Regret Minimization Algorithms for the Follower's Behavior Identification in Leadership Games

Relatore:
Prof. Marcello Restelli

Tesi di Laurea di:
Lorenzo Bisi

27/07/2017

Politecnico di Milano

Dipartimento di Elettronica Informazione e Bioingegneria

Context: Physical Security of Environments or Infrastructures

Examples

- Fighting criminal activities in cities
- Protecting airports from terrorists
- Protecting animals from poachers in natural parks

Features

- **Targets**: valuable assets, possibly with diffent values

Features

- **Targets**: valuable assets, possibly with diffent values
- **Defender**: she tries to protect targets from the opponent's attacks

Features

- **Targets**: valuable assets, possibly with diffent values
- **Defender**: she tries to protect targets from the opponent's attacks
- **Attacker**: she tries to attack the targets without being caught by the defender. She can do surveillance on the area in order to discover patterns in defender's moves.

Features

- **Targets**: valuable assets, possibly with diffent values
- **Defender**: she tries to protect targets from the opponent's attacks
- **Attacker**: she tries to attack the targets without being caught by the defender. She can do surveillance on the area in order to discover patterns in defender's moves.
- Not enough guards to cover all the targets simultaneously

Features

- **Targets**: valuable assets, possibly with diffent values
- **Defender**: she tries to protect targets from the opponent's attacks
- **Attacker**: she tries to attack the targets without being caught by the defender. She can do surveillance on the area in order to discover patterns in defender's moves.
- Not enough guards to cover all the targets simultaneously

How should the defender arrange her resources?

Security Games

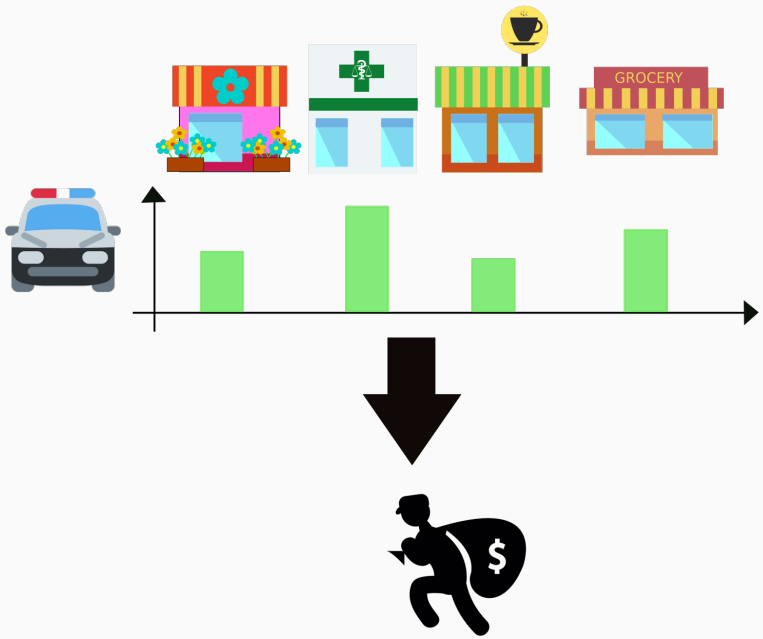
At each round:

- the defender commits to a strategy

Security Games

At each round:

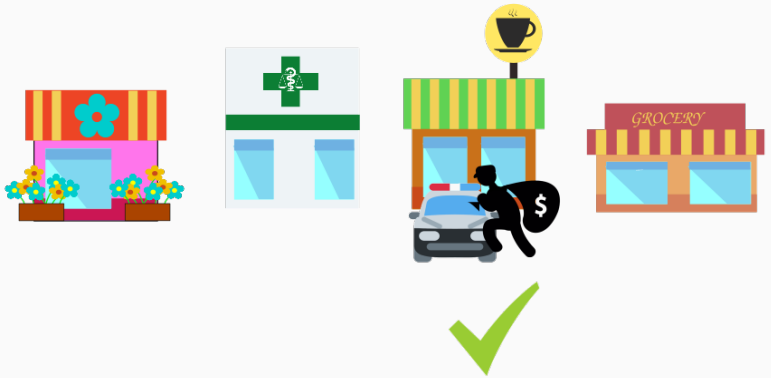
- the defender commits to a strategy
- the attacker observes it and chooses a strategy



Security Games

At each round:

- the defender commits to a strategy
- the attacker observes it and chooses a strategy
- they play and receive a payoff

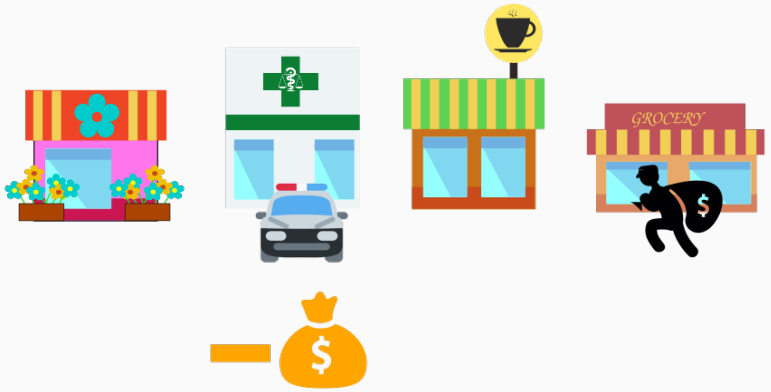


The defender does not incurr in any loss.

