

Cooperative Task Assignment of Unmanned Aerial Vehicles in Adversarial Environments

Mehdi Alighanbari and Jonathan P. How

Aerospace Controls Laboratory
Massachusetts Institute of Technology
{mehdi_a, jhow}@mit.edu

Abstract—This paper addresses the problem of risk in the environment and presents a new stochastic formulation of the UAV task assignment problem. This formulation explicitly accounts for the interaction between the UAVs – displaying *cooperation* between the vehicles rather than just *coordination*. As defined in the paper, cooperation entails coordinated task assignment with the additional knowledge of the future implications of a UAV’s actions on improving the expected performance of the other UAVs. The key point is that the actions of each UAV can reduce the risk in the environment for all other UAVs; and the new formulation takes advantage of this fact to generate cooperative assignments that achieve better performance. This change in the formulation is accomplished by coupling the failure probabilities for each UAV to the selected missions for all other UAVs. This results in coordinated plans that optimally exploit the coupling effects of cooperation to improve the survival probabilities and expected performance. This allocation is shown to recover real-world air operations planning strategies that provide significant improvements over approaches that do not correctly account for UAV attrition.

The problem is formulated as a Dynamic Programming (DP) problem, which is shown to be more computationally tractable than previous MILP solution approaches. Two DP approximation methods (the one-step and two-step look-ahead) are also developed for larger problems. Simulation results show that the one-step look-ahead can generate cooperative solutions very quickly, but the performance degrades considerably. The two-step look-ahead policy generates plans that are very close to (and in many cases, identical to) the optimal solution and the computation time is still significantly lower than the exact DP approach.

I. INTRODUCTION

Real-world air operations planners rely on cooperation between aircraft in order to manage the risk of attrition. Missions are often scheduled so that one group of aircraft opens a corridor through the anti-aircraft defenses to allow a second group to attack the protected higher value targets. The main planning challenges involve utilizing the maximum integrated capabilities of the team, especially when each UAV can perform multiple functions (*e.g.*, both destroy anti-aircraft defenses and attack high value targets). Cooperation is not just desirable, it is crucial for designing successful missions in heavily defended environments. A successful method of performing task allocations cannot assume the mission will always be executed as designed, but must account for an adversary in the environment who is actively attempting to cause failure. Simulations presented

in this paper show that ignoring the cooperation in the assignment results in mission plans that have much lower expected performance/value. Furthermore, techniques that model this probability [1], [2], but ignore its coupling to each UAV’s mission can result in very poor performance of the entire fleet. Ref. [3] introduced an approach that includes the value of the waypoints that each UAV must visit, the value of safely returning each UAV to its base, and the probability of these events given the risk in the environment. In order to maximize the mission value as an expectation, this stochastic formulation designs coordinated plans that optimally exploits the coupling effects of cooperation between UAVs to improve survival probabilities. This is done by discounting the target score with the risk of the path to that target. The path risks are not constant and change with time due to the removal of SAM sites by other UAVs. This allocation recovers planning strategies for air operations and provides significant improvements over prior approaches [1], [2]. However, the algorithm requires the solution of a MILP with a large number of integer variables, which is hard to solve, even for relatively small problems.

This paper presents a similar cooperative idea using a dynamic programming framework to solve the Weapon Task Assignment problem in adversarial environments. The full version of this problem is also difficult to solve in general, but we introduce two approximation algorithms that significantly reduce the computational complexity and still result in cooperative plans.

II. COOPERATIVE WEAPON TARGET ASSIGNMENT

The UAV task assignment problem is closely related to the Weapon Target Assignment (WTA) problem, which is a well-known problem that has been addressed in the literature for several decades [1], [5], [6], [7], [8]. The problem consists of N_w weapons and N_t targets, and the goal is to assign the weapons to the targets in order to optimize the objective, which is typically the expected accumulated value of the mission. Each target i has a value (score) of s_i and, if it is targeted by weapon j , then there is a probability p_{ij} that the target will be destroyed. Therefore the expected value of assigning weapon j to target i will be $p_{ij}s_i$. Several extensions of the general problem have been addressed and solved using different methodologies. This section looks at the WTA problem from a different

perspective, but the main idea is similar to the case of *cooperation* discussed for the UAV task assignment.

