A. Notations

$[n]$	set $\{1, \cdots, n\}$ for any $n \in \mathbb{N}$
$[L]$	ground set of size L
$\nu(i)$	reward distribution of item $i \in [L]$
$w(i)$	mean reward of item $i \in [L]$
$W_t(i)$	random reward of item i at time step t
$c_t(i)$	corruption added on random reward item i at time step t
$\tilde{W}_t(i)$	corrupted reward of item i at time step t
i_t	pulled item at time step t
C	total corruption budget
$\mathbb{P}$	probability law of the process $\{\tilde{\mathbf{W}}_t = (\tilde{W}_t(1), \ldots, \tilde{W}_t(L))\}_{t=1}^T$
$\Delta_{1,i}$	gap between mean rewards of item 1 and i
ϵ	optimality gap of item
π	non-anticipatory algorithm
i_t^π	pulled item of algorithm π at time step t
$i_{\text{out}}^{\pi,T}$	output of algorithm π
$\phi^{\pi,T}$	final recommendation rule of algorithm π
$\mathcal{F}_t$	observation history
ϵ_C	bound on $\Delta_{1,i_{\text{out}}^{\pi,T}}$
δ	failure probability
u	parameter in Algorithm 1
M	amount of phases in Algorithm 1
N	length of one phase in Algorithm 1
A_m	active set in Algorithm 1
q_m	probability to pull an active item during phase m in Algorithm 1
n_m	expected number of pulls of an active item during phase m in Algorithm 1
$\hat{w}_m(i)$	corrupted empirical mean of item i during phase m in Algorithm 1
$\tilde{H}_2(w, L, u)$	difficulty of the instance $\{w(i)\}_{i=1}^L$ for PSS(u)
$H_2(w)$	intrinsic difficulty of the instance $\{w(i)\}_{i=1}^L$
$\text{Bern}(a)$	Bernoulli distribution with parameter $a \in [0, 1]$
L_ϵ	number of item i with $\Delta_{1,i} \leq \epsilon$
Δ	equals to $\Delta_{1,2}$
λ	parameter in the analysis of corruption strategies
C_m	amount of corruptions during phase m
$\mathcal{E}_{m,i}^{(\text{U})}(a), \mathcal{E}_{m,i}^{(\text{L})}(a)$	"nice events" in the analysis of Algorithm 1
m_1	index of the phase during which item 1 turns from active to inactive
j_1	item in A_{m_1} with the least mean reward

a_i	equals to $\Delta_{1,i}/8$ for all items $2 \le i \le L$

B. Useful theorems

Theorem B.1 (Standard multiplicative variant of the Chernoff-Hoeffding bound; Dubhashi & Panconesi (2009), Theorem 1.1)**.** *Suppose that $X_1, \ldots, X_T$ are independent $[0,1]$-valued random variables, and let $X = \sum_{t=1}^{T} X_t$. Then for any $\epsilon > 0$,*

$$\Pr[\, X - \mathbb{E}X \ge \epsilon \mathbb{E}X \,] \le \exp\left(-\frac{\epsilon^2}{3}\mathbb{E}X\right), \quad \Pr[\, X - \mathbb{E}X \le -\epsilon \mathbb{E}X \,] \le \exp\left(-\frac{\epsilon^2}{3}\mathbb{E}X\right).$$

Theorem B.2 (Beygelzimer et al. (2011), Theorem 1; Gupta et al. (2019), Theorem 10)**.** *Suppose that $X_1, \ldots, X_T$ is a martingale difference sequence with respect to a filtration $\{\mathcal{F}_t\}_{t=1}^{T}$, and let $X = \sum_{t=1}^{T} X_t$. Assume that $|X_t| \le b$ for all t, and define $V = \sum_{t=1}^{T} \mathbb{E}[X_t^2 | \mathcal{F}_{t-1}]$. Then for any $\delta > 0$,*

$$\Pr\left[X \le \frac{V}{b} + b \ln \frac{1}{\delta}\right] \ge 1 - \delta.$$

Theorem B.3 (Multiplicative Chernoff Bound (Mitzenmacher & Upfal, 2017; Chen et al., 2016))**.** *Let $X_1, \cdots, X_n$ be Bernoulli random variables taking values in $\{0,1\}$ such that $\mathbb{E}[X_t | X_1, \cdots, X_{t-1}] \ge \mu$ for all $t \le n$, and $Y = X_1 + \ldots + X_n$. Then, for all $\delta \in (0,1)$*

$$\Pr[Y \le (1-\delta)n\mu] \le e^{-\frac{\delta^2 n\mu}{2}}.$$

C. Proofs of main results

In this section, we provide proofs of Lemmas 5.1 – 5.3, complete the proof of Theorem 4.1, and provide the proofs of Theorem 4.2 – 4.4.

C.1. Proof of Lemma 5.1

Lemma 5.1. *It holds that $NM \le T$ and $|A_M| = 1$.*

Proof. (i) $NM = \lfloor T/M \rfloor \cdot M \le T/m \cdot M = T$.

(ii) Since $M = \lceil \log_u L \rceil$, $|A_M| = \lceil L/u^M \rceil$, we have

$$L \le u^M \;\Rightarrow\; |A_M| \le \left\lceil \frac{u^M}{u^M} \right\rceil = 1, \qquad \frac{L}{u^{M-1}} > 0 \;\Rightarrow\; |A_M| \ge 1.$$

□

C.2. Proof of Lemma 5.2

Lemma 5.2. *Let $\overline{\mathcal{E}}$ denote the complement of any event $\mathcal{E}$. For any fixed $m, i \in A_{m-1}$ and $a \in (0,1)$,*

$$\mathbb{P}\big[\overline{\mathcal{E}_{m,i}^{(\mathrm{U})}(a)}\big] \le 2\exp\Big[-\tfrac{a^2 n_m}{3}\Big],\ \mathbb{P}\big[\overline{\mathcal{E}_{m,i}^{(\mathrm{L})}(a)}\big] \le 2\exp\Big[-\tfrac{a^2 n_m}{3}\Big].$$

