Algorithmic Preference Matching in the Department of Defense

Ian Shaw * Zac Dannelly † Drew Calcagno ‡ Richard Kuzma §

April 14, 2019

Abstract

The Department of Defense maintains a consistent matching problem in the placement of military members into jobs due to the requirement that uniformed service members rotate positions every three years. Gale and Shapely provided a stable matching process 1962 [4] that if applied to this problem would greatly improve the current, manual matching process. In this paper we go a step further noting that mixed-integer programming can be applied in the matching process to sacrifice stability in favor of military leadership goals. Additionally, we glean novel metrics from the expressed ordered preferences of both service members for jobs as well as job owners preferences for service members. These metrics are: specialization, competitiveness, similarity, and preference correlation. We formulated our prototype for the Chief of Naval Personnel. Nevertheless, our work is agnostic and can be applied not only to any service branch, but to any matching market not requiring stability, defined by obliged market participants. [1]

*Ian Shaw (Lieutenant JG, US Navy) is a cyber warfare engineer. As an alum of Stanford University, he researched sequential machine learning models at the Institute for Computational and Mathematical Engineering. He, and all authors of this piece, graduated with distinction from the US Naval Academy and together make up Naval Algorithms, the small prototype team described in the paper.

†Zac Dannelly (Lieutenant JG, US Navy) is a cryptologic warfare officer serving at Fort Meade. He has researched the effects of emerging technologies on organizations through degrees from the Judge Business School at the University of Cambridge, on the Gates Scholarship, as well as from the Management Science and Engineering Department at Stanford University.

‡Drew Calcagno (Lieutenant JG, US Navy) serves various teams in the Office of the Secretary of Defense. He writes emerging technology strategy there, particularly regarding artificial intelligence. He is an alum of the University of Oxford as a Rotary Scholar and the University of London as a Fulbright Scholar where he constructed theses on how the Department of Defense and the Intelligence Community could better coordinate counter-terrorism operations, particularly on the African continent.

§Richard Kuzma (Lieutenant JG, US Navy) is a Navy surface warfare officer passionate about how the Defense Department adapts to emerging technologies, particularly artificial intelligence. He is an alum of the Harvard Kennedy School where he wrote a thesis on how the Defense Department should structurally change to implement artificial intelligence, partnering with the Defense Innovation Unit. He is an associate at Harvards Technology and Public Purpose Project. He is on Twitter, @rskuzma.

[1] The code to demonstrate the matching algorithms, optimization, and preference-based metrics can be found in Ian Shaw's `Dynamic_Manning` Github Repository . `https://github.com/ieshaw/Dynamic_Manning`

Contents

1 Introduction

Under the backdrop of the President of the United States issuing the *Executive Order on Maintaining American Leadership in Artificial Intelligence* in February 2019, we created the algorithmic foundation for the advancement of artificial intelligence and machine learning (AI/ML) in military talent allocation. Before the US Government can accomplish its AI/ML goals, however, it must embrace the essential building blocks underlying the buzzwords.

We believe the world to be arcing toward the need for humans to grow as increasingly specialized workers, and in the vicious games of statecraft and warfare, allocating talent with hyper specificity will become a necessity for victory. A building block on that path is the institutional acceptance of data-driven human decision making and algorithmic support of those human decisions. To that end, we theorized, coded, and prototyped an mixed integer programming (MIP) algorithm that matches job seekers (sailors) and job owners (commanding officers) more optimally than can any existing Department of Defense (DoD) process, algorithmic or human.

The process of talent management through detailing in the Navy is an arena ripe for such an MIP approach. Due to the structure of many sailors applying to jobs with possibly more than one opening, this issue is a rendition of the college admission problem [4] without the need for stability [10]. The essence of our solution is completed via algorithmic linear programing, specifically the use of MIP. We explore the nuances of how algorithms greatly increase optimal outcomes. Moreover, we discuss how MIP pushes optimality beyond deferred acceptance algorithms which, until now, were considered state-of-the-art for the DoD.

2 Problem Description

Human talent allocation is a resource-intensive process for an institution. This paper outlines the mathematics necessary to lay the groundwork for such a task within one of the most complex institutions of American society – the DoD. We dove into this problem seeking to show the validity of the mathematics, as well as to discover the challenges that exist when creating a talent marketplace from the ground up. Furthermore, despite the small prototype sample, the find the algorithm to be widely scalable, particularly if combined with a proper front-end as our project is the back-end of a future full-stack system.

Market forces are largely the drivers behind a job marketplace – this concept is not new for the American economy. However, given a captive audience such as service members ordered to specific jobs, the mathematical benefit of a benevolent autocracy is immense. Rather than the suboptimal fate to which current servicemembers are subjected, the DoD is already shopping for a better way forward algorithmically. We improve upon what currently exists in all metrics except current implementation; nevertheless, we have a plan with stakeholders to quickly field the technology.

In the Navy, terminology of a detailing marketplace refers to the attempts to detail, or place, warfighters into their next job. This detailing marketplace is a system by which military members can transparently rank their job preferences and the people who own those jobs can supply their preference of incoming personnel. This problem is rather unique to the military due to the aforementioned captive market of many members obligated to remain in service due to contract, desiring to stay in for a pension, or perhaps wishing to stay due to a sense of public service. This market type is complemented by an inability for lateral entry – almost all members need to start from the entry level. Additionally, servicemembers typically change jobs every one to three years.

