A SUPPLEMENTAL MATERIAL

Theorem 2 (Expert pseudo-regret upper bound)**.** *Let us consider an instance of the FBI-SG problem and apply the FPL algorithm, where each possible profile A_k is an expert and receives, at round n, an expert reward equal to minus the loss she would have incurred observing $i_{A_{k^*},n}$ by playing the best response to the attacker A_k. Then, there always exists an attacker set $\mathcal{A}$ s.t. the defender D incurs in an expected pseudo-regret of:*

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

Proof. Let us analyse the FBI-SG problem in which the attacker profile set is $\mathcal{A} = \{Sta, Sto\}$, the true attacker $A_{k^*} = Sta$ and we use the Follow the Leader algorithm (Cesa-Bianchi and Lugosi, 2006). Assume that the best response $\boldsymbol{\sigma}^*_D(Sto)$ to the stochastic attacker Sto corresponds to the pure strategy played by the Stackelberg attacker at the equilibrium, i.e, $\boldsymbol{\sigma}^*_{Sta}(\boldsymbol{\sigma}^*_D(Sta)) = \boldsymbol{\sigma}^*_D(Sto)$. Assume the chosen target by the two strategies has value $v_{\hat{m}}$ in target $\hat{m}$, maximum value $v_{\bar{m}}$ in target $\bar{m}$ and that the stochastic attacker has strategy $\mathbf{p}$ s.t.:

$$p_m = \begin{cases} \alpha & \text{if } m = \hat{m} \\ 1-\alpha & \text{if } m = \bar{m} \\ 0 & \text{otherwise} \end{cases},$$

where $\alpha = \frac{v_{\bar{m}} - L(Sta)}{v_{\bar{m}}}$ and $\alpha v_{\bar{m}} > (1-\alpha) v_{\bar{m}}$. In this case, the defender might commit to two different strategies:

- if the defender D declares its best response to the Stackerlberg attacker $\boldsymbol{\sigma}^*_D(Sta)$ for the turn, it would provide zero loss as feedback for the stochastic attacker expert and loss equal to $-L(Sta)$ to the Stackelberg one
- if the defender D selects the best response to the stochastic attacker $\boldsymbol{\sigma}^*_D(Sto)$, the defender would gain loss equal to $-(1-\alpha)v_{\bar{m}} = -L(Sta)$ for the stochastic attacker expert and $-L(Sta)$ for the Stackelberg one. Thus, in this case the two types would receive the same feedback.

Summarizing, we have that the Stackelberg attacker expert always incurs in a loss greater or equal to the one of the stochastic one, even if the real attacker is Stackelberg. Thus, with a probability grater than 0.5 we are incurring in a loss of ΔL_k for the entire horizon, with a total regret proportional to $\Delta L_k N$. Even by resorting to randomization, thus even adopting the FPL we would have a probability of at least $0.5 - \varepsilon$ (being ε the probability with which the FPL chooses a suboptimal option) to select the wrong option, thus also the FPL algorithm would incur in a linear regret over the time horizon.

□

Theorem 3 (FB pseudo-regret upper bound)**.** *Given an instance of the FBI-SG problem s.t. $\Delta b_k > 0$ for each $A_k \in \mathcal{A}$ and applying FB, the defender incurs in a pseudo-regret of:*

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

where $\lambda_k := \max_{m \in \mathcal{M}} \max_{\boldsymbol{\sigma} \in \mathcal{S}} \ln(\sigma_{A_k}(\boldsymbol{\sigma})_m) - \min_{m \in \mathcal{M}} \min_{\boldsymbol{\sigma} \in \mathcal{S}} \ln(\sigma_{A_k}(\boldsymbol{\sigma})_m) \mathcal{I}\{\sigma_{A_k}(\boldsymbol{\sigma})_m \neq 0\}$ is the range where the logarithm of the beliefs realizations lies (excluding realizations equal to zero, which end the exploration of a profile) and $\mathcal{S} := \cup_k \boldsymbol{\sigma}^*_D(A_k)$ is the set of the available best response to the attackers profile.

