A APPENDIX

Theorem 4.3. *Assume assumption 4.1 holds, and the attacker applies (5) to perform attack. Then there exists a constant $M > 0$ such that the expected number of target arm selections satisfies $\mathbf{E}\left[N_T(a^\dagger)\right] \geq T - MT^\alpha/\rho$, and the expected cumulative attack cost satisfies $\mathbf{E}\left[C_T\right] \leq MT^\alpha/\rho$.*

Proof. Note that under attack, the bandit player is equivalently facing a new environment with loss function $\tilde{\mathcal{L}}_t$. Also note that the target arm $a^\dagger$ is the optimal arm with respect to $\tilde{\mathcal{L}}_t$, thus the regret of the bandit player is

$$\begin{aligned} R_T &= \mathbf{E}\left[\sum_{t=1}^T \tilde{\mathcal{L}}_t(a_t)\right] - \min_a \sum_{t=1}^T \tilde{\mathcal{L}}_t(a) \\ &= \mathbf{E}\left[\sum_{t=1}^T \left(\tilde{\mathcal{L}}_t(a_t) - \tilde{\mathcal{L}}_t(a^\dagger)\right)\right] \\ &= \mathbf{E}\left[\sum_{t=1}^T \mathbb{1}\left[a_t \neq a^\dagger\right]\left(\tilde{\mathcal{L}}_t(a_t) - \sum_{t=1}^T \tilde{\mathcal{L}}_t(a^\dagger)\right)\right] \\ &\geq \rho\mathbf{E}\left[\sum_{t=1}^T \mathbb{1}\left[a_t \neq a^\dagger\right]\right] \\ &= \rho\left(T - \mathbf{E}\left[N_T(a^\dagger)\right]\right) \end{aligned} \tag{13}$$

where the second-to-last inequality is due to assumption 4.1. On the other hand, since the player applies some no-regret algorithm, we must have $R_T \leq MT^\alpha$ for some constant M. Therefore, we have

$$\rho\left(T - \mathbf{E}\left[N_T(a^\dagger)\right]\right) \leq MT^\alpha, \tag{14}$$

which gives $\mathbf{E}\left[N_T(a^\dagger)\right] \geq T - MT^\alpha/\rho$.

Next we upper bound the expected attack cost. We have proved that under attack, the target arm $a^\dagger$ will be selected in $T - MT^\alpha/\rho$ rounds. Then note that by our attack design (5), the attacker only incurs attack cost when non-target arm is selected. Therefore, we have

$$\begin{aligned} \mathbf{E}\left[C_T\right] &= \mathbf{E}\left[\sum_{t=1}^T |\tilde{\mathcal{L}}(a_t) - \mathcal{L}(a_t)|\right] \\ &= \mathbf{E}\left[\sum_{t=1}^T \mathbb{1}\left[a_t \neq a^\dagger\right]|\tilde{\mathcal{L}}(a_t) - \mathcal{L}(a_t)|\right] \\ &\leq \mathbf{E}\left[\sum_{t=1}^T \mathbb{1}\left[a_t \neq a^\dagger\right]\right] \\ &= T - \mathbf{E}\left[N_T(a^\dagger)\right] \\ &\leq MT^\alpha/\rho. \end{aligned} \tag{15}$$

where we have used $|\tilde{\mathcal{L}}(a_t) - \mathcal{L}(a_t)| \leq 1$. □

Theorem 4.6. *Assume the attacker applies (6) to perform attack. Then there exists a constant $M > 0$ such that the expected number of target arm selections satisfies*

$$\mathbf{E}\left[N_T(a^\dagger)\right] \geq T - \frac{1}{\alpha+\epsilon}T^{1-\alpha-\epsilon} - MT^{1-\epsilon}, \tag{7}$$

and the expected cumulative attack cost satisfies

$$\mathbf{E}\left[C_T\right] \leq \frac{1}{\alpha+\epsilon}T^{1-\alpha-\epsilon} + MT^{1-\epsilon} + \frac{1}{\alpha+\epsilon}T^{\alpha+\epsilon}. \tag{8}$$