Proof. (i) Let $Y_t(i)$ be an indicator for item i being pulled at time step t and $\tilde{n}_m(i)$ be the number of pulls of item i during phase m. Recall that $W_t(i)$ is the stochastic reward of item i at time step t and $c_t(i) = \tilde{W}_t(i) - W_t(i)$ is the corruption added to this item by the adversary at this time step. Note that $c_t(i)$ may depend on all the stochastic rewards up to (and including) time step t, and also on all previous choices of the algorithm (though not the choice at step t). We denote $E_m := [T_{m-1}+1, \ldots, T_m]$ as the N many time steps in phase m. Then

$$\hat{w}_m(i) = \frac{1}{n_m} \sum_{t \in E_m} Y_t(i)[W_t(i) + c_t(i)].$$

For ease of analysis, let us break the sum above into two, and define

$$A_m(i) = \sum_{t \in E_m} Y_t(i) W_t(i) \quad \text{and} \quad B_m(i) = \sum_{t \in E_m} Y_t(i) c_t(i).$$

(ii) Let us first bound the deviation of $A_m(i)$. Observe that $W_t(i)$ is an independent draw from a $[0,1]$-valued r.v. with mean $w(i)$ and $Y_t(i)$ is an independent random variable drawn from $\{0,1\}$ with mean q_m. Moreover, we have that $\mathbb{E}[A_m(i)] = N \cdot [q_m w(i)] = n_m w(i) \le n_m$. Hence, for any $a_{1,m,i} > 0$, a Chernoff-Hoeffding bound (a multiplicative version thereof) as in Theorem B.1 implies that

$$\mathbb{P}\left[\frac{A_m(i)}{n_m} - w(i) \ge a_{1,m,i}\right] \le \exp\left[-\frac{a_{1,m,i}^2 \cdot n_m}{3}\right], \quad \mathbb{P}\left[\frac{A_m(i)}{n_m} - w(i) \le -a_{1,m,i}\right] \le \exp\left[-\frac{a_{1,m,i}^2 \cdot n_m}{3}\right].$$

(iii) Next, we turn to bound the deviation of $B_m(i)$. Consider the sequence of r.v.s $X_1, \ldots, X_T$ defined by $X_t = [Y_t(i) – q_m] \cdot c_t(i)$ for all t. Then $\{X_t\}_{t=1}^T$ is a martingale difference sequence with respect to the filtration $\{\tilde{\mathcal{F}}_t\}_{t=1}^T$, where

$$\tilde{\mathcal{F}}_t = \sigma(\{Y_s(i)\}_{s \le t, i \in [L]}, \{W_s(i)\}_{s \le t+1, i \in [L]}, \{c_s(i)\}_{s \le t+1, i \in [L]}).$$

According to the problem setup, the adversary obtains more information than the agent, which results in the difference between $\mathcal{F}_t$ defined in Section 2 and $\tilde{\mathcal{F}}_t$ here. Since the corruption $c_t(i)$ becomes a deterministic value when conditioned on $\tilde{\mathcal{F}}_{t-1}$ (as we assume a deterministic adversary), and since $\mathbb{E}[Y_t(i)|\tilde{\mathcal{F}}_{t-1}] = q_m$, we have

$$\mathbb{E}[X_t|\tilde{\mathcal{F}}_{t-1}] = \mathbb{E}[Y_t(i) – q_m|\tilde{\mathcal{F}}_{t-1}] \cdot c_t(i) = 0.$$

Further, we have $|X_t|, |c_t(i)| \le 1$ for all t, and we can bound the predictable quadratic variation of this martingale as

$$\begin{aligned} V &= \sum_{t \in E_m} \mathbb{E}[X_t^2|\tilde{\mathcal{F}}_{t-1}] = \sum_{t \in E_m} \mathbb{E}\big[\,[Y_t(i) – q_m]^2|\tilde{\mathcal{F}}_{t-1}\big] \cdot c_t(i)^2 \le \sum_{t \in E_m} |c_t(i)| \cdot \mathbb{E}\big[\,[Y_t(i) – q_m]^2|\tilde{\mathcal{F}}_{t-1}\big] \\ &= \sum_{t \in E_m} |c_t(i)| \cdot \mathrm{Var}[Y_t(i)] = \sum_{t \in E_m} |c_t(i)| \cdot q_m \cdot (1 - q_m) \le q_m \cdot \sum_{t \in E_m} |c_t(i)|. \end{aligned}$$

Applying a Freedman-type concentration inequality for martingales (Theorem B.2), we obtain that except with probability $\delta_{2,m,i}$ (setting $b = 1$ in Theorem B.2),

$$\frac{B_m(i)}{n_m} \le \frac{q_m}{n_m} \cdot \sum_{t \in Em} |c_t(i)| + \frac{V + \log(1/\delta_{2,m,i})}{n_m} \le \frac{2q_m}{n_m} \cdot \sum_{t \in Em} |c_t(i)| + \frac{\log(1/\delta_{2,m,i})}{n_m}.$$

Since $q_m = n_m/N$, We have

$$\mathbb{P}\left[\frac{B_m(i)}{n_m} \ge \frac{2\sum_{t \in E_m} |C_t(i)|}{N} + \frac{\log(1/\delta_{2,m,i})}{n_m}\right] \le \delta_{2,m,i}.$$

Similar arguments show that $-B_m(i)/n_m$ satisfies this bound with probability $\delta_{2,m,i}$.

(iv) Let

$$a_{2,m,i} = \frac{\log(1/\delta_{2,m,i})}{n_m}.$$

Altogether, we have

$$\mathbb{P}\left[\hat{w}_m(i) \ge w(i) + \frac{2\sum_{t \in E_m} |C_t(i)|}{N} + a_{1,m,i} + a_{2,m,i}\right] \le \exp\left[-\frac{a_{1,m,i}^2 \cdot n_m}{3}\right] + \exp[-a_{2,m,i} \cdot n_m],$$

$$\mathbb{P}\left[\hat{w}_m(i) \le w(i) - \frac{2\sum_{t \in E_m} |C_t(i)|}{N} - a_{1,m,i} - a_{2,m,i}\right] \le \exp\left[-\frac{a_{1,m,i}^2 \cdot n_m}{3}\right] + \exp[-a_{2,m,i} \cdot n_m].$$

Note that $\sum_{t\in E_m}|c_t(i)|\le C_m$. For $a_{1,m,i}=a_{2,m,i}=a\in(0,1)$, we have

$$\mathbb{P}\left[\hat{w}_m(i)\ge w(i)+\frac{2C_m}{N}+2a\right]\le 2\exp\left[-\frac{a^2\cdot n_m}{3}\right],$$
$$\mathbb{P}\left[\hat{w}_m(i)\le w(i)-\frac{2C_m}{N}-2a\right]\le 2\exp\left[-\frac{a^2\cdot n_m}{3}\right].$$