Proof. Let us analyze the regret of the FB algorithm. We get some regret if the algorithm selects a strategy profile corresponding to a type different from the real one. Thus, the regret is upper bounded by:

$$\begin{aligned} R_N(\mathfrak{U}) &= \mathbb{E}\left[\sum_{n=1}^{N} l_n\right] - L^* N \\ &= \mathbb{E}\left[\sum_{n=1}^{N} l_n - L^*\right] = \sum_{k=1}^{K} \Delta L_k \mathbb{E}[T_k(N)], \end{aligned}$$

where we recall that:

- $T_i(N) = \sum_{n=1}^{N} \mathcal{I}\{A_{k_n} = A_k\}$ is the number of times we played the best response $\boldsymbol{\sigma}_D^*(A_k)$ to attacker A_k;

- $\Delta L_k = \sum_{m=1}^{M} \sigma_A(\boldsymbol{\sigma}_D^*(A_k))_m v_m(1 - \sigma_D^*(A_k)_m) \; - L^*$ is the expected regret of playing the best response to attacker A_k when the real attacker is A.

Each round in which the algorithm selects a profile s.t. the best response is not equal to the one of A_{k^*} we are getting some regret.

Let us define variables $B_{k,n}$ and $B_{k^*,n}$ denoting the belief we have for the possible attacker A_k and of the real attacker A, respectively, of the action played by the real attacker A at turn n. Moreover, let $b_{kj,t} := \mathbb{E}_{\sigma_D^*(A_j)}[B_{k,t}]$ be the expected value of the belief we get for attacker A_k when we are best responding to A_j and the true type is $A_{k^*} \neq A_k$ at round t. Note that $b_{kj,t} < b_{k^*j,t}, \forall j$, since Δb_k is positive.

For each profile $A_k \neq A_{k^*}$, we have:

$$\mathbb{E}[T_k(N)] \leq \sum_{n=1}^{N} \mathbb{E}\left[\mathcal{I}\left\{\prod_{t=1}^{n} B_{k,t} \geq \prod_{t=1}^{n} B_{k^*,t}\right\}\right] \tag{7}$$

$$\leq \sum_{n=1}^{N} \mathbb{E}\left[\mathcal{I}\left\{\sum_{t=1}^{n} \ln(B_{k,t}) \geq \sum_{t=1}^{n} \ln(B_{k^*,t})\right\}\right] \tag{8}$$

$$= \sum_{n=1}^{N} \mathbb{P}\left(\frac{\sum_{t=1}^{n} \ln(B_{k,t})}{n} \geq \frac{\sum_{t=1}^{n} \ln(B_{k^*,t})}{n}\right) \tag{9}$$

$$= \sum_{n=1}^{N} \mathbb{P}\left(\frac{\sum_{t=1}^{n} \ln(B_{k,t})}{n} - \frac{\sum_{t=1}^{n} \ln(b_{kj_t,t})}{n} - \frac{\sum_{t=1}^{n} \ln(B_{k^*,t})}{n} + \right.$$

$$\left. + \frac{\sum_{t=1}^{n} \ln(b_{k^*j_t,t})}{n} \geq \underbrace{\left(\frac{\sum_{t=1}^{n} \ln(b_{k^*j_t,t})}{n} - \frac{\sum_{t=1}^{n} \ln(b_{kj_t,t})}{n}\right)}_{\geq \Delta b_k}\right) \tag{10}$$

$$\leq \sum_{n=1}^{N} \mathbb{P}\left(\frac{\sum_{t=1}^{n} \ln(B_{k,t})}{n} - \frac{\sum_{t=1}^{n} \ln(b_{kj_t,t})}{n} - \frac{\Delta b_k}{2} - \frac{\sum_{t=1}^{n} \ln(B_{k^*,t})}{n} + \frac{\sum_{t=1}^{n} \ln(b_{k^*j_t,t})}{n} - \frac{\Delta b_k}{2} \geq 0\right) \tag{11}$$

