A. Details on the offline attacks

A.1. Proof of Theorem 1

Proof. The optimization problem $P_1$ is a quadratic program with linear constraints in $\{a_0\}_{a \in A}$. Now it remains to show that the constraint set is non-empty.

Given any reward instance $\{y_{a}\}_{a \in A}$, any margin parameter $\xi > 0$ and any $\bar{a}$, one can check that

$$
\bar{a} = \left[(y_{\bar{a}} + \bar{a} - y_{a})/m_{a} - y_{a}'/m_{a} - \xi\right] \mathbf{1}, \quad \forall a \neq \bar{a},
$$

(27)

satisfies the constraints of problem $P_1$. That is the constraint set of problem $P_1$ is non-empty.

Thus, there exists at least one optimal solution of problem $P_1$ since $P_1$ is a quadratic program with non-empty and compact constraints. The result follows from Proposition 1.

A.2. Details on attacking Thompson Sampling

Lemma 2. Given some constants $C_i > 0$ for any $i < K$. The function $f(\bar{x}) = \sum_{i=1}^{K-1} \Phi(C_i x_i - C_i x_K)$ is convex on the domain $D = \{\bar{x} \in \mathbb{R}^K | x_i - x_K \leq 0, \forall i < K\}$.

Proof. We prove the result by checking the Hessian matrix $H$ of function $f(\bar{x})$. Note that $\Phi(x)$ is the cumulative distribution function of the standard normal distribution $\mathcal{N}(0, 1)$. For any $i < K$, we have that

$$
\frac{\partial f}{\partial x_i} = \frac{C_i}{\sqrt{2\pi}} e^{-(C_i x_i - C_i x_K)^2/2},
$$

(28)

$$
\frac{\partial^2 f}{\partial x_i^2} = -\frac{C_i^2}{2\sqrt{2\pi}} e^{-(C_i x_i - C_i x_K)^2/2}(C_i x_i - C_i x_K).
$$

(29)

On the other hand, we have that

$$
\frac{\partial f}{\partial x_K} = \sum_{i=1}^{K-1} -\frac{C_i}{\sqrt{2\pi}} e^{-(C_i x_i - C_i x_K)^2/2} = \sum_{i=1}^{K-1} -\frac{\partial f}{\partial x_i},
$$

(30)

$$
\frac{\partial^2 f}{\partial x_K^2} = \sum_{i=1}^{K-1} -\frac{C_i^2}{2\sqrt{2\pi}} e^{-(C_i x_i - C_i x_K)^2/2}(C_i x_i - C_i x_K) = \sum_{i=1}^{K-1} \frac{\partial^2 f}{\partial x_i^2}.
$$

(31)

Now, we derive the other coefficients. For any pair $(i, j)$ such that $i \neq j, i < K$ and $j < K$, we have that

$$
\frac{\partial^2 f}{\partial x_i \partial x_j} = 0.
$$

(32)

For any $i < K$, we have that

$$
\frac{\partial^2 f}{\partial x_K \partial x_i} = \frac{C_i^2}{\sqrt{2\pi}} e^{-(C_i x_i - C_i x_K)^2/2}(C_i x_i - C_i x_K) = \frac{\partial^2 f}{\partial x_i^2},
$$

(33)

$$
\frac{\partial^2 f}{\partial x_K^2} = -\frac{\partial^2 f}{\partial x_i^2}.
$$

(34)

Since the constants $C_i$ are positive, we have that $\frac{\partial^2 f}{\partial x_i^2} \geq 0$ in the domain $D$. The Hessian matrix of $f$ is the following,

$$
H = \begin{bmatrix}
\frac{\partial^2 f}{\partial x_1^2} & 0 & \ldots & 0 & -\frac{\partial^2 f}{\partial x_K^2} \\
0 & \frac{\partial^2 f}{\partial x_2^2} & \ldots & 0 & -\frac{\partial^2 f}{\partial x_K^2} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \ldots & \frac{\partial^2 f}{\partial x_{K-1}^2} & -\frac{\partial^2 f}{\partial x_K^2} \\
\frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2^2} & \ldots & -\frac{\partial^2 f}{\partial x_{K-1}^2} & \sum_{i=1}^{K-1} \frac{\partial^2 f}{\partial x_i^2}
\end{bmatrix}.
$$

(35)
Hence, for any vector $\tilde{y} \in \mathbb{R}^K$, we have that

$$
\tilde{y}^T H \tilde{y} = \tilde{y}^T \left[ \begin{array}{c}
\frac{\partial^2 f}{\partial x_1^2}(y_1 - y_K) \\
\frac{\partial^2 f}{\partial x_2^2}(y_2 - y_K) \\
\vdots \\
\frac{\partial^2 f}{\partial x_{K-1}^2}(y_{K-1} - y_K) \\
\sum_{i=1}^{K-1} \frac{\partial^2 f}{\partial x_i^2}(y_i - y_K)
\end{array} \right] = \sum_{i=1}^{K-1} \frac{\partial^2 f}{\partial x_i^2}(y_i - y_K)^2 \geq 0.
$$

(36)

Since $H$ is positive semi-definite, we show that $f(x)$ is convex on the domain $D$.

A.3. Proof of Proposition 2

Proof. By Lemma 2 and the fact that affine mapping keeps the convexity, we have the result.

