Appendix

A. Preliminaries

A.1. Table of Notation

For ease of reading, we define here key notation that will be used in this Appendix.

T	:	The horizon.
Δ_j	:	The gap between the mean of the optimal arm and the mean of arm j, $\Delta_j = \mu^* - \mu_j$.
$\tilde{\Delta}_m$	:	The approximation to Δ_j at round m of the ODAAF algorithm, $\tilde{\Delta}_m = \frac{1}{2^m}$.
n_m	:	The number of samples of an active arm j ODAAF needs by the end of round m.
ν_m	:	The number of times each arm is played in phase m, $\nu_m = n_m - n_{m-1}$.
d	:	The bound on the delay in the case of bounded delay.
m_j	:	The first round of the ODAAF algorithm where $\tilde{\Delta}_m < \Delta_j/2$.
M_j	:	The random variable representing the round arm j is eliminated in.
$T_j(m)$	:	The set of all time point where arm j is played up to (and including) round m.
X_t	:	The reward received at time t (from any possible past plays of the algorithm).
$R_{t,j}$	:	The reward generated by playing arm j at time t.
$\tau_{t,j}$	:	The delay associated with playing arm j at time t.
$\mathbb{E}[\tau]$	:	The expected delay (assuming i.i.d. delays).
$\mathbb{V}(\tau)$	:	The variance of the delay (assuming i.i.d. delays).
$\bar{X}_{m,j}$	:	The estimated reward of arm j in phase m. See Algorithm 1 for the definition.
S_m	:	The start point of the mth phase. See Appendix A.2 for more details.
U_m	:	The end point of the mth phase. See Appendix A.2 for more details.
$S_{m,j}$	:	The start point of phase m of playing arm j. See Appendix A.2 for more details.
$U_{m,j}$	:	The end point of phase m of playing arm j. See Appendix A.2 for more details.
$\mathcal{A}_m$	:	The set of active arms in round m of the ODAAF algorithm.
$A_{i,t}, B_{i,t}, C_{i,t}$	:	The contribution of the reward generated at time t in certain intervals relating to phase i to the corruption. See (11) for the exact definitions.
$\mathcal{G}_t$	:	The smallest σ-algebra containing all information up to time t, see (8) for a definition.

A.2. Beginning and End of Phases

We formalize here some notation that will be used throughout the analysis to denote the start and end points of each phase. Define the random variables S_i and U_i for each phase $i = 1, \dots, m$ to be the start and end points of the phase. Then let $S_{i,j}$, $U_{i,j}$ denote the start and end points of playing arm j in phase i. See Figure 4 for details. By convention, let $S_{i,j} = U_{i,j} = \infty$ if arm j is not active in phase i, $S_i = U_i = \infty$ if the algorithm never reaches phase i and let $S_{0,j} = U_{0,j} = S_0 = U_0 = 0$ for all j. It is important to point out that n_m are deterministic so at the end of any phase $m - 1$, once we have eliminated sub-optimal arms, we also know which arms are in $\mathcal{A}_m$ and consequently the start and end points of phase m. Furthermore, since we play arms in a given order, we also know the specific rounds when we start and finish playing each active arm in phase m. Hence, at any time step t in phase m, S_m, U_m, S_{m+1} and $U_{m,j}, S_{m,j}$ for all active arms $j \in \mathcal{A}_m$ will be known. More formally, define the filtration $\{\mathcal{G}_t\}_{t=0}^{\infty}$ where

$$\mathcal{G}_t = \sigma(X_1, \dots, X_t, \tau_{1,J_1}, \dots, \tau_{t,J_t}, R_{1,J_1}, \dots, R_{t,J_t}, J_1, \dots, J_t) \tag{8}$$

and $\mathcal{G}_0 = \{\emptyset, \Omega\}$. This means the joint events like $\{S_i \le t\} \cap \{S_{i,j} = s'\} \in \mathcal{G}_t$ for all $s' \in \mathbb{N}, j \in \mathcal{A}$.

A.3. Useful Results

For our analysis, we will need Freedman's version of Bernstein's inequality for the right-tail of martingales with bounded increments:

Theorem 10 (Freedman's version of Bernstein's inequality; Theorem 1.6 of Freedman (1975)) *Let $\{Y_k\}_{k=0}^{\infty}$ be a real-valued martingale with respect to the filtration $\{\mathcal{F}_k\}_{k=0}^{\infty}$ with increments $\{Z_k\}_{k=1}^{\infty}$: $\mathbb{E}[Z_k|\mathcal{F}_{k-1}] = 0$ and $Z_k = Y_k - Y_{k-1}$, for $k = 1, 2, \dots$. Assume that the difference sequence is uniformly bounded on the right: $Z_k \le b$ almost surely for $k = 1, 2, \dots$. Define the predictable variation process $W_k = \sum_{j=1}^{k} \mathbb{E}[Z_j^2|\mathcal{F}_{j-1}]$ for $k = 1, 2, \dots$. Then, for all $t \ge 0$,*

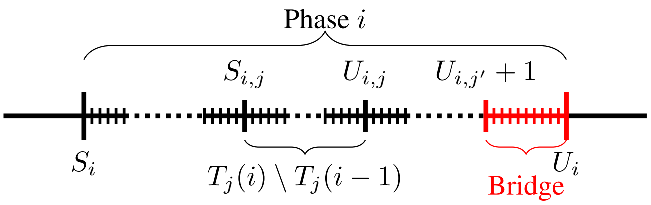


Figure 4: An example of phase i of our algorithm. Here j' is the last active arm played in phase i.

$\sigma^2 > 0$,

$$\mathbb{P}\left(\exists k \geq 0 : Y_k \geq t \textit{ and } W_k \leq \sigma^2\right) \leq \exp\left\{-\frac{t^2/2}{\sigma^2 + bt/3}\right\}.$$

This result implies that if for some deterministic constant, σ^2, $W_k \leq \sigma^2$ holds almost surely, then $\mathbb{P}(Y_k \geq t) \leq \exp\left\{-\frac{t^2/2}{\sigma^2+bt/3}\right\}$ holds for any $t \geq 0$.

