly analyzed by Wang et al. (2016), where the authors identify the incentive problem that is characterized as a “moral hazard,” that is, a lack of effort from the contractor. Specifically, a contractor under an FFP contract that is subject to TINA has the following ill incentive: The fear of being held accountable for any significant unfavorable cost discrepancy (i.e., the actual incurred cost is significantly below the ex-ante cost estimate negotiated with the DoD as the basis for contract fixed-price) would strongly motivate the contractor to shirk (i.e., reduce cost-saving effort) or even engage in cost padding (e.g., by opportunistically incurring or allocating more costs to the government contracts). Such behavior leads to deadweight welfare loss that is ultimately borne by taxpayers.

   As shrewdly pointed out by Rogerson (1994), “TINA cannot force defense contractors to reveal the lowest possible cost that they could produce at if they exerted an optimal effort. Rather, it essentially tells them that the price they negotiate must be close to the cost they actually incur.”

   It is worth noting that for ACAT I (MDAP), even at the very late stage of the acquisition cycle, d*ue to the extreme complex technology and production process*, significant information asymmetry nevertheless exists between the contractor and the DoD. Consequently, the unverifiable part of the production cost is still significant and there is plenty of room for contractors to shirk or engage in cost padding. Hence, it is very essential to realize the unintended negative consequence of enforcing TINA in this particular setting, and a lax use or even disuse of TINA is preferred here to induce the contractors to reveal their best-effort cost.

The other two red-colored cells, that is, MDAP at Pre-Milestone C (Engineering & Manufacturing Development), and Post-Milestone C (Production and Deployment), adopt CPIF and FPIF, respectively. Both CPIF and FPIF belong to incentive contracts which are designed to induce cost-saving effort from contractors. To the extent that TINA exposes compliance risk to contractors in case of ex-post unfavorable cost variance, imposing TINA in these two cells would have similar unintended consequences, as discussed in Wang et al. (2015), hence we recommend a similar fix, that is, the disuse of TINA.

3. For the yellow-colored cells, we suggest no changes to the current TINA use. These cells include:
   a. ACAT III across all the life-cycle stages

      That is, no additional TINA waiver[3] is recommended for ACAT III. The primary reason for keeping TINA in place for ACAT III is the modesty of information asymmetry between the DoD and the contractor. Therefore, the verifiability of the program cost is good. When most of the cost information is verifiable, TINA is an effective mechanism to deter defective pricing.
   b. ACAT II life-cycle stages up to Pre-Milestone C

      Under this category, CPIF and CPAF are prescribed for Pre-Milestone B and Pre-Milestone C, respectively. In general, cost-plus contracts inherently suffer from the moral hazard problem. Hence, removing TINA does not make the problem go away. However, TINA does reduce the "defective pricing" incentive by imposing the litigation risk, at least for the verifiable part of the program cost. So the net benefit of "with TINA" minus "without TINA" is positive and we suggest a "stay-put" strategy.

      For the cell that intersects ACAT II and Pre-Milestone A (Material Solution Analysis), the prescribed contract type is FFP, yet we suggest the use of TINA. This is in contrast to what we suggest for the polar case discussed in ACAT I. The major reason is that for Pre-Milestone A, which is a pre-system acquisition stage, most of the conceptual refinement work is performed through analogy or parametric estimating methods. To the extent that the estimation is based on a similar existing item or mathematical model, a big part of the cost is verifiable. As argued before, TINA is an effective way of deterring "defective pricing" when the cost information is verifiable.

[3] The current applicable TINA waiver still applies, for example, if classified as commercial items

c. ACAT I (MDAP) life-cycle stages before Pre-Milestone B

For the same reason mentioned in the last paragraph, for the FFP contract used in Pre-Milestone A MDAP, we propose to keep TINA in place. For the cell that intersects MDAP and Pre-Milestone B, TINA is also retained to mitigate the incentive of engaging in "defective pricing."

4. For the purple-colored cells, we recommend the flexible use of TINA. Use or disuse of TINA should be dependent upon individual cases. On one hand, ACAT II, even at the last two stages of the life cycle, should still demonstrate non-trivial information asymmetry between the DoD and the contractor; therefore, our worry about the contractor's ill incentive under TINA and the related "moral hazard" problem remains. On the other hand, to the extent that ACAT II is much smaller and less complex than ACAT I (MDAP), the degree of information asymmetry should be much less severe than under MDAP. If the major part of the program cost is verifiable, then enforcing TINA can effectively prevent "defective pricing" from happening. Decision makers must run a horse-racing between the two offsetting factors and accordingly choose the use or disuse of TINA to maximize social welfare. For example, one can argue that if Technology Readiness Level (TRL) reaches 8 or above, then the use of TINA is preferred.

Conclusion

TINA, as it currently stands, is a "one-size-fits-all" prescription. Specifically, TINA does not differentiate among various settings involving different acquisition category, acquisition life cycle, and corresponding preferred contract type. We propose to tailor the use or disuse of TINA to different scenarios by considering the economic incentives created by TINA enforcement. In some settings where TINA is misplaced, we propose to drop TINA to remove the ill incentives and consequent unintended negative consequences. In other settings where TINA brings more benefit than cost, we recommend keeping TINA in place. In a few settings where the judgment is not unambiguous, we propose to leave the discretion to decision makers.

References

Calhoon, B., & Sybert, J. (2012). Truth in Negotiations Act (TINA) essentials. *National Contract Management Association Webinars.*

GAO. (2015). *Defense acquisitions: Better approach needed to account for number, cost, and performance of non-major programs* (GAO-15-188). Washington, DC: Author.

Laffont, J. -J., & Tirole, J. (1986, June). Using cost observation to regulate firms*. Journal of Political Economy,* 94, 619–641.

Laffont, J. -J., & Tirole, J. (1993). *A theory of incentives in procurement and regulation*. MIT Press.

McAfee, R. P., & McMillian, J. (1987, Summer). Competition for agency contracts. *Rand Journal of Economics, 18*, 296–307.

Melumad, N., & Reichelstein, S. (1989, April). Value of communications in agencies. *Journal of Economic Theory, 47,* 334–368.

Reichelstein, S. (1992). Constructing incentive schemes for government contracts: An application of agency theory. *The Accounting Review, 67*(4), 712–731.

Rogerson, W. (1994). Economic incentives and the defense procurement process. *Journal of Economic Perspectives, 8*(4), 65–90.

Schwartz, M. (2014). *Defense acquisitions: How DOD acquires weapon systems and recent efforts to reform the process*. (Congressional Research Service Report 7-5700). Washington, DC: CRS.

Wang, C., Rendon, R., Champion, C., Ellen, M., & Walk, J. (2016). *An economic analysis of the Truth in Negotiations Act* (NPS-CM-16-014). Monterey, CA: Naval Postgraduate School, Acquisition Research Program.

THIS PAGE INTENTIONALLY LEFT BLANK

Panel 3. Improving DoD Software Design and Sustainment

Wednesday, April 26, 2017	
11:15 a.m. – 12:45 p.m.	**Chair: Rear Admiral John W. Ailes, USN,** Deputy Commander, Fleet Readiness Directorate Space and Naval Warfare Systems Command ***The Policies and Economics of Software Sustainment: DoD's Software Sustainment Ecosystem*** Forrest Shull, Software Engineering Institute Michael McLendon, Software Engineering Institute ***Software Vulnerabilities, Defects, and Design Flaws: A Technical Debt Perspective*** Robert L. Nord, Carnegie Mellon Software Engineering Institute Ipek Ozkaya, Carnegie Mellon Software Engineering Institute Forrest Shull, Carnegie Mellon Software Engineering Institute ***Applying the Fundamentals of Quality to Software Acquisition*** Steve Bygren, The MITRE Corporation Greg Carrier, The MITRE Corporation Tom Maher, The MITRE Corporation Patrick Maurer, The MITRE Corporation David Smiley, The MITRE Corporation Rick Spiewak, The MITRE Corporation Christine Sweed, The MITRE Corporation

Rear Admiral John W. Ailes, USN—is a 1985 graduate of Oregon State University with a Bachelor of Science in Computer Science, and he was commissioned through the Naval Reserve Officers Training Corps (NROTC) program. He holds a Master of Science in Electrical Engineering from the Naval Postgraduate School.

His sea tours include strike warfare/communications officer, USS *Bunker Hill* (CG 52); combat systems officer, USS *Elliot* (DD 967); combat systems officer, USS *Russell* (DDG 59), and executive officer, USS *Lake Erie* (CG 70). Ailes was the commissioning commanding officer, USS *Chafee* (DDG 90).

Ashore, Ailes has served as the test and evaluation manager for the Navy Theater Ballistic Missile Defense Program; Aegis Theater Ballistic Missile Defense warfare area manager; Naval Integrated Fire Control project officer and surface ship combat systems manager in the Program Executive Office, Integrated Warfare Systems; commander, Naval Surface Warfare Center, Port Hueneme Division; major program manager, Littoral Combat Ship Mission Modules; and chief engineer, Space and Naval Warfare Systems Command.

Ailes assumed his current position as deputy commander, Fleet Readiness Directorate, Space and Naval Warfare Systems Command, in March 2016.

Ailes' decorations include the Legion of Merit Medal (two awards), Meritorious Service Medal (three awards), Navy and Marine Corps Commendation Medal (two awards), Navy and Marine Corps Achievement Medal, and various campaign, unit, and service awards.

The Policies and Economics of Software Sustainment: DoD's Software Sustainment Ecosystem

Forrest Shull—is Assistant Director for Empirical Research at Carnegie Mellon University's Software Engineering Institute. His role is to lead work with the U.S. DoD, other government agencies, national labs, industry, and academic institutions to advance the use of empirically grounded information in software engineering, cybersecurity, and emerging technologies. Dr. Shull has been a lead researcher on projects for the U.S. DoD, NASA's Office of Safety and Mission Assurance, the Defense Advanced Research Projects Agency (DARPA), the National Science Foundation, and commercial companies. He serves on the IEEE Computer Society Board of Governors and Executive Committee. [fjshull@sei.cmu.edu]

Michael McLendon—is an Associate Director, Software Engineering Institute, Carnegie Mellon University. Previously, he served as Senior Advisor, Office of the Assistant Secretary of Defense for Systems Engineering. He also was a principal in the Office of the Assistant Secretary of Defense for Program Analysis & Evaluation and in the Office of the Under Secretary of Defense for Policy. Dr. McLendon was Professor at the Defense Systems Management College and a career Air Force officer in a range of leadership and management positions in system and technology development and acquisition, as well as the federal level and the private sector. [mmclendon@sei.cmu.edu]

Abstract

Software is the foundational building material for the engineering of systems, enabling almost 100% of the integrated functionality of cyber physical systems—especially mission- and safety-critical software reliant systems—to the extent that these systems cannot function without software. As a result, it is imperative that the DoD has the capability and capacity to affordably sustain software-reliant systems and to continually operate and achieve mission success in a dynamic threat, cybersecurity, and net-centric environment.

The Carnegie Mellon University (CMU) Software Engineering Institute (SEI) has been performing studies to inform Departmental decisions regarding software sustainment policies and programs regarding complex weapon systems. These studies were based on interviews and discussions with sustainment centers across all of the Services, case studies on selected programs, and a literature review.

In this paper we present an overview of our initial study regarding the DoD's organic software sustainment infrastructure and its key components related to complex weapon systems, and a selection of key themes from our analysis of sustainment practices. There are two key takeaway messages. First, software sustainment is not effectively described with a model based on hardware (where sustainment can be treated as a discrete series of activities intended to restore form, fit, and function). Secondly, software sustainment is really about continuous engineering in which the software undergoes a series of engineering activities intended to deliver the latest capability to the warfighter, a task which is never "done."

Motivation

Software is the foundational building material for the engineering of systems, enabling almost 100% of the integrated functionality of cyber-physical systems—*especially* mission- and safety-critical software-reliant systems—to the extent that these systems cannot function without software. There is no plateau in sight for the advancement of software technology and its use by the DoD in new systems, as well as to enhance the capabilities of legacy systems and extend their operational value far beyond their designed service life.

Software can also be a major a source of defects and potential security vulnerabilities with potentially fatal consequences, due to the increased complexity of interactions among embedded software, the hardware platform, and its associated subsystems. The dynamics of the cyber environment and the constantly changing nature of the cyber threat mean that with software we are never “done.”

The issues surrounding sustainment become increasingly complex as the DoD’s reliance on software increases. For example, the ever-expanding reliance on software means that an increasing portion of the acquisition cost as well as the sustainment cost of systems is driven by software (NRC, 2010a). There is evidence that it is three to 10 times more expensive to mitigate software defects/vulnerabilities in sustainment rather than early in acquisition and development. Successfully mitigating this software cost trend, while still enhancing warfighter capability, must be an essential element in the DoD’s affordability strategy.

Therefore it is imperative that software sustainment be a priority in defining system requirements, design, and development. This means that the software sustainment community must be an active participant early in the requirements and engineering process and that the Product Support Manager in acquisition programs must be knowledgeable and proactive in representing software sustainment equities.

Another critical challenge is the magnitude of the DoD’s software sustainment inventory. The inventory is immense, but there is limited visibility and understanding at the enterprise level of the total size, complexity, and characteristics of the DoD’s software inventory, which may be trending toward one billion lines of custom developed software code or more.[1] Additionally, the engineering of systems also relies on an extensive portfolio of commercial-off-the-shelf (COTS) software, government-off-the-shelf (GOTS) software, and increasingly on free and open-source software (FOSS). The use of non-custom development software is pervasive across all DoD system domains and its use comes with significant technical and management challenges. The DoD relies on comprehensive and complex information systems to provide almost real time visibility and management of its wholesale and retail inventory of parts and supplies, but it has no similar capability for software.

It is imperative that the DoD have the capability and capacity to affordably acquire and sustain software-reliant systems to continually operate and achieve mission success in a dynamic threat, cyber, and net-centric environment. However, the DoD’s ability to produce high-quality software more affordably and efficiently across the system life cycle is a strategic challenge (NRC, 2010a). The acquisition and sustainment of software, particularly for distributed real-time and embedded systems, remains high risk and more problematic as system complexity continues to grow.

Research Goal, Scope, and Methodology

Our research goal was to *characterize* the factors that affect the effectiveness and cost of software sustainment in the DoD. For this initial step of our research, we did not intend to produce value judgments as to the efficacy or cost-effectiveness of different

[1] This is an estimate based on the limited data available and expert judgment. However, we note that several experts in the DoD software engineering community have expressed the opinion in discussions that the number may be even higher.

sustainment choices; rather, we intended to identify and describe the major factors that DoD organizations must manage, in order to impact software sustainment performance.

To provide a manageable focus for our work, we concentrated on describing the software sustainment ecosystem as it relates to complex weapon systems. Embedded software presents the most technically difficult and resource-intensive software engineering challenge because of tightly coupled interfaces, integration with unique hardware, real-time requirements, and very high reliability and assurance needs due to life-critical and mission-critical demands. Of course, the DoD's software sustainment challenge is broader than the software embedded in complex weapon systems. Mission critical non-embedded systems, mission support systems (e.g., test equipment, mission planning, engineering models, and simulations), and the range of business systems also present significant software sustainment challenges. While our initial results can be used to understand the software sustainment ecosystem for other types of software intensive systems, a more detailed description of how the factors apply to those domains will be a subject of future work.

Our team leveraged multiple streams of data and information for this study.

- Literature Search—The body of knowledge related to software engineering is extensive. The formal body of knowledge, academic research, and practitioner publications has evolved so that there are now various communities of interest and professional organizations that focus on software engineering.[2] However, there has been limited systematic study focused on DoD software sustainment, so there is no organized set of literature and ongoing study or research agenda to create and refresh a software sustainment body of knowledge.
- SEI DoD Engagements—The SEI has been actively engaged with the military services for three decades to provide technical expertise to enhance organizational capabilities (processes, practices, and competencies) for software engineering across the life cycle and to solve technical challenges for specific weapon system and information system programs. This has included continuous engagement with the principal Army, Navy, and Air Force software sustainment centers to include the provision of knowledge and practice to institutionalize CMMI.
- DoD IPT Report—The SEI had access to the data and results of the DoD UAS Software Sustainment Integrated Product Team (IPT) Report. This IPT effort, which concluded in 2015, was led by the office of the DASD (MP&P). The SEI was invited to serve as an ex officio member of the IPT, which made visits to a number of Army, Navy, and Air Force organic software sustainment organizations.
- Interviews with Key Leaders—It was critical that the SEI inform its analysis based on the views of decision-makers who influence a range of software

[2] For example, the Software Engineering Body of Knowledge (SWEBOK), a community-driven approach using an open consensus model to document generally accepted software engineering knowledge under the leadership of the IEEE Computer Society (https://www.computer.org/web/swebok). A DoD-specific example is the Software Assurance Community of Practice.

sustainment policies, programs, and resource allocation. To that end, the SEI complemented its research with information from meetings with key leaders across all three Services, including those in the Senior Executive Service (SES), senior managers and staff in OSD, and from industry. This study was conducted at the unclassified level, and our interviews with DoD sustainment staff were conducted under the conditions of non-attribution to enable an open exchange of perspectives with senior leaders, managers, and staff engaged in software sustainment. These conversations provided context for understanding the evolution of the DoD's current software sustainment posture and enabled the SEI to refine its model of the software sustainment ecosystem.

DoD's Organic Software Sustainment Organizational Infrastructure

The DoD's organic software sustainment organizations successfully respond to a range of customer needs and deliver critical software updates and enhancements, often under the intense schedule pressure of wartime operations, to deliver critical warfighter capability. This organic infrastructure is composed of a number of principal organizations and a myriad of other smaller organizations and offices that have not been fully identified and characterized. To a large degree, in our view, the critical role and functions of these organizations are not well understood or visible. These organic organizations are structured and resourced in different ways by each Service, each performing software sustainment utilizing a variety of government and contract staffing strategies. The Services employ a variety of business model strategies in making decisions about allocating sustainment workload across their organic software development capabilities and the defense industrial base (DIB), as well as structuring public–private sector partnerships. These decisions are made within the context of a number of statutory requirements and DoD policies, such as determination of core requirements and the 50% ceiling, measured in dollars, on the amount of depot maintenance workload that may be performed by a contract with industry for a military department or defense agency during a fiscal year.

The Nature of Software in Systems

One of the keys to addressing the software sustainment challenge is to understand the nature of software in DoD systems. The characteristics of software relative to hardware are generally not well appreciated, especially in relation to how the DoD traditionally uses the term maintenance.

The critically important and growing role of software in defense systems has been noted in many prior studies (DSB, 2000; NRC, 2010a, 2010b). This growth is due in many ways to the unique characteristics of software, as summarized eloquently in a study by the National Academy of Sciences:

> This growth is a natural outcome of the special engineering characteristics of software: Software is uniquely unbounded and flexible, having relatively few intrinsic limits on the degree to which it can be scaled in complexity and capability. Software is an abstract and purely synthetic medium that, for the most part, lacks fundamental physical limits and natural constraints. For example, unlike physical hardware, software can be delivered and upgraded electronically and remotely, greatly facilitating rapid adaptation to changes in adversary threats, mission priorities, technology, and other aspects of the operating environment. The principal constraint is the human intellectual

> capacity to understand systems, to build tools to manage them, and to provide assurance—all at ever-greater levels of complexity. (NRC, 2010a)

These aspects of software are not always well understood or at least addressed in practice. For example, much of DoD depot policy (and industrial base policy) remains hardware-centric, despite software enabling an increasingly large percentage of system functionality. Due to its "uniquely unbounded and flexible" nature, the sustainment of software operates very differently from that of other building materials of contemporary systems.

Software is not a "physics of failure" domain, which is to say that software itself does not wear out or degrade over time. Maintenance at any organizational level for hardware typically focuses on returning components categorized as repairable items, such as avionics line-replaceable units (LRUs), to their original functional condition and configuration by replacing parts, using smaller electronic components, or treating corrosion. This typically involves applying standardized processes and procedures for diagnostics and repair. In the case of software, sustainment takes the form of making intentional changes to the software source code and related work products for many different reasons, not exclusively (or in many cases even primarily) driven by correction of failures. These changes are driven by a number of goals, such as to correct a flaw, to mitigate a security vulnerability, to make fact-of-life changes due to systems and system-of-system interface and interoperability impacts, and to incorporate system enhancements that deliver greater warfighter capability.

Demand and funding requirements for software sustainment do not scale with operational tempo or the size of the force structure. From a hardware or weapon system platform perspective, depot maintenance or sustainment demands and funding are routinely forecasted on the basis of the number of reparable units anticipated, taking into account certain factors such as reliability, flying hours, miles driven, engine hours, number of landings, or calendar time since last overhaul. From a software perspective, sustainment is about applying the disciplines of systems and software engineering (knowledge, processes, practices, and skills) each time the software is touched.

Due to the complexity of software, the great majority of the software sustainment effort is spent on the analysis of the specific need for a change and then designing, implementing, and testing a unique change. Once implemented, it is trivial to make additional copies of the new configuration version of the software system, and generally it is inexpensive to push out updates for all the instances of the weapon system in the force structure inventory. Further, the number of a particular type of weapon system that is in the force structure is not the driver of software sustainment. In other words, relying on a "cost per asset" analysis can be hugely misleading. Since costs are independent of the quantity of a given system in the inventory or force structure, dividing over a fairly small fleet like the B-2s is a misleading comparison with other systems. Another factor to consider is the differences in complexity of systems and the associated complexity of sustainment from system to system. Addressing cybersecurity issues is another distinction.

Software quality is related not only to operational failures but also to technical debt—that is, the reflection of inadequate attention to the design of the software architecture coupled with developers optimizing short-term goals (like the ability to deliver code on time) over longer term impacts (such as the need to create clean, well-organized code that is easy to maintain). Technical debt, as generally understood, affects the internal quality of the code and its extensibility to more easily accommodate change. However, technical debt does not necessarily impact behaviors of the software that would be visible to the end user. Technical debt's impact on architecture and internal software quality directly affects the scope, magnitude, and complexity of software sustainment.

The scope and complexity of the technical debt in an individual program is also driven by the complexity of the software supply chain. The number of different suppliers in the software supply chain for a weapon system program can be extensive. A program often has limited visibility and understanding of the architecture considerations and software practices (not only for development but for assurance as well) that each vendor employs and the degree to which there is a consistent approach to software development for the entire program. From a software sustainment perspective, organic sustainment organizations inherit the cumulative technical debt generated from the multiplicity of software development efforts on one program.

An implication of the points highlighted above is that software sustainment is more usefully viewed as *continuous engineering* rather than a set of discrete maintenance activities. Software sustainment enables an ongoing evolution of system capability to address the changing environments in which DoD systems are deployed, especially related to ongoing changes in cyber threats.

Policy Context

Software sustainment organizations plan and execute their missions within the context of the existing depot maintenance and associated governance environment. As highlighted in Figure 1, the DoD, and in turn the Services, promulgate depot maintenance policy and guidance based on a number of statutory requirements. These mandates then drive decisions relative to planning and executing software sustainment.

The overall direction and guidance for software sustainment are based on statutory requirements in Title 10 USC and DoD policies for depot-level maintenance. Figure 1 summarizes the relevant Title 10 USC statutes that influence product support and depot maintenance decisions. As a result, the legacy of the DoD's depot maintenance paradigms and policies is rooted in a hardware-centric paradigm. In turn, each Service has developed its own guidance for implementing the DoD's policy to address Service software sustainment needs within the depot maintenance framework.

- Title 10 USC § 2208 Working Capital Fund
- Title 10 USC § 2320 Data Rights
- Title 10 USC § 2302 Definitions (a.k.a Section 805, Pub. Law 111-84)
- Title 10 USC § 2337 Life Cycle Management and Product Support
- Title 10 USC § 2460 Depot Maintenance Level and Repair
- Title 10 USC § 2464 Core Depot Maintenance Level and Repair
- Title 10 USC § 2464 Ready and Controlled Source of Technical Competencies
- Title 10 USC § 2466 Limitations of the Performance of Depot Level Repair (50/50)
- Title 10 USC § 2472 Prohibition on Management of Depot Employees by End Strength
- Title 10 USC § 2474 Private Sector2474 CITE, Public-Private Partnerships
- Title 10 USC § 2476 6% Capital Investment in Organic Depots
- Title 10 USC § 2563 Sales to Persons Outside DoD
- Title 10 USC § 2667 Lease of Government Property (Enhanced Use Lease)

Figure 1. Title 10 USC Laws Influencing Sustainment

The DoD's Depot Source of Repair (DSOR) practices also drive software sustainment decisions. In practice, the Program Manager conducts a level of repair analysis (LORA) to determine if there is a depot level requirement (DAU, 2017). The PM also conducts a core logistics assessment (CLA) to determine in accordance with Title 10 USC § 2464 if there is a requirement to establish an organic (core) depot maintenance capability (i.e., government owned and government operated [GOGO]). This practice evolved from a focus on hardware and is now applied to software.

A key factor that drives software sustainment is the program manager-centric nature of decisions about sourcing strategies for the sustainment of specific weapon system programs. These program manager decisions ripple through and impact virtually every component of the software sustainment ecosystem. These program-specific policy decisions, in our view, may not be balanced with considerations for optimizing the DoD software sustainment enterprise to contribute to greater enterprise affordability and productivity.

The Software Sustainment Ecosystem Factors

Based on our findings, we believe that the software sustainment infrastructure is best described and understood as an *ecosystem* composed of interrelated elements. We found over and over that the factors that drive software sustainment are highly interrelated. For example, it is difficult to discuss the workforce needed to perform necessary sustainment activities without first understanding the business model in terms of public-private partnerships, which activities can be done by contractors, and which activities need to remain in the organic DoD workforce. Decisions about the nature and types of these business models may also be influenced by the degree to which the government has provisioned for and exercised its technical data rights for a given program at the time of developing an acquisition strategy and contract. These decisions have implications for the scope of the software sustainment system. Because of the high degree of connectivity that exists among the drivers and factors, we use the metaphor of an "ecosystem" to describe the interdependencies among these elements; decisions made at any point are affected by and affect whole series of other decisions.

There are many variables that are inherent in this ecosystem, not the least of which is time. The time variable is one of the key factors that makes this ecosystem dynamic. There is a time dependency among and between certain software sustainment demand drivers and the critical elements. For example, demands for software changes are frequent, and the underlying technology of software changes rapidly. Failure to invest in software quality up front during initial system development creates a bow wave of risk and technical debt that may continue for decades. Similarly, inadequate investments early in software workforce capital, tools, and engineering processes will increase the cost of sustainment. *In operation, the software sustainment ecosystem is dynamic.*

Based on our research, we created a framework that describes the software sustainment ecosystem, depicted in Figure 2. We abstracted the issues raised in our discussions with DoD sustainment stakeholders into six *demand drivers* and 10 *ecosystem elements.*

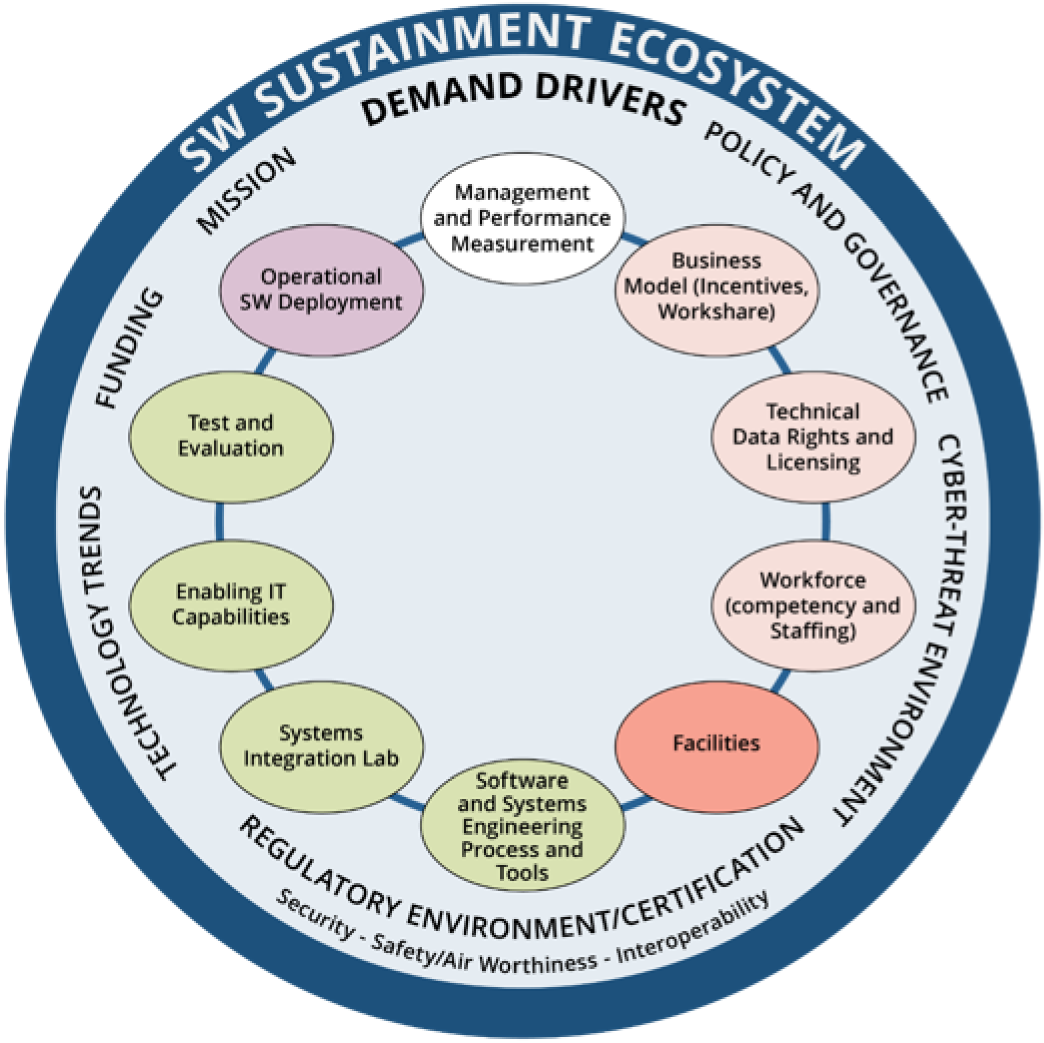


Figure 2. The DoD Software Sustainment Ecosystem Framework

These six **demand drivers** capture the fact that DoD systems exist in an environment that is highly dynamic, where there is a need to respond to constantly changing threats and mission needs. This dynamism drives many of the system changes that need to be made during software sustainment. For many of these changes, the most cost-effective way of implementing the new capability relies on the unique flexibility of software.

Our work with the DoD software sustainment community continually highlighted an array of what some called constraints, factors, or “outside the organization influences” that directly impact software sustainment planning and execution. We mapped these considerations into the six higher level demand drivers. These demand drivers include policy (which may include formal guidance such as the Defense Acquisition Guidance and standards) and governance, the nature of the mission, the cyber and mission threat environment, funding, technology trends, and regulatory/certification requirements.

The 10 **ecosystem elements**, shown as interconnected “bubbles” within Figure 2, are the tightly interconnected factors that sustainment organizations need to manage in order to effectively and continuously engineer the software. The drivers and elements of this ecosystem represent a virtual spider web of linkages and relationships. The ecosystem elements shown in the figure are as follows:

The four **infrastructure** elements are the basic, fundamental resources that are necessary for the sustainment activities to occur.

- **Systems and Software Engineering Process and Tools—**The engineering practices to be applied to plan and execute the work.
- **Enabling IT Infrastructure—**The information technology environment and assets upon which the work must be conducted.
- **Test and Evaluation (T&E)—**The mechanisms by which changes made during software sustainment are verified as ready to be rolled out to users. For DoD weapons systems, significant investments in program-specific hardware may be required.
- **Systems Integration Laboratory (SIL)—**The SIL is a specific type of T&E equipment, providing accurate analysis of the impact of changes, and is increasingly important to DoD sustainment practice.

The three **knowledge and expertise** elements include the factors that describe how the necessary skill sets are brought to bear for sustainment activities and how the government grows its organic workforce and gets access to necessary technical information—perhaps with some level of interaction with the private sector—in order to deliver and deploy the capabilities that need to go to the warfighter.

- **Workforce (Competency and Staffing—**The means of accessing a sufficient organic workforce with appropriate skill sets.
- **Business Model (Incentives, Workshare)—**The strategic decision regarding which parts of the work will be done by the organic workforce and which by contractors, and how the overall work is managed both technically and contractually.
- **Technical Data Rights and Licensing—**The tactical decisions governing what technical information is necessary to be accessed by the organic workforce, and the mechanisms by which they have access.

Three ungrouped elements complete the ecosystem.

- **Facilities—**The physical location that meets the needs of the work (providing sufficient space, security levels, etc.).
- **Operational Software Deployment—**The mechanisms and strategy by which new versions of the software under sustainment are delivered to users.
- **Management and Performance Measurement—**The management function necessary to organize and monitor the work being conducted to ensure that it is executing as planned, and to identify any problems that need to be resolved.

Conclusion

The DoD's ability to continually evolve warfighter capability to address the dynamics of the vulnerability and thread environment is driven more and more by the affordable and timely continuous engineering of a system's software. However, there has been limited enterprise visibility and management of the DoD's critical organic software sustainment infrastructure. This paper provides insights into this complex issue, and we expect to provide more detailed information describing the software sustainment ecosystem when our report is cleared for broader distribution.

We also hope that the current summary can be useful for multiple stakeholders in the DoD, as a way to understand the unique issues related to the sustainment of software. Software is continuing to provide a greater percentage of the capabilities to be found in DoD weapons systems—and providing an increasing percentage of system cost as well. For both of these trends, no plateau is in sight. The unique flexibility and usefulness of software will make it central to sustainment strategies for adapting systems to the ever-changing mission needs and cyber-threat environment for the foreseeable future. We hope that the brief synopsis in this paper and the discrete ecosystem factors that we identified help to articulate many of the software-specific issues that are needed to do that effectively.

References

Defense Acquisition University (DAU). (2017.) Level of repair analysis (LORA). In *Acquipedia*. Retrieved from https://shortcut.dau.mil/acq/lora

Defense Science Board (DSB). (2000, November.) *Report of the Defense Science Board Task Force on defense software*. Washington, DC: Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA385923

National Research Council (NCR). (2010a.). *Critical code: Software producibility for defense.* Washington, DC: National Academies Press. doi: https://doiorg/10.17226/12979

National Research Council (NCR). (2010b.) *Achieving effective acquisition of information technology in the Department of Defense.* Washington, DC: National Academies Press. doi: https://doi.org/10.17226/12823

Acknowledgments

During the course of this study, the SEI interacted with a range of senior leaders, key managers, and staff across the DoD's software sustainment organizations. Universally, the Services and those we engaged were generally responsive to our data and information requests and candidly shared their experiences and perspectives on this critical and complex issue. We are exceedingly grateful for the many thoughtful conversations and insights we received in the course of this work.

This material is based upon work funded and supported by the DoD under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

Disclaimer & Distribution Statement

Software Vulnerabilities, Defects, and Design Flaws: A Technical Debt Perspective

Robert L. Nord—is a Principal Researcher at the Carnegie Mellon University's Software Engineering Institute (SEI). He is engaged in activities focusing on managing technical debt, agile and architecting at scale, and works to develop and communicate effective practices for software architecture. He is co-author of *Applied Software Architecture and Documenting Software Architectures: Views and Beyond* and lectures on architecture-centric approaches. Dr. Nord is a recognized leader in the software engineering community, both as a thought leader in software architecture and through his work helping industry and government customers at the SEI. He is a distinguished member of the ACM. [rn@sei.cmu.edu]

Ipek Ozkaya—is a Principal Researcher and Deputy Technical Lead of Software Architecture Practices at Carnegie Mellon University's Software Engineering Institute (SEI). With her team at the SEI, she works to help government and industry organizations improve their software development efficiency and system evolution through better use of agile architecting and technical debt management. Dr. Ozkaya serves as the chair of the departments of the *IEEE Software* magazine and as an adjunct faculty member for the Master of Software Engineering Program at Carnegie Mellon University. She is the co-author of articles on software architecture and technical debt management. [ozkaya@sei.cmu.edu]

Forrest Shull—is Assistant Director for Empirical Research at Carnegie Mellon University's Software Engineering Institute. His role is to lead work with the U.S. DoD, other government agencies, national labs, industry, and academic institutions to advance the use of empirically grounded information in software engineering, cybersecurity, and emerging technologies. He has been a lead researcher on projects for the DoD, NASA's Office of Safety and Mission Assurance, the Defense Advanced Research Projects Agency (DARPA), the National Science Foundation, and commercial companies. He serves on the IEEE Computer Society Board of Governors and Executive Committee. [fjshull@sei.cmu.edu]

Abstract

Technical debt describes a universal software development phenomenon: "Quick and easy" design or implementation choices that linger in the system will cause ripple effects that make future changes more costly. Although DoD software sustainment organizations have routine practices to manage other kinds of software issues, such as defects and vulnerabilities, the same cannot be said for technical debt. In this work, we discuss the relationships among these three kinds of software anomalies and their impact on software assurance and sustainable development and delivery. Defects are directly linked to external quality, and vulnerabilities are linked to more specific security concerns, but technical debt concerns internal quality and has a significant economic impact on the cost of sustaining and evolving software systems. Emerging research results and industry input demonstrate there are clear distinctions that call for different detection and management methods for defects, vulnerabilities, and technical debt. We draw from concrete examples and experience to offer software development practices to improve the management of technical debt and its impact on security.

Introduction

Software engineers face a universal problem when developing and sustaining software: weighing the benefit of an approach that is expedient in the short-term, but which can lead to complexity and cost over the long-term. In software-intensive systems, these tradeoffs can create technical debt (Kruchten, 2012), which is a design or implementation construct that is expedient in the short-term, but which sets up a technical context that can make future changes more costly or even impossible. Accumulating technical debt in the form of design shortcuts can be a strategic approach for software developers to accelerate development and optimize resource management without impacting overall quality—as long as the debt is eventually paid off (i.e., the time is taken to improve the software quality). The results of recent practitioner-focused empirical industry studies reveal that most systems are also suffering from *unintentional* technical debt, that is, the quick and dirty choices that accumulate with no strategic thought.

An increasingly important cost driver for DoD systems is the effort that is put into developing, acquiring, and sustaining software-intensive systems. Most DoD systems are in operation for extensive periods of time—likely for multiple decades—and continue through sustainment to evolve to incorporate new functionality. Even in mature systems, ongoing sources of software changes may include changing mission profiles, the need to incorporate more effective or efficient technologies into the system, or the need to repair newly discovered software vulnerabilities. For all of these issues, software tends to be the logical and cost-effective way to make the change, meaning that in effect, software is never "done." Consequently, dealing with technical debt is an unavoidable phenomenon for the DoD. For systems on which technical debt has been allowed to accumulate (or said another way, sufficient emphasis has not been placed on maintaining software quality), dealing with this never-ending stream of changes becomes increasingly less cost-effective, as more and more effort is required to comprehend and work within a poor quality system rather than on focusing on implementing new capabilities. This can result in cost and schedule slippage or diminished abilities to field new capabilities for the same amount of effort.

The conventional approach many organizations take to managing such cost and schedule issues is to assess software project and process performance through metrics. An important class of these metrics focuses on software defects, since correcting defects (especially late in the software life cycle) can represent significant expenditures of unplanned effort. Cyber vulnerabilities, once detected, are typically candidates for focused effort to repair or mitigate quickly. Our overarching research question is this: **Will DoD systems see improved outcomes if they manage technical debt explicitly, along with these other classes of software anomalies?**

Impacts of Technical Debt

Indeed, a growing body of research indicates that focusing on defect management provides insight to only one perspective of the schedule, cost, and quality management problem. Empirical studies have shown that if technical debt is not paid back in a timely manner, it correlates with greater likelihood of defects (Falessi, 2015), unintended rework (Li, 2014), and increased time for implementing new system capabilities in software (Kazman, 2015).

For example, code quality issues such as dead code or duplicate code add to the technical debt. They do not affect the functionality seen by the end user but can impede progress and make development more costly over time. Software architecture plays a significant role in the development of large systems; flaws in a software system's design, such as a frequently changing interface between two classes (an unstable interface; Xiao, 2014), can also add significantly to the technical debt.

Technical debt can have observable adverse consequences on software security as well, an issue of high priority for DoD software-intensive systems, meaning that allowing debt to accumulate may be even more costly. Some vulnerabilities may be inadvertently introduced as the result of technical debt: for example, if a vulnerability is fixed in one location but is not fixed in a similar duplicated code fragment, or if overly complex code makes it harder to reason about whether a dangerous corner-case condition is feasible or not. Alternatively, as we will show, technical debt can also be caused by addressing a vulnerability's symptoms rather than its root cause. Both of these relationships motivate a better understanding of the complex relationship between software vulnerabilities and technical debt. Therefore, we advocate that in order to get a better handle on their software quality, DoD programs should move toward also tracking their technical debt, similar to how they may be tracking defects and vulnerabilities (Figure 1).

In the remainder of this paper, we review the state of practice in managing technical debt and illustrate the need to manage all three types of software anomalies by summarizing results from our previous study, looking at the relationship between software vulnerabilities (Nord, 2016) and technical debt to address the following question: Are software components with accrued technical debt more likely to be vulnerability-prone? We present findings from a study of the Chromium open source project that motivates the need to examine a combination of evidence: quantitative static analysis of anomalies in code, qualitative classification of design consequences in issue trackers, and software development indicators in the commit history. Understanding this relationship can provide DoD programs with (a) ideas for improving their software engineering practices in better facilitating software quality assessment initiatives through data-driven analysis as presented in this work, and (b) an approach to better take advantage of existing analysis tools to help them focus what areas of their software to improve.

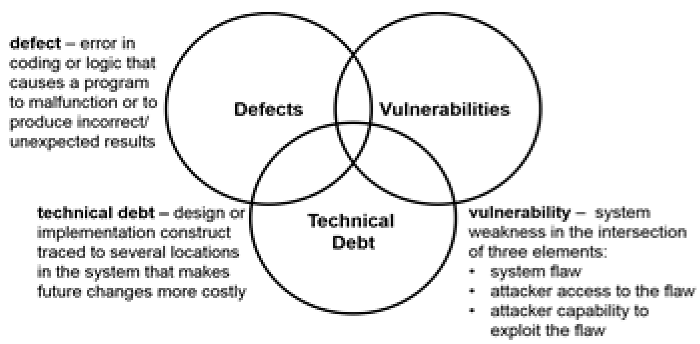


Figure 1. Software Anomalies

State of Practice in Managing Technical Debt

The technical debt metaphor is widely used to encapsulate numerous software quality problems. The metaphor is attractive to practitioners as it communicates to both technical and nontechnical audiences that if quality problems are not addressed, things may get worse. It is also very applicable in government sustainment contexts, as often the organizations that deal with the debt and those that take on the debt are not the same. While there has been significant progress made in creating an empirical and theoretical basis for identifying, quantifying, and managing technical debt (Spinola, 2012), there is still a lot of opportunity for improvement (Avgeriou, 2016).

Major software failures—for example, the recent United Airlines failure and New York Stock Exchange glitch or the National Security Agency's call data collection discrepancy—are being recognized in the popular media as the result of accumulating technical debt (Felten, 2014; Tufekci, 2015). In 2012, researchers conservatively estimated that for every 100 KLOC, an average software application had approximately US$361,000 of technical debt, the cost to eliminate the structural-quality problems that seriously threatened the application's business viability (Curtis, 2012). The undeniable message is that technical debt is real and significant. Industry and government organizations have started to respond to this message, most significantly demonstrated by increasing initiatives focusing on analyzing code quality. Yet repeatable, data-driven studies that can help quantify this understanding, especially in the context of government systems, still lag behind.

The results of our recent, broad practitioner survey of 1,831 software engineers and managers, including industry and government participants, demonstrate that they share a common understanding of the concept of technical debt (Ernst, 2015). According to participants, the lack of proven tool support to accurately identify, communicate, and track technical debt is a key issue and remains a gap in practice. More than half of the participants of our survey reported using issue trackers to communicate technical debt either explicitly ("technical debt" is mentioned) or implicitly (the concept of "technical debt" is discussed but not explicitly mentioned). This is consistent with anecdotal feedback from our own experiences of working with organizations, as well as case studies represented in literature on technical debt (Zazworka, 2013). In the absence of validated tools to concretely communicate technical debt and its consequences, developers resort to practices they are familiar with. Our work in this paper contributes to closing the gap between system analysis and understanding of observed problems of technical debt, demonstrated as security issues.

Analysis Approach

To understand whether software components with accrued technical debt are more likely to be vulnerability-prone, we need to take into account data from multiple sources: quantitative static analysis of anomalies (faults, vulnerabilities, design flaws) in code, qualitative classification of design consequences in issue trackers, and software development indicators in the commit history. In this paper we present results from our analysis with Chromium open source project (Barth, 2008; Camillo, 2015). This is a complex web-based application that operates on sensitive information and allows untrusted input from both web clients and servers. We use it as a representative test bed of typical technical debt issues and types of vulnerabilities. The Chromium open source project released Version 17.0.963.46 (referred to as Chromium 17 from here on) on February 8, 2012. This release contained 18,730 files. From February 1, 2010, to February 8, 2012, there were 14,119 bug issues reported as fixed (Chromium 2017).

A challenge we observe with some DoD programs is that this type of data is not always available and different development parties are not incentivized correctly to share this information in a timely way with key decision-makers. Therefore, replicating this study with a DoD software-intensive system has its challenges, although we expect the underlying relationships would hold equally well in the DoD context.

Our analysis approach is as follows:

1. Identify software vulnerabilities.
    a. Enumerate issues in the Chromium issue tracker (Chromium 2017) that have the `security` label.
    b. Classify each issue in terms of its Common Weakness Enumeration (CWE) using the issue's description, comments, metadata, and patch.
    c. For each issue, identify the set of files changed by commits that reference the issue.
2. Identify technical debt.
    a. Classify issues for technical debt.

b. Classify the type of design problem and rework based on the issue description, comments, and metadata.
c. Detect design flaws that co-exist in the same files changed to fix the issues labeled `security`.

3. Model the relationships between technical debt issues and vulnerabilities in the common artifacts they represent (code files, issues, commits).
   a. Extract concepts related to vulnerability types.
   b. Test whether technical debt indicators (e.g., number and type of design flaws, number of traditional bugs, number of bugs labeled security, and the lines of code that change to fix a bug) correlate with the number of vulnerabilities reported.
   c. Manually investigate how selected vulnerabilities are influenced by the correlated technical debt indicators.

We will show how design knowledge can help identify other related issues and files so that developers can more efficiently diagnose the root cause of vulnerabilities and provide a long-term fix.

Analysis Results

To identify vulnerabilities, we used the issues labeled security. Using the Chromium project's issue tracker, we identified 79 software vulnerability issues, which were related to 289 files in which we detected design flaws (described in the next section). An issue labeled security may have a well-identified security bug, such as a null pointer exception. Such an issue may not represent technical debt but could simply be an implementation oversight. On the other hand, some issues may manifest themselves with multiple symptoms. This can hint that technical debt contributed to the vulnerability.

Following this exercise we classified whether each issue was technical debt or not using the classification approach we developed (Bellomo, 2016).

To classify source code files, we used the results of a study that analyzed Chromium 17 and reported 289 files associated with design flaws that can be detected in the code. The approach analyzes a project's repositories—its code and its revisions—to calculate a model of the design as a set of design rule spaces (DRSpaces; Xiao, 2014). These DRSpaces are automatically analyzed for design flaws that violate proper design principles. Four types of design flaws can be identified from the DRSpace analysis: modularity violation, unstable interface, clique, and improper inheritance. A modularity violation occurs when files with no structural relation frequently change together. This suggests that those files share some secret or knowledge and that information has not been encapsulated or modularized. An unstable interface occurs when there is an important class or interface that many other files depend on, and this class is buggy and changes frequently, requiring its "followers" to also change. Clique refers to a cross-module cycle that prevents groups of modules from being independent of each other. Improper inheritance occurs when the parent class depends on the child or when another file depends on both a parent and its child class. We consider these flaws as indicators of technical debt.

Table 1. Design Flaws and Issues Classified as Technical Debt

	Classified Not TD	Classified TD
No Design Flaw	8	6
Design Flaw	50	15

Table 1 shows our results where we found 15 issues classified as technical debt that also demonstrate design flaws. When we analyze these results for correlations using Pearson correlation coefficient, we see promising results. Design flaws demonstrate correlations with number of bugs (0.921), bug churn (0.908), number of security bugs (0.988) and security churn (0.826). Our further analysis shows that for three of the four types of design flaws—modularity violation, clique, and improper inheritance—files with vulnerabilities are also more likely to have design flaws. The more types of design flaws a file is involved in, the higher the likelihood of it also having vulnerabilities. We look at the design concepts represented in these issues related to vulnerabilities to better understand overarching correlations. Table 2 summarizes the vulnerabilities of those issues that also reported design problems from the 79 issues we classified, in the form of CWE categories (CWE 2017).

Table 2. Affinity Groups of Vulnerability Types

Affinity	CWE	#Issues
interface	200: Information Exposure	1
resource arbitration	362: Concurrent Execution using Shared Resource with Improper Synchronization	3
	400: Uncontrolled Resource Consumption	3
invalid result	20: Improper Input Validation	2
	451: User Interface (UI) Misrepresentation of Critical Information	2
	476: NULL Pointer Dereference	1
	704: Incorrect Type Conversion or Cast	1
	825: Expired Pointer Dereference	1
boundary conditions	125: Out-of-bounds Read	1
	703: Improper Check or Handling of Exceptional Conditions	4
	787: Out-of-bounds Write	2
privilege	250: Execution with Unnecessary Privileges	2
	269: Improper Privilege Management	1
	285: Improper Authorization	1

Our study revealed that developers are already using concepts related to technical debt when investigating security issues, including the following:

- getting to the *root cause*
- understanding the *underlying design* issues
- recording symptoms where changes are taking *longer than usual* or problems are reoccurring
- predicting consequences for the *longer term*
- building evidence for a more *substantial fix*

Furthermore, when we studied the issues in detail, we observed that finding the true design root cause of the problems, i.e., the underlying technical debt, took a substantial amount of resources of the developers. DoD software deals with these challenges, where many small issues like these add up daily to accumulate to not only jeopardize sustainment resources, but also operational issues such as vulnerabilities that cause significant risks for the DoD.

Conclusions and Future Work

Software system vulnerability and technical debt are high priority concerns for our DoD software base and industry alike, leading us to address research questions such as:

- Are software components with accrued technical debt more likely to be vulnerability-prone?
- Does understanding the difference and similarities between technical debt, defect, and vulnerabilities lead to their better management?

Our studies on open source, industry, and government software data demonstrate that a conscious focus on understanding design issues that accumulate consequences in the form of vulnerabilities, extensive rework, and maintenance issues create the most risky technical debt items. The state of the practice in industry and the DoD alike is that such issues are not explicitly tracked and understood. Our results demonstrate that it is those areas that in the long run create both the highest operational and sustainment risks.

Understanding and calling out similarities and differences between defects, vulnerabilities, and technical debt has a direct impact on acquisition practices such as software risk management and technical tradeoffs that impact contracting decisions. Measurement and analysis techniques for managing technical debt in the long run improves sustainability of systems and impacts better buying power. When they address security issues, software developers use technical debt concepts to discuss design limitations and their consequences on future work. One time-consuming relationship between vulnerabilities and technical debt is tracing a vulnerability to its root cause when it is the result of technical debt. Introducing technical debt measurement and analysis potentially improves finding such root causes.

Our ongoing and future work focuses on creating intelligent mechanisms to extract and analyze this data for software development professionals and provide guidelines that DoD decision-makers can use to allocate their scarce resources most appropriately.

References

Avgeriou, P., Kruchten, P., Nord, R., Ozkaya, I., & Seaman, C. (2016). Reducing friction in software development. *IEEE Software, 33*(1), 66–73.

Barth, A., Jackson, C., & Reis, C. (2008). The security architecture of the Chromium browser. Stanford Web Security Research.

Bellomo, S., Nord, R. Ozkaya, I., & Popeck, M. (2016). Got technical debt? Surfacing elusive technical debt in issue trackers. In *Proceedings of the 13th International Conference on Mining Software Repositories* (pp. 327–338).

Camilo, F., Meneely, A., & Nagappan, M. (2015). Do bugs foreshadow vulnerabilities? A study of the Chromium Project. In *Proceedings of the 12th Working Conference on Mining Software Repositories* (pp. 269–279).

Chromium issues. (2017). Retrieved from https://code.google.com/p/chromium/issues/list

Common weakness enumeration. (2017). Retrieved from https://cwe.mitre.org/about/sources.html

Curtis, B., Sappidi, J., & Szynkarski, A. (2012). Estimating the principal of an application's technical debt. *IEEE Software, 29*(6).

Ernst, N., Bellomo, S., Ozkaya, I., Nord, R., & Gorton, I. (2015). Measure it? Manage it? Ignore it? Software practitioners and technical debt. In *Proceedings of the 10th Joint Meeting on Foundations of Software Engineering* (pp. 50–60). New York, NY: ACM.

Falessi, D., & Reichel, A. (2015). Towards an open-source tool for measuring and visualizing the interest of technical debt. In *Proceedings of the Seventh International Workshop on Managing Technical Deb*t.

Felten, E. (2014). Technical debt in the NSA's phone call data program? Retrieved from https://freedom-to-tinker.com/2014/02/10/technical-debt-in-the-nsas-phone-call-data-program/

Kazman, R., Cai, Y., Mo, R., Feng, Q., Xiao, L., Haziyev, S., et al. (2015). A case study in locating the architectural roots of technical debt. In *Proceedings of the 37th IEEE International Conference on Software Engineering* (pp. 179–188).

Kruchten, P., Nord, R., & Ozkaya, I. (2012). Technical debt: From metaphor to theory and practice. *IEEE Software, 29*(6), 18–21.

Li, Z., Liang, P., Avgeriou, P., Guelfi, N., & Ampatzoglou, A. (2014). An empirical investigation of modularity metrics for indicating architectural technical debt. In *Proceedings of the 10th International ACM SIGSOFT Conference on Quality of Software Architectures* (pp. 119–128).

Nord, R., Ozkaya, I., Schwartz, E., Schull, F., & Kazman, R. (2016). Can knowledge of technical debt help identify software vulnerabilities? CSET @ USENIX Security Symposium.

Spínola, R. O., Zazworka, N., Vetro, A., Seaman, C., & Shull, F. (2012). Investigating technical debt folklore: Shedding some light on technical debt opinion. MTD@ICSE 2012, 1–7.

Tufekci, Z. (2015). Why the Great Glitch of July 8th should scare you. *The Message*. Retrieved from https://medium.com/message/why-the-great-glitch-of-july-8th-should-scare-you-b791002fff03

Xiao, L., Cai, Y., & Kazman, R. (2014). Design rule spaces: A new form of architecture insight. In *Proceedings of the 36rd International Conference on Software Engineering* (pp. 967–977).

Zazvorka, N., Spínola, R., Vetro, A., Shull, F., & Seaman, C. (2013). A case study on effectively identifying technical debt. In *Proceedings of the 17th International Conference on Evaluation and Assessment in Software Engineering* (pp. 42–47). Porto de Galinhas, Brazil.

Disclaimer & Distribution Statement

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

Applying the Fundamentals of Quality to Software Acquisition

Steve Bygren—The MITRE Corporation

Greg Carrier—The MITRE Corporation

Tom Maher—The MITRE Corporation

Patrick Maurer—The MITRE Corporation

David Smiley—The MITRE Corporation

Rick Spiewak—The MITRE Corporation

Christine Sweed—The MITRE Corporation

Abstract

Historically, software developed under government contracts often does not stand up under real-world use, and defects frequently result in cost and schedule overruns. While proposed development activities from contractors commonly list measures to improve quality, these descriptions cannot be used to select a winning bidder if they are not part of the evaluation criteria. By making software quality requirements explicit at the proposal stage, contractor selection can be influenced by criteria based on best practices in software development.

If we want to improve the quality of our software, a “Quality in Depth” approach is needed—introducing quality-related measures at every stage of software acquisition. In a previous article, one of the authors provided recommendations for improving software quality at the construction phase (Spiewak & McRitchie, 2008). This article discusses how to apply these same principles to the source selection process.

In order to find a way to include software practices as selection criteria, the authors set out to identify and recommend changes to Sections L and M of a government RFP (Request for Proposal) or IFPP (Instructions for Proposal Preparation) and EC (Evaluation Criteria) in an attempt to improve software and system quality. These changes will enable selection teams to identify contractors whose software development processes and compliance with software quality standards are more likely to produce the desired results.

Background

What Is Software Quality?

Quality is often thought of as an absence of defects. With many software products however, “defect” does not adequately describe the range of phenomena that affects software quality as perceived by the customers, end users, and other stakeholders. Using Crosby’s philosophy, we define the term “software quality” to mean conformance to the requirements of the software product’s users and other stakeholders (Crosby, 1979). The more closely a software product conforms to these requirements, the higher its quality.

We are particularly interested in software quality as it affects the acquisition process for defense-related software. While end user requirements are of prime importance, poor software development and quality monitoring practices in early- and mid-stage acquisition can result in failure to provide the desired results. These failures range from unwanted or missing features to cost and schedule overruns to critical flaws in system security or reliability.

How Do You Measure Software Quality?

Software quality as an outcome is best measured by the number of defects encountered after development is complete as the numerator, divided by the “size” of the software as the denominator. One could also argue that if two different products were to be compared, some sort of “difficulty factor” could be applied, as well as references to the software language used or development environment employed (e.g., assembly code versus high order languages, or object-oriented versus functional languages, etc.).

Metrics exist which can be used to estimate the potential defects in code. These are based on the use of function points as the measure of “size.” Function points can also be (loosely) correlated with the commonly used measurement SLOC, or Source Lines of Code.

Approach

This article is the outcome of a study the authors conducted at MITRE. Our approach was to gather information from subject matter experts (SMEs), contracting officers, and acquisition experts for recommendations for additions to proposal documents. Part of this study was conducted through interviews and SME email group lists. Reference materials from the Air Force and Navy were found which provided recommendations from prior work (USAF, 2008; Office of the Assistant Secretary of the Navy [ASN(RD&A)], 2008). We then adapted the suggestions to Sections L and M to more thoroughly describe software quality related criteria for source selection. Some of these criteria are aimed at the technical evaluation team, while some can be used by cost evaluators and past performance evaluators, as well as the technical team.

Recommendations for Section L (Instructions for Proposal Preparation)

1. The offeror’s proposal shall include a proposed Software Development Plan (SDP) which describes their approach to software development, including the tools, techniques, and standards to be used for development; unit testing and component testing; integration tools and techniques (including configuration management) used to ensure the integrity of system builds; the number and type of reviews that are part of the development process; and the methods and tools used to manage defect reports and analysis, including root cause analysis as necessary. The proposed SDP will form the basis for a completed SDP to be available after contract award as a CDRL (Contract Deliverable Requirements List) item, subject to government review and approval.
2. The offeror shall describe their plan for effective code reuse in order to minimize the amount of new code to be developed. Reused code can come from any origin, including previous efforts by the offeror or as provided by the government in the bidders’ library.
3. The offeror shall provide a Basis of Estimate (BOE) describing the rationale for the proposed staffing. The detail of the BOE shall include labor hours for each labor category (e.g., system engineering staff versus software engineering staff) for the identified tasks in the Work Breakdown Structure (WBS) as it relates to the Statement of Work (SOW).
4. The offeror shall describe the process for orientation and training for all project employees (e.g., certification and training in software best practices including Information Assurance [IA] and risk management).
5. The offeror shall describe related systems experience, including a description of previous experience developing software of the same nature, and a

description of the extent to which personnel who contributed to these previous efforts will be supporting this effort.

6. The offeror shall describe proposed development practices. For example, if spiral / incremental development, they shall describe the number, duration, and scope of spirals, as well as how the use of your approach would result in improved product quality and user satisfaction over time.[1]
7. The offeror shall provide an Integrated Master Schedule (IMS) and accompanying narrative that describes all significant program activities that are aligned with the proposed program staffing profile. Include a timeline for completion of each activity identified in the proposed program. Provide details that clearly describe the purpose for and importance of key activities. Identify all critical path elements and key dependencies.

Recommendations for Section M (Evaluation Criteria)

The proposed SDP shall show a complete and comprehensive software development process, which incorporates best practices as well as standards such as IEEE 12207-2008. The contractor will be evaluated based on how their processes, as described in the SDP, incorporate the use of software best practices.

Evaluation criteria related to the SDP include the following:

- The number and type of peer reviews
- The use of automated unit testing including test coverage requirements
- The use of automated syntax analysis tools and adherence to the rules incorporated by them (Jones, 2010)
- The comprehensiveness of integration and test methods, including continuous integration tools if used
- The use of readiness requirements such as unit test and syntax analysis for code check-in
- Configuration management and source code control tools and techniques
- The extent to which root cause analysis of defects is part of the development process
- The selection of software source code to be reused, replaced, or rewritten from previous implementations or other origins, including a description of how it will be ensured that reused code meets or is brought up to the same standards as newly developed code. Risks associated with reused software shall also be discussed. Such software shall include government rights to the source code.

The IMS and accompanying narrative will be evaluated for level of detail and relevance of significant program activities, degree of alignment, the proposed program staffing profile, and integration of the proposed SDP into the IMS. Additionally, critical path

[1] Note that while not part of the technical evaluation, the government evaluation team will examine Contractor Performance Assessment Reports [CPARs] for relevant performance by the respondent on other contracts.

elements and key dependencies will be assessed for relevance, completeness, and the manner and level of risk containment.

Table 1 delineates a sample rating scale for SDP evaluation criteria.

Table 1. Sample Rating Scale for SDP Evaluation Criteria

Parameter/rating	Unacceptable	Marginal	Acceptable	Superior
The number and type of peer reviews	none	1 (any)	2 (design, code)	3 or more (requirements, design, code, test)
The use of automated unit testing including test coverage requirements	none	unit tests written after manual testing or only on selected code	automated tests 75% code coverage on new or modified code	automated tests 85% or more code coverage on all delivered code. The use of Test Driven Development.
The use of automated syntax analysis tools and adherence to the rules incorporated by them	none	used selectively or with heavily modified rules	used consistently with standard rules	additional rules or tools specific to security analysis
The comprehensiveness of integration and test methods including continuous integration tools if used	ad-hoc	formal integration and test	automated processes applied periodically	continuous integration including syntax analysis and unit tests
The use of readiness requirements such as unit test and syntax analysis for code check-in	none	individual manual testing	integrated testing by developer	automated part of check-in and continuous integration process
Configuration management and source code control tools and techniques	manual/paper trail	by individual developer	system-wide repository	managed tool with pre-check-in requirements
The extent to which root cause analysis of defects is part of the development process	none	"red-team" only	serious defects	routine periodic analysis of defect pool
The selection of software source code to be reused, replaced, or rewritten from previous implementations	none or no response	replacement with contractor's previous work	rework of selected items showing good knowledge of base software	innovative approach to maximum reuse and modernization

Note. The categories provided in Table 1 were suggested in a conversation with Jeff Pattee, Chief, Product Definition, Airspace Mission Planning Division, Electronic Systems Center, USAF.

Incorporating Software Quality Measures in Contracts

The contract development process includes several steps at which information can be gathered and requirements set to include software quality as a measure of vendor performance.

Sections L & M or equivalent from the RFP

- Add software quality measures as a discriminating factor in selecting the contractor

- Enumerate expectations in this area:
  - Types of methods used
  - Evidence to be provided

TRD (Technical Requirements Document), SOO (Statement of Objectives), and SOW (Statement of Work)

Add requirements in the form of deliverable items—as CDRLs or DAL (Data Accession List) items as appropriate. Examples include the following:

- Output of automated unit tests showing code coverage at or above required minimum
- Output of automated syntax analysis showing conformance to pre-determined rules
- Evidence of accomplishing required peer reviews
- Itemized list of tools with version numbers used to produce output from each source module
- Programmer's reference manual with examples
- Interface definitions
- List of all software components with the following information:
  - Purpose and function
  - Interfaces provided
  - Language/version for each module
  - Complete source code
- Source from architectural design tool where available
- Use cases (text and diagrams)
- Class diagrams where applicable
- Complete list of any third-party components with version numbers
- Contact information for any outside dependencies
- Build procedures, including documentation for building all software components from source code
- Test procedures—including any automated unit tests with source code, test scripts

Rationale for Incorporating Recommended RFP Language

The recommended RFP language was derived by the authors from a variety of sources including MITRE acquisition subject matter experts, existing guidance documents from the Navy and Air Force, and also from the authors' experience. We've tried to provide a succinct rationale as to why the language asks for specific information from the contractor in the RFP:

- The Software Development Plan (SDP) is a maturity indicator of the bidder's development process. By evaluating this, and then putting its provisions under contract, it becomes possible to select a contractor on the basis of development methodology and then obligate them to perform as proposed.
  - Automated unit tests & comprehensive peer reviews are widely used best practices. Capers Jones (2008a) has noted that these are among the required steps to achieve effective defect removal.

- Continuous Integration (CI) often includes the automated invocation of tests and code analysis during the build process. CI and static analysis expose problems sooner in the development process. The sooner problems are discovered, the lower the cost to resolve them.
- Root cause analysis prevents the introduction of defects and is a recognized best practice in all approaches to process improvement. It is a Capability Maturity Model Integration (CMMI) Level 5 practice area. Prevention is more cost effective than detecting and fixing defects after they are introduced.

- The Basis of Estimate (BoE) helps the evaluator understand the bidder's cost to compare against industry averages and government cost models. By examining proposed labor categories, this can be checked against predicted labor distributions from government cost models as well.
- The Integrated Master Schedule (IMS) can be checked for alignment with required milestone dates, and it supports an independent estimate.

Guidance for Evaluation Team Experience

The government's evaluation team must have relevant software engineering experience. The experience should cover the full life cycle of software development from design to development, integration, testing, and delivery. If the proposal is seeking a particular style of development methodology (e.g., waterfall, spiral/incremental, agile), then the evaluation team should have experience in that methodology in order to evaluate the RFP response.

Since a significant portion of the suggested contract language relates to software quality monitoring, the evaluators should be familiar with unit testing, peer reviews, continuous integration (CI), static code analysis, and metrics. Finally, evaluators should have some knowledge of various practices and approaches of applying these techniques, for example, when it comes to test-driven development.

The field of software engineering is diverse. It is insufficient to simply have general software engineering experience on the evaluation team without further having experience in the applicable domain(s). Examples of these domains include real-time/embedded, kernel/operating systems, numerical/digital signal processing, web applications, SOA, information retrieval/search, security, and human-computer interface.

Finally, the evaluation team should have an understanding of the CMMI process and rating criteria.

Guidance for Evaluating Technical Responses

The recommended contract language in this article includes Section M of the RFP, also appearing as Evaluation Criteria. The language is not very specific so as to elicit responses that are more original than simply claiming the ability to do a long list of things that the government requires. In this section, we discuss more specific guidance for the evaluation team in evaluating the responses.

In advance, the team should define objectives that are sought after and then define measurable criteria. The more objective the criteria, the better, though it is recognized that coming up with this criteria can be a challenge. After defining criteria, they are prioritized and then weighted in a scheme the team deems appropriate.

The following are some general evaluation tips:

- If key staff are identified in the proposal, how likely are they to be available during contract execution?
- In reference to quality assurance processes, does the proposal language favor or at least mention "empowerment" of the QA team over engineering processes?
- Regarding the contractor's approach to Automated Unit Testing, does the contractor require that unit tests be passed and cover a reasonable percentage of code before code can be checked in? Does the contractor use Test Driven Development?
- Regarding the contractor's approach to automated syntax analysis, does the contractor require that syntax analysis be performed and that all required rules are followed before code can be checked in?
- Regarding development build and integration, does the contractor use an automated build process which incorporates syntax analysis and automated unit testing?

You can expect that the response is going to claim appraisal at a specific CMMI maturity level (commonly at least Level 3). This can be verified with the Appraisal Disclosure Statement (ADS) document. Another source is the Standard CMMI Appraisal Method for Process Improvement (SCAMPI). For the larger contractors, particularly when work is further sub-contracted out, look for further CMMI level compliance information on the specific division/unit and sub-contractor(s) as applicable.

Development Process

If the proposal declares that a development process will be used that will involve multiple iterations/spirals/increments (which is standard practice), then the evaluation team should look for further details on the process including the following:

- What is the duration and scope of each increment?
- Are lessons and obstacles from one increment reviewed for improvement to a subsequent increment?
- Is user (customer) feedback interaction only up front or do most increments incorporate this? And how is that feedback prioritized?
- Are multiple increments planned in sufficient detail, or are only the present and possibly next increment planned?

Software Engineering

One key thing to look for in a proposal is the degree to which the contractor has experience in the technology the RFP calls for them to deliver. The more complex the system, the more important applicable contractor experience is.

Many DoD systems have a degree of interoperability and integration required of them. For integration with particular systems, verify if the contractor has experience with that system or has relationships with third parties with integration capabilities that will be used. The contractor should also participate in applicable Communities of Interest (COIs).

Testing processes and technologies that support them are important. Look for information on a test plan or strategy. If the proposal is serious about continuous integration and use of supporting tools, then listing the software to be used for this is a promising sign.

Information on how the tools are used (e.g., by exception and/or monitored on a periodic basis—and what period) is also telling. If the proposal includes information on the proposed system design, then the evaluators could look to see how "testable" the design is, particularly as it is incrementally built.

Conclusions

While it is important to implement quality measures in software construction, this is undertaken after a contractor has been selected. The authors recommend an in-depth approach, beginning with the process of selecting the contractor. It can be easy to overlook the importance of including specific language in the proposal documents in order to be able to select the right contractor from those responding to a Request for Proposal. In order to accomplish this goal, it is critical to specify the instructions in Section L (or the IFPP) and the evaluation criteria in Section M (or the EC) so that these can be used to assign strengths or weaknesses appropriately. This is an early, but often neglected, piece of the puzzle involved in building quality software products for defense applications.

Reference List

Crosby, P. B. (1979). *Quality is free: The art of making quality certain*. New York, NY: McGraw-Hill.

DoD. (2010, August 5). *Defense acquisition guidebook*. Retrieved from http://at.dod.mil/docs/DefenseAcquisitionGuidebook.pdf; current version available at https://dag.dau.mil/Pages/Default.aspx

Jones, C. (2008a, June). Measuring defect potentials and defect removal efficiency. CrossTalk. Retrieved from http://www.crosstalkonline.org/storage/issue-archives/2008/200806/200806-Jones.pdf

Jones, C. (2008b, August). *Software quality in 2010: A survey of the state of the art.* Retrieved from http://www.spr.com

Jones, C. (2010). *Software engineering best practices*. New York, NY: McGraw-Hill.

Office of the Assistant Secretary of the Navy (ASN[RD&A]). (2008, September). *Guidebook for acquisition of Naval software intensive systems*. Retrieved from https://acquisition.navy.mil/rda/content/download/5657/25845/version/1/file/Guidebook+for+Acquisition+of+Naval+Software+Intensive+SystemsSEP08.pdf

Spiewak, R., & McRitchie, K. (2008, December). Using software quality methods to reduce cost and prevent defects. CrossTalk. Retrieved from http://www.crosstalkonline.org/storage/issue-archives/2008/200812/200812-Spiewak.pdf

United States Air Force (USAF). (2008, August 15). *Weapon systems software management guidebook*. Retrieved from https://acc.dau.mil/adl/en-US/24374/file/49721/USAF%20WSSMG%20%20ABRIDGED.pdf

THIS PAGE INTENTIONALLY LEFT BLANK

Panel 4. Defense Acquisition After Shocks to the System

Wednesday, April 26, 2017	
11:15 a.m. – 12:45 p.m.	**Chair: Todd Harrison,** Director of the Defense Budget Analysis, Director, Aerospace Security Project, Center for Strategic and International Studies ***Post–Milestone B Funding Climate and Cost Growth in Major Defense Acquisition Programs*** David L. McNicol, Institute for Defense Analyses (CARD) ***Measuring the Impact of Sequestration and the Defense Drawdown on the Industrial Base*** Rhys McCormick, Center for Strategic and International Studies Andrew Hunter, Center for Strategic and International Studies Gregory Sanders, Center for Strategic and International Studies ***The Improving DOD/Silicon Valley Relationship*** James Cross, Franklin Templeton

Todd Harrison—is the Director of the Aerospace Security Project and the Director of Defense Budget Analysis at CSIS. As a senior fellow in the International Security Program, he leads the Center's efforts to provide in-depth, nonpartisan research and analysis of space security, air power, and defense funding issues. He has authored publications on trends in the overall defense budget, military space systems, civil space exploration, defense acquisitions, military compensation, military readiness, nuclear forces, and the cost of overseas military operations.

Harrison frequently contributes to print and broadcast media and has appeared on CNN, CNBC, NPR, Al Jazeera English, C-SPAN, PBS, and Fox News. He teaches classes on military space systems and the defense budget at the Johns Hopkins School of Advanced International Studies and a class on the defense budget at George Washington University's Elliott School of International Affairs. He is a term member of the Council on Foreign Relations, a member of the National Oceanic and Atmospheric Administration's Advisory Committee on Commercial Remote Sensing, and a member of the Defense News Advisory Board.

Harrison joined CSIS from the Center for Strategic and Budgetary Assessments, where he was a senior fellow for defense budget studies. He previously worked at Booz Allen Hamilton where he consulted for the Air Force on satellite communications systems and supported a variety of other clients evaluating the performance of acquisition programs. Prior to Booz Allen, he worked for a small startup (AeroAstro Inc.) developing advanced space technologies and as a management consultant at Diamond Cluster International. Harrison served as a captain in the U.S. Air Force Reserves. He is a graduate of the Massachusetts Institute of Technology with both a BS and an MS in aeronautics and astronautics

Post–Milestone B Funding Climate and Cost Growth in Major Defense Acquisition Programs

David L. McNicol—joined the Department of Defense (DoD) in 1982. From 1988 until 2002, he was a Deputy Director of Program Analysis and Evaluation (PA&E). Earlier, Dr. McNicol taught at the University of Pennsylvania and the California Institute of Technology. He holds a BA in Economics from Harvard and an MS (Management) and PhD (Economics and Finance) from MIT. Employed at the Institute for Defense Analyses (IDA) since his retirement from the DoD, he became director of the Cost Analysis and Research Division in 2006. Still at IDA, Dr. McNicol stepped down in 2012 to return to his previous role as a Research Staff Member. [dmcnicol@ida.org]

Abstract

This paper is the fourth in a series that examines the association between outcomes of Major Defense Acquisition Programs (MDAPs) and changes in acquisition policy and process and funding climate. Like an earlier paper in the series, it finds that quantity normalized Program Acquisition Unit Cost (PAUC) growth measured from Milestone (MS) B is significantly higher in programs that passed MS B in bust climates than in boom climates. The new finding in this paper is that among MDAPs that passed MS B in a bust phase, only those that continued into a boom climate showed significantly higher PAUC growth than programs that passed MS B in a boom climate. This conclusion is important because it implies that much of the observed PAUC growth may have causes other than flaws in MS B baselines. The conclusion tells us less than might be hoped, however. This is so because the PAUC growth associated with the boom climate may reflect the purchase of capability beyond that specified in the MS B baseline or, alternatively, may reflect PAUC increases that occur when programs take advantage of a boom climate to “get well.”

Introduction

This paper examines whether Major Defense Acquisition Programs (MDAPs) that entered a boom climate for procurement funding some time after passing Milestone (MS) B on average had higher unit cost growth than programs whose acquisition cycles did not extend into a boom climate. While this conjecture seems plausible, possibly even obviously correct, it has not been recognized in the cost growth literature.

The topic is worth pursuing because it bears on why unit cost growth was significantly higher for MDAPs that passed MS B in bust periods than it was for those that passed in boom periods. This observation was reported by the first paper in this series, McNicol and Wu (2014; hereafter referred to as P-5126). The explanation offered there was a version of the “camel’s nose” hypothesis—that unrealistic cost, programmatic, and technological assumptions are made in the hope that, by making the program appear to be lower in cost or more capable, they will increase the odds that the program will be successful in competing for funds. P-5126 goes further by suggesting that the incentives for adopting very optimistic assumptions are stronger for programs that pass MS B in bust funding climates than they are for programs that pass in boom periods, and that, consequently, unrealistic MS B baselines are more common in programs that passed MS B in bust climates.

For present purposes, the key point to note is that the explanation offered by P-5126 supposes that most of the growth in unit cost shown by programs that pass MS B in a bust funding climate is “baked into” the baselines established at MS B. Other possibilities exist, however, one of which is that a significant part of these programs’ cost growth might be due to increases in program content made during a post-MS B boom climate, when funding is more readily available. To the extent that is the case, we may be mistaking the costs of decisions to improve the capabilities of an existing system for growth in the costs of acquiring the capabilities specified in the MS B baseline.

Framework

The topic of this paper requires distinguishing between bust funding and boom funding climates. The period Fiscal Year (FY) 1965–FY 2009 considered here spans two bust-boom cycles in Department of Defense (DoD) procurement funding: (1) The bust climate for modernization of weapon systems that began in the mid-1960s (as discussed in Appendix A of McNicol, Tate, Burns, & Wu [2016], hereafter referred to as P-5330 [Revised])[1] and lasted until the Carter–Reagan buildup of the early to mid-1980s, and (2) the long post–Cold War bust climate followed by the post-9/11 boom. The rationales for the break points between the funding climates are provided in P-5330 (Revised).

A measure of cost growth also is required. One option is based on Program Acquisition Unit Cost (PAUC). PAUC is the sum of Research, Development, Test, and Evaluation (RDT&E) cost and procurement cost, divided by the number of units acquired. For this paper, PAUC growth is computed by comparing the MS B baseline value of PAUC—which can be thought of as a goal or a prediction—to the actual PAUC reported in the last Selected Acquisition Report (SAR) for the program, normalized to the MS B quantity. Both the MS B baseline and the actual value of PAUC are stated in constant dollars. The alternative to PAUC growth is growth in Average Procurement Unit Cost (APUC), which does not include RDT&E cost.[2] The effects of changes in the capabilities procured may be more likely to show up clearly in APUC growth, which is an advantage, but it is a less comprehensive measure of unit cost growth. We compute the results for both cost growth measures, and report the results for APUC only in the one instance in which they differ in an important way from those obtained using PAUC. Note that PAUC growth and APUC growth are adjusted for quantity but not for changes in the capabilities the program is directed to acquire.

In what follows, the term *PAUC growth* means PAUC growth from the MS B baseline, with the final SAR PAUC normalized to the MS B quantity. Similarly, the term *APUC growth* means APUC growth from the MS B baseline, with the final SAR APUC normalized to the MS B quantity. Appendix B of P-5330 (Revised) provides the conventions used in assembling the database, the sources of the data used, and the quantity normalization computations. The unit cost growth estimates were updated to the most recent comprehensive information available, that in the December 2015 SARs. Only completed programs (defined as programs with an end date of FY 2016 or earlier) are used in this analysis because some costs associated with a program may not be fully reflected in its SAR until the program is completed.

[1] The DoD budget was high during the years of the Vietnam War, but much of the acquisition budget went for munitions and to weapon systems lost in combat. Consequently, funding for major system new starts was relatively constrained.

[2] PAUC and APUC growth measures used for purposes of Nunn-McCurdy Act reporting are not quantity normalized. The median MDAP that passed MS B in the period FY 1988–FY 2007 acquired 100% of MS B baseline quantity, and the average program acquired 111%. Compared to the PAUC growth measures used in Nunn-McCurdy reporting, quantity adjustment decreased measured PAUC growth for about half of the programs in the sample and increased it for the other half.

Average PAUC growth reported in Table 1 for programs that passed MS B in bust climates is significantly higher (43%) than it is for programs that passed MS B in boom periods (15%).[3] This observation serves only to confirm, for the data used in this research, the result mentioned above from P-5126.

Table 1. Average PAUC Growth for Completed MDAPs by MS B Funding Climate

Bust		Boom	
FY 1964–FY 1980	47% (64)	FY 1981–FY 1986	18% (35)
FY 1987–FY 2002	37% (44)	FY 2003–FY 2009	0% (9)
Total	**43% (108)**	**Total**	**15% (44)**

Note. Numbers of MDAPs that passed MS B and were completed by the December 2015 SARs are shown in parentheses.

Finally, it is necessary to recognize changes over time in acquisition policy and process configurations because they are associated with significant difference in average PAUC growth. P-5330 (Revised) distinguished the following six policy and process configurations:

1. McNamara-Clifford (FY 1964–FY 1969)
2. Defense Systems Acquisition Review Council (DSARC, FY 1970–FY 1982)
3. Post-Carlucci DSARC (P-C DSARC, FY 1983–FY 1989)
4. Defense Acquisition Board (DAB, FY 1990–FY 1993)
5. Acquisition Reform (AR, FY 1994–FY 2000)
6. DAB Post AR (DAB Post AR, FY 2001–FY 2009)

Average PAUC growth does not differ significantly among DSARC, P-C DSARC, DAB, and DAB Post AR within a budget climate.[4] Their statistical similarity permits these periods to be combined into a single acquisition policy and process configuration, which will be referred to as DSARC/DAB. The main text is concerned only with the DSARC/DAB. P-5330 (Revised) found that average APUC growth was significantly higher in the McNamara-Clifford and AR configurations, which for that reason are treated separately.[5] Appendix A presents results for the McNamara-Clifford and AR configurations.

---

[3] P < 0.001 for the Mann-Whitney U (M-W U) test (U = 1261.5, n_1 = 108, n_2 = 44)

[4] For bust climates, Analysis of Variance (ANOVA) fails to reject the null hypothesis that APUC growth for completed programs in each of these bins has the same normal distribution (P = 0.996). Kolmogorov-Smirnov (K-S), Anderson Darling (A-D), and an F-test of the variances indicate that the assumptions of ANOVA are satisfied. For boom climates, K-S and A-D find that the observations for the boom portion of DSARC and the DAB periods are consistent with a normal distribution, but K-S rejects normality for the boom portion of P-C DSARC. The M-W U test does not detect a significant difference between the means of the (1) DSARC-Boom and P-C DSARC-Boom (P = 0.968, U = 88.5, n_1 = 29, n_2 = 6); (2) DSARC-Boom and DAB Post AR-Boom (P = 0.317, U = 36, n_1 = 9, n_2 = 6); or (3) the P-C DSARC-Boom and the DAB Post AR-Boom (P = 0.215, U = 94, n_1 = 29, n_2 = 9).

[5] Appendix C of P-5330 (Revised) provides a Bayesian analysis using APUC growth data. That result also probably holds for the PAUC data

Appendix A of P-5330 (Revised) provides brief descriptions of the acquisition configurations as defined here. Readers who are not generally familiar with the Office of the Secretary of Defense (OSD)–level acquisition process and various acquisition reform efforts may wish to consult that source or Fox (2011).

Evidence of a Boom Effect

The term *boom effect* is used here to label a feature observed in the unit cost growth data—MDAPs that passed through a boom climate post MS B had a higher average unit cost growth than those that did not.

Many MDAPs that passed MS B in one of the bust climates continued into a boom climate, and some programs that passed MS B during the Carter–Reagan defense buildup continued into the post-9/11 boom. A two-part naming convention is used to label two bins of programs: those that did—and those that did not—pass through a boom climate post MS B. The first part of the label gives the funding climate prevailing when the program passed MS B—bust or boom. The second part—0, 1, or 2—denotes the number of boom climates a program passed through post MS B. For example, programs that were completed entirely within a single bust phase will be referred to as Bust0—Bust because they passed MS B in a bust funding climate and zero because they were completed without entering a boom climate. Programs that passed MS B in a bust period and continued into or through a subsequent boom period are called Bust1.

A detailed evaluation of content changes for programs that did and did not experience a boom funding climate after passing MS B would be the best approach to exploring the importance and character of boom. This type of analysis would require greater resources than were available, however. Instead, this paper uses a statistical approach that relies on data that are comparatively easy to acquire—PAUC growth from the MS B baseline, the year programs passed MS B, and the year the programs were completed. In the language of medical testing, the plan is to compare unit cost growth for a treatment group—programs that experienced a boom climate post MS B—with that of a control group—programs that did not. The question asked in this section is whether the observed boom effects are statistically significant. We look first at the two bust climates and then at the two boom climates.

PAUC growth for Bust0 and Bust1 is presented in Table 2 for each of the two bust periods of DSARC/DAB. In both periods, average PAUC growth for the treatment group (Bust1) is higher than it is for the control group (Bust0)—42% compared to 16% for the first period, and 51% compared to 13% for the second. These differences are statistically significant.[6] For programs that passed MS B in a bust period, subsequent entry into a boom period is then associated with higher PAUC growth.

---

[6] K-S and A-D find the PAUC growth data in each of the two bins of the first bust period to be consistent with a normal distribution. An F-test found the two variances to be significantly different. A two-tailed t-test assuming unequal sample variances found the means of Bust1 and Bust0 for the first period to be significantly different (P = 0.011). K-S and A-D also find the PAUC growth data in each of the two bins of the first bust period to be consistent with a normal distribution. Again, an F-test found the two variances to be significantly different. A two-tailed t-test assuming unequal sample variances found the means of Bust1 and Bust0 for the first period to be significantly different (P = 0.004).

Table 2. Average PAUC Growth for Completed MDAPs in DSARC/DAB Bust by the Number of Boom Periods Experienced

Bin	1st Bust Period FY 1970–FY 1980	2nd Bust Period FY 1987–FY 1993
Bust0	16% (6)	13% (8)
Bust1	43% (39)	51% (17)
Bust2	19% (3)	none

Bust2 does not follow this pattern: Average PAUC growth for Bust2 is slightly higher than that of Bust0 but less than that of Bust1. The number of programs in this bin (N=3), however, is so small that there is no point in speculating about why it does not fit the pattern.[7] While no attempt is made to explain the observation for Bust2, it is included in an analysis discussed below that includes all of the MDAPs that passed MS B during the two DSARC/DAB bust periods.

APUC growth also does not entirely follow the pattern of PAUC growth for Bust0 and Bust1 of the two DSARC/DAB bust periods. In particular, APUC growth for programs initiated in the first DSARC/DAB bust period does not show a statistically significant boom effect in APUC growth. (The six programs of Bust0 have an average APUC growth of 21%, which is not significantly different from the 42% average APUC growth for the 39 programs of Bust1.[8]) APUC growth in the second bust period does follow the pattern—43% for Bust1, which is significantly higher than the 17% average APUC growth for Bust0.[9]

Table 3 presents data on PAUC growth for the two DSARC/DAB boom periods. The nomenclature used for the boom periods parallels that used for bust periods. Boom0 programs passed MS B in a boom climate and were completed in that boom or the succeeding bust climate. Boom1 programs passed MS B during the Carter–Reagan defense buildup and were completed during the post-9/11 boom or during the following three years. There is no treatment group (i.e., Boom1) for the second boom period and hence no experiment to examine.

Table 3. Average PAUC Growth for Completed MDAPs in DSARC/DAB-Boom by the Number of Boom Periods Experienced

Bin	1st Boom Period MS B FY 1981–FY 1986	2nd Boom Period MS B FY 2003–FY 2009
Boom0	12% (28)	0% (9)
Boom1	45% (7)	none

[7] The programs in Bust2 are the CNV 68, with a PAUC growth of 7%; the NAVSTAR GPS (85%); and ATCCS-MCS (-34%).

[8] K-S and A-D find the APUC growth data in each of the two bins of the first bust period to be consistent with a normal distribution. An F-test found no significant difference between the two variances. A two-tailed t-test of the APUC data found the means of Bust1 and Bust0 for the first period not to be significantly different (P = 0.241).

[9] M-W U P = 0.041 (U = 32.5, $n_1 = 17$, $n_2 = 8$).

Average PAUC growth for the Boom1 programs of the first boom period (45%) is significantly higher than that for the Boom0 programs (12%).[10] This finding is somewhat unexpected, since the relevant programs passed MS B in a boom funding climate and presumably had realistic baselines and were robustly funded at least initially. In fact, the finding may be spurious. Average PAUC growth for the Boom1 bin of the first boom period is dominated by three MDAPs, each of which had PAUC growth of more than 40%: C-17 (57%), T-45 Goshawk (70%), and JSTARS (123%). These programs had the essential features of Total Package Procurement (TPP; McNicol, 2004). Acquisition reforms adopted in mid-1969 ruled out use of TPP and fixed-price development contracts because they typically resulted in severe cost growth and schedule problems (McNicol, 2004; McNicol et al., 2016; Tyson et al., 1992; O'Neil & Porter, 2011). During the Reagan Administration, however, TPP-like contracts were used for a few MDAPs, including the three programs noted here. (The other four of the seven programs in Boom1 had conventional cost plus incentive fee contracts for Engineering and Manufacturing Development [EMD].) The PAUC growth of the C-17, T-45, and JSTARS programs was on a par with that of TPP programs that passed MS B during FY 1965–FY 1969 and did not continue into the Carter–Reagan boom. Their contracting strategy, not their continuation into a boom funding climate, could then account for their high PAUC growth. If the three programs are excluded, the average PAUC growth for Boom1 is 17%, which is not significantly higher than the average for Boom0.[11]

Table 4 combines data from Table 2 and Table 3. The 73 MDAPs of the DSARC/DAB bust climates had an average PAUC growth of 38%, which was significantly higher than the 9% average of the 41 MDAPs in DSARC/DAB that passed during boom climates.[12] Average PAUC growth of MDAPs in Bust0 is not significantly different from the average PAUC growth of DSARC/DAB boom, and therefore has little effect on this result.[13] Instead, the higher average of DSARC/DAB bust is mainly due to the programs in Bust1. This adds an important point to the narrative of P-5126: The higher PAUC growth of MDAPs that passed MS B in bust climates largely reflects a subset associated with those programs—those that passed MS B in a bust climate and continued on into a boom climate.

Table 4. PAUC Growth for the Combined Bust and the Combined Boom Phases of DSARC/DAB

Bust0	14% (14)	Boom0	9% (37)
Bust1	45% (56)	Boom1	17% (4)*
Combined Bust**	**38% (73)**	**Combined Boom**	**9% (41)**

* Excludes C-17, T-45, and JSTARS.
** Includes the three programs in Bust2, which have an average PAUC growth of 19%.

[10] K-S found the distribution of APUC growth of the 28 Boom0 programs that passed MS B in the first bust phase to be non-normal. M-W U found the difference between average APUC growth of Boom0 and Boom1 for the first boom phase to be significant ($P = 0.007$, $U = 164.5$, $n_1 = 28$, $n_2 = 7$).
[11] M-W U $P = 0.117$ ($UA = 83.5$, $UB = 28.5$, $n_1 = 28$, $n_2 = 4$).
[12] M-W U $P < 0.001$ ($U = 633$, $n_1 = 73$, $n_2 = 4$).
[13] M-W U $P = 0.121$ ($U = 367.5$, $n_1 = 41$, $n_2 = 14$).

Funding Climate, Program Duration, and the Boom Effect

This section takes an additional step towards explaining why the data show boom effects. Table 5 presents rearranged data from Table 2 and Table 3 and, in addition, shows average program duration for each bin. Average PAUC growth is greater in Bust1 than in Bust0 for each of the two bust periods and greater for Boom1 than for Boom0 for the first boom period. (The second boom period is excluded because there are no programs in Boom1.) The programs in Bust1 and Boom1, however, also had a longer average duration than the programs in the corresponding Bust0 and Boom0 bins. Consequently, we need to examine the extent to which longer average duration in addition to an encounter with a boom period account for their higher PAUC growth. Note that including the three programs of Bust2 (of the first bust period) and the three programs excluded from Boom1 would not change this conclusion.

Table 5. Average PAUC Growth and Average Program Duration by Number of Boom Periods Encountered for Completed Programs in DSARC/DAB

Bin		Average PAUC Growth	Average Duration †
1st Bust Period ‡	Bust0	16% (6)	6.7
	Bust1	42% (39)	14.2
2nd Bust Period	Bust0	13% (8)	7.5
	Bust1	53% (17)	15.8
1st Boom Period	Boom0	12% (28)	9.2
	Boom1§	17% (4)	22.3

† From MS B through the year in which the program's last SAR was filed.
‡ Excludes the three programs of Bust2.
§ Excludes C-17, T-45, and JSTARS.

We approach this problem by dividing the duration of the program into two parts:

1. T_{boom} = number of years post MS B spent in boom climates
2. T_{bust} = number of years post MS B spent in bust climates

These two variables are hypothesized to have distinct linear relationships to PAUC growth (abbreviated as PAUC):

$$PAUC_i = a_0 + a_1T_{boomi} + a_2T_{busti} + e_i$$

In this equation, the subscript i denotes the ith MDAP in the sample and e_i is the error term, which is assumed to be a normally distributed random variable. The coefficient a_1 is the change in PAUC for each year the program spends in a boom climate. Similarly, a_2 is the change in PAUC per year in a bust climate. The estimated intercept term a_0 is the average net effect of excluded variables. The coefficients of the model are estimated (using multiple regression) separately for programs that passed MS B in bust periods of DSARC/DAB and those that passed MS B in its first boom climate. (The second boom climate is excluded because it has no programs.) The estimates obtained are presented in Table 6.[14]

Table 6. Years in Bust Climates and Years in Boom Climates and PAUC Growth for MDAPs in the DSARC/DAB Acquisition Policy and Process Configuration

	Passed MS B in Bust Period†		Passed MS B in Boom Period‡	
	Estimate	P-Value	Estimate	P-Value
Intercept	3.4%	0.719	3.7%	0.608
Years in Boom	5.0%/yr***	< 0.001	3.7%/yr**	0.039
Years in Bust	1.6%/yr**	0.042	0.05%/yr	0.937

** Statistically significant at less than the 5% level.
*** Statistically significant at less than the 1% level.
† R-Square = 0.22 F = 9.445 (P < 0.001) N= 70. Estimated by Ordinary Least Squares (OLS). Excludes the three MDAPs in the Bust2 bin of DSARC/DAB.
‡ R-Square = 0.20 F = 5.563 (P = 0.002) N= 32. Estimated by OLS. Excludes C-17, T-45, and JSTARS.

Programs that passed MS B in a bust climate characteristically experienced PAUC growth of 1.6% for each year spent in a bust climate. PAUC growth for each year spent in a boom climate post MS B was three times that level—about 5% per year. Each of these estimates is statistically significant.

The effect of boom years for programs that passed MS B in boom periods is smaller (about 3.7% per year). This is reasonable, as we expect programs that passed MS B in boom climates to have realistic baselines and to be adequately funded (at least initially). The estimated effect per bust year on PAUC growth for programs that passed MS B in boom periods is very small and statistically not significant, which also seems reasonable.

A sense of the importance of the boom periods entered into post MS B is provided by Table 7. The table shows the estimated relationship evaluated at the sample means for T_{Boom} and T_{Bust} for Bust0 and Bust1, respectively. Programs in Bust0 have an average PAUC growth of about 14%. Of this, about 11.4 percentage points are associated with years spent in bust climates, and, of course, none for continuation into a boom climate. For Bust1 programs, boom years post MS B account for about 26 percentage points of the Bust1 average PAUC growth of 45%; the years spent in bust climates account for 15.2 percentage points.

[14] An alternative to the model above posits two categories of MDAPs, one that tends to short duration and low unit cost growth and another that tends to long duration and higher unit cost growth. Modifications and upgrades would seem to be examples of the first category and major platforms an example of the second. The “short duration” and “long duration” programs were defined, respectively, as the 20% of programs in the bin with the shortest durations, and the 20% with the longest durations. The results for all of the forms of this model considered rejected the hypothesis that shorter vice longer duration is a statistically significant factor in PAUC growth.

Table 7. Amount of PAUC Growth in Boom Climates and Bust Climates for MDAPs in DSARC/DAB That Passed MS B in Bust Climates

Period	Intercept	T_{Boom}	T_{Bust}	Average PAUC Growth
Bust0	3.4%	0	11.4%	14.3% (14)
Bust1	3.4%	26%	15.2%	45% (56)

Note. Evaluated at the sample means for T_{Boom} and T_{Bust}

Conclusions and Limitations

This paper, like earlier papers in the series, finds that PAUC growth measured from MS B is significantly higher in programs that passed MS B in bust climates than in boom climates. Moreover, among MDAPs that passed MS B in a bust phase of DSARC/DAB, only those that continued into a boom climate showed PAUC growth significantly higher than that of programs that passed MS B in a boom climate. This conclusion is important because it implies that much of the observed PAUC growth may have causes other than flaws in MS B baselines. The conclusion tells us less than might be hoped, however. This is so because the PAUC growth associated with the boom may reflect the purchase of capability beyond that specified in the MS B baseline or, alternatively, PAUC increases that occur when programs take advantage of a boom climate to "get well."

The Global Broadcast System (GBS) provides an example of a program whose content was increased early in the post-9/11 boom:

> The current GBS architecture is based on Asynchronous Transfer Mode (ATM) technology. … In December 2002, DoD directed GBS's migration to a more sustainable commercial and standards-based open architecture, based upon the Internet Protocol (IP). Also, the GBS program received FY03 Iraqi Freedom Funds (IFF) supplemental funding for IP Acceleration of production units to replace deployed ATM units. Based upon extensive warfighter inputs, the accelerated IP production effort included design and development of a new, single case version of the Receive Suite (88XR) for the Army, Navy, and Marine Corps. (*Selected Acquisition Report*, 2003)

Space Based Infrared Satellite-High (SBIRS-High) is a convenient and useful contrast to GBS, even though it passed MS B in 1997 and hence is not included in DSARC/DAB. As of the December 2015 SARs, funding for the Baseline SBIRS-High program was expected to end in FY 2018. A large portion of the growth in SBIRS-High unit procurement cost for the baseline program—roughly one-third—occurred before FY 2003, while most of the other two-thirds occurred during FY 2003–FY 2009. This increase was not driven by increased capability, however, but by the unrealistic cost estimate in the MS B SBIRS-High baseline (Kim et al., 2015; Porter et al., 2009; Younossi et al., 2008).

In the GBS example, it seems clear that capabilities beyond those in the MS B baseline were added to the program. While unit cost did increase, that was a matter of paying more for more. For SBIRS-High, in contrast, it appears that the advent of a boom funding climate provided a program experiencing severe problems an opportunity to "get well." In effect, in such cases, what otherwise would have been capability shortfalls were converted into cost growth and, relative to MS B, the DoD eventually paid more for the MS B capability than had been anticipated. The boom effect includes both of these cases. So does accretion of PAUC growth during bust years.

The average PAUC growth of all DSARC/DAB bust programs is 38%. Without making a specific estimate, P-5126 suggested that most of this PAUC growth stemmed from flawed MS B baselines. In the language of the present paper, if all of the unit cost growth actually is a matter of "getting well," the PAUC growth due to flawed MS B baseline problems remains at 38%. It is less than 38% to the extent that PAUC growth of MDAPs in Bust1, in the years they spent in both bust climates and boom climates, is due to decisions to acquire capabilities beyond those of the MS B baselines. Parts of PAUC growth in years spent in both boom and bust climates post MS B very probably do reflect acquisition of capabilities beyond that of the MS B baseline. Unfortunately, we do not have a way to differentiate between PAUC growth due to acquisition of additional capability and that due to an increase in the actual costs of the MS B capability. Further statistical analysis along the lines of that presented here seems unlikely to be useful in untangling these two elements. Instead, progress on the question of why some programs but not others in Bust1 experienced a boom effect probably will require detailed examination of changes in the relevant programs post MS B.

References

Fox, J. R. (2011). *Defense acquisition reform, 1969 to 2009: An elusive goal.* Washington, DC: U.S. Army Center of Military History.

Kim, Y., Axelband, E., Doll, A., Eisman, M., Hura, M., Keating, E. G., Libicki, M. C., … Shelton, W. (2015). *Acquisition of space systems, Volume 7: Past problems and future challenges* (MG-1171/7-OSD). Santa Monica, CA: RAND. Retrieved from http://www.rand.org/pubs/monographs/MG1171z7.html.

McNicol, D. L. (2004). *Cost growth in major weapon procurement programs* (2nd ed.). Alexandria, VA: Institute for Defense Analyses.

McNicol, D. L., Tate, D. M., Burns, S. K., & Wu, L. (2016). *Further evidence on the effect of acquisition policy on cost growth of major defense acquisition programs* (IDA Paper P-5330 [Rev.]). Alexandria, VA: Institute for Defense Analyses.

McNicol, D. L., & Wu, L. (2014). *Evidence on the effect of DoD acquisition policy and process on cost growth of major defense acquisition programs* (IDA Paper P-5126). Alexandria, VA: Institute for Defense Analyses.

O'Neil, W. D., & Porter, G. H. (2011). *What to buy? The role of Director of Defense Research and Engineering (DDR&E)—Lessons from the 1970s* (IDA Paper P 4675). Alexandria, VA: Institute for Defense Analyses.

Porter, G., Gladstone, B., Gordon, C. V., Karvonides, N., Kneece, R. R., Jr., Mandelbaum, J., & O'Neil, W. D. (2009). *The Major causes of cost growth in defense acquisition: Volume I—Executive summary* (IDA Paper P 4531). Alexandria, VA: Institute for Defense Analyses.

Selected Acquisition Report: Global Broadcast System. (2003). Defense Acquisition Management Information Retrieval (DAMIR) System.

Tyson, K. W., Om, N. I., Gogerty, D. C., & Nelson, J. R. (1992). *The effects of management initiatives on the costs and schedules of defense acquisition programs, Vol. I: Main report* (IDA Paper P-2722). Alexandria, VA: Institute for Defense Analyses.

Younossi, O., Lorell, M. A., Brancato, K., Cook, C. R., Eisman, M., Fox, B., … Sollinger, J. M. (2008). *Improving the cost estimation of space systems: Past lessons and future recommendations* (MG-690-AF). Santa Monica, CA: RAND. Retrieved from http://www.rand.org/pubs/monographs/MG690.html.

Acknowledgments

Valuable comments reflected in the paper were provided by Dr. David Sparrow, Dr. David Tate, Dr. Prashant Patel, Dr. Philip Lurie, Dr. Brian Gladstone, and Dr. Sarah Burns, all of the Institute for Defense Analyses (IDA), and Dr. Mark Husband of the Defense Acquisition University. Ms. Linda Wu of IDA managed data acquisition and the database.

Appendix A: Boom Effects for McNamara-Clifford, Acquisition Reform (AR), and the Bust Phase of the DAB Post AR

Table A-1 presents average Program Acquisition Unit Cost (PAUC) growth and average program duration data for the McNamara-Clifford and the Acquisition Reform (AR) periods.

Table A-1. Average PAUC Growth and Program Duration for Completed Programs for McNamara-Clifford and AR

		Average PAUC Growth	Average Duration
McNamara-Clifford	**Bust0**	87% (12)	9.3
	Bust1	34% (4)	20.5
Acquisition Reform	**Bust0**	2% (1)	7
	Bust1	38% (18)	14.6

* Quantity APUC from the MS B baseline

** From MS B through the year in which the program's last SAR was filed

In contrast to what was found for the DSARC/DAB-Bust period, for McNamara-Clifford, average PAUC growth for Bust0 programs is about two and one-half times that of Bust1 programs. The difference is statistically significant.[15] This may be due to the fact that the Bust1 programs continued into at least the early 1980s and therefore presumably were more strongly influenced by the 1969 Packard acquisition reforms, which are associated with a significant reduction in PAUC growth.

The cost growth data for AR are not useful for statistical analysis because only one program that passed MS B during that period (AV-8B Remanufacture) had been completed by the December 2015 SARs.

Appendix B. RDT&E Cost Growth for the DSARC/DAB Period

Table B 1 presents data on Research, Development, Test, and Evaluation (RDT&E) cost growth and duration in the DSARC/DAB period that parallel the PAUC and duration data presented in Table 2, Table 3, and Table 4. The number of observations in some cells differs from that given for PAUC because the database does not have an RDT&E estimate for all programs for which there is a PAUC growth estimate.

[15] K-S and A-D find the distributions of PAUC growth in Bust0 and Bust 1, respectively, to be consistent with a normal distribution. P = 0.048 for a two-tailed t-test with correction for unequal variances.

Table B-1. Average RDT&E Growth and Average Program Duration by Number of Boom Periods Encountered for Bust and Boom Climates

Bin		Average RDT&E Growth[a]	Average Duration[b]
1st Bust Period	Bust0	5% (5)	6.7
	Bust1	53% (38)	14.1
2nd Bust Period	Bust0	37% (7)	8.0
	Bust1	45% (17)	15.8
1st Boom Period	Boom0	41% (26)	9.6
	Boom1	65% (7)	22.3
2nd Boom Period	Boom0	1% (9)	66.1
	Boom1	n/a (0)	n/a

a Quantity APUC from the MS B baseline
b From MS B through the year in which the program's last SAR was filed

The pattern of growth in RDT&E in the first bust period is consistent with that observed for PAUC growth: (1) Average RDT&E growth for programs in Bust1 is significantly higher than the average for Bust0; and (2) the proportion of programs of Bust1 that fall into the right tail of the distribution also is significantly higher than it is for Bust0.[16]

Average RDT&E growth in the second bust period is noticeably higher in Bust1 than in Bust0, but the difference is not statistically significant. The proportion of programs with RDT&E cost growth of more than 40% also is not significantly higher in Bust1 than in Bust0.[17]

In the first boom period, average RDT&E cost growth is significantly higher for MDAPs in Boom1 than for those in Boom0, and the proportion of MDAPs with RDT&E growth of at least 40% also is significantly higher in Boom1 than in Boom0.[18]

[16] M-W U P = 0.025 (U = 35.5, n_1 = 38, n_2 = 5). P = 0.051 for Fisher's Exact Test (FET) using the number of programs in Bust0 and Bust1 with an RDT&E growth of at least 40%.
[17] M-W U P = 0.308 (U = 43, n_1 = 17, n_2 = 7). P = 1.000 for FET using the number of programs in Bust0 and Bust1 with an RDT&E growth of at least 40%.
[18] M-W U P = 0.075 (U = 132, n_1 = 26, n_2 = 7). P = 0.027 for FET using the number of programs in Bust0 and Bust1 with an RDT&E growth of at least 40%.

Measuring the Impact of Sequestration and the Defense Drawdown on the Industrial Base

Rhys McCormick—is a Research Associate with the Defense-Industrial Initiatives Group (DIIG) at CSIS. His work focuses on unmanned systems, global defense industrial base issues, and U.S. federal and defense contracting trends. Prior to working at DIIG, he interned at the Abshire-Inamori Leadership Academy at CSIS and the Peacekeeping and Stability Operations Institute at the U.S. Army War College. He holds a BS in security and risk analysis from the Pennsylvania State University and an MA in security studies from Georgetown University.

Andrew Hunter—is a Senior Fellow in the International Security Program and Director of the Defense-Industrial Initiatives Group at CSIS. From 2011 to 2014, he served as a senior executive in the Department of Defense, serving first as chief of staff to under secretaries of defense (AT&L) Ashton B. Carter and Frank Kendall, before directing the Joint Rapid Acquisition Cell. From 2005 to 2011, Hunter served as a professional staff member of the House Armed Services Committee. Hunter holds an MA degree in applied economics from the Johns Hopkins University and a BA in social studies from Harvard University.

Greg Sanders—is a Fellow in the International Security Program and Deputy Director of the Defense-Industrial Initiatives Group at CSIS, where he manages a research team that analyzes data on U.S. government contract spending and other budget and acquisition issues. In support of these goals, he employs SQL Server, as well as the statistical programming language R. Sanders holds an MA in international studies from the University of Denver and a BA in government and politics, as well as a BS in computer science, from the University of Maryland.

Abstract

The presence of a technologically superior defense industrial base has been a foundation of U.S. strategy since 1945. While the implementation of the budget cuts in the Budget Control Act of 2011 has caused concerns for the industrial base, the resulting debate has been lacking in empirical analysis. The purpose of this research is to measure the impact of the current defense drawdown across all the tiers of the industrial base. The technical approach analyzes prime and subprime DoD contract data to measures the impacts of the drawdown by sector to better understand how prime and subprime contractors have responded to this external market shock.

Introduction

The United States has long recognized the importance of supporting and sustaining an advanced defense industrial base to maintain global technological superiority. Maintaining a technologically superior industrial base requires a wide vendor pool from which to produce products, conduct research and development, and provide services for the Department of Defense (DoD). This vendor pool includes both the defense contractors awarded prime contract obligations but also the lower subcontracting tiers of the industrial base.

However, since the implementation of the 2011 Budget Control Act's (BCA's) mandatory reductions to the federal budget, Congressional, DoD, government oversight, and industry officials have all expressed concerns over the health and future of the defense industrial base. These cuts affect not only the top tier of the industrial base (the prime contractors), but also the more numerous lower-tier suppliers (subcontractors) that are so often the sources of critical technological advances. Heavily dependent on subcontract awards from the prime contractors, some of these subcontractors face the risk of going out of business due to the drawdown. Funding associated with the wars in Afghanistan and Iraq has also declined steeply since 2011, further reinforcing and magnifying the effect of the

BCA reductions. The combined effect of these reductions is what is referred to as the current defense drawdown, or the drawdown, for purposes of this project.

The current public discussion surrounding the impact of the drawdown on industry is largely based on anecdotes lacking empirical support. Through analysis of publicly-available contract data, this research effort measures the impacts of the drawdown by sector to better understand how prime and subprime contractors have responded to this external market shock.

Literature Review

As stated previously, the public discussion surrounding the impact of budget drawdown trends on industry is often based on anecdotes, absent of empirical evidence. To better assess the validity of some of these claims, the study team looked to the academic literature to help ground the analysis in general historical principles of industrial base evolution. Where similarities exist, the academic literature permits comparing whether the challenges of sequestration, and subsequent responses, are like similar historical external market shocks seen in the private sector. Reviewing the academic literature further illuminates research variables that, while present in business and academic journals, have been underexplored in the defense context.

DoD Component

The DoD faced the largest overall reductions of any department in the U.S. federal government during sequestration. These reductions had significant but uneven effects on DoD spending and affected each service in differing ways. Though the defense industrial base is effectively a monopsony in which the U.S. federal government is ultimately the only buyer, many acquisition decisions are not made by a singular decision-making organization, but by the major DoD components. While a topline budget and overall/cross-department acquisition trends are somewhat out of the components' control, lower-level trends are likely to reflect the component's top priorities and not just standardized cuts across the board. For example, given these dynamics, it would not be surprising to see the Navy limit, to the extent possible, cuts to its shipbuilding budget even if it meant taking sharper cuts elsewhere.

The policy guidance for responding to budgetary cuts coming out of the components leading up to and throughout the defense drawdown reflects this dynamic. Each of the different components had its own set of priorities and varied plans for addressing the budgetary challenges. For example, the Navy's choices are seen in this 2014 Quadrennial Defense Review statement: "To sustain investment in critical force structure and modernization, the Navy will reduce its funding for contractor services by approximately $3 billion per year to return to 2001 levels of contractor support" (DoD, 2014b). Meanwhile, the Air Force planned to address the budgetary challenge by making "near-term capacity reductions in mission areas such as lift, command and control, and fighters" to prioritize its top three modernization programs: F-35 Fighter, B-21 Bomber, and KC-46A Tanker (DoD, 2014b). Furthermore, the Army announced that it would take an approach different from either the Air Force or the Navy, electing to protect funding for readiness at the expense of modernization and force structure.

Vendor Size

A critical question asked prior to and throughout sequestration and the drawdown was whether smaller defense contractors would be able to survive the sequestration and continuing drawdowns (Samuelsohn, 2013). Furthermore, Sen. Mary Landrieu, Chairwoman of the Small Business and Entrepreneurship Committee, speculated that "small businesses are going to be the ones that feel the most immediate effects" of spending cuts originating from the BCA (Samuelsohn, 2013). Due to the number of contracts held by smaller defense contractors and their specialized niche capabilities, some argued that it seemed almost inevitable that the negative impacts of sequestration will "disproportionately" affect smaller contractors (Eaglen, 2012). Without having a large and diversified portfolio of defense contracts that reduce the impact of spending cuts in one line of business, small defense contractors looked to be unable to withstand the reductions in military spending (Homan, 2014).

Within the academic literature, the relationship between vendor size and its success during a downturn is less clear. Even though commentators tend to give credit to larger businesses having more success than small business during an economic downturn, the literature suggests that success is more dependent on strategies available to a company, not its size alone (Sivy, 2012). The role of vendor size is indirect but can still be critical; the size of a vendor influences what business strategies are available for pursuit. Vendors of different sizes pursue different strategies during periods of market shock, such as economic downturns.

Smaller businesses and non-profits may have their strategic options limited because they face significantly higher obstacles to other strategies, like raising money, during an economic downturn (Banjo & Kalita, 2010). Due to their associated risk, small businesses were often denied needed external financing from banks during the 2008 recession (Guo, 2014). Without the revenue of a growing market and no access to external financing, small business were left with higher rates of unemployment compared to large businesses (Guo, 2014). Additionally, during the recent recession, it was common for organizations to immediately seek the means to reduce their operating costs in order to stay afloat (Gulati, Nohria, & Wohlgezogen, 2010). Larger companies typically rely on their ability to consolidate and reduce significant amounts of operating costs to survive an economic downturn (Kambil, 2008). While this option may be available to larger companies who have multiple lines of business and substantial reserves to pull from, small businesses do not have the same quantity of cash flow or large reserves available (Bossaller & Kammer, 2009).

Although small businesses generally faced increasingly more difficult challenges during the downturn, they also retained certain benefits that large companies did not have access to (Lai et al., 2016). When reducing operating costs, large companies often undergo substantial structural changes that force larger lay-offs (Lai et al., 2016). Small firms, on the other hand, have a notable strength in flexibility and adaptability to a rapidly changing market (Lai et al., 2016). Without the levels of bureaucracy in a large company, small companies retain a shorter timeline for decision-making which allows them to respond quickly and efficiently to their customer base (Bossaller & Kammer, 2009).

Vendor Count: "Consolidation Theory"

Both the academic literature and historical examples suggest that the DoD should expect to see consolidation within the defense industrial base under sequestration and the subsequent drawdown. Since the end of the Cold War, defense contractors have resorted to consolidation amid budgetary drawdowns (Gholz & Sapolsky, 2000). As the defense budget fell sharply throughout the 1990s, defense contractors turned to horizontal mergers, acquisitions, and divestitures in order to prevent themselves from going under, setting off "a wave of consolidation" that reduced the number of American-based large prime defense contractors from 16 in 1993 to only six in 2000 (Alfieri et al., 2014; Kovacic & Smallwood, 1994).

After the BCA was enacted in 2011, and with the prospect of sequestration looming on the horizon, many defense contractors were worried about their imminent future (Scully, 2011). Although history suggests that we would expect to see an increase in consolidation in such circumstances, this may not be the case at the top tier of defense contracting, given that the already high-level of consolidation during the post–Cold War drawdown left little room for the large prime defense contractors to acquire additional market share (Thompson, 2010). Nonetheless, in the period leading up sequestration, large primes such as Lockheed Martin, L-3, and Exelis were vocal about seeking the means to consolidate and waiting to "take any available piece of a shrinking pie" (Banham, 2013).

The academic literature supports the argument that we might expect to see further consolidation within the defense industry under market shocks such as sequestration and the defense drawdown. One strategy for improving profit and revenue during a recession has been to effectively consolidate certain aspects of a business (Kambil, 2008). A recessionary period offers a unique opportunity for businesses to capitalize on competitors' vulnerabilities and increase value through consolidation (Rhodes & Stelter, 2009). In a recession, consolidation through a merger has been shown to generate 15% more value than in "normal conditions" (Rhodes & Stelter, 2009). Furthermore, the relationship between market shocks such as recessions to higher rates of consolidation was also recently demonstrated by the higher rate of consolidation in the banking industry during the 2008 recession.

Competition

An evergreen top DoD priority is the presence of a competitive defense industrial base. In the "Guidelines for Creating and Maintaining a Competitive Environment for Supplies and Services in the Department of Defense," the DoD lays out seven different reasons competition is important in the defense marketplace (DoD, 2014a):

1. Competition creates an incentive for contractors to provide goods and services at a lower price (economic efficiency);
2. Competition spurs innovation of transformational technologies, which allows the Department to field the best weapon systems for our warfighters quickly;
3. Competition yields improvements in the quality of products delivered and services rendered (firms that turn out low quality are driven out of the market and are unable to effectively compete);
4. Competition affords the Department the opportunity to acquire performance improvements (e.g., faster, lighter, more sustainable) by using "best value" source selection criteria;
5. Competition provides opportunities for capable small businesses to enter new markets;
6. Competition enhances (or maintains) a strong defense industrial base which provides an operational surge capability to handle demand spikes, and;
7. Competition curbs fraud by creating opportunities to re-assess sources of goods and services reinforcing the public trust and confidence in the transparency of the Defense Acquisition System.

Given the importance of competition, the DoD tracks and publishes the share of contract obligations in its annual "Competition Report." In the DoD's *FY 2015 Competition Report*, it reported that the share of contract obligations awarded after competition had been falling, with the exception of FY 2014, each year since FY 2009. Whereas 60.7% of FY 2009 contract obligations had been awarded after competition, only 55.4% of FY 2015 contract obligations were awarded after competition (Defense Procurement and Acquisition Policy, n.d.). However, CSIS analysis, supported by data contained in the *FY 2015 Competition Report*, shows that the declines in the overall competition rates are a result of policies reducing conditions in which contracts are awarded after an open competition, but receive only offer. Therefore, while the overall rate of competition may have technically fallen, the rate of effective competition has remained relatively steady (Hunter et al., 2017).

The academic literature on consolidation is also relevant here as market shocks can further reduce competition by encouraging consolidation. In a consolidated market, a smaller number of firms have a greater market share, which reduces the number of potential competitors for any given project. While the decline in competition predates sequestration, its continuation during most of the drawdown years seems to show that at the Department-wide level, the literature and the DoD's metrics are aligned.

However, while the annual DoD competition report provides important data at the topline, it insufficiently measures the rate of competition at lower levels, particularly sector-by-sector. Beyond the topline, the annual competition reports provide data on the rate of competition within the major DoD components, but each service reports its data differently, and these reporting frameworks do not always align for comparative purposes.

Methodology & Study Design

This report leverages and builds upon the methodology used in previous CSIS reports on federal contracting. [1] To measure the impact of sequestration and the defense drawdown on different sectors of the defense industrial base, the study team first created a dataset of prime and sub-prime contract awards from 2010–2015 using the Federal Procurement Data System (FPDS) and Federal Subaward Reporting System (FSRS). From this dataset, the CSIS study team separated the defense industrial base into 10 distinct "platform portfolios."[2] To create these platform portfolios, the study team first classified contract obligations by their listed DoD Claimant Program Code. Second, for instances where the "DoD Claimant Program Code was missing or not platform specific (e.g., Services or Subsistence) obligations were classified using the Product or Services Code" (Berteau, McCormick, & Sanders, 2014).

Having created these 10 unique platform portfolios, the CSIS study team decided to focus its analysis in this paper on the three sectors of the defense industrial base that Frank Kendall, former Under Secretary of Defense (AT&L), had previously identified as of the most concern: ground combat, high-performance aircraft, and surface combatant ships (Bertuca, 2014). These three sectors largely align to the Aircraft and Drones, Land Vehicles, and Ships and Submarines platform portfolios respectively. In the final technical report, the CSIS study team will expand its analysis to include the other platform portfolios beyond these initial three sectors.

---

[1] For the full CSIS FPDS methodology, see: http://csis.org/program/methodology

[2] The 10 unique CSIS platform portfolios are as follows: Aircraft and Drones; Ships and Submarines; Land Vehicles; Missiles and Space Systems; Weapons and Ammunition; Other Products; Electronics and Communications; Facilities and Construction; Other Services; Other R&D and Knowledge Based; Unlabeled

For these three platform portfolios, the CSIS study team focused on the four variables identified in the literature review that are observable through FPDS data:

- DoD Component: Did the DoD components respond differently to sequestration and the defense drawdown?
- Vendor Size: How did the share of contract obligations change among vendors of differing sizes, particularly small businesses?
- Vendor Count: How did the number of vendors change?
- Competition: Did the share of contract obligations awarded after effective competition change?

