Proceedings of the 1994 Winter Simulation Conference
ed. J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila

SIMULATION-BASED DYNAMIC OPTIMIZATION: PLANNING UNITED STATES COAST GUARD LAW ENFORCEMENT PATROLS

Michael P. Bailey
and
Robert F. Dell
Operations Research Department
Naval Postgraduate School
Monterey, California 93943-5000

Kevin D. Glazebrook
Department of Mathematics and Statistics
University of Newcastle Upon Tyne
Newcastle NEI7RO UK

Abstract

A primary mission for the United States Coast Guard (USCG) operating in coastal United States waters is to interdict contraband. The USCG schedules a fleet of cutters to meet this mission and seeks a way to determine the operational efficiency of a particular schedule. This paper develops a methodology based on generating a sequence of finite horizon dynamic programs (DPs), where each DP differs only in the way the smuggling vessels and the cutters interact. The DP takes the point of view of the smuggler who wishes to develop the smuggling strategy which maximizes some characteristic (*e.g.*, the mean) of the profit attained. The DP explicitly accounts for a smuggler who must combine his short-run profit goals with his need to gain future information about the configuration of the cutters. We develop a Monte Carlo sampling procedure to generate estimates of the random variables used in the DP.

1 INTRODUCTION

The United States Coast Guard (USCG) patrols coastal waters under the United States' jurisdiction. In recent years, this mission has increasingly called on the USCG to interdict contraband in the form of illicit drugs and refugees. Accompanied by the Department of Defense, the USCG develops plans for locating cutter patrols in an attempt to interdict smuggling. Each time period, cutters may move from patrol to patrol.

Cutter schedules are typically not adaptive, schedulers plan cutter use more than a quarter-year in advance, striving to meet steaming and patrolling goals. Through satisfaction of these goals the USCG attempts to prevent, disrupt, and punish acts of smuggling. An optimization model guided by a patrolling goal for each tuple of (patrol, day, cutter type) in its objective function can produce a schedule. The USCG seeks a way to determine the operational efficiency of a particular schedule.

This paper develops a methodology for doing just this. The method involves generating a sequence of finite horizon dynamic programs (DPs), where each DP differs only in the way the smuggling vessels and the cutters interact. The DP takes the point of view of the smuggler, and the objective is to develop the smuggling strategy which maximizes some characteristic (*e.g.*, the mean) of the profit attained. In the future, we will use a higher fidelity model of cutter actions, LESIM (1994), to generate seizure data. Our goal in the present work is to provide background and structure for the integration of the DP with a simple stochastic seizure model, with an eye toward enhancement in the sequel.

Figure 1 summarizes the approach taken in this paper. The DP formulation captures system characteristics we call *Realities* below. A state of the DP is the relative likelihood that the cutters are configured in certain patrols. The action available from any state is the single time period smuggling strategy pursued. The resulting outcome of a strategy from a state is the new likelihood of a cutter configuration. This likelihood is computed using the information about, seizures, successful smuggling attempts, and a cutter fleet motion model. We use a diverse set of possible multi-period strategies to populate our DP network, where both the single time period profit and the new state depend on the random seizure outcomes.

Upon the successful completion of many iterations of each trial strategy for a time horizon (T), we can produce empirical distributions for the value of being in a particular state and taking a particular action. We can enforce constraints on the stochastic properties of the evolution by removing action/state pairs which violate constraints. For example, we can eliminate actions which cause half of the total contraband shipped to be confiscated with probability 0.3 or more. Finally, we can test any T-period strate-

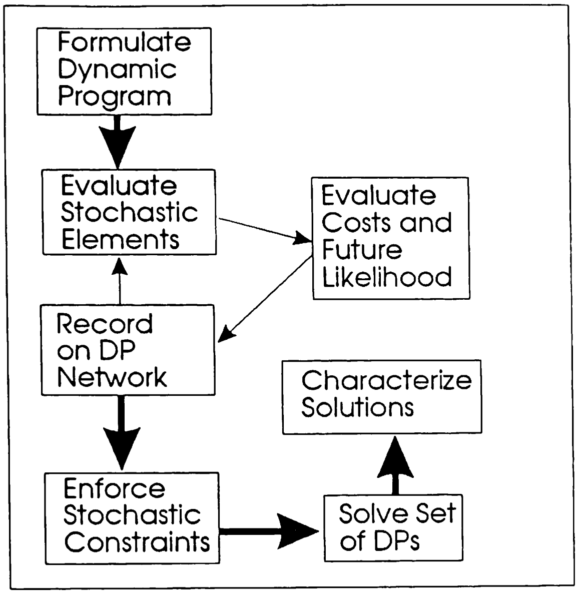

Figure 1: Hybrid Simulation Analytic Model Diagram reflecting the Use of Monte Carlo Samples in a Stochastic Dynamic Program

