

Title: **Collaborative Optimal Control Algorithms for Low-Cost Autonomous Weapons**  
Visiting Researcher: **Kyle Volle**  
Academic Advisor: Jonathan Rogers, Assistant Professor  
Robotics, Georgia Institute of Technology  
Proposed Tenure: **May 12, 2014 – August 15, 2014**

## 1 Objective

The problem under consideration is to develop a framework of algorithms that enables multiple target engagement with a swarm of low-cost, heterogeneous autonomous weapons systems. This problem falls under the scope of Munitions Aero GN&C (Navigation, Guidance & Avionics, Control, and Munitions Aerodynamics Sciences). Given a set of targets with known sizes and respective priorities, these distributed weapons would form coalitions to cooperatively engage these targets. The Air Force frequently faces asymmetrical conflict scenarios in which expensive, highly precise weapons are used against low-cost targets that may be easily replaced or are even considered expendable. The fundamental mismatch between the cost to destroy a target and the cost to replace it creates an environment of diminishing returns in terms of Air Force resources and investment.

Specific issues to be considered include distributed decision making with limited computational resources, multi-agent cooperation with restricted communication ability, mobile targets, weapon attrition and the balancing of offline pre-planning with online re-planning to better adapt to changing circumstances.

The problem is a dynamic, stochastic, resource allocation problem. It can be formulated in any number of ways, but the two that appear the most promising at this stage are as a partially-observable Markov decision process (POMDP) and as a distributed constraint optimization problem (DisCOP). These frameworks are described below and the proposed methods to solve them are detailed in the next section.

In the POMDP formulation, an agent's state is a tuple with the first term being the target currently assigned to it and the rest of the terms being the number of agents assigned to each target. POMDPs also require a transition function that describes the results of actions. Since the actions are not entirely deterministic and communication ranges are limited the world state is only partially observable, otherwise the problem could be reduced to just a Markov decision process. However belief states can be created based on most likely results of actions and what sensor information is available and assign probability distributions to these belief states. Finally, POMDPs require a reward function that is a function of the state and corresponds to how desirable being in that state is. Solving a POMDP returns an optimal policy which gives the best action to take in a given belief state.

The distributed constraint optimization framework is largely the same as described above but stated in a canonical form that allows established algorithms for this problem class to be applied. DisCOPs are defined by a tuple  $(A, V, \mathcal{D}, f, \eta)$ . Traditionally, DisCOPs have an additional term,  $\alpha$ , that maps variables to agents but for this cas, it is a trivial one-to-one mapping.

- A The set of agents. Dimension of set is  $n$ .
- V The set of variables.  $\{v_1, v_2, \dots, v_n\}$
- $\mathcal{D}$  The set of Domains.  $\{D_1, D_2, \dots, D_n\}$
- f A function that defines the constraints between variables.
- $\eta$  An overall cost function

In the DisCOP framework, each agent sets the value of one and only one variable and there are costs associated with each violated constraint. Constraint optimization problems have two major advantages over constraint satisfaction problems (CSPs). The first is that a CSP solver returns the first solution it finds even though better solutions might exist but a DisCOP solver returns the best solution. Secondly, if no solution exists a CSP solver will return a failure and a DisCOP solver will return the best approximate solution. In this case, the agents are the individual weapons, the variables are the targets assigned to each agent. Each variable has a domain which is the set of possible values that the variable can be set to. In this context the domain for a variable consists of all the targets visible to the agent associated with that variable. The constraints are simply expressed as a limit on the number of agents that can have the same values for their associated variables.

The methods proposed to solve the problem in either framework are described below. Objectives for the summer term are to develop robust algorithms for the agents to execute autonomously and to use simulations to evaluate and validate these algorithms. Findings will be documented in conference and journal papers.

## 2 General approach

There are two techniques to solving POMDPs. The first, value iteration, is more time intensive but gives a very precise measure of the expected utility of a state. The second method, policy iteration, converges faster in general but isn't as discriminatory if there isn't a fairly large difference in the utilities of some states.