Finally, the study team sought to evaluate the availability and quality of subcontracting data across the different sectors of the defense industrial base. This effort builds off a 2014 study conducted by Nancy Moore at RAND, which concluded for FSRS data from FY 2010 to FY 2012, FSRS data was often incomplete or missing, but was improving each year (Moore, Grammich, & Mele, 2014).

Did the DoD Components Respond Differently?

Across the DoD, the response to the market shock imposed by sequestration and the budget drawdown differed among the various major DoD components, both in magnitude and response strategy. At the top line, average overall DoD annual contract obligations from 2012–2015 fell by 21% compared to the pre-down period. Of the major DoD components, the Army bore the brunt of these cuts, suffering a 39% decline in average annual contract obligations over that same period while the Air Force (-15%) fell at rates below the overall DoD rate of decline. Finally, the Defense Logistics Agency (DLA) (-7% decline), Navy (-12%), and Other DoD (-6%) fell at rates significantly slower than the overall DoD rate of decline, while the Missile Defense Agency (MDA) grew 12% over that same period.

Within the major DoD components, their response to sequestration and the defense drawdown differed. The Air Force, more so than any other component, balanced the distribution of the cuts but still elected to fund certain platform portfolios over others. As annual average Air Force contract obligations declined 15% during the drawdown, most Air Force platform portfolios fell at rates similar to the overall rate. For example, average annual Air Force contract obligations for Aircraft and Drones, Missile and Space Systems, and Electronics and Communications fell by 11%, 20%, and 14% respectively. The Air Force made cuts greater than the overall rate of decline to its Facilities and Construction (-26%) and Other Products (-42%) allowing for the 11% increase in Weapons and Ammunition.[3]

The Army made cuts to every platform portfolio, but those cuts were not distributed evenly across the platforms. The Army Aircraft and Drones platform portfolio saw the smallest cut (-8% decline in average annual contract obligations), followed by Missiles and Space Systems (-22%), and Other R&D (-27%). To limit the cuts made in these platform portfolios, more severe cuts were made to Land Vehicles (-67%), Other Products (-51%), Other Services (-49%) and Weapons and Ammunition (-49%).

---

[3] Air Force Ships & Submarines (44%) and Land Vehicles (61%) platform portfolios experienced growth in average annual contract obligations during the drawdown, but represent just 0.2% of total Air Force contract obligations.

Finally, similar to the Army, the Navy elected to protect certain platform portfolios over others. However, unlike the Army, the Navy increased funding over previous levels for certain platform portfolios. As overall average annual Navy contract obligations decreased by 12%, average annual Navy contract obligations for Aircraft and Drones and Missiles and Space increased by 11% and 0.1% respectively. Additionally, average annual Navy Ships & Submarines contract obligations decreased just 4% during the defense drawdown. Funding for these three platform portfolios was offset by more severe cuts in Electronics and Communications (-23%), Facilities and Construction (-24%), Other Products (-41%), Land Vehicles (-79%) and Other R&D (-36%).

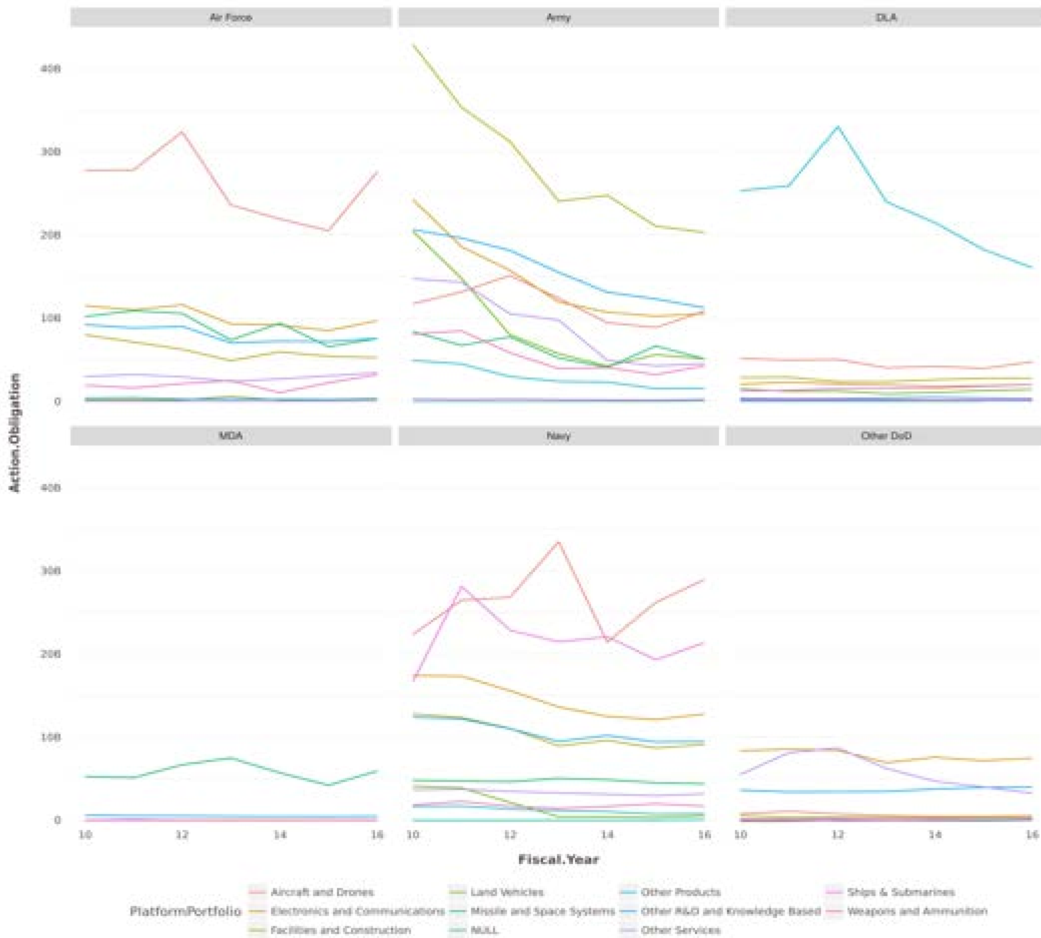


Figure 1. Platform Portfolio by Major DoD Components, 2010–2016
(Source: FPDS; CSIS analysis)

Platform Portfolio Case Study 1: Aircraft and Drones

For the first platform portfolio, Aircraft and Drones, average annual contract obligations during the defense drawdown declined 4% as compared to pre-drawdown average contract obligations. The predominant source of that decline was the 41% decline in annual average contract obligations for Aircraft and Drones R&D contract obligations. Annual average contract obligations for Products remained steady, and Services (10%) grew over that period.

Aircraft: Vendor Size

Under sequestration and the defense drawdown, the Big 5 have only further increased their market share of this sector at the expense of other large contractors.[4] Prior to the drawdown, the Big 5 accounted for 57% of total Aircraft and Drones contract obligations, compared to 30% for Large vendors. Between 2012 and 2015, the Big 5 increased their market share to 61% of total Aircraft and Drones contract obligations as Large vendors fell to 26%. These trends continued into the reversal of the contracting drawdown in 2016 with Big 5 vendors rising to 64% and Large vendors falling to 24% of total Aircraft and Drones contract obligations.

The share of contract obligations going to Small and Medium vendors remained relatively steady. Before the drawdown, small and medium vendors were awarded 5% and 8% of total Aircraft and Drones contract obligations respectively and remained at that rate throughout the drawdown period. This outcome is in line with the finding from the literature that business strategy more than business size drives results.

Figure 2 shows Aircraft and Drones by Vendor Size from 2010 to 2016.

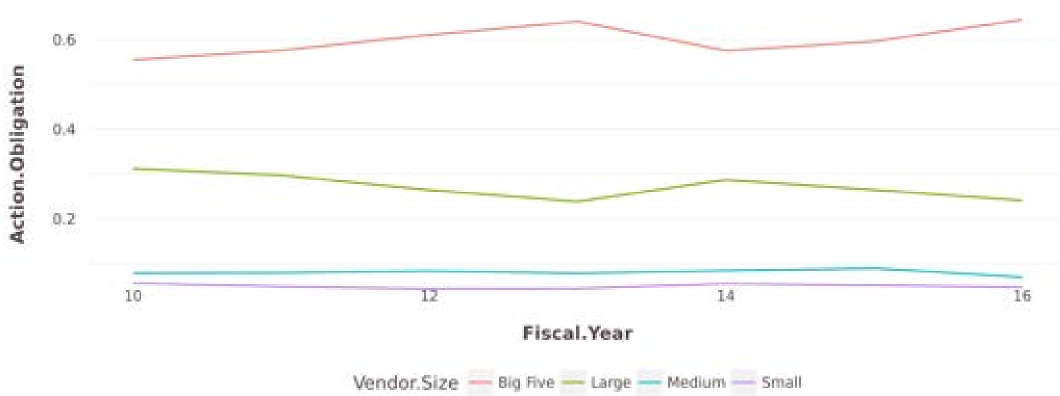


Figure 2. Aircraft and Drones by Size of Vendor, 2010–2016
(Source: FPDS; CSIS analysis)

Aircraft: Competition

Under sequestration and the defense drawdown, the historical trends for increasing contract obligations awarded without competition in the Aircraft and Drones sector accelerated. Between 2000 and 2010, only 26% of Aircraft and Drones contract obligations were awarded after effective competition, while 67% of contract obligations were awarded after no competition.[5] Between 2012 and 2015, the share of annual average contract obligations awarded after effective competition fell to 18%, while the share of annual average contract obligations awarded without competition increased to 79%. Throughout the

---

[4] CSIS defines the Big 5 as Lockheed Martin, Boeing, Raytheon, Northrop Grumman, and General Dynamics.
[5] CSIS uses the term *effective competition* to refer to competition with two or more offers.

2012 to 2015 period, annual average contract obligations awarded without competition grew from $51.9 billion prior to the drawdown, to $53.6 billion, a 3% increase even as overall Aircraft and Drones contract obligations fell 4%.

When the contract drawdown began to reverse in 2016, these trends only further continued as the share of Aircraft and Drones contract obligations awarded after no competition increased to 81%. Only 18% of Aircraft and Drones contract obligations were awarded after effective competition in 2016.

Figure 3 shows Aircraft and Drones by competition classification from 2010 to 2016.

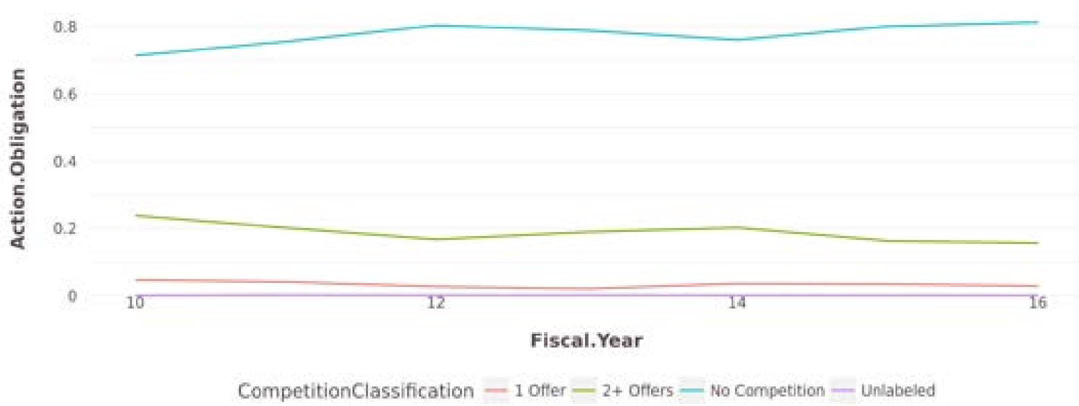


Figure 3. Level of Competition for Aircraft and Drones Contract Obligations, 2010–2016
(Source: FPDS; CSIS analysis)

Aircraft: Vendor Count

As shown in Figure 5, the number of vendors in the Aircraft and Drones sector increased over the course of the defense drawdown after initially declining. This result is somewhat surprising given that since 2005, except for 2008, the number of vendors in the Aircraft and Drones sector had been declining compared to the previous year, reaching approximately 6,100 vendors in 2010. This decline continued until 2014, when there were under 5,700 vendors in 2014, a 7% decline from 2010. However, beginning in 2015 and continuing into 2016, the number of vendors in the Aircraft and Drones sectors increased from the previous year. In 2016, there were approximately 6,250 vendors in the Aircraft and Drones sector, a 10% increase as compared to 2014. The speed with which the number of vendors rebounded and the steady market share for small and medium vendors is consistent with the observation in the literature that smaller players can prove nimble in response to market shocks.

Figure 4 shows the number of vendors in the Aircraft and Drones platform portfolio from 2005 to 2016.

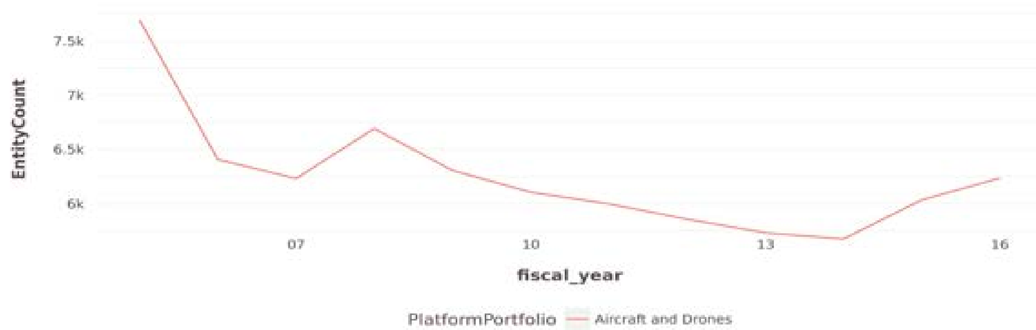


Figure 4. Aircraft and Drones by Vendor Count, 2005–2016
(Source: FPDS; CSIS analysis)

Aircraft and Drones: Subcontract Data Availability

In analyzing FSRS data, the CSIS study team found a large discrepancy in the availability of subcontracting data for the Aircraft and Drones platform portfolio between the major DoD components. Shown in Figure 5, of the major DoD components, the Navy and Other DoD had the greatest share of prime contract obligation dollars in FSRS, both near approximately 45%. Comparatively, both the Air Force and the Army lag the Navy and Other DoD, reporting less than one-third of prime contract obligations in FSRS annually.

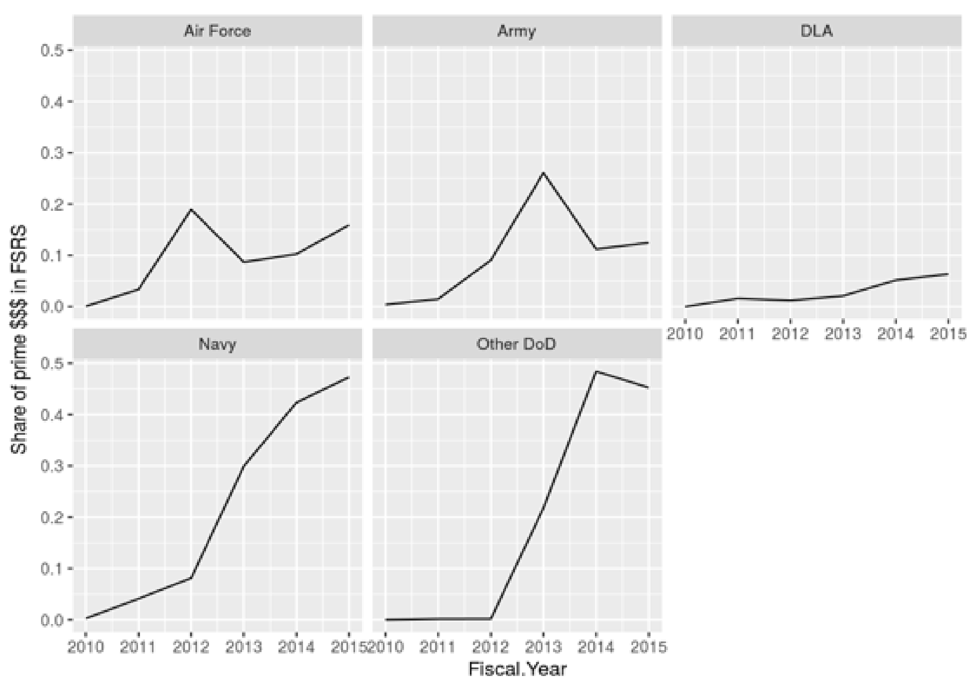


Figure 5. Aircraft and Drones FSRS Data Availability by SubCustomer, 2010–2015
(Source: FPDS; CSIS analysis)

Platform Portfolio Case Study 2: Land Vehicles

In recent years, the Land Vehicles platform portfolio underwent greater percentage declines than any other platform portfolio as average annual contract obligations fell -65% compared to before the drawdown. The collapse reflects not just that the components priorities were elsewhere, but also a fall from favored status during the period of large scale contingency operations and rapid acquisition of highly protected tactical vehicles such as mine resistant ambush protected vehicles. The 65% decline was nearly double the next closest percentage decline (Weapons and Ammunition; -34%), and significantly higher than the 21% overall DoD decline.

Land Vehicles: Vendor Size

As shown in Figure 6, throughout the 2013–2015 period, the Land Vehicles sector experienced significant change in the share of contract obligations awarded by vendor size. Prior to the start of the defense drawdown, Land Vehicle contract obligations were awarded as follows: 58% to Large vendors, 14% to the Big 5, 20% to Medium vendors, and just 7% to Small vendors. Throughout the drawdown, the Big 5 and Small vendors saw increased market share at the expense of Large vendors. During this period, the share of annual average contract obligations going to Large vendors fell to 45%, while the share going to the Big 5 and Small Vendors increased to 23% and 14% respectively. The increased share of contract obligations going to the Big 5 (-44%) and Small (-34%) vendors can be accounted for by their slower average annual contract obligation decline than the overall rate of decline (-65%).

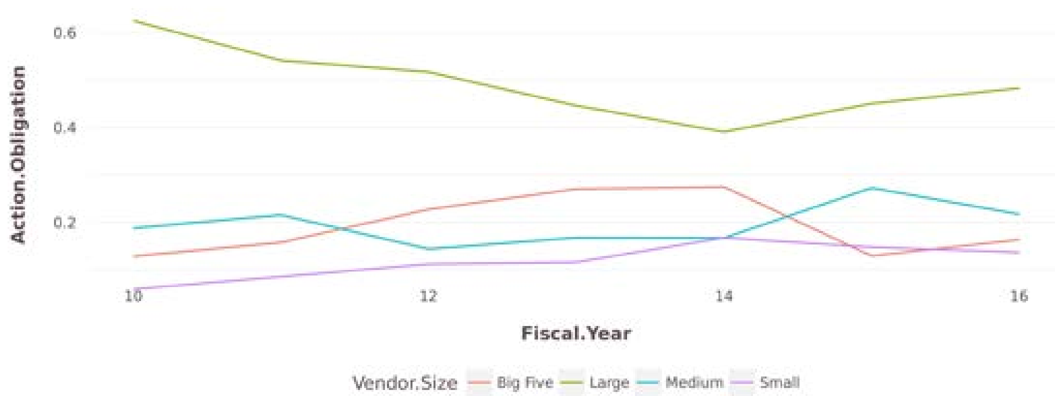


Figure 6. Land Vehicles by Size of Vendor, 2010–2016
(Source: FPDS; CSIS analysis)

Land Vehicles: Competition

Throughout the defense drawdown, the Land Vehicle sector saw a decline in the rate of effective competition and increase in the share of contract obligations awarded without competition. As shown in Figure 7, at the start of the drawdown, 32% of Land Vehicle contract obligations were awarded after effective competition, and 52% were awarded with zero competition. During the drawdown, the share of contract obligations awarded after effective competition fell slightly to 30%, as the share of contract obligation awarded without effective competition increased to 69% from 65%.

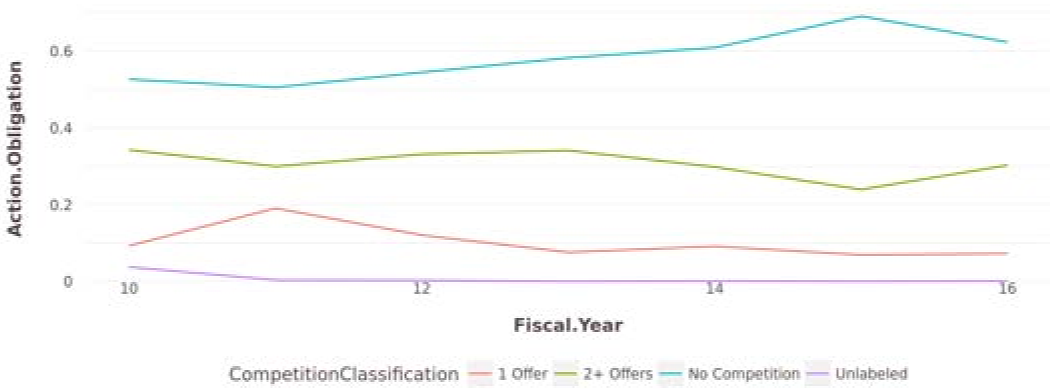


Figure 7. Level of Competition for Land Vehicles Contract Obligations, 2010–2016
(Source: FPDS; CSIS analysis)

These trends began to reverse themselves in 2016 as the share of contract obligations awarded after effective competition increased to 30%, and the share of contract obligations awarded without competition fell to 62%. New contracts appear to be driving the increasing share of effective competition; the sector saw a 9% rise in contract obligations in 2016, as compared to average annual contract obligations during the drawdown period. Land Vehicle contract obligations awarded after effective competition rose from an average of $2.0 billion from 2013–2015 to $2.2 billion in 2016.

Land Vehicles: Vendor Count

At the start of this defense drawdown, the downward trend in the number of vendors in the Land Vehicles sector continued declining before eventually flattening out and slowly rebounding near the end of the study period. After spiking in 2009 at approximately 5,900 vendors, 2010 marked the start of the decline in the number of Land Vehicles vendors as the wars in Afghanistan and Iraq and subsequent war-related vehicle funding declined. This trend continued until 2013, when there were just under 3,950 vendors, a 33% decline from 2009. However, that trend began to slowly reverse in 2014, with the number of Land Vehicles vendors growing on average, 1.5% per year since 2014, as shown in Figure 8.

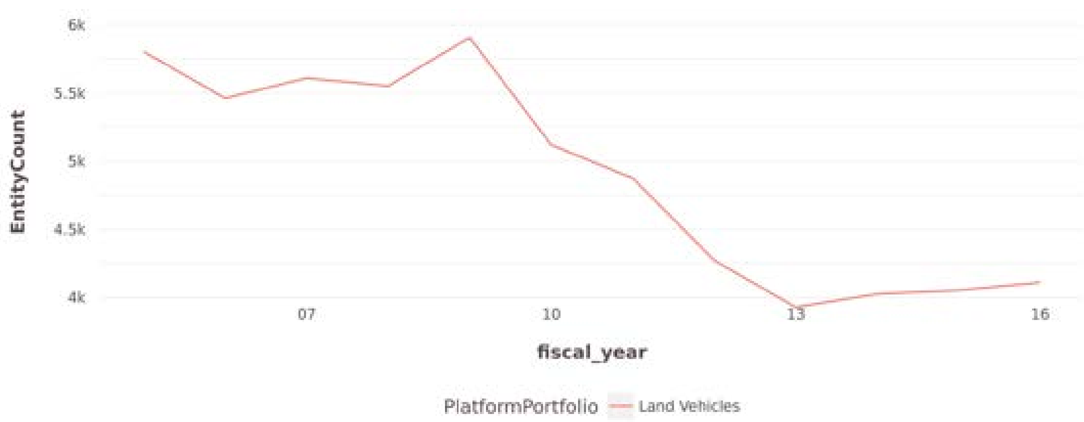


Figure 8. Land Vehicle by Vendor Count, 2005–2016
(Source: FPDS; CSIS analysis)

Land Vehicles: Subcontract Data Availability

The FSRS data show that the subcontracting data for Land Vehicles is more available than Aircraft and Drones, but significant gaps exist. At peak data availability in 2014, over 40% of prime dollars was available in FSRS. However, as prime contract obligations for Land Vehicles increased in 2015, FSRS data availably fell.

The Land Vehicle data validate Nancy Moore's findings on the manner in which old, large contracts drag down the availability rate of FSRS subcontract data. Given the requirement to report subcontract awards to FSRS is relatively new and not retroactive, many large older contracts will never show up in the database. Therefore, as those contracts are canceled, expired, or not renewed, one should expect to see FSRS data availability go up, as in the case of Land Vehicles during the defense drawdown.

Figure 9 shows the share of prime dollars available in FSRS compared to the share of prime dollars not available in FSRS from 2010 to 2015.

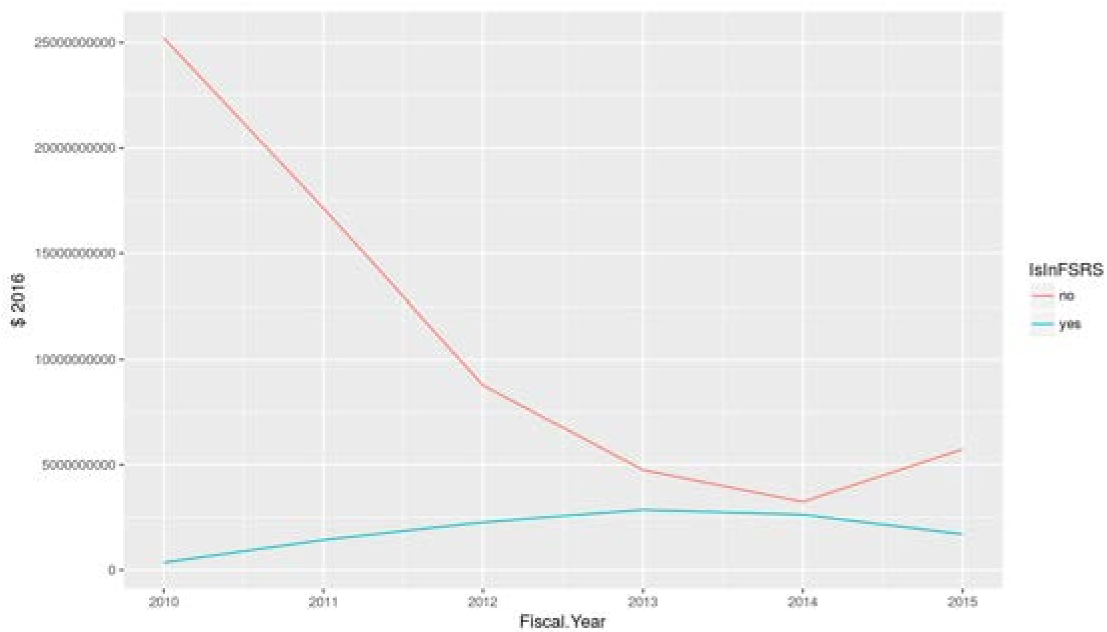


Figure 9. Land Vehicles FSRS Data Availability, 2010–2015
(Source: FPDS; CSIS analysis)

Platform Portfolio Case Study 3: Ships & Submarines

During the recent defense drawdown, the Ships & Submarines platform portfolio saw the smallest decline of all the platform portfolios, falling just 3% from the pre-drawdown levels. During the drawdown, average annual Ships & Submarines R&D contract obligations declined by 49%, compared to the 3% decline in Products and 2% increase in Services contract obligations.

Ships & Submarines: Vendor Size

Over the study period, the Ships & Submarines platform portfolio saw a significant shakeup in the share of contract obligations awarded by vendor size as the Big 5 saw significant decreases in market share. However, this trend was largely driven by Northrop Grumman's decision to spin off its shipbuilding sector into Huntington Ingalls Industries (HII) in 2011. Given shipbuilding's low profit margins and then-uncertainty about future defense budgets at that time, Northrop decided to re-prioritize investment in other sectors of the defense industrial base and spin off its shipbuilding assets, creating HII effective halfway through FY 2011 (Drew, 2011).

Figure 10 shows that at the start of the drawdown and prior to the formation of HII, the Big 5 accounted for 58% of all Ships & Submarines contract obligations. However, by the end of the defense drawdown, the Big 5 accounted for just 37% of contract obligations, as the share of contract obligations awarded to Large vendors increased from 20% to 41%. Both the share and sum of contract obligations awarded to Medium-sized vendors. As a share of contract obligations, Medium-sized vendors increased from 9% to 11%, while total contract obligations increased by 13%, rising from an annual average of $2.2 billion pre-drawdown, to $2.5 billion during the drawdown. The share of contract obligations going to Small vendors fell slightly, but remained relatively steady going from 12% to 11%.

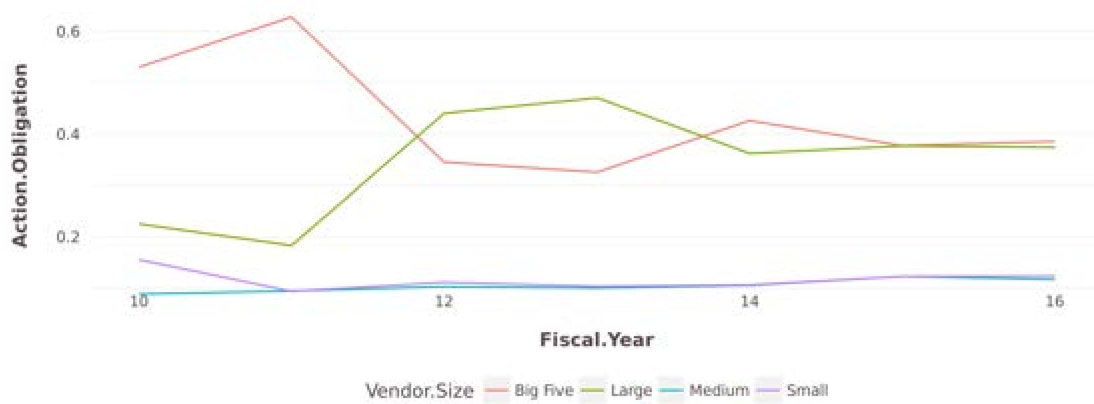


Figure 10. Ships & Submarines by Size of Vendor, 2010–2016
(Source: FPDS; CSIS analysis)

Ships & Submarines: Competition

During the most of this recent drawdown, the Ships & Submarines platform portfolio saw an increase in the share of contract obligations awarded after effective competition as shown in Figure 11. In the years prior to the start of the drawdown, 32% of contract obligations were awarded after effective competition. Throughout the drawdown, the percentage of contract obligations awarded after effective competition rose from 32% to 38%. As the share of contract obligations awarded after effective competition increase, the share of contract obligations awarded without competition and the share of contract obligations awarded with just one offer fell from 61% to 56% and 7% to 6% respectively. The rise in shipbuilding competition contrasts with the fall in harder hit sectors, suggesting that rates of competition may be one of the consequences of the shock.

Of note, the increase in the share of contract obligations awarded after effective competition was not from declines in average annual contract obligations awarded without competition (-10%), but by increasing average annual contract awarded after effective competition as the overall platform portfolio decreased. Prior to the drawdown, annual average Ships & Submarines contract obligations awarded after effective competition totaled $7.7 billion. During the drawdown, that number increased by 14%, rising from $7.7 billion to $8.8 billion.

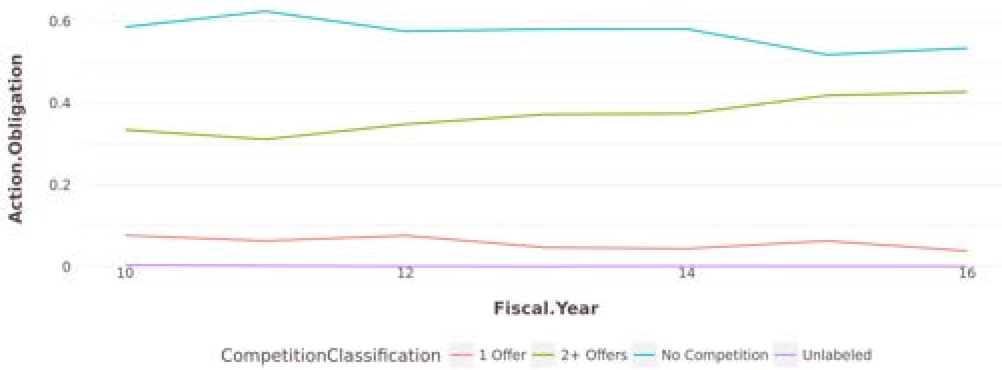


Figure 11. Level of Competition for Ships & Submarines Contract Obligations, 2010–2016
(Source: FPDS; CSIS analysis)

Ships & Submarines: Vendor Count

In the years leading up to the defense drawdown, the number of vendors in the Ships & Submarines sector had been slowly increasing after a previous decline, peaking at approximately 5,300 in 2011. After peaking in 2011, the number of vendors in the Ships & Submarines sector declined slightly, approximately 1%, for two years until 2013. Since 2013, the number of vendors in this platform portfolio has increased by 2.7% annually, totaling an approximately 5,600 vendors.

Figure 12 shows the number of vendors in the Ships & Submarines platform portfolio from 2005 to 2016.

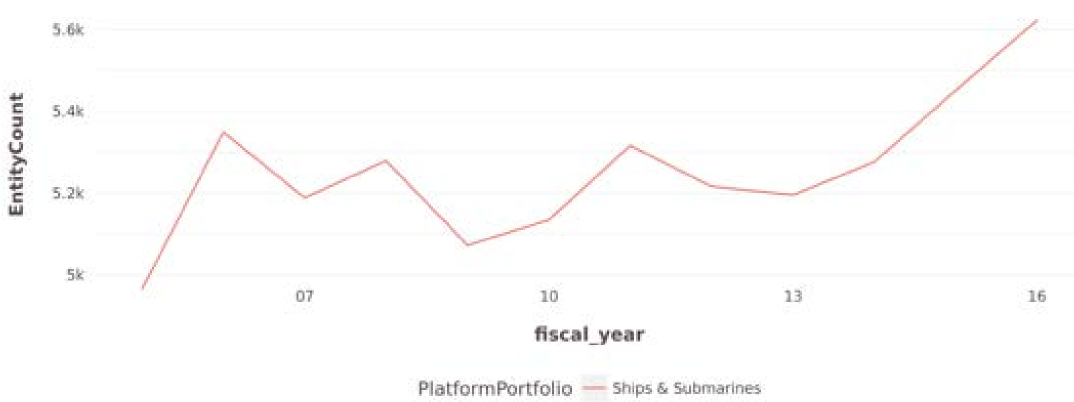


Figure 12. Ships & Submarines by Vendor Count, 2005–2016
(Source: FPDS; CSIS analysis)

Ships & Submarines: Subcontract Data Availability

Of the platform portfolios analyzed, the Ships & Submarines FSRS data was the most incomplete. As shown in Figure 13, after continual increases in data availability from 2010 to 2013, the share of prime dollars available in FSRS peaked around 25% in 2013.

Since 2013, the share of prime dollars available in FSRS has continued to fall and sits at approximately 13% in 2015.

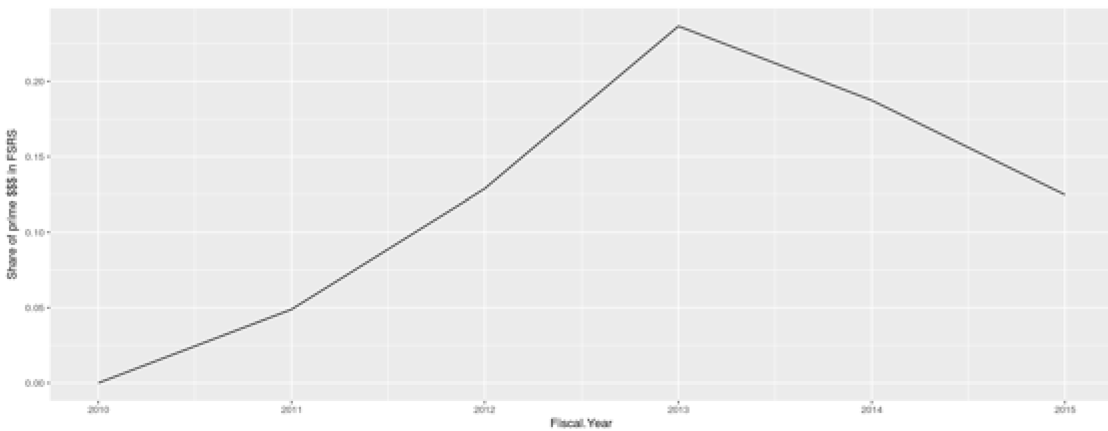


Figure 13. Ships & Submarines FSRS Data Availability, 2010–2015
(Source: FPDS; CSIS analysis)

Conclusion

The results of this preliminary data analysis show that sequestration and the broader defense drawdown have made a measurable impact on the defense industrial base. Furthermore, the data show that the impact of sequestration and the defense drawdown has not been uniform across the entire defense industrial base, with each sector analyzed in this paper responding differently.

DoD Component: Did the DoD components respond differently to sequestration and the defense drawdown?

The results of the CSIS analysis demonstrate that the Army, Navy, and Air Force each took a different approach for responding to the challenges imposed by sequestration and the defense drawdown. The Army, facing the most significant budgetary declines, elected to distribute uneven cuts across all platform portfolios. In the Army's contracting account, the Aircraft and Drones and Missile and Space Systems platform portfolios saw smaller cuts than the overall rate of Army decline, at the expense of other platform portfolios such as Land Vehicles and Weapons and Ammunition. The Air Force took a more distributed approach with only a few platform portfolios seeing cuts larger than the overall rate of decline. Finally, the Navy prioritized three platform portfolios (Aircraft and Drones; Missiles and Space Systems; Ships & Submarines) at the expense of more severe cuts in five other platform portfolios.[6]

---

[6] These five Navy Platform Portfolios that experienced more severe cuts were: Electronics and Communications; Facilities and Construction; Other Products; Land Vehicles; and Other R&D.

Vendor Size: How did the share of contract obligations change among vendors of differing sizes, particularly small businesses?

The data show that the changes in the share of contract obligations among vendors of differing sizes depended on the sector of the defense industrial base. In the Aircraft and Drones sector, the Big 5 vendors maintained and expanded their market share during the recent defense drawdown at the expense of Large vendors. Small vendors maintained their pre-drawdown, albeit small, share of this sector market seeing little change in either direction throughout the study period.

In the Land Vehicles sector, the Big 5 and Small vendors increased their share of the market throughout the defense drawdown at the expense of Medium and Large vendors.

Finally, in the Ships & Submarines platform portfolio, the Big 5's decreased market share throughout the defense drawdown is attributable to the spinoff of HII from Northrop Grumman. Contract obligations that had previously been going to Northrop were now being awarded to HII, a Large vendor. Throughout the study period, Small vendors remained relatively steady around 12% of overall Ships & Submarine contract obligations.

Vendor Count: How did the number of vendors change?

Across all three platform portfolios analyzed, the data show that at the start of the defense drawdown the number of vendors in these sectors was decreasing, before starting to rebound near the end of the defense drawdown. For example, the Aircraft and Drones sector experienced a 7% loss in vendors from 2010 to 2014, before growing by 10% from 2014 to 2016. This is similarly matched by the 1.5% and 2.7% annual growth in the number of vendors in the Land Vehicles and Ships & Submarines platform portfolio respectively.

Competition: Did the share of contract obligations awarded after effective competition change?

The data show two very different results across the three platform portfolios analyzed here. In both the Aircraft and Drones and Land Vehicles platform portfolios, the rate of effective competition declined throughout the course of the defense drawdown. In both cases, the rise in the share of contract obligation awarded without effective competition was driven by increases in the share of contract obligations awarded without competition.

Meanwhile, the Ships & Submarines industry is often anecdotally referred to as one of the least competitive sectors of the industrial base. However, the data show that share of contract obligations awarded after effective competition increased during the drawdown. This trend was driven by two separate factors. First, throughout the drawdown, annual average contract obligations awarded without competition increased at a rate higher than the overall rate of decline. Second, annual average contract obligations awarded after effective competition increased by 14% as the overall platform portfolio decreased by 3%. The rise in shipbuilding competition contrasts with the fall in harder hit sectors, suggesting that declining rates of competition may be one of the consequences of the shock.

Figure 14 summarizes the CSIS platform portfolio analysis results.

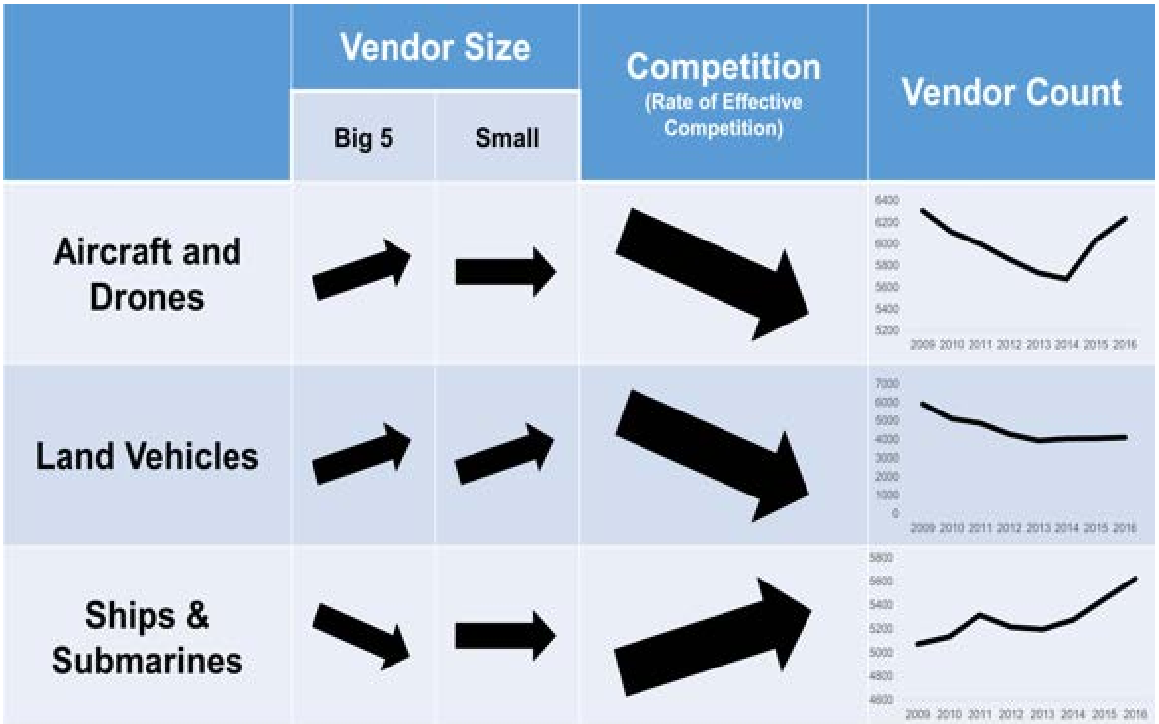


Figure 14. Summary of CSIS Platform Portfolio Analysis Results

Is the FSRS Data Available?

The data show that the FSRS database remains too incomplete to draw top-level trends across every platform portfolio category. Alarmingly, the rate of reporting also appears to be plateauing after early years of steady improvements. Of the three platform portfolios analyzed above, Aircraft and Drones and Land Vehicles had roughly the same availability at around 25 to 30% of prime contract obligations dollars appearing in FSRS in recent years. For Ships & Submarines, only 10 to 20% of contract obligation dollars appeared in FSRS in recent years. These three platform portfolios also represent three of the more complete platform portfolios with regards to data availability. Only the Missile and Space Systems and Weapons and Ammunition platform portfolios have relatively similar shares of prime contract obligations in FSRS.

The data show that the FSRS database is inadequate for top-level trend analysis; the data presented above show that certain sub-sectors are more mature than others, even within the same platform portfolio. For example, in the Aircraft and Drones platform portfolio, the Navy had much better coverage in FSRS than either the Air Force or the Army. While FSRS data may be inadequate for the top-level analysis, the maturity of certain sub-sectors makes analysis of the subcontracting trends in that sub-sector possible. When combined with data from FPDS, other mature sub-sectors, and qualitative interviews, it is possible to glean important insights into the sub-contracting dynamics in action in various parts of the defense industrial base.

References

Alfieri, J., Anderson, D., Bailey, J., Barras, W., Boran, M., Fowler, D., … Johns, B. (2014). *Spring 2014 industry study final report private sector support & services industry.*

Banham, R. (2013). Contractors are consolidating in anticipation of drawdowns. *CFO*.

Banjo, S., & Kalita, M. (2010, February 2). Once-robust charity sector hit with mergers, closings. *Wall Street Journal*.

Berteau, D., McCormick, R., & Sanders, G. (2014). Defense contracting trends by platform portfolio. In *Proceedings of the 11th Annual Acquisition Research Symposium* (Vol. II, pp. 129–143). Monterey CA: Naval Postgraduate School. Retrieved from http://www.acquisitionresearch.net/files/FY2014/NPS-AM-14-C11P14R01-054.pdf

Bertuca, T. (2014, September 3). Kendall highlights industrial base areas where he sees most risk. *Inside Defense*. Retrieved from https://insidedefense.com/inside-pentagon/kendall-highlights-industrial-base-areas-where-he-sees-most-risk

Bossaller, J. S., & Kammer, J. (2009, January). On the pros and cons of being a small firm in an economic downturn. *Small Business Update*, (61), 1–3.

Defense Procurement and Acquisition Policy. (n.d.). *Department of Defense competition report for FY 2015.* Retrieved from http://www.acq.osd.mil/dpap/cpic/cp/docs/DoD_FY_2015_Competition_Report.pdf

DoD. (2014a). *Guidelines for creating and maintaining a competitive environment for supplies and services in the Department of Defense*. Retrieved from http://bbp.dau.mil/docs/BBP%202-0%20Competition%20Guidelines%20(Published%2022%20Aug%202014).pdf

DoD. (2014b). *Quadrennial defense review 2014*. Washington, DC: Author. Retrieved from http://www.defense.gov/pubs/2014_Quadrennial_Defense_Review.pdf

Drew, C. (2011, March 15). Northrop to spin off shipyards. *The New York Times*. Retrieved from http://www.nytimes.com/2011/03/16/business/16ship.html

Eaglen, M. (2012, October). Defense sequestration targets small business. American Enterprise Institute.

Gholz, E., & Sapolsky, H. M. (2000). Restructuring the U.S. defense industry. *International Security, 24*(3), 5–51.

Gulati, R., Nohria, N., & Wohlgezogen, F. (2010). Roaring out of recession*. Harvard Business Review, 88*(3), 62–69. Retrieved from http://doi.org/BookReview

Guo, J. (2014, November 26). Why was the recession so much worse for small businesses? Blame lending. *The Washington Post*.

Homan, T. (2014, April). Defense cuts increase risks for smallest contractors. *The Fiscal Times.*

Hunter, A., McCormick, R., Ellman, J., Cohen, S., Johnson, K., & Sanders, G. (2017). *Defense acquisition trends, 2016: The end of the contracting drawdown*. Washington, DC.

Kambil, A. (2008). What is your recession playbook? *Journal of Business Strategy, 29*(5).

Kovacic, W. E., & Smallwood, D. E. (1994). Competition policy, rivalries, and defense industry consolidation. *Journal of Economic Perspectives, 8*(4), 91–110. http://doi.org/10.1257/jep.8.4.91

Lai, Y., Saridakis, G., Blackburn, R., & Johnstone, S. (2016). In a recession , large firms are more likely than SMEs to resort to personnel cuts. *LSE Business Review.*

Moore, N. Y., Grammich, C. A., & Mele, J. D. (2014). *Findings from existing data on the Department of Defense industrial base*. Santa Monica, CA: RAND. Retrieved from http://www.rand.org/pubs/research_reports/RR614.html

Rhodes, D., & Stelter, D. (2009). Seize advantage in a downturn. *Harvard Business Review, 87*(2). http://doi.org/Article

Samuelsohn, D. (2013, April). Key lawmakers say the situation seems bad enough to warrant a closer look. *Politico*.

Scully, M. (2011, July). Upcoming Pentagon budget cuts worry defense contractors. *National Journal*. Retrieved from http://doi.org/http://dx.doi.org/10.1108/17506200710779521

Sivy, M. (2012, January 18). The big winner of the great recession is … *TIME*.

Thompson, L. B. (2010, June). Defense industry consolidation: This time will be different. *Lexington Institute*.

Disclaimer

The Center for Strategic and International Studies (CSIS) does not take specific policy positions; accordingly, all views expressed in this presentation should be understood to be solely those of the author(s).

THIS PAGE INTENTIONALLY LEFT BLANK

Panel 5. Contemporary Issues in Contracting

Wednesday, April 26, 2017	
1:45 p.m. – 3:15 p.m.	**Chair: Brigadier General Cameron G. Holt, USAF,** Commander, Air Force Installation Contracting Agency ***Overseas Contingency Operations Contracts After Iraq: Enabling Financial Management Research and Transparency Through Contract Labeling*** Gregory Sanders, Center for Strategic and International Studies Andrew Hunter, Center for Strategic and International Studies James Ruedlinger, Center for Strategic and International Studies Shivani Pandya, Center for Strategic and International Studies Michael Ridings, Center for Strategic and International Studies ***Use of Incentives in Performance-Based Logistics Contracting: Initial Findings*** Andrew Hunter, Center for Strategic and International Studies Jesse Ellman, Center for Strategic and International Studies Andrew Howe, Center for Strategic and International Studies ***Price Analysis on Defense Logistics Agency (DLA) Contracts*** Janie Maddox, Naval Postgraduate School Ralucca Gera, Naval Postgraduate School

Brigadier General Cameron G. Holt, USAF—is the Commander, Air Force Installation Contracting Agency, Air Force Installation and Mission Support Center, Air Force Materiel Command, Wright-Patterson Air Force Base, OH. He leads an over 700-personnel agency with a total contract portfolio of $33 billion. In this capacity, he directs enterprise-wide installation strategic sourcing efforts for the Air Force and oversees $3.9 billion in annual obligations in mission and installation requirements. His contracting authority extends worldwide across AFICA in support of eight major commands and their 77 units. Additionally, he is designated as the Commander of a Joint Theater Support Contracting Command upon activation. General Holt also directs the contract execution in support of the Defense Technical Information Center, Air Force Medical Support Agency, Air Force Medical Operations Agency, and Air Force Civil Engineer Center.

Prior to this assignment, he served as the Director of Staff, Air Force Installation Contracting Agency, Air Force Installation and Mission Support Center, Air Force Materiel Command, Wright-Patterson AFB, OH. General Holt oversaw a team of 87 officers, enlisted and civilian personnel who coordinated the agency workflow, policies, procedures, resources, and directives for 77 units worldwide with over 800 personnel. Additionally, he led the team responsible for finding strategic sourcing opportunities within the Air Force.

General Holt received his commission through the Reserve Officer Training Corps at the University of Georgia in 1990. He has experience in the full spectrum of acquisition and contract management across four major commands, Headquarters U.S. Air Force, U.S. Air Forces Central Command, and the Joint Staff. General Holt is a joint qualified officer with multiple deployments in support of Operation Enduring Freedom.

Overseas Contingency Operations Contracts After Iraq: Enabling Financial Management Research and Transparency Through Contract Labeling

Greg Sanders—is a Fellow in the International Security Program and Deputy Director of the Defense-Industrial Initiatives Group at CSIS, where he manages a research team that analyzes data on U.S. government contract spending and other budget and acquisition issues. In support of these goals, he employs SQL Server, as well as the statistical programming language R. Sanders holds an MA in international studies from the University of Denver and a BA in government and politics, as well as a BS in computer science, from the University of Maryland. [gsanders@csis.org]

Andrew Hunter—is a Senior Fellow in the International Security Program and Director of the Defense-Industrial Initiatives Group at CSIS. From 2011 to 2014, he served as a senior executive in the Department of Defense, serving first as chief of staff to under secretaries of defense (AT&L) Ashton B. Carter and Frank Kendall, before directing the Joint Rapid Acquisition Cell. From 2005 to 2011, Hunter served as a professional staff member of the House Armed Services Committee. Hunter holds an MA degree in applied economics from the Johns Hopkins University and a BA in social studies from Harvard University. [ahunter@csis.org]

Contributing Authors: James Ruedlinger, Shivani Pandya, and Michael Ridings

Abstract

Contracts relying on crisis funds (including emergency funds) may bypass many safeguards built into normal spending processes. This study examines the literature on how these contracts are fulfilled for both civilian and defense crisis funds, primarily focusing on the American Recovery and Reinvestment Act (ARRA), disaster funds, and Overseas Contingency Operations (OCO) funds, beginning with contracts awarded in 2012 and using publicly available data.

This paper discusses the challenges and contradictions that make identifying OCO-funded contracts difficult and then presents a methodology for classifying them. The paper then analyzes trends in contracting from the post-Iraq withdrawal period. This analysis focuses on three areas where the literature review showed that crisis contracting diverges from conventional contracting: noncompetitive awards, undefinitized contract actions, and reachback contracts. The dataset created for this study will be made publicly available to allow for analysis of this data by other researchers and to close an important transparency gap.

Introduction

Contracting during a crisis is replete with challenges. Speed and flexibility are essential because delay means that urgent needs go unmet. However, uncertainty is commonplace, whether the crisis is prompted by natural disasters, military conflicts, or economic disturbances. These conditions are vulnerable to the infamous trifecta of waste, fraud, and/or abuse, but even setting those extremes aside, many justifiable crisis contracts cannot or should not be sustained in ordinary times.

This century has already seen a range of high-profile crisis contracting: contingency contracting during the invasion and subsequent occupation in Afghanistan and Iraq, the American Recovery and Reinvestment Act (Recovery Act) pursuit of shovel-ready projects in response to the global financial crises, and government responses to range of disasters such as Hurricane Katrina. Important work has been done to provide oversight and transparency by the Government Accountability and Transparency Board and the Commission on Wartime Contracting, as well as inspectors general (IGs) and others.

However, when the news moves on to a new set of crises and the final reports are filed, lessons identified in one domain may never be transferred to another. Worse yet, as attention fades, there is risk of backsliding because it becomes increasingly challenging to determine whether recommendations were followed and whether they succeeded in mitigating the risks that drove reform efforts. This paper is focused on Overseas Contingency Operations (OCO) funded contracting after the initial withdrawal from Iraq, a period that benefits from efforts by the DoD to improve data transparency; this situation is also comparatively understudied, in no small part, because of the opaqueness and ambiguity surrounding the OCO budget.

While this portion of the project is focused specifically on Department of Defense contracting, the study team has conducted a literature review that also includes studies of civilian efforts, such as the Recovery Act and disaster response efforts. Despite their differences, the many concerns about crisis contracting apply across domains. Likewise, the publicly available Federal Procurement Data System (FPDS) provides a common window through which these distinct crisis contracts cases can be observed and compared.

Following the literature review, this paper discusses the challenges and contradictions that make identifying OCO-funded contracts difficult and then presents a methodology for classifying them.[1] The paper then analyzes trends in contracting from the post-Iraq withdrawal period. This analysis focuses on three areas where the literature review indicated that crisis contracting diverges from conventional contracting: noncompetitive awards, undefinitized contract actions, and reachback contracts. The paper concludes by summarizing initial findings from the contingency contracting dataset.

What Is Contingency Contracting?

Handling crises is an important part of the job of the United States military, so it comes as no surprise that there are explicit legal categories for crisis contracting. McMillon (2000) provided a helpful glossary, including contingency contracting itself:

> Direct contracting support to tactical and operational forces engaged in the full spectrum of armed conflict and Military Operations Other Than War, both domestic and overseas. It includes Major Regional Conflicts, Lesser Regional Conflicts, Military Operations Other Than War, and Domestic Disaster/Emergency Relief. (pp. 5–7).

This paper also includes a similar category of operations that fall under a different portion of the U.S. legal code: “humanitarian or peacekeeping operations.”[2]

The U.S. government extensively relied on contingency contracting after the 9/11 attacks and the wars in Afghanistan and Iraq. This was not a new phenomenon; with the move to an all-volunteer military, contractors had an important role to play from the Gulf War

[1] The study team has published a dynamic web tool for visualizing OCO-funded contracts starting in 2012. For researchers wishing to replicate the results of this study or conduct their own research, the study team is sharing a complete list of the procurement identifiers and key characteristics of contracts in the dataset.

[2] For the full definition of contingency operations, see 10 U.S.C. 101(a)(13). For the full definition of humanitarian operations and peacekeeping, see 10 U.S.C. 2302(8).

to the war in Kosovo.[3] Nonetheless, the wars in Afghanistan and Iraq and the subsequent occupations prompted steady increases in spending on contingencies. From 2002 until 2008 approximately, $159 billion in contracts were awarded in contingency contracts (GAO, 2012). Specifically, emergency supplemental appropriations, which later evolved into the OCO budget, rapidly grew and focused on difficult to predict wartime expenses including contingency contracts. As Sharon Pickup and Asif Khan (2009) noted, this growth continued in 2007 when the "DOD revised its Financial Management Regulation, expanding the definition of acceptable maintenance and procurement costs and directing the military services to begin including 'longer war on terror' costs in their OCO funding requests" (p. 11).

The tide turned as the Iraq war wound down and President Barack Obama took office in 2009. The GAO had already encouraged the DoD to "shift certain contingency costs into the annual base budget to allow for prioritization and trade-offs among DoD's needs and to enhance visibility in defense spending" (Pickup & Khan, 2009, p. 7). The changes are described in Table 1. The Budget Control Act, implemented by Congress in 2011, reversed the trend of transferring OCO funds into the base budget request (Epstein & Williams, 2017). The BCA caps limited the funds available in the base budget, but OCO funds were not subject to caps. As a result, there was an opportunity and temptation to use OCO spending to supplement the forced decreases in the base budget (Epstein & Williams, 2017). While the OCO budget has de facto not always been limited by these definitions, the study team employs the Fiscal Year (FY) 2010 guidelines as part of contract labeling because they are compatible with a specific focus on crisis-funded contracts, rather than longer term and more persistent efforts.

[3] See McMillon (2000, pp. 13–23) for a summary of contracting operations in the 1990s and some of the challenges encountered.

Table 1. Fiscal Year 2010 OMB Guidance on What Qualifies as OCO Spending
(Pickup & Khan, 2009, p. 14)

Area	Prior OCO Funding Guidance	FY2010 OCO Funding Guidance
Geographic Theater of Operations	Does not specify locations, which allowed for funding such items as home station needs to support contingency operations.	Includes U.S. Central Command, the Horn of Africa, the Indian Ocean, and the Philippines, among others.
Equipment	Does not specify obligation time frames.	Specifies stricter definitions of replacement, repair, modification, and procurement of equipment; new criteria specify a 12-month time frame for obligating funds.
Research, Development, Test, and Evaluation (RDT&E)	No time frame restrictions.	Funding for research and development must be for projects required for combat operations in the theater that can be delivered in 12 months.
Personnel	Included pay and allowances for end strength above level requested in budget.	Excluded.
Family Support Initiatives	Included family support initiatives that would endure after U.S. forces redeploy to home stations.	Excluded.
Base Realignment and Closure	Included.	Excluded.

Literature Review

Regulatory Environment

With crisis funding continuing to grow to compensate for BCA caps, it is important to conduct a thorough review of the positive and negative aspects of crisis contracting. Both civilian and military crises covered by this paper share a key trait: time is of the essence. When a national emergency is present, or an impending military conflict requires rapid acquisition, the typical procedures defined by regulation can become a hindrance. Without the ability to bypass them, the regulations could prevent many solutions from being implemented within the time frame driven by the crisis (Britt & Miles, 1985). In anticipation of this problem, acquisition regulations offer a range of exceptions to allow for the speed of acquisition called for by crisis situations. However, this approach inherently leads to concerns that contingency contracts do not operate in the same environment of the standard federal contracting process (McMillon, 2000).

Regulatory Exemptions

Competition has been longstanding in its presence within federal procurement practices. The Competition in Contracting Act of 1984 (CICA) requires that procurements must enter into a full and open competition (Manuel, 2011). However, CICA also designates specific exemptions to competition requirements. CICA establishes seven instances when a contracting officer may engage in a noncompetitive procurement process (Manuel, 2011). Included within these exemptions are circumstances for unusual and compelling urgency, national security, and contracts necessary for the public interest (Manuel, 2011). Likewise, during a natural disaster, funds for procurement of services may disregard competition in cases of “urgent and compelling” situations (GAO, 2015).

In addition to the option to bypass full and open competition, contingency contracts are currently exempt from the requirement restricting undefinitized contracts and from having to wait until a protest is resolved to award emergency requirements (McMillon, 2000). Other exemptions simply involve raised thresholds. In 2000, the simplified acquisition threshold was twice as high for contingency contracts, increasing from $100,000 to $200,000 (McMillon, 2000, pp. 9–10). For other parts of government, crisis measures may allow for greater use of forms of contracting that the DoD already regularly relies on, such as cost-based contracts. Within the first reporting to Recovery.gov, the Recovery Act spent $7.8 billion on contracts that were noncompetitive or were not fixed price (Lipowicz, 2009).

Limitations on Crisis Contracting

However, while crisis contracting may employ a range of regulatory exemptions, it faces heightened scrutiny in other areas, particularly time frames. Certain crises may have even shorter time frames, depending on their expected duration. During Operation Restore Hope in Somalia, contracts were limited to 90 days (McMillon, 2000, pp. 16–18). These limitations are a measure to reduce the period that the United States is committed to deals that are hastily made by necessity. Without the competition requirements, crisis funding runs the risk of being unable to validate presented contract data with competitors or government sources (Dodaro, 2009). In 2009, time limitations were extended to all contracts using the urgency exemption:

> In 2008, the Duncan Hunter National Defense Authorization Act for Fiscal Year 2009, Pub. L. No. 110-417, § 862, amended certain laws to require that contracts awarded using the urgency exception not exceed the time necessary to meet the unusual and compelling requirements and for the agency to enter into another contract, and may not exceed 1 year unless the head of the agency determines exceptional circumstances apply. (Marvin, 2014, p. 13)

An area of dispute within the policy literature is whether these restrictions should be further institutionalized. The Commission on Wartime Contracting (2011) was severely critical of non-competed contracts extended without competition, even if the original contract was competed:

> $36.3 billion Defense (Army) LOGCAP III contract—The Army has awarded a number of contracts under its worldwide Logistics Civil Augmentation Program (LOGCAP). Of these contracts, the largest is the LOGCAP III contract supporting the wars in Iraq and Afghanistan. The base contract for LOGCAP III was awarded competitively, but lasted for 10 years without competition on any of its task orders. … As sole provider, without the discipline of task-order competition, KBR proposals included large amounts of questioned and unsupported costs identified by the Defense Contract Audit Agency (DCAA). (p. 75)

The Contingency Contracting Reform Act did not become law, but it usefully illuminates arguments on these issues. Section 201 of that bill sought to limit the duration of contingency contracts across the board by default. The bill would have limited contingency contracts that were not competed or that received only one offer to one year and competed contracts to three years (GAO, 2012). The Professional Service Council, a government services industry association, objected to the proposal on multiple grounds. Their primary point was that even in contingency contracting, shorter does not necessarily mean better:

> Primarily, the limitation on contract length fails to recognize the benefits and efficiencies that can be achieved by longer contract lengths. One of the key

> lessons learned from the Special Inspector General for Iraq Reconstruction was that short periods of performance significantly increased the contract price and added to the government's burden to award new contracts and administer existing ones. (Professional Services Council, 2012, p. 6)

Negative Outcomes of Crisis Contracting

Regulatory exceptions and limitations on contracting officers are worth studying, but it is the outcomes of crisis contracting that have drawn so much negative attention to the area. The first challenge is that the circumstances and requirements limit the ability to confirm contract, grant, or loan information prior to the disbursement of funds (Dodaro, 2009). The Commission on Wartime Contracting (2011) also raised this issue with recommendation 11 which cited a need to "improve contractor performance-data recording and use" (p. 10).

This challenge can extend over the entire life of these contracts. Crisis funding for natural disasters can lead to increased levels of incomplete documentation, a lack of contract closeouts, and little to no evidence of higher level contract reviews (GAO, 2015). An example of this is when hotels received contracts to house those affected by the disaster during Hurricane Sandy. The hotels received noncompetitive contract awards through the urgent need justification, but the joint field contracting offices were often left unaware of these contract awards until the contract was closed out and they received the vendor invoices (GAO, 2015).

Due to the urgency and need for a significant number of contracts in a short period of time, the contract closeouts can often become backed up and delay the documentation from being properly completed (GAO, 2015). A portion of these contracts require further approval from a level above the contracting officer. Of the nine contracts reviewed that required this approval, the GAO found only one that had received the appropriate justification (GAO, 2015). The Recovery Act, with its emphasis on oversight and considering the comparatively straightforward operating environment of an economic crisis, gives a sense of what the baseline failure rate may be for crisis contracting. Within the grants and contracts awarded to broadband services under the Recovery Act, 14% were terminated before they were completed (Goldstein, 2014). When these contracts and grants are terminated or sustained with cost overruns, the lost funds can present a larger issue to the efficiency of crisis funds being awarded for stimulus purposes (Goldstein, 2014).

Worse yet, as Comptroller Gene Dodaro (2009) succinctly put it, "experience tells us that the risk for fraud and abuse grows when billions of dollars are going out quickly" (p. 6). Compounding the challenges of gaps in documentation, staff are exposed to higher rates of fraud without the ability to conduct system edit checks or time to identify problems prior to disbursement of funds (Dodaro, 2009).

Specifically within contingency contracting, fraud has been a very present issue (Gordon, 2014). Operating under a time-stressed environment where the need for a solution is overwhelming can create many opportunities for fraud (Gordon, 2014). Citing specific numbers for waste and fraud is always controversial, and subjective determinations of what constitutes waste can easily overshadow cases of outright corruption or criminality. Nonetheless, the magnitude of these challenges is tremendous, as the Commission on Wartime Contracting (2011) argued that "at least $31 billion, and possibly as much as $60 billion, has been lost to contract waste and fraud in America's contingency operations in Iraq and Afghanistan" (p. 1).

Past Reform Efforts Have Led to Increased Transparency

Due to its inherent challenges, crisis contracting is an area where regulation and practice steadily evolve in reaction to past challenges. As with defense acquisition writ large, there will likely be no final equilibrium solution, but instead the system will evolve and reprioritize in response to the successes, or more often the failures, of past efforts. However, ongoing challenges do not mean that reform efforts were fruitless. This richness of data that enables this study is possible in no small part because of past reforms. The Recovery Act set a high standard for transparency, with President Obama insisting "every taxpayer dollar spent on our economic recovery must be subject to unprecedented levels of transparency and accountability" (Gaffney & Berger, 2009, p. 1). While disagreements about the Recovery Act persist, after stimulus funds were dispersed, Sam Rosen-Amy of OMB Watch argued, "I think it helped show Congress that there is a use for and a need for more information on where federal money is going and how it's being used" (Holeywell, 2012, p. 2).

The DoD has also made great strides in tracking crisis contract data, with financial tracking systems and contingency contract databases such as the "Synchronized Pre-deployment and Operational Tracker (SPOT)" (Swan, 2012, p. 17). However, unlike FPDS or the Recovery Act dataset, those tools are not available to the public. Improving the ability of contracting officers and others within the government to make more informed award decisions and track contract performance plays an important role in mitigating the data gaps that can mask problems. However, in Laura Dickinson's (2011) book on wartime contracting, she explained why the benefit of transparency regarding contracts is of direct interest to the public:

> As this example [regarding a Dyncorp Police Training Contract] illustrates, foreign affairs contracting raises serious concerns about public participation and transparency (which for simplicity's sake I will often refer to collectively as public participation). Significantly, public participation is simultaneously a value in and of itself—reflecting the view that people affected by an activity should have some input into how that activity is carried out—and a mechanism for either accountability or constraint. For example, if various populations are able to participate in the formulation and critique of future plans of action, such participation may well impact the actions ultimately undertaken. Just as contractual arrangements may be structured to protect and promote public law values, so too public participation may be harnessed to restrain governments from abuses and help to protect other public values, such as human dignity and anticorruption. (p. 104)

There are logical reasons for the different levels of public transparency between the Recovery Act's public dataset and the restricted tools such as SPOT. First and foremost, sharing too much data when operating in conflict environments could reveal operational details that place U.S. personnel, vendors, or the civilian population in danger. In addition, this public participation role is partially fulfilled by Inspector Generals, and the Special Inspector General for Afghanistan has remained active during the study period. Nonetheless, Dickinson's argument suggests that there is value in making the vetted and sometimes anonymized contingency contracting data more accessible, in no small part because "governments may outsource foreign affairs precisely to avoid oversight" (Dickinson, 2011, p. 105).

Factors That Aggravate or Mitigate the Risk of Crisis Contracting

The prior sections have touched on a range of the ways in which crisis contracting operates in a unique operational and regulatory environment. During the review, the study team evaluated various factors apparent in regular contract reporting that aggregate or mitigate the inherent risks of crisis contracting. Three key criteria were applied: do multiple sources, ideally in multiple domains, point to this factor as a significant source of risk; is this factor something at least partially under the U.S. government's control; can it be tracked using FPDS? By these criteria, three factors stood out: the risks of noncompetitive awards, the risk of UCAs, and the opportunity for expeditionary contracting offices to support home contracting offices, called reachback contracting.

Noncompetitive Awards

The option to bypass competition for urgency reasons is one of the better documented aspects of crisis contracting. From 2010 until 2012, only 3% of the DoD's contracts were awarded in a noncompetitive environment under the urgency exception, but this 3% still accounts for $12.5 billion worth of funds. During this same time, State's contracting efforts under contingency contracting accounted for 12.5% of contract awards (Marvin, 2014). An early report after the Recovery Act debuted reported that at least $7.8 billion was awarded to noncompetitive contracts (Lipowicz, 2009). That said, this use of noncompetitive contracts in part was a result of relying on existing contracts. Of the 32% of new contracts that were awarded through the Recovery Act, 11% were awarded without competition (Needham, 2010). That said, these numbers should be put in context of the range of other forms of noncompetitive contracting employed by the government. In 2013 alone, 36% of funds for procurement of goods and services, approximately $164 billion, were not competed (Marvin, 2014).

Trade Off Between Speed and the Benefits of Competition

The rate of competition for crisis-funded–contracting is not unusually high; instead, critics emphasize noncompetitive contracts because competition is often more important in a crisis. Higher prices can qualify as reasonable in disaster relief contracting, due to the significant and immediate increase in demand for a product offered by a contractor. Relief items in a natural disaster experience such high demands that prices significantly increase on goods such as water, lumber, and generators (Gordon, 2014). Marvin (2014) extended this finding to other forms of crisis contracting, arguing that "promoting competition—even in a limited form—increases the potential for quality goods and services at a lower price in urgent situations" (p. 1). In addition to the risk of higher prices or lower quality, noncompetitive contracts are also at greater risk of misconduct when compared to the standard procurement process (Manuel, 2011).

Of course, the challenge is that competition does not necessarily create quickly. For contingency contracting, delays can undermine a unit's effectiveness, morale, and ability to complete its mission (McMillon, 2000). Likewise, for the sake of the affected population in a natural disaster, the need to provide goods and services as soon as possible is of utmost importance (Mackin, 2015). While economic recession presents an easier operating environment, considering the primary goal of Recovery Act was to act quickly on high priority needs, contracting officers relied heavily on avenues that presented the fewest opportunities for competition to arise (Needham, 2010).

Urgency is also not the only constraint on competition. Built into the Recovery Act were guidelines specific to small business programs, which effectively encouraged the use of noncompetitive contracts to ensure they had equal opportunities to receive assistance. In May 2010, approximately 80% of the noncompetitive contracts were awarded to small businesses through these guidelines (Needham, 2010). Similarly, natural disaster contracting further allows a preference in noncompetitive contracts for local area firms in the affected area which can aid in economic recovery (Gordon, 2014). Bontjer, Holt, and Angle

(2009) applied this idea to contingency contracting when they studied the impact of such measures in Afghanistan:

> Using local goods and services to carry out project work, for instance, allows a development dollar to be spent twice—providing much needed services to Afghan citizens and communities while simultaneously creating jobs, generating revenue, and promoting a more sustainable marketplace—all of which can ultimately reduce the likelihood of a relapse into conflict. (p. 39)

Competition advocates do acknowledge competing needs, but given the benefits of even limited competition, they nonetheless urge prioritizing maximizing competition within those constraints (Office of Inspector General, 2016). This approach is mandated by the Federal Acquisition Regulations (FAR) which allows for urgency exceptions but still requires contracting officers to solicit responses from as many contractors as possible under these circumstances (Gordon, 2014). In the case of disaster relief, such regulations are not always followed. After a new competitive requirement was enacted, FEMA contracting officers reported that they were still instructed to treat every disaster relief contract as urgent and could therefore award contracts without competition. This problem created an opportunity for $32 million of procurement costs going unreported in noncompetitive disaster relief contracts in FY2013 (Mackin, 2015). The Commission on Wartime Contracting (2011) similarly believed that there was room for more competition, and it proposed the government should "set and meet annual increases in competition goals for contingency contracts" (p. 10).

Duration Limits on Noncompetitive Contracts

As was discussed in the section *Limitations on Crisis Contracting,* noncompetitive contracts that use the urgency exception are limited only one year to reduce the risk of overspending (Marvin, 2014; Office of Inspector General, 2016). With that said, the cost and benefits of a shorter contract are disputed (Professional Services Council, 2012). Reform efforts after Hurricane Katrina resulted in an even stricter 150-day limit to disaster relief contracts awarded in a noncompetitive environment (Mackin, 2015). Contingency contracting, on the other hand, is allowed to award contracts for up to a year in a noncompetitive environment (Office of Inspector General, 2016).

Upon the GAO's review of noncompetitively awarded contracts, more than half exceeded the 150-day time limit. This is not necessarily a problem; the agency can waive that requirement under certain conditions. However, in each of these contracts that violated the time limit, FEMA did not approve the extension, and some went beyond the regulation by a year and a half (Mackin, 2015).

Undefinitized Contract Actions

Undefinitized contract actions (UCAs) are a type of contract that differs from standard procurement methods in allowing the production to start without defining all the terms of the contract (Actions, 2017).[4] In crisis funding situations, these contracts can be seen as advantageous because they allow the production of goods or the allocation of services to be immediately received (Federal Audit Executive Council, 2010). Circumstances created by crisis funding certainly qualify as circumstances of urgent need that can allow for

[4] Letter Contracts are a subset of UCAs in that they specifically seek to start production of the goods immediately (Calvaresi-Barr, 2007).

UCAs (Marvin, 2014). Contingency contracts often utilize UCAs, and they can be coupled with the risk of awarding them without competition. Contingency contracts often utilize UCAs which can be coupled with the risk of awarding them without competition (Federal Audit Executive Council, 2010).

UCAs are entered under cost reimbursement contracts until later defined. This allows the vendor to be reimbursed for all reasonable costs within the procurement up until the point of defining the contract terms (Federal Audit Executive Council, 2010). While the initial award of the contract can be obligated without the terms set, the FAR still requires that within 180 days or when 40% of the work has been completed that the contract terms must be defined (Federal Audit Executive Council, 2010).This allows the vendor to be reimbursed for all reasonable costs within the procurement process up until the point of defining the contract terms (Federal Audit Executive Council, 2010). The vendor, not the customer, is responsible for determining a “reasonable” price for this initial work (Calvaresi-Barr, 2007). UCAs are to have, at the least, a “not to exceed” price amount stated at the beginning. However, upon awarding the UCA, up to 50% of the “not to exceed” amount can be paid without any approval or review (Calvaresi-Barr, 2007).

Unfortunately, UCAs also create a very high risk of overpaying for goods and services and at times make the contracting officer beholden to the vendor (Commission on Wartime Contracting in Iraq and Afghanistan, 2011). In cases of disaster relief contracting, they present an even higher risk of cost overruns. Gordon (2014) mentioned that when natural disasters occur, the price of needed materials significantly increases as the demand for these products skyrocket. Entering into a UCA through a noncompetitive award furthers the risk of the government overpaying for needed goods and services to provide relief to the affected areas (Gordon, 2014).

Historically, the use of UCAs presented high risk with contingency contract awards and led to schedule delays coupled with high cost overruns. The GAO reviewed 77 UCA awards for contingency contracting within the DoD and in 10 cases found that other contracting methods would have sufficed and promoted cost savings. In 2007, 60% of these cases DoD contracting officers failed to definitize contract award terms by the 180-day FAR regulation (Calvaresi-Barr, 2007). By 2008, although still a concerning number, the amount of cases failing to meet the definitize timeline decreased to 51%. Furthermore, the GAO found that out of 83 reviewed UCAs, 66 resulted in paying the awardee 45% or more of the not to exceed estimate at the award (Hutton, 2010).

From 2001 to 2005, obligations awarded under UCAs increased from $5.98 billion to $6.53 billion (Calvaresi-Barr, 2007).[5] UCA data collection was not centralized within the DoD, leading the DoD to have a significant lack of data to properly evaluate how much is truly being spent under UCA conditions (Calvaresi-Barr, 2007). Since 2007, the DoD has taken measures to require centralized reporting of UCAs, but in 2010, the GAO found that many UCAs are not being properly reported to the centralized offices (Hutton, 2010). On average, DoD UCA contracts overran the 180-day definitization requirement by two months (Calvaresi-Barr, 2007). The Air Force was the only branch at the time to have requirements to report UCAs, but despite reporting requirements, the GAO found nine UCA contracts in

[5] It is important to note that these costs do not include obligations awarded to undefinitized task order contract or UCA modifications (Calvaresi-Barr, 2007).

the Air Force that overran the 180-day requirement by at least a full year (Calvaresi-Barr, 2007). A majority of the UCA contracts were awarded to maintain program schedules and directly and indirectly support war efforts (Hutton, 2010). While there is the opportunity to waive the 180-day requirement, the GAO found only two of the contingency contracts that met the requirements necessary to waive the regulation in 2007 (Calvaresi-Barr, 2007).

The Office of the Inspector General had similar findings to the GAO on UCAs in 2012. Out of 251 UCAs reviewed, the Inspector General's Office found that 132 cases failed to meet the timeline for definitization (Office of Inspector General, 2012). 118 of the cases highlighted noncompliance with requirements on the impact of allowable profit on the undefinitized period (Office of the Inspector General, 2012). 64 of the cases resulted in an obligation of funds significantly above the allowable amounts (Office of the Inspector General, 2012).

Reachback Contracting

At least as important as the methods used in contingency contracting are the contracting officers charged with managing the system. McMillon in 2000 reviewed four different military contingencies since the end of the Cold War and found that "consistent problems for all components during contingencies have been the lack of experienced personnel, restrictive regulations, and a lack of proper supplies such as computers and contracting SOPs and forms" (McMillon, 2000, p. 23). The 9/11 attacks and the subsequent wars in Afghanistan and Iraq were dramatically different operating environments than the prior decade's humanitarian operations or even the first Gulf War. Nonetheless, in 2011, the Commission on Wartime Contracting (2011) reached similar conclusions, recommending that the government "provide adequate staffing and resources, and establish procedures to protect the government's interests" (pp. 4, 11).

Given the inherent challenges of deploying people and resources to the field, one straightforward approach to this problem is to rely on those not on the battlefield. One prominent implementation of this idea is reachback contracting, a unique method that allows contracting officers in the field to "reachback" to domestic contracting offices for contracting support in contingency operations (Dunn, 2016). In 2007, the Reachback Division was originally set up to offer contracting support to those in theater in Kuwait (Adrian, 2010). This idea was not entirely novel, particularly in the later stages of an operation. McMillon (2000) noted that the Air Force instructed contracting officers to "consider support from the unit's home base" in addition to a range of other options outside of the deployment area (p. 34). Within three years, the division grew to a team of 62 people supporting contracting officers in the field in Afghanistan, Iraq, Kuwait, and Qatar (Adrian, 2010). After years of successful trials and results, the Reachback Division grew to include the Air Force and then added members from the Expeditionary Contracting Command Contingency Contracting Team (Adrian, 2010).

Reviews of this approach were positive. Commanding General Michael Hoskin of U.S. Army Expeditionary Contracting Command referred to reachback as a "very effective tool" in the contracting officers' arsenal (Dunn, 2016). Reachback contracting can result in fewer deployed contracting officers because the workload is shifted back to domestic contracting offices (Dunn, 2016). Utilizing reachback methods, contracting officers could improve their strategic buying and develop greater expertise within their source selection (Ausink, Castaneda, & Chenoweth, 2011). Furthermore, reachback contracting can provide continuity to workflow management and create better standardization for contingency contract reporting (Dunn, 2016).

Reachback's intention was to help ease the challenges faced by field contracting officers in attempts to support the warfighter (Calhoun & Larssen, 2013). Reachback is able to provide support in the Financial Services Division, Contracting Policy, Property Expertise, and the Army Sustainment Command Counsel (Calhoun & Larssen, 2013). Specializing in logistics, warehousing, transportation, base operations, security, counterinsurgency, telecommunications service, and supply acquisitions, reachback provides needed support to contingency contracting (Calhoun & Larssen, 2013).

In its review of reachback capabilities, RAND found that most contingency contracting officers cared more about the advantages in workflow, standardization of requirements, and concentration of contracting expertise then the reduction in deployments (Ausink et al., 2011). Reachback contracting has the potential to lower costs and reduce risks by not having to incur the same transportation and hazardous duty pay (Ausink et al., 2011). Workflow continuity could help increase the efficiency as well, since the contracting officers do not experience the same amount of turnover that deployed CCOs experience (Ausink et al., 2011). Reachback has been used in a multitude of ways. From small commodity purchases to cradle-to-grave large contract support, reachback methods have been successfully implemented (Ausink et al., 2011).

Although reachback methods can be used in various applications, the RAND study noted that reachback provides the greatest benefit when used for commodities, highly technical items, the use of a government-wide purchasing card, theater-wide purchases, and long-term contracts (Ausink et al., 2011). Each of these areas received multiple sources of agreement and near universal government-wide support of benefitting from reachback practices (Ausink et al., 2011). In the case of urgent and local projects, many turned away from the benefits that reachback practices could offer (Ausink et al., 2011). Limitations can also arise from policies applied to specific contingencies: the Iraqi first and Afghan first policies prevented field contracting officers from utilizing reachback practices due to the local requirements (Ausink et al., 2011).

If reachback had been utilized for the areas mentioned above in FY2008, 40 field contracting officers would not have needed to have been deployed (Ausink et al., 2011). Beyond reducing deployments, reachback methods provide greater concentrations of contracting expertise and continuity of the contracting officials maintaining the contracts (Ausink et al., 2011). The RAND study concludes that when used in the appropriate categories, reachback can mitigate risk, save cost, and provide greater efficiency in contingency contracts (Ausink et al., 2011).

Identifying Contingency Contracts

The Commission on Wartime Contracting (2011) reported that spending on contracts and grants performed in Iraq and Afghanistan in support of operations in those countries was expected to exceed $206 billion through the end of FY2011 (p. 2). During that same period, transactions directly labeled as contingency contracts could only account for less than $30 billion in obligations. The study team identified three different ways that data fields available in FPDS could be used to classify contingency contracts.

- The Contingency, Humanitarian, or Peacekeeping Operation column, which has better coverage in earlier years and makes explicit reference to the relevant statutes.
- The National Interest Action field, which has seen more frequent use in recent years and includes designators for natural disasters as well as

contingencies. Only those Actions pertaining to contingencies by the U.S. military were considered labeled as contingency contracts.

- Some transactions indirectly label themselves by employing a waiver to Central Commercial Registry reporting requirements that are only available to deployed contracting officers deployed into a contingency. As a result, FPDS lacks Registry information for these transactions, but we do learn that they were conducted in a contingency environment.

Perhaps the most surprising characteristic of these data fields is how little they overlap with one another. Both identify tens of billions of contracts, yet less than a quarter of contracts are considered contingency efforts by both criteria. While these fields are contradictory and only captured a portion of the universe of contingency contracting, they were a valuable launchpad for creating the heuristic score by which other contracts were classified. This study uses a 10-point scale to label contracts, based on a range of contingency-related characteristics. For each of the criteria below, when the history of officially labeled contingency contracts being referenced covers the period from 2000 to 2016 and all obligation amounts are in constant 2015 dollars.

OCO Funding (Maximum 4 Points)

- 0 to 4 points: corresponding to the percent of the contract's funding account that was made up by enacted OCO spending. A funding account with no OCO funding would receive 0 points, a funding account with full funding provided through OCO would receive 4 points, those in between are rounded to the nearest whole number.

The study team relied on spreadsheets provided by the DoD comptroller to classify the percentage of OCO appropriation in each DoD funding account for each year. This data was then linked to the FPDS using treasury account fields, which have been reliably available in FPDS since 2012. These funding accounts are only partially standardized, requiring the team to make judgment calls on classification when treasury account symbols did not align.

Place of Performance and Contracting Office (Maximum 4 Points)

Table 2. Contracting Office and Place of Performance

Contracting Office	***Place of Performance***
• 3 points: Contracting office where 50% or more of obligations are officially contingency contracts. • 2 points: Office where 25 to 49% of obligations are labeled as contingency contracts. • 1 point: Office where 10 to 15% of obligations are labeled as contingency contracts. • -1 point: Office with at least 1 billion in obligations over the period but no labeled contingency contracts.	• 3 points: Contract being performed in Iraq or Afghanistan. • 2 points: Contract being performed in the Philippines, which are explicitly included in OMB's guidance thanks to joint counterterrorism efforts. • 1 point: Contracts being performed in a country included in the U.S. Military's Central Command area of responsibility. Much of the direct support for the wars in Iraq and Afghanistan came from infrastructure in neighboring countries. • -1 point: Contracts being performed domestically.

These two criteria work in tandem to overcome one of the main sources of ambiguity in the contract data. Bases in neighboring countries are often key to contingency efforts, but without proper labeling, it can be challenging to identify what activity any given contract is supporting. Past CSIS research efforts have also uncovered major expenditures being classified as performed far away from the battlefield. Specifically, a surprising amount of contracting obligations were "performed" in Switzerland to deliver supplies to Afghanistan.

The Product or Service Being Purchased (Maximum 2 Points):

- 2 points: Product or service codes with at least 25% of obligations labeled as contingency contracts.
- 1 point: Product or service codes with between 10 and 25% of obligations labeled as contingency contracts.
- 1 point: Product or service codes with at least $1 billion in labeled OCO expenditures from.
- -1 point: Product or service codes with at least $1 billion in total expenditures, but no OCO labeled obligations.
- -1 point: Contracts funded by Procurement or RDT&E accounts with an anticipated contract duration of more than 1 year. Such contracts are excluded from OCO funding by OMB guidance (Pickup & Khan, 2009, p. 14).

Many of the items categorized through this process are consistent with McMillon's list of "Examples of supplies … include bottled water, food, office and field supplies, construction items, repair parts, and medical supplies. Contracted services may include construction, laundry, food service, transportation, billeting, utilities maintenance, and sanitation services" (McMillon, 2000, p. 9). Unexpectedly, security services were not captured by these automatically generated lists. Private security contractors have been among the most controversial recipients of contingency contractors, and the Commission on Wartime Contracting (2011) recommended that they be "phased out for certain functions" (2011, p. 4).

Initial Results

After assigning scores to transactions during the study period, the team found that entries scoring 6 points or more reliably shared traits with contingency contracts. As the above graph shows, these contracts have reduced in their prevalence during the downturn but have also continued to shrink in value even as total contracting began to rebound in 2016.

One of the larger drivers of this decline has been the continued reduction in contract spending, first due to the withdrawal from Iraq and then to a lesser degree reinforced as the footprint of U.S. operations in Afghanistan was reduced. OCO spending has not gone away during this period, and U.S. operations in Iraq have resumed. However, Figure 1 shows the same pattern reduction with confirmed OCO contracts and those identified via the heuristic method.

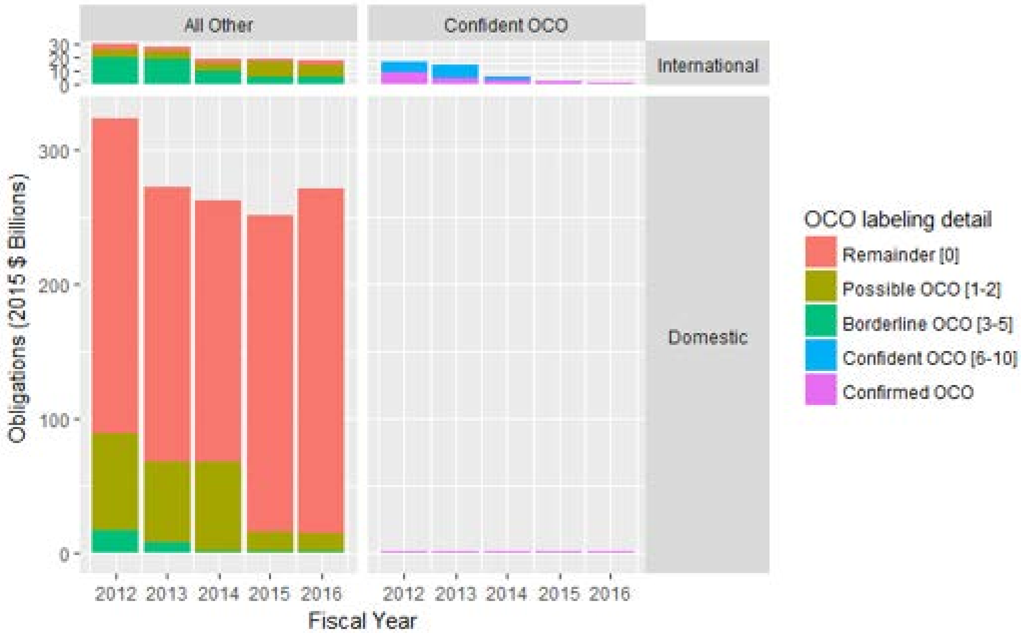


Figure 1. Contract Spending by OCO Evaluation and Place of Performance

Analysis of competition trends within the contingency dataset confirms findings from the literature review. While competition is an area of concern for contingency contracts, crisis-funding has not prevented contracting officers from already achieving higher rates of competition than across facility-related services and constructions (FRS&C), other services, and supplies in most years. The finding that most noncompetitive contracts employ an only one source exception in contrast to the urgency exception (which is reported in the no competition (other) category) departs from the literature review.

Figure 2. Competition Across Facility-Related Services and Construction, Other Services, and Supplies

While competition is more prevalent in contingency contracting, the trend has changed for the worse. Figure 2 also shows that noncompetitive contracting has become more prevalent in all three categories. The trend is most alarming in FRS&C where competition rates have cratered, which is particularly troubling because this is classically a sector where a range of different vendors can easily enter to provide services. The trend for other services involves a slow decline, which is typical in contracting markets. However, the rise in noncompetitive contracts in supplies is not necessarily as problematic as it first appears. Much of the decline has been in competition that results in only a single offer. Noncompetitive procedures do have tools to help when only one vendor is available, so going directly to these arrangements may be preferable to competitions that only ever attract one vendor. Nonetheless, the goods acquired in contingency contracting are often simpler than the advanced weapon systems acquired domestically. There may be an opportunity in this sector for greater use of competition.

When it comes to UCA contracting, the concerns raised in the literature appear to have been addressed. Usage is declining across the board and in recent years there are no reported UCA contracts in FRS&C or for supplies. The takeaway here is that perhaps the acquisition community should be less concerned about UCAs in contingency contracting and more focused on ensuring that they do not return to the past prevalence level in the category of conventional supply contracting.

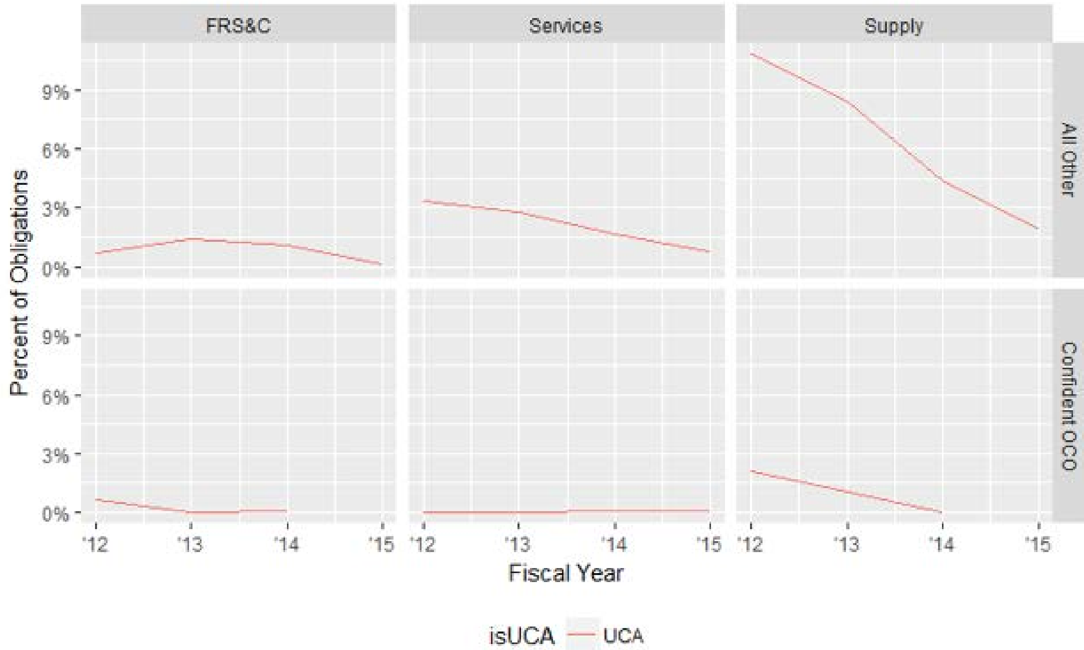


Figure 3. Undefinitized Contract Actions Across Facility-Related Services and Construction, Other Services, and Supplies

Closing Thoughts

Competitive trends appear to be in keeping with past results, but it does show the risk of relapse as contingency contracts are less used and are further from the public eye. While the sums being discussed are much smaller than in prior years, there may be room to regain previous rates of competition. By comparison, the dramatic reduction in UCA contract usage in recent years is a laudatory trend and may mean that reformers should focus on the remaining pockets of UCA elsewhere, although the study team will watch this space carefully in the next steps of this project, which consider the civilian forms of crisis contracting. The strongest signal from this research is the increasing divergence between the spending on OCO budget accounts and related contingency contracts. Future iterations of this study will examine this discrepancy and what it means for this challenging form of contracting.

References

Actions, U. C. (2017). (Revised December 9, 1998) 217.7400, 74–76.

Adrian, L. (2010). Reachback support: Providing stateside support to deployed contracting elements. *ACC Today*, 14–15.

Ausink, J. A., Castaneda, L. W., & Chenoweth, M. E. (2011*). Air Force contingency contracting: Reachback and other opportunities for improvement.*

Bontjer, R. A., Holt, J. P., & Angle, S. (2009). *Spending the development dollar twice: The local economic impact of procurement in Afghanistan*. Retrieved from https://buildingmarkets.org/sites/default/files/spending_the_development_dollar_twice_report_july_2009.pdf

Britt, C. D., & Miles, J. L. (1985). *Contracting under conditions of national emergency/full mobilization*. Montgomery, AL: Department of the Air Force, Air University.

Calhoun, A., & Larssen, M. (2013). *Implications and constraints of fiscal laws contingency contracting.* Monterey, CA: Naval Postgraduate School.

Calvaresi-Barr, A. (2007). *Use of undefinitized contract actions understated and definitization time frames often not met*. Washington, DC: Author.

Commission on Wartime Contracting in Iraq and Afghanistan. (2011). *Transforming wartime contracting controlling costs, reducing risks: Iraq.*

Dickinson, L. A. (2011). *Outsourcing war and peace: Preserving public values in a world of privatized foreign affairs* (Kindle Ed.). New Haven, CT: Yale University Press.

Dodaro, G. L. (2009). *American Recovery and Reinvestment Act: GAO' s role in helping to ensure accountability and transparency highlights*. Washington, DC: GAO.

Dunn, L. (2016, January 4*). Reachback contracting can provide a wide variety of advantages.* Association of the United States Army.

Epstein, S. B., & Williams, L. M. (2017*). Overseas contingency operations funding : Background and status.* Washington, DC: Congressional Research Service.

Gaffney, S., & Berger, B. T. (2009, June). Accountability and transparency in the American Recovery and Reinvestment Act. *Government Finance Review*, 78–82.

GAO. (2012). *Contingency contracting: Agency actions to address recommendations by the Commission on Wartime Contracting in Iraq and Afghanistan*. Washington, DC: Author.

GAO. (2015, September). *Disaster contracting: FEMA needs to cohesively manage its workforce and fully address post-Katrina reforms*. Washington, DC: Author.

Goldstein, M. L. (2014). *Recovery Act: USDA should include broadband program's impact in annual performance reports.* Washington, DC: GAO.

Gordon, D. I. (2014). *Emergency acquisitions guide*.

Holeywell, R. (2012, February). Did the stimulus do anything for transparency? *Governing*.

Hutton, J. P. (2010). *Defense contracting: DoD has enhanced insight into undefinitized contract action use, but management at local commands needs improvement.* Washington, DC: GAO.

Lipowicz, A. (2009, October 15). Agencies spent billions of stimulus money on noncompetitive contracts. *The Business of Federal Technology*.

Mackin, M. (2015). *Disaster contracting: FEMA needs to cohesively manage its workforce and fully address post-Katrina reforms*. Washington, DC: Author.

Manuel, K. (2011). *Competition in federal contracting: An overview of the legal requirements.* Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA497721

Marvin, B. (2014). *Noncompetitive contracts based on urgency need additional oversight*. Washington, DC: GAO

McMillon, C. L. (2000). *Contingency contracting within the Department of Defense: A comparative analysis*. Monterey, CA: Naval Postgraduate School. Retrieved from http://calhoun.nps.edu/handle/10945/9207

Needham, J. (2010). *Recovery Act: Contracting approaches and oversight used by selected federal agencies and states*. Washington, DC.

Office of Inspector General. (2012). *Summary report on DoD's management of undefinitized contractual actions.* Washington, DC: Author.

Office of Inspector General. (2016). *Summary guide to overseas contingency operations oversight lessons learned.* Washington, DC: Author.

Pickup, S. L., & Khan, A. A. (2009). *Overseas contingency operations: Funding and cost reporting for the Department of Defense*. Washington, DC: GAO.

Professional Services Council. (2012). *Testimony on the Comprehensive Contingency Contracting Reform Act of 2012*. Washington, DC: Professional Services Council.

Swan, C. (2012). "SPOT ON" contracting: Improving transparency and contractor personnel monitoring in contingency operations. *Journal of Contract Management*.

Undefinitized contract actions. (n.d.). Retrieved from https://www.ignet.gov/sites/default/files/files/caguideuac.pdf

Disclaimer

The Center for Strategic and International Studies (CSIS) does not take specific policy positions; accordingly, all views expressed in this presentation should be understood to be solely those of the author(s).

Use of Incentives in Performance-Based Logistics Contracting: Initial Findings

Andrew Hunter—is a Senior Fellow in the International Security Program and Director of the Defense-Industrial Initiatives Group at CSIS. From 2011 to 2014, he served as a senior executive in the Department of Defense, serving first as chief of staff to under secretaries of defense (AT&L) Ashton B. Carter and Frank Kendall, before directing the Joint Rapid Acquisition Cell. From 2005 to 2011, Hunter served as a professional staff member of the House Armed Services Committee. Hunter holds an MA degree in applied economics from the Johns Hopkins University and a BA in social studies from Harvard University. [ahunter@csis.org]

Jesse Ellman—is an Associate Fellow with the Defense-Industrial Initiatives Group (DIIG) at CSIS. He specializes in U.S. defense acquisition policy, with a particular focus on contracting trends in the Department of Defense, the Department of Homeland Security, and government-wide services; sourcing policy and cost estimation methodologies; and recent U.S. Army modernization efforts. Ellman holds a BA in political science from Stony Brook University and an MA with honors in security studies, with a concentration in military operations, from Georgetown University. [jellman@csis.org]

Andrew Howe—is a Research Intern with the Defense-Industrial Initiatives Group (DIIG) at CSIS. [ahowe@csis.org]

Abstract

Performance-based logistics (PBL) contracts, which have been used by private industry for decades, (particularly in the airline industry as a way to manage complex fleets) have only relatively recently begun to be used in the public sector worldwide. Research on PBL application indicates that PBLs can be successful in lowering costs and improving performance in both government and private contracting. In both cases, PBL contracts depend on the ability of the customer to properly structure and implement contract incentives to promote vendor behavior that reduces costs and improves performance while delivering the customer's desired outcomes.

This report examines how such incentives are used in PBL contracting and looks further towards how incentives can best be utilized in a PBL contracting environment. This report is structured in three parts: a review of the available literature on the use of incentives in PBL contracting, a data analysis of where and how PBL contracts are used in the DoD, and a summary of initial findings from the experts CSIS has interviewed on the subject.

Introduction

Performance-based logistics (PBL) contracts, which have been used by private industry for decades, (particularly in the airline industry as a way to manage complex fleets) have only recently begun to be used in the public sector worldwide. Research on PBL application indicates that PBLs can be successful in lowering costs and improving performance in both government and private contracting. In both cases, PBL contracts depend on the ability of the customer to properly structure and implement contract incentives to promote vendor behavior that reduces costs and improves performance while delivering the customer's desired outcomes.

In order to examine the question of incentive use in PBL contracts, CSIS has undertaken a research effort focused around interviews with PBL experts among Department of Defense (DoD) PBL vendors, private sector PBL vendors, and government customers (both domestic and foreign). The objective of this research effort is to better understand how incentives are used in PBL contracting and how incentives can best be utilized in a PBL contracting environment.

This report is structured in three parts: a review of the available literature on the use of incentives in PBL contracting, a data analysis of where and how PBL contracts are used in the DoD, and a summary of initial findings from the experts CSIS has interviewed on the subject, which are primarily focused on the experiences of DoD PBL vendors at this stage of the research effort.

Literature Review

In the current resource environment, the DoD has become increasingly interested in performance-based logistics (PBL) contracts due to their potential for cost saving and improved outcomes. PBLs are a form of performance-based contracting, something that the DoD has had an interest in since the 1960s (Hildebrandt, 1998). At the most basic level, PBLs alter the normal incentive and risk structure of a contract to more strongly incentivize improvements in performance and quality of service from a contractor. This report uses the broad economics definition of the term incentives, which is not limited to fee structure but includes approaches like longer contract periods to incentivize up front investments or granting the contractor more control over process as an incentive to also take on more risk. While PBLs are currently in use broadly in the private sector, and to a more limited extent in the DoD, the effect of the incentives built into PBLs needs to be better understood. This review will examine incentives based on time, cost, and scope, and will discuss other potential incentives and challenges to designing incentives.

Performance-Based Contracting and Performance-Based Logistics

Performance-based contracting is a type of contracting that calls for contracts to be structured in such a way as to enable and reward better performance on the part of the service provider or contractor. PBLs are the DoD's performance-based contracts and are specifically agreements that are "usually long term, in which the provider … is incentivized and empowered to meet overarching customer oriented performance requirements … in order to improve product support effectiveness while reducing" total ownership costs (Estevez, 2011). While definitions do vary between sources, the DoD's PBL Guidebook states that PBL is "synonymous with performance-based life cycle product support, where outcomes are acquired through performance-based arrangements that deliver Warfighter requirements and incentivize product support providers to reduce cost through innovation" (Assistant Secretary of Defense for Logistics and Materiel Readiness, 2016a). This type of performance-based contract has been used in the private sector for decades, particularly in the aviation industry (Assistant Secretary of Defense for Logistics and Materiel Readiness, 2016a; A. Hunter et al., 2015). Its popularity is due to its design, which aligns incentives between customers and suppliers (Guajardo et al., 2012). PBLs differ from other forms of contracted support because the contracted outcomes are logistical and because a PBL must include a service component (A. Hunter et al., 2015).

Traditionally, product acquisition and sustainment have been treated as separate considerations, with the government granting a greater priority to acquisition. The recent shift to placing a greater emphasis on sustainment has helped to increase the value of systems purchased by the DoD (Berkowitz et al., 2005). The DoD began using PBLs in 1999 when the Air Force reached an agreement with Lockheed Martin to provide support for the F-117 Nighthawk. While initially intended as a way to improve readiness, the DoD has since begun using PBLs to “deliver needed reliability and availability, reduce total cost, and encourage and reward innovative cost reduction initiatives” (Assistant Secretary of Defense for Logistics and Materiel Readiness, 2016a).[1] While the DoD’s *PBL Guidebook* does not specify the difference between “reduc[ing] total cost” and implementing “innovative cost reduction initiatives,” for the purposes of this report, the former is interpreted as taking known steps to reduce costs and the latter is finding new ways to reduce costs. Currently, the DoD describes PBLs as “the Department of Defense’s preferred product support strategy to deliver improved weapons systems readiness at the same or lower total cost” (Center for Executive Education, 2012). Since PBLs came into use, they have helped the DoD achieve both cost reductions and higher availability rates for systems (A. Hunter et al., 2015).

In a guide to best practices regarding PBLs, the Center for Executive Education from the University of Tennessee (2012) identified three success factors that define good PBL contracts. The first success factor is “alignment,” which can be best understood as ensuring that the government and the contractor have both embraced PBLs as a new way of structuring the provider-client relationship and not just a variant of business as usual. The second success factor is “contract structure.” The report defines a good contract structure as one that appropriately balances risk and asset management, establishes an environment that allows for creativity and shared success, and uses a pricing model that takes incentive types into account. These incentive types can take many forms, as discussed in the next section. The final success factor is performance management, which involves establishing and aligning desired outcomes and establishing metrics for reporting and improving. These points are all echoed in the DoD’s *PBL Guidebook* (Assistant Secretary of Defense for Logistics and Materiel Readiness, 2016a).

This paper focuses on incentives as the key to achieving good contract structure.

Incentives

Every business arrangement involves incentives. An incentive can be defined as a “stimulus to a desired action” or “anything that encourages or motivates somebody to do something” (Assistant Secretary of Defense for Logistics and Materiel Readiness, 2016a; Office of the Under Secretary of Defense, 2016). In the context of PBLs, an incentive is a “term or condition that encourages the desired product support integrator and/or provider behavior to deliver the relevant Warfighter outcome” (Assistant Secretary of Defense for Logistics and Materiel Readiness, 2016a). While incentives can be a part of any type of contract, they are particularly integral to PBLs. In fact, the DoD considers the “key to a successful PBL arrangement [to be] the use of incentives to elicit desired behaviors and outcomes” (Assistant Secretary of Defense for Logistics and Materiel Readiness, 2016a).

When articulating the PBL business model, Kleemann, Glas, and Essig (2012), like the Center for Executive Education (2012), included incentive payments as one of the three

[1] Some recent examples of DoD programs that include PBL contracts are the C-17 Globemaster III Sustainment Partnership, the T-45 Goshawk Contractor Logistics Support, the High Mobility Artillery Rocket System Life Cycle Contract Support I/II, the E-8 Joint Surveillance & Target Attack Radar System Total System Support Responsibility, the F/A-18 Hornet F/A-18 Integrated Readiness Support Teaming, and the F-117 Nighthawk Total System Performance Responsibility & Total System Support Partnership (Gardner et al., 2015)

key components of the compensation part of the model. After reviewing the literature on the experiences of organizations that implemented PBLs, Sols and Johannesen (2013) found broad consensus in the existing literature that aligning incentives with performance achievements is a main factor for PBL success. Therefore, while incentives are not required for a PBL, they are an integral component of contract structure and often make it work better (Kleemann, Glas, & Essig, 2012; Straight, 2006). Their importance was highlighted by the Proof Point study, which found that incentives “drive the behavior, actions, and investment decisions” of product support providers (Assistant Secretary of Defense for Logistics and Materiel Readiness, 2016a). It follows that the appropriate use of incentives can lead to preferable outcomes for the government.

Yet effective incentives are not as simple as just offering money in exchange for desired behavior. As recently as the early 2000s, the DoD was found to be giving firms award fees that were not linked to outcomes (GAO, 2005). This finding potentially calls into question the efficacy of incentives. In other words, if a firm knows it will be paid its award fee regardless of whether it achieves its performance targets, the award fee is no longer an incentive. A more recent report from the Government Accountability Office (GAO) found that, although incentives are a key part of PBLs, many of the contracts they reviewed lacked “effective incentives,” a circumstance that both lowers the ability of firms to reduce support costs and lowers their incentive to do so. For example, of the 29 PBL arrangements GAO reviewed, only four contained incentives intended to control or reduce costs (GAO, 2008). This finding by the GAO suggests that a better understanding of the effects of incentives could improve the outcomes of PBLs.

When included in contracts, incentives “encourage contractors to meet specified objective and subjective outcome metrics, resulting in explicit … or implicit … financial benefits to industry” (Assistant Secretary of Defense for Logistics and Materiel Readiness, 2016b). With traditional contracts, a contractor profits from selling increasing numbers of its given product or service and has little incentive to improve that product beyond staying ahead of a competing contractor, and even less incentive if the contractor has a monopoly on the product or service. With PBLs, the focus is on performance, not on the quantity produced, meaning that contractors are incentivized to provide products and services that perform well regardless of potential competition. If done well, PBLs can increase profits for the contractor, but they do shift risk from the government to the contractor when compared to more traditional contracts. In a traditional contract, the government purchases a number of components and thus risks having to pay in spite of a higher than anticipated failure rate or even equipment becoming obsolete. With PBLs, the government is purchasing a performance output, meaning that these risks are shared between the government and the firm (Gardner et al., 2015; Gupta, et al., 2010). This is part of their appeal to the government. However, while firms are certainly willing to enter into PBL contracts, this change in the balance of risk means that firms must have the capacity to attain a greater reward in return for greater risk. In the case of PBLs, this is through incentives, with the caveat that those incentives must promote behaviors and outcomes that benefit both the DoD and the firm (Assistant Secretary of Defense for Logistics and Materiel Readiness, 2016a). Incentives can take multiple forms, each of which is discussed subsequently.

When considering incentives, it must be remembered that contractors and the government have different priorities when it comes to risk. Vendors care primarily about financial risk, meaning concern about their return on investment. In contrast, the DoD is primarily concerned with operational risk, meaning its ability to meet mission objectives. In the face of these competing interests, PBLs strike a balance between risk to the vendor and risk to the government, with vendors accepting higher risks (i.e., having to make expenditures to react to the DoD's use of equipment, which is outside the control of the vendor) in return for the premium of higher potential profitability (Doerr, Lewis, & Eaton, 2005; Gardner et al., 2015). A further complication is that, in addition to aligning incentives for the government and the contractor, incentives must sometimes align with the components of the contractor or subcontractors that will be working on the project (Boyson et al., 2008). This potentially leads to an increase in the complexity of creating effective incentives.

Time-Based Incentives

Time-based incentives involve the initial length of a contract and altering contracts with a given contractor to extend their life. After conducting a series of interviews, Gupta et al. (2010) found that the main incentive for contractors is the continuation of the contract. The authors recommend that initial contracts should be for at least five years, which allows contractors to recover their initial investment in a project and solidifies expectations for needed employees and equipment. For example, the contract for support for the F-117 was for five years with the option to extend for an additional three, a feature that was considered a key to the success of the program (J. Hunter, n.d.). However, it should be noted that contracts for relatively simple subsystems or arrangements can be shorter, as they require less investment (Assistant Secretary of Defense for Logistics and Materiel Readiness, 2016a). In practice, the Navy specifies that its PBL contracts are long term, as is the case with the Consolidated Automatic Support System (Klevan, 2008; Stailey, n.d.).

In comparison, the UK's Ministry of Defense negotiates through-life capability management contracts that are similar to PBLs, but can be much longer. These lengthier contracts incentivize more long-term investments than shorter contracts and have been credited with billions of pounds saved for the UK government (Gansler, Lucyshyn, & Harrington, 2012). However, as is discussed in the *Cost-Based Incentives* section, lengthier contracts are not currently an option for the U.S. government under the current Federal Acquisition Regulation (FAR) and various related statutes. Another comparison can be made to the Australian approach, which involves using contract duration as the primary incentive. While a contract may initially be for a period of five years, the government can begin review in the second year to determine whether to add to the length of the contract if vendors can demonstrate that they have met performance benchmarks. This would face the same challenges as the UK approach, but potentially could be done through the use of indefinite contract vehicles, which have previously been used by the DoD (A. Hunter et al., 2015).

For the contract to be continued, and thus have the benefit of the incentive realized, the contractor must meet certain requirements related to cost, quality, or delivery. As should be evident, if a contractor cannot meet the requirements specified in the contract, the contractor runs the risk of not having the contract extended. In this case, either the incentive could be inadequate or the contractor could be incapable of reaching the agreed-upon goal.

Gardner et al. (2015) conducted a survey of six existing PBL programs and conducted interviews with PBL experts from both the DoD and industry. Like Gupta et al. (2010), they found that there was a "high level of satisfaction" with contracts that lasted five years with the option for continuation. Those interviewed said that the length ensured that risks were shared in an acceptable manner. The authors found that the ability to continue the contract past its initial period strengthened the relationship between a contractor and the government because it allowed for flexibility to make changes to the contract.

In addition, among those interviewed by Gardner et al. (2015), those who were party to a contract with multiple guaranteed years felt the most satisfied with their incentive to invest. One interviewee also told the authors that long-term contracts are one of the most important factors for contractors to accomplish weapon systems affordability improvements. In determining the optimal length of contract, the report from the Center for Executive Education (2012) found that the best practice was to have the contract last as long as the payback period for the contractor's investments.

Another question Gardner et al. (2015) sought to answer was whether the limits on contract length set by the Federal Acquisition Regulation (FAR) and related statutes limited the desired contract length for projects. The FAR regulates the acquisition of supplies and services by all federal executive agencies (GSA, DoD, & NASA, 2005). Generally, the individuals interviewed did not think the limits set by FAR were a major problem and the issue was secondary to other concerns, though some did express a desire for the ability to negotiate longer contracts. The authors did find that one of the main concerns among those they interviewed was that funding was not guaranteed over the years of a contract due to the nature of the congressional appropriations process (Gardner et al., 2015). As noted previously, one way to mitigate these challenges is to use indefinite contract vehicles such as IDVs, which do not make future work automatic but do ensure that a mechanism is already in place to allow it (A. Hunter et al., 2015).

Cost-Based Incentives

Cost-based incentives are those that are focused on contractor profits. When thinking about cost incentives, the most important consideration is the type of contract and types of fees the government will offer the contractor (Gupta et al., 2010). The FAR identifies a spectrum of contract types that fit into these categories based on the fee-type of the contract. The fees include fixed fees, incentive fees, and award fees. The primary difference between these different contract types and fee types is what criteria are used to adjudicate contractor fee and the resulting profits or losses (GSA et al., 2005).

One important factor when considering contract types is profit sharing. Typically, if there was an increase in efficiency in a cost-plus contract, the government would use this as an opportunity to lower costs, meaning that the DoD would enjoy all of the return and the contractor would not be incentivized to improve performance. In fixed-price contracts, the contractor receives the financial benefit of any gains in efficiency without the DoD cutting costs. The area of the spectrum between these two ends is filled by various types of contracts with incentive fees.

PBLs have typically been either firm-fixed-price or fixed-price incentive firm, but can also take the form of other types of fixed-price contracts (Assistant Secretary of Defense for Logistics and Materiel Readiness, 2016a). While fixed-price is not required, it is the DoD's preferred type of contract (Assistant Secretary of Defense for Logistics and Materiel Readiness, 2016a). Other forms of PBLs, such as fixed-price incentive fee, allow for profit sharing, so that both the DoD and the contractor benefit from cost reductions and increases in efficiency (Gardner et al., 2015). However, a firm-fixed-price contract may be picked deliberately to further strengthen the incentive for the firm to save money and thus come in below budget (Gupta et al., 2010).

Another approach is to use cost-based incentives that are based on performance metrics. And just as there are different types of PBL contracts for various circumstances, each type of performance-related incentive makes sense in different contexts. For example, the DoD's *PBL Guidebook* (2016) says that "shorter-term cost-type incentive arrangements are appropriate" until sufficient information has been collected on the program. In an instance where there is a single metric for defining success, the government and firm can adopt a model described by Sols, Nowick, and Dinesh (2007). The authors described a model with a “dead zone” at its center. They defined this as normal system performance, with the bottom edge and top edge of the dead zone representing the lower and upper limits of normal system performance respectively. If performance remains in this zone, the contractor will receive no reward and will not be assessed a penalty. If performance falls below the dead zone, then a penalty should be incurred by the contractor. If performance rises above the dead zone, the contractor should be awarded a bonus for exceeding normal performance. The key consideration, according to the authors, is that the contractor and government must agree on linking awards and penalties to given performance parameters. An example of a performance metric that could be used is average number of backorders and average total downtime of a system. Mirzahosseinian and Piplani (2011) found that a compensation model based on the time average of backorders leads to lower amounts of both backorders and downtime. Sols et al. (2007) also note that this could be harder if several metrics are needed, a scenario that they consider more likely than having a single parameter. Their model for a single metric is represented in a two-dimensional space. Two metrics would require a three-dimensional space. The DoD has five parameters for assessing logistic performance (operational availability, mission reliability, logistics response time, logistics footprint, and cost per unit usage), which would require a six-dimensional representation. This presents challenges when designing metrics for a contract.

When used, cost-based incentives appear to have a positive effect. In one analysis, the DoD found that performance increased for 12 out of 14 PBL projects with cost reduction incentives (Office of the Deputy Assistant Secretary of Defense Materiel Readiness, 2011). A commonly cited example of this is the set of F-117 sustainment contracts. These were cost plus incentive fee/award fee contracts. The performance incentive fee was awarded based on seven objectively measured metrics. The award fee was based on four subjectively evaluated categories. This number of metrics is mostly in keeping with the PBL Guidebook’s suggestion that three to five is the “effective number” of metrics (Assistant Secretary of Defense for Logistics and Materiel Readiness, 2016a). In total, 80% of the contract dollars were incentivized (J. Hunter, n.d.). The contracts are also Total System Performance Responsibility (TSPR) contracts, which raises the concern that within any given year they are “must-pay” obligations for the Air Force. TSPR contracts entail the government obligating the agreed-upon funds at the start of each year, which ensures that funding is stabilized. This means that the funds must be paid, even if operations requirements were to change (GAO, 2000; Gardner et al., 2015). In spite of these concerns, the operating cost for the F-117 increased minimally. In other words, the contracts largely controlled costs to the government (J. Hunter, n.d.).

Scope

Scope-based incentives take advantage of the inherent profit structure of PBL contracts. Whether there is a firm-fixed-price contract or a fixed-price incentive contract, that firm-fixed price will be based on government estimates of cost plus an allowance for contractor profits. The contractor generates additional profits by providing the agreed-upon outcome for a lower cost than was achieved in the past. A contractor’s ability to wring out further efficiencies is theoretically proportional to the portion of the process it controls. Because of this, greater scope means greater revenue and greater chance for profits for the contractor, and it means increased efficiency for the DoD (A. Hunter et al., 2015). Gupta et al. (2010) argued that another way to approach scope-based incentives is to use them as a mechanism for giving the contractor both more responsibility and larger incentives through changes in the contract based on performance. In other words, an increase in scope can be a reward for good performance.

However, Gupta et al. (2010) noted that, because of the government's requirement for a competitive procurement process, it is challenging to employ scope-based incentives. While it may make sense for the same contractor to cover multiple responsibilities for a system, if another contractor can perform some of those functions for a lower price then it will receive the contract. In addition, even without this concern, it can be challenging to determine the appropriate scope of a project. For example, A. Hunter et al. (2015) examined the Industrial Product-Support Vendor contract, which provides support for several Air Force Air Logistics Centers. The authors found that the scope of the contract was very narrow, creating the potential for duplicative efforts on the part of the contractor, the Defense Logistics Agency, and the Air Logistics Center, and limiting the contractor's ability to provide improved support by restricting its ability to leverage usage information to achieve efficiencies. Because increased scope for the contractor means reduced scope for government organizations, there are inherent limits to how easily scope can be shifted between the two. Although this situation has been improved over time, it does illustrate the difficulty in determining the appropriate scope of a contract, never mind scope-based incentives.