There are ontological parallels to other communities, too. This process is shared in some regards by the medical school graduates applying to the residency stage of their training. Those graduates apply to a pool of US residency programs. This pool is also rather narrow, mostly coming from US based medical schools, and the specificity of skill set required for success is better understood and measured than success as a military officer.

This general similarity between markets is the reason our initial matching algorithm (Gale-Shapley deferred acceptance algorithm) resembled that of the National Residency Match Program. For context, the Gale-Shapley algorithm's application to the residency matching problem earned the 2012 Nobel Prize in Economic Sciences.

This paper also proposes an optimization-based solution to the matching process. The optimization is not found in the National Residency Match Program due to the need for stability. Stability here is exemplified by medical students ability to reject their assigned position, a choice rarely given to military members, if at all. Thus, optimization in the military context does not require stability. As a result, mathematical optimization, unburdened [7] [8] [11] by the luxury of stability, is at its peak in our formulation – a notable improvement. We found that a traditional algorithm, like Gale-Shapley deferred acceptance, was unconducive to the creation of modular constraints. A constraint may be the introduction of a rule based on Navy policy that the algorithm would have to satisfy before conducting all matches (e.g. at least 95% of dual-military couples ought to be stationed within 50 miles of their partner). For that reason, we transitioned to mixed integer programming, a more agile linear programming approach to the problem. We later show that MIP outperforms the deferred acceptance iteration outright, in addition to its ease of introducing constraints without massive re-coding.

Acknowledging the difficulty of wrangling disparate and dated personnel data, this paper also explores helpful metrics that can be gleaned simply from submitted, ordered preferences by job seekers and job owners. These are specialization, competitiveness, similarity, and preference correlation. These metrics can be used to describe both sides of the market, on the seeker or the owner alike.

Further, the paper ends with suggested metrics that would be gleaned if personnel data beyond mere preferences were accessible, clean, and structured. These include a similarity measure based on quality encodings and a suggested ordering of possible jobs or applicants. AI/ML, specifically deep learning, could enable the latter. The suggested ordering would not make decisions on placement, but rather provide job seekers and job owners with an intelligent ordering, allowing a human participant to more easily distill the vast amount of job information relevant to her.

2.1 Solution Description

Our solution to the talent allocation problem is to optimize the matching of preferences from both sides: job owners (commanding officers) and job seekers (sailors). Inspired by the Gale-Shapley algorithm, we pivoted to mixed integer programming, allowing for greater flexibility in creating constraints without largely rewriting the algorithm.

2.2 Algorithm

A basic algorithm, like Gale-Shapleys set of instructions for stable marriages, is an important first step in talent allocation. We moved onto a linear programming concept, oriented around solvers: algorithms that are adaptable. These solvers have varying types like linear, convex, and mixed integer. While Gale-Shapleys algorithm can be solved in n^2 time, ours is np-complete, indicating polynomial time completion. What we lose in speed, we gain in the ability to alter the formulation without total overhaul. These qualities make the approach particularly conducive to constraint creation, a vital requirement for stakeholders utilizing our solution.

Ours is a unique subset of MIP, binary optimization, to reflect the nature of either placement in a specific job, or not (i.e. only 1 or 0).

The constraints we apply are the same as the college admissions problem [9], where each job can fill up to but not exceed the allocated number of positions and each sailor can only be assigned to one position with the added constraint that the solution is as good or better than deferred acceptance when compared by the post-match metrics in Section 3.3. We chose this to show how this approach can improve upon deferred acceptance and to guarantee feasibility (since the deferred acceptance output is a feasible solution);s the explicit program is in Appendix (B). We also provide alternated formulations in the appendices that may be of more interest to military leaders. These include co-location (Appendix (C)), quality spread (Appendix (D), weighting certain command's preferences over others (Appendix (E)), and retention (Appendix (F)).

2.3 Strategy Proofness

We quite notably solved a problem previously unexplored in mathematical and economic academia. That problem is implied preferences, the notion that a market participant can only rank so many jobs in her day-to-day as a human. To remedy the vast volume issue, we mathematically predict preference rankings based on previous choices without the need for AI/ML. In the future, extensive personnel data will enable deep learning insights to improve upon this implied preference algorithm.

An important note is the attempt at strategy proofness of the algorithm [3] [1], thus creating an incentive for honesty [6]. It is necessary to ensure that there is no benefit for a participant to provide incomplete preferences. Likewise, we must acknowledge that there can be difficulty associated with a market participant ranking an incredibly large number of opportunities. To complete this necessity, we leverage the completion of incompletely provided preferences, the aforementioned implied preference solution [13].

The complete strategy proofness investigation of this formulation is left as an aspect of future work, but we acknowledge that there could be benefits to coalitions or awareness of the preference landscape of competitors. There is also a possibility of rejecting the implied preferences, as one market maker concluded unexpressed preferences indicate indifference [5] – an expression of preference all its own. Nevertheless, the strategy proofness of Irvings approach in an MIP matching process has yet to be explored.

