IDA | RESEARCH NOTES

WELCH AWARD 2019

Fall 2019

IDA operates three Federally Funded Research and Development Centers (FFRDCs) that answer the most challenging U.S. security and science policy questions with objective analysis leveraging extraordinary scientific, technical, and analytic expertise.

The articles in this issue of *IDA Research Notes* were written by researchers affiliated with divisions of IDA's Systems and Analyses Center. Their directors, named below, would be glad to respond to questions about the specific research topics discussed in these articles as well as any other topics related to their work.

Cost Analysis and Research Division (CARD)
Dr. David E. Hunter
703.575.4686, dhunter@ida.org

Information Technology and Systems Division (ITSD)
Dr. Margaret E. Myers
703.578.2782, mmyers@ida.org

Joint Advanced Warfighting Division (JAWD)
Dr. Daniel Y. Chiu
703.845.2439, dchiu@ida.org

Operational Evaluation Division (OED)
Mr. Robert R. Soule
703.845.2452, rsoule@ida.org

Science and Technology Division (STD)
Dr. Leonard J. Buckley
703.578.2800, lbuckley@ida.org

Strategy, Forces and Resources Division (SFRD)
ADM John C. Harvey, U.S. Navy (retired)
703.575.4350, jharvey@ida.org

System Evaluation Division (SED)
Dr. Stephen M. Ouellette
703.845.2443, souellet@ida.org

IDA

4850 Mark Center Drive
Alexandria, VA 22311-1882
703.845.2000
ida.org

IDA | Welch Award 2019

The Larry D. Welch Award is named in honor of former IDA president and U.S. Air Force Chief of Staff, General Larry D. Welch (retired). The award recognizes IDA researchers who exemplify General Welch's high standards of analytic excellence through their external publication in peer-reviewed journals or other professional publications, including books and monographs.

This issue of *IDA Research Notes* is dedicated to the nominees of the 2019 Larry D. Welch Award for best external publication. The articles in this issue are summaries derived from the best of the publications that were nominated in 2019.

This year's winner and finalists are named below, along with a link where available.[1] Authors whose names appear in bold type have current or former affiliations with IDA as researchers or consultants.

Winner

The paper recognized this year as the best example of high-quality, relevant research published in the open literature is "Standardized Down-Looking Ground-Penetrating Radar (DLGPR) Data Collections," by Science and Technology Division (STD) researchers **Erik M. Rosen** and **Phillip T. Koehn**, with coauthor and former colleague, **Marie E. Talbott**. Their paper was published in *Proceedings of SPIE, The International Society for Optics and Photonics, Vol. 10628, SPIE Defense + Security, 2018*.

Finalists

"A Central Limit Theorem for Correlated Variables with Limited Normal or Gamma Distributions," by System Evaluation Division (SED) researcher **Dennis F. DeRiggi** was published in *Communications in Statistics—Theory and Methods*, December 2018, https://doi.org/10.1080/03610926.2018.1536212.

"Complexity in an Unexpected Place: Quantities in Selected Acquisition Reports," by Cost Analysis and Research Division (CARD) researchers **Gregory A. Davis**, **Margaret L. Giles**, and **David M. Tate**, was published in *Proceedings of the 15th Annual Acquisition Research Symposium, Naval Postgraduate School, Volume I, Acquisition Research: Creating Synergy for Informed Change*, April 30, 2018, https://calhoun.nps.edu/handle/10945/58801.

"Extending CryptDB to Operate an ERP System on Encrypted Data," by Information Technology and Systems Division (ITSD) researchers **Kevin E. Foltz** and **William R. (Randy) Simpson**, was published in *Proceedings of the 20th International Conference on Enterprise Information Systems (ICEIS 2018), Volume 1*, March 2018, http://www.scitepress.org/PublicationsDetail.aspx?ID=TYtRRYBqMV8=&t=1.

[1] IDA assumes no responsibility for the persistence of URLs for external and third-party internet websites referred to in this publication. Further, IDA does not guarantee the accuracy or appropriateness of these websites' content now or in the future.

"Methods in Macroeconomic Forecasting Uncertainty Analysis: An Assessment of the 2015 National Defense Stockpile Requirements Report," by Strategy, Forces and Resources Division (SFRD) researchers **Wallice Y. Ao** and **Eleanor L. Schwartz**, and their former colleagues **Justin M. Lloyd** and **Amrit Romana**, was published in *Mineral Economics, Raw Materials Report* 31, no. 3, October 2018, https://link.springer.com/article/10.1007%2Fs13563-017-0127-6.

"Multidomain Battle: Time for a Campaign of Joint Experimentation," by Joint Advanced Warfighting Division (JAWD) researchers **Kevin M. Woods** and **Thomas C. Greenwood**, was published in *Joint Forces Quarterly (JFQ 88)*, 1st Quarter 2018, https://ndupress.ndu.edu/Publications/Article/1411615/multidomain-battle-time-for-a-campaign-of-joint-experimentation/.

"On Scoping a Test that Addresses the Wrong Objective," by Operational Evaluation Division (OED) researchers **Thomas H. Johnson** and **Rebecca M. Medlin**, their former colleague **Laura J. Freeman**, and OED consultant **James R. Simpson**, was published in *Quality Engineering*, November 2018, available from https://doi.org/10.1080/08982112.2018.1479035.

Noteworthy

The Welch Award Selection Committee named three other nominated publications as being worthy of note given their success in the open literature and the quality of research they reflect. Authors whose names appear in bold type have current or former affiliations with IDA as researchers, members of division management, or consultants.

"Have Changes in Acquisition Policy Influenced Cost Growth of Major Defense Acquisition Programs?" by CARD researcher **David L. McNicol**, was published in *Proceedings of the 15th Annual Acquisition Research Symposium, Naval Postgraduate School, Volume I, Acquisition Research: Creating Synergy for Informed Change*, April 30, 2018, http://hdl.handle.net/10945/58731.

"'It's Either a Panda or a Gibbon': AI Winters and the Limits of Deep Learning," by JAWD researcher **Robert F. Richbourg**, was published in *War on the Rocks Blog*, May 10, 2018, https://warontherocks.com/2018/05/its-either-a-panda-or-a-gibbon-ai-winters-and-the-limits-of-deep-learning/.

"A Monte Carlo Tradeoff Analysis to Guide Resource Investment in Threat Detection Systems: From Forensic to Prospective Investigations," by STD researchers **Shelley M. Cazares**, **Jeffrey A. Snyder**, **Joan F. Cartier**, and **Felicia D. Sallis-Peterson**, and former colleague **John M. Fregeau**, was published in *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, 1–24, March 17, 2017, https://doi.org/10.1177/1548512917694966.

Standardized Data Collection Helps Refine Algorithms for Detecting Buried Targets[1]

Erik M. Rosen, Phillip T. Koehn, and Marie Talbott

Down-looking ground-penetrating radar is used extensively for buried mine and improvised explosive device (IED) detection. Comparing detection performance across different test sites and soil compositions is challenging given that targets vary in size, composition, and burial depth. A joint effort between the United States, Australia, and Canada uniformly collected data from various test sites in Australia and Canada using a standard set of target types, layouts, and depths. The primary objective of the effort was to provide diverse data for use in algorithm development.

Introduction

Detection performance of down-looking ground penetrating radar (DLGPR) depends on the type and condition of the soil in which targets are buried at the time of data collection (Rhebergen et al. 2004). When soil conditions are similar, the characteristics of the targets themselves matter most when it comes to DLGPR performance—with size, composition, and burial depth being the primary variables. Wilson et al. (2007) showed that supervised learning algorithms improve detection performance when applied to DLGPR data. To explore the potential of adaptive algorithm approaches to detection, the U.S. Army Night Vision and Electronic Sensors Directorate, Australian Defence Science and Technology Group, and Defence Research and Development Canada collected suitable data on a standardized set of targets. The goal of this effort was to have uniformly collected data sets for use in improving algorithms for automatic detection of buried targets and to capture performance in a range of soil types. Preliminary analysis reveals that the data also has potential for alerting an operator about whether the environment in which the DLGPR system is operating is favorable or unfavorable to detection.

We explored the concept of a discrete meter that might be used by an operator to determine if detection conditions are degraded or favorable.

Data Collection

The standardized target set consists of relevant threats that are difficult to detect. In each data collection lane, there are three instances of each target class, type, and burial depth. In addition to a standardized target layout, the construction of all target classes and types are identical across lanes and sites to ensure there are not subtle variations among targets of the same type.

Data collection in Australia was executed in fall 2016 at four sites with differing soil properties and different degrees of surface vegetation and roughness. In Canada, data collection took place at two sites located about 30 minutes apart in New Brunswick. The first collection was in fall 2017 in a temperate environment, and the second, in winter 2018 in frozen ground and fresh snow conditions. This article focuses on detection results from the Australian data.

---

[1] Based on "Standardized Down-Looking Ground-Penetrating Radar (DLGPR) Data Collections," *Proceedings of SPIE, The International Society for Optics and Photonics*, Vol. 10628, SPIE Defense + Security, 2018.

The data collection platform was a small four-wheeled off-road vehicle that was modified to accept a DLGPR payload and a differential global positioning system so that all ground-penetrating radar (GPR) sensor data would be tagged with accurate coordinates. The number of passes the GPR took through each data collection site—a pass being travel in one direction on a given lane—ranged from 16 to 40.

Detection Results

Receiver Operator Characteristic Curves

Top-level detection performance results are given in the form of receiver operator characteristic (ROC) curves by site and lane. ROC curves are formed by first matching system declarations (also known as alarms) to surveyed target locations. The ROC curves are mapped from low probability of detection (PD) to high PD by rank ordering the alarms by magnitude of the primary decision statistic, and then continuously lowering the decision threshold such that more and more alarms are considered. The maximum PD and maximum false alarm rate (FAR) occur when all the alarms are used. When the PD does not reach 1.0, it means the sensor and algorithm did not generate an alarm near some percentage of the targets.

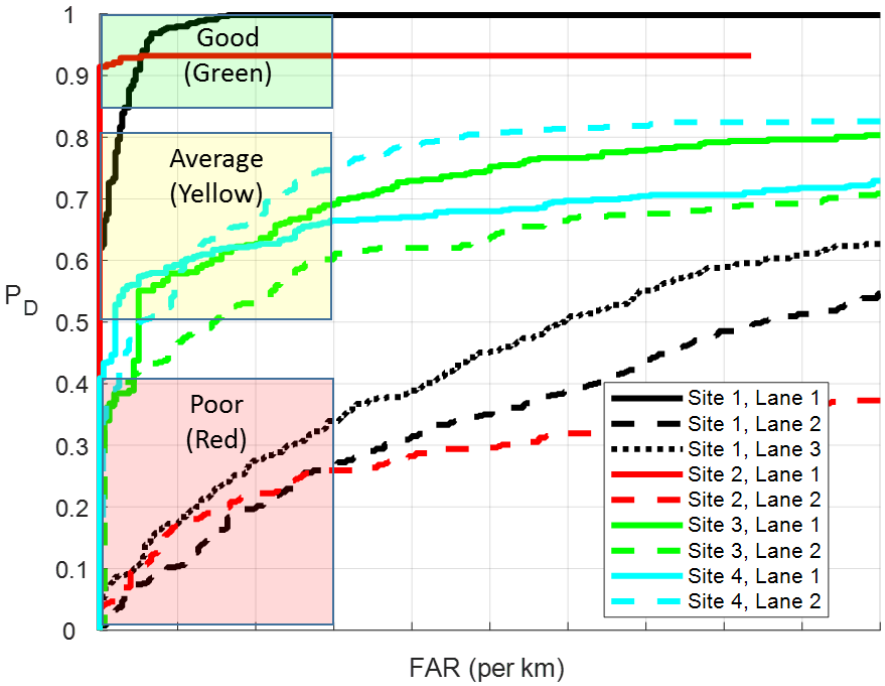


Figure 1. Detection performance ROC curves by Australian site and lane data

Figure 1 shows ROC curves for each lane at each site for the data collected in Australia. The x-axis FAR values are intentionally removed since the interest is primarily to study performance variability across the different sites and lanes. Note the significant variability in detection performance, where we have divided the lanes into three regimes—good (green), average (yellow), and poor (red). In some cases, the two extremes were experienced at the same site. Results were good in lane 1 at site 1 (black solid line), while they were poor in lane 3 at site 1 (dotted black line). At site 2, results were good in lane 1 (solid red line), but poor in lane 2 (dashed red line). The right edge of the colored boxes intersects and separates the ROC curves into Good performance

(PD about 0.95 to 1.0), Average performance (PD between 0.60 and 0.75), and Poor performance (PD between 0.25 and 0.35).

Performance can be separated by target type and depth, and even by particular instance of a target at a fixed location in one of the lanes. During data collection, the DLGPR system traversed each of the lanes several times to permit study of the variance in sensor and algorithm responses for each target type, depth, and instance.

Confidence Value as Scatter Boxplots

We examined the confidence value of alarms to understand drivers of detection performance in the lanes of interest. The confidence value of an alarm is the magnitude of the primary decision statistic. The ideal decision statistic would assign higher confidence values to targets and lower confidence values to false alarms, so that a threshold could be set beyond which all alarms would be target detections. Poor detection performance could be due to target detections being assigned a low confidence value, false alarms being assigned a high confidence value, or both. The larger the separation of confidence value distributions for targets and false alarms, the better the detection performance.

Figure 2 shows scatter boxplots (on the right) of the confidence values for all target classes and false alarms that correspond to ROC curves (on the left) for each target class by type in lane 1 at site 1 and lane 2 at site 2. A green vertical line indicates the

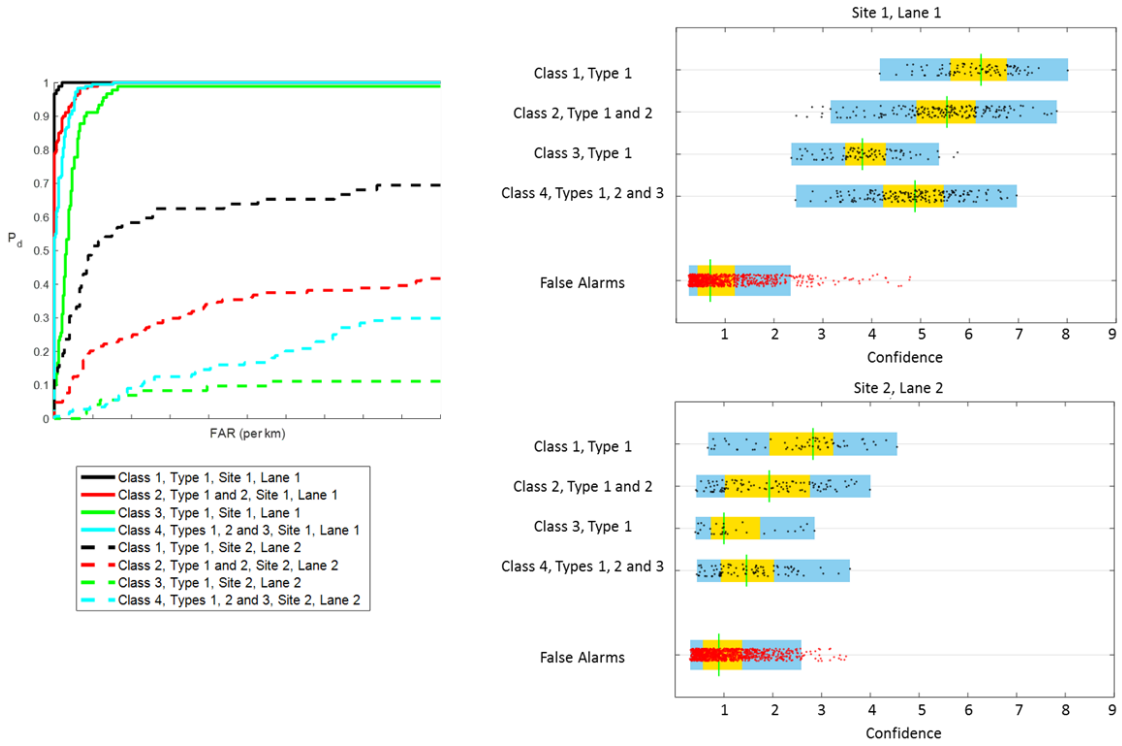


Figure 2. ROC curve (left) and scatter boxplot of confidence values (right) for target classes and false alarms in lane 1 at site 1 and lane 2 at site 2 in Australia

median, while the left and right edges of the yellow box indicate the 25th and 75th percentiles, respectively. The blue box encompasses the remainder of the data, with outliers falling outside the blue box. Each individual target detection is a black dot in the scatter box plot and each individual false alarm is a red dot in the scatter boxplot. We see that the poor detection performance in lane 2 at site 2 is not due to high confidence value alarms, but is due to low confidence target detections.

Probability of Detection and False Alarm Rate as a Function of Confidence Value

In addition to plotting confidence values, Figure 2 also shows that PD and FAR are a function of confidence value. Figure 3 shows a plot of the relationship between confidence value, PD, and FAR. This plot has three axes: the left y-axis is PD, the right y-axis is FAR, and the x-axis is confidence value. We have intentionally removed the numerals on the FAR axis since their specific values are not relevant to the discussion. The thick lines correspond to plots of PD versus confidence value, and the thin line corresponds to FAR versus confidence value. We included blue and green guidelines as an example of how to read the plot.

