Multi-Target Tracking
via Mixed Integer Optimization

Dimitris Bertsimas, Zachary Saunders, and Shimrit Shtern

Abstract—The field of multi-target tracking (MTT) faces two primary challenges: (i) data association and (ii) trajectory estimation. Many algorithms, such as the MHT and JPDAF, attempt to solve the MTT problem via probabilistic approaches, while few employ the use of optimization models. In this paper we take a new approach to the MTT problem by modeling it as a mixed integer optimization (MIO) model, making no probabilistic assumptions on the data. We model the basic case where there is no detection ambiguity, *i.e.*, no false alarms or missed detections, and extend it to the more difficult robust case where there exists detection ambiguity. These models solve the data association and trajectory estimation problems simultaneously by minimizing an easily interpretable global objective function. Additionally, we propose a greedy heuristic that scales efficiently and finds high quality solutions in milliseconds, and we show that these heuristic solutions can be used effectively as a warm start for the MIO, enabling the MIO to achieve good solutions in seconds. In the basic case, the model has no tunable parameters and the extension to the ambiguous case requires only two easily understood parameters which serve as objective function penalties. Furthermore, we introduce a new metric for measuring complexity in the data association problem as well as a metric for measuring the performance of trajectory estimation. Numerical results show that in the case of the basic model we obtain a good reconstruction of the original tracks for a range of scenarios, while in the robust case, for small to medium sized scenarios we are able to both estimate the correct number of targets as well as correctly reconstruct the trajectories.

Index Terms—data association; mixed integer optimization; multi-target tracking; optimization; trajectory estimation

I. INTRODUCTION

Multi-target tracking is the problem of estimation the state of multiple dynamic objects, referred to as *targets* over a fixed window of time. At various points of time within the window, the targets are observed in a *scan*, resulting in set of *detections*. From these detections, the multi-target tracking problem aims to extract information about target dynamics.

Solutions to this problem are sought across many civilian and military applications including but not limited to ballistic missile and aircraft defense, space applications, the movement of ships and ground troops, autonomous vehicles and robotics, and air traffic control. Each application has unique attributes and assumptions, and various algorithms have been developed for each. As a result, the field of multi-target tracking has expanded to numerous research venues, and there is a wide range of literature on the topic. A complete overview of all MTT methods, including the classes of algorithms and their variants as well as additional methods not discussed in this paper, can be found in [1]. For a more exhaustive overview of estimation techniques, filtering, gating, and more please see [2] and [3].

The field of multi-target tracking faces two primary challenges: (i) data association and (ii) trajectory estimation. Given a set of sensor detections, the data association problem consists of assigning the detections to a set of targets. Alternatively, this can be viewed as a labeling problem in which each detection needs to be labeled with a target identifier. The association problem is further complicated when sensors fail to report detections (missed detection) or incorrectly report detections (false alarm), resulting in ambiguity in the number of existing targets. The trajectory estimation problem consists of estimating the state space of a target (*i.e.*, position, velocity, acceleration, size, etc.) from the associated detections of the aforementioned assignment problem. Even when all of the associations are known, the estimation problem is challenging due to the presence of measurement noise. As can be seen, the two problems of data association and trajectory estimation are closely related and dependent on one another.

Some classical algorithms treat the data association and trajectory estimation problems separately using a combination of probabilistic approaches to determine data associations and filters to estimate trajectories. One such algorithm is the global nearest neighbor (GNN). The GNN algorithm is a naive 2-D assignment algorithm, which evaluates one scan of detections at a time, globally assigning the nearest detection at each scan [4]. Once the data association has been determined, the detections are often passed through one of numerous filters, most commonly a Kalman filter [5], which updates the trajectory estimates before the algorithm progresses forward to the next scan. This process repeats sequentially through each scan of data.

Modern algorithms in the field of multi-target tracking are most commonly statistical based, often relying on heavy probabilistic assumptions about the underlying target dynamics or detection process. The two most prevalent statistical algorithms in the field of multi-target tracking are the Multiple Hypothesis Tracker (MHT) and the Joint Probability Data Association Filter (JPDAF) and their numerous variants and extensions. Both classes of algorithms attempt to solve the data association problem by generating a set of potential hypotheses, or possible detection-to-track assignments. Here a *track* is a set of labeled detections belonging to the same target. Probabilities are assigned to each hypothesis based on the likelihood of the trajectory's existence, and numerous approaches for accomplishing this task have been proposed.

The MHT, first proposed by Reid in [6], assigns likelihood values to hypotheses using a Bayesian maximum a posteriori estimator, which requires probabilistic assumptions on both object dynamics and detection process. This algorithm is

generally considered to be the modern standard for solving the data association problem. Many variants have been proposed for implementation which leverage techniques such as clustering, gating, hypothesis selection, hypothesis pruning, and merging of state estimates. Many of these methods are summarized in [7].

While the MHT has seen various forms of success, it faces several key challenges. Namely, the curse of dimensionality and complexity. The number of possible hypotheses grows exponentially with the number of potential tracks and the number of scans. Consequently, it is considered intractable for large scenarios. Moreover, the MHT might require extensive tuning and thus may be difficult to implement in practice, in addition to being computationally expensive. For these reasons, it is generally considered to be one of the most complex MTT algorithms.

A Probability Data Association (PDA) takes a Bayesian approach to solving the data association problem by finding detection-to-target assignment probabilities via a posterior PDF, which again requires heavy assumptions on object dynamics and the detection process. In similar fashion, a Joint PDA (JPDA) assigns probabilities that are computed *jointly* across all targets. The JPDAF is an algorithm which implements the JPDA along with filters and estimation methods as discussed previously in [2].

A limited number of optimization based algorithms have been applied to solve the MTT problem, most of which attempt to solve by mapping the measurement set onto a trellis and seek the optimal measurement association sequence. Some examples include the Multi-Target Viterbi [8] and an extension in [9] which formulates [8] as a network flow, reducing the solve time from exponential to polynomial. Still others, in particular [10], have suggested adaptations of this approach that output a single best set of tracks, or a list of best sets of tracks, similarly to the MHT.

Compared to the number of statistically based algorithms in the MTT literature, optimization based algorithms are relatively lacking. In fact, most occurrences of optimization in the MTT literature propose the use of optimization to leverage statistical algorithms, in particular, the MHT. For example, integer optimization has been used to improve MHT hypothesis selection by solving an assignment problem which chooses the best hypothesis, but only after costs have been assigned (statistically based) and hypotheses have been pruned [11]. Somewhat similarly, linear optimization has also been used to assist in the hypothesis selection process for the MHT [12]. Still, other attempts aim to improve the MHT hypothesis selection process via Lagrangian relaxation [13].

More recently, Andriyenko and Schindler have proposed formulating the MTT problem as a minimization of a continuous energy [14], and then again as a minimization of discrete-continuous energy [15]. These algorithms aim to more accurately represent the nature of the problem, but sacrifice interpretability for complexity in the process. Rather than formulating the problem to lend it easily to traditional global optimization methods, the authors intend to leverage the use of optimization techniques to find strong local minima of their proposed energy objective, and they achieve strong results in doing so. However, this approach calls for the use of several parameters that must be tuned and few recommendations are provided for how to go about such a tuning process. Additionally, these methods require initialization heuristics to begin the solving process, which is in itself complicated to implement and is not directly connected to the optimization problem solved.

In this paper, we propose the use of mixed integer optimization (MIO) to formulate and solve the multi-target tracking problem. Although MIOs are generally thought to be intractable (NP-Hard), in many practical cases near optimal solutions and even optimal solutions to these problems can be obtained very efficiently [16]. This can be attributed to the fact that MIO solvers have seen significant performance improvements in recent years due to advancements in both methodology and hardware. The development of new heuristic methods, discoveries in cutting plane theory, and improved linear optimization methods have all contributed to improvements in performance [17]. Modern solvers such as Gurobi and CPLEX have been shown to perform extremely well on benchmark tests. In the past six years alone, Gurobi has seen performance improvements by a factor of 48.7 [18]. CPLEX saw improvements by a factor of 29,000 from 1991 to 2007 [19]. From 1994 to 2014, the growth of supercomputing power as recorded by the TOP500 list has improved by a factor of 567839 [20]. Thus, the total combined effective improvement of software and hardware advancements is on the scale of 800 billion times in the past 25 years.

The literature is also lacking in performance metrics for the evaluation of MTT algorithms. There is no standard method of measuring scenario complexity or algorithm performance as a function of this complexity. In many cases, only the sensor's detection noise is taken into account and other factors such as target density are negated. Recent work [21] proposes a mathematically rigorous performance metric for measuring the distance between ground truth and estimated track, but there is not much attention given to the complexity of generated scenarios. In this paper, we also suggest measures of complexity and performance which are related to the ones suggested in [21], but we show the value in relating a complexity measure to performance measures, namely that it allows us to evaluate the data association and trajectory estimation problems separately. We evaluate the methods suggested in this paper using these complexity and performance measures on two simulated experiments.

The main contributions of this paper are as follows:

(i) We introduce a simple interpretable MIO model which solves the data association and trajectory estimation problems simultaneously for a sensor with no detection ambiguity. The model does not require assumptions on data generation or any tuning of parameters. This MIO is practically tractable, in the sense that it obtains high quality solutions in reasonable amount of time for the considered applications.

(ii) We propose a simple heuristic, motivated by the optimization problem, which provides feasible solutions to this problem and show how it can be used as warm start to the MIO in order to improve the quality of

the solutions obtained as well as the running time. This heuristic is highly scalable and parallelizable, solving in milliseconds.

(iii) We extend this basic MIO model and corresponding heuristic initialization algorithm for the case of detection ambiguity, i.e., the case where there are both missed detections and false alarms, keeping interpretability while only adding two tunable parameters, as well as provide general guidelines as how to tune these parameters.

(iv) We present a new measure of complexity for the data association problem and show how it allows scenario generation and performance to be measured separately in each of their own natural demands. We also discuss a simplified measure of performance for the trajectory estimation problem.

The paper structure is as follows. We begin with a description of the MTT problem as we wish to model it in Section II. In Section III we introduce a simple MIO formulation for a sensor with no detection ambiguity and extend it to a generalized formulation. Then we present a randomized local search heuristic in Section IV, which we use as a warm start for the MIO. In Section V we discuss extensions to both the MIO model and the heuristic for the case of detection ambiguity. In order to quantify the performance of our suggested methods, we develop metrics for measuring scenario complexity and algorithm performance in Section VI. Experimental methods and computational results are presented in Section VII, including results for scenarios both with and without detection ambiguity. Finally, we summarize our contributions and identify future work in Section VIII.

General Notations: Unless specified otherwise, $\|\cdot\|$ is used to indicate the ℓ_1 norm, and $|\cdot|$ refers to element wise absolute value.

II. PROBLEM DESCRIPTION

In this paper, we restrict our exploration of the MTT problem to the automatic tracking of multiple, independent point targets using a single sensor. A *target* is the object of interest. A point target's only identifiable attributes are its state space, which we restrict to position and velocity. The state space fully defines the field of *trajectories*, or paths along which targets travel. A *detection* is collected from each target at sequential scans. Detections are subject to noise. We treat two general scenarios: with and without detection ambiguity.

When there is no detection ambiguity, the sensor produces exactly one detection for each target in each scan, and there is no other source of detections. Therefore, the number of detections in each scan exactly equals the number of targets in existence. Under these conditions, the data association problem reduces to a one-to-one assignment problem. Our basic optimization model, presented in Section III aims to model this variant of the MTT problem.

Detection ambiguity refers to a more complex case where the sensor generates both false alarms and missed detections. A *false alarm* occurs when a detection is collected when in fact no target exists. This could be the result of measurement error or difficulties in the sensor's signal processing. A *missed*

detection occurs when a data point is not collected in a given scan where a target does in fact actually exist. Therefore, in presence of detection ambiguity, the number of detections in each scan could be higher or lower than the actual number of existing targets, and the number of targets can not be immediately deduced from the number of detections. Under these conditions, each detection can be assigned to either a target, in the same manner as before, or classified as a false alarm. Furthermore, we wish to identify the location (scan and target ID) of a missed detection. In Section V we make extensions of the formulation of our basic optimization model to a robust formulation that deals with this detection ambiguity, and we will refer to this formulation as the robust MIO model.

Throughout the paper we make the following assumptions:

Assumption 1.

(i) All targets have constant velocity. i.e., targets do not maneuver and no outside forces act on them.
(ii) Each target's dynamics are independent of one another.
(iii) The number of targets remains constant throughout the window of observation, i.e., there is no birth/death of targets.
(iv) The detection errors are independent of one another.

Notation: We observe P targets over a fixed time window in which T scans are collected. Without loss of generality, and for ease of notation, we assume the scans arrive at a fixed rate of 1Hz, such that the set of scans can be time stamped by $\{1, 2, ..., T\}$. The ith detection of the tth scan is indicated by x_{it}, such that a scan of data at time t is the unordered set of detections $\mathcal{X}_t = \{x_{1t}, x_{2,t}, ..., x_{P,t}\}$. The data for the problem is the ordered set of scans $\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2, ..., \mathcal{X}_T)$. The state space of target trajectories is paramaterized by a true initial position α_j^{true} and a true constant velocity β_j^{true}.