We will also make use of the following technical lemma which combines the Hoeffding-Azuma inequality and Doob's optional skipping theorem (Theorem 2.3 in Chapter VII of Doob (1953))):

Lemma 11 *Fix the positive integers m, n and let $a, c \in \mathbb{R}$. Let $\mathcal{F} = \{\mathcal{F}_t\}_{t=0}^n$ be a filtration, $(\epsilon_t, Z_t)_{t=1,2,\ldots,n}$ be a sequence of $\{0,1\}\times\mathbb{R}$-valued random variables such that for $t \in \{1,2,\ldots,n\}$, ϵ_t is $\mathcal{F}_{t-1}$-measurable, Z_t is $\mathcal{F}_t$-measurable, $\mathbb{E}[Z_t|\mathcal{F}_{t-1}] = 0$ and $Z_t \in [a, a+c]$. Further, assume that $\sum_{s=1}^n \epsilon_s \leq m$ with probability one. Then, for any $\lambda > 0$,*

$$\mathbb{P}\left(\sum_{t=1}^n \epsilon_t Z_t \geq \lambda\right) \leq \exp\left\{-\frac{2\lambda^2}{c^2 m}\right\}. \tag{9}$$

Proof: This lemma appeared in a slightly more general form (where $n = \infty$ is allowed) as Lemma A.1 in the paper by Szita & Szepesvári (2011) so we refer the reader to the proof there. □

B. Results for Known and Bounded Expected Delay

B.1. High Probability Bounds

Lemma 1 *Under Assumption 1 and the choice of n_m given by (2), the estimates $\bar{X}_{m,j}$ constructed by Algorithm 1 satisfy the following: For every fixed arm j and phase m, with probability $1 - \frac{3}{T\tilde{\Delta}_m^2}$, either $j \notin \mathcal{A}_m$, or:*

$$\bar{X}_{m,j} - \mu_j \leq \tilde{\Delta}_m/2\,.$$

Proof: Let

$$w_m = \frac{4\log(T\tilde{\Delta}_m^2)}{3n_m} + \sqrt{\frac{2\log(T\tilde{\Delta}_m^2)}{n_m}} + \frac{3m\mathbb{E}[\tau]}{n_m}. \tag{10}$$

We first show that with probability greater than $1 - \frac{3}{T\tilde{\Delta}_m^2}$, $j \notin \mathcal{A}_m$ or $\frac{1}{n_m}\sum_{t\in T_j(m)}(X_t - \mu_j) \leq w_m$.

For arm j and phase m, assume $j \in \mathcal{A}_m$. For notational simplicity we will use in the following $\mathbb{I}_i\{H\} := \mathbb{I}\{H \cap \{j \in \mathcal{A}_i\}\} \leq \mathbb{I}\{H\}$ for any event H. If $j \in \mathcal{A}_m$ for a particular experiment ω then $\mathbb{I}_i(H)(\omega) = \mathbb{I}(H)(\omega)$. Then for any phase $i \leq m$ and time t, define,

$$A_{i,t} = R_{t,J_t}\mathbb{I}\{\tau_{t,J_t} + t \geq S_i\}, \quad B_{i,t} = R_{t,J_t}\mathbb{I}\{\tau_{t,J_t} + t \geq S_{i,j}\}, \quad C_{i,t} = R_{t,J_t}\mathbb{I}\{\tau_{t,J_t} + t > U_{i,j}\}, \tag{11}$$

and note that since $S_{i,j} = U_{i,j} = \infty$ if arm j is not active in phase i, we have the equalities $\mathbb{I}_i\{\tau_{t,J_t} + t \geq S_{i,j}\} = \mathbb{I}\{\tau_{t,J_t} + t \geq S_{i,j}\}$ and $\mathbb{I}_i\{\tau_{t,J_t} + t > U_{i,j}\} = \mathbb{I}\{\tau_{t,J_t} + t > U_{i,j}\}$. Define the filtration $\{\mathcal{G}_s\}_{s=0}^\infty$ by $\mathcal{G}_0 = \{\Omega, \emptyset\}$ and

$$\mathcal{G}_t = \sigma(X_1, \ldots, X_t, J_1, \ldots, J_t, \tau_{1,J_1}, \ldots, \tau_{t,J_t}, R_{1,J_1}, \ldots R_{t,J_t}). \tag{12}$$

Then, we use the decomposition,

$$\begin{aligned}
\sum_{i=1}^{m}\sum_{t=S_{i,j}}^{U_{i,j}}(X_t-\mu_j) &\leq \sum_{i=1}^{m}\Bigg(\sum_{t=S_{i-1,j}}^{S_{i,j}-1} R_{t,J_t}\mathbb{I}_i\{\tau_{t,J_t}+t\geq S_{i,j}\}+\sum_{t=S_{i,j}}^{U_{i,j}}(R_{t,J_t}-\mu_j)-\sum_{t=S_{i,j}}^{U_{i,j}} R_{t,J_t}\mathbb{I}_i\{\tau_{t,J_t}+t>U_{i,j}\}\Bigg)\\
&\leq \sum_{i=1}^{m}\Bigg(\sum_{t=S_{i-1,j}}^{S_{i}-1} R_{t,J_t}\mathbb{I}\{\tau_{t,J_t}+t\geq S_{i}\}+\sum_{t=S_{i}}^{S_{i,j}-1} R_{t,J_t}\mathbb{I}\{\tau_{t,J_t}+t\geq S_{i,j}\}\\
&\qquad +\sum_{t=S_{i,j}}^{U_{i,j}}(R_{t,J_t}-\mu_j)-\sum_{t=S_{i,j}}^{U_{i,j}} R_{t,J_t}\mathbb{I}\{\tau_{t,J_t}+t>U_{i,j}\}\Bigg)\\
&=\sum_{i=1}^{m}\Bigg(\sum_{t=S_{i-1,j}}^{S_{i}-1} A_{i,t}+\sum_{t=S_{i}}^{S_{i,j}-1} B_{i,t}+\sum_{t=S_{i,j}}^{U_{i,j}}(R_{t,J_t}-\mu_j)-\sum_{t=S_{i,j}}^{U_{i,j}} C_{i,t}\Bigg)\\
&=\sum_{i=1}^{m}\sum_{t=S_{i,j}}^{U_{i,j}}(R_{t,J_t}-\mu_j)+\sum_{t=1}^{S_{m,j}} Q_t-\sum_{t=1}^{U_{m,j}} P_t\\
&=\underbrace{\sum_{i=1}^{m}\sum_{t=S_{i,j}}^{U_{i,j}}(R_{t,J_t}-\mu_j)}_{\text{Term I.}}+\underbrace{\sum_{t=1}^{S_{m,j}}(Q_t-\mathbb{E}[Q_t|\mathcal{G}_{t-1}])}_{\text{Term II.}}+\underbrace{\sum_{t=1}^{U_{m,j}}(\mathbb{E}[P_t|\mathcal{G}_{t-1}]-P_t)}_{\text{Term III.}} \qquad (13)\\
&\qquad +\underbrace{\Bigg(\sum_{t=1}^{S_{m,j}}\mathbb{E}[Q_t|\mathcal{G}_{t-1}]-\sum_{t=1}^{U_{m,j}}\mathbb{E}[P_t|\mathcal{G}_{t-1}]\Bigg)}_{\text{Term IV.}},
\end{aligned}$$