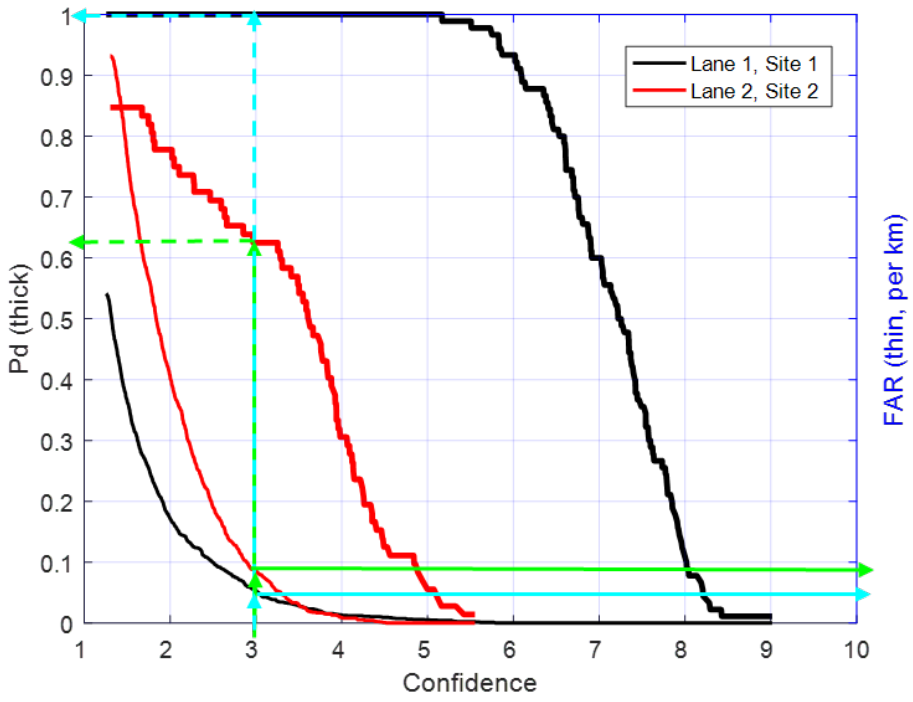


Figure 3. PD and FAR as a function of confidence value for the same target in lane 1 at site 1 and lane 2 at site 2 in Australia

To determine the FAR in lane 1 at site 1 at the confidence value of 3, begin at 3 on the x-axis and follow the blue line until it intersects the thin black curve. Map this intersection to the right y-axis, shown by the solid blue line, to find the FAR per kilometer (km) for lane 1, site 1 at the confidence value of 3. To determine the PD in lane 1, follow the dashed blue line from 3 on the x-axis until it intersects the thick black curve. Then map this intersection to the left y-axis to determine that the PD is 1 in lane 1, site 1 at a confidence value of 3. Replicate this process to find the FAR (solid green line) and PD (dashed green line) at a confidence value of 3 in lane 2 at site 2. In our example, the PD is just over 0.6 for lane 2, site 2 at a confidence value of 3, and the FAR in lane 2 at site 2 is about twice what it is in lane 1 at site 1.

For optimal detection performance, the FAR and PD lines should have as much separation as possible in the x-axis, which would indicate that the confidence values assigned to false alarms are lower than the confidence values assigned to targets. This behavior is shown in Figure 3 for lane 1 at site 1. If a user wanted to choose a confidence value to set as a threshold for system operation, 4.5 would be an ideal choice for lane 1 at site 1, since all targets would be detected with no false alarms.

B-Scans

To evaluate various approaches to target detection, we studied the DLGPR sensor data in its fundamental form, adopting the terminology used in Daniels's book on GPR (Daniels 2004). The fundamental response of any DLGPR system is the A-scan, which is the radar response as a function of time, where time of response corresponds to depth in the ground, and is associated with a particular down-track location and a particular across-track channel.

Examining the DLGPR sensor data in the form of B-scans, in which the y-axis is time/depth and the x-axis is down-track scan or across-track channel, is insightful for understanding detection performance. Figure 4 shows down-track B-scan examples of the same target at various depths for lane 1 at site 1 and lane 2 at site 2, with the target centered in the B-scan plot. Confidence values are listed at the top of each plot.

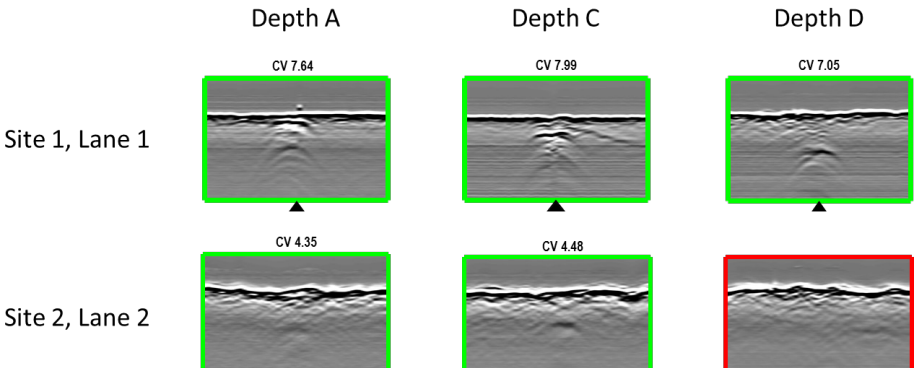


Figure 4. B-scan examples of the same target in lane 1 at site 1 and lane 2 at site 2 in Australia

The GPR response to the ground is the nearly horizontal white/black line in the upper part of the B-scan. The target responses in the B-scans are the parabolas of varying intensity shown under the ground. The depths correspond to the burial depth to the top of the target. A green box indicates the target was detected, and a red box indicates that the target was not detected. In lane 1 at site 1, all examples of the target are detected and the GPR response to the target is visible at all depths. As the burial depth increases from A to C to D, the GPR response to the target appears deeper in the B-scan. It is interesting to note that at depth A the top of the target response blends into the GPR response to the ground.

The target is detected at depth A and depth C in lane 2 at site 2, but not at depth D. The target response is faint in the B-scan for depth A, more pronounced in the B-scan for depth C and not visible in the B-scan for depth D. The scans show more clutter at the ground location and just below the ground in lane 2 at site 2. This subsurface

clutter is potentially why the GPR response to the target is fainter in the B-scan at depth A than at depth C, and why the confidence value of the target detection is slightly higher at depth C than at depth A. The target responses in the B-scans for lane 1 at site 1 are more crisp and sharp than those for the same target in lane 2 at site 2—another observation from the B-scans that could affect detection performance.

Correlation of Performance with Data Characteristics

Thus far, we have used the collected data to show that detection performance over the same set of targets varied significantly from site to site and lane to lane. But the data can be leveraged in ways that actually improve performance. One fairly straightforward approach would be to add the new data to sets of old data collected at other test sites, retrain features and classifiers using all the data, and arrive at an algorithm that works best using all the diverse data combined. This robust algorithm might be the best of all algorithms for all the data available, but it may underperform compared with other algorithms when data is restricted to a particular site or lane or soil type. Thus, the choice is either developing a one-size-fits-all algorithm or adopting a several-algorithm solution in which a particular algorithm is essentially tuned to specific soil/terrain types and conditions.

The first step in determining if a several-algorithm approach has the potential for success was finding correlations between GPR data metrics and performance. We examined data from one of the good lanes and one of the poor lanes to see if particular characteristics of the GPR data gave rise to either good or poor performance.

If for every A-scan we compute standard deviations over different time/depth regimes, we create multiple types of C-scans where for every scan and channel we have a positive scalar value. Figure 5 shows C-scans for 50-meter samples of data taken from

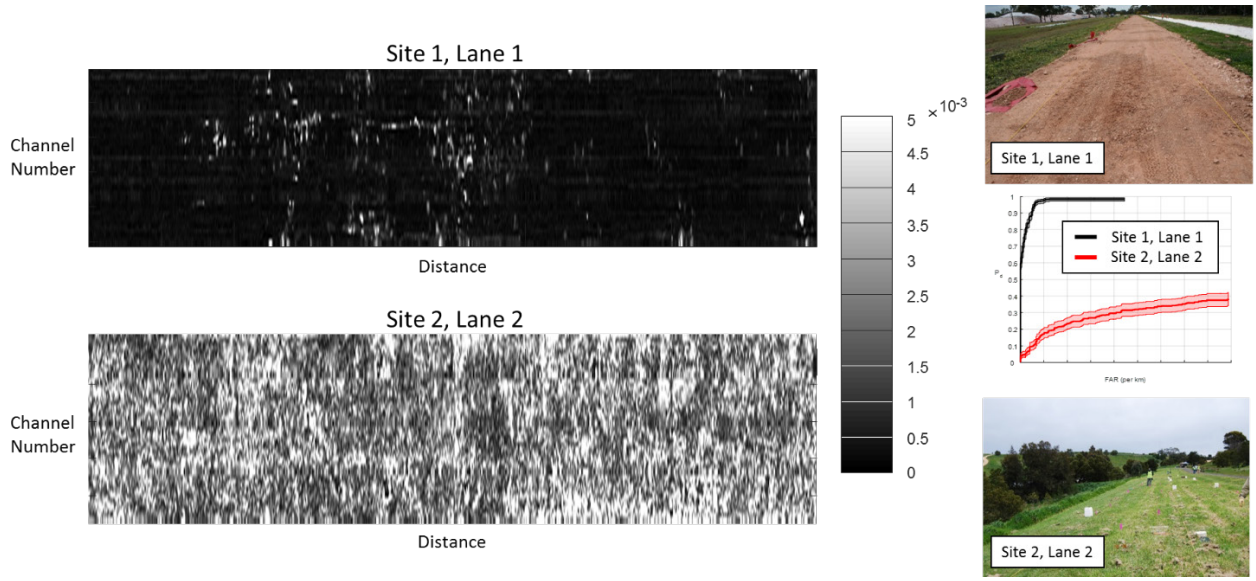


Figure 5. C-scans of the in-air standard deviations for two lanes resulting in the extremes of detection performance

site 1, lane 1 and from site 2, lane 2 for the portion of the GPR response prior to the ground bounce peak (the in-air response). The elevated levels of clutter from site 2, lane 2 are obvious. It is likely that scattering from the grass is the cause of the elevated clutter in the in-air response on the grass lane at site 2. And it is possible that the elevated shallow subsurface clutter in the grass lane is caused by root structures that are not present in the dirt lane at site 1. But correlation is not necessarily causation here, so our next step is to determine if in-air noise and PD are related.

To predict what the maximum PD would be in a given lane, we used a linear least squares fit to a plot of the mean in-air responses on a given lane against the maximum PD for that lane. Figure 6 shows a comparison of the actual maximum PD to the predicted maximum PD. The results indicate that the PD can be predicted within 10 percent or less by computing the standard deviation of the A-scan response prior to the ground bounce, for this set of data.

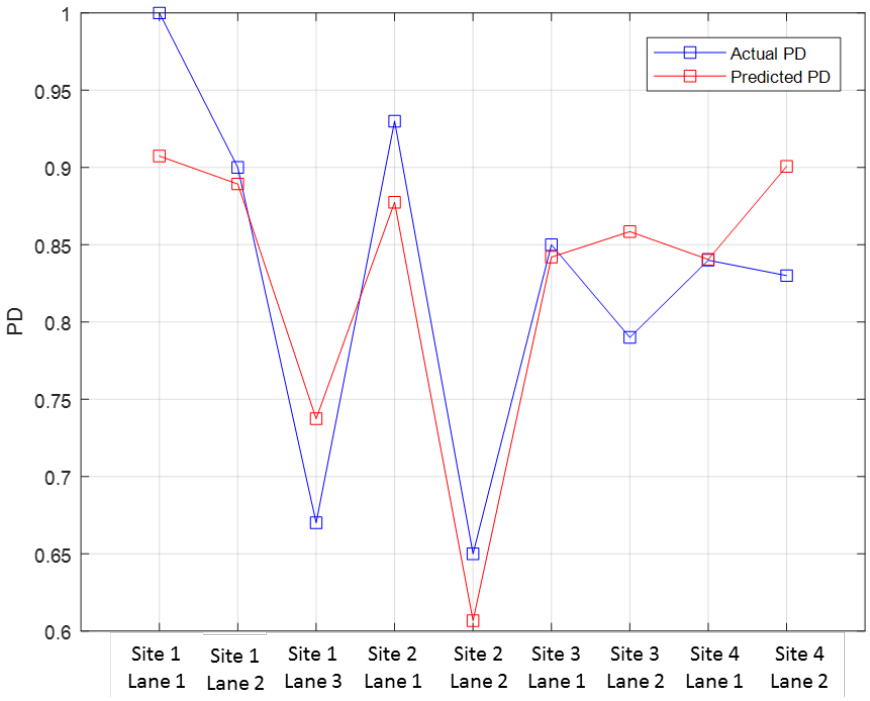


Figure 6. Actual and predicted maximum PD for in-air noise

We explored the concept of a discrete meter that might be used by an operator to determine if detection conditions are degraded or favorable. We chose thresholds below and above in which we color each C-scan pixel green for low in-air noise, red for high in-air noise, and yellow for in-air noise level between. It is not clear how well these thresholds will translate to the remainder of the data, but we simply demonstrate the possibilities here. Figure 7 shows what an operator might see scrolling by as the system is driven down a roadway, where green suggests favorable detection conditions and red warns of degraded conditions.

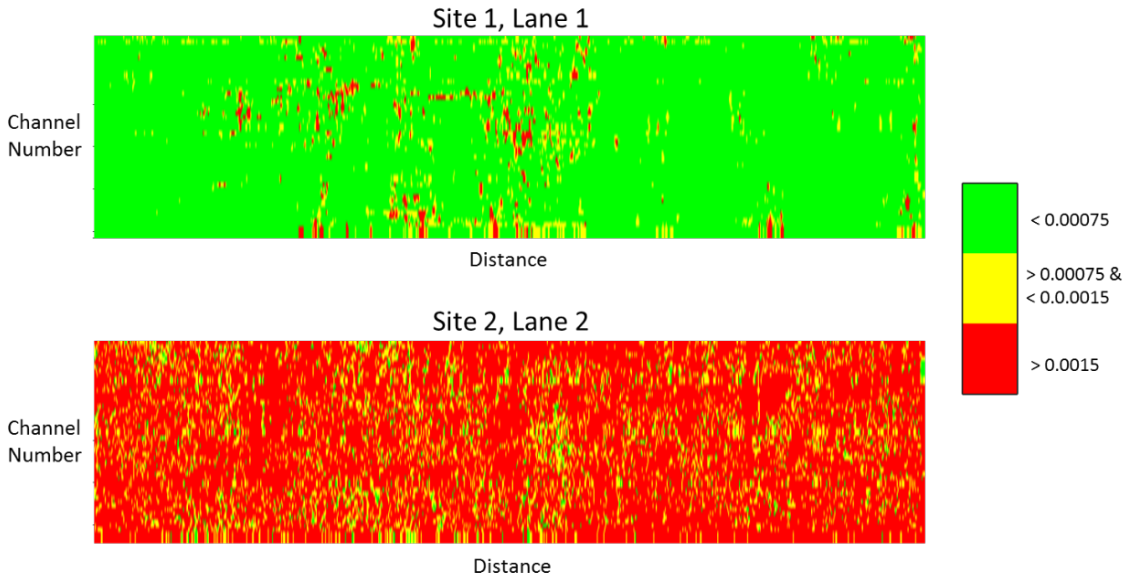


Figure 7. Mapping of in-air noise to discrete green/yellow/red meter for operator notification of potential degraded detection

Future Work

The ultimate objective of a DLGPR detection system/algorithm is to be so highly adaptable that it essentially senses its environment in real-time and adjusts thresholds and parameters in such a way that detection performance is optimized. There have been efforts in context-dependent algorithm development using DLGPR data (Ratto, Torrione, and Collins 2009), but no algorithm was ever adopted due in part to the limitations of the available data sets. The more diverse data collected for this effort may renew context dependent algorithm approaches.

References

Daniels, D. J. (ed.) 2004. *Ground Penetrating Radar.* 2nd Edition, London, England: The Institution of Engineering and Technology.

Ratto, C. R., P. A Torrione, and L. M. Collins. 2009. "Context-Dependent Feature Selection for Landmine Detection with Ground-Penetrating Radar." *Proceedings of SPIE* 7303: 730327.

Rhebergen, J. B., H. A. Lensen, R. Wijk, J. M. H. Hendrickx, R. van Dam, and B. Borchers. 2004. "Prediction of Soil Effects on GPR Signatures." *Proceedings of SPIE* 5415: 705–715.

