Missile Defense Strategy: Towards Optimal Interceptor Allocation

Dane Hankamer
Computer Science
Stanford University
Stanford, United States of America
$[dhank]$

James Myerson
Management Science & Engineering
Stanford University
Stanford, United States of America
$[jmyerson]$

Galle Smagghe
Management Science & Engineering
Stanford University
Stanford, United States of America
$[gsmagghe]$

***Abstract*—Missile defense is a field of paramount importance in military operations and strategic defense. Whether employed in the context of defending civilian populations from conventional or nuclear weapons, or for the protection of warships in a battle group, improving technologies are both increasing missile defense capabilities and augmenting the challenges such systems face from adversarial weaponry. This paper seeks to explore methods for developing and analyzing strategies for missile interceptor allocation in the face of uncertain enemy resources and actions.**

I. Introduction

We consider the Ballistic Missile Defense (BMD) problem of allocating anti-ballistic missiles (ABMs) when faced with an adversarial missile attack. In the event of an attack, a blue actor (the defender) is unsure of the strategy by which the red actor (aggressor) will prosecute the attack. In a compressed timeline, blue decision-makers must act quickly to formulate and execute a defensive response. We are interested in approaches for assigning finite ABM resources to counter a missile attack in the face of uncertain enemy actions.

This problem is of great interest to nations around the globe. The U.S. Missile Defense Agency projects a \$500M budget in FY 2019 for an integrated master test plan that informs BMD policy [1]. However, there is little public research on Markov Decision Process (MDP) approaches to solving BMD problems. Instead, heuristics are largely used to inform ABM allocation. A common BMD heuristic strategy is to launch two interceptors at each incoming missile [2]. Our goal is to explore and evaluate general methods to construct more optimal interceptor assignment strategies across attack scenarios.

We frame this problem as an MDP using Monte Carlo Tree Search (MCTS). We selected an MCTS approach because our large state space leads to a high branching factor. Implementing MCTS allows us to toggle exploration and generate policies without the computational expense of a minimax algorithm. We begin by defining the state space with seven vectors of variables: number of ABMs fired per launcher, number of ABMs remaining per launcher (in chamber (launchable) and in magazine (reserve)), health status of defensive target, time until impact of each enemy missile, target of each enemy missile, and enemy missile type. Note that, in order to keep the dimension of the state space constant, we impose a continuous track of 100 missiles; we set the value of each track to null when it is not assigned to a live missile. Actions include number of missiles to fire at each tracked missile from each launcher, and an automatic reload if possible (launcher cannot fire, and must have reserve interceptors remaining). The simulation calculates rewards with a function of the change in health points, the cost of firing missiles, and rewards for killing incoming missiles.

We define a successful policy as one that outperforms the baseline heuristic of assigning 2 midcourse ABMs and 2 terminal ABMs per attacking missile if needed. We also examine the effect of changing problem parameters to analyze algorithm behavior for undermatched, overmatched, and roughly equivalently resourced opponents. Finally, we conclude with policy recommendations for ABM allocation.

II. Related Work

We were inspired by Jason Reinhardt's PhD thesis in which he breaks down the risk of nuclear deterrence failure using infinite horizon Partially Observable Markov Decision Processes (POMDPs). He applies Incremental Policy Iteration (IPI) to derive approximately optimal policies from uncertain antagonist strategies [3]. Our motivation is to examine a physical component of deterrence strategy, BMD, in its many defense applications, both strategic and tactical. Our approach differs by using MCTS for exploring approximately optimal policies and limiting the focus to a finite horizon scenario where the end of an attack is known.

Research on the Weapon Target Assignment (WTA) problem also provides generalized approaches for exact and heuristic weapon target assignment. These methods could be used to plan an approximately optimal red strategy for targeting blue assets [4] [5]. Our BMD model dynamically responds to a series of red strategies, assigning ABM resources based on our state action space.

III. Problem Formulation

Our simulation simplifies the world to a battlefield containing red and blue teams along with their assets. In reality, the BMD allocation problem becomes much more complex when humans and coordinated systems work together across a nearly infinite state space.

