Patrolling in Stochastic Environments

Sui Ruan*
Candra Meirina*
Feili Yu*
Prof. Krishna R. Pattipati*
Dr. Robert L. Popp

*Dept. of Electrical and Computer Engineering
University of Connecticut
Contact: krishna@engr.uconn.edu (860) 486-2890

10th International Command and Control Research and Technology Symposium
June 13 - 16, 2005

Outline

- Introduction
- Stochastic Patrolling Model
- Our Proposed Solution
- Simulation Results
- Summary and Future Work

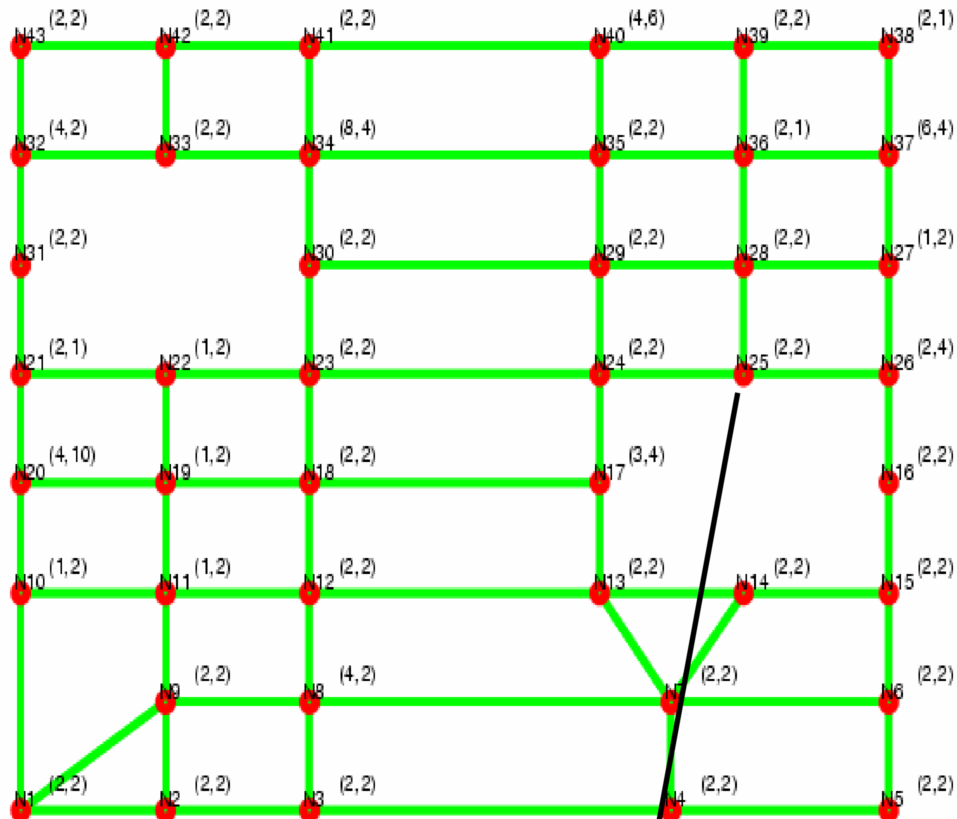
Introduction

- **Motivation**
  - Preventive patrolling is a major component of stability operations and crime prevention in highly volatile environments
  - ***Optimal resource allocation and planning*** of patrol effort are critical to effective stability and crime prevention due to limited patrolling resources

- **Model and Design Objective**
  - Introduce a model of patrolling problems that considers patrol nodes of interest to have ***different priorities*** and ***varying incident rates***
  - Design a patrolling strategy such that the net effect of ***randomized patrol routes*** with immediate call-for-service response allows limited patrol resources to ***provide prompt response to random requests***, while ***effectively covering the entire nodes***