$$\leq \underbrace{\sum_{n=1}^{N} \mathbb{P}\left(\frac{\sum_{t=1}^{n} \ln(B_{k,t})}{n} \geq \frac{\sum_{t=1}^{n} \ln(b_{kj_t,t})}{n} + \frac{\Delta b_k}{2}\right)}_{R_1} +$$

$$+ \underbrace{\sum_{n=1}^{N} \mathbb{P}\left(\frac{\sum_{t=1}^{n} \ln(B_{k^*,t})}{n} \leq \frac{\sum_{t=1}^{n} \ln(b_{k^*j_t,t})}{n} - \frac{\Delta b_k}{2}\right)}_{R_2}, \tag{12}$$

where j_t is the index of the attacker A_{j_t} we selected at round t and we defined $\Delta b_k := \min_{j|A_j \in \mathcal{A}} \ln(b_{k^*j,t}) - \ln(b_{kj,t})$, i.e., the minimum w.r.t. the best response for the available attackers of the difference between the expected value of the loglikelihood of attacker A_{k^*} and A_k if the true profile is A_{k^*}. Equation (9) has been obtained from Equation (8) since $\mathbb{E}\left[\mathcal{I}\left\{\cdot\right\}\right] = \mathbb{P}\left(\cdot\right)$ while Equation (10) has been computed from Equation (9) adding $\left(\frac{\sum_{t=1}^{n} \ln(b_{k^*j_t,t})}{n} - \frac{\sum_{t=1}^{n} \ln(b_{kj_t,t})}{n}\right)$ to both l.h.s. and r.h.s. of the inequality. We would like to point out that Δb_k does not depend on t since the distribution of $B_{k,t}$ and $B_{k^*,t}$ is the same over rounds.

Let us focus on R_1. We use the McDiarmid inequality (McDiarmid, 1989) to bound the probability that the empirical

estimate of the loglikelihood expected value is higher than a certain upper bound as follows:

$$\begin{aligned} R_1 &= \sum_{n=1}^{N} \mathbb{P}\left(\frac{\sum_{t=1}^{n} \ln(B_{k,t})}{n} \geq \frac{\sum_{t=1}^{n} \ln(b_{kj_t,t})}{n} + \frac{\Delta b_k}{2} \right) \\ &\leq \sum_{n=1}^{\infty} \mathbb{P}\left(\frac{\sum_{t=1}^{n} \ln(B_{k,t})}{n} \geq \frac{\sum_{t=1}^{n} \ln(b_{kj_t,t})}{n} + \frac{\Delta b_k}{2} \right) \\ &\leq \sum_{n=1}^{\infty} \exp\left\{ -\frac{(\Delta b_k)^2 n}{2\lambda_k^2} \right\} \leq \frac{2\lambda_k^2}{(\Delta b_k)^2}, \end{aligned}$$

where we exploited $\sum_{x=1}^{\infty} e^{-\kappa x} \leq \frac{1}{\kappa}$. We define $\lambda_k := \max_{m \in \mathcal{M}} \max_{\boldsymbol{\sigma} \in \mathcal{S}} \ln(\sigma_{A_k}(\boldsymbol{\sigma})_m) - \min_{m \in \mathcal{M}} \min_{\boldsymbol{\sigma} \in \mathcal{S}} \ln(\sigma_{A_k}(\boldsymbol{\sigma})_m) \mathcal{I}\{\sigma_{A_k}(\boldsymbol{\sigma})_m \neq 0\}$ as the range where the beliefs realizations lie (excluding realizations equal to zero which ends the exploration of a profile), where we used the fact that $\mathbb{E}[B_{k,t}] = b_k \forall k, t$ and $\mathcal{S} := \cup_k \boldsymbol{\sigma}_D^*(A_k)$ is the set of the available best response to the attackers profile.