□

C.3. Proof of Lemma 5.3

Lemma 5.3. *Conditioned on $\mathcal{E}^{(\mathrm{L})}_{m,1}(a_i)$ and $\mathcal{E}^{(\mathrm{U})}_{m,i}(a_i)$, where $a_i=\Delta_{1,i}/8$ for each $2\le i\le L$, we have*

$$\{1\in A_{m-1},1\notin A_m,i\in A_m\}\subset\left\{\Delta_{1,i}\le\frac{8C_m}{N}\right\}.$$

Proof. First of all,

$$\begin{aligned}\{1\in A_{m-1},1\notin A_m,i\in A_m\}&\subset\{1,i\in A_{m-1},i\in A_m,\hat{w}_m(1)\le\hat{w}_m(j)\ \ \forall j\in A_m\}\\&\subset\{1,i\in A_{m-1},\hat{w}_m(1)\le\hat{w}_m(i)\}.\end{aligned}$$

Assume $\mathcal{E}^{(\mathrm{L})}_{m,1}(a_i)$ and $\mathcal{E}^{(\mathrm{U})}_{m,i}(a_i)$ hold. We have

$$w(1)-\frac{2C_m}{N}-2a_i<\hat{w}_m(1)\le\hat{w}_m(i)<w(i)+\frac{2C_m}{N}+2a_i.$$

In other words,

$$w(1)-w(i)<\frac{4C_m}{N}+4a_i.$$

Note that $a_i=\Delta_{1,i}/8$, we have

$$\Delta_{1,i}<\frac{4C_m}{N}+\frac{\Delta_{1,i}}{2}\ \Rightarrow\ \Delta_{1,i}<\frac{8C_m}{N}$$

as desired.

□

C.4. Final steps to prove Theorem 4.1

(i) Assume $\mathcal{E}^{(\mathrm{L})}_{m_1,1}(a_{j_1})$ and $\mathcal{E}^{(\mathrm{U})}_{m_1,j_1}(a_{j_1})$ hold.

Case 1: $1\neq i_{\mathrm{out}}$. Lemma 5.3 implies that for any realization of j_1, m_1,

$$\Delta_{1,j_1}\le\frac{8C_{m_1}}{N}.$$

Since $i_{\mathrm{out}}\in A_m$ for all $1\le m\le M$, we have $i_{\mathrm{out}}\in A_{m_1}$. In addition, since

$$j_1:=\arg\min_{i\in A_{m_1}}w(i),$$

we have $\Delta_{1,i_{\mathrm{out}}}\le\Delta_{1,j_1}$. Therefore,

$$\Delta_{1,i_{\mathrm{out}}}\le\Delta_{1,j_1}\le\frac{8C_{m_1}}{N}\le\frac{8C}{N}.$$

Case 2: $1=i_{\mathrm{out}}$. It is trivial to see $\Delta_{1,i_{\mathrm{out}}}\le 8C/N$.

Hence, when $\mathcal{E}^{(\mathrm{L})}_{m_1,1}(a_{j_1})$ and $\mathcal{E}^{(\mathrm{U})}_{m_1,j_1}(a_{j_1})$ hold, we always have $\Delta_{1,i_{\text{out}}} \le 8C/N$.

(ii) Altogether, for any realization of m_1, j_1,

$$\mathbb{P}\Big[\Big\{\Delta_{1,i_{\text{out}}} > \frac{8C}{N}\Big\}\bigcap\{m_1 = m, j_1 = i\}\Big] \le \mathbb{P}\Big[\Big(\overline{\mathcal{E}^{(\mathrm{L})}_{m_1,1}(a_{j_1})\bigcap\mathcal{E}^{(\mathrm{U})}_{m_1,j_1}(a_{j_1})}\Big)\bigcap\{m_1 = m, j_1 = i\}\Big]. \tag{C.1}$$

In addition, we have

$$\begin{aligned}
&\mathbb{P}\Big[\Big(\overline{\mathcal{E}^{(\mathrm{L})}_{m_1,1}(a_{j_1})\bigcap\mathcal{E}^{(\mathrm{U})}_{m_1,j_1}(a_{j_1})}\Big)\bigcap\{m_1 = m, j_1 = i\}\Big] \\
&\le \mathbb{P}\big[\,\overline{\mathcal{E}^{(\mathrm{L})}_{m_1,1}(a_{j_1})}\bigcap\{m_1 = m, j_1 = i\}\big] + \mathbb{P}\big[\,\overline{\mathcal{E}^{(\mathrm{U})}_{m_1,j_1}(a_{j_1})}\bigcap\{m_1 = m, j_1 = i\}\big] \\
&\le \mathbb{P}\big[\,\overline{\mathcal{E}^{(\mathrm{L})}_{m_1,1}(a_{j_1})}\bigcap\{u\cdot i \ge |A_{m-1}|\}\big] + \mathbb{P}\big[\,\overline{\mathcal{E}^{(\mathrm{U})}_{m_1,j_1}(a_{j_1})}\bigcap\{u\cdot i \ge |A_{m-1}|\}\big] && \text{(C.2)} \\
&\le 4\exp\left[-\frac{a_i^2\cdot n_m}{3}\right]\cdot\mathbb{I}\{u\cdot i \ge |A_{m-1}|\,\} && \text{(C.3)} \\
&= 4\exp\left[-\frac{\Delta_{1,i}^2\cdot N}{192|A_{m-1}|}\right]\cdot\mathbb{I}\{u\cdot i \ge |A_{m-1}|\,\} && \text{(C.4)} \\
&\le 4\exp\left[-\frac{\Delta_{1,i}^2\cdot N}{192\cdot\min\{u\cdot i, L\}}\right] && \text{(C.5)} \\
&\le 4\exp\left[-\frac{N}{192\tilde{H}_2(w,L,u)}\right]. && \text{(C.6)}
\end{aligned}$$

Line (C.2) results from the definitions of j_1, A_m $(1 \le m \le M)$, which implying that

$$j_1 \ge |A_{m_1}| = \left\lceil\frac{L}{u^{m_1}}\right\rceil = \left\lceil\frac{L}{u\cdot u^{m_1-1}}\right\rceil \ge \frac{\left\lceil\frac{L}{u^{m_1-1}}\right\rceil}{u} = \frac{|A_{m_1-1}|}{u}.$$

Line (C.3) follows from Lemma 5.2. Line (C.4) applies the definitions of a_i and n_m for all i, v, m:

$$a_i = \frac{\Delta_{1,i}}{8}\ \forall 2 \le i \le L, \quad \text{and} \quad n_m = \frac{N}{|A_{m-1}|}\ \forall 1 \le m \le M.$$

Lines (C.5) and (C.6) result from the fact that $|A_m| \le L$ for all m and the definition of $\tilde{H}_2(w,L,u)$ in (4.2), i.e.,

$$\tilde{H}_2(w,L,u) = \max_{i\ne 1}\frac{\min\{u\cdot i, L\}}{\Delta_{1,i}^2}.$$

(iii) Combining (C.1) and (C.6), we have

$$\mathbb{P}\left[\Delta_{1,i_{\text{out}}} > \frac{8C}{N}\right] \le \sum_{m=1}^{M}\sum_{i=2}^{L}\mathbb{P}\left[\Big\{\Delta_{1,i_{\text{out}}} > \frac{8C}{N}\Big\}\bigcap\{m_1 = m, j_1 = i\}\right] \le 4M(L-1)\exp\left[-\frac{N}{192\tilde{H}_2(w,L,u)}\right].$$

We complete the proof of Theorem 4.1 with $N = \lfloor T/M\rfloor$, $M = \lceil\log_u L\rceil$.

