7. Appendix

7.1. Positive mean-gap

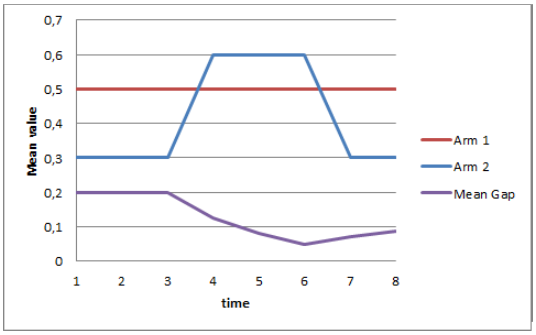

Figure 3: Mean gap versus time. Assumption 4 holds: the mean gap stays positive.

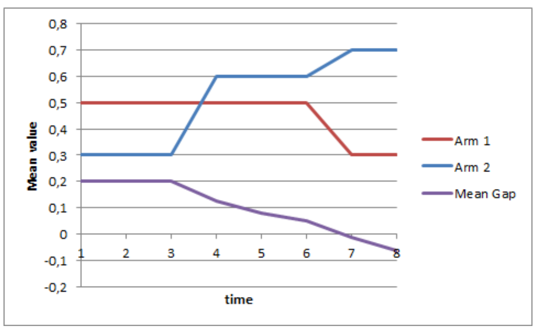

Figure 4: Mean gap versus time. Assumption 4 does not hold: a switch occurs a time 7.

Assumption 4 trivially holds when the mean rewards do not change. When the mean rewards change, Assumption 4 parts the small changes that do not imply a change of mean gap (see Figure 3) from major changes where the mean gap changes (see Figure 4). For more details see (Allesiardo et al, 2017).

7.2. Additional Experiments

MEDIAN ELIMINATION is designed to be order optimal in the worst case: its sample complexity is in $O(K \log \frac{1}{\delta})$. However, in practice it is clearly outperformed by SUCCESSIVE ELIMINATION or UGAPEC on both problems (see Figures 5, 6).

7.3. Proofs

Theorem 1. *Using any* ArmSelection *subroutine,* DECENTRALIZED ELIMINATION *is an* (ϵ, η)*-private algorithm, that finds an* ϵ*-optimal arm with a failure probability* $\delta \leq \eta^{\lfloor \frac{\log \delta}{\log \eta} \rfloor}$ *and that exchanges at most* $\lfloor \frac{\log \delta}{\log \eta} \rfloor K - 1$ *messages.*

Proof. The proof of Theorem 1 is composed of three parts.

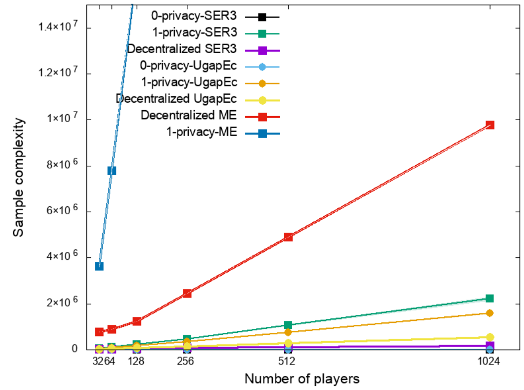

Figure 5: Problem 1:Uniform distribution of players

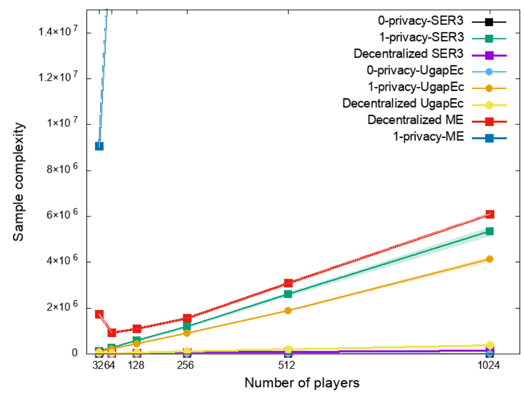

Figure 6: Problem 2:50% of players generates 80% of events

Part 1: (ϵ, η)**-privacy.** Let $E_{l^n} = \{\mathcal{K}^n(l^n) \cap \mathcal{K}_\epsilon = \emptyset\}$ be the event denoting that there is no ϵ-optimal arm in the remaining set of arm $\mathcal{K}^n(l^n)$ at epoch l^n, and $\neg E_{l^n}$ be the event denoting that there is at least an ϵ-optimal arm in the remaining set of arm $\mathcal{K}^n(l^n)$ at epoch l^n.