Stochastic Patrolling Problem Model

- Consider a finite set of nodes of interest: $\mathbf{N} = \{i;\ i=1,\ldots,n\}$
- Each node i has the following attributes:
  - Fixed location: (x_i, y_i)
  - Incident rate: λ_i (incidents/hour)
    $\Rightarrow$ assume a Poisson process
  - Important index: δ_i
    $\Rightarrow$ indicate relative importance of node i in the patrolling area
- Assume r patrol units
  $\Rightarrow$ each with average speed v



(incident rate, important index)

$(\lambda_i,\ \delta_i)$

Stochastic Patrolling Problem Methodology

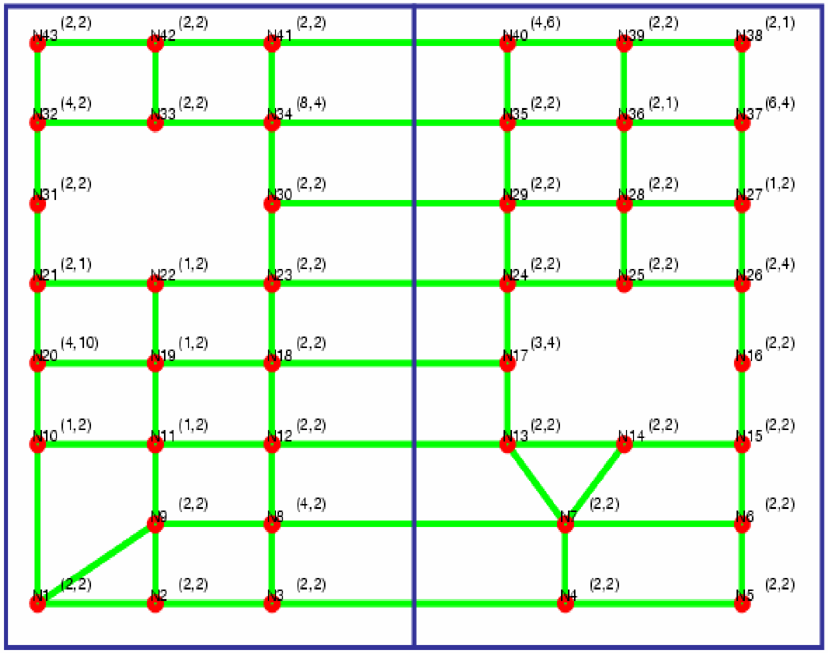
- **Step 1:** Partition the set of nodes of interest into ***sectors*** – subsets of nodes. Each sector is assigned to one patrol unit.
  $\Rightarrow$ Sector partitioning sub-problem

- **Step 2:** Utilize a response strategy of ***preemptive call-for-service response*** and design ***multiple*** off-line patrol ***routes*** for each sector
  - **Step 2.1:** Response strategy
    - Put higher priority to call-for-service requests $\Rightarrow$ stop current patrols and respond to the requests
    - Resume suspended patrols after call-for-service completion
  - **Step 2.2:** Off-line route planning sub-problem
    - Optimal routing in a sector $\Leftarrow$ Similar State Estimate Update (SSEU) in Markov Decision Process framework
    - Strategy for generating multiple patrol routes $\Leftarrow$ randomized (“softmax”) action selection method

Step 1: Sector Partitioning Sub-problem

The problem is formulated as a ***political districting problem***:

- Let the finite set of nodes of interest form a ***region***
- Each node in the region is ***center***ed at (x_i, y_i), and has an importance ***value*** of $\varphi_i = \lambda_i \, \delta_i$
- Define r areas (commensurate to the number of patrol units) over the region such that:
  - $\Rightarrow$ All nodes are covered with minimum overlaps
  - $\Rightarrow$ Similar sums of importance values between areas
  - $\Rightarrow$ Geography of the areas must be compact and contiguous



Sector a **Sector b**

δ_i : Important index of node i

λ_i : Incident rate

This problem has been extensively studied in combinatorial optimization [Garfinkel1970].