III. MIO MODEL

In this section, we deal with the case of no detection ambiguity. Therefore, we add the following, more restrictive assumptions, to those presented in Assumption 1:

Assumption 2.

(i) The sensor generates exactly one detection for each target at each scan i.e., no missed detections.
(ii) The sensor does not generate any additional detections i.e., no false alarms.

A consequence of Assumption 2 is that the number of detections at each scan will be constant and equal to the number of targets. This seemingly simple point is critical to developing models in the case of no detection ambiguity. We begin constructing our MIO model by introducing decision variables that define data associations as well as estimated trajectories. Using these decision variables, we then develop an objective function that mathematically quantifies the value of the model decisions. Finally, we restrict these variables using constraints that force the model to find solutions that are feasible for the MTT problem as we have defined it. A simple model is first developed step by step in the coming sections before generalized formulation is presented.

A. Decision Variables

The data association and trajectory estimation problems require unique decision variables. Because these two problems lie in different domains, the variables we use to represent these decisions also differ. First, we introduce *continuous* decision variables $\alpha_j \in \mathbb{R}^n$ and $\beta_j \in \mathbb{R}^n$ to represent the estimated initial position and velocity, respectively, of each trajectory j. In our interpretation of the MTT problem we allow the trajectory parameters to lie anywhere in the real-continuous domain. For the data association problem, we wish to assign detections to trajectories, a naturally discrete problem. Therefore, we introduce binary decision variables y_{itj} to indicate whether detection x_{it} is assigned to trajectory j or not:

$$y_{itj} = \begin{cases} 1, & \text{if detection } x_{it} \text{ is assigned to trajectory } j, \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Next, we use these decision variables to develop an objective function to score the solutions found by the model.

B. Objective Function

Here, we would like to develop a function that quantifies the quality of a feasible solution. Our goal is to produce a single measure for both the data association and the trajectory estimation problems. For a given assignment and a given estimated trajectories we define the quality of the estimation as the distance of each detection from the estimated position of its associated trajectory. Let $\hat{x}_{jt}$ denote the estimated location of target j at time t, then, if at scan t detection x_{it} is associated with trajectory j then, the distance

$$\|x_{it} - \hat{x}_{jt}\|,$$

is the measure of the quality of estimation for trajectory j at scan t, and the total estimation quality for a given association will be given as

$$\sum_{j=1}^{P} \sum_{t=1}^{T} \left\| \sum_{(i,j)\in\mathcal{A}_t} x_{it} - \hat{x}_{jt} \right\|, \quad (2)$$

where $\mathcal{A}_t$ is pairs of detection-trajectory assignments made for scan t.

We can now separate the problem into two parts: given an assignment finding the estimated trajectories which minimizes (2), and finding the assignment which results in the best estimated trajectories. Recall that each trajectory is defined by two parameters, α_j^{true} and $\beta_j^{\text{true}}t$ such that the true location is given as

$$\bar{x}_{jt} = \alpha_j^{\text{true}} + \beta_j^{\text{true}}t. \quad (3)$$

Thus, an estimated trajectory can analogically be defined by α_j and β_j such that its estimated location at the time of scan t is given by

$$\hat{x}_{jt} = \alpha_j + \beta_j t. \quad (4)$$

Therefore, given a full assignment $\mathcal{A} = (\mathcal{A}_1, \ldots, \mathcal{A}_T)$, the trajectory which has the best estimation error is given by the solution of the following optimization problem

$$\underset{\alpha_j, \beta_j}{\text{minimize:}} \sum_{j=1}^{P} \sum_{t=1}^{T} \left\| \sum_{(i,j)\in\mathcal{A}_t} x_{it} - (\alpha_j + \beta_j t) \right\|. \quad (5)$$

Notice that under the current assumptions, in which there is no detection ambiguity, (5) is the cost of assignment $\mathcal{A}$.

Now we turn to the problem of choosing the assignment, based on this measure. To this end we formulate the assignment cost (5) in terms of our decision variables. Note that $(i,j) \in \mathcal{A}_t$ if and only if $y_{itj} = 1$, and because all detections will be assigned to a trajectory and vice versa, the following equivalence holds

$$\sum_{(i,j)\in\mathcal{A}_t} x_{it} = \sum_{i=1}^{P} y_{itj} x_{it}. \quad (6)$$

Making the appropriate substitutions, the cost of an assignment described by variables y_{itj} is given as

$$\underset{\alpha_j, \beta_j}{\text{minimize:}} \sum_{j=1}^{P} \sum_{t=1}^{T} \left\| \sum_{i=1}^{P} y_{itj} x_{it} - (\alpha_j + \beta_j t) \right\|. \quad (7)$$

Therefore, in order to find the assignment with the lowest cost, we are left to minimize cost (7) over all assignments, and obtain the following final objective:

$$\underset{y_{itj}, \alpha_j, \beta_j}{\text{minimize:}} \sum_{j=1}^{P} \sum_{t=1}^{T} \left\| \sum_{i=1}^{P} y_{itj} x_{it} - (\alpha_j + \beta_j t) \right\|. \quad (8)$$

At this point it is necessary to discuss the advantages and disadvantages of the two natural distance measures (norms) that will be considered: the ℓ_1 and the ℓ_2 norms. The ℓ_1 norm has the advantage that it can be reformulated using linear optimization (through the addition of continuous variables and constraints), and it is well known to be more robust to outliers. Furthermore, existing algorithms for MIO are more well developed for linear rather than quadratic optimization. However, the ℓ_2 norm squared form, which is equivalent to the residual sum of squares (RSS), has the advantage that it can be quickly computed using simple linear algebra, making it more amenable to a heuristic. This concept will be discussed further in Section IV.

Because of the computational benefits of linear optimization over quadratic optimization, we choose to formulate the objective using the ℓ_1 norm. Therefore, we now show how (8) can be reformulated using linear optimization in the case of the ℓ_1 norm by introducing continuous variables $\psi_{jt} \in \mathbb{R}^n$ and the following constraints:

$$\sum_{i=1}^{P} y_{itj} x_{it} - \alpha_j - \beta_j t \le \psi_{jt}, \qquad \forall j, t, \quad (9)$$

$$-\left(\sum_{i=1}^{P} y_{itj} x_{it} - \alpha_j - \beta_j t \right), \ge \psi_{jt} \qquad \forall j, t. \quad (10)$$

The resulting objective function for the case of the ℓ_1 norm would then be:

$$\underset{\psi_{jt}}{\text{minimize:}} \sum_{j=1}^{P} \sum_{t=1}^{T} \psi_{jt}. \tag{11}$$

C. Constraints

In addition to the constraints used to linearize the objective function, we also require standard assignment constraints to ensure that only one detection is assigned to a target and vice versa. Specifically, for each scan, each detection x_{it} must be assigned to exactly one target j:

$$\sum_{j=1}^{P} y_{itj} = 1 \qquad \forall i, t. \tag{12}$$

Similarly, for each scan, each target must be assigned exactly one detection:

$$\sum_{i=1}^{P} y_{itj} = 1 \qquad \forall j, t. \tag{13}$$

D. Simple Formulation

Integrating all of these elements together, we arrive at the following MIO model:

$$\underset{\psi_{jt}}{\text{minimize:}} \sum_{j=1}^{P} \sum_{t=1}^{T} \psi_{jt} \tag{14}$$

$$\text{subject to: } \sum_{j=1}^{P} y_{itj} = 1 \qquad \forall i, t$$

$$\sum_{i=1}^{P} y_{itj} = 1 \qquad \forall j, t$$

$$\sum_{i=1}^{P} y_{itj} x_{it} - \alpha_j - \beta_j t \leq \psi_{jt} \qquad \forall j, t$$

$$-\left(\sum_{i=1}^{P} y_{itj} x_{it} - \alpha_j - \beta_j t \right) \geq \psi_{jt} \qquad \forall j, t$$

$$y_{itj} \in \{0, 1\} \quad \forall i, t, j$$

$$\alpha_j \in \mathbb{R}^n \quad \forall j, \quad \beta_j \in \mathbb{R}^n \quad \forall j.$$

This formulation is simple in the sense that it involves few variables and constraints, making it highly interpretable and easily implementable. However, it has the disadvantage of being ill suited for extensions to detection ambiguity because it heavily relies on the fact that exactly one of the detections at each scan is associated to a target. This forces the term $\sum_{i=1}^{P} y_{itj} x_{it}$ to always be equal to one of the detections. However, in the case of detection ambiguity, this no longer holds true, since there might be trajectories which are not associated with a detection in a given scan, resulting in an unintended cost to the assignment. Therefore, in the following section we present a more generalized formulation, which is amenable to scenarios with detection ambiguity.

E. Generalized Formulation

Here we modify (14) so that it can be easily extended to handle false alarms and missed detections that occur in scenarios with detection ambiguity. We previously identified that the main issue with (14) arises from the fact that $\sum_{i=1}^{P} y_{itj} x_{it}$ will always incur a cost in the objective function. However, we only wish to account for this assignment cost when a target has actually been assigned a detection. To this end, we introduce a new variable z_{jt} to substitute for this term. We force this variable to take the value x_{it} when $y_{itj} = 1$ and some arbitrary number when $y_{itj} = 0$.

$$z_{jt} = \begin{cases} x_{it}, & \text{if } y_{itj} = 1, \\ \textit{free}, & \text{otherwise.} \end{cases}$$

In the case of no detection ambiguity, $\sum_{i=1}^{P} y_{itj} = 1$ will always hold true as forced by (13), and as a result, we recover the original objective function exactly because z_{jt} will always take on the value of exactly one of the x_{it}.

Additionally, this logical condition can be translated to a model constraint through the following constraint:

$$M_t(1 - y_{itj}) \geq |z_{jt} - x_{it} y_{itj}| \qquad \forall i, t, j. \tag{15}$$

where $M_t = \max_j |x_{it}|$ for each scan. Furthermore, we can linearize (15) by substituting it for the following two linear constraints:

$$x_{it} y_{itj} + M_t(1 - y_{itj}) \geq z_{jt} \qquad \forall i, t, j,$$
$$x_{it} y_{itj} - M_t(1 - y_{itj}) \leq z_{jt} \qquad \forall i, t, j.$$

We must also adjust (8) to account for this change as follows:

$$\underset{z_{jt}, \alpha_j, \beta_j}{\text{minimize:}} \sum_{j=1}^{P} \sum_{t=1}^{T} \|z_{jt} - \alpha_j - \beta_j t\|. \tag{16}$$

This objective can be linearized in the same fashion as before by again introducing continuous variables ψ_{jt} and additional constraints as follows:

$$\underset{\psi_{jt}}{\text{minimize:}} \sum_{j=1}^{P} \sum_{t=1}^{T} \psi_{jt} \tag{17}$$

$$z_{jt} - \alpha_j - \beta_j t \leq \psi_{jt} \qquad \forall i, j, t, \tag{18}$$
$$-(z_{jt} - \alpha_j - \beta_j t) \geq \psi_{jt} \qquad \forall i, j, t. \tag{19}$$

Again, we consolidate these elements together and arrive at the following generalized MIO model:

$$\text{minimize: } \sum_{j=1}^{P}\sum_{t=1}^{T} \psi_{jt} \qquad (20)$$

$$\text{subject to: } \sum_{j=1}^{P} y_{itj} = 1 \qquad \forall i, t$$

$$\sum_{i=1}^{P} y_{itj} = 1 \qquad \forall j, t$$

$$x_{it}y_{itj} + M_t(1 - y_{itj}) \geq z_{jt} \qquad \forall i, t, j$$

$$x_{it}y_{itj} - M_t(1 - y_{itj}) \leq z_{jt} \qquad \forall i, t, j$$

$$z_{jt} - \alpha_j - \beta_j t \leq \psi_{jt} \qquad \forall i, j, t$$

$$-(z_{jt} - \alpha_j - \beta_j t) \geq \psi_{jt} \qquad \forall i, j, t$$

$$y_{itj} \in \{0, 1\} \quad \forall i, t, j$$

$$\alpha_j \in \mathbb{R}^n \quad \forall j, \quad \beta_j \in \mathbb{R}^n \quad \forall j, \quad z_{jt} \in \mathbb{R}^n \quad \forall j, t.$$

Note that (14) and (20) are exactly identical formulations when detection ambiguity does not exist. Throughout the remainder of this paper, we refer to (14) as the *basic* MIO. In Section V we will extend (20) to account for false alarms and missed detections under sensor ambiguity.

IV. HEURISTIC

In this section, we present a detailed description of a heuristic that finds high quality feasible solutions for problem (20). These solutions can be used as a warm start to the MIO, providing a performance boost to the MIO. The heuristic leverages the power of randomized local search methods to find locally optimal solutions.