A.4. Another relaxation of P for Thompson Sampling

We may find a sufficient constraint to equation (17) as

$$
\Phi \left( \frac{\tilde{\mu}_a(T) - \tilde{\mu}_{a^*}(T)}{\sigma^3 \sqrt{1/m_a + 1/m_{a^*}}} \right) \leq \frac{\delta}{K - 1}, \ \forall a \neq a^*.
$$

(37)

Then, we derive another relaxation of $P$ as

$$
P_4 : \min_{\epsilon_a, a \in A} \sum_{a \in A} ||\epsilon_a||^2_2 \text{ s.t. } \tilde{\mu}_a(T) - \tilde{\mu}_{a^*}(T) \leq \sigma^3 \sqrt{1/m_a + 1/m_{a^*}} \Phi^{-1} \left( \frac{\delta}{K - 1} \right), \ \forall a \neq a^*
$$

(38)

(39)

Note that problem $P_4$ is a quadratic program with linear constraints.

B. Details on the online attacks

B.1. Proof of Proposition 4

Proof. By equation (5), a logarithmic regret bound implies that the bandit algorithm satisfies $\mathbb{E}[N_a(T)] = O(\log T)$ for any suboptimal arm $a$. Note that the oracle constant attack shifts the expected rewards of all arms except for the target arm $a^*$. Since $C_a > [\mu_a - \mu_{a^*}]^+$, $\forall a \neq a^*$, the best arm is now the target arm $a^*$. Thus, the expected number of pulling the target arm is

$$
\mathbb{E}[N_{a^*}(T)] = T - \sum_{a \neq a^*} \mathbb{E}[N_a(T)] = T - o(T).
$$

(40)

Since the attacker does not attack the target arm, we have that

$$
\mathbb{E}[C(T)] = \mathbb{E} \left[ \sum_{t=1}^{T} \epsilon_t \right] = \sum_{a \neq a^*} C_a \mathbb{E}[N_a(T)] = O \left( \sum_{a \neq a^*} C_a \log T \right).
$$

(41)

On the other hand, suppose there exists an arm $i \neq a^*$ such that $C_i \leq [\mu_i - \mu_{a^*}]^+$, then the attack is not successful. In the case that $C_i < [\mu_i - \mu_{a^*}]^+$, the arm $i$ is the best arm rather than the target arm $a^*$ in the shifted bandit problem. That is the expected number of pulling arm $a^*$ is $\mathbb{E}[N_{a^*}(T)] = O(\log T)$. In the case that $C_i = [\mu_i - \mu_{a^*}]^+$, the arm $i$ and $a^*$ are both the best arms. That is the expected attack cost is $\mathbb{E}[C(T)] = T - o(T)$. In neither case is the attack successful. This concludes the proof.
B.2. Proof of Theorem 4

Proof. Given any \( \delta > 0 \), we have that \( \mathbb{P}(E) > 1 - \delta \) by Lemma 1. Under the event \( E \), we have that at any time \( t \) and for any arm \( a \neq a^* \),

\[
\mu_a - \mu_{a^*} < \hat{\mu}_a(t) - \mu_{a^*} + \beta(N_a(t)) \tag{42}
\]

\[
< \hat{\mu}_a(t) - \mu_{a^*}(t) + \beta(N_a(t)) + \beta(N_{a^*}(t)), \tag{43}
\]

which implies that

\[
[\mu_a - \mu_{a^*}]^+ < [\hat{\mu}_a(t) - \mu_{a^*}(t) + \beta(N_a(t)) + \beta(N_{a^*}(t))]^+. \tag{44}
\]

By the same argument in the proof of Proposition 4, we have that under event \( E \), the attacker is taking an effective attack for any bandit algorithm.

Recall that the bandit algorithm has a high-probability bound such that the regret is bounded by \( O(\log T) \) with probability at least \( 1 - \delta \). Under event \( E \), we have that \( N_a(T) = O(\log T) \) for any \( a \neq a^* \) with high probability. Thus, with probability at least \( 1 - 2\delta \), we have that \( N_{a^*}(T) = T - o(T) \). It remains to bound the cost of the attacker, i.e., \( \sum_t |\epsilon_t| \).

Given any arm \( a \neq a^* \), any time \( t \) and under the event \( E \), we have that

\[
\hat{\mu}_a(t) - \mu_{a^*}(t) < \mu_a - \hat{\mu}_{a^*}(t) + \beta(N_a(t)) \tag{45}
\]

\[
< \mu_a - \mu_{a^*} + \beta(N_a(t)) + \beta(N_{a^*}(t)). \tag{46}
\]

This implies that

\[
[\hat{\mu}_a(t) - \mu_{a^*}(t) + \beta(N_a(t)) + \beta(N_{a^*}(t))]^+
\]

\[
< [\mu_a - \mu_{a^*} + 2\beta(N_a(t)) + 2\beta(N_{a^*}(t))]^+
\]

\[
\leq [\mu_a - \mu_{a^*}]^+ + 2\beta(N_a(t)) + 2\beta(N_{a^*}(t)). \tag{47}
\]

Thus, the first statement follows. By the fact that \( \beta(n) \) is a decreasing function, we have that

\[
\sum_{t=1}^T |\epsilon_t| \leq \sum_{t=1}^T ([\hat{\mu}_a - \mu_{a^*}]^+ + 4\beta(1)) \mathbb{I}\{a_t \neq a^*\} \tag{50}
\]

\[
= \sum_{a \neq a^*} ([\mu_a - \mu_{a^*}]^+ + 4\beta(1)) N_a(T) \tag{51}
\]

\[
\leq O \left( \sum_{a \neq a^*} ([\mu_a - \mu_{a^*}]^+ + 4\beta(1)) \log T \right) \tag{52}
\]

\( \square \)