2.2: Off-line Route Planning Sub-problem
Markov Decision Process (MDP) Representation

- **States $\{s\}$:**
  - A state is denoted by $s = \{i, \underline{w}\}$
  - i represents the node that has been most recently cleared by a patrol unit (and i is also the current location of the patrol unit)
  - $\underline{w} = \{w_k\}^n_{k=1}$ denotes elapsed time of all nodes since last visits from the patrol unit
- **Action $\{a\}$:**
  - An action is denoted by $a = (i,j)$
  - $j\ (\neq i)$ is an adjacent node of i, the next node to be visited
- **Reward $g(s,a,s')$:**

  Define the reward for taking action $a = (i,j)$ at state $s = \{i, \underline{w}\}$ to reach next state $s' = \{j, \underline{w}'\}$
- **Discount mechanism:**
  - The reward **g potentially earned at time t'** is **valued as $ge^{-\beta(t'-t)}$ at time t**, where β is the discount rate
  - Encourage prompt actions
- **Objective:**

  Determine an **optimal policy**, i.e., **a mapping from states to actions**, that maximizes the overall expected reward

2.2: Off-line Route Planning Sub-problem Linear State Value Structure

- Arbitrary MDP problems are intractable
- Fortunately, our patrolling problem exhibits a special structure: linearity
  - For any deterministic policy in the patrolling problem, the state value function has the property:

State Value function, $V^{\Pi}(s)$, is the expected reward starting from state s, under policy Π.

$$V^{\Pi}(s=(i,\underline{w}))=(\underline{c}_i^{\Pi}(s))^T\underline{w}+d_i^{\Pi}(s)\quad \forall i\in \mathbf{N}$$

linear w.r.t. $\underline{w}$ (elapsed time of nodes since last visits from a patrol unit)

  - Thus, a linear approximation of state value function for optimal policy is:

$$\boxed{\tilde{V}^{*}(s=(i,\underline{w}))=(\underline{c}^{*}_{i})^T\underline{w}+d^{*}_{i}}$$

  - The problem becomes one of finding $\underline{c}^*_i$, d^*_i, $\forall i\in\mathbf{N}$ $\Rightarrow$ determine the optimal policy

Similar State Estimate Update Method -1

Introduce a variant of Reinforcement Learning (RL) method, ***Similar State Estimate Update (SSEU)*** method, to learn the optimal parameters $\underline{c}^*_i$ and d^*_i, $\forall i \in \mathbf{N}$

- Reinforcement learning is a simulation-based learning method, which requires only experience, i.e., sample of sequences of states, actions and rewards from on-line or simulated interaction with the system environment
- Given an arbitrary policy, Π, policy iteration method of RL iteratively improves the policy to gradually approach Π^* as follows:

$$k^* = \arg \max_{\forall a \in (i,k),\ k \in adj(i)} \underbrace{\alpha(s,s')}\{\underbrace{E[g(s,a=(i,k),s')]}+\underbrace{V^{\Pi}(s')}\}$$

$$V^{\Pi}(s') = (\underline{c}_k)^T \underline{w} + d_k$$

$$\alpha(s,s') = e^{-\beta * \frac{dist(i,j)}{v}}$$

Discount from s to s'

Reward for taking action a at state s, and reaching state s'

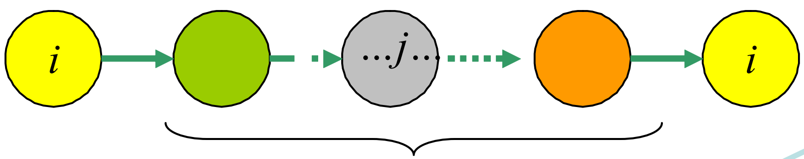
State value of s' under Π

β: discount rate
v : average speed
a : action

2.2.a: Optimal Routing in a Sector Similar State Estimate Update Method -2

- Generate a trajectory via policy iteration utilizing current parameter estimates, $\underline{c}^t_i$ and d^t_i, for two adjacent similar states of node i, state $s=\{i,\underline{w}\}$, $\tilde{s}=\{i, \underline{w}'\}$:

Similar States: same node location, different visitation time



$\underline{w}$ (elapsed time of nodes since last visits from a patrol unit)

j represents a node along the trajectory

- Evaluate new values of c_{ij}^{new} and d_i^{new}:

t^1_j denotes the first time node j is visited in the trajectory; and

$$c_{ij}^{*} = \delta_j \lambda_j e^{-\beta(t_j^{1*}-t_0)}$$

$$c_{ij}^{new} = \delta_j \lambda_j e^{-\beta(t_j^{1}-t_0)}$$

$$d_i^{new} = \sum_{j=1}^{m} g_j e^{-\beta(t_j-t_0)} + \alpha(s,s')V^t(s') - (\underline{c}_i^{new})\underline{w}^T$$

$$V^t(s') = (\underline{c}^t_i)^T \underline{w} + d^t_i$$

δ_j: Important index of node j
λ_j : Incident rate
β: discount rate

- Thus

$$\boxed{c_{ij}^{t+1} = c_{ij}^t + \frac{c_{ij}^{\text{new}} - c_{ij}^t}{m_{ij}^c}}$$ and $$\boxed{d_i^{t+1} = d_i^t + \frac{d_i^{\text{new}} - d_i^t}{m_i^d}}$$

m^c_{ij} : number of c_{ij} previous updates m^d_i : number of d_i previous updates

2.2.b: Strategy for Generating Multiple Patrolling Routes

- **Why multiple patrolling routes?**
  - To impart virtual presence and unpredictability to patrolling
    $\Rightarrow$ the patrol unit randomly selects one of many patrol routes
- **Softmax: random action selection method**
  - At each state,
    - The best action is given the highest selection probability
    - The second best action is given lesser probability
    - The third best action is given even less and …
  - Temperature – tunable parameter – decides probability differences among the actions
    - High temperatures $\Rightarrow$ virtually equal probability
    - Low temperatures $\Rightarrow$ greater difference in selection probabilities for actions having different value estimates

Simulation Results

- Results from the Illustrative Patrol Problem

Range	Method	Expected Reward	Reward per Unit Distance
Whole Region	SSEU	2,330	17.4
	Greedy	1,474	6.0
Sector a	SSEU	1,710	19.43
	Greedy	1,455	13.8
Sector b	SSEU	1,471	13.8
	Greedy	1,107	10.9

⇑ Reward:
- ⇑ Number of cleared incidents
- ⇑ Incident importance
- ⇓ Latency

Greedy refers to one-step greedy strategy, i.e., for each state, select the neighboring node with best instant reward

- Patrol routes obtained by the SSEU method are highly efficient compared to the one-step greedy strategy
- Net reward from two patrolling units (for sectors a and b) is 36% higher with the SSEU method when compared to that of one patrol unit in the whole region

Summary and Future Work

- Present an analytical model of patrolling problem with varying incident rates and priorities
- Propose a solution approach in two steps:
  - Step 1: Solve the sector partitioning sub-problem via Political Districting Method $\Rightarrow$ assign each sector to one patrol unit
  - Step 2: Utilize a response strategy of preemptive call-for-service and define an optimal and near-optimal patrol routes for each sector via SSEU and “softmax”-based method, respectively
- Future work:
  - Incorporate incident processing time and resource requirements for each node
  - Include patrol unit’s resource capabilities and workload constraints
  - Introduce dynamic rerouting in the presence of changes in the incident rates and node priorities

References

D. J. Kenney, *Police and Policing: Cotemporary Issues*, Praeger, 1989.

R. C. Larson, “Urban Police Patrol Analysis”, The MIT Press, 1972.

J. Tsitsiklis and B. Van Roy, “Analysis of Temporal-Difference Learning with Function Approximation”, *IEEE Transactions on Automatic Control*, Vol. 42(5) 1997, pp. 674-690.

R.S. Garfinkel and G.L. Nemhauster, “Optimal Political Districting by Implicitly Enumeration Techniques”, *Operations Research*, Vol. 16, 1970, pp. 495-508.

R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, The MIT Press, 1998.

Thank You !