# Estimating Force Mix Lower Bounds Using a Multi-objective Evolutionary Algorithm

Fred Ma
Defence Research & Development Canada
Ottawa, Canada
Fred.Ma@forces.gc.ca

Slawomir Wesolkowski
Defence Research & Development Canada
Ottawa, Canada
s.wesolkowski@ieee.org

***Abstract*—Nations will always experience conflicting pressures to reduce both (i) the funding of militaries and (ii) the probability that they will not be able to respond to scenarios that may arise. We develop a multiobjective evolutionary algorithm (MOEA) to generate force mix options that trade-off between lower bounds for objective (i) versus objective (ii). A set of military assets or force mix is evaluated against multiple instances of the future, each composed of a mix of stochastically generated realistic scenarios based on historically derived parameters. Scenario success is evaluated by matching each occurrence with a course of action (CoA) whose force element (FE) demands can be met. The lower bound on (i) comes from the assumption that a nation has complete flexibility to engage in scenarios at times that minimize simultaneous demand on FEs. The results are compared with the results from Tyche, a discrete event Simulator, which provides an more realistic, though pessimistic, point estimate of objective (ii). Results confirm the expected relative behavior of both models.**

## I. Introduction

The military are tasked with a number of essential missions, such as search and rescue, emergency assistance in the face of natural disasters, peacekeeping-type activities (e.g., under a United Nations mandate), peace enforcement roles, and, most importantly, protecting a nation from aggression. Those missions require the use of various assets and trained personnel, called force elements (FEs). The mix of all assets and personnel constitute, in military parlance, a force structure or force mix. The costs related to both the personnel and the acquisition, maintenance and operation of military assets can be very high. Such high costs, coupled with limited (and diminishing) military budgets, have driven the use of computational force structure studies to determine the best force structure for carrying out future military missions.

The problem we address in this paper is to determine the trade-off between large, costly force mixes, with high probabilities of being able to send our forces to scenarios arising in the next five years, versus small force mixes having lower probabilities of being able to respond to all scenarios that arise. How well a nation can respond to a range of future scenarios, especially concurrently, depends on a nation's level of ambition. We tackle a simpler version of this problem: estimating the lower bound on the force mix size, given a nation's level of ambition for the fraction of scenarios that can be responded to. The lower bound is an optimistic estimate which assumes that we can schedule scenarios so as to minimize conflicting demands on the resources that make up a nation's force mix. It represents a clear and absolute limit, below which a nation's assets should not drop, regardless of fiscal pressures. Furthermore, national forces do have discretion in when they engage in scenarios, so some freedom to schedule does exist, albeit limited.

Previously, two models have been developed to carry out force mix evaluations: the Stochastic Fleet Estimation under Steady State Tasking (SaFESST) [1] and Tyche [2], [3] models. SaFESST carries out a specialized constrained strip-packing (i.e., one-dimensional bin-packing [4]) over multiple resources. This problem addressed by SaFESST and our problem are similar to a sparse multi-capacity bin packing application for job scheduling [5], [6]. However, the number of each FE type used is not directly proportional to other FE types for each scenario, meaning that methods normally applied to multi-capacity bin packing problems would not be applicable to our problem. We set an upper limit on the quantity of each FE type. Any tasks or portions of tasks that fall outside of the constrained bins would be considered not achievable or "unscheduled." Furthermore, by placing a limit on the size of the force mix (i.e., a threshold on the number of strips of each FE type or on the height of each FE bin), SaFESST evaluates this force mix by providing the list of scenarios that cannot be accommodated in the mix of all concurrent scenarios. Because SaFESST tightly packs all scenarios together, it may underestimate the scenario failure rate and provide a lower bound on the force structure. Tyche, on the other hand, is a discrete event simulator which evaluates force structures in detail over time. Decisions about the use of individual assets are recorded in operational schedules, based on detailed rules that are meant to mimic the decisions of a military scheduler. Although the computational complexity of both models is high, their outputs can provide an assessment of capability gaps and outline scenarios that cannot be completed. Since Tyche does not consider any latitude in when a nation engages scenarios, it tends to form upper bounds on a force mix [7].

The work reported here expands on SaFESST by using a multiobjective evolutionary algorithm (MOEA) to determine a set of trade-offs between minimal force mix and the failure to respond to scenarios (the original objective in SaFESST). Specifically, the objectives being minimized are: (i) force mix cost, which depends on the required quantities of various asset types, and (ii) the number of failures to respond to scenarios, given a maximum limit on the quantity of various asset types. To minimize both objectives, an MOEA attempts to schedule the scenario occurrences to minimize peak demands for the various FEs. In addition to the timing of the scenarios, however, the search space also includes: (i) the selection of a course of action (CoA), a possible response to a scenario, for each occurrence, which affects the assets needed, and (ii) the order with which scenario occurrences are allotted assets.

We compare the MOEA results to those of a Tyche analysis using the same stochastic scenario parameters, force mix, and scenario CoAs [8]. Beyond its greater modelling detail, Tyche differs from the model reported here in that it does not seek a trade-off curve for the lower bound estimate of required assets.