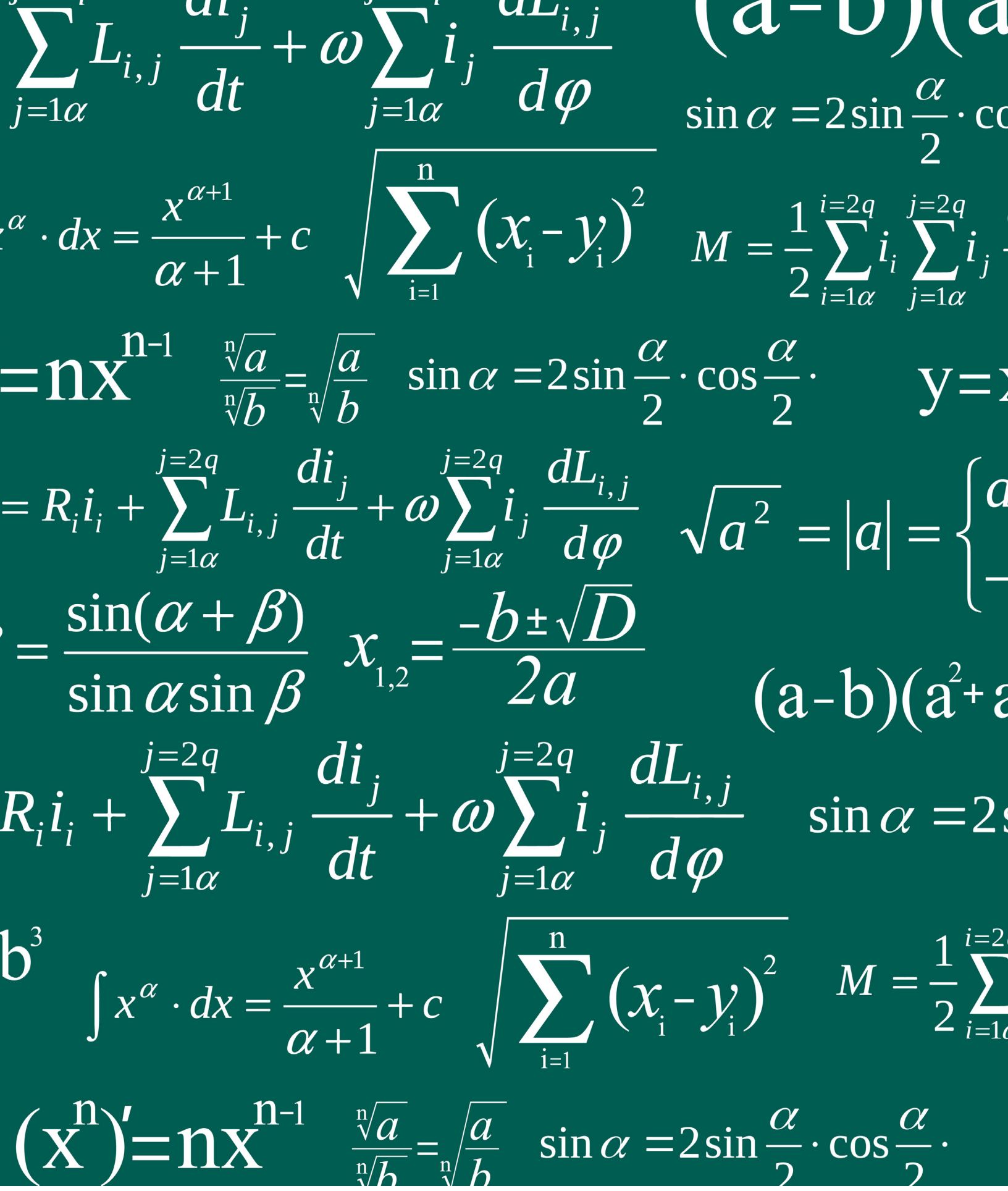
Wilson, J. N., P. Gader, W. Lee, H. Frigui, and K. C. Ho. 2007. "A Large-Scale Systematic Evaluation of Algorithms Using Ground-Penetrating Radar for Landmine Detection and Discrimination." *IEEE Transactions on Geoscience and Remote Sensing* 45, no. 8: 2560–2572.

About the Authors

Phil Koehn is a research staff member in the Science and Technology Division (STD) of IDA's Systems and Analyses Center. He received his master's and doctoral degrees in physics from the University of Rochester. He joined IDA in 2003. Phil is being recognized for the first time as a participant in the annual Welch Award competition.

Erik Rosen is also a member of the research staff in STD. He holds a bachelor's degree in physics from Virginia Tech and a master's degree in applied and engineering physics from George Mason University. He joined IDA in 1997 and served as an assistant director in STD in 2011–12. This is Erik's first recognition in the annual Welch Award competition.

Marie Talbott was a research associate in STD from January 2012 until September 2018. She holds a master's degree in electrical engineering from George Mason University. Her first association with IDA was as a summer associate in 2011. This is Marie's first time being honored as a coauthor of a Welch Award–winning publication.


$\sum_{j=1\alpha} L_{i,j} \frac{di_j}{dt} + \omega \sum_{j=1\alpha} i_j \frac{dL_{i,j}}{d\varphi}$
$\sin\alpha = 2\sin\frac{\alpha}{2} \cdot \cos$
$\int x^{\alpha} \cdot dx = \frac{x^{\alpha+1}}{\alpha+1} + c$
$\sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$
$M = \frac{1}{2} \sum_{i=1\alpha}^{i=2q} i_i \sum_{j=1\alpha}^{j=2q} i_j$
$= nx^{n-1}$
$\frac{\sqrt[n]{a}}{\sqrt[n]{b}} = \sqrt[n]{\frac{a}{b}}$
$\sin\alpha = 2\sin\frac{\alpha}{2} \cdot \cos\frac{\alpha}{2} \cdot$
$y =$
$= R_i i_i + \sum_{j=1\alpha}^{j=2q} L_{i,j} \frac{di_j}{dt} + \omega \sum_{j=1\alpha}^{j=2q} i_j \frac{dL_{i,j}}{d\varphi}$
$\sqrt{a^2} = |a| =$
$= \frac{\sin(\alpha+\beta)}{\sin\alpha \sin\beta}$
$x_{1,2} = \frac{-b \pm \sqrt{D}}{2a}$
$(a-b)(a^2+$
$R_i i_i + \sum_{j=1\alpha}^{j=2q} L_{i,j} \frac{di_j}{dt} + \omega \sum_{j=1\alpha}^{j=2q} i_j \frac{dL_{i,j}}{d\varphi}$
$\sin\alpha = 2$
b^3
$\int x^{\alpha} \cdot dx = \frac{x^{\alpha+1}}{\alpha+1} + c$
$\sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$
$M = \frac{1}{2} \sum$
$(x^n)' = nx^{n-1}$
$\frac{\sqrt[n]{a}}{\sqrt[n]{b}} = \sqrt[n]{\frac{a}{b}}$
$\sin\alpha = 2\sin\frac{\alpha}{2} \cdot \cos\frac{\alpha}{2} \cdot$

Central Limit Theorem for Correlated Variables with Limited Normal or Gamma Distributions[1]

Dennis DeRiggi

Physical phenomena, such as the concentrations of compounds in plumes of gas, are sometimes represented mathematically by limited normal distribution functions. An early example of such an application can be found in an early research paper by Lewellen and Sykes (1986, 1145–1154), where the authors employed a limited normal distribution to represent power-plant plume concentrations. As toxic plume concentrations are potential causes of casualties, whether due to accidents or terrorism, the ability to quantify their cumulative effect (i.e., total dosage) is relevant. The purpose of this article is to demonstrate that, under certain strict conditions imposed on correlation, a variant of the Central Limit Theorem applies to a collection of correlated random variables with a limited normal or gamma distribution.

Introduction

The premise underlying this work is that concentrations can be represented as a strictly stationary process $\{Y_n\}$ with common mean $\mu > 0$ and standard deviation σ. In this discussion, the variables in the process $\{Y_n\}$ are taken to be limited normal, *clipped normal* in the Lewellen and Sykes terminology, random variables. (The third section of the full article addresses the *limited gamma* distribution case.) That is, they are bounded non-negative random variables with common distribution function F, given by:

$$\begin{aligned} F(x) &= 0, \quad x < 0 \\ &= \frac{1}{\tilde{\sigma}\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(y-a)^2/2\tilde{\sigma}^2}\, dy\,, \quad 0 \le x < C \\ &= 1, \quad x \ge C, \end{aligned}$$

where $C, \tilde{\sigma} > 0$, and a is an arbitrary real number. The $\{Y_n\}$ variables are *atomic* random variables (in the sense that the probability of assuming the value zero or C is positive) with common mean μ and standard deviation σ, which are, of course, determined by $C, \tilde{\sigma}$ and a. The limited gamma is defined analogously.

[U]nder certain strict conditions imposed on correlation, a variant of the Central Limit Theorem applies to a collection of correlated random variables with a limited normal or gamma distribution.

Occasionally in this discussion, it is useful to refer to the normal random variables underlying the limited normal collection. These normal variates are designated $\{Z_n\}$ to distinguish them from their limited normal counterparts. In particular, it is assumed that the underlying normal variates have a multivariate distribution. The imposed condition on correlation referred previously is that the underlying normal variables have correlations that are exponentially decreasing. Specifically, if Z_i and Z_j are two normal

[1] Based on "A Central Limit Theorem for Correlated Variables with Limited Normal or Gamma Distributions," published in *Communications in Statistics—Theory and Methods*, December 2018, https://doi.org/10.1002/2017SW001626.

variates corresponding to Y_i and Y_j, respectively, then the correlation coefficient of the normal variates is $e^{-v|i-j|} = \exp(-v|i-j|)$ for some fixed, $v > 0$.

Motivation for assuming exponentially decreasing correlation can be found in Sykes et al. (2011), where concentrations of gaseous materials were modeled as having exponentially decreasing correlations when indexed by time. That is, the correlation between concentrations at times t_1 and t_2 is $e^{-v|t_1-t_2|}$. Here, the focus is instead on a countable collection of limited normal variates indexed by the positive integers (e.g., concentrations indexed by discrete time steps), and for which the underlying normal variates have exponentially decreasing correlations. (A numeric example appears at the end of the Supplemental Materials of the online version of this article.)

Mixing and Ibragimov's Theorems

In Theorem 2.1 of Ibragimov (1975), the author proved that given a strictly stationary *zero mean* sequence—a condition referred to as ρ-mixing—and $E\left(\left|X_n^{2+\delta}\right|\right) < \infty$ for some $\delta > 0$ and $\sigma_n \longrightarrow \infty$ are sufficient to imply that the asymptotic distribution of

$$S_n/\sigma_n = \sum_{k=1}^{n} X_k/\sigma_n$$

is standard normal, where σ_n is the standard deviation of the sum. (See Appendix A of the full article for proof of $\sigma_n \to \infty$ as n increases for sufficiently large v. Also note that $E\left(\left|X_n^{2+\delta}\right| < \infty\right)$ for limited variates.) The definition of ρ-mixing is if $\Im_j^k$ denotes the sigma algebra generated by random variables $\{X_j, X_{j+1}, \cdots, X_k\}$ and $\Im_{k+n}^{\infty}$ denotes the sigma algebra generated by $\{X_{k+n}, X_{k+n+1}, \cdots\}$, then the collection of random variables satisfies the ρ-mixing (Kolmogorov 1960) condition whenever

$$\rho(n) = \sup|\text{Correlation}\,(u, v)| \to 0 \text{ as } n \to \infty$$

where the supremum (or least upper bound) is taken over all u in $L_2\left(\Im_1^k\right)$ and v in $L_2(\Im_{k+n}^{\infty})$. Ibragimov (1975) also shows that a condition known as φ-mixing implies ρ-mixing. The collection of random variables satisfies the φ-mixing condition whenever

$$\varphi(n) = \sup\{|P(A|B) - P(A)|, A \in \Im_1^k, B \in \Im_{k+n}^{\infty}\} \to 0 \text{ as } n \to \infty\,.$$

To show that φ-mixing holds in the limited normal case, choose $\delta > 0$, select a sequence of measurable sets $\{A_n, B_n\}$ such that $A_n \in \Im_1^{k_n}, B_n \in \Im_{n+k_n+1}^{\infty}$, and

$$|P(A_n|B_n) - P(A_n)| > \delta.$$

Following the development in Lamperti (1966, 13–15), any $A_n \in \Im_1^{n_1}$ can be represented as $\left\{(U_1, U_2, \cdots, U_k, \cdots)\middle| i \le n_1 \Rightarrow U_i \in \Im_i^i\right\}$. Similarly, any $B \in \Im_{n_1+n+1}^{\infty}$ can be represented as the union of sets of the form

$$\left\{(\cdots, U_{n_1+n+1}, U_{n_1+n+2}, \cdots, U_{n_1+n+n_2-1}, \cdots)\middle| n_1 + n + 1 \le i \le n_1 + n + n_2 - 1 \Rightarrow U_i \in \Im_i^i\right\}.$$

A series of computations largely focused on the covariance matrices in the limited normal and gamma cases demonstrated that "φ-mixing" holds for both distributions. These, and other intermediate results, demonstrate that a sequence of correlated random variables $\{Y_1, Y_2, \cdots, Y_n\}$ with either a limited normal or gamma distribution and a sufficiently rapidly exponentially decreasing correlation ensures that φ-mixing implies ρ-mixing. This, in conjunction with results due to Ibragimov's observations, implies that the asymptotic distribution

$$S_n/\sigma_n = \sum^{n} X_k/\sigma_n \text{ is the standard normal.}$$

Thus, this is another example (there are several in the literature) of how the Central Limit Theorem can be extended, under certain circumstances, beyond the realm of independent variables.

References

Ibragimov, I. 1975. "A Note on the Central Limit Theorem for Dependent Random Variables." *Theory of Probability and Its Applications* 20, no. 1: 135–141. https://doi.org/10.1137/1120011.

Kolmogorov, A. N., and Y. A. Rozanov. 1960. "*On Strong Mixing Conditions for Stationary Gaussian Processes.*" *Theory of Probability and Its Applications* 5, no. 2: 204–208. https://doi.org/10.1137/1105018.

Lamperti, J.1966. *Probability: A Survey of the Mathematical Theory*. 1st Edition. New York, Amsterdam: W.A. Benjamin, Inc.

Lewellen, W. S., and R. I. Sykes. 1986. "Analysis of Concentration Fluctuations from Lidar Observations of Atmospheric Plumes." *Journal of Climate and Applied Meteorology* 25 (August): 1145–1154. https://doi.org/10.1175/1520-0450(1986)025<1145:AOCFFL>2.0.CO;2.

Sykes, R. I., S. F. Parker, D. S. Henn, and B. Chowdhury. 2011. SCIPUF Version 2.7 Technical Documentation, p. 33, Princeton, N.J.: Sage Management.

About the Author

Dennis DeRiggi has been a member of the research staff at IDA's Systems and Analyses Center since 1989, first in the Strategy, Forces and Resources Division and now in the System Evaluation Division. His doctoral degree is in mathematics from the University of Maryland. His undergraduate degree, also in mathematics, is from MIT. This marks the first time Dennis has been associated with a Welch Award–nominated publication.

Dennis acknowledges **Nathan Platt** and **Michael A. Ambroso**, both members of the research staff in the System Evaluation Division, for suggesting the topic and providing insight to the associated physical phenomena.

Complexity in an Unexpected Place: Quantities in Selected Acquisition Reports[1]

Gregory A. Davis, Margaret L. Giles, and David M. Tate

The Selected Acquisition Reports (SARs) that the Department of Defense (DoD) annually produces and submits to the Congress are a primary data source for studying Major Defense Acquisition Programs (MDAPs). But how reliable are the SAR data? This research looks at how quantities and associated unit costs are reported in the SARs. We discovered many examples in which the definition of a unit is non-intuitive, inconsistent, or both. For example, the current U.S. Army CH-47 Chinook helicopter program actually has four different variants of the CH-47, and units of each variant change over the years. Units produced 10 years ago are significantly different from new units coming off the line today. The CH-47 program is not unusual in this kind of unit variation; in fact, we have found few programs in which counting quantities is straightforward.

We found that significant changes over time and mix type issues are more than merely common for MDAPs; they are nearly universal.

Selected Acquisition Reports

The SAR dataset has many appealing characteristics for purposes of cost analysis. It reports funds from all different appropriations related to an acquisition program in one place, whereas appropriations in the budget submissions are scattered throughout different exhibits and military departments. Each SAR also reports all funding in both base-year and then-year dollars, and quantities from the beginning of the program until its planned conclusion. Analysts use data from these reports to calculate cost growth and other types of analysis related to the MDAPs.

Quantity Reporting Is Not Simple

Most analyses of these data are based on the assumption that each unit is essentially identical to every other unit of the same type, but this is often not the case. In our review of SAR data, we grouped reasons for differences among units into one of three categories: changes over time, mixed types, and reporting accidents. We found that significant changes over time and mix type issues are more than merely common for MDAPs; they are nearly universal. We also found instances of significant accidents in SAR reporting of MDAPs. Few programs are entirely devoid of these issues and many show more than one.