where,

$$\begin{aligned}
Q_t &= \sum_{i=1}^{m}(A_{i,t}\mathbb{I}\{S_{i-1,j}\leq t\leq S_i-1\}+B_{i,t}\mathbb{I}\{S_i\leq t\leq S_{i,j}-1\})\\
P_t &= \sum_{i=1}^{m} C_{i,t}\mathbb{I}\{S_{i,j}\leq t\leq U_{i,j}\}.
\end{aligned}$$

Recall that the filtration $\{\mathcal{G}_s\}_{s=0}^{\infty}$ is defined by $\mathcal{G}_0=\{\Omega,\emptyset\}$, $\mathcal{G}_t=\sigma(X_1,\ldots,X_t,J_1,\ldots,J_t,\tau_{1,J_1},\ldots,\tau_{t,J_t},R_{1,J_1},\ldots R_{t,J_t})$ and we have defined $S_{i,j}=\infty$ if arm j is eliminated before phase i and $S_i=\infty$ if the algorithm stops before reaching phase i.

Outline of proof We will bound each term of the above decomposition in (13) in turn, however first we need to prove several intermediary results. For term II., we will use Freedman's inequality so we first need Lemma 12 to show that $Z_t=Q_t-\mathbb{E}[Q_t|\mathcal{G}_{t-1}]$ is a martingale difference and Lemma 13 to bound the variance of the sum of the Z_t's. Similarly, for term III., in Lemma 14, we show that $Z'_t=\mathbb{E}[P_t|\mathcal{G}_{t-1}]-P_t$ is a martingale difference and bound its variance in Lemma 15. In Lemma 16, we consider term IV. and bound the conditional expectations of $A_{i,t},B_{i,t},C_{i,t}$. Finally, in Lemma 17, we bound term I. using Lemma 11. We then combine the bounds on all terms together to conclude the proof.

Lemma 12 *Let $Y_s=\sum_{t=1}^{s}(Q_t-\mathbb{E}[Q_t|\mathcal{G}_{t-1}])$ for all $s\geq 1$, $Y_0=0$. Then $\{Y_s\}_{s=0}^{\infty}$ is a martingale with respect to the filtration $\{\mathcal{G}_s\}_{s=0}^{\infty}$ with increments $Z_s=Y_s-Y_{s-1}=Q_s-\mathbb{E}[Q_s|\mathcal{G}_{s-1}]$ satisfying $\mathbb{E}[Z_s|\mathcal{G}_{s-1}]=0, Z_s\leq 1$ for all $s\geq 1$.*

Proof: To show $\{Y_s\}_{s=0}^{\infty}$ is a martingale with respect to $\{\mathcal{G}_s\}_{s=0}^{\infty}$, we need to show that Y_s is G_s measurable for all s and $\mathbb{E}[Y_s|\mathcal{G}_{s-1}]=Y_{s-1}$.

Measurability: First note that by definition of $\mathcal{G}_s$, τ_{t,J_t},R_{t,J_t} are all $\mathcal{G}_s$-measurable for $t\leq s$. Then, for each i, either t is in a phase later than i so $S_{i-1,j}$ and S_i are $\mathcal{G}_t$-measurable, or $S_{i-1,j}$ and S_i are not $\mathcal{G}_t$-measurable, but $\mathbb{I}\{t\geq S_{i,j}\}=0$ so $\mathbb{I}\{t\geq S_{i,j}\}$ is $\mathcal{G}_t$-measurable. In the first case, since $S_{i-1,j}$ and S_i are $\mathcal{G}_t$-measurable $A_{i,t}\mathbb{I}\{S_{i-1,j}\leq t\leq S_i-\nu_i\}$ is $\mathcal{G}_t$-measurable. In the second case, $A_{i,t}\mathbb{I}\{S_{i-1,j}\leq t\leq S_i-1\}=A_{i,t}\mathbb{I}\{\{S_{i-1,j}\leq t\}\mathbb{I}\{t\leq S_i-1\}=0$ so it is also

$\mathcal{G}_t$-measurable. Similarly, if t is after S_i , S_i and $S_{i,j}$ will be $\mathcal{G}$-measurable or $\mathbb{I}\{S_i \leq t \leq S_{i,j} - 1\} = 0$. In both cases, $B_{i,t}\mathbb{I}\{S_i \leq t \leq S_{i,j} - 1\}$ is $\mathcal{G}_t$-measurable. Hence, Q_t is $\mathcal{G}_t$-measurable, and also Q_t is $\mathcal{G}_s$ measurable for any $s \geq t$. It then follows that Y_s is $\mathcal{G}_s$-measurable for all s.