## II. Model

The two minimization objectives (force mix cost and failed scenarios) drive the model definition. In this section, we describe the modelling decisions to enable the calculation of those objectives, the resulting input parameters, the free parameters to be searched, and their organization into chromosomes for the genetic algorithm (GA).

### A. Force Assets

The forces' assets are broken down by FEs. This is a generic modelling construct representing any asset, ranging from planes, tanks, and ships to Command and Control (C2) headquarters, task forces, and services. We distinguish between an FE *type* versus a specific physical *unit* of that type of FE (either required or available). One input into the problem is a nation's nominal force mix, i.e., the quantity of each type of FE. For example, 40 fighters or 15 surface combatant vessels.

### B. Scenarios

We use the term *scenarios* to refer to types of missions on which the Canadian Armed Forces (CAF) deploy. Scenarios are modelled as randomly occurring and having random duration. We also distinguish between scenario types versus specific occurrences of a type of scenario [1] Different scenario types are defined, each with its own Poisson rate and duration distribution ( e.g., triangular or boxcar). This model is not a time domain simulation per se. The Poisson distribution is used to determine the number of occurrences of each scenario type within the analysis time window, but the scenario occurrences are not randomly scattered throughout the timeline. The model seeks a lower bound, and situating the scenario occurrences in time represents some of the degrees of freedom being searched to minimize the objectives.

For each scenario type, there are different CoAs, each having slightly different requirements as FE quantities. If a specific CoA is selected for a given scenario occurrence, then the FEs needed by that CoA are needed for the entire duration of the scenario occurrence.[2]

Using the Poisson rates and duration distributions, a set of scenario occurrences is generated over a predefined number of years. This combination of scenario occurrences constitutes one possible instance of the future (IotF). To gather statistics on force mix adequacy, our MOEA was applied to multiple IotFs. In each IotF, the set of scenario occurrences is fixed during the GA optimization process.

### C. Force Mix cost

Without loss of generality, we use notional, artificial normalized costs to illustrate the method. The cost of a force is based on the *required* quantity of each type of FE. In contrast, each FE type is also treated as a "fleet" of FE units whose size is the number of FEs units *available*, i.e., the aforementioned nominal force mix. For a given FE type, if there are $N$ units in the fleet, then the cost of each unit required is set to $1/N$ times the fleet cost. We made the fleet costs uniform for all FE types and set that cost to 1.0, e.g., the cost of a fleet of 10 ships is the same as the cost of a fleet of 70 aircraft. Thus, the larger the fleet, the lower the relative cost of an asset with respect to other FEs.

Since the demand of each FE type determines its contribution to the force mix cost, how the demand is assessed is important. The current analysis takes the FE demand on the force mix to be the peak number of a given FE type needed at any time within the analysis period, i.e., the peak simultaneous demand for each FE type.[3] If this is less than the fleet size, then it means that the fleet could have been smaller without impacting the scenario failures.

### D. Objective Functions

A MOEA is used to find the trade-off between conflicting objectives: minimizing a force's cost and minimizing scenario failures. The objectives being minimized are: (i) force mix cost, which depends on the maximum simultaneous numbers of various types of FEs needed to meet the requirements of scenarios that arise, and (ii) the number of failures to respond to scenarios, given a maximum limit on the on the numbers of the various FE types that can be simultaneously deployed. The FE limits used for objective (ii) can be thought of as the nominal force mix, consisting of actual assets possessed by the forces for which there are no alternative providers.

Objective (i) is defined as

$$\sum_{i=1}^{N_F} n_i c_i \ ,$$

where

[1]Because of this distinction, the word "type" sometimes follows the term *scenario*.

[2]This is a simplification which avoids the problem of managing partial asset use in a given scenario time frame.

[3]The peak will never exceed the quantity in the force mix. Any scenario occurrence that would otherwise lead to such a situation is considered to have failed. In that case, none of the FE demands in the chosen CoA would be realized. In particular, the scenario occurrence will not consume any time on the schedules of the actual force assets.

| | | | | |
|---|---|---|---|---|
| CoA#1<br>Start date #1<br>SOK #1 | CoA#2<br>Start date #2<br>SOK #2 | CoA#3<br>Start date #3<br>SOK #3 | CoA#4<br>Start date #4<br>SOK #4 | CoA#5<br>Start date #5<br>SOK #5 |

Fig. 1. Example chromosome.

$i$ = The FE type,
$N_F$ = The number of FE types,
$n_i$ = The peak number of FE units of type $i$ used throughout the analysis time period, and
$c_i$ = The unit cost for FE $i$ (Section II-C) .

If the scheduling of a scenario occurrence requires that $n_i$ exceed the force mix limit for any FE type $i$, that occurrence is simply not scheduled. Objective (ii) is simply a count of such failed scenario occurrences. These failures are also counted on a per-scenario-type basis to enable per-scenario-type statistics.

## III. Genetic Algorithm

The GA consists of two complementary parts. The problem-dependent part includes the parameter encoding scheme and the scheme for computing objectives. The problem-independent part consists intergenerational mechanics, such as the selection scheme, crossover operator, mutation operator, and population replacement.