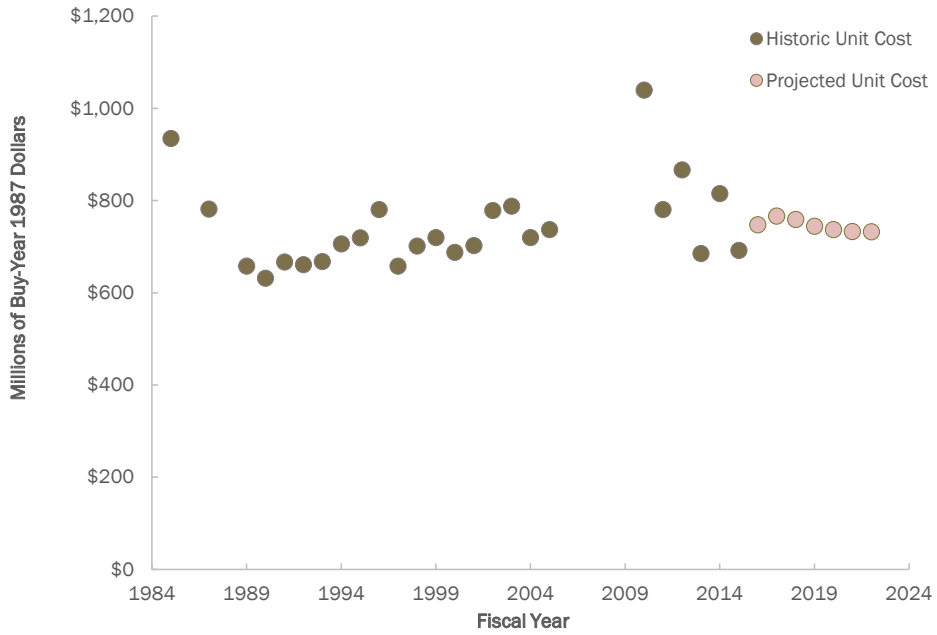
Changes over Time

Changes over time refers to cases in which a military department changes the design—but not the designation—of the items being purchased from manufacturing lot to lot. For example, the Navy bought its first DDG-51, USS *Arleigh Burke*, in 1985 and, according to the December 2015 SAR listing, plans to buy the final two ships in 2022. One analyst might expect the cost of each ship to be about the same through the

[1] Based on "Extending CryptDB to Operate an ERP System on Encrypted Data," *Proceedings of the 20th International Conference on Enterprise Information Systems* (ICEIS 2018) 1 (March 2018): 103–110, https://calhoun.nps.edu/handle/10945/58689.

years, while another might expect the cost of each successive ship to decrease due to learning. The data tell neither story.

In Figure 1, we see that the annual unit cost of these destroyers over time shows large jumps, steep falls, and periods of gradual upward slopes, with cost ranging from about $600 million to $1 billion. The usual interpretation of this pattern is that the jumps come from adopting new designs and the steep falls come from learning how to build them—a process that includes both learning by the shipyard's staff and also from investments in technology. The upward slopes that are visible in the graph are probably from gradual upgrades to the ships. The projection forecasts that the Navy will stop enhancing the ships and let the learning effect continue to dominate for the balance of the program. But history suggests that this is unlikely; the units are likely to change yet again during this period.



Source: Department of Defense, 2016, Department of Defense Selected Acquisition Reports (SARs), Release No. NR-106-16, as of December 31, 2015 (March 24, 2016), accessed March 27, 2017, "Cost Quantity Information" section.

Figure 1. Actual and Projected DDG-51 Unit Costs

Figure 2 shows four ships of the Arleigh Burke class. The visible design differences among these ships account for some of the cost differences seen in Figure 1. USS *Fitzgerald*, on the left, has a bellmouth in the stern for a towed array sonar that is missing from USS *Sampson*, on the right. *Sampson*, unlike *Fitzgerald*, also has twin hangars for helicopters. (The two other DDG-51s in the photo are USS *Michael Murphy*

and USS *Curtis Wilbur*.) In spite of these and other configuration differences, SAR reporting counts each of the four destroyers in the photo as one unit of the same type.

This issue of changes in the content of a unit over time is not unique to this program or even to shipbuilding; it permeates all of defense acquisition cost reporting.

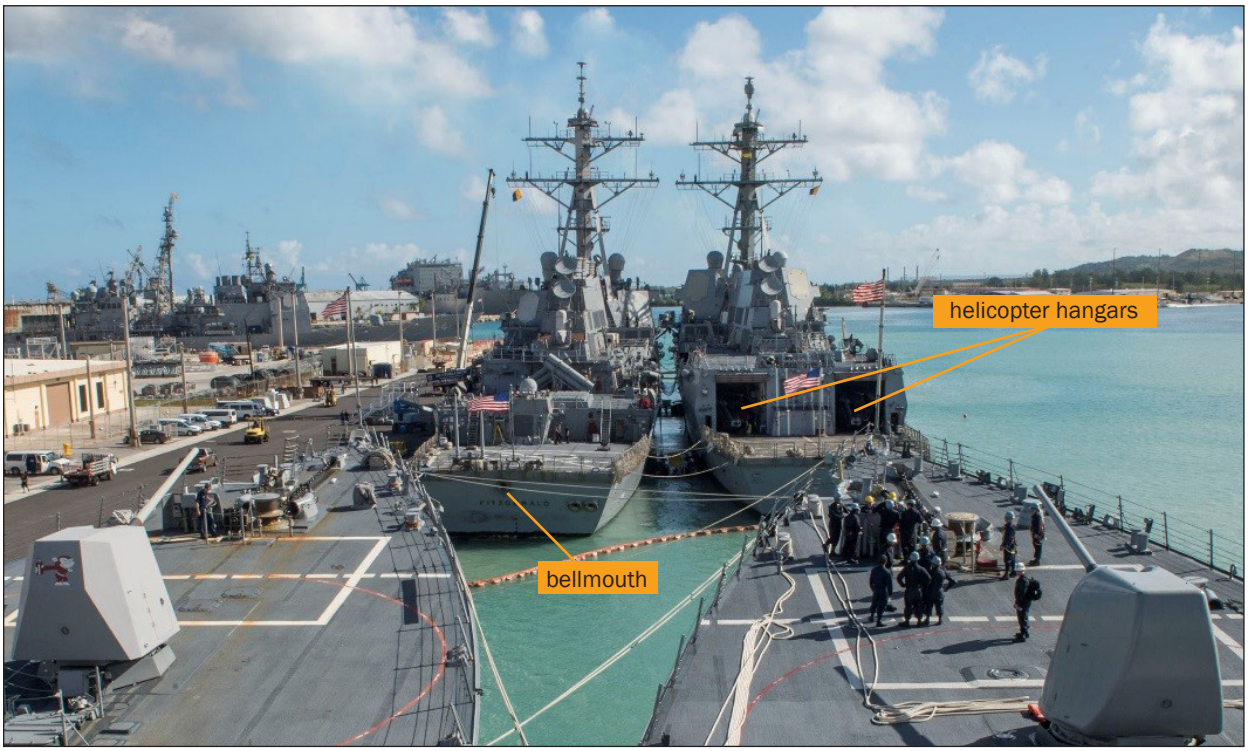


Photo by U.S. Navy Mass Communication Specialist 3rd Class Raymond D. Diaz III, courtesy of U.S. Indo-Pacific Command Image Gallery.

Figure 2. Four DDG-51 Class Destroyers in Guam

Mixed Types

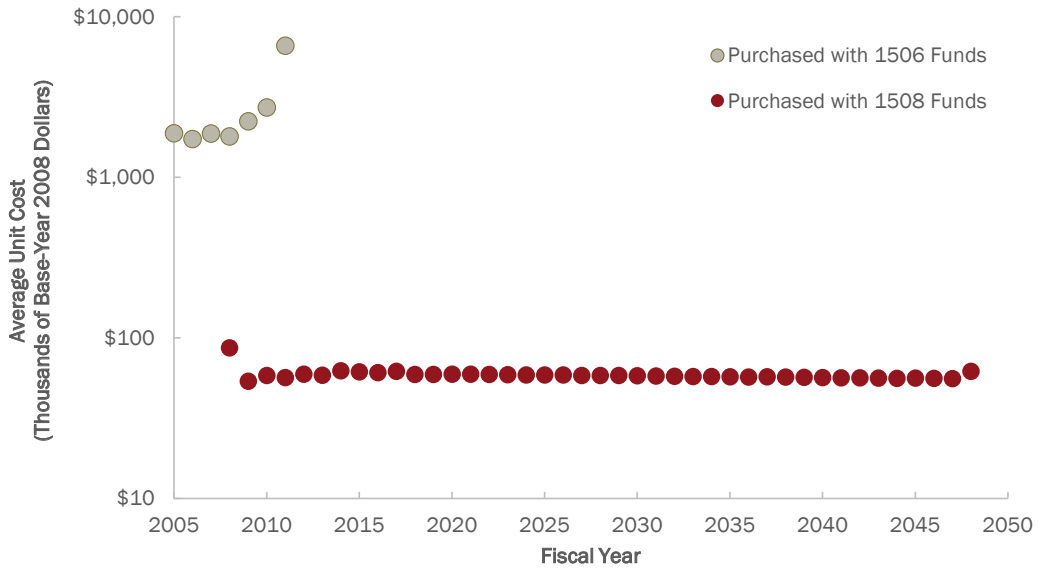
Mixed types refers to situations in which the program is purchasing multiple distinct end items but does not distinguish among them when counting units. One example of mixed types is found with the F-35 Joint Strike Fighter program. The DoD is buying three variants of the F-35 for Air Force, Marine Corps, and Navy use:

- F-35A conventional take-off and landing (CTOL)—lowest price tag and generally most capable once airborne
- F-35B short take-off and vertical landing (STOVL)—reduced payload
- F-35C carrier variant (CV)—equipped for aircraft carrier operations

The F-35 is the most well-known example of mixed types, but it is far from the only one. The Navy's Integrated Defensive Electronic Countermeasures (IDECM) program is a lesser-known program where the mix of unit types is even more diverse than those of the F-35.

The IDECM program acquires electronic suites to protect the various F/A-18 aircraft from radio frequency guided missiles. In addition to the electronic suites, the program also buys expendable decoys that are towed behind the airplanes.

The IDECM program contains two subprograms: IDECM Blocks 2/3 and IDECM Block 4. The December 2015 SAR reports the Block 4 subprogram as having an average procurement unit cost (APUC) of $2.502 million, while the Block 2/3 subprogram is reported as having a far lower APUC of $0.090 million. This cost difference is because the quantities being purchased in these blocks are so different. Block 4 has a quantity of 324 units, roughly the number of F/A-18C/D aircraft the units will be protecting. IDECM Block 2/3 has a quantity of 12,805 units, although the Navy bought fewer than 600 F/A-18E/Fs, the aircraft that these units will be mounted on. Eighty-five of the 12,805 were purchased with Navy Aircraft Procurement funds (1506 funds), and the balance were or will be bought with Procurement of Ammunition, Navy and Marine Corps funds (1508 funds). We presume that only disposable decoys are being purchased with 1508 funds. The unit costs for each year of the IDECM Blocks 2/3 subprogram by acquisition fund are presented in Figure 3.



Source: Department of Defense, 2016, Department of Defense Selected Acquisition Reports (SARs), Release No. NR-106-16, as of December 31, 2015 (March 24, 2016), accessed March 27, 2017, "End Item Recurring Flyaway" column.

Figure 3. IDECM Block 2/3 Annual Unit Cost by Appropriation Type, 2005–2050

Showing unit costs by appropriation type on the same chart required plotting them on a logarithmic scale, yet the two purchases (electronics suite and decoy) are each counted as the same type of unit for the official unit cost calculation. Just within the more

expensive 1506 units, it is clear that there have been significant changes, as the cost there does not follow a typical learning shape, which would be expected to slope down.

Reporting Accidents

The issues described thus far generally come about because leadership makes a decision about how the program should be managed and what systems it should produce, possibly without considering the impact this will have on the coherence and consistency of quantity or unit cost reporting. In contrast, *reporting accidents* seem to be outright errors in how the quantity numbers were put together, despite the quality control processes in place that are designed to prevent this. Accidents in reporting are inherently difficult to find; spotting them requires either knowing the truth or recognizing inconsistencies in separate reports that are not designed to be easily reconciled. We found three instances of accidents by comparing SAR filings and President's Budget justification books. These accidents were in the Army's CH-47F Chinook program, the Air Force's Intercontinental Ballistic Missile Fuze Modernization program, and the Air Force's Evolved Expendable Launch Vehicle program. We do not suggest that any of the reporting accidents we found were intended to confuse analysts, but they did have that effect.

Ramifications of Complex Unit Reporting

The prevalence of all of these issues poses serious challenges, both to analysts attempting to understand the causes and mechanisms of cost growth and to oversight bodies attempting to understand cost and capability changes in active programs.

Challenges for Analysts

Analysts have sought to develop predictive models of cost growth with limited success. McNicol (2017) found that reported unit cost growth in MDAPs is closely associated with periods of relatively generous defense budgets. Our findings suggest one possible mechanism for this association—namely, that generous budgets permit programs to add features and correct defects over time, so that units produced in later lots are more capable than those produced in earlier lots. There has been little effort to capture this effect with predictive models, perhaps because few analysts are aware of the need.

Similarly, acquisition analysts have long been interested in trying to predict the effect of changes in production rate on production costs. Schedule instability and production stretch-outs have long been cited as primary causes of unit cost growth, but causal models of this effect have been elusive. Our findings suggest that this may in part be due to unaccounted-for variation in the content of units being procured within a given program.

Ongoing content change also has implications for how cost analysts should model learning curves. Traditional learning curve theory assumes that the content of all units is identical, but that unit costs decline exponentially as a function of cumulative units

produced. Our findings suggest that this is a poor model. For one thing, it fails to account for increasing content (and cost) from one manufacturing lot to the next. For another, it fails to account for losses of learning due to significant design changes. It may be that the "learning and forgetting" model proposed by Benkard (2000) works fairly well precisely because the forgetting portion of the model can approximately account for both of these effects. Finally, of course, if the units being produced are actually a mix of several different designs, it is difficult to say how much production of one type of unit will drive learning for the other types.

Challenges for Oversight

The acquisition oversight community, from the Congress to the Secretary of Defense down to individual program managers, needs to be able to accurately estimate the likely impacts of changes to a program's acquisition strategy. All of these stakeholders are aware that cost growth is a significant problem for long-term planning and budgeting, but no general-purpose predictive models for program cost growth have been identified. The Congressional Budget Office applies generic cost growth factors for each category of weapon system (such as surface ship, tactical aircraft, or automated information system), but those factors are at best correct on average. Actual cost growth varies widely within each category.

Our findings suggest that some of this cost growth is deliberate, but poorly reported to at least some of the oversight stakeholders. This suggests the possibility of both predictive modeling of that portion of cost growth and improved reporting procedures to inform the Congress and the Office of the Secretary of Defense of (possibly contingent) plans for future content change within programs.

Potential Improvements

We propose a few possible adjustments to the reporting system that could make SAR data more useful both for oversight and analysis. Our primary concern was to avoid creating new destructive incentives for acquisition officials. Any change in reporting should also avoid placing too much extra burden on staff or revealing too much information to our adversaries around the world.

Different reasons for unit inconsistency call for different solutions. Correcting for *changes over time* calls for data that allow analysts to get a sense of how much units have changed. Accounting for *multiple types* calls for a better explanation of what is being purchased. And reducing *reporting accidents calls* for an examination of the process used to generate SAR data and budget justification submissions.

References

Benkard, C. L. 2000. "Learning and Forgetting: The Dynamics of Aircraft Production." *American Economic Review* 90, no. 4 (September): 1034–54. https://doi.org/10.1257/aer.90.4.1034.

Department of Defense. 2016. Department of Defense Selected Acquisition Reports (SARs). Release No. NR-106-16 (as of December 31, 2015), March 24.

https://dod.defense.gov/News/News-Releases/News-Release-View/Article/703529/department-of-defense-selected-acquisition-reports-sars/.

McNicol, D. L. 2004. *Cost Growth in Major Weapon Procurement Programs*. IDA Paper P-3832. Alexandria, VA: Institute for Defense Analyses, October.

McNicol, D. L. 2017. *Post-Milestone B Funding Climate and Cost Growth in Major Defense Acquisition Programs*. IDA Paper P-8091. Alexandria, VA: Institute for Defense Analyses, March. https://www.ida.org/research-and-publications/publications/all/p/po/post-milestone-b-funding-climate-and-cost-growth-in-major-defense-acquisition-programs.

U.S. Indo-Pacific Command. 2015. Image Gallery (150320-N-BB269-049.JPG). https://www.pacom.mil/Media/Photos/igphoto/2001027568/.

About the Authors

Margaret Giles joined the Cost Analysis and Research Division (CARD) of IDA's Systems and Analyses Center in 2016 as a research associate. She holds a master's degree in international affairs from George Washington University. Margaret is being recognized for the first time as a coauthor of a Welch Award finalist.

Gregory Davis has been a member of the research staff in CARD since 2006. He holds a doctorate in physics from the University of Rochester. This is Greg's first time being recognized for his participation in the Welch Award competition.

David Tate joined the research staff of CARD in 2000. He holds master's and doctoral degrees in operations research from Cornell University. He was recognized in 2015 as a coauthor of the Welch Award finalist publication "A Technical Review of Software Defined Radios: Vision, Reality, and Current Status."