The problem addressed is that of weapon target assignment in a risky environment. Two formulations will be presented. The first is simple to solve, but the objective function ignores the effect that the tasks performed by some of the weapons can have on the risk/performance of the other weapons. The resulting targeting process is shown to be *coordinated*, but because it ignores this interaction, it is *non-cooperative*. The second formulation accounts for this interaction and solves for the *optimal cooperative strategy* using Dynamic Programming (DP). Two approximation methods are also discussed later as an alternative approach to solve these problems and achieve an answer that is close to optimal in a reasonable computation time.

Consider the WTA problem where the targets are located in a risky environment and a weapon can get shot down while flying over these targets. (Some of these targets represent SAM sites that can shoot down UAVs or weapons). Targets have different values that get discounted with time, meaning that if the target is hit now its value is higher than if it is hit in the future. Including this time discount is particularly important for environments with targets that pop-up and then disappear/move. Since the weapons are at risk of being shot down, there is a limited probability of success for each weapon aiming at the target; this will be a function of the *risk* associated with the regions it must fly over.

The problem is to assign weapons to targets in different stages (time steps) in order to maximize the expected accumulated value. Note that “time” and “stage” are used interchangeably in this formulation. The expected value for target i , with value of s_i at time t , is $p_i(t)\lambda_i^t s_i$, where ($\lambda_i \leq 1$) is the time discount factor. $p_i(t)$ represents the probability of success in destroying target i at time t and is a function of the existing SAM sites at time t . The problem then can be formulated as

$$\max_{x_{it}} \sum_{t=1}^N \sum_{i=1}^{N_t} p_i(t) \lambda_i^t s_i x_{it} \quad (1)$$

$$\text{s.t.} \quad \sum_{t=1}^N x_{it} \leq 1 \quad \forall i \in \{1 \dots N_t\} \quad (2)$$

$$\sum_{t=1}^N \sum_{i=1}^{N_t} x_{it} \leq N_w \quad (3)$$

$$x_{it} \in \{0, 1\} \quad \forall i \in \{1 \dots N_t\}, \forall t \in \{1 \dots N\}$$

where decision variable, x_{it} equals 1 if target i is assigned to be hit at stage t . The total number of stages (time horizon) is N . The first constraint ensures that each target is assigned at most once, and the second constraint limits the number of assigned targets to the number of available weapons.

With the time discount it is typically desirable to hit the targets as soon as possible (*i.e.*, in the first stage). However, since the risk in the environment will be reduced in later stages as SAM sites are removed, the probability of success,

$p_i(t)$, will increase with time, which increases the expected score. Therefore there is a trade-off between time and risk that must be captured in the optimization problem.

A. Non-cooperative Formulation

The first formulation is defined as an assignment in which the effect of weapons on the performance of other weapons is ignored. In this case the probability of success $p_i(t)$ is not a function of time and the objective function in Eq. 1 can be rewritten as

$$\max_{x_{it}} \sum_{t=1}^N \sum_{i=1}^{N_t} p_i \lambda_i^t s_i x_{it} \quad (4)$$

Since the survival probabilities are constant in this formulation, the time discount $\lambda_i < 1$ forces the targets to be assigned in the first stage. As a result, the optimization simplifies to a sorting problem in which the targets are sorted based on their expected value $p_i s_i$ and N_w targets that have the largest expected values get assigned.

B. Cooperative Formulation

This section presents a more cooperative weapon target assignment approach that can be solved as a dynamic program. To proceed, define the *state* of the system at each time step (stage) t to be the list of remaining targets, r_t , and the number of remaining weapons, m_t . Several assumptions have been made to simplify the notation: the weapons are assumed to be similar; the time discount factor λ_i is assumed to be equal for all targets; and the risk associated with the SAM sites are assumed to be equal. However, the same algorithm can be used and a similar discussion holds for the more general case.