In our scenario, the red team represents insurgent aggressors with varying conventional missile resources. The blue team attempts to defend 4 targets using mid-course and terminal ABMs. Specifically, the blue team defends a civilian population center, a military base, and 2 terminal launchers. Each terminal launcher holds 4 ABMs in the chamber and has 12 reserve missiles. Additionally, the blue team has 24 chambered mid-course ABMs with 24 in the reserve magazine; note that the mid-course launcher (e.g. an Aegis Missile Destroyer) is beyond the red team's capacity to target due to maneuverability, location, and sophistication. The red team utilizes 2 types of munitions. Type I missiles inflict minor damage and receive little to no guidance, which results in a low hit rate. Type II missiles cause significant damage to hardened targets (the military base) and represent guided missiles with a higher accuracy.

Negative rewards are associated with firing ABMs or losing health points at any of the 4 targets, while a positive reward is given for eliminating an incoming ballistic missile. Illustrative multipliers are used to scale the negative rewards received depending on the defense target, and can be tuned. We run war simulations for 3 distinct enemy types: low resource, moderate resource, and resource rich. Each war continues for 30 time steps, where an attack missile takes up to 10 time steps to reach a target and the red team stops discharging missiles after 20 time steps (representing a blue kinetic intervention against enemy launchers). The low resource enemy has a 33% chance of launching a missile at each time step and primarily fires Type I missiles at random targets. The red team with moderate resources has a 50% chance of launching a missile for up to 2 missiles at every time step and employs a mixture of Type I and Type II missiles across military and civilian targets. The resource rich enemy has a 66% chance of launching a missile for up to 4 missiles at each time step and strategically targets military assets with Type II missiles and larger civilian centers with Type I missiles. Interceptors can be fired at missiles with specific distances to target, approximating the "mid-course" and "terminal" capabilities of these launchers. By analyzing a wide variety of red team strategies, we can use the simulation results to better inform allocation policies across a set of potential adversarial actors.

IV. Methods

While this topic requires extremely detailed and complex formulation for a realistic implementation taking into account battlefield command, control, computers, intelligence, surveillance, and reconnaissance systems, as well as the coordination of the human operators of such systems, we sought to demonstrate the viability of improving BMD strategies with MCTS through a simplified simulation scenario.

A. MDP with MCTS (Generative Model)

MCTS is a sampling-based online approach that executes many simulations from the current state while updating an estimate of the state-action value function $Q(s, a)$. Given our immense state-action space and time tracking, MCTS is advantageous because the complexity does not grow exponentially with the horizon [6], and for the same reason we implemented a generative model instead of specifying a full transition model. This allows us to consider only the actions available from the state we are in, rather than the full action space. This aids computation in our model where it is often the case that the only possible action is to continue tracking and/or reload (e.g. no missiles are incoming, or all are out of viable intercept windows).

MCTS is broken down into search, expansion, and rollout phases. In the search portion of the algorithm, we update $Q(s, a)$ for the explored state-action pairs and track the number of times an action was taken from each observed state. At each time step, we perform the action that maximizes

$$Q(s, a) + c\sqrt{\frac{logN(s)}{N(s, a)}}$$

where $N(s) = \sum_a N(s, a)$ and c is our exploration constant set to 5 for our model. In the expansion phase, we initialize $N(s, a)$ and $Q(s, a)$ to 0 and iterate over all possible actions from each new state reached. The rollout phase of our model selects the action that maximizes reward at every time step until the desired depth is reached using a random policy. We implement a decreasing depth parameter that corresponds to time until the engagement ends via intervention. Specifically, our MCTS model looks 31 actions ahead at the beginning of the war and only 1 action ahead at the end of the war. This depth parameter incentivizes the blue team to protect its launch capabilities and meter its resources at the outset of the engagement, and liberally engage incoming threats for increased protection when the end of the war is imminent. Altogether, we run 100 simulations with no defensive actions, the baseline heuristic, and the MCTS method for poor and moderate resource enemy strategies, and 50 simulations for each for the resource rich enemy.