C.5. Proof of Theorem 4.2

Theorem 4.2. *Fix $\lambda \in (0,1)$ and $\Delta \in (0,1/2)$. For any online algorithm, there is a BAI with an adversarial corruption instance in T steps, corruption budget $C = 1 + (1+\lambda)2\Delta T$, and optimality gap Δ, such that*

$$\begin{aligned}
\mathbb{P}[\Delta_{1,i_{\text{out}}} > 0] = \mathbb{P}[\Delta_{1,i_{\text{out}}} \ge \Delta] &= \mathbb{P}[i_{\text{out}} \ne 1] \\
&\ge \frac{1}{2}\cdot\left[1 - \exp\Big(-\frac{2\lambda^2\Delta T}{3}\Big)\right].
\end{aligned}$$

Proof. We fix $w = \{w(i)\}_{i\in[L]}$, where $1 > w(1) > w(2) > w(3) \geq \ldots \geq w(L) > 0$, and we define $\Delta = w(1) - w(2)$. We assume $w(2) - \Delta > w(3) > 0$. We prove the Theorem by a coupling argument between two Bernoulli instances $\mathcal{I}, \mathcal{I}'$, both on the ground set $[L]$. Both involve T time steps and corruption budget $C = (1 + \lambda)2\Delta T$.

In instance $\mathcal{I}$, the uncorrupted reward distribution of item i is Bern$(w(i))$, and the adversary corrupts the rewards of item 1 probabilistically, as detailed in the forthcoming coupling in Algorithm 2. In instance $\mathcal{I}'$ the uncorrupted reward distribution of the items are:

- Bern$(u(1))$, where $u(1) = w(2) - \Delta$, for item 1,
- Bern$(u(i))$, where $u(i) = w(i)$, for item $i \in [L] \setminus \{1\}$,

but the adversary does not corrupt any of the rewards on instance $\mathcal{I}'$. The optimal items in instances $\mathcal{I}, \mathcal{I}'$ are different, and they are item 1, item 2 respectively. Both instances have optimality gap Δ, since in instance $\mathcal{I}'$ we have $u(2) > u(1) > u(3) \geq \ldots \geq u(L) > 0$.

We denote the original and corrupted rewards of item i at time step t in instance $\mathcal{I}$ as $W_t(i), \tilde{W}_t(i)$ respectively, and the original and corrupted rewards of item i at time step t in instance $\mathcal{I}'$ as $U_t(i), \tilde{U}_t(i)$ respectively. Since there is no corruption on $\mathcal{I}'$, we have $U_t(i) = \tilde{U}_t(i)$ for all t, i always.

Fix a BAI algorithm π, and considering running π on the instances $\mathcal{I}, \mathcal{I}'$. When π is randomized, we assume that π has the same random seed in the two runs, so that π recommends the same item in both instances $\mathcal{I}, \mathcal{I}'$ if $\tilde{W}_t(i) = \tilde{U}_t(i)$ for all t, i. Now, we couple the instances as shown $\mathcal{I}, \mathcal{I}'$ in Algorithm 2.

Algorithm 2 Coupling on instances $\mathcal{I}, \mathcal{I}'$

Set remaining corruption budget $B \leftarrow C$.
for time step $t = 1, \ldots, T$ **do**
  Adversary observes $\{W_t(i)\}_{i\in[L]}$, where $W_t(i) \sim$ Bern$(w(i))$.
  Adversary generates $G_t \sim$ Bern$(2\Delta/w(1))$, independent of W_t.
  if $B \geq 1$ **then**
    if $W_t(1) = 0$ **then**
      Set $\tilde{W}_t(1) \leftarrow 0$.
    else if $W_t(1) = 1, G_t = 1$ **then**
      Set $\tilde{W}_t(1) \leftarrow 0$ ($c_t(1) = -1$).
      Update $B \leftarrow B - 1$.
    else if $W_t(1) = 1, G_t = 0$ **then**
      Set $\tilde{W}_t(1) \leftarrow 1$.
    end if
    Set $\tilde{W}_t(i) \leftarrow W_t(i)$ for all $i \in [L] \setminus \{1\}$.
    Set $U_t(i) \leftarrow \tilde{W}_t(i), \tilde{U}_t(i) \leftarrow \tilde{W}_t(i)$ for all $i \in [L]$.
  else
    Set $\tilde{W}_t(i) \leftarrow W_t(i)$ for all $i \in [L]$ ($c_t(i) = 0$).
    Set $U_t(i) \leftarrow \tilde{W}_t(i), \tilde{U}_t(i) \leftarrow \tilde{W}_t(i)$ for all $i \in [L] \setminus \{1\}$.
    Sample $U_t(1) = \tilde{U}_t(1) \sim$ Bern$((w(2) - \Delta))$ (recall $u(1) = w(2) - \Delta$).
  end if
end for

We make two crucial observation on the coupling in Algorithm 2:

1. If the corruption budget C is sufficient, that is if we have $B \geq 1$ at the start of time step T, then $\tilde{W}_t(i) = \tilde{U}_t(i)$ for all i, t, so that the algorithm π recommends the same item in both instances.

2. The coupling is valid, in the sense that:

   (a) The corruption budget is never exceeded,
   (b) We always have $W_t(i) \sim$ Bern$(w(i))$,
   (c) We always have $\tilde{U}_t(i) = U_t(i) \sim$ Bern$(u(i))$.