Security Games

At each round:

- the defender commits to a strategy
- the attacker observes it and chooses a strategy
- they play and receive a payoff



The defender incurrs in a loss equal to the value of the target.

Attacker Rationality

- **Perfect Rationality**: The attacker plays in the *best possible way* given the defender's commitment.
  e.g: LAX airport (ARMOR)

- **Bounded Rationality**: The attacker thinks in a *human* way and plays according to her limited capacities.
  e.g: Pouching (Green Security Games)

- **Unknown Rationality**: What can we do not know opponent's intelligence?

Problem Definition

Follower's Behaviour Identification in Security Games

FBI-SG problem is a tuple $(\mathcal{G}_N, \mathcal{A}, A_{k^*})$, where :

- $\mathcal{G}_N$ is a 2-players security game;
- $\mathcal{A} = \{A_1, \ldots, A_K\}$ is a set of possible attacker behavioural profiles;
- $A_{k^*} \in \mathcal{A}$ denoting the actual profile of the attacker.

The goal is minimizing the **expected pseudo regret** over a finite-time horizon of **N** rounds:

$$R_N(\mathfrak{U}) = \mathbb{E}\left[\sum_{n=1}^{N} l_n\right] - L^* N$$

FBI-SG is an **online learning** problem:

- *online* nature of repeated games
- presence of *uncertainty*

Analysed Attacker's Profiles

Let's call $\boldsymbol{\sigma}$ the defender's strategy. We consider three different kind of profiles:

- **Stackelberg Attacker** (Perfect Rationality - *strategy-aware*):

$$\boldsymbol{\sigma}_{Sta}(\boldsymbol{\sigma}) = \arg \max_{\boldsymbol{\sigma}' \in \Delta_M} \sum_{m \in \mathcal{M}} \sigma'_m \, v_m \, (1 - \sigma_m)$$

- **Subjective Utility Quantal Response Attacker (SUQR)** (Bounded Rationality - *strategy-aware*):

$$\sigma_{SUQR}(\boldsymbol{\sigma})_m = \frac{\exp\{-\alpha \sigma_m + \beta v_m\}}{\sum_{h=1}^{M} \exp\{-\alpha \sigma_h + \beta v_h\}},$$

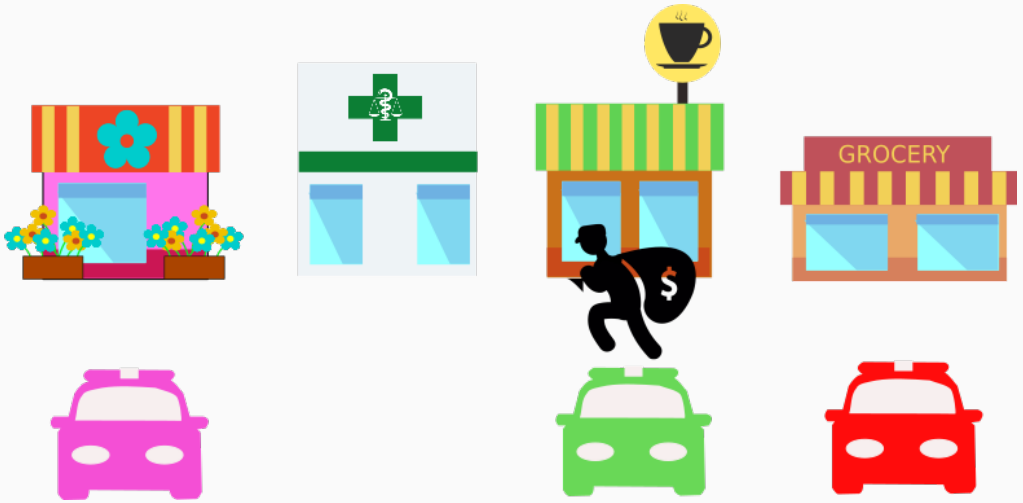
- **Stochastic Attacker** (Bounded Rationality - *not strategy-aware*)

$$\boldsymbol{\sigma}_{Sto}(\boldsymbol{\sigma}) = \mathbf{p}(Sto) \quad \forall \boldsymbol{\sigma} \in \Delta_M,$$

where $\mathbf{p}(Sto) \in \Delta_M$ is a probability distribution over the targets

Best Responding to a Profile

The **best response** for the defender against a profile is the *commitment* that **minimizes the expected loss**.



