Presence of adversarial observers should require that optimal patrolling policy exhibit trade-off between efficiency and entropy.
Motivation

Game environment

Prior work

General game model

Simplified game models

Numerical simulations of various stochastic policies

Towards proving analytical bounds

Limitations and extensions
Game Environment and Model

Both Patroller, $\mathcal{P}$, and Adversary, $\mathcal{A}$, constrained to some planar environment

$\mathcal{A}$’s location unknown to $\mathcal{P}$

Model $\mathcal{A}$’s observations as indicator function

$\mathcal{A}$ only observable when conducting attack of duration $s$

$\mathcal{P}$ captures $\mathcal{A}$ if $\mathcal{P}$ inter-arrival time $\Delta T < s$

Patroller $\mathcal{P}$ must determine optimal policy given that $\mathcal{A}$ can observe $\mathcal{P}$ for some time prior to deciding when and where to attack.
Prior Work

Optimal Search

Search Games

Security Games

[1] Paruchuri 06
[3] Agmon 09
[5] Basilico 09,10,11
[6] Huynh 10
[7] Gage 93
[8] Stone 89
[9] Bethke 08, 10
[10] Chevaleyre 04
[12] Fu 09
[14] Gatti 08
[16] Dickerson 10
Prior Work – Gal 80: Search Games

- Optimal Search
- Security Games
- Search Games


- Cold War submarine hunting
- “Princess-Monster” game
- Princess is static or dynamic
- Monster searches within an enclosed environment
- Game payoff to Monster is time required to capture the Princess
- One shot game: no learning between stages

[1] Phil Root, MIT–LIDS
Prior Work – Agmon 08-11: Patrolling Perimeter

- Optimal Search
- Security Games
- Adversary has varying degrees of information
- Policy consists of course reversal probability
- Validated with human subjects

Phil Root, MIT–LIDS
Prior Work – Basilico 09–11: Patrolling in Presence of Adversaries

[5] Basilico 09-11

- Very similar work on patrolling agents in game setting
- Shows that patroller can do no worse than leader equilibrium
- Recasts infinite-horizon game with strategic game with $H$ stages
- Patroller strategy is to maximize the probability of capture
Proposed game theoretic model establishes trade-off between efficiency versus entropy (i.e. $G$ versus $\Gamma$)
Simplified game model – First order adversary

Model used by Basilico, Agmon

<table>
<thead>
<tr>
<th>Commit Crime, C</th>
<th>Successful Crime, S</th>
<th>Unsuccessful Crime, U</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>(−1, 1)</td>
<td>(1, −1)</td>
</tr>
</tbody>
</table>

\[ \text{Probability of “Uncapture”} \]

\[ P_u = \Pr(\Delta T > s) \]

\[ E[U_\mathcal{P}] = -P_u \cdot 1 + (1 - P_u) \cdot 1 \]

\[ = 1 - 2P_u \]

\( \mathcal{A} \) cannot take advantage of all information in \( \Delta T \) distribution.
A adversary A chooses between committing crime C or deferring action D.

<table>
<thead>
<tr>
<th>Adversary A</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Commit Crime, C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Successful Crime, S</td>
<td>Unsuccessful Crime, U</td>
<td>Defer Action D</td>
</tr>
<tr>
<td>((-\Gamma, 1))</td>
<td>((\Gamma, -k))</td>
<td>((1, -\delta))</td>
</tr>
</tbody>
</table>

\(\mathcal{P}\) patrol policy \(\pi_i\)

**Probability of "Uncapture"**

\[ P_u = 1 - \int_0^s \bar{\lambda} e^{-\bar{\lambda}s} \]

\[ P_u = e^{-\bar{\lambda}s} \]

**Conditions for A to choose C versus D**

\[ E[U_C] > E[U_D] \]

\[ P_u - k(1 - P_u) > -\delta \]

\[ P_u > \frac{k - \delta}{1 + k} \]

A chooses to commit crime C based on risk aversion \(k > 0\) and time discount \(\delta > 0\).
**Simplified game – Conditional probability**