Extending CryptDB to Operate an Enterprise Resource Planning System on Encrypted Data[1]

Kevin Foltz and William R. Simpson

The Department of Defense is adopting a cloud computing model, but storing sensitive data in the cloud raises security issues. One potential solution involves the use of partial homomorphic encryption to protect the data. Such an encryption scheme allows for select computations on the data while it remains encrypted. Prior work demonstrated the feasibility of using partial homomorphic encryption as part of a database encryption scheme in which standard Structured Query Language (SQL) queries are performed on encrypted data using the CryptDB system. Our work extends this concept to work with an Oracle Enterprise Resource Planning (ERP) database to include stored procedures, views, and multiuser access controls. We show that these additional functionalities can be practically implemented using encrypted data, and they can be implemented in a way that requires no changes to the ERP application code. The time delays before transfer of data completes (latency) and computational resources necessary for operating on encrypted data are within a small factor of those for unencrypted data. These results demonstrate the feasibility of using partial homomorphic encryption to securely store and compute on ERP data in the cloud.

Homomorphic encryption… provides added security for hosting in a cloud, where raw sensitive data may be accessible to an untrusted third party.

Partial Homomorphic Encryption

Homomorphic encryption enables manipulation of encrypted data to perform computations on the underlying unencrypted information (plaintext) without decrypting the data. It provides added security for hosting in a cloud, where raw sensitive data may be accessible to an untrusted third party. By homomorphically encrypting data prior to storing it in the cloud, the data can be used for computations while remaining protected. This method stops threats to confidentiality posed by the cloud provider, its employees, and external entities in a position to compromise the cloud provider.

Full homomorphic encryption allows any encrypted computation (Gentry 2009) but is prohibitively slow for all but the simplest of computations (Gligor 2014). Partial homomorphic encryption permits only a single operation, such as addition, but this single operation is fast. In systems where this single operation is sufficient, partial homomorphic encryption can be part of a viable security solution.

Encrypting an ERP Database

Research has shown that a set of standard SQL queries can be run on a database encrypted with partial homomorphic encryption (Popa et al. 2012). The encryption is performed by the proxy in the CryptDB system, which is located between the database requester and the database. The CryptDB proxy translates queries on unencrypted data into queries on encrypted data, allowing a user to access the encrypted database as if it were not encrypted.

---

[1] Based on "Extending CryptDB to Operate an ERP System on Encrypted Data," *Proceedings of the 20th International Conference on Enterprise Information Systems (ICEIS 2018)* 1 (March 2018): 103–110, https://doi.org/10.5220/0006661701030110.

Real systems are not as simple as a single database. A typical ERP system has the following additional complications:

- Proprietary ERP code that cannot be changed
- Primary and foreign key reference integrity
- Stored procedures
- Views
- Multiple accounts with different permissions

We first rewrote CryptDB to work with the Oracle SQL database used with Oracle ERPs. CryptDB was originally written to work with MySQL, an open-source relational database management system. Then, we added new features and capabilities to address the issues listed above. We used a test ERP application to compare functionality and performance between the original unencrypted database and the encrypted database.

Results

Results include assessments of the following:

- Functionality
- Bulk encryption
- Operational performance

Functionality

To test functionality, we connected one instance of the ERP application to the original unencrypted database and a second instance through the CryptDB proxy to the encrypted database. A side-by-side comparison showed nearly identical behavior when they retrieved data from their respective databases. This confirmed that porting CryptDB to work on an Oracle database preserved the original capabilities. Further testing showed that the additional features were functioning correctly on the encrypted database.

Bulk Encryption

The conversion of an existing unencrypted database to an encrypted database was conducted on a commercial off-the-shelf laptop and desktop with four and forty cores, respectively.

The time needed for this bulk encryption scaled linearly with database size and inversely with the number of central processing unit cores available. This finding suggests the initial encryption is highly scalable and parallelizable. The time to encrypt one million database entries was about one hour, indicating real-world feasibility.

Operational Performance

We tested a million-user database with a set of 18 queries that included insert, update, delete, deterministic encryption (which allows determination of whether two encrypted values have the same plaintext values without revealing the plaintext), order-preserved encryption (which reveals the relative size of the plaintext values without revealing the values themselves), and Pallier homomorphic encryption (which allows addition of encrypted values while protecting underlying plaintext). Queries 1–6 returned large result sets, Queries 7–15 were more typical of ongoing business operations, and Queries 16–18 modified the database using insert, update, and delete operations. To compare the user experience and resource requirements for encrypted operation versus unencrypted operation, increasing numbers of active threads (objects that can run instructions in a process simultaneously) were repeatedly cycled through Queries 7–15.

The operational impact of encryption to latency was based on the average time duration between an application request and response. Changes in resource requirements were computed as the inverse of changes in maximum throughput. Maximum throughput is the throughput achieved when resources are maximally utilized, and it is observed as the value of throughput at which latency starts to rapidly rise.

The different queries showed a range of changes in latency for encrypted and unencrypted data, as Figure 1 indicates. Query 15, which invokes homomorphic addition with the Paillier cryptosystem, showed the largest increase in latency; Queries 11 and 14, which performed matching and searching, had smaller latency increases. Query 12, which is an order-preserving encryption search, actually performed better on encrypted data. We suspect that this improvement is due to internal database optimizations that more than compensated for the additional encryption processing.

Figure 2 shows the average latency and throughput performance curves for a mix of queries. Latency for encrypted data is about twice the latency for unencrypted data, and the factor of four reduction in maximum throughput translates to a factor of four increase in resource requirements. These small integer factors compare favorably with the multiple orders of magnitude penalty for using full homomorphic encryption.

Conclusion

Database features added to CryptDB to support an ERP application for an Oracle database were shown to operate correctly on encrypted data. The structure of CryptDB required no changes to the ERP application code when switching between an unencrypted and encrypted database, which allows “black box” applications to use this method. Converting an existing database to an encrypted database scaled well in database size and computational resources. Operational performance varied with the query type but for a typical operational mix showed only small integer multiple increases in latency and resource requirements compared to an unencrypted database. This shows that it is feasible to run an ERP system with an encrypted database in an untrusted cloud hosting environment using CryptDB.

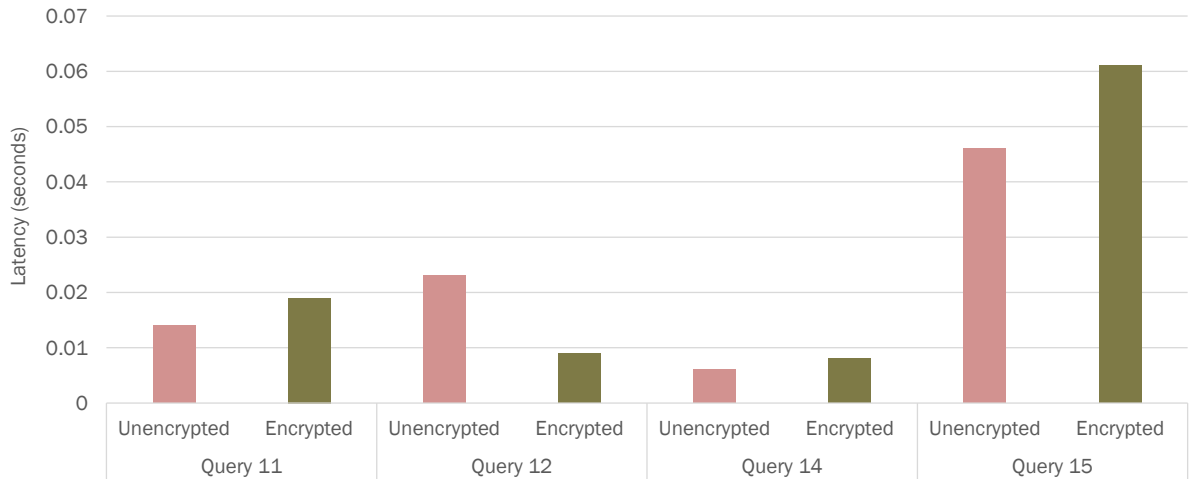


Figure 1. Latency for a Range of Queries with a Single Thread

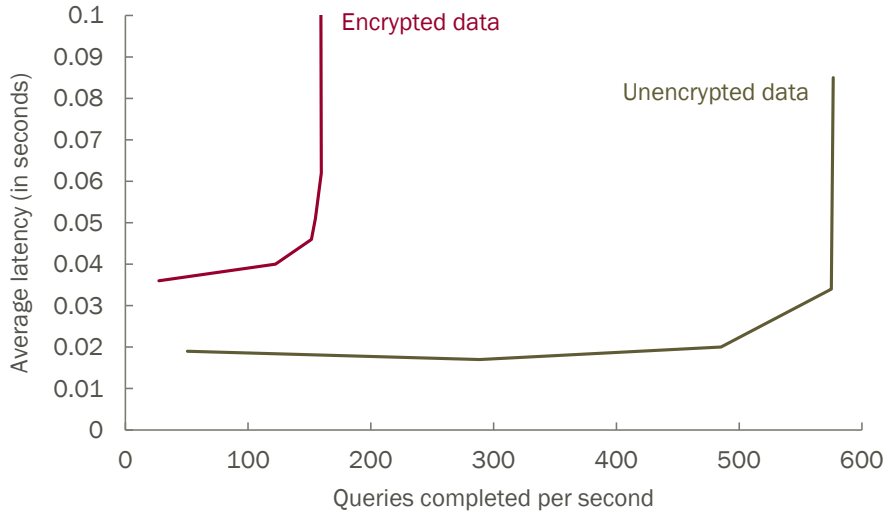


Figure 2. Performance Summary, Queries 7–15

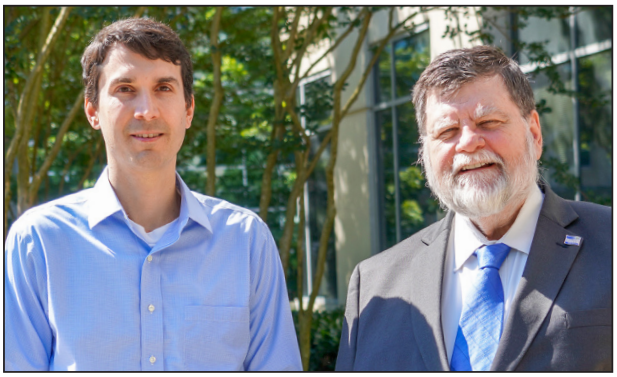
References

Gentry, C. 2009. "A Fully Homomorphic Encryption Scheme." Doctoral thesis. Stanford University, Department of Computer Science (September). https://crypto.stanford.edu/craig/craig-thesis.pdf.

Gligor, V. 2014. "Homomorphic Computations in Secure System Design," Final Report. Pittsburgh, PA: Carnegie Mellon University.

Popa, R. A., C. M. S. Redfield, N. Zeldovich, and H. Balakrishnan. 2012. "CryptDB: Processing Queries on an Encrypted Database," *Communications of the ACM* 55, no. 9 (September): 103–111. https://doi.org/10.1145/2330667.2330691.

About the Authors



Kevin Foltz is a research staff member in the Information Technology and Systems Division of IDA's Systems and Analyses Center. He holds a master's degree in strategic security studies from National Defense University and both master's and doctoral degrees in electrical engineering from the California Institute of Technology. This is Kevin's first time being recognized for his contribution in the Welch Award competition. He has been with IDA since 2002.

William (Randy) Simpson is also a research staff member in ITSD, where he has worked since 1993. He holds a master's degree in administration from George Washington University and both master's and doctoral degrees in aerospace engineering from Ohio State University. Randy is the author of the book *Enterprise Level Security: Securing Information Systems in an Uncertain World*, a Welch Award finalist in 2017, and coauthor of "High Assurance Challenges for Cloud Based Computing," a Welch Award finalist in 2012.

Economic Uncertainty and the 2015 National Defense Stockpile[1]

Justin M. Lloyd, Wallice Y. Ao, Amrit K. Romana, and Eleanor L. Schwartz

The Strategic and Critical Materials Stock Piling Act (Title 50 USC § 98) requires the Secretary of Defense to provide a biennial report to Congress on estimated requirements for the national defense stockpile (NDS) of nonfuel strategic and critical materials. The report must include analyses of the sensitivity of the requirements to variability in model assumptions and input data. IDA contributes to this biennial effort not only by helping to determine stockpile requirements but also by conducting sensitivity analyses. This article quantifies the sensitivity of the NDS report recommendations to macroeconomic forecasting errors and describes the methods used to conduct the sensitivity analysis.

Introduction

The United States relies heavily on imports of strategic and critical materials (e.g., rare earth elements, metals, and carbon fibers) from Asian, South American, African, and European countries. These materials are elements of a broad range of products from airframes, engines, and global positioning satellites to transmission lines, batteries, and pharmaceuticals, many of which are essential to sustaining critical civilian and military services. The U.S. stockpiles some of these strategic and critical materials as insurance in the event imports are disrupted. Potential shortfalls in the supply of these materials are an important consideration in the biennial report to Congress concerning the question: What materials should the U.S. stockpile contain?

The United States stockpiles… strategic and critical materials as insurance in the event imports are disrupted…. What materials should the U.S. stockpile contain?

The Risk Assessment and Mitigation Framework for Strategic Materials (RAMF-SM) is a suite of models used to help the Department of Defense (DoD) determine material stockpile requirements. The DoD-accredited RAMF-SM identifies the materials needed to meet civilian and military demand given a baseline national emergency scenario. The models calculate the difference between material demands and supplies to derive potential shortfalls. Materials subject to potential shortfall become candidates for stockpiling or other measures.

RAMF-SM integrates an extensive government-wide repository of data and policy judgments spanning the legislative and executive branches, other government offices and agencies, and the private sector. RAMF-SM results were the basis of determinations made in Strategic and Critical Materials 2015 Report on Stockpile Requirements, the 2015 report to Congress.

RAMF-SM integrates an extensive government-wide repository of data and policy judgments spanning the legislative and executive branches, other government offices

[1] Based on "Methods in Macroeconomic Forecasting Uncertainty Analysis: An Assessment of the 2015 National Defense Stockpile Requirements Report," Mineral Economics, Raw Materials Report 31, no. 3, October 2018, https://doi.org/10.1007/s13563-017-0127-6.

and agencies, and the private sector. RAMF-SM results were the basis of determinations made in *Strategic and Critical Materials 2015 Report on Stockpile Requirements*, the 2015 report to Congress.

Role of Economic Forecasting

An important consideration in the preparation of the report is the sensitivity of the report's findings to policy, strategic, military, and economic assumptions made in the report's analysis. These assumptions stem from a wide range of data sets, policies, and directives furnished by several government agencies and private organizations. Ultimately, the report delivers recommendations for the appropriate composition and quantity of the nation's strategic defense stockpile of critical materials. Logically, a major driver of these recommendations is the forecasted economic demand and output of the United States over the period being analyzed. The nominal set of assumptions, scenario definitions, and data that provide the foundation for the NDS report are referred to as the Base Case. A substantial component of the Base Case consists of the macroeconomic forecasts for the U.S. economy over the future years to which the Base Case scenario applies. These forecasts are manifested in the output of the macroeconomic simulations employed in the NDS analysis, but they also drive the macroeconomic simulations themselves.

To promote consistency with other government analyses, the NDS report leverages the forecasts of the Council of Economic Advisers (CEA) of the President of the United States. Twice a year CEA, together with the Office of Management and Budget (OMB) and the U.S. Department of the Treasury, develops macroeconomic projections of the U.S. economy. These projections reflect a given presidential administration's economic agenda and are published as a summary with the annual Economic Report of the President. They also coincide with the production of estimates for the Budget of the United States Government and the Mid-Session Review of the Budget. The CEA provides IDA with a comprehensive set of forecasts for key macroeconomic indicators compiled on a National Income and Product Accounts (NIPA) basis. While other trusted sources of macroeconomic forecast data are available from both private and alternative federal institutions, the NIPA breakdown of the CEA forecasts is consistent with the president's budget and other government analyses and is sufficiently detailed. Thus, CEA forecasts are uniquely instrumental in our analysis.

One issue arising from this process is the accuracy of the macroeconomic forecast delivered by CEA. Every fiscal year, OMB publishes a volume of analytical perspectives supplementing the Budget of the United States Government. This volume provides analysis on the historical error trends of various government and private sector forecasts. According to these analyses, the CEA macroeconomic forecasts historically exhibit a similar level of accuracy to the forecasts and an average of private sector, blue chip forecasts. Non-negligible errors exist between the actual and predicted macroeconomic status of the U.S. economy. To address these issues, we quantified the sensitivity of the NDS recommendations to these forecasting errors.

Method of Analysis

Our analysis was conducted in three major stages:

Stage 1: Calculate Base Case estimates for material shortfall. The economic models of RAMF-SM enable the systematic decomposition of aggregated macroeconomic variables into industry-level categories. These models were first calibrated to be consistent with the detailed NIPA breakdown of the CEA's economic forecasts and then applied to compute the corresponding projections of industry-level final output requirements. We used these output requirements in the subsequent calculations of raw material demands and, specifically, the types and amount of strategic and critical materials needed for national defense purposes. Shortfall for each raw material was then calculated by the difference between demand and supply.