3 Metrics

3.1 Pre-Match Metrics

3.1.1 Competitiveness

Definition

Competitiveness measures the relative desirability of a given sailor or job based on the expressed preference ranking of the other party. We wanted to consider a more sophisticated way of determining this metric beyond an average ranking for either the sailor or job as it is very susceptible to right tail bias.

For example, a position ranked first by ten individuals and hundredth by twenty individuals has the same average ranking as a job ranked seventh by all thirty individuals; yet the former is much more competitive. To attempt to compensate for this potential discrepancy, we adjust the average by a power of one half as to lessen the impact of very low preferences, thereby weighting favorable preference more in our score consideration.

In order to generalize the competitiveness score we scale the resulting average by the total number of jobs or sailors within the system. This allows for competitiveness to range from 1 being most competitive, and 0 being not competitive at all.

$$
\begin{aligned}
F_j &= 1 - \frac{a_j}{M} && \texttt{\% of openings job } j \texttt{ is of all openings in the Navy} && (1)\\
S_i^S &= 1 - \frac{1}{nm}\sum_j \sqrt{P_{ij}^O} && \texttt{sought after metric} && (2)\\
C_i^S &= S_i^S && \texttt{competitiveness of a job seeker} && (3)\\
C_j^O &= F_j S_j^O && \texttt{competitiveness of a job} && (4)\\
& && && (5)
\end{aligned}
$$

To follow the development of this metric, more details are in Appendix (**??**).

Impact

Competitiveness will be used to rapidly identify the most attractive candidates (e.g. Sailor A has a 0.98 competitiveness score and is therefore a top ranked candidate amongst all job owners). These candidates would be good fits in many jobs and therefore could be considered ideal individuals to screen for command – the selection process for being a higher ranking commanding officer. A prominent stakeholder in our study, the Chief of Naval Personnel, has consistently sought after a way to rank sailors and evaluate their competitiveness as well. This is one way.

On the job side, a low competitiveness would be a great way to decide on incentive structures. These jobs are those that very few sailors prefer and therefore ought to be incentivized to be filled, perhaps with the promise of promotion or some additional sort of compensation to target the right sailor to be willing to take an undesirable post.

3.2 Specialization

Definition

Specialization measures the extent to which the maximum desire for a sailor or job diverges from the mean. Sailor A is highly specialized if one job owner highly desires her much more than the mean job owner (e.g. Sailor A is the 1st preference choice of Job Owner X, but is on average, most job owners rank her as their 21st choice). In this example, Sailor A is a specialist to the extent that her skill set will serve a specific function well (with great benefit to Job Owner X).

This comparative advantage over her peers represents a disproportionately positive contribution to the Navy by her being placed in that role compared to any other roles for which she may be qualified, the same as the importance of selecting her in relation to other more competitive but less specialized candidates who could also work for Job Owner X.

$$\texttt{Specialization}_i^S = \frac{\text{kurtosis}\{P_i^O\}}{\text{skew}\{P_i^O\}}$$

Impact

A sailor with a high specialization has some skill set that will serve a specific job function well; the sailor has an advantage over peers for a single role compared to the other roles for which the sailor is qualified. If most job owners rate a sailor with low preference but a single owner rates that sailor with a high preference, the marginal gains of the sailor being matched with that owner are high. Matching the sailor with this particular role utilizes her specialization and produces a disproportionate positive contribution to the Navy.

Intuitively, the matching algorithm will favor creating high-specialization matches. Failing to make these matches has a more negative impact on the system than failing to make non-specialized matches, because the usefulness of the sailor goes down significantly between their optimal match and average matches. This metric may be useful in the future to examine if very particular aspects of jobs or sailors make them attractive to small subsets of the other population.

3.3 Post-Match Metrics

Preference Allocation Windows

Preference allocation indicates how many individuals received their top preference, how many individuals got a top one, top five, and top ten preference, and how many failed to be matched with any preference.

Average Preference Match

Preference match is the sum of the ordinal preference of a sailor for their assigned command, and the ordinal preference of the command for their assigned sailor. For example if a sailor is assigned their most preferred command but the command ranked the sailor their, the preference match would have a value of 4. This metric provides a high level understanding of how the matching system performs in aggregates.

4 Running the Prototype

4.1 Overview

The essential effort of this paper is to demonstrate the mathematical underpinnings of algorithmic matching for a Defense application. The pilot implementation of this algorithm, however, was a very human process. At the most foundational level, the algorithm solely requires preferences from job seekers (sailors) and job owners (commanding officers). In order for these preferences to be well informed, the system ought to have standardized job descriptions and sailor resumes. Obtaining this central information proved to be the most difficult aspect of piloting the algorithm rather than the mathematics, no matter how novel the MIP approach or exquisite the introduction of implied preferences.