Realizations of different best response strategies have different colors.

State of the Art Algorithms

Online Learning Algorithms

These algorithms choose a best response based on the received loss:

- **Partial Feedback** (UCB1 and Thompson Sampling):

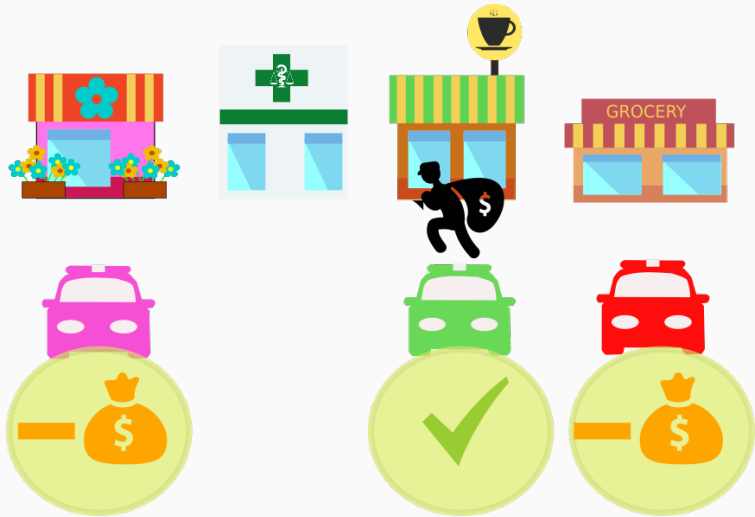
$$\lim_{N\to\infty} L_N \geq \log \mathbf{N} \sum_{k\neq k^*} \frac{\Delta L_k}{KL(\mathcal{L}_k, \mathcal{L}^*)}$$

(*logarithmic* regret)

- **Expert Feedback** (Follow the Perturbed Leader):

$$R_N(\mathfrak{U}) \propto \Delta L_k \mathbf{N}$$

(*linear* regret)



Proposed Algorithms

Current Action Likelihood

At the end of each round n we observe attacker action i, given our commitment σ_D.
We can then compute for each profile A_k the **current action likelihood**:

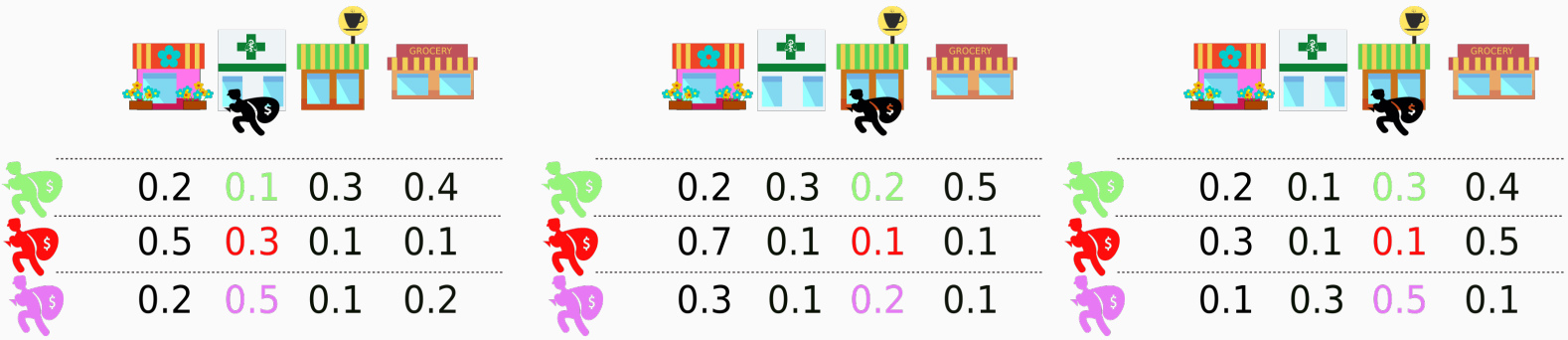
$$B_k(i, \sigma_D) := \sigma_{A_k}(\sigma_D)_i$$

0.2	0.1	0.3	0.4
0.5	0.3	0.1	0.1
0.2	0.5	0.1	0.2

Belief

Based on the previous observations it is possible to compute the **sequence likelihood**:

$$\Lambda_k^{(n)} := \prod_{j=1}^{n} B_k(i_j, s_j)$$



	0.2	0.1	0.3	0.4
	0.5	0.3	0.1	0.1
	0.2	0.5	0.1	0.2

$n = 1$

	0.2	0.3	0.2	0.5
	0.7	0.1	0.1	0.1
	0.3	0.1	0.2	0.1

$n = 2$

	0.2	0.1	0.3	0.4
	0.3	0.1	0.1	0.5
	0.1	0.3	0.5	0.1

$n = 3$

- $\Lambda_1^{(3)} = 0.1 \cdot 0.2 \cdot 0.3 = 0.006$
- $\Lambda_2^{(3)} = 0.3 \cdot 0.1 \cdot 0.1 = 0.006$
- $\Lambda_3^{(3)} = 0.5 \cdot 0.2 \cdot 0.5 = 0.05$

The sequence likelihood $\Lambda_k^{(n)}$ represents our **belief** on profile A_k.