Stage 2: Systematically perturb the baseline macroeconomic forecast that drives Base Case material demand estimates. This was done to capture the impact of economic forecasting errors, which are a proxy for forecast uncertainty. This approach focuses on quantifying the implications on NDS recommendations when the economy experiences a higher or lower than expected growth. A systematic procedure to translate the measure of forecast uncertainty from the aggregate, gross domestic product (GDP) level to the industry level follows. First, a new equilibrium GDP corresponding to the high or low economic growth case was computed. Second, a scale factor was calculated as a function of the baseline (original) equilibrium GDP, new equilibrium GDP, and defense spending, which was assumed to be held at the baseline level because of policy constraints. This scale factor was used to calculate civilian, export, and import demands for each industry under the new equilibrium. Finally, output requirements and material demands corresponding to the new equilibrium GDP were computed.

Stage 3: Use the output from Stage 2 to calculate adjusted material shortfalls using the procedure described in Stage 1. We compared these results to the Base Case to characterize the sensitivity of material shortfall estimates in response to macroeconomic forecast uncertainty. The mathematical characterization of shortfall estimate uncertainties was used to examine various stockpiling situations, such as a more conservative stockpiling environment in which worst-case planning practices are designed to hedge against maximum downside risks.

Results

To explore the potential sensitivity of the 2015 report's conclusions to the accuracy of the CEA's forecast, we observed the effect of systematic variations to the baseline economic forecast on material shortfall calculations. Specifically, we analyzed two sensitivity cases with higher and lower macroeconomic growth than the Base Case. One of these cases assumed that the annual economic growth rate was 1.1% higher than the Base Case forecast growth rate, while the other case assumed that the growth rate was 1.1% lower than in the Base Case. The results showed that a seemingly minor change in the growth rate had a significant effect on the shortfall of the strategic material supply.

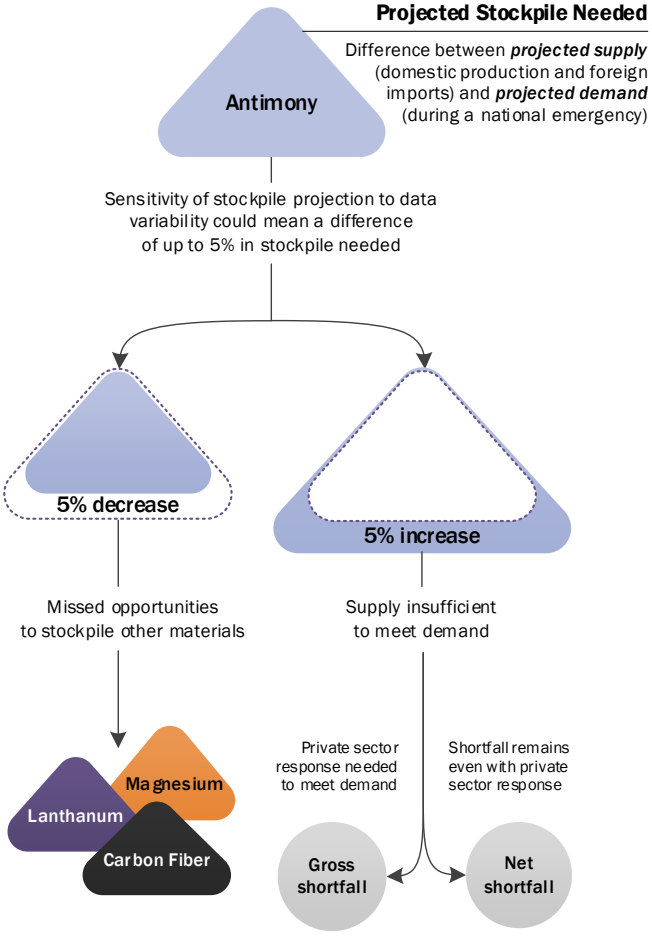


Figure 1. Effects of Shortfall Sensitivity

For the 68 materials examined, the total shortfall value in the higher economic growth case was 18% higher than for the Base Case. The number of materials with shortfalls rises. In the lower economic growth case, the total shortfall was 16% lower than the Base Case. These changes are overwhelmingly in the civilian shortfall. The defense shortfall total exhibits small changes.

Examining the sensitivity of the estimated demands for individual materials to changes in the assumed GDP growth rate provides additional insights. The NDS requirements report analysis typically addresses 70–80 individual materials. Interestingly, the impacts of both the higher and lower assumed growth rates are largely uniform from material to material. Overall, a 1.1% lower annual GDP growth rate results in an overall drop in the demand for an individual material of around 5.5%, whereas a 1.1% annual increase in the assumed GDP growth rate results in around a 6% increase in material demands. Across materials, demand sensitivities are roughly linear in changes in the assumed GDP growth rate. Figure 1 illustrates the effects of the shortfall sensitivity to uncertainties in the economic growth rate.

Conclusion

This work contributes to a better understanding of the effect of economic forecasting errors—a measure of forecast uncertainty—on material shortfall estimates. It is clear from our results that uncertainty ultimately plays an important role in policy making for U.S. material stockpiles. Policy planners and leaders can use this information to make more fully informed decisions, for example, adopting more cautious stockpiling strategies to hedge against worst-case scenarios.

Bibliography

Mas-Colell, A., M. D. Whinston, and J.R. Green. 1995. *Microeconomic Theory*. New York, NY: Oxford Printing Press.

Miller, R. E., and P. D. Blair. 2009. *Input/Output Analysis: Foundations and Extensions*. New York, NY: Cambridge University Press.

Obama, B. H. 2015. *Economic Report of the President*. https://obamawhitehouse.archives.gov/sites/default/files/docs/cea_2015_erp_complete.pdf/.

Office of the Management and Budget. 2014. *Mid-session Review, Budget of the U.S. Government, Fiscal Year 2015*. https://obamawhitehouse.archives.gov/sites/default/files/omb/budget/fy2015/assets/15msr.pdf.

———. 2014. *Analytical Perspectives, Budget of the United States Government, Fiscal Year 2015*. https://www.govinfo.gov/content/pkg/BUDGET-2015-PER/pdf/BUDGET-2015-PER.pdf.

———. 2014. *Budget of the United States Government, Fiscal Year 2015*. https://www.govinfo.gov/content/pkg/BUDGET-2015-BUD/pdf/BUDGET-2015-BUD.pdf.

U.S. Bureau of Economic Analysis. 2014. *Measuring the Economy: A Primer on GDP and the National Income and Product Accounts*. https://www.bea.gov/sites/default/files/methodologies/nipa_primer.pdf.

University of Maryland. 2008. Interindustry Forecasting Project at the University of Maryland (INFORUM). Interindustry Large-scale Integrated and Dynamic (ILIAD) model. College Park, MD: University of Maryland.

———. 2011. Interindustry Forecasting Project at the University of Maryland, University Park (INFORUM). Long-term Interindustry Forecasting Tool (LIFT). College Park, MD: University of Maryland.

About the Authors

Wallice Ao is a member of the research staff in the Strategy, Forces and Resources Division (SFRD) of IDA's Systems and Analyses Center. She joined SFRD in 2015. She holds a doctoral degree in economics from the University of Wisconsin-Madison. This is the first time Wallice has been recognized for her contributions to a finalist publication in the Welch Award competition.

Eleanor Schwartz joined IDA in 1980 and has been a member of SFRD's research staff since 1984. She holds a master's degree in management from the Massachusetts Institute of Technology. Eleanor was previously recognized as a coauthor of a "Strategic Material Shortfall Risk Mitigation Optimization Model," a finalist in the 2017 Welch Award competition.

Justin Lloyd was a research staff member in SFRD from 2012 to 2018. He holds a master's degree in mechanical engineering from Virginia Polytechnic Institute and State University and both master's and doctoral degrees in electrical engineering from Johns Hopkins University. This is Justin's first recognition for his part in the Welch Award competition.

Amrit Romana was a research associate in SFRD from 2014 to 2019. She holds a bachelor's degree in mathematics and economics from the University of Michigan. This is Amrit's first recognition in the Welch Award competition.

Approaching Multidomain Battle through Joint Experimentation[1]

Kevin M. Woods and Thomas C. Greenwood

The rapid growth of capabilities tied to the addition of space and cyber domains of warfare is forcing a re-examination of previous military concepts and doctrine. This article explores the debate around the concept of military operations across warfighting domains. *Multidomain battle* is not a new idea, but developing it beyond a slogan into a new warfighting concept is difficult. New concepts need to demonstrate that they justify the disruptive effects of the change they require. This is a high bar that is worth testing for applicability to warfare in the twenty-first century.

Introduction

Multidomain battle (MDB) holds the promise of more fluid, adaptive, and effective operations across land, sea, air, space, and cyber domains simultaneously. Although operations are conducted in and occasionally across these five domains, developing a concept that makes domain integration the norm and not the exception is a tall order.

This article advocates two approaches to exploring multidomain battle (MDB): (1) linking the concept to the existing body of available evidence and (2) generating new evidence through experimentation. We offer these approaches as ways to explore the top-down theater-level implications of MDB.

Will the application of a multidomain approach enable the Department of Defense (DoD) to overcome current warfighting challenges? Will it allow the military departments to seize new opportunities or merely distract them from restoring conventional warfighting capabilities? Perhaps more importantly, can MDB serve as a unifying concept that DoD business processes can be organized around for the development of future concepts and capabilities?

Careful attention must be paid to the data that will provide solid evidence for the conclusions reached by conducting experiments. If carefully planned and executed, discovery experimentation could be a valuable tool.

MDB is a future concept that "must be stated explicitly in order to be understood, debated and tested to influence the development process" (Schmitt 2002, 4). Any new concept must first be articulated, matured, and validated before it transitions to a capability. We argue that concepts on the scale of MDB require a campaign of experimentation that provides compelling evidence that supports fleshing out its operational and institutional contexts.

State of Debate

Proponents of the emerging MDB concept make the case that the joint force must adapt to the times. One of MDB's strongest proponents, Admiral Harry Harris, commander of U.S. Pacific Command, argues that "MDB conceptualizes bringing

[1] Based on "Multidomain Battle: Time for a Campaign of Joint Experimentation," published in *Joint Forces Quarterly (JFQ 88)*, 1st Quarter 2018, https://ndupress.ndu.edu/Publications/Article/1411615/multidomain-battle-time-for-a-campaign-of-joint-experimentation/.

jointness further down to the tactical level [by] allowing smaller echelons to communicate and coordinate directly while fighting in a decentralized manner" (Harris 2017, 19). Regardless of the operating theater and specific mission, tactical-level MDB operations will drive the departments to change "to a culture of inclusion and openness, focusing on a purple (or joint) first mentality" (Brown 2017). Rhetorically, at least, the emerging MDB concept is progressing from the often stated but little realized goal of reducing conflict and increasing interdependency among the military departments. The most optimistic version of MDB would have operations seamlessly integrated across domains. (For example, see Joint Staff 2012a, 2012b, and 2015.)

Critics dismiss MDB by arguing that it is old wine in a new bottle (Sinnreich 2016), but a more fundamental challenge is posed by the argument that categorizing future war by domain—especially the cyber domain—is neither logical nor practical. One observer notes that *domain* "contains some built-in assumptions regarding how we view warfare that can limit our thinking…[and] could actually pose an intractable conceptual threat to an integrated joint force" (Heftye 2017).

Some cynics see MDB's real purpose as a ploy to preserve force structure by returning land power to the tip of the spear in joint operations (Pietrucha 2016); others see it as requiring institutional reforms that are simply unattainable (Shattuck 2017). At one end of the spectrum is formation of separate departments for the space and cyber domains, and at the other end is creation of a single force that eliminates the independent service branches altogether (Davies 2017).

Running parallel to the ongoing MDB debate are distinct theater versions of the concept. Because practice trumps theory in the application of military force, how the MDB concept evolves will be strongly influenced by how the operating theaters find a way to employ its promise.

Given the multiple lenses through which the emerging MDB concept is viewed, the concept development challenge is to generate credible evidence that is relevant to decision makers from across the tactical-operational and conceptual-institutional divides.

Emerging Concept

An Army–Marine Corps white paper posits three interrelated components of an MDB solution: force posture, resilient formations, and converging joint force capabilities (U.S. Army Training and Doctrine Command 2017, 23). While these components provide a useful framework for institutional considerations of the concept, they do not capture some of the explicit and tacit implications of MDB's potential utility in a theater or joint campaign. To that end, we offer the following four attributes, derived from the current MDB concept:

1. Despite the *battle* suffix, MDB may have more to do with campaigns than tactical actions. Various descriptions point to an operational-level concept designed to maneuver friendly forces—and direct their kinetic and nonkinetic fires or effects—simultaneously across five domains.

2. Overmatch in one domain may trigger cross-domain multiplier effects that theater commanders can leverage to bypass, unhinge, and defeat an enemy. This, of course, works in both directions.

3. Cyber and space domains may become tomorrow's most valued battlespace given U.S. force dependence on the electromagnetic spectrum and satellite-enabled intelligence and communications. Continued development of sophisticated cyber weapons and means of their employment could exacerbate this trend.

4. MDB implies the need to reexamine the U.S. approach to joint command and control. The authorities needed by geographic combatant commanders across five domains will increasingly challenge the concept of boundaries and the traditional relationships used to conduct joint campaigns.

These attributes could be useful in developing a joint campaign of experimentation to better understand the MBD concept and to develop evidence for or against its military utility in the joint force. More aspirational than practical at this point, the concept needs to demonstrate that it is both more than the sum of its parts and better than the status quo.

Applying Existing Evidence

Examples of multidomain operations of the past provide insight into how cross-domain capabilities, applied primarily at the tactical level, can have outsize operational implications. Here we look at use of MDB in the Battle of Guadalcanal and the Falkland Islands War.

Battle of Guadalcanal

Shutler (1987) portrays U.S. operations against the Japanese in air, sea, and undersea domains (which he calls regimes) during the 1942 South Pacific campaign during World War II:

- U.S. land forces created an antiair warfare shield at Guadalcanal to protect the island Espiritu Santo from Japanese land-based aircraft. The mission then shifted from antiair warfare to enabling U.S. land-based aircraft to support subsequent island-hopping battles and the eventual reduction of the Japanese strongpoint on the island of Rabaul (Shutler 1987, 20)

- Preventing Japanese ground forces from reinforcing Guadalcanal required U.S. submarines, surface ships, and naval aviation to establish maritime and aviation "shields" that the Japanese were ultimately unable to penetrate (Shutler 1987, 23–25). This enabled U.S. Marines to preserve their tactical overmatch ashore. Finally, U.S. naval forces attacked and sank seven Japanese troop transports trying to reinforce Guadalcanal (Edson 1988, 51).

- A multiplier effect occurred once U.S. air operations began at Guadalcanal's Henderson Field. The Japanese fleet was largely restricted to night operations, partially because of U.S. airpower being projected from ashore and U.S. fleet interference with Japanese shipping during daylight hours. The implications went well beyond the tactical area of operations, marking the start of the U.S. island-hopping campaign.

Except for its value to the air domain, Guadalcanal had only marginal tactical utility in the Pacific theater. The airfield was the operational lynchpin that was denied to the enemy by adroit integration of multidomain (land, sea, and air) activities.

Falkland Islands War

The same multiplier effect occurred in a more modern campaign in the 1982 Falkland Islands War between the United Kingdom and Argentina. As the U.S. fleet had done in the South Pacific, the United Kingdom established maritime and antiair shields around the Falklands to isolate the objective area, protect amphibious operations of the Royal Navy and Royal Marines, and deny the ability for Argentina to reinforce its forces. The following examples of multidomain actions in the Falklands campaign indicate the effects these actions had on the campaign's outcome:

- A British submarine attacked and sank the Argentine cruiser *General Belgrano*, forcing the Argentine surface navy to remain inside its territorial waters for the duration of the campaign, which had a cross-domain effect (Woodward 1992, 246).
- The removal of *General Belgrano* relieved naval surface pressure on Great Britain's fleet in the Falkland littorals. This, in turn, allowed Royal Navy vessels to detect Argentine aircraft launched from the mainland and alert the British Task Force.
- A British amphibious raid on Pebble Island forced Argentine aircraft to fight at their maximum operating radius with reduced time on station and limited aerial refueling capability. This raid, conducted by special operations forces supported by naval gunfire, relieved Great Britain's amphibious fleet and embarked ground forces of their concerns about Argentine air superiority during the amphibious landing.

Conclusion

It is worth considering how multiple domains were integrated in these examples from the previous century. The process (including technical, conceptual, and instructional efforts) of integrating what at the time were new-fangled flying machines into the traditional warfighting domains of land and sea began decades before a mature concept evolved. It was not a straight line or a preordained outcome. The associated technologies and tactical concepts were leavened by decades of peacetime "experimentation" and wartime adaptation. The resulting capabilities for presenting an adversary with multiple, simultaneous dilemmas across domains changed the way the United States fights at both the tactical and operational levels of war.

Developing New Evidence

The second source of evidence for the viability of the MDB concept is through a rigorous campaign of joint experimentation—even as the specific capabilities are still being developed. In this context, *joint experimentation* indicates the exploration of ideas, assumptions, and crucial elements of nascent MDB capabilities. It covers a range of activities and should be undertaken in parallel with development of specific capabilities or tactical employment concepts.

Only through an experimentation campaign of iterative activities with learning feedback loops (including workshops, wargames, constructive and virtual simulation, and live field events) will evidence be sufficient to genuinely assess what it will take to realize, adapt, or abandon the MDB idea.