A similar reasoning can be applied to R_2 getting an upper bound of the following form:

$$R_2 \leq \frac{2\lambda_{k^*}^2}{(\Delta b_k)^2}.$$

The regret becomes:

$$R_N(\mathfrak{U}) = \sum_{i=1}^{K} \Delta L_k \mathbb{E}[T_k(N)] \leq \sum_{i=1}^{K} \Delta L_k \left(\frac{2\lambda_k^2}{(\Delta b_k)^2} + \frac{2\lambda_{k^*}^2}{(\Delta b_k)^2} \right) \leq \sum_{i=1}^{K} \frac{2(\lambda_k^2 + \lambda_{k^*}^2)\Delta L_k}{(\Delta b_k)^2},$$

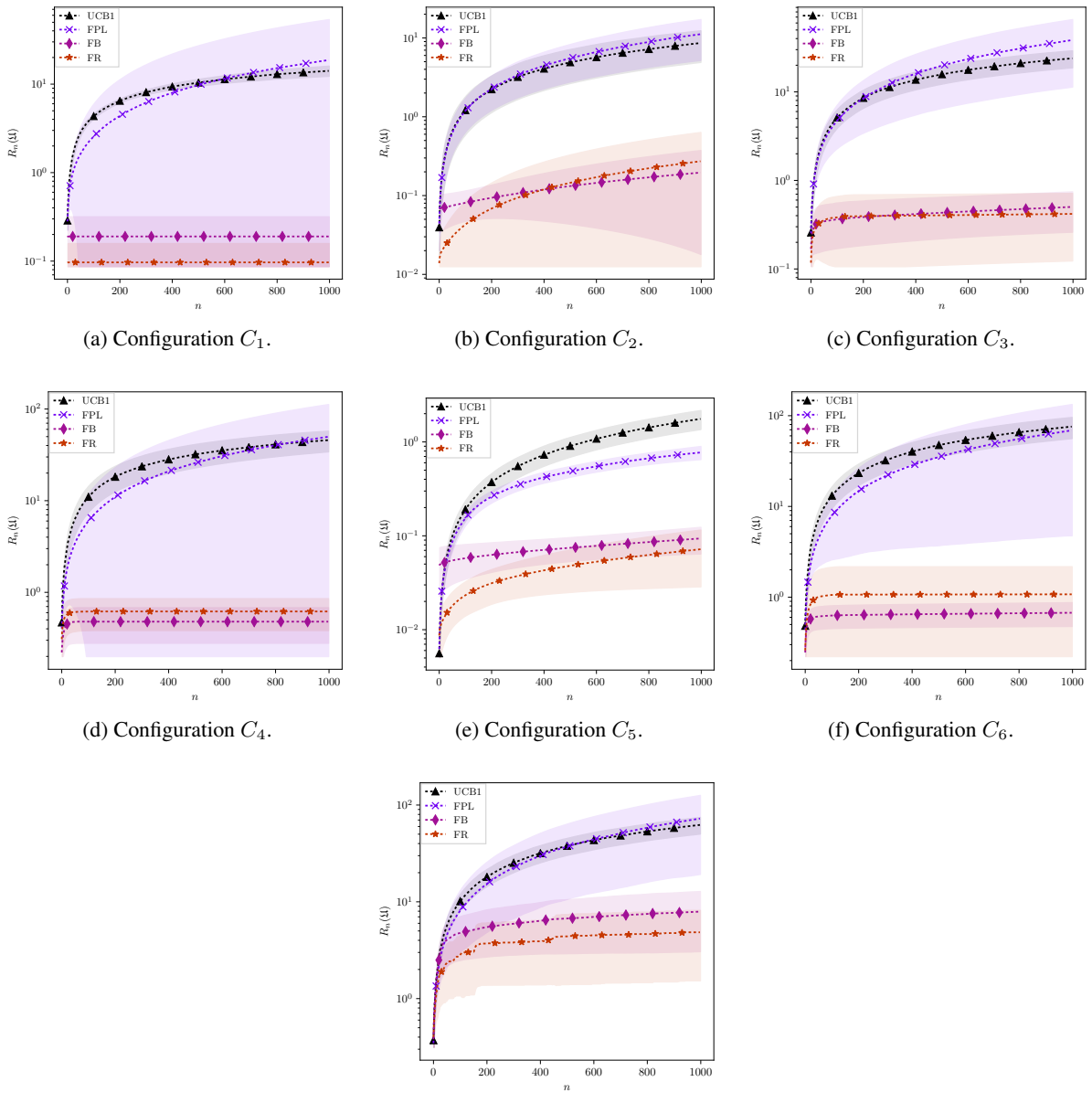
which concludes the proof.