Value iteration uses an equation called the Bellman Equation to find what states (or belief states) have the greatest expected value and generates a policy that steers the agent towards those states. Policy iteration requires an initial policy and the policy is updated on every cycle until it converges and doesn't change any more. Policy iteration often converges faster because the policy often remains constant while the value function is still settling to its exact values. Ultimately, the policy is what is desired and the exact utility values are not required.

The downside of POMDPs is that the state space can quickly grow intractable and use up large amounts of memory and take too long for practical purposes, particularly for value iteration. In fact, if unreachable states are not culled, the state space for this problem with  $n$  weapons and  $t$  targets has  $t(n+1)^t$  states. However more thoughtful methods of enumerating states can reduce this by almost two orders of magnitude. If mission details are known in advance, the optimal policy can be calculated offline and distributed to the agents in advance. Making the assumption that the model used to generate the policy is sufficiently close to the ground truth, the policy generated will be close to the optimal policy for the actual situation. This policy can be input into the policy iteration algorithm to make minor corrections on the fly.

As mentioned above, an alternative method of solving this problem is with the use of distributed constraint optimization problems. There are several published methods of solving DisCOPs, as they remain an active area of research. Most artificial intelligence algorithms are evaluated for completeness, correctness/optimality, time complexity and space complexity. When evaluating DisCOP algorithms, a fifth criteria is added to the typical four: the number of required messages between agents. Given that the problem under consideration includes greatly reduced communication, the algorithm with the best performance in this category is a good candidate. The Distributed Pseudotree Optimization Procedure (DPOP) algorithm is linear in the number of messages required and has been proven complete and correct.

## 3 Impact of technology to USAF

In asymmetrical conflict scenarios, the use of expensive weapons are deployed against low-cost or even expendable targets results in a sub-optimal use of resources. A strategy of multiple target engagement

by coalitions of low-cost weapons systems has the potential to alleviate some of this mismatch between the cost to destroy a target and the cost to replace it.

Even in more symmetric scenarios, this type of inexpensive distributed weapons system has a role to play. These low cost munitions can be used to overwhelm air defenses at much less cost than using existing weaponry for the same purpose. The reliability of each individual engagement vehicle may suffer when compared to expensive and more precise weapons systems, however cooperative engagement strategies can maximize kill probabilities and remain robust against weapon attrition.

Finally, the autonomous nature of these proposed weapons systems significantly reduces the required operator oversight, which in turns lowers the costs and risks associated with deploying them.

## 4 Biographical sketch

### 4.1 Professional Preparation

Undergraduate Institution(s):	University of Kansas	Mechanical Engineering	BSc, 2012
	Washburn University	Physics	BSc, 2012
Graduate Institution(s):	Georgia Tech	Robotics	2012-Present

### 4.2 Appointments

Graduate Research Assistant in robotics under Dr Jonathan Rogers

### 4.3 Publications

List the publications of both the visiting researcher and academic advisor.

#### Publications most closely related to the proposed project (Limit 5)

1. None

#### Significant publications outside of proposed project (Limit 5)

1. J. Rogers, Stochastic Model Predictive Control for Guided Projectiles Under Impact Area Constraints, *Journal of Dynamic Systems, Measurement, and Control*, Under Review, 2013.
2. J. Rogers, N. Slegers, Robust Parafoil Terminal Guidance Using Massively Parallel Processing, *Journal of Guidance, Control, and Dynamics*, Vol. 36, No. 5, September-October 2013, pp. 1336-1345.
3. Z. Sunberg, J. Rogers, A Belief Function Conflict Metric for Orderable Sets, *Information Fusion*, Vol. 14, No. 4, October 2013, pp. 361-373.
4. J. Rogers, M. Costello, Smart Projectile State Estimation Using Evidence Theory, *Journal of Guidance, Control, and Dynamics*, Vol. 35, No. 3, May-June 2012.
5. J. Rogers, M. Costello, D. Hepner, A Roll Orientation Estimator for Smart Projectiles Using Thermopile Sensors, *Journal of Guidance, Control, and Dynamics*, Vol. 34, No. 4, July-August 2011, pp. 688-697.