Other Incentives

The literature on other types of incentives for PBLs is limited. Other types of incentives that could be considered are those based on scale of the contract, flexibility of the contract, and prestige accrued by the contractor.

Challenges to Designing Incentives

One of the main challenges to adopting any form of performance-based contracting (the more generic term for what the DoD calls PBL) is achieving what Selviaridis and Norrman (2015) call a joint intent between the two parties involved in the contract. Their research found that customers were reluctant to offer extra rewards and providers were concerned about agreeing to performance-based incentives, perceiving them to be risky. While the authors were not examining defense contractors, the same challenges apply to PBLs.

Another potential issue arises when more than one contractor is involved in fulfilling the contract, such as when a contractor uses subcontractors (Selviaridis & Norrman, 2014). As noted previously, each contractor may react to incentives differently or incentives designed for the main contractor may not incentivize changes in behavior by the subcontractors. Yet another issue is that if incentives are poorly designed and overseen, they can also lead to unintended behavior that is beneficial for the contractor but detrimental to their client (Koning & Heinrich, 2013). The authors of this study found that in some contexts, such as when the risk of failing to meet contract expectations is greater, contractors can exhibit gaming behavior to avoid losing out on funding. However, it should be noted that the authors found this behavior to have little impact on outcomes.

An additional concern is that it is possible for a system to exceed expectations based on one parameter while underperforming based on another parameter (Sols, Nowick, & Verma, 2007). This creates a challenge when designing incentives, as the award of the benefits of incentives is based on measurable metrics. This scenario creates some complexity in determining whether the award should be given.The Current State of DoD PBL Contracting[2]

This section of the report examines how PBL contracts are currently used within the DoD to provide context for the analysis that follows. Data for this analysis is drawn from the publicly-accessible Federal Procurement Data System.

The use of PBL contract structures within the DoD grew steadily through much of the 2000s. From less than $400 million in 2000, obligations under PBL contracts rose to over $2 billion by 2004, and just under $6 billion by 2010. Use of PBL contracts has surged since then, reaching a high of nearly $9 billion in 2014, before falling off since. Figure 1 shows total DoD contract obligations under PBL contracts, broken out by major DoD component.

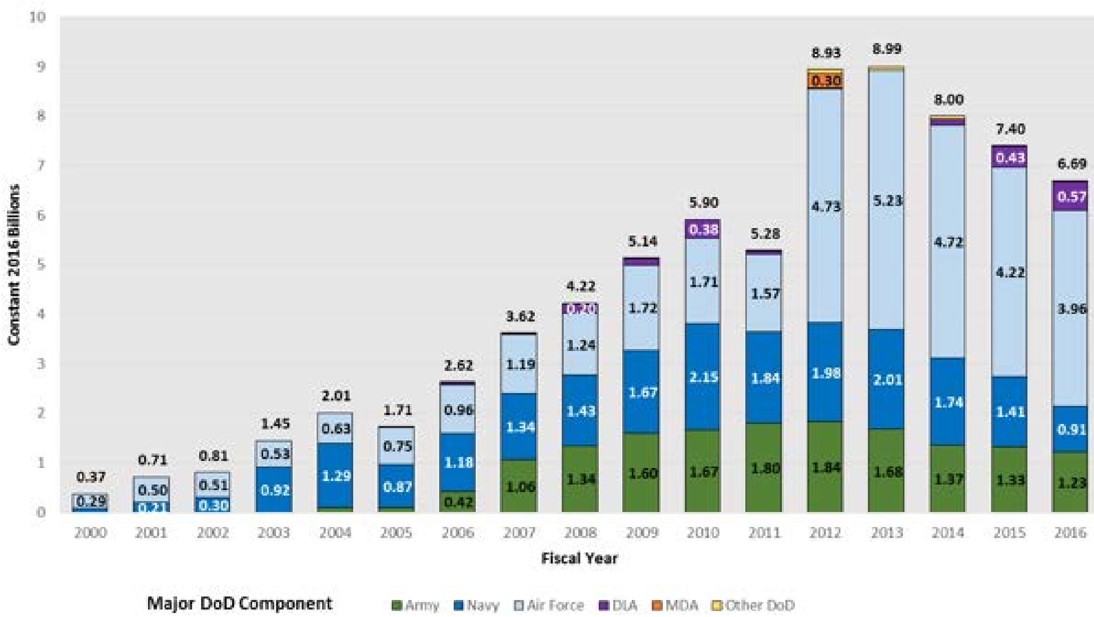


Figure 1. DoD PBL Contract Obligations by Component, 2000–2016
(Source: FPDS; CSIS analysis)

Overall DoD PBL contract obligations were nearly four times higher in 2016 than they were in 2005, indicative of the degree to which acceptance of the utility of PBL contract structures has grown within the DoD. As a share of overall DoD contract obligations, PBL contracts have risen from just over 1% in 2009 to nearly 3% between 2013 and 2015, before declining slightly in 2016 to 2.3%.

---

[2] The Federal Procurement Data System, which CSIS uses as its primary source for government contract data, does not have a field that can be used to broadly identify PBL contracts, CSIS has attempted to fill this gap with a number of data sources, including reviews of contract solicitations. While some smaller PBLs may not have been captured in this effort, CSIS is confident that it has identified a sufficient share of DoD PBL contracts to meaningfully inform an analysis of trends.

Army, which had less than $100 million in PBL contract obligations in 2005, saw PBL contract obligations rise to a high of over $1.8 billion by 2012, driven by large PBL contracts related to the UH-72A light utility helicopter and the RQ-7 Shadow tactical UAV.

Navy, meanwhile, was at the forefront of the adoption of PBL contract structures in the early- to mid-2000s, with nearly $1.3 billion in PBL contract obligations in 2004, spread among a number of PBL programs not readily identifiable in FPDS. Obligations peaked in 2010 at nearly $2.2 billion and remained near that level until 2013. Air Force, meanwhile, saw significant obligations for PBL contracts as early as 2000, related to the B-2 bomber platform, and steady growth between 2003 and 2010, due to increasing obligations related to that same platform. Air Force PBL contract obligations more than tripled between 2011 and 2012, primarily driven by $2.2 billion in obligations in 2012 under a PBL contract related to the C-17A transport aircraft.

Since DoD PBL contract obligations peaked in 2013, total obligations have declined by 26%, over three times as steeply as the decline in overall DoD contract obligations between 2013 and 2016. Both Army (-27%) and Air Force (-24%) have seen declines roughly in line with the overall rate of decline for DoD PBL contract obligations, but Navy has declined at more than double that rate (-55%), with significant declines across the range of platforms and systems that the Navy maintains under PBL contract structures.

Notably, despite what seems to be the end of a period of decline for DoD contracts, with overall DoD contract obligations rising by 7% in 2016 after sustained declines between 2009 and 2015, DoD PBL contract obligations fell by 10% in 2016.

What the DoD Uses PBL Contracts For

Because PBL contracts often involve purchasing a mix of multiple products and services, the usual FPDS categorization schema that CSIS uses to track what is being contracted for—Product or Service Code—is less useful here. Instead, Figure 2 looks at platform portfolios, a categorization schema developed by CSIS, using a combination of the *ProductorServiceCode* and *ClaimantProgramCode* fields in FPDS, that aggregates all product, service, and R&D contracts by the type of platform the contracts are associated with.

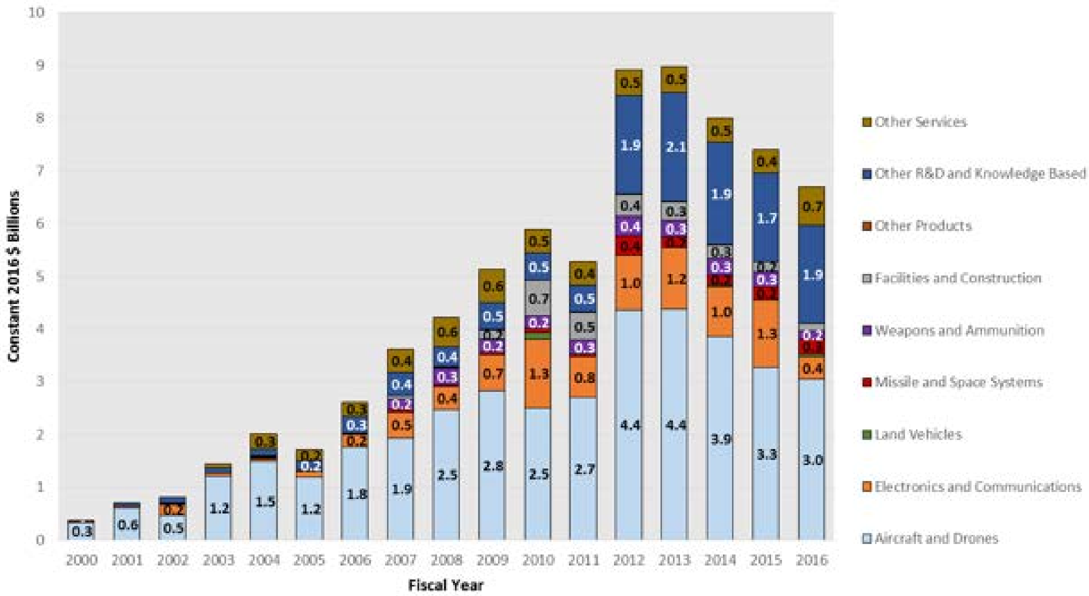


Figure 2. DoD PBL Contract Obligations by Platform Portfolio, 2000–2016
(Source: FPDS; CSIS analysis)

Unsurprisingly, the Aircraft & Drones platform portfolio has been the biggest driver of growth in PBL contracting over the 2000 to 2016 period. Aircraft & Drones accounted for over 54% of DoD PBL contract obligations in every year from 2000 to 2009, and have accounted for at least 43% in every year since. That decline is largely maintained by the growth in obligations under the "Other R&D and Knowledge Based" category, but further investigation by CSIS into the individual contracts has revealed that much of the growth in that category is for PBLs that are related to Aircraft & Drones platforms, but are not identifiable as such using the *ProductorServiceCode* and *ClaimantProgramCode* fields. CSIS is presently testing ways to improve the accuracy of the Platform Portfolio categorization schema based on this discovery, and will integrate that revision into the final analysis of this research effort.

Despite that minor data issue, the available data does show Aircraft & Drones PBL contract obligations returning to prior levels after a notable spike in 2012 and 2013, driven heavily by the growth of the C-17A PBL program. The decline since 2013 has been broad-based, with a number of PBL programs seeing reduced contract obligations.

Electronics & Communications has consistently been one of the larger categories of PBL contract obligations, accounting for more than 10% in every year from 2007 to 2015. Land Vehicles, meanwhile, have only accounted for more than 1% of DoD PBL contract obligations in a single year (2% in 2010) during the 2000 to 2016 period. Missiles & Space Systems had never accounted for more than 1% until 2012, but have accounted for between 2% and 4% since. Similarly, Weapons & Ammunition, which had never accounted for more than 1% of PBL contract obligations from 2000 to 2006, accounted for between 4% and 7% in every year from 2007 to 2015.

Interestingly, there have been almost no PBL contract obligations for Ships & Submarines, with total PBL contract obligations of less than $40 million over the entire 2000 to 2016 period. While the maintenance and repair needs of ships and submarines differ greatly from those of most other platforms in the DoD's inventory, it is nonetheless surprising to see that virtually no PBL work has been tried, even for smaller surface ships or shipboard systems.

How the DoD Structures PBL Contracts

Unsurprisingly, the vast majority of DoD PBL contracts are structured as Firm Fixed Price contracts, which follows generally accepted best practices for PBL contracting. Since 2000, 68% of DoD PBL contract obligations have been awarded under Firm Fixed Price contract, as seen in Figure 3.

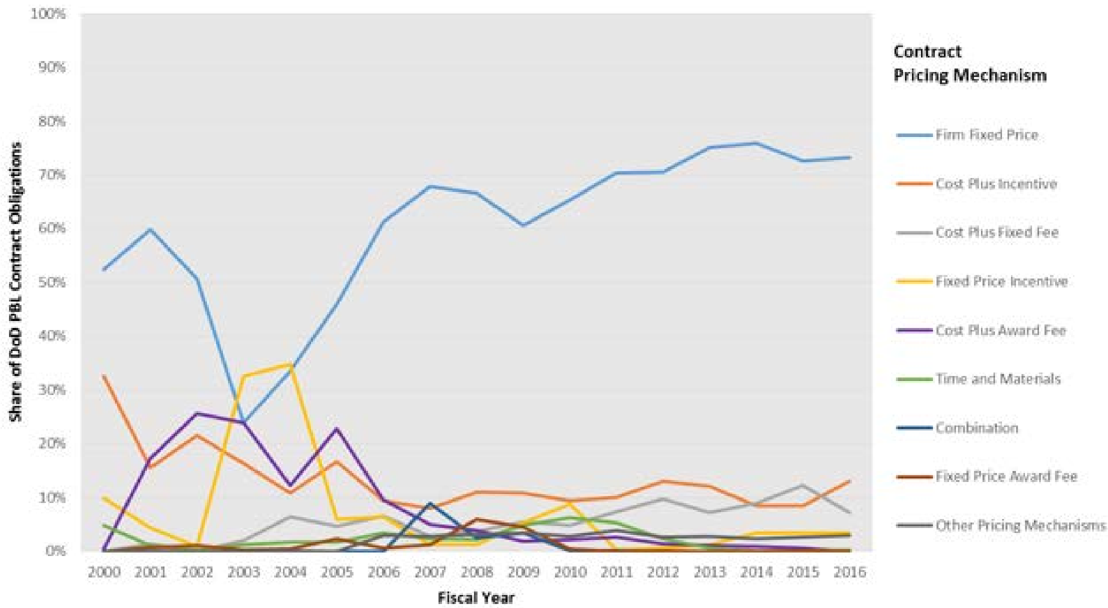


Figure 3. DoD PBL Contract Obligations by Contract Pricing Mechanism, 2000–2016

(Source: FPDS; CSIS analysis)

Aside from a brief dip in the early- to mid-2000s, when the share of obligations awarded under Fixed Price Incentive, Cost Plus Incentive, and Cost Plus Award Free briefly surged. While both Fixed Price Incentive Fee and Cost Plus Award Fee contract types have not been a significant factor in DoD PBL contracting since those brief spikes in usage, a surprisingly large share of PBL contracts are still structured as Cost Plus Incentive; between 8% and 13% of PBL contract obligations were structured as Cost Plus Incentive in every year since 2006. Cost Plus Fixed Fee, which was not used significantly for PBL contracts in the early 2000s, grew to account for between 3% and 7% of DoD PBL contract obligations from 2004 to 2011, and between 7% and 12% from 2012 to 2016. Both Cost Plus contract types seem to be primarily used for PBL contracts that are more transactional in nature, but CSIS is consulting with experts to better understand how and why the decision is made to structure some PBLs as Cost Plus, rather than Fixed Price.

Competition for DoD PBL Contracts

While about half of overall DoD contract dollars in recent years have been awarded after effective competition,[3] DoD PBL contracts are far less competitive, as seen in Figure 4.

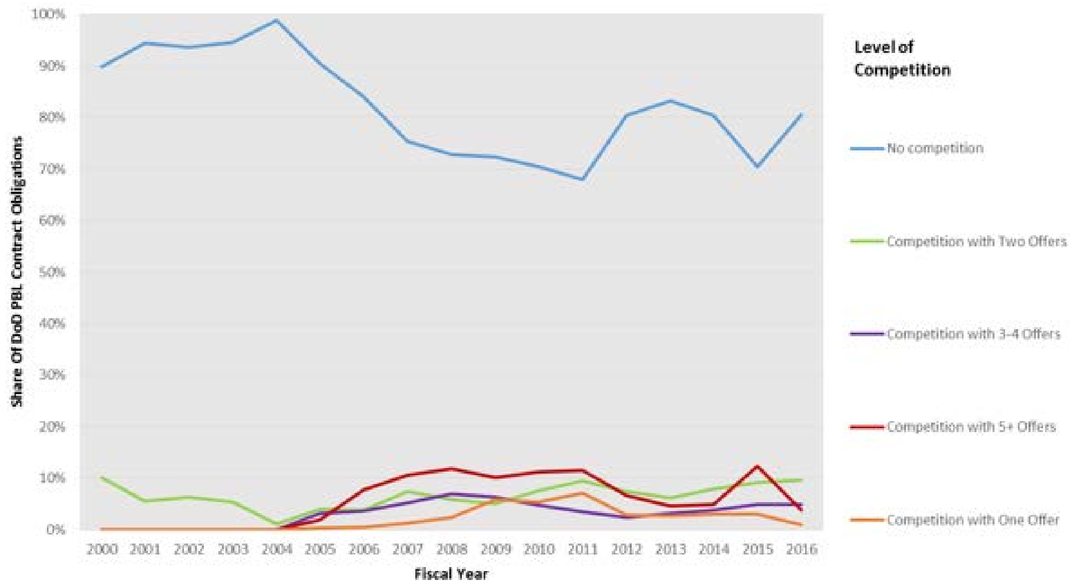


Figure 4. Level of Competition for DoD PBL Contract Obligations, 2000–2016
(Source: FPDS; CSIS analysis)

For the 2000 to 2016 period, 78% of DoD PBL contract obligations have been awarded without competition. This is not surprising, since most PBLs for platforms and systems go to the original manufacturer for a number of reasons, including the following:

- Most manufacturers retain the technical data rights to their platforms and systems, without which it is impossible for another vendor to perform the functions under a PBL. (Even in cases where the original manufacturer might be willing to sell those data rights, the cost is likely to be more than the DoD is willing to pay.)
- Original manufacturers have supply chains already developed, whereas anyone competing to take over a PBL would have to build a new supply chain from scratch.
- In discussion with experts, some mentioned their hesitance to try to compete to take over an existing PBL even when one was potentially going to be put up for competition, due to the large advantage that the incumbent vendor is perceived to have.

[3] CSIS defines “effective competition” as a competitively-sourced contract which receives at least two offers, and which excludes competitions where only one offer is received.

Nonetheless, there has been a significant increase in the share of PBL contract obligations awarded after effective competition since the early 2000s. While only 1% of PBL contract obligations were awarded after effective competition in 2004, that share rose to between 23% and 25% between 2007 and 2011, with competitions receiving five or more offers making up the largest portion of those effectively competed PBL contracts. That share has declined in recent years, mostly hovering in the mid to high teens, but nonetheless remains notably higher than in the early 2000s.

For both the Navy and Air Force, the share of PBL contract obligations awarded on a sole source basis has remained in the low to mid 80% range in recent years, which, while higher than the overall DoD PBL rate, is an improvement over the rates seen in the early- to mid-2000s. The Army and DLA, by contrast, have always seen lower rates of sole source awards, with 52% and 62%, respectively, awarded on a sole source basis since 2000. This difference is primarily a factor of the fact that DLA and the Army spend a greater share of their PBL contract obligations on subsystem- and component-level PBLs, which are more likely to have multiple vendors able to potentially perform.

Who Performs DoD PBL Contract Obligations

The industrial base that performs PBL contracts for the DoD is heavily concentrated, which is not surprising, given that many of the large PBL contracts are tied to major platforms and systems, which are in turn produced by a small number of the largest defense vendors. Table 1 shows the top 15 DoD PBL vendors between 2009 and 2016, with both their respective contract obligations and their shares of overall DoD PBL contract obligations for that period.

Table 1. Top 15 DoD PBL Vendors, 2009–2016

(Source: FPDS; CSIS analysis)

Rank	Vendor	Total DoD PBL Contract Obligations 2009-2016	Share of Total DoD PBL Contract Obligations 2009-2016
1	Boeing	14.5	26%
2	L3 Communications	7.4	13%
3	Northrop Grumman	6.5	12%
4	Lockheed Martin	4.4	8%
5	General Electric	3.3	6%
	Top 5 Total	**36.1**	**64%**
6	Airbus	3.0	5%
7	General Dynamics	2.2	4%
8	Rolls Royce	2.2	4%
9	Maritime Helicopter Support [Lockheed Martin/Sikorsky Joint Venture]	1.8	3%
10	Bell-Boeing Joint Program Office	1.5	3%
11	Textron	1.5	3%
12	Raytheon	1.3	2%
13	General Atomics	1.2	2%
14	Honeywell	0.9	2%
15	Dyncorp International	0.5	1%
	Top 15 Total	**52.2**	**93%**
	Overall DoD PBL	**56.3**	

The top five DoD PBL vendors accounted for 64% of the total DoD PBL contract obligations between 2009 and 2016, and the top 15 accounted for 93%. Both of those figures have increased significantly over the 2009 to 2016 period: The share going to the top five PBL vendors has increased from 55% in 2009 to a high of 71% in 2015, before falling back to 66% in 2016, while the share going to the top 15 has risen from 87% in 2010 to 95% in 2016.

Northrop Grumman accounted for the largest shares of DoD PBL contract obligations in 2009 through 2011, but since 2012, Boeing has received nearly 75% more obligations than the second-ranked vendor, L3 Communications.

Initial Interview Findings

The core of this research effort is a series of interviews with experts on PBL contracting within vendors who perform PBLs for the DoD, vendors who perform PBLs for the private sector, and government entities (both foreign and domestic) that contract for PBLs. At this stage of the research effort, CSIS has conducted interviews with multiple experts that manage PBLs for the DoD, covering the range of PBL projects, from component-level PBLs to system-level PBLs to full platform PBLs. While CSIS plans to conduct more interviews in the coming months, to gain the broadest range of perspectives on the issue of incentives in PBL contracting, the experts that CSIS has already spoken to have provided a few key insights into how they approach the issue of incentives in PBLs.

There are three key initial findings from discussions with these experts:

- Contract length is the most powerful incentive.
- Negative monetary incentives are effective, even down to the subcontractor level.
- Positive monetary incentives are not seen as effective or desirable.

Contract Length Is the Most Powerful Incentive

Virtually every expert that CSIS has interviewed thus far has cited contract length as the most powerful incentive in a PBL environment. This consistency is likely the result of the nature of how vendors operate in a PBL environment. As discussed briefly in the literature review section, PBLs generate savings and performance improvements because vendors are incentivized to invest up-front in equipment and process improvements that allow them to meet performance targets and reduce costs. In theory, these up-front investment costs will be offset by profits in later years, but that assumes that there are later years to the contract.

In some cases, vendors performing PBLs for the DoD have found themselves on year-to-year contracts, and those experts cited the uncertainty in those structures as a powerful disincentive to invest in equipment and process improvements. After all, if the basic business model for PBLs is for up-front costs to be justified by long-term profits, and there is no guarantee that the contract will still be active long-term, it is difficult to make a business case to justify the up-front investments. Longer-term contracts also allow vendors to fund their suppliers long-term, which can help generate significant savings. In a year-to-year contract environment, or any one with particularly short contract terms, the risk to vendors is likely to be too high for them to tolerate in order to make the sorts of investments necessary for a successful PBL.

Even in cases where the contract length is at least five years, which experts cited as the bare minimum necessary in order for them to feel that the risk inherent in up-front investments is justified by the long-term rewards, the experts that CSIS spoke with cited other factors that disincentivize investment. Even with a five year contract, which many contracting entities within the DoD are hesitant to award, the single year nature of federal budgeting means that a contract is no guarantee of future work. If a vendor has a five year contract to ensure availability of a platform, invests money up-front to improve availability and drive down costs, and then, two years into the contract, Congress decides not to appropriate the funds necessary to conduct work on the platform at the previously understood levels, the vendor can find themselves in a bad situation. Even if the contract isn't outright cancelled, if the work level is scaled back significantly in a PBL where payment is based on the volume of work (as happened to some programs during the budget drawdown and sequestration), a vendor can find themselves without enough profit over the course of the contract to offset the up-front investments. (That same dynamic can act as a disincentive to government customers as well; experts cited cases where firm-fixed-price PBLs based on assumptions of workload ended up with lower workloads than expected, which left the government customer feeling like they had significantly overpaid.)

Experts that CSIS spoke to cited 10 years as an ideal length for a PBL contract; while contracts of that length are not an option under U.S. federal contracting regulations and related statutes, other countries such as the UK and Australia have had positive experiences with longer-term PBL contracts. In the UK, they have also used triggered option year contract structures, where a contract is awarded for a base length, and then future years are triggered as long as performance metrics are continually met. Australia also uses a rolling contract extension approach. A contractor performing well may receive a sixth year of performance during year three of the contract as a reward. A contractor not performing to the government's satisfaction may receive a warning in year three but have a chance to turn around their performance and still earn the extension in year four.

Experts among DoD PBL vendors indicated that these sorts of arrangements helped mitigate some of the risk and uncertainty of shorter-than-desirable base contract lengths, but they noted that these triggered option year arrangements have notable limitations. Most significantly, according to industry experts, they are most effective in competitive environments, which are a distinct minority of the PBL market; in a sole-source environment, where there is no threat of losing the contract to another vendor, the option years don't alleviate the fundamental concerns about future-year funding and workloads. This skepticism of length as an incentive in a sole-source environment has also been expressed by other U.S. experts in discussion of earlier CSIS work on this topic. This discrepancy between U.S. and international views of the efficacy of extensions merits further study in subsequent interviews.

Negative Monetary Incentives Are Effective, Even Down to the Subcontractor Level

Just as there was broad agreement about the efficacy of contract length as both a positive and negative incentive in a PBL environment, there was consensus among the experts that CSIS has spoken to about the effectiveness of negative monetary incentives. These sorts of incentives can take a number of forms, but at their core, they are fairly simple: if a vendor fails to meet a contractually-mandated performance metric over a particular period of time, the amount of money they receive under the contract is reduced by a pre-determined amount. The experts agreed that this sort of incentive was effective, primarily when it was something that was within their ability to control, and was something they could plan around. To the degree that negative performance incentives were tied to metrics that the vendor had less control over, or were harder to predict, the risk level inherent in those negative incentives would be greatly increased.

Some of the experts mentioned that these sorts of negative monetary incentives were effective even down to the subcontractor level. In a PBL environment, some vendors hold their larger subcontractors responsible for their role in meeting performance metrics, such that if they are responsible for the vendor not meeting the metric, they also share in the penalty. It was emphasized that, when this sort of shared responsibility is implemented, it only extends to the largest subcontractors, who have the ability to weather the potential penalties without it threatening their stability as a business. In some cases, penalties might not flow down to the subcontractor level in the initial years of a PBL project, but would start to be enforced later in the contract. The interviewees also noted that, in a well-constructed supply chain, the subcontractors should already exist as part of a team with the prime vendor, and that a sense of shared responsibility for meeting performance metrics should already be assumed, even absent shared penalties.

Positive Monetary Incentives Are Not Seen as Effective or Desirable

One common theme among the experts that CSIS had not seen any indication of, either in the literature or in prior research on PBLs, is the consensus that positive monetary incentives are neither effective nor desirable for vendors. Most positive monetary incentives take the form of additional money for meeting performance metrics targets above the contractual baselines, but the experts within DoD PBL vendors expressed a number of concerns about these contract structures. Most fundamentally, there was broad agreement that the additional money was rarely worth the cost of meeting the higher metric target. In cases where the experts had managed or worked on PBL contracts with positive monetary incentives, they had rarely seen cases where the work to meet the higher metric target resulted in a net profit. One aggravating factor was that these positive incentives were sometimes combined with cost sharing measures such as fixed-price incentive fee contracts. In these cases, the cost sharing mechanism proved more important towards driving contractor decisions than the possibility of receiving a performance reward.

Interviewees noted that it was particularly difficult to predict the cost of meeting those higher targets at the start of the contract, which meant that properly pricing the positive monetary incentive was a challenge. Additionally, in cases where the vendors could properly price the higher metric target, it was difficult to get the government to agree to incentive levels high enough to make hitting the increased target potentially profitable.

Other Findings of Note

In addition to those three key findings from the experts that CSIS has spoken to thus far, the following are points of interest that have been raised by one or two experts, but which are interesting enough that CSIS will pursue them in future interviews:

- the government incentive to keep a certain percentage of work in-house;
- hesitancy of vendors to try to compete for existing PBLs because of the perceived difficulty of dethroning incumbents;
- control as an incentive and a risk factor—government-furnished equipment, requirements to use depots (which the vendor has minimal ability to manage) as subcontractors, and other features that lessen the scope of what factors of a PBL the vendor can exert influence over; and
- skepticism of “power-by-the-hour” PBL arrangements, due to the number of hours frequently coming in below projections.

Final Thoughts

In the final stages of this research effort, CSIS will continue to interview experts from across the spectrum of PBL contracting experience. This will help CSIS gather the broadest possible picture of how incentives are currently used in PBLs and how incentives should be used. Additionally, CSIS will identify specific PBL contracts as case studies and examine the incentive structures (both contractual and implicit) of those contracts to illustrate the real-world consequences of the choices made in structuring those PBL contracts. The final report under this research project will provide lessons learned on using incentives in a PBL environment.

References

Assistant Secretary of Defense for Logistics and Materiel Readiness. (2016a). *PBL guidebook: A guide to developing performance-based arrangements*. Retrieved from http://bbp.dau.mil/docs/PBL_Guidebook_Release_March_2016_final.pdf

Assistant Secretary of Defense for Logistics and Materiel Readiness. (2016b*). Product support manager guidebook.* Retrieved from https://acc.dau.mil/adl/en-US/440507/file/82239/PSMGuidebookUpdate2016(6-24-16).pdf

Berkowitz, D., Gupta, J. N. D., Simpson, J. T., & McWilliams, J. B. (2005). Defining and performance-based logistics in government. *Defense Acquisition Review Journal, 11*(3), 255–267.

Boyson, S., Corsi, T., Harrington, L., & Schoolfield, B. (2008). *Building dual-use supply chain management capabilities: Experiences to date of system integrators, original equipment manufacturers & third party logistics firms*. Retrieved from https://www.dau.mil/cop/pbl/DAU%20Sponsored%20Documents/UMD%20Report%20LMCO%20Dual%20Use%20Report%20November%202008%20FINAL.pdf

Center for Executive Education. (2012). *The tenets of PBL: A guidebook to the best practices elements in performance-based life cycle product support management* (2nd ed.).

Doerr, K., Lewis, I., & Eaton, D. R. (2005). Measurement issues in performance-based logistics. *Journal of Public Procurement, 5*(2), 164–186.

Estevez, A. F. (2011). *DoD product support business case analysis guidebook.* Washington, DC: DoD.

Federal Acquisition Regulation (FAR), 48 C.F.R. ch. 1 (2015).

Gansler, J. S., Lucyshyn, W., & Harrington, L. H. (2012). *An analysis of through-life support: Capability management at the UK's Ministry of Defence*. Retrieved from http://www.cpppe.umd.edu/publications/analysis-through-life-support-capability-management-uk%E2%80%99s-ministry-defense

GAO. (2000). *Defense logistics: Air Force report on contractor support is narrowly focused.* Washington, DC: Author.

GAO. (2005). *Defense acquisitions: DOD has paid billions in award and incentive fees regardless of acquisition outcomes*. Washington, DC: Author.

GAO. (2008). *Defense logistics: Improved analysis and cost data needed to evaluate the cost-effectiveness of performance based logistics*. Washington, DC: Author.

Gardner, C. P., Ogden, J. A., Kahler, H. M., & Brady, S. (2015). Balancing incentives and risks in performance-based contracts. *Defense Acquisition Review Journal , 22*(4), 472–506. Retrieved from http://www.dau.mil/publications/DefenseARJ/ARJ/ARJ75/ARJ75-Gardner.pdf

Guajardo, J. A., Cohen, M. A., Kim, S.-H., & Netessine, S. (2012). Impact of performance-based contracting on product reliability: An empirical analysis. *Management Science, 58*(5), 961–979. https://doi.org/10.1287/mnsc.1110.1465

Gupta, J. N. D., Eagan, M. C., Jones, J. N., & Platt, J. C. (2010). Contractor incentives for success in implementing performance-based logistics: A progress report. In *Proceedings of the Seventh Annual Acquisition Research Symposium*. Monterey, CA:

Naval Postgraduate School. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a530064.pdf

Hildebrandt, G. G. (1998). The use of performance incentives in DoD contracting. *Acquisition Review Quarterly*, 217–234.

Hunter, A., Riley, M., Sanders, G., & Ellman, J. (2015). *Performance-based logistics: A process analysis for the Defence Logistics Agency*. Washington, DC.

Hunter, J. (n.d.). F-117 performance based logistics. U.S. Air Force.

Kleemann, F. C., Glas, A., & Essig, M. (2012). Public procurement through performance-based logistics: Conceptual underpinnings and empirical insights. *Journal of Public Procurement, 12*(2), 151–188.

Klevan, P. (2008). Navy success with PBL. Naval Supply Systems Command.

Koning, P., & Heinrich, C. J. (2013). Cream skimming, parking, and other intended and unintended effects of high-powered, performance-based contracts. *Journal of Policy Analysis and Management, 32*(3), 461–483.

Mirzahosseinian, H., & Piplani, R. (2011). Compensation and incentive modeling in performance-based contracts for after-market service. In *Proceedings of the 41st International Conference on Computers & Industrial Engineering* (pp. 739–744). Singapore.

Office of the Deputy Assistant Secretary of Defense Materiel Readiness. (2011*). Proof Point Project: A study to determine the impact of performance based logistics (PBL) on life cycle costs.*

Office of the Under Secretary of Defense. (2016). Guidance on using incentive and other contract types. Retrieved from http://www.acq.osd.mil/dpap/policy/policyvault/USA001270-16-DPAP.pdf

Selviaridis, K., & Norrman, A. (2014). Performance-based contracting in service supply chains: A service provider risk perspective. *Supply Chain Management: An International Journal, 19*(2), 153–172.

Sols, A., & Johannesen, L. H. (2013). The role of transition contracts in performance-based logistics initiatives. *Systems Engineering, 16*(4), 453–463. https://doi.org/10.1002/sys.21242

Sols, A., Nowick, D., & Verma, D. (2007). Defining the fundamental framework of an effective performance-based logistics (PBL) contract. *Engineering Management Journal, 19*(2), 40–50. https://doi.org/10.1080/10429247.2007.11431730

Stailey, B. (n.d.). *Consolidated Automatic Support System (CASS): PBL success.*

Straight, R. L. (2006). Performance-based contracting: Results, performance standards, incentives. In *91st Annual International Supply Management Conference*.

Disclaimer

The Center for Strategic and International Studies (CSIS) does not take specific policy positions; accordingly, all views expressed in this presentation should be understood to be solely those of the author(s).

Price Analysis on Defense Logistics Agency (DLA) Contracts

Janie Maddox—Lecturer, Center for Civil-Military Relations, Naval Postgraduate School, has 30 years of experience working in the government contract management and educational fields. In 1994, she won the National Contract Management Association's (NCMA's) prestigious Blanche Witte Award. This recognized her leadership in the development of the Contract Pricing Reference Guides referenced in the FAR to guide government contract pricing. Her government teaching and educational management experience exists between the private and government sectors. She performed all the data collection for the current research but analyzed the data with Professor Gera to produce the report. [jlmaddox@nps.edu]

Ralucca Gera—is a tenured Associate Professor of mathematics in the Applied Mathematics department of the Naval Postgraduate School and has been teaching since 2005. Most of her publications are in applied mathematics, including over 30 papers in reputable journals in the area of graph theory and one in the Acquisition Research Symposium. Her awards include the NPS-wide Hamming award, the GSEAS teaching awards, and several department level awards. Dr. Gera's research grants feature collaboration with the Computer Science department and Operations Research department, as well as the Graduate School of Business at the NPS. She analyzed the data with Janie Maddox to produce the report.

Introduction

Over the last decade, the federal acquisition workforce has had to adapt to the need for new skill sets. Procurement reforms in the late 1990s required contracting specialists to have a greater knowledge of market conditions, industry trends, and market prices. Using market forces to determine reasonable prices has required an increase in both market research and price analysis methods. Most contracts pricing of acquisitions required cost analysis before these reforms became part of contracting regulations. These new regulations dictate a skill set for conducting price analysis that is largely missing from both the workplace and Defense Acquisitions University (DAU) existing curriculum. Since 2001, the Department of Defense Inspector General (DoDIG) and the Government Accountability Office (GAO) have issued several reviews of selected agencies discussing concerns about commercial and noncommercial prices of spare parts and services. Most of these reports have identified situations in which contracting officers failed to obtain adequate pricing information for justifying price reasonableness. In conclusion, both the DoDIG and the GAO have found that the contracting officers need improved expertise in both understanding and conducting price/cost analysis. The current gap in knowledge contributes to agencies missing cost saving opportunities as well as ventures to improve acquisition outcomes. Exercising appropriate price analysis methods that come from adequate price analysis guidance and training would address this gap.

Purpose

The purpose of this research is to identify the price analysis techniques being used and documented in the contracting file, and to explore potential improvements in conducting price analysis within the Defense Logistics Agency (DLA). This project builds on research previously conducted on Department of Defense (DoD) contracts (Redfern, Nelson, & White, 2013; Gera & Maddox, 2013; Maddox, Fox, & Gera, 2014).

Findings and Analysis

1a. Do Pricing Memoranda Deviate From Federal Acquisition Regulation (FAR), Defense Federal Acquisition Regulation Supplement (DFARS) Requirements, and DFARS Procedures, Guidance, and Information (PGI) Procedures?

We look specifically at how the contract file pricing memoranda deviated from Federal Acquisition Regulation (FAR), Defense Federal Acquisition Regulation Supplement (DFARS) requirements and DFARS Procedures, Guidance, and Information (PGI) procedures.

This question generated multiple findings. Rather than grouping our findings into one answer, the authors have addressed them individually below:

1a(i). Inadequate Documentation Finding

A number of contract files that we reviewed did not demonstrate that prices paid were reasonable due to inadequate FAR price analysis methods, as depicted in Table 1: Summary of Inadequately Justified Price Analysis Documentation in the Files by FAR Price Analysis Technique.

Table 1. Summary of Inadequately Justified Price Analysis Documentation in the Files by Price Analysis Technique

Price Analysis Techniques	Total	Supplies	Services	Construction
Inadequate price competition	10 of 25	0	22	3
Acceptance of prior prices without establishing reasonableness and/or appropriate adjustments made for differences	6 of 14	0	6	0
Incomplete references to current competitive price list	6 of 16	1	15	0
Incomplete comparison with IGCE or use of unreliable IGCEs	25 of 38	0	36	2
Incomplete statements based on references to market research	5 of 7	0	7	0
Offeror did not provide data that was appropriate	0 of 4	0	0	0
None of the above techniques used in pricing documentation	2 of 2	1	0	1
Totals of inadequate price analysis documentation for price reasonableness justification	54 of 106			

1a(i). Analysis

We noted that the file review verified the type of price analysis documented. Poor documentation in the files influenced our decisions to rate a pricing memo as justifiable or not. From this data, we determined that the personnel involved in performing these contract actions did not include sufficient documentation to support the price analysis method used as required by FAR and DFARS. Considering the number of inadequate price analysis justifications found in the memos sampled, it appears that contracting personnel do not know how to appropriately perform and document price analysis. In particular, two types of price analysis—references to market research and IGCEs—were performed and documented incorrectly more than 50% of the time. Previous price documentations were unjustified 43% of the time and 40% of price competition was found to be inadequate.

1a(ii). Comparison of Current Offered Pricing Findings

In 10 cases of the 25 currently offered prices, evidence showed that the proposed prices were not truly competitive.

Five pricing memorandums included some comparisons of current proposed (offered) prices when the lowest price was less than 80% of the next lowest price. For example, if the prices are $10, $50, and $55, respectively, then less than 80% would be anything lower than 80% x $50, which equals $40. So, the lower quote of $10 would be considered smaller.

In two files, a price from a technically unacceptable offeror was nonetheless used to make a price comparison.

There were nine instances in the sample of 66 where factors other than price determined the source selection, but price remained a substantial factor in 100% of those cases. Three contracts in this category were awarded to an offeror that was not the lowest offeror; however, two of the three had a statement of price reasonableness that *did not justify* the choice of the higher offer in the file.

The inappropriate comparisons certainly raise the issue that although competition is present and sought, is there actual price competition?

1a(ii). Analysis

Proposed prices that are not within 20% of the next lowest price raise questions about the reliability of the proposed prices, and the existence of actual price competition. This could indicate a mistake in the offered price, a misunderstanding of the contract requirements, etc. In few cases, documentation included some determination of why such a large gap separated the lowest price from the next valid price or a price verification request by the CO to the lowest offeror.

According to the FAR Part 6, the award of a contract to a supplier based on the lowest evaluated price alone can be a false economy if there is subsequent default, late deliveries, or other unsatisfactory performance resulting in additional contractual or administrative costs. While it is important that Government purchases be made at the lowest price, this does not require an award to a supplier solely because that supplier submits the lowest offer. The price from any offer that would not be considered for contract award such as technical unacceptability should not be used as a basis for price analysis. According to FAR 15, in awarding to any source other than the lowest priced offeror, the perceived benefits of a higher priced proposal shall merit the additional cost, and the rationale for tradeoffs must be documented in the file. Adequate price competition does not necessarily in of itself make a price reasonable.

1a(iii). Comparison of Proposed Price to Previous Price (Historical) and Sole Source Commercial Findings

Fourteen contract actions compared proposed prices to previous (historical) prices paid. In the six instances that had an invalid previous price documented in the file, a previous price could not be validated for one or more of the reasons displayed in Figure 1 (some had multiple disqualifying reasons). In two of the six, the previous price had a significant time lapse between the current and previous price. In three cases of the six, significant changes affected the terms and conditions. In four of the six cases, the reasonableness of the previous price seemed uncertain.

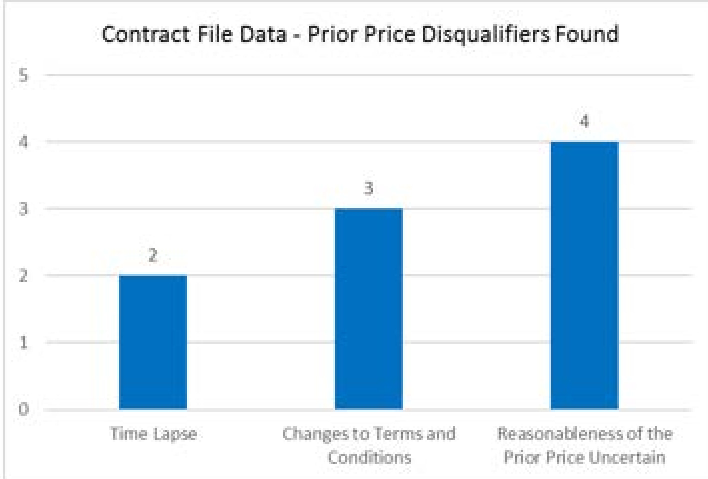


Figure 1. Contract File Data—Prior Price Disqualifiers

We found 26 items that were both Commercial and considered a sole source (see Figure 2). Of the 26, 10 had sufficient data to determine price reasonableness. In the 16 in which data did not substantiate price reasonableness, the CO requested data from the offeror only four times. Offeror provided the requested data all four times. In one case, the requested data had not been reviewed for contract award.

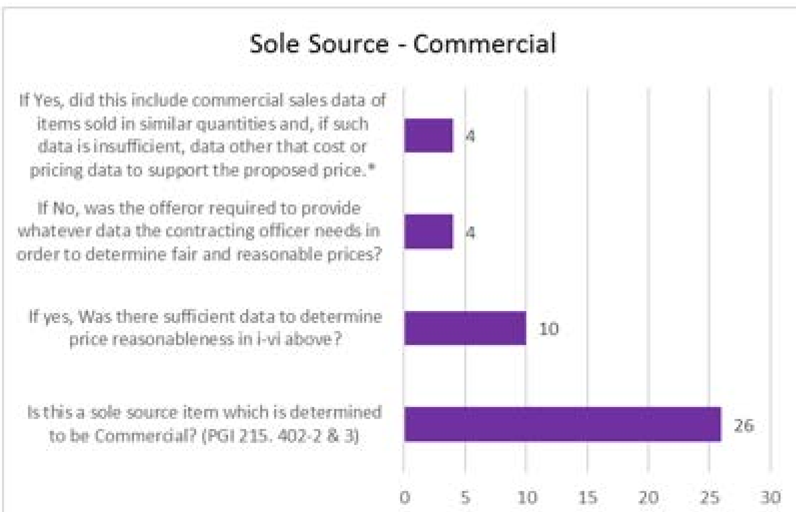


Figure 2. Commercial Sole Source

1a(iii). Analysis

If using invalid previous prices, then price reasonableness has not been determined. For example, it would not be sufficient to use price(s) from a database paid for by another contracting officer without understanding the type of analysis that was performed to determine the price. The DoD strengthened guidance on this subject in PGI 215.403–3(4) per below:

> *Reliance on prior prices paid by the Government*. Before relying on a prior price paid by the Government, **the contracting officer must verify and document that sufficient analysis was performed to determine that the prior price was fair and reasonable**. Sometimes, due to exigent situations,

> supplies or services are purchased even though an adequate price or cost analysis could not be performed. The problem is exacerbated when other contracting officers assume these prices were adequately analyzed and determined to be fair and reasonable. **The contracting officer also must verify that the prices previously paid were for quantities consistent with the current solicitation. Not verifying that a previous analysis was performed, or the consistencies in quantities, has been a recurring issue on sole source commercial items reported by oversight organizations.** Sole source commercial items require extra attention to verify that previous prices paid on Government contracts were sufficiently analyzed and determined to be fair and reasonable. At a minimum, a contracting officer reviewing price history shall discuss the basis of previous prices paid with the contracting organization that previously bought the item. These discussions shall be documented in the contract file.

Since previous price comparison is one of the two preferred price analysis techniques, contracting personnel often use it in determining price reasonableness. This method is effective, provided the validity of the comparison (similar items, categories, quantities, quality, qualifications, and/or circumstances), and the reasonableness of the previous price(s) can be established.

In this sample, more than 40% of the previous price comparisons made were invalid since the previous price was not verified. This illustrates why the authors determined that the contracts sampled do deviate from FAR/DFARS/PGI requirements and procedures. Further, If COs are not diligent in validating previous prices prior to using them for current pricing actions, then unreasonable prices can continually perpetuate themselves into future contracting actions.

Adding to the 43% previous price comparisons we found that were invalid, all six were identified as sole source commercial. Current guidance requires extra attention to verify previous prices paid in looking at Sole Source Commercial Items, as stated from PGI 215.403–3(4):

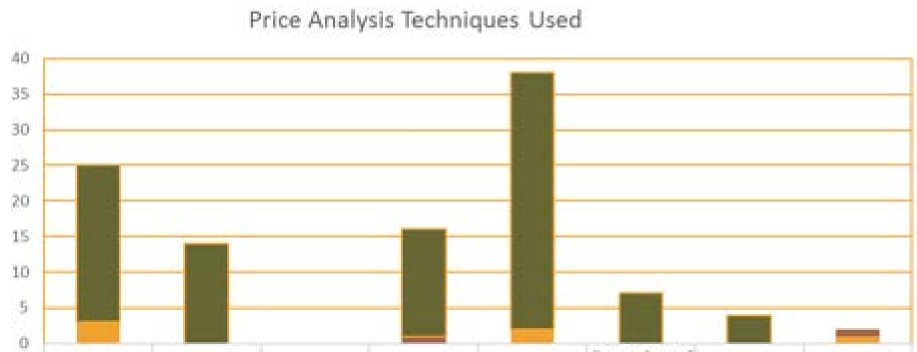
> Sole source commercial items require extra attention to verify that previous prices paid on Government contracts were sufficiently analyzed and determined to be fair and reasonable. At a minimum, a contracting officer reviewing price history shall discuss the basis of previous prices paid with the contracting organization that previously bought the item.

Of the six comparisons to previous price that were Sole Source Commercial and were considered unjustified, the CO requested and reviewed additional sales data in only one case. These six were part of the 16 total Sole Source commercial found to have insufficient data to determine price reasonableness.

1b. Do Pricing Memoranda Document the Type of Price Analysis Used in Determining Price? What Price Analysis Methods Are Being Used?

Findings

All of the pricing memos documented some type of price analysis used in determining that the price was reasonable. The research findings show that 25% of files used current competitive prices as a price analysis method (which is 24% of the total files). Comparison with the IGCE was documented in 38 pricing memos out of the 66 files, namely, 36% of the files. Comparison to Competitive Price Lists and through market research were present in 23 of the 66 files, totaling 35%. Previous prices (historical) documentation were present in 14 of the 66 files, totaling 13%. Comparison to another name was present in two of the 66 files representing 3% of the files, as seen in Figure 3.



	Current Offered Prices	Previous prices	Parametric estimating	Competitive published price lists	Comparison of proposed prices with independent government cost estimates	Comparison of proposed prices obtained through market research	Offerors Data	None of above
Services	22	14	0	15	36	7	4	
Supply	0	0	0	1	0	0	0	1
Construction	3	0	0	0	2	0	0	1

Figure 3. Contract File Data—Price Analysis Techniques Used

Analysis

According to the contract files sampled, comparison with the IGCE led the price analysis techniques used, with current offered prices ranking closely behind. IGCEs are not as reliable as current offered prices, therefore, their usefulness is questioned. Further price analysis techniques should supplement an IGCE.

The application of price analysis techniques is notable. Contracting personnel within the offices sampled recognize the importance of price analysis in determining price reasonableness. However, as discussed in 1ai. (Inadequate Documentation), contracting personnel did not include sufficient documentation to support the price analysis method used as required by FAR and DFARS. Without the proper supporting documentation, the value of the techniques is questionable.

1c. Do Pricing Memoranda Refer to Market Research?

1c. Market Research Findings

Seven contract pricing memoranda reviewed in the contract files used market research to establish price reasonableness, and a majority of the files in the sample contained market research reports as shown in Figure 4. Of the 32 market research reports, 17 (53%) of those contracts addressed the type of pricing data collected, as shown in Figure 5. Fifteen contracts in the sample did not address the type of pricing data collected in the market research report, and 23 contracts in the sample did not have a market research report that should have. In several files, the IGCE and market research report were combined into one document. Note: 11 samples were delivery orders without a requirement for market research, so only 55 files would have required a market research report.

Figure 4. Contract File Data—Market Research Reports

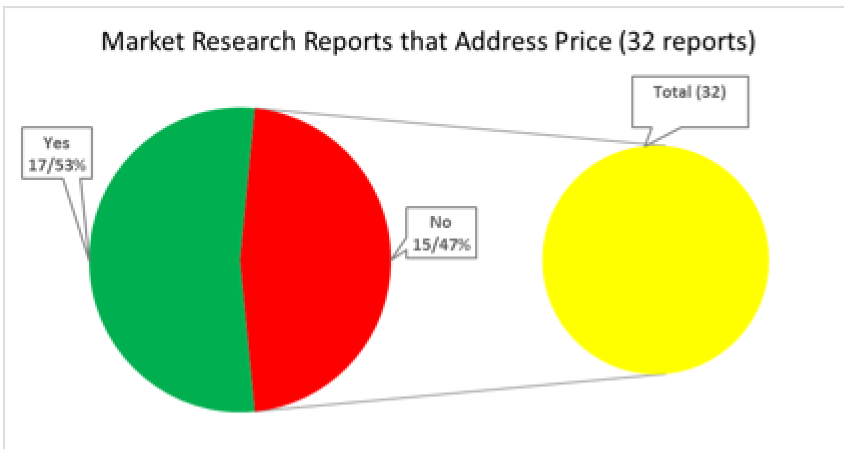


Figure 5. Contract File Data—Market Research Reports That Address Price

Notable Incomplete Statements Based on References to Market Research

- A few memos referred to Market Research as to the type of price analysis used; however, information came from a competitive price list for the price analysis comparison. COs seemed confused on the difference between market research and a competitive price list comparison.
- In using this comparison, buyer stated that he used partner agreements for justifying price. No evidence of partner agreements in file or reference to what specific information from the partner agreement proved useful.
- Buyer makes a price comparison and states that the basis of price justification is through market research. However, no information related to the market since the price used for comparison came from the IGCE. The IGCE was based on GSA schedule labor rates, but no support given for labor hours/mix or travel costs. So, a somewhat confusing PNM with unjustified price comparisons.
- A market survey had been conducted and one quote had been received from the same offeror who won the contract. The buyer used the market survey to justify the price by saying it was a fair market price obtained through

comparison of prices obtained through market research. No other price analysis had been used to justify the price. Three options were exercised from this contract and the buyers state the option prices were based on competition and therefore reasonable. There was no competition on the initial contract. Note: No IGCE appeared in the file.

1c. Analysis

Knowledge of marketplace suppliers and prices can be critical to the government's ability to negotiate a reasonable price. Poorly done market research lessens an activity's ability to achieve fair and reasonable prices. The authors found that market research appeared in a majority of the files we reviewed and were generally customer/requirements personnel generated. Market research does improve the buyers' understanding of pricing in the marketplace. The authors didn't look in depth at the quality of the market research reports but did note that market research reports addressed price in 53% of those examined. So, we conclude that the market research reports that examined pricing should have improved the buyers' understanding of pricing in the marketplace.

1d. What Are the Validity of the IGCEs and Contracting Officer's/Specialist's Interpretation and Use of the IGCEs in Pricing Memoranda?

1d. IGCE Findings

The IGCE has two roles: First, it supports what the customer and contracting offices believe is the "should price" and should be completed before the receipt of the price proposal and second, as a price analysis technique per FAR, parts 13 and 15. We will examine both here since they work together.

Developing and documenting an IGCE by its creator is a critical phase in the planning of the acquisition. The customer in the requiring activity is responsible for these actions. It must be substantiated with valid supporting documentation in order to be useful as a "should price," or a pricing technique, or both. The COs must be concerned with the reliability of the IGCE since it can be used as a proposal analysis comparison to determine a proposed price as fair and reasonable according to IAW 15.404–1(b)(v). When the IGCE is not substantiated, it should not be used as a pricing technique in validating a proposed price.

e looked at each of the IGCEs and concluded whether the "should price" was substantiated. In determining substantiation, we looked primarily at the source of data and the estimator's assessment of that data. We also looked at how the CO or buyer assessed the reliability of the IGCE. Just because we found an IGCE substantiated didn't absolve the CO from determining its reliability. Lastly, we assessed the validity of the COs' comparison of the IGCE to the proposed price. The number of substantiated IGCEs are not comparable to whether the CO assessed the reliability of the IGCE. In looking at the reliability of the IGCE, we only looked at IGCEs fulfilling the second role as a pricing technique.

The documentation stated that an IGCE was in 63 of the 66 contracts sampled (95%). Forty-two IGCEs were substantiated (63.6%) by previous purchase, catalogs, published price lists, contact with a vendor, or other, typically a government technical report as seen in Figures 6 and 7. Essentially, the developer of the IGCE explained the sources of information used to make the estimate. Fifty-seven of the 58 service contracts had an IGCE in the file; one of the supply contracts had an IGCE in the file, and all six construction contracts had an IGCE in the file.

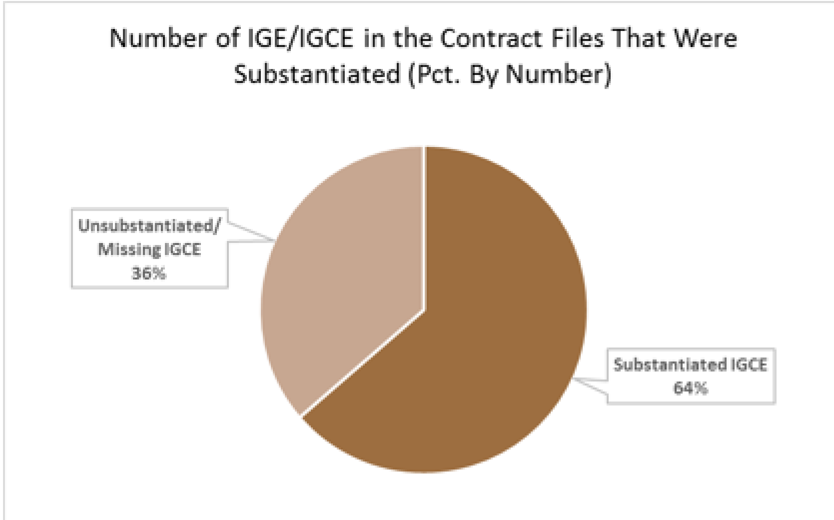


Figure 6. Number of IGE/ IGCEs Substantiated in the Contract Files

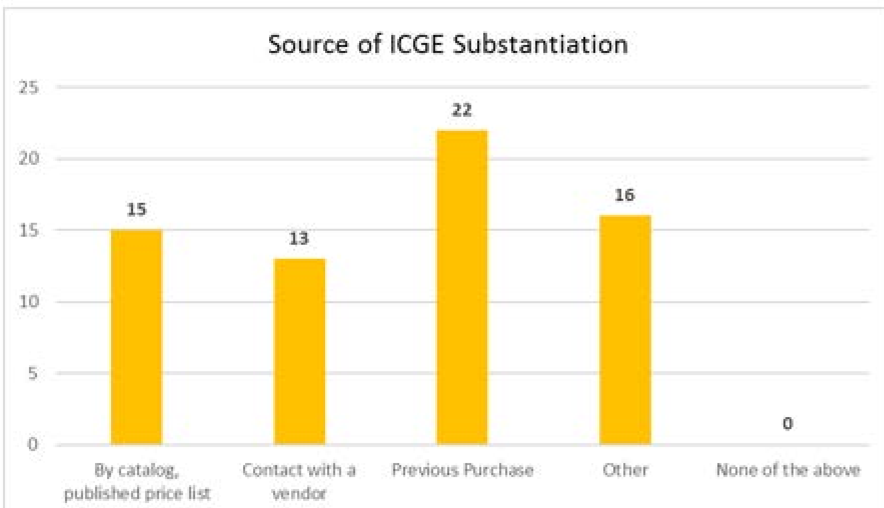


Figure 7. Source of IGCE Substantiation in the Contract Files

Notable Findings From Unsubstantiated IGCEs Reviewed

Lack of justification of labor hour estimates in the following:

- Only a statement that an engineer with experience estimated them.
- Several estimates of labor hour efforts based on historical contracts, but no historical data in the file as back up or even a contract number for reference.
- Task orders reviewed lacked evidence of estimates for labor hour effort. Since labor rates were already agreed to in the base ID/IQ contract, no perceived need to estimate labor mix and effort.
- Labor hour estimate based on a reference to non-identified historical work and a reduction of that historical effort based on a consolidation. No details about what is meant by consolidation. From reviewing the file, it appears that

tasks in the SOW has been reduced from previous efforts that should have been explained in the IGCE.

Lack of justification in estimating the labor rates as follows:

- Statements that historical rates were used without reference to any contract or data in file to back it up.
- Escalation rates were applied to future years with no reference to the source of the escalation rate.
- One escalation factor used was simply based on a quote in the DoD COTR handbook that stated "escalation between 2 and 3% is generally considered reasonable."
- IGCE creator used rates from a schedule with similar job titles, not similar services.
- Unusual quantitative method used to determine an acceptable range of labor rates. Estimator took 4 quotes, averaged them, and then created a range by adding 20% to the average price, and subtracting 20% from the average price. No details why estimator used a +/- 20%. Made the range too large and not useful.
- Only provided an estimated total dollar amount without a break down of labor mix, hours, or rates.

Thirty-eight of the 66 files highlighted in Table 2 used IGCEs as the basis for the price reasonableness of current prices more frequently than any other technique, essentially 37% of the contract actions reviewed. However, we found that only 13 of the 38 IGCEs used for determining price reasonableness could be determined reliable for use as a comparison.

Table 2. Answers to the Question: "What was the Documented Justification for Price Reasonableness?"

Price Justification in Pricing Memos		Construction	Supplies	Services
Comparison to current offered price?	25	3	0	22
Comparison to previous prices paid?	14	0	0	14
Parametric estimating?	0	0	0	0
Competitive published price lists?	16	0	1	15
Comparison of proposed prices with independent government cost estimates?	38	2	0	36
Comparison of proposed prices obtained through market research?	7	0	0	7
Analysis of offeror data?	4	0	0	4
None of the Above	2	1	0	1

Examples of the incomplete comparison with IGCE or use of unreliable IGCEs found in the file reviews:

- Though a construction contract used RS Means to substantiate the IGCE, the winning price came in at $265k versus the IGCE estimate of $452K. The winning price only represents 58% of the IGCE. No documentation in the file justified why the IGCE was so high, despite plenty of offers alongside the winning price to justify the lower price.
- The source of data in the IGCE is the sole source vendor's quote and referred to no history. Buyer used the unsubstantiated/unreliable IGCE for justifying price reasonableness.
- Documentation stated that price reasonableness was based on the comparison of the proposed price to an independent government estimate, but did not include any comments that would indicate the reliability of the IGCE in several pricing memos.

- Price justification relied on a weak IGCE. A quote from the previous offeror of the same services formed the basis of the IGCE. The IGCE was not substantiated; therefore, not reliable for comparison.
- Pricing memo mentioned that the IGCE was based on market research and historical data, but none of that was referenced in the IGCE.
- Though an IGCE was substantiated and could be used in justifying the reasonableness of the offered price, it had not been used. In the pricing memo, the IGCE is incorrectly stated as RS Means.
- A pricing memo discussed how the IGCE justifies the reasonableness of the offered price; however, it is incorrectly stated in the memo as pursuant to 15.404–1(c)(2)(iii)(D) and not to 15.404–1(c)(2)(v).
- Though the IGCE was substantiated and considered reliable to use, the offered price of $217k was only 40% of the IGCE, which was $553k, or essentially 60% lower. No other price analysis supported this lower price. Offeror negotiations took place. The buyer's objective was based on lowering the offered price by 10%, though it was well under the IGCE. The contractor conceded 1%. Some other data for comparison should have been sought.
- The PNM contains a statement, "In addition, the offeror's price was below the IGCE" as one of the justifications that the price is reasonable. No mention as to whether the IGCE was used for comparison or determined reliable or even why just being lower was a justification.
- A substantiated IGCE was used as the sole technique for price comparison. The buyer did not discuss why there was a 23% (significant) difference between the IGCE and the price and/or why other price analysis techniques had not been done to determine the reasonability of the price.

1d. Analysis of IGCEs

The use of an IGCE to determine price reasonableness is frequent, and the documentation of the reliability of IGCEs is not consistent. In contrast, only 24% of the IGCEs in the contract files identified as a price analysis technique in determining a fair and reasonable price could be validated as reliable; see Figure 8. Having more than 76% of a customer's IGCEs used without documentation for reliability is discouraging because it creates doubt about the price reasonableness determination.

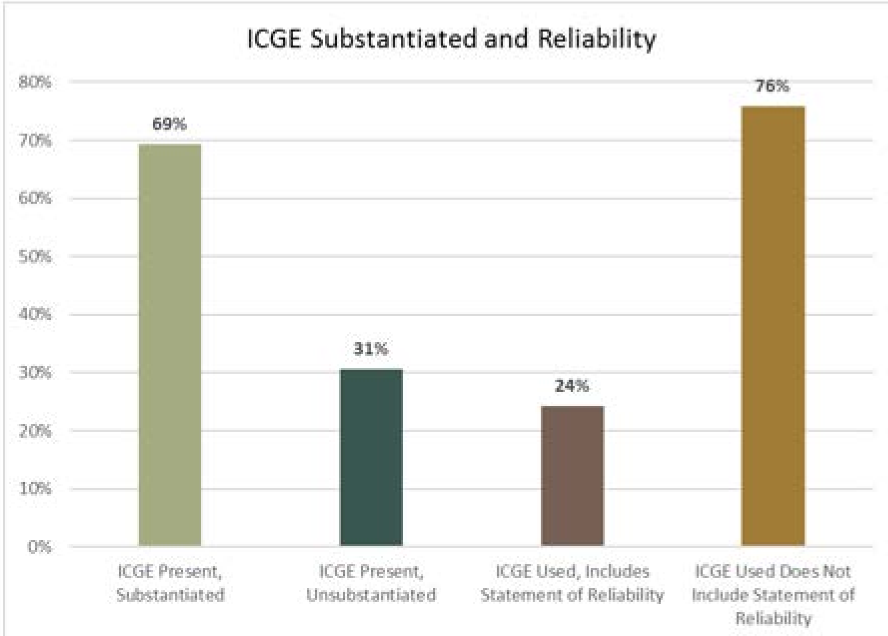


Figure 8. Contrast between IGCE Reliability and IGCE Substantiated

1e. If Pricing Memoranda Show Deviations, Do They Differ by the Same Characteristics and/or by Different Characteristics?

1e. Findings

In 45.5% of the files, deviations in pricing memoranda do exist as depicted as inadequate justification in Table 3. The files reviewed contained some consistent deviations and other unique ones.

Table 3. Contract File Data—Adequate vs. Inadequate Justification for Price Reasonableness

	Adequate Justification	Inadequate Justification	Total	Percent Justified
Contract Files	36	30	66	54.5%
Services	32	26	58	55.1%
Supplies	0	2	2	0%
Construction	4	2	6	66%

A majority of the pricing memoranda do deviate by two consistent characteristics: the lack of supporting documentation to justify the techniques used to establish price reasonableness based on previous prices and IGCEs. See findings that support the acceptance of prior prices without establishing their reasonableness and/or appropriate adjustments made for differences under the answers to 1a(iii) and findings behind the incomplete comparison with IGCE in the answers to 1d in the previous section.

The pricing memoranda in the sample reviewed established that the contracting officer determined price reasonableness as well as listed the technique used; however, substantiating documentation (e.g., calculation sheets, reference materials such as catalog data found online, copy of previous price documentation, and methodology) are not always included to support the source of their recommendations. A very small percentage of the IGCEs in the contract files used as a price comparison for price reasonableness included a statement of reliability or sought additional information from the IGCE creator to support the reasonableness of the offered price. Contracting personnel listed the IGCE as justifying the price of 38 contracts. Only 15 of the 38 were judged as reliable.

Some of the files reviewed contained unique deviations in the pricing memos. The following notes the assortment of unique problems uncovered in the pricing memos:

Notable Findings From Unjustified Pricing Memos Reviewed

1. Offeror sales data requested, received, but not reviewed.
2. Pricing memos that do not discuss the types of proposal analysis used in justifying price.
3. Despite an acceptable total evaluated price, unbalanced pricing involved the price of one or more contract line items being significantly over or understated, as indicated in the price analysis techniques applied.
4. The actual pricing memo left out details that had to be found under other tabs. The efforts made to justify price appeared adequate but were not recorded accurately in the PNM.
5. Buyer accepted a discount off a vendor’s price as the justification for accepting the price. The discount of 12% off the commercial sales price was the same as the price given the vendor by the manufacturer. The vendor has no commercial sales, making it difficult to determine whether the government received the best price. Though the offeror has multiple sales within the government, it has no commercial sales to the public. No comparison made as to whether the commercial priced items are fair and reasonable. No comparison to other vendors providing pricing for similar items to see whether the discount is reasonable. Nothing to say manufacturer prices are reasonable, other than to say they are commercial items sold in substantial quantities.
6. Offered price was reasonable because it was in line with competitive offers from recent years, yet no specific data provided justification.
7. The buyer used the IGCE and competitive price lists for price justification. However, the IGCE was the vendor’s quote, and the price list was simply a price list from the vendor. No adequate comparisons made. Also confusing is the use of MFR based on FAR 13.106–3(a) (2) statements for determining price reasonableness. Then in award summary, COR quotes FAR 15.404 IV and V (competitive price lists and IGCE) for determining price reasonableness. Should be either FAR 13 or 15, not both. Though in either case, COR still would not have justified the price adequately.
8. A buyer stated that an IGCE appeared in the PNM, but did not use it for the price comparison to the proposed price. The IGCE was substantiated and could have been used as an appropriate comparison. Unfortunately, the buyer used an invalid previous price for price analysis, instead of the IGCE. Had he done both, at least the IGCE would have supported the price reasonableness determination. Possibly buyers are not aware that more than

one price comparison is appropriate and sometimes necessary to justify price reasonableness. Three memos with similar issue.

9. Though an adequate price competition, the buyer only stated price was reasonable as it conformed with GSA schedule pricing, and included the application of discounts.
10. Very confusing: Buyer used a determination of an FMP as the basis for use in negotiating a final price, which was then considered fair and reasonable. The PNM did not specify the type of price analysis used for price comparison. A price analyst assisted the buyer and used competitive price lists of similar vendors to build the fair market price. Nothing was documented about how the labor effort or labor categories used in the FMP estimate were determined by the buyer or price analyst.

    In the researcher's review of the IGCE, the IGCE estimator used a DISA contract to estimate the hours/categories. However, the buyer did not discuss that the IGCE was used in determining the FMP. Then, the FMP, which was built solely by the price analyst upon competitive price lists, was used to negotiate the final price.

    The justification for price reasonableness was the negotiated price. The researcher believes the data available support the price, but it was not written up correctly, so the price was not justified. The statement in the PNM said only the following: "FMP based on a GSA schedule;" "Based on 15.405 a, price negotiations, the CO determined price fair and reasonable based on the negotiation that met the FMP."

    Researcher concludes that CO believed that the negotiations of an FMP allowed the CO to justify price without conducting or documenting price analysis since other information that helped justify the price was missing from file.

    Researcher notes some confusion on fair market pricing, especially for services. Not only does FMP need to determine the rates are fair and reasonable, but also evaluate the hours and labor mix, except for historical 8(a) as noted in the FAR citation that follows. Also as a type of analysis, fair market prices still need to be justified.

    According to FAR 19.807, in estimating the fair market price, "The CO shall estimate the fair market price of the work to be performed by the 8(a) contractor. In estimating the fair market price, the CO shall use cost or price analysis and consider commercial prices for similar products and services, available in-house cost estimates, data (including certified cost or pricing data) submitted by the Small Business Administration (SBA) or the 8(a) contractor, and data obtained from any other Government agency. In estimating a fair market price for a repeat purchase, the contracting officer shall consider recent award prices for the same items or work if there is comparability in quantities, conditions, terms, and performance times. Comparison of commercial prices for similar items may also be used."

1f. What Are the Most Predominant Price Analysis Techniques Used in Purchasing Services?

Figures 9 and 10 depict responses to the question by percentages, then numbers for Contract File Data–Answers to the Question: "What Was the Documented Justification for Price Reasonableness for Services?" This offers insight into the predominant type of price analysis techniques exercised in purchasing services.

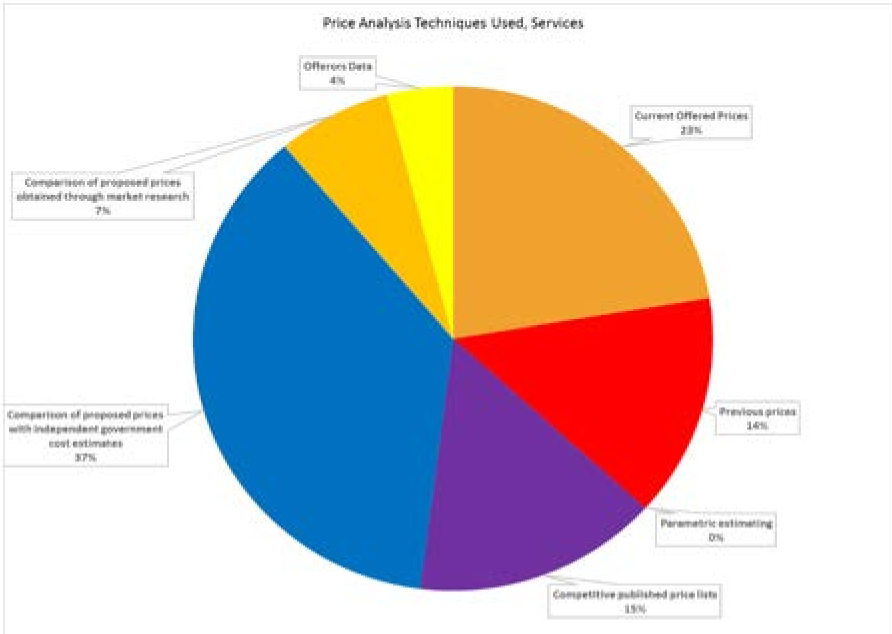


Figure 9. Contract File Data—Price Analysis Techniques Used, Services (Pie Chart)

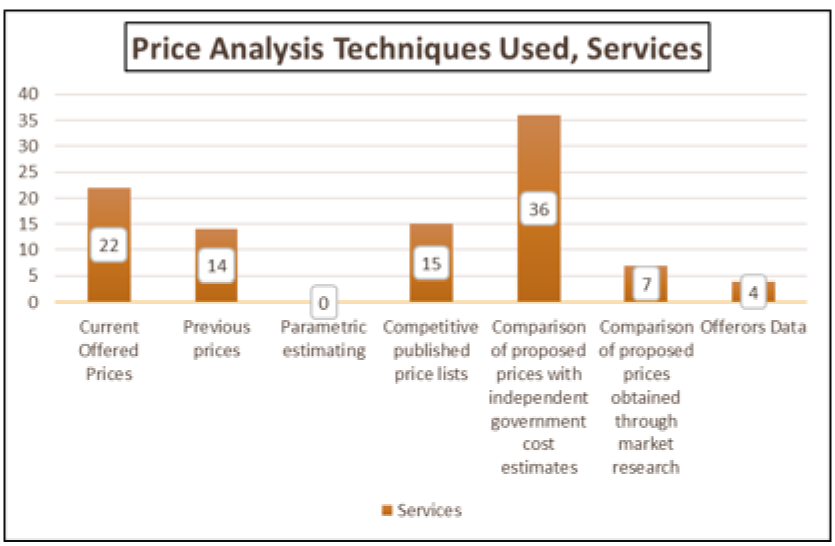


Figure 10. Services—Price Analysis Techniques Used

Findings

According to the contract file data sampled, the services reviewed compared previous prices and competitive price lists equally as price analysis techniques.

Services were very dependent on IGCEs. Thirty-six of 98 cumulative techniques used for services, or 37% of the time, service contracts tapped IGCEs as a primary price analysis technique. The next highest was current offered prices with 22 documented cases in the service files, or tapped 23% of the time.

Analysis

Out of the 58 service files we reviewed, 36 (62%) of the files used the IGCE for comparison. For services, there is more dependence on IGCEs to make price comparisons since IGCEs generally include an estimate of labor hours by the type of effort required. IGCEs are more effective for justifying the price of services than other price analysis techniques outside of two or more currently offered prices.

Summary

Overall, the use of price analysis techniques is common, but serious deficiencies hamper the correct use of those techniques and limit proper supporting documentation. Poor documentation to support the price reasonableness determination was the biggest weakness in the files examined. Competition was limited in establishing price reasonableness. The most frequently used techniques for determining price reasonableness within the files reviewed were comparisons to competitive price lists, comparison through market research, comparison to previous pricing, and comparison to IGCEs. The use of indexing and a statistically stratified sample appeared in a couple of files, but not regression and parametric analysis or other quantitative methods; however, contracting professionals interviewed know the techniques and have been trained to use them.

Consistent with DoDIG report findings, it appeared DLA contracting activities are concerned with high workloads and shortages of qualified personnel. The reviewers can appreciate how the workload and shortages may compound pricing inaccuracies and poor IGCEs. However, the number of unjustified pricing memos we reviewed is worrisome. Table 4 shows that over $61 million in services were not adequately justified for price reasonableness.

Table 4. Summary of Unsubstantiated Pricing Memos by Service, Supply, and Commercial Value

	Percent of actions justified	Potential value not justified	Potential value justified	Percent of value not justified
Contract Files	54.5%	$68,926,782	$65,778,678	49%
Services	55.1%	$61,765,782	$59,135,156	51%
Supplies	0%	$7,000,000	$0	100%
Construction	66%	$161,000	$6,643,522	2%

Overall Recommendations

In a sole-source environment, determining commercial item prices for services to be fair and reasonable can be very challenging. However, contracting personnel should be able to obtain enough information to determine price reasonableness. If not through data available, then from each offeror. The limited technical evaluations reviewed under the auspices of cost analysis were not an evaluation, but more of acceptance. The following overall recommendations may be considered for implementation:

- A well-written checklist would be helpful to both contracting personnel and file reviewers. It would improve consistency by defining exactly what needs to be in the pricing documentation. A checklist should include a section on pricing. Reviewers saw other checklists provided to contracting personnel as a means to check off any FAR/DFARS/DLAD requirements, pre-award administration policies and procedures, but little on price analysis. The use of a checklist makes it easier for contracting personnel to at least identify the type of price analysis used in an award decision and pricing memo, instead of just writing it in.
- Consider examining what is preventing contracting personnel from performing price analysis properly, such as the following:
    - Determine whether current assessment methods consistently follow price reasonableness standards in accordance with the FAR/DFARS-PGI.
    - Train and retrain contracting personnel on price analysis techniques in determining price reasonableness along with what is proper support documentation for pricing.
    - Determine why offeror data is not requested more often. Only four files contained data requests from the offeror.
- Eliminate or reduce the challenges that contracting personnel have in executing proper price reasonableness as discussed in the interviews.
- Provide guidance to contracting personnel on how to assist and guide their personnel in preparing IGCEs and market research reports. This should be in line with any guidance provided to IGCE personnel, such as the IGCE Memo for distribution in the DLA entitled "Documenting the Independent Government Cost Estimate."
- Confusion about the use of GSA Federal Supply Schedules and compliance with FAR 8.404. Contracting personnel did not seem to follow 13 March 2014 DPAP policy directing COs to make price reasonableness determinations using FAR 15 in lieu of FAR 8.404. See http://www.acq.osd.mil/dpap/policy/policyvault/USA001004-14-DPAP.pdf
- Author suggested solicitation language to request additional price data that will help contracting personnel make a fair and reasonable determination when competition is not expected.

References

Defense Federal Acquisition Regulations Supplement (DFARS)/Procedures, Guidance, and Information (PGI), ch. 215 (2016).

Federal Acquisition Regulation, 48 C.F.R. ch. 15 (2016).

Federal Acquisition Regulation, 48 C.F.R. ch. 13 (2016).

Federal Acquisition Streamlining Act of 1994, Pub. L. No. 103–355, §§ 8001–8003, 108 Stat.

3243 (1994).

Fox, P., Gera, R., & Maddox, J. (2014). *Price analysis on commercial items purchases within the Department of the Navy.* Unpublished manuscript, Acquisition Research Program, Naval Postgraduate School, Monterey, CA.

GAO. (2009, March 9). *Cost estimating and assessment: Best practices for developing and managing capital program cost*s (GAO–09–3SP). Retrieved from http://www.gao.gov/products/GAO–09–3SP

Gera, R., & Maddox, J. (2013). *Price analysis on commercial items purchases within the Department of Defense.* Unpublished manuscript, Acquisition Research Program, Naval Postgraduate School, Monterey, CA.

Office of the Deputy Director of Defense Procurement and Acquisition Policy for Cost, Pricing, and Finance. (2012). *Contract pricing reference guides*. Retrieved from https://acc.dau.mil/CommunityBrowser.aspx?id=406579&lang=en-US

Office of the Secretary of Defense (2007, September). *Contracting practices—Independent government estimates, government surveillance, and contractor quality assurance* [Memorandum]. Washington, DC: Author.

Office of the Secretary of Defense (2013, January). *Mission installation and contracting command: Independent government cost estimate guide.* Retrieved from http://www.acq.osd.mil/dpap/sa/docs/MICC%20IGCE%20GuideApril2013[1].pdf

Ginman, R. (2014, March 13). *Class deviation determination of fair and reasonable prices when using federal supply schedule contracts* [Memorandum]. Washington, DC: Office of the Under Secretary of Defense for Acquisitions, Technology, & Logistics.

Redfern, A., Nelson, E., & White, M. (2013*). Price analysis on commercial item purchases within the Department of Defense* (Joint applied project). Montery, CA: Naval Postgraduate School. Retrieved from http://calhoun.nps.edu/public/handle/10945/37743

Rumbaugh, M. (2010). *Understanding government contract source selection*. Vienna, VA: Management Concepts.

THIS PAGE INTENTIONALLY LEFT BLANK

Panel 6. Enabling Technology Transition

Wednesday, April 26, 2017	
1:45 p.m. – 3:15 p.m.	**Chair: Michael McGrath,** Consultant and Senior Technical Advisor, McGrath Analytics LLC ***Accelerating Innovation—From S&T Labs to Acquisition Programs*** Venkat V. Rao, Defense Acquisition University ***Effectively Implementing Policies That Mandate the Use of Technology—A Grounded Theory Study*** Elke Drennan, Carnegie Mellon Software Engineering Institute ***CREATE: Enabling Innovation With Computational Prototyping and Supercomputing*** Douglass E. Post for the CREATE Team

Michael McGrath—is an independent consultant. As a former Vice President at Analytic Services Inc. (ANSER), he led business operations in systems and operations analysis. He previously served as the DASN (RDT&E), where he was a strong Navy proponent for improvements in technology transition, modeling and simulation, and test and evaluation. In prior positions he served as Vice President for Government Business at the Sarnoff Corporation (former RCA corporate lab); ADUSD for Dual Use and Commercial Programs in the Office of the Secretary of Defense (OSD) with responsibility for industrial base and commercial technology investment programs; Program Manager at the Defense Systems Research Projects Agency (DARPA) where he managed manufacturing technology programs; and Director of the DoD Computer-aided Acquisition and Logistics Support program, automating the interface between DoD and industry for technical data interchange and access. His early government career included positions in Logistics Management at Naval Air Systems Command and in Acquisition Management in OSD. He has served on Defense Science Board and National Academies studies, and is an active member of the National Defense Industrial Association (NDIA), the National Materials and Manufacturing Board, the Board on Army Science and Technology, and several university and not-for-profit advisory boards.

Dr. McGrath holds a BS in Space Science and Applied Physics and an MS in Aerospace Engineering from Catholic University and a doctorate in Operations Research from George Washington University (where he also served as adjunct faculty).

Accelerating Innovation—From S&T Labs to Acquisition Programs

Venkat V. Rao—serves as Professor of Acquisition Management and currently teaches Acquisition and Program Management courses for the Defense Acquisition University (DAU). Rao is part of the faculty of the Midwest Region, which is responsible for providing professional education and training to approximately 30,000 Defense department personnel in the 12 states within the Midwest. Rao is also the site lead for the DAU-Midwest Sterling Heights office and has broad program management and product development experience in the technology industry spanning 30 years. [venkat.rao@dau.mil]

Abstract

The current emphasis on innovation in the DoD and the imperative to maintain our technology edge in DoD weapons and vehicle systems requires that all avenues to meet this objective be reviewed. In this paper, we look in the DoD research labs and, specifically, the Tank Automotive Research, Development Engineering Center (TARDEC), to understand the challenges and successes of integrating and transitioning promising technologies into fielded programs. The paper examines several core activities and TARDEC's role in bringing new technologies into programs. The paper describes the successes and challenges of technology integration and transition, and abstracts the systemic issues in the process. The objective is to identify process changes to address these issues, thus providing another path to maintaining, and in some cases, establishing the DoD advantage in various performance aspects of weapon and vehicle systems.

Introduction

The DoD and the Army are focused on technology innovation, especially with respect to weapons systems, to maintain the overwhelming superiority we have achieved over the past 75 years. There is concern that our technology superiority is decreasing and that new cyber and other threats are emerging, thus driving the urgency to maintain superiority in systems with a focus on driving innovative technology into programs. Innovative solutions might originate from basic research conducted at universities, the Defense Advanced Research Projects Agency (DARPA), and the Army laboratories. These are then refined and further developed in the Research and Development Engineering Centers (RDECs). The developments from the RDECs are transitioned to programs via requirements and prototypes to program offices, which are then reflected in solicitations to OEMs and prime contractors to be realized in programs.

This topic has been studied over the years, and many best practices have emerged and are in use across the systems and technology enterprise. In this paper, we are focused on the Tank Automotive Research, Development Engineering Center (TARDEC) and have used TARDEC and RDEC interchangeably. TARDEC supports the PEO Ground Combat Systems (GCS) and PEO Combat Support and Combat Service Support (CS&CSS) and is focused on combat and tactical vehicles and systems. In this paper, we have studied and are reporting on experiences of personnel from TARDEC and program offices to understand best practices that have been used and challenges encountered in innovation and transitioning technology to programs.

Methodology

The approach included a questionnaire that covered several topics:

- Current state of technologies developed in labs

- Technology development in universities and other organizations
- Changes required to drive technology from labs to programs
- Program success in technology transition
- Participation in technology transfer programs
- Management practices in technology integration
- Aids and barriers to successful integration
- Integration issues
- OEM/Contractor dependencies
- Communication and organizational alignments
- Crossing the chasm from technology development to programs

The interviewees included senior-level leaders or directors, chief engineers from PEO offices, mid-level systems engineers, and integration engineers with significant experience in RDEC and program offices. Traditional studies have focused on measuring results by, for example, the number of programs transitioned. In this study, we attempted to capture from the practitioners what has worked and issues they have encountered to get an unfiltered view of technology innovation and technology transitions.

This is a very broad field, and we have tried get an in-depth view of a small portion of the enterprise. As such, we make no attempt to broadly generalize these results to other RDECs even in the Army because programs and underlying technologies can vary significantly across PEOs and RDECs. However, some of the communication, planning, and alignment aspects discussed in this study should be applicable to other organizations as well.

Discussion of Interview Results

Current State of Technologies Being Developed in Labs

To establish a baseline on the usefulness and utility of technologies being developed in the labs, we attempted to understand the interview subjects' views on the value of these technologies with specific examples.

The consensus view was that significant technology is being developed in the labs and that many specific technologies are tied to programs of record (POR). Examples include lithium ion batteries in the form factor required by army vehicles, light weight track, transport armor on JLTV, combat transmission systems, advanced combat engines, active development of a combat vehicle prototype, lightweight armor, integrated modular occupant protection, manufacturing techniques for welding improvements, and composite armor.

One view is that labs and TARDEC in particular are very good at niche technologies such as armor, but that they place less emphasis on the development and integration of emerging technologies such as mobility solutions in the commercial industry automotive industry. That said, recent initiatives suggest that technologies from the automotive industry are being explored for prototype development.

Innovation and Development in Universities and Other Organizations

We next compared innovation and development in the labs with innovation and development in other organizations such as universities, non-governmental labs, and defense contractors.

Some views stated that basic research and some level of applied research originated from universities but development and effort was required to transition this research to RDECs. Others noted that universities supported science and technology in other ways, for example, by providing tools for systems engineering. An example of a tool is the Vehicle Health Management capability at Wayne State University that was supported by the program office.

Visibility into IRAD spending by defense contractors is limited, so it is not clear whether innovation is coming from defense contractors. In addition, it appears that defense contractors' spending on IRAD was focused on requirements in programs of record and not specifically on innovation. Therefore, unless innovation was reflected in the requirements, it would not make it to programs of record. Some interviewees suggested following up the requirements development with prototypes or demonstrators by the labs to further inform the requirements. Another point of view suggests that, in some instances, it is difficult for TARDEC to take innovation risk. To accelerate innovation, then, TARDEC has had to use non-traditional defense partners under non-Federal Acquisition Regulations (FAR).

Another factor that impedes innovation is cost, both for development and unit cost of production. This also leads to risk minimization both by the program office and the defense contractor.

One conclusion that can be drawn from these responses is that universities, Army research labs, and DARPA provide basic research. Applied research, however, must come from RDECs, which must also oversee early phases of the acquisition cycle, while program offices manage engineering and development in the later phases of the acquisition cycle via defense contractors. In this way, RDECs would provide the valuable input to programs with well-informed requirements derived from advanced development and prototypes. Program offices would then be better positioned to solicit solutions from contractors.

Changes Required to Drive Technology Developed Into Programs

A recommended strategy to drive technology from the RDECs to programs is by getting the Army Training and Doctrine Command (TRADOC) to drive requirements to both the RDECs and program offices. This would allow the development of technology and prototypes in the RDECs, and, because program offices would be working to the same requirements, the transition from the RDECs to programs would be easier. A second potential feature of the above approach would be to get results from RDEC efforts to feed requirements in RFPs.

There are other systemic issues however. First, PMs are focused on incremental changes to programs managing risk and cost and shorter term EMD goals, while the RDECs are focused on technology goals and revolutionary changes, and the Army is focused on strategic initiatives. These viewpoints need to be aligned to enable an efficient approach to innovation. Second, funding requests need to be aligned. Currently, PMs' Program Objectives Memoranda include funding requests for what they know now and not for what may be coming down the pike. However, on occasion, PMs also require technology upgrades or better "gizmos" to address threat and priority changes and need a quick turnaround; while the RDECs are willing to respond, the technology maturation can be a lengthy process. This is especially true of technology changes to the underlying vehicle platform.

So with little coordination and no targeted funding, transitioning from the RDECs to programs presents serious challenges.

Program Success in Technology Transition

This topic revealed several views on transitioning technology. One view is that there is no discernible trend in the rate of successful integration. This then leads to a question about the availability of metrics that measure successful integration, the lack of which makes it difficult to understand the roadmap to successful integration. Another view is that integration success has been opportunity-driven, especially over the last two years, exemplified by the Armored Fighting Vehicle, which has seen several integration efforts that have informed the requirements process. A different approach to measuring the success of integration must start with the definition of success. For example, with respect to armor capability, the integration efforts in TARDEC informed the requirements for final design, which led to an improved vehicle. Thus RDEC provided a feasible solution, which drove to a better product even though the specific work effort may not have transitioned.

The strategy of modernizing rather than initiating new development programs led to modernizing at the lower level subsystems rather than the system level. The focus on the subsystems has led to successful integration efforts; however, when contingency requirements drive modernization, the enhancements are fielded, but sustainment efforts are challenged due to lack of program ownership. This lack of program ownership also limits additional quantity buys and further improvements. Finally, there is the view that integration efforts are hampered by the level of oversight at various leadership levels and contribute to lower success rates in technology integration.

The summary of the above discussion leads to several conclusions:

- Successful integration needs to be defined to include informing requirements in addition to transitioning specific development into programs.
- Appropriate metrics to capture all the value of RDEC efforts need to be defined and captured.
- Integration and transitions must have a program owner identified even for efforts driven by contingency requirements.

Aids and Barriers to Successful Integration

Successful integration has been driven by demand from the acquisition or PM functions for risk-reduction efforts or capability improvements. Several examples of successful integration include armor protection, Victory architecture, throttle control software, and lightweight track among many others. In addition, contingency requirements, urgent fielding requests, and a focus on controlling sustainment costs also drive integration. Natural conflicts exist between RDEC leadership and PEO/PMs, as they operate in different environments with different objectives, but strong personal relationships can overcome these conflicts and lead to successful integrations.

Challenges to successful integration include an understanding of the technology development effort as either being exploratory in nature, which would require close cooperation between the RDEC and program office to mature and succeed, versus a specific deliverable that can be driven by a transition agreement. Transition agreements ensure that the RDEC capability meets schedule and functional requirements and that the program is ready to accept the deliverable. Another challenge to integration, expressed by one interviewee, was that PMs focus on thresholds and “do not lose sleep on objectives,” while TARDEC focuses on objectives. With limited interaction between PM engineers and TARDEC, technology is not incorporated into programs. If PMs can focus on objectives and drive the process, TARDEC can help reduce the risk in programs; so, while a specific technology integration may not be the final result, informed requirements and risk mitigation might be the benefits that need to be recorded. Finally, several interviewees articulated that limited funding was a common challenge and a significant barrier to integration.

To summarize, lack of funding can be a significant barrier to integration and transitioning of RDEC efforts to the programs. PM demand for improvements, risk mitigation, and contingency requirements are the primary drivers for successful integration. In all cases, a thorough understanding of the specific technology development and integration effort versus conceptual deliverables must be defined and agreed to between the RDEC and the program office.

Participation in Technology Transfer Programs

Several technology development and transfer programs have been in existence along with various funding avenues. The following are some of the programs:

- Joint Concept Technology Demonstration
- Foreign Comparative Testing
- Quick Reaction Fund
- Rapid Reaction Fund
- Small Business Innovative Research (SBIR)/Small Business Technology Transfer(STTR)
- Agile Integration and Development
- Collaborative Technology Alliances
- Technology Transfer Initiatives
- Technology Enabled Capability Demonstration (TECD) and Army Technical Objective (ATO)

All of the above programs have been used on different technology development efforts; SBIR and STTR have been used extensively. The JCTD was used for the Trailer program; Foreign Comparative Testing was used on the Howitzer program; and Agile Integration and Development was used on the lightweight track development. Several Technology Transfer Initiatives have been used at TARDEC.

These initiatives are a necessary condition for technology innovation and transition to be successful, and their widespread use must be viewed as an advantage for TARDEC and the PMOs.

Management Practices in Technology Integration

Management practices also bear on the possibility of technology transition. The following are some of the common practices:

- Roadmap reviews of technology plans between organizations
- Technology requirements and alignment with program requirements
- Formal collaboration between TARDEC and Program Offices
- Technology Transfer Agreements or equivalent
- Metrics to measure success of technology integration

We determined that several of these practices are in use, including requiring twice-a-year requirements reviews, a 30-year plan review, strategic engagement at the leadership level, and soliciting long-range input from TRADOC. Systems agreements are another mechanism to align TARDEC and PMO offices. One caution was sounded on TTA (Technology Transition Agreements), which are binding on the RDEC to deliver and on the

PM to implement except when PMs are not sure of the risks and challenges of integration. TTAs can be signed without guarantees, which can impact future transition efforts.

These practices are necessary conditions for successful transitions but need a greater emphasis on the alignment to ensure PMOs and RDECs are working towards the same program goals. In some instances, these efforts will focus on emerging requirements and may not result in capabilities that can be transferred to programs. In other instances, specific technologies are ready to be integrated into programs of record. The key factor in successful transitions is the agreement and alignment between the RDEC and the program offices.

Integration Issues

A discussion on integration issues identified that a propensity for low risk drives both RDEC and PMOs. PMs focus on low technical risk to ensure programs can meet objectives, while the S&T manages to a TRL 6, resulting in few game changers and fewer revolutionary improvements. In some instances, a 5% improvement in combat vehicles can take up to 10 years. A second issue that arises in integration is the lack of clarity around the integrator's role and who plays this role. Informal requirements communicated to the RDECs can lead to technology demonstrators, but the work effort to convert these technology demonstrators to a capability with specific platform-integration goals is a gap that needs to be filled to facilitate successful transitions. The lack of alignment of RDEC strategic priorities with PM requirements results in funding issues to integrate and transition technology from the RDEC to programs. The lack of coordination results in some efforts getting funded and in other instances the PM turning to the OEM or prime contractor to meet the requirement.

OEM/Contractor Dependencies

Innovations from RDECs go through multiple development efforts and eventually require integration by OEMs or prime contractors. The TARDEC prototype integration facility and systems integration lab have led to clearer requirements and informed the integration efforts by the prime contractors. The efforts on the Active Protection System and double V hull based on an open systems architecture resulting in transition specification are examples of TARDEC development efforts that have or will result in a handoff to the contractor or OEM for integration into a vehicle platform for production. Integration into vehicle platforms will require knowledge and expertise in manufacturing and high-volume production. The OEM or prime contractor is the appropriate resource. This discussion highlights the interdependencies between the program office and the RDEC in ensuring that funding is available for the RDECs to develop the technology and prototypes for integration. Solicitations can then reflect the design requirements that emerged from the RDEC efforts, and funding can be programmed for development and production to include these transition efforts.

The consensus opinion of the interviewees was that 90% of the transition efforts in the RDECs will require integration by the prime contractors and the OEMs, and funding and resources must therefore be planned.

Communications and Organizational Alignments

Organizational changes such as reorganizing TARDEC research groups to focus on programs and the appointment of CIEs (Chief Integration Engineer) have supported technology transition efforts. Formal communications between PMO and TARDEC are also critical for continued integration success. The Active Protection System is an example of how significant input from the PMO has influenced the project's development in TARDEC. Senior leadership summits, review of 30-year planning documents, subject-matter-expert communications, and formal exchanges supplemented by personal relationships reinforce the importance of communication in supporting technology transitions.

Crossing the Chasm

In this section, we cover the changes that are required to successfully cross from the RDECs to programs. Greater interaction between subject matter experts in the labs and program offices on an ongoing basis, including participation in preliminary and critical design reviews should benefit the technology transition process. The use of TRADOC to drive requirements to both the RDECs and PMOs with a short, medium, and long-term horizon should assist in directing RDEC efforts appropriately, and when coordinated with PMOs, should support the development of a strategic and tactical plan.

There is a recognition that mobility initiatives in the commercial automotive and transportation industries are driving innovation that should be explored for use in programs and by TARDEC. TARDEC is using an innovative business approach to allow a consortium consisting of small technology companies, larger defense companies, and automotive companies to develop prototypes based on technologies emerging in the auto industries, such as autonomous vehicles. These prototypes can then inform requirements, capabilities, and designs for future solicitations. Similar consortiums are also being used to explore other technology prototypes in areas such as sensors and robotics that have more immediate applicability to ground vehicles, both combat and tactical. These consortia respond to requests for technology prototypes under Other Transaction Authority (OTA) provided by Congress to the DoD specifically to attract nontraditional defense and technology companies. The OTA provides more flexibility in dealing with smaller companies in the areas of intellectual property rights and speed to market when compared to contracting under the FAR regulations.

Funding program transitions explicitly is a key driver of ensuring transfer of RDEC efforts to programs. Leadership plays a key role in the realization of successful integration efforts by supporting the successes and challenges of technology transitions and stressing the importance of these efforts.

Current Literature on Technology Transitions

A limited survey of literature on technology transitions suggests similar views to those discussed in this paper. In *The Future of Army Science and Technology Requires Punctuated Equilibrium*, Col. John R. Cavedo (n.d.) describes a broad strategy for reorganizing and managing Army Science and Technology with a new business model. One of the key recommendations from this paper is that S&T should "focus less on technology transition and more on proving the value of technology through prototyping and requirements validation. This will require additional 6.4 funding." The paper also refers to the Deputy Assistant Secretary Research and Technology (DASA RT) goal of "Align S&T and develop strategies which provide technology insertion points to Programs of Record." The paper posits that "by focusing too much on technology transition there is a high probability that S&T managers will avoid risk and won't push the boundaries of technology advancement," and that "transitioning an S&T effort to a program of record is fraught with blind spots."

DASA policy on Transition Agreements (TA) for Army Science and Technology Projects requires TAs be developed for all Advanced Technology Development (6.3), Advanced Component Development and Prototypes (6.4), and Manufacturing Technology (6.7) executed projects. The TA captures RDEC responsibilities and deliverables, PEO or Recipient responsibilities and mutual responsibilities. This template provides the basis for an organized formal process for transition agreements that should lead to success.

In "Bridging the Valley of Death," Anthony Davis, Director of Agile Acquisition for U.S. Special Operations Command, and Tom Ballenger, aviation systems analyst with JHNA Inc., have outlined a Transition Confidence Level Scale from 1 to 9 similar to the Technology Readiness Level model to track the steps for transition from uncertainty to a completed transition (Davis & Ballenger, 2017). "Like the TRL chart, the steps enable status scoring for a project, and form a roadmap for progress and coordination typically needed for transition success." This promising approach enables a data-driven standardized approach to measuring the progress of technology transitions. Similar models have been recommended in the RDECs, and these models allow for agreements with program offices on metrics to support the transition, and these can also be added to the TA template from DASA RT.

This study and the above literature have focused on the Army S&T enterprise, but future studies should review the best practices in the Air Force and Navy for possible use by Army RDECs.

Conclusions

- Universities, laboratories, DARPA, and defense contractors play a role in technology innovation, but primarily RDECs support the realization of the innovation and its transfer to programs.
- OEM contractors perform best when requirements are informed by advanced development and prototypes from RDECs developed internally or through non-traditional technology development organizations. There exists a significant interdependency between the RDECs and OEM contractors, with PMO offices interfacing with both; 90% of integration efforts will require OEM contractor participation.
- RDECs need to extend their reach by partnering with non-traditional companies to drive technologies and capabilities from smaller companies or large non-defense contractors into programs.
- Integrated efforts by TRADOC, RDECs and program offices to coordinate a view of the future both in the short term and medium term would benefit innovation and transition efforts.
- POM funding requests should be aligned so that they include not only the short- to medium-term requests but also include funding to translate innovation into prototypes and transitioning into programs.
- RDEC value is realized via many different avenues: transitioning technology prototypes to programs, informing requirements to improve solicitations, and engineering or manufacturing process improvements. Appropriate metrics to measure this value must developed. Transition Confidence Levels or other similar measures are a valuable tool to assess transition efforts and direct resources.
- Lack of funding is an obvious roadblock to technology innovation and transition, while risk-mitigation-based demands from program offices, contingency requirements, and threat changes seem to be significant drivers of innovation.
- Technology Transfer programs like SBIR, STTR, JCTD, TECD, and ATO are a necessary condition to foster innovation.
- OTA can be used to engage non-traditional companies to access the latest technology from smaller companies and the commercial marketplace.

- Integration and transitions must have program owners in both the RDEC and program offices who are aligned and in agreement to deliver successful transitions.
- Communications at the strategic level, communication of long range plans, senior leadership summits, and exchanges between subject matter experts are all important activities to support effective transitions.

Crossing the chasm from technology innovation and development to programs successfully requires many organizations and activities to come together. The conclusions above reflect many practices that are currently in place. The challenge, however, is to align the practices and operate them as part of a system for effective transition of innovations.

Figure 1 depicts an operational view of a process for technology transfers

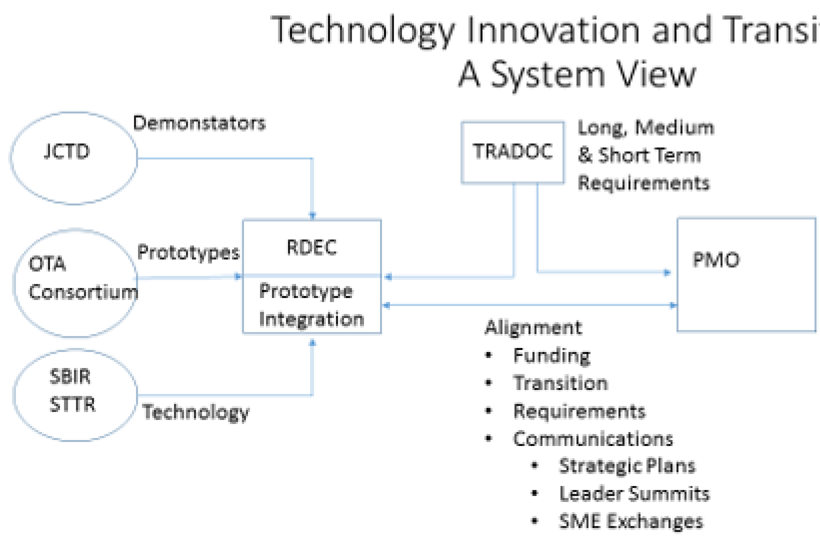


Figure 1. Technology Innovation and Transition: A System View

References

Cavedo, J. R., Jr. (n.d.). *The future of Army science and technology requires punctuated equilibrium*. United States Army War College.

Davis, A., & Ballenger, T. (2017, January–February). Bridging the "Valley of Death." *Defense AT&L.*

Effectively Implementing Policies That Mandate the Use of Technology—A Grounded Theory Study[1]

Elke Drennan—works at the CERT division of Carnegie Mellon University's Software Engineering Institute, a federally funded research and development center. Dr. Drennan speaks three languages and has degrees in the following subject areas: business, psychology, computer information systems, and a doctorate of management with a specialization in information systems technology.

Abstract

A challenge faced by organizations globally is a workforce reluctant to use collaboration tools. Leaders invest large percentages of their budgets in information technology (IT) solutions, but often see little in return (Tirgari, 2012). The purpose of this grounded theory study was to explore how employee perceptions about organizational policies that mandate the use of technology affect the acceptance, use, and perceived productivity thereof. Eighteen participants of a major IT command responded to nine open-ended interview questions. Data analysis involved open, axial, and selective coding of the participants' responses, which produced three major themes and 13 sub-themes. The three major themes were leadership, policy, and mandated tool. The findings from this study offer leaders a theory that proposes numerous ways to more effectively implement organizational policies that mandate the use of technology. By following the recommendations of this study, leaders can expect gains in compliance and worker productivity.

Introduction

In today's global economy, organizations are becoming increasingly geographically dispersed and, therefore, have come to rely on technology for communication and collaboration (Kirkman et al., 2002; Saraswat, 2012). Interest and research have shifted specifically to knowledge management (KM) and collaboration tools for effective information sharing within organizations (Hew & Hara, 2007; Kim & Lee, 2006). Leaders must understand how their policies affect employees' use of these tools and their resultant perceived productivity (Garicano & Heaton, 2010). Although information technology implementations may by themselves increase productivity, if they are not complemented by organizational policies, these productivity gains are less significant (Garicano & Heaton, 2010). Conversely, IT implementations have been shown to hinder productivity if complementing policies are not in place (Thielst, 2007). Therefore, organizational policies play an important role in how productive employees will be when using technology (Tirgari, 2012).

In the organization that was studied, a large Department of Defense (DoD) IT service provider pseudo-named ITCOM (IT Command), leaders established formal policies that require employees to use the collaboration tool Microsoft SharePoint for specific tasks such as routing documents, sharing intellectual property within the community, daily check-ins, and posting announcements. This study explored, qualitatively, how these policies affected employee use of technology and the workforce's perceived productivity. Understanding how policies affect their workforce will provide leaders with the necessary insight to implement policies for maximal effectiveness. This knowledge will allow leaders to adapt how they

[1] This paper is a summary of Dr. Drennan's dissertation.

structure their policies and implement them in a manner that will improve worker receptiveness and, thus, increase worker productivity.

Background of the Problem

To become more efficient, organizations are continually automating tasks that were once accomplished manually, often in a face-to-face fashion (Austerberry, 2011; Reichley, 1997). For example, instead of routing a form through a lengthy approval process by carrying it to each approving individual's desk for signature, workflows can be implemented in SharePoint that automatically route forms electronically in a more efficient matter. Introducing technology into every day work results in a significant change for employees, not just in how they do their work, but also in their organizational culture (Borck, 2001; Malik & Danish, 2010; Nunamaker, Reinig, & Briggs, 2009; Walsh & Maloney, 2007). People have difficulty with accepting change, and changes that involve technology are even more complex to manage (Sun & Zhang, 2005; Yi et al., 2005).

Because of people's inherent resistance to change, employees are not likely to embrace newly automated processes (Long & Spurlock, 2008). To ensure that employees do, in fact, make use of new technologies, some employers are forced to mandate the use of newly implemented tools (Sun & Zhang, 2005). After making significant investments in technology, leaders expect to see a return on their investment, such as increases in efficiency or productivity (Reichley, 1997; Sun & Zhang, 2006). However, employers who mandate the use of technology can expect to be faced with further resistance (Sun & Zhang, 2005). Therefore, leaders may create organizational policies that require employees to use technology for specific job functions (Garicano & Heaton, 2010; Nyström, 2006).

These policies may be distributed via technology, such as e-mail or a notification portal like SharePoint, or be shared face-to-face. The manner in which these policies are implemented can significantly impact employees' use or refusal to use technology (Sun & Zhang, 2005). If employees feel that they were not involved in the change, or do not see the value in automating a process, they will likely resist the implementation (Chin, 1998). However, if policies were implemented in a more effective manner, employees will likely perceive them more positively and will be more likely to comply with them (Tirgari, 2012).

Purpose of the Study

The general problem that created the need for this study was the desire to understand the role that policy plays in technology usage. Organizations worldwide rely on collaboration tools for both knowledge management and knowledge sharing, and although the correct technology may be implemented in an organization, employees do not use these systems effectively, if at all (Germain, 2011; Nyström, 2006; Qureshi, Liu, & Vogel, 2006; Workman, 2007; Zivick, 2012). A number of hindrances contribute to this situation, such as a lack of management commitment, the employees' expectations, reward systems, and training (Kim & Lee, 2006; Nyström, 2006). Although previous studies have thoroughly explored a number of these obstacles, very little research exists on the relationship between organizational policies and IT implementations. To fill the gap in the body of knowledge, this grounded theory study explored how organizational policies that mandate the use of technology impact employees' use of said technology and information sharing and whether employees perceived themselves to be more productive.

Significance of the Study

To achieve the maximum benefit from technology implementations and to be globally competitive, leaders must understand how their policies affect employees' usage of tools and resultant perceived productivity. Every year organizations spend a large amount of their budgets on IT, but they do not get the expected return on investment (Tirgari, 2012). The results of this study may help in the recovery of some or all of this investment.

Previous studies have examined obstacles to the use of technology (Zawawi et al., 2011), but they have not explored how employees' perceptions of policies can affect their use of technology and resultant productivity. Existing research has identified that employees may be reluctant to use the tool for a variety of reasons. Employees may not have been involved in the implementation, fear losing their job, or lack the skills necessary to interact with the system (Kotter & Schlesinger, 2008; Nyström, 2006). These studies have also identified that management support is essential for employee buy-in, but they have not examined the employees' perceptions of implemented policies or how those perceptions affect productivity (Garicano & Heaton, 2010; Kotter & Schlesinger, 2008; Long & Spurlock, 2008; Malik & Danish, 2010; Nyström, 2006). Previous studies also have not used a grounded theory approach to study this particular phenomenon. The grounded theory (GT) method has become more relevant to IS research in recent times and should be the likely choice when researchers see truths as socially constructed and when representation is depicted as a distributed systems phenomenon (Bryant, 2002; Urquhart, Lehmann, & Myers, 2010).

A literature review revealed that previous studies fail to provide a deeper understanding of the role that organizational policies play in the interaction with and use of technology. Individual perceptions about policies, whether positive or negative, may impact the effective use of tools. Understanding the relationship between perceptions of policies and perceived productivity is critical to better managing employees. Leaders must create policies that effectively encourage the use of technology and, more specifically, collaboration tools (Tirgari, 2012). Thus, this study contributed to the body of knowledge by developing a theory about how leaders can create and implement organizational policies that will achieve a high level of compliance and encourage the productive use of collaboration and knowledge management tools (Tirgari, 2012).

Literature Review

The literature review is divided into the following sections: globalization, information technology, and virtual collaboration; global virtual teams (GVTs); barriers to IT implementations and information-sharing; successfully implementing collaboration tools; and factors for successful policy implementations.

Globalization, Information Technology, and Virtual Collaboration

As organizations have become increasingly globally dispersed, work team structures have changed from being co-located to being virtual and thereby reliant on technology for communication and collaboration (Germain, 2011; Qureshi et al., 2006; Workman, 2007; Zivick, 2012). Instead of traditional face-to-face meetings, colleagues use video teleconferencing to meet virtually, independent of time and space (DeRosa et al., 2004; Lomas, Burke, & Page, 2008). Fifty to 80% of organizations have teams and at least 61% of employees within large organizations have been part of a virtual team (Germain, 2011).

Team members must collaborate to share information and produce quality work (Durugbo et al., 2011). Collaboration changes the structure and behavior of organizations as a result of pooled expertise and standardized work patterns (Durugbo et al., 2011). Instead of solving problems individually, employees must work together to solve a problem or achieve a common goal (Durugbo et al., 2011; McShane & Von Glinow, 2004). For example, to successfully launch global products, prices, associated services, and technical support, as well as the development of the products themselves, must be coordinated amongst distributed teams (Harvey & Griffith, 2007). The most effective way to do this is through virtual global teams (Harvey & Griffith, 2007). Because these teams consist of culturally diverse, geographically dispersed members, they have a globally diverse perspective (Harvey & Griffith, 2007). However, they have to think in concert to achieve their goals (Harvey & Griffith, 2007). Global virtual teams allow organizations to respond quickly to changes in the marketplace (Beagrie, 2005; Zivick, 2012). Since GVTs are geographically independent, the most skilled workers can be assembled to solve problems (Zivick, 2012). Organizations also save significantly on travel costs by implementing GVTs (DeRosa et al., 2004; Zivick, 2012).

Thus, the way information is shared has also changed significantly. Although e-mail is still a common form of communication, collaboration tools have become widely used for sharing and distributing information (Lomas et al., 2008). Collaboration tools offer a number of benefits to globally dispersed teams. Teams may easily share documents, enforce versioning control, communicate via instant messaging and receive instant feedback, and access their tools from any place at any time (Lomas et al., 2008). Lomas et al. found that collaboration tools, such as instant messaging and video chat, can lead to an increase in sharing personal information amongst students, which can enhance their ability to work together effectively as a team and increase productivity. In virtual business environments, team building can therefore also occur through use of these tools, leading to more effective teams.

Turban, Liang, and Wu (2010) found that social tools, which are composed of wikis, blogs, social networking platforms, and discussion forums, are also used by virtual teams within businesses for decision-making. When using these tools, no team member has to be physically present, and, thus, the decision-making process is expedited, compared to decision-making in traditional settings, where discussions are held in a conference room (Turban et al., 2010). Turban et al. confirmed that collaboration tools do indeed provide numerous benefits for group decision-making, such as expedited information sharing, increased individual input, accelerated decision making, prioritizing and analyzing of solutions, and greater participation.

In a series of in-depth studies of students' perceptions and use of technology, Conole et al. (2008) found that when multiple types of technology are available in learning environments, students will choose the technology most appropriate to their learning needs. Distributed collaborations available from mobile devices, the Internet, and social software changes the way students interact and learn, as well as how they consume and share knowledge (Conole et al., 2008). As these students enter the workforce, they will communicate and collaborate much differently than will their older counterparts (Conole et al., 2008). For these individuals, the time it takes to train on how to use collaboration tools will be significantly shortened; however, they may not be as effective in communicating with their colleagues who are from an older generation (Conole et al., 2008).

Tang and Austin (2009) determined that different types of technologies have varying effects on individuals. In a study of business students, video offered the greatest level of enjoyment, Microsoft PowerPoint improved motivation and learning, and Internet usage was most applicable to future jobs. Further, individual preferences vary by age (Tang & Austin, 2009). Younger generations prefer video, while older generations prefer lectures (Tang & Austin, 2009). By using a mix of technologies, professors can reach the widest audience and increase motivational and learning levels (Tang & Austin, 2009). Although this study was conducted at a university, it is relevant to business leaders. Understanding how different types of technologies affect workers' perceptions can help them address problems in the use of technology and its effectiveness (Tang & Austin, 2009). For example, if instructional videos are used to train users on technology, managers may find that instruction through PowerPoint might be more effective (Tang & Austin, 2009). Being aware of technology preferences by age groups can also help leaders better understand resistance to technology and address employees' needs in a more meaningful way (Tang & Austin, 2009).

Global Virtual Teams

Challenges for GVTs

GVTs face a number of challenges, from creating trusting relationships (Germain, 2011) to general confusion, employee isolation, cultural differences, language barriers, and technological breakdowns (Holland, Malvey, & Fottler, 2009). Some of the ways that managers can address these challenges are through building team identity with an initial face-to-face meeting; promoting open lines of communication with all team members; being easily accessible during working hours; building individual and team trust; developing a sense of team community by identifying and recognizing cultural differences; becoming familiar with the background and experience of each team member; and offering one-on-one meetings with team members, if necessary (Holland et al., 2009).

Although technology seems to offer numerous benefits to virtual teams, there are more than a few disadvantages. Galleta and Zhang (2006) found that technology can be an impediment to good communication if team members do not understand how to use a tool correctly. Additional studies show that employees may lack the proper training or skills to successfully use technology (Bergiel, Bergiel, & Balsmeier, 2006; Chong, 2005), or their personalities may not be amenable to this type of work environment (Bergiel et al., 2006). They may also be resistant to its use because they consider it a threat to their livelihood or because their buy-in to the technology was never obtained, according to Long and Spurlock (2008).

Borck (2001) discovered that because knowledge management tools significantly change employees' work habits, their implementation may be unwelcome. Further, older business leaders may not understand technology well enough to truly make it useful within their organizations (Bergiel et al., 2006). Lanubile et al. (2010) found that new tools must address concerns for incompatibility, be introduced stepwise, and need to be well-supported to be effective.

According to Bushnell (1999), incorrect implementations or functionality problems can hinder the use of technology. Configurations may be wrong or availability may be limited. Based on a survey of IT engineering managers, only 15% were satisfied with their tool's performance, largely due to incomplete implementations of capabilities (Chin, 1998). Brown et al. (2002) found that ease of use and overall usefulness are the primary reasons for users to adopt technology. However, Brown et al. (2002) determined that the mandated use of technology changes the employee/technology relationship, affecting the underlying reasons for technology acceptance. In voluntary-use environments, technology's perceived usefulness is the primary reason for adoption (Brown et al., 2002; He, Fang, & Wei, 2009). When technology use is mandated, individual feelings are irrelevant (Brown et al., 2002). Johnson and Howell (2005) found that when students were required to use a specific type of technology tool, they had a more favorable attitude towards technology and were more likely to use other types of computer-based applications.

Elmholdt (2004) criticized KM tools because he found them to be incompatible with normal knowledge and learning activities. He believed that tacit knowledge cannot be explicated through the application of IT (Elmholdt, 2004). Elmholdt posited that KM technology lacks content-rich information that is present in collegial networks. Further, he believed that the control and ownership associated with this type of technology subjectifies employees as replaceable resources (Elmholdt, 2004). Employees will therefore resist the implementation and use of KM technologies (Elmholdt, 2004).

Creating Successful GVTs

Brake (2008) discovered that because collaborating online is very different from collaborating in person, certain rules for procedures and purpose must be established and agreed upon for teams to work as cohesive units. The way people express themselves in writing has different requirements than do in-person interactions. Team members must also be aware of cultural differences to ensure a respectful relationship exists (Brake, 2008). For example, when addressing someone from Germany, a more direct approach is effective. That same approach would be disrespectful if used when communicating with an Arab.

A study by Rahmati, Darouian, and Ahmadinia (2012) showed that the organization's culture, consisting of values, beliefs, practices, and behavior, shape the behavior of team members. The behavioral norm within an organization directly relates to the norm within teams, dictating which types of behaviors are acceptable. Numerous studies have shown that proper management of organizational cultures is critical for teams to succeed (Brake, 2008; Chong, 2005; Nemiro et al., 2008).

Dubé and Paré (2001) found that different cultural backgrounds and communication styles of individuals in global virtual teams often clash with organizational management styles. To address this problem, the researchers recommended that cultural diversity training be a requirement for GVT members and believe that this training is essential for the success of GVTs (Dubé & Paré, 2001). The training should include many basic elements, such as accountability, expected behaviors, normal working hours, level of involvement, and performance requirements (Dubé & Paré, 2001). Further, trainees must learn how decisions are made, how conflicts are resolved, and how work will be reviewed and approved (Dubé & Paré, 2001). Leaders must also address language and IT proficiency, as well as the technology's accessibility, reliability, compatibility, and its appropriate use (Dubé & Paré, 2001). Distributed teams cannot function effectively if they rely on technology and said technology is only available during limited hours (Galleta & Zhang, 2006). Virtual teams need 24/7 availability of tools to function effectively.

Workman (2007) determined that the differences between the culture in virtual teams and the local organizational culture can also challenge team members. While trying to address situations in their local organization, team members have to work on global problems (Workman, 2007). Formalizing their process structure by making it means-focused versus ends-focused increases both the quality and quantity of work performed by virtual teams (Workman, 2007). Political structures only mildly increased quantity but did not affect quality, but teams that sought information/clarification about their roles and responsibilities improved on both quality and quantity (Workman, 2007). Further, when interpersonal relationships were closely linked to the teams, they had higher quality and quantity output (Workman, 2007). Finally, more tightly controlled teams outperformed loosely controlled ones (Workman, 2007).

Zivick (2012) posited that leaders need sufficient resources to fund their teams and all their requirements to include technology, training, and support. Further, leaders must establish a clear linkage from the teams' goals to the organizations' goals and missions (Zivick, 2012). Doing so gives teams legitimacy and lessens confusion amongst team members (Zivick, 2012). Managers must empower their team members, provide regular feedback, and create a positive, trusting, and structured work environment (Zivick, 2012). By establishing standardized work practices, team socialization norms, and explicit roles and responsibilities, managers can further ensure the success of their teams (Zivick, 2012).