Follow the Belief (FB)

- Identification of the most likely attacker profile through **Maximum Likelihood Estimation**
- FB best responds to the profile with the highest belief:

$$\boldsymbol{\sigma}_{FB}^{(n)} := \boldsymbol{\sigma}_{D}^{*}(\arg\max_{k|A_k \in \mathcal{A}} \Lambda_k^{(n)})$$

Example

In the previous situation, FB would have best responded to A_3

- $\Lambda_1^{(3)} = 0.1 \cdot 0.2 \cdot 0.3 = 0.006$
- $\Lambda_2^{(3)} = 0.3 \cdot 0.1 \cdot 0.1 = 0.006$
- $\Lambda_3^{(3)} = 0.5 \cdot 0.2 \cdot 0.5 = 0.05$

Theoretical Results

We recall the following results for state-of-the-art algorithms:

- Follow the perturbed leader (exploiting expert feedback):

$$R_N(\mathfrak{U}) \propto \Delta L_k \mathbf{N} \quad (\textit{linear} \text{ regret})$$

- MAB algorithm (exploiting bandit feedback):

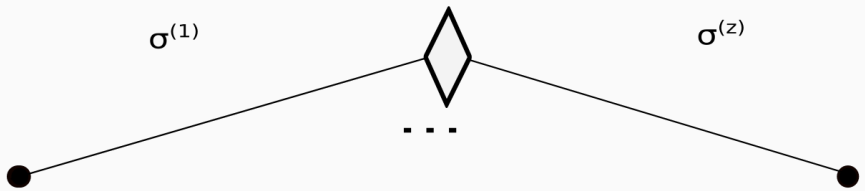
$$\lim_{N\to\infty} L_t \geq \log \mathbf{N} \sum_{k\neq k^*} \frac{\Delta L_k}{KL(\mathcal{L}_k, \mathcal{L}^*)} \quad (\textit{logarithmic} \text{ regret})$$

Instead, using FB with *known* profiles, under mild assumptions we obtain:

$$R_N(\mathfrak{U}) \leq \sum_{k=1}^{K} \frac{2(\lambda_k^2 + \lambda_{k^*}^2)\Delta L_k}{(\Delta b_k)^2} \quad (\textbf{constant} \text{ regret})$$

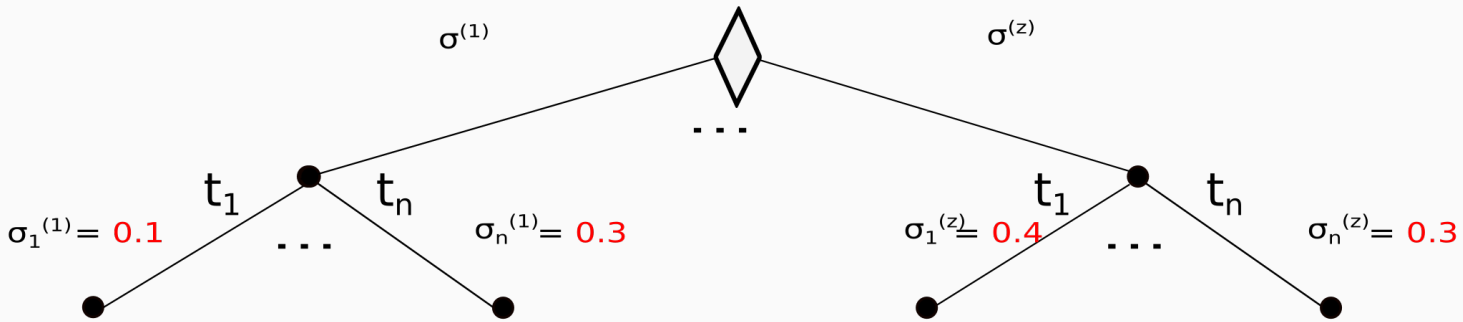
Follow the Regret (FR)

FR simulates execution paths for every choice up to $\mathbf{h_{max}}$ rounds. It recursively calls the function, weights results and chooses the best response that **minimizes the expected regret**.