B. POMDP with MCTS

In the initial MDP analysis, we programmed complete observation of the state space and the parameters, as a proxy for perfect sensors and intelligence. The blue actor's coordinated systems "knew" the intercept probabilities of his weapons on each enemy track, the accuracy of enemy missiles in all permutations of type and target, and action outcomes. Illustrative initializations were chosen based on approximate translations of systems to our extremely simplified scenario.

We relax these ideal assumptions in the second phase of our analysis, and utilize a POMDP. Our POMDP method expands the state space to include beliefs over the probability of an enemy missile successfully hitting each defensive target and beliefs about the success rates of mid-course and terminal ABMs. We initialize Dirichlet distributions to uniform priors over the probability of success for both red and blue team missiles. The new state space adds 24 elements to the initial MDP state space: 16 states for success and failure counts of Type I and Type II missiles at each of the 4 targets and 8 states for success and failure counts of mid-course and terminal ABMs aimed at Type I and Type II ballistic missiles.

The probability of a red team missile i hitting a blue team target j is represented by

$$P([i]\ hits\ target\ [j]) = \frac{\alpha_{ij}}{\alpha_{ij}+\beta_{ij}}\ \forall i,j\ \in \{1,2\}\times\{1,2,3,4\}$$

where α and β represent success and failure counts. The counts of α and β for Type I and Type II missiles are increased by 1 for each observation of a red missile reaching a blue target. Similarly, the probability of a blue team missile k eliminating a red missile type i is given by

$$P([k]\ kills\ [i]) = \frac{\alpha_{ki}}{\alpha_{ki}+\beta_{ki}} \forall k,i\ \in \{1,2\}\times\{mid, ter\}$$

where α and β represent success and failure counts. Success counts are increased by 1 for each red missile destroyed, and failure counts increase corresponding to the number of ABMs launched against an incoming missile.

For each time step, the current state is updated according to the set model parameters. However, we act based on our belief about the effectiveness of red and blue missile types. We run 100 simulations for each enemy strategy and blue policy on the low and moderate resource red teams.

V. Model Evaluation

A. MDP with MCTS

For the low resource enemy, the mean rewards over 100 simulations were:

- no action policy: -173.5
- baseline: -30.6
- MCTS: -20.3

Thus the MCTS model outperformed both the no action policy and the baseline heuristic policy. The rewards observed at each time step are displayed in the graphs below for each of the scenarios (figures 1,2,4).

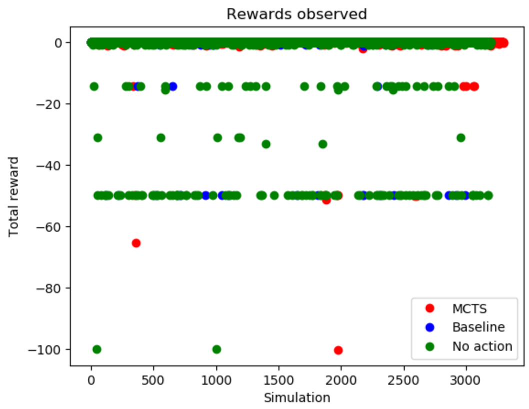

Fig. 1. Low Resource Enemy: Rewards per Sim. Step

For the moderate resource enemy, the mean rewards over the 100 simulations were:

- no action policy: -458.4
- baseline: -101.7
- MCTS: -47.9

Here again the MCTS model outperformed both comparison strategies. However, the negative reward was $\approx 47\%$ of that of the baseline heuristic strategy, an improvement over the $\approx 66\%$ metric observed in the low resource scenario. As the volume of missiles closer to being comparable to the number of interceptors available, the model performed better relative to the benchmark. Interestingly, as can be seen from figure 3, the MCTS model closely follows the heuristic strategy of firing two terminal missiles at each targetable threat. The mid-course launcher is far more heavily relied upon, and we found that it is predisposed to primarily preserve terminal launch capabilities.