As DECENTRALIZED EXPLORATION $(\mathcal{A})$ performs an ArmSelection subroutine on each player, Property 1 ensures that for any player at epoch l^n:

$$\mathbb{P}\left(E_{l^n} | \mathcal{H}_{t^n}, \mathcal{A}, \neg E_{l^n}\right) \leq \eta \times f(l^n).$$

For the sake of simplicity, in the following we will omit the dependence on $\mathcal{A}$ of probabilities.

The message λ_k^n is sent by player n as soon as the arm k is eliminated from $\mathcal{K}^n(l^n)$ (see lines $17-18$ algorithm 2). Hence, we have:

$$\mathbb{P}\left(E_{l^n} | \mathcal{M}_n, \neg E_{l^n}\right) = \mathbb{P}\left(E_{l^n} | \mathcal{H}_{t^n(l^n)}, \neg E_{l^n}\right) \leq \eta \times f(l^n),$$

where $t^n(l^n)$ is the time where epoch l^n has begun.

To infer what arm is an ϵ-optimal arm for player n on the basis of $\mathcal{M}_n$ and $\mathcal{A}$, we first consider the favorable case for the adversary, where player n has sent $K-1$ elimination messages which corresponds to epoch $l^n = L$. Using Property 1 of the subroutine used by $\mathcal{A}$ and the set of messages $\mathcal{M}_n$ the adversary can infer that:

$$\mathbb{P}\left(\{\mathcal{K}^n(L) \not\subseteq \mathcal{K}_\epsilon\} | \neg E_{L-1}\right) = \sum_{l^n=1}^{L} \mathbb{P}\left(E_{l^n} | \mathcal{M}_n, \neg E_{l^n}\right) \leq \eta \sum_{l^n=1}^{L} f(l^n) = \eta.$$

The previous equality holds since if at epoch l^n the event $\{\mathcal{K}^n(l^n) \not\subseteq \mathcal{K}_\epsilon\}$ holds, then it holds also for all following epochs. Then the inequality is obtained by applying Property 1 to each element of the sum. Hence, if $l^n = L$ knowing the set of messages $\mathcal{M}_n$ and Property 1, the adversary cannot infer what arm is an ϵ-optimal arm for player n with a probability higher that $1-\eta$.

Otherwise if $l^n < L$ then $\mathcal{K}^n(L) \subset \mathcal{K}^n(l^n)$, which implies that:

$$\mathbb{P}\left(\{\mathcal{K}^n(l^n) \not\subseteq \mathcal{K}_\epsilon\} | \mathcal{M}_n, \neg E_{l^n}\right) \geq \mathbb{P}\left(\{\mathcal{K}^n(L) \not\subseteq \mathcal{K}_\epsilon\} | \mathcal{M}_n, \neg E_{L-1}\right).$$

Hence, if $l^n < L$ the adversary cannot infer what arm is an ϵ-optimal arm with a probability higher that $1-\eta$.

Part 2: Low probability of failure. An arm is eliminated when the events $\{k \notin \mathcal{K}^n(l^n)\}$ occur for $\lfloor \frac{\log \delta}{\log \eta} \rfloor$ independent players. Assumption 3 ($\forall n \in \mathcal{N}$, $P_x(x=n) \neq 0$) and Property 2 ensures that it exists a time $t = \sum_{n=1}^{N} t^n$ such that for $K-1$ arms, there are $\lfloor \frac{\log \delta}{\log \eta} \rfloor$ voting players. Moreover, Property 1 implies that $\forall n \in \mathcal{N}$, $\forall l^n$:

$$\mathbb{P}\left(\{\mathcal{K}^n(l^n) \not\subseteq \mathcal{K}_\epsilon\} | \mathcal{M}_n, \neg E_{l^n}\right) \leq \eta \times f(l^n).$$

Hence, the $\lfloor \frac{\log \delta}{\log \eta} \rfloor$ independent voting players eliminate the ϵ-optimal arm with a probability at most:

$$\mathbb{P}\left(\{\mathcal{K}(l) \not\subseteq \mathcal{K}_\epsilon\} | \mathcal{M}, \neg E_l\right) \leq (\eta \times f(l))^{\lfloor \frac{\log \delta}{\log \eta} \rfloor},$$

where $\mathcal{K}(l)$ denotes the shared set of remaining arms at elimination epoch l (see line 7 of Algorithm 2), and $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2 \cup ... \cup \mathcal{M}_N$.