Barriers to IT Implementations and Information Sharing

As with any organizational change, the implementation of collaboration tools will encounter employee resistance. The biggest barriers to IT implementations are not their cost or problems with the technology itself, but rather user resistance, according to Chin (1998). Long and Spurlock (2008) discovered that when implementing IT, leaders face additional challenges. Numerous studies stressed the importance of clear communication and training as being key success factors in managing the acceptance of technology-driven change (Borck, 2001; Chin, 1998; Cogburn & Levinson, 2008; Long & Spurlock, 2008). Long and Spurlock posited that, as with any communication, each worker will interpret any given message differently, based on his or her frame of reference. To effectively communicate, the sender must understand how to best approach the receiver of the message (Long & Spurlock, 2008). According to Long and Spurlock (2008) and Yu (2009), good leaders will learn how to build relationships with their employees, and this relationship must be based on trust. From these relationships, leaders can gain insight into individual preferences, motivations, and resistances to change which they can then address appropriately (Long & Spurlock, 2008).

Hew and Hara (2007) identified that different types of knowledge exist and are shared and that motivators and barriers to information sharing vary amongst career fields. Previous studies have explored online knowledge sharing, but failed to categorize the types of knowledge shared (Hew & Hara, 2007). In their case study, Hew and Hara found that the types of knowledge shared, as well as the barriers and motivators to knowledge sharing, differed by the three professions they studied: Web development, advanced nursing, and literacy education. By observing their study's participants while they were engaged online and by using semi-structured interviews, the researchers found that practical knowledge was most commonly shared (Hew & Hara, 2007). Hew and Hara also identified the most common motivator to knowledge sharing to be reciprocity; the six other motivators were personal gain, collectivism, altruism, respectful environment, personal interest, and technology. The eight barriers were lack of time, a negative attitude, technology, unfamiliarity with the subject, confidentiality concerns, not wanting to cause a fight, and the perceived inability to make use of knowledge (Hew & Hara, 2007). Surprisingly, technology was both a motivator and a barrier to knowledge sharing. By understanding how their employees are affected, what motivates them to share, and what causes them to withhold from sharing, leaders can address these factors so that successful collaboration will occur. Further, as IT matures, user satisfaction and service quality increase and a more positive team-oriented culture results (Hartman et al., 2009). Nevertheless, Mohd and Mohamed (2009) noted that users' resistance to technology does not affect their performance.

Paghaleh, Shafiezadeh, and Mohammadi (2011) identified cultural and political perspectives as impediments to knowledge sharing. They stated that although technology may be readily available in organizations, it may not be used at all (Paghaleh et al., 2011). Therefore, simply implementing it without properly addressing the motives for resistance is a worthless endeavor. Paghaleh et al. found that workers guard knowledge closely because it can be used to gain a competitive advantage. Further, even though technology makes knowledge sharing easier by crossing hierarchical boundaries, informal individual and social networks dictate what is shared and with whom (Paghaleh et al., 2011). By creating an organizational culture that encourages cooperation and sharing and instituting motivational programs that reward team achievements rather than individual efforts, organizations can overcome resistance to information sharing (Paghaleh et al., 2011).

Successfully Implementing Collaboration Tools

As work environments change as a result of IT implementations, organizational cultures must change accordingly. Research shows that organizational cultures must become learning cultures that embrace change (Alcantara, 2009; Malik & Danish, 2010). Workers must be open to new ways of communicating, learn new ways of performing work through the use of technology, continue to develop new skills, and not live in fear of technology. Organizational cultures that support these elements must then be complemented by organizational policies that reflect a changing culture, according to Alcantara (2009) and Garicano and Heaton (2010).

Borck (2001) posited that employees must feed knowledge management solutions to make them useful, and the enterprise KM (EKM) will infuse an organization's culture with knowledge. EKMs control and consolidate data into intellectual assets that bridge the flow of intellectual capital within an organization (Borck, 2001). Yukl (2006) found that transformational leadership inspires followers to act morally and ethically. Transformational leadership also nurtures innovation and makes virtual teams thrive, according to Senge (1998). Cogburn and Levinson (2008) determined that the most effective teams all trust one leader who is culturally sensitive. However, even if a leader is ineffective, virtual teams can still achieve some level of success as long as the teams are cohesive and each member makes contributions (Cogburn & Levinson, 2008). In such instances, work is independent instead of interdependent, and the team's success-level is significantly limited (Cogburn & Levinson, 2008).

According to Bergiel et al. (2006), when leaders implement collaboration tools, they also need to be aware of how both older workers and their younger counterparts will react to and interact with them. The system needs to be accessible and useful to the entire workforce to be effective within an organization (Bergiel et al., 2006). Igbaria and Guimaraes (1994) also found that user involvement during implementation, as well as technology friendliness, led to successful implementation efforts.

Smart and Desouza (2007) discovered numerous ways that managers of small to medium-sized organizations can best address technology resisters and gain their acceptance. For employees who only see how technology affects them on an operational versus a strategic level, managers should create success metrics that apply to the operational level (Smart & Desouza, 2007). This will help employees understand the value of IT from their own perspectives. Managers must, of course, understand the underlying reasons for employee resistance, which varies by employee (Long & Spurlock, 2008). Further, implementations should not be rushed, timing must be appropriate, and leaders must help employees understand the labor-saving value IT brings to the workforce (Smart & Desouza, 2007). Managers can also use social awards and feedback from those employees who embrace technology to make the implementation effort more successful (Smart & Desouza, 2007).

Upon investigating how employed MBA students used collaboration tools, Westerfelt (2010) found that they indeed benefited from using them at work. In her study, Westerfelt had participants use three different online collaboration tools, then gathered feedback about their preferences via questionnaires. Students liked the tools to varying degrees, depending on their practical applications, whiteboard features, document sharing capabilities, and user-friendliness (Westerfelt, 2010). The participants felt that time-savings were the biggest advantage of using collaboration tools, while technological illiteracy, technical issues, and personal comfort level were major impediments (Westerfelt, 2010). Similar to some of the previously discussed studies, Westerfelt also identified user-friendliness as a significant participant concern. When leaders select a collaboration tool, they must remember that regardless of how many special features a tool may offer, if it is not easy to use, it will not be accepted by the user community. Lomas et al. (2008) also found that tools that are the most user-friendly and natural feeling are more likely to be used. Additionally, if a tool is released before its time or is too radically different from previous tools, it will not be adopted (Lomas et al., 2008).

Factors for Successful Policy Implementations

As organizational cultures evolve and begin to embrace the changes resulting from technological implementations, organizational policies must be updated accordingly to reflect the current culture. Research shows that when leaders develop policies, they should involve employees and seek their input throughout the process (Boer, 2012; Kapsali, 2011). Employees must clearly understand the intent and meaning of policies to accept and comply with them (Boer, 2012; Kotter & Schlesinger, 2008; Long & Spurlock, 2008; Nyström, 2006). Kapsali (2011) also found that managers must be flexible when implementing policies, customizing them to specific systems thinking constructs, user groups, and local realities. Further, by including the workforce in the development of policies, managers can ensure that the policies will achieve their intent (Alcantara, 2009; Tirgari, 2012). For example, policies could prescribe a minimum requirement for training and organizationally-funded continued education (to reflect a learning culture) and for changing which former manual tasks must gradually be performed using technology, according to findings by Malik and Danish (2010) and Witte (2002). This could be a phased approach, where either certain tasks or a specific number of tasks must be accomplished using IT. By having prior knowledge of the technology, employees will be better able to determine which changes can be achieved within a specific time period.

Nyström (2006) found that as organizational policies continue to evolve, managers must ensure that they clearly communicate changes or revisions to policies to their workforce. If possible, managers should continue to involve employees in the development of policies to ensure their acceptance. According to Peckover, Hall, and White (2009), as both culture and policies must stay aligned with each other, managers must ensure that collaboration and knowledge sharing are supported by both entities. The organizational culture must be reflective of the open sharing of ideas, and policies should align with collaboration through electronic means. Managers could offer incentives for those departments that notably reduce their consumption of paper and, instead of printing items, use collaboration tools to share knowledge. Further, departments with well-organized and easily accessible sites could also be rewarded, and the winner’s site could become the new standard for the organization. Additional policies could be created that prescribe the development of sites using only a minimal number of components, specifically ones that have shown to be effective and time-saving.

Policymakers must use a process-oriented organizational approach when developing policies and investing resources into training (Alcantara, 2009; Maier & Remus, 2003; Witte, 2002). Policies and implementations must be closely connected, and those using technology must clearly understand organizational policies and be involved in the implementation process (Boer, 2012; Spetz, Keane, & Curry, 2009; Thielst, 2007). Finally, Witte (2002) determined that mandatory training should be addressed as a separate requirement within policies to ensure that end-users can gain maximum benefits and efficiencies from the implementations.

Garicano and Heaton (2010) found that if IT implementations are studied in isolation, they may not reflect productivity gains. To truly increase productivity, both organizational policies and management practices must be aligned with IT implementations (Garicano & Heaton, 2010). Once IT is integrated in an organization, the quality and type of data available to be studied changes, which has obscured the findings of some studies on productivity (Garicano & Heaton, 2010). Thus, when attempting to study productivity gains, researchers must change the way they measure them in an electronic versus a manual environment (Garicano & Heaton, 2010).

Research Method and Design

This study focused on how individuals use technology, how organizational policies affect that usage, and the resultant perceived productivity. Because this study focused on perceptions, personal experiences, and perceived productivity, a qualitative approach was appropriate (Cheseboro & Borisoff, 2007). Data were gathered from interviews about individual experiences, personal views, and details of situations; data did not consist of generalizations made from standardized questionnaires (Cheseboro & Borisoff, 2007).

Grounded theory designs are popularly used for IT studies because data are systematically gathered and analyzed (Bryant & Charmaz, 2007; Matavire & Brown, 2013; Urquhart et al., 2010). GT studies allow researchers to develop and build theories about phenomena when either none exist or when existing ones are inadequate (Goulding, 2002). The aim of this study was to understand the perceptions of individuals and to generate a theory about how leaders can implement more effective policies that mandate the use of technology. No other qualitative design was suitable for achieving this goal.

Unlike other qualitative studies, grounded theory studies are the only ones in which researchers do not work from existing theoretical frameworks. The collected data guide the research and lead to the development of categories, relationships, attributes, and ultimately, a theory. Because limited research exists on the topic under investigation, a grounded theory approach allowed for the development of a theory about a phenomenon which is currently not well-understood. This theory will help global leaders more effectively manage their organizations.

This study was conducted in the natural setting in which the phenomenon occurred, not in a laboratory or otherwise sterile facility (Pratt, 2007). When asking questions, the interviewer used a semi-structured format of open-ended questions and developed theories as data were collected (Cheseboro & Borisoff, 2007). Data were analyzed and compared as they were collected via constant comparison. As categories began to emerge, they drove theory generation (Bryant & Charmaz, 2007).

Using initial observations and one-on-one interviews of 18 participants, data were collected and analyzed through the constant comparative method. This method revealed emerging themes and conceptual patterns about how organizational policies can be implemented to obtain the desired outcome and achieve maximum effectiveness (through maximal compliance).

Research Questions

Several research questions guided this study; however, one was central: How can leaders more effectively implement organizational policies that mandate the use of technology? Four sub-questions assisted in answering the main research question:

1. How do employees' perceptions of ITCOM's organizational policies affect their use of technology and perceived productivity?
2. How can leaders adapt the manner in which policies are worded to best reach organizational goals; i.e., what policies would be perceived as encouraging to employees in increasing their use of technology to become more productive?

3. How should policies best be implemented to gain maximum receptiveness; for example, should a phased approach be used?
4. What other factors (intrinsic and extrinsic) might affect employees' inclination to use technology/comply with policy?

Population and Sampling Frame

All of the study's participants were of grades GS-12 to GS-14 and non-supervisors. The objective was to address only this category of employees to understand the effects of policies on the workers instead of on the senior leaders who participated in implementing the policy. To be eligible to partake in the research, individuals must also have a requirement to use Microsoft SharePoint at least weekly in performing their jobs. Each candidate was asked about his or her GS civilian grade, weekly usage of SharePoint, and an awareness of the requirement to use SharePoint. Once subjects were identified as meeting the requirements for participation in the study, they were informed about their rights as participants. In addition to the informed consent form, each participant received a detailed explanation of the contents of the form to avoid any potential misunderstanding.

The grounded theory sampling strategies used were snowball and convenience sampling, which both fall under the purview of purposeful sampling. The participants who were interviewed are all employees of ITCOM who used Microsoft SharePoint daily as part of their jobs. Participants responded to nine open-ended questions addressing their reaction to the released organizational policy that required them to use the new enterprise tool, Microsoft SharePoint, to accomplish at least part of their daily work activities.

Because of the organization's military status, each candidate was informed that senior leadership had approved that the study be conducted at ITCOM and that his or her time away from normal duties was authorized. This was important information to share, as some participants were worried about whether they could partake in interviews during the work day.

Grounded theory studies do not have a stated required number for sample sizes or data saturation (Glaser & Strauss, 1967; Corbin & Strauss, 1998). A number of factors drive the required number of samples, such as the sensitivity of the phenomenon, the scope of the study, the experience and skill of the researcher, and the participants' familiarity with the phenomenon (Corbin & Strauss, 1998; Morse, 2000). Because data drive the study and sample size, the study is complete when theoretical saturation has been reached. Researchers cannot know ahead of time when this will occur, but by choosing participants that are very familiar with the phenomenon, they can limit the number of required interviews (Glaser & Strauss, 1967; Corbin & Strauss, 1998). If subject matter experts are interviewed initially, researchers can use these initial interviews as a guide to narrow the focus of the study (Corbin & Strauss, 1998).

When interviews begin producing the same data, theoretical saturation has been reached. Researchers must collect enough data to clearly discern concepts, patterns, categories, and properties (Glaser & Strauss, 1967; Strauss & Corbin, 1998). Researchers will know that they have reached data saturation when no new data emerge on a category, the category is well-developed, and category relationships are established and validated (Corbin & Strauss, 1998).

Data Collection

Data collection began with a pilot study to validate the adequacy and clarity of the interview questions and the value of the observations. Resultantly, a supplemental question (#5) was added to the list for the purpose of eliciting from the participants additional thoughtfulness and potential solutions, versus only an identification of associated problems. The observation portion of the study entailed watching participants as they used Microsoft SharePoint. After being observed, these individuals were interviewed about their experiences when using the tool and how their perceptions of organizational policies affected their use of the tool and resultant productivity. The observation period of the pilot study identified that observations were far less useful than the interviews themselves, and were, thus, excluded from the study.

After participants were identified, interviews were scheduled and conducted in an empty office space within the departments in which the individuals worked. Some of the participants found the interview to be a welcomed break from their work day and desired to have short, informal discussions either before or after the formal interview. The interviews were recorded using a tape recorder to ensure the accuracy of the notes taken during the sessions. Participants were guaranteed that no one within their chain of command would be able to access the notes or the tapes. Anonymity was maintained by assigning a number at the top of each interview sheet. The numbers corresponded to their sequence amongst the participants; i.e., the third subject received a “#3” at top of his interview questionnaire.

During the interviews, the goal was to understand how organizational policies affected the use of the technology. For example, if in the past individuals were able to submit documents through manual processes, but now policies mandate automated processing via SharePoint, how did this change affect productivity? More importantly, how did the subjects’ perceptions of the policies affect their use of the technology, their performance, and their resultant productivity level? During the interviews, the participants were asked a number of questions about how they reacted to the policies (see Appendix A), including how they believed policies could be better implemented to encourage the effective use of technology.

Interviews were conducted until saturation was reached. Saturation occurs when additional data collected does not add anything new to the existing categories (Goulding, 2002). In this study, data saturation appeared evident after 14 interviews, at which point responses duplicated the previously emerged categories; i.e., no new categories emerged as the interviews continued. However, additional interviews had already been scheduled, and to ensure that no new categories would emerge, data collection continued until the 18th interview. Data were reviewed and re-analyzed for meaning and implications until no new categories were discovered, at which point theoretical saturation had been reached (Glaser & Straus, 1967; Rubin & Rubin, 1995).

Besides the interviews relevant to the participants’ use of technology, data were also gathered from policy documentation (Fraser, 2008). Participants were asked about the content of these documents and how they perceived their messages. The intent was to capture the participants’ views and perceptions about policies and how they affected their willingness to use technology (Fraser, 2008). At the end of the interview sessions, the interviewees had the opportunity to review their responses and modify them to more accurately represent the intended meaning.

Additional data were collected via a thorough review of existing, scholarly literature in the field. For example, peer-reviewed studies about how to best implement policies and technology within organizations and how to manage change effectively were significant in understanding best practices. Using data gathered from existing literature in combination with the observations and interviews provided a more holistic understanding of the problem and aided in the generation of a theory.

Data Analysis

GT researchers use the following methods for data analysis: open coding, axial coding, and selective coding (Goulding, 2002). Open coding involves condensing data into meaningful units (West, 2007). Goulding advocates that open coding involve analyzing transcripts (of interviews, for example) line-by-line to identify all possible codes. This process continues until a pattern emerges across data sets (Goulding, 2002). Once categorized, data sets are examined for specific attributes and subcategories for each category (Goulding, 2002). Axial coding entails making interconnections between categories and subcategories. During axial coding, various aspects of categories are more clearly defined, which refines them and their interconnections. Data collection, open coding, and axial coding are iterative processes in GT research (Goulding, 2002). Selective coding consists of unifying all categories around core categories and adding descriptive detail (Corbin & Strauss, 1990) and usually occurs in the later phases of the study (Corbin & Strauss, 1990). The end result is a theory that is based on the collected data. The theory may be a statement, model, or hypotheses about the phenomenon (Goulding, 2002).

As in any grounded theory study, data collection and analysis occur simultaneously in this study. As data were gathered through interviews, they were analyzed concurrently and concepts began to emerge (Fraser, 2008). As concepts emerged, they drove further data collection (Fraser, 2008). Categories and attributes of categories were defined during the data analysis process (Goulding, 2002). As more data were collected and analyzed, these categories were modified and refined (Cheseboro & Borisoff, 2007).

The notes taken during the interviews were typed in and printed from Microsoft Office. To organize the vast amounts of data collected and analyzed during the initial analysis, the responses to interview questions were aligned under each research question. This approach provided a systematic way to understand the collected data, categorize them, and examine their relationships. Subsequent analysis procedures involved re-organization, review, and re-analysis of the data by themes, corresponding sub-themes, attributes, and interrelationships amongst categories and attributes.

Data were further analyzed through the use of open coding, which involved analyzing text line-by-line to discover key phrases and words (Goulding, 2002). By doing so, concepts, or units of related data, developed. Each transcript from the observation and interview underwent this type of analysis to identify codes, causing patterns and group-related codes to become visible (Goulding, 2002). By linking codes together, categories emerged. Categories are higher order codes that consolidate concepts into a theoretical framework. As data continued to be analyzed until saturation, groupings were also verified and corrected, as necessary (Goulding, 2002).

Further, each interview was compared to the previous one(s), answer by answer. Individual word and line analyses offered a way of providing insight to the meaning behind the participants' responses. Some comments and recollections were examined separately to assess their relevance to the participants' answers, interview questions, and research questions. As text was constantly compared and reviewed via open coding, several patterns, or common themes, began to emerge. The various subcategories that emerged from the data review and comparison were assigned to corresponding themes.

Using axial coding, categories were reviewed and reassembled to identify the relationships amongst them. Related themes were then placed under higher level concepts (Goulding, 2002). Open and axial coding are methods used to condense data into categories, or themes, and to understand the relationships amongst the categories and subcategories (Goulding, 2002).

Subsequent coding phases entailed continuous analysis, refinement, and review as data surfaced from additional interviews. A constant comparison of data ensured that relationships were recognized, as were new themes that developed from the analysis. Selective coding involved using the previously identified categories, defining, developing, and refining them further, then assimilating them to tell a story.

Validity and Reliability

Researchers strive to perform valid and reliable studies, but perfect validity and reliability are impossible goals (Neuman, 2006). Reliability refers to the repeatability and consistency of the study and its findings (Neuman, 2006). In other words, if the study were repeated and the measurement instruments are reliable, the findings should be similar. Validity questions the truthfulness of the measurements—do instruments measure what they were intended to measure (Fraser, 2008)?

In qualitative studies, validity and reliability are addressed in terms of the trustworthiness of the study (Lincoln & Guba, 1985). Trustworthiness is comprised of four criteria: credibility (internal validity), transferability (external validity), dependability (reliability), and confirmability (researcher objectivity; Lincoln & Guba, 1985). To be credible, the study's results must be believable. Credibility was established by testing the measurement instrument, or the interview questions, via the pilot study to ensure that they measured what they were intended to measure. Additionally, participants' responses were verified with the individuals to ensure that their intended meanings were accurately captured. Transferability addresses the degree to which results can be generalized (Lincoln & Guba, 1985). Generalizability is not a goal in qualitative studies, but qualitative studies can be made more transferable if researchers address specific elements of the context, or environment, in which the study occurred, as well as thoroughly describing the limitations and assumptions of the study.

Dependability refers to reliability—if the study were repeated, would the findings be the same (Lincoln & Guba, 1985)? Confirmability relies on the researcher's objectivity (Lincoln & Guba, 1985). Dependability and confirmability were addressed by mitigating potential researcher bias. Researcher bias can be mitigated if one constantly questions oneself and one's objectivity. In this particular case, the researcher was very familiar with the organization and with the subjects, so a sense of self-awareness and self-questioning became especially important in mitigating possible instances of bias. Additionally, by ensuring that only the study's collected data were used during analysis, in lieu of any personal opinions held by the researcher, potential bias was further eliminated. In this study, a significant amount of data were collected. Since a GT design was used, these data drove the study, which helped address both the dependability and confirmability criteria of trustworthiness. Finally, data were checked and rechecked to ensure that analysis was performed correctly, that data were categorized correctly, and that nothing was accidentally omitted or overlooked. By documenting, checking, and re-checking data collection and analysis procedures, researchers can further increase confirmability. Dependability is enhanced if researchers describe changes that occur in the environment and how those changes were addressed.

Researchers constantly attempt to discover new categories of evidence until data saturation is reached (Jones, Kriflik, & Zanko, 2005). Data saturation is what makes GT studies robust—researchers do not stop collecting data until saturation is reached (Jones et al., 2005). Interviews are often complemented by observations to strengthen a study (Goulding, 2002), but, ultimately, data saturation, the constant comparative method, and the diligent data analysis procedures are what make GT studies rigorous.

Results

Three primary categories, or common themes, emerged from the data analysis process and are listed in no order of significance: leadership, policy, and mandated tool (see Appendix B).

Theme 1: Leadership

The leadership theme consisted of five sub-themes, indicative of what leadership responsibilities were most important to the study's participants in regard to gaining their acceptance of the new organizational policy: communications strategy, involvement with and commitment to the policy, policy enforcement, training, and stakeholder involvement.

Communications Strategy

Seventy percent of the participants felt very strongly about senior leadership not having adequately communicated a plan for change, addressed the policy, and distributed the policy. Based on the number of times participants referenced this sub-theme, it can be considered one of the most significant problems faced in regard to employee acceptance of the new organizational policy. One-third of the participants felt that both acceptance of the mandated tool and the policy, in general, would have significantly increased if leadership had just informed them about what was going to happen and why the change was necessary. Twenty-two percent believed that being informed in the very beginning instead of at the end would have also made a difference, while one-third expressed concern about how the policy was published. They felt that the policy should have been easy for anyone to find and that senior leadership had a responsibility for ensuring that the policy was received by all employees, at every level. Because policy distribution was performed inadequately, 22% of the participants said they had never even seen it. One individual mentioned that he knew of the policy's existence, but was unaware of its content, while another felt that the policy should have been shared via all available communication channels, such as town hall meetings, e-mail notifications, and a posting on the portal. Using more than one communication channel would have ensured that individuals who missed e-mail notifications, for instance, would have received the information through another channel, according to the participant. Further, he stated that direct supervisors should have taken responsibility for ensuring the dissemination of the policy to their employees.

Involvement With and Commitment to the Policy

One-third of the participants believed that the changes encompassed by the policy were not supported by senior leadership. Besides keeping everyone informed, senior leaders should have led by example and used the tool first, they stated. Further, senior leaders should have been advocates for the tool that was prescribed for use.

According to 17% of the participants, management commitment and involvement entails changing the organization's culture. In order for the workforce to be receptive to the proposed changes, these participants felt that their organizational culture needed to reflect a more open-minded and flexible attitude. The traditional views that employees held were neither conducive to change nor to accepting and using modern technology. Eleven percent believed that by offering change management courses, leadership could have helped individuals adapt to the new way of doing things.

Policy Enforcement

Twenty-two percent of the participants thought that without enforcement, many individuals would fail to comply with the current policy. Non-compliance would result in redundant processes and numerous inefficiencies, due to either improper usage or simply non-usage of the tool. According to one participant, improper use would not only affect individual productivity, but also any general productivity gains that he believed leaders were seeking to achieve through the new policy.

Training

Next to communication problems, training was the most heavily cited concern by nearly 80% of the participants. The respondents felt that if leadership did not provide proper training for the new and complicated tool, they could never be truly efficient or effective at using it. Also, without a proper understanding of the functionality and the tool's benefits, tool acceptance would be reduced and resistance would increase, turning it into a time-waster. Nearly half of the participants believed that training via small groups would have been the most effective way to teach employees what they needed to know. They suggested that different levels of training were necessary, based on individual job roles. Some employees were chosen to fill the role of content manager for their team. Participants believed that these individuals should have received specialized training to address the additional responsibilities they would have in managing specific group sites and resolving issues. According to one individual, the absence of specialized training for content managers delayed trouble-ticket resolutions and site management duties that these individuals were required to make daily as part of their assigned role. Further, another participant explained that content managers lost important departmental data because "they didn't know what they were doing." According to the interviewee, training should have occurred before the tool's use was made mandatory via the policy, and training should have been staggered.

Stakeholder Involvement

This sub-theme differs from the communications strategy sub-theme in that it addresses involving employees in the change process from the start, versus communicating changes that they are not involved in making at the end.

Nearly 40% of the participants believed that their involvement in the decision-making process would not have changed their attitude toward the practice, independent of whether they liked or did not like the mandated tool. One individual believed that he would have been more prepared and supportive of the tool and the policy had he been involved from the beginning. Seventeen percent stated that the need to be involved and acceptance of the tool and the policy depended mostly on an individual's familiarity with the technology. For example, if someone already knew SharePoint, then being told to use the tool would not be as challenging as its usage would be for someone who lacked IT skills or familiarity with the tool. Eleven percent of the participants believed that being involved in the policy's development and having an understanding of the change would have increased its acceptance. Finally, one participant would have chosen a different tool, based solely on personal preference, had he been involved in the decision-making process.

Theme 2: Policy

In addition to the distribution problems, those who had seen the policy felt that it was lacking significant elements. The four sub-themes for policy are general content, diction, compliance, and time to comply. The policy theme is comprised of answers provided in response to all of the research questions.

General Content

This sub-theme developed mostly from responses to SRQ3. Seventeen percent of the participants criticized the policy for being unclear and wordy. Another 22% felt that explaining the necessity of the change (i.e., the policy) was extremely important for its acceptance. In doing so, the policy would have also explained the benefits of the tool, which half of the participants believed to be a critical missing element.

The policy should have contained rules and guidelines for structured use, as well as processes and procedures, according to almost 40% of the participants. Further, 17% of the participants believed that downloadable standard operating procedures (SOP) on tool usage and system functionality would have been especially helpful in providing guidance on tasks, such as moving old data to the new system. Roles and responsibilities also needed to be defined, according to several participants; one individual believed that the inclusion of definitions would have led to increased acceptance of the policy. Seventeen percent suggested that including a meaningful measurement of effectiveness, or metrics, would have provided a means for measuring potential efficiencies gained, tangible results for monetary savings, and productivity gains.

More importantly, the policy should have established a training requirement for the workforce, according to more than half of the participants. In doing so, employees would have been guaranteed to receive training that would have helped them learn to use the newly mandated tools. One individual thought that keeping the policy updated was important, as the organization's needs/requirements changed, while another stated that the policy needed to match the readiness or availability of the tool: "Upon the policy's release, SharePoint was far from being operational. Thus, we were forced to use a tool with limited functionality that could not be used to perform our work functions."

Diction

Responses to interview question #7, *How did the wording of the policy affect you? For instance, how would you have reacted differently if the policy had been phrased in a different way, perhaps in a more positive fashion?,* were mostly responsible for producing this sub-theme. Responses to these questions were quite varied. For instance, half of the participants felt that directive wording was very much appropriate for an organizational policy and that word choices themselves were irrelevant. Two individuals stated that they were used to being told what to do. Because they had worked for military organizations as both civilians and soldiers, being directed to comply was a normally occurring event for them. Others (17%) stated that word choices should have been more positive or consensual in nature to improve employee receptiveness to and acceptance of the policy. Another participant added that the policy, "Should sound intelligent," and two others stated that a value-added perspective versus a directive tone would have been more appropriate. Finally, one participant felt that diction was irrelevant, but later added that he would have reacted more favorably if he were given some choices within the policy's requirements.

Compliance

When participants were asked interview question #2, *Why do you think the SharePoint policy was created?*, their responses varied widely. Nearly one-third of the participants believed that policies were a necessary means to achieve compliance. Eleven percent posited that employees would only use new tools if they were forced to do so and that the way to force people to comply is by issuing a policy. One individual added that employees would be reluctant to use SharePoint because they were unfamiliar with the tool. Therefore, leadership had to create a policy to ensure compliance. A different participant believed that the policy was a way to move the organization toward data consolidation, regardless of the tool selected for accomplishing this goal, while a third individual thought that standardization was a goal. A total of 22% of the participants thought the goal was collaboration. Some (11%) felt the policy's intent was to do both.

Reducing storage, administration, and software licensing costs were other management goals that 22% of the participants believed the policy was intended to achieve. One participant also thought that the policy was created purely from a management perspective, intended to achieve managerial goals. Alternate views were expressed by participants who believed that the policy was created to mandate a new and better tool, which would increase performance and lead to efficiency gains, reduce network bandwidth requirements, provide a greater data storage capacity, and improve accessibility. Eleven percent of interviewees felt that the policy adequately provided guidelines for a more structured use of the new tool.

Time to Comply

When asked interview question #6, At what point in time between the policy release and its enforcement did you comply with the policy? What affected your acceptance of the policy?, responses were much less varied. More than half of the participants stated that they complied immediately because they were mandated to do so. Some stated that their understanding of the value of and familiarity with the tool also affected their compliance. One individual explicated that change is a necessary part of progress, which increased his willingness to comply with the policy, in addition to having been mandated to do so. Other participants faced some intervening forces that slightly delayed their compliance. For example, although one participant complied quickly, he had to research the tool himself, which delayed his ability to use it effectively. Because he had not received training, it was a trial and error exercise. The lack of a customer support forum or published help desk numbers further delayed his usage of the tool. Another participant stated that in the absence of training, his ability to use the tool was significantly limited. Finally, one participant complied as quickly as he could, but competing priorities contributed to a slight delay. He stated, "We have so many competing priorities. Nothing ever falls off the plate. Everything just becomes a number one priority in this command."

According to 17% of the participants, some of the previous tools were disabled immediately, which left them no choice but to use SharePoint. In other words, their compliance was forced. One individual complied after attending training, which gave him a better understanding of the value of the new tool. Twenty-two percent did not comply immediately because of the lack of functionality encountered when using the new tool. Unable to move data from the previous tool due to file type and file size restrictions, one participant felt forced to use both tools to do his job. This created redundant data and workflows, as well as additional work for him. Others continue to avoid using SharePoint because they find the tool difficult to use. However, one participant believed that a better layout and organization of the tool would lead to increased general compliance.

Theme 3: Mandated Tool

This category repeatedly emerged from responses to multiple interview questions related to research questions 1, 3, and 4. The mandated tool theme was referred to more than any other category or subcategory during the study's procession. Four major sub-themes emerged under mandated tool: perceptions of tool functionality, initial reactions to the requirement, phased implementation, and effect on productivity.

Perceptions of Tool Functionality

Numerous participants felt strongly about the limited functionality of SharePoint at the time its use was mandated by the organizational policy. They believed that at least the major problems should have been resolved by the time they were required to use the tool. Nearly one-third of the participants mentioned performance issues that made the tool slow and cumbersome, discouraging their use of SharePoint. These issues also led to redundant sites being used (some on the old system), the creation of unnecessarily complex workflows, permissions issues that prevented them from modifying sites, and accessibility problems. One participant referred to these problems as "pilot pains." He thought that because ITCOM was the first organization to implement SharePoint at an enterprise-level, the organization's employees were the guinea pigs who were made to suffer through the problems.

The tool should have been easy to use, and sites should have been well-organized—this would have increased tool acceptance, according to 22% of the participants. The tool also should have had greater functionality than the tool it replaced. Having full functionality would have increased the acceptance rate, according to 17%. Conversely, 22% of the participants believed that previous knowledge and understanding of the tool's capabilities is what would have most affected acceptance. They felt that because they had past experiences with SharePoint, they were much more receptive to using the tool than were their peers. One individual commented that the tool should have been more carefully selected, then evaluated later to ensure that it met its intent.

Because the tool was administered at the wrong level and by an external organization, additional problems arose, stated 45% of the participants. As functionality problems were identified, the time to resolve them became extraordinarily long. Further, the way the tool was configured and implemented significantly decreased the number of available features. Seventeen percent of the participants would have been much more receptive to using SharePoint had the tool been installed "straight out-of-the box."

Initial Reactions to the Requirement

When the policy to mandate the use of SharePoint was first released, employee reactions varied notably. Nearly 40% of the participants were happy and excited about receiving a new tool with which to perform their duties. A number of participants commented on the positive features of the tool, such as improved processes, better collaboration, better standardization, better overall functionality than the previous tool, efficiency increases, centralized storage and better security, reduced waste, the ease of finding documents and no longer having to rely on email, and not having to use the old tool anymore. Others felt dread and fear. They were worried about the tool not being user-friendly, losing control, and not having defined processes, procedures, roles, and responsibilities. One participant was also distraught by the bad performance of the tool, while two others complained about bad or limited functionality. Another individual expressed concern about users' readiness to use such a complex tool and the resultant user resistance. Some participants were also displeased about the amount of time they would have to invest in learning to use the new tool. Finally, one participant stated that he did not know about the policy or the requirement to use SharePoint.

Phased Implementation

When asked whether their perception and acceptance of the policy would have improved if the tool had been implemented using a phased approach, 56% of the participants stated, "Yes." They felt that a phased approach would have limited disruptions, while offering many benefits to the workforce, such as more time: to adapt to the change, for administrators to resolve initial problems with the tool and provide better functionality, and for training. Using a phased approach would have encouraged the use of the tool and increased compliance, according to 11% of the participants. One individual stated, "This tool is too big to roll-out all at once." Nearly 40% of the participants felt that a phased approach would have led to greater acceptance. One participant believed that the implementation was phased and that this approach did not help reduce the number of problems the workforce encountered. The remaining ca. 40% of the participants felt that a phased implementation would have led to more problems. They believed that a phased approach would have decreased user acceptance, tool functionality, compliance, and would have led to duplicate data and confusion.

Effect on Productivity

Interview question #8 addressed perceived productivity gains that resulted from using the new tool. The responses to this question varied—45% of the participants felt that there were definite productivity gains, but 40% believed that gains were limited. Seventeen percent of participants stated that a lack of metrics meant that there was no way to measure actual increases in productivity. For one-third of the participants, documents were easy to find and share, both externally and internally—that alone was a significant efficiency gained through the new tool. Some participants welcomed not having to carry an external storage device to access data, and centrally located data led to better collaboration, according to others. Seventeen percent believed that productivity was further improved through better accessibility and standardization.

Those who believed that productivity gains were limited cited training and functionality problems as the main reasons. The way the sites were organized made information difficult to find. Seventeen percent thought that once the layout improved, productivity gains would rise significantly. One participant noted that too many people are still relying on e-mail to share files and that the resultant limited use of the tool is hampering organization-wide productivity gains. Further, leaders are not using SharePoint or enforcing the use of the tool, which is why the usage problems persist.

Discussion

Theme 1: Leadership

This theme addresses those things that are important for leaders to do when implementing organizational policies. As presented in the Results section, 70% of ITCOM employees desired their leaders to better communicate with them, inform them of upcoming changes early on, and explain why changes are necessary. According to some participants, communication should occur via a number of venues to ensure that everyone receives the message. If possible, leaders should schedule town hall meetings, send e-mails, and post information to collaboration portals. Once a policy is published, leaders should ensure that they distribute it to everyone. Again, leaders should use any means necessary to inform their employees of a policy, according to participant responses. Copies could be printed and distributed during meetings, they could be disseminated electronically, or they could be posted to a centralized location to which everyone has access.

As indicated by one-third of the study's participants, ITCOM's organizational policy was not well-distributed. This is not an uncommon problem for the organization. Although a specific department exists that has the responsibility for distributing policies throughout ITCOM via e-mail, policies often do not reach the entire workforce, as evidenced by participant responses. By using every means of communication available to them, leaders can ensure that their workforce is in receipt of the policy.

In addition to conquering distribution issues, leaders should also enforce what is mandated within a policy. According to some participants, ensuring compliance will work toward the general good of all, as it prevents employees from using previous tools and creating redundant data repositories and workflows.

Training was found to be an important component of tool use not only in this study, but in previous studies conducted by Nyström (2006) and Dubé & Paré (2001). By providing training, leaders give employees the opportunity to understand a new tool and all of its features. Training also gives employees insight into how a tool could help improve their work processes, thereby increasing acceptance. Demonstrating the functionality of the tool could show how it will create efficiencies and reduce workloads, instead of becoming an impediment.

Theme 2: Policy

This theme addresses those things that make a policy effective. It covers policy content, diction, compliance, and the timeframe within which to comply. In order for employees to perceive a policy with a positive and responsive attitude, the policy must be structured in a specific way, as evidenced by participant responses. Polices should be succinct and clear and explain why a transition to a new tool is necessary. This finding corresponds with studies about policy acceptance conducted by Boer (2012), Kotter and Schlesinger (2008), Long and Spurlock (2008), and Nyström (2006). Some of the participants believed that in addition to explaining the necessity of the change, the policy should be explicit in describing the benefits of the new and mandated tool. They believed that an explanation of user benefits was critical for acceptance.

Further, the policy should contain the following sections: detailed rules and guidelines for structured use of the tool, processes and procedures, roles and responsibilities, and metrics. Participants identified the inclusion of standard operating procedures as a preferred means of explaining how to use the new tool, addressing specific tasks such as how to move data from the old tool to the new tool, as well as general instructions for use. Finally, outlining roles and responsibilities helps clarify who is responsible for which tasks.

The establishment of metrics in the policy provides a way for leaders to assess the effectiveness of a new tool. Several employees felt that being able to provide tangible means of measuring improvements, such as gains in productivity or monetary savings, would help leaders demonstrate the utility of the tool. Finally, the most important item to include in the policy is the establishment of a training requirement. A training requirement ensures employees will learn how to properly use a tool, which would accelerate acceptance and reduce the employees' time to comply with the policy. Conducting training prior to a tool's implementation also ensures familiarity with a tool, which a number of participants felt was paramount to acceptance.

Theme 3: Mandated Tool

Acceptance of both the policy and SharePoint were strongly affected by perceptions of the tool's functionality. Many participants felt that the lack of functionality at the time that SharePoint's use was mandated discouraged employees from using the tool. They faced a number of problems, such as gaining access, permissions issues, broken links, slow response times, and disorganized site layouts. If these initial problems had been resolved prior to employees using the tool, the workforce would have had a much more positive experience and perception of SharePoint, according to many of the participants. The tool should have been easy to use and offer better functionality than the previous one. This finding corresponds with previous studies that addressed the factors behind technology acceptance in voluntary-use environments, but differs from findings when use is mandated (Brown et al., 2002; He et al., 2009). This incongruence might be explained by the differences in organization types and cultures. Unlike the referenced studies, this study took place in a military organization where employees are used to receiving orders. The differences in culture between private industry and military organizations may explain the inconsistent perceptions regarding the mandated use of technology. Finally, SharePoint should have been administered at the right organizational level, either internally or externally, via outsourcing to an organization capable of quickly resolving technical problems. The number of problems that arose were unnecessarily high because of administration issues, according to 45% of the participants.

Nearly 40% of the participants were excited when they were told that they would get a new and better tool with which to do their jobs. This positive attitude could have been prolonged if SharePoint had functioned correctly. Ensuring proper functionality leads to greater use and acceptance of tools and compliance with policies, according to the study's participants, as well as findings from previous studies (Bushnell, 1999; Chin, 1998).

When implementing new tools, a phased approach was preferred by over 55% of the workforce. This finding also corresponds to discoveries made in previous studies (Malik & Danish, 2010; Witte, 2002). Phased implementations give administrators more time to resolve initial problems, employees more time to attend training, and is less disruptive to the workday. Again, proper adherence to these factors would have led to greater acceptance of the tool, according to approximately 40% of the participants.

Nearly 45% of the study's participants felt that they became more productive by using SharePoint. This finding is supported by previous studies about productivity increases gained from collaboration tools (Lomas et al., 2008), especially when they are complemented by organizational policies (Garicano & Heaton, 2010). Productivity increased as tool functionality problems were resolved and, potentially, could continue to increase as the tool is improved and sites become better organized. Finally, if leaders enforced the use of SharePoint and more people used it, some participants believed that productivity would increase not only for those individuals but for the entire workforce, as redundant processes would be eliminated.

Limitations

Because this study was conducted in a military organization, some of the employees' perceptions may differ from what one would find if the study were conducted in the private sector. For example, ITCOM employees are used to receiving orders. This may not be true for individuals who have not worked for military or government organizations. Some perspectives are unique to this demographic.

Further, private industry may not use formal policies to mandate the use of technology. Although about one-third of the study's participants felt that policies were necessary to ensure compliance, the private sector may not follow the same procedure. Thus, the results from this study may only be applicable to organizations that use policies to mandate technology use.

Finally, this study's participants included non-supervisory employees between the grades of 12 to 14. The intent was to understand how the general workforce at ITCOM perceived the organizational policies, not those who were involved in mandating the policy. Had the latter group been included, the responses likely would have changed notably.

Implications for Leaders

The way that ITCOM currently implements policies is ineffective in achieving compliance, as evidenced by participant responses. Leaders must change how policies are written, distributed, implemented, and communicated. Currently, employees are displaying a high level of resistance to the policies, based on all of these factors. The presence of resistance indicates that policies are ineffective because they are not meeting their intent. Leaders must show that they are committed to their policies and develop measures to enforce compliance. Kim and Lee (2006) and Nyström (2006) found that a lack of leadership commitment to technology implementations and training hinder the use of and acceptance of tools.

Participants believed that leaders must make appropriate training available to the workforce to ensure that employees understand how to best use newly mandated tools. Previous studies by Bergiel et al. (2006) and Chong (2005) also found that employees often lack the proper training or skills to successfully use technology, which leads to increased resistance (Cogburn & Levinson, 2008; Long & Spurlock, 2008). Zivick (2012) identified how important it is for leaders to have enough funding to properly train their work teams on technology. Alcantara (2009), Maier and Remus (2003), and Witte (2002) corroborated this view, especially in situations where new technologies are linked with organizational policies. Analogous to participant responses, Witte (2002) found that training should be a requirement addressed within the policy itself. Optimally, training will occur prior to the mandated use of the tool.

Another important point to recognize is that organizational cultures must be ones that are adaptive to change. Some of the participants posited that the existing organizational culture is not flexible, which makes the workforce less receptive to proposed changes. Previous studies corroborate that when policies do not align with organizational cultures, they can be counter-productive in their attempt to reach organizational goals (Alcantara, 2009; Boer 2012; Calhoun 2002; Spetz, et al., 2009; Thielst, 2007). According to Alcantara (2009) and Witte (2002) organizational cultures must be learning cultures that support the use of technology. Nemiro et al. (2008) and Rahmati et al. (2012) identified the criticality of cultural management in establishing behavioral norms for successful collaboration. Paghaleh et al. (2011) and Hartman et al. (2009) added that cultures must be ones that encourage teamwork and cooperation to enable employees to overcome their resistance to information sharing via collaboration tools. In light of the findings from this and previous studies, leaders would do well to develop an organizational culture that encourages open-mindedness and flexibility. By doing so, policies that mandate changes will receive less resistance from the workforce, according to participants.

Leaders must also ensure that policies contain the critical elements identified in this study and that they are clear and concise. Again, this finding corresponds with those of previous studies about policies being clear in intent and meaning to be accepted by the workforce (Boer, 2012; Kotter & Schlesinger, 2008; Long & Spurlock, 2008; Nyström, 2006). Kapsali (2011) also found that if leaders are flexible when implementing policies, customizing them to specific user groups, acceptance will increase. Alcantara (2009) and Tirgari (2012) posited that by including the workforce in the policy development process, policies will be more effective. This finding is contradictory, as 40% of ITCOM participants believed that their involvement in the policy's creation would not have changed their reaction to it. An additional 16% of respondents felt that the need to be involved in the policy's creation is directly linked to individual knowledge about the technology that is being implemented. The different type of culture that exists within a military organization may account for the incongruous finding.

Leaders should take measures to ensure that the tools for which they mandate use are fully functional and appropriate for the employees in performing their work. This finding aligns with Bushnell's (1999) study, in which he identified a lack of functionality as a hindrance to technology acceptance. Dubé and Paré (2001) and Galleta and Zhang (2006) would agree, adding that proper functionality is especially important in virtual environments. Further, Brown et al. (2002) and He et al. (2009) attributed usefulness or perceived usefulness as primary reasons for technology acceptance, which also correspond to the findings of this study.

Leaders should consider using a phased implementation approach and creating metrics to measure cost savings and productivity increases. Studies by Malik and Danish (2010) and Witte (2002) also found a phased implementation approach to be more effective. By following all of the aforementioned guidelines, leaders will be able to more effectively implement their policies, according to participants.

Conclusion

The purpose of this grounded theory study was to develop a theory about how leaders can implement organizational policies that mandate the use of technology with maximum effectiveness. This study provided invaluable insight into employees' perceptions about policies and the resultant impact on the use of collaboration tools and perceived productivity, giving leaders the insight and understanding to effectively implement policies to achieve the following organizational goals: encourage the use of technology, increase productivity, and stay competitive in a global economy.

Recommendations for Future Research

The findings and theory of this study created a foundation for future research on the topic of how perceptions of policies affect compliance of the use of mandated technology and perceived productivity. Future researchers could explore the same phenomenon in the private sector, performing an equivalent study and comparing the results. Additionally, a quantitative study could be conducted to test the theory generated in this study, using a larger population. By doing so, researchers could test the strengths and weaknesses of this theory using a different research methodology and potentially increase the generalizability of the theory.

References

Alcantara, R. E. (2009). *A learning model of policy implementation: Implementing technology in response to policy requirements.* Retrieved from ProQuest Dissertations and Theses. (AAT 3355192)

Austerberry, D. (2011). Workflow automation. *Broadcast Engineering, 53*(7), 8–11.

Beagrie, S. (2006). How to manage well from afar. *ICIS Chemical Business, 1*(26), 40.

Bergiel, B. J., Bergiel, E. B., & Balsmeier, P. W. (2006). The reality of virtual teams. *Competition Forum, 4*(2), 427–432.

Boer, P. J. (2012). *Implementation of an education technology policy in Namibia's high schools: Through the eyes of the teacher*. Retrieved from ProQuest Dissertations and Theses. (AAT 1240600739)

Borck, J. R. (2001). Information building. *Infoworld, 23*(40), 47–48.

Brake, T. (2008). *Where in the world is my team? Making a success of your virtual global workplace.* San Francisco, CA: Jossey-Bass.

Brown, S. A., Massey, A. P., Montoya-Weiss, M., & Burkman, J. R. (2002). Do I really have to? User acceptance of mandated technology. *European Journal of Information Systems, 11*(4), 283–295.

Bryant, A. (2002). Re-grounding grounded theory. *JITTA: Journal of Information Technology Theory and Application, 4*(1), 25–42.

Bryant, A., & Charmaz, K. (Eds.). (2007). *The SAGE handbook of grounded theory*. Thousand Oaks, CA: SAGE.

Bushnell, R. (1999). Detecting 20/20 enterprise-wide vision. *Modern Materials Handling, 54*(8), 35–35.

Calhoun, A. L. (2002). *Implementation of educational policy: Technology implementation in a California middle school*. Retrieved from ProQuest Dissertations and Theses. (AAT 3117239)

Chesebro, J. W., & Borisoff, D. J. (2007). What makes qualitative research qualitative? *Qualitative Research Reports in Communication, 8*(1), 3–14. doi:10.1080/17459430701617846

Chin, K. (1998). Information technology meets resistance. *Chemical Engineering, 105*(3), 50.

Chong, S. C. (2005). Critical factors in the successful implementation of knowledge management. *Journal of Knowledge Management Practice 6*, 1–23.

Cogburn, D. L., & Levinson, N. S. (2008). Teaching globalization, globally: A 7-year case study of South Africa–U.S. virtual teams. *Information Technologies & International Development, 4*(3), 75–88.

Conole, G., de Laat, M., Dillon, T., & Darby, J. (2008). 'Disruptive technologies,' 'pedagogical innovation': What's new? Findings from an in-depth study of students' use and perception of technology. *Computers & Education, 50*(2), 511–524. doi:10.1016/j.compedu.2007.09.009

Corbin, J. M., & Strauss, A. L. (1990). Grounded theory research: Procedures, canons and evaluative criteria. *Zeitschrift für Soziologie, 19*(6), 418–427.

Corbin, J. M., & Strauss, A. L. (1998). *Basics of qualitative research*. Thousand Oaks, CA: SAGE.

DeRosa, D. M., Hantula, D. A., Kock, N., & D'Arcy, J. (2004). Trust and leadership in virtual teamwork: A media naturalness perspective. *Human Resource Management, 43*(2), 219–233.

Dubé, L., & Paré, G. (2001). Global virtual teams. *Communications of the ACM, 44*(12), 71–73.

Durugbo, C., Hutabarat, W., Tiwari, A, & Alcock, J. R. (2011). Modelling collaboration using complex networks. *Information Sciences, 181*(15), 3143–3161. doi:10.1016/j.ins.2011.03.020

Elmholdt, C. (2004). Knowledge management and the practice of knowledge sharing and learning at work: A case study. *Studies in Continuing Education, 26*(2), 327–339. doi:10.1080/158037042000225281

Fraser, K. C. (2008*). A virtual network organization business management model applied to retail banking: A grounded theory study*. Retrieved from ProQuest Dissertations and Theses.

Galletta, D., & Zhang, P. (Eds.). (2006). *Human-computer interaction and management information systems: Applications*. Armonk, NY: M.E. Sharpe.

Garicano, L., & Heaton, P. (2010). Information technology, organization, and productivity in the public sector: Evidence from police departments. *Journal of Labor Economics, 28*(1), 167–201.

Germain, M. (2011). Developing trust in virtual teams. *Performance Improvement Quarterly, 24*(3), 29–54. doi:10.1002/piq.20119

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New Brunswick, NJ: AldineTransaction.

Goulding, C. (2002). *Grounded theory: A practical guide for management, business, and market research*. Thousand Oaks, CA: SAGE.

Hartman, S. J., Fok, L. Y., Li, J., & Fok, W. M. (2009). Exploring the use of various technologies, perceptions of success and organizational culture in Jamaica. *Competitiveness Review, 19*(3), 224–238. doi:10.1108/10595420910962098

Harvey, M. G., & Griffith, D. A. (2007). The role of globalization, time acceleration, and virtual global teams in fostering successful global product launches. *Journal of Product Innovation Management, 24*(5), 486–501. doi:10.1111/j.1540-5885.2007.00265.x

He, W., Fang, Y., & Wei, K. (2009). The role of trust in promoting organizational knowledge seeking using knowledge management systems: An empirical investigation. *Journal of the American Society for Information Science & Technology, 60*(3), 526–537. doi:10.1002/asi.21006

Hew, K. F., & Hara, N. (2007). Knowledge sharing in online environments: A qualitative case study. *Journal of the American Society for Information Science & Technology, 58*(14), 2310–2324. doi:10.1002/asi.20698

Holland, J. B., Malvey, D., & Fottler, M. D. (2009). Health care globalization: A need for virtual leadership. *Health Care Manager, 28*(2), 117–123. doi: 10.1097/HCM.0b013e3181a2cb63

Igbaria, M., & Guimaraes, T. (1994). Empirically testing the outcomes of user involvement in DSS development. *Omega, 22*(2), 157–172. doi:10.1016/0305-0483(94)90076-0

Johnson, G. M., & Howell, A. J. (2005). Attitude toward instructional technology following required versus optional WebCT usage. *Journal of Technology and Teacher Education, 13*(4), 643–654.

Jones, M. L., Kriflik, G. K., & Zanko, M. (2005). Grounded theory: A theoretical and practical application in the Australian film industry. In A. Hafidz Bin Hj (Eds.), *Proceedings of International Qualitative Research Convention 2005 (QRC05).* Malaysia: Qualitative Research Association of Malaysia.

Kapsali, M. (2011). How to implement innovation policies through projects successfully. *Technovation, 31*(12), 615–626.

Kim, S., & Lee, H. (2006). The impact of organizational context and information technology on employee knowledge-sharing capabilities. *Public Administration Review, 66*(3), 370–385.

Kirkman, B., Rosen, B., Gibson, C. B., Tesluk, P. E., & McPherson, S. O. (2002). Five challenges to virtual team success: Lessons from Saber, Inc. *Academy of Management Executive, 16*(3), 67–79.

Kotter, J. P., & Schlesinger, L. A. (2008). Choosing strategies for change. *Harvard Business Review, 86*(7/8), 130–139.

Lanubile, F., Ebert, C., Prikladnicki, R., Vizcaino, A. (2010). Collaboration tools for global software engineering. *IEEE Software,* 3/4, 52–55.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. Newbury Park, CA: SAGE.

Lomas, C., Burke, M. & Page, C. L. (2008). Collaboration tools*. Educause Learning Initiative.* Retrieved from http://net.educause.edu/ir/library/pdf/ELI3020.pdf

Long, S., & Spurlock, D. G. (2008). Motivation and stakeholder acceptance in technology-driven change management: Implications for the engineering manager. *Engineering Management Journal, 20*(2), 30–36.

Maier, R., & Remus, U. (2003). Implementing process-oriented knowledge management strategies. *Journal of Knowledge Management, 7*(4), 62–74.

Malik, M. E., & Danish, R. Q. (2010). Impact of motivation to learn and job attitudes on organizational learning culture in a public service organization in Pakistan. A Research *Journal of South Asian Studies, 25*(2), 217–235.

Matavire, R., & Brown, I. (2013). Profiling grounded theory approaches in information systems research. *European Journal of Information Systems, 22*(1), 119–129. http://dx.doi.org/10.1057/ejis.2011.35

McShane, S. L., & Von Glinow, M. (2004). *Organizational behavior: Emerging realities for the workplace revolution* (3rd ed.). New York, NY: McGraw-Hill.

Mohd, D. N., & Mohamed, I. S. (2009). Evaluating technology resistance and technology satisfaction on students' performance. *Campus-Wide Information Systems, 26*(4), 298–312. http://dx.doi.org/10.1108/10650740910984637

Morse, J. (2000). Determining sample size. *Qualitative Health Research, 10,* 3–5.

Nemiro, J., Beyerlein, M., Bradley, L., & Beyerlein, S. (Eds.). (2008). San Francisco, CA: Jossey-Bass. *The handbook of high-performance virtual teams: A toolkit for collaborating across boundaries*

Neuman, W. L. (2006). *Social research methods: Qualitative and quantitative approaches* (6th ed.). Boston, MA: Allyn & Bacon.

Nunamaker, J. F., Reinig, B. A, & Briggs, R. O. (2009). Principles for effective virtual teamwork. *Communications of the ACM, 52*(4), 113–117.

Nyström, C. A. (2006). Philosophy of technology in organizations: A foundation for intranets. *Systems Research & Behavioral Science, 23*(6), 839–844. doi:10.1002/sres.702

Paghaleh, M., Shafiezadeh, E., & Mohammadi, M. (2011). Information technology and its deficiencies in sharing organizational knowledge*. International Journal of Business & Social Science, 2*(8), 192–198.

Peckover, S., Hall, C., & White, S. (2009). From policy to practice: The implementation and negotiation of technologies in everyday child welfare. *Children & Society, 23*(2), 136–148. doi:10.1111/j.1099-0860.2008.00143.x

Pratt, G. L. (2007). *Cultural and operational process issues of information technology support in academia*: A grounded theory study. Retrieved from ProQuest Dissertations and Theses.

Qureshi, S., Liu, M., & Vogel, D. (2006). The effects of electronic collaboration in distributed project management. *Group Decision and Negotiation, 15,* 55–75. doi:10.1007/s10726-005-9006-6

Rahmati, V., Darouian, S., & Ahmadinia, H. (2012). A review on effect of culture, structure, technology and behavior on organizations. *Australian Journal of Basic & Applied Sciences, 6*(3), 128–135.

Reichley, S. (1997). Small businesses use technology to up sales. *The Business Journal, 12*(36), 35.

Rubin, H., & Rubin, I. (1995). *Qualitative interviewing: The art of hearing data*. Thousand Oaks, CA: SAGE.

Saraswat, S. P. (2012). A phenomenological investigation of information and communications technology at a multinational enterprise from India. *Journal of Global Information Technology Management, 15*(4), 43–63.

Senge, P. (1998). The practice of innovation. *Leader to Leader, 9,* 16–22.

Smart, B. A., & Desouza, K. C. (2007). Overcoming technology resistance. *Business Strategy Review, 18*(4), 25–28. doi:10.1111/j.1467-8616.2007.00495.x

Spetz, J., Keane, D., & Curry, S. A. (2009). Information technology implementation in a rural hospital: A cautionary tale. *Journal of Healthcare Management, 54*(5), 337–347.

Sun, H., & Zhang, P. (2005). The role of moderating factors in user technology acceptance. *International Journal of Human Computer Studies, 64*, 53–78. doi:10.1016/j.ijhcs.2005.04.013

Tang, T. L., & Austin, M. J. (2009). Students' perceptions of teaching technologies, application of technologies, and academic performance. *Computers & Education, 53,* 1241–1255. doi:10.1016/j.compedu.2009.06.007

Thielst, C. B. (2007). Effective management of technology implementation. *Journal of Healthcare Management, 52*(4), 216–219.

Tirgari, V. (2012). Information technology policies and procedures against unstructured data: A phenomenological study of information technology professionals. *Academy of Information and Management Sciences Journal, 15*(2), 87–106.

Turban, E., Liang, T., & Wu, S. (2011). A framework for adopting collaboration 2.0 tools for virtual group decision making. *Group Decision & Negotiation, 20*(2), 137–154.

Urquhart, C., Lehmann, H., & Myers, M. D. (2010). Putting the 'theory' back into grounded theory: guidelines for grounded theory studies in information systems. *Information Systems Journal, 20*(4), 357–381. doi:10.1111/j.1365-2575.2009.00328.x

Walsh, J. P. & Maloney, N. G. (2007). Collaboration structure, communication media, and problems in scientific work teams. *Journal of Computer-Mediated Communication, 12*(2), 378–398.

West, C. J. (2007). *Workforce development in the context of organizations, leadership, and postmodern thought.* Retrieved from ProQuest Dissertations and Theses.

Westerfelt, D. (2010). An investigation of online collaborative tools by business professionals in the MBA environment. *Business Education Innovation Journal, 2*(2), 14–21.

Witte, K. A. (2002). *Technology policy implementation in a California central valley adult school.* Retrieved from ProQuest Dissertations and Theses. (AAT 305562165)

Workman, M. M. (2007). The effects from technology-mediated interaction and openness in virtual team performance measures. *Behaviour & Information Technology, 26*(5), 355–365. doi:10.1080/01449290500402809

Yi, M. Y., Jackson, J. D., Park, J. S., & Probst, J. C. (2005). Understanding information technology acceptance by individual professionals: Toward an integrative view. *Information & Management, 43*, 350–363. doi:10.1016/j.im.2005.08.006

Yu, M. (2009). Employees' perception of organizational change: The mediating effects of stress management strategies. *Public Personnel Management, 38*(1), 17–32.

Yukl, G. (2006). *Leadership in organizations*. (6th ed.). Upper Saddle River, NJ: Pearson.

Zawawi, A., Zakaria, Z., Kamarunzaman, N., Noordin, N., Sawal, M., Junos, N., & Najid, N. (2011). The study of barrier factors in knowledge sharing: A case study in public university. *Management Science & Engineering, 5*(1), 59–70.

Zivick, J. (2012). Mapping global virtual team leadership actions to organizational roles. *The Business Review, Cambridge, 19*(2), 18–25.

Appendix A: Interview Questions

1. What was your reaction when you were told that your organization had released a policy that would require you to use SharePoint to do your normal every-day work processes? Why?
2. Why do you think the SharePoint policy was created?
3. How would your reaction have been different if you had been involved in the decision-making process to make the practice mandatory?
4. How would you have reacted differently if the process change had been more gradual/incremental?
5. How would you have done things differently?
6. At what point in time between the policy release and its enforcement did you comply with the policy? What affected your acceptance of the policy?
7. How did the wording of the policy affect you? For instance, how would you have reacted differently if the policy had been phrased in a different way, perhaps in a more positive fashion?
8. What other factors affected your response to the policy?
9. What effect do you feel SharePoint has had on your productivity?

Appendix B: Table 1

Themes and Sub-Themes

Theme Number	Theme	Sub-Theme
1	Leadership	i. Communications Strategy ii. Involvement with and Commitment to the Policy iii. Policy Enforcement iv. Providing Training v. Stakeholder Involvement
2	Policy	i. General Content ii. Diction iii. Compliance iv. Time to Comply
3	Mandated Tool	i. Perceptions of Tool Functionality ii. Initial Reactions to Requirement iii. Phased Implementation iv. Effect on Productivity

CREATE: Accelerating Defense Innovation With Computational Prototypes and High Performance Computers

Douglass Post—established and leads the DoD High Performance Computing Modernization Program (HPCMP) Computational Research and Engineering Acquisition Tools and Environments (CREATE) Program[1] (2005–present). He received a PhD in Physics from Stanford University and is the Associate Editor-in-Chief of the AIP/IEEE publication *Computing in Science and Engineering.* He led Lawrence Livermore National Laboratory (LLNL) A and X-Division and the Los Alamos National Laboratory (LANL) X-Division nuclear weapon and inertial fusion code development programs (1998–2005). He led the International Thermonuclear Experimental Reactor (ITER) Physics Team (1988–1991) and the ITER In-Vessel Physics group (1992–1998), for which he received the American Nuclear Society's Outstanding Technical Achievement Award for fusion science and engineering in 1992. He established and led the tokamak modeling group at Princeton University's Plasma Physics Laboratory (1975–1993). Dr. Post is a Fellow of the American Nuclear Society, the APS, and the IEEE. He received the American Society of Naval Engineers 2011 Gold Medal Award for the CREATE Program. He has written over 250 publications with over 7,000 citations. [douglass.post@hpc.mil]

Abstract

Today, rapid innovation in product development is essential to be competitive in any field. It is true for DoD acquisition as well as everyone else. To investigate the potential for computational prototypes and High Performance Computing (HPC) to enable product innovation in DoD acquisition programs, the U.S. Department of Defense HPC Modernization Program (DoD HPCMP) Office initiated the Computational Research and Engineering Acquisition Tools and Environments (CREATE) Program in 2006 (Post et al., 2016). The CREATE goal is to develop and deploy physics-based HPC software applications for the design and analysis of military air craft, ships, and radio frequency antenna systems (and more recently ground vehicles) to enable DoD acquisition programs to improve acquisition outcomes through the construction and analysis of virtual prototypes for those systems. Development of the software applications began in 2008. Ten years later, the CREATE software tools are already beginning to enable DoD engineering organizations (government and industry) to accelerate the rate of innovation in major defense systems, and reduce the cost, time, and risks of acquisition programs for those systems. One aspect of this paradigm is that it enables the DoD to employ features of the Silicon Valley culture that facilitate rapid product innovation.

Introduction

The enabling, disruptive technology that enables accelerated innovation through the use of virtual prototypes is the rapid growth of high performance computing over the last 60 years. Since the end of World War II, the calculating power of computers has grown exponentially from ~1 Floating Point Operation/second (FLOP/s) to over 10^{16} FLOP/s

[1] The CREATE Team consists of Douglass Post, John N. D'Angelo, Saikat Dey, Larry N. Lynch, Robert L. Meakin, Richard L. Vogelsong, S. Morton, A. Wissink, J. Livingston, A. Mackenna, J. Gorski, E. Moyer, Robert Keane, Richard P. Kendall, L. Votta, P. Gibson, D. Borovitcky, L. Miller, S. Allwerdt, Oscar A. Goldfarb, Nathan S. Hariharan, Kevin P. Newmeyer, and approximately 160 other CREATE scientists, engineers, and staff.

(http://www.top500.org/). This means that—for the first time in history—it is possible to make accurate predictions of the behavior of many, many complex physical systems (e.g., the weather, chemical systems, airplanes, ships, automobiles, etc.). We can now develop and deploy science-based software applications for high performance computers that

1. Include the major physical effects that determine the performance of the system,
2. Utilize highly accurate mathematical and numerical solution algorithms,
3. Are verified and experimentally validated,
4. Can predict the performance of a full-scale system (e.g., an entire ship or airplane),
5. Enable multidimensional design of experiments to generate large trade-spaces for a full scale system, and
6. Can complete a high-fidelity, time-dependent, three-dimensional multi-physics calculation for a maneuvering system in a few days that took weeks in 2005, and months (if even possible at the same level of fidelity) in 1995.

This capability enables design engineers to construct realistic virtual prototypes of physical systems (ships, microprocessors, earth moving equipment, etc.) and make accurate predictions of their performance by solving the physics equations that govern their behaviour.

In the past, it was necessary to construct real prototypes for these systems and use live tests to assess their performance and find the design flaws. With simple systems, and incremental changes, there was time in the past to follow the traditional product development paradigm of “design, build, test, fail, re-design” iterated cycles that had proved so successful since the beginning of the industrial revolution. For the standard system engineering product development process (Kossiakoff & Sweet, 2003; see Figure 1), the design of new products is based on “rule of thumb” extrapolations of existing products. Sub-system physical prototypes are developed and experimentally tested during the engineering design phase, and full system physical prototypes are developed and tested just before and during full scale production. With today’s more complex weapon systems such as fighter airplanes, aircraft carriers, tanks, submarines, and so forth, these live tests occur too late to provide timely data on design defects and performance shortfalls. Expensive and time-consuming rework is required to fix the problems uncovered by live testing. The DoD 5000 acquisition process (Carter, 2013) is very similar to the standard systems engineering product development process depicted in Figure 1.

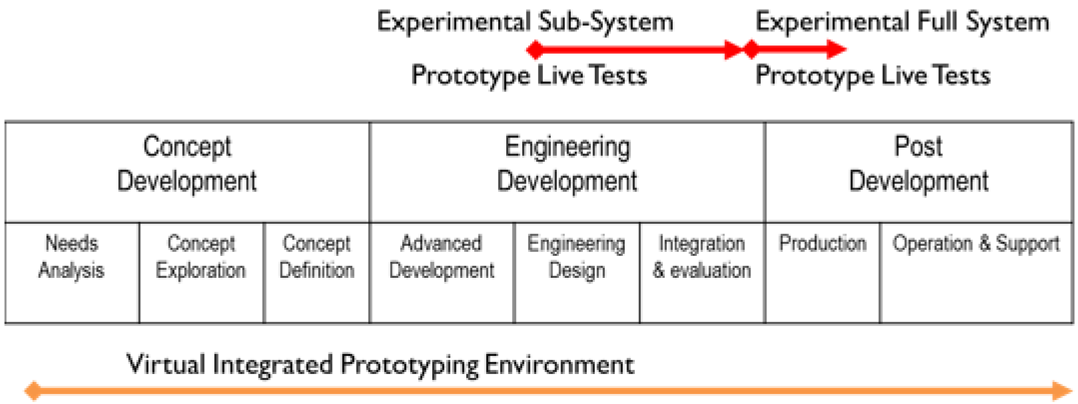


Figure 1. System Engineering Product Development Life Cycle Model
(adapted from Kossiakoff & Sweet, 2003)

The effective use of virtual prototypes requires a high performance computing ecosystem consisting of the appropriate software applications, supercomputers, and a high speed network to connect the users to the supercomputers. For the DoD, that ecosystem is provided by the U.S. DoD HPC Modernization Program (HPCMP) Office, which supplies high-performance computing resources to the Science and Technology, Test and Evaluation, and Acquisition Engineering communities of the Armed Services and DoD agencies. The HPCMP Office provides continuous modernization of five DoD Supercomputing Resource Centers (DSRCs), the network that connects them (DREN), and the associated physics-based simulation software applications in response to prioritized needs of the Services. Recognizing the need to help spur innovation for major defense systems, the HPCMP Office in 2006 launched the Computational Research and Engineering Acquisition Tools and Environments (CREATE) Program. CREATE was chartered to reduce the cost, time, and risks of DoD acquisition programs by developing and deploying multi-disciplinary, physics-based software applications for the design and analysis of military air craft, naval ships, and radio frequency antenna systems by DoD engineering organizations. In 2012, the scope of CREATE was expanded to include ground vehicles.

Using the CREATE tools, engineers can construct virtual prototypes and computationally analyze product performance at any stage of the development process, supplementing or substituting for data from live tests. For conceptual design of new systems, “rule-of-thumb” extrapolations of existing designs can be replaced with physics-based generation of design options allowing extensive trade-space exploration, and rapid assessment of the feasibility and advantages of all the design options with physics-based analysis tools. These engineers can consider many thousands of design options instead of a few. For detailed design development, high-fidelity analysis of virtual prototypes can replace “failure data from live tests” with “physics-based accurate predictions of virtual prototype performance” to drive design iteration and innovation. As the design matures, the digital model can be matured accordingly in terms of completeness of geometry and governing physics. Live tests can then be used to validate the final designs. This provides timely decision data that enables engineers to identify design flaws and performance shortfalls early, allowing problems to be fixed before metal is cut and minimizing rework that cause schedule delays and cost growth. The tools can be applied at all stages of the product development process, from early concept design through operation, support, sustainment, and modernization (Figure 1).

CREATE code development began in 2008, and 10 years later, CREATE is beginning to accomplish those goals. The CREATE tools are being used by more than 160 DoD engineering organizations (50% government, 40% industry, and 10% other) for the assessment of more than 70 DoD acknowledged weapon systems and platforms. There are over 1,400 active CREATE user software licenses, and the number of users is continuing to increase.

The CREATE tools can also help make the testing process more productive, more effective, and more efficient. The tools can be used to identify the most sensitive and uncertain operating conditions so that testing programs can concentrate on those areas, allowing the total test data requirements to be reduced a factor of five or more with a concomitant reduction in the required testing time (Kraft, 2010). In addition, it frees the testing community to address the basic scientific and engineering issues that determine weapon system performance as well as enabling a much greater number of test events with the existing test facilities. The software allows the DoD Test and Evaluation (T&E) community to rehearse testing events and design the test experiments.

Many industries and federal agencies have successfully adopted the virtual prototype paradigm (Council on Competitiveness, 2010; Francis, 1995; Miller, 2010) to obtain a competitive advantage in their market place. Until recently, however, there has been little public information on the value of the paradigm, chiefly because it gives organizations that use it a significant competitive advantage, and their use of the paradigm is considered a trade secret. A few examples are publically available. The Partnership for Advanced Computing in Europe (PRACE) established an Automotive Simulation Center (ASC) in Stuttgart, with projects in vehicle drive, vehicle structure, vehicle physics, and so forth. ASC director, Alexander Walser, noted in 2014 that "numerical simulation made its way in the design phase of automotive development and productions (as) a useful tool for faster problem analysis and reduction of cost and product design time" (Walser, 2014). Specific benefits, however, were not detailed. In an aerospace industry example, however, Doug Ball, then Boeing's chief engineer for enabling technology and research, stated, "When we were designing the 767 back in the 1980s, we built and tested about 77 wings. By using supercomputers to simulate the properties of the wings on recent models such as the 787 and the 747-8, we only had to design (*and test*) seven wings, a tremendous savings in time and cost" (Ball, 2009).

Starting in 1992, Goodyear developed a physics-based design tool for tires that became the core of their "innovation engine" (Miller, 2010). Goodyear CEO Richard Kramer noted in the Q4 2010 Earnings Call, "Our *innovation engine* again delivered in 2010. The percentage of new products in our overall lineup is the highest ever. … Our innovative new products continue to accumulate an impressive list of test wins and third-party endorsements" (Kramer, 2011). In the *2009 Annual Report,* then Goodyear CEO Robert Keagan stated that "Our new product engine is poised to take advantage of the demand for high-value-added tires and to do so with unmatched speed to market" (Keagan, 2010, pp. 2–3). With this approach, Goodyear reduced its product development time from three years to as little as eight months and reduced its new product prototyping and testing costs from 40% to 15% of the R&D budget, an annual savings of $100 million (Engardio, 2008).

The CREATE Program

CREATE code development began in 2008 as a set of four projects (Air Vehicles, Naval Ships, Radio Frequency Antennas, and Meshing and Geometry). A fifth project was added in 2012 for Ground Vehicles. Combined, the projects are developing and deploying 11 individual software applications (Table 1). The choice of the projects was dictated by the relative size and importance of major defense acquisition programs and the potential for HPC and physics-based software applications to predict the performance of associated systems.

The CREATE projects have two types of software products: (1) concept development tools (viz., DaVinci, RSDE/IHDE, a module of SENTRi, and MAT) to generate conceptual designs and analyze their feasibility and performance using fast, but lower-fidelity analytics; and (2) high-fidelity, multi-disciplinary tools (Kestrel, Helios, NESM, NavyFOAM, SENTRi, and Mercury) to provide accurate predictions of the system performance. The Meshing and Geometry project is subordinate to the other four CREATE projects and provides capability for concept design tools to generate numerical representations of the platform and then to generate the meshes needed for computational analysis. The Meshing and Geometry tool (Capstone) can also produce meshes from geometry representations generated by the most commonly used commercial CAD tools. Capstone provides a numerical representation of the geometry of the weapon system of interest, a digital prototype of the weapon system.

The CREATE tools have been developed, deployed, and supported by the DoD (DoD employees and contractors). They are validated with DoD experimental data for DoD use cases, and are "owned" by the DoD. The DoD has government purpose distribution rights to all of the CREATE software. The CREATE tools give DoD acquisition program engineers the ability to make independent assessments of proposed designs and contractor deliverables. In this role, the tools are directly helping the DoD acquisition engineering community grow its organic engineering capability, a DoD priority for technical workforce development.

The Eleven HPCMP CREATE™ Software Applications

CREATE Air Vehicles

DaVinci—Concept Design Tool for Air Vehicles

DaVinci will allow engineers to populate design option spaces of fixed and rotary wing aircraft and provide an initial assessment of the performance of the design. Choosing from previously engineered components, the tool will allow engineers to select and modify wings, fuselage, and propulsion components. Aircraft designers then can select, adjust, and rearrange internal components in aircraft designs. It is "Model Centric." It will support highly efficient construction and maintenance of air vehicle models, including geometry that is parametric and includes water-tight external geometry, internal structure, subsystem layout, volumes, and mass properties. It will provide a multi-disciplinary, physics-based analysis capability that is variable fidelity, but is consistent across disciplines. It provides the ability to persist design data and intent throughout acquisition and program life. The virtual model can be tested virtually through its flight envelope to assess basic performance characters, including mission performance, and decision support with uncertainty quantification and sensitivity analysis.

Table 1. CREATE Projects and Products

Projects	Products
Air Vehicles (CREATE-AV)	DaVinci (Rapid conceptual design)
	Kestrel (High-fidelity, full-vehicle, multi-physics analysis tool for fixed-wing aircraft)
	Helios (High-fidelity, full-vehicle, multi-physics analysis tool for rotary-wing aircraft)
Naval Ships (CREATE-SH)	Rapid Ship Design Environment (RSDE; Rapid Design and Synthesis Capability)
	Navy Enhanced Sierra Mechanics (NESM; Ship Shock & Shock Damage Assessment)
	NavyFOAM (Hydrodynamic-predict hydrodynamic performance)
	Integrated Hydro Design Environment (IHDE; Facilitates hydrodynamic performance evaluations in early stage ship design)
RF Antennas (CREATE-RF)	SENTRi (Electromagnetics antenna design integrated with platforms)
Ground Vehicles (CREATE-GV)	Mercury (physics-based tool for M&S of terrain mechanics and vehicle systems and components. Incorporates suspension, tire and track, soil modeling, and powertrain simulation)
	Mobility Analysis Tool (MAT; converts physics-based vehicle performance data and terrain information into mission-based analysis of performance over large areas of terrain)
Meshing & Geometry (CREATE-MG)	Capstone (Components for generating geometries and meshes needed for analysis)

The current version of DaVinci (V3.1) has the capability to build parametric designs for fighter, transport, and surveillance aircraft that include the Outer Mold Line (OML) and structural envelope. It supports geometric analysis including areas, volumes, centroids and moments, and high fidelity aerodynamic analysis using the high-fidelity tool Kestrel. It can design and build a geometric model of a platform, and build a mesh that captures this geometry with Capstone for high fidelity analysis with Kestrel. Surrogate geometry models have been built for the support of the KC-46, the new Air Force Tanker based on the Boeing 767, and the Joint Surveillance and Target Attack Radar System (JSTARS).

Kestrel—High-Fidelity Tool for Fixed Wing Air Vehicle Performance Prediction

Kestrel has the capability to provide accurate predictions of the performance of DoD air vehicles, with a specific focus on the fixed-wing community. It integrates computational fluid dynamics, structural dynamics, propulsion, and control for sub-sonic through supersonic aircraft operation. In detail, the capabilities available in Kestrel v5.0 include (1) Aerodynamics (Navier-Stokes solvers and a full suite of boundary conditions and turbulence models), (2) Structural Dynamics (Modal models or Finite Element Analysis for aero-structure interactions), (3) Flight Control Systems (Control surface movement—deforming geometry or overset), and (4) Propulsion (Engine “cycle-decks” for propulsion effects, or direct engine simulation including inlet and rotating machinery, nozzle, and moving walls).

This set of capabilities is unique in the international aerospace community. It provides the capability to develop major innovations in the design of next generation aeronautical weapon systems. With Kestrel, engineers can verify designs prior to key decision points (and prior to fabrication of test articles or full-scale prototypes), plan and rehearse wind-tunnel and full-scale flight tests, evaluate planned (or potential) operational use scenarios, perform flight certifications (e.g., airworthiness, flight envelope expansion, mishap investigation, etc.), and generate response surfaces usable in DaVinci, flight-simulators, and other environments that require real-time access to performance data. Kestrel has been applied to the analysis of over 30 fixed-wing DoD aviation systems including store separation, A-10, F-18E, F-15, B-52, E-2D, P-3, and many others. Dr. Theresa Shafer, an engineer at the Patuxent River Naval Air Station, was awarded the American Society of Naval Engineers (ASNE) 2014 Rosenblatt Young Naval Engineer Award for her career accomplishments, including her work using Kestrel to produce flight certification and airworthiness data for seven small unmanned aerial vehicles that enabled seven small aerospace companies to bid on NAVAIR development contracts.

Helios—High-Fidelity Tool for Rotary Wing Air Vehicle Performance Prediction

Helios is a high-fidelity, full-vehicle, multi-physics analysis tool for rotary-wing aircraft. Helios v5.0 can calculate the performance of a full sized rotorcraft, including the fuselage and rotors. It can handle arbitrary rotor configurations (e.g., conventional main rotor/tail-fan, co-axial main rotor/pusher propeller, tandem main rotors, tiltrotors, quad-tiltrotors, etc.). It has the capability to analyze and predict prescribed maneuvers with tight coupling of rotor aero-structural dynamics. A highly accurate treatment of the vortex shedding from the rotor blade tips using adaptive mesh refinement gives Helios a unique capability to assess the interaction of these vortices with the fuselage and nearby rotor blades. Helios can provide all the benefits for rotary-winged aircraft that Kestrel can for fixed-wing aircraft.

There have already been important examples of the use and value of Helios. The Army Rotorcraft Program (AMRDEC/AED) used Helios with Boeing to generate early design stage predictions of helicopter performance for a proposed rotor blade upgrade for the CH-47F helicopter (Chinook) to achieve up to an estimated 2,000 pounds improved hover thrust for 400+ Chinooks with limited degradation of forward flight performance. The Army Joint-Multi-Role Technology Demonstrator (JMR-TD) Program used Helios to provide decision data on the proposals from four vendors for the JMR-TD program. Helios enabled government engineers to provide the government the ability to conduct an independent analysis of the contractor proposals. The Army Rotorcraft Program (AMRDEC/AED) is using Helios to assess the H-60 tail rotor effectiveness for providing directional control of aircraft in combination with increased engine power and main rotor performance.

CREATE Ships

RSDE—Rapid Ship Design Environment (Rapid Concept Design for Ships)

RSDE is a concept design tool that allows engineers and naval architects to assess the tradeoffs inherent in designing ships to meet a spectrum of competing key performance parameters. Employing the concept of design space exploration, engineers and naval architects can provide data for decision makers on the impact of tradeoffs in range, speed, armament, aviation support, etc. on the size and, in large measure, the cost of a proposed ship concept. RSDE can generate tens of thousands of candidate ship designs with varying hullforms, subdivision, and machinery arrangements. An initial assessment of the intact and damaged stability and resistance, and an initial structural design and analysis is done for each candidate ship design. RSDE has been used to enable set-based design (Singer, Doerry, & Buckley, 2009) on Navy acquisition programs. This design method allows down-selection of a ship design to occur later in the process when the tradeoffs are more fully understood. It has been applied to numerous ship design studies including the Amphibious Landing Craft LX(R) Analysis of Alternatives, and the Small Surface Combatant Trade Study. Dr. Adrian Mackenna, the team leader for the RSDE tool, was awarded the 2014 American Society of Naval Engineers (ASNE) Gold Medal for his work developing and applying the RSDE tool.

NESM—Navy Enhanced Sierra Mechanics (Ship Shock & Shock Damage Assessment)

NESM builds on the Department of Energy's Sandia National Laboratory shock analysis tool Sierra Mechanics to provide a means to assess ship and component response to external shock and blast using accurate high performance computational tools. NESM can reduce the time and expense required for physical shock testing of ship classes and also improves the initial ship design process by assessing planned component installations for shock performance prior to final arrangement and installations decisions. The tightly coupled multi-physics capabilities include (1) Structural Dynamics (Implicit linear-elastic solver: static, modal, transient, acoustics, and more), (2) Solid Mechanics (Explicit plasticity solver: failure, high-strain, multi-grid, and more), (3) Fluid Dynamics (Euler solver: shock propagation, load environments, and threat modeling), and (4) Fluid-Structure Interaction. The solution algorithms in NESM can exploit massively parallel computers, and can scale to thousands of cores, enabling efficient computer use and the ability to address full-sized naval vessels up and including next generation aircraft carriers and submarines.

NESM will materially contribute to the design of next generation naval weapon systems and platforms, support planning and rehearsal of ship tests prior to Life Fire Testing (more "bang" per test dollar), and the evaluation of planned (or potential) operational use scenarios. NESM has been officially adopted by the Navy for these uses. "The NAVSEA Technical Warrant (for Shock/Ships) concurs that NESM is the appropriate and technically acceptable modeling and simulation (M&S) tool which meets the M&S requirements to support current and future surface ship shock applications." NESM was previously approved for "Full Ship Shock Trials (FSST) Alternative R&D Programs (PEO Ships & PEO Carriers)," which led to the release of OPNAVINST 9072.2A, providing future ship classes with an alternative to Full Scale Shock Trials. NESM has been used to support Littoral Combat Ship (LCS) Live-Fire Test & Evaluation (LFT&E) and the USS *Cole* Validation Study, and to provide support for Live Fire Test and Evaluation for the Navy's next generation Nuclear Aircraft Carrier (CVN-78 and 79). Dr. Thomas Moyer, the NESM team leader, was awarded the ASNE 2015 Soldberg Award for his pioneering research modeling shock effects in naval systems, including leading the NESM team.

NavyFOAM—High Fidelity Predictions of Ship Hydrodynamic Performance

NavyFOAM is based on the OpenFOAM (http://www.openfoam.org) libraries and code architecture. To that base, we have added a number of features and capabilities that enable simulation of the air-sea interface (e.g., surface waves) and other effects important for naval vessels. NavyFOAM is a fully parallelized, multi-physics computational fluid dynamics (CFD) framework developed using modern object-oriented programming (OOP). The code enables high-fidelity hydrodynamic analysis and prediction of ship performance such as resistance, propulsion, maneuvering, seakeeping and seaway loads. It has demonstrated accuracy against experimental data for a number of target applications such as resistance, propeller characteristics, hull/propulsor interaction, and six-degree-of-freedom ship motion of underwater vehicles and surface ships. Offering a suite of Navier-Stokes–based flow solvers tailored to specific applications including single- and multi-phase solvers, NavyFOAM allows assessment of alternative hull and propulsor designs. With NavyFOAM users can evaluate a ship's performance in a wide array of operating conditions and sea-states including both subsea and surface operations. Its modularity and software architecture expedites coupling with third-party software and collaborative multi-disciplinary software development (e.g., fluid-structure interaction, hydroacoustics). It has been applied to many naval systems including assessment of the safe operating envelope of the DDG-1000, propeller designs, the USMC Amphibious Combat Vehicle, the Columbia Ballistic Missile Submarine Program (a $100 billion procurement to replace the aging Ohio-class ballistic missile submarines) and many other systems of interest to the U.S. Navy.

IHDE—Integrated Hydrodynamic Design Environment (Facilitates Hydrodynamic Performance Evaluations for Early Stage Ship Design)

IHDE is a desktop application that integrates a suite of Navy hullform design and analysis tools allowing a user to perform evaluations of performance, including visualization, in a simplified and timely manner from a single interface. Prior to the development of IHDE, naval architects and marine engineers often had to learn how to use a dozen or more individual design tools, each with a different user interface and input format. IHDE provides a single interface for access to all of the tools. In a few days to weeks, a single user with IHDE can finish projects that used to take several highly experienced users many months to complete. Current capabilities are geared toward surface ships, both monohulls and multihulls—including catamarans and trimarans. Typical uses include predicting (1) resistance in calm water, (2) seakeeping behavior in waves, (3) hydrodynamic loads due to wave slamming, and (4) operability (percentage of time a ship can carry out its particular mission in various parts of the world based on historic sea state data).

The U.S. Navy's Center for Innovation in Ship Design (CISD) has used IHDE to assess the performance of many ship designs, including (1) T-AGOS-19 Ocean Surveillance ship; (2) Hospital ship (Mercy) replacement design; (3) Salvage Tow & Rescue (T-STAR); (4) Green Arctic Patrol Vessel (GPAV); (5) Medium Affordable Surface Combatant (MASC); and (6) an optimized MASC. IHDE is also an important adjunct capability for medium fidelity analysis of ship designs developed with RSDE. It was used with RSDE as part of the Amphibious Landing Craft LX(R) Analysis of Alternatives and the Small Surface Combatant Trade study. Its use with high performance computers will allow engineers to rapidly assess the major performance parameters of thousands to hundreds of thousands of candidate design options.

SENTRi—Electromagnetic Tools for DoD Systems

SENTRi is a robust and high-fidelity Full Wave electromagnetic prediction code for Radio Frequency (RF) modeling of antennas, microwave circuits, and radar cross-section prediction. SENTRi is designed for the modeling of complex structures—including highly heterogeneous material structures with multi-scaled features. A key goal is the calculation of the simultaneous performance of multiple-antenna systems embedded on a platform. The key features for electromagnetics are based on solutions of Maxwell's equations with advanced hybrid finite-element boundary-integral techniques. This provides high accuracy with the ability to solve large, complex problems. SENTRi is continuously validated with DoD measurements. SENTRi is being used for antenna design, antenna in-situ analysis, RF signature prediction, Electromagnetic Interference (EMI), Electromagnetic Compatibility (EMC), material modeling, microwave device analysis (i.e., waveguides, filters, circulators, power dividers), phased array antenna systems, and apertures (i.e., radomes, windows, frequency selective surfaces). SENTRi is being used by approximately 60 DoD organizations (government and industry).

CREATE Ground Vehicles (GV)

Mercury—Modeling and Simulation of Terrain Mechanics and Ground Vehicle Systems

Mercury incorporates suspension, tire and track, soil modeling, and powertrain simulation, and also integrates physics domains for powertrains, vehicle dynamics (wheels and tracks), and tire-soil and track-soil interaction. It simulates multiple performance tests used in vehicle acquisition: driver comfort (Ride/Shock), soft soil mobility (VCI1, sand-slope), maximum speed, vehicle stability (lane changes, circular turns). Mercury allows thousands of design concepts where the user can vary spring and damper properties, vehicle mass and inertia, tire properties, and axle spacing and location to be tested in a single simulation and provides performance metrics.

MAT—Mobility Analysis Tool

MAT is a computational tool for analyzing HPC physics data and producing mobility performance metrics required for trade exploration and systems engineering. It incorporates soil condition, vehicle performance and configuration, vegetation density, average surface roughness, average slope, and so forth. MAT converts physics-based vehicle performance data and terrain information into mission-based analysis of performance over large areas of terrain to predict percent GO/NOGO across selected terrains of interest and mission rating speeds. MAT interfaces with Mercury to use simulated performance data to provide performance metrics for concept designs.

Capstone—Rapid Geometry and Mesh generation

Capstone is a CAD-neutral application that provides two distinct capabilities. The first is the capability to develop numerical representations of a DoD weapon system (i.e., a Nonuniform Rational B-Splines [NURBS]–based digital product model consisting of the platform geometry with the associated attributes). The second is the capability to generate a mesh from the geometry. Valid and easily produced meshes with the required accuracy are the essential starting point for the other CREATE (solver) tools for detailed analysis. In addition, a number of non-CREATE groups use Capstone for its geometry and mesh generation capability for their applications.

A digital product model has many advantages for acquisition. It enables automated design optimization. It facilitates the transfer of design information between the government and contractors, eliminating much of the reliance on paper documents, and improving the accuracy and speed of information flow. It provides a permanent, analysable description of the platform through all stages of the acquisition process. Copies of the product model can be generated and assigned to individual airplanes and other systems allowing the DoD to track the history, performance and maintenance of entire life cycle of each individual platform. Together with DaVinci, which builds on top of the Capstone platform, it enables the recent Air Force initiatives of the Digital Thread and the Digital Twin (Kraft, 2015).

CREATE Program Organization and Management

To develop and deploy the CREATE software applications, we worked with the DoD organizations responsible for overseeing the design and analysis of air vehicles, naval vessels, RF antennas, and ground vehicles (Figure 2) to empower them to develop and deploy the tools. The CREATE Program first formed five projects with a total of 11 multi-disciplinary teams of DoD subject matter experts. Then we jointly identified a team leader within a DoD customer organization for each software product who possessed the right mix of subject matter and high performance computing expertise, and the required leadership, program, and project management skills. Then we helped the leader build a multidisciplinary software development team with approximately 10–15 members. Each core development group is located within a customer organization (e.g., the Navy's ship design groups are located at the Naval Surface Warfare Center, Carderock Division, Bethesda, MD, so the CREATE Ships team members are also Carderock employees and contractors and located there). Additional developers are drawn from other organizations as needed to provide required expertise not available in the customer organization. For instance, the structural dynamics modules for NESM and Kestrel are being supplied by the Sierra Mechanics group at the U.S. Department of Energy's (DOE's) Sandia National Laboratory as part of a highly productive DOE/DoD inter-agency collaboration. Generally about one third of each team resides at the core organization, and the remaining two-thirds are located at other organizations. The CREATE staffing mix is about 90 DoD employees and about 90 DoD support contractors. The team members are distributed across ~30 collaborating organizations.

Embedding the code teams and the team leaders in the relevant DoD customer organization greatly improves our ability to recruit the most capable talent in the DoD for each technical area. It also helps ensure adoption and ownership of the CREATE tools by the relevant Service since their experts are responsible for developing the tools and are "trusted agents" of that Service. In many cases the design engineers for the relevant weapon systems are collocated with CREATE tool developers. This helps the developers get rapid feedback on the usability and accuracy of the code, and a sense of satisfaction from directly seeing the impact that the code is having on the DoD.

The CREATE Program leaders sponsor each team by providing funding and active management and oversight of the code development process. We developed a set of software project management and software engineering practices for the CREATE Program and promulgated them to the teams as guidance. We sought a balance between a very agile code development process to allow the code development teams the flexibility to accomplish technically difficult tasks while ensuring adequate accountability together with an organized code development process. The HPCMP and CREATE Program Office actively manages the 11 teams cooperatively with the hosting Service organizations. While clear lines of authority and obligation have been formally established between the HPCMP CREATE program and each executing and hosting organization, both groups have developed a high degree of trust and work together to resolve conflicts. There is a strong degree of alignment on the technical aspects of the CREATE Program between the CREATE Program leadership and the Service organizations.

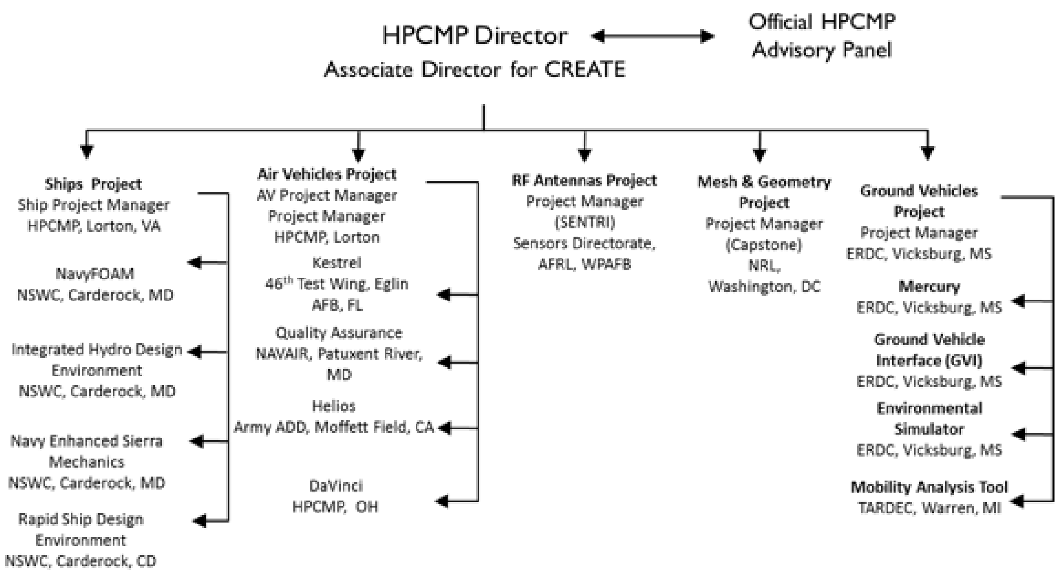


Figure 2. CREATE Program Organization Chart

Each project has a board of directors consisting of senior members of the relevant customer organization (Navy, Air Force, and Army acquisition engineering communities). Each board meets at least once a year. The boards review the progress during the prior year, advise the project about new requirements, and serve as liaisons between the CREATE projects and their DoD customer community. The board members are members of the Senior Executive Service, or other senior staff of the Navy, Air Force, or Army. The boards provide an additional mechanism for ensuring that the CREATE tools are aligned with the needs of the Services.

The annual CREATE budget is about $30 million/year. The total investment in CREATE by the DoD HPCMP from 2008 through 2015 is ~$200 million. The Services provide "in-kind" contributions of another ~$11 million/year, a testament to the value of CREATE to them. The Service contributions include office space and supplies, administrative support, network and other host services, additional professional staff for the code development teams, and access to validation experiments and data.

Building the Right Software Right

At the beginning of the CREATE Program, we developed a vision for how the DoD could implement virtual prototypes with physics-based high performance computing engineering software within its own processes to "modernize" its acquisition process. We then fleshed out that vision through joint assessments with each service of their detailed acquisition processes to identify the specific tools needed to reduce the time, cost, and risk and improve the system performance for the Service's acquisition programs. For instance, the CREATE Air Vehicles Program assessed 27 different acquisition workflows to develop its requirements. This vision is captured in an "Initial Capability Document" (ICD) for each project. The ICDs were reviewed and approved by the Board of Directors for each project and are reviewed periodically.

Although CREATE was proposed to be a 12-year program, the CREATE tools are designed for a 30 to 40-year life since that is the expected life span of successful engineering codes. Since the DoD spends roughly $200 billion/year acquiring, maintaining, sustaining, and modifying major weapon systems, the expectation has been that if the CREATE tools were successful in enabling the DoD to significantly improve acquisition outcomes for an expenditure of ~$30 million/year (0.00015 of $200 billion/year), the tools would be supported and continue to be modernized until they were no longer needed.

Fiscal Year	FY2011				FY2012				FY2013				FY2014				FY2015				FY2016				FY2017				FY2018 Planned				FY2019 Planned			
Quarter	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
AV-DaVinci					1							2								3				4				5				6				7
AV-Helios					2				3			4					5				6					7				8				9		
AV-Kestrel			2						3			4				5			6					7			8				9				10	
MG-CAPSTONE	1				2				3				4				5				6				7				8				9			
RF-SENTRi				2					3				4				5				6					7				8				9		
Ships-IHDE	2				3				4				5				6				7				8				9				10			
Ships-NavyFOAM	1				2				3					4			5				6				7				8				9			
Ships-NESM	1				1.1					2					2.1				3				4				5				6				7	
Ships-RSDE					0.5				1.0				1.1				1.2					2						3				4				5
GV-Mercury																												1				2				3

Figure 3. CREATE Annual Releases With Version Numbers

Note. The FY2017 3rd quarter and subsequent releases are planned. Some CREATE products had releases prior to 2011.

Each CREATE code development team follows a highly disciplined software development process. There is a strong emphasis on software quality. To facilitate development by non-collocated teams, the CREATE Program, through the DoD HPCMP, provides a supportive code development environment with virtual clusters, central servers with code and document repositories, issue trackers, user and developer forums, configuration management services, and access to high performance computers for testing and performance enhancement, and high quality video conferencing. Guided by the ICD, a 12-year product roadmap and feedback from the BODs and customers, each product team issues a new release each year with the upgraded capability and new features needed by the customer communities. This places the upgraded tool into the hands of the user community and gives the code development team rapid feedback on the quality and usefulness of the upgraded code. The annual release cadence (Figure 3) adds discipline and accountability to the development process and is a key factor in the success of the CREATE development process.

The CREATE teams generally use an “agile” development process tailored for their environment and code. The releases are designed and tested for all the standard Linux and Unix HPC operating systems, as well as MacOS and Windows where appropriate. Each release is extensively tested both during the development process, and as an integral part of the release process. Each software release is documented with a (1)Product Technical Description that describes the physics and engineering capabilities (including the equations) in the code, the computer science approaches, software architecture, and solution algorithms; (2) Developers Manual that describes the source code in detail, provides an index/table of contents, and other information essential for understanding the source code; (3) Users’ Manual to help the users set up their problems, run the code, and analyze the results; and (4) Test Plan with an archive of test problems with the input and test results. In addition, there are tutorials and a user forum on the CREATE server, all backed up by a user support group.

Information Assurance for the CREATE Products

Information assurance for computational engineering applications can be understood from its role in the high performance computing (HPC) ecosystem. The CREATE tools are part of an ecosystem consisting of (1) Subject Matter Experts (SMEs) who use the tools on high performance computers over high speed networks to generate virtual prototypes and analyze their performance; (2) software applications like the CREATE tools that can be used to generate virtual prototypes and predict their performance; (3) Experimental testing organizations to generate validation data to establish the validity of the models that are the basis of the software applications; (4) high speed computer networks that provide the SMEs access to supercomputers; (5) supercomputers; and (6) sponsors who need the results of the calculations and provide the funds to generate the results.

CREATE is designed to provide the DoD with a military competitive advantage. Thus the CREATE Program and the DoD must control the distribution and use of the codes to sustain that advantage. The CREATE codes are unclassified but are subject to the International Traffic in Arms Regulations (ITAR). To be effective, DoD users must be able to run the CREATE codes securely and access their proprietary data on a high performance computer. The codes and data are encrypted at rest and accessed and transferred over a secure encrypted network with two-factor authentication. To ensure integrity of the codes and the users' data, the codes and the user's data are archived on the secure HPCMP supercomputers and backed up frequently to several remote secure data repositories.

However, many, if not most, DoD users (engineers) have access only to a Windows Personal Computer (PC) with Microsoft Office and a browser. Usually no other software is allowed on these PCs so those engineers cannot easily access the HPCMP supercomputers and the CREATE tools. To remove this barrier to access, the DoD HPCMP and the CREATE program have developed a "portal" that allows DoD users to access the DoD HPCMP supercomputers through their browser. The "portal" features two-factor authentication and encrypted data transfer. It allows users to securely set up their job; run it; and store, analyze, and visualize the results through their browser.

To prevent unauthorized access to the CREATE codes and to the intellectual property of the users, we limit access to codes to DoD employees or DoD contractors who have a valid reason for access to the CREATE software. They must sign a software distribution agreement that describes the limitations of their use (not to redistribute the code, reverse engineer it, etc.) and their intended use. They also agree to abide by the ITAR procedures which have civil and criminal penalties if violated. The CREATE source code is only accessible to the development team. The ideal is a "Software as a Service" model where the user can only execute the code, but not get a copy, even an executable. However, that's not practical for some users, and those users are handled on a case-by-case basis.

Intellectual Property Rights

The DoD HPCMP must have "government purpose rights" to be able to distribute the CREATE software to users. Even a single line of code for which the DoD doesn't have these rights would leave the DoD vulnerable to lawsuits and large financial settlements for copyright or patent infringement and theft of intellectual property. Legal reviews of the ~40 CREATE support contracts have determined that the DoD does have "government purpose rights." Remedial action was necessary for some of the contracts.

It is DoD policy to share DoD RDT&E results with the U.S. defense industry if it is in the interests of the DoD because a strong U.S. defense industry is essential for national security (Weiss, 2014). Several defense industries have expressed a strong interest in adopting the CREATE tools for use in their commercial as well as military design work. This requires that the DoD have “unrestricted rights” for the relevant CREATE software. As the result of further extensive legal reviews of the CREATE AV contracts over the last three years, together with additional remedial action, the DoD now has “unrestricted rights” to the CREATE AV tools. We anticipate that several large defense companies will be using CREATE AV tools for both military and commercial systems, and a few key defense industries have already expressed interest in this type of use.

Transition and Adoption

Transition of research results to applications and products has been a historic challenge for much of the science and technology research done by the DoD research community. The approach adopted by the CREATE Program of embedding the CREATE development teams in the DoD customer organizations responsible for the design and development of the relevant weapon system has been very successful in overcoming that challenge. The CREATE teams are trusted agents of their DoD organization. For example, to assess various options for a follow-on to the Littoral Combat Ship, or to develop a concept for the Ohio replacement submarine, the Navy turned to the trusted organizations responsible for those tasks, the Naval Architecture and Engineering Department at the Naval Surface Warfare Center at Carderock. These were the same groups developing the ship design tool, RSDE and the hydrodynamics tool NavyFOAM. The transition was almost automatic.

In general, computational design tools are very effective methods for capturing the corporate knowledge and new research results in a field, and giving design engineers access to that knowledge in a tool that the engineer can use to design a new system. This greatly facilitates the development of innovative designs. It also gives a design engineer the opportunity to compare the impact of new research results in the context of present practice, and allows the engineer to answer many “What if?” questions and define the benefit of the research for the system of interest. It facilitates a successful transition from research to practice, a transition across the “Valley of Death.” This has been recognized by a number of groups in the DoD science and technology research community. Some of those groups are beginning to work with the CREATE team to incorporate their research results into the CREATE codes as a means to transition their research results to acquisition programs. For instance, the CREATE Kestrel group is working with the DoD hypersonics R&D community to develop a version of Kestrel that incorporates the most recent research results on the behaviour of hypersonic air platforms. Aerospace engineers are familiar with Kestrel as a validated, user-friendly multi-physics aerospace engineering tool for sub-sonic, transition, and super-sonic flight conditions. In a few years, they will have access to the same mature engineering tool that has been upgraded to handle supersonic flight.

Accelerating the Development of Innovative Weapon Systems With Virtual Prototyping and HPC

The CREATE tools enhance the ability of DoD engineers to develop innovative weapon systems. With the CREATE tools, these engineers can develop thousands, even hundreds of thousands of design options for potential weapon systems, capture their properties in digital prototypes, assess the feasibility and capability of each prototype with physics-based tools, perform trade-space studies of the prototypes, and develop optimized designs. Then high-fidelity tools such as Kestrel and NavyFOAM can be used to assess the performance of the final selected design option, a “virtual test.” This process can be done in weeks to months for hundreds of thousands of design options, much faster than is the case with physical prototypes that are currently proving to take tens of years to design, develop, and test prototypes for only a few candidate design options. This allows the rapid identification of design defects and performance shortfalls, well before metal is cut. In addition, changes in requirements can be inserted into the design process at almost any point until very late in the design process.

History convincingly illustrates that continual innovation in military technology is necessary for achieving and sustaining the competitive military technological advantage needed for national survival (Colinvaux, 1980; Kennedy, 1987). By most accounts, the worldwide rate of technological change will remain high, or even increase (Desilver, 2014). In this context, the United States would benefit from a faster and more agile major weapon development process. DoD leadership has recognized this and launched an effort to tap the innovation skills of Silicon Valley (the Defense Innovation Unit Experimental; see https://www.diux.mil) for DoD acquisition. Virtual prototypes and high performance computing offer the DoD an additional opportunity to apply a number of the features of the innovation culture of Silicon Valley (Table 2) identified in several short papers published by the *Harvard Business Review* (Anthony, 2013; Fox, 2014; Martins, Dias, & Khanna, 2016) to accelerate innovation in the acquisition of major DoD weapon systems.

Innovation requires the right people working in the right environment and with the right tools. In Silicon Valley the product development teams at successful companies are small, able to take risks, rapidly develop and try many new product features, and go through many trials and failures until success is achieved. The development and design teams own the development process. They have the autonomy to be resourceful and make decisions. They are able to learn from failures and adapt as the product design evolves. They have a day-to-day determination to see something through despite near-constant failure. The corporate environment is generally flat. Corporate management is accessible to the teams for advice, support, and requests for resources. All the expertise in the company is accessible to the teams. There is continuity in the corporate leadership, leading to continuity in the corporate memory.

Table 2. Some Key Features of Silicon Valley Innovation Culture

- Small teams with significant autonomy that are empowered to take risks and make decisions
- Generally flat organizations, but with a clear hierarchy
- Early development and testing of many options and alternatives
- Early identification of things that work and those that do not
- Fanatical pursuit of promising options to a successful conclusion
- Leadership continuity
- Close working relations and connections with customer communities
- Emphasis on incremental improvements and modifications, as opposed to huge leaps
- Silicon Valley mostly focused on information technology with small sized products (smartphones, integrated circuits, music players, computers, calculators, etc.), but—
  - New aerospace and automotive start-ups such as Space-X, Blue Origin, Virgin Galactic, Facebook Aquila, Tesla, Google Car, etc. with industrial scale products retain many of their Silicon Valley roots and features.

The best teams are typically small and almost fanatically focused on producing features that attract users and customers. The team members work collaboratively with each other with a minimum of structure and formality. They move around the company and the industry, learning new skills and encountering new ideas. Organizations and companies are fairly flat, not rigidly hierarchal. Management typically sets goals and enforces accountability, but the teams have the agility and flexibility to develop the product. Rigid planning is not conducive to inventing something new. This is not what one finds in large aerospace industries or the DoD, even if the operational part of the DoD does share many of these features (Reinertsen, 2009).

Another feature common in Silicon Valley is that successful innovation is often more an evolutionary improvement than a discontinuous and revolutionary advance over present capabilities. New products in Silicon Valley are often incremental improvements of yesterday's products. Today's iPhone7 is the 20th generational descendant of the Apple Newton introduced in 1993 (Table 3). In contrast, many large defense acquisition projects are structured to produce products that embody very large advances over the capabilities of existing systems with few or no intermediate steps. While the F-22 was a major step forward in terms of stealth, it took 19 years from the start of the project to the delivery of the first operational F-22. During the 26 years of the F-22 program, the Soviets/Russians fielded six generations of surface-to-air missiles. In contrast, innovation in Silicon Valley is usually the result of many failures that lead to a series of small successes and small evolutionary advances (e.g., iPhone 5 to iPhone 6), ending up in revolutionary advances in capability (i.e., Newton to the iPhone 7 over the course of 23 years).

Table 3. History of the iPhone

("History of the iPhone," n.d.)

1993	Newton
1994	120
1998	Apple Message Pad 2100
2005	ROKR
2006	
2007	iPhone (1st generation)
2008	iPhone 3G
2009	iPhone 3GS
2010	iPhone 4
2011	iPhone 4S
2012	iPhone 5
2013	iPhone 5S iPhone 5C
2014	iPhone 6/6+
2015	iPhone 6S/6S Plus
2016	iPhone SE iPhone 7/7 Plus

This incremental approach may not appear to be immediately applicable to the latter stages of the development of large-scale weapon systems, particularly aircraft carriers, submarines, large surface ships, or many other complex weapon systems. It now takes at least 10 to 15 years to design and build a nuclear aircraft carrier for ~$13 billion or more. The cost and national importance of this type of system provides considerable incentives to minimize risk, so innovations must be and are introduced cautiously. The infrastructure to build large systems also takes time to construct, either by modification of existing facilities or construction of new ones. However, as noted above, many aerospace and automotive start-ups that can trace their roots to Silicon Valley start-ups (e.g., Space-X), are building large-scale, complex systems, and are certainly among the most innovative members of their industry. Many strongly emphasize the use of computational prototypes with high performance computing. The advantages of computational prototyping are still very applicable to the early stages of design concept development and detailed engineering design. Also, even though most IT products are small in physical size compared to ships and airplanes, the infrastructure required to build them, such as chip fabrication facilities, can cost many billions of dollars. Finally, the integrated circuit vendors (e.g., INTEL) rely strongly on computational prototyping. They construct a computational prototype for every new chip. It is the only way they can design the layout and test it to put billions of components on a single chip successfully (Colwell, 2005).

Summary and Future

The CREATE Program is successfully developing and deploying a suite of physics-based computational engineering software tools with the design and analysis capabilities needed by the DoD Air Vehicle, Ship, RF, and Ground Vehicle acquisition engineering communities to reduce the cost, schedule, and risk of acquisition programs. The CREATE tools enable DoD engineers to generate and analyze virtual prototypes of DoD Air Vehicles, Ships, RF antennas and, in the future, Ground Vehicles, and to accurately predict the performance of the weapon systems. This approach to product development accelerates the development of innovative systems because it enables design engineers to rapidly develop, analyze, assess, optimize, and test many design options, without having to construct and test physical prototypes until late in the product development process. Design defects and performance shortfalls can be detected and fixed well before metal has been cut.

At the latest count, over 160 DoD acquisition engineering organizations (government, industry, and academia) are using the tools to design and assess over 70 DoD weapon systems. Acquisition community interest and customer use is growing exponentially (AF, Navy & Army engineers, Boeing, LMC, NG, Raytheon, Sikorsky, Bell, Pratt & Whitney, AFLCMC, AMRDEC, NAVAIR, NAVSEA, C-130/C-17 Cargo Release, F/A-18E, ARL, SPAWAR, Ball Aerospace, etc.). The CREATE tools are already enabling the design and analysis of many important DoD systems (e.g., CH-47 rotor-blade retrofit, Ohio replacement submarine, CVN-78 shock test, NAVAIR UAV flight certification, Air Force next-generation aircraft). The CREATE Program has made significant progress to successfully overcome major challenges to provide user support, resolve intellectual property issues, achieve successful deployment of software, and implement sound software engineering and software project management practices that lead to a high level of software quality, software that is usable, maintainable, verified and validated, extensible, scalable, and documented.

After 10 years of development and deployment, the CREATE Program is beginning to achieve the goal of revolutionizing the way the DoD procures major weapon platforms through the use of virtual prototypes. These tools will enable the DoD acquisition engineering community to develop innovative weapon systems by allowing DoD engineers to generate and evaluate thousands of design options, rather than the handful that was previously possible.

The CREATE tools are government-developed, government-owned, and government-supported so that the DoD can independently evaluate contractor deliverables. The tools are designed to be sufficiently robust and useable that experienced engineers with good judgment can utilize the tools with confidence. The tools are designed and built for a

~30-year plus life cycle. The tools are on the verge of being adopted by the defense industry for commercial use, so that they will contribute to a strong U.S. economy, as well as to a strong U.S. defense. The tools are also beginning to be used to improve the effectiveness and efficiency of DoD T&E enterprises.

By 2019, each CREATE tool will deliver the capability promised in its 12-year vision. That capability, however, offers a foundation to fill many other DoD capability gaps and is an important part of the HPCMP mission for continuous modernization of hardware (DSRCs), networks (DREN), and defense-specific software applications. The CREATE tools are foundational elements of the OSD Engineered Resilient Systems S&T Initiative (Goerger, Madni, & Eslinger, 2014) and the Air Force Digital Thread and Digital Twin Programs (Kraft, 2015).

The CREATE team is proposing enhancements and upgrades to the existing CREATE that would greatly increase their range of applicability and impact. These include upgrades to Kestrel to address hypersonic design tasks, an addition of the ability of RSDE and DaVinci to estimate life cycle cost, the ability to produce a full conceptual ship design including integrated compartment arrangements and hullform optimization, operational assessments, and other aspects of multi-hull surface ships and submarines. With additional funding it would be possible to start new CREATE projects to address DoD capability gaps such as prediction and design of space satellite performance, rocket propulsions systems, structural performance design, combat power and electrical systems layout, and electronic warfare systems. The CREATE development process has proved very successful, and points the way for the DoD to develop tools to generate virtual prototypes of many different DoD systems and to predict their performance with physics-based computational tools.

References

Anthony, S. (2013, July 30). Five ways to innovate faster. *Harvard Business Review*, 3.

Ball, D. (2009). *Boeing catches a lift with high performance computing*. Paper presented at the Council on Competitiveness, HPC Success Stories.

Carter, A. (2013). *Defense acquisition* (DoD Instruction 5000.02). Retrieved from http://www.acq.osd.mil/fo/docs/DSD%205000.02_Memo+Doc.pdf

Colinvaux, P. (1980). *The fates of nations: A biological theory of history*. New York, NY: Simon and Schuster.

Council on Competitiveness. (2010). *Council on Competitiveness: HPC success stories.* Retrieved from http://www.compete.org/about-us/initiatives/hpc/success%20stories

Desilver, D. (2014). Chart of the week: The ever-accelerating rate of technology adoption. Retrieved from http://www.pewresearch.org/fact-tank/2014/03/14/chart-of-the-week-the-ever-accelerating-rate-of-technology-adoption/

Colwell, R. P. (2005) *The Pentium Chronicles: The people, passion, and politics behind Intel's landmark chips.* Wiley-IEEE Computer Society.

Engardio, P. (2008, September 20). Chawla, S.K. Los Alamos and Sandia: R&D treasures by S.K. Chawla. *Businessweek Magazine.*

Fox, J. (2014, December 5). What still makes Silicon Valley so special. *Harvard Business Review.*

Francis, S. (1995). *Warhead politics, Livermore and the competitive system of nuclear weapon design* (UCRL-LR-124754).

Goerger, S. R., Madni, A. M., & Eslinger, O. J. (2014). *Engineered resilient systems: A DoD perspective*. Paper presented at the 2014 Conference on Systems Engineering Research.

History of the iPhone. (n.d.). In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/History_of_iPhone)

Keagan, R. J. (2010). *The road before us: 2009 Goodyear Annual Report.*

Kennedy, P. (1987). *The rise and fall of the great powers: Economic change and military conflict from 1500 to 2000*. New York, NY: Random House.

Kossiakoff, A., & Sweet, W. (2003). *Systems engineering principles and practice*. Hoboken, NJ: John Wiley and Sons.

Kraft, E. M. (2010, September). After 40 years why hasn't the computer replaced the wind tunnel. The *ITEA Journal of Test and Evaluation, 31*, 329–346.

Kraft, E. M. (2015). *HPCMP CREATE™-AV and the Air Force digital thread.* Paper presented at the Proceedings of the 53rd AIAA Aerospace Sciences Meeting, AIAA 2015-0042, Kissimmee, FL.

Kramer, R. J. (2011). *Goodyear Tire & Rubber Q4 2010 earnings call*. Retrieved from http://www.nasdaq.com/aspx/call-transcript.aspx?StoryID=252155&Title=goodyear-tire-rubber-s-ceo-discusses-q4-2010-results-earnings-call-transcript

Martins, H., Dias, Y. B., & Khanna, S. (2016, April 26). What makes some Silicon Valley companies so successful. *Harvard Business Review*.

Miller, L. K. (2010). Simulation-based engineering for industrial competitive advantage. *Computing in Science and Engineering, 12*(3), 14–21.

Post, D., Atwood, C., Newmeyer, K., Meakin, R., Hurwitz, M., et al. (2016). CREATE: Software engineering applications for the design and analysis of air vehicles, naval vessels, and radio frequency antennas. *Computing in Science and Engineering, 11*, 14–24.

Reinertsen, D. (2009). *The principles of product development flow*. Redondo Beach, CA: Celeritas.

Singer, D., Doerry, N., & Buckley, M. (2009). What is set based design? *Naval Engineers Journal, 121*(4), 31–43.

Walser, A. F. (2014). *Drive safe, green and smart: HPC applications for sustainable mobility.* Paper presented at the When Science Meets Industry PRACE Scientific and Industrial Conference. Retrieved from http://www.prace-ri.eu/IMG/pdf/pracedays14_walser.pdf

Weiss, L (2014). *America, Inc.: Innovation and enterprise in the national security state*. Ithaca, NY: Cornell University Press.

Acknowledgements

The success of the CREATE Program is due to the dedicated work of the ~180 members of the 11 CREATE development teams. The authors are grateful for the support of the DoD HPCMP; the Army Corps of Engineers Engineer Research and Development Center; the Carnegie Mellon University Software Engineering Institute; the Office of the Secretary of Defense; the Army, Navy, and Air Force acquisition engineering communities; the chairs of the CREATE Boards of Directors: Dr. Edward Kraft, Dr. Robert Clemens, and Ms. Robin White; and the members of the Boards, Mr. Robert Rassa and the National Defense Industrial Association; and the Sierra Mechanics Group at the Sandia National Laboratory. This material is based upon work supported by, or in part by, the Department of Defense High Performance Computing Modernization Program. Any opinions, findings, and

conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the DoD High Performance Computing Modernization Program or the U.S. Government.

Panel 7. Evolving Cost Estimating Methodology

Wednesday, April 26, 2017	
1:45 p.m. – 3:15 p.m.	**Chair: Nancy Spruill,** Director, Acquisition Resources & Analysis, OUSD (AT&L) ***Empirical Cost Estimation Tool (PMS 320)*** Johnathan C. Mun, Naval Postgraduate School Thomas J. Housel, Naval Postgraduate School ***Estimating the Estimate: Toward a Quick and Inexpensive Method for Weapons System Cost Estimation*** Charles K. Pickar, Naval Postgraduate School Kevin G. Feely, KCR Business Innovations ***Determining the Value of a Prototype*** Zachary S. McGregor-Dorsey, Institute for Defense Analyses (CARD)

Nancy Spruill—is the Director for the Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics, Acquisition Resources and Analysis (ARA). Dr. Spruill integrates the diverse aspects of defense acquisition into a balanced and coherent program that supports the national strategy and makes the most effective use of resources provided. The Director, ARA also serves as the Executive Secretary to the Defense Acquisition Board; oversees the Defense Acquisition Executive System (DAES); manages AT&L's participation in the Planning, Programming, Budgeting, and Execution System (PPBE); and is responsible for the timely and accurate submission to Congress of Selected Acquisition Reports and Unit Cost Reports for Major Defense Acquisition Programs.