### A. Chromosome Structure

A chromosome represents a candidate solution in the sense that a schedule of the stochastically generated scenario occurrences can be constructed using the parameter values within the genes. To schedule each scenario occurrence, the genes must contain the start date/time, duration, and the specific CoA selected. Since the required FEs are defined for each CoA, a chromosome also determines the schedule of demands for the various FE types. The next section describes how values for the objectives are determined from such a schedule of FE usage.

Figure 1 shows an example chromosome for a hypothetical case in which there is a total of five scenario occurrences (of the various scenario types) throughout the period being simulated. A chromosome consists of a series of one gene per scenario occurrence. Each gene contains three fields that govern the scheduling of the corresponding scenario occurrence: (i) its start date, (ii) its CoA, which determines its FE demands, and (iii) a SOK (sort order key, a real number) that determines the order in which the scenarios have their FE demands scheduled onto the FE timelines. Given $n$ scenario occurrences for a given IotF, there are $3n$ parameters for each chromosome.

The SOK is compared with the SOKs of the other genes in the chromosome to determine the order in which the genes' associated FE demands are scheduled onto the timelines for the FEs in the force mix. This scheduling is not meant to represent the nation's priorities; it is a degree of freedom being searched by the GA to minimize the objectives. Different SOK values yield different force mix costs and scenario failures. Using an explicit scheduling order parameter for each scenario occurrence gives greater freedom to the GA search compared to using the gene position as a scheduling order parameter. The gene positions also determine which genes are swapped in crossover. Using an explicit scheduling order parameter decouples this function from the function of scheduling ordering. As well, using the gene position for scheduling ordering involves handling different portions of the chromosome differently in order during crossover to preserve a full set of genes. This is avoided by using an explicit SOK.

The use of a specific start date/time for each scenario occurrence differs from the more common scheme of best-fit bin-packing the scenario occurrences [9]. Forgoing left packing speeds up the construction of the schedule, especially when the FE usages for a given scenario occurrence need to be time-aligned. It also introduces more free parameters into the search space. The latter admits potentially better solutions, but also potentially increases the cost of searching the space. This dual effect was further amplified by the early decision not to discretize in the time domain. In the details of the GA design, therefore, a heavy emphasis was placed on exploration rather than exploitation.

The duration of each scenario occurrence is needed to construct the schedule of FE usage for each chromosome. This parameter is stochastically generated when the scenario occurrence counts are generated, and is not a degree of freedom to be searched. Hence, it does not form a field within the genes of a chromosome.

### B. Chromosome Evaluation

Keeping in mind that each gene corresponds to a scenario occurrence, the objective functions are evaluated by scheduling the demands of the scenario occurrences onto the timelines of the FE units in the force mix. The scheduling starts with empty timelines for all FE units. The scenario occurrences are sorted by their SOKs and processed in the resulting order. For each scenario occurrence, empty portions on the FE timelines of the proper type are sought to meet each FE unit demanded. The timelines are always scanned for openings in the same order, starting with the lowest numbered FE unit of the required type so that any empty timelines are always numbered the highest. The positions of the openings on the timelines, and their widths, must be able to fit the start date and duration of the scenario occurrence.

A scenario failure occurs if any of the CoA demands can't be met by timelines of the right FE type. If this happens, the scenario occurrence is discarded and the timelines are restored to their state prior to attempting to schedule the failed scenario occurrence.

An attempt is next made to schedule the next scenario occurrence, as determined by the genes' SOK fields. This repeats until all the scenario occurrences are either scheduled or have failed. The two minimization objectives are then tallied and stored with the chromosome.

### C. Intergenerational Mechanics

We chose to use the Nondominated Sorting Genetic Algorithm II (NSGA-II) [10] for the intergenerational mechanics due to its popularity in the literature. Our adaptation of NSGA-II is structured as follows.

The selection for crossover is based on a chromosome's rank. As described for NSGA-II [10], the rank is determined first by the nondominated front to which it belongs, and then by its proximity to other solutions on the same front. The crossover scheme was uniform crossover, i.e., a pre-specified portion of the genes are swapped between a pair of parent chromosomes. For genes selected for swapping, the three fields of a gene were kept together and exchanged as a unit.

Mutation was performed on chromosomes drawn from the population with uniform probability, with replacement. The probability of mutating the CoAs, start date, and SOK fields were specified independently. Hence, there was a small probability of more than one field being mutated, e.g., if $p_{\mathrm{SD}}$ and $p_{\mathrm{COA}}$ were the probabilities of mutating the start date and CoA fields, respectively, then the probability of a chromosome being mutated in both fields was $p_{\mathrm{SD}}p_{\mathrm{COA}}$. If a particular field was to be mutated, then there was a uniform probability of mutating that field in 1 to $N_g$ genes, where $N_g$ is specified separately for each field. For a given field, a mutated value was chosen with uniform probability. If no fields end up being mutated for a particular chromosome, then another chromosome must be drawn. This is repeated until the quota for mutated genes (discussed next) is met.