Proof. Under attack, the bandit player is equivalently facing loss sequence $\tilde{\mathcal{L}}_{1:T}$. Note that $a^\dagger$ is the optimal arm with respect to $\tilde{\mathcal{L}}_{1:T}$, thus the regret is

$$\begin{aligned} R_T &= \mathbf{E}\left[\sum_{t=1}^{T}\left(\tilde{\mathcal{L}}_t(a_t) - \tilde{\mathcal{L}}_t(a^\dagger)\right)\right] \\ &= \mathbf{E}\left[\sum_{t=1}^{T}\mathbb{1}\left[a_t \neq a^\dagger\right]\left(\tilde{\mathcal{L}}_t(a_t) - \tilde{\mathcal{L}}_t(a_t^\dagger)\right)\right] \\ &= \mathbf{E}\left[\sum_{t=1}^{T}\mathbb{1}\left[a_t \neq a^\dagger\right]\left(1 - \tilde{\mathcal{L}}_t(a_t^\dagger)\right)\right] \\ &\geq \mathbf{E}\left[\sum_{t=1}^{T}\mathbb{1}\left[a_t \neq a^\dagger\right] t^{\alpha+\epsilon-1}\right], \end{aligned} \tag{16}$$

where we have used that $\tilde{\mathcal{L}}_t(a_t^\dagger) \leq 1 - t^{\alpha+\epsilon-1}$. Now note that since $\epsilon < 1-\alpha$, $t^{\alpha+\epsilon-1}$ is monotonically decreasing as t grows, thus we have

$$\begin{aligned} \sum_{t=1}^{T}\mathbb{1}\left[a_t \neq a^\dagger\right] t^{\alpha+\epsilon-1} &\geq \sum_{t=N_T(a^\dagger)+1}^{T} t^{\alpha+\epsilon-1} \\ = \sum_{t=1}^{T} t^{\alpha+\epsilon-1} &- \sum_{t=1}^{N_T(a^\dagger)} t^{\alpha+\epsilon-1}. \end{aligned} \tag{17}$$

Next, by examining the area under curve, we obtain

$$\sum_{t=1}^{T} t^{\alpha+\epsilon-1} \geq \int_1^T t^{\alpha+\epsilon-1}dt = \frac{T^{\alpha+\epsilon}-1}{\alpha+\epsilon}. \tag{18}$$

Similarly, we can also derive

$$\sum_{t=1}^{N_T(a^\dagger)} t^{\alpha+\epsilon-1} \leq \int_0^{N_T(a^\dagger)} t^{\alpha+\epsilon-1}dt = \frac{\left(N_T(a^\dagger)\right)^{\alpha+\epsilon}}{\alpha+\epsilon}. \tag{19}$$

Therefore, we have

$$\begin{aligned} \sum_{t=1}^{T}\mathbb{1}\left[a_t \neq a^\dagger\right] t^{\alpha+\epsilon-1} &\geq \frac{1}{\alpha+\epsilon}\left(T^{\alpha+\epsilon} - \left(N_T(a^\dagger)\right)^{\alpha+\epsilon}\right) - \frac{1}{\alpha+\epsilon} \\ &= \frac{T^{\alpha+\epsilon}}{\alpha+\epsilon}\left(1 - (1 - \frac{T - N_T(a^\dagger)}{T})^{\alpha+\epsilon}\right) - \frac{1}{\alpha+\epsilon} \\ &\geq \frac{T^{\alpha+\epsilon}}{\alpha+\epsilon}\frac{T - N_T(a^\dagger)}{T}(\alpha+\epsilon) - \frac{1}{\alpha+\epsilon} \\ &= T^{\alpha+\epsilon} - T^{\alpha+\epsilon-1}N_T(a^\dagger) - \frac{1}{\alpha+\epsilon}. \end{aligned} \tag{20}$$

The inequality follows from the fact $(1-x)^c \leq 1 - cx$ for $x, c \in (0,1)$. Plug back in (16) we have

$$\begin{aligned} R_T &\geq \mathbf{E}\left[T^{\alpha+\epsilon} - T^{\alpha+\epsilon-1}N_T(a^\dagger) - \frac{1}{\alpha+\epsilon}\right] \\ &= T^{\alpha+\epsilon} - T^{\alpha+\epsilon-1}\mathbf{E}\left[N_T(a^\dagger)\right] - \frac{1}{\alpha+\epsilon}. \end{aligned} \tag{21}$$