Expectation: Since Q_t is $\mathcal{G}_s$ measurable for all $t \leq s$,

$$\begin{aligned}
\mathbb{E}[Y_s|\mathcal{G}_{s-1}] &= \mathbb{E}\bigg[\sum_{t=1}^{s}(Q_t - \mathbb{E}[Q_t|\mathcal{G}_{t-1}])|\mathcal{G}_{s-1}\bigg] \\
&= \mathbb{E}\bigg[\sum_{t=1}^{s-1}(Q_t - \mathbb{E}[Q_t|\mathcal{G}_{t-1}])|\mathcal{G}_{s-1}\bigg] + \mathbb{E}[(Q_s - \mathbb{E}[Q_s|\mathcal{G}_{s-1}])|\mathcal{G}_{s-1}] \\
&= \sum_{t=1}^{s-1}(Q_t - \mathbb{E}[Q_t|\mathcal{G}_{t-1}]) + \mathbb{E}[Q_s|\mathcal{G}_{s-1}] - \mathbb{E}[Q_s|\mathcal{G}_{s-1}] \\
&= \sum_{t=1}^{s-1}(Q_t - \mathbb{E}[Q_t|\mathcal{G}_{t-1}]) = Y_{s-1}
\end{aligned}$$

Hence, $\{Y_s\}_{s=0}^{\infty}$ is a martingale with respect to the filtration $\{\mathcal{G}_s\}_{s=0}^{\infty}$.

Increments: For any $s = 1, \ldots,$ we have that

$$Z_s = Y_s - Y_{s-1} = \sum_{t=1}^{s}(Q_t - \mathbb{E}[Q_t|\mathcal{G}_{t-1}]) - \sum_{t=1}^{s-1}(Q_t - \mathbb{E}[Q_t|\mathcal{G}_{t-1}]) = Q_s - \mathbb{E}[Q_s|\mathcal{G}_{s-1}].$$

Then,

$$\mathbb{E}[Z_s|\mathcal{G}_{s-1}] = \mathbb{E}[Q_s - \mathbb{E}[Q_s|\mathcal{G}_{s-1}]|\mathcal{G}_{s-1}] = \mathbb{E}[Q_s|\mathcal{G}_{s-1}] - \mathbb{E}[Q_s|\mathcal{G}_{s-1}] = 0.$$

Lastly, since for any t, there is only one i where one of $\mathbb{I}\{S_{i-1,j} \leq t \leq S_i - 1\} = 1$ or $\mathbb{I}\{S_i \leq t \leq S_{i,j} - 1\} = 1$ (and they cannot both be one), and since $R_{t,J_t} \in [0,1]$, $A_{i,t}, B_{i,t} \leq 1$, so it follows that $Z_s = Q_s - \mathbb{E}[Q_s|\mathcal{G}_{s-1}] \leq 1$ for all s. □

Lemma 13 *For any t, let $Z_t = Q_t - \mathbb{E}[Q_t|\mathcal{G}_{t-1}]$, then, for any $s < S_{m,j}$,*

$$\sum_{t=1}^{s}\mathbb{E}[Z_t^2|\mathcal{G}_{t-1}] \leq 2m\mathbb{E}[\tau].$$

Proof: First note that

$$\begin{aligned}
\sum_{t=1}^{s}\mathbb{E}[Z_t^2|\mathcal{G}_{t-1}] &= \sum_{t=1}^{s}\mathbb{V}(Q_t|\mathcal{G}_{t-1}) \leq \sum_{t=1}^{s}\mathbb{E}[Q_t^2|\mathcal{G}_{t-1}] \\
&= \sum_{t=1}^{s}\mathbb{E}\bigg[\bigg(\sum_{i=1}^{m}(A_{i,t}\mathbb{I}\{S_{i-1,j} \leq t \leq S_i - 1\} + B_{i,t}\mathbb{I}\{S_i \leq t \leq S_{i,j} - 1\})\bigg)^2\bigg|\mathcal{G}_{t-1}\bigg].
\end{aligned}$$

Then, given $\mathcal{G}_{t-1}$, all indicator terms $\mathbb{I}\{S_{i-1,j} \leq t \leq S_i - 1\}$ and $\mathbb{I}\{S_i \leq S_{i,j} - 1\}$ for all $i = 1, \ldots, m$ are measurable and only one can be non zero. Hence, all interaction terms in the expansion of the quadratic are 0 and so we are left with

$$\begin{aligned}
\sum_{t=1}^{s}\mathbb{E}[Z_t^2|\mathcal{G}_{t-1}] &\leq \sum_{t=1}^{s}\mathbb{E}\bigg[\bigg(\sum_{i=1}^{m}(A_{i,t}\mathbb{I}\{S_{i-1,j} \leq t \leq S_i - 1\} + B_{i,t}\mathbb{I}\{S_i \leq t \leq S_{i,j} - 1\})\bigg)^2\bigg|\mathcal{G}_{t-1}\bigg] \\
&= \sum_{t=1}^{s}\mathbb{E}\bigg[\sum_{i=1}^{m}(A_{i,t}^2\mathbb{I}\{S_{i-1,j} \leq t \leq S_i - 1\}^2 + B_{i,t}^2\mathbb{I}\{S_i \leq t \leq S_{i,j} - 1\}^2)\bigg|\mathcal{G}_{t-1}\bigg] \\
&= \sum_{i=1}^{m}\sum_{t=1}^{s}\mathbb{E}[A_{i,t}^2\mathbb{I}\{S_{i-1,j} \leq t \leq S_i - 1\}|\mathcal{G}_{t-1}] + \sum_{i=1}^{m}\sum_{t=1}^{s}\mathbb{E}[B_{i,t}^2\mathbb{I}\{S_i \leq t \leq S_{i,j} - 1\}|\mathcal{G}_{t-1}]
\end{aligned}$$

$$\leq \sum_{i=1}^{m} \sum_{t=S_{i-1,j}}^{S_i-1} \mathbb{E}[A_{i,t}^2|\mathcal{G}_{t-1}] + \sum_{i=1}^{m} \sum_{t=S_i}^{S_{i,j}-1} \mathbb{E}[B_{i,t}^2|\mathcal{G}_{t-1}].$$