At any stage t , the decision (control), u_t is defined to be the list of targets to be hit at that stage. Bellman’s equations for this problem can be written as

$$J_t^*(r_t, m_t) = \max_{u_t, |u_t| \leq m_t} \{S(u_t) + \lambda J_{t+1}^*(r_t - u_t, m_t - |u_t|)\} \quad (5)$$

$$t \in \{0, \dots, N-1\}$$

$$J_N^*(r_N, m_N) = 0$$

$$\text{where} \quad S(u_t) = \sum_{s_i \in u_t} p_i(t) s_i \quad (6)$$

and $|u_t|$ is the size of u_t (*i.e.*, the number of targets assigned at stage t). $p_i(t)$ represents the survival probability associated with the path that the weapon takes to the target. Note that it can be an arbitrary function of this path (*e.g.*, simply proportional to the time that weapon is inside each SAM range) or it can also be a function of the distance from the center of SAMs.

Solving the DP in Eq. 5 for r_0 equal to the list of all the targets, and m_0 equal to the number of available weapons, gives a sequence of optimal u_t^* that defines which targets to hit at each stage. $J_0^*(r_0, m_0)$ is the optimal expected score. Note that the horizon in the above DP problem, N , is finite and is less than the number of targets ($N \leq N_t$). It is trivial

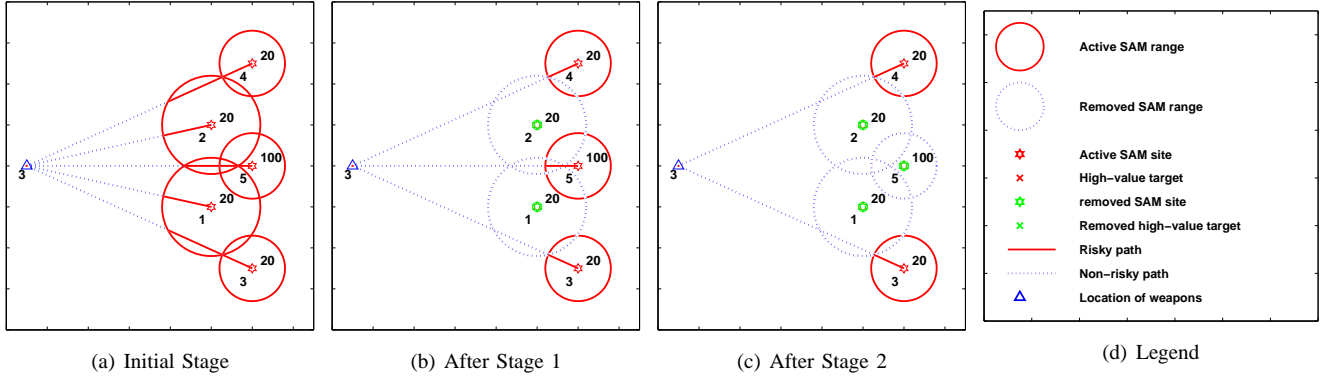


Fig. 1. The solution to the cooperative weapon target assignment for a problem of 3 weapons and 5 targets in a risky environment.

to show that in any optimal assignment all the targets are targeted before stage N . In this work, $N = N_t$. Because $p_i(t)$ is a function of time, the benefit of removing SAM sites in reducing the risk for other weapons is captured in the formulation. The DP solution will thus provide the optimal balance between risk and time. Furthermore, since weapons will be assigned to targets specifically to reduce the risk for other UAVs, the solutions will be both coordinated and cooperative.

C. A Simple Example

The first example is used to show the effectiveness of the cooperative assignment. The problem in Figure 1 has 5 targets (all SAM sites) with different values and ranges (the circles around the SAM sites show their range – the score of each target is shown next to it). The dotted lines show the trajectory (assumed to be straight lines between weapon and target) to each target from the weapon site, and the solid portion corresponds to the risky part of the path that passes over the SAMs. The position of the weapons is shown by \triangle which is labeled with the total number of available weapons at the beginning of the mission ($N_w = 3$ in this example).