Empirical Cost Estimation Tool (PMS 320)

Johnathan C. Mun—is a research professor at NPS and teaches executive seminars in quantitative risk analysis, decision sciences, real options, simulation, portfolio optimization, and other related concepts. He received his PhD in finance and economics from Lehigh University. He is considered a leading world expert on risk analysis and real options analysis. Dr. Mun has authored 12 books and is the founder and CEO of Real Options Valuation Inc. [jcmun@realoptionsvaluation.com]

Thomas J. Housel—specializes in valuing intellectual capital, knowledge management, telecommunications, information technology, value-based business process reengineering, and knowledge value measurement in profit and non-profit organizations. He is a tenured full professor for the Information Sciences (Systems) Department at NPS. He has conducted over 80 knowledge value added (KVA) projects within the non-profit, Department of Defense (DoD) sector for the Army, Navy, and Marines. Dr. Housel also completed over 100 KVA projects in the private sector. The results of these projects provided substantial performance improvement strategies and tactics for core processes throughout DoD organizations and private sector companies. [tjhousel@nps.edu]

Introduction

This research project pertains to the identification, review, and potential development of existing and alternative ship cost modeling methodologies. Most ship cost modeling has been traditionally weight-based. This approach drives the U.S. Navy decision makers to acquire smaller ships that require custom-designed shipboard components.

Current, and future, Department of Defense (DoD) acquisition budgeting processes require identifying, modeling, and estimating the costs of shipbuilding. The purpose of this research project is to determine if there is a more accurate way to empirically predict and model ship acquisition costs. The cost modeling tool developed in this study is intended to support development of ship cost forecasts. The proof of concept example for using this cost modeling tool included herein will provide a roadmap for other new ship acquisition cost modeling. The outcome of this research will likely increase cost savings.

The focus of this research is a comprehensive review of the most promising cost modeling methodologies. Notional cost data, or rough order of magnitude values, will be collected or generated to support this review of the cost methodologies. These data will be generated by the researchers using archival cost data from ship maintenance projects of various destroyer (i.e., DDG) acquisitions. We will identify these extrapolations, and we will use the resulting notional data to help evaluate the efficacy of the various cost models. This approach allows readers and study sponsors to see the various types of cost models, approaches, and sample data variables that are required to run the cost models and to examine sample results, as well as review the pros and cons of each approach. This study may require a follow-on project if there is a method that is of interest or that the sponsors feel might be applicable for a given ship acquisition context. The required data variables as well as sample results will be listed in the report, so the sponsors will know what to expect prior to engaging in any new research project. A follow-on study would allow us to obtain real-life cost data that could be plugged into the desired cost model.

The selected cost model will likely include the standard parametric models, nonparametric methods, systems dynamics based on project management task-based schedule and cost models; semiparametric Monte Carlo simulation models; curve fitting, time-series, and cross sectional models; nonlinear models, and so forth that have proven useful in forecasting costs in other acquisition contexts.

The Theory of Predictive Modeling in Cost

Different Types of Forecasting Techniques

The review of standard forecasting logic, in what follows, is useful in understanding the foundations of the various cost modeling techniques assessed in this study. Generally, forecasting can be divided into quantitative and qualitative approaches. Qualitative forecasting is used when little to no reliable historical, contemporaneous, or comparable data exist. Several qualitative methods exist such as the Delphi or expert opinion approach (a consensus-building forecast by field experts, marketing experts, or internal staff members), management assumptions (target growth rates set by senior management), as well as market research or external data or polling and surveys (data obtained through third-party sources, industry and sector indexes, or active market research). These estimates can be either single-point estimates (an average consensus) or a set of prediction values (a distribution of predictions). The latter can be entered into the Risk Simulator software tool, used in this study, as a custom distribution and the resulting predictions can be simulated; that is, running a nonparametric simulation using the prediction data points as the custom distribution. This approach can leverage experts' knowledge by combining it with available quantitative data to arrive at more reliable ship building cost estimates.

Expert knowledge can be leveraged using the software by including qualitative estimates with quantitative analysis techniques. We provide several ship cost modeling case examples that are designed to demonstrate how the various cost modeling tools can be used in estimating ship building costs. That will also be helpful in learning how to apply the Risk Simulator software to develop more robust ship building cost estimates. The appendix provides a quick review of the quantitative methodologies that are available in the software.

Case Application: DDG 51 FLT III

This section provides a detailed illustration of the proposed integrated cost estimation modeling approach. As this is only an illustration, and due to a lack of proprietary data for this first phase of the analysis, the input assumptions are only high-level approximations based on publicly available information and publicly available subject matter expert estimates. Therefore, the results generated are not designed to be used in any specific decision making. Nonetheless, the approach presented has proven to be robust and valid, and with the correct input assumptions, can be rerun to generate accurate and reliable ship cost estimates. Information and data were obtained via publicly available sources and were collected, collated, and used in an integrated risk-based cost and schedule modeling methodology. The objective of this case study was to develop a comprehensive cost modeling strategy and approach, and as such, notional data were used. Specifically, we used the Arleigh Burke Class Guided Missile Destroyer DDG 51 Flight I, Flight II, Flight IIA, and Flight III (Figure 1) as a basis for the cost and schedule assumptions, but the modeling approach is extensible to all other ship building cost contexts within the U.S. Navy.

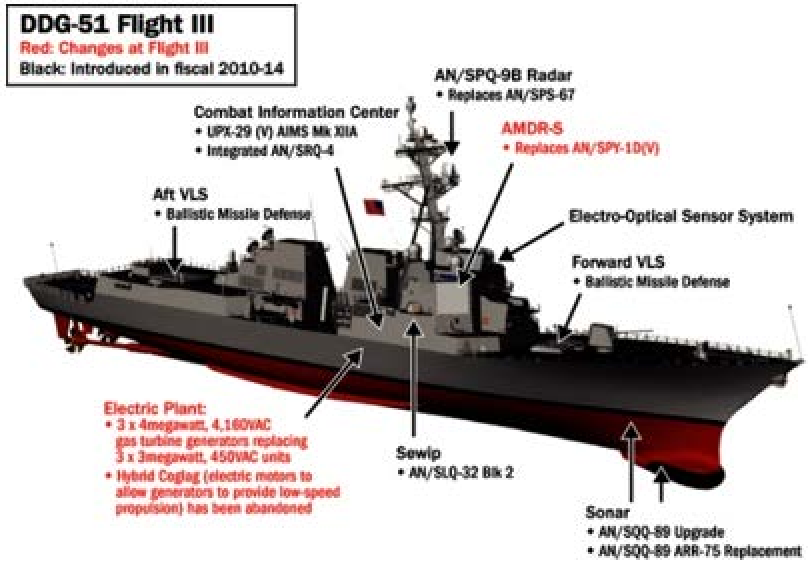
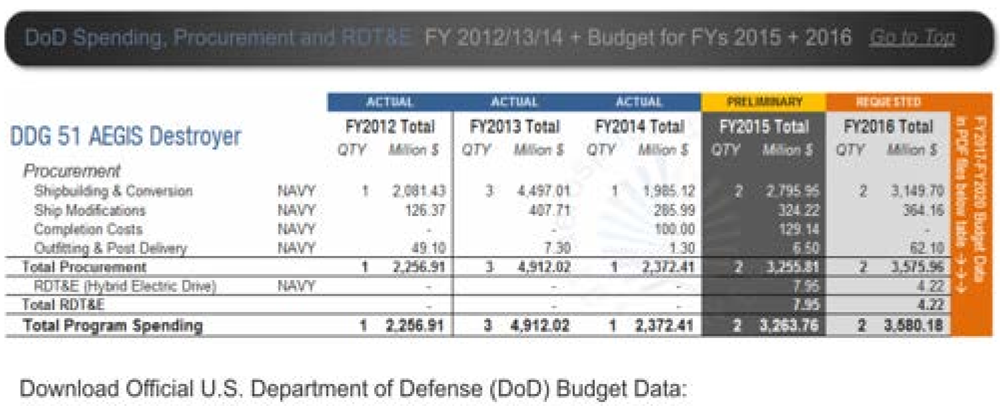


Figure 1. Overview of DDG 51 Flight III

DoD Spending on the Aegis Destroyer in FY 2012–2014

Figure 2 shows some sample acquisition budgets for DDG 51 Aegis destroyers from FY 2012 through FY 2016. The comprehensive DoD budget was downloaded and analyzed in the current research.



DoD Spending, Procurement and RDT&E FY 2012/13/14 + Budget for FYs 2015 + 2016 Go to Top

DDG 51 AEGIS Destroyer		ACTUAL FY2012 Total QTY	Million $	ACTUAL FY2013 Total QTY	Million $	ACTUAL FY2014 Total QTY	Million $	PRELIMINARY FY2015 Total QTY	Million $	REQUESTED FY2016 Total QTY	Million $
Procurement											
Shipbuilding & Conversion	NAVY	1	2,081.43	3	4,497.01	1	1,985.12	2	2,795.96	2	3,149.70
Ship Modifications	NAVY		126.37		407.71		285.99		324.22		364.16
Completion Costs	NAVY		-		-		100.00		129.14		-
Outfitting & Post Delivery	NAVY		49.10		7.30		1.30		6.50		62.10
Total Procurement		1	2,256.91	3	4,912.02	1	2,372.41	2	3,255.81	2	3,575.96
RDT&E (Hybrid Electric Drive)	NAVY		-		-		-		7.95		4.22
Total RDT&E			-		-		-		7.95		4.22
Total Program Spending		**1**	**2,256.91**	**3**	**4,912.02**	**1**	**2,372.41**	**2**	**3,263.76**	**2**	**3,580.18**

FY2017-FY2020 Budget Data in PDF files below table → → →

Download Official U.S. Department of Defense (DoD) Budget Data:

Figure 2. DoD Spending and Procurement for FY 2012–2014

High-Level Shipbuilding Process

Figure 3 shows the high-level process flow of building ship hulls and sections.

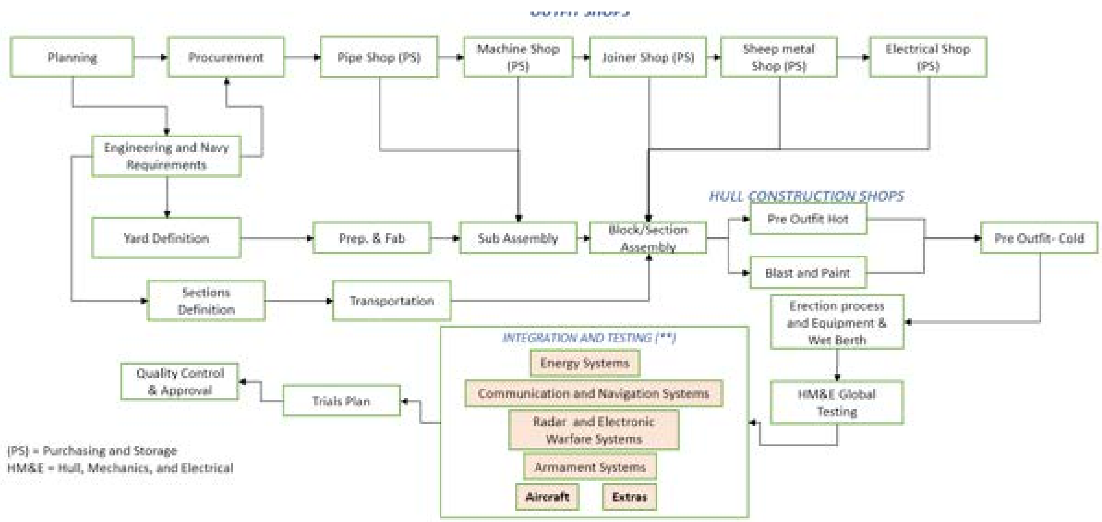


Figure 3. High-Level Process Flow (Hull and Sections)

Information, Communication, and Technology Subprocess

Figure 4 shows the ship's subprocess for information, communication, and technology (ICT).

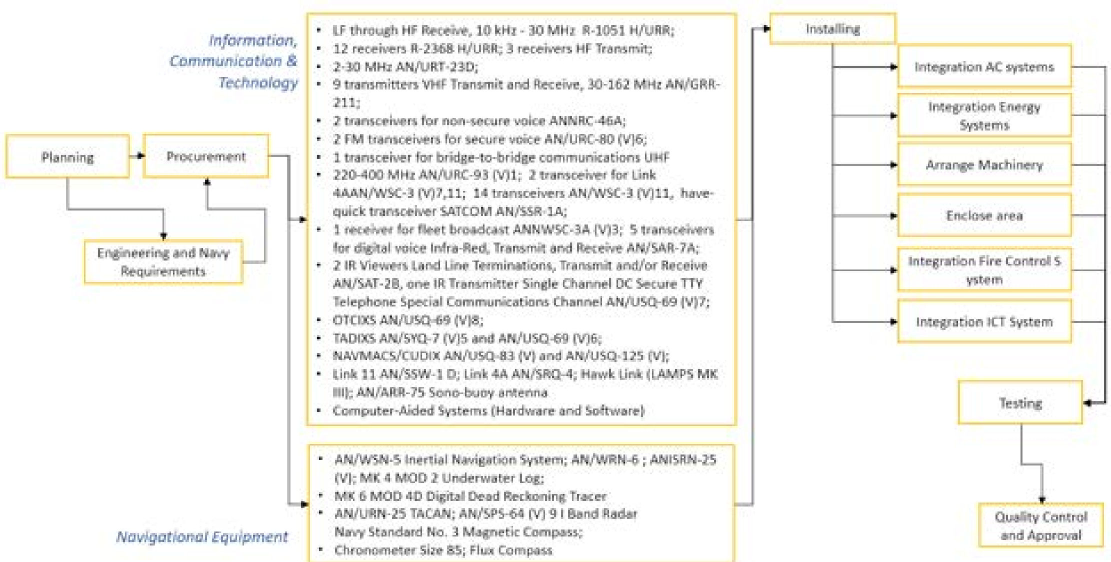


Figure 4. Subprocess for Information, Communication, and Technology (ICT)

Weapons System Subprocess

Figure 5 shows the ship’s subprocess for weapons systems.

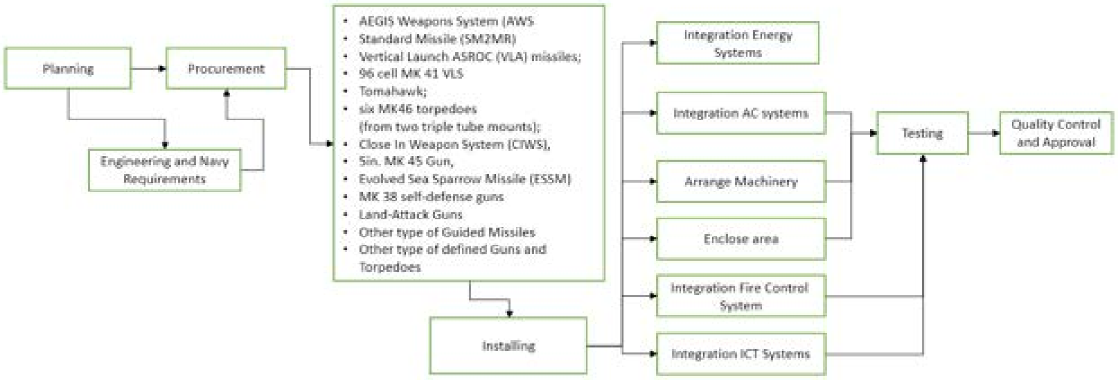
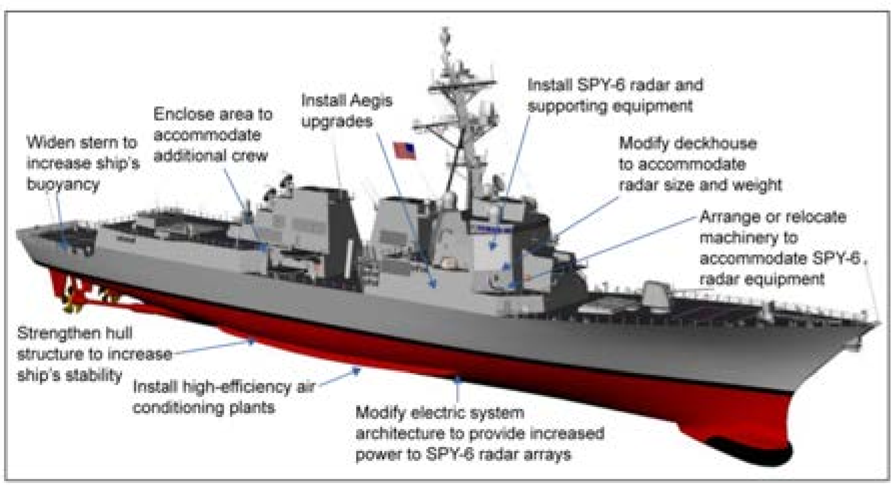


Figure 5. Subprocess for Weapons Systems

SPY-6 Radar System

Figure 6 shows the ship’s radar subsystem’s process.



Source: GAO (analysis); Navy (image and data). | GAO-16-613

ARLEIGH BURKE DESTROYERS:
Delaying Procurement of DDG 51 Flight III Ships Would Allow Time to Increase Design Knowledge
GAO-16-613: Published: Aug 4, 2016. Publicly Released: Aug 4, 2016.

What GAO Found

The Air and Missile Defense Radar (AMDR) program's SPY-6 radar is progressing largely as planned, but extensive development and testing remains. Testing of the integrated SPY-6 and full baseline Aegis combat system upgrade—beginning in late 2020—will be crucial for demonstrating readiness to deliver improved air and missile defense capabilities to the first DDG 51 Flight III ship in 2023. After a lengthy debate between the Navy and the Department of Defense's (DOD) Director of Operational Test and Evaluation, the Secretary of Defense directed the Navy to fund unmanned self-defense test ship upgrades for Flight III operational testing, but work remains to finalize a test strategy.

Figure 6. SPY-6 Radar System and Rework

DoD Extras: Electronic Warfare, Decoys, Extra Capabilities

Figure 7 shows the ship's Electronic Warfare, Decoys, and Extra Capabilities subprocesses.

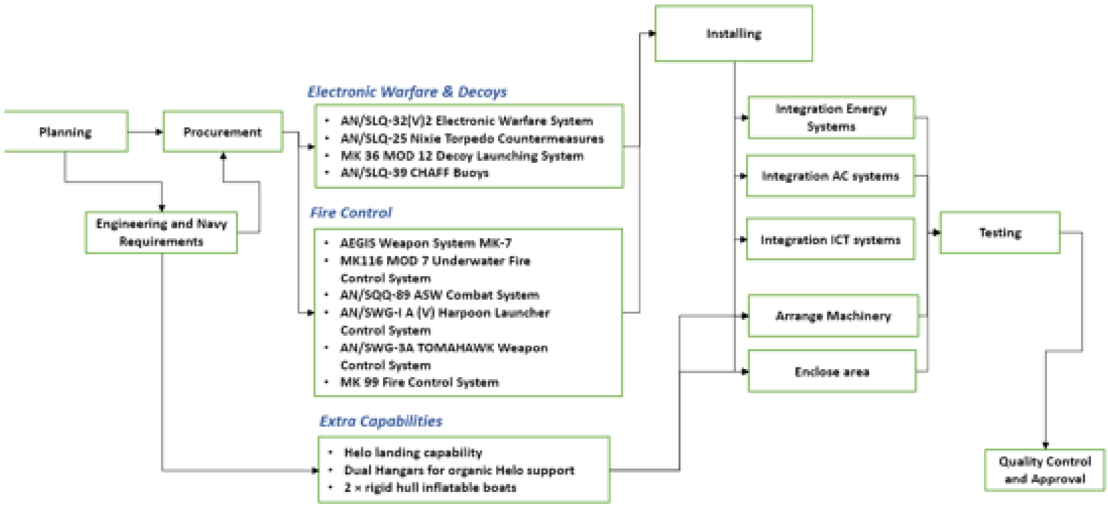


Figure 7. Subprocesses and Examples of DoD Extras

Risk-Based Schedule and Cost Process Modeling

Figures 8 illustrates how the project management tasks are incorporated into the Project Economics Analysis Toolkit (PEAT) software application. It includes all the high-level tasks required to build the ship along with their attendant costs with one million simulation trials that provide the possible distributions of the costing data. The parallel development of tasks 20–25 is where the ship's various subsystems are incorporated into the cost and schedule model.

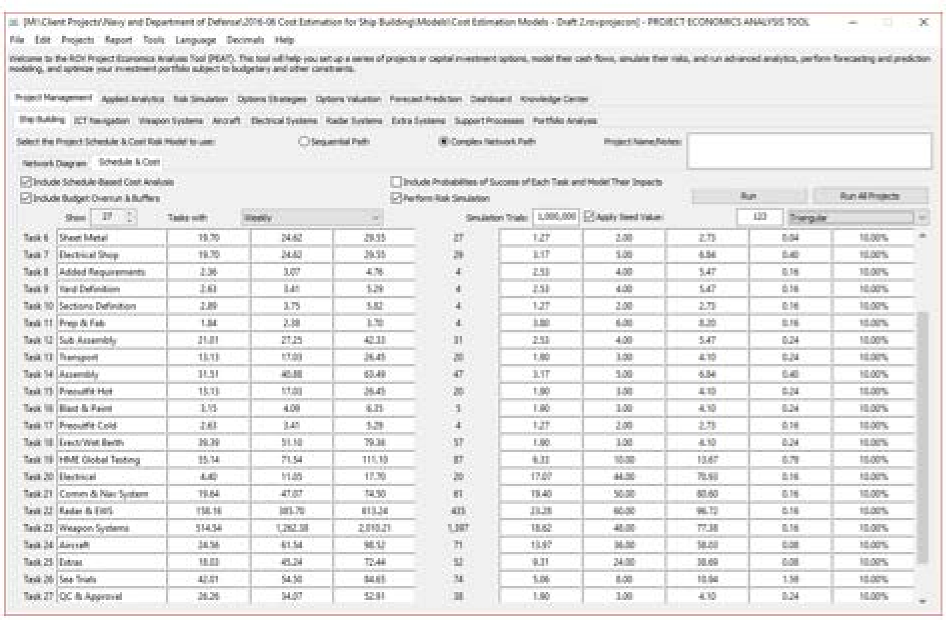


Figure 8. Input Assumptions

Similarly, using the cost and schedule modeling approach, we can zoom into various tasks and model each task in more detail. This permits us to use the results by reinserting the more detailed data values back into the more comprehensive model as required to improve accuracy. For instance, Figure 9 shows the ship's weapons subsystem, with Figure 10 showing its cost and schedule assumptions. This model's result can be inserted back into Task 23 in the comprehensive model (Figure 5).

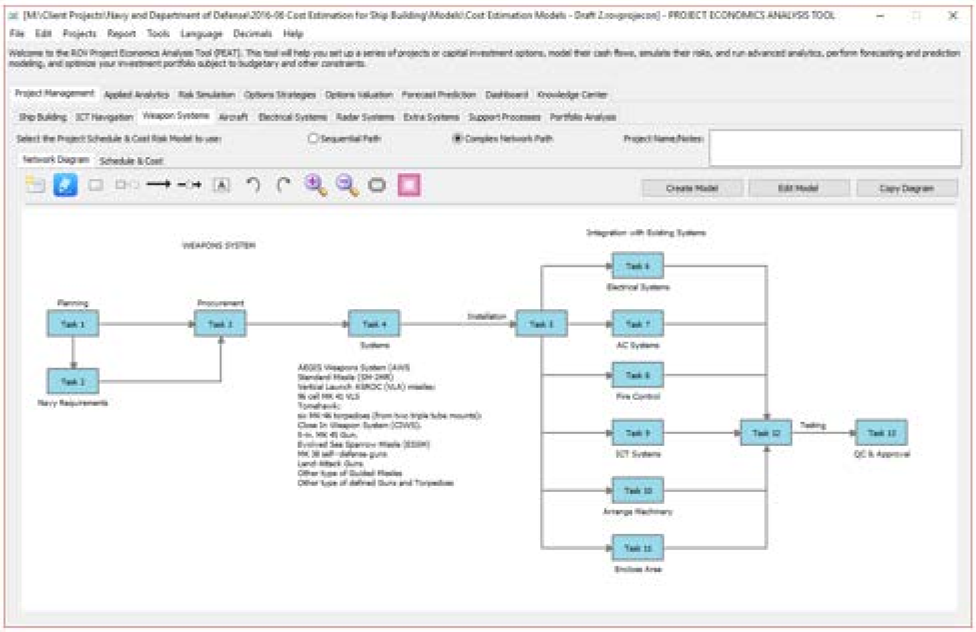
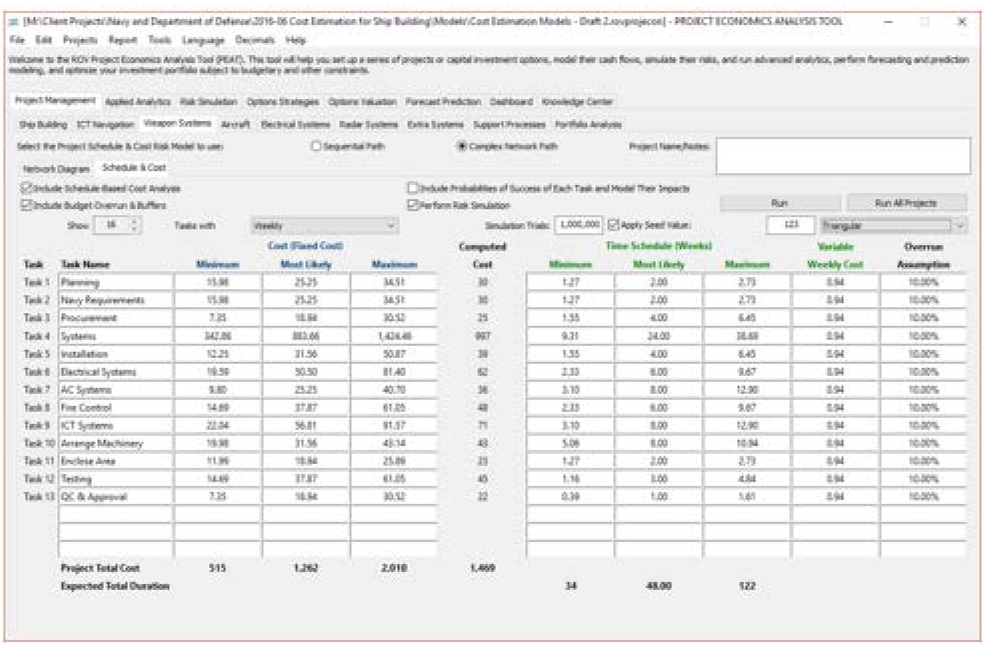


Figure 9. Weapons Subsystem Process Development



Task	Task Name	Cost (Fixed Cost) Minimum	Most Likely	Maximum	Computed Cost	Time Schedule (Weeks) Minimum	Most Likely	Maximum	Variable Weekly Cost	Overrun Assumption
Task 1	Planning	15.98	25.25	34.51	30	1.27	2.00	2.73	0.94	10.00%
Task 2	Navy Requirements	15.98	25.25	34.51	30	1.27	2.00	2.73	0.94	10.00%
Task 3	Procurement	7.35	18.84	30.52	25	1.55	4.00	6.45	0.94	10.00%
Task 4	Systems	342.86	883.66	1,424.46	997	9.31	24.00	38.69	0.94	10.00%
Task 5	Installation	12.25	31.56	50.87	39	1.55	4.00	6.45	0.94	10.00%
Task 6	Electrical Systems	19.59	50.50	81.40	62	2.33	6.00	9.67	0.94	10.00%
Task 7	AC Systems	9.80	25.25	40.70	36	3.10	8.00	12.90	0.94	10.00%
Task 8	Fire Control	14.69	37.87	61.05	48	2.33	6.00	9.67	0.94	10.00%
Task 9	ICT Systems	22.04	56.81	91.57	71	3.10	8.00	12.90	0.94	10.00%
Task 10	Arrange Machinery	19.98	31.56	43.14	43	5.06	8.00	10.94	0.94	10.00%
Task 11	Enclose Area	11.99	18.94	25.89	23	1.27	2.00	2.73	0.94	10.00%
Task 12	Testing	14.69	37.87	61.05	45	1.16	3.00	4.84	0.94	10.00%
Task 13	QC & Approval	7.35	18.94	30.52	22	0.39	1.00	1.61	0.94	10.00%
	Project Total Cost	515	1,262	2,010	1,469					
	Expected Total Duration					34	48.00	122		

Figure 10. Weapons Subsystem Cost and Schedule Assumptions

Critical Success Factors in Cost and Schedule Estimates

Tornado analysis is a powerful analytical tool that captures the model's sensitivity to fluctuations in the critical success factors values for cost and schedule. This is done by identifying the static impacts of each variable on the outcome of the model; that is, the tool automatically perturbs each variable in the model a preset amount, captures the fluctuation on the model's forecast or final result, and lists the resulting perturbations ranked from the most significant to the least. Figures 11 and 12 illustrate the application of a tornado analysis. Tornado analysis answers the question: "What are the critical success drivers that affect the model's output the most?"

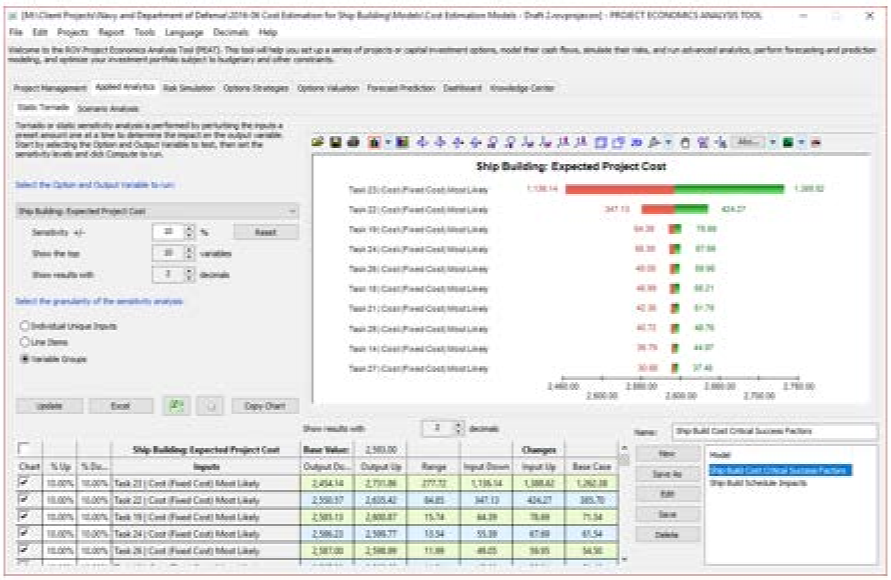


Figure 11. Tornado Analysis of Critical Success Factors (Cost Factors)

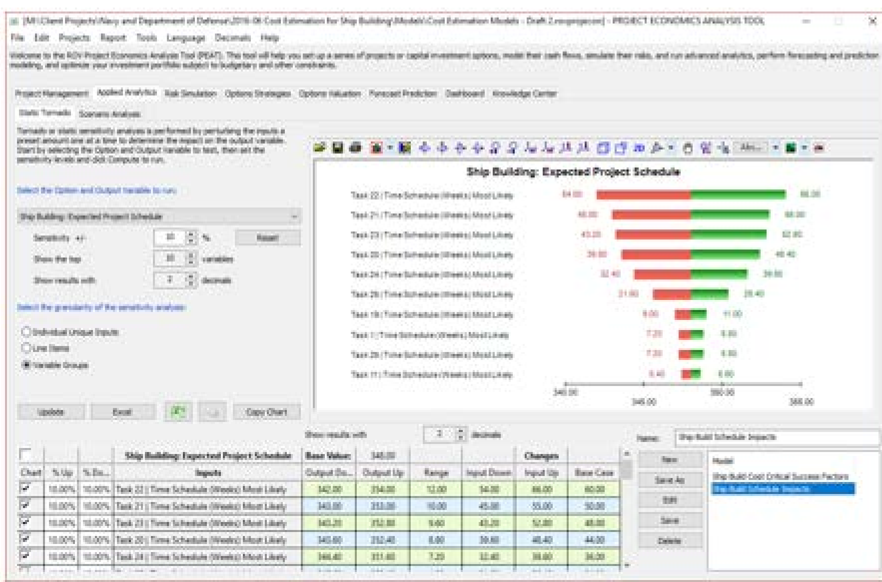


Figure 12. Tornado Analysis of Critical Success Factors (Schedule Factors)

Risk-Based Schedule and Cost Process Simulation

Next, the Monte Carlo Risk Simulation capability of the tool was used to create artificial futures by generating hundreds of thousands of sample paths of outcomes and analyzing their prevalent characteristics. In the Monte Carlo simulation process, triangular distributions (i.e., best-case, most-likely case, and worst-case scenarios) were used on the previously identified critical inputs. Figure 13 shows the values for a sample distributional spread used in Monte Carlo Risk Simulations per the *Air Force Cost Analysis Handbook* (AFCAH). These probability spreads were applied to each of the task's cost and schedule inputs, and each of the tasks was simulated tens of thousands to hundreds of thousands of trials.

Figure 14 shows a sample representation of the results from the simulation process. For instance, the figure shows a 90% confidence interval for the total acquisition cost of a full-complement ship (fully built ship delivered after tests and sea trials, complete with ICTS, weapons systems, electrical systems, SPY-6 radar, and other add-ons). The 90% confidence interval pegs the total acquisition costs to be between $2.0 billion and $3.2 billion for a single ship. ***Clearly, these results are only for illustration purposes and are not meant to be definitive.*** Figure 15 shows the probability that there will be a budget overrun. For instance, if the acquisition budget is $2.2 billion, then we see that there is an approximately 12% probability of the cost coming in at or under budget, which means that there is an 88% probability of a budget overrun, with a mean or average actual acquisition cost of $2.6 billion.

Similarly, Figure 16 shows the total schedule from the initial contracting phase to delivery of the ship, complete with all subsystems installed and tested. The 90% confidence interval pegs the total schedule at between 110 and 146 weeks, averaging at 127 weeks.

Alternatively, the modeling approach allows us to look at the ship's subsystems. For example, Figure 17 shows the 90% confidence interval for weapons systems costs ($1.1 to $1.8 billion), while Figure 18 shows modeling the cost of building the ship without any subsystems. Each individual system or combinations of systems can be similarly modeled and analyzed (Figure 19), or overlaid on one another, as shown in Figures 20, 21, and 22. The probability distributions in these three figures allow you to compare how one system's cost and uncertainties compare to one another. Finally, Figure 23 shows how the individual task's schedule and cost elements impact and are correlated to each other, by way of dynamic sensitivity analysis.

U.S. Air Force Cost Analysis Handbook (AFCAH)

Distribution	PEI	Probability	15%	Mode	85%	Fitted Distributions Min	Fitted Distributions Likely	Fitted Distributions Max
Triangular Low Left	Mode	1.0 (75%)	0.695	0.878	1.041	0.482	0.878	1.247
Triangular Low	Mode	1.0 (50%)	0.834	1	1.166	0.633	1.000	1.367
Triangular Low Right	Mode	1.0 (25%)	0.959	1.122	1.305	0.753	1.122	1.518
Triangular Medium Left	Mode	1.0 (75%)	0.492	0.796	1.069	0.137	0.796	1.412
Triangular Medium	Mode	1.0 (50%)	0.723	1	1.277	0.388	1.000	1.612
Triangular Medium Right	Mode	1.0 (25%)	0.931	1.204	1.508	0.588	1.204	1.863
Triangular High Left	Mode	1.0 (75%)	0.347	0.754	1.103	0.000	0.754	1.550
Triangular High	Mode	1.0 (50%)	0.612	1	1.388	0.142	1.000	1.858
Triangular High Right	Mode	1.0 (25%)	0.903	1.236	1.711	0.442	1.236	2.225
Triangular EHigh Left	Mode	1.0 (75%)	0.3	0.745	1.15	0.000	0.745	1.657
Triangular EHigh	Mode	1.0 (50%)	0.509	1.004	1.5	0.000	1.004	2.100
Triangular EHigh Right	Mode	1.0 (25%)	0.876	1.367	1.914	0.258	1.367	2.553

Figure 13. Sample Distributional Spread per the U.S. Air Force Cost Analysis Handbook

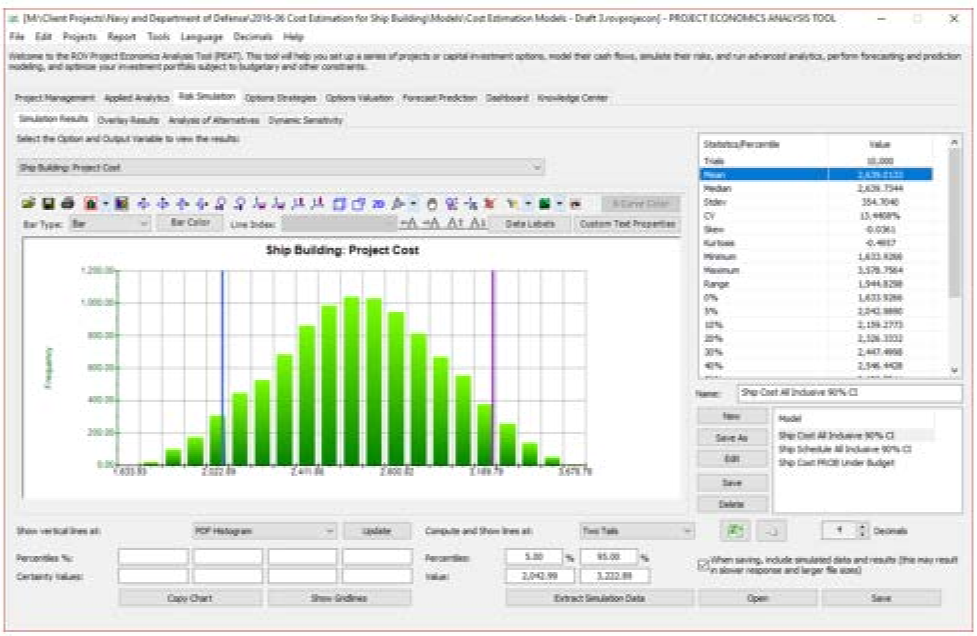


Figure 14. Simulation Results on Shipbuilding Cost (90% Confidence Interval)

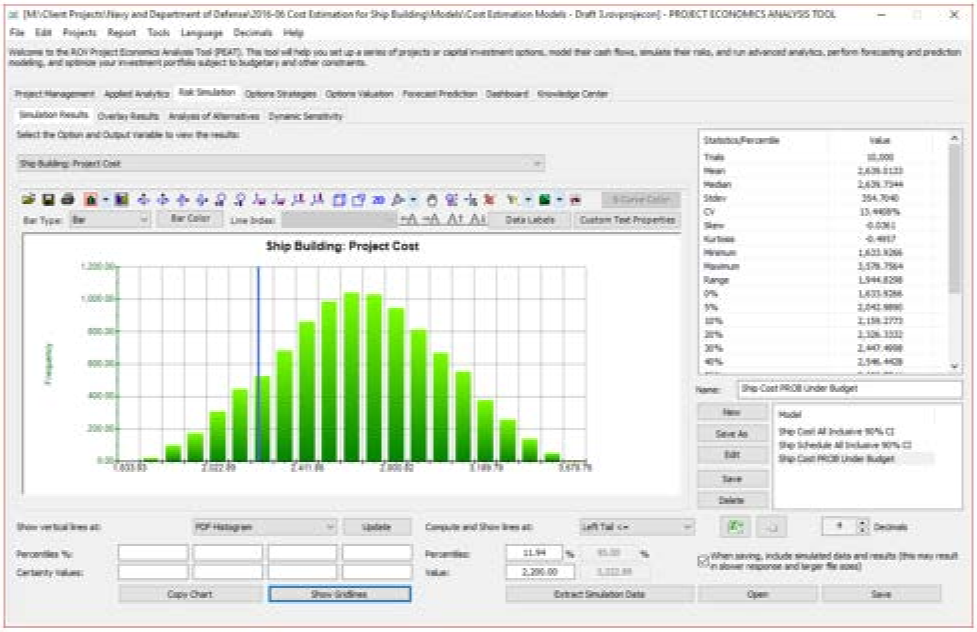


Figure 15. Probability of Cost Exceeding Budget

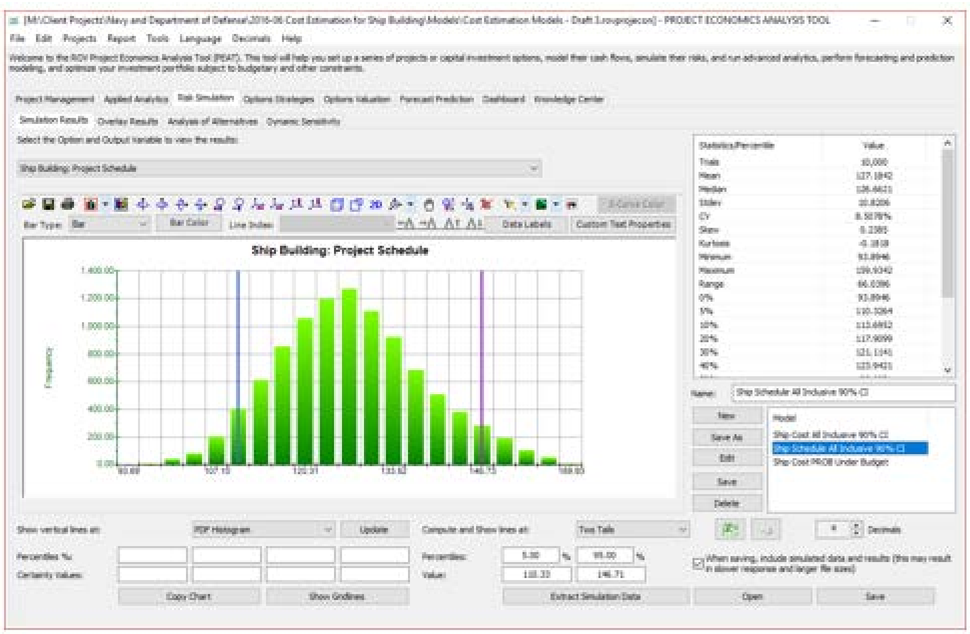


Figure 16. Schedule Risk (90% Confidence Interval)

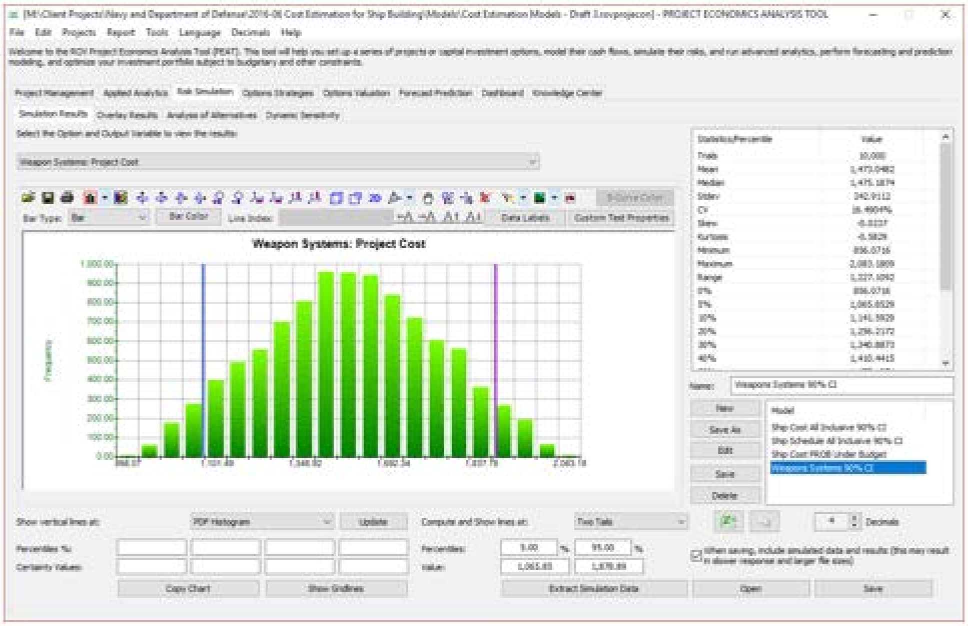


Figure 17. Cost of Weapons Systems (90% Confidence Interval)

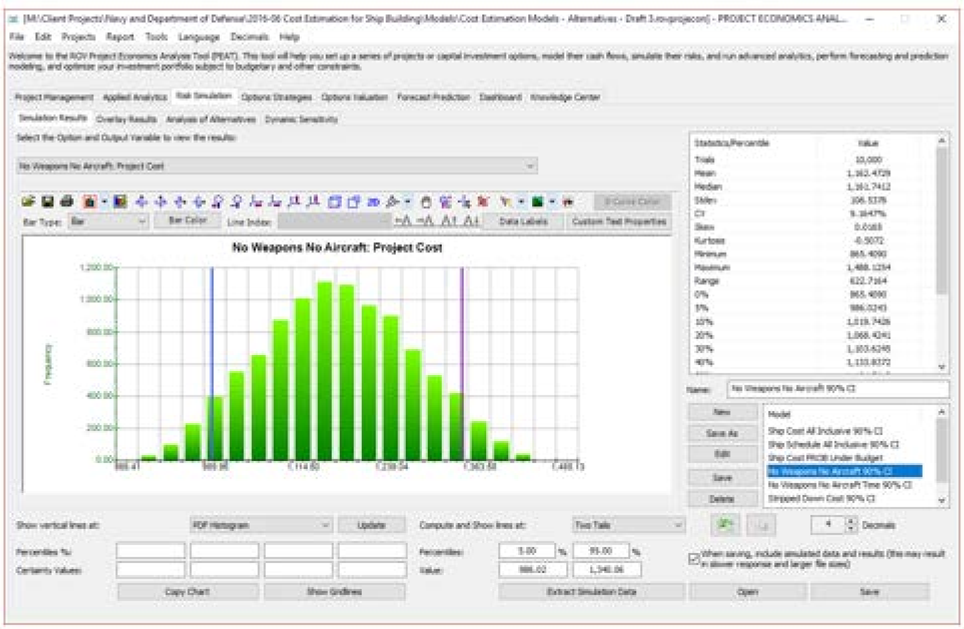


Figure 18. Simulated Cost of No Weapons and No Aircraft

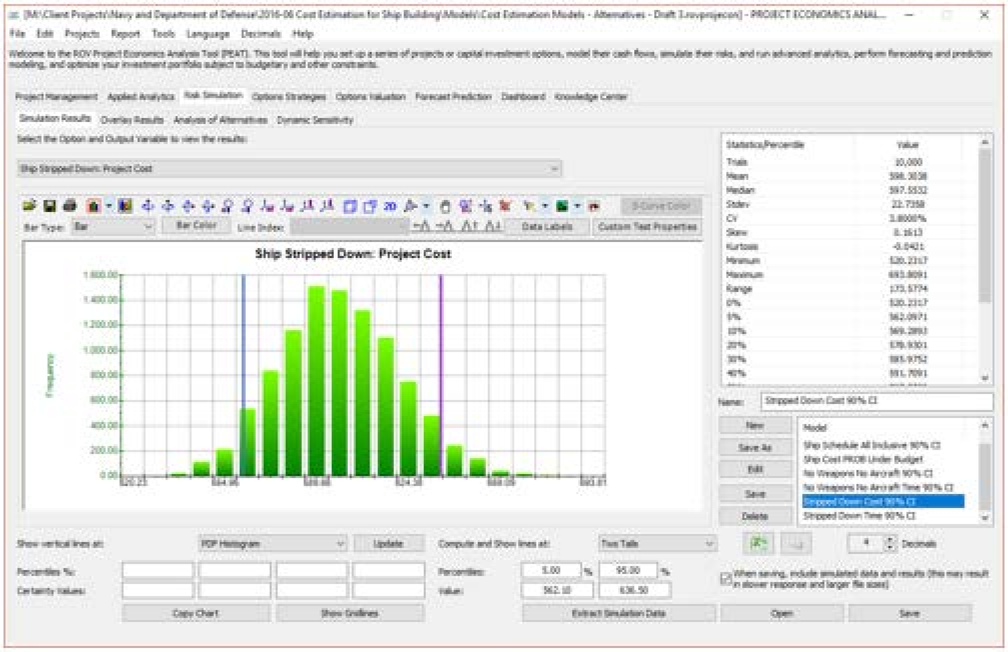


Figure 19. Simulated Cost of Stripped-Down Ship Build

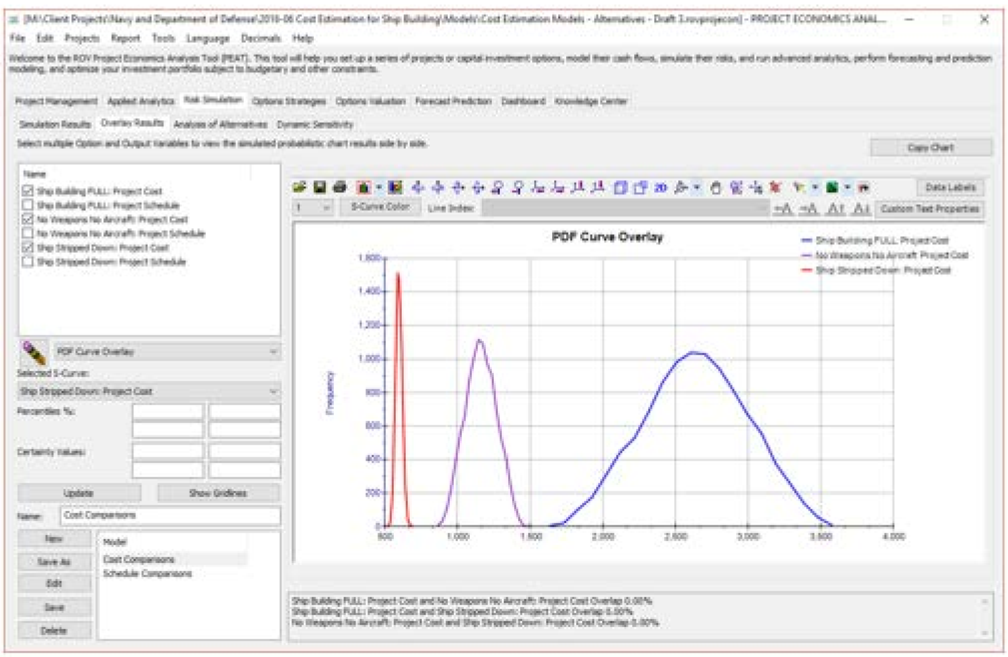


Figure 20. Comparative Analysis of Ship Configurations

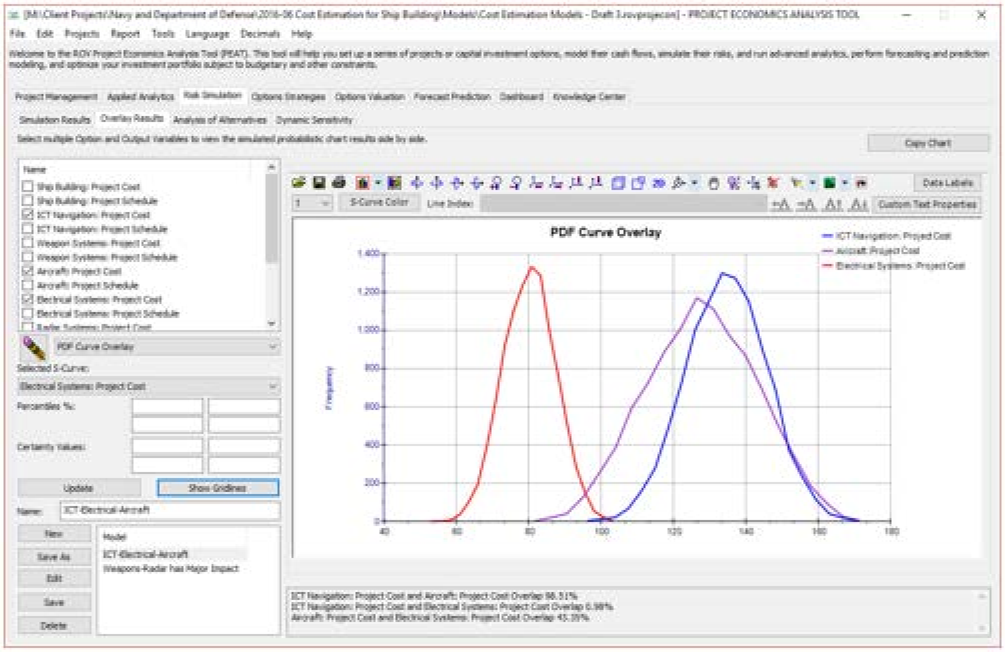


Figure 21. Overlay of Simulated Probability Distributions (Subsystems)

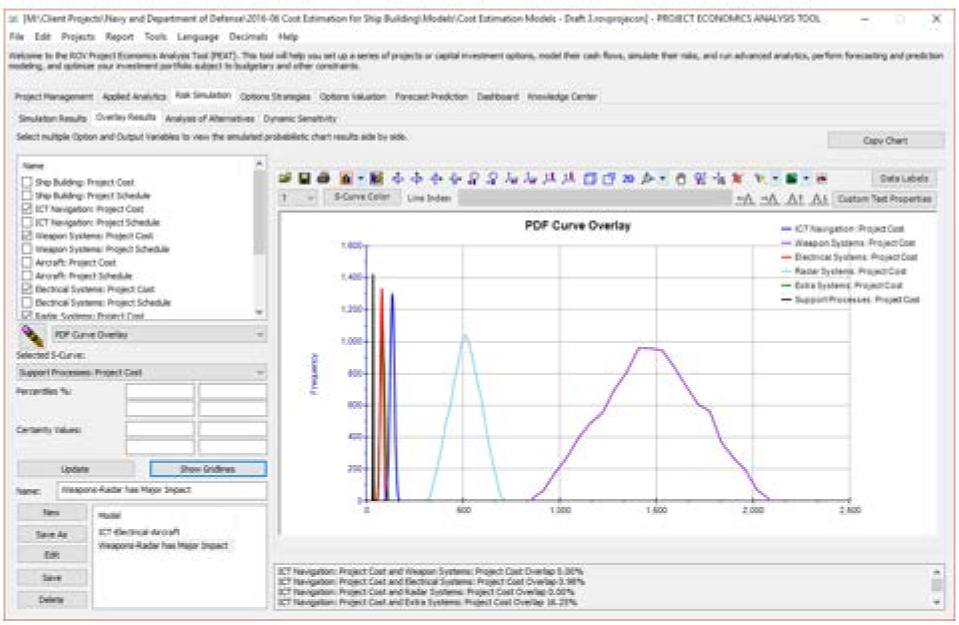


Figure 22. Overlay of Simulated Probability Distributions (All Subsystems)

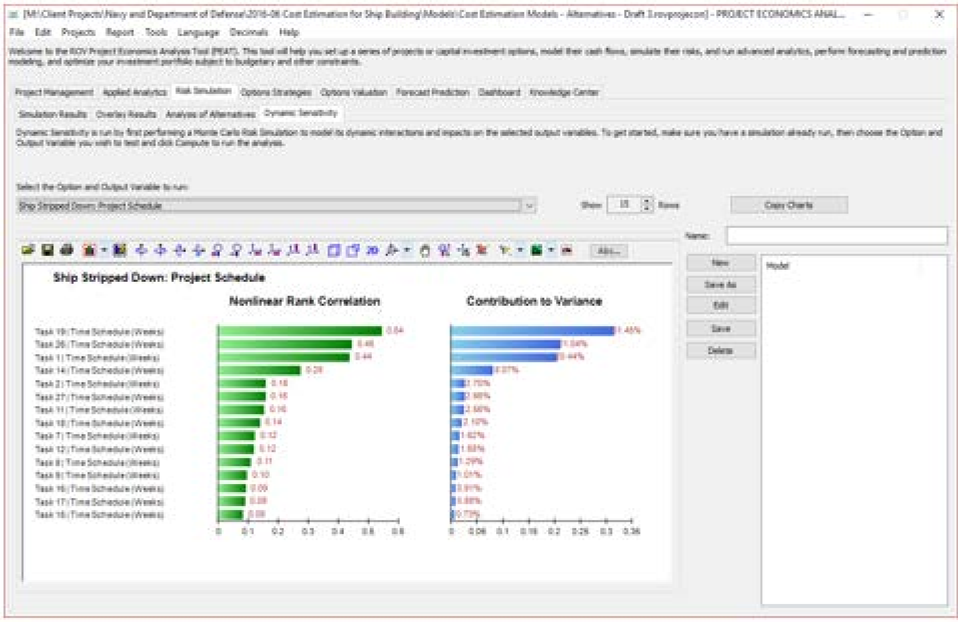


Figure 23. Dynamic Sensitivities of Stripped-Down Ship Build

Parametric Cost Models With Historical Data

A complementary approach to generate additional input cost assumptions includes the use of parametric modeling. To run parametric models, historical data is first required. Figure 24 shows an example dataset obtained via various defense agencies' publicly available information. The dataset shows various ship types, the unit costs (in millions), displacement in tons, speed, length, crew size, and year the ships were delivered.

Parametric models were developed and tested using simple multiple regression analysis, nonlinear regression, and econometric models. For instance, the following shows a simple linear parametric regression model and its results, where the functional form tested was

$$y = \beta_0 + \beta_1 \mathrm{X}_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

$$Cost = -11837 - 0.10\, Tons + 80.44\, Speed + 55.56\, Length + 6.09\, Crew$$

Although the model looks good, with statistically significant *p*-values (e.g., 0.0097) that are lower than the standard 0.05 or 0.10 significance cutoffs and coefficients of determination (R-squared) that are relatively high at 82.60%, the model is flawed. For instance, the coefficient for displacement is negative, which defies conventional logic, where typically the heavier the ship, the higher the cost. This means the model's specification is incorrect and another model is required. Figure 25 shows a mixed nonlinear parametric model with the following specification:

$$y = \beta_0 + \beta_1 \ln(X_1) + \beta_2 \ln(X_2) + \beta_3 X_3 + \beta_4 X_4$$

$$Cost = -40271 + 3351 \ln(Tons) + 3952 \ln(Speed) - 26.37\, Length - 2.18\, Crew$$

This model makes slightly more sense in that tonnage and speed have a positive relationship to cost and their effects are nonlinear. However, some of the other independent variables such as crew and length still show negative effects, albeit all modeled variables have the statistical significance of low *p*-values and a higher adjusted R-squared coefficient.

Multivariate Analysis (Warship Prices)

ID	Navy Ship	Unit Cost ($M)	Displacement (Tons)	Speed (KMH)	Length (M)	Crew	Year	Value	Q
1	DDG 51	2133	9648	56	155.3	276	2012	2,133	1
2	DDG 51	1553	9648	56	155.3	276	2012	3,106	2
3	DDG 51	1884	9648	56	155.3	276	2012	1,884	1
4	DDG 51	1423	9648	56	155.3	276	2013	4,269	3
5	DDG 51	2372	9648	56	155.3	276	2014	2372	1
6	DDG 51	1615	9648	56	155.3	276	2015	1,615	1
7	DDG 51	1330.5	9648	56	155.3	276	2016	2,661	2
8	DDG 1000	3554	15730	56	185.9	148	2007	3554	1
9	DDG 1000	3010	15730	56	185.9	148	2008	3010	1
10	Joint High Speed Vessel (JHSV)	185	2397	80	103	41	2010	185	1
11	Joint High Speed Vessel (JHSV)	184	2397	80	103	41	2011	184	1
12	Joint High Speed Vessel (JHSV)	376	2397	80	103	41	2012	376	1
13	Joint High Speed Vessel (JHSV)	207	2397	80	103	41	2013	207	1
14	LHA 6 America	3204	45695	37	114.91	1,687	2007	3,204	1
15	LHA 6 America	3213	45695	37	114.91	1,687	2011	3,213	1
16	Littoral Combat Ship	1077	3292	87	115.3	45	2010	1,077	1
17	Littoral Combat Ship	1147	3293	87	115.3	45	2011	1,147	1
18	Littoral Combat Ship	1858	3294	87	115.3	45	2012	1,858	1
19	Littoral Combat Ship	1821	3295	87	115.3	45	2013	1,821	1
20	LPD 17 San Antonio Class	1903	25300	39	208.5	360	2009	1,903	1
21	LPD 17 San Antonio Class	2088	25300	39	208.5	360	2012	2,088	1
22	USS Ticonderoga (CG 47)	1000	9754	56	173	30	2008	1,000	1
23	DD-21 Zumwalt	2700	16000	56	170	150	1996	2,700	1
24	Nimitz Class Aircraft Carrier (CVN 68)	4045	99800	56	332.8	558	2009	4,045	1
25	Nimitz Class Aircraft Carrier (CVN 68)	3421.3	99800	56	332.8	558	2011	3,421	1
26	Nimitz Class Aircraft Carrier (CVN 68)	4568.8	99800	56	332.8	558	2012	4,569	1
27	Nimitz Class Aircraft Carrier (CVN 68)	4738.2	99800	56	332.8	558	2016	4,738	1

Similar methodology in "Why Has the Cost of Navy Ships Risen?" RAND National Defense Research Institute 2006
Data Source: http://www.bga-aeroweb.com/Defense/DDG-51-AEGIS-Destroyer.html
http://www.globalsecurity.org/military/systems/ship/ddg-51.htm
http://www.defenseindustrydaily.com/adding-arleigh-burkes-northrop-grumman-underway-06007/

Figure 24. Sample Dataset for Parametric Modeling

The econometric-based parametric model shown in Figure 25 is the best model both in significance as well as logic. For instance, there are polynomial functions and first order versus second order interactions of the independent variables. Specifically, the functional form producing the best-fitting mixed nonlinear parametric cost model is

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_3 + \beta_3 X_4 + \beta_4 \ln(X_1) + \beta_5 \ln(X_2) + \beta_6 \ln(X_3)$$

$$Cost = 86373 - 0.37\,Tons + 302.18\,Length + 4.39\,Crew + 7108.91\ln(Tons) + 9778.02\ln(Speed) - 46327.8\ln(Length)$$

Clearly these are only illustrations based on sample publicly available data. Nonetheless, the approach is similar with actual data. The only difference would be to use datasets that pertain to the ship that is being modeled to prevent out-of-sample biases. Additional independent variables will need to be collected, and various econometric tests will need to be performed (e.g., see Appendix 4 of the primary report for an example list of specifications, data integrity, and error tests that will be performed, such as heteroskedasticity, multicollinearity, non-sphericity, nonlinearity, and so forth).

Basic Econometrics Analysis Report

Regression Statistics	
R-Squared (Coefficient of Determination)	0.9027
Adjusted R-Squared	0.8850
Multiple R (Multiple Correlation Coefficient)	0.9501
Standard Error of the Estimates (SEy)	437.4818
Number of Observations	27

The R-Squared or Coefficient of Determination indicates that 0.90 of the variation in the dependent variable can be explained and accounted for by the independent variables in this regression analysis. However, in a multiple regression, the Adjusted R-Squared takes into account the existence of additional independent variables or regressors and adjusts this R-Squared value to a more accurate view of the regression's explanatory power. Hence, only 0.89 of the variation in the dependent variable can be explained by the regressors.

The Multiple Correlation Coefficient (Multiple R) measures the correlation between the actual dependent variable (Y) and the estimated or fitted (Y) based on the regression equation. This is also the square root of the Coefficient of Determination (R-Squared).

The Standard Error of the Estimates (SEy) describes the dispersion of data points above and below the regression line or plane. This value is used as part of the calculation to obtain the confidence interval of the estimates later.

Regression Results

	Intercept	LN(VAR2)	LN(VAR3)	VAR4	VAR5
Coefficients	-40271.8660	3351.7927	3952.3622	-26.3671	-2.1820
Standard Error	7260.7364	596.1151	828.0072	6.9614	0.6903
t-Statistic	-5.5465	5.6227	4.7733	-3.7876	-3.1609
p-Value	0.0000	0.0000	0.0001	0.0010	0.0045
Lower 5%	-55329.7116	2115.5257	2235.1804	-40.8042	-3.6136
Upper 95%	-25214.0205	4588.0597	5669.5439	-11.9300	-0.7504

Degrees of Freedom	
Degrees of Freedom for Regression	4
Degrees of Freedom for Residual	22
Total Degrees of Freedom	26

Hypothesis Test	
Critical t-Statistic (99% confidence with df of 22)	2.8188
Critical t-Statistic (95% confidence with df of 22)	2.0739
Critical t-Statistic (90% confidence with df of 22)	1.7171

Figure 25. Parametric Model With Nonlinear Regression

Parametric Probability Distribution and Curve Fitting

Another powerful cost modeling approach is distributional fitting; that is, how does an analyst or engineer determine which distribution to use for a particular task's input cost or schedule variable? What are the relevant distributional parameters? If no historical data exist, we can make assumptions about the variables in question using the qualitative Delphi method, where a group of subject matter experts are tasked with estimating the behavior of each variable. Then, these values can be used as the variable's input parameters (e.g., uniform distribution with extreme values between 0.5 and 1.2). When testing is not possible (e.g., a new or novel weapon subsystem), management can still make estimates of potential outcomes and provide the best-case, most-likely case, and worst-case scenarios, whereupon a triangular or custom distribution can be created.

However, if reliable historical data are available, distributional fitting can be accomplished. Assuming that historical patterns hold and that history tends to repeat itself, then historical data can be used to find the best-fitting distribution with their relevant parameters to better define the variables to be simulated. Figure 26 illustrates a distributional-fitting example of the costs shown previously (Figure 24).

The null hypothesis (H_o) being tested is such that the fitted distribution is the same distribution as the population from which the sample data to be fitted came. Thus, if the computed *p*-value is lower than a critical alpha level (typically .10 or .05), then the distribution is the wrong distribution. Conversely, the higher the *p*-value, the better the distribution fits the data. Roughly, you can think of *p*-value as a percentage explained, that is, if the *p*-value is 0.9849 (Figure 26), then setting a normal distribution with a mean of 1990 and a standard deviation of 1290 explains about 98.49% of the variation in the data, indicating an especially good fit. The results from the Risk Simulator software also rank all the selected distributions and how well they fit the data. The fitted distribution can now be set up to run a simulation. The results from the simulation (tens to hundreds of thousands of simulation trials can be run) can be interpreted accurately (Figure 27).

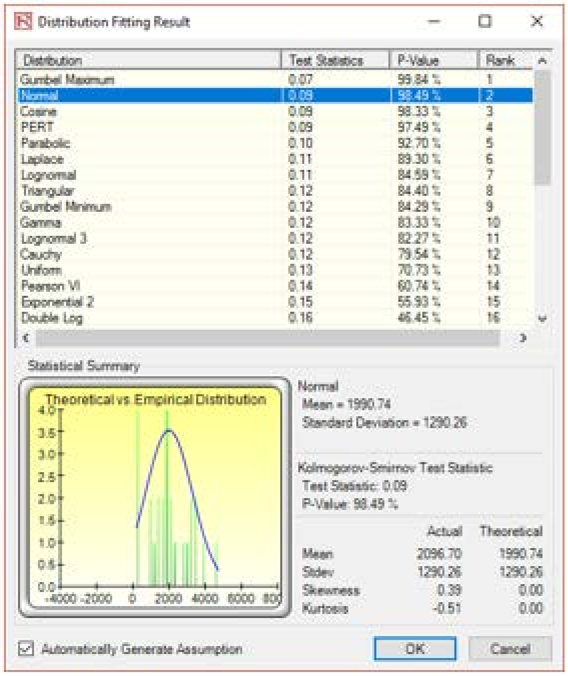


Figure 26. Parametric Monte Carlo Simulation Model Distributional Fitting

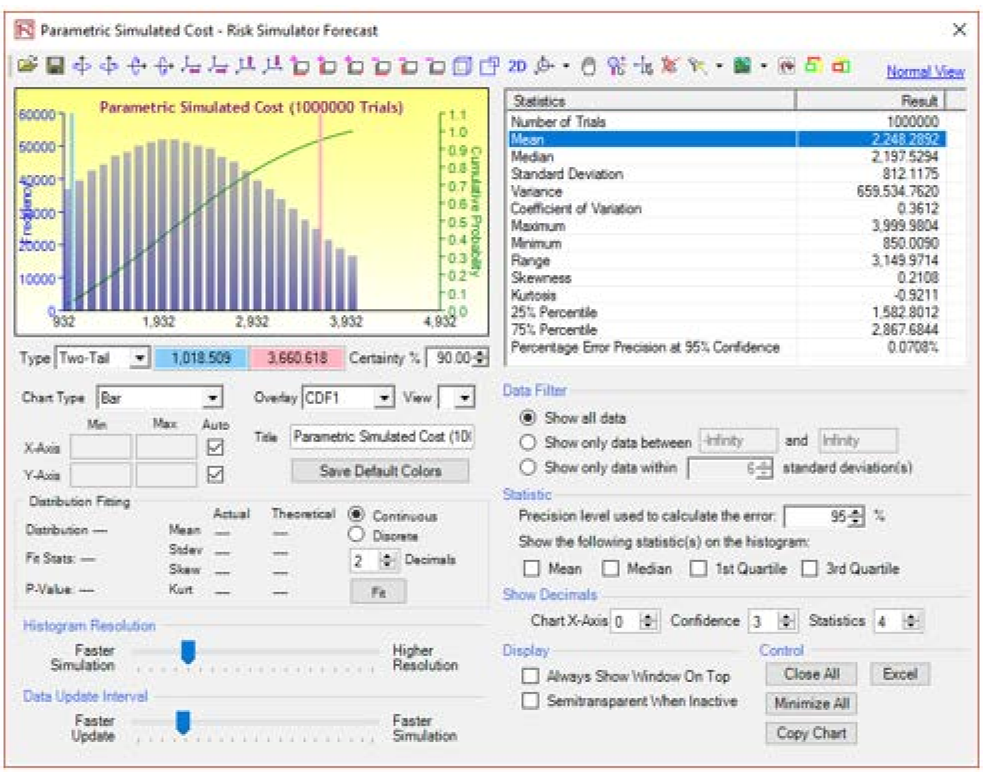


Figure 27. Parametric Simulated Cost Results

Conclusions and Next Step Recommendations

Based on this preliminary analysis and review of the alternatives, we conclude that the risk-based cost and schedule simulations as well as parametric econometric models can be applied to modeling the cost of current and future U.S. Navy warships. It is evident in the analysis that any cost modeling must also include schedule risk because schedule delays can cause significant cost creep and budget overruns. Using the project process diagrams and task-based cost modeling coupled with Monte Carlo simulations to account for uncertainties in input assumptions and estimates and risks of overruns, a comprehensive methodology was developed.

We therefore recommend the following:

- Collect and use actual cost data and develop more accurate cost estimates going forward in order to better calibrate the inputs based on real-life conditions. (We can provide suggestions on how to generate a database and methods to capture said required data.)
- Use the Risk Simulator–based simulated probability distributions to determine how well the vendors are performing (e.g., running at 92% efficiency, etc.), thus creating a common set of agreed upon performance metrics for the organization.
- Use control charts (based on simulated results) to determine if processes and tasks are in-control or out-of-control over time.
- Identify critical success factors to start collecting cost and schedule data for more accurate estimates.
- Incorporate learning curves and synergies when more than one ship is on order and the unit cost per ship would be lower.

The next phase of this research will focus on collecting actual cost and schedule data from a specific ship with subject matter experts' inputs to obtain the qualitative values. The resulting simulations will provide an alternative to the existing cost and schedule forecasting models that can be compared for accuracy over the course of the ship build. If complete archival cost and schedule data are available for a specific ship build along with the forecasted costs and schedule, this data can be applied to the ship cost model forecasting approaches suggested by the current study for purposes of comparison to the existing models that were used during the ship build.

References

Brown, A., & Neu, W. (2008). *Naval surface ship design optimization for affordability: Phase I.*

Congressional Budget Office. (2015, October). *An analysis of the Navy's Fiscal Year 2016 shipbuilding plan.* Retrieved from https://www.cbo.gov/sites/default/files/114th-congress-2015-2016/reports/50926-Shipbuilding.pdf

Deegan, C. (2005). *NAVSEA Cost Estimation Handbook.* Washington, DC: Naval Sea Systems Command.

Deschamps, L., & Greenwell, C. (n.d.). *Integrating cost estimating with the ship design process.*

Cost Assessment Data Enterprise (CADE). (2014). *Joint agency cost schedule risk and uncertainty handbook* (CSRUH). Retrieved from http://cade.osd.mil/cade/CSRUH.aspx

Kaluzny, B. L., Barbici, S., Berg, G., Chiomento, R., Derpanis, D., Jonsson, U., & Ramaroson, F. (2011). An application of data mining algorithms for shipbuilding cost estimation. *Journal of Cost Analysis and Parametrics, 4*(1), 2–30. doi:10.1080/1941658X.2011.585336

Lee, U. (2014). *Improving the parametric method of cost estimating relationships of naval ships* (Master's thesis). Cambridge, MA: Massachusetts Institute of Technology.

Miroyannis, A. (2006). *Estimation of ship construction costs* (Master's thesis). Cambridge, MA: Massachusetts Institute of Technology.

Mislick, G. K., & Nussbaum, D. A. (2015). *Cost estimation: Methods and tools*. Hoboken, NJ: Wiley.

Moore, G. W., & White, E. D. (2005). A regression approach for estimating procurement cost. *Journal of Public Procurement, 5*(2), 187–209.

Mulligan, R. F. (2008, September 1). A simple model for estimating newbuilding costs. *Maritime Economics & Logistics 10*(3), 310–321.

Mun, J. (2015). *Modeling risk: Applying Monte Carlo risk simulation, strategic real options, stochastic forecasting, portfolio optimization, data analytics, business intelligence, and decision modeling*. Dexter, MI: Thompson-Shore.

Mun, J. (2016). *Real options analysis: Tools and techniques*. Dexter, MI: Thompson-Shore.

Naval Sea Systems Command (NAVSEA). (2015). Cost estimating. In *SUPSHIP Operations Manual (SOM)* (NAVSEA S0300-B2-MAN-010; ch. 6). Washington, DC: Department of the Navy.

Ross, J. M. (n.d.). *A practical approach for ship construction cost estimating*. Retrieved from http://ds-t.com/trade_shows-cd/ProteusEngineeringCOMPITPaper2004.pdf

Smith, M. B. (2008). *Updating MIT's cost estimation model for shipbuilding* (Master's thesis). Cambridge, MA: Massachusetts Institute of Technology.

SPAR Associates. (2015). *Independent cost estimating services* [Presentation]. Retrieved from http://www.sparusa.com/Presentations/Independent%20Cost%20Estimating%20Services-Overview.pdf

Truver, S. C. (2001, January–February). Navy develops product oriented design and construction cost model: PODAC emerges as critical element in achieving operationally superior, affordable naval forces. *Program Manager*, 38–41.

Walcott, J. (2012). Budget Office questions Navy shipbuilding cost estimates. *Bloomberg*. Retrieved from http://www.bloomberg.com/news/articles/2012-07-25/budget-office-questions-navy-shipbuilding-cost-estimate

Appendix: Most Common Forecast and Predictive Modeling Techniques

- *ARIMA*. Autoregressive integrated moving average (ARIMA, also known as Box–Jenkins ARIMA) is an advanced econometric modeling technique. ARIMA looks at historical time-series data and performs back-fitting optimization routines to account for historical autocorrelation (the relationship of a variable's values over time, that is, how a variable's data is related to itself over time). It accounts for the stability of the data to correct for the nonstationary characteristics of the data, and it learns over time by correcting its forecasting errors. Think of ARIMA as an advanced multiple regression model, where time-series variables are modeled and predicted using its historical data as well as other time-series explanatory variables. Advanced knowledge in econometrics is typically required to build good predictive models using this approach. Suitable for time-series and mixed-panel data (not applicable for cross-sectional data).
- *Auto-ARIMA*. The Auto-ARIMA module automates some of the traditional ARIMA modeling by automatically testing multiple permutations of model specifications and returns the best-fitting model. Running the Auto-ARIMA module is like running regular ARIMA forecasts; the differences being that the required P, D, Q inputs in ARIMA are no longer required and that different combinations of these inputs are automatically run and compared. Suitable for time-series and mixed-panel data (not applicable for cross-sectional data).
- *Basic Econometrics*. Econometrics refers to a branch of business analytics, modeling, and forecasting techniques for modeling the behavior or forecasting certain business, economic, finance, physics, manufacturing, operations, and any other variables. Running Basic Econometrics models is similar to regular regression analysis except that the dependent and independent variables are allowed to be modified before a regression is run. Suitable for all types of data.
- *Basic Auto Econometrics.* This methodology is similar to basic econometrics, but thousands of linear, nonlinear, interacting, lagged, and mixed variables are automatically run on your data to determine the best-fitting econometric model that describes the behavior of the dependent variable. It is useful for modeling the effects of the variables and for forecasting future outcomes, while not requiring the analyst to be an expert econometrician. Suitable for all types of data.
- *Combinatorial Fuzzy Logic.* Fuzzy sets deal with approximate rather than accurate binary logic. Fuzzy values are between 0 and 1. This weighting schema is used in a combinatorial method to generate the optimized time-series forecasts. Suitable for time-series only.
- *Custom Distributions*. Using Risk Simulator, expert opinions can be collected and a customized distribution can be generated. This forecasting technique comes in handy when the dataset is small, the Delphi method is used, or the goodness-of-fit is bad when applied to a distributional fitting routine. Suitable for all types of data.
- *GARCH*. The generalized autoregressive conditional heteroskedasticity (GARCH) model is used to model historical and forecast future volatility levels of a marketable security (e.g., stock prices, commodity prices, oil prices, etc.). The dataset has to be a time series of raw price levels. GARCH will first convert the prices into relative returns and then run an internal optimization to

fit the historical data to a mean-reverting volatility term structure, while assuming that the volatility is heteroskedastic in nature (changes over time according to some econometric characteristics). Several variations of this methodology are available in Risk Simulator, including EGARCH, EGARCH-T, GARCH-M, GJR-GARCH, GJR-GARCH-T, IGARCH, and T-GARCH. Suitable for time-series data only. This technique can be used with cost data in the current ship costs context by forecasting ship cost volatility.

- *J-Curve*. The J-curve, or exponential growth curve, is one where the growth of the next period depends on the current period's level and the increase is exponential. This phenomenon means that over time, the values will increase significantly, from one period to another. This model is typically used in forecasting biological growth and chemical reactions over time. Suitable for time-series data only. It can be used in the current cost context by forecasting cost growth data.
- *Markov Chains*. A Markov chain exists when the probability of a future state depends on a previous state and when linked together forms a chain that reverts to a long-run steady state level. This approach is typically used to forecast the market share of two competitors. The required inputs are the starting probability of a customer in the first state returning to the same state in the next period, versus the probability of switching to a competitor's state in the next state. Suitable for time-series data only.
- *Maximum Likelihood on Logit, Probit, and Tobit*. Maximum likelihood estimation (MLE) is used to forecast the probability of something occurring given some independent variables. For instance, MLE is used to predict if a credit line or debt will default given the obligor's characteristics (30 years old, single, salary of $100,000 per year, and total credit card debt of $10,000), or the probability a patient will have lung cancer if the person is a male between the ages of 50 and 60, smokes five packs of cigarettes per month or year, and so forth. In these circumstances, the dependent variable is limited (i.e., limited to being binary 1 and 0 for default/die and no default/live, or limited to integer values such as 1, 2, 3, etc.) and the desired outcome of the model is to predict the probability of an event occurring. Traditional regression analysis will not work in these situations (the predicted probability is usually less than zero or greater than one, and many of the required regression assumptions are violated, such as independence and normality of the errors, and the errors will be fairly large). Suitable for cross-sectional data only.
- *Multivariate Regression*. Multivariate regression is used to model the relationship structure and characteristics of a certain dependent variable as it depends on other independent exogenous variables. Using the modeled relationship, we can forecast the future values of the dependent variable. The accuracy and goodness-of-fit for this model can also be determined. Linear and nonlinear models can be fitted in the multiple regression analysis. Suitable for all types of data.
- *Neural Network*. This method creates artificial neural networks, nodes, and neurons inside software algorithms for the purposes of forecasting time-series variables using pattern recognition. Suitable for time-series data only.
- *Nonlinear Extrapolation*. In this methodology, the underlying structure of the data to be forecasted is assumed to be nonlinear over time. For instance, a

dataset such as 1, 4, 9, 16, 25 is considered to be nonlinear (these data points are from a squared function). Suitable for time-series data only.

- *S-Curves*. The S-curve, or logistic growth curve, starts off like a J-curve, with exponential growth rates. Over time, the environment becomes saturated (e.g., market saturation, competition, overcrowding), the growth slows, and the forecast value eventually ends up at a saturation or maximum level. The S-curve model is typically used in forecasting market share or sales growth of a new product from market introduction until maturity and decline, population dynamics, and other naturally occurring phenomenon. Suitable for time-series data only.
- *Spline Curves*. Sometimes there are missing values in a time-series dataset. For instance, interest rates for years 1 to 3 may exist, followed by years 5 to 8, and then year 10. Spline curves can be used to interpolate the missing years' interest rate values based on the data that exist. Spline curves can also be used to forecast or extrapolate values of future time periods beyond the time period of available data. The data can be linear or nonlinear. Suitable for time-series data only.
- *Stochastic Process Forecasting*. Sometimes variables are stochastic and cannot be readily predicted using traditional means. Nonetheless, most financial, economic, and naturally occurring phenomena (e.g., motion of molecules through the air) follow a known mathematical law or relationship. Although the resulting values are uncertain, the underlying mathematical structure is known and can be simulated using Monte Carlo risk simulation. The processes supported in Risk Simulator include Brownian motion random walk, mean-reversion, jump-diffusion, and mixed processes, useful for forecasting nonstationary time-series variables. Suitable for time-series data only.
- *Time-Series Analysis and Decomposition*. In well-behaved time-series data (typical examples include sales revenues and cost structures of large corporations), the values tend to have up to three elements: a base value, trend, and seasonality. Time-series analysis uses these historical data and decomposes them into these three elements, and recomposes them into future forecasts. In other words, this forecasting method, like some of the others described, first performs a back-fitting (backcast) of historical data before it provides estimates of future values (forecasts). Suitable for time-series data only.
- *Trendlines*. This method fits various curves such as linear, nonlinear, moving average, exponential, logarithmic, polynomial, and power functions on existing historical data. Suitable for time-series data only.

Estimating the Estimate: Toward a Quick and Inexpensive Method for Weapons System Cost Estimation

Charles K. Pickar—DBA, is a member of the faculty of the Naval Postgraduate School, where he teaches project management, defense acquisition, and systems engineering. He is a retired Army officer with extensive experience in the U.S. defense industry serving in program manager, systems engineer, and business development roles at the director and VP levels at Lockheed Martin, Northrop Grumman, and SAIC. Before joining NPS, he led the Applied Systems Engineering Program Area in the National Security Analysis Department of the Johns Hopkins University Applied Physics Laboratory. He is the current Chair of the Systems Education Technical Committee of the IEEE Systems Council. His research and published work focuses on applying systems engineering and system dynamics approaches to defense acquisition problems. [ckpickar@nps.edu]

Kevin G. Feely—is the President and CEO of KCR Business Innovations, a consulting company that provides decision analysis models on cost and schedule for defense companies. He is also a licensed CPA. He is a cost estimation expert, having served over 25 years in cost estimating and accounting roles at defense companies including Boeing, SAIC, and most recently at Aerovironment. He is an expert on developing and building algorithm-based decision models and has used that expertise in both costing as well as merger and acquisition (M&A) decision-making. [kefeely@kcrbi.com]

Abstract

This paper describes a process to provide low-cost and timely cost estimation for weapons system and services procurements based on data analytics. This will enable quick estimation that can serve as a sanity check for more formal estimation methods. Specifically, this cost estimation technique is built on the price-to-win methodology used by defense companies to respond to DoD solicitations. The process is based on market clearing prices as reflected by actual winning bids, expert knowledge to validate the scope of work, and an algorithm developed to incorporate these aspects into an estimate. The process is an "estimate of the estimate" and meant as an adjunct to formal cost estimating processes. The value lies in the ability to create quick, inexpensive estimates responsive to management needs while the formal cost estimating process proceeds.

Introduction

Cost estimation is as much science as it is art. And, like most science and art, it is resource intensive and time-consuming. The science is driven by the desire to accurately capture the elements of cost—including the amount of labor required, labor rates, overhead rates, and so forth—to provide an accurate starting point for budgeting and program management activities. The art of cost estimation is the fine line between accepting provided data and critically examining that data before accepting it. The most basic inputs to a cost model are the cost elements defined and captured by accountants: direct labor, materials, and overhead. However, identifying these cost elements is just the science. The art demands experience with both the data and the requirements of systems development. Good cost estimates at every stage of the development process must have this mixture of art and science. The goal of this paper is to present a tool that uses the art and science to provide inexpensive and timely initial cost estimates for weapons systems and services procurements based on data analytics.

Realistic cost estimation is a necessity for successful weapon systems development and services contracts. However, cost estimates are frequently wrong, or unrealistic, and in many cases, time-consuming and unresponsive. Further, as the GAO states, "bias and over optimism creep into estimates that advocates of weapon systems prepare, and the estimates tend to be too low" (GAO, 2009). The science of cost estimation depends on accurate data consisting of unambiguous definitions of the tasks, standardized work breakdown structures, recognition of the uncertainties of system development, and the recognition that significant program changes will cause changes in the estimate (GAO, 2009). The art is found in the familiarity and know-how of both the cost estimators as well as the cost engineers and others involved in applying knowledge of developments to develop estimates.

Current challenges with the cost estimation process include

- known costs being excluded without adequate justification
- invalid historical cost data
- inconsistent consideration of inflation
- cost and time necessary to accomplish

It is this last issue of cost and time necessary to accomplish that is the focus of this effort. Whether cost estimator, cost engineer, or project manager, the need for quick, ROM (rough order of magnitude) cost estimates is an ever-present requirement of the DoD acquisition workforce. Users and contractors routinely ask for more capability from existing systems, as well as proposing new systems. Young (Young & Markley, 2008) describes the scenario:

> "How soon can you get me a rough order of magnitude [ROM] on the cost?" The project engineer does a mental retrieval and concludes that a full bottoms up engineering estimate is needed, but that will take too long—about three to four months. The project engineer knows it has to be faster, so he throws a number out. "I need a month to develop a ROM." "Give me a ROM in two weeks if you really want any chance of funding this initiative," is the reply.

Unfortunately, the time and money to act on these requests is in short supply. We need a tool that can react to the demands of present-day acquisition, yet be accurate enough to satisfy the standards of cost estimating.

The proposed tool was developed for the defense industry while the authors were employed by one of the major defense contractors. The development of this tool initially focused on one business-focused cost estimating activity, determining the price-to-win (PTW) in defense markets. We believe this process, however, offers the means to address the scenario described above, and those similar situations that occur every day across the DoD acquisition space. It is important to note this process is not a formal cost estimation process, nor is it meant to supplant the recognized cost estimating processes. This "estimating the estimate" is an adjunct to formal cost estimating to enable cost estimators, cost engineers, and the PMO a quick way to develop an estimate—of the formal cost estimate. The process is based on basic economic theory—the law of supply and demand and market clearing theory, the point where producers' products and consumers' demand are equal. This quick estimation process can serve as a sanity check for the more validated and formal estimation methods. It can also help address the challenge of responding to requests for ROM pricing in case of extra funding availability, end of year funding, etc.

Cost estimation is serious business with very real fiscal and operational consequences. This tool builds on the *science* of cost estimation by incorporating both the actual winning bids for DoD competitions and the associated costs throughout the development. In other words, it models the market using prices that were successfully bid. This becomes the market clearing price. This approach offers the possibility of providing macro-level estimates based on microeconomic theory and historical trends and weapon systems development using a data analytics approach. The *science* still defines the determinants or drivers of cost such as productivity or labor rates. The *art* is in modeling the market.

The following is the research question to be examined:

Can market pricing provide a reasonably accurate cost estimating methodology that is quick and less expensive to execute that provides actionable information to government program managers?

Specifically, we seek to examine whether a data analytics approach based on market pricing will yield actionable cost estimates.

Cost Estimation

The field of cost estimation is rich in research. However, of late the preponderance of the effort appears to be focused on software cost estimation. This is understandable given the importance of software in the modern world, and the opaque nature of estimating software development. And, of course, the fact that an ever-larger percentage of capability in weapons systems comes from the software-hardware integration, rather than hardware itself. Regardless of the focus, the basic quantitative and qualitative methods are used in all cost estimating.

Cost estimation for both software, hardware, and service projects is a formal, documented process. Review of the field identifies the broadly defined methods to estimate costs. These include the following (Boehm, 1984; Evans, Lanham, & Marsh, 2006; Jorgensen, 2005; Leung & Fan, 2002):

- analogy;
- top-down;
- bottom-up;
- Parkinson;
- algorithmic models;
- expert judgment; and
- Price-to-Win (PTW).

The most commonly used techniques in the DoD are analogy, parametric (top-down), and engineering (bottom-up) estimating (Mislick & Nussbaum, 2015). Analogy simply compares similar developments using historically captured cost information. A parametric or top-down estimate builds a cost estimate for the development project from historical data comparing variables through a statistical relationship. Finally, an engineering or bottom up estimate is a comprehensive cost estimate starting at the work package level and aggregating costs to build a more complete estimate.

The Parkinson estimation is based on the Parkinson principle that “work expands to fill the available volume.” It is only mentioned to acknowledge that while not a rigorous estimating tool, there are times when the cost is determined by available resources, rather than a defined end-state.

Algorithmic cost estimation models use one or more algorithms to analyze variables considered to be the major cost drivers for the weapons system. The algorithmic methods are based on mathematical models that produce cost estimates as a function of a number of variables that are considered to be the major cost factors (Leung & Fan, 2002).

Expert judgment simply acknowledges that engineering experts should be able to estimate the effort necessary to accomplish development tasks and translate those estimates to costs in technological activities—where they have experience. Thus, expert judgment is defined as the consultation of one or more experts (Hughes, 1996).

Some disagree on whether the final category, price-to-win (PTW), is actually a cost estimation methodology, categorizing it as a cost management process rather than a cost estimating process (Boehm, 1984). In the PTW, the cost estimate is equated to the price believed necessary to beat competitors. In other words, PTW is a market-focused estimate focused on identifying the price necessary to win a government competition. Defense companies regularly use PTW as a target price to drive internal cost-savings measures as well as to drive the design-to-cost (DTC) target. Notwithstanding this characterization, we propose that certain aspects of the PTW process can be used to approximate initial system costs. The proposed tool combines two of these cost estimation methodologies, expert judgment to define an algorithm as a PTW to provide an initial cost estimate.

Price-to-Win

Regardless of stated evaluation criteria, price is a significant factor in most government contract decisions. PTW is used in industries that have limited customers—monopsonies including the U.S. defense industry. While often mentioned in the broad category of pricing methodologies, it is often dismissed as "the price believed necessary to win the contract," thus not acceptable for formal cost estimating. This definition has evolved into developing a strategy that fits the customer budget rather than the effort required to complete the work (Leung & Fan, 2002). It is true the defense industry uses PTW for competitive reasons. Industry competitors want to present the government customer with their lowest price, while ensuring adherence to RFP requirements, at least in comparison to other competitors. The PTW is part of a decision-making process that includes an assessment of the firms' ability to develop a cost-competitive offer within their risk tolerance. PTW is focused on ensuring the industry solution meets the government needs, while emphasizing the competitive advantage of individual companies. It is worth noting that PTW is widely practiced in the U.S. defense industry. In fact, as a matter of process, many firms require a PTW determination before deciding to spend the money necessary to prepare a proposal, and throughout the proposal development.

The PTW approach consists of estimating the price for each competitor, a potentially expensive process in that the current practice of PTW requires accurate assessments and analysis of competitive intelligence on the competition. Defense companies seek to understand in detail both competitor companies' strengths and weaknesses, as well as their pricing structure. PTW also analyzes the nuances of the government customer, specifically what the award history is, as well as any trends in reasons for selections. The competitive intelligence is based on open source materials to try to determine both any competitive advantages individual companies may have, as well as any unique approaches to solving the government's problem. A type process of PTW includes the steps shown in Table 1.

Table 1. The Price-to-Win Process

Task	Description
Strategic Assessment of the Market/Opportunity	An ongoing process that tracks winning bids, usually done by the business development organization of a company
Competitive Analysis	Competitor company capabilities/services/competitive advantage Specifications of products Estimation of rates Estimation of solution
Customer Assessment	Customer contract award history Bottom up cost modeling
Risk Assessment	Assessment of cost and technical risk

A consideration in competitive intelligence assessments is that the U.S. defense industry is more or less balanced in capability across like companies. For instance, Raytheon and Lockheed Martin have legacy radar and missile businesses and compete in those areas. Boeing and Lockheed Martin compete in the high-performance fighter aircraft market and so on. The drawdowns and consolidations of the defense industry in the 1990s, as well as more recent consolidation, make competition in those areas fierce. Losing a government contract could mean exiting that line of business, thus determining what the competitor is going to do, and deciding an offer price is a high-stakes effort.

By industry capability we also mean intellectual property, manufacturing efficiency, and human resources. Intellectual property is driven by the investment firms make in Independent Research and Development (IRAD) efforts. Manufacturing efficiency tends to mirror the overall industry and remains a source of potential profit if managed cost-effectively. Human resources refer to engineering talent—the product of quality education and individual potential. The defense industry capability is driven by the labor market, and manufacturing efficiency is determined by the overall state-of-the-art. The defense industry draws from the same talent pool. Differences in competitive pricing originate from specific qualifications or competitive advantage of intellectual property from self-developed research and development programs, not necessarily widely varying labor rates. The PTW reflects the market clearing price.

For the defense firm, the PTW analysis should yield a value that addresses customer need that also will be successful against competition. A finalized PTW analysis reflects a schedule-performance tradeoff that becomes a pacing item for development of the proposed system.

The PTW approach is a macro-economic examination of an existing DoD program. While not specific enough to address the actual development of a cost estimate, it could serve as both ROM and as an indicator in the continued pursuit of a detailed cost estimate. These estimates would not replace the detailed parametric cost estimates, or IGCE, nor be a substitute for market research. Instead, the solution would provide the PM/contracting officer a means to validate/confirm the results of more in-depth cost analysis, while providing program office personnel a starting point for budgeting, and cost realism.

The PTW Process Translated

As noted, this tool leverages the science and art of cost estimation through the application of expert judgment and algorithmic data modeling. Expert judgment is an estimate based on the expertise of one or more people familiar with the costs and scope of similar system developments (Keeney & Winterfeldt, 1989; Morris, 1974). In the case of

PTW, the expert judgment comes from people familiar with the market for that particular product. In defense firms this includes cost engineers, specific engineering domain engineering experts, and the business development staff. In use as a cost estimation tool for the DoD, expert judgment would include government cost engineers and domain experts, as well as results of the market research activities (RFI—Requests for Information).

The data used for the tool consists of the actual winning bids and the associated scope of past programs, plus a mechanism to track the inevitable changes to the cost estimate as changes occur throughout the development. This approach could be more accurately described as an "algorithmic" method. It offers the possibility of providing macro-level estimates based on microeconomic theory and historical trends and weapon systems development using a data analytics approach.

Three data elements—scope, budget, and contract award price—were used to develop an algorithm to explain the winning price ranges for the competitive solicitations chosen. Although focused on the price necessary to win a contract, we believe these price ranges should correspond to the initial cost estimates provided by the government. Cost is estimated as a mathematical function of product, project, and process attributes whose values are estimated by project managers.

An essential factor of this analysis is the ability to identify both initial costs against specific SOW tasks, as well as track cost-growth/scope increase as the product developed. As we continue to develop the tool, we expect this analysis to show relationships between similar contracts of similar value scope.

In developing the tool, we started with three basic macro-economic assumptions:

1. In non-commodity markets the equilibrium price is the mean of a range of prices which are normally distributed about the equilibrium price. The government contracting market is a non-commodity market.
2. The equilibrium price represents the balancing of costs, risks, and margin for the government and the contractor. If one of these three elements is negatively skewed for a specific supplier, they would exit the market. If one of these three elements is positively skewed for a specific supplier, competition would respond and the price would adjust accordingly.
3. In the government contracting market, a monopsony, supply exceeds demand and the price for the goods or services will be below the equilibrium price.

The first step in this cost estimation process is to identify an equilibrium price (EP). The EP is approximated by developing two extreme estimates for a given government provided statement of work (SOW). This first task depends on expert judgment to establish an initial range of possible prices, a low-price estimate (LPE) and a high-price estimate (HPE). These expert estimates are considered from both the government and contractor perspective, acknowledging the different ways government and contractor cost estimators consider a system cost. The low end of the estimate reflects the expert's opinion on the cost associated with meeting the minimums of schedule and performance. The low-price estimate is the absolute minimum the government estimator believes is necessary for schedule and performance execution, and therefore reflects a price that represents the extreme risk for a contractor to execute. Figure 1 shows the normal distribution of the LPE-HPE estimates.

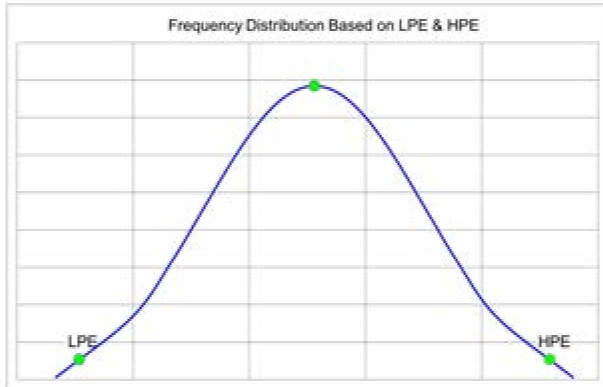


Figure 1. Equilibrium Price

On the other end of the spectrum, the HPE represents a price the government believes addresses the risk for both government and contractor. These two extremes should define the possible market for the system to be developed, with the mean representing the equilibrium price in balanced market where the price reflects an optimal balance between costs, risks, and contractor margin.

As noted, the normal market clearing distribution will tend to be skewed because of monopsony in defense markets. In this case, supply will always exceed demand; therefore, the mean will tend to be lower than the equilibrium price. Figure 2 represents that distribution.

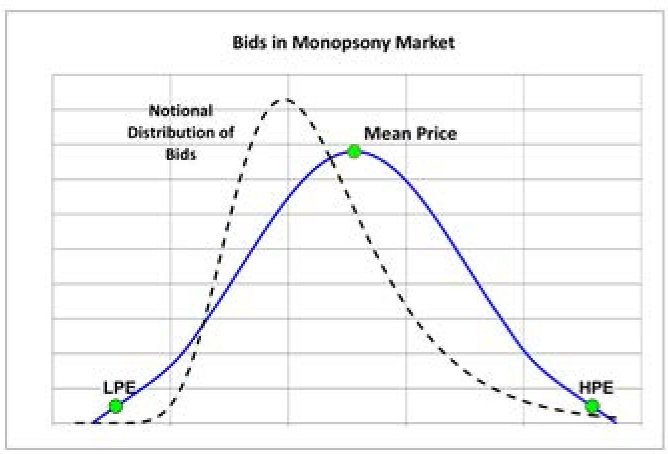


Figure 2. Bids in a Monopsony Market

Obviously, the government expert estimator needs access to the contractor perspectives on price, but that is normal. We believe government marketing research RFI could assist in this data collection. There are two steps to solicit these estimates. First, as is often done today, the PMO would request a ROM price as part of the market research effort. The challenge with ROM pricing estimates provided by potential contractors is the contractor concern their ROM will become the government target price. Thus, contractors will always add a “pad” to the ROM to reflect unknown and unforeseen risk. The government expert should be the arbiter of these estimates. In planning the RFI, the PMO should also request an LPE and HPE using the definitions provided above. Assuming multiple responses to the market research, the government expert would be able to develop a reasonable estimate of the LPE and HPE. The government expert could use those estimates in developing a government LPE/HPE or maintain separate analysis and weigh the results prior to the next step.

The next step determines the cost estimate range and uses the second aspect of cost estimation, the algorithmic cost model. This step consists of two sequential activities built on historical contracting data (at this point proprietary in nature). The first part, which is ongoing, is to collect data on the defense industrial market in the United States. This data collection captures the open source information available on defense contracts. The data include the DoD request for proposal (RFP) details, including project scope and the government's budget. This information is matched to contract award price. Initially the emphasis has been on major programs (ACAT I); however, we intend to continue to gather as much data as possible. The initial dataset is small and reflects only the past three years. As more data becomes available over time, we believe the accuracy of the model will be improved.

The second part of this step uses a statistically relevant number of prior procurements (aligned to scope) to estimate the range of costs for a specific SOW type. The intent is to create a frequency distribution of the actual bids received. This information forms the basis for the algorithm used to identify the competitive price ranges for the sampled procurements. From the contractor perspective (PTW) the winning price should be below the equilibrium price by bidders altering the balance of cost, risk, and margin to win the contract—and reflecting the market clearing price. Thus, from a market clearing perspective, using historical data for like-system procurements, an initial estimate of the cost could be derived. Figure 3 represents that range. The final price is then determined by using statistical tools and applying the algorithm.

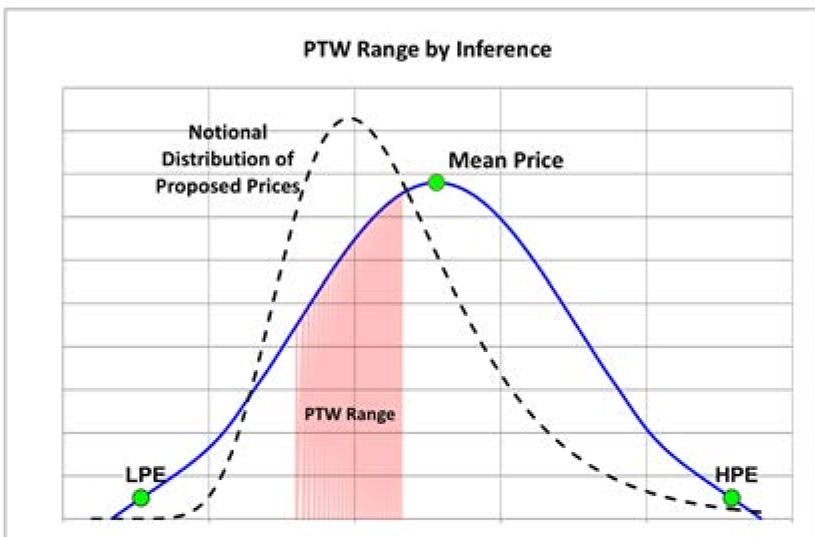


Figure 3. Algorithm Applied

Estimating the Estimate in Practice

To demonstrate the process, the following example is presented. The example project is a communications/electronics retrofit solution for surface ships for the U.S. Navy. The desired vehicle is a firm, fixed price contract, and the evaluation criteria is LPTA (Lowest Price Technically Acceptable). This effort is for a build-to-print production contract for 89 systems for three different ship types. The expert judgment estimate determined by RFI for the LPE and HPE is shown in Table 2.

Table 2. Cost Estimates Example

Ship	LPE ($M)	HPE ($M)
Ship Type A	$7.0	$9.5
Ship Type B	$7.5	$12.0
Ship Type C	$2.5	$5.0

The first step is to estimate the market price for each ship type. (For purposes of this example, we will estimate Ship A only.) The LPE and HPE represent the extremes of the market pricing for the solution fitted on the respective ship. LPE is defined as the estimate theoretically technically compliant (in this case LPTA), which minimizes labor costs and technical and programmatic risk. HPE is the estimate capturing all reasonable labor costs and factors in all programmatic and technical risk. Other factors not apparent in this example are the quantities of ships and the corresponding communications/electronics system solution.

Using proprietary historical information, the following calculations in Table 3 represent a range of LPTA estimates for the Ship Type A work.

Table 3. LPTA Example

	LPTA Range Probabilities	
10%	$7,330,402	$7
16%	$7,373,314	$7
20%	$7,396,158	$7
40%	$7,484,087	$7
50%	$7,521,955	$8
60%	$7,559,822	$8
80%	$7,647,751	$8
90%	$7,713,507	$8

Using the same process, but approaching the problem from a best value approach, the results of the analysis are shown in Table 4. Figure 4 is a graphical representation of the plotted ranges for the Ship A estimate. In this example, the 90% probability value was within 3% of the actual winning bid.

Table 4. Best Value Example

	Best Value FFP Range Probabilities	
10%	$7,844,924	$8
16%	$7,894,477	$8
20%	$7,920,857	$8
40%	$8,022,393	$8
50%	$8,066,121	$8
60%	$8,109,849	$8
80%	$8,211,386	$8
90%	$8,287,318	$8

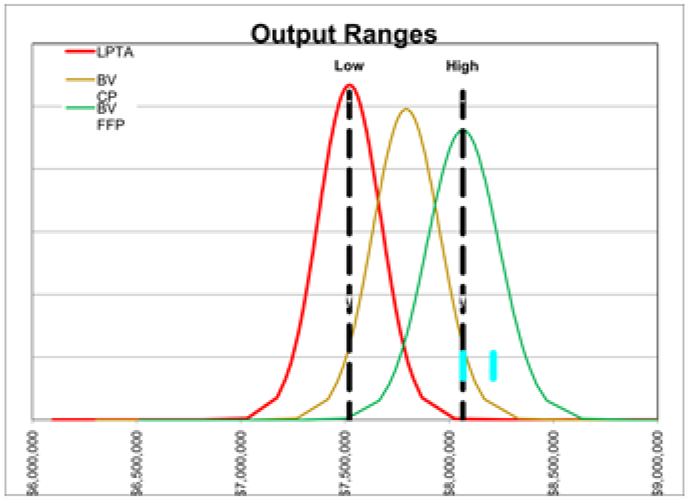


Figure 4. Plotted Ranges

Figure 5 shows the S-curves and the range of estimates for the different contract types, LPTA< CP and FFP (Best Value). The end state of this "estimate of the estimate" is a statistical range of pricing that provides the key layers, project engineers, cost estimators, program office, and contracting officials a starting point. To return to the scenario, instead of an expert judgment-only guided "guestimate" or worse, the project engineer can provide an empirically based estimate. To be sure, it will not be the final estimate, but it will provide a starting point. More importantly it will provide an answer to a very tough question.

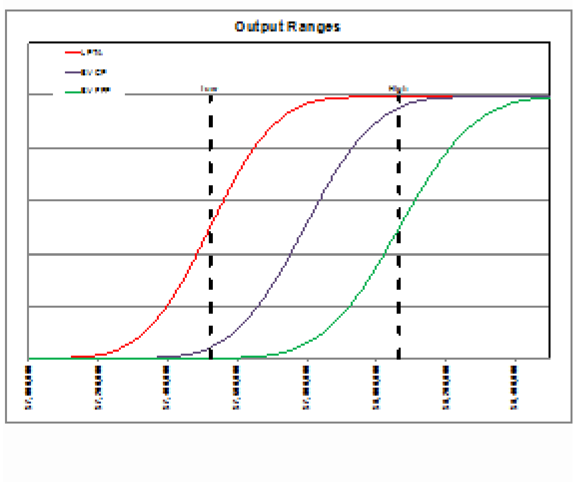


Figure 5. S-Curves for Example

Conclusion

This paper proposes using a defense industry tool, the price-to-win, to assist the cost estimating process for the DoD. Using market pricing and accepting the idea that while costs differ between competitors, their costs are generally similar, we suggest that applying economic market clearing ideas can provide a quick, inexpensive, and reasonably accurate cost estimate for most efforts.

This approach provides a macro-economic estimate of DoD programs. While not specific enough to address the actual development of a cost estimate, it could serve as both ROM, as well as an indicator in the continued pursuit of detailed cost proposals. The estimates would not replace the detailed parametric cost estimates, or IGCE, nor be a substitute for market research. Instead, the solution would provide a means to validate/confirm the results of more in-depth cost analysis, while providing program office personnel a starting point for budgeting and cost realism.

There are both pragmatic and theoretical limitations to this approach. Pragmatically, the technique must be tested using real data available only to the government. A second potential limitation is confidence in the ability of the government cost engineers (expert judgment) to define the LPE and HPE estimates. A major assumption of this approach is that government cost engineers and project managers are, in fact, experts and have a reasonable understanding of the range of costs for similar development projects. Our experience reinforces this belief, but it is clear that the better the estimates for the LPE and HPE, the better the overall estimate.

Theoretically, there must be a spread of at least one standard deviation but not too close to the end of the distribution in order to address the entire market. Theoretical limitations also include the available market data both in system and software procurements, and the amount of variance assumed to be in the distribution.

Finally, this tool can be improved and constantly updated by linking the DoD information on winning bids and the associated scope with the existing algorithm. A next step in this research is to request the use of said data and formally establish a validation effort to determine the quality of the results of the tools' computations.

References

Boehm, B. W. (1984). Software engineering economics. *IEEE Transactions on Software Engineering, SE-10*(1), 4–21. http://doi.org/10.1109/tse.1984.5010193

Evans, D. K., Lanham, J. D., & Marsh, R. (2006). Cost estimation method selection: Matching user requirements and knowledge availability to methods. University of West of England.

GAO. (2009). *GAO cost estimating and assessment guide*. Washington, DC: Author.

Hughes, R. T. (1996). Expert judgement as an estimating method. *Information and Software Technology, 38*(2), 67–75. http://doi.org/10.1016/0950-5849(95)01045-9

Jorgensen, M. (2005). Practical guidelines for expert-judgment-based software effort estimation. *IEEE Software, 22*(3), 57–63. http://doi.org/10.1109/MS.2005.73

Keeney, R. L., & von Winterfeldt, D. (1989). On the uses of expert judgment on complex technical problems. *IEEE Transactions on Engineering Management, 36*(2), 83–86. http://doi.org/10.1109/17.18821

Leung, H., & Fan, Z. (2002). Software cost estimation. *Handbook of Software Engineering*, 1–14.

Mislick, G. K., & Nussbaum, D. A. (2015). *Cost estimation: Methods and tools.*

Morris, P. A. (1974). Decision analysis expert use. *Management Science, 20*(9), 1233–1241. http://doi.org/10.1287/mnsc.20.9.1233

Young, J. A., & Markley, T. (2008). A better cost estimating tool: The key to not going over budget. *Defense AT&L*, 23–26.

Determining the Value of a Prototype

Zachary S. McGregor-Dorsey—received his degree in mathematics from the University of Colorado Boulder. He has worked in the DoD cost-related fields for nearly a decade, working on everything from research to detailed life cycle cost estimates. His current interests lie in graph theory, machine learning, and banjo playing, and he is always looking for applications of those interests to the cost world. [zmcgrego@ida.org]

Abstract

Most major defense acquisitions require a technology that is not yet fully developed, introducing a non-negligible amount of risk to the program and its cost. Developing a prototype or technology demonstrator prior to execution of the program can be useful in mitigating this risk, yet these demonstration programs also have associated costs. This paper develops a method to value this risk mitigation, setting an appropriate maximum cost for the demonstration program. This novel application of Value of Information theory and properties of the Bayesian preposterior distribution requires only a program cost estimate distribution and some estimate of possible results of the demonstration program. The method is broadly applicable to programming with varying amounts of technological uncertainty. We describe the method, then show how actual cost overruns of historical programs with and without prototypes can be used to estimate a value of prototype efforts relative to estimated program cost. We conclude with a discussion of other applications and how to explain the method and results to decision-makers.

Introduction

With any new defense program, the decision to begin its research, development, test, and evaluation (RDT&E) phase is a significant commitment of resources. It is a commitment that—due to contracts, regulations, and momentum—is not easily annulled. Even programs that experience cancellation tend to do so after appreciable investment. The decision-maker faces this weighty determination at a time when information about the program and its challenges is largely speculative and difficult to measure, making any estimate of its cost subject to a high degree of uncertainty.

In this situation, a decision-maker may appreciate additional information that helps to refine their estimate of the resources necessary to complete the program. A possible source of this information is a prototype or technology demonstrator.[1] A key attribute of a prototype program is its possible existence outside a formal RDT&E program; the information gained can then be used to make the riskier decision on moving ahead with the program.[2] However, prototype development itself can be costly, so the decision-maker would prefer to know in advance the value of the information that would be provided by the prototype effort.

This paper is intended to address this wish, providing a method to value the prototype effort using the decrease in uncertainty it would provide to the cost estimate. To employ the method, the decision-maker needs only an initial cost estimate (presented as a

---

[1] Hereafter, we refer to both prototypes and technology demonstrators as simply *prototypes* because we are only interested in their role as information augmenter. Indeed, any program or project that can reduce the uncertainty of a cost estimate is a good candidate for the method herein.

[2] Even if the prototype occurs as part of RDT&E, it still may offer a natural point for reassessment, depending on the general shape of the program.

distribution of possible costs), a cost constraint for the program, and some estimate of the decrease in uncertainty the prototype effort would provide. We introduce the method through increasingly complex illustrations of its application, first employing Value of Information (VoI) theory to establish a method for comparing the prototype/no prototype options, and then using properties of the *preposterior distribution* (defined below) to compare the values. We end the paper by using the method to create a rule of thumb, developed from historical data, for the value of a typical prototype.

Illustrations of Method

In this paper, we are concerned with the particular application to determining prototype value in the case of defense acquisition. We use language that reflects this focus, even though much of the work below is more broadly applicable. We also note that this paper presents a nontechnical description of the method; some detail and mathematics are omitted and can be found in the references.

Framework and Terminology

Before a defense acquisition program begins, a decision-maker is presented some estimate of the final cost[3] of the program. This estimate will be a probability distribution of possible final costs; we call this the prior cost distribution and denote the final cost subject to it as the random variable *C*. The decision-maker also has some cost constraint *b* for the overall program, presumably related to the relative importance of the associated mission and the priority of the program within its portfolio. The cost constraint *b* can be interpreted in several ways. It might simply be the maximum amount the decision-maker is willing to spend to address the mission. More realistically, it might be the known cost of a lower-risk alternative that addresses the same mission need, or the point at which the opportunity cost of choosing the program over a lower-risk, less capable system is greater than the capabilities gained.[4] For simplicity, we assume the cost constraint to be the cost of satisfying the mission the program addresses in some other way.

The cost constraint informs the decision directly: If the final cost of the program is expected to exceed *b*, i.e., $C > b$, the decision-maker should not embark on the program; otherwise, when $C \leq b$, they should. Of course, the final cost *C* has some probability of being greater than *b* and some probability of being less, but the decision-maker cannot make a fraction of a decision. For simplicity, we assume the decision-maker relies on the expected value[5] of $C, \mathrm{E}[C]$, to make their decision.[6] That is, the program will only be considered if E[C] ≤ b, as illustrated in Figure 1.

---

[3] While by no means necessary, it is helpful to think of this final cost as the cost of the associated RDT&E program, because the relationship between RDT&E and a prototype is more direct.
[4] As may be inferred from the interpretations, it is possible that the cost constraint itself is a probability distribution of costs. This is a possible extension and we will discuss letting *b* vary. However, for simplicity, we assume the cost constraint is a single value.
[5] Recall that the expected value of a random variable is the average value resulting from (infinite) repeated sampling of the random variable. This is calculated as the sum of all possible values of the random variable weighted by their associated probabilities.
[6] Other constraints may be used, such as $Pr(C \leq b)$ for some probability p. However, as will be seen, to do so requires assumptions about the shape of the resulting cost distributions that the expected value assumption does not.

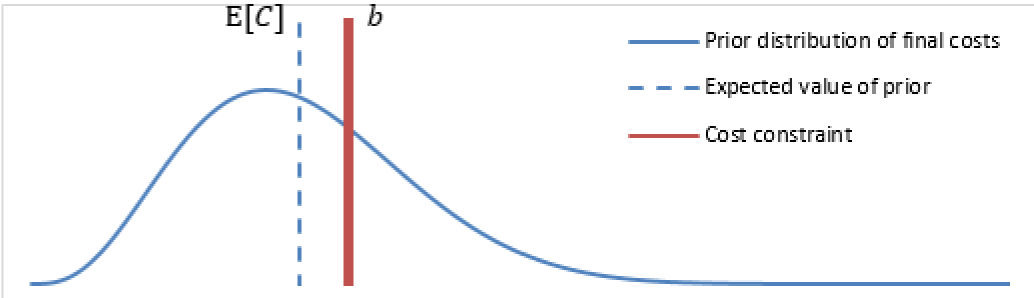


Figure 1. Example of a Program That Should Proceed

We introduce the option of obtaining more information about the distribution of final costs through prototype development. The decision-maker now has three choices:[7]

1. Start the program (with full intention to complete it);
2. Pursue the alternative program at cost *b*; or
3. Invest in a prototype, after which another decision will be made to start or not start the program.

The decision-maker can choose between the first two options using the cost constraint as discriminator. However, the closer in value E[*C*] and *b* are, the greater the probability that *b* lies between E[*C*] and the actual observed final cost, meaning the decision was not optimal. As these values approach each other, the decision-maker then loses confidence in their decision and becomes more interested in obtaining more information. At some point, they instead choose the third option.

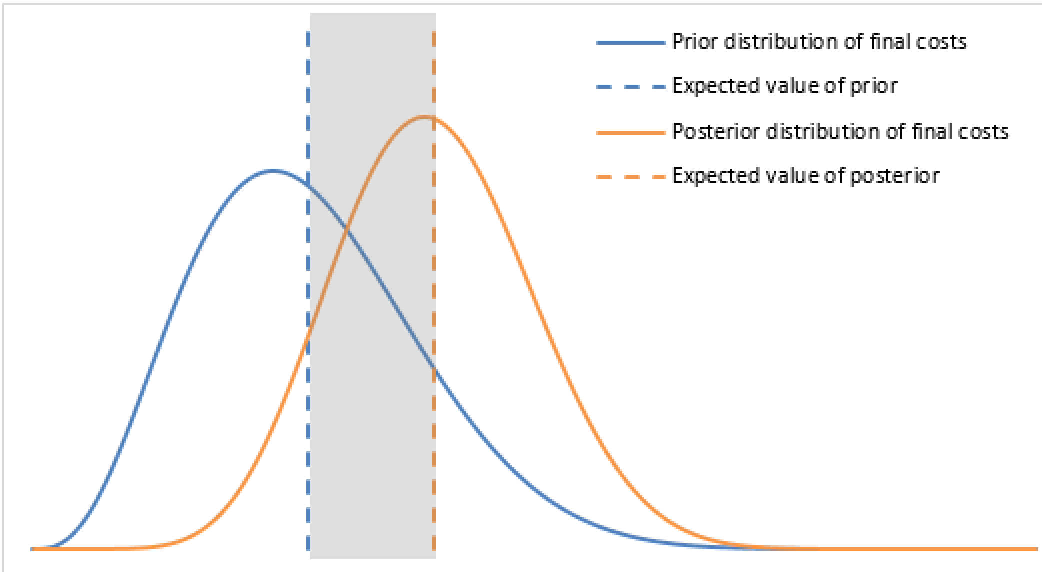
The third option will, at some investment cost *I* of the prototype program, provide the decision-maker with a new probability distribution of possible final program costs. We call this a *posterior cost distribution* and denote the final (uncertain) program cost subject to it by the random variable *D*, adding a subscript when more than one posterior cost distribution is referenced. We assume *D* is the remaining cost of the program after completing the prototype, i.e., the cost *I* is not reflected in *D*. After a prototype program is pursued, the decision-maker will make the decision to continue the main program similarly to the one described previously, but now comparing E[*D*] and $b - I$.

Note that before the prototype is developed, its resulting *D* is unknown and, indeed, many different posterior cost distributions D_i might arise. Each of these distributions has some probability of occurring, so we can consider the probability distribution over posterior cost distributions, to which we associate the random variable Δ. Abusing notation a bit, we can say the values of Δ are all the possible random variables D_i. We will actually be more interested in the probability distribution of the expected values of $\mathrm{E}[D_i]$ of posterior cost distributions. We associate the random variable $\bar{\Delta}$ with this distribution of means.

The decision-maker is interested in pursuing a prototype only when the result is likely to change their decision. If their decision is not changed, they stay on the same path, as

[7] Of course, specific real world situations are more complicated, requiring some adjustment of the model.

initially they gain nothing more than confirmation and the cost *I* of the prototype is wasted.[8] For example, consider Figure 2. While the decision-maker does not know what the posterior cost distribution will be, they do know it is likely to have an expected value somewhere in the thick part of the prior cost distribution (blue line). Otherwise, we could not consider the prior cost distribution to be an accurate reflection of our prior knowledge, and should change it accordingly. The example posterior cost distribution (orange line) given in the figure is plausible, though less likely than one in the thicker part of the distribution.