Then note that $R_T \leq MT^\alpha$, thus we have

$$\mathbf{E}\left[N_T(a^\dagger)\right] \geq T - \frac{T^{1-\alpha-\epsilon}}{\alpha+\epsilon} - MT^{1-\epsilon}. \tag{22}$$

We now analyze the attack cost. Note that when $a_t \neq a^\dagger$, the per-round attack cost is $|\tilde{\mathcal{L}}_t(a_t) - \mathcal{L}_t(a_t)| \leq 1$. On the other hand, when $a_t = a^\dagger$, the per-round attack cost is

$$|\tilde{\mathcal{L}}_t(a^\dagger) - \mathcal{L}_t(a^\dagger)| \leq t^{\alpha+\epsilon-1} \tag{23}$$

Therefore, the expected attack cost is

$$\begin{aligned}
\mathbf{E}\left[C_T\right] &= \mathbf{E}\left[\sum_{t=1}^{T} |\tilde{\mathcal{L}}_t(a_t) - \mathcal{L}_t(a_t)|\right] \\
&\leq \mathbf{E}\left[\sum_{t=1}^{T} \mathbb{1}\left[a_t \neq a^\dagger\right]\right] + \mathbf{E}\left[\sum_{t=1}^{T} t^{\alpha+\epsilon-1}\right] \\
&\leq T - \mathbf{E}\left[N_T(a^\dagger)\right] + \sum_{t=1}^{T} t^{\alpha+\epsilon-1} \\
&\leq \frac{T^{1-\alpha-\epsilon}}{\alpha+\epsilon} + MT^{1-\epsilon} + \frac{1}{\alpha+\epsilon} T^{\alpha+\epsilon},
\end{aligned} \tag{24}$$

where we have used (22). □

Theorem 4.9. *Let $\rho \in (0, 1]$ be any constant. Define $\mathcal{T}_\rho = \{t \mid \mathcal{L}_t(a^\dagger) > 1 - \rho\}$, i.e., the set of rounds where $\mathcal{L}_t(a^\dagger)$ is within distance ρ to the maximum loss value. Let $|\mathcal{T}_\rho| = \tau$. Also assume that the attacker applies (6) to perform attack, then there exists a constant $M > 0$ such that the expected number of target arm selections satisfies*

$$\mathbf{E}\left[N_T(a^\dagger)\right] \geq T - \rho^{\frac{1}{\alpha+\epsilon-1}} - \tau - MT^\alpha/\rho, \tag{9}$$

and the cumulative attack cost satisfies

$$\mathbf{E}\left[C_T\right] \leq \rho^{\frac{1}{\alpha+\epsilon-1}} + \tau + MT^\alpha/\rho. \tag{10}$$

Proof. Let $t_0 = \rho^{\frac{1}{\alpha+\epsilon-1}}$ and define $\mathcal{T}_0 = \{t \mid t \geq t_0 \text{ and } t \notin T_\rho\}$. Note that $\epsilon < 1 - \alpha$, thus $t^{\alpha+\epsilon-1}$ is a monotonically decreasing function of t when $t \geq 1$, thus we have

$$t^{\alpha+\epsilon-1} \leq \rho, \forall t \in \mathcal{T}_0. \tag{25}$$

Therefore $\forall t \in \mathcal{T}_0$, $1 - t^{\alpha+\epsilon-1} \geq 1 - \rho$. Furthermore, note that $\forall t \in \mathcal{T}_0$, we must have $t \notin \mathcal{T}_\rho$, which means $\mathcal{L}_t(a^\dagger) \leq 1 - \rho$, thus the loss function prepared by the attacker (6) satifies

$$\forall t \in \mathcal{T}_0, \tilde{\mathcal{L}}_t(a) = \begin{cases} \mathcal{L}_t(a^\dagger) \leq 1 - \rho & \text{if } a = a^\dagger, \\ 1 & \text{otherwise,} \end{cases} \tag{26}$$

As a result, $\forall t \in \mathcal{T}_0$, whenever the bandit player selects a non-target arm $a_t \neq a^\dagger$, the player incurs at least regret ρ. Next note that by our assumption $|\mathcal{T}_\rho| = \tau$, thus

$$|\mathcal{T}_0| \geq |\{t \mid t \geq t_0\}| - |\mathcal{T}_\rho| \geq T - t_0 - \tau \tag{27}$$