Then, for any $i \geq 1$,

$$\begin{aligned}
\sum_{t=S_{i-1,j}}^{S_i-1} \mathbb{E}[A_{i,t}^2|\mathcal{G}_{t-1}] &= \sum_{t=S_{i-1,j}}^{S_i-1} \mathbb{E}[R_{t,J_t}^2 \mathbb{I}\{\tau_{t,J_t} + t \geq S_i\}|\mathcal{G}_{t-1}] \\
&\leq \sum_{t=S_{i-1,j}}^{S_i-1} \mathbb{E}[\mathbb{I}\{\tau_{t,J_t} + t \geq S_i\}|\mathcal{G}_{t-1}] \\
&= \sum_{s=0}^{\infty} \sum_{s'=s}^{\infty} \mathbb{I}\{S_{i-1,j} = s, S_i = s'\} \sum_{t=s}^{s'-1} \mathbb{E}[\mathbb{I}\{\tau_{t,J_t} + t \geq S_i\}|\mathcal{G}_{t-1}] \\
&= \sum_{s=0}^{\infty} \sum_{s'=s}^{\infty} \sum_{t=s}^{s'-1} \mathbb{E}[\mathbb{I}\{S_{i-1,j} = s, S_i = s', \tau_{t,J_t} + t \geq S_i\}|\mathcal{G}_{t-1}] \\
&\qquad\qquad\qquad\qquad \text{(Since } \{t \geq S_{i-1,j}, S_i = s'\} \in \mathcal{G}_{t-1}\text{)} \\
&= \sum_{s=0}^{\infty} \sum_{s'=s}^{\infty} \sum_{t=s}^{s'-1} \mathbb{E}[\mathbb{I}\{S_{i-1,j} = s, S_i = s', \tau_{t,J_t} + t \geq s'\}|\mathcal{G}_{t-1}] \\
&= \sum_{s=0}^{\infty} \sum_{s'=s}^{\infty} \mathbb{I}\{S_{i-1,j} = s, S_i = s'\} \sum_{t=s}^{s'-1} \mathbb{P}(\tau_{t,J_t} + t \geq s') \\
&\qquad\qquad\qquad\qquad \text{(Since } \{t \geq S_{i-1,j}, S_i = s'\} \in \mathcal{G}_{t-1}\text{)} \\
&\leq \sum_{s=0}^{\infty} \sum_{s'=s}^{\infty} \mathbb{I}\{S_{i-1,j} = s, S_i = s'\} \sum_{l=0}^{\infty} \mathbb{P}(\tau > l) \\
&\leq \mathbb{E}[\tau].
\end{aligned}$$

Likewise, for any $i \geq 1$,

$$\begin{aligned}
\sum_{t=S_i}^{S_{i,j}-1} \mathbb{E}[B_{i,t}^2|\mathcal{G}_{t-1}] &= \sum_{t=S_i}^{S_{i,j}-1} \mathbb{E}[R_{t,J_t}^2 \mathbb{I}\{\tau_{t,J_t} + t \geq S_{i,j}\}|\mathcal{G}_{t-1}] \\
&\leq \sum_{t=S_i}^{S_{i,j}-1} \mathbb{E}[\mathbb{I}\{\tau_{t,J_t} + t \geq S_{i,j}\}|\mathcal{G}_{t-1}] \\
&= \sum_{s=0}^{\infty} \sum_{s'=s}^{\infty} \mathbb{I}\{S_i = s, S_{i,j} = s'\} \sum_{t=s}^{s'-1} \mathbb{E}[\mathbb{I}\{\tau_{t,J_t} + t \geq S_{i,j}\}|\mathcal{G}_{t-1}] \\
&= \sum_{s=0}^{\infty} \sum_{s'=s}^{\infty} \sum_{t=s}^{s'-1} \mathbb{E}[\mathbb{I}\{S_i = s, S_{i,j} = s', \tau_{t,J_t} + t \geq S_{i,j}\}|\mathcal{G}_{t-1}] \\
&\qquad\qquad\qquad\qquad \text{(Since } \{t \geq S_i, S_{i,j} = s'\} \in \mathcal{G}_{t-1}\text{)} \\
&= \sum_{s=0}^{\infty} \sum_{s'=s}^{\infty} \sum_{t=s}^{s'-1} \mathbb{E}[\mathbb{I}\{S_i = s, S_{i,j} = s', \tau_{t,J_t} + t \geq s'\}|\mathcal{G}_{t-1}] \\
&= \sum_{s=0}^{\infty} \sum_{s'=s}^{\infty} \mathbb{I}\{S_i = s, S_{i,j} = s'\} \sum_{t=s}^{s'-1} \mathbb{P}(\tau_{t,J_t} + t \geq s') \\
&\qquad\qquad\qquad\qquad \text{(Since } \{t \geq S_i, S_{i,j} = s'\} \in \mathcal{G}_{t-1}\text{)} \\
&\leq \sum_{s=0}^{\infty} \sum_{s'=s}^{\infty} \mathbb{I}\{S_i = s, S_{i,j} = s'\} \sum_{l=0}^{\infty} \mathbb{P}(\tau \geq l)
\end{aligned}$$

$$\le \mathbb{E}[\tau].$$

Hence, combining both terms and summing over the phases m gives the result. □

Lemma 14 *Let $Y'_s = \sum_{t=1}^s (\mathbb{E}[P_s|\mathcal{G}_{s-1}] - P_s)$ for all $s \ge 1$, $Y'_0 = 0$. Then $\{Y'_s\}_{s=0}^\infty$ is a martingale with respect to the filtration $\{\mathcal{G}_s\}_{s=0}^\infty$ with increments $Z'_s = Y'_s - Y'_{s-1} = \mathbb{E}[P_s|\mathcal{G}_{s-1}] - P_s$ satisfying $\mathbb{E}[Z'_s|\mathcal{G}_{s-1}] = 0, Z'_s \le 1$ for all $s \ge 1$.*

Proof: The proof is similar to that of Lemma 12. To show $\{Y'_s\}_{s=0}^\infty$ is a martingale with respect to $\{\mathcal{G}_s\}_{s=0}^\infty$, we need to show that Y'_s is G_s measurable for all s and $\mathbb{E}[Y'_s|\mathcal{G}_{s-1}] = Y'_{s-1}$.