For each generation, a pre-specified fraction of the population (the lowest ranking chromosomes) was replaced by newcomers, i.e., mutated chromosomes and offspring from crossover. The proportion of newcomers consisting of offspring versus mutated chromosomes was also pre-specified. To emphasize exploration of the search space, the replacement by newcomers was unconditional.

Exploration was also emphasized by mitigating the selective pressure of ranked selection, which preserves diversity in the genes. The mitigation is illustrated in Figure 2 for a notional population of 10. The convention used is that better chromosomes have higher rank numbers. The bar heights are proportional to the probability of being selected. The dark bars represent the normal selective pressure from ranked selection, while the yellow bars show the pre-specified adjustment to reduce selective pressure. Under this modification, selection is done as follows. If we let $f_0$ be the lighter-coloured offset for rank 1, we can use the same approach as the formula for sum of an arithmetic series to show that the selected rank should be $\lceil\sqrt{2\xi + (f_0 + 1/2)^2} - f_0 - 1/2\rceil$, where $\xi$ is a random number between 0 and $(N + 1 + 2f_0)N/2$, and $N$ is the population size [11]. We set $f_0 = N/10$.

The ranking depicted in Figure 2 is simplified in that there are no ties in the ranking of chromosomes. In practice, however, there could be possibly many chromosomes with the same objective function values, which means that they are tied. The lower numbered chromosomes within the tie will be disadvantaged. We address this effect by randomly ordering the chromosomes before ranking the population. This ensures that any tied chromosomes are equally likely to be placed in any one of the tied positions.

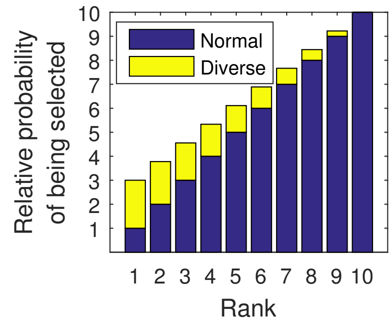



Fig. 2. Mitigating selective pressure in ranked selection.

TABLE I
SCENARIO FREQUENCIES AND DURATIONS

| Index | Poisson rate Annual | Duration [days] | | |
|---|---|---|---|---|
| | | min | peak | max |
| 1 | 0.2 | 2 | 14 | 16 |
| 2 | 0.08 | 2 | - | 27 |
| 3 | 1.4 | 1 | 10 | 188 |
| 4 | 0.3 | 173 | 360 | 4130 |
| 5 | 0.3 | 223 | 2555 | 4751 |
| 6 | 0.09 | 94 | - | 250 |
| 7 | 1.2 | 1 | 14 | 133 |

## IV. RESULTS

### A. Experiment Setup

This section describes the two groups of parameter values used to populate our model. The first group pertains to the military problem as described in Section II. The second group of the parameters pertains to the GA's intergenerational mechanics, as described in Section III-C.

To enable comparison of the results, the military problem parameters were chosen to match Tyche's evaluation of force mix performance using discrete event simulation [8]. To reduce computational requirements, however, we excluded FEs that were modelled in the Tyche study as always available, e.g., external services, as they do not impact the scheduling. Our resulting nominal force mix had 39 FE types and a total of 342 FEs units across all FE types.

Seven scenario types from the Tyche study were used. Their Poisson rates and triangular/boxcar duration distributions are shown in Table I. The boxcar duration distributions are those without a *peak* value under *Duration*. The stochastic parameters in Table I were determined from a survey of CAF operations from 1990 to 2012 [8]. The period over which to generate scenario occurrences was also chosen to be five years in order match the Tyche study.

For the CoAs available to each scenario, Figure 3 gives a rough sense of the variability in demand between CoAs. The CoAs available to each scenario type are grouped together along the CoA axis.

Each possible IotF consisted of a set of stochastically generated scenario occurrences over a five year time window. There were seven scenario types (see Table I), and each could

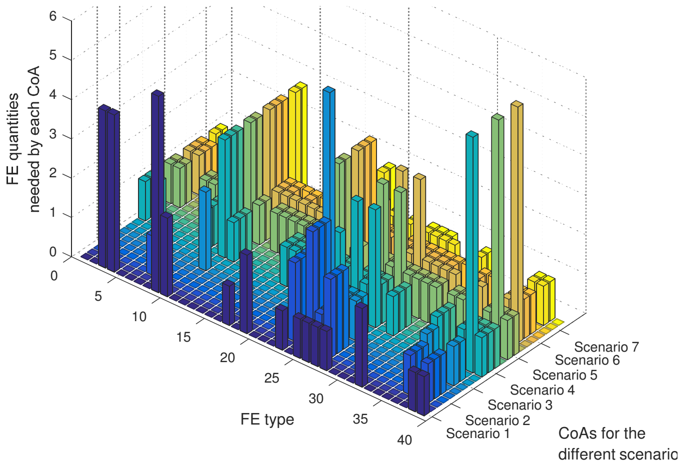



Fig. 3. Force element quantities needed by the CoAs.