gy composed of segments of the trial strategies, we can chain backward to maximize expected summed profit, or we can search for strategies which produce desirable quantiles of the value distribution.

Section 2 discusses literature and motivates our development. Section 3 describes modeling the smuggler's problem as a DP. Section 4 provides a description of Monte Carlo sampling to build a DP network. Section 5 contains conclusions and the Appendix has a summary of notation.

2 APPROACHES

The pristine smuggler's problem is famous in the literatures of stochastic analysis, sequential decision making and game theory (Owen, 1982). Here are some recent treatments of the smuggler-interdictor problem:

2.1 Stationary Shipping

Many models, including the Law Enforcement Simulation (LESIM, 1994) use a simple filtered Poisson process to model the traffic attempted along each route. This model assumes that the smuggler ignores the presence of the interdictor or any information he might gain by succeeding or failing an attempted shipment. LESIM combines this simple shipping model with a high-fidelity model of the actions taken by a cutter during patrol, detection, boarding, and seizure.

2.2 Gaming

Clearly, smuggling is a game of competing strategies. Examining this approach, Wood and Washburn (1994) explore game theoretic methods to calculate strategies which give both sides maximum expected equilibrium benefit. Unfortunately for the Coast Guard, the opponents are not equally facile in adapting as the game is played.

2.3 SOAR

The Simulation of Adaptive Response (SOAR) model (Caulkins *et al*, 1993) models the dynamic fluctuations of shipping prices over time, where the smuggler calculates the route with the minimum expected cost at each time period. The perception of cost is based on shipping expenses and profits, and on perceived probability of seizure. The smuggler's memory of a captured shipment on a route, r, time units in the past fades like $e^{-\beta r}$. The smuggler's actual shipping is again a filtered Poisson process, with the splitting probabilities proportional to the perceived costs on each route. SOAR uses static interdiction probabilities on each route, and allows the user to manipulate these to experiment with interdictor schedules.

2.4 The Realities

- The USCG schedules cutters on a quarterly or yearly basis and does not deviate from this schedule except in times of crisis. Hence the USCG typically does not react to sudden changes in traffic by reallocating cutters.

- The Coast Guard schedules cutters to enter and leave patrol areas by designating an area and duration of stay. This duration is generally between one and six weeks, and more than one cutter can be assigned to an area simultaneously.

- The smuggler strives to provide a consistent supply of contraband. For our purposes we assume that the smuggler attempts to deliver a constant amount of contraband in each time period.

- The smuggler doesn't know the location of the cutters with any certainty, but does get feedback in the form of confiscated shipments. Hence the smuggler is doing two things at once:

  1. shipping contraband to accumulate immediate profit;
  2. collecting information about the location of the cutters so that future decisions are made better.

3 WHAT THE SMUGGLER KNOWS, AND WHAT HE ESTIMATES

We assume a single smuggler who knows many of the cutter scheduling constraints. Here we outline what the smuggler knows, and what information he estimates as operations evolve.

3.1 Smuggler's Knowledge

We assume that the smuggler has access to the following:

1. the number of ships in the cutter fleet and the locations of the R patrol routes the cutters may occupy;
2. the number of shipments attempted ($s(t) = (s_1(t), s_2(t), \ldots, s_R(t))$ and confiscated ($n(t) = n_1(t), n_2(t), \ldots, n_R(t)$) at time t;
3. the maximum and minimum number of cutters allowed to patrol the same route; and
4. the assumed probability that a cutter remains at its current station for the next time period.