Therefore, the total regret after attack is

$$\begin{aligned}
R_T &= \mathbf{E}\left[\sum_{t=1}^{T} \left(\tilde{\mathcal{L}}_t(a_t) - \tilde{\mathcal{L}}_t(a^\dagger)\right)\right] \\
&\geq \mathbf{E}\left[\sum_{t \in \mathcal{T}_0} \mathbb{1}\left[a_t \neq a^\dagger\right] \left(\tilde{\mathcal{L}}_t(a_t) - \tilde{\mathcal{L}}_t(a_t^\dagger)\right)\right] \\
&= \mathbf{E}\left[\sum_{t \in \mathcal{T}_0}^{T} \mathbb{1}\left[a_t \neq a^\dagger\right] \left(1 - \mathcal{L}_t(a_t^\dagger)\right)\right] \\
&\geq \mathbf{E}\left[\sum_{t \in \mathcal{T}_0}^{T} \mathbb{1}\left[a_t \neq a^\dagger\right] \rho\right] (\mathcal{L}_t(a^\dagger) \leq 1 - \rho) \\
&\geq \rho \mathbf{E}\left[\sum_{t \in \mathcal{T}_0}^{T} \mathbb{1}\left[a_t \neq a^\dagger\right]\right].
\end{aligned} \tag{28}$$

Since $R_T \leq MT^\alpha$ for some constant M, we have

$$\mathbf{E}\left[\sum_{t\in\mathcal{T}_0} \mathbb{1}\left[a_t \neq a^\dagger\right]\right] \leq MT^\alpha/\rho. \tag{29}$$

Therefore,

$$\begin{aligned}\mathbf{E}\left[\sum_{t=1}^{T} \mathbb{1}\left[a_t \neq a^\dagger\right]\right] &\leq T - |\mathcal{T}_0| + \mathbf{E}\left[\sum_{t\in\mathcal{T}_0} \mathbb{1}\left[a_t \neq a^\dagger\right]\right] \\ &\leq t_0 + \tau + MT^\alpha/\rho \\ &= \rho^{\frac{1}{\alpha+\epsilon-1}} + \tau + MT^\alpha/\rho.\end{aligned} \tag{30}$$

Thus we have

$$\mathbf{E}\left[N_T(a^\dagger)\right] = T - \mathbf{E}\left[\sum_{t=1}^{T} \mathbb{1}\left[a_t \neq a^\dagger\right]\right] \geq T - \rho^{\frac{1}{\alpha+\epsilon-1}} - \tau - MT^\alpha/\rho. \tag{31}$$

We now upper bound the attack cost.

$$\begin{aligned}\mathbf{E}\left[C_T\right] &= \mathbf{E}\left[\sum_{t=1}^{T} |\tilde{\mathcal{L}}(a_t) - \mathcal{L}(a_t)|\right] \\ &= \mathbf{E}\left[\sum_{t\notin\mathcal{T}_0} |\tilde{\mathcal{L}}(a_t) - \mathcal{L}(a_t)|\right] + \mathbf{E}\left[\sum_{t\in\mathcal{T}_0} |\tilde{\mathcal{L}}(a_t) - \mathcal{L}(a_t)|\right] \\ &\leq T - |\mathcal{T}_0| + \mathbf{E}\left[\sum_{t\in\mathcal{T}_0: a_t\neq a^\dagger} |\tilde{\mathcal{L}}(a_t) - \mathcal{L}(a_t)|\right] \left(\tilde{\mathcal{L}}_t(a^\dagger) = \mathcal{L}_t(a^\dagger)\right) \\ &= T - |\mathcal{T}_0| + \mathbf{E}\left[\sum_{t\in\mathcal{T}_0: a_t\neq a^\dagger} (1 - \mathcal{L}_t(a_t))\right] \\ &\leq T - |\mathcal{T}_0| + \mathbf{E}\left[\sum_{t\in\mathcal{T}_0: a_t\neq a^\dagger} 1\right] \\ &= t_0 + \tau + \mathbf{E}\left[\sum_{t\in\mathcal{T}_0} \mathbb{1}\left[a_t \neq a^\dagger\right]\right] \\ &\leq \rho^{\frac{1}{\alpha+\epsilon-1}} + \tau + MT^\alpha/\rho,\end{aligned} \tag{32}$$

where we have used (29) in the last inequality. □

We now derive a lower bound on the cumulative attack cost for the Exp3 victim algorithm. The Exp3 algorithm is described in Algorithm 1.