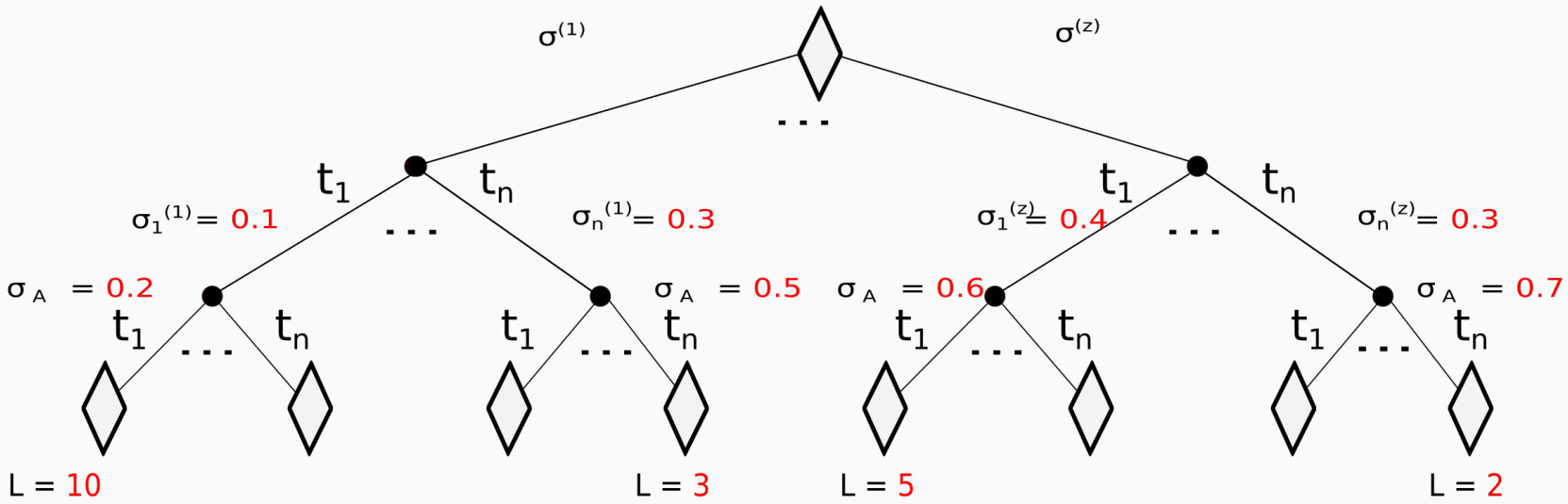
Follow the Regret (FR)

FR simulates execution paths for every choice up to $\mathbf{h_{max}}$ rounds. It recursively calls the function, weights results and chooses the best response that **minimizes the expected regret**.



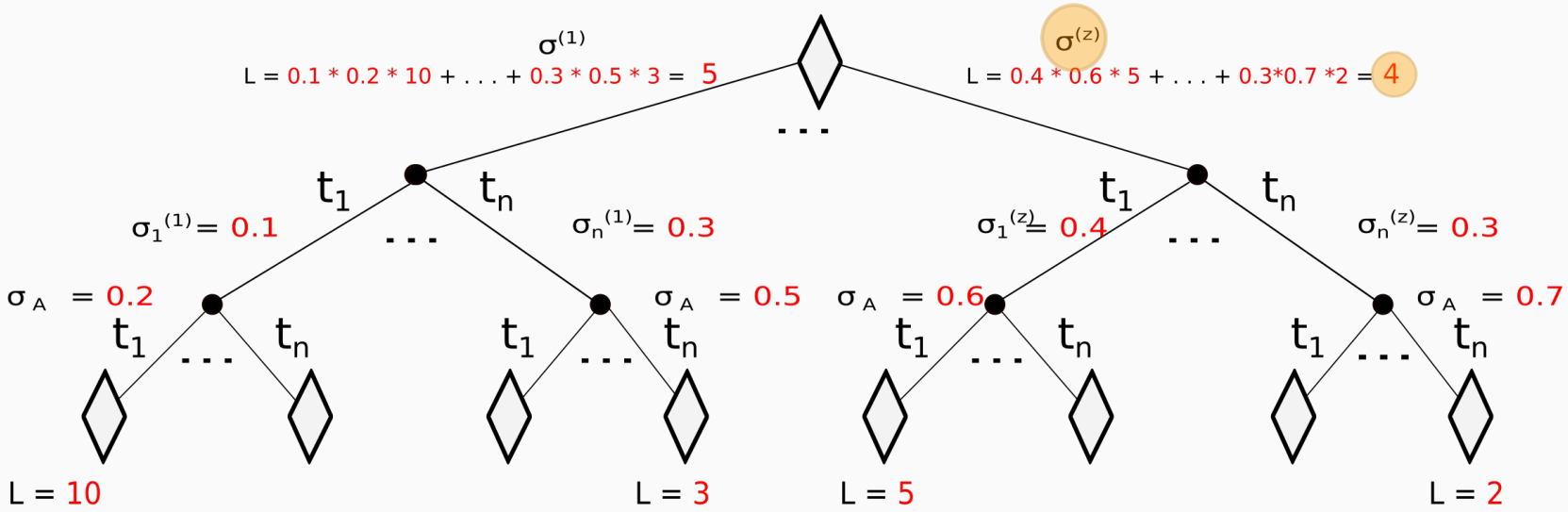
Follow the Regret (FR)

FR simulates execution paths for every choice up to $\mathbf{h_{max}}$ rounds. It recursively calls the function, weights results and chooses the best response that **minimizes the expected regret**.