**Conditional Probability of “Uncapture”**

\[ P_u(T) = \Pr(\Delta T > T + s | \Delta T > T) \]

\[ P_u(T) = \frac{\int_{T+s}^{\infty} f_\pi(\Delta T) d\Delta T}{\int_{T}^{\infty} f_\pi(\Delta T) d\Delta T} \]

**Time Discounted \( P_u, W_u \)**

\[ W_u(T) = \frac{1}{\hat{t}(T)} P_u(T) \]

\[ \hat{t}(T) = E[\text{time before } \Delta T_i > T] = \frac{p}{1 - p} \hat{t}(T) + T \]
Objective is to determine optimal patrolling policy given two-player infinite-horizon strategic game by developing analytic performance bounds using time-discounted probability of “uncapture”, $W_u(T)$ as metric.

Novel Contributions

- Model $A$’s observations as indicator function
- Allow $A$ to use information from sequence $\langle \Delta T_i \rangle$
- Allow $A$ to choose optimal waiting time, $T^*$
- Introduce framework to address dynamic learning phase
“Persistent Patrol with Limited-range On-board Sensors”
Huynh, Enright, Frazzoli (2010)

Numerical simulation

1. Create paths from policy
2. Determine sequence $\langle \Delta T_i \rangle$
3. Plot $W_u$

Patroller $\mathcal{P}$ seeks policy with **minimum of maximum** $W_u$
Uniform point sampling demonstrates clear central tendency as expected
Perimeter point sampling demonstrates a marked uniformity compared to uniform sampling.
Stochastic Reflection Policy

Adapted from Lalley and Robbins, 1988, “Stochastic Search in a Convex Region”

Stochastic reflection policies do not achieve uniformity in this environment.
Brownian motion with small step size (low variance) can lead to poor exploration.
"Lévy Flight are optimal for the location of stationary targets that are randomly and sparsely distributed, and once visited are not depleted but instead remain targets for future searches."
Lévy Spirals (aka “Lollipop Search”) 
Adapted from Reynolds (2009) “The Lévy Flight Paradigm: Random Search Patterns and Mechanisms”

Lévy Flight is optimal in relocating to a food patch, and spiral search is optimal within a food patch given an ideal sensor.
Markov Analysis to Explore Performance Bounds

Bresenham’s Algorithm used to discretize influence of line segment in environment

Uniform Point Policy

Perimeter Point Policy

Markov models demonstrate appropriate trends and may be useful in determining performance bounds of various patrolling policies.
## Limitations and Future Work

### Limitations

<table>
<thead>
<tr>
<th>Limitation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-uniform adversary utility</td>
<td>Considers only uniform adversary target preference over environment.</td>
</tr>
<tr>
<td>Game Equilibrium</td>
<td>May be difficult to characterize; can likely only bound expected performance.</td>
</tr>
</tbody>
</table>

### Future Work

<table>
<thead>
<tr>
<th>Future Work</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adversary can learn temporal dependencies</td>
<td>We assume that the series of $\Delta T$ are not drawn i.i.d. from the probability distribution, and $A$ can learn any temporal dependencies of this Markov chain renewal process.</td>
</tr>
<tr>
<td>Adversary has dynamic learning phase</td>
<td>$A$ is not given $\Delta T$ probability distribution, rather $A$ must learn this empirical distribution and choose to attack or not given some optimal stopping rule.</td>
</tr>
</tbody>
</table>
References I


Derivation of Time Discounted $P_u$ (1 of 2)