If the algorithm fails, then the following event occurs : at stopping time, $\exists k \in \mathcal{K}(L), k \notin \mathcal{K}_\epsilon$. Using the union bound, we have:

$$\mathbb{P}\left(\{\mathcal{K}(L) \not\subseteq \mathcal{K}_\epsilon\} | \mathcal{M}, \neg E_{L-1}\right) \leq \sum_{l=1}^{L} (\eta \times f(l))^{\lfloor \frac{\log \delta}{\log \eta} \rfloor} \leq \eta^{\lfloor \frac{\log \delta}{\log \eta} \rfloor}.$$

Finally notice that:

$$\eta^{\lceil \frac{\log \delta}{\log \eta} \rceil} \leq \delta = \eta^{\frac{\log \delta}{\log \eta}} \leq \eta^{\lfloor \frac{\log \delta}{\log \eta} \rfloor}.$$

Part 3: Low communication cost. The index of each arm is sent to other players no more than once per player (see line 17 of the algorithm 2). When $\lfloor \frac{\log \delta}{\log \eta} \rfloor$ messages have been sent for an arm, this arm is eliminated for all players (see lines $4-9$ of the algorithm 2).

Thus $\lfloor \frac{\log \delta}{\log \eta} \rfloor (K-1)$ messages are sent to eliminate the suboptimal arms. Then, at most $\lfloor \frac{\log \delta}{\log \eta} \rfloor - 1$ messages have been sent for the remaining arm. Thus, the number of sent messages is at most $\lfloor \frac{\log \delta}{\log \eta} \rfloor K - 1$.

□

Theorem 2. *Using any* $\mathrm{ArmSelection}(\epsilon, \eta, \mathcal{K})$ *subroutine, with a probability higher than* $(1-\delta)\left(1 - I_{1-p^*}\left(T_{P_{x,\mathbf{y}}} - T_{P_{\mathbf{y}}}, 1 + T_{P_{\mathbf{y}}}\right)\right)^{\lfloor \frac{\log \delta}{\log \eta} \rfloor}$ DECENTRALIZED ELIMINATION *stops after:*

$$\mathcal{O}\left(\frac{1}{p^*}\left(T_{P_{\mathbf{y}}} + \sqrt{\frac{1}{2}\log\frac{1}{\delta}}\right)\right) \textit{ samples in } P_{x,\mathbf{y}},$$

where $I_a(b,c)$ denotes the incomplete beta function evaluated at a with parameters b, c.

Proof. Let T_n be the number of samples of player n at time $T_{P_{x,\mathbf{y}}}$ when the algorithm stops. T_n is a binomial law of parameters $T_{P_{x,\mathbf{y}}}, P_x(x=n)$. We have:

$$\mathbb{E}_{P_x}[T_n] = P_x(x=n) T_{P_{x,\mathbf{y}}}.$$

Let $\mathcal{B}_{\delta,\eta}$ be the set of players that have the $\lfloor \frac{\log \delta}{\log \eta} \rfloor$ highest T_n. The algorithm does not stop, if the following event occurs: $E_1 = \{\exists n \in \mathcal{B}_{\delta,\eta}, T_n < T_{P_{\mathbf{y}}}\}$.

Applying Hoeffding inequality, we have:

$$\mathbb{P}\left(T_n - P_x(x=n) T_{P_{x,\mathbf{y}}} \leq -\epsilon\right) \leq \exp(-2\epsilon^2)$$

When $\neg E_1$ occurs, $\forall n \in \mathcal{B}_{\delta,\eta}$ we have with a probability at most δ:

$$T_{P_{\mathbf{y}}} - P_x(x=n) T_{P_{x,\mathbf{y}}} \leq -\sqrt{\frac{1}{2}\log\frac{1}{\delta}}.$$

Then, when $\neg E_1$ occurs we have with a probability at most δ:

$$T_{P_{x,\mathbf{y}}} \geq \frac{1}{p_{\delta,\eta}}\left(T_{P_{\mathbf{y}}} + \sqrt{\frac{1}{2}\log\frac{1}{\delta}}\right),$$

where $p_{\delta,\eta} = \min_{n \in \mathcal{B}_{\delta,\eta}} P_x(x=n)$.