Algorithm 1 The Exponential Weighted Exploration Exploitation (Exp3) Algorithm

Parameters: $w_1 = (1, ..., 1)$, total horizon T, and a constant learning rate η.
for $t = 1, 2, \ldots, T$ **do**
  Define $\pi_t = \frac{w_t}{||w_t||_1}$
  Draw $a_t \sim \pi_t$, and observe loss $\ell_t = \mathcal{L}_t(a_t)$
  for $a = 1, ..., K$ **do**
    if $a \neq a_t$ **then**
      $w_{t+1,a} = w_{t,a}$
    else
      $w_{t+1,a} = w_{t,a} \exp(-\eta \frac{\ell_t}{\pi_{t,a}})$
    end if
  end for
end for

Lemma 5.1. *Assume the bandit player applies the Exp3 algorithm with parameter η (see (34) in the appendix) and initial arm selection probability π_1. Let the loss functions be $\mathcal{L}_{1:T}$. Then $\forall a \in \mathcal{A}$, the total number of rounds where a is selected, $N_T(a)$, satisfies*

$$\mathbf{E}\left[N_T(a)\right] \geq T\pi_1(a) - \eta T \sum_{t=1}^{T} \mathbf{E}\left[\pi_t(a)\mathcal{L}_t(a)\right], \tag{11}$$

where π_t is the arm selection probability at round t. Furthermore, since $\pi_t(a) \leq 1$, we have

$$\mathbf{E}\left[N_T(a)\right] \geq T\pi_1(a) - \eta T \sum_{t=1}^{T} \mathcal{L}_t(a). \tag{12}$$

Proof. The Exp3 algorithm maintains a weight $w_t(a)$ for each arm $a \in \mathcal{A}$, which is often initialized as $w_1(a) = 1, \forall a$. The probability of selecting any arm a in round t is computed as

$$\pi_t(a) = \frac{w_t(a)}{\sum_{a'} w_t(a')}. \tag{33}$$

The player selects an arm a_t by sampling according to π_t. After observing the loss $\ell_t = \mathcal{L}_t(a_t)$, the Exp3 updates the weights as below.

$$w_{t+1}(a) = w_t(a)\exp(-\eta\hat{\ell}_t(a)), \forall a, \tag{34}$$

where η is some constant to be selected later.

$$\hat{\ell}_t(a) = \begin{cases} \frac{\ell_t}{\pi_t(a_t)} = \frac{\mathcal{L}_t(a_t)}{\pi_t(a_t)} & \text{if } a = a_t \\ 0 & \text{otherwise} \end{cases} \tag{35}$$

Note that $\forall a \in \mathcal{A}$, we have

$$\begin{aligned} \pi_{t+1}(a) &= \frac{w_t(a)\exp(-\eta\hat{\ell}_t(a))}{\sum_{a'} w_t(a')\exp(-\eta\hat{\ell}_t(a'))} \\ &\geq \frac{w_t(a)\exp(-\eta\hat{\ell}_t(a))}{\sum_{a'} w_t(a')} \\ &= \pi_t(a)\exp(-\eta\hat{\ell}_t(a)). \end{aligned} \tag{36}$$