Follow the Regret (FR)

FR simulates execution paths for every choice up to $\mathbf{h_{max}}$ rounds. It recursively calls the function, weights results and chooses the best response that **minimizes the expected regret**.



Experiments

Setting

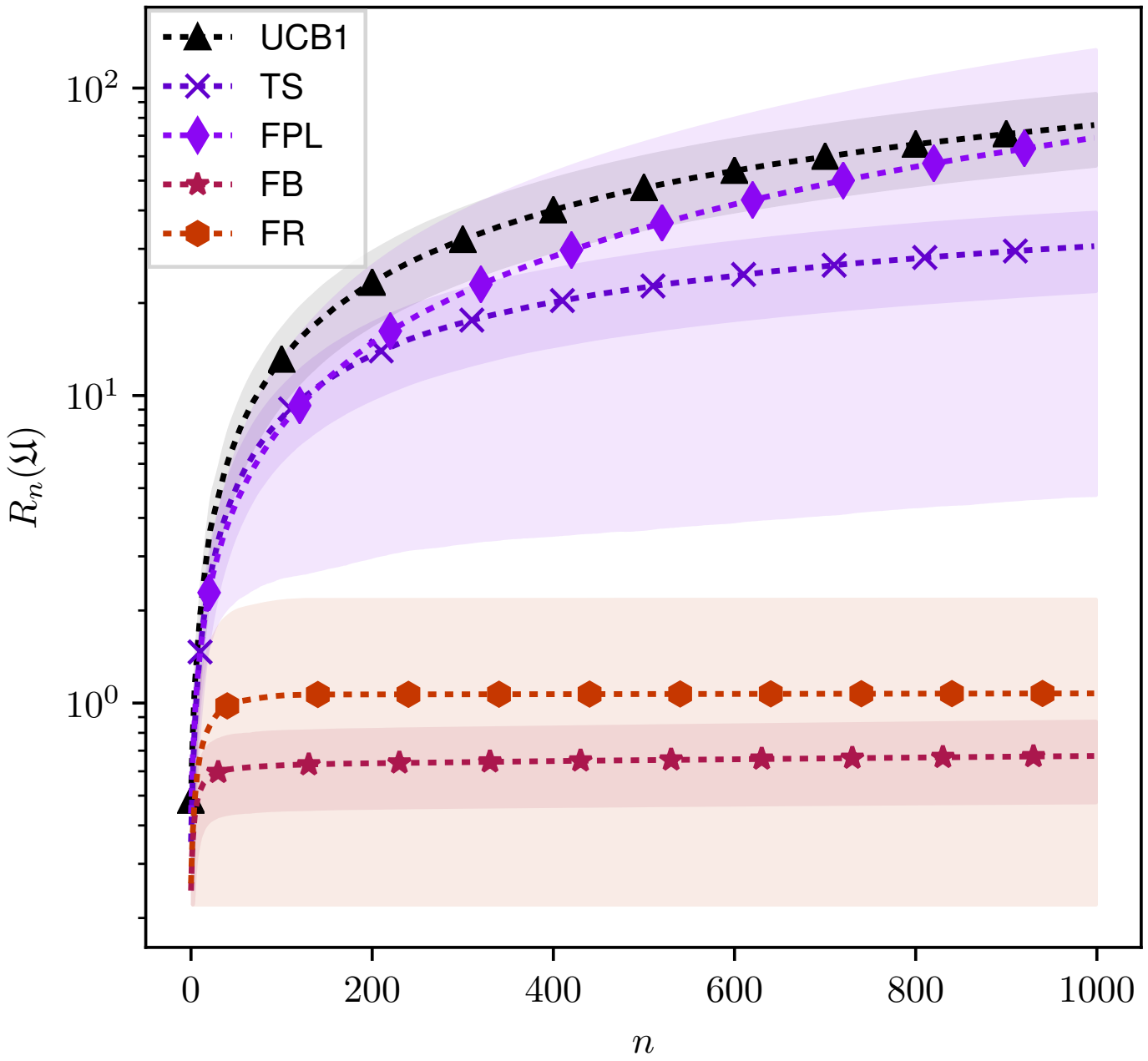
- Different combinations of profile types;
- For each combination, 10 FBI-SG configurations;
- For each configuration, 100 independent experiments;
- Time Horizon $N = 1000$;
- Performance Index: **expected pseudo regret**.

	Sta	*Sto*	*SUQR*	K
C_1	1	1	-	2
C_2	1	-	1	2
C_3	1	1	1	3
C_4	1	5	-	6
C_5	1	-	5	6
C_6	1	5	5	11

Some of the tested combinations.