To calculate the survival probability of flying over SAM site j for d_j units of distance, the following equation is used

$$\tilde{p}_j = p^{d_j} \quad (7)$$

where $0 \leq p \leq 1$ and $1 - p$ is the probability of getting shot down for flying over the SAM site for 1 unit of distance. Eq. 7 is based on the assumption that the survival probability decreases with the distance (time) traveled over the hostile environments. The overall survival probability for a weapon flying over a set J of SAM sites to target i

TABLE I: Comparison of the cooperative and non-cooperative assignment for different values of λ and p_s .

p_s	Cooperative solution				Non-coop. solution
	$\lambda = 0.6$	$\lambda = 0.8$	$\lambda = 0.9$	$\lambda = 1.0$	
0.80	12.8	16.2	17.8	19.5	2.8
0.90	37.1	45.7	50.0	54.3	13.9
0.95	61.4	74.7	81.3	88.0	38.6
0.98	82.4	99.4	108.0	116.5	81.2

can be calculated as

$$p_i = \prod_{j \in J} \tilde{p}_j \quad (8)$$

The survival probability, p is set to 0.95 and the time discount coefficient, λ , is set to 0.9 for this example. Figure 1 shows the optimal DP solution to this problem. Figure 1(a) is the initial state of the environment. In stage 1 (Figure 1(b)), SAM sites 1 and 2 are removed, reducing the risk along the path to SAM site 5, which has a much higher value (e.g., a command post). The dotted circles show that the SAM site has been removed and there is no risk associated with flying over that region. In stage 2 (Figure 1(c)), the last weapon is assigned to the high value target through a low-risk path.

To analyze the advantages of cooperation in this formulation, the expected value of this assignment is compared to the first formulation in Eq. 4. This approach assigns the three highest value targets in a single stage. The expected value for the two assignments for different values of time discount factor, λ , and survival probability, p , are shown in Table I. For a fixed value of λ , as the survival probability decreases, the difference between the expected value of cooperative and non-cooperative assignments increases. This shows that cooperation is crucial in high risk environments. For a fixed p , as the value of λ decreases the difference between the two assignments decreases, showing that when time is very important in the mission, planning in stages is not as attractive. Figure 2 shows the same results for a range of p and λ values. The y -axis is the difference between the performance of cooperative and not-cooperative assignments. The plot shows that the advantage of cooperative assignment versus non-cooperative assignment increases as the survival probability, p , decreases or time discount factor, λ increases. Note that an increase in λ is equivalent to decrease in the importance of time.

D. Larger Simulations

This section presents a larger problem ($N_t = N_w = 10$) to more clearly show the cooperation achieved by this formulation. The survival probability is set at $p = 0.9$ and time discounted factor, $\lambda = 0.9$. Figure 3 shows the result of

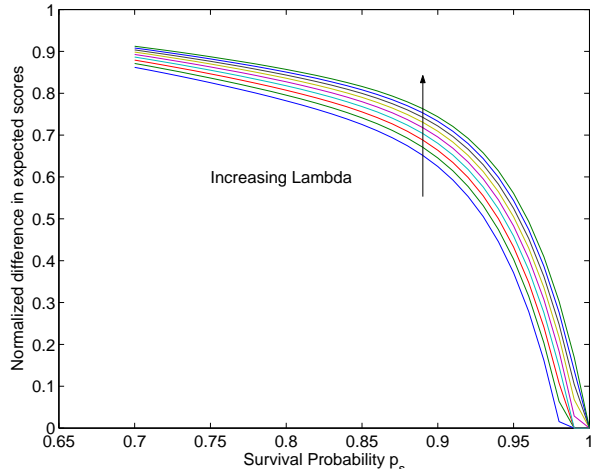


Fig. 2. Effect of survival probability p_s and time discount factor λ

the optimal cooperative assignment using the DP algorithm. Figure 3(a) illustrates the initial stage of the environment. In this example, targets 4 and 7 are high value targets and the rest are SAM sites with different ranges and values. Figure 3(b) shows the environment after stage 1. At this stage, all of the SAM sites that add risk to the paths to the high value targets are removed. Note that SAM sites 9 and 10, which are not threatening any paths, are also removed because postponing the assignment of these targets will just reduce their expected value. In stage 2 (Figure 3(c)) the remaining weapons are assigned to the remaining targets to complete the mission.