The fundamental concept behind randomized local search methods is to begin with a randomized starting point and through local improvements converge to a locally optimal solution. By applying this scheme to a growing number of randomized starting points, the probability of reaching high quality solutions, or even the globally optimal solution, increases.

We now detail the heuristic mechanism for a single starting point. The heuristic *starting point* is a randomized solution to the assignment problem, that satisfies the assignment equations (12) and (13). This is done for each scan, by choosing a permutation over the detections uniformly at random, and associating to trajectories by the order of the permutation, i.e., if in the random permutation detection i is at location j than it will be assigned to trajectory j. The assignment cost, temporarily denoted by f, of this starting point is then calculated by solving (7). After initializing the heuristic starts a *sweep* through the scans, i.e., a single pass through the scans. At each stage of the sweep two detections are randomly selected from the current scan and they exchange assignments, an operation we call a *swap*, generating a new feasible solution. The assignment cost of this new solution is calculated, and if it improves the current cost it is kept and otherwise it is rejected. The heuristic will continue to conduct sweeps, until a full sweep is completed without accepting a

single swap, in which point it terminates. The full description of the heuristic algorithm for a single starting point is given in Figure 1, which is also referred to as the *basic* heuristic throughout the course of this paper.

Algorithm 1 Randomized local search with heuristic swaps

Input: $\mathcal{X}$, P, T
Output: f, y
   Initialization : Assign random initial assignment to y^0
1: Calculate $\alpha_j, \beta_j \quad \forall j$
2: Calculate f^0 - the cost of assignment y^0
3: swapped $\leftarrow true$
4: $k \leftarrow 1$
5: **while** swapped **do**
6:    swapped $\leftarrow false$
7:    **for** t in $\{t_1, t_2, ..., T\}$ **do**
8:       Randomly choose $j, m \in \{1, \ldots, P\}$
9:       Find i, l such that $y_{itm}^{k-1} = 1$ and $y_{ltj}^{k-1} = 1$
10:       Swap such that $y_{itj}^k \leftarrow 1$, $y_{itm}^{k-1} \leftarrow 0$, $y_{ltm}^k \leftarrow 1$ and $y_{ltj}^k \leftarrow 0$
11:       Calculate f^k the cost of assignment y^k as well as $\alpha_j, \beta_j, \alpha_m, \beta_m$
12:       **if** $(f^k \geq f^{k-1})$ **then**
13:          $y^k \leftarrow y^{k-1}$
14:       **else**
15:          swapped $\leftarrow true$
16:       **end if**
17:    **end for**
18:    $k \leftarrow k + 1$
19: **end while**
20: **return** f^k, y^k

Fig. 1: Pseudocode for heuristic for a single starting point.

The goal of any heuristic is to find good feasible solutions in an efficient manner. In particular, the goal of our heuristic is to find good feasible solution for the MIO formulation, which can serve as a warm start for the MIO. In Section III we discussed our choice to use the ℓ_1 norm over the ℓ_2 norm for use in the objective of our MIO models. We now turn to discuss why the ℓ_2 is the preferred choice for use in this heuristic. The two main areas of concern are 1) efficiency of the algorithm and 2) quality of the solution.

In the case of the MIO, the preferred objective function was the ℓ_1 norm because it lent itself easily to linear optimization solvers which have known performance advantages over quadratic optimization solvers. However, in the case of the heuristic, the objective function is computed given an assignment, i.e., we need to solve problem (7), to obtain the specific assignment's cost, rather than the more difficult problem (8). Furthermore, this objective needs to be recomputed after each swap, even if it is eventually rejected, and hundreds to thousands of swaps may be carried out for a single starting point. This fact makes the computational cost of this calculation critical to the scalability of the heuristic. Computing the ℓ_1 norm objective would require solving an linear optimization

problem. Even though this linear optimization problem can be computed quite quickly by state of the art optimization solvers, the ℓ_2 norm squared, or RSS, can be computed by simple linear algebra in a fraction of the time, as shown in [22]. Therefore, with respect to efficiency, the ℓ_2 norm is the clear choice for use in the heuristic.

When judging the quality of the heuristic solution, we must look at its purpose. Since it will serve as a warm start for the MIO, which uses the ℓ_1 norm in its objective, we would assume that better solutions will be obtained by using the same norm for the heuristic objective. However, the heuristic will converge to high quality solutions, and since both norms represent measures of distance, and hence are correlated, we assume that a high quality solution as measured by the ℓ_2 norm is also a high quality solution as measured by the ℓ_1 norm. Thus, the choice of the ℓ_2 over the ℓ_1 norm might not significantly degrade the solution quality.

Therefore, although the use of ℓ_2 norm runs the risk of obtaining solutions which are not necessarily extremely good solutions for the ℓ_1 norm objective the potential loss in solution quality is far outweighed by the guaranteed efficiency improvements afforded by the ℓ_2 norm. Finally, the heuristic described above can be parallelized by running partitions of the starting points on separate cores, meaning that a large number of starting points can run in a relatively short amount of time. Increasing the number of starting points greatly reduces the potential for the heuristic to get stuck at a poor quality solution. Therefore, we make the choice to use the ℓ_2 norm in the objective function of the heuristic.

V. EXTENSIONS TO DETECTION AMBIGUITY

We transition to treat the case of detection ambiguity. Specifically, we now allow for false alarms, the instance in which a detection is triggered when no target exists, and missed detections, the instance in which a target exists but no detection is generated. Consequently, the number of detections at each scan varies, and the number of targets we wish to track is now ambiguous. To state this explicitly, we introduce additional notation for the case of detection ambiguity. Let n_t be the number of detections at scan t. We denote

$$N_0 = \min_t n_t$$

as the largest number of detections across all scans, and similarly, we denote

$$N_1 = \max_t n_t$$

as the smallest number of detections across all scans. The only assumption we make in the case of detection ambiguity is that the true number of targets falls somewhere in the range of $[N_0, N_1]$. Particularly, we replace Assumption 2 with the following less restrictive assumption.

Assumption 3.

(i) *The sensor generates at most one detection for each target at each scan i.e., there can be missed detections.*

(ii) *The sensor can generate detections which do not originate from any target i.e., there can be false alarms.*

(iii) *The number of true targets P satisfies $N_0 \leq P \leq N_1$.*

Several new challenges emerge in the case of detection ambiguity. First, we now need to estimate the number of targets. We denote the number of estimated targets as P_{est} Second, now each detection must either be assigned to a target, as before, or classified as a false alarm, not both. Finally, for each target we want to identify the scans in which it was not detected.

In this section, we develop an approach that reduces the difficulty of estimating the number of targets inherent to this new problem by two different approaches. In the first approach, the number of targets can be determined via the optimization model itself, through the use of additional decision variables and constraints. Alternatively, the second approach we formulate an MIO model that assumes a fixed number of targets P and then use the power of parallelization to run that model with all possible values of P which satisfy Assumption 3, choosing the solution for P that minimizes the total objective function value. These two approaches are equivalent, in the sense they have the same optimal solution set.

However, as it turns out, the first approach leads to a model which is not tractable for practical use, while the second results in smaller more tractable models, which as we stated can be solved in parallel. Therefore, we turn our focus to discuss the second approach and defer the interested reader to Appendix A for a complete discussion of the first approach. We begin by extending our basic MIO model to the case of detection ambiguity before discussing necessary adaptations to the basic heuristic, which will also assume a fixed number of targets and provide good quality warm start solutions to its corresponding MIO.

A. Robust MIO with Fixed Number of Targets

In this section, we extend (20) to account for missed detections and false alarms through the addition of decision variables and constraints. The objective function is also updated to reflect the necessary additions.

1) Decision Variables: We need to establish new variables for identifying false alarms as well as missed detections. Toward this goal, we introduce new binary decision variables F_{it} to indicate whether or not a detection x_{it} is a false alarm.

$$F_{it} = \begin{cases} 1, & \text{if detection } i \text{ at time } t \text{ is a false alarm,} \\ 0, & \text{otherwise.} \end{cases}$$

Likewise, we introduce binary decision variables M_{jt} to indicate whether or not trajectory j has a missed detection at time t.

$$M_{jt} = \begin{cases} 1, & \text{if detection for trajectory } j \\ & \text{at time } t \text{ is a missed detection,} \\ 0, & \text{otherwise.} \end{cases}$$

2) Objective Function: We can easily extend (16) to account for detection ambiguity by introducing penalties θ and ϕ for each missed detection and false alarm, respectively. This implies a linear penalty function, meaning that every missed

detection (false alarm) is contributes the same penalty to the objective function. Therefore, we simply need to penalize the total number of false alarms and missed detections in the objective function. If we denote the total number of false alarms TF and total number of missed detections TM, the resulting objective function takes the form:

$$\underset{\psi_{jt}}{\text{minimize:}} \sum_{j=1}^{P}\sum_{t=1}^{T} \psi_{jt} + \theta \cdot TF + \phi \cdot TM.$$

As a general rule, we would expect to increase θ or ϕ, as the number of expected false alarms or missed detections decreases, respectively. A more exhaustive discussion on the insight behind these penalties, in addition to recommendations for tuning them can be found in Appendix A.

3) Constraints: Finally, we must restrict the set of feasible solutions to satisfy the assumptions made earlier. We accomplish this by simply modifying (12) and (13) to account for false alarms and missed detections, respectively. Specifically, all detections must either be assigned to a trajectory j or to a false alarm:

$$\sum_{j=1}^{P} y_{itj} + F_{it} = 1 \qquad \forall i, t. \tag{21}$$

All trajectories j must either be assigned a detection or a missed detection:

$$\sum_{i=1}^{n_t} y_{itj} + M_{jt} = 1 \qquad \forall j, t \tag{22}$$

Here we see why it was necessary to generalize the simple model to (20). When $M_{jt} = 1$, we have $\sum_{i=1}^{n_t} y_{itj} = 0$ by (22). Therefore (15) does not restrict z_{jt} at all and it is obvious, from the structure of the objective function, that in this case $z_{jt} = \alpha_j - \beta_j t$ is the optimal solution, since it results in no change to the objective score (no estimation penalty), which is precisely the desired effect. To state this explicitly, we have:

$$z_{jt} = \begin{cases} x_{it}, & \text{if } y_{itj} = 1, \\ \alpha_j - \beta_j t, & \text{otherwise}, \end{cases}$$

in the case of detection ambiguity.

Lastly, to properly penalize false alarms and missed detections in the objective function, we must force TF (TM, respectively) to equal the sum of all false alarms (missed detections, respectively).

$$\sum_{i=1}^{n_t}\sum_{t=1}^{T} F_{it} = TF$$

$$\sum_{j=1}^{P}\sum_{t=1}^{T} M_{jt} = TM \tag{23}$$

4) Full Formulation: Merging all of these elements together we arrive at our MIO model for a case of detection ambiguity and a fixed number of targets. We refer to this model as the *robust* MIO.

$$g(P) = \underset{\psi_{jt}}{\text{minimize:}} \sum_{j=1}^{P}\sum_{t=1}^{T} \psi_{jt} + \theta \cdot TF + \phi \cdot TM \tag{24}$$

$$\text{subject to: } \sum_{j=1}^{P} y_{itj} + F_{it} = 1 \qquad \forall i, t$$

$$\sum_{i=1}^{n_t} y_{itj} + M_{jt} = 1 \qquad \forall j, t$$

$$\sum_{i=1}^{n_t}\sum_{t=1}^{T} F_{it} = TF$$

$$\sum_{j=1}^{P}\sum_{t=1}^{T} M_{jt} = TM$$

$$x_{it}y_{itj} + M_t(1 - y_{itj}) \geq z_{jt} \qquad \forall i, t, j$$

$$x_{it}y_{itj} - M_t(1 - y_{itj}) \leq z_{jt} \qquad \forall i, t, j$$

$$z_{jt} - \alpha_j - \beta_j t \leq \psi_{jt} \qquad \forall j, t$$

$$-(z_{jt} - \alpha_j - \beta_j t) \leq \psi_{jt} \qquad \forall j, t$$

$$y_{itj} \in \{0, 1\} \quad \forall i, t, j$$

$$\alpha_j \in \mathbb{R}^n \quad \forall j, \quad \beta_j \in \mathbb{R}^n \quad \forall j$$

$$z_{jt} \in \mathbb{R}^n, \quad \forall j, t.$$

P_{est} would then be given by

$$P_{\text{est}} = \min_{N_o \leq P \leq N_1} g(P),$$

and the trajectories and assignments will correspond with the optimal solution of the corresponding model.

B. Heuristic with Fixed Number of Targets

The heuristic for the scenario with ambiguity follows closely from the heuristic developed under the scenario without ambiguity, with a few key differences. In the first place, the process for establishing a random starting point during the initialization process requires refinement. Initial solutions should allow for false alarms and missed detections. With equal probability, each detection is randomly classified as a false alarm or selected to receive a target assignment. Once a detection has been identified for assignment, assignments are randomly selected uniformly across all targets. The remaining scans for which trajectories do not receive an assignment are identified as missed detections.