Now using the inequality $e^{-x} \geq 1 - x$, we have

$$\pi_{t+1}(a) \geq \pi_t(a)(1 - \eta\hat{\ell}_t(a)). \tag{37}$$

Taking expectation on both sides of (37), we have

$$\begin{aligned} \mathbf{E}\left[\pi_{t+1}(a)\right] &\geq \mathbf{E}\left[\pi_t(a)(1 - \eta\hat{\ell}_t(a)\right] \\ &= \mathbf{E}\left[\pi_t(a)\right] - \eta\mathbf{E}\left[\pi_t(a)\hat{\ell}_t(a)\right] \\ &= \mathbf{E}\left[\pi_t(a)\right] - \eta\mathbf{E}\left[\mathbf{E}\left[\pi_t(a)\pi_t(a)\frac{\mathcal{L}_t(a)}{\pi_t(a)} \mid \pi_t\right]\right] \\ &= \mathbf{E}\left[\pi_t(a)\right] - \eta\mathbf{E}\left[\mathbf{E}\left[\pi_t(a)\mathcal{L}_t(a) \mid \pi_t\right]\right] \\ &= \mathbf{E}\left[\pi_t(a)\right] - \eta\mathbf{E}\left[\pi_t(a)\mathcal{L}_t(a)\right] \end{aligned} \tag{38}$$

Now by telescoping, we have $\forall a \in \mathcal{A}$

$$\begin{aligned} \mathbf{E}\left[\pi_t(a)\right] &\geq \mathbf{E}\left[\pi_{t-1}(a)\right] - \eta\mathbf{E}\left[\pi_{t-1}(a)\mathcal{L}_{t-1}(a)\right] \\ &\geq \mathbf{E}\left[\pi_{t-2}(a)\right] - \eta\mathbf{E}\left[\pi_{t-1}(a)\mathcal{L}_{t-1}(a)\right] - \eta\mathbf{E}\left[\pi_{t-2}(a)\mathcal{L}_{t-2}(a)\right] \\ &\geq \ldots \geq \pi_1(a) - \eta\sum_{h=1}^{t-1}\mathbf{E}\left[\pi_h(a)\mathcal{L}_h(a)\right] \end{aligned} \tag{39}$$

For all a, the total number of rounds where a is selected is $N_T(a) = \sum_{t=1}^{T} \mathbb{1}[a_t = a]$. We have

$$\begin{aligned}\mathbf{E}[N_T(a)] &= \sum_{t=1}^{T} \mathbf{E}[\pi_t(a)] \\ &\geq T\pi_1(a) - \eta \sum_{t=1}^{T} \sum_{h=1}^{t-1} \mathbf{E}[\pi_h(a)\mathcal{L}_h(a)] \\ &\geq T\pi_1(a) - \eta T \sum_{h=1}^{T} \mathbf{E}[\pi_h(a)\mathcal{L}_h(a)].\end{aligned} \tag{40}$$

□

Theorem 5.2. *Assume some victim-agnostic attack algorithm achieves* $\mathbf{E}\left[N_T(a^\dagger)\right] = T - o(T)$ *on all victim bandit algorithms that has regret rate* $O(T^\alpha)$*, where* $\alpha \in [\frac{1}{2}, 1)$*. Then there exists a bandit task such that the attacker must induce at least expected attack cost* $\mathbf{E}[C_T] = \Omega(T^\alpha)$ *on some victim algorithm. Specifically, one such victim is the Exp3 algorithm with parameter* $\eta = \Theta(T^{-\alpha})$*.*

Proof. Since we want to derive a lower bound on the expected attack cost for *victim-agnostic* attackers, it suffices to choose a special bandit task, and a victim bandit algorithm that guarantees $O(T^\alpha)$ regret, such that any victim-agnostic attacker must induce at least $\Omega(T)$ expected attack cost on the chosen task and the victim algorithm. The proof consists of three steps as below.

(1). We first construct the following special bandit task. The player has two arms s a_1, a_2. The loss functions are the following.

$$\mathcal{L}_t(a) = \begin{cases} 0 & \text{if } a = a_1 \\ 0.5 & \text{if } a = a_2 \end{cases} \tag{41}$$

The attacker target arm is $a^\dagger = a_2$. Note that this is an easy attack scenario since $\mathcal{L}_t(a^\dagger) = 0.5 < 1, \forall t$. Let the loss functions manipulated by the attacker be $\tilde{\mathcal{L}}_t$. Suppose the attack is successful, i.e., $\mathbf{E}\left[\tilde{N}_T(a_2)\right] = T - o(T)$, where $\tilde{N}_T$ is the number of arm selections under the manipulated loss $\tilde{\mathcal{L}}_t$. Then it must be the case that $\mathbf{E}\left[\tilde{N}_T(a_1)\right] = o(T)$.