Measurability: As before, by definition of $\mathcal{G}_s$, τ_{t,J_t}, R_{t,J_t} are all $\mathcal{G}_s$-measurable for $t \le s$. Also, we can reduce measurability again to measurability of $\mathbb{I}\{\tau_{s,J_s} + s \ge U_{i,j}, S_{i,j} \le s \le U_{i,j}\}$. But, $\{U_{i,j} = s'\} \cap \{S_{i,j} \le s\} \in \mathcal{G}_s$ for all $s' \in \mathbb{N}$ and Y'_s is adapted to $\mathcal{G}_s$.

Increments: For any $s \ge 1$, we have that

$$Z'_s = Y'_s - Y'_{s-1} = \sum_{t=1}^{s} (\mathbb{E}[P_t|\mathcal{G}_{t-1}] - P_t) - \sum_{t=1}^{s-1} (\mathbb{E}[P_t|\mathcal{G}_{t-1}] - P_t) = \mathbb{E}[P_s|\mathcal{G}_{s-1}] - P_s.$$

Then,

$$\mathbb{E}[Z'_s|\mathcal{G}_{s-1}] = \mathbb{E}[\mathbb{E}[P_s|\mathcal{G}_{s-1}] - P_s|\mathcal{G}_{s-1}] = \mathbb{E}[P_s|\mathcal{G}_{s-1}] - \mathbb{E}[P_s|\mathcal{G}_{s-1}] = 0.$$

Lastly, since for any t and $\omega \in \Omega$, there is at most one i for which $\mathbb{I}\{S_{i,j} \le t \le U_{i,j}\} = 1$, and by definition of R_{t,J_t}, $C_{i,t} \le 1$, so it follows that $Z'_s = \mathbb{E}[P_s|\mathcal{G}_{s-1}] - P_s \le 1$ for all s. □

Lemma 15 *For any t, let $Z'_t = \mathbb{E}[P_t|\mathcal{G}_{t-1}] - P_t$, then*

$$\sum_{t=1}^{U_{m,j}} \mathbb{E}[Z_t'^2|\mathcal{G}_{t-1}] \le m\mathbb{E}[\tau].$$

Proof: The proof is similar to that of Lemma 13. First note that

$$\begin{aligned} \sum_{t=1}^{U_{m,j}} \mathbb{E}[Z_t'^2|\mathcal{G}_{t-1}] &= \sum_{t=1}^{U_{m,j}} \mathbb{V}(P_t|\mathcal{G}_{t-1}) \le \sum_{t=1}^{U_{m,j}} \mathbb{E}[P_t^2|\mathcal{G}_{t-1}] \\ &= \sum_{t=1}^{U_{m,j}} \mathbb{E}\left[\left(\sum_{i=1}^{m} (C_{i,t}\mathbb{I}\{S_{i,j} \le t \le U_{i,j}\}\right)^2 |\mathcal{G}_{t-1}\right]. \end{aligned}$$

Then, given $\mathcal{G}_{t-1}$, all indicator terms $\mathbb{I}\{S_{i,j} \le t \le U_{i,j}\}$ for $i = 1, \dots, m$ are measurable and at most one can be non zero. Hence, all interaction terms are 0 and so we are left with

$$\begin{aligned} \sum_{t=1}^{U_{m,j}} \mathbb{E}[Z_t'^2|\mathcal{G}_{t-1}] &\le \sum_{t=1}^{U_{m,j}} \mathbb{E}\left[\left(\sum_{i=1}^{m} (C_{i,t}\mathbb{I}\{S_{i,j} \le t \le U_{i,j}\}\right)^2 |\mathcal{G}_{t-1}\right] \\ &= \sum_{i=1}^{m} \sum_{t=1}^{U_{m,j}} \mathbb{E}[C_{i,t}^2 \mathbb{I}\{S_{i,j} \le t \le U_{i,j}\}|\mathcal{G}_{t-1}] \\ &\le \sum_{i=1}^{m} \sum_{t=S_{i,j}}^{U_{i,j}} \mathbb{E}[C_{i,t}^2|\mathcal{G}_{t-1}] && \text{(since the indicator is } \mathcal{G}_{t-1}\text{-measurable)} \\ &= \sum_{i=1}^{m} \sum_{t=S_{i,j}}^{U_{i,j}} \mathbb{E}[R_{t,J_t}^2 \mathbb{I}\{\tau_{t,J_t} + t > U_{i,j}\}|\mathcal{G}_{t-1}] \end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{i=1}^{m} \sum_{t=S_{i,j}}^{U_{i,j}} \mathbb{E}[\mathbb{I}\{\tau_{t,J_t} + t > U_{i,j}\}|\mathcal{G}_{t-1}] \\
&= \sum_{i=1}^{m} \sum_{s=0}^{\infty} \sum_{s'=s}^{\infty} \mathbb{I}\{S_{i,j} = s, U_{i,j} = s'\} \sum_{t=s}^{s'} \mathbb{E}[\mathbb{I}\{\tau_{t,J_t} + t > U_{i,j}\}|\mathcal{G}_{t-1}] \\
&= \sum_{i=1}^{m} \sum_{s=0}^{\infty} \sum_{s'=s}^{\infty} \sum_{t=s}^{s'} \mathbb{E}[\mathbb{I}\{S_{i,j} = s, U_{i,j} = s', \tau_{t,J_t} + t > U_{i,j}\}|\mathcal{G}_{t-1}] \\
&= \sum_{i=1}^{m} \sum_{s=0}^{\infty} \sum_{s'=s}^{\infty} \sum_{t=s}^{s'} \mathbb{E}[\mathbb{I}\{S_{i,j} = s, U_{i,j} = s', \tau_{t,J_t} + t > s'\}|\mathcal{G}_{t-1}] \\
&= \sum_{i=1}^{m} \sum_{s=0}^{\infty} \sum_{s'=s}^{\infty} \mathbb{I}\{S_{i,j} = s, U_{i,j} = s'\} \sum_{t=s}^{s'} \mathbb{P}(\tau_{t,J_t} + t > s') \\
&\leq \sum_{i=1}^{m} \sum_{s=0}^{\infty} \sum_{s'=s}^{\infty} \mathbb{I}\{S_{i,j} = s, U_{i,j} = s'\} \sum_{l=0}^{\infty} \mathbb{P}(\tau > l) \\
&\leq \sum_{i=1}^{m} \mathbb{E}[\tau] = m\mathbb{E}[\tau].
\end{aligned}
$$

