Proceedings of the 2014 Winter Simulation Conference
A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, eds.

USING SIMULATION AND OPTIMIZATION TO INFORM ARMY FORCE STRUCTURE REDUCTION DECISIONS

Jason Southerland

Center for Army Analysis
6001 Goethals Road
Fort Belvoir, VA 22307, USA

Andrew Loerch

George Mason University
4400 University Drive MSN 4A6
Fairfax, VA 22030, USA

ABSTRACT

Given constraints dictated by the current fiscal environment, the Army has been directed to reduce its total personnel strength from around 1.05 million across the active duty, Army National Guard, and Army Reserve, to a maximum of around 980 thousand personnel. In particular, the active duty Army will have to drawdown to around 450 thousand personnel. In this paper we discuss a methodology the Army is using to help inform decisions about how to execute this drawdown. We describe a simulation-based optimization that identifies potential cuts to a large subset of the active duty Army's total strength.

1 INTRODUCTION

Given the current fiscal environment, the United States Army is being directed to reduce its overall size. This overall size reduction will be achieved by a reduction in both individual personnel and collective units. The portfolio of collective units is known as the force structure. This paper discusses a simulation based optimization methodology the Army is using to inform the reduction of a subset of its force structure, specifically, the active duty, rotational, operating force.

At the time of the writing of this paper, the Army has not made any final decisions regarding the shape of its reductions. And any such results would likely be sensitive information, until shared with Congress and announced publicly. As such, this paper does not aim to present any modeling results. The focus of this paper will be the application of an existing simulation model, coupled with a binary integer program, to broaden the existing base of analytic support to a decision the Army regularly revisits.

The rest of this paper is organized into four additional sections. Section 2 provides relevant background information, including an overview of the mechanism through which the Army makes force structure decisions. Section 3 describes the methodology. Section 4 provides a discussion of key takeaways from this modeling effort and section five summarizes this paper.

2 BACKGROUND AND RESEARCH OBJECTIVE

In this section we describe the contextual information necessary to understand our application of simulation and optimization to the problem of informing Army force structure reductions. In particular, we orient the reader to the organization of the Army; the model describing the mechanism the Army uses to prepare units for missions; and a simulation model the Army uses to simulate the application of this model to collections of missions that the Army may be asked to execute in the future.

2.1 Army Force Structure

The total strength of the Army is distributed across three components—the Active Component (AC); the Army National Guard (ARNG); and the Army Reserve (USAR). Each of these three components is

978-1-4799-7486-3/14/$31.00 ©2014 IEEE

composed of an operating force (OF), which is, generally speaking, responsible for executing military missions; and a generating force (GF), that is responsible for recruiting, training, preparing, and sustaining the operating force for those missions. In addition, the active component and the Army National Guard have a portion of their total strength set aside for persons in training, in between assignments, etc., in an account known as ‘Transients, Trainees, Holdees, and Students,’ or TTHS.

Taken collectively, the OF, GF, and TTHS describe the aggregate number of personnel authorized for the Army. However, these distinctions are much too vague to inform decisions. If reductions to the Army are described in these terms, there are still details required to determine, what, exactly, is to be reduced. In particular, the OF and GF are composed of a number of units, each unit adhering to a specific design. GF units are generally described in a Table of Distribution and Allowances (TDA). In the case of the OF, these designs are known as Tables of Organization and Equipment (TOEs). Each TOE is referred to by a unique Standard Requirements Code (SRC), with each SRC representing a single unit type. In this paper, we use SRC and unit type interchangeably.

In both the TDA and TOE cases, the designs specify the collection of equipment a unit is authorized and the numbers of personnel by rank and specialty resident in the unit. Depending on the level of aggregation at which these TDAs and TOEs are viewed, the complexity of the Army structure ranges from a few hundred unit types to a few thousand.

Individual TDAs and TOEs define what might properly be called micro-force structure. In that sense, macro-force structure is the number of units with each design that reside within the Army. Force structure at the macro-level is essentially the portfolio of capabilities resident within the Army.