The results of such an assessment will help identify MDB similarities and differences between the theaters. It will also inform future doctrine, organization, training, materiel, leadership and education, personnel, facilities, and policy initiatives that must be addressed before MDB becomes a deployable set of capabilities.

The nature of *jointness* as practiced in a post–U.S. Joint Forces Command (USJFCOM) environment is a complex challenge. USJFCOM developed a generally top-down approach to joint concept development and experimentation, which often resulted in excessively large experiments. When USJFCOM was disestablished in 2011, joint concept development reverted to the Joint Staff (J7), whose time and resources for experimentation were more limited. Efforts to develop and experiment with new joint concepts in a bottom-up, collaborative effort. While this approach has many practical advantages over the top-down approach, it is not without challenges.

As the two historical case studies indicate, cross-domain overmatch and multiplier effects are often discovered and subsequently leveraged in the course of operations. Early discovery experimentation with some level of joint analysis and sponsorship is essential. Not only will such early experiments increase the capacity to do joint experimentation, but they can also help co-develop service branch concepts within a joint context.

One potentially lucrative approach would be to embark on a series of parallel joint discovery experiments designed to identify the specific characteristics, demands, and challenges associated with assessing the feasibility of MDB transcending theater-specific applications to serve as a more universal warfighting concept. Such a joint discovery experiment has historically been at the heart of military experimentation (Murray 2000).

The ability to use experimentation to explore the utility of emerging technologies and concepts is a force multiplier. Technology cannot be optimized until its impact on warfighting concepts and doctrine is fully appreciated.

Bridging the large gap between the envisioned operating environment in the MDB concept and the availability of validated models and simulations is a major challenge.

Any effort to explore MDB in a joint context must include an effort to integrate existing military department modeling and simulation tools (in the same bottom-up approach discussed here). This will help the departments operate across new domains in support of specific joint priorities.

It is time to subject the MDB concept to discovery experimentation. Discovery experimentation allows operators to interact with new or potential concepts and capabilities to explore their military utility—something that is not often supported through traditional studies or hypothesis-based experiments. Careful attention must be paid to the data that will provide solid evidence for the conclusions reached by conducting experiments. If carefully planned and executed, discovery experimentation could be a valuable tool.

References

Edson, J. J. 1988. "The Asymmetrical Ace," *Marine Corps Gazette*, April.

Joint Staff. 2012a. *Capstone Concept for Joint Operations*. Washington, DC.

———. 2012b. *Joint Operational Access Concept*. Washington, DC.

———. 2015. *Joint Concept for Rapid Aggregation*. Washington, DC.

Harris Jr., H. B. 2017. Statement of Admiral Harry B. Harris, Jr., U.S. Navy, Commander, U.S. Pacific Command before the Senate Armed Services Committee on U.S. Pacific Command Posture. April 27. https://www.armed-services.senate.gov/imo/media/doc/Harris_04-27-17.pdf.

Heftye, E. 2017. "Multi-Domain Confusion: All Domains Are Not Created Equal," *Real Clear Defense*. May 26. https://www.realcleardefense.com/articles/2017/05/26/multi-domain_confusion_all_domains_are_not_created_equal_111463.html.

Murray, W. 2000. *Experimentation in the Period between the Two World Wars: Lessons for the Twenty-First Century*. Alexandria, VA: Institute for Defense Analyses, November 2000)

Pietrucha, M. 2016. "No End in Sight to the Army's Dependence on Airpower." *War on the Rocks*. December 13. https://warontherocks.com/2016/12/no-end-in-sight-to-the-armys-dependence-on-airpower/.

Schmitt, J. F. 2002, *A Practical Guide for Developing and Writing Military Concepts*, Defense Adaptive Red Team Working Paper #02-4. McLean, VA: Hicks & Associates (December). http://www.navedu.navy.mi.th/stg/databasestory/data/youttasart/youttasarttalae/bigcity/United States/1.dart_paper.pdf.

Shattuck, A. J. 2017. "The Pipe Dream of (Effective) Multi-Domain Battle," *Modern War Institute at West Point* (March 28). https://mwi.usma.edu/pipe-dream-effective-multi-domain-battle/.

Shutler, P. D. 1987. "Thinking About Warfare." Marine Corps Gazette. November.

Sinnreich, R. H. 2016. "'Multi-Domain Battle': Old Wine in a New Bottle." *The Lawton Constitution*. October 30.

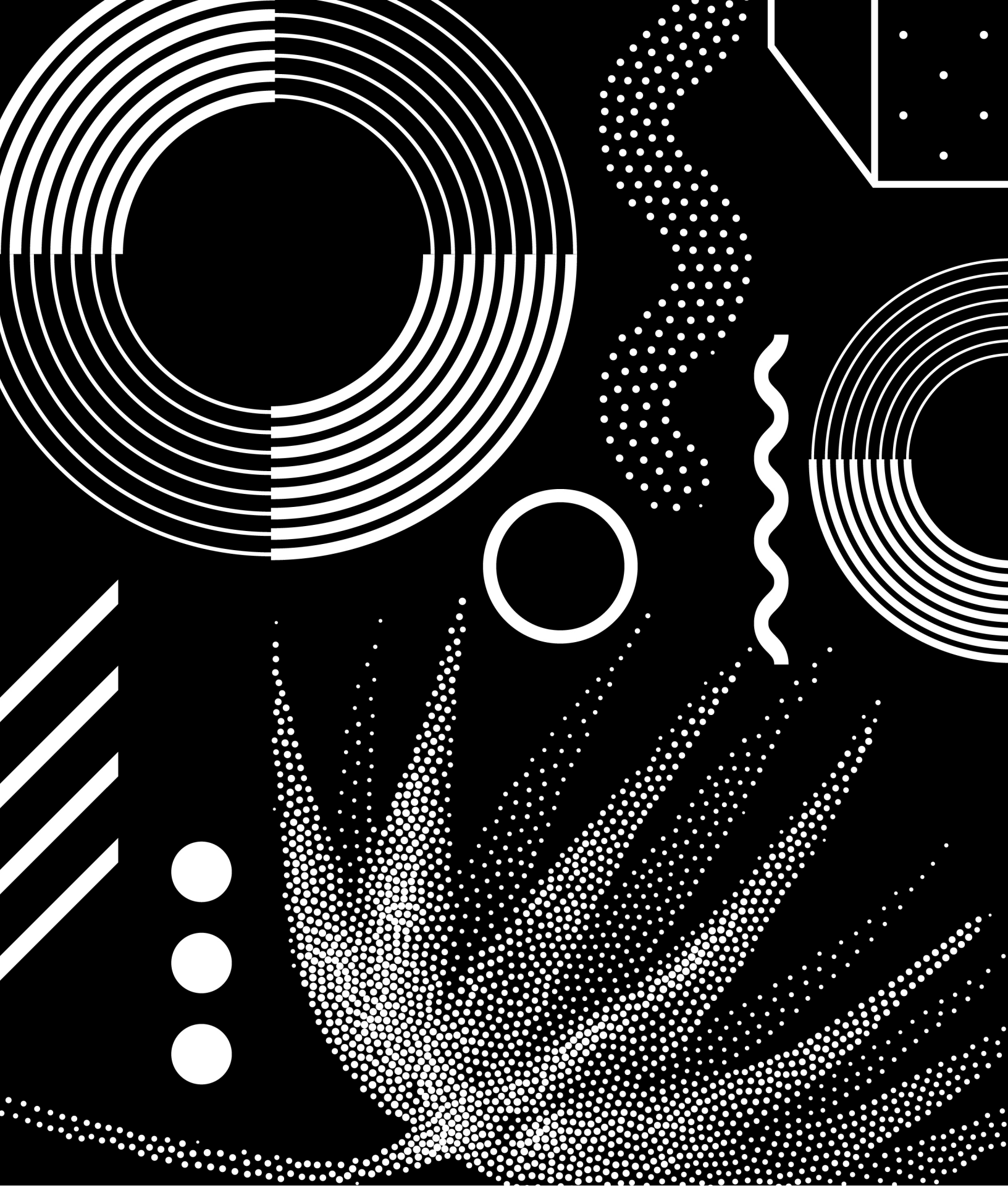
U.S. Army Training and Doctrine Command. 2017. "Multi-Doman Battle: Evolution of Combined Arms for the 21st Century: 2025–2040." December. https://www.tradoc.army.mil/wp-content/uploads/2020/10/MDB_Evolutionfor21st.pdf.

Woodward, S. 1992. *One Hundred Days: The Memoirs of the Falklands Battle Group Commander*. Annapolis, MD: U.S. Naval Institute Press.

About the Authors

Kevin Woods is the deputy director of the Joint Advanced Warfighting Division (JAWD) of IDA's Systems and Analyses Center, where he served as a member of the research staff since 2004. He holds a master's degree in national security and strategic studies from the Naval War College and a doctorate degree in history from the University of Leeds in the UK. Kevin was twice recognized for his contributions to Welch Award-nominated publications, first as a coauthor of the 2012 winner, *The Saddam Tapes: the Inner Workings of a Tyrant's Regime, 1978–2001*, and then as a coauthor of a 2015 finalist, *The Iran-Iraq War: A Military and Strategic History*.

Thomas Greenwood joined JAWD as a research staff member in 2016. Tom earned a master's degree from Georgetown University in government and national security studies. This marks the first time he is being recognized for his part in the Welch Award competition.



Scoping a Test That Has the Wrong Objective[1]

Thomas H. Johnson, Rebecca M. Medlin, Laura J. Freeman, and James R. Simpson

The Department of Defense test and evaluation community uses power as a key metric for sizing test designs. Power depends on many elements of the design, including the selection of response variables, factors and levels, model formulation, and sample size. The experimental objectives are expressed as hypothesis tests, and power reflects the risk associated with correctly assessing those objectives. Statistical literature refers to a different, yet equally important, type of error that is committed by giving the right answer to the wrong question. If a test design is adequately scoped to address an irrelevant objective, one could say that a Type III error occurs. We focus on a specific Type III error that test planners might commit to reduce test size and resources.

Despite the perceived increase in power and decrease in test resources that comes from reparameterization, we conclude that it is not a prudent way to gain test efficiency.

Introduction

Design of experiments is becoming more widely used when testing military systems to aid in planning, executing, and analyzing a test. In the planning phase, critical questions about the system under test are identified and the experimental objectives are set. These questions and objectives guide the development of the response variables, factors, and levels (Freeman et al. 2013).

Equally important in the planning phase is the evaluation of the experimental design. An assortment of measures is available to assess the adequacy of a design prior to data collection. Hahn et al. (1976) call these *measures of precision*, which include the standard error of predicted mean responses, standard error of coefficients, correlations metrics, and optimality criteria values. Measures of precision are affected by many aspects of the plan, including choice of factors and levels, assumed model form, combination of factor settings from run to run, and total number of runs in the experiment.

Power is an additional and widely used measure of precision, especially in the Department of Defense test and evaluation community. When the objective of an experiment—perhaps determining whether a new weapon system is better than an old system—is expressed as a hypothesis test, power informs risk associated with correctly assessing that objective. Because power increases with sample size, it is a useful metric for determining test length and test resourcing.

An adequate experiment requires sufficient power, but more importantly, it requires the hypothesis tests to reflect accurately the test objectives. If an experiment provides adequate power, but addresses the wrong objective, we might say an error is committed. Kimball (1957) refers to this Type III error *as an error of the third kind*, describing it as "the error committed by giving the right answer to the wrong question."

[1] Based on "On Scoping a Test That Addresses the Wrong Objective," *Quality Engineering* 31, no. 2 (November 2018): 230–239, https://doi.org/10.1080/08982112.2018.1479035.

Problem Statement

We focus on a Type III error that test planners might commit in an attempt to reduce test size and test resources. We provide an example that shows how reparameterization of the factor space (i.e., redefining it) from fewer factors with more levels per factor to more factors with fewer levels per factor fundamentally changes the hypothesis tests, which may no longer be aligned with the original objectives of the experiment.

Consider an experiment that plans to characterize a military vehicle's vulnerability against a particular type of mine. The test program has a limited number of vehicles and mines at their disposal to run a series of destructive tests to characterize the vehicle's vulnerability.

The measured response variable is the static deformation of the vehicle's underbody armor plate after interaction with the blast wave and ejecta from the buried charge. In other words, deformation is a direct measurement of the vehicle's armor shape change with respect to a reference point.

The engineering team believes that the non-uniform placement of structural elements, armor plates, and hardware on the vehicle's underbody may result in different deformations depending on where the mine is detonated. Thus, the team identifies six underbody detonation locations that may provide unique deformations (Figure 1). The program would like to be able to detect a difference in deformation between any two of the six detonation locations. The necessity for making these comparisons was driven by careful consideration of how mines affect vehicles

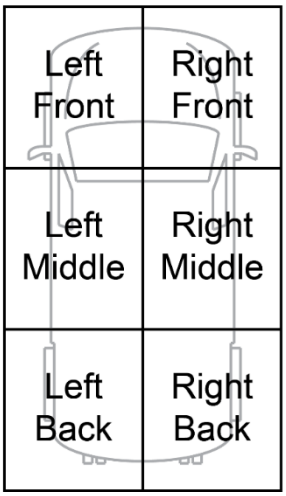


Figure 1. Vehicle underbody detonation locations

Intelligence analysts believe that the vehicle is most likely to encounter two variants of the mine type (variant A and B). The program would like to discover if mine variant significantly affects deformation. This discovery could be critical in informing military

tactics. Additionally, the engineering team would like to determine whether the effect of mine variant on deformation changes as detonation location changes. Thus, the test objectives are as follows: detect a difference in deformation between any two of the six detonation locations, detect a difference in deformation due to mine variant, and detect a change in the effect of mine variant as location changes.

After the test program agrees on these objectives, the test planner sets out to design the experiment. Given the established test objectives, the test planner constructs a twice-replicated factorial experiment using a two-level factor for mine variant and a six-level location factor. Based on the program's initial allocation of test resources, which accommodate 24 blast events, the test planner finds that the 38% power associated with testing the significance of the location factor and mine variant by location interaction is unsatisfactorily low (as we show in the case study).

Then, the test planner discovers a cost-cutting measure whereby reparameterization of the six-level factor into two factors, a two-level side factor and a three-level position factor (Figure 2), substantially increases power to 89% for the same number of test runs (also shown in the case study).

How could this happen? What information was lost? Is the tester still addressing the original objectives?

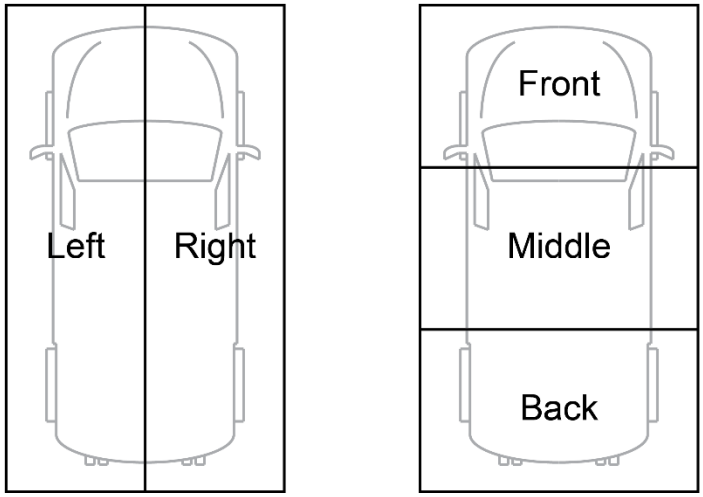


Figure 2. Reparameterization of the six-level location factor into a two-level side factor and a three-level position factor

The reparameterization of the factor space changed the hypothesis tests and thus the test objectives. The main effect hypothesis tests in the second proposal no longer correspond to detecting a difference in any two of the six detonation locations on the vehicle, nor does the inclusion of a three-way interaction term in the model correspond to detecting a change in the effect of mine variant as location changes. The tests on the side and position main effects only allow for the detection of a difference between the left and right side, and a difference between any two of the three positions.

The interaction between these two factors allows for detecting only the effect of one factor (for example, side) differing among the levels of the other factor (for example, position).

In the next section, we show how the test planner's reparameterization results in a different set of hypothesis tests and different test objectives. (The full version of this article provides additional details on the theory.)

Case Study Example

In this section, we investigate the two test design proposals. In each proposal, the test planner selects static deformation as the response variable of interest, which is normally distributed and measures the deformation of the vehicle's armor due to the mine blast. Each experimental run consists of a single detonation event, resulting in one measurement of deformation, measured in inches. We refer to the first proposal as the *Location Proposal*, which differentiates between the six detonation locations using a single factor. We refer to the second proposal as the *Side-by-Position Proposal*, which changes the six-level location factor into two separate factors. The two proposals include the same mine variant factor. Table 1 lists the factors and levels for each proposal, as well as model parameters for the analysis of variance (ANOVA).