□

Lemma 16 *For $A_{i,t}, B_{i,t}$ and $C_{i,t}$ defined as in (11), let $\nu_i = n_i - n_{i-1}$ be the number of times each arm is played in phase i and j_i' be the arm played directly before arm j in phase i. Then, it holds that, for any arm j and phase $i \geq 1$,*

(i) $\displaystyle\sum_{t=S_{i-1,j}}^{S_i-1} \mathbb{E}[A_{i,t}|\mathcal{G}_{t-1}] \leq \mathbb{E}[\tau]$

(ii) $\displaystyle\sum_{t=S_i}^{S_{i,j}-1} \mathbb{E}[B_{i,t}|\mathcal{G}_{t-1}] \leq \mathbb{E}[\tau] + \mu_{j_i'} \sum_{l=0}^{\nu_i} \mathbb{P}(\tau > l)$

(iii) $\displaystyle\sum_{t=S_{i,j}}^{U_{i,j}} \mathbb{E}[C_{i,t}|\mathcal{G}_{t-1}] = \mu_j \sum_{l=0}^{\nu_i} \mathbb{P}(\tau > l)$

Proof: We prove each statement individually. Several of the proofs are similar to those appearing in Lemmas 13 and 15.

Statement (i):

$$
\begin{aligned}
\sum_{t=S_{i-1,j}}^{S_i-1} \mathbb{E}[A_{i,t}|\mathcal{G}_{t-1}] &\leq \sum_{t=S_{i-1,j}}^{S_i-1} \mathbb{E}[\mathbb{I}\{\tau_{t,J_t} + t \geq S_i\}|\mathcal{G}_{t-1}] \\
&= \sum_{s=0}^{\infty} \sum_{s'=s}^{\infty} \mathbb{I}\{S_{i-1,j} = s, S_i = s'\} \sum_{t=s}^{s'-1} \mathbb{E}[\mathbb{I}\{\tau_{t,J_t} + t \geq S_i\}|\mathcal{G}_{t-1}] \\
&= \sum_{s=0}^{\infty} \sum_{s'=s}^{\infty} \sum_{t=s}^{s'-1} \mathbb{E}[\mathbb{I}\{S_{i-1,j} = s, S_i = s', \tau_{t,J_t} + t \geq S_i\}|\mathcal{G}_{t-1}] \\
&\qquad\qquad\qquad\qquad (\text{Since } \{t \geq S_{i-1,j}, S_i = s'\} \in \mathcal{G}_{t-1}) \\
&= \sum_{s=0}^{\infty} \sum_{s'=s}^{\infty} \sum_{t=s}^{s'-1} \mathbb{E}[\mathbb{I}\{S_{i-1,j} = s, S_i = s', \tau_{t,J_t} + t \geq s'\}|\mathcal{G}_{t-1}]
\end{aligned}
$$

$$= \sum_{s=0}^{\infty}\sum_{s'=s}^{\infty}\mathbb{I}\{S_{i-1,j}=s, S_i=s'\}\sum_{t=s}^{s'-1}\mathbb{P}(\tau_{t,J_t}+t\geq s')$$

$$\text{(Since } \{t\geq S_{i-1,j}, S_i=s'\}\in\mathcal{G}_{t-1}\text{)}$$

$$\leq \sum_{s=0}^{\infty}\sum_{s'=s}^{\infty}\mathbb{I}\{S_{i-1,j}=s, S_i=s'\}\sum_{l=0}^{\infty}\mathbb{P}(\tau>l)$$

$$= \sum_{l=0}^{\infty}\mathbb{P}(\tau>l)=\mathbb{E}[\tau].$$

Statement (iii):

$$\sum_{t=S_{i,j}}^{U_{i,j}}\mathbb{E}[C_{i,t}|\mathcal{G}_{t-1}] = \sum_{t=S_{i,j}}^{U_{i,j}}\mathbb{E}[R_{t,J_t}\mathbb{I}\{\tau_{t,J_t}+t>U_{i,j}\}|\mathcal{G}_{t-1}]$$

$$= \sum_{s=0}^{\infty}\sum_{s'=s}^{\infty}\mathbb{I}\{S_{i,j}=s, U_{i,j}=s'\}\sum_{t=s}^{s'}\mathbb{E}[R_{t,J_t}\mathbb{I}\{\tau_{t,J_t}+t>U_{i,j}\}|\mathcal{G}_{t-1}]$$

$$= \sum_{s=0}^{\infty}\sum_{s'=s}^{\infty}\sum_{t=s}^{s'}\mathbb{E}[R_{t,J_t}\mathbb{I}\{S_{i,j}=s, U_{i,j}=s', \tau_{t,J_t}+t>U_{i,j}\}|\mathcal{G}_{t-1}]$$

$$\text{(Since } \{S_{i,j}=s, U_{i,j}=s'\}\in\mathcal{G}_{t-1}\text{ for } s\leq t\text{)}$$

$$= \sum_{s=0}^{\infty}\sum_{s'=s}^{\infty}\sum_{t=s}^{s'}\mathbb{E}[R_{t,J_t}\mathbb{I}\{S_{i,j}=s, U_{i,j}=s', \tau_{t,J_t}+t>s'\}|\mathcal{G}_{t-1}]$$

$$= \sum_{s=0}^{\infty}\sum_{s'=s}^{\infty}\mathbb{I}\{S_{i,j}=s, U_{i,j}=s'\}\sum_{t=s}^{s'}\mu_j\mathbb{P}(\tau_{t,J_t}+t>s')$$