Using the lower bound (11) in Lemma 5.1, we have that for arm a_1,

$$\mathbf{E}\left[\tilde{N}_T(a_1)\right] \geq T\tilde{\pi}_1(a_1) - \eta T \sum_{t=1}^{T} \mathbf{E}\left[\tilde{\pi}_t(a_1)\tilde{\mathcal{L}}_t(a_1)\right], \tag{42}$$

where $\tilde{\pi}_t$ is the arm selection probability under attack. Therefore, we must have

$$T\tilde{\pi}_1(a_1) - \eta T \sum_{t=1}^{T} \mathbf{E}\left[\tilde{\pi}_t(a_1)\tilde{\mathcal{L}}_t(a_1)\right] = o(T) \tag{43}$$

which results in

$$\begin{aligned}\sum_{t=1}^{T} \mathbf{E}\left[\tilde{\pi}_t(a_1)\tilde{\mathcal{L}}_t(a_1)\right] &= \frac{T\tilde{\pi}_1(a_1) - o(T)}{\eta T} \\ &= \frac{\tilde{\pi}_1(a)}{\eta} - \frac{o(T)}{\eta T}.\end{aligned} \tag{44}$$

Note that in the RHS of (44), as $T \to \infty$, we have

$$\left(\frac{o(T)}{\eta T}\right) / \left(\frac{\tilde{\pi}_1(a)}{\eta}\right) = \frac{o(T)}{T\tilde{\pi}_1(a)} \to 0. \tag{45}$$

Therefore, as $T \to \infty$, we have

$$\sum_{t=1}^{T} \mathbf{E}\left[\tilde{\pi}_t(a_1)\tilde{\mathcal{L}}_t(a_1)\right] \to \frac{\tilde{\pi}_1(a)}{\eta}. \tag{46}$$

(2). We now choose a particular victim bandit algorithm that guarantees $O(T^\alpha)$ regret rate. Specifically, we choose the Exp3 algorithm that uses learning rate $\eta = \beta T^{-\alpha}$ for some constant $\beta > 0$ and $\alpha \geq \frac{1}{2}$. In the standard analysis of Exp3 algorithm, the regret bound is $R_T \leq \frac{1}{\eta}\log K + \frac{\eta}{2}TK$. Plug in $\eta = \beta T^{-\alpha}$, the regret of the chosen victim Exp3 algorithm is

$$R_T \leq \frac{1}{\beta}T^\alpha \log K + \frac{\beta}{2}T^{1-\alpha}K = O(T^\alpha) + O(T^{1-\alpha}) = O(T^\alpha), \tag{47}$$

where the last equality is due to $\alpha \geq \frac{1}{2}$ and thus $O(T^{1-\alpha})$ is negligible compared to $O(T^\alpha)$. Therefore, the victim bandit algorithm guarantees regret rate $O(T^\alpha)$.

(3). Finally, we prove a lower bound on the attack cost if some victim-agnostic attacker performs attack on the bandit task and the victim algorithm chosen above. Note that since $\mathcal{L}_t(a_1) = 0$ and $\tilde{\mathcal{L}}_t(a_1) \geq 0$, thus we always have $\tilde{\mathcal{L}}_t(a_1) = |\tilde{\mathcal{L}}_t(a_1) - \mathcal{L}_t(a_1)|$. Therefore, the expected attack cost is

$$\begin{aligned}
&\mathbf{E}\left[\sum_{t=1}^{T}\sum_{a}\tilde{\pi}_t(a)|\tilde{\mathcal{L}}_t(a) - \mathcal{L}_t(a)|\right] \\
&\geq \mathbf{E}\left[\sum_{t=1}^{T}\tilde{\pi}_t(a_1)|\tilde{\mathcal{L}}_t(a_1) - \mathcal{L}_t(a_1)|\right] \\
&= \sum_{t=1}^{T}\mathbf{E}\left[\tilde{\pi}_t(a_1)\tilde{\mathcal{L}}_t(a_1)\right] \to \frac{\tilde{\pi}_1(a)}{\eta} \\
&= \frac{\tilde{\pi}_1(a)}{\beta}T^\alpha = \Omega(T^\alpha),
\end{aligned} \tag{48}$$

where we have used (46). □