Note. This decision changes with the added information of the posterior distribution whenever the cost constraint *b* lies in the shaded region.

Figure 2. When the Posterior Cost Distribution Is Useful

Suppose the constraint *b* is very high, far to the right in the chart. Here, the decision-maker would start the program, almost without question. If the initial distribution *C* is accurate, it would be very unlikely that a prototype program would reveal a new expected value above that constraint, so a prototype is unlikely to change the decision. In this case, the value of a prototype is very low. Similarly, if *b* is very low, far to the left in the chart, the prototype is very unlikely to change the decision to not start the program. In both cases, for the specific case of *D* shown in the chart, the value of the prototype is 0.

On the other hand, when the constraint *b* more evenly splits the prior cost distribution, it is more likely that a prototype could change the *a priori* decision. In Figure 2, if *b* took any value in the shaded region, the decision would change after that particular prototype result. We can quantify the expected value of the prototype effort using the difference in the expected costs to the decision-maker between making a prototype or not.

[8] It is likely that the program will benefit from prototype spending if the program is continued, so the entire cost of the prototype may not be lost, but we do not address this directly. The benefit from the prototype is assumed to be captured in the increased fidelity of the cost distribution.

For this calculation to be intuitive, we need to look at all possible posterior cost distributions together, using VoI theory.

Doing so gives us a distribution of values for the prototype determined by Δ, *C*, and the constraint *b*. We call the expected value of this distribution the value of the prototype. We will assume *C* and *b* are given, and use the *preposterior distribution* of program costs together with assumptions about the amount of information a prototype provides, in order to determine this value.

We make the following assumptions:

- Decisions regarding the continuation of the program are based on expected values.
- The distribution of final costs *C* is reliable, i.e., unbiased with unbiased estimation of standard deviation.
- Any uncertainty in the cost of the prototype program is small compared to the cost of the entire program, and thus can be treated as a point value.
- A firm affordability cap for the program is known and expressible as a constant bound.
- Modifications can be made to the method to relax these requirements, but this is not addressed here.

One assumption requires specific attention. The requirement that the distribution of final costs is known and accurate is strong. For example, this implies the program is technically feasible and will, if attempted, succeed. Additionally, there is historical evidence that initial cost estimates are systematically low. (See the section titled A General Application for some evidence of this.) Even if the expected value of the estimate is correct, the variance (risk) is not all known and can therefore be vastly underestimated. We have not tested the robustness of this method against inaccurate initial cost estimates, although we hope to in the future. For that reason, any application requires great care and the most conservative of initial cost estimates.

Value of Information (VoI)

VoI is a concept from decision analysis. In its simplest form, the theory examines two possibilities, one in which a decision is made without a certain piece of information and one in which it is made with the information. The decisions have costs and benefits, depending on some unknown state of the future, which the information describes in some way. The difference between the expected values of these decisions is the value of the information.

A Simple Illustration

For illustration, suppose the decision-maker has the information that a program has a 40% chance of costing $9 billion ($9B), and a 60% chance of costing $14 billion ($14B). The expected cost of the program is thus $0.4 * \$9\text{B} + 0.6 * \$14\text{B} = \$12\text{B}$. If their cost constraint *b* is $10 billion, they will not initiate with the program as it stands.

However, they are then told that a prototype effort would give them the information they need to determine whether the full program is a $9 billion or $14 billion program. How much should they be willing to spend on the prototype?

Recall that the cost constraint of $10 billion is the value of satisfying the mission, which, for purposes of explication, we assume to be the cost of satisfying the mission in some other way. If the prototype is not pursued, the decision-maker will pursue the $10 billion alternative, and so the expected cost of meeting the mission need is $10 billion.

If the prototype is pursued, there are two possibilities. The prototype will reveal—40% of the time—that the program will cost $9 billion. The decision-maker will then spend $9 billion to execute the program. The remaining 60% of the time, the prototype will reveal a cost of $14 billion. The decision-maker will then stick with the $10 billion alternative. Therefore, when the prototype is pursued, the expected cost of satisfying the mission is $0.4 * \$9\text{B} + 0.6 * \$10\text{B} = \$9.6\text{B}$.

By VoI, we conclude the value of the prototype is $\$10\text{B} - \$9.6\text{B} = \$0.4\text{B}$ to the decision-maker, the difference between the two decision paths. That is, the decision-maker should be willing to spend up to $0.4 billion dollars on a prototype that can provide this level of information.

This illustration, however, is far from realistic. The distribution of final costs *C* is not, as a rule, discrete, and the prototype effort will typically not provide perfect information on the final cost. These observations can be accommodated with a small amount of generalization, as we will see in the next illustration. However, we will assume for now that we know what the distributions in Δ look like. This assumption will be addressed in the next section.

Illustration of More Complexity

Now suppose the decision-maker is presented with any distribution of final costs *C* such that E[*C*]=$11 billion. Suppose further that the prototype will demonstrate two critical technologies, A and B, and each technology will be determined to be easy or difficult to mature. Therefore, from the prototype, four new cost distributions can result. Letting the random value indicate its distribution, we get Δ = { D_1, D_2, D_3, D_4}, where

D_1 corresponds to both technologies proving easy to mature,

D_2 corresponds to A being easy and B being difficult,

D_3 corresponds to A being difficult and B being easy, and

D_4 corresponds to both being difficult.

Prior to the prototype development, the conditional expected cost under each case was determined, and the probabilities of A and B being easy to mature are found to be 0.6 and 0.1, respectively, resulting in Table 1.

Table 1. Properties of the Posterior Cost Distributions in an Illustration of More Complexity

Distribution	Probability of Distribution	$E[C\|D_i]$
D_1	0.06	$8 billion
D_2	0.54	$9 billion
D_3	0.04	$13 billion
D_4	0.36	$14 billion

Given the cost constraint *b* of $10 billion, the decision-maker can now compute the value of the prototype. If they decide to not pursue the prototype, the mission cost is $10 billion, as before. With the prototype, note that they will pursue the program if distributions D_1 or D_2result, and cancel the program if distributions D_3 or D_4 result. The expected mission cost given a prototype effort is thus

$$0.6 * \$8B + 0.54 * \$9B + 0.04 * \$10B + 0.36 * \$10B = \$9.34B.$$

Therefore, the value of the information from the prototype is, in billions of dollars,

$$\$10B - \$9.34B = \$0.66B.$$

We note here that it is straightforward to apply sensitivity analysis to the cost constraint. In Figure 3, we see what happens as the cost constraint varies. The closer the constraint is to the $\mathrm{E}[C]$of \$11 billion, the greater the value of the prototype, because the constraint is more likely to be between $\mathrm{E}[C]$ and $\mathrm{E}[D]$, for any given posterior cost distribution of final costs *D*. We also see that as the cost constraint approaches the extremes, the prototype has no value, because the decision-maker would not change their decision from that based on *C* alone. The values in this figure together with the distribution of the cost constraint give the expected value of the prototype.

The above illustration is instructive, but it is unrealistic in one important way. In general, we will know what the distributions in Δ look like or how they are distributed, i.e., what their various probabilities of occurrence are. However, given our assumptions about how the decision will be made, we really only need to know the distribution of the *means* of the distributions in Δ. The next section will illustrate this.

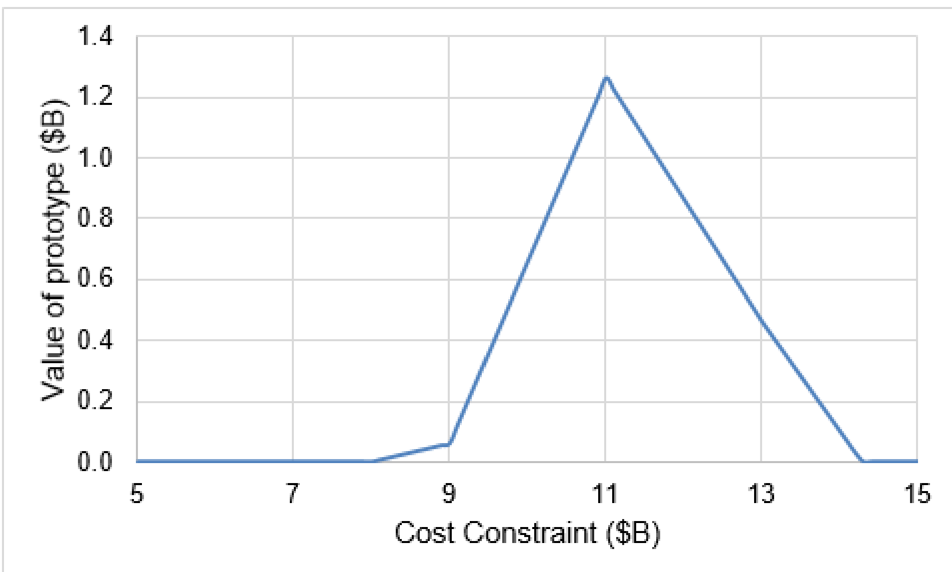


Figure 3. Value of Prototype as the Cost Constraint Varies

Preposterior Distribution

Our main goal in this section is to describe the distribution of means of posterior cost distributions $\bar{\Delta}$ which we call the *preposterior* cost distribution. (The initial distribution of final costs *C* is the *prior* cost distribution and a posterior cost distribution of final costs *D*, when known, is the *posterior* cost distribution.) Using this notation in application to an illustration of more complexity (discussed in the previous section), we have, in billions of dollars,

$$E(\bar{\Delta}) = 0.6 * \$8B + 0.54 * \$9B + 0.04 * \$13B + 0.36 * \$14.3B = \$11B.$$

Before we present the properties technically, we examine them intuitively. This section is largely derived from Section 5.4 of Raiffa and Schlaifer (1961); any additional justifications may be found there.

Creating a prototype provides information about the technical hardship of the program, which should have some effect on the prior cost distribution of *C* of that program. There is no reason, *a priori*, to expect this new information will increase or decrease the mean of the initial estimate; if there were, the initial estimate should be adjusted accordingly. We emphasize here that a specific posterior cost estimate might have a different mean from the prior cost distribution, but the mean of all possible posterior estimates must match that of the prior estimate if it is reliable. That is, the expected value of $\bar{\Delta}$ is the same as that of *C*, because *C* summarizes all of the information we have about what $\bar{\Delta}$ could look like.

The new information gleaned from the prototype will also produce a tighter cost distribution (lower variance) than that of the prior cost distribution.[9] Let us consider the extreme possibilities to gain intuition for this tightening. If the prototype provides no new information at all, any value of Δ is just *C*—the posterior is the same as the prior. So $\bar{\Delta}$ would consist of a single value, E[*C*], and clearly E [$\bar{\Delta}$]=E[*C*]. Thus, the variance of $\bar{\Delta}$ would be 0 and the variance of every possible value Δ takes on (every *D*) is equal to the variance of *C*.

On the other hand, if the prototype provides perfect information (i.e., tells us the exact cost of the program), the prototype gives the same information as actually doing the program. Any value *D* of Δ is the final cost. It is a degenerate distribution with all mass at that final cost, and these values must be random draws of *C*, since the prior cost distribution is accurate. The expected values of these *D* are the final cost, so $\bar{\Delta}$ consists of expected final costs taken from the prior cost distribution, i.e., $\bar{\Delta}$=*C*. Clearly, then, E [$\bar{\Delta}$]=E[*C*] and the variance of $\bar{\Delta}$ and *C* are equal. The variance of the degenerate distributions Δ definitionally have variance 0.

Taken together, we see that as the variance of the resulting distributions decreases (i.e., as the prototype provides more information), the variance of $\bar{\Delta}$ increases and is bounded by that of the variance of *C*. Indeed, the more information provided, the more $\bar{\Delta}$ looks like *C* as a distribution.

The observations above can be summarized with the following three theorems. Stating these theorems exactly requires some technical complexity beyond the scope of this paper. See Raiffa and Schlaifer (1961) for proofs.

> *Theorem 1. The expected value of the means of new cost distributions resulting from the information gained by a prototype is equal to the expected value of the prior cost distribution, i.e.,*
>
> $$E(\bar{\Delta}) = E(C)$$
>
> *Theorem 2. The variance of the means of new cost distributions resulting from the information gained by a prototype is equal to the variance of the prior cost distribution less the expected value of the variances of the new cost distributions, i.e.,*

---

[9] Intuitively, this makes sense; the more the decision-maker knows, the more specific they can make their estimate. However, this fact actually relies on the prior cost distribution being accurate, meaning no new information can be added that adds new possibilities. If something has not been thought of, this is captured by a wider variance and new information will only discount some of those possible outcomes.

$$Var(\bar{\Delta}) = Var(C) - E(Var(D))$$

Theorem 3. As the information provided by the prototype approaches perfect information, the distribution of $\bar{\Delta}$ approaches the distribution of C, i.e.,

$$\lim_{n \to \infty} Pr(\bar{\Delta} < d) = Pr(C < d)$$

where n is the amount of information.

The third theorem depends on our assumption that *C* is a reliable distribution, unbiased in mean and standard deviation. It also assumes that we can always add information that will sharpen our estimate. The second assumption is true because we can always run the experiment of completing the program, at which point the cost will be known.

The open problem at the end of the previous section was the shape of the distribution Δ, without which our VoI calculation would be impossible. While we are not able to describe this distribution completely, we now know some key attributes that can help us estimate the value of the prototype.

Putting It Together

For our final illustration, assume the decision-maker has a particular goal in mind for their prototype. Suppose they have a cost estimating relationship (CER) based, in part, on a particular technical attribute that is not precisely known, and their prototype will determine this attribute nearly perfectly.[10] Using the CER, they use Monte Carlo simulation with multiple possible values of the technical attribute (and the standard error of the CER) to estimate the cost distribution *C* of their program. For this example, assume that *C* has a scaled beta distribution with support on (0, $40B), mean $11B, and variance 40.[11] They then apply the CER to specific values of the technical attribute and find that the variance of the posterior, on average, is 10.[12]

From Theorem 1, the decision-maker knows the expected value of $\bar{\Delta}$ to be $11 billion. From Theorem 2, they can compute the variance of $\bar{\Delta}$ to be 30, the difference of the variance of *C* and the mean of variances of elements of $\bar{\Delta}$. Finally, from Theorem 3, lacking other information, they assume that $\bar{\Delta}$ is the same type of distribution as *C.* They can now estimate the value of the prototype.

Again, we assume the cost constraint is $10 billion. Without the prototype, they do not proceed with the program, so the cost of satisfying the mission is $10 billion. With the prototype, they consider two situations. In the first, the prototype reveals a cost less than the constraint. They then sum these costs, weighted by the probability they occur (given by the distribution of $\bar{\Delta}$), i.e., they calculate

$$\int_0^{10} x \, \mathrm{f}(\mathrm{x}) dx = 3.1,$$

---

[10] For example, models may predict a hypothetical aircraft design has a lift-to-drag ratio in a certain range, but the exact ratio is not known. A prototype may demonstrate the actual lift-to-drag ratio.

[11] The correct units for variances here is squared billion dollars. We omit the units for clarity.

[12] In this situation, the decision-maker could just generate some representative subsample of Δ and apply VoI directly. We have simplified this example from one in which the reduction in variance after the prototype can be estimated but the distributions are not known, for purposes of illustration.

where $f(x)$ is the probability density function of Δ. In the second, the prototype reveals a cost greater than the constraint, so in each case, the cost of satisfying the mission is $10 billion. Thus, they calculate

$$\int_{10}^{\infty} 10\ \mathrm{f(x)}dx = 5.2.$$

The sum of these is the expected cost of satisfying the mission after the prototype. In our case, the expected cost is $8.3 billion. Thus, the decision-maker should be willing to pay up to $1.7 billion for the prototype.

We again consider what happens when we vary the cost constraint in Figure 4. As expected, the value of the prototype is greatest when the constraint is near the expected value of the prior cost distribution.

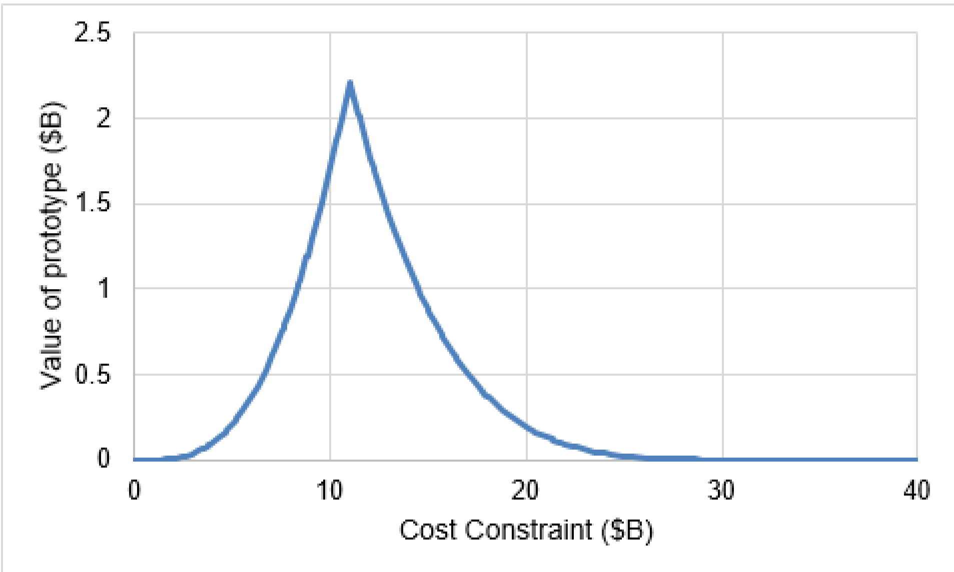


Figure 4. Value of Prototype as the Cost Constraint Varies

A General Application

Usually, the decision-maker does not have access to the decrease in variance provided by a prototype, so it is useful to develop a rule of thumb for the general case. To do so, we use historical cost overruns of development programs[13] with and without prototypes to estimate the change in variance. We take our data from Tyson, Nelson, Gogerty, Harmon, and Salerno (1991), which examined the development costs of 51 historical aircraft and tactical munition programs, 35 of which did not have prototypes and 16 of which did. The report found that, on average, final development costs for programs without prototypes was 1.62 ± 0.96 the cost of their initial cost estimate, whereas programs with prototypes were 1.17 ± 0.17 their initial cost estimate.

[13] We look at development programs here, not the total program, mainly due to data limitations. The effect of the prototype program on production and sustainment also needs consideration.

While the cost overruns are important, our goal is simply to compare the variances to find the value of a prototype. Suppose we have a new (possible) program. We treat the set of programs without prototypes, their costs scaled to their initial estimates to make comparable, as describing its initial cost estimate distribution of *C.* In doing this, we are treating these programs as observations of possible outcomes of our hypothetical program. Note this does not imply we expect this program to overrun by 62%; we are using this database as a reasonable proxy for an accurate prior cost distribution. We can choose our own expected value, so for simplicity, we assume E[*C*] is 1 of some unit. Scaling the variance accordingly, we find this distribution has a variance of 0.34.

We then treat the set of programs with prototypes as a sample of the set $\bar{\Delta}$ of possible means of distributions resulting from a prototype. Because $\bar{\Delta}$ has the same mean as *C*, we normalize these data to 1 as well and find $\bar{\Delta}$ has a variance of 0.02.

Using the above, we can now estimate the value of a prototype for different cost constraints. We assume a beta distribution with expected value of 1 and respective variances for *C* and $\bar{\Delta}$. (We note that the distribution of the data is well approximated by the beta distribution.) In Figure 5, we present the value assuming the expected value of *C* is 1. In Figure 6, we present the value of the prototype as a percent of cost constraint. Note that while the figures are similar, Figure 6 shows that the value of the prototype is relatively higher for smaller cost constraints.

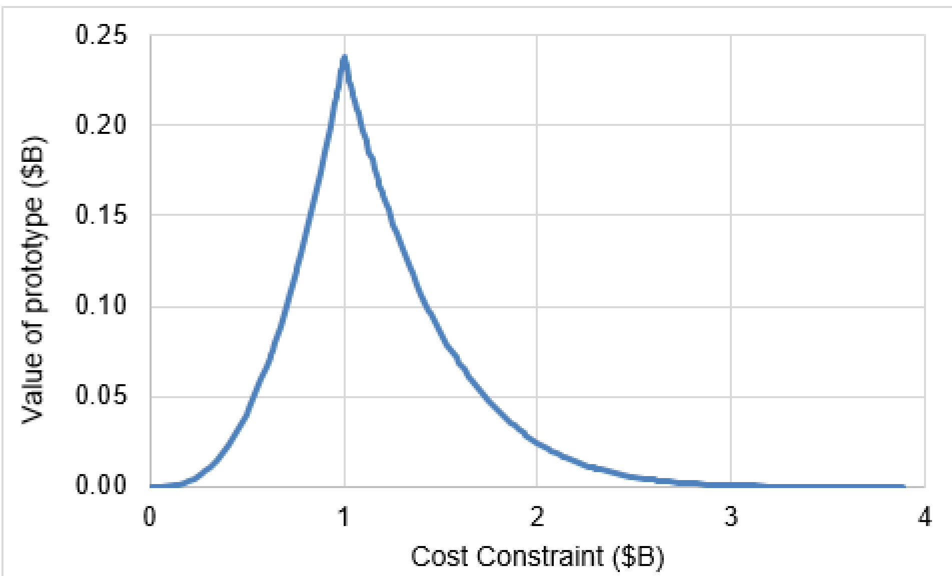


Figure 5. The Value of Prototype for a $1 Billion Development Program, Using Historical Development Cost Distributions

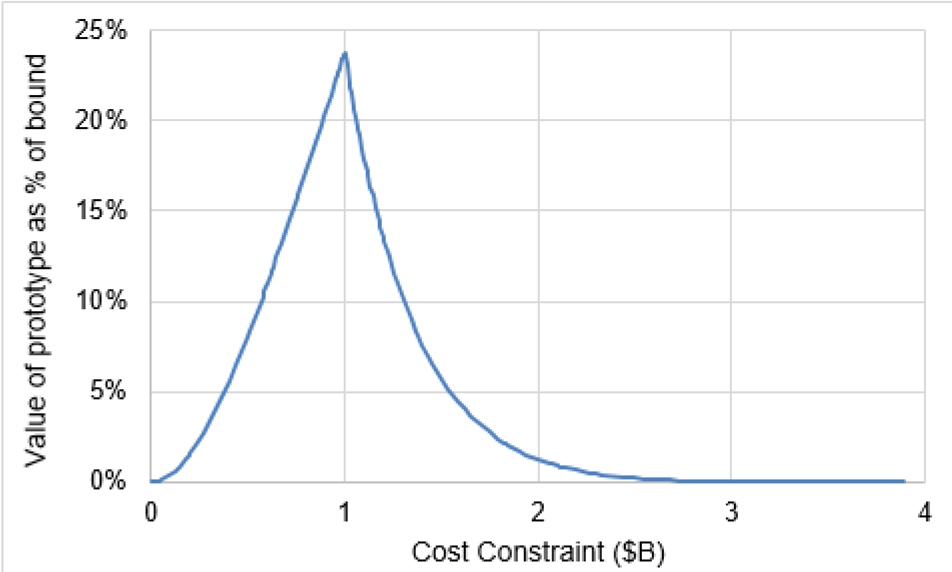


Figure 6. The Value of the Prototype Relative to the Cost Constraint, Using Historical Cost Distributions

A possible problem with the above formulation is that we only consider programs that successfully start and do not experience cancellation. The underlying data can be improved by adding failed programs and should be addressed in future research. Also, while it is safe to assume that none of the programs in the historical data benefitted from this research, it is unlikely that the decision of whether or not to build a prototype was random; that decision process could introduce a bias. On the other hand, it is possible that the decision on building a prototype was primarily determined by the political situation at the time and may therefore be entirely independent of the technical merits, in which case the sample would be unbiased.

Conclusion

This work needs expansion in several directions. Our immediate interest is in determining the effect of a biased prior. That is, assuming the distribution of C is unbiased in mean and standard deviation is a very strong assumption. When that assumption is violated, how much is the final value of the prototype affected? To this end, we also want to incorporate more historical data, particularly the actual cost of prototype programs, to test and refine the method in less than perfect cases.

We also would like to allow the decision-maker to use values other than the expected value for decisions. We do not see this as impossible, but it may be technically difficult and rely even more on the assumption that the distribution of $\bar{\Delta}$ is similar in shape to the distribution of C. Other directions of research should allow more generality in the method, such as allowing the cost constraint to also come from a distribution. Specific application to commodity type should also be studied to find likely reductions in cost variance due to prototype development. As always, additional historical data would also be a benefit.

The method presented above is a relatively simple way of determining the value of a prototype. The assumptions, while strong, are not unreasonable. In general, if we cannot assume that an initial cost estimate distribution is reliable, we should adjust the estimate to reflect our uncertainties. Estimating the amount of variance reduction provided by a prototype program is likely possible in cases where the prototype is geared to answering specific technical questions, which is most common. Indeed, the effect on the variance may provide a guide to what goals the prototype should have. In the cases where this variance reduction is not estimable, we can use historical programs to estimate the variance reduction. As shown above, the value of a prototype can reach more than 20% of the value of the entire program in those cases where the affordability of the program is most uncertain.

The author has developed several models in the course of this work, using both Excel and Python, that are available upon request.

References

Raiffa, H., & Schlaifer, R. (1961). *Applied statistical decision theory*. Cambridge, MA: MIT Press.

Tyson, K. W., Nelson, J. R., Gogerty, D. C., Harmon, B. R., & Salerno, A. W. (1991). *Prototyping defense systems* (IDA Document D-1097). Alexandria, VA: Institute for Defense Analyses.

Acknowledgments

The research for this paper was funded by the Air Force Research Laboratories. We thank them for their support.

Panel 8. Cultivating the Future Acquisition Workforce

Wednesday, April 26, 2017	
3:30 p.m. – 5:00 p.m.	**William Mark Deskins,** Director, Acquisition Career Management, ASN (RDA) ***Army Contracting Workforce Organizational Climate Assessment*** Rene G. Rendon, Naval Postgraduate School Edward (Ned) H. Powley, Naval Postgraduate School ***Understanding Defense Acquisition Workforce Challenges*** Colleen Murphy, The MITRE Corporation Adam Bouffard, The MITRE Corporation ***2016 Assessment of the Civilian Acquisition Workforce Personnel Demonstration Project*** Jennifer Lamping Lewis, RAND Corporation Laura Werber, RAND Corporation Cameron Wright, RAND Corporation Irina Danescu, RAND Corporation Jessica Hwang, RAND Corporation Lindsay Daugherty, RAND Corporation

William Mark Deskins—rejoined the ASN (RDA) staff in May 2015 to serve as the Director, Acquisition Career Management (DACM), the Navy and Marine Corps' lead for the professional development and management of the DoN acquisition workforce.

With over 25 years of professional experience, Deskins's spectrum of experience includes private industry, Navy field, headquarters, and program office assignments. From February 2009 to May 2015, Deskins was the Deputy Program Manager for Strategic and Theater Sealift (PMS 385), which included two ACAT programs: the Joint High Speed Vessel (JHSV) and the Mobile Landing Platform (MLP). During his tenure both programs went from Milestone B to Initial Operating Capability, delivering five JHSVs and two MLPs. Prior to joining PMS 385, Deskins was the Chief of Staff for Team Ships and was responsible for front office operations and executive coordination and oversight of the vast array of issues related to surface ship acquisition, maintenance, modernization, and disposal. From 2004 to 2007, he was a member of ASN (RD&A) staff and was responsible for executive oversight of issues related to surface ship, submarine and carrier maintenance, modernization, and disposal.

From 2001 to 2007, Deskins was the Deputy Program Manager for Inactive Ships (PMS 333) in PEO (Ships), where he was responsible for the planning, programming, budgeting, and execution of the U.S. Navy's inactivation and disposal of conventionally powered surface ships. As the last stage of the acquisition life cycle, he oversaw the decommissioning of 20 ships and the disposal of 46 ships through foreign military sales (six), scrapping and recycling (12), fleet support through SINKEX (26), artificial reefing (one) and the donation of the aircraft carrier ex-MIDWAY that operates today as a highly successful ship museum in San Diego.

Deskins's previous jobs include working five years as an industrial engineer in private industry in North Carolina and beginning his government career as an industrial engineer at the Indian Head

Division of the Naval Surface Warfare Center (NSWC). Promotions brought him to the NSWC headquarters staff, where he worked strategic planning and corporate operations. He was selected for NAVSEA's Commander's Development Program where he held highly responsible positions working on combat systems, strategic and business planning at the Division and NAVSEA corporate level, and PEO Program Management assignments.

His awards include two Navy Superior Civilian Service Awards in 2004 and 2014, two Navy Meritorious Civilian Service Awards in 1994 and 2000, and several Special Act Awards and Outstanding Performance Awards. He is an acquisition professional, DAWIA Level III certified in Program Management, and has held memberships in the International Council on Systems Engineering and the Institute of Industrial Engineers.

Deskins's education includes a bachelor's in industrial engineering from West Virginia University and a master's in technology management from the University of Maryland University College. His executive education includes Harvard Business School and University of North Carolina.

Army Contracting Workforce Organizational Climate Assessment

Rene G. Rendon—is an Associate Professor at the Graduate School of Business and Public Policy, NPS, where he teaches defense acquisition and contract management courses. He also serves as the Academic Associate for the MBA specialization in contract management. Prior to joining the NPS faculty, he served for over 20 years as an acquisition contracting officer in the United States Air Force. His career included assignments as a contracting officer for the Peacekeeper ICBM, Maverick Missile, and the F-22 Raptor. He was also a contracting squadron commander and the director of contracting for the Space Based Infrared Satellite program and the Evolved Expendable Launch Vehicle rocket program. Rene has published in the Journal of Public Procurement, the *Journal of Contract Management*, the *Journal of Purchasing and Supply Management*, and the *International Journal of Procurement Management.* [rgrendon@nps.edu]

Edward (Ned) H. Powley—is an Associate Professor of Management in the Graduate School of Business and Public Policy at NPS. [ehpowley@nps.edu]

Abstract

The DoD obligated approximately $273.5 billion in contracts for major weapon systems, supplies, and services in fiscal year 2015. The DoD contracting workforce professionals are responsible for managing the millions of contract actions for the procurement of critical supplies and services, ranging from commercial-type supplies, professional and administrative services, highly complex information technology systems, and major defense weapon systems. The DoD's organizational climate is a significant contributor to the success of the contracting workforce. An analysis of an organization's climate and its various components can provide its leadership with a road map for developing a healthier climate, and thus improve performance.

The purpose of this study is to conduct an organizational climate assessment of the Army contracting workforce. Using a web-based survey, we assessed the Army contracting workforce on the various components of organizational climate. Based on the number of survey responses and response rate, we used quantitative data analysis methods to analyze the survey data and identify research findings. This research benefits the Army by establishing a baseline climate of the Army's contracting workforce. It also identifies the dimensions that need to be addressed in order to improve the Army's contracting organizational climate. These research findings can then guide the DoD, as well as the federal government contracting community, in developing a road map for improving its contracting organizational climate.

Research Approach

In coordination with the Army Deputy Assistant Secretary (Procurement), we developed the survey instrument and deployed the survey to the Army contracting workforce. We developed the survey on the NPS Lime Survey system and provide the survey link to the Army Deputy Assistant Secretary (Procurement) for deployment throughout the Army contracting workforce.

Based on the number of survey responses and response rate, we used appropriate quantitative data analysis methods to analyze the survey data and identify research findings.

This study will provide a baseline measurement of the Army's contracting organizational climate and address the following research questions:

1. What is the baseline climate of the Army's contracting workforce in relation to the following dimensions: work relationships, employee recognition,

employee commitment, supervision, leadership, job satisfaction, organizational commitment, employee characteristics, and job stress.

2. Is a change in the Army's contracting organizational climate necessary?
3. What dimensions need to be addressed in order to improve the Army's contracting organizational climate?

Benefits of Research

This research will benefit the Army by establishing a baseline climate of the Army's contracting workforce in relation to the following dimensions: work relationships, employee recognition, employee commitment, supervision, leadership, job satisfaction, organizational commitment, employee characteristics, and job stress. It will also identify the dimensions that need to be addressed in order to improve the Army's contracting workforce climate. These research findings can then guide the DoD, as well as the federal government contracting community, in developing a road map for increasing its contracting workforce climate.

Organizational Climate

As noted above, we examined a number of dimensions that, as a whole, are indicative of organizational climate and culture and have been used in similar settings (Gerbich, 2017; Doelling, 2005). Table 1 outlines the key dimensions captured in this study, a brief description of the construct, and sample scale items (see also McKeithen, 2016).

Table 1. Organizational Climate Dimensions

Dimension	Description	Sample Items
Job Satisfaction	Employee's affective attachment to a job; involves both extrinsic and intrinsic features to job (Cook et al., 1981)	• Considering your skills and the effort you put into your work, how satisfied are you with your pay? • How satisfied do you feel with your chance for getting ahead in this organization in the future?
Supervisor-Related Commitment	Employee's commitment to supervisor (manager) and internalization of supervisor's values (Becker et al., 1996)	• When someone criticizes my supervisor, it feels like a personal insult. • My supervisor's successes are my successes.
Job Role Ambiguity	Provides employee clear set of responsibilities so managers are able to give appropriate guidance and hold individuals accountable (Fields, 2002)	• I know what is the best way (approach) to go about getting my work done. • My job is such that I know when I should be doing a given work activity.
Job Characteristics	Characteristics of jobs that increase internal motivation and for which an employee has some level of control (Fields, 2002; Wayne, Shore, & Liden, 1997)	• In the positions that I have held at my current work center, I have often been assigned projects that have enabled me to develop and strengthen new skills. • Besides formal training and development opportunities, to what extent have your managers helped to develop your skills by providing you with challenging job assignments?
Job Stress	Aspects of job affecting employees' stress levels and undesirable constraints and demands (Davey, Kinicki, & Scheck, 1995; Fields, 2002)	• My supervisor places demands on me that aren't placed on coworkers. • Personal concerns have interfered with my job performance.
Work-Family Conflict	Inter-role conflict between work and family that are mutually incompatible; demands from one increase conflict in the other (Thomas & Ganster, 1995)	• After work, I come home too tired to do some of the things I'd like to do. • On the job, I have so much work that it takes away from my other interests.
Commute Stress and Safety	Cognitive and affective assessment of stress incurred due to employee's commute to and from work (Kluger, 1998; Fields, 2002)	• I resent the hassles my commute causes me. • My commute affects my productivity on the job in the following ways: It takes work time out of my day.
Organizational Justice	Employees' perceptions of fairness of procedures, outcomes, and information sharing; and interactions in the workplace (Dulebohn & Ferris, 1999)	• The supervisor considered the important aspects of your work when rating you. • The supervisor rated you on how well you did your job, not on his/her personal opinion of you.
Job Fit	Employees' perceived ability to control and meet job demands (Xie, 1996)	• My job gives me a chance to do the things I feel I do best. • I feel that my job and I are well matched.
Workplace Values	Employees' perceptions about the level of importance an organization places on values such as quality, innovation, cooperation, and so forth (Van Dyne, Graham, & Dienesch, 1994)	• Individual employees recognized and rewarded for superior performance. • Reputation for innovation surpasses Army contracting agencies. • Procedures facilitate widespread participation in decision-making.
High Quality Relationships	Employees' view of the quality of connections and relationships in the workplace (Carmeli & Gittell, 2009; Dutton, 2003). HQCs are highly correlated with job satisfaction, team learning, other measures of organizational effectiveness.	• Whenever anyone at work expresses an unpleasant feeling, she/he always does so in a constructive manner. • We cope well with the pressures experienced at work. • We are attentive to new opportunities that can make our system more efficient and effective. • I feel that my co-workers and I try to develop meaningful relationships with one another. • There is a sense of empathy among my co-workers and myself.

Methodology

We designed the survey to capture professionals' perceptions of climate, broadly speaking. The survey incorporates the dimensions outlined previously. We used a standard 7-point Likert scale where 1 = Strongly Disagree, 2 = Somewhat Disagree, 3 = Disagree, 4 = Neither Agree nor Disagree, 5 = Agree, 6 = Somewhat Agree, and 7 = Strongly Agree. There were a total of 136 items and four additional open-ended response questions at different points in the survey. These open-ended questions afforded respondents the opportunity to offer written comments and feedback for improvement on certain dimensions.

The survey was administered using NPS's Lime Survey tool. The structure of the survey comprised four sections, which included several scales based on the dimensions indicated: (1) job satisfaction, perceptions of supervisors, job role ambiguity, and job characteristics; (2) job stress—personal and work related stress, work-family conflict, and commute stress and safety; (3) organizational justice, job fit, workplace values, and high quality relationships; and (4) job-related demographics (certification levels, organization type, and so forth). Personally identifiable information was not collected from participants.

Approval for the survey was obtained through the Deputy Assistant Secretary of the Army (Procurement) and was launched by the Workforce Development Directorate Office of the Deputy Assistant Secretary of the Army (Procurement) office. Human subjects approval was secured at the Naval Postgraduate School.

Our population of interest was contracting professionals within the U.S. Army. The survey we developed was sent to approximately 10,000 military, civilian, and Army Corps of Engineers in 1102, 1105, 0800, and 51C job categories. Non-acquisition professionals—those that serve in assistive vice direct contracting roles—have been excluded from the study. That is, we included only those professionals with appropriate and valid authorization to obligate government funds. The survey was open for approximately two weeks. We obtained 1,455 responses; due to incomplete surveys, the final count was 998 surveys, for a 9.9% response rate.

The majority of respondents were civilian contracting professionals (89%); 0.9 percent was military. Six percent held DAWIA Level 1 Certification; 27% were at DAWIA Level 2; and 55.5% were certified at DAWIA Level 3 (10% responded Other). In terms of the commands represented, most respondents were from the Army Contracting Command (ACC) (44%), 12% were from Army Materials Command (AMC), and 21% were from the Army Corps of Engineers. Finally, approximately 60% of respondents were non-warranted, while just fewer than 40% were warranted contracting officers.

Results

The following sections outline some of the topline results from the survey responses. We highlight job satisfaction, job role ambiguity, job stress, organizational justice, and quality of connections.

In terms of job satisfaction, contracting professionals report a moderate degree. There are no appreciable differences between the civilian and military samples (Figure 1), nor are there significant differences between the warranted and non-warranted contracting officers (Figure 2).

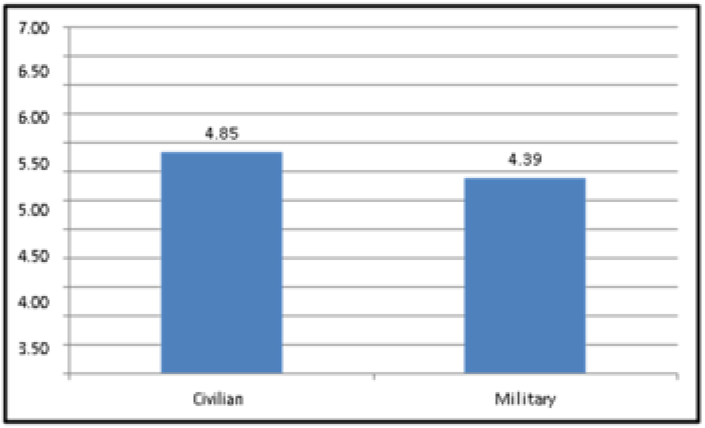


Figure 1. Job Satisfaction by Civilian and Military

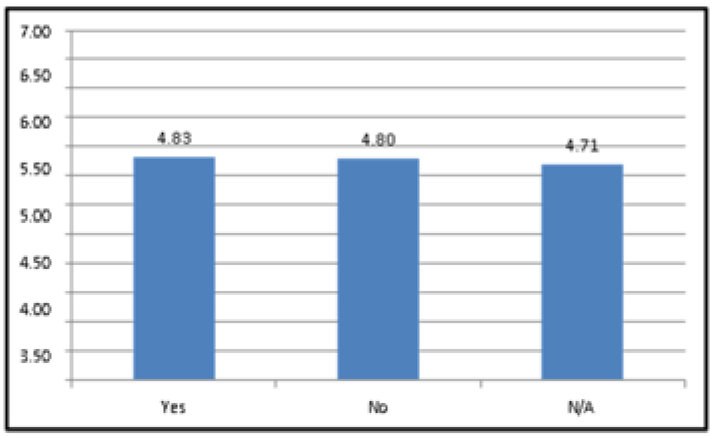


Figure 2. Job Satisfaction by Warranted vs. Non-Warranted Contracting Officers

We also looked at job role ambiguity. Higher ratings for these items suggest that contracting professionals are confident about their work and sense low degrees of ambiguity associated with their job roles (Figure 3). Uncertainty about job roles appears minimal, though higher DAWIA levels show less ambiguity than lower levels (Figure 4).

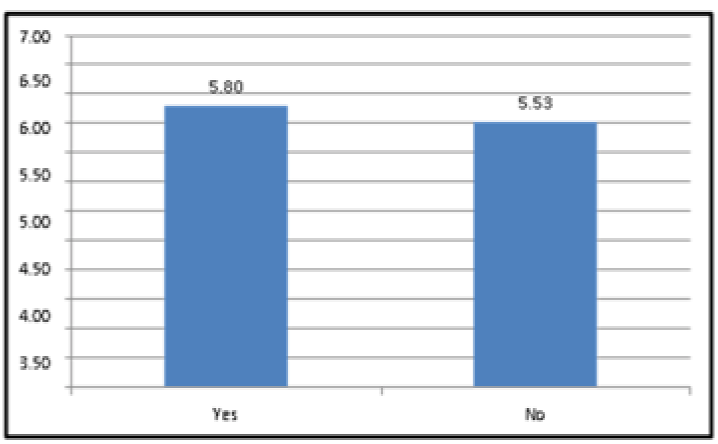


Figure 3. Job Role Ambiguity by DAWIA Levels

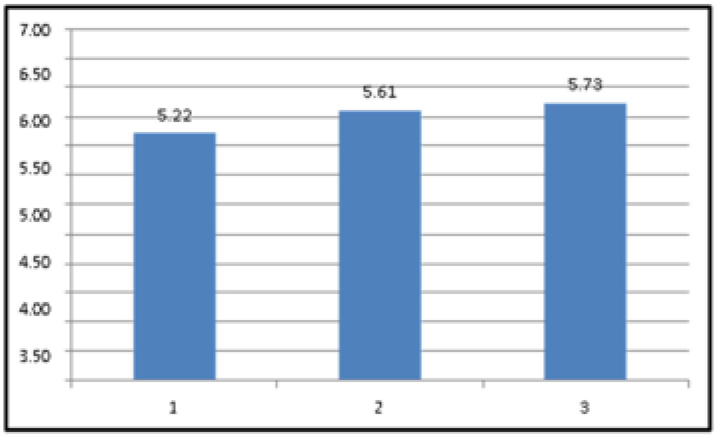


Figure 4. Job Role Ambiguity by DAWIA Levels

Job stress scores were low, suggesting low levels of stress associated with day-to-day work (Figure 5). Other factors associated with stress are work-life conflicts and commute stress and strain. The responses to these scales offer a similar picture, though work-life conflict was highest.

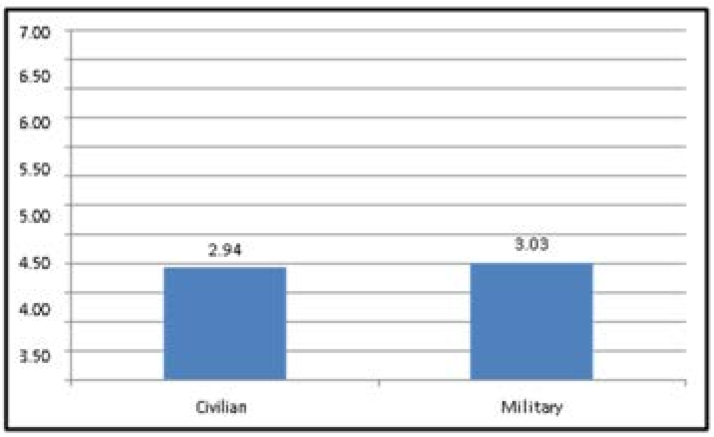


Figure 5. Job Stress by Civilian vs. Military

Organizational justice measures the perceptions of fairness about job processes such as performance evaluation. Respondents report lower degrees of organizational justice, suggesting a need to focus on better performance evaluation processes (Figure 6). As reported by McKeithen (2016), we find a negative correlation between organizational justice and job satisfaction. This “suggests that when organizations foster environments where employees are evaluated based on their own merits, and employees believe supervisors are using accurate information when conducting performance appraisals, job stress is low. Conversely, when employees perceive that supervisors are not fully gathering accurate assessment information when conducting evaluations and/or appraisals, an employee’s level of job- related stress is high” (McKeithen, 2016, p. 101).

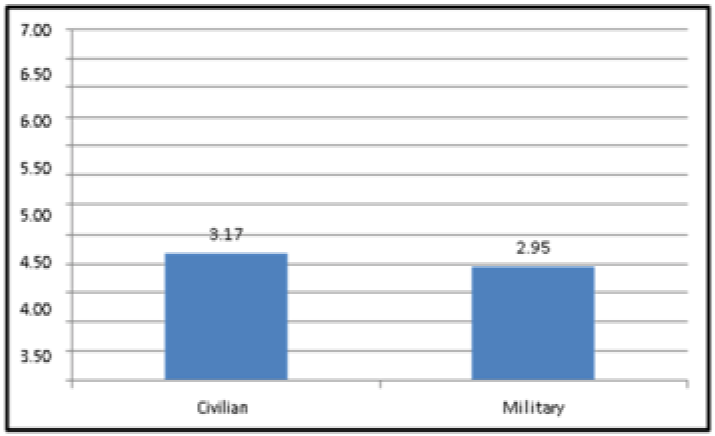


Figure 6. Organizational Justice by Civilian and Military

We also examined the quality of connections. Contracting professionals report that they have moderately high quality of connections (Figure 7). In addition, we include a table that reports all categories and average scores by command (Figure 8).

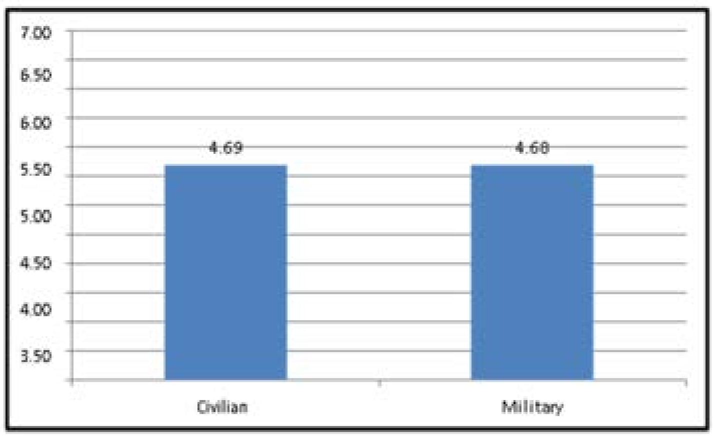


Figure 7. High Quality Connections by Civilian and Military

Command	Job Satisfac-tion	Super-visor Commit-ment	Job Charac-teristics	Job Role Ambi-guity	Job Strain	Work-Family Conflict	Commute Stressos Combined	Organiz-ational Justice	Job Fit	Work-place Values	HQC Capacity	HQC Experience
ACC	4.90	4.22	4.70	5.58	2.86	3.42	2.65	3.19	3.66	4.54	4.74	5.30
AMC	4.39	3.89	3.86	5.67	3.15	2.75	2.82	3.00	3.57	3.73	3.79	4.38
ECC	4.32	4.33	4.27	5.76	3.10	3.56	2.54	3.08	3.68	4.38	4.78	5.23
Medical Command	5.16	4.44	5.00	5.92	2.87	3.10	2.81	3.18	3.90	4.45	4.71	5.33
MICC	4.56	4.25	4.63	5.66	3.04	3.60	2.63	3.14	3.70	4.38	4.66	5.16
National Guard Bureau	4.57	4.47	5.00	5.55	3.05	4.17	2.56	3.06	3.77	4.42	5.07	5.53
US Army Corps of Engineers	4.97	4.34	4.75	5.70	2.99	3.59	2.54	3.18	3.81	4.45	4.60	5.19
Other	4.67	4.16	4.32	5.45	2.95	3.44	2.67	3.08	3.63	4.12	4.39	5.00
No repsonse	5.20	4.35	4.90	5.86	2.21	2.03	2.81	2.67	3.84	4.42	5.03	5.26
Dimension Average	4.81	4.27	4.68	5.63	2.94	3.51	2.62	3.15	3.71	4.44	4.69	5.24

Figure 8. Scores by Command

Recommendations

While there are not significant low ratings on the dimensions we captured, our analysis is limited given a sole data time point. That said, there are several possible opportunities.

1. *Job Stress*: Contracting professionals report low levels of job stress, but when taking commute stress into account, stress was more pronounced. One option may be to "consider incorporating more opportunities for employees to telework from home when appropriate" (McKeithen, 2016, p. 103). Open-ended responses indicated a desire for increased use of telework (McKeithen, 2016).
2. *Organizational Justice*: Contracting professionals report lower degrees of organizational justice, particularly when asked about performance management practices. One option might include examining reward structures and procedures for evaluation purposes.
3. *Quality of Connections:* Contracting professionals indicate they have a moderate degree of positive connections with work colleagues. Developing high quality connections may indeed have a positive effect on job satisfaction (Dutton, 2003), but also may be highly related to building resilience among unit members (Challburg & Brown, 2016).

Future Research

The most recent survey to our knowledge was in 2005 (Doelling, 2005), nearly 12 years ago. The time lag between the initial work and this current work is too great to make meaningful comparisons. Additional analyses for making population estimates based on the sample current sample would provide a stronger argument for the findings. Moreover, longitudinal research will invariably yield comparative data and thereby generate potential insights about the direction the contracting field. We recommend ongoing assessment of the contracting profession. Future assessments may be benchmarked against past assessments and provide a dashboard to evaluate the contracting workforce.

References

Becker, T. E., Billings, R. S., Eveleth, D. M., & Gilbert, N. L. (1996). Foci and bases of employee commitment: Implications for job performance. *Academy of Management Journal, 39*(2), 464–482.

Chalburg, M., & Brown, C. C. (2016). *Resilience among naval recruits: A quantitative and qualitative analysis of interventions at Recruit Training Command and implications on fleet readiness* (Master's thesis). Monterey, CA: Naval Postgraduate School.

Cook, J. D., Hepworth, S. J., & Wall, T. D. (1981). *The experience of work: A compendium of 249 measures and their use.*

Davey, J. A., Kinicki, A. J., & Scheck, C. L. (1997). A test of job security's direct and mediated effects on withdrawal cognitions. *Journal of Organizational Behavior*, 323–349.

Dickenson, M. W., Smith, D. B., Grojean, M. W., & Ehrhart, M. (2001). An organizational climate regarding ethics: The outcome of leader values and the practices that reflect them. *The Leadership Quarterly* (12), 197–217.

Doelling, M. C. (2005). A *work environment climate assessment of an Army acquisition center* (Master's Joint Applied Project). Monterey, CA: Naval Postgraduate School.

Dulebohn, J. H., & Ferris, G. R. (1999). The role of influence tactics in perceptions of performance evaluations' fairness. *Academy of Management Journal, 42*(3), 288–303.

Dutton, J. E. (2003). *Energize your workplace*. San Francisco, CA: Berrett-Koehler.

Fields, D. L. (2002). *Taking the measure of work: A guide to validated organizational research and diagnosis*. Thousand Oaks, CA: SAGE.

Gerbich, L. M. (2017). *An organizational climate assessment of the Navy contracting workforce* (Master's thesis). Monterey, CA: Naval Postgraduate School, Monterey, CA.

GAO. (2017). *High-risk series: Progress on many high-risk areas, while substantial efforts needed on others*. Washington, DC: Author. Retrieved from http://www.gao.gov

Kluger, A. N. (1998). Commute variability and strain. *Journal of Organizational Behavior, 19*, 147–165

McKeithen, M. L. (2016). *Organizational climate assessment of the Army contracting workforce* (MBA professional report). Monterey, CA: Naval Postgraduate School.

Moran, E. T., & Volkwein, J. F. (1992). The cultural approach to the formation of organizational climate. *Human Relations, 45*(1), 19–47.

Thomas, L. T., & Ganster, D. C. (1995). Impact of family-supportive work variables on work-family conflict and strain: A control perspective. *Journal of Applied Psychology, 80*(1), 6.

Van Dyne, L., Graham, J. W., & Dienesch, R. M. (1994). Organizational citizenship behavior: Construct redefinition, measurement, and validation*. Academy of Management Journal, 37*(4), 765–802.

Wayne, S. J., Shore, L. M., & Liden, R. C. (1997). Perceived organizational support and leader-member exchange: A social exchange perspective. *Academy of Management Journal, 40*(1), 82–111.

Xie, J. L. (1996). Karasek's model in the People's Republic of China: Effects of job demands, control, and individual differences. *Academy of Management Journal, 39*(6), 1594–1619.

Understanding Defense Acquisition Workforce Challenges[1]

Colleen Murphy—is a Contract Analyst at MITRE, supporting DoD and Department of Homeland Security acquisitions. She is a project lead responsible for providing objective and independent acquisition management analysis and recommendations to enable successful development and evolution of strategies for the acquisition of cybersecurity and other critical IT capabilities. Previously, she worked in industry and began her career as a contracting professional in the United States Air Force. She supported Operations Enduring Freedom and Iraqi Freedom as a contingency contracting officer. Murphy holds an MBA from Virginia Tech and is DAWIA Level III certified in contracting. [cmmurphy@mitre.org]

Adam Bouffard—is a Contract Analyst at MITRE, supporting DoD acquisition, program management, and contracts for three different projects. Adam is also a contributor to MITRE acquisition research projects. Previously he worked as a Contract Specialist and Contracting Officer for Naval Sea Systems Command and the Office of Naval Research. He holds a BS in Psychology from Virginia Tech, an MBA from the University of Maryland, University College (UMUC), and an MS in Acquisition and Supply Chain Management from UMUC. He is DAWIA Level III certified in contracting and is currently studying for his MS in Systems Engineering at George Washington University. [abouffard@mitre.org]

Executive Summary

The DoD acquisition system is a complex enterprise requiring professionals with many years of experience to execute the process expertly. However, the acquisition workforce faces several key challenges. First, the number of experienced acquisition professionals in the DoD is declining. They are being replaced by a young generation facing a long learning curve. Second, the acquisition workforce lacks the experience, knowledge, and tools necessary to digest and apply the wealth of information related to acquisition. Third, the workforce struggles to keep pace with the increasing complexity of the federal acquisitions.

The MITRE Corporation conducted this research to validate the significance of the key challenges facing the workforce and provide a foundation for "next steps." MITRE developed an independent survey to document the issues facing practitioners in the field.

Through our research, we validated that the workforce believes they are ill-equipped to meet the demands of the acquisition environment. We determined that the workforce needs solutions to assist them to shorten the learning curve, modern tools that appeal to the changing workforce demographic, and cultural changes that support and encourage the workforce to think critically to successfully operate in a complex environment. The recommendations proposed as part of this research include a digitized work environment, tailored acquisition models, and workforce cultural changes.

A digitized work environment is necessary to appeal to the junior acquisition workforce. They are accustomed to technology at their fingertips to answer questions and solve problems. Currently, acquisition policy and guiding documents are mainly a collection of static pdf documents spread across a variety of federal and organizational level websites. Digesting volumes of information to understand how to navigate acquisition processes is

[1] This technical data deliverable was developed using contract funds under Basic Contract No. W15P7T-13-C-A802. Approved for Public Release; Distribution Unlimited. Case Number 17-1585.

neither efficient nor appealing to a young and digitally focused workforce. Digitized policy provides centralized access to current policy and reference material to simplify complex processes.

Tailored acquisition models provide streamlined approaches to address direction from acquisition leadership to tailor acquisition strategies when appropriate. Tailored models include only the required elements for a specific type of acquisition and enable less-experienced professionals to successfully navigate a tailored approach.

Workforce cultural changes include recommendations to address implementation at the workforce level. For example, policy changes to approve tailored models, implementing a coaching environment to transition the knowledge of the departing experienced workforce to the junior workforce, and implementing modern tools and apps to deliver digital capabilities.

Workforce Challenges With Experience, Knowledge, and Tools

The government constantly calls upon the federal acquisition workforce to deliver acquisition solutions in an increasingly complex environment. Successfully accomplishing this task requires an extensive understanding of the acquisition system, a wide awareness of best practices and exemplars, and access to state-of-the-art digital tools to develop and leverage solutions across the DoD enterprise.

Overview of Workforce Challenges

Our research discussed in this paper shows that federal acquisition requires a unique skill set to navigate successfully so that government agencies can deliver systems and services that meet mission needs. Our research also supports that, in many cases, the acquisition workforce lacks the requisite experience, knowledge, and tools to keep pace with the demands of this environment.

Early career acquisition professionals receive Defense Acquisition University (DAU) classroom and online training to gain an overview of the core acquisition elements. Yet they cannot apply the knowledge in practice until they accumulate actual on-the-job experience. Going through the core activities to guide programs through the acquisition life cycle represents the ultimate development of an acquisition professional.

Ideally, acquisition professionals would be exposed to a broad array of acquisition types early in their career. While some are so fortunate, such as military professionals who transfer every two to four years, not all are exposed to the variety of acquisitions and the skills necessary to develop a broad knowledge base. For example, the process for acquiring large weapons systems subject to DoD Instruction (DoDI) 5000.02 is clearly much more complex and rigorous than that for acquiring general services. Agile development acquisitions present even more unique challenges. Nonetheless, acquisition professionals are expected to have the capability to operate in any of these environments. At the very least, they are expected to turn to the massive amounts of statutory, regulatory, and guidance information currently available through online resources and figure out what they need to know to execute the acquisition successfully.

Our research illustrates that changing demographics present another challenge. Half of the DoD acquisition workforce is eligible to retire within the next 10 years—depriving countless programs of decades of experience. Compounding this loss of expertise, evidence shows that 40–50% of the workforce has less than five years of experience. With the workforce peaks at the early and late career stages, DoD reports often cite the huge shortfall of mid-career professionals who anchor most program offices. While the DoD has made progress over the past few years in addressing this shortfall, it still poses a major risk to the acquisition workforce.

One of the most significant challenges in this area is an information gap. The existing DAU classroom and online training can provide the basic information, but without years of experience gaining the knowledge needed to successfully execute the variety of DoD acquisitions, or access to professionals with the requisite knowledge, young professionals will be at a significant disadvantage.

Our research shows that acquisition involves a long learning curve. It takes many years of experience to develop the depth and breadth of skills and acquire adequate knowledge to execute the acquisition process for all types of requirements. Acquisition professionals are expected to have a broad knowledge base, but those practical skills come only with hands-on experience.

It is impossible to curb the pending retirement of experienced acquisition professionals. Therefore, the acquisition community needs advances in technology and tools to enable the next generation of the acquisition workforce to rapidly digest and synthesize the vast amounts of information in the acquisition environment. This is especially critical as the up-and-coming acquisition workforce is accustomed to digital technology and social media that enable instantaneous access to current and accurate information to solve problems. The federal government will find it difficult to hire, retain, and train a developing workforce capable of tackling complex acquisitions without incorporating advances in the digital space to replace the lack of long experience and knowledgeable professionals to consult.

The difficulties confronting the acquisition workforce are serious and have the potential to negatively impact the government's future ability to effectively execute acquisitions. Changes are desperately needed to adequately prepare and arm the workforce for the task at hand.

Research Methodology

The MITRE team first conducted a literature review to understand current assessments of the federal acquisition workforce. The review focused on the demographics of the current workforce, the rising complexities of federal acquisition, and the availability and access to acquisition knowledge and training.

Based on this literature review, the MITRE team developed hypotheses regarding workforce perceptions about achieving proficiency, the changing complexity of federal acquisition, and the availability of relevant and helpful tools and resources. MITRE then conducted a survey focused on acquiring data and metrics to test those hypotheses. MITRE analyzed this data to identify areas where improvement is needed to empower and assist the acquisition workforce to succeed in meeting federal acquisition needs.

MITRE also assembled three focus groups to assist in the analysis of our survey data and the data gathered in terms of the tools associated with learning acquisition. The focus groups were asked questions to assess the data gathered in terms of tool availability, tool design, and creating an efficient learning environment for acquisition professionals.

Literature Review

The literature review examined existing research and surveys of the acquisition workforce performed in recent years. MITRE found several recurrent themes in the existing data.

The federal government must meet wide-ranging and dynamic mission imperatives with limited budgets and resources. Programs are regularly faced with budget reductions and cost saving/efficiency targets, but are also challenged to think critically to develop innovative solutions to quickly acquire and deliver services and solutions to users. To address these challenges, the government relies on acquisition professionals who are knowledgeable of the acquisition system and capable of developing creative and innovation solutions in a constrained environment.

To understand the specific factors preventing the workforce from successfully developing and implementing innovation solutions, the MITRE team reviewed the 2016 government-wide survey of federal acquisition executives conducted by Grant Thornton Public Sector and the Professional Services Council (PSC), *Aligning for Acquisition Success: Overcoming Obstacles to Results*. One of the areas explored by this survey is barriers to innovation in acquisition. The survey asked about the perceived challenges preventing the acquisition workforce from successfully developing and implementing innovative solutions. Figure 1 depicts the breakout of the top reported barriers to innovation in acquisition.

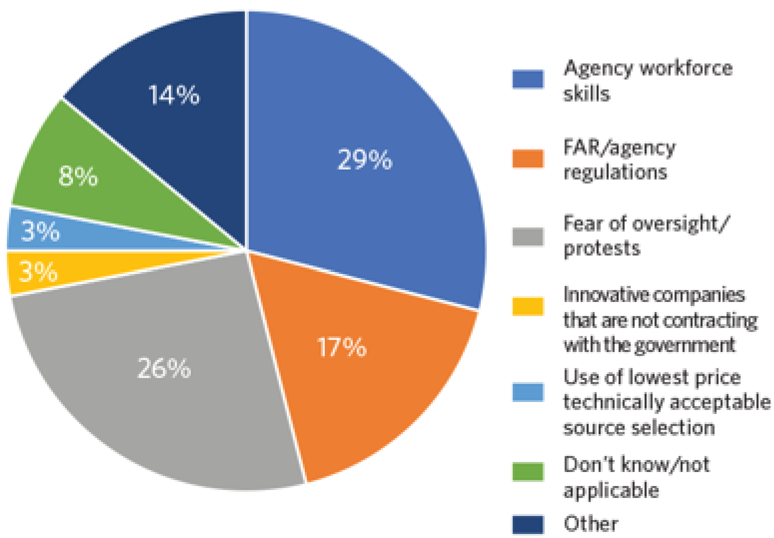


Figure 1. Reported Barriers to Innovation in Acquisition
(Grant Thornton & PSC, 2016)

Workforce skills was reported as the top inhibitor to innovation. The fear of oversight and protests was the second highest barrier to innovation. These two factors combined make up approximately 55% of the barriers reported. This is important as these responses suggest the workforce believes it is not adequately skilled to develop and acquire innovative solutions nor adequately supported to take risks. Grant Thornton and PSC (2016) also go on to say that “these factors contribute to a confidence gap, as workers remain unprepared or unwilling to take well-reasoned risks to achieve potential innovations or cost savings, instead defaulting to familiar, often suboptimal, strategies.”

Grant Thornton and PSC (2016) made an important observation that “these inhibitors are interconnected. Acquisition workers’ inexperience means they tend to focus on compliance and don’t understand the flexibilities in the FAR. Thus, they tend to be overly risk-averse out of fear of protests or punishment, rather than trying new and different things.”

An unskilled acquisition workforce has severe impacts on the government and its ability to execute its missions effectively. The Center for the Study of Democratic Institutions at Vanderbilt University, in cooperation with Princeton University and the Volcker Alliance, conducted a survey on the future of government service, garnering responses from over 3,500 federal executives that found very similar results. When asked if the executives worry about an inadequately skilled workforce as a significant obstacle to fulfilling their agency's core mission, 39% responded "agree" or "strongly agree" (Center for the Study of Democratic Institutions, 2015).

These results highlight the need to improve workforce skills and to develop and shape a workforce confident in their abilities and in having the right tools and access to information to develop and execute innovative solutions to achieve mission objectives.

MITRE also wanted to understand the impact of the changing acquisition workforce demographic on the government's ability to effectively execute acquisitions. Figure 2 shows that approximately 50% of the acquisition workforce within the DoD is eligible to retire within the next 10 years. Jeffrey Koses, a Federal Acquisition Institute board member, notes that "fully developing an acquisition professional takes five to 10 years." A transitioning workforce combined with a steep learning curve for acquisition has led to a knowledge gap. Koses says it best in terms of the future acquisition workforce, noting that the government should be "rethinking training to match the way these digital natives are accustomed to receiving and consuming information" (McCabe & Laurent, 2015).

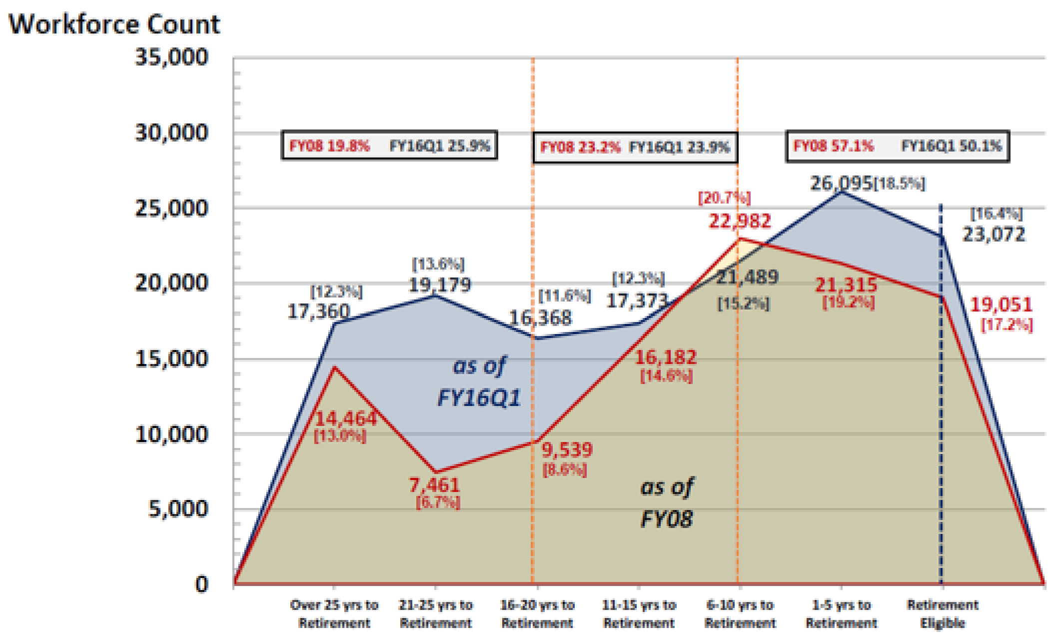


Figure 2. Civilian Acquisition Workforce Demographics (FY 2008–2016Q1)
(USD[AT&L], 2016)

Grant Thornton and PSC (2016) elaborate on this, stating that generational issues present a major challenge for the future. This includes replacing the retiring baby boomers who make up 50% of the retirement-eligible workforce and training and retaining the millennials who will replace them. A Grant Thornton and PSC (2016) respondent commented, "The federal employment construct of a 'job for life' is not the mindset of millennials.” This makes revitalizing acquisition workforce training and tools imperative not just to educate the new workforce but also to retain them.

McCabe and Laurent (2015) note that keeping and maintaining a workforce from the millennial generation is very difficult because they tend to stay in jobs for only about two to three years. The Under Secretary of Defense for Acquisition, Technology, and Logistics (USD[AT&L]) *Performance of the Defense Acquisition System: 2016 Annual Report* highlighted the need for a focus on younger generations given the amount of retirement eligible workforce members (see Figure 2).

As the experienced workforce retires and moves on, federal acquisitions are challenged to bridge the knowledge gap and prepare the next generation of professionals to face the challenges of an increasingly complex acquisition system. We believe the changing workforce demographic is a significant impact to federal acquisitions.

Also of interest to our research is the workforce thoughts on existing acquisition training and tools. Acquisition workforce training and tools haven’t modernized much over the past two decades. Stan Soloway, PSC’s former President and Chief Executive Officer, states that training is “bound to traditional models and assumptions far more relevant to a hardware-dominated, single-customer market of limited commerciality.” Existing training and tools provide a basic level of understanding necessary to establish a solid acquisition foundation, but they remain relatively focused on tried and true methods that don’t always lend to the innovative thinking and problem-solving. The dated and status quo focus on the FAR being rigid and unforgiving discourages new hires from thinking outside of the box and doesn’t develop skills to take advantage of the flexibilities it offers. The government must encourage the future acquisition workforce to change acquisition and its culture, and it must embrace the innovative ways this generation learns (McCabe & Laurent, 2015).

Additionally, more agile training methods are required to address the changing nature of acquisition, innovation, and the marketplace. The typical large hardware system acquisition rubric doesn’t translate well when acquiring rapidly changing technologies or services. Acquisition techniques need to evolve from those established decades ago to result in a successful acquisition workforce.

MITRE Acquisition Workforce Survey Methodology

MITRE conducted a survey to collect information to test the below hypotheses and to assess the need for development of solutions that enable the acquisition workforce to better navigate the acquisition processes, access acquisition information, and improve knowledge and skills to operate in the complex federal acquisition environment.

The survey was designed to explore the following theories:

- Federal acquisition has a long learning curve. Only those with five or more years of experience have adequate knowledge and confidence to execute their role in the acquisition process for any size acquisition and any type of product or service without having to rely heavily on the expertise of others in their field. More than 25% of the current DoD acquisition workforce has less than five years of experience. Therefore, the DoD does not have enough experienced acquisition professionals to adequately execute acquisitions (USD[AT&L], 2016).
- Acquisition professionals rely heavily on the expertise of more experienced colleagues to learn acquisition skills. As the number of proficient acquisition professionals continues to decrease (due to attrition), those with less than

five years of experience will not have enough experienced professionals available from whom to learn how to execute their role in the acquisition process.

- Acquisition has become increasingly complex over the past five years. If this trend continues and acquisition complexity continues to increase, it will become even more challenging to train an inexperienced workforce, especially if existing tools are inadequate and not updated in a timely manner. This will decrease the likelihood that programs can successfully navigate the acquisition life cycle to deliver within cost, schedule, and performance requirements.
- Acquisition professionals do not have the necessary knowledge, tools, and training to tailor the acquisition process for the specific solution they are acquiring. Furthermore, the available information is not organized in an easily navigable fashion.

Survey Results

Survey responses were received from over 250 individuals supporting the DoD, civilian agencies, and the intelligence community. Figure 3 depicts the survey respondents' demographics.

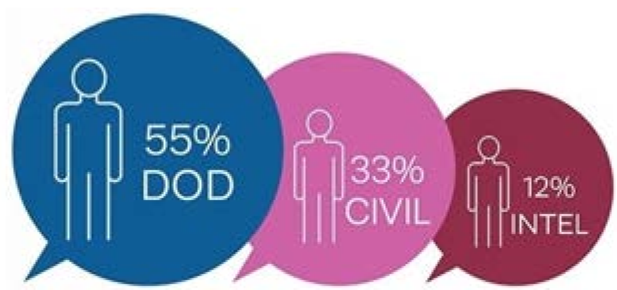


Figure 3. Respondent Community Breakdown

Acquisition Has a Long Learning Curve

Of the total respondents, 64% stated that it takes 10 years or more to become fully proficient in acquisition. Furthermore, the longer the respondent had been in the acquisition workforce, the more years of experience the respondent believed that it took to become proficient: 18% of respondents believed it took upwards of 15 years to become proficient and 10% believed that it takes 20 years or more to become proficient. These responses are shown in Figure 4 and validate MITRE's hypotheses. MITRE's hypotheses are further supported by a Harvard Business Review (HBR) article stating it takes time to become an expert, a minimum of 10 years, and in some cases 15 to 25 years of steady practice (Ericsson, Prietula, & Cokely, 2007).

Figure 4. Respondent Proficiency Responses

Nearly 50% of the DoD acquisition workforce lacks the necessary years of experience to be considered proficient in acquisition using this measurement (see Figure 2). This data supports the findings in the literature review of a sub-optimally experienced workforce. Taking into consideration the hypothesis that the learning curve associated with acquisitions is steep, the lack of experience equates to a sub-optimally skilled workforce as well.

Acquisition Professionals Rely Heavily on Expertise of More Experienced Colleagues and Reference Tools

First the study aimed to identify respondents' preferred sources of information about federal acquisition. Figure 5 shows how respondents ranked formal training, more experienced colleagues, and reference tools.

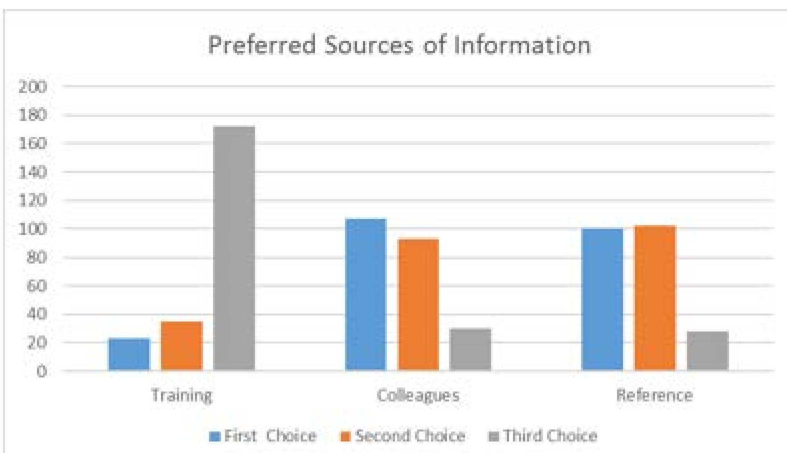


Figure 5. Preferred Sources of Information

Only 10% of the respondents cited formal acquisition training as their primary avenue for learning. While classroom training supplies foundational knowledge about acquisition, it does not deliver the specific acquisition knowledge that easily translates into day-to-day practice. Respondents preferred to use either their more experienced colleagues or reference tools to answer their acquisition questions.

Approximately 46% of respondents confirmed that experienced colleagues are their primary source of acquisition information, while 44% use reference tools as their first choice. This confirms our hypothesis that acquisition professionals rely heavily on the expertise of more experienced colleagues, but also that professionals rely almost as much on reference materials. As the number of experienced professionals continues to decrease, the need for tools to bridge the knowledge gap becomes even more critical.

Acquisition Has Become Increasingly More Complex

The survey specifically asked if respondents believed acquisition has become more complex over the last five years; 74% of respondents believe that acquisition has become slightly more or significantly more complex. Along with this, respondents reported the results shown in Figure 6 regarding rating the most prevalent challenges to learning and executing acquisitions.

Figure 6. Rank Order of Challenges

Respondents felt pressure to “think outside of the box” and to be innovative, but did not feel supported in doing so and encountered considerable resistance when they sought to deviate from the status quo. The results also show that the government releases a large amount of policy and guidance that respondents are unable to appropriately adjudicate due to uncertainty of application or time to process the new changes. This confirms the hypothesis that acquisition continues to become more complex. This data also illustrates that the acquisition workforce is neither adequately empowered nor equipped to successfully execute acquisitions. If the government continues trying to “fix” acquisition by releasing additional policies and guidance, the risk that programs will miss cost, schedule, and performance goals rises.

Acquisition Professionals Do Not Have the Necessary Knowledge, Tools, and Training to Tailor the Acquisition Process for the Specific Solution They Are Acquiring

The survey then focused on the current tools and guidance that respondents noted they rely on to learn a new process and stay current with new information, for example, DAU Portal, or agency/organization-specific tools. The survey asked respondents to rate on a scale of 1 to 5 the following aspects of the existing tools and guidance:

- Currency: Is the available information accurate and up to date?
- Relevancy: Does the information help you execute your acquisition?
- Availability: Are you able to easily find the information that you need when you need it?
- Usability: Are you able to easily navigate the tools (e.g., search capabilities)?
- Digestibility: Is the information conveyed in a way that is easy to understand and apply to your acquisition?

Respondents rated four out of the five aspects as poor, marginal, or fair. Figure 7 shows that the workforce believes current tools lack relevance, usability, digestibility, and availability. Most survey respondents believe the existing tools are difficult to navigate to find the relevant information sought.

Figure 7. Current Tool Ratings

Finally, the survey asked respondents to rate the adequacy of existing tools, information, and training to help the acquisition workforce to tailor processes to fit unique acquisition needs. Of the respondents, 55% believe existing tools, information, and training are inadequate or only somewhat adequate to conduct tailoring activities.

These ratings reinforce MITRE's hypothesis regarding the state of existing tools for accessing the knowledge to execute acquisitions. In fact, a review of the major acquisition sites that provide the tools and information to the workforce revealed broken links, scattered information that is often outdated, misaligned guidebooks and recommendations with new policy, and scattered exemplars and best practices. Many of these websites were developed and designed in the 1990s/early 2000s and have never been updated.

Overall, the survey results, literature review, and data analysis confirmed the hypotheses that MITRE developed about the acquisition workforce. MITRE used this information and generated recommendations for further research and development.

Analysis

MITRE found a disconnect between the expectations of DoD acquisition leadership and the capabilities of the workforce. For example, the workforce is constantly encouraged and challenged to be "innovative," yet policy and processes for executing an acquisition program remain unchanged. The current acquisition environment is largely confined to executing acquisitions in an outdated manner that does not align with the demands for a workforce that thinks critically and is equipped to develop innovative strategies and solutions to meet warfighter needs faster than ever.

The acquisition workforce is key in meeting Third Offset Strategy objectives that call for delivering innovative capabilities to ensure U.S. military superiority, which is being challenged by our "pacing competitors" (Pellerin, 2016). Yet the workforce is not currently trained to think critically or provided tools to enable such thinking. Fostering innovative acquisition thinking and strategies is essential to addressing acquisition challenges and keeping pace with technological challenges. The government must also institute cultural and organizational changes to incentivize the workforce to deliver innovative solutions.

Frank Kendall, Under Secretary for Defense, Acquisition, Technology & Logistics, has long encouraged critical thinking through the Better Buying Power 1.0–3.0 initiatives. DoD acquisition leadership also encourages tailoring of policy, when appropriate, to streamline acquisition timelines. For example, DoDI 5000.02 contains more than 12 references to tailoring. The recent DoDI 5000.75, *Business Systems Requirements and Acquisition*, issued in February 2017, includes the following guidance regarding tailoring:

> **Tailoring**. The procedures used to develop business capability requirements and supporting systems will be tailored to the characteristics of the capability being acquired. Tailoring will focus on application of best practices to the totality of circumstances associated with the program, including affordability, urgency, return on investment, and risk factors. The functional sponsor, MDA, and CAE or designee will collaborate to tailor program strategies and oversight, including: program information, acquisition phase content, and the timing and scope of decision reviews and decision levels. (DoD, 2017, Sec. 4: Procedures)

This policy gives acquisition professionals the latitude to tailor requirements and use best practices to address affordability, urgency, return on investment, and risk factors. But it assumes the workforce is familiar with business systems best practices and has the in-depth knowledge to apply tailoring to a unique acquisition program. Much of the acquisition workforce does not have the deep understanding of acquisitions necessary to know which processes and documents can be tailored to achieve effective and efficient results.

The government must make revolutionary changes to align with the changing workforce demographic. Despite the massive amount of information available through online resources, the current acquisition environment lacks tools that enable the workforce to sort through all the policy, guidance, and available information to develop tailored solutions that satisfy warfighter needs.

Conclusions and Recommendations

The DoD is heavily burdened by bureaucratic processes, policies, and culture that prevent it from effectively exploiting leading technologies for military advantage. Former Defense Secretary Carter proclaimed that the DoD has a strategic imperative to innovate, with speed and agility the key factors, and championed investment in agile, innovative organizational structures/constructs. He chartered the Defense Innovation Board, composed of executives from the leading Silicon Valley companies and academia, to infuse innovation in the DoD. General Selva, Vice Chair of the Joint Chiefs of Staff, says the DoD is not organized for innovation and needs more operational experimentation.

These expectations are lofty given the trajectory of the acquisition workforce. To achieve the success metrics established by DoD leadership, the DoD must make fundamental changes at the workforce level. The workforce needs modern tools to maneuver through the acquisition system and fill the gaps left by the retiring experienced personnel. To meet demands to identify, develop, and integrate game-changing technologies for our warfighters to maintain a technical superiority, the DoD must focus attention on the acquisition workforce to ensure it can respond to the ever-changing threat environment and the associated demands.

A Digitized Work Environment

The government needs new tools to shorten the learning curve and appeal to a younger workforce accustomed to digital solutions. Although many online resources are currently available, they are limited in functionality and still require an extensive baseline knowledge of the acquisition system.

Digitized policy allows the acquisition workforce, especially a more digitally focused workforce, to quickly access what is currently a collection of hundreds of pages of static .pdf documents with pages of reference documents. It provides instant clarification of terminology, links to relevant sections of up-to-date policy documents, and access to reference documentation without the user's having to leave the site.

Tailored Acquisition Approaches

HBR cites an argument from 13th-century philosopher and scientist Roger Bacon that it would be impossible to master mathematics in less than 30 years. Today, individuals can master complex frameworks such as calculus in their teenage years. HBR attributes this to a change in the organization and accessibility of material and states, "Students of mathematics no longer have to climb Everest by themselves; they can follow a guide up a well-trodden path" (Ericsson, Prietula, & Cokley, 2007). This concept of repacking and improving accessibility can be applied to acquisitions through tailored models.

Tailored acquisition models give users the benefit of pre-tailored, approved acquisition solutions for a variety of acquisitions. For example, tailored models for agile acquisitions can offer streamlined processes to guide implementation of agile solutions. Chang & Modigliani (2017) liken tailored acquisition models to "Google maps for acquisition":

> Today, acquisition professionals are expected to tailor the DoDI 5000.02 on their own. This can be compared to handing them a map and telling them to figure out the best way to drive from New York City to Los Angeles. If this is their first time traveling this route, it would take a lot of time to study the map, plan the route, talk to others about shortcuts, and encounter traffic and detours along the way. Perhaps they will reach their destination, but not without wasting significant time and fuel. Proactively tailored models are the Google Maps for acquisition. Routes are optimized for the type of product or service being acquired with turn by turn guidance for each acquisition phase. Tailored acquisition models provide the acquisition workforce with a pre-chartered route that guide users on a path for success.
>
> Tailored acquisition models are pre-filtered to provide only the information, processes, documentation, and reviews that are relevant for that type of acquisition. If a Service or Portfolio Acquisition Executive approves these models for their organization, programs no longer should request tailoring permission and obtain waivers from multiple oversight organizations. Programs can operate with pre-authorization to streamline specific procedures and documents based on the type of product or service being acquired.

Tailored models aid acquisition professionals who lack a full understanding of the life cycle DoDI 5000.02 process to quickly determine the applicable processes, artifacts, and milestones necessary to execute a specific type of acquisition, such as an acquisition for an agile IT system. These tailored models include only the necessary artifacts and consolidated processes and milestones to eliminate unnecessary requirements. The government cannot expect the workforce to move toward innovative and streamlined acquisition when they are clearly struggling with the basics.

Workforce Cultural Changes

In addition to new tools, such as tailored models and digitized policy, cultural changes are necessary to support and accomplish innovative solutions and translate high-level policy into actual implementation at the workforce level. For example, tailored models can only succeed if acquisition executives and leadership at the Program Executive Organization (PEO) levels approve the use of tailored models and the accompanying modified processes.

Accelerated learning can also be addressed by optimizing learning through on-the-job training opportunities. For example, rather than assigning a crushing workload to acquisition veterans, organizational leaders could instead refocus the priority of seasoned veterans on growing the junior staff. Given the overwhelming workload that plagues many organizations, this may be difficult to implement. Per HBR, having expert coaches makes a big difference in the learning process to include accelerating the learning process (Ericsson, Prietula, & Cokely, 2007). But not attempting to identify opportunities to implement organizational cultural changes focused on developing the junior acquisition workforce present a long-term threat to the ability to execute acquisitions.

The increasing complexity of acquisition also must be addressed. Business models, as well as technology, are changing. Continuing to rely upon traditional tools and training doesn't support the acquisition workforce to "think outside the box" and "think critically." The government has an opportunity to reshape the way it approaches acquisitions by leveraging the inexperience of the junior workforce. To do this, the government must create training opportunities that embrace the complex and changing landscape. For example, if the government were to simplify access to statutory and regulatory information through modern digital tools and implement tailored models, or even apps that speak to upcoming generations, training could instead focus around thinking critically and developing innovative solutions.

FFRDC Digital Acquisition Capability

To accelerate the learning curve through digital tools, MITRE recently launched ACQUIRE, a digital platform for acquisitions (http://acquire.mitre.org) to accelerate the learning curve through digital tools. ACQUIRE presents a new way of thinking about and executing acquisitions.

- ACQUIRE provides a digital capability for acquisition professionals to understand the mechanics of implementing a type of acquisition with which they may be unfamiliar.
- The platform contains digitized policy and tailored acquisition models that enable less-experienced professionals to quickly navigate the complex acquisition environment by consolidating and simplifying the vast amounts of available information.

ACQUIRE gives members of the workforce one-stop access to current acquisition policy and enables them to quickly navigate to helpful and relevant information about the problem they are attempting to solve.

Since MITRE is a Federally-Funded Research and Development Center, the ACQUIRE capability was developed for the public good and is available for use and application to the acquisition work environment.

Federal acquisitions can also leverage leading Artificial Intelligence (AI) tools, like IBM's Watson, to digest vast amounts of structured and unstructured data from policies, guides, and program documentation to help programs apply the vast information sources and provide executives better visibility into their enterprise.

References

Center for the Study of Democratic Institutions. (2015). *Survey on the future of government service*. Washington, DC: Author.

Chang, S., & Modigliani, P. (2017). *Proactively tailored acquisition models*. Defense Acquisition University.

DoD. (2017). *Business systems requirements and acquisition* (DoD Instruction 5000.75). Washington, DC: Author.

Ericsson, K., Prietula, M., & Cokely, E. (2007). Managing people: The making of an expert. *Harvard Business Review*.

Grant Thornton & PSC. (2016). *Aligning for acquisition success: Overcoming obstacles to results.* Washington, DC: Author.

McCabe, K., & Laurent, A. (2015, April 3). The real problem with acquisition training. Retrieved from http://www.govexec.com/contracting/2015/04/real-problem-acquisition-training/109138/

Pellerin, C. (2016, October 31). Deputy secretary: Third offset strategy bolsters America's military deterrence. *DoD News Defense Media Activity*.

Under Secretary of Defense for Acquisition, Technology, and Logistics (USD[AT&L]). (2016). *Performance of the defense acquisition system: 2016 annual report*. Washington, DC: DoD.

The Volcker Alliance. (2016). *Doing the people's business: Key competencies for effective public procurement*. New York, NY: Author.

2016 Assessment of the Civilian Acquisition Workforce Personnel Demonstration Project

Jennifer Lamping Lewis—is an economist at the RAND Corporation and a member of the Pardee RAND Graduate School faculty. She has applied her expertise in strategy, incentives, and quantitative analysis to a number of issues in defense acquisition and workforce management, including a study exploring the relationship between the characteristics of the defense acquisition workforce and the performance of defense acquisition programs. Lewis has co-led or contributed to a number of other studies on workforce management issues, ranging from extending military tour lengths to facilitating military-to-civilian conversions to using auctions to induce voluntary separations and set incentive pays. Lewis formerly served as the associate director of RAND's Forces and Resources Policy Center. Prior to joining RAND, she spent six years on the faculty of the University of Colorado at Boulder. Lewis received her doctorate in economics from Columbia University.

Additional Authors:

Laura Werber—is a senior management scientist at the RAND Corporation and a professor at the Pardee RAND Graduate School. She has experience applying a variety of research methods to defense workforce management and business process improvement concerns and has been recognized for her expertise with qualitative data in particular. Werber co-led the fiscal year 2012 AcqDemo assessment and has led or contributed to studies pertaining to the need for STEM-degreed personnel in the Air Force, barriers to promotion for minorities within an intelligence agency, and best practices for recruiting and retaining members of the Air Force civilian acquisition workforce. Werber received her doctorate in organizational behavior from Stanford University.

Cameron Wright—is a doctoral candidate at the Pardee RAND Graduate School and an assistant policy analyst at the RAND Corporation. His work includes devising and implementing a quality ranking system for over 500 engineering colleges in India, helping the U.S. Army improve deployment processes, and conducting a survey analysis of U.S. Special Operations Forces on their attitudes regarding the integration of women into their units. Wright received his master's degree in business administration from Ohio University and a bachelor's degree in economics from Bowling Green State University.

Irina Danescu—is a research assistant at the RAND Corporation. She has contributed to a number of studies in defense acquisition and workforce management, including an examination of whether Australia should procure ships from foreign shipbuilders and an analysis of the U.S. Air Force's capacity to provide intelligence analysis. Danescu received a bachelor's degree in economics and a bachelor's degree in public policy from Duke University.

Jessica Hwang—is a doctoral candidate in the Statistics Department at Stanford University. Her contributions to RAND studies have included estimating installation-specific rates of sexual assault and sexual harassment in the U.S. military. Hwang received her master's degree and bachelor's degree in statistics from Harvard University.

Lindsay Daugherty—is a policy researcher at the RAND Corporation. She has experience with survey research and qualitative methods and has contributed to several analyses of the DoD civilian acquisition workforce, including the fiscal year 2012 AcqDemo assessment. Daugherty received her doctorate in policy analysis from the Pardee RAND Graduate School.

Key Findings

- Salary levels, salary growth, and retention outcomes in the Department of Defense Civilian Acquisition Workforce Personnel Demonstration Project (AcqDemo) were similar to those in the General Schedule (GS) system, after controlling for an array of other factors.
- AcqDemo employees earned higher starting salaries but were promoted less frequently relative to equivalent employees in the GS system.

- Within AcqDemo, higher levels of contribution to the organizational mission were associated with higher salaries, more-rapid salary growth, more promotions, and a greater likelihood of retention.
- However, less than half of AcqDemo survey respondents perceived a link between contribution and compensation. Possible explanations include a perceived lack of transparency regarding how performance ratings are calculated and translated to pay; difficulty measuring employee performance objectively and inclusively, particularly for managers; and a narrowing spread in salaries resulting from pay caps and AcqDemo business practices.
- A majority of AcqDemo survey respondents expressed concerns about the transparency and fairness of various aspects of the appraisal and compensation system.

The executive summary can be found at

http://www.rand.org/pubs/research_reports/RR1783z1.html

The full report can be found at

http://www.rand.org/pubs/research_reports/RR1783.html

THIS PAGE INTENTIONALLY LEFT BLANK

Panel 9. Enabling Cybersecurity

Wednesday, April 26, 2017	
3:30 p.m. – 5:00 p.m.	**Chair: Rear Admiral John W. Ailes, USN,** Deputy Commander, Fleet Readiness Directorate Space and Naval Warfare Systems Command ***Cybersecure Modular Open Architecture Software Systems for Stimulating Innovation*** Walt Scacchi, University of California, Irvine Thomas A. Alspaugh, University of California, Irvine ***Security Measurement—Establishing Confidence That System and Software Security Is Sufficient*** Carol Woody, Software Engineering Institute ***Decision Support for Cybersecurity Risk Assessment*** Hanan Hibshi, Carnegie Mellon University Travis D. Breaux, Carnegie Mellon University

Rear Admiral John W. Ailes, USN—is a 1985 graduate of Oregon State University with a Bachelor of Science in Computer Science, and he was commissioned through the Naval Reserve Officers Training Corps (NROTC) program. He holds a Master of Science in Electrical Engineering from the Naval Postgraduate School.

His sea tours include strike warfare/communications officer, USS *Bunker Hill* (CG 52); combat systems officer, USS *Elliot* (DD 967); combat systems officer, USS *Russell* (DDG 59), and executive officer, USS *Lake Erie* (CG 70). Ailes was the commissioning commanding officer, USS *Chafee* (DDG 90).

Ashore, Ailes has served as the test and evaluation manager for the Navy Theater Ballistic Missile Defense Program; Aegis Theater Ballistic Missile Defense warfare area manager; Naval Integrated Fire Control project officer and surface ship combat systems manager in the Program Executive Office, Integrated Warfare Systems; commander, Naval Surface Warfare Center, Port Hueneme Division; major program manager, Littoral Combat Ship Mission Modules; and chief engineer, Space and Naval Warfare Systems Command.

Ailes assumed his current position as deputy commander, Fleet Readiness Directorate, Space and Naval Warfare Systems Command, in March 2016.

Ailes' decorations include the Legion of Merit Medal (two awards), Meritorious Service Medal (three awards), Navy and Marine Corps Commendation Medal (two awards), Navy and Marine Corps Achievement Medal, and various campaign, unit, and service awards.

Cybersecure Modular Open Architecture Software Systems for Stimulating Innovation

Walt Scacchi—is Senior Research Scientist and Research Faculty Member at the Institute for Software Research, University of California, Irvine. He received a PhD in information and computer science from UC Irvine in 1981. From 1981 to 1998, he was on the faculty at the University of Southern California. In 1999, he joined the Institute for Software Research at UC Irvine. He has published more than 200 research papers and has directed more than 70 externally funded research projects. In 2011, he served as co-chair for the 33rd International Conference on Software Engineering Practice Track, and in 2012, he served as general co-chair for the Eighth IFIP International Conference on Open Source Systems. [wscacchi@ics.uci.edu]

Thomas A. Alspaugh—is a Project Scientist at the Institute for Software Research, University of California, Irvine. His research interests are in software engineering, requirements, and licensing. Before completing his PhD, he worked as a software developer, team lead, and manager in industry. He also worked as a computer scientist at the Naval Research Laboratory on the Software Cost Reduction, or A-7, project. [thomas.alspaugh@acm.org]

Abstract

Our interest is to stimulate the development of innovative approaches to continuously assuring the cybersecurity of open architecture (OA) software systems. We focus on exploring the potential for using blockchains and smart contract techniques and how these techniques can be applied to support acquisition efforts for software systems for OA command and control, or business enterprise (C2/B) systems. We further limit our focus to examining the routine software system updates to OA software configuration specifications that arise during the development and evolution processes arising during system acquisition. We discuss new ways and means by which blockchains and smart contracts can be used to continuously assure the cybersecurity of software updates arising during OA software system development and evolution processes. We present a case study examining the software evolution process that updates an OA C2/B system to describe these details. We then discuss some consequences that follow for what emerges from these innovations in the expanded scope of cybersecurity assurance of not just the delivered OA C2/B software systems, but also in the engineering processes which create, transform, or otherwise update technical data that is central to the acquisition of OA software systems.

Overview

How might we stimulate the development of innovative approaches to continuously assuring the cybersecurity of open architecture (OA) software systems? This is the acquisition research challenge we are addressing. In particular, we are interested in investigating innovations that represent either incremental improvements or substantial departures from our current acquisition practices for such systems. We target our efforts to practical OA software system production, deployment, and sustainment, for applications like command and control or business enterprise (C2/B) systems that are central to the mission and operations of military or industrial enterprises. We seek to stimulate significant innovations that employ emerging concepts and technologies to problems observable in the acquisition, development, and evolution of modern C2/B systems.

Problem

The particular problem we investigate here is how best to develop and demonstrate a new conceptual approach to providing continuous cybersecurity assurance (DoD & GSA, 2013) with OA C2/B software systems in response to evolutionary updates to currently installed software configurations that routinely arise during the technical development and maintenance, upkeep, and sustainment in the field—what we call “software evolution.”

Solution

The innovations we focus our attention to are the concepts, techniques, and technologies that denote blockchains and smart contracts, along with how they can be used to continuously assure the cybersecurity of software updates arising during OA software system development and evolution processes.

Approach

Our efforts focus on an innovative utilization of blockchains and smart contracts within the technical software development and evolution processes that arise within the acquisition of complex OA C2/B software systems. We are not focusing attention at this time to software purchasing activities or financial transactions, though blockchains and smart contracts are likely to stimulate innovations in this aspect of OA software system acquisition.

Why This Approach?

Based on prior studies of issues and challenges arising in the development and evolution of OA software systems for C2/B system applications (Guertin, Sweeney, & Schmidt, 2015; Scacchi & Alspaugh, 2012–2017; Womble et al., 2011), we have already drawn attention to technical problems that arise in the software engineering processes that software producers, system integrators, and customer end-users (both enterprises and individuals therein) experience. But we recognize these processes are partially-ordered sets of activities whose completion often entails technical data transactions like creation of digital system design documents, composition and integration of software components (e.g., applications, mobile apps, plug-in widgets), and deployed software executable/update packages that are stored, installed, and tracked in different online repositories across a network environment. At present, these transactions often lack a common or centralized repository for tracking these diverse transactions across networked platforms that span an OA software system ecosystem (a supply chain network from producers to system integrators to customer enterprises/individuals). We believe blockchains are a candidate for this. These transactions similarly lack a common and potentially reusable specification for how to manage and track such software engineering transactions in forms that are open to independent validation and audit. We believe smart contracts are a candidate to address this.

Background: Blockchains and Smart Contracts as Ledgers and Contractual Agreements for Tracking and Managing Transactions

Blockchains are a 21st century computational mechanism for realizing the equivalent of the traditional bookkeeping ledger utilized in finance and accounting. Such ledgers record and track the assignment of incoming (budget authorization or revenue) and outgoing (allocations and expenses) enterprise transactions and denominated amounts, whether in a monetary currency, bartered trade, or some other transactional resource (e.g., gold bullion, Bitcoins, original artworks; DuPont & Maurer, 2015). Such transactions are grouped in blocks; for example, a set of interrelated OA software system updates may be grouped together into a block that denotes a transformation of the current system configuration into an evolved system configuration. Both transactions and blocks are serialized, logged, timestamped, and tracked in ways that are open to internal, external, or independent verification and audit by decentralized third-parties (“Blockchain,” n.d.). Updates to the blockchain are allowed only by consensus of remote mechanisms and proofs of work by anonymous, untrusted service providers (called miners) who collect a modest execution fee for their efforts. The payment and deposit of an execution fee also mitigates against the actions of unknown others who might act to corrupt the blockchain state. Finally, blockchains can be realized as persistent databases or cloud-based repositories (“Blockchain,” n.d.). Figure 1 displays a traditional centralized ledger versus a decentralized blockchain ledger.

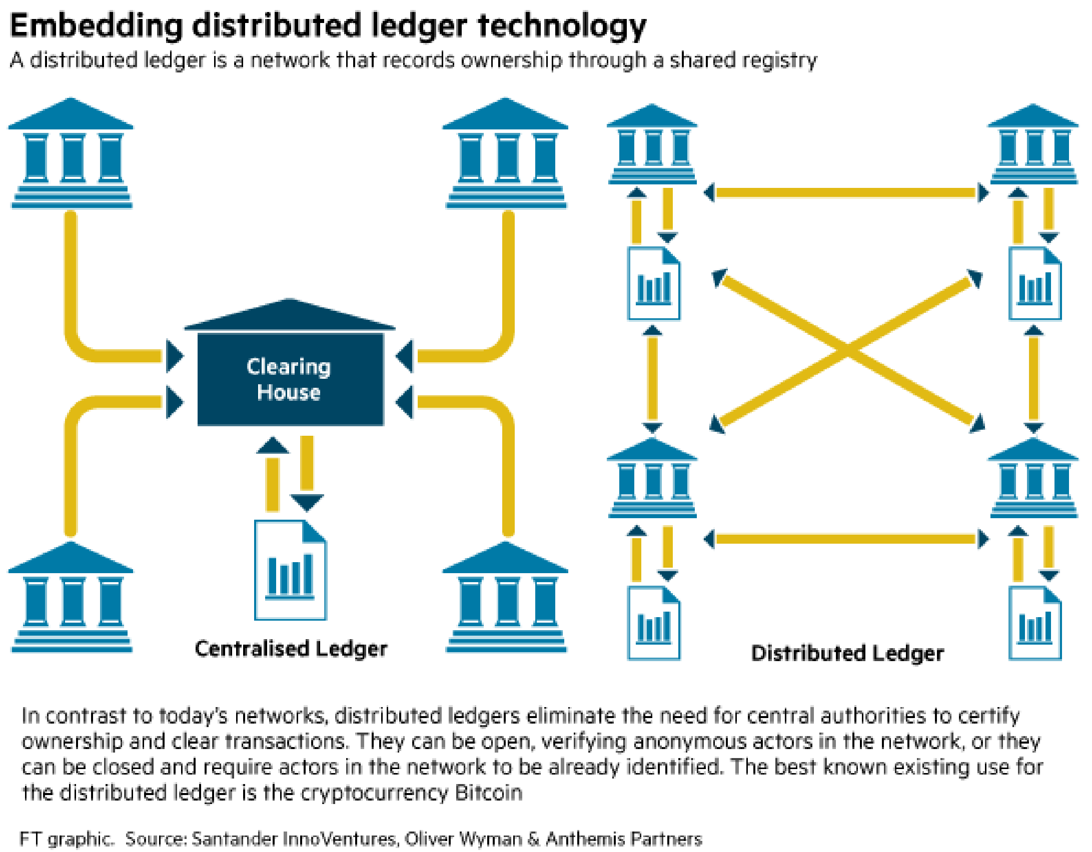


Figure 1. Traditional Ledger Network (Left) and Decentralized Blockchain Ledger Network (Right)

Blockchains operate as an append-only data structure or database maintained by a decentralized collection of mutually distrusting computational nodes participating in a peer-to-peer network. Blockchains are secure by design (“Blockchain,” n.d.). Blockchain ledgers are updated (appended) as a result of recorded transactions, much like a personal bank account is updated through deposit, withdrawal, credit, or debit transactions made by the account holder, through a third-party (the bank or transaction system processor), who may charge a fee for transactions. Much like bank account transactions, blockchain update transactions are distributed over a network, time-stamped, persistent, and verifiable. However, the peer-to-peer network of blockchain nodes is a decentralized autonomous authority without legal standing, compared to the centralized authority taken by a bank or credit/debit card transaction processor.

Smart contracts are similarly the computational counterparts of traditional paper contracts for how a group of interrelated transactions will be governed to assure fulfillment of terms, conditions, rights, and obligations. Such transactions, for example, may be associated with the acquisition of a complex system or with the ongoing procurement of retail supply purchasing agreements. These smart contracts denote networked software system protocols that facilitate, verify, or enforce the negotiation or performance of a specified contract, and thus denote which transactions to process (where, when, how, and for what parties) in what order ("Smart Contracts," n.d.). They are realized using computer-based, formal specifications of transaction-based processes that can be codified into executable computer programs. Such computational support allows for modeling, analysis, and simulation of transactions or processes that can be enacted, verified, and validated at Internet-time speeds, with precision and automated recall of transaction details well beyond what enterprises have traditionally performed. Smart contracts also allow for the establishment and operation of decentralized autonomous services that allow for cooperating parties to enact and fulfill the details of a shared contract through only automated means. Next, smart contracts are automatically enforced by the consensus mechanism associated with the blockchain. Smart contracts are thus attractive to use to securely manage recurring transactions between known or unknown parties, such as those associated with updating the technical data, source code, repositories, and related artifacts associated with the software development and evolution processes of large, long-term software acquisition efforts.

Blockchains are being extended to accommodate smart contracts that allow for the formation of virtual, decentralized autonomous organizations that act to govern, enforce, and assure the integrity and validity of complex or idiosyncratic blockchain update transactions ("Smart Contracts," n.d.). For example, multi-party agreements whereby two or more program offices or other enterprises can act to share the procurement costs of a new C2/B system application or component of mutual interest to the participating parties (Reed et al., 2012; Reed et al., 2014; Scacchi & Alspaugh, 2015). Similarly, smart contracts can govern transactions between mutually distrusting participants that are automatically enforced by automated consensus mechanisms associated with blockchain updates. This capability thus provides a mechanism for detecting, rejecting, or preventing unauthorized update transactions to the blockchain, as might be attempted via a cyber attack during OA software system development or evolution. Accordingly, our interest is to investigate how blockchains, smart contracts, and related technologies can be utilized to improve cybersecurity, specifically to manage and track software engineering development and evolution processes that entail process transactions that update the configuration of OA software systems.

So how might we utilize blockchains and smart contracts to innovate the continuous development and evolution of OA systems? How can this be conceived and applied in ways that are not specifically limited to financial transactions commonly associated with system acquisition?

Blockchains and Smart Contracts for Installed Software Configurations

How might we utilize blockchains and smart contracts to record, track, and verify updates to OA software system configurations as they evolve over time? We examine this question in this section.

Ledgers of Installed Software Configurations

We envision a new kind of ledger: one that records executable computational updates to the specification of the current installed, operational configuration of C2/B systems of interest. The executable computational updates are similar to scripts in a declarative scripting language, like those used to direct the invocation of utilities on an operating system, procedural scripts involved in building (compiling and integrating) a targeted software executable, or for customizing the functional display and navigation operations within a Web browser. We call the repository in which this specification is recorded the *installed software configuration* (ISC) ledger. Such a specification is a kind of technical data to be managed with an acquisition effort.

The ISC ledger records the transactions that update the software applications, including their components, interconnections, interfaces, or licenses for such installed on each machine of interest, such as a desktop PC, smartphone, or central computation server within a mission command or enterprise data center. The installation is enacted via an installation (update) transaction, which may be enabled using an "installation wizard" for a standalone PC application, or using a ready-to-install packaged software app acquired from an online app store. For each application installed, the ledger lists the repository from which the software app or update was acquired, the version of the application or update, and some information with which to confirm/verify the version, such as the size of that version of the app, meta-data about where it resides in storage on the machine, other information, or a combination of these. How do we ensure that the repository's copy is safe, has not been unintentionally modified, and has not been attacked or unknowingly compromised? How do we ensure that attacks are not falsely recorded in the ledger?

In order for a ledger to be up-to-date, each approved installation must be recorded there. How do we ensure this is the case for approved installations? If a ledger is up to date, then an auditor can verify the approved installations by examining the ISC specification for the machine of interest (e.g., a smartphone or laptop PC). Furthermore and most importantly, the blockchain can be queried to identify non-approved or non-compliant installations and whether these are apps or updates that were innocently installed but not recorded in the ledger or attacks—maliciously injected software for some nefarious purpose, which would not be recorded in the ledger. In either case, the auditor can then institute for each application that does not match the ledger a rollback to a known safe ISC state matching what has previously been verified on the ledger.

The following issues must be managed appropriately for the ledger scheme to succeed.

- ***How is it ensured that the origination or destination repository's copy is safe and has not been attacked?***
  This is a separate concern, and one that is equally problematic with or without a ledger system. We do not discuss it further here, merely noting that it must be ensured for devices to remain secure. But in normal operation, the ISC specification has a unique identifier in the hashcode value associated with the current system when last updated and verified by miners, and this hashcode may reveal whether the ISC specification copy's hashcode matches the one checked during audit or subsequent miner verification activities. If the hashcode values are different, then something has altered the copy, and thus it may be rolled back to a prior verified state or ISC specification.
- ***How is it ensured that every approved installation or update is recorded in the ledger?***
  The ledger system must be integrated with whatever system manages installations and updates for the machines in question. We note that unapproved installations or updates can be automatically detected and can be rolled back or reverted at the next audit point/event, so there will be a strong motivation to ensure that desired transactions are recorded.
- ***How do we ensure that attacks are not falsely recorded in the ledger?***
  Obviously this is a key concern. As discussed later, changes to the ledger are

validated by multiple autonomous parties (miners) using several sources of information, and each particular copy of a ledger competes with all others for accuracy as part of the blockchain scheme.

Transactions for Installed Software Configurations

Each transaction in a ledger records an installation or update of an app on a specific machine. How do we ensure that all valid installations or updates are presented? Every time a new application is installed, or an existing application is updated, the appropriate information is recorded in the ledger. If an application is installed or updated without being recorded in the ledger, that installation or update is recognized as unverified, and thus rolled back the next time the machine is audited. Audits may simply involve checking a hashcode value (a long, non-guessable string of characters that is generated within the blockchain system) associated with the current ISC specification on the target machine, with the corresponding value in the blockchain (this is a simple match-checking query that can be performed periodically), or by enterprise policy. When the audit reveals a mismatch, then a rollback may be triggered that reverts the ISC on the machine to a previously trusted ISC, and then removes, deprecates, or flags the unverified ISC as suspect, along with distributing a notification to relevant parties of such action following enterprise policy. But how do we ensure that only valid installations or updates are presented? Transactions that would record an invalid installation or update, fraudulently misrepresenting the repository's version's size or hash or from an untrusted repository, are identified by comparison with the set of trusted repositories, with the size and hash information recorded there for the installation or update in question, and for the data calculated from the destination machine afterwards. Accordingly, we are acting to use blockchain techniques as intended, but for a new kind of use case, namely that of ISC specification update, verification, and reconciliation.

Smart Contracts for Installed Software Configurations

A smart contract works within the framework of the blockchain ledger and transaction system, ensuring that the required obligations for each transaction are met before the transaction is enacted, verified, and then recorded in the ledger. These obligations are associated with those we have previously identified and specified as security requirements for ensuring access and update rights encoded in a software system's security license (Alspaugh & Scacchi, 2012).

An Example Ledger, Transaction, Smart Contract Implementation System

Ethereum is being used to implement smart contracts, transactions, and a blockchain ledger ("Ethereum," n.d.). Ethereum is a set of technologies: a general-purpose programming language, open application program interfaces (APIs), and an open transaction/blockchain repository associated with the APIs. Ethereum uses a cryptocurrency called *ether*, and users of Ethereum can transfer money, ownership, or control of exchanged resources whose (fungible) value is denominated in the form of ether between each other and to contracts to hold in escrow. Online currency exchange markets can exist for converting ether to a traditional currency like U.S. dollars. Users of Ethereum send transactions to it in order to create contracts, invoke existing contracts, and transfer ether. The transactions are public and permanently recorded in the blockchain, unless access to the blockchain is restricted/private to an authorized set of known parties who must be granted permission to access or update the blockchain.

Ethereum is decentralized, with a network of blockchains for which each transaction is processed by a number of *miners*, possibly anonymous actors, who perform computations on the blockchain that collectively verify the validity of a transaction of data/value between the participating parties. These miners are mutually-untrusted peers who are paid fees (in ether) for the work of processing each transaction and its contract provisions. A miner groups transactions into blocks and performs a calculation (or "solves a puzzle") that takes as inputs the previous block in the blockchain and the transactions in the new block. A *valid* block, one whose puzzle has been solved and which meets certain other conditions, can be appended to the blockchain. The miner broadcasts the new valid block to the network and receives the ether paid for each of the transactions by their originators. In this way, Ethereum-based smart contracts are validated by decentralized miners who receive payment when contracted transactions they verify are successfully appended by consensus to the blockchain.

A transaction may appear in a number of different blocks, produced by different miners and appended to different blockchains. Ethereum pays miners somewhat more to append a block to a longer blockchain, which has the effect, over time, of converging the ledger to the blocks and thus transactions that the majority of miners agree are valid.

Continuous Software Development and Evolution Processes for Open Architecture Software Systems

In previous work, we have identified and substantiated seven types of software evolution process update transactions, shown in Figure 2. We further observe that a given software evolution process may entail either (a) one type of transaction per update, or (b) multiple concurrent types of updates per transaction. This may be due to current-to-evolved transformations where the evolved system version of the OA configuration involves the replacement of more than one component arising from the availability of a new technology that represents a departure from the current system architecture, or that integrates functionally similar capabilities through a new mix of components, interfaces and interconnections (e.g., when combining multiple widgets into mashups; Endres-Niggemeyer, 2013). The purpose may be to reduce software maintenance complexity and extend the sustainability of a deployed current (or legacy) system through adoption and integration of remote (cloud-based) services that are functionally similar to the capabilities formerly available in multiple components. For example, replacing legacy office productivity applications (word processor, email, calendar) with browser-based remote networked services (Google Docs, Microsoft Office 365), can provide end-users with functionally-similar processing capabilities, but with fewer application components installed on the end-user's desktop PC system. Furthermore, subsequent updates to remote services may by policy be integrated and deployed automatically for minor functionally equivalent evolutionary updates (e.g., bug fixes), or be deployed only by request or authorization when functionally similar system version updates are made available (Scacchi & Alspaugh, 2013a, 2015, 2016, 2017).

Blockchains and Smart Contracts for Managing Software Development and Evolution Process Transactions

How might we utilize blockchains and smart contracts to manage software development or evolution updates to OA software system configurations over time? We examine this question in the following section.

Ledger: What Versions of What Software Components and Connectors Are Integrated in What OA Configuration Topology

A ledger records and defines through the design-time OA specification, the ecosystem in which the OA is evolving (Scacchi & Alspaugh, 2012). The OA is represented using an architecture description language, and successive ledger entries record successive configurations of the OA system as it evolves. The ledger as a whole presents the history of the OA's evolution, and as long as the components and connectors remain available from their repositories, an instance of any stage of the OA can be rebuilt as needed. At a minimum the ledger records every release of the OA system.

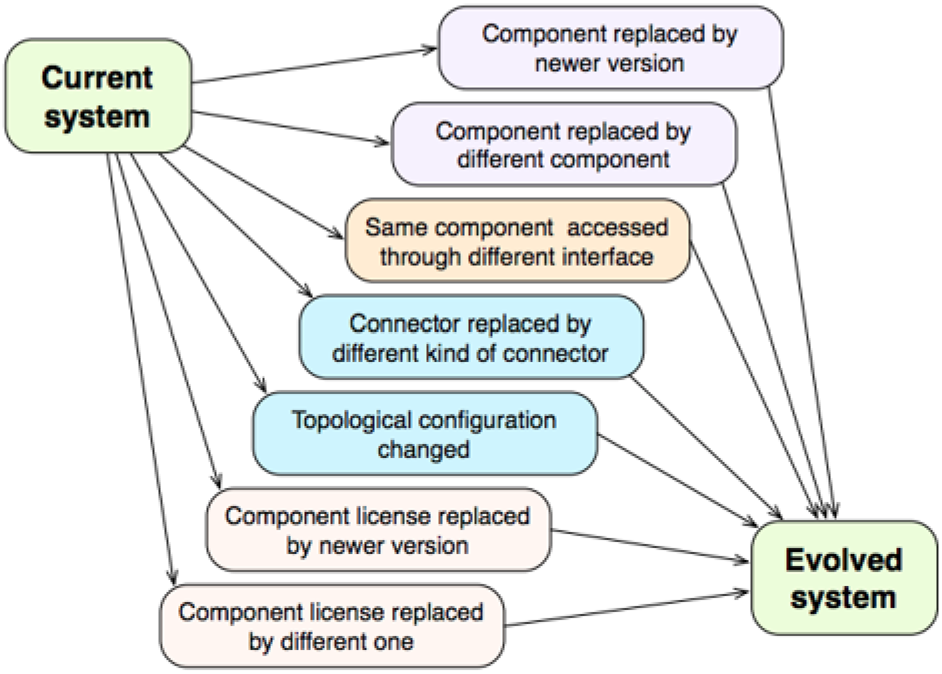


Figure 2. Seven Types of Software Evolution Update Transactions
(Scacchi & Alspaugh, 2013)

If a machine on which the OA ISC is installed needs to be rolled back to an earlier configuration, the desired version of the ISC can be rebuilt guided by the corresponding ledger entry.

Transactions: OA Evolution Steps

Each transaction corresponds to one (or several) of the seven types of OA evolution, stating which component, connector, or license is being changed or what change is being made to the OA topology. In total, the sequence of all transactions for an OA system represents the history of its evolution. The ledger summarizes the system's evolution, based on the transactions made to it, and presents each of the versions that the evolution has proceeded through.

Not everyone can record a transaction with the ledger, and each actor that can record a transaction may be restricted in precisely what sorts of transactions can be recorded. These restrictions ensure that the OA ISC is evolved through steps that preserve its security. It also accommodates actors who may or may not have been vetted and authorized so that they are trusted to preserve the system's security through their transactions.

Smart Contracts: Enforcing Obligations for Each OA Evolution Step

Smart contracts restrict the transactions that may occur to those believed to preserve the OA system's security as the system evolves. A transaction may only be enacted if the actor doing to has been vetted and authorized for it, and has presented credentials identifying himself appropriately; and also only if the current state of the OA system development and the evolution step(s) proposed meet the conditions imposed by a smart contract associated with the ledger. The smart contract in essence states obligations that the actor, the evolution step, and the OA system must meet in order for the transaction to occur; if the obligations are not met, then the transaction cannot be performed, at least not with this smart contract. The obligations declared in a smart contract indicate which parties or actors can access/update what OA system elements or other technical data arising during software development or evolution processes. As before, these process obligations are similar to those previously identified for controlling software system/data usage obligations, along the rights to access and update the system/data provided to developer, system integrators, or end-users (Alspaugh & Scacchi, 2012).

It is possible that more than one smart contract may potentially allow a specific transaction, each contract presenting a different set of obligations. But in any case the transaction cannot proceed until a smart contract for the ledger allows it to do so.

To help make clear what we are looking to accomplish through our efforts to stimulate innovation in securing the development and evolution of OA software systems, we now turn to present a case study focusing on updating the installed software configuration of a deployed current OA C2/B software system.

Case Study: OA C2/B Software System Evolution Process Updates

In this case study, we describe how blockchains and smart contracts can be employed to model and analyze cybersecurity requirements for OA software systems that arise during the software evolution processes. As described previously, there are seven types of software evolution process updates that take a current system, transform it one of the seven ways, which produces an evolved system. This evolution process iteratively cycles through software development processes that build, release, and deploy (Scacchi & Alspaugh, 2013b, 2017) installed software configurations once the development life cycle starts. The process continues to (slowly) cycle over time, until the system is retired or abandoned. Our focus further narrows to evolving OA C2/B systems that incorporate multiple end-user computing platforms, such as smartphones, tablets, or other Web-compatible “edge” devices (Zheng & Carter, 2015), as we have addressed before (Scacchi & Alspaugh, 2015, 2016).

Blockchain ledgers serve to verify in a decentralized manner the proper sequencing of valid transactions for a user/device account. Such an account operates like a personal bank account that can be used to deposit and withdraw funds (e.g., through account transactions associated with a debit/credit card that is bound to the account). The enterprise that manages accounts for users may charge a fee for account transactions, though such fees may be assigned to a third-party (e.g., the party who receives payment via a card that has been authorized to possess sufficient funds balance to cover the payment in the future). The current “balance of funds” in a software evolution process account indicates the name, size, and other meta-data that identify executable software applications (including mobile apps, plug-in widgets, or other installed software). At present, computing platforms or devices do not maintain software process transaction accounts, but in our scheme they would.

Next, the blockchain ledger as a decentralized database would be distributed across a (virtual private) network of computing systems, such as those with restricted, authenticated access to a centralized C2/B system host/sub-network. Said differently, if we have smartphones or mobile/laptop PCs that can roam in the wild, and intentionally or unintentionally acquire software updates (e.g., known app updates but with revised access rights; new social media apps; or cyber-penetration attack vectors via misdirected access to a remote server), we want all such evolutionary software update transactions to be reconciled and validated against the corresponding virtual private network's blockchain ledger in ways that maintain device/user autonomy, but reveal and can reject unvalidated evolutionary updates. The ways and means for how valid or invalid transactions are revealed (externally documented on the blockchain) or rejected (e.g., enforced automated uninstallation, external network access blocked, or notify user of problematic update) are determined by enterprise cybersecurity policies encoded into an associated smart contract (a functional software program logically isolated from end-user application software).

Let us consider the following usage scenario. Suppose we have a mission platform like a battleship or a multi-ship flotilla (or, alternatively, an aircraft flight wing, a ground-based command post, or remote enterprise business office) assigned to operate within an international location. Such a location may be in a region known to have a history of prior cybersecurity attacks on personal computers, mobile, or Web-based devices that access the public Internet. Mission personnel are restricted by policy from using their enterprise mobile devices outside the cybersecurity perimeter of the mission platform. However, personnel may also possess and use private personal devices, such as low-cost smartphones that are used for non-mission purposes.