□

B ADDITIONAL RESULTS

For the sake of completeness, we report in Figures 8 and 9 all the graphs regarding the regret for all the running configurations $C_1, \ldots, C_7$ and for the two dimensions of the target space, namely $M \in \{5, 10\}$. By inspecting these additional set of figures are in line with what has been presented in Section 6 of the main paper, where the proposed techniques, namely `FB` and `FR`, are able to outperform the literature methods. Even here, there is not a clear method providing statistical evidence that it is able to outperform the other.

Moreover, we also provide in Figure 10 the results for configuration C_6 with a number of target $M = 40$. In this configuration, we were able to run only the `FB` algorithm for computational time constraints. The results show that the `FB` has performance similar to the ones experienced with smaller target space, thus it is able to scale without significant loss in terms of expected pseudo-regret $R_N(\mathfrak{U})$.



(a) Configuration C_1.

(b) Configuration C_2.

(c) Configuration C_3.

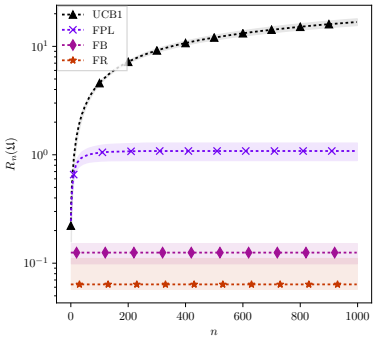
(d) Configuration C_4.

(e) Configuration C_5.

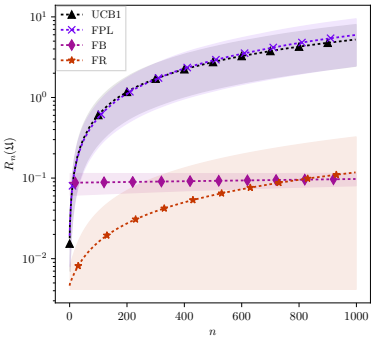
(f) Configuration C_6.

(g) Configuration C_7.

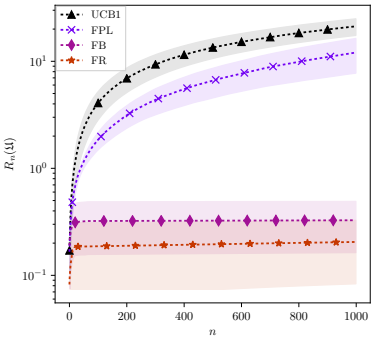
Figure 8: Expected pseudo-regret for the different configurations with $M = 5$ targets.



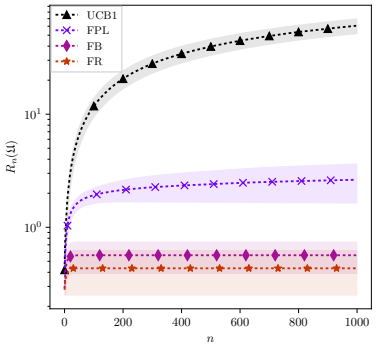
(a) Configuration C_1.



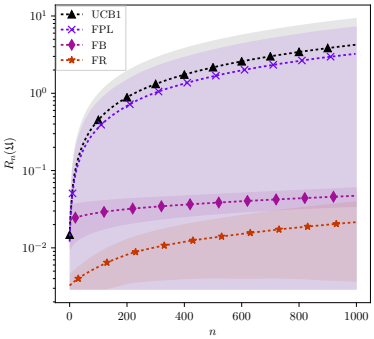
(b) Configuration C_2.