Even Alvin Roth, who won a Nobel Prize for his foundational matching algorithm, experienced the same human and bureaucratic difficulties. Roth suggests, "overall, one lesson from the [matching project] is that mechanism design in a political environment requires that not only policy makers themselves be persuaded of the virtues of a new design, but that they be able to explain and defend the mechanism to the various constituencies they serve." [1] The explainability of our algorithm is relatively simple in that it matches preferences from two sides – seekers and owners – yet any sort of deviation from existing processes creates concern and risk for senior stakeholders, real or imagined. Like building a boat in a bottle, the art of the prototype was not so much the algorithm itself, but the delicacy needed to create it within a debilitatingly constrained environment.

To best alleviate these issues, our team leveraged a web-based polling tool that allows both sailors and commanding officers to input their preferences. Sailors were given job descriptions and commanding officers were given sailor bios in order to inform their respective preferences. In order to test it across larger data sets, however, we introduced a slew of sources described in the following section.

4.2 Data Sources

Our algorithm was run against three total data-sets.

Naval Cyber Officers

The Navy has various types of service. Some officers work with computers rather than operating aircraft or driving ships. We worked with one subset of cyber-oriented cryptologic warfare officers to attempt a bottom-to-top pilot of the matching algorithm. That required an novel data collection effort because the previous method of matching sailors to jobs was completed non-optimally, largely based on who was available at the time. This dataset proved to be one of the most difficult to collect because there was no collection infrastructure or existing practice of doing so.

Naval Explosive Ordinance Disposal Officers (EOD)

EOD officers are Naval expeditionary special forces who deal primarily with the handling, disarming, and disposing of explosive materials across the world. Given the specialization of their work, the small communitys leadership have expressed a desire for a matching mechanism to identify and manage its peoples talents. Working closely with that leadership, our team was able to gather a large amount of preferences on either side of the marketplace (seekers and owners).

Tangentially, as evidenced by the near impossibility of obtaining data in the cyber community due to a lack of precedent, the EODs precedent extended to their funding a front end collection mechanism. Later in this piece we will advise the Navy to combine our algorithmic back end to the EOD communitys budding front end to create a full stack system with phenomenal potential for scalability.

Naval Doctors (Medical Corps)

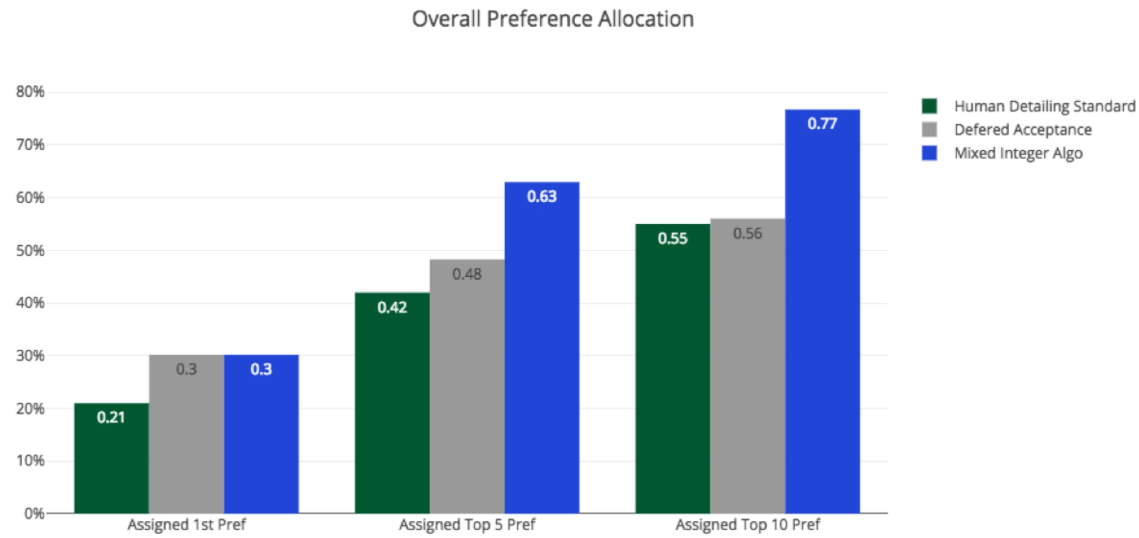
The Navy has its own doctors. These doctors may go to civilian medical schools or the governments own medical school (Uniformed Services University Hebert School of Medicine, or USUHS). Additionally, they conduct their residencies at public or private hospitals. As a result of that process, they are involved in the National Resident Matching Program, a foundational aspect of this paper.

Their familiarity with the Gale-Shapley matching algorithm, used by the Navy as well as National Residency match in parallel processes, the culture of the medical corps is such that data is collected and a nearly identical process is run frequently. As a result, our request for data from this community was met willingly and completely, offering a great opportunity to run clean data as well as provide our algorithm back to that excited customer most seamlessly.

4.3 Results

4.3.1 Algorithm Results

Given that our algorithm does not need to maintain stability, we are able to further optimize as compared to a deferred acceptance approach (the one used for hospital residents). Our algorithm is shown to measurably improve placement at each incremental preference level. The improvements are even more notable when compared to the current non-algorithmic matching effort conducted by humans with imperfect information and an overwhelming amount of data.



Above is the plot for the Medical Corp data, the plots for all our data sets can be found in Appendix (REF APPENDIX).