Finally if E_1 does not occur, then we have with a probability at least $1-\delta$:

$$T_{P_{x,\mathbf{y}}} \leq \frac{1}{p_{\delta,\eta}}\left(T_{P_{\mathbf{y}}} + \sqrt{\frac{1}{2}\log\frac{1}{\delta}}\right).$$

Let $\mathcal{N}_M$ bet the set of the $M = \lfloor \frac{\log \delta}{\log \eta} \rfloor$ most likely players. Let $n^* = \arg\min_{n \in \mathcal{N}_M} P_x(x = n)$, and $p^* = \min_{n \in \mathcal{N}_M} P_x(x = n)$.

Now, we consider the following event: $E_2 = \{n^* \notin \mathcal{B}_{\delta,\eta}\}$. By the definition of $\mathcal{B}_{\delta,\eta}$, the event E_2 is equivalent to the event $\{T_{n^*} < T_{P_{\mathbf{y}}}\}$. Then, we have:

$$\mathbb{P}\left(T_{n^*} < T_{P_{\mathbf{y}}}\right) = I_{1-p^*}\left(T_{P_{x,\mathbf{y}}} - T_{P_{\mathbf{y}}}, 1 + T_{P_{\mathbf{y}}}\right),$$

where $I_a(b, c)$ denotes the incomplete beta function evaluated at a with parameters b, c.

Finally, with a probability at least $(1 - I_{1-p^*}\left(T_{P_{x,\mathbf{y}}} - T_{P_{\mathbf{y}}}, 1 + T_{P_{\mathbf{y}}}\right))^{\lfloor \frac{\log \delta}{\log \eta} \rfloor}$, we have $p_{\delta,\eta} = p^*$.

□

Corollary 1. *With a probability higher than* $(1 - \delta)\left(1 - I_{1-p^*}\left(T_{P_{x,\mathbf{y}}} - T_{P_{\mathbf{y}}}, 1 + T_{P_{\mathbf{y}}}\right)\right)^{\lfloor \frac{\log \delta}{\log \eta} \rfloor}$ DECENTRALIZED MEDIIAN ELIMINATION *stops after:*

$$\mathcal{O}\left(\frac{1}{p^*}\left(\frac{K}{\lfloor \frac{\log \delta}{\log \eta} \rfloor \epsilon^2} \log \frac{1}{\delta} + \sqrt{\frac{1}{2} \log \frac{1}{\delta}}\right)\right) \textit{ samples in } P_{x,\mathbf{y}}.$$

Proof. We have:

$$\eta^{\lfloor \frac{\log \delta}{\log \eta} \rfloor} \leq \delta = \eta^{\frac{\log \delta}{\log \eta}} \leq \eta^{\lfloor \frac{\log \delta}{\log \eta} \rfloor}$$
$$\Rightarrow \frac{1}{\delta} \geq \frac{1}{\eta^{\lfloor \frac{\log \delta}{\log \eta} \rfloor}}$$
$$\Leftrightarrow \log \frac{1}{\eta} \leq \frac{1}{\lfloor \frac{\log \delta}{\log \eta} \rfloor} \log \frac{1}{\delta}$$

MEDIAN ELIMINATION algorithm (Even-Dar et al, 2006) finds an ϵ-optimal arm with a probability at least $1 - \eta$, and needs at most:

$$T_{P_{\mathbf{y}}} = \mathcal{O}\left(\frac{K}{\epsilon^2} \log \frac{1}{\eta}\right) \leq \mathcal{O}\left(\frac{K}{\lfloor \frac{\log \delta}{\log \eta} \rfloor \epsilon^2} \log \frac{1}{\delta}\right) \text{ samples in } P_{\mathbf{y}}$$

Then the use of Theorem 2 finishes the proof.