To show the effect of the discount factor in the results, the same problem is solved for $\lambda = 0.97$. The optimal answer in this case assigns weapons to targets in 4 stages (Figure 4). Since the time discount is very close to 1, the effect of time on the values of targets is very small and therefore the algorithm assigns the weapons to targets in order to maximize their expected value $p_i(t)s_i$. This situation forces the weapons to be assigned to targets sequentially. In the first stage (Figure 4(a)), SAM sites 6, 8, 9, and 10 that are on the way to the rest of the SAM sites are removed. In stage 2 (Figure 4(b)), SAM sites 2, 3, and 5, whose paths were cleared in the previous stage, are assigned to be hit. Figure 4(c) shows the 3rd stage where high value target 7, which now has a no-risk path, is removed. SAM site 1 is also removed in this stage to clear the path to high-value target 4. These two examples clearly show cooperation in the assignment, in which the objective of the assignment is not only to achieve value for each weapon, but also to increase the probability of success for other weapons. This cooperative approach results in an assignment with a much higher overall expected value.

III. APPROXIMATE DYNAMIC PROGRAMMING

The DP algorithm generates an optimal cooperative weapon target assignment in a risky environment, but as the dimension of the problem (number of targets N_t) grows, the computation time grows exponentially for this approach.

This section presents two approximation methods developed to resolve the computation issues for larger problems.

A. One-step Lookahead

In order to reduce the computation required by DP, an effective way is to reduce the horizon at each stage based on the lookahead of a small number of stages [9]. This idea is very similar to the receding horizon task assignment [10] in which the planning horizon is limited to reduce the computation. The simplest possibility is to use a one-step lookahead where at stage t and state r_t the control u_t maximizes the expression

$$\max_{u_t, |u_t| \leq m_t} \{S(u_t) + \lambda \bar{J}_{t+1}(r_t - u_t, m_t - |u_t|)\} \quad (9)$$

$$t \in \{0, \dots, N-1\}$$

where \bar{J}_{t+1} is an approximation of the true cost-to-go function, J_{t+1}^* , with $\bar{J}_N = 0$. In the one-step lookahead, having the approximation \bar{J} , the calculation reduces to one maximization per stage, which is a significant savings compared to an exact DP. The performance of the one-step lookahead policy depends on how well \bar{J} approximates the true cost-to-go. A good cost-to-go can be calculated using complex algorithms and results in a close to optimal answer, but the computation complexity associated with calculating the cost-to-go itself might defeat the purpose. Therefore, while a good approximate cost-to-go is desirable, the calculation must be simple. A simple approximation of cost-to-go for the problem of the weapon target assignment is introduced in the following that can be calculated very quickly.

At stage t and state (r_t, m_t) , $\bar{J}_t(r_t, m_t)$ is the solution to the non-cooperative formulation in Eq. 4.

$$\bar{J}_t(r_t, m_t) = \max_{x_{it}} \sum_{t=1}^N \sum_{i=1}^{N_t} p_i \lambda_i^t s_i x_{it} \quad (10)$$

$$\text{s.t.} \quad \sum_{t=1}^N x_{it} \leq 1, \quad \forall i \in r_t \quad (11)$$

$$\sum_{i=1}^{N_t} \sum_{t=1}^N x_{it} \leq N_w \quad (12)$$

This cost-to-go approximation assumes that all the remaining weapons are assigned to targets in the next stage. This is a simple approximation cost-to-go that can be calculated very easily and as a result, the computation time required to generate the assignment is much lower than the exact DP algorithm. To compare the result of the one-step lookahead approximation with the optimal solution from the exact DP algorithm, the problem of 10 weapons and 10 targets discussed in Section II-D is used. The results of the approximation method for $\lambda = 0.9$ are shown in Figure 5 and are compared to the optimal result in Table II. In the optimal solution, the mission is accomplished in 2 stages while in the one-step lookahead solution it is accomplished in 4 stages. This assignment