Once a starting point has been initialized, the heuristic progresses in much the same manner as before. Again, it sweeps through all scans continuously making random swaps. However, the swapping process also change.Specifically, false alarms and missed detections must be included in the switches. but the robust heuristic randomly chooses from the following options when making swaps:

1) Switch detection assignments between two existing targets.
2) Switch the detection assignment of an existing target with a false alarm.

3) Switch the detection assignment of an existing target with a missed detection for a different existing target.
4) Move the detection assignment of an existing target to a false alarm and replace it with a missed detection.
5) Move a false alarm into the location of a missed detection for an existing target.

Similar to the basic heuristic, this robust extension will accept the switch/move if the objective score improves, and reject the switch/move otherwise, and terminates once it completes a full sweep without accepting a single swap.

The framework of this new heuristic, also referred to as the *robust* heuristic, is the same as the one presented in Figure 1 with the appropriate modifications in the initialization and swapping steps.

Due to the increase in potential combinations of solutions, we expect this variant of the heuristic to run slightly slower. Though this effect is mitigated by the fact that this variant is just as parallelizable as the basic heuristic. Next, we shift our discussion to focus on measuring the complexity and performance of our approaches.

VI. SCENARIO COMPLEXITY & PERFORMANCE METRICS

In order to measure the quality of solution obtained by any MTT algorithm one needs to have a measure of both performance and scenario complexity. Unfortunately, as stated in [1], there does not exist a unified approach for measuring scenario complexity, nor does there exist clear measures of performance for each of the trajectory estimation and data association problems. In this paper, we argue that the data association problem has a natural performance metric but lacks a measure of complexity, while the trajectory estimation problem has a natural measure of complexity but lacks a clear performance metric, and thus we construct the missing measures for each.

It is natural to assume that the difficulty of a scenario is highly correlated with a sensor property which quantifies the deviation of the detections from the true targets. Thus in order to discuss these measures we must first define σ to be a measure of this sensor property, quantifying how noisy the sensor detections are, and in most cases is the standard deviation of the detection error. We will show how σ can be used to define complexity measures for different scenarios.

A. Data Association

In the case of the data association problem, the preferred performance metric often used in practice is % accuracy, *i.e.*, the number of correct detection assignments out of the number of possible correct assignments. For the case without sensor ambiguity, the number of possible assignments is simply the total number of detections, or equivalently, the number of targets multiplied by the number of scans:

$$Accuracy = \frac{\text{\# correct assignments}}{\text{Total \# of detections}} = \frac{\text{\# correct assignments}}{PT}.$$

In the case of sensor ambiguity, however, the number of possible correct assignments requires a deeper explanation. To develop a better understanding, we consider our goal, which is to correctly assign detections to targets and identify both false alarms and missed detections. With this in mind, we define the number of possible correct assignments as the number of targets multiplied by the number of scans plus the number of false alarms:

$$Accuracy = \frac{\text{\# correct assignments}}{PT + \text{\# False Alarms}}.$$

Whereas accuracy serves as a good measure of performance for data association, there does not exist a corresponding measure of complexity which comparatively measures the difficulty of the data association problem. We argue that σ alone is not the best measure of difficulty for the data association problem. For example, for a scenario with very close target trajectories it may be difficult to ascertain data associations even for small σ values, and similarly with high enough σ values even widely spaced targets could be difficult to differentiate. Therefore, we introduce a metric ρ to quantify this complexity. For ease of notation in developing this metric we first define D_{ijt} as the distance between one true trajectory i and another true trajectory j:

$$D_{ijt} = \|\alpha_i^{\text{true}} + \beta_i^{\text{true}}t - \alpha_j^{\text{true}} + \beta_j^{\text{true}}t\|.$$

Additionally, we define a variable c_{ijt} that will take the value of 1 if the distance between trajectory i and trajectory j is greater than some monotonically increasing function of σ which we will denote by $h(\sigma)$:

$$c_{ijt} = \begin{cases} 1, & \text{if } D_{ijt} > h(\sigma), \\ 0, & \text{otherwise.} \end{cases}$$

Then the difficulty of a scenario in the sphere of the data association problem is quantified by the complexity measure ρ, which is the proportion of detection pairs that fall within a closely defined proximity to each other:

$$\rho = \frac{\sum\limits_{t=1}^{T} \sum\limits_{i<j} c_{ijt}}{\binom{P}{2}T}.$$

This metric has several desirable attributes. First and foremost, it falls within the range of $[0, 1]$, identical to the range of accuracy, making it easily comparable. Secondly, it is easy to understand and interpret. Higher values of ρ indicate easier scenarios because fewer targets are within close proximity for a shorter amount of time, and vice versa. Finally, as we have defined it, ρ has an inverse relationship with σ, which means that it serves as a connection between scenario generation and performance measuring processes. While σ can be used more naturally for scenario generation, where it is useful as a parameter for signal noise, ρ can be calculated after the fact and used to quantify the difficulty of the scenario as it pertains to the data association problem.

B. Trajectory Estimation

In the case of the trajectory estimation problem, the preferred complexity metric often used in practice is σ itself. Increasing the noise may often lead to stronger bias in the trajectory estimation, especially in scenarios with fewer

scans, and results in a deteriorated quality of the estimation. Therefore, we believe that σ is the correct metric for use in measuring the difficulty of the trajectory estimation problem.

However, establishing a performance metric for the trajectory estimation problem is necessary. We choose to implement a metric which captures the core goal of the trajectory estimation problem, that is to estimate a trajectory as close as possible to the true ground track.

$$\delta = \frac{\sum\limits_{t=1}^{T} \sum\limits_{j=1}^{P} \|\bar{x}_{jt} - \hat{x}_{jt}\|}{PT}.$$

We match the true trajectories to the estimated trajectories using an one-to-one assignment problem which can be formulated using linear optimization. See Appendix C for more details. Lower values of δ correspond to higher performance because the distance between the estimated and true ground trajectories is smaller.

In the next section, we will see how these measures of complexity and performance are useful in quantifying the strengths and weaknesses of our methods.

VII. Experimental Simulations & Computational Results

We evaluate our approaches on a wide variety of simulated scenarios, comparing the results against two benchmark solutions. For the first benchmark, we randomly generate detection assignments, including randomly assigning false alarms and missed detections for the case of detection ambiguity. This solution will be referred to as the *random* solution. In the second benchmark, the detection assignments are known perfectly, meaning that all assignments are exactly correct including the classification of false alarms and identification of missed detections. This solution is referred to as the *ideal* solution, however, it is only ideal in relation to the data association problem, but it provides a means for bounding the expected error in the trajectory estimation problem. Note that we do not compare our methods to any known MTT algorithms, such as the MHT or JPDAF, due to the complexity in tuning and the overhead of implementation of these algorithms.

There does not exist among the literature a clearly defined comprehensive set of standard test scenarios as pointed out by [1], which also notes that two types of scenarios of particular importance include crossing trajectories and parallel trajectories. Because we would like to test our methods across scenarios with a wide range of complexity, for both the data association and trajectory estimation problems, it is necessary to create scenarios using both methods. With this in mind, we choose to generate scenarios of both trajectory types using a simple methodology that will be outlined next in our discussion on experimental methods.

We run two separate experiments, one with detection ambiguity and one without. Both experiments, including the scenario generation process, heuristic, and MIO, were implemented in the development software *julia* 0.4.3 [23] using the optimization package *JuMP* [24]. The optimization software Gurobi 6.5.0 [25] was used to solve the MIOs, and the optimization processes was restricted to the use of a single core. Each simulation was run on a single compute node of the unclassified TX-Green cluster located at Lincoln Laboratories. The cluster utilizes DL165 G7 compute nodes, consisting of 2.2 GHz compute cores, with 8 GB of RAM each, for a total peak performance of 77.1 TFLOPS [26].

We begin by outlining our experimental methods for scenarios without detection ambiguity and discuss the results of our approaches on these scenarios.

A. Scenarios without Detection Ambiguity

In order to evaluate scalability of our algorithms we test our methods across a range of scenarios with varying numbers of targets and scans. In particular we consider: $P \in \{4, 6, 8, 10\}$ and $T \in \{4, 6, 8, 10\}$ seconds. Scans are collected at a rate of 1 Hz. The cartesian product of P and T creates 16 unique scenario sizes. We generate 10 unique crossing scenarios and 10 unique parallel scenarios of each size.

To generate trajectories, a grid size of width τ is selected and discretized into ω points. Trajectories are defined by two of these points, the first referred to as the *initial position* and second the *final position*. For crossing trajectories, the initial and final positions have no restrictions, while for parallel trajectories they are restricted to disallow overlaps. This is accomplished by repeating the process for crossing trajectories, but with neighboring grids of smaller grid widths. - I do not understand this sentence This ensures trajectories do not overlap, but they will remain in very close proximity due to the decrease in grid width. For our experiments, we elected for $\tau = 20$ and $\omega = 25$ for crossing scenarios and $\tau = 2$Is it 2 or 20? and $\omega = 5$ for parallel scenarios only 5? so what happens when there are 10 targets?.

For each scenario, we randomly generate 10 realizations of data by first perturbing each true position measurement by an error $\epsilon \sim \mathcal{N}(0, \sigma)$ with $\sigma \in \{0.1, 0.5, 1.0, 2.0, 3.5, 5.0\}$, where σ represents the noise parameter. Adding the detection error to the true position results in a detection:

$$x_{it} = \alpha_i^{\text{true}} + \beta_i^{\text{true}} t + \epsilon.$$

Scans $\mathcal{X}_t$ are simulated by randomizing the order of x_{it} for each t. Each unique $\mathcal{X}$ generated is referred to as a *simulation*. For each such simulation, we run the heuristic with a range of number of starting points $N \in \{100 \;\; 1,000 \;\; 10,000\}$, and use each of these solutions as a warm start for the MIO. The optimization process is set to terminate after 3T seconds, with solutions collected at intervals of $\{1, T, 2T, 3T\}$ seconds.

At the conclusion of the experiment, we calculated the difficulty of each scenario, the accuracy of each solution, and the trajectory estimation error δ. When measuring the difficulty of scenarios in terms of ρ, we propose the use of $h(\sigma) = 2\sigma$, since it is difficult to distinguish detections which lie between target trajectories that are closer.

1) Scenario Generation: We begin with a discussion on the relationship between ρ and σ and show how this relationship benefits both scenario generation and complexity measuring by allowing each to occur in their own natural domain. Figure 2 shows the relationship between σ and ρ for the 20 scenarios generated simulated in our experiments. The plot is

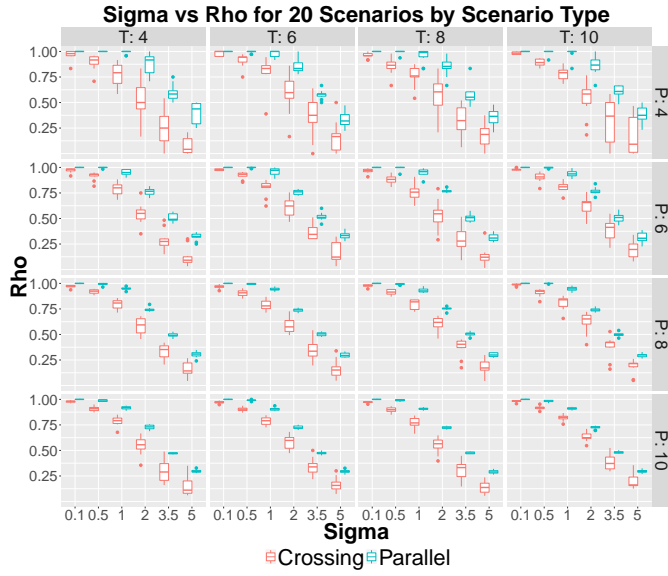broken down by scenario type between crossing and parallel trajectories.

Fig. 2: Relationship between σ and ρ summarized by scenario type for all 20 generated scenarios in this experiment.

Remember that higher values of ρ indicate a lower proportion of detections within very close proximity to one another. We note that the parallel method of scenario generation clearly generates easier scenarios, as measured by ρ. This suggests that in general, it may be more difficult to ascertain correct associations for crossing scenarios than for parallel scenarios. In addition, we can conclude from Figure 2 that σ and ρ are highly correlated, and the higher σ is the lower ρ becomes, which in to be expected from the definition of ρ. Furthermore, we note that while σ is restricted to a small range of six values, a wider range of possible ρ values is captures, and in some instances the entire $[0,1]$ range of ρ is captured. ??? why is this result beneficial? the cause and effect are not clear??? This result is beneficial to both scenario generation and complexity measuring because it suggests that σ can be used in its natural domain of the data generation process, and ρ can be back calculated as a measurement of difficulty for the data association problem. As a result, we gain a highly interpretable performance metric for the data association problem without sacrificing the ability to generate scenarios in their natural domain. Still missing - refer to a specific graph, give numerical examples of what you claim to be true

2) Basic Heuristic: We transition to discuss performance of the heuristic, beginning with an examination of the experimental run times. Table I summarizes the minimum, mean, and maximum run times of the heuristic for a single starting point, arranged by the number of targets (P) and number of scans (T). Times are shown in milliseconds.