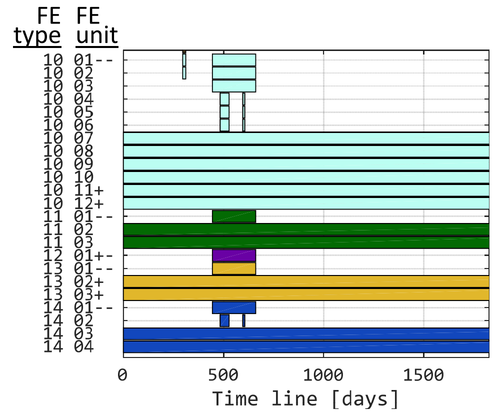



Fig. 4. A portion of the schedule for the force elements in the force structure.

be responded to with one to three CoAs. For each IotF, the GA created schedules for all FEs to minimize the objectives. Scenario occurrences were analyzed over 1495 IotFs to gather statistics on failed scenarios for comparison with Tyche.[4]

The GA parameters are as follows. The population size and generational count are both set to 400. For the uniform crossover, 50% of the genes were swapped. For ranked selection with less selective pressure, the probability of selecting the worst chromosome was set to one tenth of the probability selecting the best one.

Mutation was set to be high to emphasize exploration. The experimentally set probabilities of mutating the CoAs, start date, and SOK were {20%,40%,10%}. For these three fields, there was uniform probability of mutating from 1 to 2 genes, 1 to 4 genes, and only 1 gene, respectively. More mutation was allowed for in the start date to better explore the extra degrees of freedom due to our chromosome representation. Less mutation was allowed for in the SOK because its direct effect is not local to the gene containing the mutated SOK; the order in which the demands of the scenario occurrences are scheduled onto the FE timelines depends on how the SOKs compare to each other.

The generational replacement scheme was also chosen to emphasize exploration. For each generation, the lowest ranking 60% of the population members were unconditionally replaced by newcomers. 40% of the newcomers were created by crossover, while 60% were mutated chromosomes.

## B. General Observations

Visually assessing the efficacy of the GA was challenging due to the large number of timelines (342) on which the scenario occurrences may collide. Figure 4 shows a small portion of the schedule for the FEs in the force mix. Each set of vertically aligned blocks represents the FE demands of a single scenario occurrence, and the entire schedule contains 342 timelines, one for each FE unit in the force mix. This schedule was obtained from one of the nondominated solutions for one of the IotFs, then constructing the schedule using the genetic information in the chromosome, but *without* disallowing overages. The GA search itself does not allow overages in constructing the schedule to evaluate the chromosome, but we constructed the schedule *with* overages to look for improvements that the GA missed.[5] Overages are shown by "+" signs along the y-axis labels (the dashes demarcate the beginning of a new platform type, as does a change in the colors of the scheduled blocks of time).

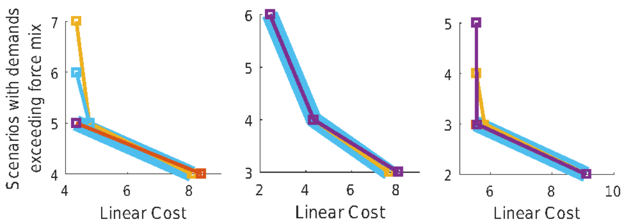
The challenge to visual inspection being illustrated by Figure 4 is as follows. There is an apparent failure of the GA to deconflict the three scenario occurrences in the 500-700 day time frame. With careful inspection, however, the explanation can be found on one of the 342 timelines: the burgundy block, FE 12, unit 01. It is shown as an overage, yet it is the first unit of that FE type used. This is because the quantities of some FEs types are zero. It may be that a nation needs to acquire some of these FEs. When overages are disallowed, however, the scenario occurrence that requires such FEs is never placed in the timelines; there is no actual conflict, therefore, and the schedule is actually appropriate for the given objectives.

Another prominent feature in Figure 4 is the swaths of time consumed by long running scenario occurrences. We expected the GA to favour the scheduling of long scenario occurrences and scenarios with demanding CoAs after other scenario occurrences. This is because failing such scenarios frees much of the schedule for many more small or less demanding scenarios, which helps to minimize the objective for failed scenarios (see Figure 4). The lengthy usages of the timeline are scheduled after the shorter blocks, as seen from the higher FE unit numbers occupied by the large blocks.

Finally, we found that much mutation of the scenario occurrence start dates was needed to find openings in the timelines. This was not surprising, since the scheme does not tightly pack scenarios to guarantee some degree of abutment.

[4]The number of IotFs was determined by the time available for this study and holds no particular significance.

[5]Without overages, many of the scheduled scenario demands would not be present on the timelines, and we would be looking at a fairly sparse schedule without much information about the absent scenarios and why they failed.




(a) Typical (b) Almost nondominated (c) Marginally nondominated

Fig. 5. Example fronts from three IotFs.

