Bandit Algorithms for Data-Driven Resolution/Field-of-View Tradeoffs in Multi-Mode Sensing and Intelligence Collection

Jefferson Huang

Assistant Professor

Operations Research Department

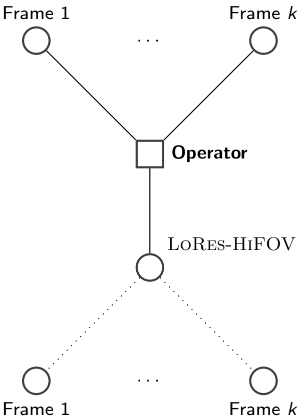
Naval Postgraduate School

31st European Conference on Operational Research

University of West Attica
Athens, Greece

14 July, 2021

Plan of the Talk



Outline:

I. **Motivation:** Cyber Intrusion Model

II. **Motivation:** Routing Imaging Assets

III. **Background:** Multi-Armed Bandit Problems

IV. A Bandit Model of the Resolution/Field-of-View Tradeoff

V. The Router-Then-Commit (RTC) Algorithm

VI. A UCB-Type Algorithm

VII. Summary & Extensions

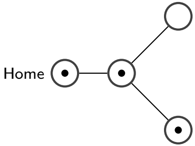
Motivation: Cyber Intrusion Model

D. Kronzilber (NPS Master's Thesis, 2017) proposed a model for the optimal infiltration of, and intelligence collection from, a computer network.

- The intelligence yield from each computer is random, with unknown mean.

Question: Given a set of computers that have been infiltrated, how can the infiltrator simultaneously:

- **Learn** which computers have the highest mean intelligence yield?
- **Maximize** the rate of intelligence collection?



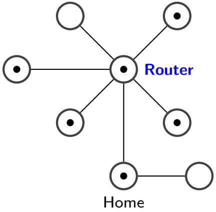
Example: Infiltrated network nodes are denoted by a •

In addition to nodes representing computers, Kronzilber suggested the inclusion of **routers** as special nodes.

- Using a router, the infiltrator can observe **small/partial bits of intelligence** from all computers connected to it.
- The router provides a wider **field-of-view (FOV)**, at the cost of additional noise.

Resolution/FOV Tradeoff

How can the infiltrator effectively employ the router, versus individual infiltrated computers?



Example: Infiltrated network nodes are denoted by a •

Motivation: Routing Imaging Assets

An operator is tasked with collecting imagery intelligence over an **area of interest**.

Collected image frames are fed into an image processing module that assigns a score reflecting the **intelligence value** of the image.

- lower-resolution image $\Longrightarrow$ noisier intelligence score

Two **imaging modes** are available:

LoRes-HiFOV: Collect from all frames at once, at low resolution

HiRes-LoFOV: Collect from a single frame, at high resolution

Resolution/FOV Tradeoff

How can the operator effectively employ LoRes-HiFOV, versus HiRes-LoFOV?

Background: Multi-Armed Bandit Problems

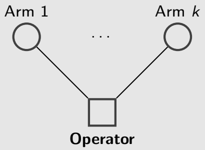
In the old days, a "one-armed **bandit**" referred to a lever-operated slot machine.

Multi-Armed Bandit (MAB)

There are k reward sources, referred to as **arms**.

- Rewards are random.
- The **mean reward** μ_a for each arm a is unknown.

In each **round** $t = 1, \ldots, n$, the operator can collect from exactly one arm.



The operator's **objective** is to minimize their **regret** R_n; letting X_t be the reward earned in round t,

$$R_n = \mathbb{E}\left[\sum_{t=1}^{n}\left(\max_a \mu_a - X_t\right)\right]$$

Exploration/Exploitation Tradeoff

In each round, the operator must balance two considerations:

- **Exploration:** Learn about the mean rewards.
- **Exploitation:** Maximize the reward earned.