3.2 Information Estimated

At any time, the smuggler estimates the state of the cutter fleet expressed as one of the possible configurations of the cutters on the R routes. Denote the number of cutters on route r in configuration c as $d_{c,r}$, $c = 1, 2, \ldots, C$, where $\sum_{r=0}^{R} d_{c,r}$ equals the cutter fleet size and $d_{c,\cdot}$ is a specific configuration vector. We use route 0 as the location of cutters which are not on patrol.

3.2.1 Cutter Fleet Motion Model

From the smuggler's point of view, the configuration of the cutters on the routes evolves as a discrete time Markov chain. Given the perceived probability of a cutter changing patrol route for the next time period, we can calculate the probability $P_{c,c'}$ of transition from configuration $d_{c,\cdot}$ to $d_{c',\cdot}$ via complex counting arguments. This matrix estimates the likelihood that the cutters are in a given sequence of configurations. This model is tempered by the outcomes of smuggling operations to produce the likelihood that the cutters are arranged in a particular configuration in the next time period.

3.2.2 Seizures

Given the cutter configuration is $d_{c,\cdot}$ at time t and an assumption that the cutters detect smugglers independently, the (random) number of seizures $N_r(t)$ on route r is distributed as binomial random variable

$$P[n_r(t)|d_{c,r}] = P[N_r(t) = n_r(t)|d_{c,r}] = \binom{s_r(t)}{n_r(t)} (1-p_r)^{d_{c,r}(s_r(t)-n_r(t))}(1-(1-p_r)^{d_{c,r}})^{n_r(t)}; \quad (1)$$

where p_r is the probability of detection of one smuggling vessel by one cutter on route r.

3.2.3 Likelihood Updates

In the smuggler's view, the configuration of cutters on the routes at time t is a random variable $D(t)$. Let $\phi_c(t)$ be the smuggler's perceived likelihood that the cutter fleet is in configuration c at time t. Based on the realized seizures $n(t) = (n_1(t), n_2(t), \ldots, n_R(t))$, the prior density $\phi(t) = (\phi_1(t), \phi_2(t), \ldots, \phi_C(t))$, and the cutter motion model $P_{c,c'}$, the smuggler can calculate the likelihood that $d_{c',\cdot}$ is the next configuration he faces:

$$\hat{\phi}_{c'}(t+1) = P[D(t+1) = d_{c',\cdot}] = \frac{\mathcal{L}[d_{c',\cdot}|n(t)]}{\sum_{c=1}^{C} \mathcal{L}[d_{c,\cdot}|n(t)]} \quad (2)$$

$$\begin{aligned}\mathcal{L}[d_{c',\cdot}|n(t)] &= \sum_{c=1}^{C} P[d_{c',\cdot}|d_{c,\cdot}]P[n(t)|d_{c,\cdot}]P[d_{c,\cdot}] \quad (3)\\ &= \sum_{c=1}^{C} P_{c,c'}[\prod_{r=1}^{R} P[n_r(t)|d_{c,\cdot}]]\phi_c(t); \quad (4)\end{aligned}$$

where $\mathcal{L}$ is the likelihood function.

The smuggler's goal is to control the flow of contraband through the system to meet delivery goals by exploiting his estimate ϕ and by manipulating $s(t)$.

3.3 Dynamic Programming

The above updating process lends itself directly to sequential optimization for a finite time horizon with T as the planning horizon for smuggling operations. Let C_1 be the immediate cost of an interdicted shipment, including lost equipment, legal fees, and the shipment itself. Let C_2 be the immediate profit realized from a completed delivery. The (random) value of occupying state $\phi(t)$ at time t and using future strategy $s(t), s(t+1), \ldots, s(T)$ can be stated as

$$V(\phi(t)|s,t) = -\sum_{r=1}^{R} C_1 N_r(t) + \sum_{r=1}^{R} C_2(s_r(t) - N_r(t))$$

$$+V(\hat{\phi}(t+1)|s,t+1) \qquad (5)$$

where s,t is shorthand for $\{s(t'), t \le t' \le T\}$ and $\hat{\phi}(t+1)$ is the state resulting from the likelihood update described above in (2). Expected value maximization of profit can be accomplished directly by choosing s to maximize $V(\phi(t)|s,t)$ at each stage:

$$E[V(\phi(t)|\text{o},t)] =$$

$$max_{s(t)} - \sum_{r=1}^{R} C_1 E[N_r(t)] + \sum_{r=1}^{R} C_2(s_r(t) - E[N_r(t)])$$
$$+ E[V(\hat{\phi}(t+1)|\text{o},t+1)] \qquad (6)$$

where the o signifies the optimal strategy. Using

1. discretized set of possible likelihood configurations $\phi^1, \phi^2, \ldots, \phi^L$ which form a mesh with uniform spacing Δ on $[0,1]^R$;
2. interpolation between the ϕ^l;
3. $E[N_r(t)] = s_r(t) - s_r(t)(1-p_r)^{d_{c,r}}$ for each t;
4. $E[V(\phi^l|\text{o},T)] = 0$ for $l = 1,2,\ldots,L$;

We can solve (6) using backward recursion to produce an optimal strategy, $(s(1), s(2), \ldots, s(T))$, with maximum expected profit. In what follows, we explore the use of Monte Carlo methods for exploring the behavior of (5) when the goal is other than maximizing summed expected profit.

4 MONTE CARLO METHODS

In order to get a stochastic characterization of the possible smuggler operations, we generate a sequence of random outcomes of seizures under different smuggling strategies. We use these outcomes to populate the arc lengths on a network connecting DP states to one another. We then exploit methods for examining the behavior of networks with stochastic arc lengths to characterize the capabilities and tendencies of the smuggler.

4.1 Building the Dynamic Programming Network

Let $s^1, s^2, \ldots, s^I$ be a set of strategies for trial, $s_r^i(t)$ being the number of shipments on route r at time period t for strategy i. We will select these strategies so that they reflect methods that will likely be successful, and so that they represent a diverse set of choices. Let $\hat{\phi}(\phi^l|s(t), t+1)_k$ be the k^{th} sample of the cutter configuration likelihood $\hat{\phi}(t+1)$ produced when we attempt smuggling strategy $s(t)$ from likelihood ϕ^l at time period t. What varies between samples $\hat{\phi}(\phi^l|s(t), t+1)_k$ and $\hat{\phi}(\phi^l|s(t), t+1)_{k+1}$ results from differences in the number of seizures made.

```
FOR strategy s^i, i = 1,2,...,I
  FOR time period t, t = 1,2,...,T
    Configure cutters as d_c.
    FOR replication k, k = 1,2,...,K
      FOR route r, r = 1,2,...,R
      Sample n_r(t) from s_r^i(t)
        FOR likelihood φ^l, l = 1,2,...,L
          Calculate ô(φ^l, t|s(t))_k and
            φ̂(φ^l, t+1|s(t))_k
```

Figure 2: The Monte Carlo Sampling Procedure to Generate Samples of $\hat{\phi}$ and $\hat{o}$, Resulting Likelihoods and Immediate Costs, Respectively.

Let $\hat{o}(\phi^l|s(t),t)_k$ be the k^{th} sample of the immediate value of the strategy $s(t)$ used against likelihood ϕ^l at time t:

$$\hat{o}(\phi^l|s(t),t)_k =$$
$$-\sum_{r=1}^{R} C_1 n_r(t) + \sum_{r=1}^{R} C_2(s_r(t) - n_r(t)); \qquad (7)$$

$n_r(t)$ is the observation of seizures from the k^{th} replication.

We develop the set of $\hat{\phi}$ and $\hat{o}$ for each time period t, likelihood ϕ^l, and each candidate strategy s^i as shown in the algorithm in Figure 2.

4.2 Measuring Performance and Constructing Strategies

Thus, we can now produce an empirical distribution of the value of pursuing any strategy which is a combination of segments of the $s^1, s^2, \ldots, s^I$.

4.2.1 Optimizing Summed Profit

To collect samples of the optimal summed profit for the T time units, we select the strategy $s(t)_k^*$ at each stage which optimizes $V_k(\phi^l|\text{o},t)$ for each ϕ^l. The optimal strategy is found by chaining backward from $V_k(\phi^l|\text{o},T)$:

$$V_k(\phi^l|s,t) =$$
$$\hat{o}(\phi^l|s(t),t)_k + V_k(\hat{\phi^l}(\phi^l|s(t),t+1)_k|s(t),t+1) \qquad (8)$$