Before we discuss the process the Army uses to manage the preparation of units for missions, there is one more critical distinction in Army force structure that bears on our problem—OF units may be either rotational or non-rotational. Rotational units are those units managed under standard Army Force Generation rules, and non-rotational units are those units managed using different mechanisms. Individual SRCs may have any combination of rotational and non-rotational units.

2.2 Army Force Generation

Army Force Generation (ARFORGEN) is the process through which the Army prepares most units for missions. The ARFORGEN model describes three discrete phases through which units progress prior to deployment—Reset, Train/Ready, and Available. Specific activities occur in each phase, building additional mission readiness as units progress through the phases. As such, units in the Available phase of their cycle are at higher levels of readiness than those in Train/Ready, etc. Units in their Available phase will either deploy to execute a mission or return to the Reset phase. While the Army prefers to only deploy Available units, in certain situations, the Army will deploy less ready units in the Train/Ready phases of their readiness cycle.

ARFORGEN cycles are defined by a number of parameters. These parameters include, but are not limited to, maximum deployment length, commonly referred to as boots on the ground (BOG) time; total cycle length; and transition times between phases. An additional parameter describes the absolute minimum time into a unit’s cycle a unit may deploy in emergency circumstances.

As a matter of convenience, cycles are often referred to by a “Dwell:BOG ratio”, where dwell is (heuristically) calculated as the difference between total cycle length and BOG. For example, in a cycle intended to last 24 months with a 9-month maximum deployment, the Dwell:BOG ratio would be 15:9. As a matter of practice, this short-hand ratio is seldom achieved. Units deploy sooner or later in their cycle, resulting in the former case in a realized ratio less than indicated by short-hand notation, and in the latter case in a realized ratio greater than indicated by the short-hand notation.

2.3 Total Army Analysis

Total Army Analysis (TAA) is the Army’s analytical venue for determining its force structure at the macro level. The objective of TAA is to determine and justify its force structure consistent with

Department of Defense guidance and other Army plans (United States Army 1995). The force structure decision resulting from TAA provide the basis for nearly all of the Army's annual budget submission. In a typical TAA, conducted annually, the Army will be directed to grow certain capabilities in certain quantities. Given that end-strength is, by law, constrained, one task in TAA is to determine how to offset this directed growth by divesting of capacity in other capabilities, generally at a resolution that considers between 250 and 350 unit types, a rather complex task. Another task in TAA is to improve the overall ability of the Army to execute missions. One mechanism for achieving this is to increase the number of units in some SRCs, beyond those numbers dictated by directed growth. Given the zero-sum nature of Army force structure decisions, this growth must also be offset by reductions elsewhere.

In recent TAAs, the Army has used output data from the Marathon model to help inform force structure decisions, in particular with respect to the rotational operating force. This simulation model is described below. The data from Marathon help focus the task on SRCs with relatively greater capacity as candidates for reduction and on those with relatively less capacity as candidates for growth, where capacity is measured in terms of ability to satisfy demand at some realized dwell to BOG ratio. For example, an SRC whose inventory can satisfy all demands at a 2:1 ratio has greater capacity than an SRC that requires a 1:1 dwell to BOG ratio to satisfy all demands.

To further manage the complexity of the force structure decision, the force structure is divided into classes of SRC, for example, infantry and aviation units. Each class of SRC is considered by a separate resourcing panel. Each of these resourcing panels receive class-specific guidance in the shape of total number of personnel growth or reduction, which comes from Senior Army Leadership and is informed by strategy and policy. Based on this guidance, each resourcing panel produces a recommended force structure for all SRCs considered in that panel across all three components.

As such, the approach to TAA can be described as a series of sub-problems. The high-level problem determines guidance with respect to cuts and growth for each resourcing panel. Each resourcing panel, in turn, determines how to meet its class-specific guidance. The approach we describe in section three takes a more holistic approach to informing these decisions.

2.4 Marathon

Marathon is a simulation model that applies ARFORGEN rules to units in order to fill demands. Inputs include demand; supply, or Army units that can satisfy demands; and policy, which governs state transitions of units in supply along a spectrum from undeployable to deployable. The following discussion describes each of these inputs and the internal logic Marathon uses to match supply to demand.