- Given sequence $\langle \Delta T_i \rangle$, assume sequence is i.i.d. from distribution $f_\pi$
- Define:

  \[ p(T) = \Pr(\Delta T < T) \]
  \[ \bar{t}(T) = E[\Delta T | \Delta T < T] \]

\[ \Pr(\Delta T_n \geq T, \Delta T_k < T \ \forall k = 1, \ldots, n-1) = (1 - p(T))p(T)^{n-1} \]

\[ E[\text{elapsed time} | \Delta T_n \geq T, \Delta T_k < T \ \forall k = 1, \ldots, n-1] = (n-1)\bar{t}(T) + T \]

\[ E[\text{elapsed time} | \Delta T > T] = \hat{t}(T) = \sum_{k=1}^{\infty} \left( (k-1)\bar{t}(T) + T \right) (1 - p(T))p(T)^{k-1} \]
Derivation of Time Discounted $P_u$ (2 of 2)

\[
\hat{t}(T) = \sum_{k=1}^{\infty} p^{k-1} T(1 - p) + \sum_{k=1}^{\infty} p^{k-1} (k - 1) \bar{t}(1 - p)
\]
\[
= T(1 - p) \sum_{k=0}^{\infty} p^k + \bar{t}(1 - p) \sum_{k=0}^{\infty} kp^k
\]
\[
= T(1 - p) \frac{1}{1 - p} + \bar{t}(1 - p) \lim_{k \to \infty} p \frac{1 - (k+1)p^k + kp^{k+1}}{(1-p)^2}
\]
\[
= T + \bar{t}(T) p(T) \frac{1}{1 - p(T)}
\]

$W_u(T) = \frac{1}{\hat{t}(T)} P_u(T)$
Markov Transition Matrix
\[ P \in \mathbb{R}^{n \times n} \]
\[ |\text{destination state}| = n \]
\[
\begin{bmatrix}
0 & p_{12} & \cdots & p_{1n} \\
p_{21} & 0 & \cdots & p_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
p_{n1} & p_{n2} & \cdots & 0
\end{bmatrix}
\sum_j p_{ij} = 1
\Rightarrow
\begin{bmatrix}
0 & p_{12} & \cdots & p_{1n} \\
p_{21} & 0 & \cdots & p_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
p_{n1} & p_{n2} & \cdots & 0
\end{bmatrix}
\sum_j \tilde{p}_{ij} = 1
\]

Skewed Transition Matrix
\[ \tilde{P} \in \mathbb{R}^{n \times n^2} \]
\[ |\text{transitions}| = n^2 \]
\[
\begin{bmatrix}
0 & p_{12} & \cdots & p_{1n} \\
p_{21} & 0 & \cdots & p_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
p_{n1} & p_{n2} & \cdots & 0
\end{bmatrix}
\sum_j \tilde{p}_{ij} = 1
\]

Influence Matrix
\[ L \in \mathbb{R}^{n^2 \times n} \]
\[ |\text{influenced states}| = n \]
\[
\begin{bmatrix}
1 & 0 & 0 & \cdots & 0 \\
1 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 1 & 1 & \cdots & 1 \\
0 & 0 & 0 & \cdots & 1
\end{bmatrix}
\sum_j L_{ij} \leq n
\]
\[
\begin{bmatrix}
q_{11} & q_{12} & \cdots & q_{1n} \\
q_{21} & q_{22} & \cdots & q_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
q_{n1} & q_{n2} & \cdots & q_{nn}
\end{bmatrix}
\sum_j q_{ij} \geq 1
\]

For stationary distribution \( \pi \), \( \sigma \) is the stationary distribution of influenced states.

\[ \pi P = \pi \Rightarrow \pi Q = \sigma \]
Proof of Exponential Distribution Optimality (1 of 4)

Let \( \pi \rightarrow f \) be a mapping such that
\[
\int_a^b f(T) \, dT = \Pr(a \leq \Delta T \leq b)
\]

\[P_c(T) = \frac{\int_T^{T+s} f \, dT}{\int_T^\infty f \, dT} = \frac{F(T+s) - F(T)}{1 - F(T)} \approx \frac{(F(T) + F'(T)s) - F(T)}{1 - F(T)} \approx \frac{F'(T)s}{1 - F(T)} \]

\[F^* = \arg \max_{F \in \mathcal{F}} \min_{T \in [0,t_f]} \frac{F'(T)s}{1 - F(T)} \]