We compared the results from applying state-of-the art **baselines** (Follow the Perturbed Leader (FPL), UCB1 and Thompson Sampling) and our **proposed solutions** (`FB` and `FR`).

Experiments - Combination C6



Introducing Unknown Profiles

Unknow Profiles

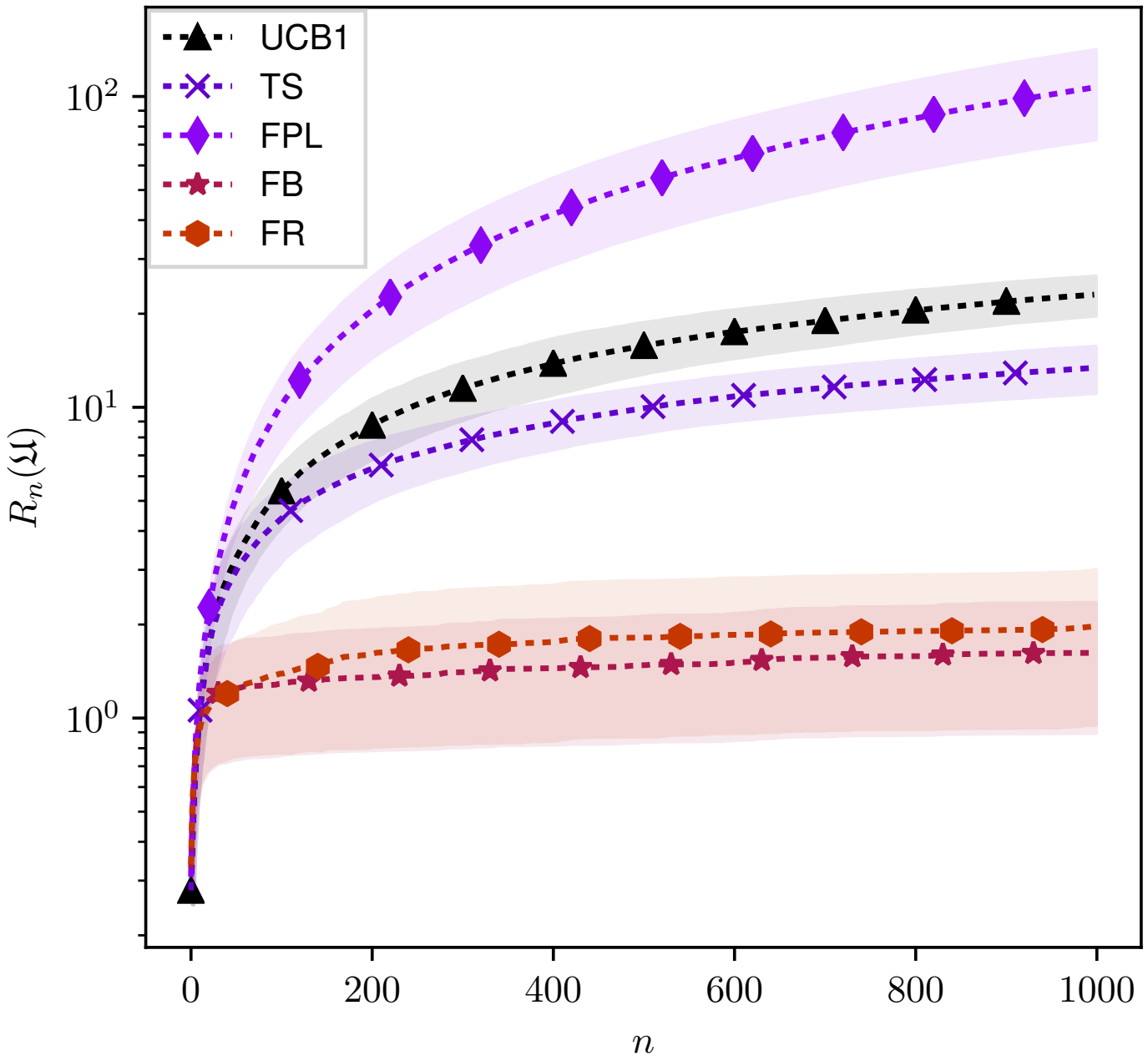
- **Unknown SUQR Attacker**: α and β are *unknown* to the defender.
- **Unkwown Stochastic Attacker**: she has a fixed distribution over the targets as a strategy, but it is *unknown* to the defender.

Using **Maximum Likelihood Estimation** we estimate the missing parameters and we can compute action *likelihoods* and *best responses*.