The claims (a, b) are clearly true, and for claim (c), we need to verify that $\tilde{U}_t(1) = U_t(1) \sim \text{Bern}(u(1))$. Indeed, at a time step t:

- If $B < 1$, then Line 19 imposes that $U_t(1) \sim \text{Bern}(u(1))$.
- If $B \geq 1$, then by the **if** loop in Line 5, we have

$$\begin{aligned}
&\mathbb{P}[U_t(1) = \tilde{U}_t(1) = 1 | B \geq 1 \text{ at the start of time step t}] \\
&= \mathbb{P}[W_t(1) = 1, G_t = 0) = \mathbb{P}[W_t(1) = 1] \cdot \mathbb{P}[G_t = 0] \\
&= w(1) \cdot \left(1 - \frac{2\Delta}{w(1)}\right) = w(1) - 2\Delta = w(2) - \Delta = u(1).
\end{aligned}$$

The key to the proof is that the optimal item in instances $\mathcal{I}, \mathcal{I}'$ are 1, 2 respectively which are different item. By observation 1, if $B \geq 1$ at the start of time T, then the algorithm π cannot identify the optimal item in both instances. Denote events $\mathcal{A}_1 = \{\pi \text{ outputs item 1 in } \mathcal{I}\}$ and $\mathcal{A}_2 = \{\pi \text{ outputs item 2 in } \mathcal{I}'\}$, and denote $\mathbb{P}$ as the probability measure under the coupling in Algorithm 2 and the algorithm π. Now,

$$\begin{aligned}
&\mathbb{P}[\mathcal{A}_1 \cap \mathcal{A}_2] \\
&\leq \mathbb{P}[\ \pi \text{ outputs different items on } \mathcal{I}, \mathcal{I}'\ ] \\
&\leq \mathbb{P}[\ \tilde{W}_t(1) \neq \tilde{U}_t(1) \text{ for some } t \in [T]\ ] \\
&\leq \mathbb{P}[\ \text{At the start of time step } T, \text{ we have } B < 1\ ] \\
&= \mathbb{P}\left[\ \sum_{t=1}^{T-1} \mathbb{I}\{W_t(1) = 1, G_t = 1\} > C - 1\right] \\
&\leq \mathbb{P}\left[\ \sum_{t=1}^{T} \mathbb{I}\{W_t(1) = 1, G_t = 1\} > (1+\lambda)2\Delta T\right].
\end{aligned}$$

To this end, note that the random variables in $\{\mathbb{I}\{W_t(1) = 1, G_t = 1\}\}_{t=1}^T$ are i.i.d. with mean

$$\mathbb{E}[\mathbb{I}\{W_t(1) = 1, G_t = 1\}] = \mathbb{E}[\mathbb{I}\{W_t(1) = 1\}] \cdot \mathbb{E}[\mathbb{I}\{G_t = 1\}] = w(1) \cdot \frac{2\Delta}{w(1)} = 2\Delta.$$

By applying Theorem B.3, we have

$$\mathbb{P}[\mathcal{A}_1 \cap \mathcal{A}_2] \leq \mathbb{P}\left[\sum_{t=1}^{T} \mathbb{I}\{W_t(1) = 1, G_t = 1)\} > (1+\lambda)2\Delta T\right] \leq \exp\left(-\frac{2\lambda^2 \Delta T}{3}\right).$$

Finally, we have

$$\mathbb{P}[\mathcal{A}_1] + \mathbb{P}[\mathcal{A}_2] = \mathbb{P}[\mathcal{A}_1 \cup \mathcal{A}_2] + \mathbb{P}[\mathcal{A}_1 \cap \mathcal{A}_2] \leq 1 + \exp\left(-\frac{2\lambda^2 \Delta T}{3}\right),$$

so that

$$\min\{\mathbb{P}[\mathcal{A}_1], \mathbb{P}[\mathcal{A}_2]\} \leq \frac{1}{2}\left[1 + \exp\left(-\frac{2\lambda^2 \Delta T}{3}\right)\right],$$

completing the proof of the theorem. □

C.6. Proof of Theorem 4.3

Theorem 4.3. *Fix $L > 1$, $\lambda \in (0, 1)$ and $\Delta \in (0, 1/4)$. For the SH algorithm, there is a BAI with adversarial corruption instance with T time steps, corruption budget $C = (1+\lambda)2\Delta T/(L \log_2 L)$, and optimality gap Δ, such that if T is sufficiently large,*

$$\mathbb{P}[\Delta_{1,i_{\text{out}}} > 0] = \mathbb{P}[\Delta_{1,i_{\text{out}}} \geq \Delta] = \mathbb{P}[i_{\text{out}} \neq 1] \geq 1/2.$$

Proof. Consider Theorem 4.2's attack strategy, but applied to SH in phase 1.

We claim that there is a BAI instance $\mathcal{I}$ with T time steps, gap Δ and $C = (1+\lambda)2\Delta T/(L\log_2 L)$, such that

$$\Pr(\Delta_{1,i_{\text{out}}} \geq \Delta) = \Pr\left(\Delta_{1,i_{\text{out}}} \geq \frac{C \cdot L\log_2 L}{2(1+\lambda)T)}\right) \geq 1/2$$

when T is sufficiently large. This is a matching lower bound for SH in Table 1.

Consider a Bernoulli instance $\{w(i)\}_{i\in[L]}$ with $w(1) \in [1/2, 1]$ and $w(i) = w(1) - \Delta$ for $i \in [L] \setminus \{1\}$. In phase 1, SH pulls each $i \in [L]$ for $\tau = \lceil T/(L\log_2 L)\rceil$ times, computes the empirical means $\{\hat{w}_1(i)\}_{i\in[L]}$, and removes the $\lceil L/2\rceil$ items with smallest $\hat{w}_1(i)$ from consideration.

During phase 1, SH pulls item 1 at fixed time steps $\{t_s\}_{s=1}^{\tau}$. When the adversary determines $\{c_{t_s}(i)\}_{i\in[L]}$, *he knows* $\{W_{t_s}(i)\}_{i\in[L]}$, *and knows that item* 1 *will be pulled at time* t_s. The adversary attacks by solely corrupting item 1 solely at times $\{t_s\}_{t=1}^{\tau}$.

If the corruption budget is not exhausted, set

$$\begin{aligned}
&\Pr(\tilde{W}_{t_s}(1) = 0 | W_{t_s}(1) = 1) = \frac{2\Delta}{w(1)} = 1 - \Pr(\tilde{W}_{t_s}(1) = 1 | W_{t_s}(1) = 1),\\
&\Pr(\tilde{W}_{t_s}(1) = 0 | W_{t_s}(1) = 0) = 1,
\end{aligned}$$

which implies that

$$\begin{aligned}
&\Pr(\tilde{W}_{t_s}(1) = 1) = w(1) - 2\Delta,\\
&\Pr(c_{t_s}(1) = -1) = \Pr(\tilde{W}_{t_s}(1) = 0 | W_{t_s}(1) = 1)\Pr(W_{t_s}(1) = 1) = 2\Delta.
\end{aligned}$$

If exhausted, no corruption.