4.3.2 Post-Match Metrics Results

The plot of Specialization vs. Competitiveness has a negative correlation. The negative correlation is expected and thus further validates the formulation. A highly competitive market participant is desired by multiple firms, thus their skills generalize. A highly specialized market participant is desired strongly by few firms, making them un-competitive across the market when looked at in its entirety. Both competitive sailors and specialized sailors have a place in the Navy, and this metric validates their best placement.

Above is the plot for the Medical Corp data, the plots for all our data sets can be found in Appendix (REF APPENDIX).

5 Future Work

5.1 Inspired from Academia

Ongoing work should continue to explore the best ways in which markets can be created. This includes consideration of partial matches or lotteries in MIP. [2] One may also consider the ability to explore the viability of leveraging Budish's wagering formulation of approximate competitive equilibrium from equal incomes. [3] Additional research would include how to incorporate synergy of selection preferences by Job owners if allowed more than one person. [8] Further, there could be that service members could be place in multiple jobs, such as selecting their main role and their collateral duties in the same matching process. [6] Lastly, an open question the authors are curious about is the opportunity for unsupervised learning (such as K-means clustering) on the preference data to see if there are clusters of service members with regards to their preferences, and what theres clusters indicate.

5.2 Asks from DoD Personnel

From interviewing DoD leadership there are several institution specific asks. Many of our considered markets have an on-line placement procedure [12], such as that which happens at many information-sensitive commands due to the trickle of clearance issuance as opposed to bulk assignment.

In the same spirit of temporal considerations, military members have set time-lines at each command and are given expected rotation dates. Consideration of overlapping rotation and end strength could give another direction to objective function formulation.

Sometimes military members have the option to take orders or separate from the service. A future formulation could allow service members to express ordinal preference up to a point, and then indicate if they do not receive any of those they would opt to separate from the service. A constraint, we would imagine, would need to be added with respect to retention on each detailing cycle.

Tailored Compensation decisions could be made for uncompetitive, unspecialized positions the DoD needs filled.

Explore strategic importance of positions. This can either be an iteration where billets in priority tranche's are run iterative (tier 1 billets all matched, those sailors and billets are taken out of the pool, then tier 2 billets all matched, etc.), or a weighting where in a single matching optimization the preferences of higher tiered billets are given a greater weighting.

Forthcoming work on verifying and validating our implementation of inferred preferences will continually be bolstered as well [13].

References

[1] P. Pathak A. Abdulkadiroglu, A. Roth and Sonmez T. Changing the boston school choice mechanism. Technical report, Boston Public Schools, January 2006.

[2] U. Rothblum A. Roth and J. Vande Vate. Stable matchings, optimal assignments, and linear programming. *Mathematics of Operations Research*, 18(4), November 1993.

[3] E.Budish. The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes. *Journal of Political Economy*, 119(6), 2011.

[4] D. Gale and L. S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69:9–15, January 1962.

[5] R. Irving. Stable marriage and indifference. *Discrete Applied Mathematics*, 48:261–272, 1994.

[6] A. Roth. The economics of matching: Stability and incentives. *Mathematics of Operations Research*, 7(4), November 1982.

[7] A. Roth. Stability and polarization of interests in job matching. *Econometrica*, 52(1):47–57, 1984.

[8] A. Roth. The college admissions problem is not equivalent to the marriage problem. *Journal of Economic Theory*, 36:277–288, 1985.

[9] A. Roth. Conflict and coincidence of interest in job matching: Some new results and open questions. *Mathematics of Operations Research*, 10(3), August 1985.

[10] A. Roth and M. Sotomayor. Interior points in the core of two-sided matching markets. *Journal of Economic Theory*, (45):85–101, 1988.

[11] A. Roth and M. Sotomayor. The college admissions problem revisited. *Econometrica*, 57(3):559–570, May 1989.

[12] S. Mitchell S. Khuller and V. Vazirani. On-line algorithms for weighted bipartite matching and stable marriages. *Theoretical Computer Science*, 127:255–267, May 1994.

[13] I. Shaw. Completing incomplete ordinal preferences. Paper is in the process of being published., March 2019.

Appendices

A Notation

Throughout the paper, we will reference the notation listed in this section.