Table 1. Test design factors, levels, and ANOVA model parameters, shown in parentheses, for the two proposals

	Location Proposal		Side-by-Position Proposal		
Factors	Mine	Location	Mine	Side	Position
Levels	A (m_1)	Left/Back (l_1)	A (m_1)	Left (s_1)	Back (p_1)
	B (m_2)	Left/Middle (l_2)	B (m_2)	Right (s_2)	Middle (p_2)
		Left/Front (l_3)			Front (p_3)
		Right/Back (l_4)			
		Right/Middle (l_5)			
		Right/Front (l_6)			

The test planner chooses a main effects plus two-factor interaction model for the Location Proposal and includes an additional three-factor interaction in the Side-by-Position Proposal. The ANOVA model for the Location Proposal is:

$$\mu_{ij} = \mu_0 + m_i + l_j + ml_{ij}, \quad i = 1, 2, \; j = 1, \ldots, 6$$

The ANOVA model for the Side-by-Position Proposal is:

$$\mu_{ijk} = \mu_0 + m_i + s_j + p_k + ms_{ij} + mp_{ik} + sp_{jk} + msp_{ijk},$$
$$i = 1, 2, \; j = 1, 2, \; k = 1, 2, 3.$$

The effect sizes for the power analysis are first defined in terms of the parameters of the ANOVA model and are then converted into regression model coefficients. In each proposal, the coefficient vector $\boldsymbol{\beta}$ is of size 12×1. Each coefficient vector can be

partitioned by model effects comprising main effects and two-way interactions. The coefficient vector for the Location Proposal is:

$$\boldsymbol{\beta} = [\beta_0 \quad \beta_m \quad \boldsymbol{\beta}_l^T \quad \boldsymbol{\beta}_{ml}^T]^T.$$

The coefficient vector for the Side-by-Position Proposal is:

$$\boldsymbol{\beta} = \left[\beta_0 \quad \beta_m \quad \beta_s \quad \boldsymbol{\beta}_p^T \quad \beta_{ms} \quad \boldsymbol{\beta}_{mp}^T \quad \boldsymbol{\beta}_{sp}^T \quad \boldsymbol{\beta}_{msp}^T\right]^T.$$

The design size for each proposal is the same, which is a duplicated full factorial experiment resulting in 24 runs. Let A_l and A_{sp} denote the single replicate full factorial model matrix for the Location Proposal and Side-by-Position Proposal, respectively, each of size 12×12. The model matrices $\boldsymbol{X}_l = [\boldsymbol{A}_l^T | \boldsymbol{A}_l^T]$ and $\boldsymbol{X}_{sp} = \left[\boldsymbol{A}_{sp}^T | \boldsymbol{A}_{sp}^T\right]$ are where the single replicate full factorial model matrices are constructed according to the following equations:

$$\boldsymbol{A}_l = (\boldsymbol{J}_2 \otimes \boldsymbol{J}_6 | \boldsymbol{\Delta}_2^T \otimes \boldsymbol{J}_6 | \boldsymbol{J}_2 \otimes \boldsymbol{\Delta}_6^T | \boldsymbol{\Delta}_2^T \otimes \boldsymbol{\Delta}_6^T)$$

$$\boldsymbol{A}_{sp} = (\boldsymbol{J}_2 \otimes \boldsymbol{J}_2 \otimes \boldsymbol{J}_3 | \boldsymbol{\Delta}_2^T \otimes \boldsymbol{J}_2 \otimes \boldsymbol{J}_3 | \boldsymbol{J}_2 \otimes \boldsymbol{\Delta}_2^T \otimes \boldsymbol{J}_3 | \boldsymbol{J}_2 \otimes \boldsymbol{J}_2 \otimes \boldsymbol{\Delta}_3^T | \boldsymbol{\Delta}_2^T \otimes \boldsymbol{\Delta}_2^T \otimes \boldsymbol{J}_3 | \boldsymbol{\Delta}_2^T \otimes \boldsymbol{J}_2 \otimes \boldsymbol{\Delta}_3^T | \boldsymbol{J}_2 \otimes \boldsymbol{\Delta}_2^T \otimes \boldsymbol{\Delta}_3^T | \boldsymbol{\Delta}_2^T \otimes \boldsymbol{\Delta}_2^T \otimes \boldsymbol{\Delta}_3^T)$$

After defining the models, the test planner is ready to calculate the power associated with assessing the test objectives. Recall, the test objectives are to detect a difference in deformation between any two of the six detonation locations, detect a difference in deformation due to mine variant, and detect a change in the effect of mine variant as location changes. Following typical procedures using statistical software, the test planner assumes that calculating power for main effects and interactions in each proposal addresses the test objectives.

The test planner constructs the hypothesis tests using the equation of the general linear hypothesis test:

$$H_0: \boldsymbol{C\beta} = \boldsymbol{t} \quad , \quad H_1: \boldsymbol{C\beta} \neq \boldsymbol{t},$$

where $\boldsymbol{\beta}$ is the $(k+1)\times 1$ vector of coefficients, and k is the number of coefficients in the model excluding the intercept. The $\boldsymbol{t}$ vector specifies the hypothesized constant value of the effect tested and has size $q \times 1$. In practice, $\boldsymbol{t}$ is almost always set to $\boldsymbol{0}$. The $\boldsymbol{C}$ matrix isolates the coefficients or combination of coefficients tested and has size $q \times (k+1)$, where $q \leq k+1$. In other words, q is the number of simultaneous hypotheses being tested. The power of the hypothesis test is equal to

$$P(F \geq \hat{F}_\alpha) = 1 - \tilde{F}(\hat{F}_{\alpha,q,n-k-1}, q, n-k-1, \lambda),$$

where $\hat{F}$ is the F central quantile function that provides the critical F value evaluated at the $(1 - \alpha)$th quantile, and $\tilde{F}$ is the non-central F distribution function evaluated at the critical F value.

Table 2 shows the $\boldsymbol{C}$ matrix associated with each hypothesis test on the main effects and interactions for the two proposals. The hypothesis tests in this table assume that $\boldsymbol{t}$ is equal to $\mathbf{0}$.

Table 2. Hypothesis Test for Each Proposal

Location Proposal		Size-by-Position Proposal	
Hypothesis	C	Hypothesis	C
$H_0: \beta_m = 0$	$\begin{bmatrix} 0 & 1 & \mathbf{0}_{1\times 10} \end{bmatrix}$	$H_0: \beta_m = 0$	$\begin{bmatrix} 0 & 1 & \mathbf{0}_{1\times 10} \end{bmatrix}$
$H_0: \boldsymbol{\beta}_l = \mathbf{0}$	$\begin{bmatrix} \mathbf{0}_{5\times 2} & \boldsymbol{I}_5 & \mathbf{0}_{5\times 5} \end{bmatrix}$	$H_0: \beta_s = 0$	$\begin{bmatrix} \mathbf{0}_{1\times 2} & 1 & \mathbf{0}_{1\times 9} \end{bmatrix}$
$H_0: \boldsymbol{\beta}_{ml} = \mathbf{0}$	$\begin{bmatrix} \mathbf{0}_{5\times 7} & \boldsymbol{I}_5 \end{bmatrix}$	$H_0: \boldsymbol{\beta}_p = \mathbf{0}$	$\begin{bmatrix} \mathbf{0}_{2\times 3} & \boldsymbol{I}_2 & \mathbf{0}_{2\times 7} \end{bmatrix}$
		$H_0: \beta_{ms} = 0$	$\begin{bmatrix} \mathbf{0}_{1\times 5} & 1 & \mathbf{0}_{1\times 6} \end{bmatrix}$
		$H_0: \boldsymbol{\beta}_{mp} = \mathbf{0}$	$\begin{bmatrix} \mathbf{0}_{2\times 6} & \boldsymbol{I}_2 & \mathbf{0}_{2\times 4} \end{bmatrix}$

Next, the test planner defines the effect sizes. An effect size is defined for each hypothesis test and represents the value of the coefficients tested assuming H_0 is false and H_1 is true. Using the unified approach (see the original article for details), the test planner first defines the effect size in terms of parameters of the ANOVA model and then converts those parameters into coefficients for the regression model.

The test planner selects an effect size of 1 inch. In the Location Proposal, letting $\boldsymbol{m}$, $\boldsymbol{l}$, and $\boldsymbol{ml}$ be the vectors of the ANOVA model parameters for m_i, l_j, and ml_{ij}. The effect size definition implies that the range of $\boldsymbol{m}$, $\boldsymbol{l}$, or $\boldsymbol{ml}$ is equal to 1 inch. Using a similar approach in the Side-by-Position Proposal, the effect size implies the range of ***m,s,p,ms,mp,sp,or msp*** is equal to 1 inch.

Part two of the unified approach requires a search among the candidate set of effect sizes for the particular effect size that yields minimum power. The candidate search is unnecessary in this case study because the test designs are completely balanced. For balanced designs, power for each effect size within a candidate set is identical. It is only with unbalanced designs that the candidate search is necessary to provide a unique estimate of power.

The test planner defines the individual ANOVA model parameter vectors to satisfy part one of the unified approach. To illustrate for l in the first proposal, the test planner arbitrarily chooses the configuration shown below. (Another configuration could have been selected because each gives the same power because the design is balanced. See Table 2 in the original article for all design configurations.)

$$\boldsymbol{l} = [1/2 \quad 0 \quad 0 \quad 0 \quad 0 \quad -1/2]^T$$

The equation that converts the ANOVA model parameter to the regression model coefficients is

$$\boldsymbol{\beta}_l = (\boldsymbol{\Delta}_6^T)^{-1}\boldsymbol{l} = [1/2 \quad 0 \quad 0 \quad 0 \quad 0\,]^T.$$

The conversion for two- and three-factor interactions follows a similar process. Take, for example, the ***ml*** interaction in the first proposal. The test planner arbitrarily chooses the configuration, and the ANOVA model parameter vector is

$$\boldsymbol{ml} = [1/2 \quad 0 \quad -1/2 \quad 0 \quad 0 \quad 0 \quad -1/2 \quad 0 \quad 1/2 \quad 0 \quad 0 \quad 0]^T.$$

The equation that converts the ANOVA model parameter to regression model coefficients is

$$\boldsymbol{\beta}_{ml} = (\boldsymbol{\Delta}_2^T \otimes \boldsymbol{\Delta}_6^T)^{-1}\boldsymbol{ml} = [1/2 \quad 0 \quad -1/2 \quad 0 \quad 0]^T.$$

The test planner repeats this calculation process for each hypothesis test.

Having defined the effect sizes, the test planner calculates power. The assumed confidence level is 95% ($\alpha = 0.05$). Next, the test planner calculates the non-centrality parameter λ (see Johnson et al., (2018) for more detail). In the non-centrality parameter equation, σ is the root mean-squared error, representing the overall "noise" in the experiment. Based on observations from previous testing that had been executed under similar conditions, the test planner assumes σ is equal to 0.5 inches, which implies a "signal-to-noise" ratio equal to 2 (recall the effect size or "signal" is 1 inch). Finally, the test planner calculates power. The numerator degrees of freedom, non-centrality parameter, and the power for each hypothesis test and proposal are shown in Table 3.

Table 3. Power for Each Proposal

Location Proposal				Side-by-Position Proposal			
Hypothesis	q	λ	Power	Hypothesis	q	λ	Power
$H_0: \beta_m = 0$	1	24	.99	$H_0: \beta_m = 0$	1	24	.99
$H_0: \boldsymbol{\beta}_l = 0$	5	8	.38	$H_0: \beta_s = 0$	1	24	.99
$H_0: \boldsymbol{\beta}_{ml} = 0$	5	8	.38	$H_0: \boldsymbol{\beta}_p = 0$	2	16	.89
				$H_0: \beta_{ms} = 0$	1	24	.99
				$H_0: \boldsymbol{\beta}_{mp} = 0$	2	16	.89
				$H_0: \boldsymbol{\beta}_{sp} = 0$	2	16	.89
				$H_0: \boldsymbol{\beta}_{msp} = 0$	2	16	.89

After completing the power calculations for both proposals, the test planner prepares briefing slides and presents the results to a room of non-statistically oriented engineers, managers, and military personnel. The premise of the results, although omitted from the presentation, is that both proposals address the test objectives. Without any discussion about the connection between the hypothesis tests and test objectives, the test planner quickly arrives at the power results for each proposal. The choice of proposal becomes clear.

For the same number of runs, the Side-by-Position Proposal provides no less than 89% power for all main effects and interactions, compared to the 38% power for the

Location Proposal. Impressed by the savings in test resources, the test program agrees to the Side-by-Position Proposal and commits a Type III error.

Conclusion

The test planner mistakenly believed that the main effects and interaction hypothesis tests in the Side-by-Position Proposal addressed the test objectives. In truth, the tests on the main effects allow for the detection of a difference between the left and right sides, and a difference between any two of the three positions, respectively. Neither of these hypothesis tests, nor the interaction test, allows for the detection of a difference between any two of the six locations on the vehicle. The perceived high power associated with this proposal led the team to believe the test was adequately resourced. Despite the perceived increase in power and decrease in test resources that comes from reparameterization, we conclude that it is not a prudent way to gain test efficiency.

In this particular example, the high power was not used to argue for a smaller test design, but this could and does happen in practice. The proper decision would have been to select the Location Proposal, and recognize that the only solution to adequately assessing the test objectives is to add more experimental runs.

The case study highlights potential negative consequences of redefining a factor space, but the approach should not be completely discredited. If the test planner and the test program had renegotiated the objectives, such that the objectives aligned with the hypothesis tests in the Side-by-Position Proposal, reparameterization could have been a shrewd cost-saving strategy. It is the lack of careful planning that leads to a Type III error.

The unified effect size approach, coupled with the ability to convert effect sizes between the ANOVA and regression model formulations, facilitated the diagnosis of the Type III error. These tools were useful for comparing reparameterizations of a factor space for a fixed experimental design, but they can also be useful for comparing competing designs that have different sample sizes or different degrees of imbalance. The consistent estimate of power from the unified approach enables such comparisons.

References

Freeman, L. J., A. G. Ryan, J. L. K. Kensler, R. M. Dickinson, and G. G. Vining. 2013. "A Tutorial on the Planning of Experiments." *Quality Engineering* 25, no. 4: 315–332. https://doi.org/10.1080/08982112.2013.817013.

Hahn, G. J., W. Q. Meeker Jr, and P. I. Feder.1976. "The Evaluation and Comparison of Experimental Designs for Fitting Regression Relationships." *Journal of Quality Technology* 8, no. 3: 140–157. https://doi.org/10.1080/00224065.1976.11980735.

Kimball, A. W. 1957. "Errors of the Third Kind in Statistical Consulting." *Journal of the American Statistical Association* 52, no. 278: 133–142. https://doi.org/10.1080/01621459.1957.10501374.

About the Authors

Thomas Johnson is a member of the research staff of the Operational Evaluation Division (OED) of IDA's Systems and Analyses Center. Tom joined IDA in 2011. Both his master's degree and his doctorate degree are in aerospace engineering from Old Dominion University. Tom was twice recognized as a coauthor of finalist Welsh Award publications, for "Power Approximations for Generalized Linear Models Using the Signal-to-Noise Transformation Method" in 2018 and for "A Comparison of Ballistic Resistance Testing Techniques in the Department of Defense" in 2015.

Rebecca Medlin joined IDA in 2015 as a member of OED's research staff. She earned both her master's and doctorate degrees in statistics from Virginia Polytechnic Institute and State University. Rebecca is being recognized for the first time this year as a finalist coauthor in the Welch Award competition.

Laura Freeman was first a research staff member and then an assistant director in OED during her tenure at IDA, which ended in 2019. She earned her master's and doctoral degrees in statistics, both from Virginia Polytechnic Institute and State University. Laura was a coauthor of two other Welch Award finalist publications: "Power Approximations for Generalized Linear Models Using the Signal-to-Noise Transformation Method" in 2018 and "A Comparison of Ballistic Resistance Testing Techniques in the Department of Defense" in 2015.

James Simpson, an OED consultant, holds a doctorate in industrial engineering from Arizona State University. He was recognized as a coauthor of the 2018 Welch Award finalist publication "Power Approximations for Generalized Linear Models Using the Signal-to-Noise Transformation Method."

Past Welch Award Winners

2018

"Deterrence Is Not a Credible Strategy for Cyberspace (and What Is)"

Orbis

Michael P. Fischerkeller and Richard J. Harknett

2017

"Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review"

Review of Educational Research

J. A. Kulik and J. D. Fletcher

2016

"Mining Measured Information from Text"

Proceedings of the 38th International ACM CIGR Conference on Research and Develoment in Information Retrieval

A. Maiya, D. Visser, and A. Wan

2015

"Visible Signatures of Hypersonic Reentry"

Journal of Spacecraft and Rockets

J. Teichman and L. Hirsch

2014

"Mixed Models Analysis of Radar Residuals Data"

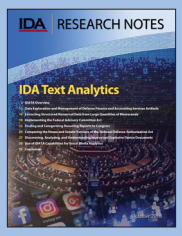
IEEE Access

C. Gaither, D. Loper, C. Jackson, and J. Pozderac

Past Issues

2019 | Challenges in Cyberspace: Strategy and Operational Concepts

2018 | Challenges in Cyberspace: The Human Dimension



2018 | IDA Text Analytics

2018 | Multidisciplinary Research for Securing the Homeland

2016 | Acquisition, Part 2

IDA | RESEARCH NOTES

4850 Mark Center Drive • Alexandria, VA 22311-1882
ida.org

@IDA_org