Let $X_1, \ldots, X_\tau \sim \text{Bern}(w(1) - 2\Delta)$, $Y_1, \ldots, Y_\tau \sim \text{Bern}(2\Delta)$ be i.i.d. random variables and event

$$\mathcal{E} := \{\text{corruption budget not exhausted at end of phase } 1\}.$$

Then

$$\begin{aligned}
\Pr\left(\hat{w}_1(i) \leq w(1) - \frac{3\Delta}{2}\right) &\geq \Pr\left(\hat{w}_1(i) \leq w(1) - \frac{3\Delta}{2}\middle|\mathcal{E}\right)\cdot\Pr(\mathcal{E})\\
&= \Pr\left(\frac{1}{\tau}\sum_{s=1}^{\tau} X_s \leq w(1) - \frac{3\Delta}{2}\right)\cdot\Pr\left(\sum_{s=1}^{\tau} Y_s \leq \frac{(1+\lambda)2\Delta T}{L\log_2 L}\right),
\end{aligned}$$

which exceeds $1 - \exp\left[-\frac{T\Delta^2}{2L\log_2 L}\right] - \exp\left[-\frac{2\lambda^2 T\Delta}{3L\log_2 L}\right]$. Thus, for all $i \neq 1$,

$$\Pr\left(\hat{w}_1(i) > w(1) - \frac{3\Delta}{2}\right) \geq 1 - \exp\left[-\frac{T\Delta^2}{2L\log_2 L}\right].$$

Lastly, by the union bound,

$$\Pr(\text{Item 1 removed after phase 1}) \geq 1 - (L+1)\exp\left[-\frac{T\max\{1,\lambda^2\}\Delta^2}{2L\log_2 L}\right].$$

We complete the proof by noting that

$$1 - (L+1)\exp\left[-\frac{T\max\{1,\lambda^2\}\Delta^2}{2L\log_2 L}\right] \geq 1/2$$

for T large enough.

□

C.7. Proof of Theorem 4.4

Theorem 4.4. *Fix any $\lambda, \epsilon \in (0,1)$. If $C \geq L \cdot \{1-(1-\lambda)[1-w(1)]\} \cdot T$, Strategy (I)'s attack results in*

$$\mathbb{P}[\Delta_{1,i_{\text{out}}} > \epsilon] \geq 1 - \frac{L_\epsilon}{L} - \exp\bigg[-\frac{\lambda^2 TL[1-w(1)]}{2}\bigg].$$

If instead $C \geq L \cdot [1-(1-\lambda)w(L)] \cdot T$, Strategy (II)'s attack results in

$$\mathbb{P}[\Delta_{1,i_{\text{out}}} > \epsilon] \geq 1 - \frac{L_\epsilon}{L} - \exp\bigg[-\frac{\lambda^2 TLw(L)}{2}\bigg].$$

Proof. **Part (a).** Let $X_1, \cdots, X_n$ be Bernoulli random variables taking values in $\{0,1\}$ such that $\mathbb{E}\left[X_t | X_1, \cdots, X_{t-1}\right] \leq \mu$ for all $t \leq n$, and $Y = X_1 + \ldots + X_n$. Let $X_t' = 1 - X_t$ for all $1 \leq t \leq n$, $\mu' = 1-\mu$, $Y' = X_1' + \ldots + X_n'$. Then, Theorem B.3 indicates for all $\delta \in (0,1)$

$$\Pr[Y' \leq (1-\delta)n\mu'] \leq \exp\bigg(-\frac{\delta^2 n\mu'}{2}\bigg) \;\Rightarrow\; \Pr[n - Y \leq (1-\delta)n(1-\mu)] \leq \exp\bigg[-\frac{\delta^2 n(1-\mu)}{2}\bigg]$$
$$\Rightarrow\; \Pr[Y \geq n - (1-\delta)n(1-\mu)] \leq \exp\bigg[-\frac{\delta^2 n(1-\mu)}{2}\bigg].$$

(i) For all $i \in [L]$, $W_t(i)$ denotes the random reward of item i at time step t. Fix any $\lambda \in (0,1)$. We can apply the inequality above with $\mu = w(1)$, $n = TL$ to get

$$\mathbb{P}\bigg[\sum_{t=1}^{T}\sum_{i=1}^{L} W_t(i) \geq TL - (1-\lambda)TL[1-w(1)]\bigg] \leq \exp\bigg[-\frac{\lambda^2 TL[1-w(1)]}{2}\bigg]$$
$$\Rightarrow\; \mathbb{P}\bigg[\sum_{t=1}^{T}\sum_{i=1}^{L} \mathbb{I}\{W_t(i)=1\} \geq TL \cdot \{1-(1-\lambda)[1-w(1)]\}\bigg] \leq \exp\bigg[-\frac{\lambda^2 TL[1-w(1)]}{2}\bigg].$$

(ii) Let

$$\mathcal{E}_{\lambda,0} := \bigg\{\sum_{t=1}^{T}\sum_{i=1}^{L} \mathbb{I}\{W_t(i)=1\} < TL \cdot \{1-(1-\lambda)[1-w(1)]\}\bigg\}.$$

When $\mathcal{E}_{\lambda,0}$ holds, throughout the whole horizon (T time steps), there are less than $TL \cdot \{1-(1-\lambda)[1-w(1)]\}$ random rewards that equal to 1. If we additionally have

$$C \geq TL \cdot \{1-(1-\lambda)[1-w(1)]\} := C_{\lambda,0},$$

the adversary can shift the random reward to 0 whenever it equals to 1, which implies that the agent get a corrupted reward equals to 0 at each time step.