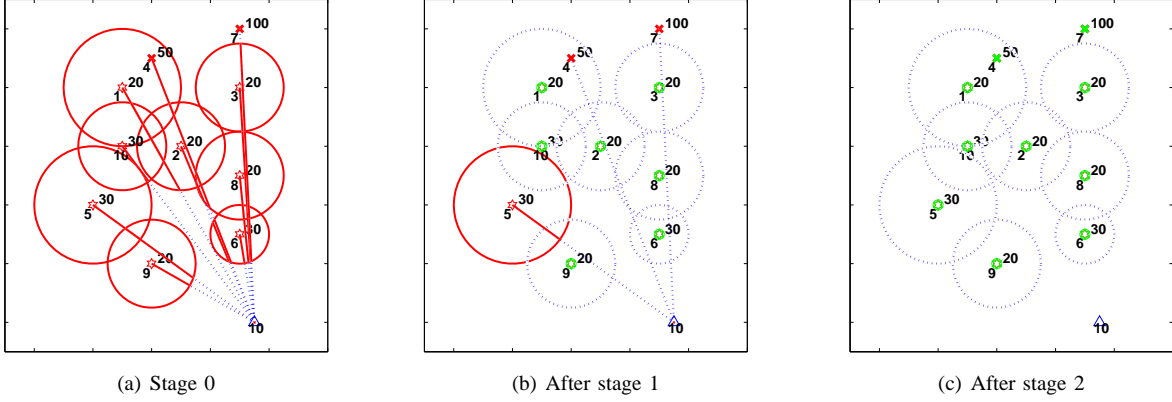


Fig. 3. Optimal solution for problem with 10 weapons and 10 targets ($p_s = 0.9$, $\lambda = 0.9$). Mission implemented in 2 stages and expected value 160.

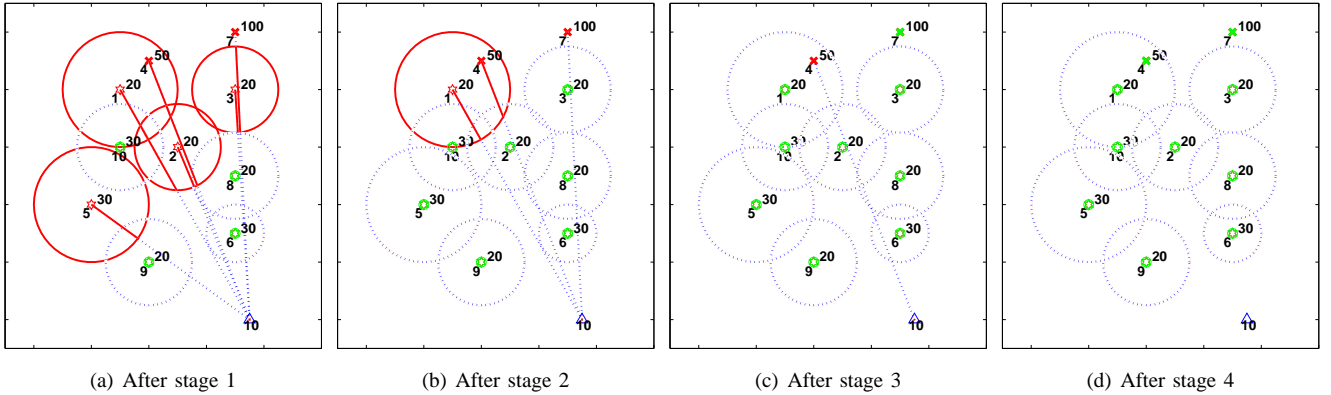


Fig. 4. Optimal solution for problem similar to Fig. 3 ($p_s = 0.9$, $\lambda = 0.97$). Mission implemented in 4 stages and expected value 174.

has resulted in lower performance compared to the optimal solution, but the computation time is considerably reduced. In the next section, the two-step lookahead algorithm will be discussed to increase the performance compared to one-step lookahead.

B. Two-step Lookahead

In the two-step lookahead policy (at stage t and state (r_t, m_t)), we use the control u_t , which attains the maximum in the expression

$$\max_{u_t, |u_t| \leq m_t} \{S(u_t) + \lambda \tilde{J}_{t+1}(r_t - u_t, m_t - |u_t|)\} \quad t \in \{0, \dots, N-1\} \quad (13)$$

where \tilde{J}_{t+1} is obtained from the one-step lookahead approximation

$$\begin{aligned} \tilde{J}_{t+1}(r_{t+1}, m_{t+1}) = & \max_{u_{t+1}, |u_{t+1}| \leq m_{t+1}} \{S(u_{t+1}) \\ & + \lambda \tilde{J}_{t+2}(r_{t+1} - u_{t+1}, m_{t+1} - |u_{t+1}|)\} \end{aligned} \quad (14)$$

and \tilde{J}_{t+2} is an approximation of the true cost-to-go function J_{t+2}^* . The approximation discussed in Eq. 12 for the one-step lookahead is also used for the two-step lookahead policy. This method was compared with the one-step lookahead and exact DP solutions for the problem of 10 targets and 10 weapons. The result is identical to the optimal solution shown in Figure 3.