At a glance, it is evident that the heuristic scales more efficiently with increases in the number of targets than increases in the number of scans. Notice that the average run time roughly doubles with each addition of two scans, for a fixed number of targets. again missing numerical examples - talk in numbers

		Heuristic Run Times (in milliseconds)		
P	T	Min	Mean	Max
4	4	0.07	0.10	0.18
4	6	0.18	0.24	0.38
4	8	0.34	0.45	0.62
4	10	0.58	0.76	1.02
6	4	0.11	0.15	0.25
6	6	0.31	0.39	0.58
6	8	0.64	0.81	1.05
6	10	1.24	1.56	2.02
8	4	0.14	0.19	0.30
8	6	0.46	0.57	0.86
8	8	0.95	1.24	1.58
8	10	2.07	2.53	3.37
10	4	0.19	0.25	0.41
10	6	0.63	0.80	1.03
10	8	1.44	1.84	2.44
10	10	2.96	3.73	4.56

TABLE I: Heuristic run times (in milliseconds) for a single starting point.

an the other hand, the average run time never increases more than 50% for an increase of two targets, given a fixed number of scans. Therefore, we conclude that the heuristic would be capable evaluating scenarios with a surprisingly large number of targets, but measures should be taken to avoid significant increases in the number of scans.

Implementing the heuristic as sliding window algorithm would very likely mitigate this scaling effect in regards to T. Rather than solve all scans in a single batch at once, a sliding window algorithm solves a subset of scans, or a smaller window, and advances through all scans sequentially. As the window progresses forward through the scans, "soft" decisions are made meaning that the heuristic would begin with the decisions from the previous solution. As scans pass beyond the horizon and out of the sliding window, the decisions become fixed and we refer to them as "hard" decisions. This process continues until all scans have been processed. The run times of a sliding window variant of the heuristic would not exhibit the curse of dimensionality in T since the number of scans remains constant. Additionally, the heuristic is likely to produce higher quality solutions as a result of these "soft" decisions of previous steps since it is starting from a solution which is likely to be better than a completely random solution.

Furthermore, the full strength of the heuristic isn't realized until we remember that the heuristic is extremely parallelizable. By running the heuristic on several processors, we can reduce the number of starting points run on each processor, reducing the total processor running time. As a result, given enough processors, the heuristic can run several thousand starting points and still find solutions in a fraction of a second. To illustrate, say that we intend to run 50,000 heuristic starting points for a scenario with six targets and six scans. The average run time for a single starting point of this size is about 0.4 milliseconds. Running all of these starting points in sequence would require approximately 20 seconds of run time, however, the same amount of starting points parallelized onto 100 processors would only require a run time on the order of 0.2 seconds. Thus, the run time of the heuristic can be reduced to meet the efficiency needs of the system subject only to the

limitation of available processors.

We have seen that the heuristic scales efficiently in regards to the number of targets, and we have also seen the power of the heuristic when parallelized, but *does the heuristic provide good feasible solutions to serve as a warm start to the MIO?* In order to answer this question appropriately, we compare the heuristic solution to the ideal solution, which as a reminder is the solution in which the data association problem is exactly correct. Toward this goal we compute the MIO objective value of both the heuristic solution and the ideal solution, and report the ratio between the first value and the second. Figure 3 plots this ratio against σ. The y axis title is misleading, I would call the MIO objective by a letter, say f and compare f_H vs. f_I where H signifies Heurtic and I Ideal solutions.
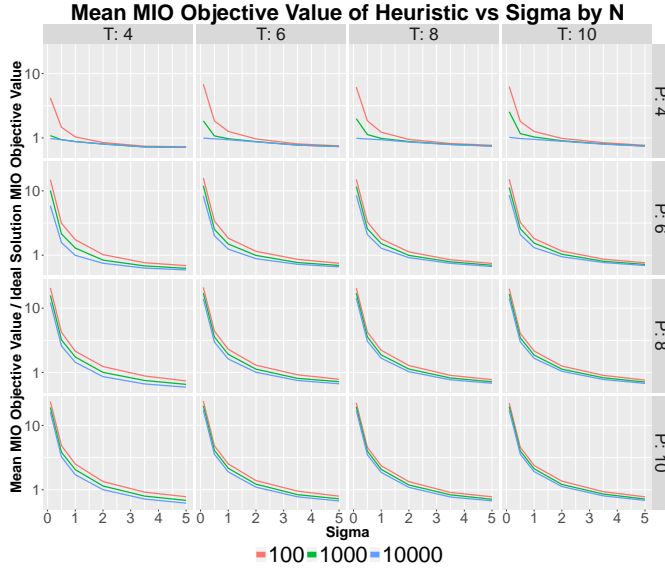

Fig. 3: Quality of heuristic solution as compared to the ideal solution's MIO objective value summarized by number of starting points.

On the whole, we see that increasing the number of starting points improves the quality of heuristic solutions as compared to the ideal solution's MIO objective value, though the magnitude of improvement diminishes as the number of targets increases. Yet for scenarios with only four targets, we see that increasing the number of starting points has a significant effect colored quantify. This suggests that in order to maintain solution quality, the number of necessary starting points increases with P, likely due to the exponential increase in combinatorial solutions as the number of targets increases. Additionally, there does not appear to be a significant loss in quality as the number of scans increases, suggesting that solution quality scales efficiently with increases in T. Altogether, this suggests that the heuristic solution quality scales better in regards to increases in T than P. We remind the reader that this easily mitigated due to the fact that a very large number of starting points can be evaluated efficiently through the power of parallelization.

It is important to note that the heuristic outperforms the ideal solution's MIO objective value for larger values of σ again give an example, state the outperforms = ratio smaller than 1. Remember that the ideal solution is simply ideal in the sphere of data association, while the MIO objective function serves to quantify solutions in regards to both data association and trajectory estimation simultaneously. This result suggests that the heuristic gives a better fit to the noisy measurements but this fit does not necessarily lead to better estimated trajectories but we later see that in fact it does , it is simply the true association does not lead to the estimated trajectories which are closer to the true ones because of the noise. With this in mind, we conclude that it may be necessary to tradeoff correct data associations in order to improve the trajectory estimation in cases of higher noise.

We continue our evaluation of the basic heuristic by measuring its performance on the data association problem. Figure 4 plots the mean accuracy of each of the three quantities of starting points against ρ.
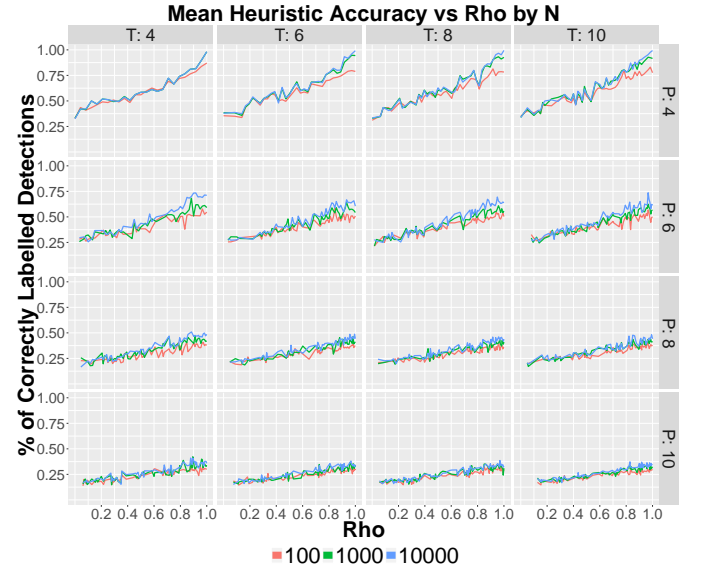

Fig. 4: Accuracy of basic heuristic by number of heuristic starting points.

First of all, we see that the heuristic finds good solutions to the data association problem, especially not especially define exactly in what situations +quantify for scenarios with fewer targets and higher values of ρ. While Figure 3 may have instigated cause for concern of the scalability of the heuristic in regards to P, Figure 4 reassures that the heuristic is adding value and finding nontrivial solutions, at least as it pertains to the data association problem this is not exactly true for P=10 for example, explain what you means by scale, the sentence is too general. Perhaps this is further evidence that the ideal solution of perfect data associations does not always correspond to the best solution when balancing both data association and trajectory estimation.

Overall, we conclude that the heuristic run times scale especially well in regards to increases in the number of targets, though the solution quality may not scale quite as well. In any event, we have seen that there is not a significant difference in heuristic performance for the range of N values that we

explored. Therefore, for simplification as we move forward in our analysis, we will restrict our discussions of the heuristic to $N = 1,000$. Next, we will see how the MIO performs.

3) Basic MIO: We shift our focus to the MIO by first measuring its performance on the data association problem. Figure 5 plots the mean accuracy of the MIO, initialized by the $N = 1,000$ heuristic solutions, after 1,T, and 2T seconds against ρ. Note that we excluded the data for the MIO after 3T seconds for the sake of clarity as it showed little to no improvement over the MIO solution after 2T seconds. Also note that the ideal solution, which always achieves an accuracy of 1.0, has also been excluded. For comparison, we have included the accuracy of random solutions.
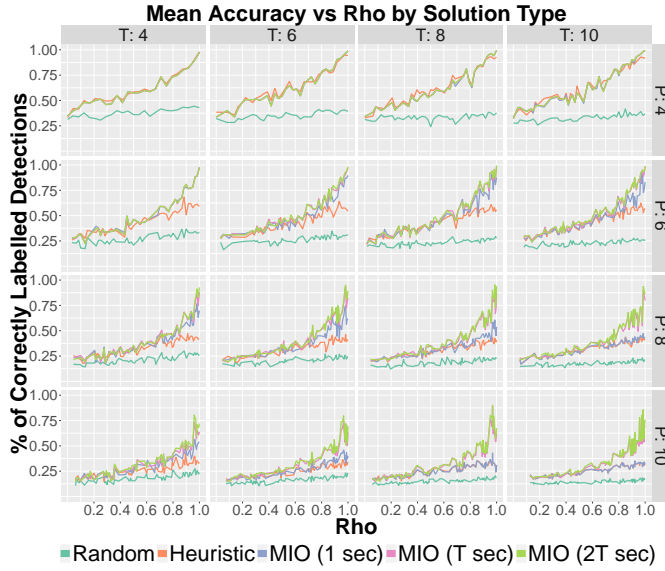

Fig. 5: Accuracy of MIO compared against the heuristic and a randomized solution.

The MIO finds high quality solutions to the data association problem, and it appears to scale fairly well in regards to increases in both P and T. For all scenarios with eight or fewer targets, accuracy tops out at or near "near" not a good word, quantify, in which cases? in 1.0, and reduces to about 0.85 for ten target scenarios. Furthermore, scenarios with ten scans appear to slightly outperform scenarios with only four scans, suggesting that additional scans benefit the MIO solution quality, likely due to the added information gathered from an increase in the number of detections.

While not seen in Figure 5, we note that for scenarios with four targets, many of the MIO solutions were actually proven to be the optimal solution. The fact that the heuristic performance is almost identical to the MIO again suggests that the heuristic finds high quality solutions. This is further supported by the fact the heuristic performs almost as well as the MIO after 1 second even in scenarios with ten targets. Equally as important, Figure 5 shows that in nearly all scenarios, the MIO achieves its best or near best solutions after T or fewer seconds, suggesting the usefulness of the MIO as an online algorithm with a sliding window. As mentioned previously in regards to the heuristic, why are you repeating the same thing?

maybe condense or just talk about the MIO implementation in this case a sliding window algorithm would make decisions on a subset of scans, and these decisions will be fixed before accepting a new set of scans. In regards to the MIO, this would be implemented by adding constraints to restrict the values of y_{itj} to match that of the subsetted solution. The fact that the MIO finds very good solutions in T or fewer seconds means that a sliding window algorithm would be able to solve each subset in real time before advancing to the next subset of scans. Furthermore, the MIO would likely benefit from the fixed decisions of the preceding windows, since the fact we have not utilized it does not explain why it will benefit the MIO and what is "our approaches" this is added knowledge that has not utilized by our approaches.