2.4.1 Demand

In Marathon, we represent individual military missions as demands. These demands are pre-specified in data. Demands are characterized by a start time, duration, and a force list. Each force list details the capabilities required for the specified mission. A collection of demands arrayed over time is known as a demand signal. Thus a demand signal is essentially a time series of SRC demands over the duration of the simulation horizon. Typical demand signals detail demands over a horizon of between ten and fifteen years.

The Army uses Office of the Secretary of Defense (OSD) approved demands to build their demand signals, and obtain OSD approval for any demand signal that is created prior to using it for force structure shapig decisions.

2.4.2 Supply

We represent the inventory of various capabilities as supply. Supply is characterized by a unique SRC description, the number of each SRC in the inventory, and the initial ARFORGEN cycle location. Supply can be specified either at the SRC level of detail, or at the individual unit level of detail.

2.4.3 Policy

Policy is the mechanism by which we govern unit behavior in the simulation and can change across time. Atomic policies are time invariant while composed policies govern the use of atomic policies over specified time intervals. Policies are a parametric representation of ARFORGEN, describing the three parameters we discussed in section 2.2—the minimum amount of non-deployed time before a unit is deployable; the maximum deployment length; and the maximum amount of non-deployed time a unit may accumulate in a single cycle before returning to reset. For example, an atomic policy might specify, using the short-hand policy notation, a 1:1 dwell to BOG ratio with a 9-month deployment, while a composed policy might specify a 1:1 policy for the first three years of the simulation, and 2:1 for the next five years, and a 3:1 for the remainder of the simulation. By applying policies, Marathon determines when units are deployable or not.

2.4.4 Internal Logic

Marathon assigns deployable units of supply to demand. At each event step, Marathon uses a myopic, greedy heuristic to perform this assignment. Unfilled demands are rank-ordered by priority. The internal logic attempts to fill the highest priority demands first, assigning deployable units, progressively exhausting the priority list, until either all demands have been filled or there are no remaining deployable units to assign to demands.

2.5 Research Objective

Our research is intended to produce a list of SRC inventory levels for all Active Component, rotational, operating force SRCs. This list is not intended to provide the answer to the force structure reduction problem. Instead, we intend this effort to provide information to inform the various resourcing panels. The methodology must be flexible enough to allow us to provide responsive feedback to decision makers and adaptable to incorporating greater analytic capabilities.

3 METHODOLOGY

As discussed previously, the methodology we developed is a simulation-based optimization. In this section, we describe the setup of the Marathon simulation; output data processing; and the formulation of the optimization model.

3.1 General Approach

Our approach is best summarized as measuring SRC performance at various potential inventory levels and using these data to find an optimal portfolio of SRC inventories. We discuss the meaning of “optimal portfolio” in section 3.4. To measure the performance at various inventory levels, we performed numerous Marathon simulations. We assumed that all capabilities were independent, i.e. no SRC could substitute for another SRC, and we thus could vary inventory quantities for all capabilities in a single instance of the simulation. Relaxing this independence assumption would substantially increase the complexity of the problem. We discuss this issue in section 4.

3.2 Marathon Simulation Setup

Recall that our research objective is focused on active duty force structure reductions given a specified structure for the Army National Guard and Army Reserve. In light of these our approach with the simulation was to apply a consistent, realistic policy for the ARNG and USAR, and to develop a less constrained policy for the active duty units.

This developed policy, known as "MaxUtilization" applies a finite maximum deployment length, but has no minimum dwell before deployment and no maximum dwell before reset. The practical effect of this policy is that active duty units can deploy whenever needed to satisfy demand. While this policy is impractical in reality, with some exceptions, it allows us to measure the capacity of the active duty force as a single number, the realized dwell to BOG ratio. Given that our research objective is to determine how to reduce the overall strength of the Active Component by determining which SRCs to reduce, it makes sense to reduce SRCs with greater capacity, as measured by the realized dwell to BOG ratio, relative to demands.