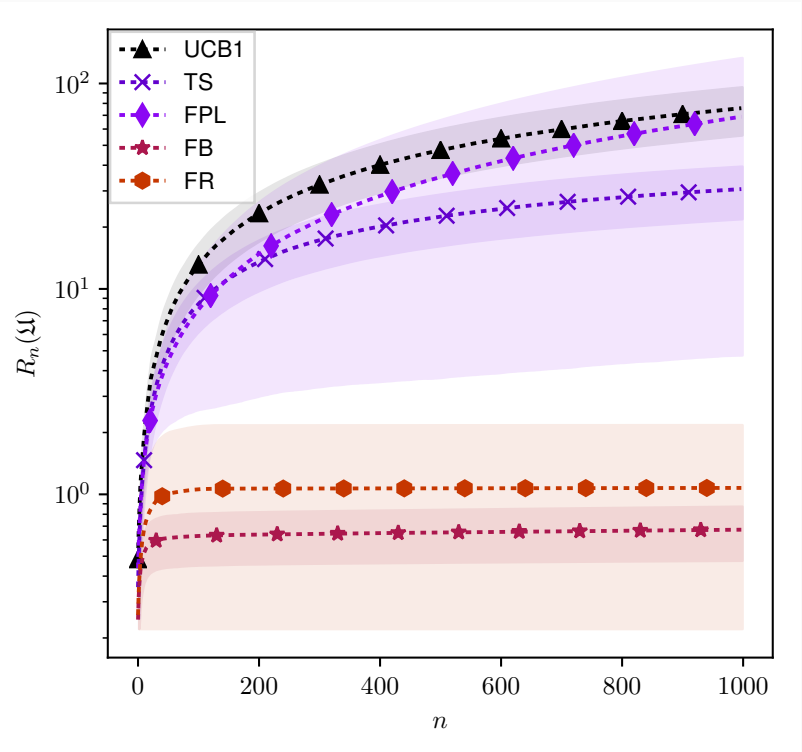
	Sta	*Sto*	*SUQR*	U_{Sto}	SU_{SUQR}	K
E_1	1	0	0	1	1	3
E_2	1	5	5	1	1	13

Sets of attacker profiles $\mathcal{A}$ used for the experiments.

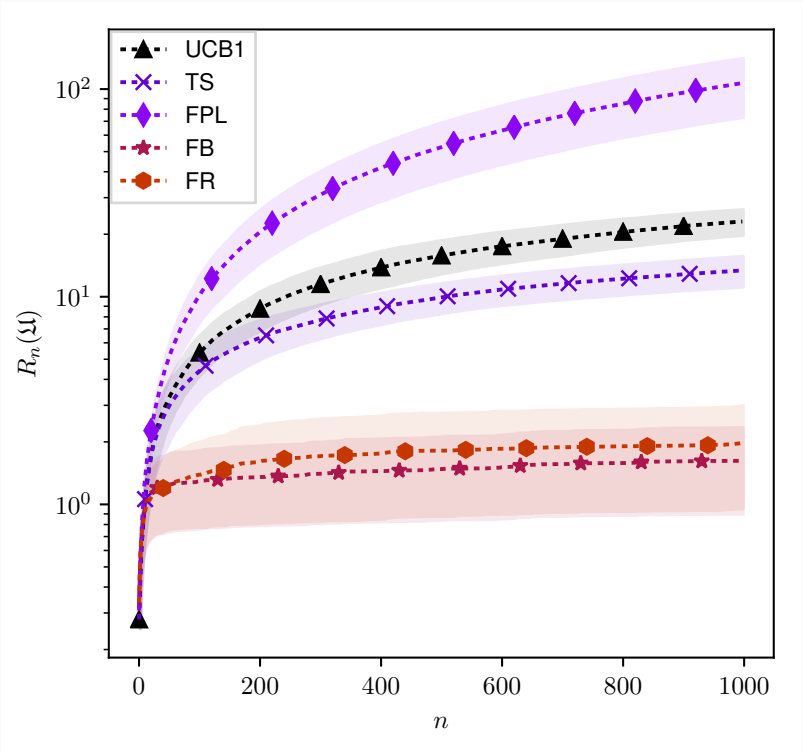
Experiments - Combination E1



The Value of Information



C6: $\mathcal{A} = \{Sta, Sto_1, \ldots, Sto_5, SUQR_1, \ldots, SUQR_5\}$



E1: $\mathcal{A} = \{Sta, USto, USUQR\}$

Conclusions and Future Work

- **Definition of a new problem**: FBI-SG.
- **Realizations of novel algorithms**: `FB` and `FR`.
- **Theoretical Results**: `FB` is always better than Expert and MAB algorithms with Known Profiles.
- **Experimental Results**: `FB` and `FR` outperform classical online learning algorithms.

A paper on this work has been accepted at UAI 2017.

Future Work:

- FB Bound for Unknown profiles
- Generalizing FR: policy gradient
- partial feedback
- multiple resources

Regret Minimization Algorithms for the Follower's Behavior Identification in Leadership Games

Relatore:
Prof. Marcello Restelli

Tesi di Laurea di:
Lorenzo Bisi

27/07/2017

Politecnico di Milano

Dipartimento di Elettronica Informazione e Bioingegneria

appendix1