Next, we evaluate the performance of the basic heuristic and MIO through the lens of trajectory estimation. As discussed previously, we are interested in comparing δ, our proxy for ground track error, against σ, our measure of difficulty for trajectory estimation, in order to analyze performance of in the sphere of estimation. Figure 6 plots σ against δ for each of the solution types. In addition to the random solution shown on the previous plot, we also add a comparison to the ideal solution, which was previously defined.
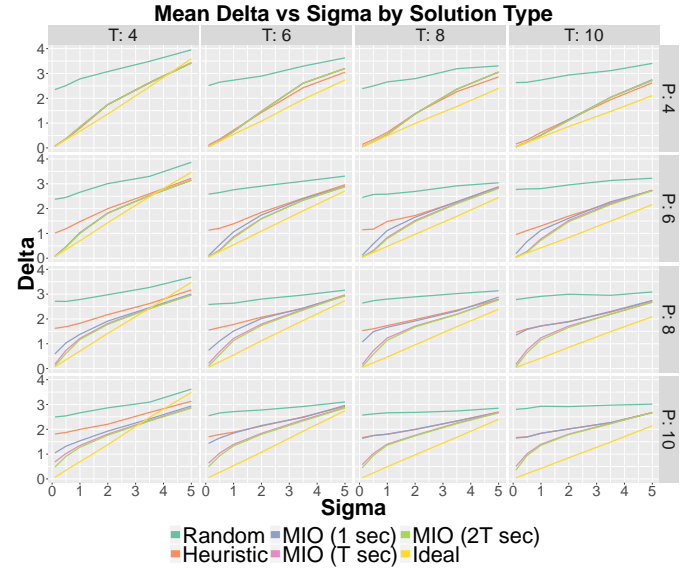

Fig. 6: Trajectory estimation performance

Remember that lower values of delta correspond to trajectory estimations that are closer to that of the true ground track. We see that the performance of the heuristic converges to that of the MIO for scenarios with few targets, as well as for large values of σ quantify. Additionally, we see that as the number of targets increases we begin to see stronger improvements by the MIO over the heuristic. Interestingly, we see that for the scenarios with the largest number of targets and scans, the MIO after one second is not much better than the heuristic, while the MIO after T seconds provides significant improvement over that of the heuristic and MIO after 1 second, there is little further improvement in running the MIO for 2T seconds. explain why this happens, why only for large number

of targets

Again, we see that ~~occasionally~~ the heuristic and/or MIO outperform(s) the ideal solution ~~especially for~~ in scenarios with only four scans and high values of σ. This effect increases as the value of σ increases. This result suggests that as the number of scans approaches infinity, the estimated trajectories given by the ideal association approaches the true ground tracks. Put differently, as more and more data is available, it becomes easier to estimate the trajectories even in the event of ~~large~~ low signal to noise ratio, and so the trajectory estimates that result from the ideal solution converges to the true ground track. but what does this have to do with the MIO? and the phenomenon you just described?

In summary, we have shown that in the case of no detection ambiguity, the heuristic is scalable, particularly in regards to increases in the number of targets, and it finds high quality solutions in milliseconds. The MIO is also scalable, appearing relatively robust to both increases in the amount of targets and scans. Both approaches show promising signs for use in a sliding window algorithm.

B. Scenarios with Detection Ambiguity

Here we extend the discussion from the previous section to analyze the performance of our methods on scenarios with detection ambiguity. We first summarize our experimental methods before discussing performance of both the robust heuristic and the robust MIO in the spheres of both the data association and trajectory estimation problems.

This experiment serves as an extension of the basic one, in order to test the performance of our algorithms under detection ambiguity. We use the same scenarios generated from the basic experiment, but due to the additional difficulty inherent with detection ambiguity, we limit the range of signal noise to $\sigma \in \{0.1, 0.5, 1.0, 2.0\}$, excluding choosing to exclude the extreme cases of signal noise. In addition, we simulate both missed detections and false alarms. A detection is removed with probability, γ, and we consider $\gamma \in \{0.2, 0.15, 0.1, 0.05\}$. We do not allow empty scans. For each scan, we generate false alarms according to a poisson distribution with parameter, λ, and detection locations are then randomly selected uniformly within the state space. We consider $\lambda \in \{0.1, 0.5, 0.1, 2.0\}$. The false alarms are then added to $\mathcal{X}_t$ and the detection order of $\mathcal{X}_t$ is again randomly shuffled as before.

Once the data has been generated, we follow the same sequence as before in the basic experiment, running the heuristic first and then feeding the solution into the MIO as a warm start. Note that the heuristic is given 1,000 starting points only, as concluded from the results of the basic experiment. Once again, the optimization process was set to terminate after 3T seconds, with solutions collected at intervals of $\{1, T, 2T, 3T\}$ seconds. Prior to the running of this experiment, we performed a mini experiment and used the results to tune the penalties θ and ϕ. A summary of the exact penalties used along with an explanation of the insight behind them, can be found in Appendix A.

We will now evaluate the performance of the robust heuristic and MIO. We begin with a discussion on the run times of the robust heuristic, following the number of targets estimation, accuracy and trajectory estimation.

1) Robust Heuristic Run Times: Table II summarizes the minimum, mean, and maximum run times of the heuristic from Experiment 2 for a single starting point, arranged by the number of estimate targets ($P_{estimated}$) and number of scans (T). Times are shown in milliseconds.

$P_{\text{estimated}}$	T	Heuristic Run Times (in milliseconds)		
		Min	Mean	Max
2	4	0.15	0.23	0.41
2	6	0.42	0.56	0.93
2	8	0.77	1.04	2.24
2	10	1.27	1.73	20.23
4	4	0.15	0.34	1.04
4	6	0.50	0.94	2.69
4	8	1.09	1.88	3.87
4	10	2.12	3.25	13.51
6	4	0.14	0.42	0.96
6	6	0.57	1.29	4.45
6	8	1.33	2.66	75.28
6	10	2.53	4.61	18.69
8	4	0.16	0.50	1.10
8	6	0.60	1.59	3.46
8	8	1.38	3.37	6.87
8	10	2.63	5.84	12.40
10	4	0.18	0.55	1.10
10	6	0.72	1.82	3.98
10	8	1.53	3.96	8.18
10	10	3.42	6.93	13.93
12	4	0.16	0.56	0.99
12	6	0.99	1.95	3.96
12	8	1.74	4.33	8.69
12	10	3.40	7.71	15.10

TABLE II: Robust heuristic run times (in milliseconds) for a single starting point.

Comparing Table I and Table II we see that the run times for an estimated number of targets in the robust heuristic range from roughly double to three and four times that of a comparable number of targets in the basic heuristic. Due to the increase in combinatorial solutions in the robust heuristic over the basic heuristic, this was an expected result. The robust heuristic, however, can be parallelized in the same many as the basic heuristic, meaning that the robust heuristic can actually recover the speed of the basic heuristic with the introduction of additional processors. More importantly, we see that the robust heuristic scales very efficiently with $P_{\text{estimated}}$, and this the scaling actually improves as the number of estimated targets and scans increases. Increasing from two to six estimated targets for four scans roughly triples the run time, while increasing from eight to twelve estimated targets for four scans increases the run time by only 12%. This is a great result because it means that a relatively wide range of estimated targets can be parallelized without fear of one subset requiring a substantially longer run time than another.

Although the robust heuristic scales well with the number of estimated targets, it is more sensitive to increases in the number of scans. However, this effect is no worse than what we saw for the basic heuristic. It appears that on average increasing the number of scans by 2 results in a doubling of the run time for a fixed number of estimated targets. We conclude that these results again support the use of the robust heuristic

in an online algorithm with a sliding window, as discussed in the previous section.

2) Evaluating the Number of Targets: Next, we continue our analysis of the robust approaches by quantifying the algorithms' ability to estimate the correct number of targets. This is perhaps the most important goal of a MTT algorithm and so we begin our analysis here. To this end, we define

$$P_{\text{diff}} = P_{\text{true}} - P_{\text{est}} \tag{25}$$

where $P_{\text{estimated}}$ is the number of estimated targets and P_{true} is the number of true targets, and we plot the distribution of $P_{\text{estimated}}$. Figure 7 shows the distribution of $P_{\text{difference}}$ for scenarios with four targets and eight scans, and for comparison, Figure 8 plots the same result for scenarios of eight targets and eight time scans.
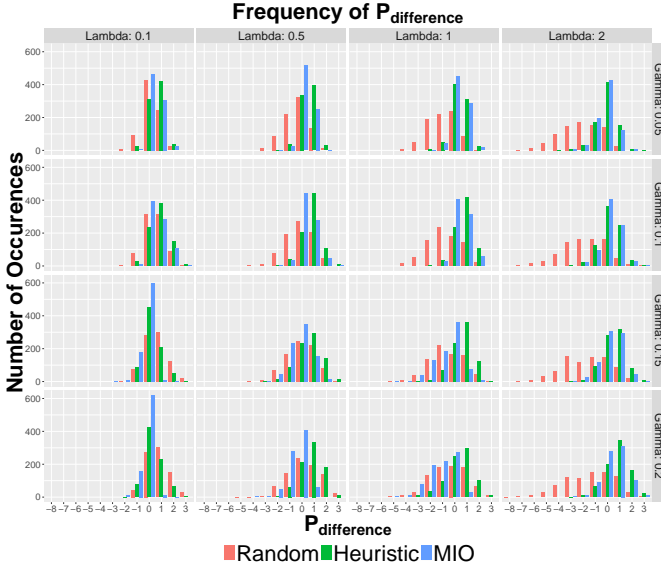

Fig. 7: Distribution of the difference in true and estimated number of targets for scenarios with 4 targets and 8 scans, arranged by γ and λ.

Note that the algorithms have correctly estimated the number of targets when $P_{\text{difference}} = 0$. When $P_{\text{difference}} < 0$, we have overestimated the number of targets, and when $P_{\text{difference}} > 0$, we have underestimated the number of targets. We see that both the robust heuristic and the robust MIO estimate the number of targets correctly a high proportion of the time in the scenario with four targets, particularly for smaller values of λ. As λ increases, though, both algorithms tend to underestimate. The same trend persists in the larger scenario. This suggests that either 1) the false alarm penalty needs further tuning and likely was not set high enough in the experiment or 2) the missed detection penalty set too high. In the case where θ is set too low, the algorithms would prefer to classify detections as false alarms rather than create additional trajectories for the detections. In the case of ϕ set too high, the algorithms would opt out of creating additional trajectories in order to decrease the need to fill smaller scans with missed detections. Furthermore, because the effect of underestimation is more prominent in the scenario with more targets, we
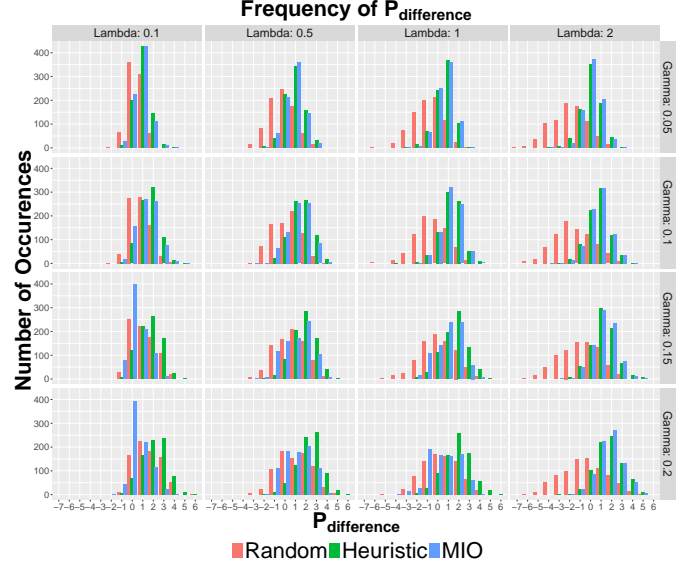

Fig. 8: Distribution of the difference in true and estimated number of targets for scenarios with 8 targets and 8 scans, arranged by γ and λ.

conclude that both penalties should probably take into account the number of targets that it is currently estimating.

3) Data Association: Knowing that we tend to underestimate the number of targets with the given penalties, we move on in our analysis to measuring the accuracy of our robust approaches. Figures 9 and 10 plot the accuracy performance metric against the difficulty metric, ρ, for scenarios of four and eight targets, respectively. Both scenarios have eight scans and both Figures have been arranged by γ and λ.
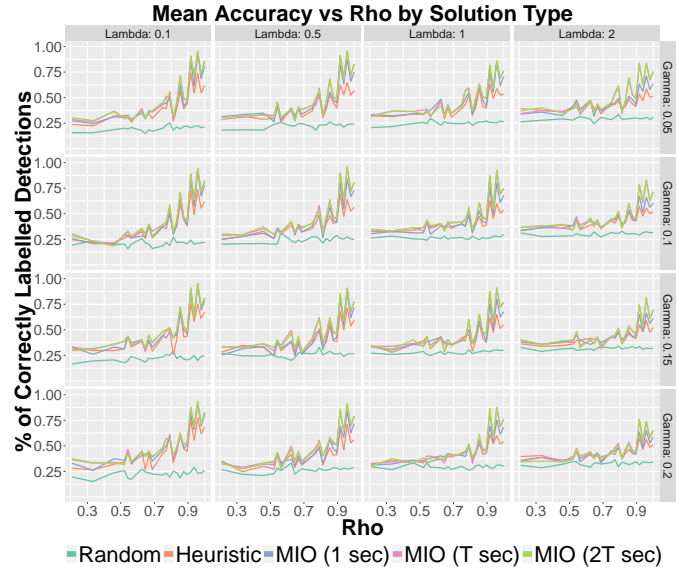

Fig. 9: Accuracy of robust heuristic and MIO as compared to random solutions for scenarios of 4 targets and 8 scans, arranged by γ and λ.