Table II presents further comparisons of the various methods. Computation time is substantially reduced compared

to the exact DP case, but as expected, is higher than the one-step lookahead. On the other hand, the performance increases from the solution of the one-step lookahead, and in this case is identical to the optimal solution. To see if these results hold for other cases, the three algorithms (exact DP, one-step and two-step lookahead) were used to solve many randomly generated scenarios. In any set of these scenarios, the number of targets, N_t , number of weapons, N_w , time discount factor, λ , and survival probability, p , are kept constant and the position and value of targets and the range of SAM sites are randomly generated. Figure 6 illustrates the results of these simulations for three values of λ . The horizontal axis in this graph shows the degree of sub-optimality (as a percentage) defined as

$$100 \times \frac{J_{\text{optimal}} - J_{\text{approximation}}}{J_{\text{optimal}}} \quad (15)$$

The vertical axis shows the cumulative percentage of the cases that are within the interval of sub-optimality indicated on the horizontal axis. Note that it is desirable for the graph to be close the (0, 100). These results clearly demonstrate that two-step lookahead policy outperforms the one-step lookahead policy, and that the performance of the two-step lookahead is close to the optimal performance.

IV. CONCLUSIONS

This paper discussed the problem of risk in the environment and a stochastic formulation of UAV task assignment

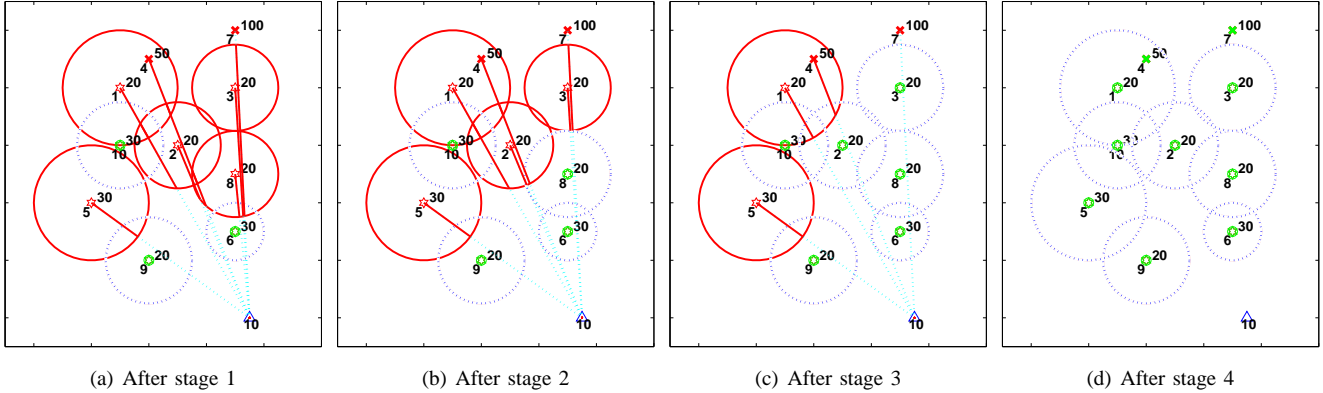


Fig. 5. One-step lookahead solution to problem similar to Fig. 3 ($p_s = 0.9$, $\lambda = 0.9$). Mission implemented in 4 stages and expected value 133.

TABLE II: Comparing the result of the non-cooperative, DP, one-step lookahead and two-step lookahead solutions for the problem of 10 weapons and 10 targets.