As anyone who possesses and routinely uses a mobile/edge device like a smartphone or laptop PC now frequently experiences, software (evolution) updates are common, sometimes one or more per week across the 30–60+ apps found on such devices. Sometimes mistakes are made by personnel regarding which device to use for accessing remote services like making phone calls to home, to informally coordinate with friends in allied forces, to check for local restaurants offering interesting local cuisines, or to post data for sharing on social media. Access control to some devices may be misconfigured due to a prior update or unintentionally left open in a discoverable device pairing mode, so that other unknown devices or remote computers can quietly/covertly make network connections that enable data/files upload, download, or remote control. Mobile or web-based edge devices will be relentlessly targeted for cyber attack, so when a cyber attack vulnerability is in the hands of opposing forces or hostile competitors, we assume they will seek out and attack these vulnerabilities at some time and place. It is therefore these invalid software evolution updates to installed software configurations that denote potential cyber attacks that we seek to detect, isolate, trace, expunge or prevent, using the capabilities of blockchains and smart contracts. In this way, our use of blockchains and smart contracts is innovative, original, and not previously associated with software evolution process transactions.

Consider a desktop PC with apps/widgets acquired from either a restricted-access enterprise-specific app store, a Defense app store (George et al., 2014; George, Morris, & O'Neil, 2014), or else from a public-access app store or OSS component repository. Web browser-based apps like cloud-based word processors, calendars, and email app services are frequently included in such stores. However, open access app stores (like those operated by Apple, Google, Microsoft, and others) also offer free/low-cost apps that offer many other remote, cloud-based services. In either situation, these remote service apps may operate downloaded software code that runs within a platform-based Web browser that accesses public or (virtual) private networks. Enterprise end-users with computer programming expertise may even create and integrate multiple apps/widgets into mashups as a kind of end-user software evolution process update (Endres-Niggemeyer, 2013; Scacchi & Alspaugh, 2015). These mashups may enable the participating apps/widgets to interoperate, exchange or update local data, or transfer data/files to/from remote networked repositories (Scacchi & Alspaugh, 2015, 2016).

If our mobile device is a laptop PC, its current (or legacy) OA software configuration may include open source software (OSS) or proprietary closed source software (CSS) versions of a common Web browser, word processor, email, calendar, and more hosted on the PC's operating system. For instance, a laptop may have a Firefox web browser (OSS), AbiWord (OSS) or Microsoft Word (CSS) word processor, Gnome Evolution (fOSS) or Outlook (CSS) for email and calendaring, and host a PC operating system like a Fedora/Linux distribution (OSS), Microsoft Windows (CSS), or Apple OSX (CSS and OSS). The deployed, run-time executable version of this OA ISC system on the laptop PC may appear to an end-user as an array of loosely-coupled applications, such as displayed in Figure 3. Now, suppose a decision has been made to update this OA ISC system, to evolve it from the current configuration to one where the word processor, email, and calendaring applications hosted on the laptop PC are to be replaced with functionally similar remote Web services that will operate within the existing Web browser. These remote services thus entail reliance and usage of browser-based software components that are hosted in the cloud and downloaded on user demand. This transition can simplify and reduce the costs of corresponding software update services associated with locally hosted applications (e.g., recurring license fees for CSS elements). The resulting deployed and evolved laptop PC software system may appear to the end-user as shown in Figure 6.

Each type of software evolution process update can have a smart contract associated with it. Each such contract programmatically specifies what computational actions need to be performed to complete the transaction with the affected technical data and associated data repositories, and similarly, what actions need to be performed on the blockchain. Let us consider the following transformation of a current ISC shown in Figure 3 to an evolved ISC seen in Figure 6. Figure 3 corresponds to its ISC model visualized in Figure 4, which is derived from its specification in an architectural description language (ADL), as we have established before (Alspaugh, Asuncion, & Scacchi, 2013a; Alspaugh, Scacchi, & Asuncion, 2010). As the current system, we assume for this moment, that it has previously been submitted via an earlier transaction on the blockchain that was verified by miners and thus is now a recorded part of the blockchain. Thus we can determine the provenance of the current ISC system and its specification. This blockchain contains a record of the ISC specification and the results (e.g., blockchain hashcode values) that the miners computed and agreed by anonymous vote to denote the ISC installed and operational on the target machine/platform. The transformation from this current system to the evolved system thus entails enaction of the associated smart contracts associated with a set of embedded evolution update transactions that collectively denote what updates must be verified as a block for the evolved ISC specification to be appended to the blockchain.

For example, we may elect to use a predefined smart contract (an executable software script) whose transactions transform a component-based C2/B system with a Web browser installed, into a remote service-based C2/B system, where Web/cloud-based services provide functionally similar capabilities to end-users. This might entail a smart contract that performs the following transactions (described in English for simplicity): (1) check that the ISC blockchain hashcode value(s) match those for the current system; if matching, then proceed; (2) deprecate and replace designated software application components with remote service apps/widgets; (3) replace deprecated component licenses with remote services licenses (e.g., ToS); (4) replace ISC interconnection topology with the evolved ISC; (5) send request to miners to independently compute and verify the evolved ISC specification hashcode value on the target machine/platform denotes the ISC and associated meta-data they independently build to compute the evolved ISC hashcode; (6) if miners' vote independently verifies the ISC specification, then assert into the blockchain the evolved ISC specification value as denoting the new current ISC ready for use; and (e) perform end of contract transactions. Many low-level details are not described here, but would need to be in a smart contract. These details can include, for instance, the installation parameter settings that are selected or configured by either the end-user or installation script, in line with a security technical implementation guide (STIG) for the targeted machine/platform.

The software evolution conveyed in the smart contract example will change the topological configuration of software components found in the system integration build specification, release, and deployed run-time architectures. Here we see that in Figure 5, the configuration model of the evolved OA system still incorporates the same kind of components as the current system model (shown in Figure 4), but now the topology of components interconnections and interfaces has been updated to realize the deployed, run-time desktop software. Last, a transformation from the current software components with their respective licenses, to the evolved configuration will also entail an update to new licenses (e.g., Google Terms of Service), and how these components will be secured (from end-user level assurance of locally installed components to end-user agreement with remotely provided component security that is mostly invisible to end-users).

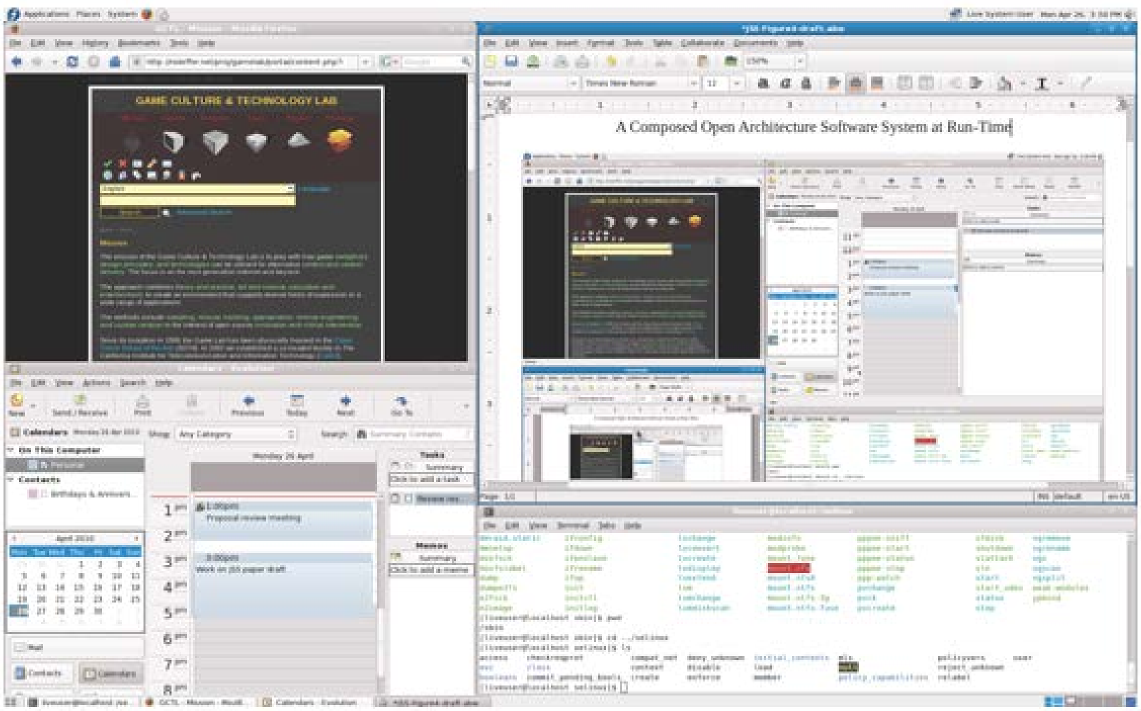


Figure 3. Current Deployed OA ISC Corresponding to Figure 4, Utilized by End-Users

Note. Firefox Web Browser (Upper Left), Evolution Calendar (Lower Left), AbiWord Word Processor (Upper Right), and Fedora/Linux Desktop Operating System Platform (Lower Right)

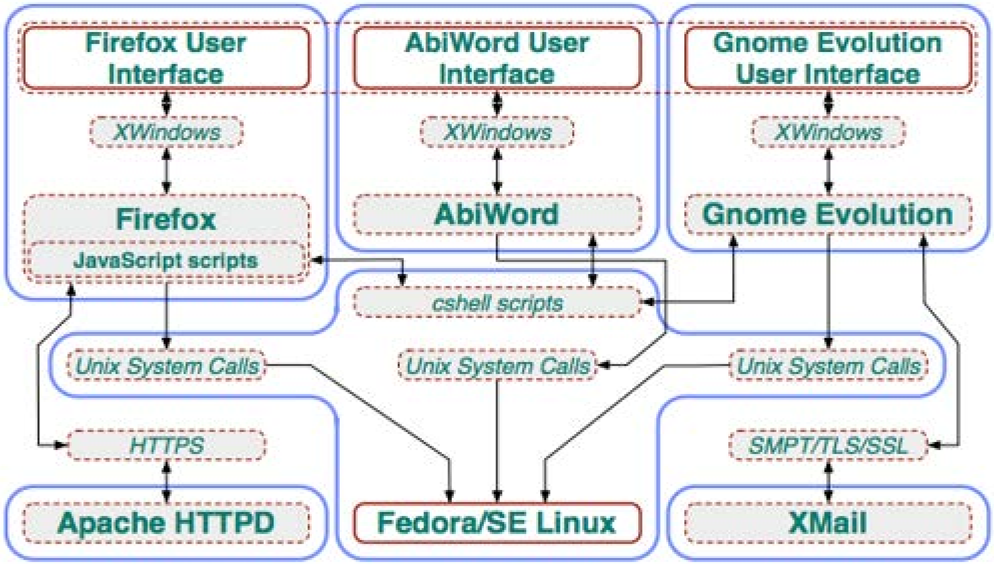


Figure 4. Current ISC Specification for OA C2/B System
(Scacchi & Alspaugh, 2013, 2017)
Note. This is the current ISC specification for an OA C2/B system within security containers at build-time, intended to denote a record on the blockchain for which components need to be included during integration (and testing) of the software components and code APIs within the released and deployed ISC.

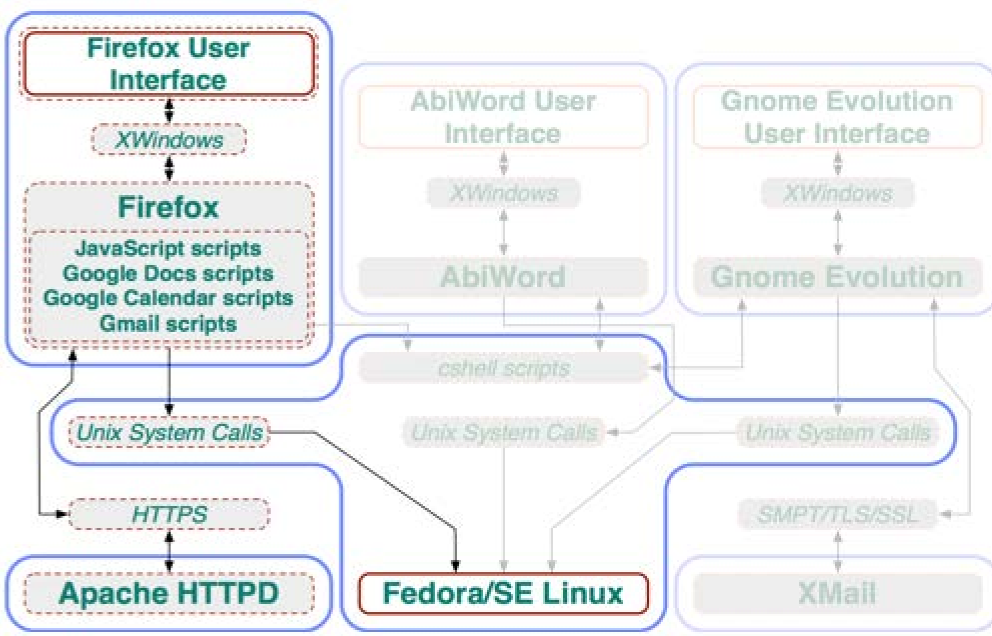


Figure 5. The Evolved OA ISC Specification at Build-Time
Note. The topology of the ISC has evolved, including where now legacy components have been deprecated and likely marked for eventual removal, so as to eliminate any residual vulnerability pathway still present.

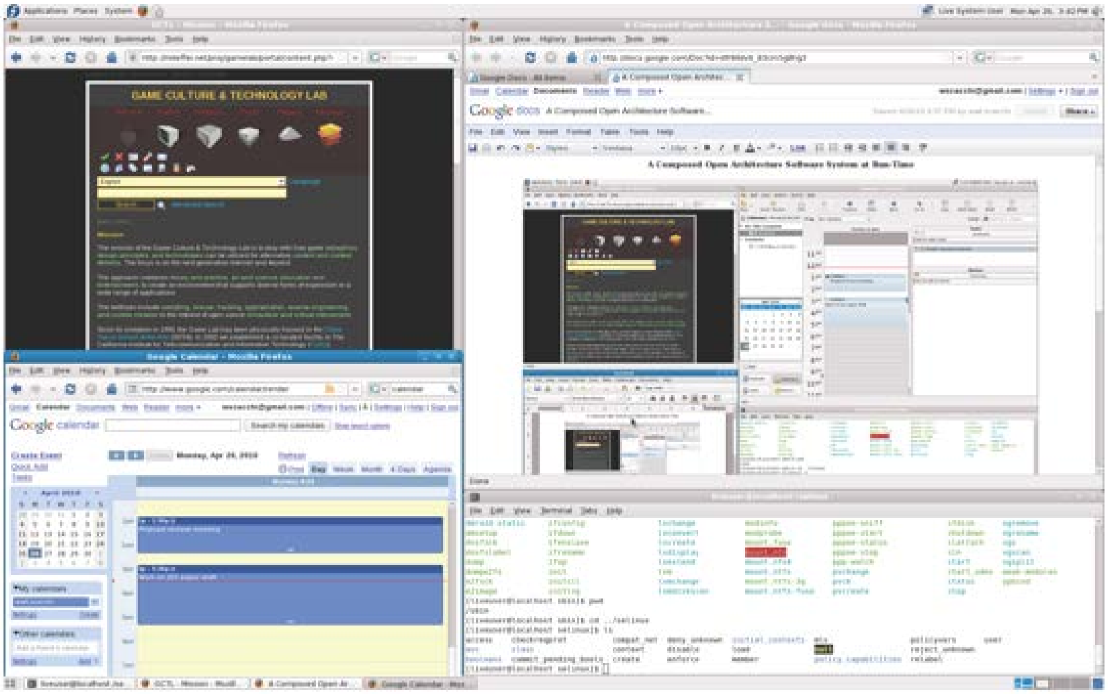


Figure 6. Evolved OA ISC Corresponding to Figure 5, Installed for Utilization by End-Users

Note. Firefox Web Browser as Before (Upper Left), Google Calendar (Lower Left), Google Docs (Upper Right), and Fedora/Linux Operating System Platform as Before (Lower Right)

The transformation of the current system in Figure 3 and Figure 4 to the evolved system in Figure 5 and Figure 6 entails multiple types of software system evolution updates. But now we must consider whether and how such evolution process transactions potentially allow for introduction of cybersecurity vulnerabilities or attack vectors. This can happen, for instance, in the following ways: If the current system is trusted, because its components have individually had their security tested for known vulnerabilities and have passed assurance checks, then evolution process update transactions may introduce unintended vulnerabilities, either within the components replaced, within the new topological configuration, via shifts in the obligations or rights (added, subtracted, revised) in the new components, or via the overall incorporation of all of these evolutionary updates. So we need to assure the security of the update transactions acquired from the component producers and from the system integrators.

As these transactions entail request-response transactions with remote parties across a network, then they may be vulnerable to “man-in-the-middle” attacks, as well as to mistakes made in selecting the appropriate component versions for the specific edge device platform. So we want these transactions to be coordinated and tracked using blockchains and smart contracts, so that we can better trust the security of the evolution process updates. Said differently, we want any and all updates that affect the OA software system components, interconnections and interfaces, or licenses to be mediated and verified by remote parties via blockchain transactions. This entails that each edge device or system platform must be able to periodically (e.g., daily, after an application program exits, or by mission-specific policy) identify itself and assert the “value” of its current ISC elements and configuration specification, in a way that can be reconciled against the last known corresponding verified values on the blockchain. If a discrepancy between the value of the last known (and trusted) current system configuration, and the system evolved configuration is detected, then some unknown evolution update has occurred, such that system security is now unknown and may no longer be trusted. Such a condition may then produce a notification of such discrepancy, automatically revert to the last known trusted current system, or some other intervention action, depending on the evolution process update security policies expressed in the corresponding smart contract. Subsequently, we now have new ways and means for assuring, detecting, or preventing authorized/unauthorized evolutionary changes to an OA ISC during the software development and evolution processes which occur routinely during a system acquisition effort.

Overall, the purpose of this case study is to help describe and reveal that common and widespread acquisition processes associated with the development, usage, or evolution of OA software systems supporting C2/B mission applications is not necessarily secure, and thus can allow for unknown or poorly understood evolutionary updates that are intended or not. Our efforts begin to characterize the need to continuously secure and assure these software engineering process updates and their provenance. Such continuous assurance capabilities are needed in addition to other techniques that focus on assuring the security and integrity of the individual software components acquired from diverse producers or integrators through software ecosystems that release deployable run-time software applications or remote services.

Discussion

There are three topics we find merit consideration, given what now appears possible in the use of blockchains and smart contracts as mechanisms for assuring software development and evolution process update transactions for OA C2/B systems. These are (a) how cyberattacks that may potentially arise in traditional software engineering processes can now be prevented, detected or marked for action; (b) innovations in acquisition research that may follow; and (c) future extensions of this line of research and study.

Cyberattacks on Software Evolution, Release, and Update Processes

The types of software evolution updates in Figure 2 also classify comparable types of attacks on OA systems during their development, build, deployment, and run time processes. The difference being that cyberattacks on software denote unauthorized or unverified updates from the current ISC during design-time, build-time, and deployment-time software engineering activities, to an evolved ISC. This implies that covert software evolution changes by an attacker may follow the same steps as those by a trusted software producer or system integrator, namely replacement of a component by a newer version or by a different component, access to a component through a different interface, replacement of a connector, or replacement of the topological configuration. (We are presently unaware of attacks involving replacement of a component license, but such attacks that change/rewrite IP or security license obligations and rights are clearly possible [Scacchi & Alspaugh, 2012, 2015, 2016].) The result is a compromised version of the system that is functionally similar to the current (trusted) ISC system, but masquerading as one that is authorized, validated, and functionally equivalent intended not to be recognized as something different.

When the attack is made on a deployed instance of the ISC system, its presence can be identified by the change in the size or hashcode value of the compromised system, compared to the current system's provenance established in the blockchain. The window of time during which the attacked system may take effect is limited by the frequency with which the edge device's software is compared with what the blockchain ledger recorded as being installed, as after any change is discovered the edge system's software can be rolled back to its (prior, now current) trusted configuration.

The process is more complex for attacks during development, build, and deployment, because the context is more complex. Here we wish to prevent insecure components, connectors, and configurations from being incorporated into the OA system, but an OA system is by its nature typically the result of a distributed, decentralized development, with components coming from other projects and developed and evolved by parties distant and often unknown to the OA system's integrators. We foresee the use of blockchains, transactions, and smart contracts to record each component and connector's provenance, vetting, and authorization. Smart contracts restrict the possible transactions (evolution steps) to those believed to preserve the OA system's security. When an unexpected change is discovered in an edge device system's software, it is rolled back to a safe version; when a security fault is discovered in a version of the system, a process that may be much more involved, the components, connectors, and topology involved may be rolled back to a trusted safe version, and the smart contracts through which the fault was introduced may be updated to prevent a "similar" evolution in the future. This may be done either by withdrawing authorization from actors involved, by blacklisting a component repository whose vetting was careless, or by similar means. The blockchain ledger records the information needed to take such steps.

This points to two further areas of research. First, the blockchain ledger system now becomes a locus against which attackers will wish to operate, and further study is needed to examine how to resist such attacks, isolate their effect, and to the extent possible reject them through the blockchain and transaction mechanism itself. Second, can the ledger be used as a database of information for effectively distinguishing fraudulent or corrupted evolution steps? Further research will be necessary.

The only allowed OA evolution updates of the secure system are those that are first verified as valid updates, from known trusted parties, and that satisfy a contract for the blockchain ledger. In cases where a vulnerable or corrupted component, connector, or topology successfully runs this gauntlet, the ledger provides a means for rolling back transactions to a secure version of the system that can be deployed in place of the insecure later version.

We note that in contrast to a procedural programming language such as the Solidity language used for Ethereum contracts, a declarative scripting language mitigates against recently discovered vulnerabilities of smart contract technologies, such as those found for the Ethereum run-time interpreter (Atzei, Bartoletti, & Cimoli, 2016).

Innovation for Acquisition Research

The work prior to this paper in software cybersecurity is primarily focused on making a particular version of the software system itself, as a product, secure. In this paper, we are expanding our view to include the ecosystem within which the system evolves, the software architecture specification that defines and constrains that ecosystem, the evolution of the components and connectors that are integrated into the system, and the OA evolution process by which any OA system evolves from version to version. To this, we are adding the ability to record, track, verify, and maintain the security of the OA system throughout its development and evolution processes.

We are proposing the use of blockchains and smart contracts to assure the security of software engineering process update transactions. We are not at this time investigating how blockchains and smart contracts may be used as potential mechanisms that support the financial transactions or new business models for purchasing the services or products associated with a OA software system acquisition (Scacchi & Alspaugh, 2016). That is a topic for future research. Similarly, though blockchains and smart contracts are relatively new, they also entail their own set of vulnerabilities associated with their different technological implementations (Atzei, Bartoletti, & Cimoli, 2016) that must be addressed. Whether or how such vulnerabilities may manifest within acquisition processes is also a topic for future research.

Future Extensions and Elaborations of This Approach

We have discussed the application of a blockchain system for coordinating and steering the evolution of an OA software system that is produced or integrated by a single party. But a blockchain system is by its nature a distributed system, and though its distributedness does not in itself give extra benefit in multi-producer, multi-integrator software ecosystems, clearly it is as effective in recording evolution and provenance in them, and it is already adapted to the challenges of interactions with many parties.

In our prior research, we have called for a declarative domain-specific language (DSL) for specifying the obligations and rights incorporated into IP and security licenses for OA software (Alspaugh & Scacchi, 2012; Scacchi & Alspaugh, 2013a). Now we see that such a DSL can be extended to incorporate software engineering process transactions using a process modeling language like PML (Noll & Scacchi, 2001; Scacchi, 2001) or a similar notation and that such extension is advantageous for managing OA software security system and engineering process challenges. The design and incorporation of these extensions into the DSL is thus a next step for us to research, develop, and refine.

Last, we have also called for research and development of software obligations and rights management systems (SORMS) as a core capability for the DoD, government agencies, and other enterprises to help manage and improve their OA software system buying power (Scacchi & Alspaugh, 2015, 2016). We envision a SORMS that interprets and evaluates DSLs for software licensing as an essential tool for enterprises that manage OA software systems, such as those found in most large organizations in industry, government, and defense. Thus, we call for effort to add capabilities that extend the SORMS to accommodate blockchain ledger repositories, as decentralized or centralized databases, on which are enacted smart contracts for handling software development and evolution process update transactions.

Conclusions

In this paper, we sought to stimulate the development of innovative approaches to continuously assuring the cybersecurity of open architecture (OA) software systems. We focused on exploring the potential for using blockchains and smart contract techniques and how they can be applied to support acquisition efforts for software systems for OA command and control or business enterprise (C2/B) systems. We further limited our focus to examining the routine software system updates to OA software configuration specifications that arise during the development and evolution processes arising during system acquisition. Our efforts described through our case study and related efforts thus denote a promising line of work in progress. Much remains to be done, but the direction forward appears robust and productive. We welcome questions and comments that identify possible oversights, and we suggest complementary capabilities that enhance the potential of blockchain and smart contract tools, techniques, and technologies for continuously assuring the cybersecurity of modular open architecture software systems.

References

Alspaugh, T. A., Asuncion, H., & Scacchi, W. (2013). The challenge of heterogeneously licensed systems in open architecture software ecosystems. In S. Jansen, S. Brinkkemper, and M. Cusumano (Eds.), *Software ecosystems: Analyzing and managing* *business networks in the software industry* (pp. 103–120). Northampton, MA: Edward Elgar.

Alspaugh, T. A., & Scacchi, W. (2012, September). Licensing security. In *Proceedings of the Fifth International Workshop on Requirements Engineering and Law* (pp. 25–28).

Alspaugh, T. A., Scacchi, W., & Asuncion, H. A. (2010). Software licenses in context: The challenge of heterogeneously licensed systems, *Journal of the Association for Information Systems, 11*(11), 730–755.

Atzei, N., Bartoletti, M., & Cimoli, T. (2016). *A survey of attacks on Ethereum smart contracts*. Retrieved from http://eprint.iacr.org/2016/1007.pdf

Blockchain. (n.d.). In *Wikipedia*. Retrieved March 15, 2017, from https://en.wikipedia.org/wiki/Blockchain

DoD & GSA. (2013, November). *Improving cybersecurity and resilience through acquisition*. Retrieved from https://www.gsa.gov/portal/getMediaData?mediaId=185367

DuPont, Q., & Maurer, B. (2015, June 23). Ledgers and law in the blockchain. *King's Review.* Retrieved from http://kingsreview.co.uk/articles/ledgers-and-law-in-the-blockchain/

Endres-Niggemeyer, B. (Ed.). (2013). The mashup ecosystem. In *Semantic mashups: Intelligence reuse of web resources* (pp. 1–50). Springer.

Ethereum. (n.d.). In *Wikipedia*. Retrieved March 15, 2017, from https://en.wikipedia.org/wiki/Ethereum

George, A., Galdorisi, G., Morris, M., & O'Neil, M. (2014, June). DoD application store: Enabling C2 agility? In *Proceedings of the 19th International Command and Control Research and Technology Symposium* (Paper 104). Alexandria, VA.

George, A., Morris, M., & O'Neil, M. (2014). Pushing a big rock up a steep hill: Lessons learned from DoD applications storefront. In *Proceedings of the 11th Annual Acquisition Research Symposium* (Vol. 1, pp. 306–317). Monterey, CA: Naval Postgraduate School.

Guertin, N. H., Sweeney, R., & Schmidt, D. C. (2015, April). How the Navy can use open systems architecture to revolutionize capability acquisition: The Naval OSA strategy can yield multiple benefits. In *Proceedings of the 12th Annual Acquisition Research Symposium* (Vol. 1, pp. 107–116). Monterey, CA: Naval Postgraduate School.

Noll, J., & Scacchi, W. (2001). Specifying process-oriented hypertext for organizational computing. *Journal of Network and Computer Applications, 24*(1), 39–61.

Reed, H., Benito, P., Collens, J., & Stein, F. (2012, June). Supporting agile C2 with an agile and adaptive IT ecosystem. In *Proceedings of the 17th International Command and Control Research and Technology Symposium* (Paper 44). Fairfax, VA.

Reed, H., Nankervis, J., Cochran, J., Parekh, R., & Stein, F. (2014, June). Agile, adaptive IT ecosystem: Results, outlook, and recommendations. In *Proceedings of the 19th International Command and Control Research and Technology Symposium* (Paper 11). Arlington, VA.

Scacchi, W. (2001). Redesigning contracted service procurement for internet-based electronic commerce: A case study. *Journal of Information Technology and Management, 2*(3), 313–334.

Scacchi, W., & Alspaugh, T. (2012, July). Understanding the role of licenses and evolution in open architecture software ecosystems. *Journal of Systems and Software, 85*(7), 1479–1494.

Scacchi, W., & Alspaugh, T. (2013a, February). Advances in the acquisition of secure systems based on open architectures. *Journal of Cybersecurity & Information Systems, 1*(2), 2–16.

Scacchi, W., & Alspaugh, T. (2013b, May). Processes in securing open architecture software systems. *In Proceedings of the International Conference on Software and Systems Process* (pp. 126–135). San Francisco, CA.

Scacchi, W., & Alspaugh, T. (2014). Achieving Better Buying Power through cost-sensitive acquisition of open architecture software systems (NPS-AM-14-C11P07R01-036). In *Proceedings of the 11th Annual Acquisition Research Symposium*. Monterey, CA: Naval Postgraduate School.

Scacchi, W., & Alspaugh, T. (2015). Achieving Better Buying Power through acquisition of open architecture software systems for web and mobile devices (NPS-AM-15-005). In *Proceedings of the 12th Annual Acquisition Research Symposium*. Monterey, CA: Naval Postgraduate School.

Scacchi, W., & Alspaugh, T. (2016). Achieving Better Buying Power for mobile open architecture software systems under diverse acquisition scenarios (SYM-AM-16-033). In *Proceedings of the 13th Annual Acquisition Research Symposium* (pp. 163–183). Monterey, CA: Naval Postgraduate School.

Scacchi, W., & Alspaugh, T. A. (2017). Issues and challenges in the operations and maintenance of open architecture software systems. *CrossTalk: The Defense Software Engineering Journal,* 10–15.

Smart Contracts. (n.d.). In *Wikipedia*. Retrieved March 25, 2017, from https://en.wikipedia.org/wiki/Smart_contract

Womble, B., Schmidt, W., Arendt, M., & Fain, T. (2011). Delivering savings with open architecture and product lines. In *Proceedings of the Eighth Annual Acquisition Research Symposium* (Vol. 1, pp. 8–13). Monterey, CA: Naval Postgraduate School.

Zheng, D., & Carter, W. (2015). *Leveraging the internet of things for a more efficient and effective military.* Center for Strategic & International Studies. Retrieved from https://www.csis.org/analysis/leveraging-internet-things-more-efficient-and-effective-military

Acknowledgments

The research described in this report was supported by grant #N00244-16-1-0053 from the Acquisition Research Program at the Naval Postgraduate School, Monterey, CA. No endorsement, review, or approval implied. The concepts, topics, and materials presented are the sole responsibility of the authors.

Security Measurement—Establishing Confidence That System and Software Security Is Sufficient

Carol Woody—PhD, has been a senior member of the technical staff at the Software Engineering Institute since 2001. She is the Technical Manager of the CERT Cybersecurity Engineering team which addresses security and survivability throughout the development and acquisition life cycles, especially in the early stages. Her research focuses on building capabilities for measuring, managing, and sustaining cybersecurity for highly complex networked systems and systems of systems. She coauthored *Cyber Security Engineering: A Practical Approach for Systems and Software Assurance*, published in November 2016 by Pearson Education, InformIT, as part of the SEI Series in Software Engineering. [cwoody@cert.org]

Abstract

Evaluating the software assurance of a product as it functions within a specific system context involves assembling carefully chosen metrics that demonstrate a range of behaviors to establish confidence that the product functions as intended and is free of vulnerabilities. The first challenge is to establish that the requirements define the appropriate security behavior and the design addresses these security concerns. The second challenge is to establish that the completed product, as built, fully satisfies the specifications. Measures to provide assurance must, therefore, address requirements, design, construction, and test. We know that software is never defect free. According to Jones and Bonsignour (2012), the average defect level in the United States is 0.75 defects per function point or 6,000 per million lines of code (MLOC) for a high-level language. Thus, software, on average, cannot always function perfectly as intended. Additionally, we cannot establish that software is completely free from vulnerabilities based on our research which indicates that 5% of defects should be categorized as vulnerabilities. So how can we establish reasonable confidence in software security? To answer this question, the Software Engineering Institute (SEI) is researching how measurement can be used to establish confidence in software security. This paper will share our progress to date.

Is the System Secure?

The Department of Defense (DoD) has followed a well-structured acquisition and development life cycle for decades. The rules for these activities are clearly laid out in Department of Defense Instruction (DoDI) 5000.2 (DoD, 2003)[1]. There are a series of activities and milestone reviews that require a program office to demonstrate to management and oversight groups that good engineering is underway as the acquisition of a system moves from an idea through to final implementation. Can we leverage this process focus to establish confidence that software assurance is well-integrated into each of the life cycle steps? The current acquisition focus is on establishing appropriate requirements and ensuring these are met in the delivery of the system.

The evaluation of the security of the same system is equally as well-structured using the guidance provided in DoDI 8510 (DoD, 2014). These guidelines were recently rewritten to broaden the focus beyond the security technology controls to include protection of critical

[1] The most recent version effective February 2, 2017, is available at http://www.acqnotes.com/wp-content/uploads/2014/09/DoD-Instruction-5000.02-The-Defense-Acquisition-System-2-Feb-17-Change-2.pdf

assets based on potential risks. If selected controls are implemented to protect these key assets and these are validated through the certification and accreditation reviews, is the system secure? It is hard to say since the engineers are focused on requirements and the security analysis is focused on security controls. How sufficient are the security requirements? Requirements are subject to change as the system progresses into the life cycle, but the security controls are not typically reviewed after the initial selection.

The acquisition is focused on delivery of a system, and software is typically viewed as a necessary component. In reality, software is quickly becoming the major component. Software handled 8% of the functionality of the F-4 Phantom fighter in 1960, but by 2000, its role grew to 80% for the F-22 Raptor, and the trend does not appear to be abating (National Research Council, 2010). We know that software is never defect free. According to Jones and Bonsignour (2012), the average defect level in the United States is 0.75 defects per function point or 6,000 defects per million lines of code (MLOC) for a high-level language. Very good code levels would be 600 to 1,000 defects per MLOC, and exceptional levels would be below 600 defects per MLOC. Thus, software, cannot always function perfectly as intended since there will always be defects. Additionally, we cannot establish that software is completely free from vulnerabilities based on our research which indicates that 5% of defects should be categorized as vulnerabilities (Woody, Ellison, & Nichols, 2014). Hence, as the role of software increases, the availability of software vulnerabilities also increases. It is impossible to avoid the constant news that systems are under attack and vulnerabilities are so prevalent that attackers are successful. Software security is a growing concern and needs to be effectively managed as part of an acquisition. Can we do this?

The Program Protection Plan[2] assigns a software assurance reporting responsibility to the program office, but they are typically not the group building and maintaining the software, so they pass this responsibility to the contractor. How will a program office know if there is sufficient software security in the system the contractor delivers? Few engineers in a program office are trained in security, and even with training, will they be able to directly evaluate the product? Maybe not, but they should be able to evaluate the quality of the processes used by the contractor in building, integrating, and verifying the system. Higher quality with fewer defects, along with a focus on software security, should result in fewer defects and fewer vulnerabilities.

The program office can ask the contractor to report on a wide range of metrics. There are metrics for cost, schedule, quality, complexity, resiliency, and technical debt just to list a few of the categories. Capers Jones (2015) reports that over 3000 different metrics are in use, and most of them are inconsistent at best and typically misleading. In his 2015 report, Jones notes “the software industry labors under a variety of non-standard and highly inaccurate measures compounded by very sloppy measurement practices.” What are the metrics for software security? We can count vulnerabilities just as we count defects. Can we assume all code will be of high quality so that defects and vulnerabilities are at a minimum? High quality development requires resources with the capability to deliver high quality as well as effective processes to ensure the results delivered meet expectations. What kind of useful measurements can we apply to these?

[2] See Defense Acquisition University (DAU; n.d.) for a description of the Program Protection Plan.

Using Engineering Evidence to Reduce Fear, Uncertainty, and Doubt (FUD) for Software

For a measurement to be useful in engineering, there are three questions that need to be addressed:

- Are we measuring the right things at the right time?
- Are our measurements trending in the right direction?
- Do we collect information soon enough to react to problems within other constraints?

As an example, let's consider how these questions are addressed in the acquisition of an airplane. How do we establish that the plane will fly when it is delivered? The planes are engineered to meet requirements that are defined for the expected use. The Dreamliner is designed to carry up to 330 passengers on long distances in comfort (Boeing, n.d.). It is reported to have a range of 11,910 kilometers, or 6,430 nautical miles. The wing span is 197 feet and 4 inches, height is 55 feet and 10 inches, and the length is 224 feet. All of these characteristics are determined based on lift and speed and other aerodynamic characteristics to allow the plane to meet its flight requirements. The F35 has a very different set of requirements as a single-seat, single-engine, all-weather stealth fighter plane (Lockheed Martin, n.d.). It is designed for a maximum speed of Mach 1.5 at altitude, with a range of approximately 1,620 nautical miles using internal fuel. The wing span is 32.78 feet, height is 13.33 feet, length is 50.5 feet, and wing area is 450 feet. Each of these planes is built in a highly structured manufacturing operation using best practices for constructing and testing the parts and validating the assembled whole. Reviews are conducted at scheduled times throughout the development and testing of prototypes is extensive to ensure requirements are met. There should be no surprises at the point of delivery about the plane's ability to meet its flight requirements (NASA, 2009). To apply this approach to software security, we need effective processes and a means of measuring how well they are working.

Building Blocks for Engineering Software Security

After determining that there were many software security practices, but nothing structured for evaluating software assurance, SEI undertook the task of developing the Software Assurance Framework (SAF)[3]. The SAF defines a set of cybersecurity practices that programs should apply across the life cycle and supply chain. The SAF can be used to assess a program's current practices to identify gaps and chart a course for improvement. By verifying and identifying improvements for a program's cybersecurity practices relevant to software, the SAF helps to (1) establish confidence in the program's ability to acquire software-reliant systems that are sufficiently secure, and (2) reduce the cybersecurity risk of deployed software-reliant systems. When developing the SAF, we leveraged the software acquisition and cybersecurity expertise of the SEI's technical staff and also referenced a variety of acquisition, development, process improvement, and cybersecurity documents including the following:

[3] A technical note, CMU/SEI-2017-TN-001, that provides a detailed description of the key practices selected for the SAF, is in the publication process and is expected to be available on the SEI website this spring.

- National Institute of Standards and Technology (NIST) Special Publication 800-53, titled *Security and Privacy Controls for Federal Information Systems and Organizations* (NIST, 2013)
- NIST Special Publication 800-37, titled *Guide for Applying the Risk Management Framework to Federal Information Systems: A Security Life Cycle Approach* (NIST, 2010)
- DoDI 5000.2, titled *Operation of the Defense Acquisition System* (DoD, 2003)
- Capability Maturing Model® Integration (CMMI®; CMMI, 2007)
- Build Security In Maturity Model (BSIMM; McGraw, Migues, & West, 2015)

The selected practices fall into four general focus areas: process management, engineering, project management, and support. Within each of these general areas we have grouped practices within subcategories.

Process management would include the following categories of practices:

- process definition
- infrastructure standards
- resources
- training

Engineering would include the following categories of practices:

- product risk management
- requirements
- architecture
- implementation
- testing, validation, and verification
- support documentation and tools
- deployment

Project management would include the following categories of practices:

- project plans
- project infrastructure
- project monitoring
- project risk management
- supplier management

Support would include the following categories of practices:

- measurement and analysis
- change management
- product operation and sustainment

In order to link the detailed practices from the SAF framework to a specific program to address software assurance, we have used a standard software management technique, Goal-Question-Metric (GQM) Approach developed in the 1980s as a structuring mechanism (Basili, 1984). This is a well-recognized and widely used metrics approach. It requires the establishing of a goal for which we structure questions with associated metrics that begin to answer each question. In the following section, we will show an example of how these building blocks can be used to structure an answer to our question about whether a system is secure.

Engineering for Software Security

If we apply the engineering approach described earlier about the plane's flying to security, we must start with establishing our engineering goal for software security. We would like to establish a requirement that software critical to mission and flight functions have no vulnerabilities, but this is not feasible (Woody, Ellison, & Nichols, 2014). However, we can structure a goal for the airplane that the critical software be of the highest quality such as:

> Mission- and flight-critical applications executing on the plane or used to interact with the plane from ground stations shall be high quality, with no more than 600 defects per MLOC and vulnerability levels below 30 per MLOC.

As the saying goes, quality cannot be tested in, it must be built into the product (Koch, n.d.). In order to meet this goal, the contractor would have to ensure they are building the system using high quality engineering processes at each step of the life cycle. This system goal would need to flow down to each step in building the software. To define how we might measure and monitor these processes to ensure high quality, we can create sub-questions for each major area of software development that reflect the contribution to be made to the system as follows:

- **Software Requirements.** Does the program/project define and manage software security requirements?
- **Software Architecture**. Does the program/project appropriately address security in its software architecture and design?
- **Implementation**. Does the program/project minimize the number of vulnerabilities inserted into the code?
- **Testing**, **Validation**, **and Verification**. Does the program/project test, validate, and verify security in its software components?
- **Support Tools and Documentation**. Does the program/project develop tools and documentation to support secure configuration and operation of software components?
- **Deployment**. Does the program/project consider security during the deployment of software components?

Each of these questions could be addressed within the software life cycle through the selection of appropriate practices, outputs, and metrics that demonstrate quality results. For each of the software development areas, we would want to confirm that best practices are performed and these are producing expected outputs along with metrics appropriate to expected results.

As an example, since much of the concern with software security is tied to vulnerabilities, consider how we could be confident that the number of vulnerabilities introduced into the critical code are minimized. There are several best practices in secure coding that we would expect to be performed as follows:

- Secure coding standards are applied.
- Code developers are trained in the use of secure coding standards.

- Evaluation practices (e.g., code reviews and apply tools) are applied to identify and remove vulnerabilities in delivered code (including code libraries, open source, and other reused components).

In addition, we should expect to see outputs and metrics that reflect that these practices are appropriately addressed as shown in Table 1.

Table 1. Activities/Practices, Outputs, and Metrics

Activities/Practices	Outputs	Candidate Metrics
Secure coding standards are applied	Policy that requires the use of secure coding standards Contract language to ensure vendor(s) practices require use of secure coding standards	% of vendor contracts including requirements for the use of secure coding standards % of system developed using secure coding standards % of code verified for secure coding standard conformance
Code developers are trained in the use of secure coding standards	Competency standards for code developers require training in secure coding standards Hiring qualifications require training in secure coding standards Contract language requires use of developers trained in secure coding standards	% of software developers trained in secure coding standards % of code supported by developers trained in secure coding standards
Evaluation practices (e.g., code reviews and apply tools) are applied to identify and remove vulnerabilities in delivered code (including code libraries, open source, and other reused components)	Policy that requires the use of evaluation practices to identify and remove vulnerabilities and reporting of metrics Output of evaluations Corrections documented Contract language requires use of evaluation practices to identify and remove vulnerabilities and metrics reporting	% of vendor contracts requiring use of evaluation practices and reporting of vulnerability metrics Code coverage: % of code evaluated (total and by each type of review) Vulnerabilities per MLOC identified and removed Unaddressed vulnerabilities per MLOC % code libraries evaluated % open source evaluated % legacy components evaluated Count of high priority vulnerabilities identified and count of those removed

Subsequent steps in the development process should continue to confirm that vulnerabilities are at a minimum through testing, validation, and verification.

Conclusion and Next Steps

Returning to our initial question about determining if the system is secure, we have established that our evaluation must focus sufficient attention on the quality and security built into the software which makes up over 80% of the functionality of most systems. If this software is well-defined, well-built, and well-implemented using best practices for engineering software with good security, we should be able to review outputs and confirm through metrics with reasonable confidence that the system is secure. There is information we can collect and evaluate all along the life cycle about the product, processes, and practices to give us confidence in achieving our goal that the final product will be sufficiently secure. System engineering reviews can be used to confirm progress as follows:

- **Initial Technical Review (ITR).** Assess the capability needs (including software security) and materiel solution approach.
- **Alternative Systems Review (ASR).** Ensure that solutions will be cost-effective, affordable, operationally effective. Ensure that solutions can be developed in a timely manner at an acceptable level of software security risk.
- **System Requirements Review (SRR).** Ensure that all system requirements (including security) are defined and testable, and consistent with cost, schedule, risk (including software security risk), technology readiness, and other system constraints.
- **Preliminary Design Review (PDR).** Evaluate progress and technical adequacy of the selected design approach including software security.
- **Critical Design Review (CDR).** Determine that detail designs satisfy the design requirements (including software security) established in the specification and establish the interface relationships.

References

Basili, V. R., & Weiss, D. M. (1984). A methodology for collecting valid software engineering data. *IEEE Transactions on Software Engineering, SE-10*(6), 728–738.

Boeing. (n.d.). Boeing 787. Retrieved from http://www.boeing.com/commercial/787/#/by-design

Chrissis, M. B., Konrad, M., & Shrum, S. (2011). *CMMI for development: Guidelines for process integration and product improvement* (3rd ed.). Boston, MA: Addison-Wesley.

Defense Acquisition University (DAU). (n.d.). *Program protection plan*. Retrieved from https://acc.dau.mil/CommunityBrowser.aspx?id=549148&lang=en-US

DoD. (2003). *Operation of the defense acquisition system* (DoDI 5000.2). Retrieved March 20, 2017, from: http://www.acq.osd.mil/dpap/Docs/new/5000.2%2005-12-06.pdf

DoD. (2014). *Risk management framework for DoD information technology* (DoDI 8510.01). Retrieved from http://www.dtic.mil/whs/directives/corres/pdf/851001_2014.pdf

Jones, C. (2015, January 26). *The mess of software metrics V5.0* (Namcook Analytics).

Jones, C. & Bonsignour, O. (2012). *The economics of software quality* (Addison Wesley).

Koch, A. S. (n.d.). Testing is not quality assurance. PMP. Retrieved from http://www.projectconnections.com/articles/080306-koch.html

Lockheed Martin. (n.d.). F-35 Lightning II. Retrieved from https://www.f35.com/about

McGraw, G., Migues, S., & West, J. (2015). Building Security in Maturity Model (Version 6). Cigital. Retrieved October 10, 2016, from https://www.bsimm.com/

NASA. (2009). *Final report: NASA study on flight software complexity*. Retrieved from http://www.nasa.gov/pdf/418878main_FSWC_Final_Report.pdf

National Institute of Standards and Technology (NIST). (2010). *Guide for applying the risk management framework to federal information systems: A security life cycle approach* (NIST Special Publication 800-37, Rev. 1). Retrieved from http://csrc.nist.gov/publications/nistpubs/800-37-rev1/sp800-37-rev1-final.pdf

National Institute of Standards and Technology (NIST). (2013). *Security and privacy controls for federal information systems and organizations* (NIST Special Publication 800-53, Rev. 4). Retrieved from http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r4.pdf

National Research Council. (2010). *Critical code: Software producibility for defense*. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a534043.pdf

Woody, C., Ellison, R., & Nichols, W. (2014). *Predicting software assurance using quality and reliability measures* (CMU/SEI-2014-TN-026). Retrieved April 2, 2017, from the Software Engineering Institute, Carnegie Mellon University website: http://resources.sei.cmu.edu/library/asset-view.cfm?AssetID=428589

Acknowledgments

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

Decision Support for Cybersecurity Risk Assessment

Hanan Hibshi—is a PhD candidate and a Research Assistant in the Societal Computing program at Carnegie Mellon University. Hibshi's research area includes usable security, security requirements, and expert's decision-making. Hibshi's research involves using grounded theory and mixed-methods user experiments to extract rules for use in intelligent systems. Hibshi received an MS in Information Security Technology and Management from the Information Networking Institute at Carnegie Mellon University and a BS in Computer Science from King Abdul-Aziz University in Jeddah, Saudi Arabia. [hhibshi@cs.cmu.edu]

Travis D. Breaux—is an Associate Professor of Computer Science, appointed in the Institute for Software Research of the School of Computer Science at Carnegie Mellon University. Dr. Breaux's research program searches for new methods and tools for developing correct software specifications and ensuring that software systems conform to those specifications in a transparent, reliable, and trustworthy manner. This includes demonstrating compliance with U.S. and international accessibility, privacy and security laws, policies, and standards. Dr. Breaux is the Director of the Requirements Engineering Laboratory at Carnegie Mellon University. Dr. Breaux has several publications in ACM and IEEE-sponsored journals and conference proceedings. Dr. Breaux is a member of the ACM SIGSOFT, IEEE Computer Society, and USACM Public Policy Committee. [breaux@cs.cmu.edu]

Abstract

The U.S. DoD transition to a multi-tier, risk management framework aims to streamline information assurance assessments by promoting alignment with NIST information assurance control sets. While these control sets are broadly applicable and comprehensive, those responsible for accreditation will continue to struggle with assessing security risk in dynamically reconfigurable systems. Security analysts rely largely on background knowledge and experience to make security-related decisions. With increasingly dynamic software, analysts need to resolve dependencies among components and understand how those dependencies affect security requirements. Analysts need new decision-support tools based on models that predict how analysts reason about security in distributed systems. We present an approach that formalizes security expert assessments of security requirements nested in scenarios into threat mitigation rules. The assessments are collected empirically using factorial vignettes. The vignette results are statistically analyzed to yield membership functions for a type-2 fuzzy logic system. The corresponding type-2 fuzzy sets encode the interpersonal and intrapersonal uncertainties among security analysts in their decision-making. This work establishes an early foundation for a digital cyber-security decision-support service where an IT professional with any level of security background can benefit from efficiently receiving security assessments and recommendations.

Introduction

The U.S. DoD acquisition process goes through well defined and documented security guidelines. Security guidelines and checklists are widely available and well-documented, but organizations and government agencies like the DoD are still relying on human security analysts to evaluate the security of their systems and reason over these guidelines. The DoD transition to a multi-tier, risk management framework aims to streamline information assurance assessments by promoting alignment with National Institute Standards and Technology's (NIST) information assurance control sets (Marzigliano, 2014; Swenson, 2009). The DoD considers NIST security controls to be the minimum and requires an additional set of controls that vendors need to meet before they can work on classified networks (Swenson, 2009). NIST controls such as the 800-53 ("NIST/ITL Special Publication (800)," 2015) would need to go through a process of implementation to create system design and development requirements. Each control represents a class of technology aimed at mitigating a security threat.

Review of the controls is done by human security analysts who are supposed to have sufficient expertise to reason over potentially millions of scenarios that account for various permutations of controls. Under NIST SP 800-53, the analyst decides if a specific system is high, medium, or low impact and then the analyst satisfies the impact rating by selecting security controls (e.g., audit events, lock sessions, etc.).

Human security assessment can be impacted by context, where security requirements apply; *priorities* that some requirements have over other requirements; *uncertainty* due to human's experts' memory constraints; and the *stove-piped knowledge* among security experts who come from a variety of backgrounds, such as systems, networks, databases, and web applications. We try to use an approach that helps address these challenges. Figure 1 summarizes the steps of our overall approach. In each step, we use methods and techniques that would help understand and address the challenges mentioned above. We have completed two phases of this project, where in each phase we sun a series of studies. In the first phase, we ran all the steps shown in Figure 1. In the second phase, we completed steps 1 through 4 and we are still conducting more studies to refine our results. In the upcoming sections, we will explain the studies conducted, research methodology, and technical challenges of each phase.

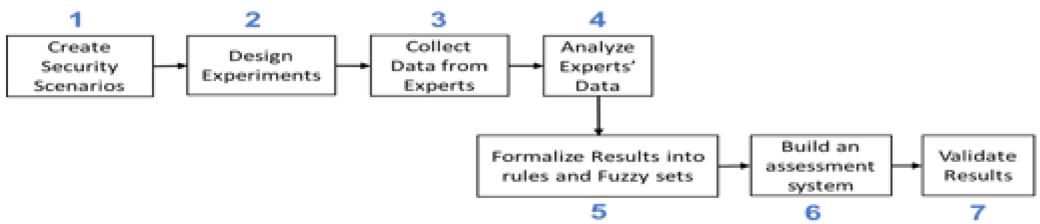
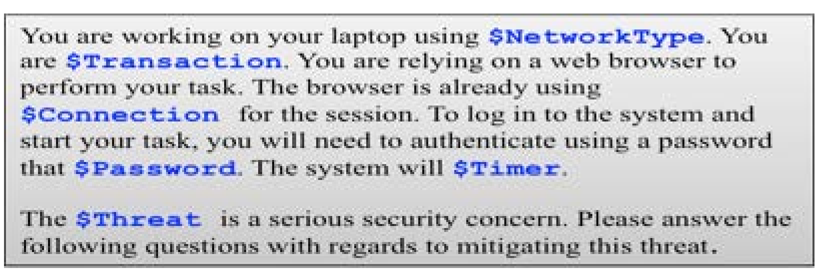


Figure 1. The Overall Process to Build Security Assessment Digital Solutions

Phase 1: Secure Web Transaction Scenarios

Step 1: Creating Security Scenarios

The main theme in this phase is *secure web browsing*. We conducted a user study using scenarios where participants are asked to rate the security of performing an online transaction (e.g., reading email or credit card purchasing) (Hibshi, Breaux, & Broomell, 2015). Participants were presented with a variety of settings (security requirements) that were manipulated throughout the study to measure which different requirements compositions contribute to the overall security of the scenario and to understand the priorities that exist among requirements. Figure 2 shows the template used to generate security scenarios or vignettes.


You are working on your laptop using $NetworkType. You are $Transaction. You are relying on a web browser to perform your task. The browser is already using $Connection for the session. To log in to the system and start your task, you will need to authenticate using a password that $Password. The system will $Timer.

The $Threat is a serious security concern. Please answer the following questions with regards to mitigating this threat.

Figure 2. Vignettes Template Used in Phase 1
(Hibshi et al., 2015)

The template shows variables (starting with $) that are each replaced by a security requirement. We call values that replace the variable *levels* of the requirements variable. We generate different vignettes by using different combinations of requirements levels. In Table 1, we show the variables used in phase 1 vignettes and their levels.

Table 1. Variables Used in Phase 1 and Their Levels

Variable Name	Level(s)
$NetworkType	(EmpNetwork) Your employer's network at your office
	(PublicWIFI) Public unencrypted Wi-Fi at a public area (restaurant, airport)
	(VPNUnencrypted) Your employer's VPN that you connected to through public unencrypted Wi-Fi
	(VPNEncrypted) Your employer's VPN that you connected to through public encrypted Wi-Fi
$Transaction	(E) Accessing your email account and replying to confidential emails
	(F) Performing a financial transaction using your credit card
$Connection	SSL
$Password	(Weak) A password that is at least 8 characters long
	(Strong) A password that is at least 16 characters and must include an uppercase and a lowercase letter, a symbol, and a number digit
$Timer	(Yes) Automatically log you off the session after 15 minutes of inactivity
	(No) Never time-out
$Threat	(Man) Man-in-the-Middle
	(Pac) Packet-Sniffing

This technique of using scenarios with discrete factors that get manipulated to study human judgment is called *factorial vignettes* (Rossi & Nock, 1982; Wallander, 2009). We chose this empirical method because it was shown to be more reliable to evaluate and collect human judgment as compared to direct questioning (Rossi & Nock, 1982; Wallander, 2009). Our purpose is to measure the effect of security requirements' composition on the analysts' risk perception and therefore their overall ratings; to identify priorities among requirements, and to understand the effect of ambiguity on analysts' security decisions (Hibshi et al., 2015).

Step 2: Designing the Experiments

We use a mixed-effect design (a combination of within-subject and between-subject factors) for the user study. We ran two separate experiments. First, we invited participants to evaluate scenarios for *Man-in-the-Middle* threat, then we re-invited them after two weeks for the second experiment where the participants evaluate the *packet-sniffing* threat. In each experiment, each participant is assigned a condition where they see four scenarios. They see all the four different values for the `$NetworkType` variable (within-subject effect), but they only see one value for all the other variables (between-subject effect; see Figure 2 and Table 1; Hibshi et al., 2015).

For each scenario, the participant is asked to rate the overall security of the scenario choosing one of the three following ratings:

- **Excessive** security measures that exceed the requirements to mitigate the threat
- **Adequate** security measures that are enough to mitigate the threat
- **Inadequate** security measures that are not enough to mitigate the threat

After rating the overall security, we ask participants to rate each individual requirement shown in the scenario. The ratings 5-point scale, where point 1 is labeled "inadequate mitigation," point 3 is labeled "adequate mitigation," and point 5 is labeled "excessive mitigation." Participants are also given the opportunity to list additional security requirements that they believe contribute to increasing the security level to adequate (Hibshi et al., 2015).

Participants in this study need to have sufficient security expertise. Therefore, we ask participants to answer a list of security knowledge questions that would help assess their level of security understanding, followed by background demographic questions about their years of expertise, number of courses in security, and their job roles. We also collect general demographics such as age, gender, and highest level of education (Hibshi et al., 2015).

Step 3: Collecting Data From Experts

In this study, we sent invitations to security mailing lists at Carnegie Mellon University and North Carolina State University, and we offered participants a $10 Amazon Gift card as a compensation. A total of 174 participants responded to the Man-in-the-Middle threat survey, of which, 116 returned to respond to the Packet-Sniffing survey. The sample has 101 graduate students, 42 undergraduate students, and 2 university professors (Hibshi et al., 2015).

Step 4: Analyzing Experts' Data

In this step, we use two methods of analysis:

- Multi-level modeling of the user security assessments. This method is suitable to analyze data from our study that instruments a mixed effect design.
- Grounded analysis (Corbin & Strauss, 2007) of additional requirements. We code the statements provided by participants, and then we categorize the codes into one of six categories: server, client, encryption, network, encryption, attack detection/prevention, and integrity and authentication.

Our study results show that security requirements' composition affect the experts' risk perception and security assessment. For example, in scenarios where the password level is strong, participants rated the overall security of the scenario to be less than adequate if the `$Networktype` is public Wifi. Participants view the network to be have higher priority than the three requirements: password, timer, and SSL. Once the `$Networktype` is raised to an adequate level, then other requirements will start impacting the risk assessments (Hibshi et al., 2015).

Step 5: Formalizing Results Into Rules and Fuzzy Sets

We formalize the results of the empirical study to derive if-then rules that we use in a security assessment system based on rule-based interval type-2 fuzzy logic (Hibshi, Breaux, & Wagner, 2016). The following is an example rule that we derived from the results:

$$R^1: IF\ \boldsymbol{NetworkType}\ is\ \boldsymbol{Inadequate}\ THEN\ \boldsymbol{OverallRating}\ is\ \boldsymbol{Inadequate}$$

Any fuzzy logic system needs fuzzy sets that can be constructed using experts' input.Type-2 fuzzy sets allow us to model the uncertainty in the data by providing *a footprint of uncertainty* (FOU; Mendel, 2001). It is important to point out that uncertainty in our data is always present because it relies on experts' input. Experts' data include interpersonal uncertainty, which is the uncertainty between different experts, and intrapersonal uncertainty, which is the uncertainty that the same expert may experience on two different occasions due to the nature of humans' memory (Hibshi et al., 2016).

To build fuzzy sets in our security assessment system, we conducted another empirical study on security experts where we asked participants to provide an interval on a range from 1 to 10 to represent linguistic labels for adequacy (Hibshi & Breaux, 2016; Hibshi et al., 2016). Figure 3 shows the results of the collected intervals from 38 security experts. Then, we use the data collected to construct type-2 fuzzy sets and their membership functions. Figure 4 shows the membership functions for the three fuzzy sets: inadequate, adequate, and excessive.

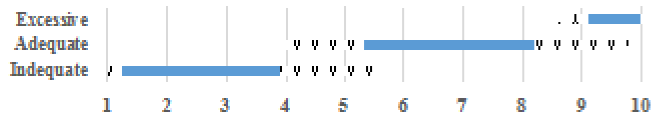


Figure 3. The Fuzzy Sets With the Start and End Means and Standard Deviation (Hibshi et al., 2016)

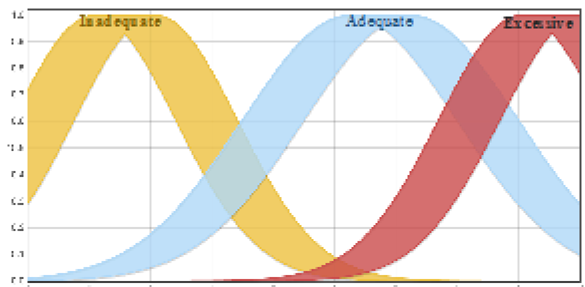


Figure 4. . Membership Function for Inadequate, Adequate, and Excessive (Hibshi et al., 2016)

Step6: Building a Security Assessment System

We will explain how we build and evaluate a security assessment system that would help security analysts evaluate their security decisions. Figure 5 shows the overall architecture of our assessment system.

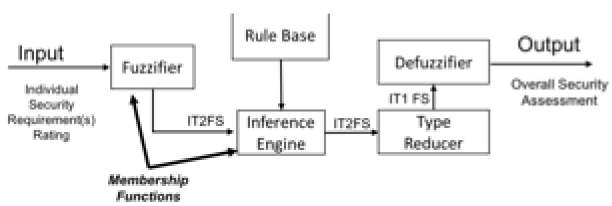


Figure 5. Mamdani IT2FLS for Security Assessment (Hibshi et al., 2016)

We use the type-2 fuzzy sets and rules formalized in step 5 of the process to build our assessment system.

Step 7: Validate Results of the Assessment System

Our goal here is to evaluate the system and to measure how well it mimics human experts' reasoning. Hence, we designed a survey similar to the survey used in the data collection step (see steps 1 and above), and we sent the survey to 13 security experts who rated 52 scenarios (four scenarios per expert). Then, we used the individual security requirements ratings as inputs to our assessment system, which will produce an overall security assessment output. Later, we interviewed the experts and asked them to provide reasoning behind their ratings. Finally, we showed them the system's rating and asked them to compare it to their overall ratings. The results show that the security analysts found the assessment system to provide reliable security ratings, generating more conservative assessments in 19% of the test scenarios compared to the experts' ratings (Hibshi et al., 2016).

Phase 2: Security Administration Scenarios

The studies conducted in phase 1 have the following limitations:

- We used a single security scenario that puts the participant in the role of a user.
- We recruited mostly graduate security and privacy students for the user study.

To address these limitations:

- We increased scenario coverage by selecting scenarios from four security domains: networking, operating systems, databases security, and web applications security.
- We used a language in the scenarios that puts the study participant in an expert role analyzing the requirements shown in the scenarios.
- We recruited security professionals from industry and government.

Figure 6 shows the template used to generate vignettes for the web applications security study.

> You are a website administrator responsible for securing a web app against cyberattacks.
> Currently, you are evaluating the following settings:
> The web app performs `$WebAuth`.
> The web app will `$StoredUserData` in a database for display to other users.
> The Cross-Site Request Forgery attack is a serious security concern. Please answer the following questions with regards to mitigating this threat.

Figure 6. Scenario Template Used in the Web Applications Security Study

The `$WebAuth` variable represents the type of authentication used in the web application and it can take on one of many values. To illustrate, we consider two extremely different values: "basic authentication," which is a weak form of web-based authentication, or "form-based authentication using encrypted credentials stored in a database," which is stronger. Similarly, the `$StoredUserData` variable represents how the user input is being collected and could take the values: "collect user-supplied content from GET request" or "require CSRF tokens and escape and validate user supplied content from POST requests before storing." Again, the latter value is stronger than the former.

In a similar fashion, we constructed scenario templates to generate vignettes for the remaining security domains: networking, operating systems, and database security. Currently, we are still collecting and analyzing data for this phase to help design and build our next security decision-support system.

Summary and Future Work

In this paper, we summarized our research that consists of a series of empirical studies where we study how security experts make their decisions. We used the data collected from experts to formalize and model the human reasoning to build decision-support tools. One of the major challenges in security decision-making is the amount of interpersonal and intrapersonal uncertainties present in the data. Hence, we choose to model experts' data using interval type-2 fuzzy logic, which can handle these uncertainties. We continue to create scenarios, design experiments, and collect data from experts so we can build decision support tools that would better assist the security experts as they make their decisions and evaluate requirements. These smart tools would help security analysts with their acquisition process, as it is a building step towards semi-automating the currently manual process of reviewing systems and evaluating them against security requirements.

References

Corbin, J., & Strauss, A. (2007). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage.

Hibshi, H., & Breaux, T. D. (2016). *Evaluation of linguistic labels used in applications* [Technical report]. Carnegie Mellon University.

Hibshi, H., Breaux, T., & Broomell, S. B. (2015). Assessment of risk perception in security requirements composition. *2015 IEEE 23rd International Requirements Engineering Conference (RE)* (pp. 146–155).

Hibshi, H., Breaux, T. D., & Wagner, C. (2016). Improving security requirements adequacy: An interval type 2 fuzzy logic security assessment system. In *2016 IEEE Symposium Series on Computational Intelligence* (pp. 1–8). IEEE. Retrieved from http://ieeexplore.ieee.org/abstract/document/7849906/

Marzigliano, L. T. (2014, March 14). Defense Department adopts NIST security standards. Retrieved from http://www.informationweek.com/government/cybersecurity/defense-department-adopts-nist-security-standards/d/d-id/1127706

Mendel, J. M. (2001). *Uncertain rule-based fuzzy logic systems: Introduction and new directions*. Prentice Hall PTR.

NIST/ITL Special Publication (800). (2015, January 2). Retrieved from http://www.itl.nist.gov/lab/specpubs/sp800.htm

Rossi, P. H., & Nock, S. L. (1982). *Measuring social judgments: The factorial survey approach*. SAGE.

Swenson, G. (2009, June 11). NIST, DOD, intelligence agencies join forces to secure U.S. cyber infrastructure. Retrieved from https://www.nist.gov/news-events/news/2009/06/nist-dod-intelligence-agencies-join-forces-secure-us-cyber-infrastructure

Wallander, L. (2009). 25 years of factorial surveys in sociology: A review. *Social Science Research, 38*(3), 505–520. https://doi.org/10.1016/j.ssresearch.2009.03.004

THIS PAGE INTENTIONALLY LEFT BLANK

Panel 10. Thoughts on Management of Complex Systems Acquisition

Wednesday, April 26, 2017	
3:30 p.m. – 5:00 p.m.	**Chair: Mark Kitz,** Director, System of Systems Engineering and Chief Engineer, Program Executive Office for Intelligence, Electronic Warfare & Sensors ***A Systems Theoretic-Based Framework to Discover Pathologies in Acquisition System Governance*** Charles B. Keating, Old Dominion University Joseph M. Bradley, Old Dominion University Polinpapilinho F. Katina, Old Dominion University Ra'ed M. Jaradat, Mississippi State University ***Optimal Selection of Organizational Structuring for Complex Systems Development and Acquisitions*** Alexandra Dukes, Purdue University Scott Parrigon, Purdue University Navindran Davendralingam, Purdue University Sang Eun Woo, Purdue University Daniel DeLaurentis, Purdue University ***A Systems Complexity-Based Assessment of Risk in Acquisition and Development Programs*** Antonio Pugliese, Stevens Institute of Technology Roshanak Rose Nilchiani, Stevens Institute of Technology

Mark Kitz—is Director, System of Systems Engineering (SOSE) at the Program Executive Office Intelligence, Electronic Warfare and Sensors (IEW&S), Chief Engineer PEO IEW&S. He was assigned in this position on May 4, 2015. Kitz was most recently the Technical Director, Project Manager Distributed Common Ground Systems–Army (DCGS–A), an ACAT IAM program within the PEO IEW&S portfolio. Prior to DCGS–A, Kitz had multiple assignments as an engineer and scientist with the Intelligence Information Warfare Directorate of the Communication Electronics Research and Development Command, including Chief Engineer and a Project Lead for TROJAN and a Director for the Intelligence Processing Laboratory. Kitz has a master's degree from the New Jersey Institute of Technology and a bachelor's degree from Lafayette College. Kitz is an avid runner and triathlete and the proud father of three young boys, aged 9, 7, and 3.

A Systems Theoretic-Based Framework to Discover Pathologies in Acquisition System Governance

Charles B. Keating[1]—(Attending) is a Professor of Engineering Management and Systems Engineering and Director of the National Centers for System of Systems Engineering (NCSOSE) at Old Dominion University (ODU) in Norfolk, VA. He received a BS in Engineering from the United States Military Academy (West Point), an MA in Management from Central Michigan University, and his PhD in Engineering Management from Old Dominion University. Dr. Keating's current research focus is on complex system governance, system of systems engineering, and management cybernetics. [ckeating@odu.edu]

Joseph M. Bradley—received his PhD in Engineering Management and Systems Engineering at Old Dominion University (ODU) in Norfolk, VA. He holds the degrees of Professional Engineer and Masters of Science in Mechanical Engineering from Naval Postgraduate School and Bachelor of Engineering from The Cooper Union. He is currently president of his own small consulting firm working with clients in government and industry. His areas of research include complex system governance, systems theory, competency models, and performance measurement systems. Dr. Bradley's research has been published in the *Systems Engineering, Naval Engineers Journal*, and *International Journal of System of Systems Engineering*. [josephbradley@leading-change.org]

Polinpapilinho F. Katina—received his PhD in the Department of Engineering Management and Systems Engineering at Old Dominion University (ODU) in Norfolk, VA. He holds an MEng in Systems Engineering (2011) and a BS in Engineering Technology (2009). He is currently a postdoctoral fellow at National Centers for System of Systems Engineering (NCSOSE). Dr. Katina recently co-edited a critical textbook on Infranomics. His research has been published in the *International Journal of Critical Infrastructure Protection, International Journal of Critical Infrastructures, International Journal of Decision Sciences, Risk and Management, International Journal of System of Systems Engineering*, and *Journal of Requirements Engineering.* [pkatina@odu.edu]

Ra'ed M. Jaradat—is an Assistant Professor at Mississippi State University and a Postdoctoral Scholar in the Department of Engineering Management and Systems Engineering at Old Dominion University (ODU). His research interests include engineering management, systems engineering, supply chain management and critical infrastructures, system dynamics, and risk management. Prior to joining the doctoral program at ODU, he worked as a Systems Analyst and Operations Officer for five years. Dr. Jaradat is currently a member of the Academy of Management and Institute of Industrial Engineers. He holds a BS in Business, an MA in Operations Management, and a PhD in Engineering Management from Old Dominion University. [jaradat@ise.msstate.edu]

Abstract

The acquisition field continues to face increasing pressures to perform under conditions of escalating complexity, uncertainty, and ambiguity. These conditions suggest that traditional approaches, practices, and acquisition technologies might be incongruent with support demands for acquisition practitioners. *This research is focused on exploiting and extending recent developments in Complex System Governance (CSG) to advance the acquisition field*. CSG is focused on the design, execution, and evolution of fundamental system functions necessary for control, communications, coordination, and integration of complex systems (e.g., acquisition). CSG is based in Systems Theory (fundamental laws governing complex systems), Management Cybernetics (the science of effective system organization), and

[1] Corresponding author

Governance (provision of direction, oversight, and accountability for systems). Recent advances in CSG (Keating, Katina, & Bradley, 2015) make this an opportune time for exploitation of this field to advance acquisition research and practice in novel ways. Following an introduction and literature review, this paper reports on efforts to (1) establish a systems theory based framework for Acquisition System Governance, (2) mapping of systems pathologies (systemic errors that degrade system performance) to a CSG Reference Model with implications for acquisition practice, and (3) suggests implications for moving CSG forward to improve acquisition practice. The paper closes with directions for bringing CSG to practice through research based application development.

Introduction

The Defense Acquisition System continues to struggle in the midst of increasing levels of complexity, uncertainty, and ambiguity. There are enormous pressures on the acquisition field and practitioners faced with the new realities of increasingly complex systems, resources become further constrained, and expectations for maintenance of cost, schedule, and performance projections escalating. These pressures for change (a.k.a. reform) in the Defense Acquisition System to evolve in response to shifting circumstances is not new or unprecedented. In fact, the history of defense acquisition has been one marked by numerous attempts at reform. For instance, Fox (2011) has traced 50 years of acquisition reform through several periods from the 1960s. Among these reform periods he identified were the following: (1) 1960s and 1970s McNamara innovations and the birth of DoD Directive 5000.1 specifying the Acquisition process; (2) 1980s Carlucci initiatives, PPBS, and weapons acquisition reform; and (3) 1990s, where the DoD 5000 series incorporated less than 50% of reform initiatives. The difficulties and results of Defense Acquisition System reform were captured by Fox's (2012) assertion that

> despite the defense community's intent to reform the acquisition process, the difficulty of the problem and the associated politics, combined with organizational dynamics that are resistant to change, have led to only minor improvements. The problems of schedule slippages, cost growth, and shortfalls in technical performance on defense acquisition programs have remained much the same throughout this period. (p. vii)

Numerous authors have echoed these sentiments. In fact, Schwartz (2013) cites that over 150 major studies examining acquisition reform have been conducted since World War II—with the state of continuing cost overruns, schedule slippages, and missed technical performance remaining. However, Swartz (2013) also cites multiple improvements in the processes used by the DoD to procure goods and services, such as (1) the creation of the Federal Acquisition Regulation; (2) the formation of the Defense Acquisition University: and (3) the institution of such changes to acquisition as multi-year procurement, independent cost estimation, joint requirements board, and movement to commercial technologies. However, despite improvements and initiation of (e.g., should-cost analyses) a recent GAO (2007) assessment of defense acquisition for selected weapon programs concluded that programs continued to experience cost increases and schedule delays. Continuing difficulties in acquisition of defense systems does not to minimize either the dedicated professionals in the acquisition workforce, the well-intentioned aspirations of acquisition institution leaders, or system modifications instituted. Instead, it invites new and novel thinking in response.

Given this seeming consensus on the state of affairs for acquisition, and corresponding attempts for reform, the question is begged, *why does this state of acquisition continue to have difficulty in meeting cost, schedule, and technical performance expectations?* There have been many attempts to explain the factors that delineate the nature and sources of failure in acquisition (Berteau, Levy, Ben-Ari, & Moore, 2011; Francis, 2008, 2009; Rascona, Barkakati, & Solis, 2008). In fact, a look at recent Government Accountability Office assessments of major acquisition programs continues to accentuate difficulties in the acquisition field (Cilli, Parnell, Cloutier, & Zigh, 2015). However, there are continuing efforts at Acquisition System reform (Bucci, Slattery, & Maine, 2015) that recognize the need to streamline the system and craft a more agile and flexible Acquisition System. While these are noble ideals, agreement on how to engage such an endeavor is far from a consensus. The outward appearance of the Acquisition System is that of a monolithic system not well-suited for the complexity, speed, uncertainty, and ambiguity that exist in warfighting needs and environments.

Although the underlying reasons for Fox's (2012) criticisms of the performance of the Defense Acquisition System might be debatable, adequately addressing the continuing difficulties in defense acquisition appears to remain an elusive goal. Providing a sampling of external examinations of past and current reports on defense acquisition program difficulties supports the conclusion that Defense Acquisition continues to struggle, suggesting that "In general, these reports call for early, robust, and rigorous assessments of a broader range of alternatives across a thorough set of stakeholder value criteria to include life-cycle costs, schedule, and performance" (Cilli et al., 2015, p. 587). From a systems perspective of the state of acquisition, we suggest five thought provoking considerations (Figure 1). These considerations provide a systemic frame of reference for the modern landscape of defense acquisition. While these characteristics are endemic to modern systems in general, the particular emphases of the Defense Acquisition System invites a different level of dialog, exploration, and systemic understanding.

1. *Sprawling Complexity Exceeding Absorptive Capacity of the System.* While the technologies and operation and maintenance demands for future weapons systems have continued to rise exponentially, so too has the complexity of systems designed to acquire them. For the Defense Acquisition System this suggests that the calls for reform, increased agility, and other such suggestions by numerous authors, is perhaps best summed up in Kendall's (2014) congressional testimony stating,

    > Our system over time accumulated excessive levels of complex regulatory requirements that are imposed on our program managers and other acquisition professionals. … One thing I hope we can all agree on is the need to simplify and rationalize the bureaucratic burdens we place on our acquisition professionals. (p. 6)

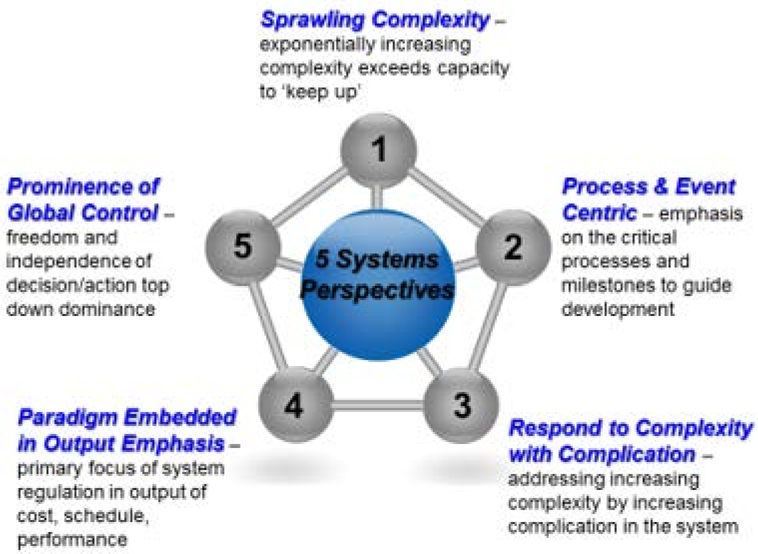


Figure 1. Five Systems Perspectives for the Defense Acquisition Field

2. *Process and Event Centric Focus.* Although the Defense Acquisition System is proclaimed to be a system, in reality it is hard pressed to meet the standard tenets that are central to classification as a system (e.g., boundary, transformation, etc.). Instead, the Acquisition System is a collection of elements for which the precise representation is not presented, operated, or evolved holistically as a system. For instance, Friar's (2015) depiction of the sprawling regulatory environment surrounding the FAR and its implementation notes the precarious relationship of related FAR elements, such as service specific FAR supplements, corresponding guidebooks, implementation instructions, and the DoD 5000 series. Even the DoD 5000 refers to the Defense Acquisition Management System as both a "framework" as well as an "event-based process." The demands placed on the systems that are the product from the Defense Acquisition System are somehow held to a different standard when it comes to *systemic grounding* (design, execution, development) expectations. Processes and events fall short in classification as a system. The ramifications of this distinction are profound and far reaching for understanding of critical issues and future development.
3. *Response to Increasing Complexity Relegated to Increasing Complication*. The original intent of the FAR was quite straightforward in attempting to provide a streamlined approach to the acquisition of material to support the effective functioning of the government enterprise. However, in the time since the inception of the FAR it has continued to elaborate in structure, volume, and complication of process and implementation controls. For instance, the Federal Acquisition Regulation expanded from 1,953 pages at introduction in 1984 to 2,193 pages by 2014, with the DFAR supplement adding another 1,554 pages and each of the services initiating a host of their own specialized implementation guides, instructions, directives, and memorandums (Friar, 2015).
4. *Driving Paradigm Embedded in an Output Emphasis*. Outputs from the Defense Acquisition System are those tangible, verifiable, and objective elements that serve as products that provide value consumed external to the

system. This provides the basis for a worldview (the system of values and beliefs through which all that is sensed is processed) which translates into the design, execution, and development of the Defense Acquisition System. It is hard to read a criticism of the current state of affairs for acquisition that is not targeted to cost, schedule, and technical performance. However, we suggest that these indicators are systemically limited in their measuring the value of the Defense Acquisition System. While these indicators (cost, schedule, performance) are necessary indicators of system performance, they alone do not provide sufficiency as a set of judgments of Acquisition System performance. For example, Cilli et al. (2015) point out the sunk costs of five programs between 2006 and 2011 in excess of $32 billion. The question for examination of paradigm consistency would need to consider whether or not the failure in these programs, and the acquisition system which permitted those failures, might be found beyond the cost, schedule, and technical performance triad.

5. *Prominence of Global Control.* From a systems perspective, control is about providing constraint of a system only to the degree to which is necessary to assure continued performance (Keating et al., 2016). Excess constraint in a system (control) wastes resources and limits local autonomy independence for decision, action, and interpretation. The common manifestation of excessive global control is what has been described aptly in much of the acquisition literature as overregulation, bureaucracy, and excessive constraint without evidence of commensurate value added to the system. The near constant state of acquisition reform (Fox, 2012; Schwartz, 2014) supports the increasing elaboration of the system in ways that do not necessarily enhance performance. This does not demean the improvements achieved in reform processes, but instead suggests that a different (systemic) viewpoint might shift the premises, and thus understanding of the complexities inherent in the system.

This systems perspective for the Defense Acquisition System is intended to suggest that a different frame of reference might be helpful. Our intention is to invite a dialog to further exploration and understanding of the current system, while offering insights into issues in design, execution, and development of the system from an alternative frame of reference. For our purposes, the alternative frame of reference is focusing on understanding system difficulties through discovery of underlying pathologies (aberrations from 'healthy' functioning of a system). To achieve our purpose, the remainder of the paper is organized around four primary objectives. First, in Literature Review for Defense Acquisition System Governance, we provide a literature review targeted to an examination of defense acquisition in relationship to the emerging Complex System Governance field. Second, in the section titled A Systems Theory Based Paradigm and Model for Complex System Governance for Defense Acquisition, we elaborate a Systems Theory based paradigm for Complex System Governance and draw the relevance for the Defense Acquisition System. Third, the System Pathologies in Complex System Governance for Acquisition section describes a set of pathologies in the governance of complex systems and projects that set to the acquisition field. Fourth, the final section concludes the paper with implications for further research and application development of CSG for the acquisition field, focusing on the Defense Acquisition System.

Literature Review for Defense Acquisition System Governance

The literature for the Defense Acquisition System is substantial. A simple search of Google Scholar (April 2, 2017) identified over 3,400 citations for "defense acquisition system." The detailed parsing of this literature is beyond the scope of this paper. However, we can engage a level of literature review to suit our purpose of this paper—*exploiting and extending recent developments in Complex System Governance (CSG) to advance the acquisition field.* To tailor the literature we have focused on three primary objectives (Figure 2) for the literature review, including (1) capture the literature of the Defense Acquisition System with respect to inclusion of the Systems Theory, Management Cybernetics, and System Governance fields; (2) examine the premier defense acquisition related journals for distribution of literature across a taxonomy of research and development areas; and (3) suggest preliminary literature review implications for CSG development in relationship to the Defense Acquisition System.

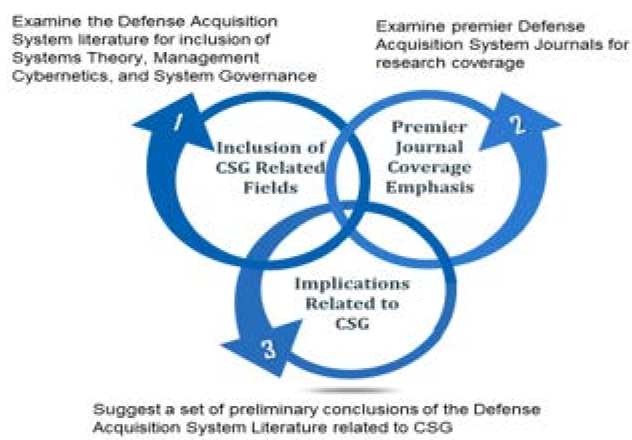


Figure 2. Organization of the Literature Review

CSG Related Field Coverage in the Defense Acquisition System Literature

The three primary informing fields for CSG are found in Systems Theory (Adams et al., 2014; Whitney et al., 2016; von Bertlanffy, 1968), Management Cybernetics (Beer, 1972, 1979, 1985), and System Governance (Calida & Keating, 2014). Several predominant scholarly databases were reviewed for each of the CSG informing fields with respect to "defense acquisition system." The results of the review of the CSG informing fields identified a scarcity in coverage in relationship to Defense System Acquisition. The results of this examination are summarized in Table 1.

Table 1. Coverage of CSG Related Field in the Defense Acquisition System Literature

CSG Informing Field	Related Works for Defense Acquisition System	Synthesizing Themes
Management Cybernetics	Vore (1990)	Examined application of management cybernetics to work measurement for weapon systems acquisition.
System Governance	Berteau, and Ben-Ari (2014)	Focused on acquisition challenges of systems of systems and offers best practice themes, including level of organizational focus, decision-making authority, and enforcement.
Systems Theory	Magee and de Weck (2004)	Classification scheme for complex systems using Systems Theory
	Sheard and Mostashari (2010)	Examined the relationship of complexity to complex systems and systems engineering (including acquisition)
	Schwenn et al. (2015)	Systems Theory basis to better explain, model, and simulate the complexity in acquisition of systems
	Willette (2014)	Application of (complex) Systems Theory to understanding incongruities within aspects of defense acquisition
	Robey (2009)	Application of Systems Theory to better understand DoD Acquisition System
	Alexander (2007)	Applied Systems Theory to identify improvement areas in defense acquisition
	Walker (2014)	Basis of Systems Theory for requirements in support of defense acquisition

From this preliminary review of literature for the Defense Acquisition System body of knowledge, we make three initial conclusions. First, the coverage of the three informing fields for CSG have not received significant levels of development. While we might conjecture as to the proximate cause of this scarcity, the fact remains that the coverage is minimal. Second, it appears that there is an opportunity to project relatively unexplored fields into the Defense Acquisition System dialog. The extension of CSG to the Defense Acquisition System might provide fruitful insights into developmental issues that continue to plague the field. Third, we must be cautious not to overextend these preliminary results. This is representative of an initial exploration of the vast literature for defense acquisition. Care must be taken not to amplify this "glimpse" as more than an invitation for further examination at this point. For example, there were other references to "governance" in acquisition literature (e.g., Rebovich, 2007; Gansler, Lucyshyn, & Rigilano, 2012). However, these articulations were tangential to the primary focus on the systemic nature of governance in relationship to the Defense Acquisition System.

Emphasis of Premier Journals for Defense Acquisition System Literature

The literature for the Defense Acquisition System is immense. For the second phase of the literature review we examined three premier sources of literature supporting development of the acquisition field. The purpose of this exploration was to examine the distribution of articles across a spectrum of the acquisition body of knowledge, ranging from conceptual/theoretical to practical tools. This provided a gestalt view of the distribution of field development. The sources for this review included (1) the *Defense AT&L Magazine*, a publication of the Defense Acquisition University serving professionals in the acquisition community; (2) the *Defense Acquisition Research Journal,* a publication of the Defense Acquisition University targeted to development of a spectrum of topics targeted to the professional acquisition community of researchers and practitioners; and (3) the *Systems Engineering Journal*, a peer reviewed journal published by Wiley publishers under the auspices of the International Council on Systems Engineering (INCOSE). INCOSE is an organization dedicated to the advancement of systems engineering, of which acquisition is a central focus for the organization. Our approach was to fit the documents from a 10-year window (2006–2016) into a taxonomy (based on the description of the classification identified in Table 2).

Table 2. Distribution of Literature Published (Focused During the 10-Year Period of 2006–2016, Total 151 Articles)

Area	Description for Classification	Defense AT&L Journal	Defense AR Journal	Systems Engineering Journal	TOTAL (number)	TOTAL (percent)
Tools	Implements used to support accomplishment of a specific task or purpose	25	35	0	60	40
Methods	Systematic approaches that are performed to achieve an objective	15	25	3	43	29
Models	Representations that capture attributes against which comparisons can be made	7	15	4	26	17
Methodologies	Generalized frameworks that guide applications for the field	3	10	1	14	9
Conceptual Foundations	The fundamental underlying philosophical, theoretical, and axiomatic (principles) basis for the field	2	4	2	8	5

We used these results to establish an initial view of the distribution across a spectrum of focal areas contributing to the body of knowledge for acquisition. This classification provided a gestalt indicator of the emphasis for the field.

Implications of Literature Related to CSG for the Defense Acquisition System

We are hesitant to draw absolute conclusions based on our preliminary review of literature. However, this analysis does suggest several indications that warrant further exploration in a more rigorous examination of the literature in the acquisition field. Among these we include the following:

1. The existence of Defense Acquisition System literature directly drawn from the underlying CSG informing fields (Systems Theory, Management Cybernetics, System Governance) is scarce. What is there is limited in depth related to the CSG fields. This is not unexpected and suggests that inclusion of these fields, and CSG which is drawn from them, might enhance the Defense Acquisition System literature.
2. There appears to be a heavy inclination toward the practice side of literature (tools, methods, models) as fully 127 articles (84%) fit into these categories. Without explanation as to why this skewed distribution exists, it does lend itself to a closer examination of the literature and proclivities of the community.
3. At the level of Methodologies and Conceptual Foundations there were 22 articles (approximately 15%). This result provides an interesting pivot point to provide a closer examination as to the degree that this relatively limited balance in acquisition field development might offer new insights.
4. These preliminary results ask many more questions than they answer. However, they do suggest some different paths of inquiry into further research based development of the Defense Acquisition field. For instance, can tools, methods, and models be more insightfully grounded in an underlying conceptual/theoretical base? Are there implications for acquisition tool/method development stemming from expanding conceptual foundations? And, do failures of tools/methods for acquisition programs suggest that underlying conceptual/theoretical foundations might be contributing to failures?

The initial results of this literature have served the purpose for this paper. The review provides sufficient motivation to pursue two further aims in establishing CSG research in relationship to Defense Acquisition System development. First, a deeper examination and more extensive/rigorous classification of literature is warranted based on initial results. Second, the scarcity of CSG and informing fields application to the acquisition field suggests an opportunity to expand the breadth and depth of the body of knowledge for the Defense Acquisition System. At this point of examination, it does not appear that CSG or its tenets have been a subject of investigation or application in Defense Acquisition System research/applications.

A Systems Theory Based Paradigm and Model for Complex System Governance for Defense Acquisition

In this section we provide the basis for CSG with respect to defense acquisition. To provide this link, three primary themes will be developed. First, we provide the underlying source for CSG as stemming from the intersection of three fields, including *Systems Theory* (propositions that define behavior and explain performance of all complex systems), *Management Cybernetics* (the science of effective structural organization of systems), and *System Governance* (provision of direction, oversight, and accountability) for a system. Second, we introduce the paradigm for CSG. Our emphasis is to provide the high-level depiction of CSG such that the ensuing detailed development of CSG will have an intellectual grounding. Third, we introduce the CSG reference model. This model is explained in relationship to the underlying CSG paradigm and implications for extension to the Defense Acquisition System and field are suggested. We close this section with CSG implications for defense acquisition and a summary of the paradigm as it relates to defense acquisition.

In broad terms, DoD Directive 5000.01 (2007, p. 4) defines the *Defense Acquisition System* as “the management process by which the Department of Defense provides effective, affordable, and timely systems to the users.” with the primary objective (DoD 5001.01, p. 3) being “to acquire quality products that satisfy user needs with measurable improvements to mission capability and operational support, in a timely manner, and at a fair and reasonable price.” At an abstract level, the ambitions of the acquisition system are certainly noble. However, as identified in the previous section, the Defense Acquisition System continues to be a source of challenge. While the precise cause-effect relationships generating difficulties in acquisition deployment are uncertain, the drive for reform, adjustment, and performance improvement in the system continue. Given the complexity in modern warfare systems, and the processes necessary to bring those systems to fruition, it is not likely that a single breakthrough will vanquish issues in the acquisition field. Instead, it is likely that there are numerous contributors to issues, and the resolution of acquisition field issues will require a continuing series of evolutionary changes. In pursuit of Defense Acquisition System improvement, we suggest that new thinking and approaches might provide an acceleration of advancement. CSG is proposed as a contribution to bring the Defense Acquisition System to a different level of understanding. This different level of understanding might provide for a corresponding different level of decision, action, and interpretation to guide evolution of the field in different directions.

As the basis for this different understanding of the Defense Acquisition System, CSG offers three important distinctions. First, CSG is built on an underlying foundation steeped in a strong conceptual grounding. The conceptual basis for CSG development and application is found in Systems Theory (Whitney et al., 2015), Management Cybernetics (Beer, 1979), and System Governance (Calida & Keating, 2014; see Figure 3).

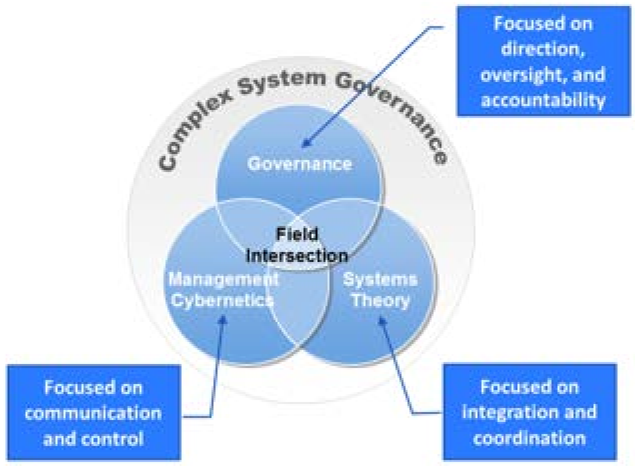


Figure 3. CSG Draws Upon Governance, Systems Theory, and Management Cybernetics

In essence, Systems Theory offers the set of propositions that have been continually developed and applied over the past eight decades. The propositions have withstood the test of time and application, defining the structure, behavior, and performance of all systems. Management Cybernetics provides a strong conceptual foundation for communication and control essential to CSG. In particular, Management Cybernetics offers CSG design cues for control through the model of a metasystem. The metasystem is a set of functions that stand above/beyond the particular systems/entities that it seeks to steer—in the cybernetic sense of providing control. Management Cybernetics also provides a set of communication channels associated with the steering functions of the metasystem. System Governance is concerned with the provision of direction, oversight, and accountability for a system such that future performance and trajectory are ensured.

Governance and systems perspectives were chosen for very specific reasons versus the management and process terms used with respect to defense acquisition. First, the focus on governance permits a distinction to be made from the management perspective. Second invoking systems allows distinction from more limited process perspectives for acquisition. Table 3 is offered as an attempt to clarify distinctions across several attributes. This listing is not intended to provide absolutes, but rather to point out that (1) the Governance/System orientation operates at a different level than more traditional Management/Process focused thinking; (2) in actuality, the table provides bookends along a spectrum—the realities of any system might vary along that spectrum for each attribute; and (3) the shift in perspective for Governance/System introduces the opportunity to foster different thinking, decision, action, and interpretation possibilities.

Table 3. Attributes of a Governance/System Perspective Versus a Management/Process Perspective

Attribute	Governance/System	Process/Management
Execution	Adaptability/Improvisation	Repeatability/Consistency
Objective	Long Range Effectiveness	Short Range Efficiency
Time Horizon	Longer/Strategic	Shorter/Operational
Definition	Malleable/Emergent/Evolving	Rigid/Absolute/Stable
Change	Global/Evolutionary/Chaotic	Local/Short term/Ordered
Primary Emphasis	Viability, Adaptation,	Cost, Schedule, Technical Performance
Dominant Paradigm	Holism/Antipositivism/Emergence	Reductionism/positivism/determinism
Relationship Understanding	Complexity/irreducible/synthesis	Cause-effect/reducible/decomposition
Interpretation	Subjective-negotiable	Objective-absolute
Boundaries	Ambiguous/Shifting	Defined/Stable
Environment	Amorphous	Coherent
Behavior Preference	Uncertain/Non-linear	Deterministic/Linear
Risk Tolerance	High (Acceptance)	Low (Avoidance)
Representation Preference	Interpretative (relationships/diagram/approximate)	Symbolic (mathematical/model/precision)
Solution Goals	Satisficing/feasible/desirable	Optimal/absolute/expected
Rationality	Anticipated irrationality (pluralism)	Expected rationality (unitary)
Development	Unique-Not Replicable-Methodology	Standardized-Replicable-Method
Design Focus	Sociotechnical/External/Strategic	Technical/Internal/Operational
Engagement	Power/Politics/Flexibility	Hierarchy/Apolitical/Repeatability

We now shift attention to the underlying paradigm that defines the essence of the shift to CSG.

The CSG Paradigm Shift

CSG is built upon an underlying paradigm intended to suggest a departure from more traditional reductionist thinking concerning the design, execution, and evolution of systems. In this sense, the paradigm construct is used intentionally to signal a shift in thinking. The CSG Paradigm is composed of three primary elements, including (1) the *Systems Philosophical, Theoretical, Conceptual* foundations which act to provide a stable theoretical/conceptual grounding for everything that follows; (2) *Metasystem Functions* which include nine essential functions, drawn from Management Cybernetics, that define what must be performed by any complex system to remain viable (continue to exist); and (3) *Implementing Mechanisms* that are specific to a particular system and define how it performs metasystem functions. Together, these three elements form the triad of CSG (Figure 4) and are invoked to produce governance. In turn this governance is responsible for the behavior/performance of a complex system and the degree to which a system continues to exist (i.e., remain viable). Continued viability will be determined by (1) the degree to which the design for CSG is capable of meeting the demands of the system environment; (2) the effectiveness of CSG execution consistent with the design; and (3) the ability of the system to evolve in response to the nature and pace of perturbations emanating from internal system flux and external environmental turbulence.

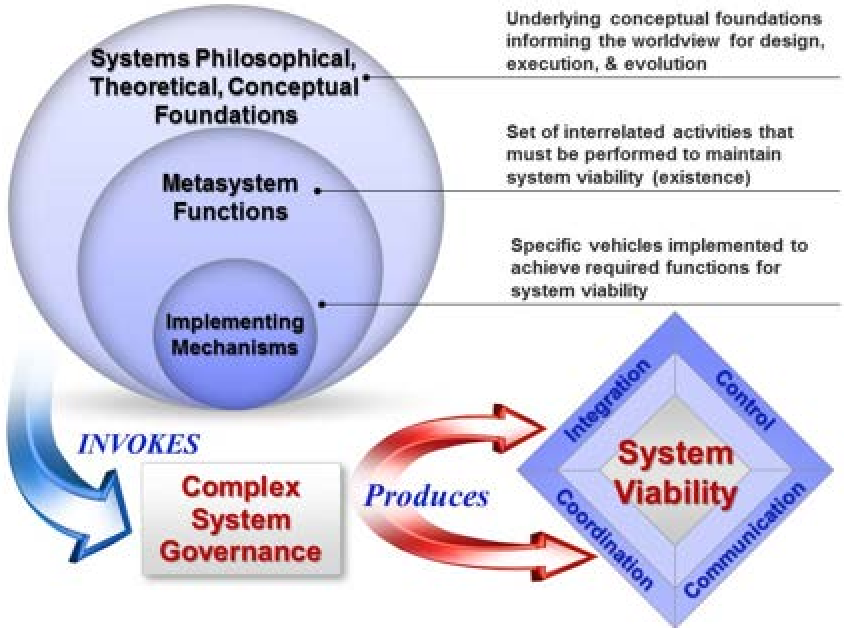


Figure 4. The CSG Paradigm

The high level CSG paradigm can be stated succinctly through the definition as "design, execution, and evolution of the metasystem functions necessary to provide control, communication, coordination, and integration of a complex system" (Keating et al., 2015, p. 274). A further elaboration of the terms in the definition provides insight into the nature of CSG:

- **Design:** purposeful and deliberate arrangement of the governance system to achieve desirable system performance and behavior.
- **Execution:** performance of the system design within the unique system context, subject to emergent conditions stemming from interactions within the system and between the system and its environment.
- **Evolution:** the change of the governance system in response to internal and external perturbations as well as revisions to system trajectory.
- **Metasystem:** the set of nine interrelated functions that provide for governance of a complex system.
- **Control:** invoking the minimal constraints necessary to ensure desirable levels of performance and maintenance of system trajectory, in the midst of internally or externally generated perturbations of the system.
- **Communication:** the flow, transduction, and processing of information within and external to the system, that provides for consistency in decisions, actions, interpretations, and knowledge creation made with respect to the system.
- **Coordination:** providing for interactions (relationships) between constituent entities within the system, and between the system and external entities, such that unnecessary instabilities are avoided.
- **Integration:** continuous maintenance of system integrity. This requires a dynamic balance between autonomy of constituent entities and the interdependence of those entities to form a coherent whole. This

interdependence produces the system identity (uniqueness) that exists beyond the identities of the individual constituents.

- **Complex system:** a set of bounded interdependent entities forming a whole in pursuit of a common purpose to produce value beyond that which individual entities are capable.

It is important to note that all systems must perform the metasystem functions at a minimal level to maintain viability. However, viability is not a guarantee of performance excellence. On the contrary, viability simply assures is that the system continues to exist. There are degrees of viability, the minimal of which is existence. *Implementing Mechanisms* are the specific vehicles (e.g., processes, procedures, activities, practices, plans, artifacts, values/beliefs, customs, mores) that implement metasystem governance functions for a system. These mechanisms are not a general set of mechanisms, but rather exist as unique to a specific system to define how a specific system performs the functions. CSG mechanisms may be explicit/tacit, formal/informal, routine/non-routine, effective/ineffective, or rational/irrational. However, all mechanisms can be articulated in relationship to the metasystem governance functions which they support.

The essence of this paradigm can be used to guide thinking about CSG for acquisition. We now shift attention to the CSG Reference Model. This model is consistent with the CSG paradigm shift and provides a detailed description for application.

The Complex System Governance Reference Model

Central to CSG is a reference model that depicts the central elements of CSG and their interrelationships (Figure 5). A brief depiction of the nature and role of the CSG functions (identified as Metasystem functions) has been previously developed in several publications (Keating et al., 2015; Keating et al., 2016; Bradley et al., 2016).

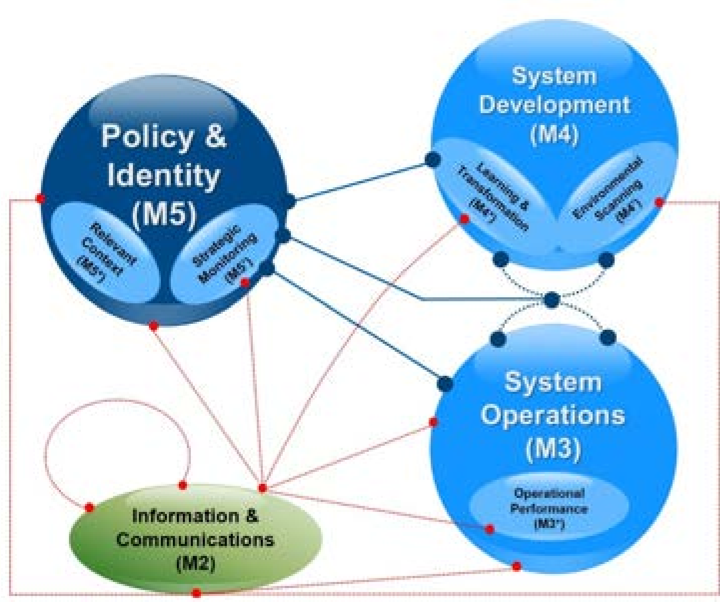


Figure 5. The CSG Reference Model

We summarize the nine critical functions performed by a complex system to maintain viability (continued existence):

- *Metasystem Five (M5)—Policy and Identity*—focused on overall steering of the system, giving policy level direction, representation of the system to external constituents, and maintaining identity for system coherence. For example, M5 would be concerned with the generation and propagation of consistency of purpose for a particular acquisition program.
- *Metasystem Five Star (M5*)—System Context*—focused on the specific context within which the metasystem is embedded. Context is taken as the circumstances, factors, conditions, trends, or patterns that constrain and enable the execution of a system. For example, the particular political dynamics in play for an acquisition program would be an element of consideration for context.
- *Metasystem Five Prime (M5')—Strategic System Monitoring*—focused on oversight of the system at a strategic level. This monitoring is essential to ensure that the trajectory of the system is consistent with the desirable future state. For example, stability of future program resources essential to maintain system development would be a possible area for strategic monitoring.
- *Metasystem Four (M4)—System Development*—focusing on the long-range development of the system to ensure future viability. This function ensures that the current and future models of the system are examined to ensure consistency with trajectory and expectations for system development. For example, this function would identify inconsistencies between existing capabilities in the program (system) and those required for realization of the future capabilities required to ensure integrity of the program (system).
- *Metasystem Four Star (M4*)—Learning and Transformation*—focused on facilitation of learning based on detection and correction of design errors in the metasystem and guiding planning to support transformation of the metasystem. For example, execution of design reviews targeting errors in material procurement.
- *Metasystem Four Prime (M4')—Environmental Scanning*—focused on sensing the environment for circumstances, trends, patterns, or events with implications for both present and future system performance. For example, the early identification of new regulatory requirements.
- *Metasystem Three (M3)—System Operations*—focused on the day to day operations of the metasystem to ensure that the system maintains performance levels. For instance, responding to schedule shifts to compensate for supplier shifts in material availability.
- *Metasystem Three Star (M3*)—Operational* Performance—focused on monitoring system performance to identify and assess aberrant or emergent conditions in the system. For example, conducting audits of resource utilization.
- *Metasystem Two (M2)—Information and Communications*—focused on the design for flow of information and consistent interpretation of exchanges (communication channels). For instance, providing forums for dissemination of information for access throughout the system.

Ultimately, effectiveness in purposeful design, execution, and evolution of the nine metasystem governance functions, via implementing mechanisms, determines system performance, including acquisition systems. We now examine the essence of CSG for defense acquisition.

Essence of CSG for Defense Acquisition

Although the underlying theory, concepts, and execution of CSG are challenging, the essence of CSG related to defense acquisition is not difficult to gasp. The essence of CSG for acquisition might be captured in the following paragraph and elaborated in the four points that follow:

> Subject to fundamental system laws, all systems engaged in acquisition (programs, projects, entities, agencies, etc.) perform essential governance functions. System performance is determined by effectiveness in achievement of nine governance functions consistent with system laws. Violation of system laws degrades system performance. System performance can be enhanced through purposeful development of governance functions and their implementing mechanisms.

1. ***All systems are subject to the laws of systems***. Just as there are laws governing the nature of matter and energy (e.g., physics law of gravity), so too are our systems subject to laws found in Systems Theory and Management Cybernetics. These system laws are always there, non-negotiable, non-biased, and explain system performance. The implication for *acquisition practice is understanding system performance in relationship to underlying systems laws*.
2. ***All systems perform essential governance functions that determine system performance.*** Nine system governance functions are performed by all systems, regardless of sector, size, or purpose. These functions define what must be achieved for governance of a system. Every system invokes a set of unique implementing mechanisms (means of achieving governance functions) that determine how governance functions are accomplished. Mechanisms can be formal-informal, tacit-explicit, routine-sporadic, or limited-comprehensive in nature. CSG produces system performance which is a function of previously discussed communication, control, integration, and coordination. *Acquisition practitioners must ask*, "*Do we understand how our system performs essential governance functions to produce performance?"*
3. ***Violations of systems laws in performance of governance functions carry consequences***. Irrespective of noble intentions, ignorance, or willful disregard, violation of system laws carries real consequences for system performance. In the best case, violations degrade performance. In the worst case, violations can escalate to cause catastrophic consequences or even eventual system collapse. *Acquisition practitioners must ask*, *"Do we understand problematic system performance in terms of violations of fundamental systems laws?"*
4. ***System performance can be enhanced through purposeful development of governance functions.*** When system performance outputs fail to meet expectations (e.g., cost, schedule, technical performance), deficiencies in governance functions can offer novel insights into the deeper systemic sources of failure. Performance issues can be traced to governance function issues as well as violations of underlying system laws. Thus, system development can proceed in a more informed and purposeful mode. *Acquisition practitioners must ask, "How might problematic performance be explained as stemming from deeper system governance issues and violations of system laws, suggesting different development directions?"*

We now shift our attention to exploration of pathologies that are representative of violations of systems laws in performance of CSG functions.

System Pathologies in the Complex System Governance for Acquisition

At a basic level, a system pathology has been described as a "circumstance, condition, factor, or pattern that acts to limit system performance, or lessen system viability, such that the likelihood of a system achieving performance expectation is reduced" (Keating & Katina, 2012, p. 253). The basis for pathologies in systems, including the acquisition system, stems from Systems Theory (the set of principles, laws, and concepts that explain behavior and performance of systems) and Management Cybernetics (focused on the structure that serves to produce system organization). Stemming from the work of Katina (2015, 2016) and Keating and Katina (2012) a set of 53 system pathologies has emerged. While each of these pathologies is mapped to the underlying systems law(s) in violation, that mapping is beyond the scope of this paper. However, in this section we present the current state of set of pathologies and project their implications for the Defense Acquisition System. We close the section with the implications that further development of pathologies can hold for improving the acquisition system.

Table 4. Metasystem Pathologies With Acquisition System Applicability

Metasystem function	***Corresponding Set of Pathologies***	***Acquisition System Applicability***
Metasystem Five (M5): Policy and Identity	M5.1. Identity of system is ambiguous and does not effectively generate consistency system decision, action, and interpretation.	Identity clarity is essential to increase the likelihood of consistency in establishment of priorities and tradeoffs. For example, given cuts in scarce resources, the tradeoffs among cost, schedule, and technical performance should be informed by a consistent and stable reference frame (identity).
	M5.2. System vision, purpose, mission, or values remain unarticulated, or articulated but not embedded in the execution of the system.	Consistency in program tradeoffs occurring throughout the acquisition system are dependent on the congruence of thinking, decisions, and interpretations held by the entity.
	M5.3. Balance between short-term operational focus and long-term strategic focus is unexplored.	Every program has a tension between long- and short-term objectives. The explicit definition and resolution of this tension over the acquisition process is essential to avoid unnecessary crises and conflicts.
	M5.4. Strategic focus lacks sufficient clarity to direct consistent system development.	Purposeful direction for strategic development of an acquisition program/entity should be by design, explicit, measurable, and dynamic—not a haphazard endeavor in response to crises.
	M5.5. System identity is not routinely assessed, maintained, or questioned for continuing ability to guide consistency in system decision and action.	Every acquisition entity has an identity that marks its uniqueness. Purposeful articulation and development left to chance misses the opportunity to exploit this advantage in dealing with complexity.
	M5.6. External system projection is not effectively performed.	Providing clarity in communication and messaging to external entities can help preclude expectation mismatches.
Metasystem Five Star (M5*): System Context	M5*.1. Incompatible metasystem context constraining system performance.	Potential incompatibilities between a program design and execution with the context within which it is embedded invites system degradation.
	M5*.2. Lack of articulation and representation of metasystem context.	Lacking an explicit mapping of contextual influences misses opportunities to influence impacts or preclude disruptions stemming from context.
	M5*.3. Lack of consideration of context in metasystem decisions and actions.	Performance impacts of contextual factors should be part of program planning and execution.
Metasystem Five Prime (M5'): Strategic System Monitoring	M5'.1. Lack of strategic system monitoring.	Strategic system performance indicators should be monitored to identify emergent variability.
	M5'.2. Inadequate processing of strategic monitoring results.	Program/system development should be in direct response to identified strategic variabilities.

	M5'.3. Lack of strategic system performance indicators.	Strategic system performance indicators, beyond cost, schedule, and technical performance should exist to guide strategic execution.
Metasystem Four (M4): System Development	M4.1. Lack of forums to foster system development and transformation.	Beyond day-to-day execution, strategic development activities should be engaged for long range, strategic system evolution.
	M4.2. Inadequate interpretation and processing of results of environmental scanning—non-existent, sporadic, limited.	Results of active scanning for potential system perturbations should be routinely processed and responses formulated.
	M4.3. Ineffective processing and dissemination of environmental scanning results.	Scanning results should be disseminated to the point where responsive decision/action can be implemented.
	M4.4. Long-range strategic development is sacrificed for management of day-to-day operations—limited time devoted to strategic analysis.	Appropriate balance in emphasis between present and future is essential to ensure that both short-term and long-term viability are maintained. This balance may dynamically shift over time for the program and/or changing circumstances.
	M4.5. Strategic planning/thinking focuses on operational level planning and improvement.	Programs that are consumed with day to day existence sacrifice long-term viability and precluding avoidable future crises.
Metasystem Four Star (M4*): Learning and Transformation	M4*.1. Limited learning achieved related to environmental shifts.	Without a continuous and explicit learning system in place, opportunities for system development can be missed.
	M4*.2. Integrated strategic transformation not conducted, limited, or ineffective.	Long-range viability of a program/system should be by active comprehensive design, not left to chance.
	M4*.3. Lack of design for system learning—informal, non-existent, or ineffective.	Vehicles to invoke continuous detection and correction of design, execution, and development issues should be in place.
	M4*.4. Absence of system representative models—present and future.	Programs/systems should have a dynamic mapping showing how value is produced, for both the current and future anticipated system.
Metasystem Four Prime (M4'): Environmental Scanning	M4'.1. Lack of effective scanning mechanisms.	Environmental scanning mechanisms should be comprehensive, focused, and integrated.
	M4'.2. Inappropriate targeting/undirected environmental scanning.	Program/system environmental scanning should be by purposeful design.
	M4'.3. Scanning frequency not appropriate for rate of environmental shifts.	Timing of environmental scanning should be appropriate and shift over the life cycle of a program.
	M4'.4. System lacks enough control over variety generated by the environment.	Filtering of environment noise and amplification of the program to the environment (e.g. external stakeholders) should be instituted by design.
	M4'.5. Lack of current model of system environment.	Programs/systems should have a dynamic explicit mapping of the environment.
Metasystem Three (M3): System Operations	M3.1. Imbalance between autonomy of productive elements and integration of whole system.	Constituent systems/entities should be given the maximum autonomy possible without degrading program/system performance.
	M3.2. Shifts in resources without corresponding shifts in accountability/shifts in accountability without corresponding shifts in resources.	For every shift in resources provided for a program/system, there should be a corresponding negotiated adjustment in expectations.
	M3.3. Mismatch between resource and productivity expectations.	Alignment between expectations for program value production and resources allocated for that production should be congruent.
	M3.4. Lack of clarity for responsibility, expectations, and accountability for performance.	Roles for system design, execution, and development should provide clarity of assignment and eliminate unnecessary redundancy.
	M3.5. Operational planning frequently pre-empted by emergent crises.	Crises must be understood in relation to underlying issues in system design, execution, or development.
	M3.6. Inappropriate balance between short-term operational versus long-term strategic focus.	Continual sacrifice of system long-term development can degrade near-term performance or sacrifice long-term viability.
	M3.7. Lack of clarity of operational direction for productive entities (i.e. subsystems).	Program/system entities producing core value should be given clear, concise, and timely direction for expectations as to what must be accomplished.

	M3.8. Difficulty in managing integration of system productive entities (i.e. subsystems).	Constituent entities of a program/system must be provided clear direction on limits for autonomy and expectations for integration into the program.
	M3.9. Slow to anticipate, identify, and respond to environmental shifts.	Robust processes must mount timely, resilient, and effective responses across a spectrum of environmental perturbations.
Metasystem Three Star (M3*): Operational Performance	M3*.1. Limited accessibility to data necessary to monitor performance.	Actionable data on performance of operation design and execution must be available.
	M3*.2. System-level operational performance indicators are absent, limited, or ineffective.	Program/system performance indicators, in addition to cost, schedule, and technical performance should exist to provide a holistic performance picture.
	M3*.3. Absence of monitoring for system and subsystem level performance.	Monitoring performance must be accomplished for variances in operational trajectory and expectations.
	M3*.4. Lack of analysis for performance variability or emergent deviations from expected performance levels or the meaning of deviations.	Programs/systems should have analysis methods in place to systemically examine sources of performance variation to derive interpretations and support informed responses.
	M3*.5. Performance auditing is non-existent, limited in nature, or restricted mainly to troubleshooting emergent issues.	Non-pejorative internal auditing of performance should be conducted routinely within the program/system to determine consistency in design and execution of the program/system.
	M3*.6. Periodic examination of system performance largely unorganized and informal in nature.	Examination of system performance across operational design and execution should be efficient with minimal disruption.
	M3*.7. Limited system learning based on performance assessments.	Results of performance assessment should be interpreted and actionable across system design, execution, and development aspects of the system.
Metasystem Two (M2): Information and Communications	M2.1. Unresolved coordination issues within the system.	Program/system entities should not continue to experience identified coordination issues.
	M2.2. Excess redundancies in system resulting in inconsistency and inefficient utilization of resources—including information.	Existence of potentially unnecessary redundancies within the program/system should be continually questioned and identified for elimination when necessary.
	M2.3. System integration issues stemming from excessive entity isolation or fragmentation.	Isolation/fragmentation of entities attributable to inadequacies in system design, execution, or development should be targeted for elimination.
	M2.4. System conflict stemming from unilateral decisions and actions.	Conflicts stemming from decision/action execution should be mapped to identify system design/execution/development inconsistencies.
	M2.5. Excessive level of emergent crises associated with information transmission, communication, and coordination within the system.	Crises should be examined for design and execution issues in communications and implications for system development.
	M2.6. Weak or ineffective communications systems among system entities (i.e. subsystems).	Communications within the program/system should be explicitly designed and execution monitored to identify performance or development issues.
	M2.7. Lack of standardized methods (i.e. procedures, tools, and techniques) for routine system level activities.	Expectations for design, execution, and development of standardized methods within the system should be clear.
	M2.8. Overutilization of standardized methods (i.e. procedures, tools, and techniques) where they should be customized.	Programs/systems should understand their unique needs and question application of standardized approach compatibility and necessity.
	M2.9. Overly ad-hoc system coordination versus purposeful design.	Program/system design should preclude repetitive coordination conflicts.
	M2.10. Difficulty in accomplishing cross-system functions requiring integration or standardization.	Interfaces between program/system entities should enhance rather than detract from cross-functional activities.
	M2.11. Introduction of uncoordinated system changes resulting in excessive oscillation.	Changes in programs/systems should be integrated by design with minimal disruption to ongoing operations or system development.

There are four primary conclusions with respect to pathologies identified for CSG functions and their implications for the Defense Acquisition System. First, these pathologies are not unique to any given acquisition entity (program, project, agency). They may certainly be present or absent and vary in degree should they be present in any system. The set of pathologies are indicative of aberrant conditions in the metasystem functions necessary to maintain viability for complex systems, including acquisition systems. Thus, the 53 pathologies (Table 4) can act to limit system performance, or lessen system viability, such that the likelihood of a system achieving performance expectations is reduced. This is particularly critical in acquisition systems that must function in conditions of increasing complexity. Second, these pathologies do not exist in a binary fashion of "present" or "not present." Rather, they may be experienced in "degrees of existence" along a continuum ranging from minimal to significant. This opens the possibility of pathologies having not only a degree related to their existence, but also the potential system degradation that may be experienced stemming from that level of existence. Third, the existence of a pathology has real consequences for performance of a given acquisition system—measured in terms of a range of possible effects. These effects may not be easily derived from the observable (e.g., crisis) surface manifestations resulting from the underlying pathology. As each acquisition system is unique, so too will be the associated pathologies that become apparent as the system operates. The pathologies will not be static over the life cycle for a given program/system. Thus, the impact of a pathology may lessen or exacerbate over time as the acquisition system is executed. Fourth, based on prior research, these pathologies should be a subject of exploration during problem/program formulation, since bringing change to the system is largely dependent on understanding the current state of the system (Dery, 1984; Katina, 2015; Quade, 1980). We suggest that in the formative stages of acquisition system exploration, knowledge of pathologies and their assessment can play a vital role in targeted system design and subsequent development. They can serve in both the design of new acquisition programs/systems or evaluation of programs/systems currently underway.

Conclusions and Implications

CSG is a systems based field in the embryonic stages of development. The contribution of CSG to the Defense Acquisition System is targeted to enable practitioners to better deal with issues stemming from the problems associated with increasing complexity in the systems and environment they face. In essence, CSG purposefully addresses system drift. System drift denotes systems/programs that, irrespective of the best intentions, have either never been purposefully designed or whose execution continually fails to meet desired performance expectations. In short, these drifting systems/programs fail to deliver anticipated value, much less produce high performance. We do not need to look far to see examples of drifting systems/programs related to defense acquisition. Consider the following examples (Table 5) and the suggestion of how the CSG perspective might view the program failures (Bradley, Katina, & Keating, 2016). While CSG provides a set of lenses from which to view program deficiencies, CSG pathologies provides a deeper examination of the deep systems sources of program failure or degradation in relationship to system functions necessary for CSG. We cannot provide assurance that CSG and discovery of pathologies in these efforts would have precluded failure. However, CSG pathologies provide a different perspective and set of insights. This might direct acquisition professionals to early identification of sources of system errors with sufficiently early identification to provide correction prior to program/system collapse.