Discretize \( F(T) \rightarrow F_k \)
Recast as optimization problem
\[
F^* = \arg \max_{F \in \mathcal{F}} c
\]
\[
s(F_{k+1} - F_k) \geq c
\]
\[\frac{1 - F_k}{1 - F_k} \geq c
\]
\[F_0 = 0
\]
\[F_\infty = 1
\]

Theorem

The optimal discrete CDF is
\[F^*_k = 1 - e^{-\lambda^*k}\]
where \( \lambda^* > 0 \) represents the highest average revisit rate (i.e. \( \lambda^* = 1/\mathbb{E}[\Delta T_{\min}] \)) as a function of geometry \( \lambda^* = \frac{|E|}{2vr_s} \).
If $F_k^* = 1 - e^{-\lambda^* k}$, then constraint (3) reduces to constant

$$\frac{s(e^{-\lambda^* k} - e^{-\lambda^*(k+1)})}{e^{-\lambda^* k}} \geq c^*$$

$$s(1 - e^{-\lambda^*}) \geq c^* \quad (4)$$

Proof by contradiction

- Let $\tilde{F}_k$ be optimal CDF such that $\tilde{c} > c^*$
- Let $\tilde{c} = s(1 - e^{-\lambda^*}) + \epsilon$
- Let $\tilde{F}_0 = 0$ by convention.
- Then $\tilde{F}_k = \frac{(1 - \tilde{F}_{k-1})s(1 - e^{-\lambda^*}) + \epsilon}{s} + \tilde{F}_{k-1}$
Assume \( \tilde{F}_{k-1} \sim \text{Exp}[\lambda^* + \Delta] \).

\[
\tilde{F}_k = (1 - (1 - e^{-(\lambda^* + \Delta)(k-1)})(1 - e^{-\lambda^* + \varepsilon/s}) + 1 - e^{(\lambda^* + \Delta)(k-1)}
\]

\[
= 1 + (-e^{-\lambda^* + \Delta})e^{-\lambda^* + \varepsilon/s} + 1 - e^{-(\lambda^* + \Delta)(k-1)}
\]

\[
\ln \left[ \frac{\tilde{F}_k - 1}{-e^{-\lambda^* + \varepsilon/s}} \right] = -(\lambda^* + \Delta)(k - 1)
\]

\[
\ln \left[ \tilde{F}_k - 1 \right] - \ln \left[ -e^{-\lambda^* + \varepsilon/s} \right] = -(\lambda^* + \Delta)(k - 1)
\]

\[
\ln \left[ -e^{-\lambda^* + \varepsilon/s} \right] \approx \ln \left[ \lambda - \frac{\varepsilon/s}{e^{-\lambda}} \right]
\]

\[
\tilde{F}_k = 1 - e^{-(\lambda^* + \Delta)k}
\]
Therefore we find that the assumption $\tilde{F}_{k-1} \sim \text{Exp}[\lambda^* + \Delta]$ satisfies the constraint equation

Requires that $\Delta = \frac{\epsilon/s}{e^{-\lambda^*}}$

But $\Delta > 0$ so $\lambda^* + \Delta > \lambda^*$

Contradiction because we stated that $\lambda^*$ is the highest possible revisit rate.

Proof Summary

We have shown that the optimal policy for this simplified game is a policy $\pi^* \mapsto f^*$ where $f^* \sim \text{Exp}[\lambda^*]$. This is a local optimum, and we cannot yet claim global optimality.

This proof demonstrates the optimality of exponential distribution $f^* \sim \text{Exp}[\lambda^*]$, but there does not exist a policy $\pi^* \mapsto f^*$. The proof does not inform us whether a policy $\pi_1 \mapsto f_1 \sim \text{Exp}[\lambda_1]$ with $\lambda_1 > \lambda^*$ is optimal versus a non-exponential distribution $\pi_2 \mapsto f$. 
Uniform point sampling on a torus demonstrates absence of clear central tendency as expected.