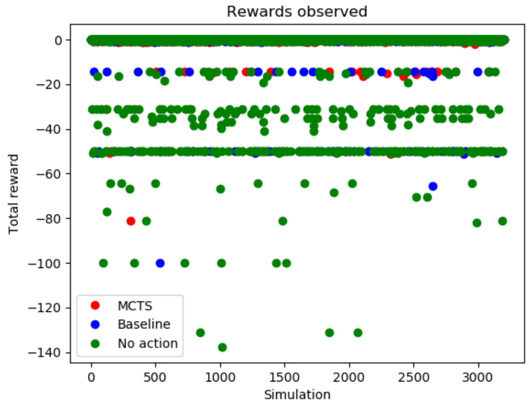

Fig. 2. Moderate Resource Enemy: Rewards per Sim. Step

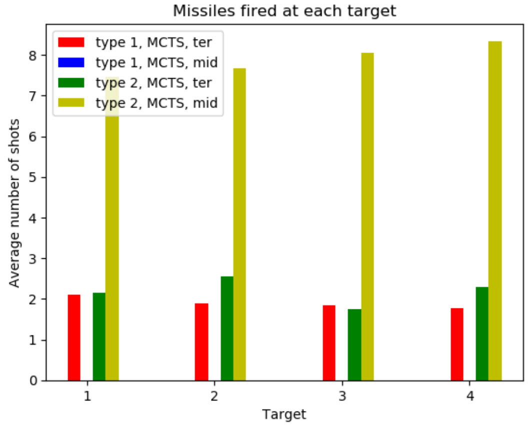

Fig. 3. Moderate Resource Enemy: Interceptor Strategy Profile

For the resource rich enemy, the mean rewards over the 100 simulations were:

- no action policy: -2943.3
- baseline: -788.1
- MCTS: -604.7

In this scenario, where the large resource enemy has overwhelming superiority in terms of missile to interceptor ratio, the performance advantage of the MCTS model is significantly narrowed: the negative reward of the MCTS model is $\approx 77\%$ of that of the baseline heuristic model.

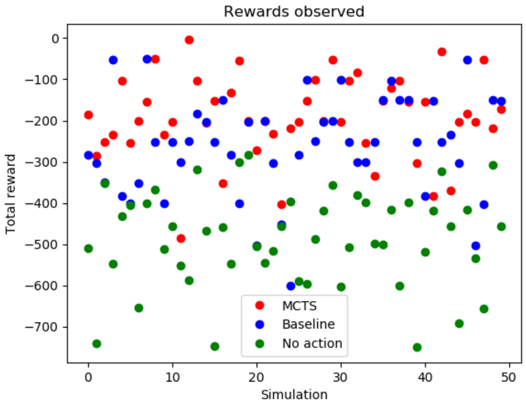

Fig. 4. Resource Rich Enemy: Rewards per Simulation

We now consider the performance of the POMDP model.

B. POMDP with MCTS

Figures 5 and 6 show the results achieved by the POMDP formulation. For the low resource enemy, the POMDP model calculated a mean reward of -11.8, which represents $\approx 58\%$ of that of the MDP model and $\approx 39\%$ of the negative reward for the baseline model. This result is not repeated in the medium resource enemy scenario, where we observed that the average negative reward was -108.9, which is larger than the -101.7 average negative reward achieved by the baseline heuristic strategy, and $\approx 2.3\times$ larger than the MCTS result in the face of similar enemy actions.

The POMDP model achieved varying levels of success in updating beliefs over the course of the simulations to be close to the values of the true parameters. We examine these results in the context of the low resource enemy, where observations were sparse. In particular, the beliefs concerning the accuracy of enemy missiles were all on the interval $[.44, .51]$, whereas the true parameters were contained on $[0.2, 0.95]$. This is due to the fact that there were very few "leakers" and "free riders" in this scenario, i.e. missiles which evaded the interceptors or were never fired at, and so very few opportunities to observe whether a missile hit or missed a target. In the medium resource enemy scenario, this issue was ameliorated to some degree. We observe a clear correspondence between belief update accuracy and number of observations in this case.