Algorithm	Expected Accum. Value	Comp. time (sec)	Num. of stages
One-step lookahead	133.7	0.4	6
Two-step lookahead	160.0	13.4	2
DP	160.0	56.2	2
Non-cooperative	21.9	0.1	1

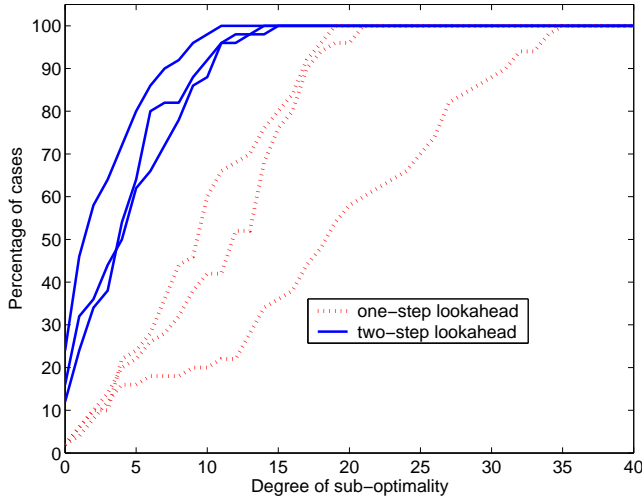


Fig. 6. Comparison of the performance of the one- and two-step lookahead policies with the optimal solution for different values of λ . Degree of sub-optimality is defined in Eq. 15.

problem was presented. This formulation explicitly accounts for the interaction between the UAVs – displaying *cooperation* between the vehicles rather than just *coordination*. Cooperation entails coordinated task assignment with the additional knowledge of the future implications of a UAV’s actions on improving the expected performance of the other UAVs. The key point is that the actions of one UAV can reduce the risk in the environment for the other UAVs; and the new formulation takes advantage of this fact to generate cooperative assignments to achieve better performance. The problem was formulated as a Dynamic Programming (DP)

problem. A comparison with other approaches showed that including this cooperation lead to a significant increase in performance. Two DP approximation methods (the one-step and two-step lookahead) were also developed for large problems where curse of dimensionality in DP is prohibitive. Simulation results showed that the one-step lookahead can generate a cooperative solution very quickly, but the performance degrades considerably. The two-step lookahead policy generated plans which are very close to (and in many cases, identical to) the optimal solution.

ACKNOWLEDGMENT

This research was funded under AFOSR grant # FA9550-04-1-0458.

REFERENCES

- [1] R. A. Murphey, “An approximate algorithm for a weapon target assignment stochastic program,” In *Approximation and Complexity in Numerical Optimization: Continuous and Discrete Problems*. Kluwer Academic Publishers, 1999.
- [2] J. L. Ryan, T. G. Bailey, and J. T. Moore, “Reactive tabu search in unmanned aerial reconnaissance simulations,” In D. J. Medeiros et al., editor, *Proceedings of the 1998 Winter Simulation Conference*, 1998, e”, IEEE, May 2002.
- [3] J. S. Bellingham, M. Tillerson, M. Alighanbari, and J. P. How, “Cooperative Path Planning for Multiple UAVs in Dynamic and Uncertain Environment,” In *Proceeding of IEEE Conference on Decision and Control*, Dec. 2002.
- [4] J. S. Bellingham, *Coordination and Control of UAV Fleets using Mixed-Integer Linear Programming*, SM Thesis, MIT Department of Aeronautics and Astronautics, Aug. 2002.
- [5] P. Hosen, and M. Athans, “The Dynamic Weapon Target Assignment Problem.” *Proc. of Symposium on C² Research*, Washington, D.C. 1989.
- [6] R. A. Murphey, “Target-Based Weapon Target Assignment Problems,” in *Nonlinear Assignment Problems: Algorithms and Applications*, P. M. Pardalos and L. S. Pitsoulis eds., Kluwer Academic Publishers, 2000.
- [7] R. K. Ahuja, A. Kumar, K. C. Jha, and J. B. Orlin. “Exact and heuristic algorithms for the weapon-target assignment problem”. Submitted to *Operations Research*, 2003.
- [8] P. Hosen, and M. Athans, “An Asymptotic Result for the Multi-Stage Weapon-Target Allocation Problem,” In *Proceeding of IEEE Conference on Decision and Control*, Dec. 1990.
- [9] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, Athena Scientific, Belmont, Massachusetts, 2000.
- [10] M. Alighanbari, *Task Assignment Algorithms for Teams of UAVs in Dynamic Environments*, SM Thesis, MIT Department of Aeronautics and Astronautics, May 2004.