Altogether, when $\mathcal{E}_{\lambda,0}$ holds and $C \geq C_{\lambda,0}$, the agent get a corrupted reward equals to 0 at each time step. Therefore, the observations of random rewards throughout the whole horizon provides no information about the mean reward $w(i)$ for any item $i \in [L]$. In this case, the best method for the agent to output an item is to randomly output any ground item with a uniform probability of $1/L$. As a result, for any item i,

$$\mathbb{P}\big[\{i_{\text{out}} = i\} \bigcap \mathcal{E}_{\lambda,0}\big] \leq \frac{1}{L}.$$

Recall that $L_\epsilon := |\{i \in [L] : \Delta_{1,i} \leq \epsilon\}|$ counts the items with mean reward at most ϵ worse than that of the optimal item, we have

$$\mathbb{P}\big[\{\Delta_{1,i_{\text{out}}} \leq \epsilon\} \bigcap \mathcal{E}_{\lambda,0}\big] \leq \sum_{i \in [L], \Delta_{1,i} \leq \epsilon} \mathbb{P}\big[\{i_{\text{out}} = i\} \bigcap \mathcal{E}_{\lambda,0}\big] \leq \frac{L_\epsilon}{L}.$$

(iii) Therefore,

$$\mathbb{P}[\Delta_{1,i_{\text{out}}} \le \epsilon] = \mathbb{P}\big[\{\Delta_{1,i_{\text{out}}} \le \epsilon\} \bigcap \mathcal{E}_{\lambda,0}\big] + \mathbb{P}\big[\{\Delta_{1,i_{\text{out}}} \le \epsilon\} \bigcap \overline{\mathcal{E}_{\lambda,0}}\,\big]$$
$$\le \mathbb{P}\big[\{\Delta_{1,i_{\text{out}}} \le \epsilon\} \bigcap \mathcal{E}_{\lambda,0}\big] + \mathbb{P}\big[\,\overline{\mathcal{E}_{\lambda,0}}\,\big] \le \frac{L_\epsilon}{L} + \exp\bigg[-\frac{\lambda^2 TL[1-w(1)]}{2}\bigg],$$

Lastly,

$$\mathbb{P}[\Delta_{1,i_{\text{out}}} > \epsilon] \ge 1 - \frac{L_\epsilon}{L} - \exp\bigg[-\frac{\lambda^2 TL[1-w(1)]}{2}\bigg].$$

Part (b). (i) For all $i \in [L]$, $W_t(i)$ denotes the random reward of item i at time step t. Fix any $\lambda \in (0,1)$. We can apply Theorem B.3 with $\mu = w(L)$, $n = TL$ to get

$$\mathbb{P}\bigg[\sum_{t=1}^{T}\sum_{i=1}^{L} W_t(i) \le (1-\lambda)TLw(L)\bigg] \le \exp\bigg[-\frac{\lambda^2 TLw(L)}{2}\bigg].$$

Meanwhile,

$$\bigg\{\sum_{t=1}^{T}\sum_{i=1}^{L} W_t(i) \le (1-\lambda)TLw(L)\bigg\} = \bigg\{\sum_{t=1}^{T}\sum_{i=1}^{L} \mathbb{I}\{W_t(i)=1\} \le (1-\lambda)TLw(L)\bigg\}$$
$$= \bigg\{\sum_{t=1}^{T}\sum_{i=1}^{L} \mathbb{I}\{W_t(i)=0\} \ge TL - (1-\lambda)TLw(L) = TL\cdot[1-(1-\lambda)w(L)]\bigg\}.$$

Therefore,

$$\mathbb{P}\bigg[\sum_{t=1}^{T} \mathbb{I}\{W_t(i)=0\} \ge TL\cdot[1-(1-\lambda)w(L)]\bigg] \le \exp\bigg[-\frac{\lambda^2 TLw(L)}{2}\bigg].$$

(ii) Let

$$\mathcal{E}_{\lambda,1} := \bigg\{\sum_{t=1}^{T}\sum_{i=1}^{L} \mathbb{I}\{W_t(i)=0\} < TL\cdot[1-(1-\lambda)w(L)]\bigg\}.$$

When $\mathcal{E}_{\lambda,1}$ holds, throughout the whole horizon (T time steps), there are less than $TL\cdot[1-(1-\lambda)w(L)]$ random rewards that equal to 0. If we additionally have

$$C \ge TL\cdot[1-(1-\lambda)w(L)] := C_{\lambda,1},$$

the adversary can shift the random reward to 1 whenever it equals to 0, which implies that the agent get a corrupted reward equals to 1 at each time step.

Altogether, when $\mathcal{E}_{\lambda,1}$ holds and $C \ge C_{\lambda,1}$, the agent get a corrupted reward equals to 1 at each time step. Therefore, the observations of random rewards throughout the whole horizon provides no information about the mean reward $w(i)$ for any item $i \in [L]$. In this case, the best method for the agent to output an item is to randomly output any ground item with a uniform probability of $1/L$. As a result, for any item i,

$$\mathbb{P}\big[\{i_{\text{out}} = i\} \bigcap \mathcal{E}_{\lambda,1}\big] \le \frac{1}{L}.$$

Recall that $L_\epsilon := |\{i \in [L] : \Delta_{1,i} \le \epsilon\}|$ counts the items with mean reward at most ϵ worse than that of the optimal item, we have

$$\mathbb{P}\big[\{\Delta_{1,i_{\text{out}}} \le \epsilon\} \bigcap \mathcal{E}_{\lambda,1}\big] \le \sum_{i\in[L],\Delta_{1,i}\le\epsilon} \mathbb{P}\big[\{i_{\text{out}} = i\} \bigcap \mathcal{E}_{\lambda,1}\big] \le \frac{L_\epsilon}{L}.$$

(iii) Therefore,

$$\mathbb{P}[\Delta_{1,i_{\text{out}}} \leq \epsilon] = \mathbb{P}\big[\{\Delta_{1,i_{\text{out}}} \leq \epsilon\} \bigcap \mathcal{E}_{\lambda,1}\big] + \mathbb{P}\big[\{\Delta_{1,i_{\text{out}}} \leq \epsilon\} \bigcap \overline{\mathcal{E}_{\lambda,1}}\,\big]$$
$$\leq \mathbb{P}\big[\{\Delta_{1,i_{\text{out}}} \leq \epsilon\} \bigcap \mathcal{E}_{\lambda,1}\big] + \mathbb{P}\big[\,\overline{\mathcal{E}_{\lambda,1}}\,\big] \leq \frac{L_\epsilon}{L} + \exp\bigg[-\frac{\lambda^2 T L w(L)}{2}\bigg],$$

Lastly,

$$\mathbb{P}[\Delta_{1,i_{\text{out}}} > \epsilon] \geq 1 - \frac{L_\epsilon}{L} - \exp\bigg[-\frac{\lambda^2 T L w(L)}{2}\bigg].$$

□

D. Additional numerical results

D.1. Details of Figures 3(a) and 3(b)

Table A.1. Comparison of PSS(u) to Other Algorithms

λ	w^*	w'	PSS(2)	SH	UP
0.5	0.4	0.2	76	42	12
0.5	0.5	0.2	91	64	13
0.5	0.5	0.3	74	51	19
0.9	0.4	0.2	72	45	9
0.9	0.5	0.2	83	64	7
0.9	0.5	0.3	60	40	12

Here, we provide the raw numbers of for Figures 3(a) and 3(b). We see that PSS(2) consistently and clearly outperform the non-robust BAI algorithms on all instances here.