$$
\begin{aligned}
m &= \text{number of different jobs available} && (6)\\
n &= \text{number of persons} && (7)\\
\vec{P_i^S} &= \text{Preference vector of job seeker } i, \ \in \mathbb{Z}^{+,m\times 1} && (8)\\
P^S &= [\vec{P_i^S}|\ldots|\vec{P_n^S}] \in \mathbb{Z}^{+,m\times n}, && (9)\\
& \quad \text{Preference Matrix of Seekers} && (10)\\
\vec{P_i^O} &= \text{Preference vector of job owner } j, \ \in \mathbb{Z}^{+,n\times 1} && (11)\\
P^O &= [\vec{P_i^O}|\ldots|\vec{P_n^O}] \in \mathbb{Z}^{+,n\times m}, && (12)\\
& \quad \text{Preference Matrix of Job Owners} && (13)\\
\vec{A} &= \text{Position Available vector } \ \in \mathbb{Z}^{+,m\times 1} && (14)\\
a_j &= \text{Amount of positions for job } j, \ \in \mathbb{Z}^{+} && (15)\\
M &= \sum_i a_j && (16)\\
&= \ \text{number of openings across Navy} && (17)\\
M &\geq m && (18)\\
X &= \text{Placement Matrix} \in \{0,1\}^{n\times m} && (19)\\
x_{i,j} &= \begin{cases} 1 & \text{if } S_i \text{ is slated for job } j \\ 0 & \text{otherwise} \end{cases} && (20)\\
C &= \text{Co-Location Matrix, upper triangular} && (21)\\
C_{ij} &= \begin{cases} 1 & j > i, \ \text{and Seeker } i \text{ requests co-location with Seeker } j \\ 0 & j \leq i, \ \text{or Seeker } e_{i,1} \text{ does not request co-location} \end{cases} && (22)\\
n_c &= \text{number of couples requesting co-location} && (23)\\
&= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij} && (24)\\
& && (25)
\end{aligned}
$$

B MIP Matching Formulation

The optimization function used in our investigation takes the form:

$$\min \quad \sum_{i=1}^{n}\sum_{j=1}^{m} f(X_{ij}) \tag{26}$$

$$\text{such that} \quad \sum_{j=1}^{m} X_{i,j} \leq 1 \quad \forall i \in \{1, \dots n\} \tag{27}$$

$$\sum_{i=1}^{i=n}\sum_{j=1}^{m} X_{ij} = \min\left(n, \sum_{j=1}^{m} a_j\right) \quad \forall j \in \{1, \dots m\} \tag{28}$$

$$\sum_{i=1}^{n} X_{ij} \leq a_j \quad \forall j \in \{1, \dots m\} \tag{29}$$

$$\sum\sum X_{ij}\mathbf{1}(X_{ij}P_i^S \leq w) \geq \sum\sum X_{ij}\mathbf{1}(X_{ij}^{DA}P_i^S \leq w) \quad \forall w \in \{1, 5, 10\} \tag{30}$$

$$\sum\sum X_{ij}\mathbf{1}(X_{ij}P_{ji}^O \leq w) \geq \sum\sum X_{ij}\mathbf{1}(X_{ij}^{DA}P_{ji}^O \leq w) \quad \forall w \in \{1, 5, 10\} \tag{31}$$

$$\tag{32}$$

The Goodness Function f is the strategic objective function of the assignment process. For the sake of this paper, we set it to value the preference of the seeker twice as much as the preference of th job owner.

$$f(X_{ij}) = X_{ij}(P_{ij}^S + P_{ji}^O)$$

B.1 Matrix Formulation

In matrix form this can be re-written (with $tr()$ indicating the trace):

$$\min \quad tr(XP^O) + tr(X^T P^S) \tag{33}$$

$$\text{such that} \quad X^T \times 1^{m\times 1} \leq 1 \tag{34}$$

$$(X \bullet 1) \bullet 1 = \min(n, A \bullet 1^{m\times 1}) \tag{35}$$

$$X \times 1^{n\times 1} \leq A \tag{36}$$

$$\sum\sum X \bullet \mathbf{1}(X \bullet P^S \leq w) \geq \sum\sum X \bullet \mathbf{1}(X^{DA} \bullet P^S \leq w) \quad \forall w \in \{1, 5, 10\} \tag{37}$$

$$\sum\sum X \bullet \mathbf{1}(X \bullet (P^O)^T \leq w) \geq \sum\sum X \bullet \mathbf{1}(X^{DA} \bullet (P^O)^T \leq w) \quad \forall w \in \{1, 5, 10\} \tag{38}$$

$$\tag{39}$$

The proof for the equivalence of the objective functions (26) and (33) is in Section B.1.1. The constraints in lines (27) and (34) ensure that each seeker receives only one job, lines (28) and (35) ensure that either all the jobs are filled or everyone has a job, lines (29) and (36) ensures that all jobs are at or below capacity (do not exceed capacity), and lines (31, 32, 38, 39) ensure that our process is just as good or better in the post-match metrics of preference windows described in Section 3.3.

B.1.1 Matrix formulation of Objective Function

$$
\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{m} f(X_{ij}) &= \sum_{i=1}^{n}\sum_{j=1}^{m} X_{ji}\left(P_{ji}^{S} + P_{ij}^{O}\right) && (40)\\
&= \sum_{j=1}^{m}\sum_{i=1}^{n} X_{ji}\left(\left(P^{S}\right)_{ij}^{T} + P_{ij}^{O}\right) && (41)\\
&= \sum_{j=1}^{m}\sum_{i=1}^{n} X_{ji}\left(\left(P^{S}\right)^{T} + P^{O}\right)_{ij} && (42)\\
&= \sum_{j=1}^{m}\left(X\left(\left(P^{S}\right)^{T} + P^{O}\right)\right)_{jj} && (43)\\
&= tr\left(X\left(\left(P^{S}\right)^{T} + P^{O}\right)\right) && (44)\\
& && (45)
\end{aligned}
$$

B.2 Formulation for CVXPY

The difficulty with the constraints in lines (31, 32, 38, 39) is that CVXPY does not have an indicator or filter atomic function that can be used in the objective function. Thus in order to code this formulation with the CWVPY we reformulated using the functions max, which is an atomic function in the CVXPY API.

$$
\begin{aligned}
x &= X_{ij} && (46)\\
p &= P_{ij}^{S} && (47)\\
g &\in \{1, 5, 10\} && (48)\\
\mathbf{1}(xp \leq g) &= \mathbf{1}(xp - g \leq 0) && (49)\\
&= \mathbf{1}(\max(xp - g, 0) == 0) && (50)\\
&= \max(1 - \max(xp - g, 0), 0) && (51)\\
& && (52)
\end{aligned}
$$

The jump to the last line is made possible by the fact that xp and g are positive integers.