A classic approach involves assigning iteratively updated **indices** $\mathcal{I}_a(t)$ to each arm a. They reflect, as of the start of round t,

- the average reward $\bar{X}_a(t)$ earned from arm a and
- the number of times $N_a(t)$ arm a has been collected from

In round t, the operator collects from arm with the **highest index**, i.e.,

$$a^* = \arg\max_{a=1,\ldots,k} \mathcal{I}_a(t)$$

Example: Upper Confidence Bound (UCB) Algorithm

$$\mathcal{I}_a(t) = \bar{X}_a(t) + \sqrt{\frac{2\ln(n)}{N_a(t)}}$$

A Bandit Model of the Resolution/Field-of-View Tradeoff

Introduce an additional arm, called LoRes-HiFOV.

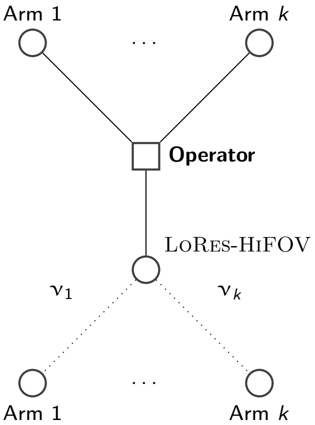
- Collecting from LoRes-HiFOV means that, **for each arm**

$$a = 1, \ldots, k$$

the operator collects from arm a, with probability ν_a.

Assumptions for this Talk

- The ν_a's are identically equal to ν.
- The operator knows ν.



The Router-Then-Commit (RTC) Algorithm

Use LoRes-HiFOV to rule out a subset of the arms.

Idea

1. Pull LoRes-HiFOV a fixed number of times τ.
2. Eliminate a subset $E \subseteq \{1, \ldots, k\}$ of the arms from further consideration.
3. Consider the resulting $(k - |E|)$-armed bandit problem.

Select τ so that, with high probability, all arms $\underline{a}$ with a "large" optimality gap

$$\Delta_{\underline{a}} := \left(\max_{a=1,\ldots,k} \mu_a \right) - \mu_{\underline{a}}$$

will be eliminated.

Lemma

Suppose LoRes-HiFOV is initially pulled t times. For any $\delta > 0$, if

$$t > \frac{\ln(2/\delta)}{\nu},$$

then

$$\mathbb{P}\left(|\bar{X}_a(t) - \mu_a| \geqslant \sqrt{\frac{1}{2} \ln\left(\frac{\nu}{\nu + (\delta/2)^{1/t} - 1} \right)} \right) \leqslant \delta$$

RTC Algorithm

1. Given a target gap $\Delta > 0$, pull LoRes-HiFOV

$$\tau = \tau(\Delta) := \left\lceil \frac{\ln(2kn^2)}{\nu \cdot (1 - e^{-\Delta^2/8})} \right\rceil$$

times.

2. Eliminate every arm $\underline{a}$ for which

$$\bar{X}_{\underline{a}}(\tau) + \sqrt{\frac{1}{2} \ln\left(\frac{\nu}{\nu + [1/(2kn^2)]^{1/\tau} - 1} \right)} < \max_{a=1,\ldots,k} \left[\bar{X}_a(\tau) - \sqrt{\frac{1}{2} \ln\left(\frac{\nu}{\nu + [1/(2kn^2)]^{1/\tau} - 1} \right)} \right]$$

3. Apply the UCB algorithm to the remaining arms.

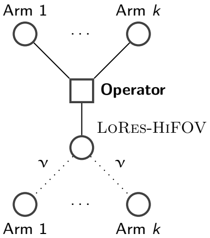
Theorem

Suppose the rewards are all between 0 and 1. If $n \geqslant \tau(\Delta)$, then the RTC algorithm incurs a regret of

$$R_n \leqslant 1 + \tau(\Delta) + 8\sqrt{kn \ln(n)} + 3k\Delta.$$

For example, if $\Delta = \sqrt{8 \ln\left(\frac{\sqrt{k}}{\sqrt{k} - 1} \right)}$, then $O\left(\sqrt{k} \ln(kn) \right)$ pulls of LoRes-HiFOV ensures $R_n = O\left(\sqrt{k} \ln(kn) + \sqrt{kn \ln(n)} + k\sqrt{\ln(\sqrt{k})} \right)$.