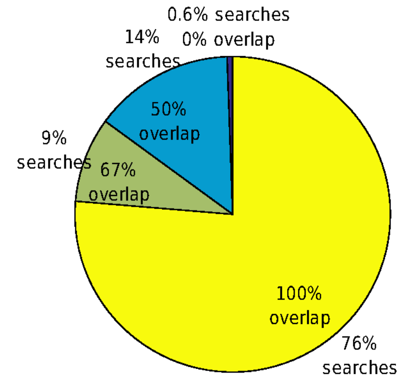



Fig. 6. Proportion of searches achieving various levels of coverages of their superfronts (all IotFs).

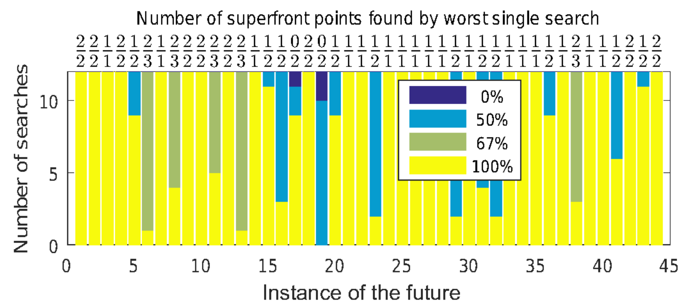



Fig. 7. Proportion of searches achieving various levels of coverages of their superfronts (per IotF).

### C. Performance Evaluation

The nondominated solutions found by the MOEA yield schedules that minimize force mix cost and/or failed scenario occurrences. Each IotF has its own set of scenario occurrences and its own true nondominated front. The GA's stochastic search finds an approximation to the true nondominated front. To evaluate the GA's convergence, we performed twelve searches on each of several dozen IotFs. For each IotF, we pooled the nondominated points from the twelve searches and identified a new nondominated front from them. Since this new "superfront" consists of the best of the nondominated points, it approximates the unknown true front at least as well as the individual search fronts (oftentimes better). If the search scheme is well tuned to the problem, then the individual search fronts for an IotF will be similar to the superfront. Figure 5 shows the twelve search fronts and the superfronts from three different IotFs. The vertices of the lines represent the nondominated solutions; the lines are drawn only as a visual aid. The superfronts are shown as the thick light-blue lines.[6]

What we found is that there are very few points on these fronts: only one to three. Subfigure 5a shows a typical case where the superfront has two points. In this case, some of the individual fronts do not coincide with the upper left most point on the superfront, allowing some of the solutions (e.g., at $y = 6$ and $y = 7$) to appear to be nondominated, even though they are suboptimal compared to the superfront. Subfigure 5b contains less disparity between the superfront and the search fronts. Subfigure 5c shows a more extreme example of the middle burgundy point being marginally off a superfront point, allowing the solution at $y = 5$ to appear marginally nondominated. We expect marginally nondominated solutions to be of limited value, since anyone inspecting such solutions would recognize the exorbitant trade-off of one objective for negligible gain in the other objective.

We crudely gauged the thoroughness of the searches by tallying up the proportion of searches that overlapped with 0%, 50%, 67%, and 100% of their respective superfronts (no searches overlapped by 33%). Figure 6 shows that the majority of the searches contained all the points of the superfronts to which they contribute. The stacked bar chart of Figure 7 further breaks these proportions down by individual IotFs. Each vertical bar shows the overlaps with the superfront for each of the 12 searches of one IotF. The high overlapping searches are shown toward the bottom of the bar, while the low overlapping searches are located near the top. For some IotFs, all 12 searches achieved 100% overlap with the superfront, e.g., instances 1-4 of the future. The fractions along the top show the number of overlapping points for the search that overlaps least with the superfront. The numerator is the number of overlapping points, while the denominator contains the size of the superfront (which is common for that entire IotF).

When interpreting the proportion of searches that attain the various percentages of overlap with their superfronts, one should keep in mind that this is a pessimistic indicator of search front quality. It penalizes a search for not finding the exact points of the superfront, ignoring that nondominated points found by a GA search can be close to the superfront.

As mentioned, a prominent feature of this problem is the limited number of points on the nondominated front. This may indicate a high degree of correlation between the optimization objectives. Another possible contributing factor is the small number of failed scenarios. There are no fractionally failed scenarios in our problem formulation. Yet another possible cause is that the number of FE timelines used can only change by whole numbers, meaning that even the cost is discretized. If

[6] On a noncolour hard copy, these may show up as shades of gray, but the thick band is the superfront.

a particular FE type has a large fleet size $N$, then the notional cost used here can change in small quanta $1/N$; however, most fleet sizes were small.

Because of the small nondominated fronts, Deb's crowding distance scheme [10] to maximize diversity on the nondominated front was essential.

### *D. Comparison with Tyche*

The GA results were compared with Tyche. This study used the same scenario parameters, FE demands, and force mix.[7] However, Tyche also required additional details, such as FE basing, travel time, and a response time (if assets can reach theatre with a time window, that a scenario occurrence would not fail). We modified the definition of scenario success from the previous study [8]. Specifically, a scenario occurrence is successfully responded to if the required FEs can be scheduled for the entire duration of the scenario occurrence.[8]

The results of the comparison are shown in Table II. Several GA configurations were tried for this comparison, but the main one uses a population of 400, an evolution of 400 generations, and unconditional replacement of the worst population members by the newly generated chromosomes. In the central columns of the table, these are highlighted with thick borders, along with the Tyche results. The results of earlier GA configurations are shown on the flanks to corroborate the GA results. We compare results at both ends of the nondominated front (i.e., the solutions with least scenario failures and those with least cost) to the Tyche failures. These correspond to columns *A–E* and *G–J*, respectively. For each column, the scenario occurrence counts and failures are aggregated across all IotFs, the number of which are given in the *Instances of the future* row.