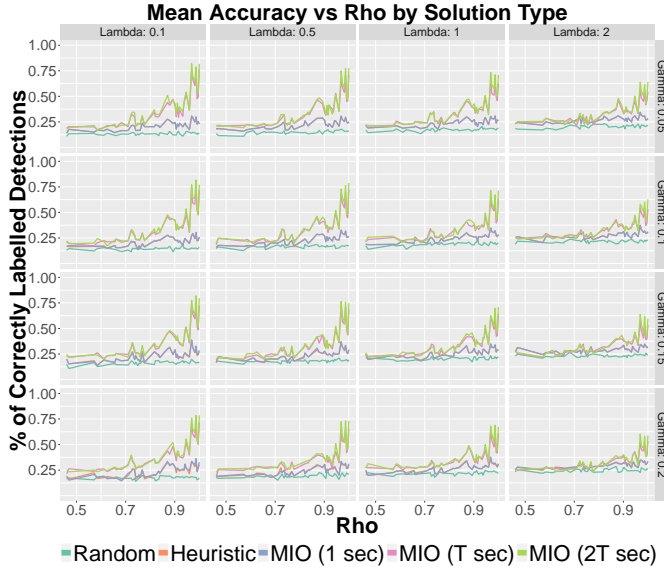Similar to the performance of the basic heuristic, we again see that the robust heuristic improves greatly over that of a

Fig. 10: Accuracy of robust heuristic and MIO as compared to random solutions for scenarios of 8 targets and 8 scans, arranged by γ and λ.

random solution, and the MIO offers even further improvement. Again, we seeing that running the MIO for 1 second offers significant improvements over the heuristic, and running the MIO for T seconds offers further improvement. However, running the MIO for 2T seconds offers little to no further improvement. Again, these results support the use of the MIO in an online algorithm with a sliding window as mentioned previously.

Comparing Figure 9 with the 4 target and 8 scan element of Figure 5, we see only a slight decrease in performance when $\gamma = 0.05$ and $\lambda = 0.1$. This is an important result because we no longer know the number of targets in the robust case, yet we achieve almost the same levels of accuracy. Furthermore, both Figure 9 and Figure 10 show that the robust algorithms are more robust to decreases in the detection probability γ than to increases in the false alarm rate λ. We conclude that the robust approaches are more sensitive to changes in the false alarm rate, in particular when it comes to making data associations.

We have shown that both the heuristic and the MIO tend to underestimate the number of targets, due to the chosen penalties. We have also shown that accuracy degrades as the false alarm rate increases. This is likely not a coincidence. It is probable that as a result of underestimation, in which fewer trajectories are generated, there is a higher rate of misclassification of detections as false alarms, which in turn directly leads to a reduced accuracy. Therefore, it is a promising result to see accuracies above 75% in Figure 10, even when Figure 8 suggests overestimation. It is entirely possible that further parameter tuning or introducing more complex penalties would lead to even great performance in the data association problem.

4) Trajectory Estimation: We conclude our analysis of the robust approaches with a discussion on their performance in the sphere of the trajectory estimation problem. Figures 11

and 12 plot the δ performance metric against the difficulty metric, σ, for scenarios of four and eight targets, respectively. Again, both scenarios have eight scans and both Figures have been arranged by γ and λ. Note that the range on σ has been reduced from $[0.1, 5.0]$ to $[0.1, 2.0]$.
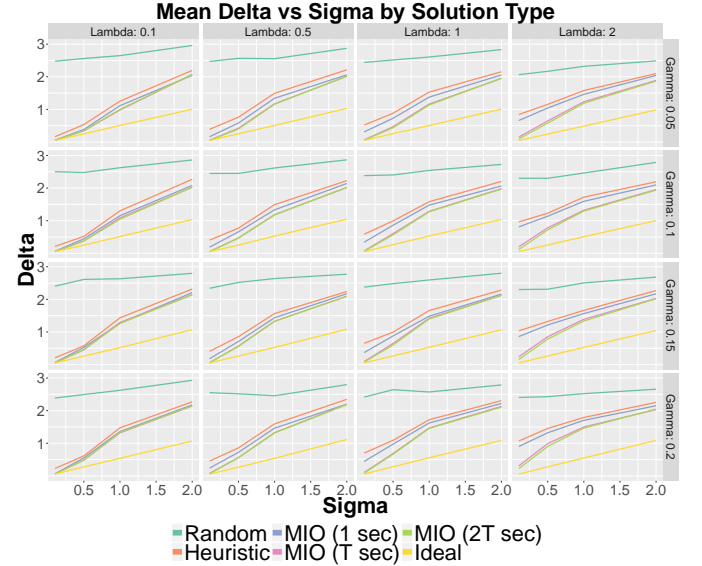

Fig. 11: δ of robust heuristic and MIO as compared to random solutions for scenarios of 4 targets and 8 scans.
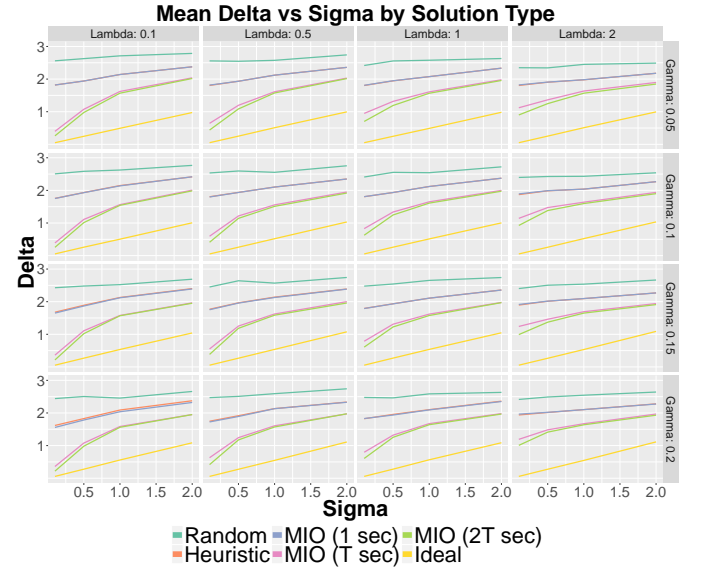

Fig. 12: δ of robust heuristic and MIO as compared to random solutions for scenarios of 8 targets and 8 scans.

Again, we measure against the basic approaches by comparing Figure 11 with the 4 target and 8 scan element of Figure 6. We see that the robust approaches do not drastically reduce in performance for the easiest robust scenario of $\gamma = 0.05$ and $\lambda = 0.1$. However, the gap in performance between the ideal solution and the solutions of the robust algorithms grows wider with increases in σ, something that is expected but was not as sizable in the basic experiment. Therefore, the robust

approaches may be less robust to increases in σ in scenarios with detection ambiguity. However, it also appears that these methods are more robust to increases in the false alarm rate λ when it comes to trajectory estimation than they were when it came to data association, especially in the larger scenario shown in Figure 12. We conclude that our robust methods are fairly robust to both increases in the false alarm rate and decreases in trajectory estimation under detection ambiguity, but increasing the signal noise degrades the performance of our methods more so in scenarios with detection ambiguity than in scenarios without detection ambiguity.

In summary,.....discussion on robust summary

VIII. Summary and Future Work

We presented multi-target tracking approaches which jointly solve the problems of data association and trajectory estimation via global optimization methods using a single objective function. To this end, we proposed the use of a randomized local search heuristic as a warm start for a mixed integer optimization model, and we did so for scenarios with and without detection ambiguity. We accomplish this without the need of a trajectory bank nor the a prior computation of trajectory hypotheses. We showed that the heuristic finds good quality feasible solutions very quickly. We also show that the MIO offers improvement over the heuristic and this improvement is found in real time for applications considered by this paper.

The proposed methods show potential for use in an online algorithm with a sliding window, suggesting one possible area of future work. Additionally, future research could explore the use of more complex penalty functions. One possible idea would be the use of piecewise linear functions which would require further expansion of our formulations. Other areas of future work could seek to relax some of the scenario based assumptions mad in Assumption 1, in particular, extensions to non-linear trajectories or the birth/death of targets.

Appendix A
Detection Ambiguity Penalty Values

Here we provide recommendations for the tuning of penalty parameters θ and ϕ. We begin with an explanation grounded in logic. It can be shown that as the false alarm rate λ increases, the number of expected false alarms also increases. Therefore, it stands to reason that as a general rule of thumb the false alarm penalty θ should decrease as λ increases. Similarly, the number of expected missed detections increases as the missed detection probability γ increases, and so too the missed detection penalty should decrease. Then it follows logically that the missed detection penalty ϕ should increase as γ decreases. Furthermore, it is convenient to reason that the value of both of these penalties should somehow be tied to the value of σ, though this is a more difficult sequence of logic to justify. Through examination we found these logical concepts to hold true across a variety of scenario sizes and difficulties. Using this insight as well as the results of a mini experiment, we arrived at the penalty values summarized in Table III.

σ	γ	λ	θ	ϕ
0.1	0.2	0.1	0.4	0.1
0.1	0.2	0.5	0.2	0.1
0.1	0.2	1.0	0.3	0.2
0.1	0.2	2.0	0.1	0.2
0.1	0.15	0.1	0.4	0.1
0.1	0.15	0.5	0.4	0.1
0.1	0.15	1.0	0.4	0.3
0.1	0.15	2.0	0.1	0.3
0.1	0.1	0.1	0.3	0.1
0.1	0.1	0.5	0.3	0.2
0.1	0.1	1.0	0.2	0.2
0.1	0.1	2.0	0.1	0.3
0.1	0.05	0.1	0.3	0.2
0.1	0.05	0.5	0.2	0.2
0.1	0.05	1.0	0.1	0.4
0.1	0.05	2.0	0.1	0.4
0.5	0.2	0.1	0.5	0.1
0.5	0.2	0.5	0.4	0.2
0.5	0.2	1.0	0.4	0.2
0.5	0.2	2.0	0.3	0.3
0.5	0.15	0.1	0.5	0.1
0.5	0.15	0.5	0.4	0.2
0.5	0.15	1.0	0.4	0.3
0.5	0.15	2.0	0.3	0.4
0.5	0.1	0.1	0.4	0.3
0.5	0.1	0.5	0.4	0.3
0.5	0.1	1.0	0.3	0.3
0.5	0.1	2.0	0.2	0.4
0.5	0.05	0.1	0.4	0.4
0.5	0.05	0.5	0.4	0.4
0.5	0.05	1.0	0.3	0.5
0.5	0.05	2.0	0.3	0.5
1.0	0.2	0.1	0.5	0.1
1.0	0.2	0.5	0.5	0.2
1.0	0.2	1.0	0.5	0.1
1.0	0.2	2.0	0.5	0.4
1.0	0.15	0.1	0.5	0.1
1.0	0.15	0.5	0.5	0.3
1.0	0.15	1.0	0.5	0.2
1.0	0.15	2.0	0.5	0.4
1.0	0.1	0.1	0.5	0.5
1.0	0.1	0.5	0.5	0.4
1.0	0.1	1.0	0.4	0.4
1.0	0.1	2.0	0.4	0.5
1.0	0.05	0.1	0.5	0.5
1.0	0.05	0.5	0.5	0.5
1.0	0.05	1.0	0.5	0.5
1.0	0.05	2.0	0.5	0.5
2.0	0.2	0.1	0.5	0.1
2.0	0.2	0.5	0.5	0.2
2.0	0.2	1.0	0.5	0.1
2.0	0.2	2.0	0.5	0.5
2.0	0.15	0.1	0.5	0.1
2.0	0.15	0.5	0.5	0.3
2.0	0.15	1.0	0.5	0.2
2.0	0.15	2.0	0.5	0.5
2.0	0.1	0.1	0.5	0.5
2.0	0.1	0.5	0.5	0.5
2.0	0.1	1.0	0.5	0.5
2.0	0.1	2.0	0.5	0.5
2.0	0.05	0.1	0.5	0.5
2.0	0.05	0.5	0.5	0.5
2.0	0.05	1.0	0.5	0.5
2.0	0.05	2.0	0.5	0.5

TABLE III: Experiment 2 penalty values.

Appendix B
Robust MIO With Number of Targets as a Decision Variable

For completeness here we present an alternative approach to solving the MTT problem with detection ambiguity. This

MIO model is an extension to (24) that directly determines the number of targets via optimization by incorporating additional variables and constraints into framework of the formulation.

A. Decision Variables

Because we do not assume a fixed number of targets, this decision must now be made by the model. Toward this goal, we introduce a new binary decision variable w_j to indicate whether or not trajectory j corresponds to an existing target. Any detections assigned to a non-existing target would be a false alarm.

$$w_j = \begin{cases} 1, & \text{if trajectory } j \text{ exists,} \\ 0, & \text{otherwise.} \end{cases}$$

Since we by Assumption 3 the number of targets will be at most N_1, the number of needed variables will be N_1.