Where $V_k(\hat{\phi^l}(\phi^l|s(t),t+1)_k|s(t),t+1)$ is calculated via interpolation as follows. Let $\Phi(\phi)$ be the adjacent likelihoods to ϕ:

$$\Phi(\phi) = \{\phi^l : |\phi_r^l - \phi_r| < \Delta, r = 1,2,\ldots R\}. \qquad (9)$$

$\Phi(\phi)$ usually has 2^R elements. For each member $\phi^l \in \Phi(\phi)$, define $\alpha_{\phi^l} = \|\phi^l - \phi\|_2$, $\alpha = \sum_{\phi \in \Phi(\phi)} \alpha_\phi$.

$$V_k(\hat{\phi}^l(\phi^l|s,t+1)_k|s(t),t+1) = \sum_{\phi \in \Phi(\hat{\phi}(\phi^l|s(t),t)_k} \frac{\alpha_\phi}{\alpha} V_k(\phi|s,t). \quad (10)$$

As can be shown with computational examples, the value of $V_k(\phi(0)|o,0)$ will be much greater than the value $E[V(\phi(0)|o,t)]$ as calculated in equation (6). This is because the maximization is taken *after* the randomness has been realized, rather than before. This result mimics many which show how network optimization routines used with expected arc lengths stray from the results produced when the network optimization is done after each arc has realized it length. The obvious similarity stems from the link between dynamic programming and network shortest (or longest) paths, see Bailey (1994) for several examples.

In the smuggler problem, the distinction between optimization before or after realization relates the distance between the formulation of the smuggling policy and the feedback the smuggler receives during operations. Before realization, using (6), corresponds to the smuggler planning all of his operations strategically – setting his plan before the T time units begin. Optimization post-realization corresponds to the smuggler having a *crystal ball*, and knowing what seizures he will realize if he follows a particular strategy. As reality lies between these two extremes, the post-realization answer gives the smuggler an upper bound on his performance, while providing the USCG with a lower bound on the performance of a cutter schedule.

4.2.2 Maximizing a Quantile

Using a particular strategy s built from segments of $s^1, s^2, \ldots, s^l$, the smuggler can produce the empirical distribution of the value from the data $V_k(\phi(0)|s,0)$ directly – $V_k(\phi(0)|s,0)$ is calculated with *no* expectations taken. From this, the smuggler can compare policies on the basis of the α^{th} quantile of these empirical distributions for different values of s. Techniques to sharpen these estimates could be employed, see Heidleberger and Lewis (1984).

Using a low quantile corresponds to the smuggler being risk averse. The smuggler might want to maximize his worst-case profit, where worst case is interpreted as a profit he is $(1-\alpha)$% sure of receiving. On the other hand, he may wish to look for policies which have high large quantiles, giving himself the chance to make a possible windfall with great risk.

The search among the possible strategies for those which produce a high α^{th} quantile is clearly problematic. One heuristic would be to maximize the quantile at each stage. Such an approach has the added benefit of minimizing the USCG's ability to disrupt short-term supply.

4.2.3 Avoiding Disruption of Short-Term Supply

In most real supplier-consumer relationships, the supplier must satisfy target delivery levels and meet short-run cash flow constraints during the evolution. Also, he seeks to maximize his total profit. Suppose that we thinned the table of feasible strategies by removing the single time period strategies $s(t)$ from likelihood configuration ϕ^l at time t such that the α^{th} quantile of $\hat{o}(\phi^l, t|s(t))_k$ lies below a prespecified value. From the remaining dynamic programming network, we could then maximize long-run expected profit or some quantile of the long-run profit for each replication, producing $s(t)^*_k$ for $t = 1, 2, \ldots, T$ and $k = 1, 2, \ldots, K$.

4.2.4 Smuggler Tendencies and Reactions

Using the table of outcomes and the $K \times T$ optimal single time period strategies, we can calculate the following quantities directly:

- the frequency that a single time period strategy is optimal;
- the distribution of the distance (measured in some way) between the likelihood configuration and the true cutter configuration;
- the smuggler's cost of reducing T, the planning horizon;
- the distribution of the number of time periods the smuggler takes to realize that the cutter configuration does not cover a particular route; and
- the distribution of the number of time periods the smuggler takes to react to drastically increased coverage of a route.