In comparing the percentages in Columns E, F and H of Table II, the GA performed as expected with a heavy bias towards preserving small-sized scenarios. On the other hand, scenario types 4, 5, and 6 involve the most heavy commitment for duration and FE requirements, and result in the largest number of failures (see similar observations in Section IV-B). Although the GA yields lower failure rates for lower commitment scenarios, they are only modestly lower. The large number of scenario occurrences (scenario types 3 and 7, column B) translate into a sizable reduction of failures. It is the absolute number of scenarios failures that drives the GA, and this is reflected in the *Failed scenario occurrences* row of columns *E* and *F*.

Scenario types 1 and 6 have 100% failure rates. These are special cases requiring FEs that will never be available, as described in Section IV-B. One reason for zero-quantity FEs is that their actual units are monopolized by continuously running scenarios that we excluded from our model.[9] Scenarios types 1 and 6 have less than 100% failure in the Tyche results because of the way in which continuous scenarios were modelled. Their monopolization of FE was modelled in a way that made it pre-emptable, albeit infrequently.

In Table II, the GA results for minimizing failures differs slightly from those for minimizing costs. As one would expect, minimizing cost leads to failure rates at least as high as those for minimizing failures. With the exception of scenario type 4, however, any increases are modest. Apart from high demand scenario types 4 and 5, the failure rates are still lower than those for Tyche. This is not surprising, since the GA optimizes the schedule of scenarios to get a lower bound, while Tyche deals with each scenario as it arises.

The results of the different GA configurations in Table II are quite consistent. The $400\times400$ results (population 400 and 400 generations in the evolution) can be compared with the $125 \times 125$ results using unconditional replacement of population members by new chromosomes. For minimizing failures, this means comparing columns $C$ and $E$; for minimizing cost, columns $H$ and $I$. The failure percentages are very similar. These results can also be compared with a scheme in which the new chromosomes compete to enter into the population for the next generation, i.e., by being combined with the population, ranked by nondominance first and crowding distance second, then having the population truncated back to its original size. This strategy is more exploitative and less exploratory. For minimizing failures, the columns for comparison are $A$, $C$, and $E$; for minimizing costs, $H$, $I$, and $J$. Again, the failure percentages seem relatively consistent despite such variations in the GA scheme.

## V. Conclusion and Future Work

We presented a multiobjective evolutionary algorithm to estimate the lower bounds of a force mix by scheduling scenario demands onto the timelines of the force elements in a force mix. The algorithm was demonstrated using empirically determined scenario occurrence patterns, a realistic force mix, and for each scenario, several courses of action to emulate variability in force element demands. The intent was to determine a trade-off curve between force mix cost and probability of failing to respond to scenarios, in the form of a nondominated front. The problem encoding involved a large search space, and special measures were taken to emphasize exploration over exploitation. The resulting nondominated fronts were found to be sparse, with few trade-off points. Most of the searches found all the points on the superfront, indicating at least convergence to a good local minimum. Our usage is tolerant of good but imperfect approximations to the Pareto front. This is because we are using the GA to identify *trends* in the force mix lower bound over many IotFs, with possibly very different sets of scenarios.

The GA results were robust against variations in GA design. The results were compared with Tyche, a high-fidelity discrete event simulator. Because of the objective to minimize overall scenario failures, the GA favoured failing large, demanding, and infrequent scenario types to drive down failures in

[7] The Tyche force mix included services provided by entities outside of the military, which were modelled as always available. Since they imposed no constraints on the simulation, we excluded them from the force mix.

[8] In reference [8], a scenario occurrence was successfully responded to if assets arrived on time, but they could be pulled away for other scenario occurrence.

[9] They increase model complexity without changing the results.

TABLE II
SCENARIO OCCURRENCE FAILURE RATES.

| | | | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Which end of ND front | | | Minimize failures | | | | | | Minimize cost | | | |
| Population x Generations | | | 125x125 | | | 400x400 | | Tyche | 400x400 | | 125x125 | |
| Replacement scheme | | | Conditional | Unconditional | | Unconditional | | | Unconditional | | Unconditional | Conditional |
| Scenario type | FDS | Name | % failed | Scenario occurences | % failed | Scenario occurences | % failed | % failed | Scenario occurences | % failed | % failed | % failed |
| 1 | 1b | Arctic | 100.0 | 233 | 100.0 | 1518 | 100.0 | 99.5 | 1518 | 100.0 | 100.0 | 100.0 |
| 2 | 3 | MTE | 0.0 | 98 | 0.0 | 606 | 0.0 | 0.2 | 606 | 0.0 | 0.0 | 0.0 |
| 3 | 4 | Vancouver | 0.0 | 1476 | 0.1 | 10436 | 0.0 | 8.1 | 10436 | 1.0 | 2.2 | 1.9 |
| 4 | 5a | Sudan | 66.3 | 334 | 69.8 | 2284 | 65.8 | 59.1 | 2284 | 100.0 | 99.4 | 100.0 |
| 5 | 5b | Somalia | 97.1 | 349 | 95.1 | 2310 | 95.2 | 57.3 | 2310 | 100.0 | 100.0 | 100.0 |
| 6 | 5c | North Korea | 100.0 | 109 | 100.0 | 607 | 100.0 | 87.4 | 607 | 100.0 | 100.0 | 100.0 |
| 7 | 6 | Haiti | 0.1 | 1294 | 0.0 | 8776 | 0.0 | 79.0 | 8776 | 3.2 | 2.3 | 3.9 |
| Instances of the alternate future | | | 161 | | 226 | | 1495 | 1000 | | 1495 | 226 | 161 |
| Scenario occurences | | | 2943 | | 3893 | | 26537 | 18700 | | 26537 | 3893 | 2943 |
| Failed scenario occurences | | | 685 | | 908 | | 5828 | 9278 | | 7109 | 1085 | 840 |
| % failed | | | 23.3 | | 23.2 | | 22.0 | 47.7 | | 26.8 | 27.9 | 28.5 |

frequent, low demand scenario types. The failure rates for infrequent, high demand scenarios were higher than those for Tyche, but lower for infrequent, low demand scenarios. The failure rates were lower overall for the GA, as expected by its targeting of a lower bound problem. These are preliminary results, and further comparative analyses are being undertaken. Comparison with other force mix analysis tools is also being pursued.

This study has identified a number of avenues for future work. For modelling operations more accurately, a scenario prioritization scheme should be developed to stop demanding scenarios from being ignored. Alternatively, failure rates could be minimized on a per-scenario-type basis rather than using the overall failure rate. Another improvement could be to change the requirement for an FE to be monopolized for the entire scenario duration, allowing currently conflicting scenarios to be co-located. The model could be also extended with a clear separation between capabilities and FEs. This would allow the examination of the effectiveness of various FEs delivering the same capabilities .

The cost/failure trade-off can be studied under force mixes that reflect different national dispositions in the spectrum of conflict, e.g. peace keeping versus war fighting. A comparison with other force mix analysis tools (such as with Tyche in this paper) would be useful to determine their strengths and identify the kinds of operational questions that each tool, or combination thereof, would be best suited to answer.

## ACKNOWLEDGMENT

We thank S. Mostapha Kalami Heris and the Yarpiz team for making their NSGA-II code publicly available, and Kevin Bartlett for his *timeline* plotting code. We thank Cheryl Eisler for assistance with Tyche.

## REFERENCES

[1] S. Wesolkowski and D. Wojtaszek, "SaFESST: Stochastic fleet estimation under steady state tasking via evolutionary fleet scheduling," in *Proceedings of the IEEE Congress on Evolutionary Computation*, Brisbane, Australia, June 2012, pp. 1–8.

[2] C. Eisler and D. Allen, "A strategic simulation tool for capability-based joint force structure analysis," in *Proceedings of the International Conference on Operations Research and Enterprise Systems*, February 2012, pp. 21–30.

[3] C. Eisler, S. Wesolkowski, and D. Wojtaszek, "The Tyche and SaFE models: Comparing two military force structure analysis simulations," in *Proceedings of the European Conference on Modelling and Simulation*, May 2014, pp. 625–632.

[4] S. Martello, M. Monaci, and D. Vigo, "An exact approach to the strip-packing problem," *INFORMS Journal on Computing*, vol. 15, no. 3, 2003.

[5] W. Leinberger, G. Karypis, and V. Kumar, "Job scheduling in the presence of multiple resource requirements," in *Proc. ACM/IEEE Conference on Supercomputing*, 1999.

[6] W. Leinberger, G. Karypis, and V. Kumar, "Multi-capacity bin packing algorithms with application to job scheduling under multiple constraints," Department of Computer Science and Engineering, University of Minnesota, Tech. Rep. TR 99-024, 1999.

[7] S. Wesolkowski and C. Eisler, "Capability-based models for force structure computation and evaluation," in *NATO Workshop on Integrating Modelling & Simulation in the Defence Acquisition Lifecycle and Military Training Curriculum*, Washington, D.C., October 2014.

[8] C. Eisler, "Force sturcture evaluation for capability-based planning," Defence R&D Canada – Centre for Operational Resesarch & Analysis, Ottawa, Canada, Tech. Rep. DRDC-RDDC-2015-R026, March 2015.

[9] C. Kenyon, "Best-fit bin-packing with random order," in *Proc. ACM-SIAM Symposium on Discrete Algorithms*, 1995.

[10] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, 2002. [Online]. Available: http://dx.doi.org/10.1109/4235.996017

[11] F. Ma, "Mapping to coarse grain reconfigurable arrays using evolutionary algorithms," Ph.D. dissertation, Carleton University, 2005.