□

Corollary 2. *With a probability higher than* $(1 - \delta)\left(1 - I_{1-p^*}\left(T_{P_{x,\mathbf{y}}} - T_{P_{\mathbf{y}}}, 1 + T_{P_{\mathbf{y}}}\right)\right)^{\lfloor \frac{\log \delta}{\log \eta} \rfloor}$ DECENTRALIZED SUCCESSIVE ELIMINATION *stops after:*

$$\mathcal{O}\left(\frac{1}{p^*}\left(\frac{K}{\epsilon^2}\left(\log K + \frac{1}{\lfloor \frac{\log \delta}{\log \eta} \rfloor} \log \frac{1}{\delta}\right) + \sqrt{\frac{1}{2} \log \frac{1}{\delta}}\right)\right)$$

samples in $P_{x,\mathbf{y}}$.

Proof. SUCCESSIVE ELIMINATION algorithm (Even-Dar et al, 2006) finds an ϵ-optimal arm with a probability at least $1 - \eta$, and needs at most:

$$T_{P_{\mathbf{y}}} = \mathcal{O}\left(\frac{K}{\epsilon^2} \log \frac{K}{\eta}\right) \leq \mathcal{O}\left(\frac{K}{\epsilon^2}\left(\log K + \frac{1}{\lfloor \frac{\log \delta}{\log \eta} \rfloor} \log \frac{1}{\delta}\right)\right)$$

samples in $P_{x,\mathbf{y}}$. Then the use of Theorem 2 finishes the proof. □

Theorem 3. *For $K \geq 2$, $\delta \in (0, 0.5]$, for the sequences of rewards where Assumption 4 holds, DSER3 is an (ϵ, η)-private algorithm, that exchanges at most $\lfloor \frac{\log \delta}{\log \eta} \rfloor K - 1$ messages, that finds an ϵ-optimal arm with a probability at least $(1 - \delta)\left(1 - I_{1-p^*}\left(T_{P_{x,\mathbf{y}}} - T_{P_{\mathbf{y}}}, 1 + T_{P_{\mathbf{y}}}\right)\right)^{\lfloor \frac{\log \delta}{\log \eta} \rfloor}$, and that stops after:*

$$\mathcal{O}\left(\frac{1}{p^*}\left(\frac{K}{\epsilon^2}\left(\log K + \frac{1}{\lfloor \frac{\log \delta}{\log \eta} \rfloor} \log \frac{1}{\delta}\right) + \sqrt{\frac{1}{2} \log \frac{1}{\delta}}\right)\right)$$

samples in $P_{x,\mathbf{y}}$.

Proof. Theorem 3 is a straightforward application of Theorem 2, where $T_{P_{\mathbf{y}}}$ is stated in Theorem 1 (Allesiardo et al, 2017). □

Theorem 4. *For $K \geq 2$, $\epsilon \geq \frac{\eta}{K}$, $\varphi \in (0, 1]$, for any sequences of rewards that can be splitted into sequences where Assumption 4 holds,* DSER4 *is an (ϵ, η)-private algorithm, that exchanges on average at most $\varphi T(\lfloor \frac{\log \delta}{\log \eta} \rfloor K - 1)$ messages, and that plays, with an expected probability at most $\delta + \varphi T I_{1-p^*}\left(T_{P_{x,\mathbf{y}}} - T_{P_{\mathbf{y}}}, 1 + T_{P_{\mathbf{y}}}\right)^{\lfloor \frac{\log \delta}{\log \eta} \rfloor}$, a suboptimal arm on average no more than:*

$$\mathcal{O}\left(\frac{1}{p^*}\left(\frac{1}{\epsilon^2}\sqrt{\frac{SK \log K + \frac{1}{\lfloor \frac{\log \delta}{\log \eta} \rfloor} \log \frac{1}{\delta}}{\delta^{\frac{1}{\lfloor \frac{\log \delta}{\log \eta} \rfloor}}}} + \sqrt{\frac{1}{2} \log \frac{1}{\delta}}\right)\right)$$

times, where S is the number of switches of best arms, φ is the probability of reset in SER4*, T is the time horizon, and the expected values are taken with respect to the randomization of resets.*

Proof. The upper bound of the expected number of times a suboptimal arm is played by SER4, is stated in Corollary 2 (Allesiardo et al, 2017). Then this upper bound is used in Theorem 2 to state the upper bound of the expected number of times a suboptimal arm is played using DSER4. The expected number of resets is φT. Theorem 2 provides the success probability of each run of DECENTRALIZED ELIMINATION, which states the expected failure probability of DSER4. Then using Theorem 1 the expected upper bound of the number of exchanged messages is stated. □