D.2. Further observations

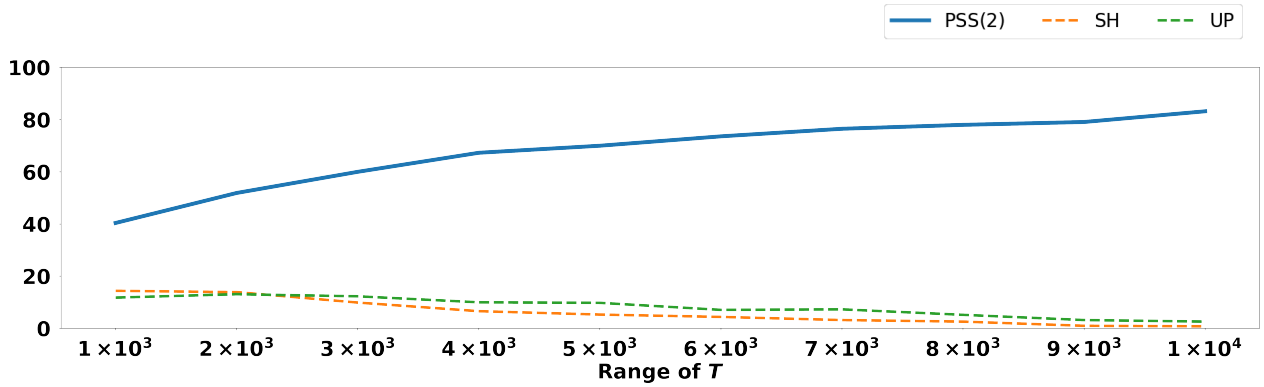
To further evaluate the impact of T, L, and λ on the success probabilities of PSS(2), SH and UP, we run each algorithm for 1000 times independently with varying sets of parameters, while keeping the MAB instance at $w^* = 0.4$ and $w' = 0.2$ fixed.

Recall that according to Theorem 4.3, we set the CPS

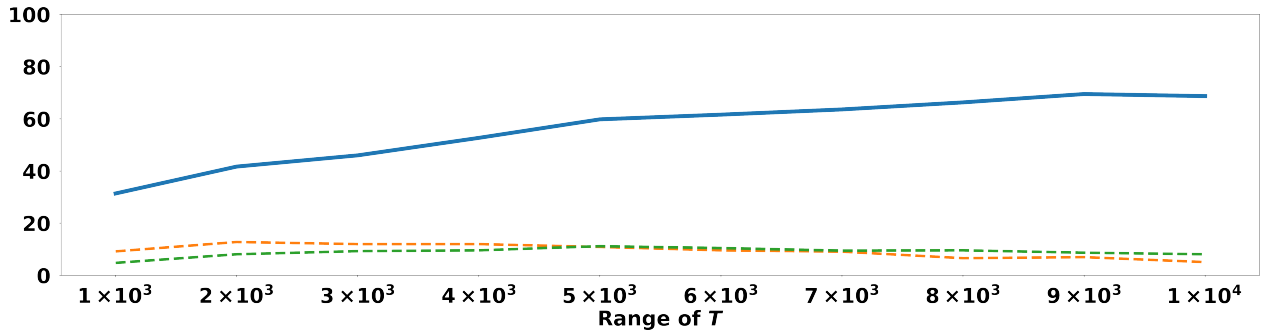
$$\frac{C}{T} = \frac{2\Delta \cdot (1+\lambda)}{L \log_2 L}. \tag{D.1}$$

This is the scaling of the CPS that ensures that SH fails with high probability as T grows. Notice that C/T grows with λ and decreases with L. We implement the attack strategy as applied in Theorem 4.3 (see Algorithm 2) and vary L and T.

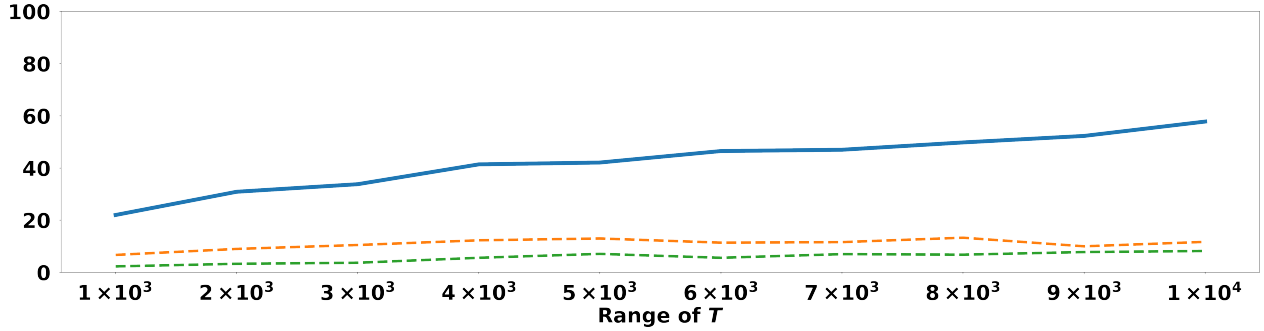
In each subplot in Figure A.1, we consider different number of items L and use different values of λ, resulting in different CPSes. We let λ grows with L, so the identification of the best arm would pose significant difficulty to SH as prescribed by Eqn. (D.1). The figures show that as T grows, the success (BAI) probabilities of PSS(2) demonstrate an increasing trend, and in the case of $L = 32, \lambda = 9$ the percentage of successful BAI converges to 100%. In stark contrast, the percentages of successful BAI for SH are always below 20%. In the case of $L = 32, \lambda = 9$, the percentage appears to converge to 0 as T increases. This implies that SH fails with high probability when T is sufficiently large, which corroborates Theorem 4.3. However, the randomization inherent in PSS(2) ensures that it remains extremely robust to the corruption strategy and it successfully identifies the best item a large fraction of times as $T \to +\infty$; this corroborates our main result—Theorem 4.1.



(a) $L = 32, \lambda = 9$



(b) $L = 64, \lambda = 19$



(c) $L = 128, \lambda = 39$

Figure A.1. Percentage of correct BAI of PSS(2), SH and UP. We fix the instance to be $w^* = 0.4, w' = 0.2$.