Table 5. Analysis of Troubled Programs Through the Lens of CSG

DOD program/ Report Source	Problem/failure appear to be governance related?	Does the language in the report indicate a similar meaning for governance as the Complex System Governance?	Is there any concrete indication that CSG would have helped this program?
Zumwalt Class Destroyers (DDG1000) GAO-08-904	Yes	No, model/framework of governance (Milestone C suggested) won't help with alignment of perspectives or understanding decisions and actions (communication channel dialog) among others	Yes—this initiative seems to lack clear vision/strategy. Report suggests that channels of communication are weak (p. 45 for example)
Ford Class Aircraft Carrier (CVN78) GAO-16-847	Yes	Yes, the report seems to identify many governance issues that can be mapped to metasystem functions within the CSG Reference Model	Yes—contextual assessment to evaluate acquisition culture. The ship is already built though, so now the asset needs to be protected and maintained.
Total Asset Visibility (Air Force) GAO-08-866	Yes	Yes, especially the "transformation plans" demonstrating initiative to evolve meta-systemic functioning	Yes—systems thinking likely not present in development; poor coordination of unsuccessful program
Major Automated Information Systems (MAIS) GAO-12-629	Yes	Yes, GAO seems to have an idea of the metasystem governance expected of a complex system, as well as realistic expectations regarding scope	Yes—some metasystem functions are clearly missing or inadequate, ex. poor coordination and communication (GAO, 2012, pp. 57–58)
National Security Cutter (Coast Guard/Navy) GAO-16-148	Yes	Yes, report seems to capture design/execution elements necessary for control/communication/coordination/integration (but possibly not sufficient?)	Yes—CSG embraces varying perspectives—the CG and Navy did not seem prepared to align perspectives and have poor communications

Further research and development directions for CSG in relationship to research in acquisition is envisioned in the application of the research and development model provided in Figure 6. This model suggests a close coupling of four primary elements to holistically engage CSG development and deployment to improve practices in the acquisition field. In a nutshell, we summarize this as *The system-science based engineering of technologies to support application development that advances practices related to design, execution, and development of complex systems*. These four elements include the following:

1. *System Science Based*. System science is a broad area that includes multiple different fields that explore the phenomena associated with explaining the behavior and performance of systems. For our purposes, we suggest that system science provides the conceptual underpinnings for all derivative developments (e.g., technologies, methods, tools) based on application of the science. For CSG, the systems science basis is found primarily in Systems Theory and Management Cybernetics.
2. *Engineering of Technologies*. Based upon the underlying systems science, engineering involves the development of implementing technologies. These technologies are developed as CSG supporting artifacts (e.g., tools, techniques, software), grounded in systems science and addressing a targeted aspect of design, execution, or development of complex systems.
3. *Application Development*. Application is focused on development of the particular methods and methodologies that bridge the divide between the engineering of technologies and preparation for deployment in practice. The application emphasis is the appropriate preparation of technologies respective of their qualification for deployment for particular purposes,

integration with other technologies and methods/methodologies, and providing for effective deployment within operational/practice contexts.

4. *Practice*. Ultimately the beneficiary of the systems science, engineering, and application triad is the practice field where deployment is targeted. This is where the different technologies, as deployed through application design, are targeted to enhance practices related to better design, execution, and development of complex systems and their problems.

A central and critical aspect of this framework is the close coupling of science, engineering, application, and practice. Each of the elements in this framework are interrelated. This suggests that their execution is not mutually exclusive or independent of one another. On the contrary, the breakthroughs in CSG for acquisition are seen as four interconnected elements operating to inform, and be informed by, the other elements. We postulate that this interdependent coupling will moderate the trajectory of each of the constituent elements in ways not accessible with their independent development.

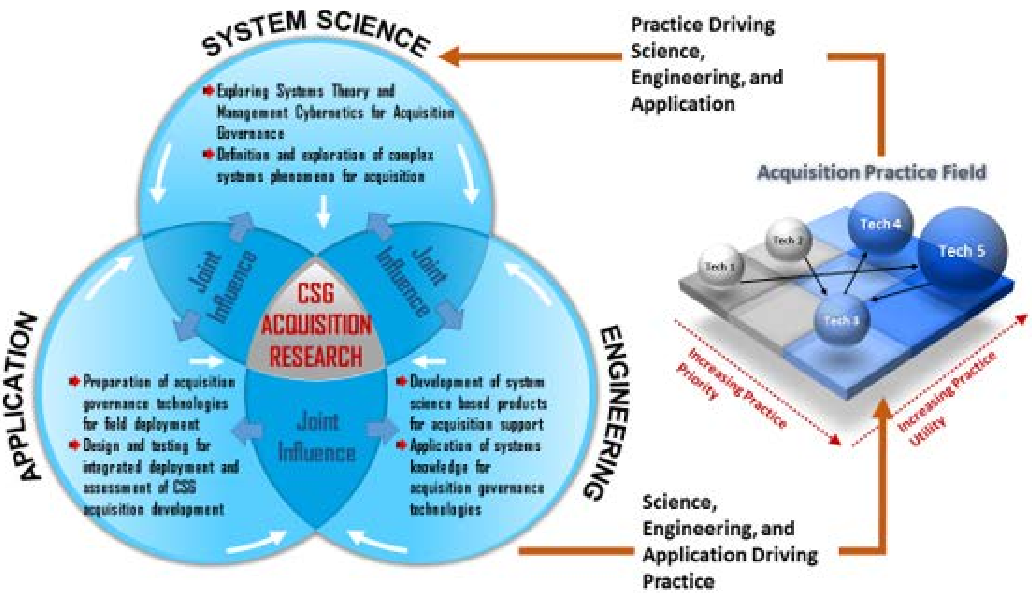


Figure 6. The CSG Research and Development Framework

Further development of CSG pathologies for acquisition systems is focused on four critical challenges:

1. *Maintenance of grounding in Systems Science*. There is a propensity in the professions (e.g., acquisition) to be pragmatic and practitioner focused. While this is expected from a practitioner's perspective, development of the CSG field will be enhanced by the appreciation and grounding of advancements in the underlying theoretical and conceptual basis. Further development of pathologies must remain grounded in the Systems Science base if the ultimate utility of this area is to be achieved.
2. *Engineering of technologies to support practical applications of CSG.* Technologies are the byproduct of engineering to address problems or fulfill needs in ways that advance practical purposes. Engineering of CSG technologies (tools, techniques, artefacts) is critical to help link the theoretical/conceptual science-based formulations of CSG in preparation for

deployment. CSG technologies should advance practical aspects of deployment for practitioners and also be appreciative to their fit within the larger landscape of the emerging CSG field for application to acquisition.

3. *Design for application and deployment of technologies*. CSG technologies are not intended for direct deployment to acquisition practice settings. Instead, emphasis must be placed on effectively bridging between technologies and their deployment. This bridge exists as the design for application. The design for application must take into account the wider perspective of the problem/need, the context, qualifying assumptions, and so forth. The building of applications of CSG from this perspective is essential to enhance the appropriate qualification of technologies to unique acquisition practice circumstances. This requires that the technologies be fit to (a) the particular context within which they will be deployed and operate and (b) the specific acquisition problems for which the technology is appropriate.
4. *Case demonstration of deployable applications*. Advancing CSG for acquisition cannot be achieved separate from the world of practice. The determination of application utility for practice can only be established from deployment in practice. The pragmatic focus of the acquisition field necessitates the development of practice enabling tools, techniques, and technologies. CSG must emphasize producing these tools to aid practitioners in the acquisition field.

The acquisition field is under tremendous pressure to increase effectiveness in delivery of on time, on budget, and on performance systems. While this has always been an objective for acquisition systems, the current nature of the problem domain has substantially increased the challenges facing the field. Further development of CSG and systemic analysis of pathologies can serve to advance the capabilities and capacity of the acquisition field and professionals to better address current and future challenges.

References

Beer, S. (1972). *Brain of the firm: The managerial cybernetics of organization*. London, England: Penguin Press.

Beer, S. (1979). *The heart of enterprise*. New York, NY: John Wiley & Sons.

Beer, S. (1985). *Diagnosing the system for organizations*. Hoboken, NJ: John Wiley & Sons.

Bertalanffy, L. V. (1950). An outline of general systems theory. *The British Journal for the Philosophy of Science, 1*(2), 134–165.

Bertalanffy, L. V. (1968). *General system theory: Foundations, development, applications* (Rev. ed.). New York, NY: George Braziller.

Berteau, D., & Ben-Ari, G. (2014). *Identifying governance best practices in systems-of-systems acquisition* (No. CSIS-AM-14-006). Washington, DC: Center for Strategic and International Studies.

Berteau, D., Ben-Ari, G., Hofbauer, J., Sanders, G., Ellman, J., & Morrow, D. (2011). Cost and time overruns for major defense acquisition programs: An annotated brief. Paper presented at the Eighth Annual Acquisition Research Symposium, Monterey, CA.

Berteau, D., Levy, R., Ben-Ari, G., & Moore, C. (2011). *Defense Industry access to capital markets: Wall Street and the Pentagon. An annotated brief.*

Bradley, J. M., Katina, P. F., & Keating, C. B. (2016). Complex system governance for acquisition. In *Proceedings of the 2016 Acquisition Research Symposium*. Monterey, CA: Naval Postgraduate School.

Calida, B. Y., & Keating, C. B. (2014). System governance: Emergence of practical perspectives across the disciplines. In *Infranomics* (pp. 269–296). Cham, Switzerland: Springer International.

Cilli, M., Parnell, G. S., Cloutier, R., & Zigh, T. (2015). A systems engineering perspective on the revised defense acquisition system. *Systems Engineering, 18*(6), 584–603.

Dery, D. (1984). *Problem definition in policy analysis*. Lawrence, KS: University Press of Kansas.

Boudreau, M. (2007). Acoustic rapid COTS insertion: A case study in modular open systems approach for spiral development. Paper presented at the IEEE International Conference on System of Systems Engineering.

Fauser, M. (2006). *Transnational migration–A national security risk? Securitization of migration policies in Germany, Spain and the United Kingdom*. Retrieved from http://pdc.ceu.hu/archive/00004804/01/rap_i_an_0206a.pdf

Fox, J. R. (2012). *Defense acquisition reform, 1960–2009: An elusive goal* (Vol. 51). Government Printing Office.

Francis, P. L. (2008). *Defense acquisitions: Zumwalt-class destroyer program emblematic of challenges facing Navy shipbuilding*. Washington, DC: GPO.

Francis, P. L. (2009). *Defense acquisitions: Cost to deliver Zumwalt-class destroyers likely to exceed budget* (1437909310). Washington, DC: GPO.

Friar, A. (2015). *Swamped by regulations: Perils of an ever-increasing burden*. Ft. Belvoir, VA: Defense Acquisition University.

Gansler, J. S., & Lucyshyn, W. (2015). Allocating national security resources. In F. Melese, A. Richter, & B. Solomon (Eds.), *Military cost–benefit analysis: Theory and practice* (Vol. 14, p. 18). New York, NY: Routledge.

Gansler, J. S., Lucyshyn, W., & Rigilano, J. (2012). Implementing system-of-systems governance. Maryland University College Park Center for Public Policy and Private Enterprise.

GAO. (2008, June 25). *Risk management: Strengthening the use of risk management principles in homeland security* (GAO-08-904). Washington, DC: Author.

GAO. (2008, August 8). *DOD business transformation: Air Force's current approach increases risk that asset visibility goals and transformation priorities will not be achieved* (GAO-08-866). Washington, DC: Author.

GAO. (2012, July 11). *Information technology cost estimation: Agencies need to address significant weaknesses in policies and practice*s (GAO-12-629). Washington, DC: Author.

GAO. (2016, January 12). *National security cutter: Enhanced oversight needed to ensure problems discovered during testing and operations are addressed* (GAO-16-148). Washington, DC: Author.

GAO. (2016, August 12). *Department of Transportation, Federal Transit Administration: Transit asset management; national transit database* (GAO-16-847). Washington, DC: Author.

GAO. (2017, March 17). *Defense acquisitions: Assessments of selected weapon programs* (GAO-16-333SP). Washington, DC: Author.

Jaradat, R. M., & Keating, C. B. (2014). Fragility of oil as a critical infrastructure problem. *International Journal of Critical Infrastructure Protection, 7*(2), 86–99.

Kadish, R. T., Abbott, G., Cappuccio, F., Hawley, R., Kern, P., & Kozlowski, D. (2006). *Defense acquisition performance assessment report.* Washington, DC: Pentagon.

Katina, P. F. (2015). Emerging systems theory–based pathologies for governance of complex systems. *International Journal of System of Systems Engineering, 6*(½), 144–159.

Keating, C. B. (2014). *Governance implications for meeting challenges in the system of systems engineering field.* Paper presented at the IEEE System of Systems Engineering Conference, Adelaide, SA.

Keating, C. B., & Bradley, J. M. (2015). Complex system governance reference model. *International Journal of System of Systems Engineering, 6*(1), 33–52.

Keating, C. B., & Katina, P. F. (2011). Systems of systems engineering: prospects and challenges for the emerging field. *International Journal of System of Systems Engineering, 2*(2–3), 234–256.

Keating, C. B., & Katina, P. F. (2012). Prevalence of pathologies in systems of systems. *International Journal of System of Systems Engineering, 3*(3–4), 243–267.

Keating, C. B., Katina, P. F., & Bradley, J. M. (2015). *Challenges for developing complex system governance.* Paper presented at the Proceedings of the 2015 Industrial and Systems Engineering Research Conference.

Kendall, F. (2014, July 10). Testimony before the House Committee on Armed Services.

Mills, S., & Goldsmith, R. (2014, September–October). Cybersecurity challenges for program managers. *Defense AT&L.*

Naphade, M., Banavar, G., Harrison, C., Paraszczak, J., & Morris, R. (2011). Smarter cities and their innovation challenges. *Computer, 44*(6), 32–39.

O'Rourke, R. (2014). *Case studies in DoD acquisition: Finding what works* (1437928390). Washington, DC: Congressional Record.

Quade, E. S. (1980). Pitfalls in formulation and modeling. In G. Majone & E. S. Quade (Eds.), *Pitfalls of analysis* (Vol. 8, pp. 23–43). Chichester, England: Wiley-Interscience.

Rascona, P. M., Barkakati, N., & Solis, W. M. (2008). *DOD business transformation Air Force's current approach increases risk that asset visibility goals and transformation priorities will not be achieved.* Washington, DC: GPO.

Rebovich, G. (2007, April). Engineering the enterprise. In *Proceedings of the First Annual IEEE Systems Conference,* 1–6.

Schwartz, M. (2014). *Defense acquisitions: How DOD acquires weapon systems and recent efforts to reform the process.*

Schwenn, K., Colombi, J., Wu, T., Oyama, K., & Johnson, A. (2015). Toward agent-based modeling of the US Department of Defense acquisition system. *Procedia Computer Science, 44,* 383–392.

Sheard, S. A., & Mostashari, A. (2010, July). 7.3. 1 A complexity typology for systems engineering. In *INCOSE International Symposium, 20*(1), 933–945.

Skyttner, L. (2005). *General systems theory: Problems, perspectives, practice* (2nd ed.). Singapore: World Scientific.

Walker, R. G. (2014). *A method to define requirements for system of systems*. Old Dominion University.

Whitney, K., Bradley, J. M., Baugh, D. E., & Chesterman, C. W., Jr. (2015). Systems theory as a foundation for governance of complex systems. *International Journal of System of Systems Engineering, 6*(1), 15–32.

Willette, S. E. (2014). *AQ methodology study of complex adaptive system emergence in the defense acquisition structure* (Doctoral dissertation, Walden University).

Optimal Selection of Organizational Structuring for Complex Systems Development and Acquisitions

Alexandra Dukes—is a Graduate Student in the School of Aeronautics and Astronautics at Purdue University. She is currently conducting research in the Center for Integrated Systems in Aerospace (CISA) led by Dr. Daniel DeLaurentis. [dukes@purdue.edu]

Scott Parrigon—is a Graduate Student in the Department of Psychological Sciences at Purdue University. He is currently part of Dr. Sang Eun Woo's research group. [spariggo@purdue.edu]

Navindran Davendralingam—is a Research Scientist in the School of Aeronautics and Astronautics at Purdue University. He is currently conducting research in the Center for Integrated Systems in Aerospace (CISA) led by Dr. Daniel DeLaurentis. [davendra@purdue.edu]

Sang Eun Woo—is an Associate Professor of Industrial and Organizational Psychology in the Department of Psychological Sciences at Purdue University. Her research focuses on how people's personality and motivation can help explain various psychological phenomena in the workplace. [sewoo@purdue.edu]

Daniel DeLaurentis—is a Professor in the School of Aeronautics and Astronautics, Purdue University. His research and teaching interests focus on design and optimization of aerospace vehicles and systems of systems. [ddelaure@purdue.edu]

Abstract

Research suggests that product designs tend to reflect the structure of the organization in which they are conceived (i.e., Conway's Law). Prior works on this topic, especially in the context of acquisitions, have been largely descriptive without prescribing tangible ways to reduce the inefficiencies resulting from possible misalignments between a product's structure and the structure of the organization that builds the product. We present a mathematical modeling framework that enables the optimal selection of an organization's structure (here, the different ways that various types of program managers are allocated) and its product structure (here, a modular, complex system structure). We leverage quantitative and qualitative methods from areas of organizational sciences, systems engineering, and operations research in a unified manner. We demonstrate application to a defense acquisition concept problem that seeks to maximize overall performance of a complex system (the "product") being developed, while minimizing risks associated with mismatches between program manager competencies and system development ("the organizational structure").

Introduction

A product's structure is strongly affected by organizational structure, communication mechanisms, and resource channels between organizational units that work together to realize an intended product. Inefficient setup in an organization's structure often results in poor requirements being set, poor understanding of interfaces between elements of the product, and ultimately, a poor return on investment due to a consequently subpar product being realized. Prior research conducted in software engineering analyzes this relationship and *concludes that product designs tend to reflect the structure of an organization in which they are conceived*, also known as Conway's Law (Conway, 1968). Work by Ulrich (1995) and Sinha (2012) explored the question of how the degree of a new product's novelty affects the structure of an organization. In more recent literature, Honda performed a comparison of information passing strategies in system-level modeling and found that the structure of information coordination, for the case of an example satellite design problem, directly impacts the drive towards an optimal design configuration (Honda, Ciucci, Lewis, & Yang, 2015). A recent article, published in *Harvard Business Review*, presents a case study of how Juniper networks, a company that provides IT routing and network solutions, utilized

HR strategies to improve business processes across its complex organizational structure (Boudreau, 2015). The strategies reduced the number of decision chains involved in product development and sought to identify “clusters” of employees with the most diverse experience in promoting healthy innovation.

While these prior literatures allude to the coupled nature between a product structure and the structure of the organization that builds it, they are mainly descriptive in nature. These literatures do not provide a framework to improve decision-making processes related to the product structure (e.g., what collection of systems to acquire and connect) and to the organizational structure (e.g., how to allocate human resources such as program managers to constituent systems). Such decision-making processes have significant implications for improving the performance of the product. It is the couplings between organization structure and product architecture, in the context of acquisition, which forms the heart of our research goal.

Motivation

Our research is motivated by a need to enable better decision-making on how to objectively select systems that comprise a complex system and allocate program managers to each of these selected systems in a manner that maximizes complex system performance, while minimizing risks associated with mismatches between program manager competencies and system development. More specifically, we refer to organizational structures based on the allocation of program manager types (based on a spectrum of program management competencies) to manage each of the selected systems in the complex system. We follow Simon’s definition of a complex system as being a hierarchical collection of systems and subsystems that are interconnected to provide some desired capability (Simon, 1962). We consider multiple collaborating systems within this definition too since complex systems are typically developed within a collaborative construct of units within and/or across an organization.

Currently, there is a lack of systematic and quantitative modeling framework to assist decision-makers in forming organizational structures that best fit the desired complex systems development and vice versa (Honda et al., 2015; MacCormack, Baldwin, & Rausnak, 2012). This lack is driven in part by difficulties associated with underlying problem of simultaneously selecting a product structure and an organizational structure in an optimal fashion. From a *product* perspective, the task of maximizing a product’s (here, complex system) performance may result in a product structure that cannot be well managed, given the population and distribution of program manager types. From an organizational perspective, on the other hand, fixing the selection of an organization’s distribution of program managers will limit the types of products that can be effectively developed. Therefore, there needs to be an objective means of selecting systems in a complex system and allocating managers in a quantitative manner.

Methodology

We first define a scope for the “product” and “organizational” components of our mathematical framework. For the organizational structure, we focus on the program manager competencies and how various skillsets and variability can impact product development. On the product side, we adopt a modular perspective on the complex system architecture where the complex system consists of a hierarchical tree of constituent systems that connect via defined interfaces and standards. We illustrate our methodology in the context of defense acquisition; here, the organizational structure is reflected by the distribution of Department of Defense (DoD) program manager types, and the complex system architecture is reflected by modular systems that are yet-to-be acquired and connected to form a complex defense system.

Our research employs a cross-discipline strategy that seeks to allocate different organizational program manager types, based on program management competency ratings, to the system acquisition life cycle architecture for optimal performance through its phases. For the organizational elements of our framework, we adopt methods from organizational psychology to translate qualitative insights from literature into a quantitative assessment of program manager competency requirements and clarify how they may relate to the execution of the defense acquisition life cycle. For the complex system architecture, we adopt the mathematical modeling techniques and abstractions as used by Davendralingam (Davendralingam, Mane, & DeLaurentis, 2012) and an optimization perspective to enable objective selection of both the complex system architecture and organizational structure.

Problem Description

We seek to address the problem of how to optimally select systems, from a candidate pool of modular systems that constitute a complex system and allocate program managers to each system in a manner that maximizes overall performance of the complex system (the "product") while minimizing risks associated with mismatches between program manager competencies and system development ("the organizational structure"). Our problem is based on a defense acquisitions and is motivated by availability of data and inputs. We first establish a model for the organizational component and a model of the complex system components of our work. The organizational model reflects the relationship of program manager competencies to defense acquisition processes that need to be executed in developing a constituent system. The product model, on the other hand, reflects how selection of different collections of constituent systems, when combined, provide a desired overarching military capability. In the following sections, we explain our modeling perspective of the organizational and product portions of our framework. We then present an optimization based approach that unifies both models within a decision-making framework. The data available for this study is derived from studies conducted on program manager competencies by Roy Wood (2010, 2014), and prior case study reports on various defense acquisition programs.

Modeling Organizational Structures (i.e., Program Manager Competency Mapping)

In modeling the organizational component of our mathematical framework, we first need to understand the context by which the organizational units (here, the program managers) perform. In the case of our defense acquisition problem, the program manager performs a series of required programmatic tasks throughout an acquisition process life cycle. The ability of the program manager to execute each of the required tasks in the life cycle, is based on a list of program manager competencies; this naturally has an impact on the end development of each system and the complex system as a whole. First, we need to identify/create a life cycle model that allows us to readily map program manager competencies. Second, we need to identify a list of program manager competencies that are relevant to our life cycle model. Lastly, we need to effectively map these program manager competencies onto the life cycle model by relating relevant subsets of these competencies to each phase of the life cycle model. In the following sections, we articulate each of the steps in the development of our organizational structure model, beginning with the identification of our life cycle model.

Life Cycle Model Identification

The first step in our organizational structures modeling process was to identify a useful model of the acquisition life cycle. For this purpose, we chose to use a swim lane process model. The decision to create a swim lane model stemmed from a qualitative analysis of life cycle models provided by the Department of Defense and the Defense Acquisition University. There are two prominent models used to describe the system acquisition life cycle of the DoD. Figure 1 is titled “Generic Acquisition Phases and Decision Points” within the literature and is presented in multiple variations throughout DoD Instruction (DoDI) 5000.02 (DoD, 2015).

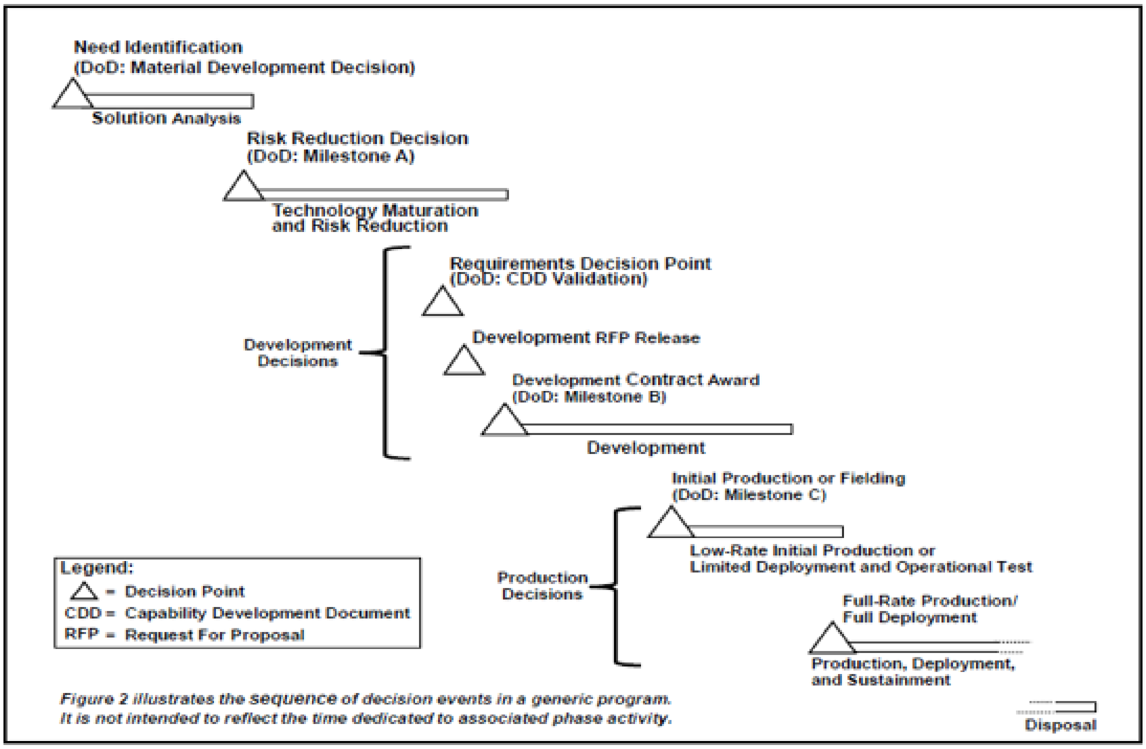


Figure 1. Generic Acquisition Phases and Decision Points
(DoD, 2015, p. 6)

For our purposes, this model does not provide enough detail to properly distinguish where the competency data would be utilized through the different phases. A significant contribution of the 5000.02 documentation is the descriptions of the phases given with Figure 1 and its ability to provide insight into the DoD program manger’s role throughout each step within the life cycle. The second model, provided by the Defense Acquisition University and presented in Figure 2, provided significantly more visual detail in the processes occurring within each phase.

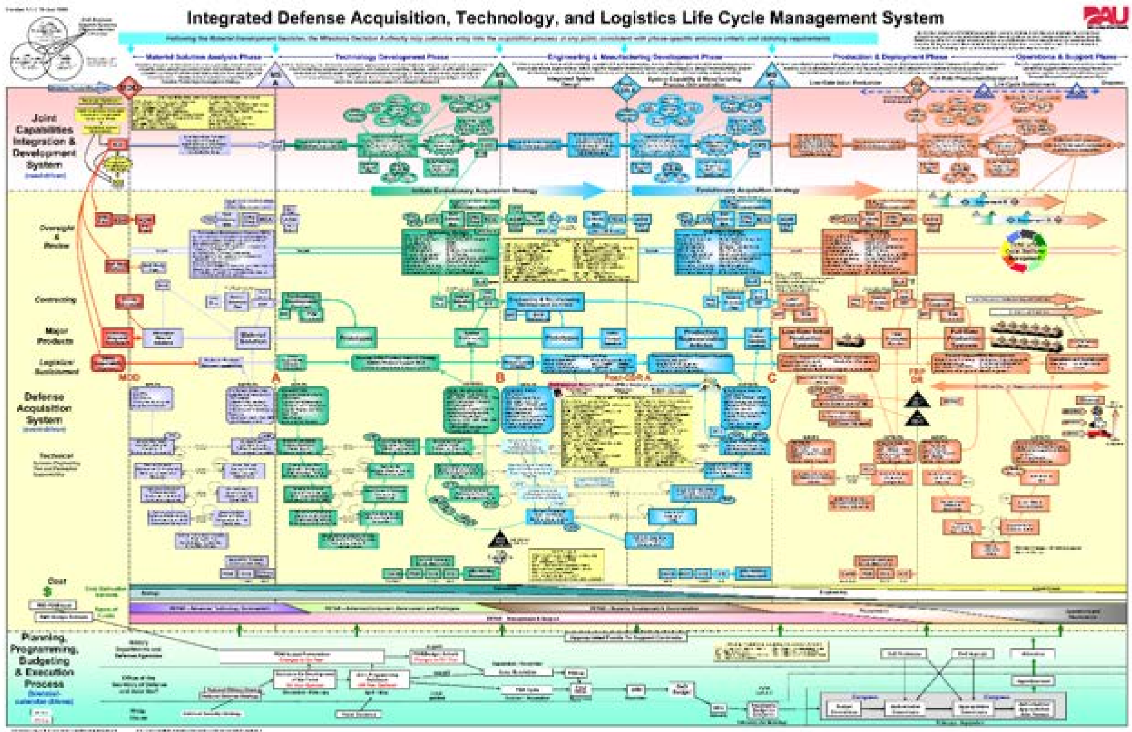


Figure 2. Integrated Defense Acquisition, Technology, and Logistics Life Cycle Management System

(Defense Acquisition University, 2009)

Due to the scope of this research, this diagram was not ideal for the time frame given to perform our analysis. Thus, we synthesized the information from both existing models forming a new model (swim lane model, Figure 3) that was executable within our given time frame.

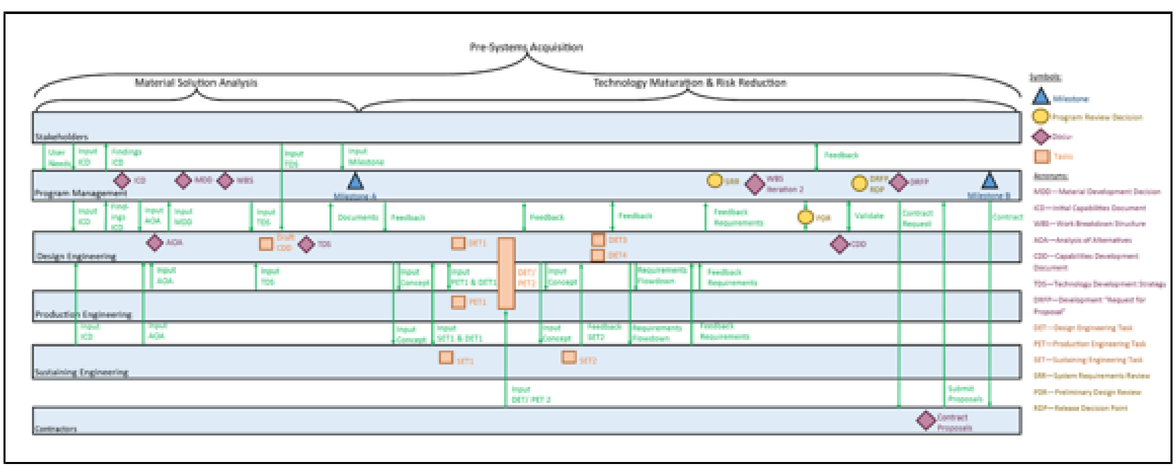


Figure 3. Swim Lane Model Depicting Processes Within the DoD System Acquisition Life Cycle

The swim lane model encompasses DoD System acquisition processes from the inception of acquisition process to Milestone B. To reduce scope for demonstration, the model was furthered reduced to processes between Milestone A and Milestone B for evaluation in the optimization problem of this paper. The swim lanes represent the tasks and interactions between the Stakeholders, Program Management, Design Engineering, Production Engineering, Sustaining Engineering, and Contractors. Each swim lane contains several actors within the DoD that were grouped within these categories based on the functions they are described to perform by DoDI 5000.02, Defense Acquisition University's Integrated Defense Life Cycle Management System visualization, and the DoD Product Support Implementation Roadmap. For example, the "Product Support Management" as stated in the DoD Integrated Product Support Implementation Roadmap diagram would fall into the "Product Management" swim lane (DoD, 2012). The elements within the swim lanes are grouped within four major categories: Milestones, Program Review Decisions, Documents, and Tasks. The Milestones, Program Review Decisions, and Documents are referenced in the instructional and GAO literature. We created the Tasks to capture steps within the life cycle that must be accomplished but are not given a formal title within the DoD literature. A description of each of the tasks are provided in Table 1.

Table 1. Swim Lane Model Task Descriptions

Task Label	Task Description
DET1	Evaluate program integration and potential risks based on Milestone A results
PET1	Evaluate potential production needs based on Milestone A results
SET1	Evaluate potential support and maintenance needs based on Milestone A results
DET/PET2	Perform competitive prototyping
SET2	Define support objectives based on competitive prototyping results
DET3	Develop system architecture
DET4	Develop technical architecture

In this study, we focus on the Technology Maturation & Risk Reduction phase. The Technology Maturation & Risk Reduction phase aims to mitigate potential risks and develop a program plan, budget, and schedule. After this phase, a contractor has been selected to pursue the program and the DoD commits its resources to the development, manufacturing, and fielding of the selected solution. The Technology Maturation & Risk Reduction was partitioned into four phases for evaluation within the optimization problem. Phase 1 begins at the conclusion of Milestone A and ends at the start of DET/PET2. Phase 2 begins at DET/PET2 and ends at the start of DET3 and DET4. Phase 3 begins at DET3 and DET4 and ends at the start of SRR. Phase 4 begins at the start of SRR and ends at the conclusion of Milestone B. The competencies addressed in Wood and the availability of qualitative data describing the program manager's role within the life cycle motivated the selection and partitioning of this phase as well as the time frame of this pilot study.

With the components of the swim lane model articulated, we can now move onto the second major phase of our organizational structure modeling—identifying the program manager competencies that can be effectively mapped onto the swim lane model. In the following sections, we articulate the competencies used, as well as the process we used to map them onto the swim lane model.

Identifying Program Manager Competencies

To map program manager competencies onto this swim-lane model, we needed to first obtain a relatively comprehensive initial list of relevant program manager competencies. For this, we utilized data collected by Wood (2010, 2014) that used a set of 35 program manager competencies indicative of the major capabilities that influenced how successful a program manager would be. Specifically, these were designed to assess the program manager competencies that "can be used in drafting project management interviewing questions, developing appraisal models to select the most qualified project managers for promotion, and designing job descriptions for project managers that can be tailored by an organization to clearly outline the roles, duties, and responsibilities of a project manager" (Golob, 2002, p. 7). These competencies were developed based upon a literature review, subject-matter expert reviews, and two surveys of program managers and the managers of program managers. A more detailed explication of these and this process can be found in Golob (2002).

These 35 competencies that resulted from this process were posited to measure 20 technical (or "hard" skills), and 15 behavioral (or "soft" skills). However, as has been posited recently in the program manager literature (Nijhuis, Vrijhoef, & Kessels, 2015) these individual program manager competencies likely are subcomponents that are attributable to more general, higher-order taxonomies of competencies from the general management/organizational psychology literatures. For example, Nijhuis et al. (2015) found that these higher-order taxonomies were effectively able to integrate the diversity of program manager competencies that had been identified in the extant literature. For example, the two soft skill competencies of project leadership (i.e., the ability to set a vision, identify the action steps, motivate others to maintain their commitment to program success and the ability to influence a team to willingly work toward predetermined program objectives) and facilitation (i.e., the ability to facilitate or guide team members through a process that helps them discover answers and overcome barriers to successful program completion) likely map onto the higher-order managerial competency of Leading and Deciding that has been well-validated within the general managerial/organizational psychology literatures (Bartram, 2005; Kurz & Bartram, 2002). Thus, while these 35 competencies are a great start, to make them practically useful for our optimization problem, as well as more theoretically parsimonious, it is important for us to map them onto these higher-order managerial competencies.

For this higher-order managerial competency mapping, we used the Great Eight model of managerial competencies (Bartram, 2005; Kurz & Bartram, 2002). These researchers defined competencies as "sets of behaviors that are instrumental in the delivery of desired results or outcomes (Bartram, Robertson, & Callinan, 2002, p. 7). The Great Eight competencies represent a parsimonious representation of the domain of managerial competencies that exist in the extant literature. The Great Eight structure has been extensively validated and refined. This refinement has created not only the broad Great Eight, but 112 component competencies that underlie the eight core dimensions. The eight core dimensions are Leading and Deciding, Supporting and Cooperating, Interacting and Presenting, Analyzing and Interpreting, Creating and Conceptualizing, Organizing and Executing, Adapting and Coping, and Enterprising and Performing.

Due to the high degree of conceptual overlap between our 35 program manager competencies and the Great Eight dimensions, we used the Great Eight as the basis for our higher-order managerial competencies. To link our 35 competencies to the Great Eight dimensions, we engaged in an iterative process of mapping the individual competencies onto the broad Great Eight. Once complete agreement of the mapping was established between all members of the research team, this mapping was finalized. With this mapping in hand, we can parsimoniously integrate these program manager competencies into our swim lane.

Deriving Baseline Great Eight Ratings From Qualitative Data

In this part, we derive a set quantitative ratings for each of Great Eight dimensions where each rating represents the degree to which each Great Eight dimension is important towards accomplishing the acquisition tasks in the swim-lane model; these ratings are considered to be baseline as they each represent an aggregate, required rating for each Great Eight dimension, based on the qualitative data from the GAO reports. To accomplish the task of generating these baseline values, it becomes necessary to properly map the program manager competencies from Wood (2010) onto the swim-lane model, through integrating the qualitative data available from the GAO reports and instructional documentation with the Wood competencies. Specifically, we utilized information regarding the tasks and competencies required at each stage of the swim-lane model to determine the importance of each competency for successful performance of the program manager at that stage in the life cycle. As articulated previously, rather than mapping each of the 35 specific competencies used within the Dr. Wood's research, we use the higher-order Great Eight dimensions that these 35 specific competencies correspond to as depicted in Table 2. This reduces our mapping from 140 ratings (i.e., 35 competencies x 4 phases) to 32 (i.e., 8 competencies x 4 phases) that is more theoretically and empirically parsimonious due to the aggregation of theoretically-redundant competencies.

Table 2. Placement of the Roy Wood Competencies to the Great Eight Dimensions

Great Eight Competencies	Roy Wood Competencies
Leading and Deciding	Document program assumptions; Implement corrective action; Project leadership; Facilitation
Supporting and Cooperating	Trustworthiness; Issue and conflict resolution; Coaching
Interacting and Presenting	Communicated program status; Negotiations; Setting and managing expectations; Communication style; Listening skills; Team building
Analyzing and Interpreting	Document program constraints; Measure program performance; Implement change control; Conduct administrative closure; Problem solving
Creating and Conceptualizing	Define program strategy; Decision making
Organizing and Executing	Determine program goals; Determine program deliverables; Quality assurance; Identify resources requirements; Develop a budget; Create a work breakdown structure (WBS); Develop a resource management plan; Establish program controls; Develop program plan; Organizational Skills
Adapting and Coping	Respond to risk; Flexibility
Enterprising and Performing	Technical ability; Sound business judgement

The process of mapping the Great Eight dimensions onto the swim-lane model was done via a systematic coding process. First, aggregated qualitative data from the GAO reports and instructional documentation were reviewed by a two-person cross-discipline team (an example of this aggregated data can be found in Table 3).

Table 3. Example of the Qualitative Aggregated Data Used to Map the Competencies to the Life Cycle Phases

Phase 3: Requirements Development Qualitative Data	
Instructional Documentation Summary	The Requirement Development effort involves tasks DET3 (develop system architecture) and DET4 (develop technical architecture), as well as, the program review decision SRR (System Requirement Review). The goal is to reduce risk and create a set of requirements which will create a baseline for the program to be presented at the PDR. The resulting requirements are additionally used in the CDD, RFP, and Milestone B.
Instructional Documentation Sources	Department of Defense, DoDI 5000.02, 2015; Department of Defense, DoD Integrated Product Support Implementation Roadmap, 2012; Defense Acquisition University, Integrated Defense Life Cycle Management System, 2004
Instructional Documentation Representative Quotes	"The point at which the major cost and performance trades have been completed and enough risk reduction has been completed to support a decision to commit to the set of requirements that will be used for preliminary design activities, development, and production (subject to reconsideration and refinement as knowledge increases)" (DoD, 2015). "Capability requirements are not expected to be static during the product life cycle. As knowledge and circumstances change, consideration of adjustments or changes may be requested by acquisition, budgeting, or requirements officials" (DoD, 2015).
GAO Reports Summary	The Requirements Development phase is hindered by the continual changing of key requirements throughout the acquisition life cycle and the lack of proper requirements development before Milestone B.
GAO Sources	GAO-08-674T, 2008; GAO-06-110, 2005; GAO-16-489T, 2016
GAO Representative Quotes	"We found four factors that have the potential to impact acquisition outcomes on individual programs: (1) unsettled requirements in acquisition programs can create significant turbulence including increased cost growth" (GAO, 2008). "Second, they (users/contractors) cannot veto new requirements. Faced with long development life cycles and promising technology advances, users often ask for new or better capabilities as a program proceeds forward. Program managers themselves are not always empowered to say "no" to demands that may overly stretch their programs, and few senior leaders above them have been willing to" (GAO, 2005). "Because DoD does not yet have approved requirements and is not planning to hold a Milestone B review, its approach for Block 4 modernization will not require the program to have such important cost, schedule, and performance reporting and oversight mechanisms in place" (GAO, 2016).

The two-person coding team consisted of one engineering graduate student with expertise in the intricacies of the program management/engineering life cycle a doctoral student in organizational psychology with expertise in leadership competencies and job performance. During the review of the aggregated GAO reports/instructional documentation, this team discussed each stage of the project life cycle, the tasks involved, how each phase fed into those which followed, and the metrics for successful performance at each phase. Once a similar frame-of-reference was created, the team discussed each of the Great Eight dimensions (considering both the general dimension, as well as the specific Roy Wood competencies underlying it) and its relevance to each phase. After the general relevance was thoroughly articulated by both members of the team, a consensus as to a numeric rating of importance (ranging from 1 to 10) for each Great Eight dimension was mapped onto each phase of the swim-lane model, for a total of 32 ratings. The team had 100%

consensus as to the final ratings. These final ratings were then used as a baseline in the development and execution of the optimization model and are presented in Table 4.

Table 4. Great Eight Mapping to Life Cycle Phases and PM Archetypes

	Acquisition Life Cycle Phase				Program Manager Archetype			
	Loop 1: Post A Concept Dev	Loop 2: Prototyping	Loop 3: Req. Dev.	Loop 4: SRR to DRFP	PM Type I	PM Type II	PM Type III	PM Type IV
Leading & Deciding	8.0	7.5	9.5	3	9	7	6	6
Supporting & Cooperating	4.5	6.5	6.5	6.5	9	6	7	3
Interacting & Presenting	9.0	9	5.5	10	7	5	4	3
Analyzing & Interpreting	2.5	5	3.5	4	5	6	3	3
Creating & Conceptualizing	2.0	8	8	6.5	5	9	9	2
Organizing & Executing	2.0	4.5	3	7.5	6	9	9	1
Adapting & Coping	2.0	4.5	2	5	3	5	5	4
Enterprising & Performing	7.0	7	8	7.5	5	5	7	3

Table 4 shows both the Great Eight Mapping assessment scores that were ascertained for each of the four studied phases of the total defense acquisition life cycle; columns 1–4 provide estimated numerical values of required level of competence, in each of the Great Eight dimensions, for the corresponding life cycle phase. Table 4 also shows a set of notional Great Eight Mapping scores for four classifications (columns 5–8) of program managers. In this example used to generate the product architecture, we *assume* that there exist four archetypes of program managers, each with a different distribution of Great Eight Mapping strengths. While the values and number of program manager archetypes in this example problem are for illustrative purposes only, we note that there are well-known quantitative methods that can be used to solicit such values in real world situations. For example, clustering algorithms such as *hierarchical clustering* can be used to quantitatively determine the number of clusters and the values of Great Eight dimensions for program managers in each cluster, given a large survey pool and survey instrument that is executed to extract relevant information.

Modeling Complex System Structures

The complex system (product) architecture portion is modelled as an interconnected set of nodes, each having a finite set of inputs and outputs. The interconnections characterize how node capabilities (outputs) feed and consequently fulfill requirements (inputs) of any connected compatible node.

Figure 4 (a) and (b) show a generalized representation of a complex system which has interdependencies between constituent systems, across multiple layers of the hierarchical structure. Each node (system) is connected to other nodes on the network, in accordance with the set of requirements needed for them to interdependently operate. The connections between nodes are also governed by a set of interaction rules. Interactions between systems are modeled as relatively simple nodal behaviors that are applicable to a wide variety of types of inter-system connections. While not exhaustive, the combinations of these nodal behaviors as modeling rules can cover a large set of real world inter-system interactions.

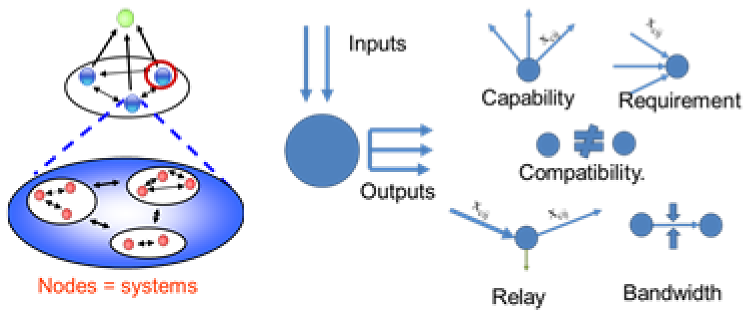


Figure 4. (a) Complex System Hierarchy (b) Nodal (System) Behaviors

Figure 4 (b) shows the five most intuitive system (node) interactions:

- Capability: systems have finite supply of capabilities that limit the number of connections they may form.
- Requirements: System requirements are fulfilled by receiving connections from other nodes that possess a capability to fulfill said requirements.
- Relay: Systems can relay capabilities between adjacent system. This can include excess input of capabilities that are used to fulfill node requirements.
- Bandwidth: Total amount of capabilities and number of connections between systems are bounded by the bandwidth of the connection linkages between systems.
- Compatibility: Systems can only connect to other systems based on a pre-established set of connection rules.

The performance of the complex system is related to the ability of the connected network of individual systems to fulfill overarching core objectives. System-wide performance is quantified by the capability of nodes that most directly contribute to the core objectives.

An Optimization Approach to Selecting Optimal Organizational and Complex System Structure

We pose the task of selecting the optimal organizational architecture and product architecture as a mathematical programming (optimization) problem involving two main segments. The first segment of an optimization problem involves an *objective function* equation that is either maximized or minimized, depending on the metric that is being used. The second segment involves a set of equations called *constraints* that reflect rules as in Figure 4 (b). A simple example of a mathematical program is the maximization of expected stock investment returns, subject to constraints on availability of funds to invest, where the decision variables are which stocks to buy, and how much to buy of each stock.

The problem of selecting an optimal complex system architecture and its organizational architecture is more specifically posed as a multi-objective optimization problem that addresses both an index that describes the level of performance for a chosen product architecture and the uncertainty in program manager performance allocated across the selected architecture. (In the simple case of the stock problem, the notion is tradeoff between expected portfolio returns and risk). The decision variables involve which systems to select in the product architecture and which program manager types to be assigned to systems that need to be developed (we explain types in a subsequent section).

Concept Application: Naval Warfare Scenario Acquisitions

Our naval warfare scenario concept application problem is based on developing a complex military system, through selection of constituent systems (from a candidate set of systems), and allocating DoD program managers in a way that maximizes the complex system performance, while minimizing risks associated with mismatches between program manager competencies and system development. The performance of the complex system is based on an aggregated performance index of its constituent systems, and risks of mismatches between program manager competencies and system development are reflected in each program manager's competencies in executing the Technology Maturation & Risk Reduction phase of the defense acquisition life cycle.

Table 5. Candidate Systems or Naval Warfare Scenario

		SoS Capabilities (Outputs)			Capabilities (Outputs)				Cost	Num Power	Num Comm	TRL
No.	System Name	SoS CAP 1	SoS CAP 2	SoS CAP 3	Power.	Comm.	Power Req.	Comm Req.	[$]	Links	Links	
1	Control Station 1	150	0	0	150	0	0	0	$10,000.00	3	3	9
2	Control Station 2	300	0	0	300	0	0	0	$20,000.00	3	3	9
3	Control Station 3	450	0	0	450	0	0	0	$300,000.00	3	3	9
4	Control Station 4	600	0	0	600	0	0	0	$400,000.00	3	3	6
5	Control Station 5	750	0	0	750	0	0	0	$500,000.00	3	3	4
6	First Satellite 1	0	0	100	0	0	75	95	$500,000.00	3	3	9
7	First Satellite 2	0	0	200	0	0	125	150	$650,000.00	3	3	9
8	First Satellite 3	0	0	300	0	0	150	250	$750,000.00	3	3	7
9	First Satellite 4	0	0	400	0	0	175	350	$850,000.00	3	3	3
10	First Satellite 5	0	0	500	0	0	185	450	$900,000.00	3	3	4
11	UAV-1	20	0	0	0	0	100	0	$200,000.00	3	3	9
12	UAV-2	30	0	0	0	0	200	0	$300,000.00	3	3	9
13	UAV-3	40	0	0	0	0	300	0	$400,000.00	3	3	4
14	UAV-4	50	0	0	0	0	120	0	$450,000.00	3	3	3
15	UAV-5	60	0	0	0	0	300	0	$500,000.00	3	3	2
16	Carrier Ship -1	0	5	0	0	0	50	0	$500,000.00	3	3	9
17	Carrier Ship -2	0	10	0	0	0	150	0	$600,000.00	3	3	9
18	Carrier Ship -3	0	20	0	0	0	200	0	$700,000.00	3	3	2
19	Second Satellite 1	0	0	100	0	100	0	0	$50,000.00	3	3	9
20	Second Satellite 2	0	0	200	0	200	0	0	$60,000.00	3	3	9
21	Second Satellite 3	0	0	300	0	300	0	0	$70,000.00	3	3	7
22	Second Satellite 4	0	0	400	0	400	0	0	$80,000.00	3	3	3
23	Second Satellite 5	0	0	500	0	500	0	0	$90,000.00	3	3	3

Table 5 lists a catalogue of systems and their hypothetical characteristics. The table shows 23 available systems that can be acquired towards development of an overarching capability, across five classes of systems (Control Station, First Satellite, UAV, Carrier Ship, Second Satellite). The first three columns (SoS CAP1, SoS CAP 2, SoS CAP 3) list outputs of system level capabilities that directly contribute to the top-level performance of the overall complex systems. For example, Control Station 1's SoS Cap1 contribution of 150 refers to a capability of 150Mbps of communication bandwidth that contributes directly to the overall performance index of the complex system in general. Columns three and four are capabilities that do not contribute directly to the top-level performance index, but contribute to satisfying constraints at a lower level of abstraction; for example, the same Control Station 1 generates 150 units of power that can be distributed to other systems that connect to it. While power is an output of Control Station 1, it is not a capability that directly contributes to the top-level capabilities of the overall complex system. Columns 5–6 are the requirements of each system. Column 7 reflects acquisition costs. Columns 8 and 9 reflect the number of other systems can link to each system; this constraint, in the case of Control Station 1, is to be able to provide power to up to three other systems that connect to it.

The last column is the Technology Readiness Level (TRL) of each system. We assume that high TRL numbers denote a commercial off-the-shelf type of system that has relatively straightforward acquisition processes in place, where as a lower TRL level system will require the assignment of a program manager to develop and mature the system towards final acquisition. We assume a finite number of each type of program managers that are available to be assigned to each system listed in Table 5. For simplicity, the measure of performance of each program manager type, in executing acquisition tasks listed in Table 1, is defined as the Euclidean norm of program managers dimensional scores (columns 5–8) that are less than the estimated required values (columns 1–4). The overall performance of the program manager in executing acquisition tasks is taken as simply the average Euclidean norm values across the four loops—here, we term this as an average risk. Values of the average risk and population of program managers for each type are tabulated in Table 6.

Table 6. Concept Problem Program Manager Population per Type

PM Type	Population	Average Risk
I	2	4.1
II	2	5.3
III	2	4.7
IV	2	10.1

Mathematical Formulation: Mixed Integer Programming (MIP)

We formulate our concept problem of maximizing a complex system's performance while minimizing program manager competency related risks as a multi-objective optimization problem. We adopt a modified version of a prior optimization model by Davendralingam that views a complex systems architecture as a collection of nodes with interdependency rules that govern their connectivity. The resulting mathematical program is as follows:

$$\max\left(\frac{\sum_i S_{ic}\cdot w\cdot x_i^B - R_c}{R_c}\right) \tag{1}$$

subject to:

$$\sum_i x_{cij} \geq x_j^B S_{rj} \tag{2}$$

$$\sum_j x_{cij} \leq x_i^B C_{ci} \tag{3}$$

$$x_1^B + \ldots + x_n^B = T \tag{4}$$

$$\sum_c x_{cij} - x_{ij}^{c_{bin}} M \leq 0 \tag{5}$$

$$M\sum_c x_{cij} - x_{ij}^{c_{bin}} \geq 0 \tag{6}$$

$$\sum_i x_{ij}^{c_{bin}} \leq n_{\max} \tag{7}$$

$$\sum_i x_{cij} - \sum_j x_{cij} - x_j^B S_{ij} = 0 \tag{8}$$

$$\sum_i x_{iq}^{PM} - x_{q \in Q} M \leq 0 \quad \forall q \in Q_{TRL<8} \tag{9}$$

$$M \sum_i x_{iq}^{PM} - x_{q \in Q'} \leq 0 \quad \forall q \in Q_{TRL<8} \tag{10}$$

$$\sum_t x_{iq}^{PM} \leq 1 \qquad \forall q \in Q \quad t=1...4 \tag{11}$$

$$C_{req}^{PM} x_{iq}^{PM} \leq C_{cap}^{PM} x_{iq}^{PM} \quad \forall q \in Q_{TRL<4} \tag{12}$$

$$C_{req}^{PM} x_{iq}^{PM} \leq E_{\max} \tag{13}$$

$$x_{cij} \in \text{real, integer}, x_q^B \in \text{binary}, x_{iq}^{PM} \in \text{binary}, x_{ij} \in \text{binary}$$

where:

S_{ic}	capability (c) of system (i)
w,	weighting factor vector of SoS capabilities (constant)
x_{ib}	binary decision variable for selecting system (i)
R_c	base SoS capability for normalization
x_{cij}	quantity of capability (c) between system (i) and (j)
x_{ij}	adjacency matrix (binary) that indicates connection between systems (i) and (j)
S_{rj}	requirement (r) of system (j)
M	Big-M constant value
Q	-set of all possible system choices (q = 1 … 23)

The mathematical model as represented by Equations 1–13 represent the formulation of a mixed integer linear programming model. The “mixed” term denotes the existence of both integer and continuous decision variables. Equation 1 is the objective function that represents the maximization of the overall complex system capability index. Here, the capability index is the normalized sum of capabilities of the complex system level capabilities (columns 1–3 in Table 5), where the normalization is done with respect to some lowest common denominator, R. Equation 2 ensures that for each system type (j) selected, there is sufficient capability type (C) being received from other connecting systems (i) that can satisfy the requirement type (R). Equation 3 ensures that the amount of capability provided by each system, type (i) for each capability type (c) does not exceed the maximum capability of type (c) for the system. Equation 4 generically defines mutual exclusivity rules for systems—for example, if selection of system 1 (x1) and system 2 (x2) is a mutually exclusive condition, then the constraint would be (X1+X2 ==1) where x1 and x2 are binary variables and constant T denotes the condition that the sum of the can only result in one system. Equations 5 and 6, more specifically, follow a “Big-M formulation” that facilitates the calculation of the number of connections that can be made to individual nodes. Equation 7 constrains the number of connections that can exist for each system type (i) and for each capability type (c) for the system. Equation 8 enforces that the total of some capability (q) that is supplied to a node (e.g., power flow or communications bandwidth), combined with its inherent capability (q) is not exceeded by demand for the capability from connected nodes.

Equations 9 and 10 jointly enforce that if a system type (q) is selected from the set of systems that have a TRL level less than 9, then a program manager must be assigned to the system. Equations 9 and 10, like Equations 5 and 6, employ the use of a Big-M formulation where the pairs of constraints act as logical conditions. Equation 11 sets the condition that only up to one program manager from the four types (t) can be assigned to each system. Equation 12 imposes the condition that for each system type (q) that belongs to the set of systems with a TRL of level 5 or below, the program manager assigned to the system needs to have a Great Eight competency score that at least meets the score for the requirements of a critical subset of the Great Eight in columns 1–3 of Table 4; these critical subsets are for the top three highest scores for the loops (1–2). Equation 13 limits the total performance error, accumulated due to assigning program managers across different systems, to a maximum value of Emax; this value is varied to generate an efficiency frontier that trades off the overall complex system performance against the uncertainty in overall program manager performance. It must be noted that while Equation 13 is a linear equation and is reflective of the relatively simple model used for our concept problem, it does not detract from more complex forms of modeling for program manager performance. With a richer collection of data, approaches that account for more explicit interdependencies between program manager interactions, when allocated to systems, can be modeled in quadratic forms (Davendralingam et al., 2012) that can be efficiently included in the current modeling framework, even under conditions of data uncertainty. Furthermore, there are a range of robust optimization techniques that can be applied to address data uncertainty as well (Davendralingam et al., 2012).

Results

The resulting optimization model as represented by Equations 1–13 is solved in MATLAB 2016b using the YALMIP toolbox with the GNU Linear Programming Kit (GLPK) solver. The problem is solved for a bounded range of values of E_{max} in Equation 13 ($5 \leq E_{max} \leq 50$) to generate the Pareto frontier that trades off the overall complex system capability index (optimal values of the objective function) against overall program manager performance; this includes Pareto filtering to only include non-dominated solutions on the efficiency frontier. Figure 5 shows the filtered Pareto frontier generated by solving the optimization model for each range value of E_{max} . Table 7 provides the breakdown of selected systems that comprise the portfolio of systems within the overall complex system, and program manager allocations across each portfolio point on the efficiency frontier.

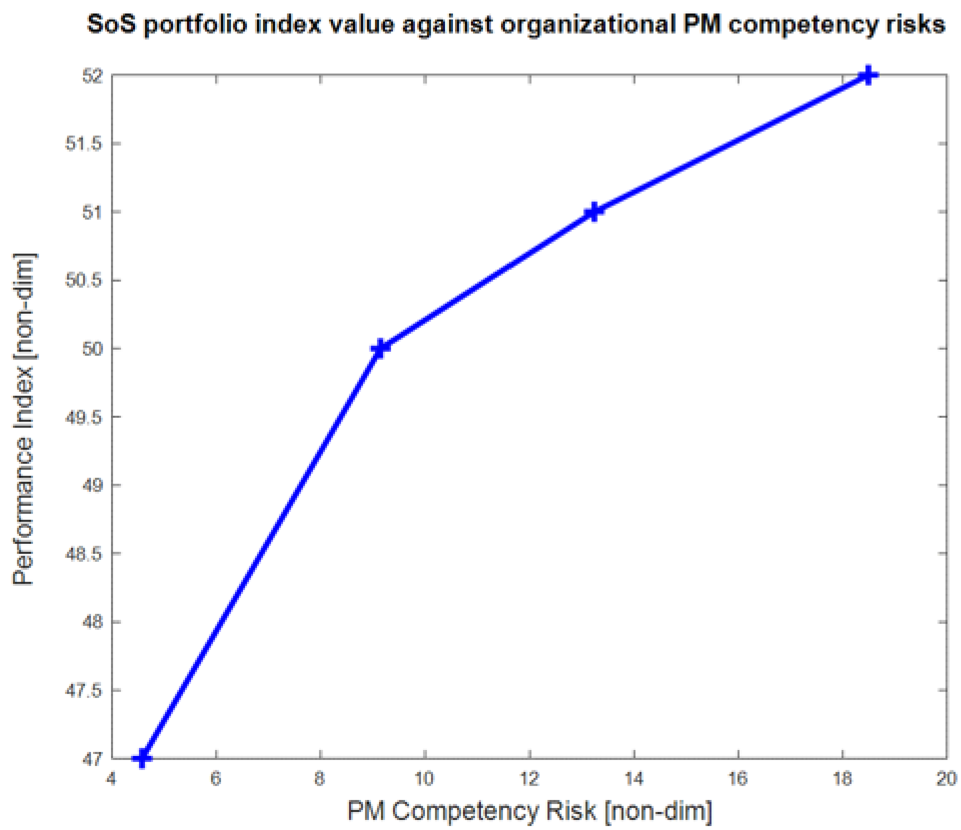


Figure 5. Efficiency Frontier of Performance Against PM Competency Risk (Risk Measured as Average Mean Squared Error)

Table 7. Portfolio of Systems and Program Manager Allocations

No.	Candidate Systems	Portfolio 1	Portfolio 2	Portfolio 3	Portfolio 4
1	Control Station 1	-	-	-	-
2	Control Station 2	-	-	-	-
3	Control Station 3	-	-	-	-
4	Control Station 4	-	-	-	-
5	Control Station 5	X	X	X	X
6	First Satellite 1	-	-	-	-
7	First Satellite 2	X	-	-	-
8	First Satellite 3	-	-	-	-
9	First Satellite 4	-	-	X	X
10	First Satellite 5	-	X	-	-
11	UAV-1	-	-	-	-
12	UAV-2	X	X	X	X
13	UAV-3	-	-	-	-
14	UAV-4	-	-	-	-
15	UAV-5	-	-	-	-
16	Carrier Ship -1	-	-	-	-
17	Carrier Ship -2	X	X	X	-
18	Carrier Ship -3	-	-	-	X
19	Second Satellite 1	-	-	-	-
20	Second Satellite 2	X	X	-	-
21	Second Satellite 3	-	-	-	X
22	Second Satellite 4	-	-	X	-
23	Second Satellite 5	X	X	X	X
Program Manager Type		# of PMs (system # PM allocated to)			
I		-	-	1 (9)	-
II		-	-	-	2(9,21)
III		1 (23)	2 (23,10)	2(22,23)	2(18,23)
IV		-	-	-	-

The results generated through solving the optimization problem of Equations 1–13 provide a way for decision-makers to assess potential tradeoffs between selecting different complex system architectures (here, portfolio of interconnected systems) and organizational architecture (here, program manager type allocations) by relegating some combinatorial aspects of the problem to the algorithm and delegating decision-making to the practitioner. The results show the progressive levels of complex system performance that can optimally be achieved, given each prescribed acceptable level of risk associated with the program manager performance, for each portfolio. As more capable systems are brought into the picture, to generate a higher performing complex system, program managers are additionally added in an optimized sense, in a manner that bounds risk the sequential increments enforced in Equation 13. The program manager allocation also adheres to the rulesets established (for example, the constraints established for allocation of program managers to systems with TRL<9 and TRL<5 as established in prior sections). While an initial instinct may be to first select program managers that are, on average, the least risky following Table 6, we see instead that the optimization selects program manager Type III in Portfolio 1 and Portfolio 2, due to the enabling effect that Type III manager has on developing low TRL systems with a higher potential to improve the complex system performance index. Another useful observation of the results presented, is that the solution generated by the optimization routine, reveals potential pathways for evolving an architecture; for example, when considering portfolios 3 and 4, we observe that a future upgrade from portfolio 3 to 4 will include retirement of Carrier-Ship 2 and a Second Satellite-2 unit, in favor of a Second Satellite-3 unit and a Carrier-Ship 3; this path of system addition and replacement is complemented by the need to replace a Type I program manager with two Type II program managers to facilitate the architectural transition. Early stage knowledge on such shifts can enable the correct requirements to be set on what type of program managers to look for or train for these future updates, thereby minimizing risks and organizational misalignments.

As the number of candidate systems increases and the dependencies increase as well, it becomes very difficult to objectively select systems that constitute a complex system and program managers that manage each of the constituent systems without the aid of quantitative means such the mathematical framework presented in this paper. The mixed integer programming formulation is efficient even for much larger instances of number of systems (and/or number of program manager types), assuming the same problem abstraction being used in this paper. Furthermore, the MIP perspective lends itself to further formulations of the problem at hand to better account for various forms of interdependencies between product and organization and data uncertainty.

Concluding Statements and Future Work

The approach presented in this paper represents a preliminary quantitative framework that facilitates the optimal selection of an organizational architecture and product architecture (in this case, a complex system architecture). The approach leverages theories from industrial organizational psychology and mathematical programming techniques from operations research to yield a unified approach that facilitates the selection of an optimal composition of systems and organizational structure (in this case, program managers) towards achieving a desired complex system performance.

We demonstrate the work for a concept problem based on a naval defense acquisition scenario and present the mathematical formulation and example solution of the problem. The concept problem utilizes a combination of qualitative and quantitative measures, driven by prior literature and *a priori* insights from program manager competency literature, to form the foundation of the organizational elements in our concept problem. The resulting mathematical programming problem is posed in a very flexible framework of a mixed integer linear programming problem—to which there are very well understood means of solution, even for large scale problems.

Potential future research may encompass extensions on the modeling techniques for capturing interdependency behaviors between interacting organizational elements (e.g., modeling interaction behaviors of program managers) and adapting the mathematical modeling to more explicitly include such interactions. Furthermore, additional elements of organizational structure, such as acquisition processes relevant to the acquisition of individual systems, can also be brought to bear within this framework.

References

Austin-Brenemann, J., Honda, T., & Yang, M. (2012). A study of student design team behaviors in complex system design. *Journal of Mechanical Design, 134*.

Bartram, D. (2005). *The great eight competencies: A criterion-centric approach to validation.*

Bartram, D., Robertson, I. T., & Callinan, M. (2002). Introduction: A framework for examining organizational effectiveness. In I. T. Robertson, M. Callinan, & D. Bartram (Eds.), *Organizational effectiveness: The role of psychology* (pp. 1–10). Chichester, UK: Wiley.

Davendralingam, N., Mane, M., & DeLaurentis, D. A. (2012). Capability and development risk management in system-of-systems architectures: A portfolio approach to decision making. In *Proceedings of the Ninth Annual Acquisition Research Symposium*. Monterey, CA: Naval Postgraduate School.

Defense Acquisition University. (2009). *Integrated defense acquisition, technology, and logistics life cycle management system*.

DeLaurentis, D. (2015). Optimal selection of organizational structuring for complex system development and acquisitions. *FY15 Acquisition Research Program*.

DoD. (2015). *Operation of the defense acquisition system* (DoD Instruction 5000.02). Washington, DC: Author.

Golob, M. P. (2002). *Implementing project management competencies in the workplace.* Unpublished doctoral dissertation, Capella University, Minneapolis, MN.

Honda, T., Ciucci, F., Lewis, K., & Yang, M. (2015). Comparison of information passing strategies in system-level modeling. *AIAA Journal, 53*(5), 1121–1133.

Kurz, R., & Bartram, D. (2002). Competency and individual performance: Modelling the world of work. In I. T. Robertson, M. Callinan, & D. Bartram (Eds.), Organizational effectiveness: *The role of psychology* (pp. 227–255). Chichester, UK: Wiley.

MacCormack, A., Baldwin, C., & Rausnak, J. (2012). Exploring the duality between product and organizational architectures: A test of the “mirroring” hypothesis. Elsevier, 41(8), 1309–1324.

Nijhuis, S. A., Vrijhoef, R., & Kessels, J. W. M. (2015). Towards a taxonomy for project management competences. *Procedia-Social and Behavioral Sciences, 194,* 181–191.

Wood, R. (2010). How well are PMs doing? Industry view of defense program manager counterparts. *Defense Acquisition University*, 206–218.

Acknowledgments

The authors would like to acknowledge Dr. Robert Kenley (Professor of Practice, School of Industrial Engineering, Purdue University) and Mr. Daniel Dumbacher (Professor of Practice, School of Industrial Engineering, Purdue University) for their invaluable insights into the defense acquisition life cycle and related processes. Further acknowledgments to Dr. Roy Wood for providing data from the survey paper on program manager competencies and deeply useful discussions on program manager assessments. The authors also wish to acknowledge the Naval Postgraduate School for funding this work under NPS Grant: N00244-16-1-0005.

A Systems Complexity-Based Assessment of Risk in Acquisition and Development Programs

Antonio Pugliese—is a PhD candidate in systems engineering at Stevens Institute of Technology in Hoboken, NJ. He received his BS and MS in aerospace engineering and a postgraduate master in systems engineering from the University of Naples Federico II in Italy. His doctoral research is on structural complexity metrics for cyberphysical systems.

Roshanak Rose Nilchiani—is an Associate Professor of systems engineering at the Stevens Institute of Technology. She received her BS in mechanical engineering from Sharif University of Technology, MS in engineering mechanics from University of Nebraska-Lincoln, and a PhD in aerospace systems from Massachusetts Institute of Technology. Her research focuses on computational modeling of complexity and systemalities for space systems and other engineering systems, including the relationship between system complexity, uncertainty, emergence, and risk. The other track of her current research focuses on quantifying, measuring, and embedding resilience and sustainability in large-scale critical infrastructure systems.

Abstract

Development and acquisition efforts of cyberphysical systems can often encounter cost or schedule overruns due to the complexity of the system. It has been shown that a certain amount of system complexity is related to the system functionalities (effective complexity), whereas excessive complexity is related to unnecessary intricacies in the design (apparent complexity). While the former is necessary, the latter can be removed through precise local redesign. One of the major challenges of systems engineering today is the development of tools, quantitative measures, and models for the identification of apparent complexity within the system.

This research has the goal of evaluating and measuring the structural complexity of the engineered system, and does it through the analysis of its graph representation. The concepts of graph energy and other spectral invariant quantities allow for the definition of an innovative complexity metric. This metric can be applied knowing the design of the system, to understand which areas are more in need of redesign so that the apparent complexity can be reduced.

Introduction

Complexity is one of the hallmarks of all engineered systems, specially a prominent feature of defense acquisition programs. Complex engineered systems are continuously exposed to various types of uncertainties, risks, and failures in their life cycle. The causes of failures and risks are either a known observed phenomenon and perhaps overlooked in the development phase, or it is a new type of failure. The former case can lead to improvements in engineering design and management and systems engineering processes of a complex system. The latter case, instead, can potentially provide useful information that can be obtained only through unfolding of these types of failures and events. Complex engineered systems design effort resides partially in the domain of known risks and uncertainties. This domain, also known as the domain of complicated systems, is characterized by known unknowns which can be addressed with time and effort, through theoretical and experimental research. This means that systems in this domain can express large epistemological emergence, given to the lack of knowledge, but a low ontological one, meaning that the knowledge can be obtained.

However, the engineered systems with high levels of ontological emergence, meaning that the system under study is so far from the current level of knowledge, show low levels of predictability in behavior. Complex engineered systems involve humans with certain levels of autonomy interacting with the engineered systems. Predicting the behavior of a complex system, characterized emergent phenomena is very challenging. To reduce various risks in design and operation of an acquisition program, systems engineers have attempted to operate within the domain of knowable risks as much as possible, through mitigation and exploitation techniques such as trade space exploration, modular designs, open architectures, redundancy, verification, and testing. However, a comprehensive, applied, and formal way to measure and capture various dimensions of complexity and risks in the life cycle of a complex engineered system or an acquisition program is lacking. In past years, the systems engineering research community has introduced some measures of complexity for engineering design; however, domain dependence and limitation in universality of use of these measures has seriously limited their use in the decision-making process. In this paper, we introduce a constructed measure of complexity that has carefully tried to address as well as to avoid many shortcomings of existing quantitative measures of complexity.

This paper begins with a literature review on state of the art complexity and emergence. The literature review also covers some existing measures of complexity and their merits as well as shortcomings and limitations. The paper continues to introduce the concepts of spectral theory of systems complexity and explains matrix energy and directed edges in our suggested complexity measure. The paper concludes by analyzing the results and sets the stage for the future work of various case studies, quantitatively connecting emergence to spectral complexity measures of an engineered system or an acquisition program.

Literature Review

What Is Complexity?

Three Types of Problems

The first hint to the role of complexity in science and engineering design has been given by Weaver (1948). He described three distinct types of problems: problems of simplicity, problems of disorganized complexity, and problems of organized complexity. Problems of simplicity are the problems with a low number of variables that have been tackled in the nineteenth century. An example is the classical Newtonian mechanics, where the motion of a body can be described with differential equations in three dimensions. In these problems, the behavior of the system is predicted by integrating equations that describe the behavior of its components. Problems of disorganized complexity are the ones with a very large number of variables that have been tackled in the twentieth century. The most immediate example is the motion of gas particles, or as an analogy, the motion of a million balls rolling on a billiard table. The statistical methods developed are applicable when particles behave in an unorganized way and their interaction is limited to the time they touch each other—which is very short. In these problems, it has been possible to describe the behavior of the system without looking at its components or the interaction among them. Problems of organized complexity are the ones that are to be tackled in the twenty-first century, and the ones that see many variables showing the feature of organization. These problems have variables that are closely interrelated and influence each other dynamically. This high level of interaction that gives rise to organization is the reason these problems cannot be solved easily. Weaver (1948) described them as solvable with the help of powerful calculators, but today's technology is not yet able yet to solve the most complex of these problems. These are the problems that nowadays we define as "complex." Predicting the behavior of a system with many interconnected parts changing their behavior in line with the state of other components is a problem of organized complexity, and the system itself is referred to as a complex system.

The Point of View of the Observer

On his blog, Rouse (2016) wrote about the absolute or relative nature of complexity. To illustrate his stand on the matter, he considers two uses for a Boeing 747: as a paperweight and as an airliner. This thought exercise allows us to understand that the 747 as a paperweight is not very complex. It does a perfect job, given its large mass, but carries no complexity in its operation. On the other side, the airliner function exposes all the operational difficulties of flying and maintaining an airplane. From this example, he concludes that complexity should be defined in terms of a relationship between the entity and an observer. Thus, complexity is relative to the point of view of the observer.

Wade and Heydari (2014) categorized complexity definition into three major groups, according to the point of view of the observer. When the observer is external to the system and can only interact with it as a black box, then the type of complexity that can be measured is called behavioral complexity, since it looks at the overall behavior of the system. When the observer has access to the internal structure of the system, such as blueprints and source code for engineered systems, or scientific knowledge for natural systems, then the structural complexity of the system is the one being measured. If the process of constructing the entity is under observation, then the constructive complexity is to be measured, which is the complexity of the building process. This definition relates complexity to the difficulty of determining the output of the system.

Fischi built a framework for the measurement of dynamic complexity entirely based on the role of the observer (Fischi, Nilchiani, & Wade, 2015). The definition of complexity used in this framework is based on the system being observed, the capabilities of the observer, and the behavior that the observer is trying to predict.

Complexity and Emergence

Often complex systems have behaviors that cannot be immediately explained, and for this reason complexity is associated with the concept of emergence. As defined by Checkland (1981), emergence is “the principle that entities exhibit properties which are meaningful only when attributed to the whole, not to its parts.” In other words, an emergent phenomenon is a phenomenon at the macro-level that was not hard-coded at the micro-level (Page, 1999), and which can be described independently from the underlying phenomena that caused it (Abbott, 2006).

Both natural and engineered systems are capable of expressing emergence. One example of emergence in natural system is wetness. Water molecules can be arranged in three different phases (i.e., solid, liquid, and gas), but only one of them expresses a certain type of behavior—that is, high adherence to surfaces. This behavior is due to the intermolecular hydrogen bonds that affect the surface tension of water drops. These bonds are also active in the solid and liquid phase, but in those cases, they are either too strong or too weak to generate wetness. As we will see for many systems, some properties emerge only when conditions are just right. In engineered systems, the system requirements and software specifications are supposed to be written in such a way that they are independent from their implementation. For this reason, the functions and properties they describe are emergent (Abbott, 2006).

These definitions of emergence often do not differentiate on whether the emergent property is expected or unexpected, and this is obvious, since not every system has a designer that is putting together components to generate the system, and therefore sometimes there is no one to expect the property. Natural systems, which are created through evolution, do not justify the classification of emergence into expected and unexpected. In engineered systems, this is of course not true. The system engineer is responsible for the identification of the properties of the system in relation to its environment. Not only the operational environment, but also assembly, integration, testing, and disposal environment. In the design process, it is customary to differentiate between the attributes of the system that are wanted, and therefore expected, and the ones that are unexpected, which can be beneficial or adverse.

Two Types of Emergence

The works of Chalmers (2008), Bedau (1997), and Kauffman (2007) identify differences between two types of emergence: epistemological and ontological.

Kauffman proposes two approaches to the nature of emergence. The reductionist approach sees emergence as epistemological, meaning that the knowledge about the systems is not yet adequate to describe the emergent phenomenon, but it can improve and explain it in future. This is the case of wetness, where knowledge about molecules and intermolecular interactions can explain the emergent phenomenon. On the other hand, there is the ontological emergence approach, which says that “not only we don’t know that will happen, [but] we don’t even know what can happen,” meaning that there is a gap to fill not only about the outcome of an experiment (or process), but also about all the possible outcomes (Kauffman, 2007). Ontological emergence is given by the enormous amount of states the system could evolve into. The evolution of the swimming bladder in fish is an example of ontological emergence (Longo, Montévil, & Kauffman, 2012). An organ that gives neutral buoyancy in the water column as its main function also enables the evolution of some kinds of worms and bacteria that will live in it. Ontological (or radical) emergence is given by the enormous amount of states the system could evolve into. In these cases, we not only are not able to predict which state will happen, but not even what are the possible states.

Chalmers (2008) provides definitions for two different types of emergence, weak and strong, based on the capabilities of the observer. At the lower level, there is weak emergence, which includes any property possessed by the whole and not its parts. A chair is an example of weak emergence since the property of allowing someone to sit is present in the whole but not in its parts. At the upper level, there is strong emergence with the example of consciousness. In the case of weak emergence, the emergent phenomenon is just unexpected, while in the case of strong emergence, it is completely non-deducible. This of course depends on the capabilities of the observer in linking the phenomena at the two levels. Chalmers, being a philosopher and cognitive scientist, implicitly assumes that the observer has the knowledge and capabilities of a human being. An example that he provides to illustrate the difference between weak and strong emergence is the high-level patterns in cellular automata. These patterns are unexpected but deducible just by looking at the low-level rules of the automaton, making them weakly emergent and not strongly emergent. The only example that Chalmers provides of strong emergence is consciousness, and he goes along to state that there is no other such phenomenon other than the ones in which the strong emergence derives “wholly from a dependence on the strongly emergent phenomena of consciousness.” Thus, the way of differentiating between a system with weak or strong emergence is to look for conscious elements within the system.

Complexity and Complication

The idea that complexity also depends on the tools and knowledge available to the observer is common to many researchers. Crawley, Cameron, and Selva (2015) use the concepts of essential and apparent complexity to make a distinction between complexity and complication. Being engineers, they only consider designed systems. Essential complexity comes from functionality and represents the minimum amount of complexity required for the desired functionalities to emerge. Apparent complexity, on the other hand, represents the unnecessary intricacies that a designed system can have. These are the architectural features that are not required from the functionality, which make the design complicated and hard to understand. The role of the system architect is to minimize the apparent complexity without affecting the essential one.

Evolved systems do not have functions, but they create advantages to their stakeholders. The presence of a heart in the cardiovascular system of many animals is only explained by the advantage it creates in the distribution of resources within the organism. The heart does various things, such as pumping blood and making the characteristic heart sound. The categorization of these behaviors into functions and side effects has no meaning in evolved systems, and therefore the concepts of essential and apparent complexity lose their validity (Longo, Montévil, & Kauffman, 2012). On the other side, the work of Chaisson suggests that evolution keeps the apparent complexity as low as possible by selecting the unfit organisms. This implies that the complexity of an evolved organism, which has thrived for a substantial amount of time in its environment, is close to its essential complexity (Chaisson, 2014). This means that evolved systems are complex but not complicated.

Gell-Mann (1995), being a physicist, has a more holistic definition of effective complexity and logical depth (apparent complexity). Effective complexity is the length of a concise description of the regularities of an entity. This quantity should not be confused with logical depth. Mandelbrot’s set has high logical depth, since being a fractal, a simple rule is applied infinite times in a recursive fashion, but a low amount of effective complexity, since the formula used to describe it is relatively short. This is in general true for all fractals.

In the decision-making field, the Cynefin framework has been proposed by Snowden (2005) to help identify the best approach to solving a specific problem.This sense-making model can be used to understand (from the data available) the characteristics of the problem at hand and which strategy will lead to a solution. As shown in Figure 1, the framework identifies five domains of knowledge: simple and complicated, which are ordered; complex and chaotic, which are unordered; and disorder.

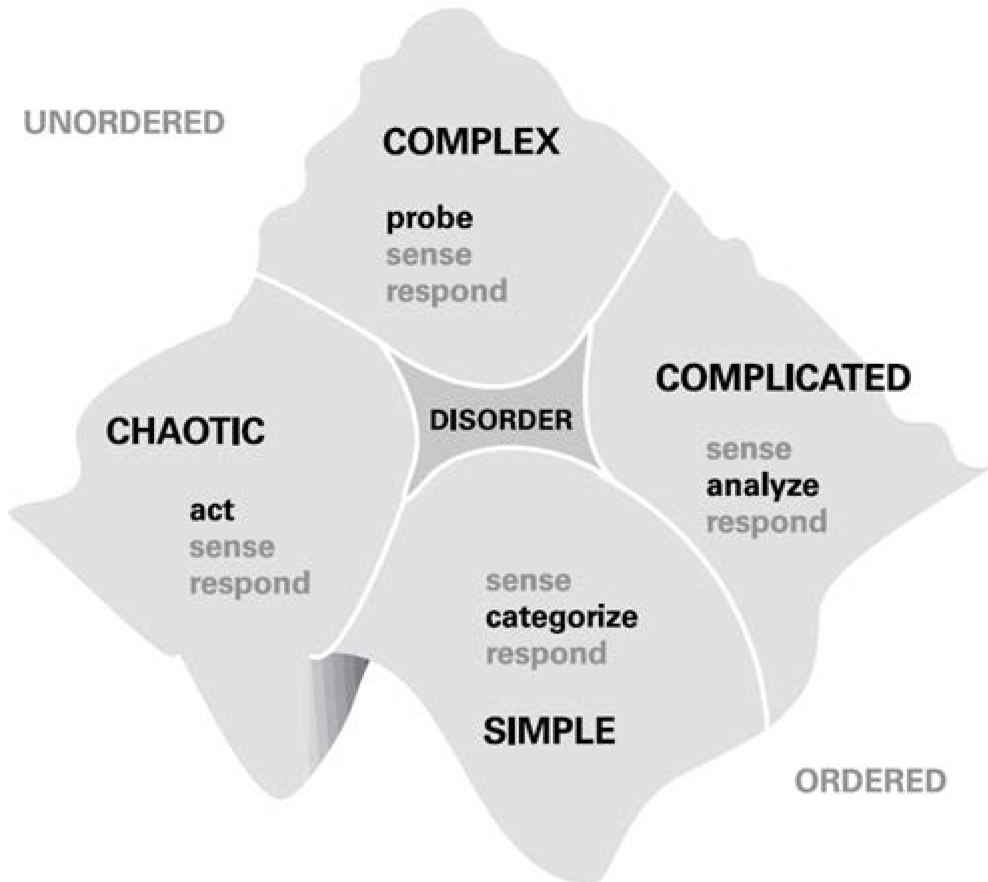


Figure 1. Representation of the Cynefin Framework
(Snowden & Boone, 2007)

In the simple domain, systems decisions can be taken unanimously with all the parties, due to the shared understanding of the matter and to the clear relationships between cause and effects. The simple domain is the domain of best practice, where once a solution to a problem has been found, it is applicable unless there is a domain shift. In the complicated domain, there still is a relationship between cause and effect, but not all the parties are able to discern it. The answer to the problem often is not best practice, and some relatively deep analysis is necessary to find a proper solution. This definition is common to the definition of complication as arising from intricacies, where the problem is made difficult because of the way it is formulated. In the complex domain, there might not be a right answer at all, and the relationships between cause and effect can be identified only in retrospect. This is due to ontological emergence, which can create higher logical structures in which feedback loops are hidden. In the chaotic domain, no patterns are discernible, and no relationships can be identified. This is the domain of emergency, where the immediate goal is not to find a solution to a problem, but to bring the system back to an ordered state, from which a solution can be found (Snowden & Boone, 2007).

Wade and Heydari (2014) provide a more technical description of the Cynefin framework. The simple and chaotic domains have both low complication, the former with low complexity, and the latter with high one. The complicated and complex domains have both high complication, the former with low complexity, and the latter with high one. Systems with high complication can be analyzed with a reductionist approach such as decomposition, while systems with high complexity cannot. This point of view on reductionism is shared by Bedau, according to whom weak (epistemological) emergence can be analyzed using reductionist techniques (Bedau, 1997). The possibility of applying a reductionist approach to the systems exhibiting only weak emergence allows to connect these two types of emergence with the definition of complicated and complex system. Complicated systems are the ones which exhibit weak emergence, which can be analyzed using reductionist techniques, and in which the emergent phenomena are unexpected but still predictable. Complex systems are the ones in which strong emergence comes in place, where the reductionist approach does not work, and where high-level behaviors are not predictable.

Structural, Dynamic, and Socio-Political Complexity

In engineered systems, complexity can be divided into six types (Sheard & Mostashari, 2010).

- Structural complexity
  - Size, or number of elements in the system, number of types of elements, instances of a certain type.
  - Connectivity, number of connections, types of connections.
  - Topology, architectural patterns, local and global patterns.
- Dynamic complexity
  - Short term, at the time scale of the system operations, behavior of the system while executing its functions.
  - Long term, at the lifetime scale, evolutionary process of the system, retirement or mission extension.
- Socio-political complexity, anything having to do with humans, cognitive limitations, social phenomena.

How to Measure Complexity?

With a better understanding of complexity, we can now look at how this quantity can be measured.

Cyclomatic Complexity

McCabe (1976) provided a complexity metric for software systems. This metric looks at the graph representation of the program, and it is defined as

$$v(G) = e - n + p$$

where *e* is the number of edges, *n* the number of vertices, and p the number of connected components in the graph. This metric is called the *cyclomatic number*. It can be demonstrated that in a strongly connected graph, the cyclomatic number is equal to the maximum number of linearly independent circuits (McCabe, 1976).

Free Energy Density Rate

Chaisson (2004) proposed a metric for the evaluation of complexity based on the amount of energy of the entity under study. More precisely, energy rate density, which is “the amount of energy available for work while passing through a system per unit time and per unit mass” (Chaisson, 2015). This metric is a boundary metric, since it considers the input and output of the system without looking at its internal structure. It has been derived through the generalization of various metrics used in various fields, such as propellant to mass ratio for engineering, or metabolic rate for biology. The metric has been evaluated for multiple entities such as galaxies, stars, planets, plants, animals, societies, and technological systems, showing a rising trend in complexity (Chaisson, 2014).

Per this metric, a system with a large intake of energy per second (i.e., power) and a low mass will be a very complex one. From the engineering point of view, this can also represent a very inefficient system. The success of Chaisson's metric is because the systems under study are mostly evolved systems or are designed with efficiency in mind. Therefore, the applicability of this metric assumes that the system has been designed, or shaped by evolution, in such a way that there is no waste of energy, or useless mass.

Propagation Cost and Clustered Cost

MacCormack presented two types of metrics for the evaluation of the complexity of software systems (MacCormack, Rusnak, & Baldwin, 2006). The directed dependency between files in the source code is the function call. The propagation cost is the average of the visibility of modifications to dependent files, while the clustered cost considers the importance of the node scaling the relative cost accordingly.

Spectral Structural Complexity Metric

Sinha presented a structural complexity metric based on the design structural matrix (DSM) of the system (Sinha & de Weck, 2012). The metric is defined as

$$C(n,m,A) = \underbrace{\sum_{i=1}^{n} \alpha_i}_{C_1} + \underbrace{\left(\sum_{i=1}^{n}\sum_{j=1}^{n} \beta_{ij} A_{ij}\right)}_{C_2} \underbrace{\gamma E(A)}_{C_3}$$

where *n* is the number of components in the system, m the number of interfaces, *A* the DSM, α_i the complexity of each component, $\beta_{ij} = f_{ij}\alpha_i\alpha_j$ the complexity of each interface, $\gamma = 1/n$ a normalization factor, and *E(A)* the matrix energy of the DSM. C_1 is the complexity contribution of the components, C_2 is the contribution of the interfaces, and C_3 is the contribution of the topology. The application of the metric sees the evaluation of α_i through expert judgment, and assumes $f_{ij} = 1$ for lack of information (Sinha & de Weck and Olivier, 2013).

Graph Energy

The metric is inspired from the Hückel Molecular Orbital (HMO) Theory, which evaluates the energy of π-bonds in conjugated hydrocarbon molecules as a solution of the time-independent Schrödinger equation

$$H\psi = E\psi$$

where *H* is the Hamiltonian matrix, and *E* the energy corresponding to the molecular orbital. This equation is an eigenvalue problem of the Hamiltonian. In 1978, Gutman defined the energy of a graph, as

$$E = \sum_{i=1}^{n} |\lambda_i|$$

where λ_i are the eigenvalues of the adjacency matrix representing the carbon substructure of the molecule (Gutman & Shao, 2011).

Instead of the eigenvalues, the approach introduced by Nikiforov (2007) and embraced by Sihna evaluates the graph energy using the singular values of the matrix. This modification extends the applicability to directed graphs where the adjacency matrix is not symmetric, while for undirected ones where the adjacency matrix is symmetric matrix, this new approach is coincident with the original eigenvalues one.

The HMO theory is applied to structures of carbon atoms, which are homogeneous. Its application to systems of heterogeneous components, this metric does not consider the role of components with different levels of complexity.

Methodology

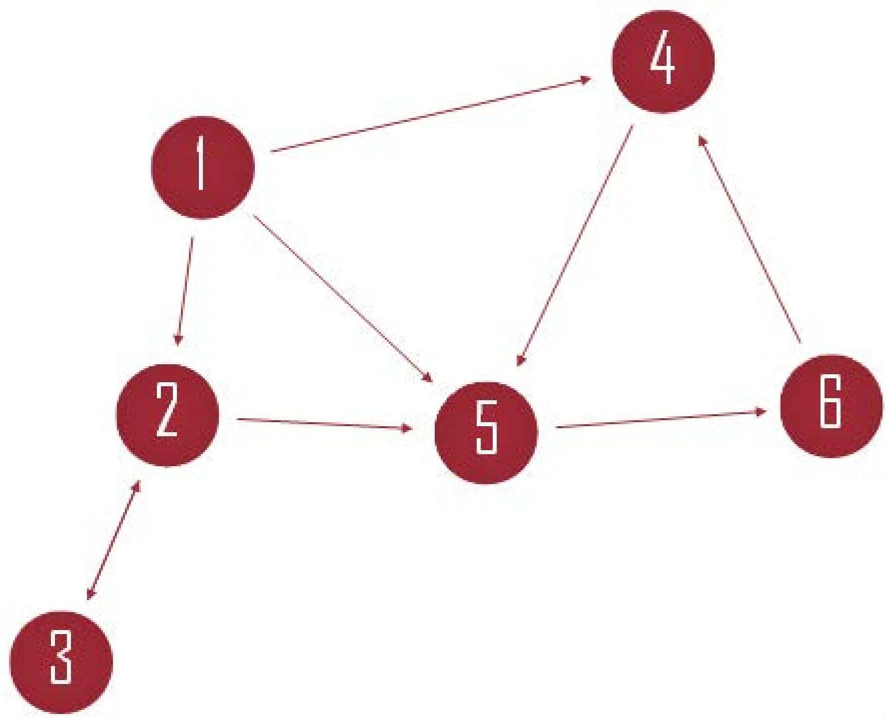


Figure 2. Graph Representation of a System

The goal of this research is to measure the structural complexity of engineered systems. The system of interest can vary from a piece of software controlling a reaction wheel to an attitude control system, a satellite, and up to a whole network of satellites. Let's consider the system represented in Figure 2. This graph is a general representation of any engineered system, in which the components are represented by the vertices and the interfaces by the edges. The generality of this approach allows one to evaluate the complexity of the more disparate engineered systems if they can be represented as a graph. The complexity metric proposed by Sinha needs the following data to be available: the complexity of each component α_i, the complexity of each interface β_{ij}, and the adjacency matrix *A*. Here we describe two limitations of Sinha's approach, which will be overcome by the newly developed metric.

Component Swap Test

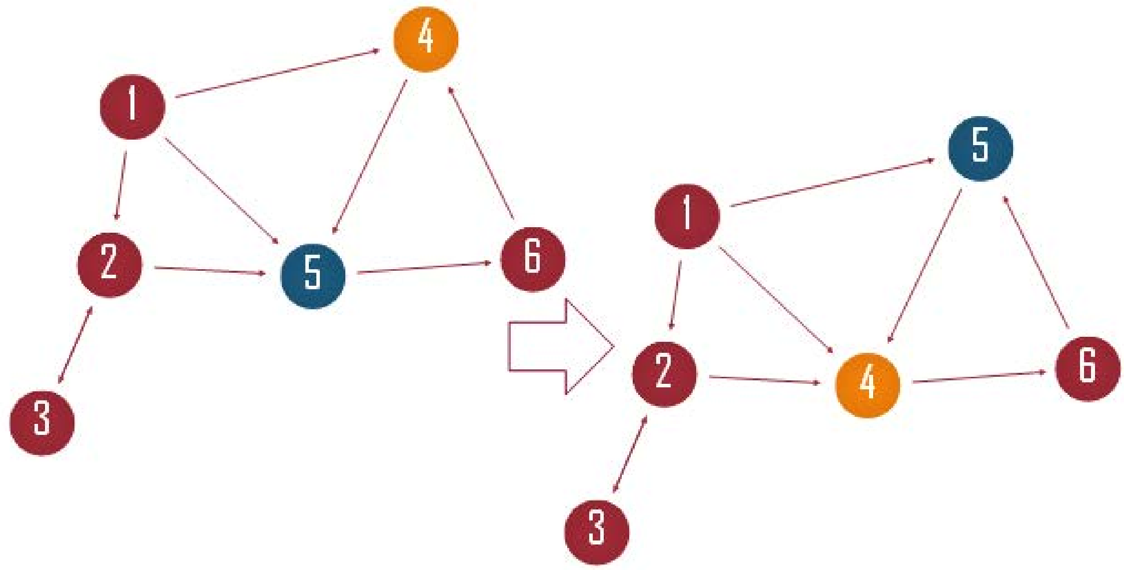


Figure 3. Swapping of Nodes Within a Graph

The component complexity $\alpha_i$represents the complexity of the irreducible components at a certain hierarchical level within the system representation. In the graph representation of the system, it can be represented as the weight of a looping edge over each vertex. Let's consider two vertices, u and v, and their weights, $w(u,u) = \alpha_u$ and $w(v,v) = \alpha_v$. Swapping the weights to have $w(u,u) = \alpha_v$and $w(v,v) = \alpha_u$ should generally reflect a change in the value of the structural complexity metric.

As an example, consider two separate temperature control systems within a building. One takes care of a conference room and the other one of a biotech laboratory. These two systems are going to have in general very different essential complexities, as it can be seen from their required level of performance (e.g., accuracy, responsiveness). The complexity of the whole building would generally be affected in case these two systems are swapped. A good complexity metric should be able to verify the component swap criterion.

The complexity metric developed by Sinha is not able to distinguish between the two systems. This is because the contribution of the components C_1 is evaluated using a sum of the component complexities, which is commutative.

Interface Swap Test

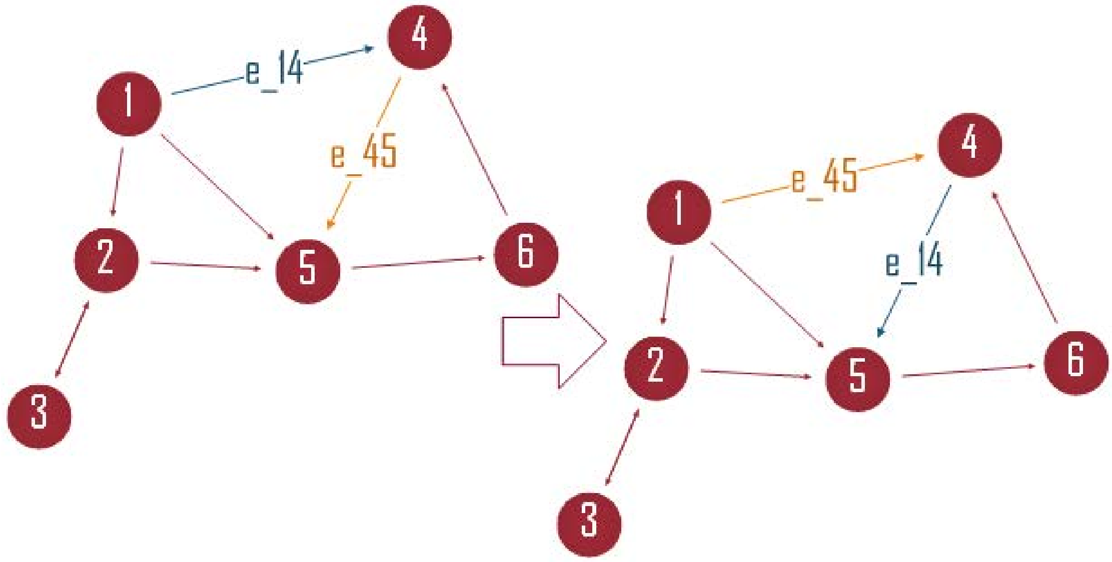


Figure 4. Swapping of Edges in a Graph

The interface complexity β_i represents the complexity of the interconnection between two components at a certain hierarchical level within the system representation. In the graph representation of the system, it can be represented as the weight of the edge between two vertices. Let's consider two edges having weights $w_1(u_1, v_1) = \beta_1$ and $w_2(u_2, v_2) = \beta_2$. Swapping the weights to have $w_1(u_1, v_1) = \beta_2$ and $w_2(u_2, v_2) = \beta_1$ should generally reflect a change in the value of the structural complexity metric.

In this case as well, the metric developed by Sinha does not reflect a change following this swap because of the commutative property of the sum.

Approach for the Development of a New Metric

Requirements for a Structural Complexity Metric

In this research, we are developing a spectral structural complexity metric that can overcome the limitations in other structural complexity measures while maintaining all the good features of the existing ones. The new metric shall be able to

1. Measure the complexity of a system with directed interfaces, in which the adjacency matrix is asymmetric.
2. Measure the complexity of a system with multiple parallel edges, in which two components can be connected via more than one edge.
3. Measure the complexity of a system with respect to its size, meaning that the complexity metric should be normalized with respect to the extension of the system.
4. Pass the component swap test.
5. Pass the interface swap test.

Directed Edges

Any engineered system can be represented through a graph in which the components are vertices and the interfaces are edges. In general, interfaces have a direction, such as for broadcasting communication systems in which one components transmits data to many receivers. Directionality can create asymmetry in the representation of the graph, in case the adjacency matrix is used, with subsequent complex eigenvalues. This approach will use the Laplacian matrix, which is Hermitian, and since we are going to use real values for the matrix, it will be symmetric. Therefore, the use of the Laplacian matrix allows us to have real eigenvalues—more precisely non-negative ones—which can be used for the definition of the metric.

Multiple Parallel Edges

Engineered systems can also have multiple interfaces between components. In a graph representation, this means that the edge (u, v) can have multiple instances, namely, $(u, v)_1, (u, v)_2, \ldots, (u, v)_k$. For example, the interfaces between two components in a cyberphysical system can be thermal, mechanical, electromagnetic, or logical (i.e., in software). In our approach, multiple interfaces are simply bundled together and considered as one. The same approach has also been used already by Sinha, even if without explicit mention (Sinha & de Weck, 2012).

Size Normalization

Since the size of the system influences its complexity, we want to adopt Chaisson's approach and normalize the metric with the size of the system. This can be done by normalizing the graph metric with the number of vertices, or by using normalized matrices such as the normalized Laplacian, which is normalized with the degree of the nodes.

Weighted Edges

The role of the graph in this application is to carry the information about complexity of components and interfaces. For this reason, the edges of the graph need to be weighted according to their complexity. The complexity of the components is represented through weights on self-looping edges.

In the following section, the theory behind the development of a new metric is presented, and a running example is used to illustrate the $\alpha_u = 1$ for all the vertices, and $\beta_{uv} = 1$ for all the edges. While the theory considers the more general case and is not based upon this assumption, its use in the illustrative example allows the reader to more easily focus on the topological contribution to the system complexity.

Spectral Theory of Systems Complexity

Spectral Graph Theory

Spectral Graph Theory is the study of graphs through the eigenvalues of their matrix representation. The set of eigenvalues is known as the spectrum. The elements of spectral graph theory here reported were published by Chung (1997) and Spielman (2007). Let's consider a graph with n vertices and m edges. If u and v are two vertices in the graph, and they are connected by an edge, we say that they are adjacent. An edge that connects a vertex to itself is called a loop. Graphs that contain no loops are called simple graphs. Edges can be associated to a direction. Directed graphs have edges with an associated direction, meaning that the edges (u, v) and (v, u) are two distinct entities. For undirected graphs, those are two representations of the same entity.

Edges in a graph can also be weighted, meaning that we can define a function

$$w(u, v) : V \times V \rightarrow \mathbb{R} .$$

where

$$w(u, v) \geq 0$$

and, in the case of undirected graphs,

$$w(u,v) = w(v,u)$$

The degree of a vertex is defined as the number of incoming edges connected to it, and in the case of weighted edges

$$d_v = \sum_u w(u,v)$$

In this section, we introduce various matrix representations of graphs that will be useful in the creation of a spectral complexity metric. To do this, we will consider the graph represented in Figure 2 as a running example.

Adjacency Matrix

The adjacency matrix is defined as

$$A(u,v) = \begin{cases} 1 & if\ u\ and\ v\ are\ adjacent, \\ 0 & otherwise. \end{cases}$$

The adjacency matrix is symmetric in the case of undirected graphs. For directed graphs, the symmetry holds only if edges appear in pairs. In the case of weighted edges, the adjacency matrix is defined as

$$A(u,v) = \begin{cases} w(u,v) & if\ u\ and\ v\ are\ adjacent, \\ 0 & otherwise. \end{cases}$$

The eigenvalues of the adjacency matrix are labeled in increasing order and represented as

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

and the following is true

$$\sum_{i=1}^{n} \lambda_i = 0$$

$$\sum_{i=1}^{n} \lambda_i^2 = 2m$$

In our example, the adjacency matrix will have the following values in case we consider the edges of the graph as directed or undirected

$$A_{dir} = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \quad \text{and} \quad A_{undir} = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Laplacian Matrix

The Laplacian matrix is defined as $L(u,v) = D(u,v) - A(u,v)$, where $D(u,v)$ is the diagonal matrix of the vertex degrees. This definition is equivalent to

$$L(u,v) = \begin{cases} d_v & if\ u = v, \\ -1 & if\ u\ and\ v\ are\ adjacent, \\ 0 & otherwise. \end{cases}$$

If edges are weighted, its definition is given by

$$L(u,v) = \begin{cases} d_v - w(u,v) & if\ u = v, \\ -w(u,v) & if\ u\ and\ v\ are\ adjacent, \\ 0 & otherwise. \end{cases}$$

In the case of directed graphs (Chung, 2005), the Laplacian matrix is defined as

$$L(u,v) = \Phi - \frac{\Phi P + P^* \Phi}{2}$$

where Φ is the diagonal matrix of the flow of a vertex $\phi(v)$ and P is the transition probability matrix. For a weighted directed graph (Butler, 2007),

$$P(u,v) = \frac{w(u,v)}{d_{out}(u)}$$

The Laplacian matrix is always symmetric, both in the case of directed and undirected graphs. The eigenvalues of the Laplacian matrix are usually labeled in a decreasing order, and are represented as

$$0 = \mu_1 \leq \mu_2 \leq \cdots \leq \mu_n$$

and the following is true:

$$\sum_{i=1}^{n} \mu_i = 2m, \sum_{i=1}^{n} \mu_i^2 = 2m + \sum_{i=1}^{n} d_i^2$$

In our example, the directed and undirected Laplacian matrices assume the following values

$$L_{dir} = \begin{bmatrix} 2.97 & -0.26 & -0.017 & -0.16 & -0.18 & -0.064 \\ -0.26 & 2.97 & -1.20 & -0.041 & -0.32 & -0.034 \\ -0.017 & -1.20 & 0.99 & -0.027 & -0.032 & -0.022 \\ -0.16 & -0.041 & -0.027 & 2.97 & -1.64 & -1.19 \\ -0.18 & -0.32 & -0.032 & -1.64 & 3.97 & -1.38 \\ -0.64 & -0.034 & -0.022 & -1.19 & -1.38 & 1.98 \end{bmatrix}$$

and

$$L_{undir} = \begin{bmatrix} 3 & -1 & 0 & -1 & -1 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 3 & -1 & -1 \\ -1 & -1 & 0 & -1 & 4 & -1 \\ 0 & 0 & 0 & -1 & -1 & 2 \end{bmatrix}$$

Normalized Laplacian Matrix

The normalized Laplacian matrix for undirected graphs is defined as $\mathcal{L} = D^{-1/2}LD^{-1/2}$ which is equivalent to

$$\mathcal{L}(u,v) = \begin{cases} 1 & if\ u = v, \\ -\frac{1}{\sqrt{d_u d_v}} & if\ u\ and\ v\ are\ adjacent, \\ 0 & otherwise. \end{cases}$$

For weighted graphs, the weighted normalized Laplacian matrix is

$$\mathcal{L}(u,v) = \begin{cases} 1 - \frac{w(u,v)}{d_u} & if\ u = v, \\ -\frac{w(u,v)}{\sqrt{d_u d_v}} & if\ u\ and\ v\ are\ adjacent, \\ 0 & otherwise. \end{cases}$$

In the case of directed graphs (Chung, 2005), the normalized Laplacian matrix is defined as

$$\mathcal{L}(u,v) = I - \frac{\Phi^{1/2} P \Phi^{-1/2} + \Phi^{-1/2} P^* \Phi^{1/2}}{2}$$

where *I* is the identity matrix, Φ is the diagonal matrix of the flow of a vertex $\phi(v)$, and *P* is the transition probability matrix. For a weighted directed graph (Butler, 2007),

$$P(u,v) = \frac{w(u,v)}{d_{out}(u)}$$

The normalized Laplacian matrix is always symmetric, and its eigenvalues are represented as

$$0 = \nu_1 \le \nu_2 \le \cdots \le \nu_n$$

In our running example, the values of this matrix for the directed and undirected case are

$$\mathcal{L}_{dir} = \begin{bmatrix} 0.99 & -0.088 & -0.0096 & -0.052 & -0.052 & -0.026 \\ -0.088 & 0.99 & -0.69 & -0.014 & -0.092 & -0.014 \\ -0.0096 & -0.69 & 0.99 & -0.016 & -0.016 & -0.016 \\ -0.052 & -0.014 & -0.016 & 0.99 & -0.47 & -0.49 \\ -0.052 & -0.092 & -0.016 & -0.47 & 0.99 & -0.49 \\ -0.026 & -0.014 & -0.016 & -0.49 & -0.49 & 0.99 \end{bmatrix}$$

and

$$\mathcal{L}_{undir} = \begin{bmatrix} 1 & -0.33 & 0 & -0.33 & -0.29 & 0 \\ -0.33 & 1 & -0.58 & 0 & -0.29 & 0 \\ 0 & -0.58 & 1 & 0 & 0 & 0 \\ -0.33 & 0 & 0 & 1 & -0.29 & -0.41 \\ -0.29 & -0.29 & 0 & -0.28 & 1 & -0.35 \\ 0 & 0 & 0 & -0.41 & -0.35 & 1 \end{bmatrix}$$

Matrix Energy

Graph Energy

Graph energy has been defined by Gutman in 1978 (Gutman, 2001; Gutman & Shao, 2011) as

$$E_A(G) = \sum_{i=1}^{n} |\lambda_i|$$

and it has the following properties

1. $E(G) \geq 0$, where equality is attained only for *m* = 0, meaning that the graph has no edges, and all the vertices are disconnected;
2. the energy of two disconnected graph components $G_1$and G_2 is $E(G) = E(G_1) + E(G_2)$
3. if one component is G_1 and all the other components are isolated vertices, then $E(G) = E(G_1)$.

Laplacian Graph Energy

Gutman also defined the Laplacian energy of a graph (Gutman & Zhou, 2006) as

$$E_L(G) = \sum_{i=1}^{n} |\gamma_i| = \sum_{i=1}^{n} \left|\mu_i - \frac{2m}{n}\right|$$

where γ_i are the auxiliary Laplacian eigenvalues defined as

$$\gamma_i = \mu_i - \frac{2m}{n}$$

Generalized Matrix Energy

A generalization of all these definitions can be given considering a general matrix (Cavers, Fallat, & Kirkland, 2010)

$$E_M(G) = \sum_{i=1}^{n} \left|\lambda_i(M) - \frac{tr(M)}{n}\right|$$

where $tr(M)$ is the trace of the matrix *M*. Thanks to this generalization it is possible to define the normalized Laplacian energy of a graph

$$E_{\mathcal{L}}(G) = \sum_{i=1}^{n} \left|\nu_i - \frac{tr(\mathcal{L})}{n}\right| = \sum_{i=1}^{n} |\nu_i - 1|$$

Spectral Structural Complexity Metrics

The advancements in spectral graph theory presented in the previous section allow us to define a series of complexity metrics based on the spectrum of a certain representation of the system. Let's start with defining the weight function as

$$w(u,v) = \begin{cases} \alpha_u & if\ u = v \\ \beta_{u,v} & otherwise \end{cases}$$

where α_u represents the complexity of the component u, and $\beta_{u,v}$ the complexity of the interface between components u and v. This function allows us to use the definitions for the weighted adjacency, Laplacian, and normalized Laplacian matrices, for both the case of directed and undirected graphs. The structural complexity evaluated using the adjacency matrix is defined as

$$C_A = \frac{E_A(G)}{n}$$

where the adjacency matrix considers the weights of the edges, and n is the number of vertices of the graph. In the case of unweighted edges, this metric is equivalent to the C_3 component of the one defined by Sinha (Sinha & de Weck, 2012). The adjacency matrix is historically the most used in systems engineering (as DSM) and in spectral graph theory. In recent years, there has been a shift in spectral graph theory, given by the interesting properties of the Laplacian eigenvalues. The second smallest eigenvalue is particularly interesting, since it represents the connectivity of the graph. Also, the multiplicity of zero in the Laplacian spectrum represents the number of connected components, used in the metric proposed by McCabe. For these reasons, we are defining the structural complexity evaluated using the Laplacian matrix as

$$C_L = \frac{E_L(G)}{n}$$

where the Laplacian matrix considers the weights of the edges, and *n* is the number of vertices of the graph. This type of normalization has an alternative, which is to normalize using the degree matrix of the system. This alternative approach brings to the normalized Laplacian matrix. The structural complexity evaluated using the normalized Laplacian matrix is defined as

$$C_{\mathcal{L}} = E_{\mathcal{L}}(G)$$

where the normalized Laplacian matrix is defined considering the weights of the edges.

In Figure 5 we report the values of these metrics for the directed and undirected case of our running example.

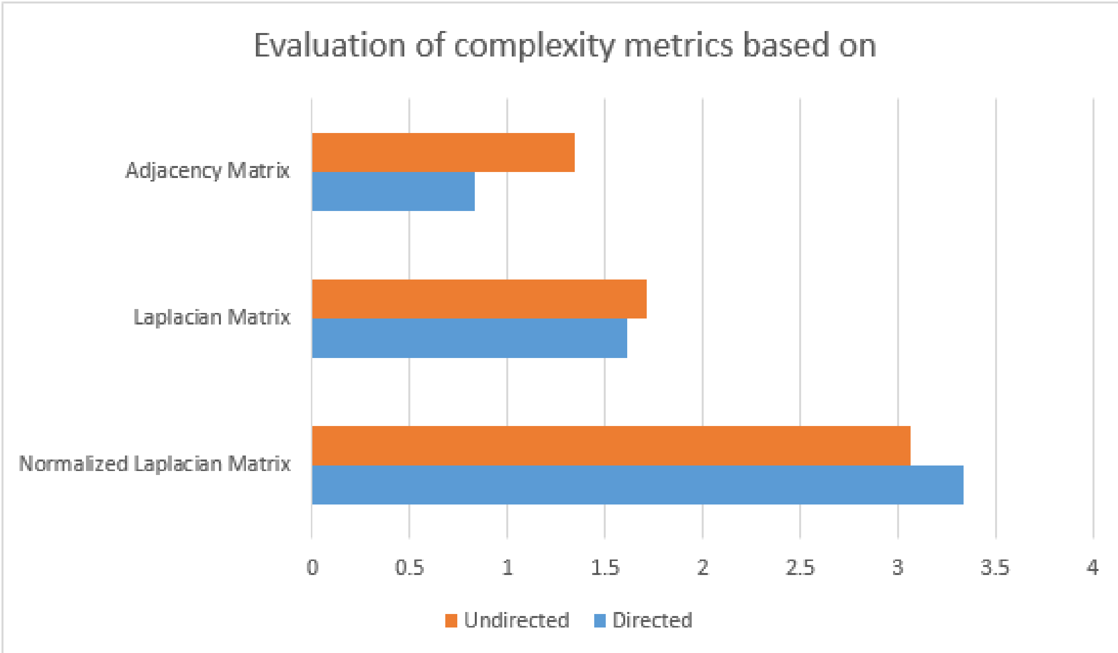


Figure 5. Evaluation of Complexity Metrics Based on the Matrix Energy of the Adjacency Matrix, Laplacian Matrix, and Normalized Laplacian Matrix, for the Directed and Undirected Graph Represented in Figure 2

Conclusion and Future Work

In this paper, we presented an alternative to existing structural complexity metrics. The purpose of the metrics is to measure the structural complexity of the system, considering the contributions of the size, the connectivity, and the topology of the system. At this stage of the research, the focus has been shifted on the topological contribution, with the plan of addressing size and connectivity in a later stage.

These metrics will subsequently be applied to real world systems, with the goal of verifying their applicability and understanding their differences in terms of features and limitations. The results will then be compared to the ones from other complexity metrics, with the goal of validating the new metrics and clarifying their possible shortcomings.

In the context of the systems engineering practice, these metrics represent a continuation of the widespread effort to introduce quantitative tools to increase objectivity of measurements. The spectral approach developed by Sinha is the starting point of possibly a series of research efforts that will gradually introduce new metrics trying to patch limitations in the existing ones. The long-term expectation is for practitioners to converge on the use of a low number of metrics that will be applied depending on the specific case.

References

Abbott, R. (2006). Emergence explained: Abstractions: Getting epiphenomena to do real work. *Complexity, 12*, 13–26.

Bedau, M. A. (1997). Weak emergence. *Noûs, 31,* 375–399.

Butler, S. (2007). Interlacing for weighted graphs using the normalized Laplacian. Electronic *Journal of Linear Algebra, 16*, 87.

Cavers, M., Fallat, S., & Kirkland, S. (2010). On the normalized Laplacian energy and general Randić index R-1 of graphs. *Linear Algebra and Its Applications, 433,* 172–190.

Chaisson, E. J. (2004). Complexity: An energetics agenda. *Complexity, 9,* 14–21.

Chaisson, E. J. (2011). Energy rate density as a complexity metric and evolutionary driver. *Complexity, 16*, 27–40.

Chaisson, E. J. (2014). The natural science underlying big history. *Scientific World Journal*, *2014*.

Chaisson, E. J. (2015). Energy flows in low-entropy complex systems. *Entropy, 17,* 8007–8018.

Chalmers, D. J. (2008). Strong and weak emergence. In *The re-emergence of emergence.* Oxford University Press.

Checkland, P. (1981). *Systems thinking, systems practice*.

Chung, F. (2005). Laplacians and the Cheeger inequality for directed graphs. Annals of *Combinatorics, 9*, 1–19.

Chung, F. R. (1997). *Spectral graph theory* (Vol. 92). American Mathematical Society.

Crawley, E., Cameron, B., & Selva, D. (2015). *System architecture: Strategy and product development for complex systems.* Pearson.

Fischi, J., Nilchiani, R., & Wade, J. (2015). Dynamic complexity measures for use in complexity-based system design. *IEEE Systems Journal*.

Gell-Mann, M. (1995). What is complexity? Remarks on simplicity and complexity by the Nobel Prize-winning author of *The Quark and the Jaguar. Complexity, 1*, 16–19.

Gutman, I. (2001). The energy of a graph: Old and new results. In *Algebraic combinatorics and applications* (pp. 196–211). Springer.

Gutman, I., & Shao, J.-Y. (2011). The energy change of weighted graphs. *Linear Algebra and Its Applications, 435*, 2425–2431.

Gutman, I., & Zhou, B. (2006). Laplacian energy of a graph. *Linear Algebra and Its Applications, 414,* 29–37.

Kauffman, S. (2007). Beyond reductionism: Reinventing the sacred. *Zygon, 42*, 903–914.

Longo, G., Montévil, M., & Kauffman, S. (2012). No entailing laws, but enablement in the evolution of the biosphere. *Proceedings of the 14th Annual Conference Companion on Genetic and Evolutionary Computation* (pp. 1379–1392).

MacCormack, A., Rusnak, J., & Baldwin, C. Y. (2006). Exploring the structure of complex software designs: An empirical study of open source and proprietary code. *Management Science, 52,* 1015–1030.

McCabe, T. J. (1976). A complexity measure. *IEEE Transactions on Software Engineering*, 308–320.

McCabe, T. J., & Butler, C. W. (1989). Design complexity measurement and testing*.* *Communications of the ACM, 32,* 1415–1425.

Nikiforov, V. (2007). The energy of graphs and matrices. *Journal of Mathematical Analysis and Applications, 326,* 1472–1475.

Page, S. E. (1999). Computational models from A to Z. *Complexity, 5,* 35–41.

Rouse, W. B. (2016). *Complexity: Absolute or relative?*

Sheard, S. A., & Mostashari, A. (2010). A complexity typology for systems engineering. *Twentieth Annual International Symposium of the International Council on Systems Engineering.*

Sinha, K., & de Weck, O. (2012). Structural complexity metric for engineered complex systems and its application. Gain Competitive Advantage by Managing Complexity: *Proceedings of the 14th International DSM Conference Kyoto, Japan* (pp. 181–194).

Sinha, K., & de Weck, O. (2013). A network-based structural complexity metric for engineered complex systems. *2013 IEEE International Systems Conference* (pp. 426–430).

Snowden, D. (2005). Multi-ontology sense making: A new simplicity in decision making. *Journal of Innovation in Health Informatics, 13*, 45–53.

Snowden, D. J., & Boone, M. E. (2007). A leaders' framework for decision making. *Harvard Business Review, 85*, 68.

Spielman, D. A. (2007). Spectral graph theory and its applications. *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS '07)* (pp. 29–38).

Wade, J., & Heydari, B. (2014). Complexity: Definition and reduction techniques. *Proceedings of the Poster Workshop at the 2014 Complex Systems Design & Management International Conference* (pp. 213–226).

Weaver, W. (1948). Science and complexity. *American Scientist, 36*, 536–544.

THIS PAGE INTENTIONALLY LEFT BLANK

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL
555 DYER ROAD, INGERSOLL HALL
MONTEREY, CA 93943

www.acquisitionresearch.net