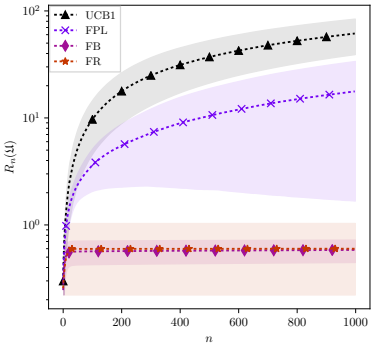
(c) Configuration C_3.



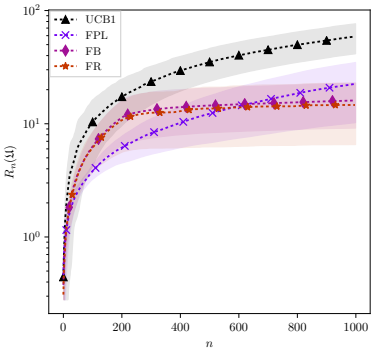
(d) Configuration C_4.



(e) Configuration C_5.



(f) Configuration C_6.



(g) Configuration C_7.

Figure 9: Expected pseudo-regret for the different configurations with $M = 10$ targets.

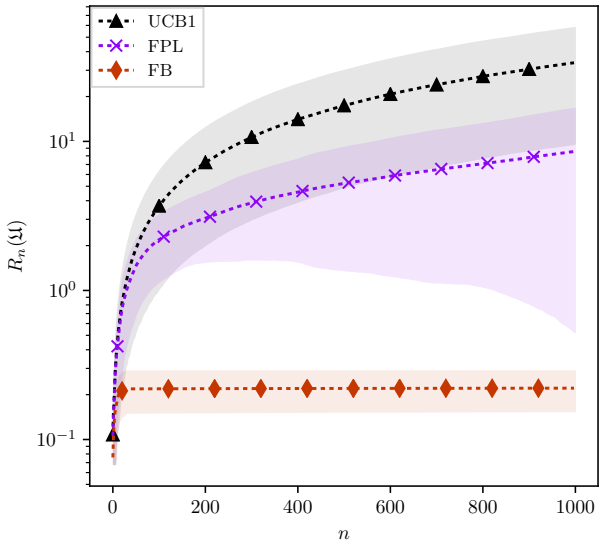


Figure 10: Expected pseudo-regret for the configuration C_6 with $M = 40$ targets.

Table 4: Computational time in seconds needed by FB and FR to solve an instance over $N = 1000$ rounds and the corresponding 95% confidence intervals.

		C_1	C_2	C_3	C_4	C_5	C_6	C_7
$M = 5$	FB	5.9 ± 1.7	11.1 ± 2.2	11.7 ± 2.9	3.5 ± 1.0	23.7 ± 2.4	14.9 ± 4.3	14.7 ± 3.2
	FR	77.0 ± 2.1	121.1 ± 3.2	170.4 ± 4.1	146.2 ± 4.7	651.7 ± 36.6	1029.2 ± 64.7	1113.7 ± 40.2
$M = 10$	FB	10.3 ± 2.6	21.9 ± 13.2	23.0 ± 17.9	7.1 ± 2.3	63.0 ± 7.4	47.22 ± 14.05	48.59 ± 13.48
	FR	356.1 ± 14.3	678.5 ± 15.9	887.0 ± 11.1	960.4 ± 13.0	4402.5 ± 14.2	7526.5 ± 189.9	7291.6 ± 23.7
$M = 20$	FB	33.5 ± 3.0	222.2 ± 126.9	137.8 ± 77.6	33.7 ± 1.2	484.5 ± 107.7	226.8 ± 45.3	229.5 ± 46.44
	FR	–	–	–	–	–	–	–
$M = 40$	FB	104.5 ± 7.1	2061.5 ± 837.2	1412.0 ± 812.1	128.9 ± 16.5	2347.9 ± 1223.2	1634.2 ± 487.6	1643.62 ± 468.8
	FR	–	–	–	–	–	–	–

We also report here Table 4, the full version of Table 3, with the time values up to the first decimal and also specifying the confidence interval.