A UCB-Type Algorithm



Notation:

- $\bar{X}_a(t)$ = average reward from arm a up to round t
- $N_a(t)$ = number of collections from arm a prior to round t
- $N_{\mathrm{LH}}(t)$ = number of LoRes-HiFOV collections
- $\mathcal{A}(t)$ = set of **active arms** at the start of round t
- $\mathcal{I}_a(t) = \bar{X}_a(t) + \sqrt{\frac{2\ln(n)}{N_a(t) + \nu N_{\mathrm{LH}}(t)}}$ = **index** of arm a
- $\mathcal{I}_{\mathrm{LH}}(t) = \nu|\mathcal{A}(t)|\sqrt{\frac{2\ln(n)}{\nu N_{\mathrm{LH}}(t)}}$ = **index** of LoRes-HiFOV

UCB-Type Algorithm

1. **Initialization:**
   - Collect from LoRes-HiFOV until each arm has been collected from at least once.
   - **Eliminate** all arms $a \in \{1, \ldots, k\}$ where

$$\mathcal{I}_a(t) < \max_{a=1,\ldots,k} \left\{ \bar{X}_a(t) - \sqrt{\frac{2\ln(n)}{N_a(t) + \nu N_{\mathrm{LH}}(t)}} \right\}$$

2. **Main:** For each round $t = 1, \ldots, n$,
   - If $\mathcal{I}_{\mathrm{LH}}(t) > \max_{a=1,\ldots,k} \mathcal{I}_a(t)$, collect from LoRes-HiFOV.
   - Otherwise, collect from arm

$$a^* = \arg\max_{a=1,\ldots,k} \mathcal{I}_a(t)$$

   - **Eliminate** all arms $a \in \{1, \ldots, k\}$ where

$$\mathcal{I}_a(t) < \max_{a \in \mathcal{A}(t)} \left\{ \bar{X}_a(t) - \sqrt{\frac{2\ln(n)}{N_a(t) + \nu N_{\mathrm{LH}}(t)}} \right\}$$

   - Update $\mathcal{A}(t)$

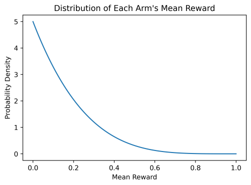
Empirical Performance of the UCB-Type Algorithm

The LoRes-HiFOV arm should be especially beneficial when there are many arms, and relatively few are good.

Example (0-1 Rewards)

Each arm yields a reward of 1 (e.g., "Useful Intelligence") or 0.

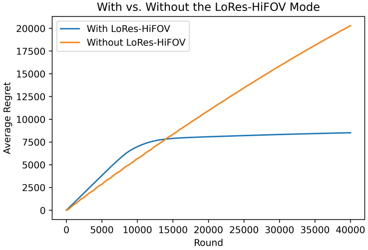
- The **mean rewards** μ_a vary according to a Beta distribution with parameters $\alpha = 1$ and $\beta = 5$:

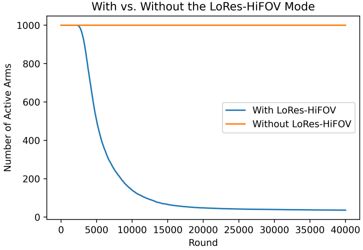


- $k = 1000$ arms.
- $n = 40,000$ rounds.
- $\nu = 0.02$

The average performance of our algorithm over 100 simulation replications, with and without the LoRes-HiFOV mode, is shown on the right.

- Without LoRes-HiFOV = UCB Algorithm with Arm Elimination





The UCB-Type Algorithm uses LoRes-HiFOV to quickly screen the less desirable arms.

Summary & Extensions

Summary:

- We proposed a model for trading off "resolution" versus "field-of-view".
- We analyzed the "router-then-commit" algorithm, where LoRes-HiFOV is pulled a number of times first to eliminate some of the arms from further consideration.
- We proposed an index-based ("UCB-type") algorithm that has good empirical performance.

Some Potential Extensions:

- Don't know the ν_a's.
- Different noise models for LoRes-HiFOV
- More than two resolution/FOV options
- Detecting **changes** in mean intelligence values.

Contact Information

Jefferson Huang, PhD

Assistant Professor
Operations Research Department
Naval Postgraduate School

✍ **Web:** http://faculty.nps.edu/jefferson.huang/
✉ **Email:** jefferson.huang@nps.edu