The exception to this utility occurs in situations where there is insufficient inventory to satisfy all demands. In such cases, the capacity is best measured by applying a more constrained policy, and measuring both the realized Dwell to BOG ratio and the number, timing, etc. of missed demands.

We used a single demand signal consistent with defense planning constructs. For a discussion of these constructs, see the 2010 Quadrennial Defense Review Report (Office of the Secretary of Defense 2010). As discussed in section 3.1, we fixed the ARNG and USAR inventories and varied the active duty inventories in a number of simulations over SRC quantity ranges as determined through a method we describe in section 3.3.

3.3 Determining Quantity Ranges for the Simulation

To determine SRC quantity ranges for the simulation, we calculated what is known colloquially as "ARFORGEN algebra," or static analysis. ARFORGEN algebra calculates the number of units of some type that are deployable at any time under ideal conditions, where all units are spaced in equal intervals across their ARFORGEN life cycle. In reality, this equal interval conditions rarely, if ever, holds. Marathon was developed to analyze force structure in precisely these less-than-ideal conditions. Nonetheless, the calculations are useful to inform the simulation, by providing bounds on SRC quantity ranges, for example.

The desired outcome of an ARFORGEN algebra calculation is a determination of the number of units of each type deployable at one time. To perform this calculation, we determine what proportion of a unit's ARFORGEN cycle a unit is deployable and multiply this proportion by the number of units in the inventory with the same cycle. For example, if a unit type has 18 units with a 6-year cycle and is deployable for one year in that cycle, $\frac{1}{6} * 18 = 3$ units would be deployable.

For our purposes we needed to determine what the ARFORGEN algebra contribution would be for the ARNG and USAR units. We calculated these quantities for all capabilities and subtracted these quantities from specified scenario requirements to determine what quantity the active duty units would have to provide in order to satisfy all scenario requirements. We then used similar algebraic calculations, $Quantity = \frac{Required}{Proportion\ Available}$, to determine quantities for each active duty SRC. By comparing these quantities to planned inventory quantities, with the added constraint of only reducing or holding constant individual SRC inventories, we were able to determine the range of quantities over which each SRC should be simulated.

3.4 Output Data Processing

In order to provide data to support our optimization model, we had to develop some data processing routines on existing Marathon output data for each of our numerous simulations. Marathon provides a broad collection of standard output, including, but not limited to, demand satisfaction over time, by scenario and SRC, and a complete history of individual unit deployments.

The unit deployment data details, for each deployment, the name of the unit; the SRC type of that unit; and the start and completion time of the deployment. We used these data to develop, for each SRC type, the minimum realized dwell to BOG ratio over any four-year period in the simulation horizon. To calculate this minimum, we first determined, for each t in the simulation horizon the total number of days the active component units of each SRC spent deployed in the time interval, [t,t+1459] where t is measured in days. We then determined the total number of "unit-days" the SRC could have in the four year window by taking the product of the SRC inventory and the length of the period. For example, a SRC with three active component units would have 3*1,460=4,380 "unit-days." The dwell to BOG ratio is then calculated as $\frac{UnitDays-BOG}{BOG} = \frac{UnitDays}{BOG} - 1$. The minimum ratio is simply the minimum realized ratio and represents the most active four year period experienced by the SRC.

3.5 Optimization Model

Having run multiple simulations in Marathon and collating the data, we then developed an optimization model. Given that we used a "MaxUtilization" policy for the active duty units and our task was to find a collection of force structure reductions for these active duty units, it made sense that we should focus our optimization on some sort of dwell to BOG objective. As a matter of practice, stated ARFORGEN policies specify a rotation rate goal, stated as a dwell to BOG rate. Violating this rate equates to soldiers experiencing more time away from home than is desirable. Thus, our objective was to minimize the total number of soldiers experiencing deployment accumulation in violation of the goal, weighted by the magnitude of the violation.

To achieve this minimization, the optimization, at its most basic, asserts two constraints—the collection of SRC decisions must meet a total personnel reduction constraint within some parametric tolerance; and only one quantity can be chosen for each SRC.