5 CONCLUSION

Some previous attempts at modeling the interaction of the smugglers and the USCG patrol schedule give too much flexibility to the cutter schedule to be realistic. Other approaches don't model the cutter schedule appropriately, so the smuggler's strategies are too simplistic. In this work, we have formulated a dynamic program where the smuggler is forced to combine his short-run profit goals with his need to gain future information about the configuration of the cutters.

Using Monte Carlo methods, we have developed a scheme to estimate stochastic properties of the smuggler's performance using a particular schedule, and shown how constraints on the smuggler's short-run performance can be enforced.

ACKNOWLEDGEMENT

The authors would like to acknowledge the help of Professor Lyn Whitaker.

APPENDIX: NOTATION

- Indices:

r	patrol or smuggler route ($r = 0, 1, \ldots, R$),
c, c'	cutter configurations ($c = 1, 2, \ldots, C$),
t, t'	time periods ($t = 1, 2, \ldots, T$),
i	strategy ($i = 1, 2, \ldots, I$),
k	replication ($k = 1, 2, \ldots, K$),
l	likelihood ($l = 1, 2, \ldots, L$).

- Data:

$s(t)$	shipment attempts at time t ($s(t) = (s_1(t), s_2(t), \ldots, s_R(t))$),
$n(t)$	confiscated shipments in time t ($n(t) = (n_1(t), n_2(t), \ldots, n_R(t))$),
$d_{c,r}$	number of cutters on patrol route r in configuration c,
$P_{c,c'}$	probability of transition from configuration c to c',
p_r	probability of detecting a smuggling vessel by a cutter on route r,
C_1	smuggler's cost of interdicted shipment,
C_2	smuggler's profit for a completed delivery.

- Random Variables:

$N_r(t)$	number of seizures using route r at time t,
$D(t)$	configuration of cutters at time t.
$V(\phi(t)\|s, t)$	value of pursuing strategy s from time t onward.

- Other Notation:

$\phi(t)$	likelihood configuration at time period t,
$\phi_i(t)$	i^{th} component of $\phi(t)$,
$\mathcal{L}$	likelihood operator,
$\hat{o}$	immediate profit minus immediate costs,
$\hat{\phi}(t)$	estimate of $\phi(t)$ generated using likelihood updates,
ϕ^l	l^{th} discrete choice of ϕ,
Δ	mesh size,
α_{ϕ^l}	weight given to ϕ^l for interpolation,
$\Phi(\phi)$	those ϕ^l which are adjacent to ϕ.

REFERENCES

Bailey, M.P. (1994). Solving a Class of Stochastic Minimization Problems. *Operations Research*, **42** (3), pp. 428-438.

Brown, G.G., R.F. Dell, and R.A. Farmer (1994). Scheduling Coast Guard District Cutters. Department of Operations Research, Naval Postgraduate School, Monterey, California, 93943-5000.

Caulkins, J.P., G. Crawford, and P. Reuter. (1993). Simulation of Adaptive Response: A Model of Drug Interdiction. *Math. Comput. Modelling* **2** , pp. 37-52.

Heidleberger, P. and P.A.W. Lewis. (1984). Quantile Estimation in Dependent Sequences. *Operations Research* **32** (1), pp. 185-209.

Owen G., 1982. *Game Theory*, Academic Press, New York.

United States Coast Guard Research and Development Center. (1994). Law Enforcement Simulation (LESIM). Groton, CT.

Wood, K. and A. Washburn. (1993). Two-Person Zero-Sum Games for Network Interdiction. Department of Operations Research, Naval Postgraduate School, Monterey, California, 93943-5000.

AUTHOR BIOGRAPHIES

MICHAEL P. BAILEY is an Associate Professor of Operations Research at the Naval Postgraduate School. His research interests include stochastic analysis and modeling, and computer simulation. He works on a variety of applications critical to the Department of Defense and the U. S. Government.

ROBERT F. DELL is an Assistant Professor of Operations Research at the Naval Postgraduate School. His interests include integer and combinatorial optimization, particularly as applied to problems of sequencing and scheduling.

KEVIN D. GLAZEBROOK is a Professor of Mathematics and Statistics at the University of

Newcastle Upon Tyne. His research focuses on optimal policies for systems in random environments, and has done extensive work developing conditions under which systems have optimal index policies.