VI. Conclusions

Both of our models performed reasonably well against the benchmark heuristic strategies encoded from public operational procedures. Both the MDP and POMDP formulation performed similarly; however, the moderate resource enemy scenario showed a decrease in the performance of the POMDP below the level of the heuristic BMD strategy. We will investigate whether encoding prior beliefs from ABM test rates rather than a uniform prior would restore the advantage of the model. We are also interested to see the effect of

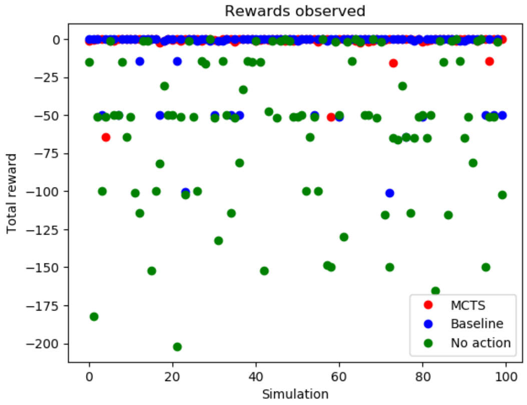

Fig. 5. Poor Resource Enemy: POMDP Rewards per Simulation

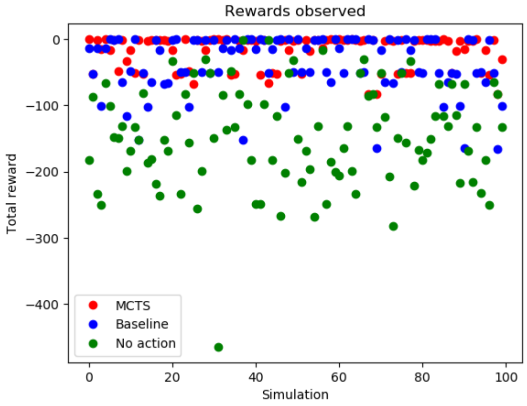

Fig. 6. Moderate Resource Enemy: POMDP Rewards per Simulation

changing the exploration constant c to find the optimal tradeoff between exploration and exploitation. In the work presented, we encountered difficulties with correspondence between our constructed world and realistic operations, as varying this constant led to the expenditure of sub-optimally enormous numbers of interceptors at single enemy tracks (esp. midcourse). Different update laws and distributions could be implemented to more closely mirror reality as well.

Moving forward, a great deal can be done to evolve our model to incorporate more realistic and complex strategic simulations. Furthermore, a number of illustrative figures were utilized to encode the costs of blue resources as well as the importance of target protection, and these parameters can be tuned to more accurately reflect the prioritizations of stakeholders evaluating the model. Additionally, with more time, it would be possible to incorporate responsive and optimal enemy strategies to challenge the capacities of our system. Overall, we are excited at the prospect of furthering understanding in this arena, and encouraged by the preliminary results we saw.

Participation

The team all participated a great deal in the project. Dane and James researched and evaluated ideas as candidates for the project. They worked to scope and describe both the initial problem (launch on warning policy evaluation) and the problem that was eventually explored for this assignment. Galle and James worked together to translate the problem into a formulation that could be evaluated with the methods from this class, as both an MDP and a POMDP. Galle worked to implement this MDP formulation in Python with James, and then translated this into Julia. Galle then implemented the POMDP while Dane and James built the report out, and then we all finished the report together.

References

[1] ”Fiscal Year 2019 Budget Estimates” (U.S. Missile Defense Agency, March 2018).

[2] Theodore A. Postal, Targeting, in Managing Nuclear Operations, ed. Ashton B. Carter, John D. Steinbruner, and Charles A. Zraket (Washington, DC: Brookings Institution, 1987), 373406.

[3] Reinhardt, J 2018, ”A Probabilistic Analysis of the Risk of Nuclear Deterrence Failure,” PhD Thesis, Stanford University, Stanford CA.

[4] Yucel et al., ”The Generalized Weapon Target Assignment Problem” (10th International Command and Control Research and Technology Symposium, 2005).

[5] Ahuja, Kumar, Jha, & Orlin, ”Exact and Heuristic Algorithms for the Weapon Target Assignment Problem” (2003).

[6] M. J. Kochenderfer, ”Decision Making Under Uncertainty: Theory and Application” (MIT Press, 2015), 102-103.