Our formulation included two sets of indices—s, indicating the set of SRCs; and o, indicating the set of options. We specified a total personnel reduction target, r, and a tolerance, e, within which this tolerance must be met. The following data were specified for each SRC independent of any reduction options—$Goal_s$, indicating the desired dwell to BOG ratio; Q_{0s}, indicating the currently planned inventory for SRC s; and $Size_s$, indicating the number of personnel in a single unit of SRC s. Our pre-simulation analysis and Marathon data processing provided us the following three data sets—Q_{so}, indicating the quantity of SRC s associated with option o; DB_{so}, indicating the realized dwell to BOG ratio, as described in section 3.4, for the quantity of SRC s and option o; and , U_{so},where , $U_{so} = max\ (0, Goal_s - DB_{so})$, and indicates the extent to which the realized ratio violates the goal.

Our decision variable was binary,x_{so} , indicating a choice of for Q_{so}SRC s, if $x_{so} = 1$.

As described above, our objective was to minimize the total number of soldiers experiencing deployment accumulation in violation of the goal, weighted by the magnitude of the violation. The following objective function captures this dynamic— $\min_x \sum_s \sum_o x_{so} * Q_{so} * Size_s * U_{so}$. In this equation, $Q_{so} * Size_s$ indicates the number of soldiers of SRC s in the force, and U_{so} describes the magnitude of the goal violation.

We formulated three constraints on the problem. First, we could only choose one option for each SRC—$\sum x_{so} = 1, \forall s$. And the collection of decision had to adhere to a total reduction target, within tolerance—$r - e \leq \sum_s Size_s * (Q_{0s} - \sum_o x_{so} * Q_{so}) \leq r + e$.

4 FINDINGS

As we discussed in the introduction, any results from the application of this methodology would be sensitive and likely not able to be discussed in public. As such, we focus our discussion in this section on lessons we learned in the course of developing the methodology and how we have been able to provide value at each stage in that development.

4.1 Complexity and Scope

As mentioned in section 3.1, our methodology leveraged an independence, or non-substitution, assumption. This assumption greatly reduced the complexity of our problem in two ways—reducing the number of simulation iterations that had to be executed and simplifying the formulation of the optimization.

From a simulation perspective, the non-substitution assumption greatly reduced the number of Marathon iterations we had to execute. While we cannot address specific numbers in this paper, the following example illustrates how we leveraged the assumption to simplify our task.

Consider three SRCs and for each of which we identified ten quantity options. In the independence case, we would have to run ten iterations of Marathon. It is true that we would have to simulate each SRC 10 times for a total of simulations, but Marathon batches these "sub-simulations" into a single simulation instance.

If we relaxed the assumption, we would have to execute iterations of Marathon. In this case, the three SRCs would interact with each other, and would thus be run in a single sub-simulation. While this sub-simulation batching might, in theory, reduce the number of sub-simulations in a single simulation instance, these potential reductions are offset by the addition of inventory from other components for a single SRC.

The second complexity reduction benefit affects the formulation of the optimization. If we relax the independence assumption, we would have to consider interaction effects in our optimization. This consideration would add non-linearity to the formulation.

4.2 Pre-Simulation

In our methodology we described using static analysis to determine the inventory ranges over we should perform the simulation. These calculations proved to be valuable exclusive of the simulation results and subsequent optimization.

From a methodological perspective, the calculations provided a sanity check on the optimization model results. Specifically, we identified a number of capabilities whose simulation range was zero. In other words, these capabilities, if reduced, would not be able to meet scenario requirements. If our methodology incorporated growing SRC inventories beyond directed growth, these SRCs would be candidates for that growth. Comparing the optimization decisions for these capabilities has proved valuable for validating the optimization model.

From an application perspective, these results also provide important information to decision makers. The intent of the optimization model is not to provide a final, fire-and-forget solution to the force structure reduction problem. Approaching our discussions with decisions makers from this perspective, the range data provide amplifying information which the decision makers can use to shape further conversation or even dictate additional constraints in the optimization.