C MIP Co-Location Formulation

The co-location formulation takes the form:

$$\min \quad \sum_{i=1}^{i=n}\sum_{j=1}^{j=m} f(x_{i,j}) \tag{53}$$

$$\text{such that} \quad \sum_{j=1}^{m} x_{i,j} \leq 1 \quad \forall i \in \{1,\dots n\} \quad \texttt{only one job per person} \tag{54}$$

$$\sum_{i=1}^{i=n}\sum_{j=1}^{j=m} x_{i,j} == \min\left(n, \sum_{j=1}^{m} a_j\right) \quad \forall j \in \{1,\dots m\} \tag{55}$$

$$\texttt{either all the jobs are filled or everyone has a job} \tag{56}$$

$$\sum_{i=1}^{n} x_{i,j} \leq a_j \quad \forall j \in \{1,\dots m\} \quad \texttt{all jobs are at or below capacity} \tag{57}$$

$$\frac{1}{n_c}\sum_{i=1}^{n_e} C(e_1, e_2) \geq 0.95 \quad \texttt{at least 95\% of couples are co-located} \tag{58}$$

C.1 Matrix Formulation

$$\min \quad f(X) \tag{59}$$

$$\text{such that} \quad \sum_{j=1}^{m} x_{i,j} \leq 1 \quad \forall i \in \{1,\dots n\} \quad \texttt{only one job per person} \tag{60}$$

$$(X^T \bullet 1) \bullet 1 == \min(n, A \bullet 1) \tag{61}$$

$$\texttt{either all the jobs are filled or everyone has a job} \tag{62}$$

$$X^T \bullet 1 \leq A \quad \texttt{all jobs are at or below capacity} \tag{63}$$

$$\frac{1}{n_c}\sum_{i=1}^{n_e} X^T C X D \geq 0.95 \quad \texttt{at least 95\% of couples are co-located} \tag{64}$$

Co-location Function C returns 1 if the couple is considered co-located, 0 if not or if single. Here we choose 50 miles between job locations to be consider co-located because that is the threshold for receiving dislocation allowance (DLA) for a permanent change of station (PCS) according to the Joint Travel Regulations (JTR). The location function $L(S_i)$ returns the lat/long location of the stationing for Seeker i.

$$C(e_1, e_2) = \begin{cases} 0 & \text{if } e_2 == 0 \\ \mathbb{1}\big(||L(S_i) - L(S_j)|| \leq 50\big) & \text{otherwise} \end{cases}$$

The inspiration and initial formulation of this optimization was done by a young Air Force officer who has since moved onto the private sector.

The extension of the formulation to include the Co-Location Function C and have more than one positions for each job a_i are the contributions of this team.

C.1.1 Explanation of Co-location constraint math

$$\frac{1}{n_c}\sum\sum X^T C X \geq 0.95$$

This works because

$$(CX)_{ij} = \begin{cases} 1 & \text{Seeker } i\text{'s mate is assigned to job } j \\ 0 & \text{otherwise} \end{cases}$$

$$(X^T)_{ij} = \begin{cases} 1 & \text{Seeker } i \text{ is assigned to job } j \\ 0 & \text{otherwise} \end{cases}$$

$$(X^T CX)_{ij} = \begin{cases} n & \text{Number of couples assigned to the } (i,j) \text{ job pairing} \\ 0 & \text{otherwise} \end{cases}$$

$$D_{ij} = \begin{cases} 1 & \text{The job pairing } (i,j) \text{ is considered co-location} \\ 0 & \text{otherwise} \end{cases}$$

$$(X^T CX)\dot{D} = \text{ The Hadamard (element-wise) multiplication of } (X^T CX) \text{ and } D$$

$$\big((X^T CX)\dot{D}\big)_{ij} = \begin{cases} n & \text{Number of co-located couples assigned to the } (i,j) \text{ job pairing} \\ 0 & \text{otherwise} \end{cases}$$

$$\sum\sum(X^T CX)_{ij} = \text{Number of couples co-located}$$

$$\frac{1}{n_c}\sum\sum(X^T CX)_{ij} = \text{Ratio of couples co-located}$$

D MIP Quality Spread Formulation

There is precedent in the DoD to do assignments in a manner by which the distribution of quality officers is even across assignments. This happens in how US Navy submariners are assigned a boat after initial training and to US Marines when selecting their Marine Occupational Specialty (MOS). The purpose if this is to ensure that no one unit or community within the force falls behind in competency due to a preponderance of low-performing personnel.