B. Constraints

Most constraints remain similar to their original counterparts in (24), except for slight adjustments needed to account for the possibility that some trajectories may not exist. Explicitly, the number of possible trajectories is now N_1 so all instances of summing over j must be changed accordingly. For example, we adjust (21) and (23) as follows:

$$\sum_{j=1}^{N_1} y_{itj} + F_{it} = 1 \qquad \forall i, t,$$

$$\sum_{j=1}^{N_1} \sum_{t=1}^{T} M_{jt} = TM.$$

By the same accord the RHS of (22) no longer equals to 1 because some trajectories may not exist. Therefore we say all *existing* trajectories must either be assigned a detection or a missed detection, which implies the following constraints

$$\sum_{i=1}^{n_t} y_{itj} + M_{jt} = w_j \qquad \forall j, t. \qquad (26)$$

Moreover, we restrict α_j and β_j to be zero if trajectory j does not exist. This ensures only existing trajectories are penalized in the objective function.

$$|\alpha_j| + |\beta_j| \le M_0 w_j \qquad \forall j.$$

This model is symmetric between various w_j variables. Such symmetry in models is an undesirable quality since the multiple optimal solutions cause solving such problems to be less efficient. We can actually reduce the symmetry that is inherent to this formulation. Since $N_0 \le P \le N_1$, we can set $w_j = 1$ for all $j = 1, \dots, N_0$, which leaves us with only $N_1 - N_0$ additional binary variables. Adding the constraints

$$w_{N_0+1} \ge \dots \ge w_{N_1},$$

further reduces the number of equivalent solution and increases the efficient resolvability of the model.

C. Full Formulation

Incorporating these additional variables and constraints, we arrive at the following complete alternative formulation.

$$\underset{\psi_{jt}}{\text{minimize:}} \quad \sum_{j=1}^{N_1} \sum_{t=1}^{T} \psi_{jt} + \theta \cdot TF + \phi \cdot TM$$

$$\text{subject to:} \quad \sum_{j=1}^{N_1} y_{itj} + F_{it} = 1 \qquad \forall i, t$$

$$\sum_{i=1}^{n_t} y_{itj} + M_{jt} = 1 \qquad \forall j = 1, \dots, N_0, t$$

$$\sum_{i=1}^{n_t} y_{itj} + M_{jt} = w_j \qquad \forall j = N_0, \dots, N_1, t$$

$$\sum_{i=1}^{n_t} \sum_{t=1}^{T} F_{it} = TF$$

$$\sum_{j=1}^{N_1} \sum_{t=1}^{T} M_{jt} = TM$$

$$w_{N_0+1} \ge \dots \ge w_{N_1}$$

$$|\alpha_j| + |\beta_j| \le M_0 w_j \qquad \forall j$$

$$x_{it} y_{itj} + M_t(1 - y_{itj}) \ge z_{jt} \qquad \forall i, t, j$$

$$x_{it} y_{itj} - M_t(1 - y_{itj}) \le z_{jt} \qquad \forall i, t, j$$

$$z_{jt} - \alpha_j - \beta_j t \le \psi_{jt} \qquad \forall j, t$$

$$-(z_{jt} - \alpha_j - \beta_j t) \le \psi_{jt} \qquad \forall j, t$$

$$y_{itj} \in \{0, 1\} \quad \forall i, t, j$$

$$\alpha_j \in \mathbb{R}^n, \quad \beta_j \in \mathbb{R}^n, \quad w_j \in \mathbb{R}^n \quad \forall j$$

$$z_{jt} \in \mathbb{R}^n, \quad \forall j, t.$$

D. Extension of Robust Heuristic

In order to initialize the above MIO we need to explore the entire state space which include specifying the number of targets. Such a heuritic will be identical to that proposed in Section V, except the number of target used will also be randomly selected during the initialization process. As in all previous cases, the warm start will be chosen among the different heuristic solutions based on the objective function value.

APPENDIX C
TRAJECTORY ASSIGNMENT PAIRING

In order to analyze the performance of a multi-target tracking algorithm, we must first find the best matching of the true trajectories of the scenario and the estimated trajectories of the algorithm solution. Put differently, we wish to find a set of assignment pairings which match true and estimated trajectories. Here we present a linear optimization model which solves for the globally optimal assignment pairings of true and estimated trajectories. In addition, we extend the assignment problem to scenarios with detection ambiguity, where the algorithm may overestimate or underestimate the number of targets. When this situation arises, a few minor adjustments are necessary to find the optimal assignment pairing.

The goal of this assignment problem is to optimally assign pairs of true trajectories i to estimated trajectories j if there exists such a pairing to be made. Only a single set of decision variables are needed to determine if the true trajectory i should be assigned to the estimated trajectory j or not.

$$y_{ij} = \begin{cases} 1, & \text{if true trajectory } i \text{ is assigned} \\ & \quad \text{to estimated trajectory } j, \\ 0, & \text{otherwise.} \end{cases}$$

Remember that we denote the true position of trajectory i at scan t with $\bar{x}_{it}$ and the estimated position of trajectory j at scan t with $\hat{x}_{jt}$. Then the cost c_{ij} of assigning true trajectory i to estimated trajectory j is the norm distance between these two trajectories as measured at each scan.

$$c_{ij} = \sum_{t=1}^{T} \|\bar{x}_{it} - \hat{x}_{jt}\|$$

If we denote the true number of targets as P_{true} and the estimated number of targets as P_{est} then the objective of the integer optimization model would be:

$$\underset{y_{ij}}{\text{minimize:}} \sum_{i=1}^{P_{\text{true}}} \sum_{j=1}^{P_{\text{est}}} c_{ij} y_{ij}$$

When the number of true targets is equal to the number of estimated targets ($P_{\text{true}} = P_{\text{est}} = P$), we simply require two equality constraints to ensure that each true trajectory i is assigned to exactly one estimated trajectory j and vice versa.

$$\sum_{i=1}^{P} y_{ij} = 1 \qquad \forall j = 1, ..., P \tag{27}$$

$$\sum_{j=1}^{P} y_{ij} = 1 \qquad \forall i = 1, ..., P \tag{28}$$

However, when the number of estimated trajectories differs from the number of true trajectories, these constraints must be modified slightly. In the case where the number of true targets exceeds the estimated number of targets ($P_{\text{true}} \geq P_{\text{est}}$), we restrict each true trajectory i to the assignment of *at most* one estimated trajectory j, and Equation 27 is modified to:

$$\sum_{i=1}^{P_{\text{true}}} y_{ij} \leq 1 \qquad \forall j = 1, ..., P_{\text{est}}.$$

On the contrary, when the number of estimated targets exceeds the true number of targets ($P_{\text{true}} \leq P_{\text{est}}$), then we restrict each estimated trajectory j to the assignment of *at most* one true trajectory i, and Equation 28 is modified to:

$$\sum_{j=1}^{P_{\text{est}}} y_{ij} \leq 1 \qquad \forall i = 1, ..., P_{\text{true}}.$$

In summary, the generalized integer optimization assignment model is presented below.

$$\underset{y_{ij}}{\text{minimize:}} \sum_{i=1}^{P_{\text{true}}} \sum_{j=1}^{P_{\text{est}}} c_{ij} y_{ij}$$

$$\text{subject to:} \sum_{i=1}^{P_{\text{true}}} y_{ij} = 1 \qquad \forall j = 1, ..., P_{\text{est}}$$

$$\sum_{j=1}^{P_{\text{est}}} y_{ij} = 1 \qquad \forall i = 1, ..., P_{\text{true}}$$

$$y_{ij} \in \{0, 1\} \quad \forall i = 1, ..., P_{\text{true}}, j = 1, ..., P_{\text{est}}$$

This model is vital to scoring the performance of an MTT algorithm's solution because it ensures the globally optimal pairing of true and estimated trajectories.

REFERENCES

[1] G. Pulford, "Taxonomy of multiple target tracking methods," *Radar, Sonar and Navigation, IEE Proceedings* -, vol. 152, no. 5, pp. 291–304, October 2005.

[2] Y. Bar-Shalom and X. Li, *Multitarget-multisensor Tracking: Principles and Techniques*. Yaakov Bar-Shalom, 1995. [Online]. Available: https://books.google.com/books?id=GfOoMQEACAAJ

[3] Y. Bar-Shalom and X.-R. Li, *Estimation with Applications to Tracking and Navigation*. New York, NY, USA: John Wiley & Sons, Inc., 2001.

[4] S. Blackman, *Multiple-target Tracking with Radar Applications*, ser. Radar Library. Artech House, 1986. [Online]. Available: https://books.google.com/books?id=Ag9TAAAAMAAJ

[5] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[6] D. Reid, "An algorithm for tracking multiple targets," *Automatic Control, IEEE Transactions on*, vol. 24, no. 6, pp. 843–854, Dec 1979.

[7] S. Blackman, "Multiple hypothesis tracking for multiple target tracking," *Aerospace and Electronic Systems Magazine, IEEE*, vol. 19, no. 1, pp. 5–18, Jan 2004.

[8] J. K. Wolf, A. M. Viterbi, and G. S. Dixon, "Finding the best set of k paths through a trellis with application to multitarget tracking," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 25, no. 2, pp. 287–296, Mar 1989.

[9] D. Castanon, "Efficient algorithms for finding the k best paths through a trellis," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 26, no. 2, pp. 405–410, Mar 1990.

[10] R. Perry, A. Vaddiraju, and K. Buckley, "Multitarget list viterbi tracking algorithm," in *Signals, Systems amp; Computers, 1998. Conference Record of the Thirty-Second Asilomar Conference on*, vol. 1, Nov 1998, pp. 436–440 vol.1.

[11] C. Morefield, "Application of 0-1 integer programming to multitarget tracking problems," *Automatic Control, IEEE Transactions on*, vol. 22, no. 3, pp. 302–312, Jun 1977.

[12] C. Carthel and S. Coraluppi, "Multi-hypothesis sonar tracking," in *Fusion 2004: Seventh International Conference on Information Fusion*, 2004.

[13] A. P. Poore and N. Rijavec, "A lagrangian relaxation algorithm for multidimensional assignment problems arising from multitarget tracking," *SIAM Journal on Optimization*, vol. 3, no. 3, pp. 544–563, 1993. [Online]. Available: http://dx.doi.org/10.1137/0803027

[14] A. Andriyenko and K. Schindler, "Multi-target tracking by continuous energy minimization," in *CVPR*, 2011.

[15] A. Andriyenko, K. Schindler, and S. Roth, "Discrete-continuous optimization for multi-target tracking," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, June 2012, pp. 1926–1933.

[16] A. Lodi, *50 Years of Integer Programming 1958-2008.From the Early Years to the State-of-the-Art*, M. Jnger, T. M. Liebling, D. Naddef, G. L. Nemhauser, W. R. Pulleyblank, G. Reinelt, G. Rinaldi, and L. A. Wolsey, Eds. Springer Berlin Heidelberg, 2010.

[17] R. E. Bixby, "Mixed-integer programming: It works better than you may think," http://www.ferc.gov/CalendarFiles/20100609110044-Bixby, %20Gurobi%20Optimization.pdf, 2010, accessed 4 April 2016s.

[18] I. Gurobi Optimization, "Gurobi 6.5 performance benchmarks," http://www.gurobi.com/pdfs/benchmarks.pdf, 2015, accessed 4 April 2016s.

[19] G. Nemhauser, "Integer programming: Global impact," https://smartech.gatech.edu/bitstream/handle/1853/49829/presentation.pdf, 2013, accessed 4 April 2016s.

[20] Top500 Supercomputer Sites, "Performance development," http://www.top500.org/statistics/perfdevel/, 2015, accessed 4 April 2016s.

[21] B. Ristic, B. N. Vo, D. Clark, and B. T. Vo, "A metric for performance evaluation of multi-target tracking algorithms," *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3452–3457, July 2011.

[22] C. Heij, P. de Boer, P. Franses, T. Kloek, H. van Dijk, and A. Rotterdam, *Econometric Methods with Applications in Business and Economics*. OUP Oxford, 2004. [Online]. Available: https://books.google.com/books?id=PJf6GUAavhAC

[23] N. J. Stor, I. Slapnicar, and J. L. Barlow, "Accurate eigenvalue decomposition of real symmetric arrowhead matrices and applications," *Linear Algebra and Its Applications*, vol. 464, pp. 62–89, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0024379513006265

[24] M. Lubin and I. Dunning, "Computing in operations research using julia," *INFORMS Journal on Computing*, vol. 27, no. 2, pp. 238–248, 2015. [Online]. Available: http://dx.doi.org/10.1287/ijoc.2014.0623

[25] I. Gurobi Optimization, "Gurobi optimizer reference manual," 2015. [Online]. Available: http://www.gurobi.com

[26] N. Bliss, R. Bond, J. Kepner, H. Kim, and A. Reuther, "Interactive grid computing at lincoln laboratory," *MIT Lincoln Laboratory Journal*, vol. 16, no. 1, 2006.