4.3 Post-Simulation

Analyzing the dwell to BOG data resulting from the numerous simulations provided value in two ways. First, comparing the realized ratios to the range data calculated pre-simulation served as additional validation of the value of those range calculations. For capabilities with ranges determined to be limited, we expected the realized ratios to be commensurately low. In most cases this proved true, though there were a few exceptions. By examining the demands for the exceptions, we were able to identify a few capabilities that warranted additional scrutiny, outside the scope of the simulation. For example, one SRC proved to have a relatively low demand for most of the simulation. However, there was a very brief period, approximately one month out of a thirteen year simulation, where the demand spiked by a factor of more than ten.

The secondary value of analyzing these data resulted from charting the data for each SRC across the various inventory quantities. A visual inspection of these charts provided a simple visual to determine the quantities below which the goal dwell to BOG ratio was violated. As we inspected the data we noticed a few idiosyncrasies—we expected the realized ratios to decrease monotonically as quantity decreased. A few violations of this expected monotonicity proved to correspond to inventory levels at which those capabilities were no longer able to satisfy all demands.

4.4 Optimization

A review of section 3.5 reveals a rather simple optimization model. The model as formulated is able to find optimal solutions in about one second. Compared to the manual process of trying to find feasible solutions, the ability of the model to quickly find optimal solutions provides a greatly increased ability for analytic excursions. We will be able to incorporate new constraints and parameters and provide timely feedback. This timely feedback will ultimately improve the richness of information decision makers are able to consider before finalizing force structure reductions.

5 CONCLUSION

In this paper we discussed a methodology the Army is using to inform its force structure reduction decisions. The methodology included leveraging an existing simulation with some purpose-developed data processing to feed a simple optimization model. We intend for the application of the methodology to serve as a start point for discussion on how best to reduce the size of the total Army.

In our discussion we identified some simplifying assumptions that allowed us to meet decision-maker needs timely. Given that Total Army Analysis is an annual decision venue, it would be prudent of us to continue to improve the methodology to better support future decisions. Critical needed improvements include substitution in Marathon, which would require using something other than a MaxUtilization policy and perhaps a more detailed experimental design to use Marathon efficiently. Given these our optimization formulation will likely need to change, though the shape of those changes will be determined as we improve our utilization of Marathon for force structure experimentation.

ACKNOWLEDGMENTS

We would like to thank Tom Spoon, Myles Miyamasu, and COL Scott Nestler at the Center for Army Analysis for the ready provision of their subject matter expertise; Dr. Robert Dell from the Naval Postgraduate School for his willingness to discuss even the most basic elements of optimization with us; and Kerry Schindler and James Chou in the Army Headquarters Force Management Directorate for the opportunity to work on this interesting problem.

REFERENCES

Office of the Secretary of Defense. 2010. Quadrennial Defense Review Report. United States Department of Defense.

United States Army. 1995. Army Regulation 71-11, Total Army Analysis. Army Publishing Directorate.

AUTHOR BIOGRAPHIES

JASON SOUTHERLAND is an operations research analyst at the Center for Army Analysis and a PhD candidate in Operations Research at George Mason University. His research interests include sequential

decision making processes and force structure analyses. He holds an MS in Operations Research from George Mason University. His e-mail address is jsouther@gmu.edu.

ANDREW LOERCH is an Associate Professor and the Associate Chair of the Department of Systems Engineering and Operations Research at George Mason University where he directs the track in Military Applications of Operations Research in the masters program in Operations Research. He holds a Master of Science in Operations Research from the Naval Postgraduate School, and a PhD in Operations Research from Cornell University. He is also a retired Army Colonel with 26 years of active federal service of which 15 years was spent as a military operations research analyst. Dr. Loerch is a Past President and Fellow of the Military Operations Research Society, is an associate editor of Military Operations Research, and is the editor of the book, Methods for Conducting Military Operational Analysis He recently received the Vance R. Wanner Memorial Award for outstanding contributions to the field of Defense Analysis. His e-mail address is aloerch@gmu.edu.