To incorporate this into the matching algorithm would be easy. We would simply add this to the objective function of Appendix (B).

$$f(X) = \gamma(P^S + P^O) + \lambda(\sigma^2(X))$$

The explanation of the objective function can be found in subsection D.1.

D.1 Explanation of Objective Function

$$f(X) = \gamma(P^S + P^O) + \lambda(\sigma^2(X)) \tag{65}$$

$$\gamma, \lambda = \texttt{weighting coefficient,} \in [0,1] \cap \mathbb{R} \tag{66}$$

$$\sigma^2(X) = \texttt{Variance of the quality assignments} \tag{67}$$

$$\sigma^2(X) = \frac{\sum_{j=1}^{m}(\mu_j^q - \bar{\mu^q})^2}{m-1} \tag{68}$$

$$\mu_j^q = \texttt{Average Quality score of person assigned to position } j \tag{69}$$

$$\mu_j^q = \sum_{i=1}^{n} \frac{X_{ij}Q_i}{A_j} \tag{70}$$

$$\bar{\mu^q} = \sum_{j=1}^{m} \frac{\mu_j^q}{m} \tag{71}$$

$$Q = \texttt{Quality scores for each seeker ,} \mathbb{R}^{n\times 1} \tag{72}$$

E MIP Importance Formulation

Suppose service leadership were to favor some sort of command. This could look like giving operational commands twice the value preference over administrative commands, or favoring certain mission sets over others. This can easily be incoporated by a tweak to the objective function.

$$f(X) = (P^S + WP^O)X$$

Here the vector $W \in \mathbb{R}^{m \times 1}$ is a weighting vector. For example, if the service leadership were to value a command #1 which is highly operational twice that of any other command then

$$W_i = \begin{cases} 2 & i = 1 \\ 1 & \text{otherwise} \end{cases}$$

F MIP Retention Formulation

Retention is often and important goal of military leaders wanting to maintain manpower numbers in order to keep the force ready for possible operations. The difficulty is, with an all volunteer force, if a service member is beyond contractual obligation, they need to be incentivized to stay in. For example, a sailor may be willing to stay in for another 3 years if they are guaranteed to stay in San Diego so that their child can finish out high school there, but if they are asked to move across the country they will opt to leave the service and seek local employment outside of the DoD.

For the purpose of controlling retention numbers, we can adapt the formulation from Appendix (B). The objective function can now take the form below, if the goal is to maximize retention (minimization is just a sign change).

$$f(X) = \gamma(P^S + P^O) - \lambda(R(X))$$

The explanation of the objective function can be found in subsection F.1.

Also we can incorporate a retention constraint. Say a service leader were to desire retention above 90%, then the following constraint would be necessary.

$$R(X) \geq \texttt{Desired retention rate between 0 and 1}$$

Here $R(X)$ is the same as that in the objective function fomulation, defined at 77.Keep in mind, there is maximum rate on retention that is not necessarily 100%, see subsection F.2 for a discussion on this. Furthermore, when employing this constrain, special care must be taken to guarantee feasibility.

We believe the implementation of either or both of these retention formulations to be strategy proof (indicatingn it is in the best interest of market participants to honestly express their ordinal preferences) so long as the output of X is enforced. Yet, we have not explored the strategy proofness regorously and leave that investigation to futrue work.

F.1 Explanation of Objective Function

$$f(X) = \gamma(P^S + P^O) + \lambda(\sigma^2(X)) \tag{73}$$

$$P_j^S = \begin{cases} m+1 & \text{Would rather separate than accept assignment to } j \\ [1, m] \cap \mathbb{Z} & \text{otherwise expressed preference} \end{cases} \tag{74}$$

$$\gamma, \lambda = \texttt{weighting coefficients}, \in [0,1] \cap \mathbb{R} \tag{75}$$

$$R(X) = \texttt{Retention rate of assignment set } X \tag{76}$$

$$R(X) = \frac{\sum_{i=1}^{n} \mathbf{1}\left(X \bullet P^S \langle m+1\right)}{n} \tag{77}$$

$$\tag{78}$$

F.2 Maximum Retention

The maximum retention rate given the constraints of the college admission formulation (defined in lines (27, 28, 29), is the value of the objective function for the program described below. Keep in mind these constraint must be feasible for this analysis to be worthwhile; this requires that there is at least one solution where there are is at least one job seeker willing to fill each position in the market.

$$
\begin{aligned}
\max \quad & R(X) && (79)\\
\text{such that} \quad & \sum_{j=1}^{m} X_{i,j} \leq 1 \quad \forall i \in \{1, \dots n\} && (80)\\
& \sum_{i=1}^{i=n} \sum_{j=1}^{m} X_{ij} = \min\left(n, \sum_{j=1}^{m} a_j\right) \quad \forall j \in \{1, \dots m\} && (81)\\
& \sum_{i=1}^{n} X_{ij} \leq a_j \quad \forall j \in \{1, \dots m\} && (82)\\
& && (83)
\end{aligned}
$$

G More Results

G.1 Pre-Match Metrics

G.2 Post-Match Metrics