$$\text{(Since } \{S_{i,j}=s, U_{i,j}=s'\}\in\mathcal{G}_{t-1}\text{ and given } \mathcal{G}_{t-1}, R_{t,J_t}\text{ and } \tau_{t,J_t}\text{ are independent)}$$

$$= \sum_{s=0}^{\infty}\sum_{s'=s}^{\infty}\mathbb{I}\{S_{i,j}=s, U_{i,j}=s'\}\mu_j\sum_{l=0}^{\nu_i}\mathbb{P}(\tau>l)$$

$$= \mu_j\sum_{l=0}^{\nu_i}\mathbb{P}(\tau>l)$$

Statement (ii): For statement (ii), we have that for $(i,j)\neq(1,1)$,

$$\sum_{t=S_i}^{S_{i,j}-1}\mathbb{E}[B_{i,t}|\mathcal{G}_{t-1}] = \sum_{t=S_i}^{S_{i,j}-\nu_{i-1}-2}\mathbb{E}[B_{i,t}|\mathcal{G}_{t-1}] + \sum_{t=S_{i,j}-\nu_{i-1}-1}^{S_{i,j}-1}\mathbb{E}[B_{i,t}|\mathcal{G}_{t-1}].$$

Then, $S_{i,j}$ is $\mathcal{G}_{t-1}$ measurable for $t\geq S_i$, so we can use the same technique as for statement (i) to bound the first term. For the second term, since we will only be playing arm j'_i for $S_{i,j}-\nu_{i-1}-1,\dots,S_{i,j}-1$, we can use the same technique as for statement (iii). Hence,

$$\sum_{t=S_i}^{S_{i,j}-1}\mathbb{E}[B_{i,t}|\mathcal{G}_{t-1}] \leq \sum_{l=\nu_{i-1}+1}^{\infty}\mathbb{P}(\tau>l)+\mu_{j'_i}\sum_{l=0}^{\nu_{i-1}}\mathbb{P}(\tau>l)\leq\mathbb{E}[\tau]+\mu_{j'_i}\sum_{l=0}^{\nu_i}\mathbb{P}(\tau>l).$$

Note that, for $(i,j)=(1,1)$, the amount seeping in will be 0, so using $\nu_0=0, \mu'_{1_1}=0$, the result trivially holds. Hence the result holds for all $i,j\geq 1$. □

Lemma 17 *For any arm $j\in\{1,\dots,K\}$ and phase m, it holds that for any $\lambda>0$,*

$$\mathbb{P}\bigg(\sum_{t\in T_j(m)}(R_{t,j}-\mu_j)\geq\lambda\bigg)\leq\exp\bigg\{-\frac{2\lambda^2}{n_m}\bigg\}.$$

Proof: The result follows from Lemma 11. When applying this lemma, we use $n = T$, $m = n_m$, for $t = 0, 1, \dots, T$ set $\mathcal{F}_t = \sigma(X_1, \dots, X_t, R_{1,j}, \dots, R_{t,j})$ and for $t = 1, 2, \dots, T$ define $Z_t = R_{t,j} - \mu_j$ and $\epsilon_t = \mathbb{I}\{J_t = j, t \leq U_{m,j}\}$. Note that $T_j(m) = \{t \in \{1, \dots, T\} : \epsilon_t = 1\}$ and hence $\sum_{t \in T_j(m)} (R_{t,j} - \mu_j) = \sum_{t=1}^{T} \epsilon_t (R_{t,j} - \mu_j)$. Further, $\sum_{t=1}^{T} \epsilon_t = |T_j(m)| \leq n_m$ with probability one.

Fix $1 \leq t \leq T$. We now argue that ϵ_t is $\mathcal{F}_{t-1}$-measurable. First, notice that by the definition of ODAAF, the index M of the phase that t belongs to can be calculated based on the observations $X_1, \dots, X_{t-1}$ up to time $t-1$. Since $t \leq U_{m,j}$ is equivalent to whether for this phase index M, the inequality $M \leq m$ holds, it follows that $\{t \leq U_{m,j}\}$ is $\mathcal{F}_{t-1}$-measurable. The same holds for $\{J_t = j\}$ for the same reason. Hence, it follows that ϵ_t is indeed $\mathcal{F}_{t-1}$-measurable.

Now, Z_t is $\mathcal{F}_t$-measurable as $R_{t,j}$ is clearly $\mathcal{F}_t$-measurable. Furthermore, by our assumptions on $(R_{t,j})_{t,j}$ and $(X_t)_t$, $\mathbb{E}[R_{t,j} | \mathcal{F}_{t-1}] = \mu_j$ also holds, implying that Z_t also satisfies the conditions of the lemma with $a = -\mu_j$ and $c = 1$. Thus, the result follows by applying Lemma 11. □

We now bound each term of the decomposition in (13) in turn.

Bounding Term I.: For Term I., we use Lemma 17 to get that with probability greater than $1 - \frac{1}{T\tilde{\Delta}_m^2}$,

$$\sum_{i=1}^{m} \sum_{t=S_{i,j}}^{U_{i,j}} (R_{t,J_t} - \mu_j) \leq \sqrt{\frac{n_m \log(T\tilde{\Delta}_m^2)}{2}}.$$

Bounding Term II.: For Term II., we will use Freedmans inequality (Theorem 10). From Lemma 12, $\{Y_s\}_{s=0}^{\infty}$ with $Y_s = \sum_{t=1}^{s} (Q_t - \mathbb{E}[Q_t | \mathcal{G}_{t-1}])$ is a martingale with respect to $\{\mathcal{G}_s\}_{s=0}^{\infty}$ with increments $\{Z_s\}_{s=0}^{\infty}$ satisfying $\mathbb{E}[Z_s | \mathcal{G}_{s-1}] = 0$ and $Z_s \leq 1$ for all s. Further, by Lemma 13, $\sum_{t=1}^{s} \mathbb{E}[Z_t^2 | \mathcal{G}_{t-1}] \leq 2m\mathbb{E}[\tau] \leq \frac{6m \times 2^m \mathbb{E}[\tau]}{12} \leq n_m/12$ with probability 1. Hence we can apply Freedman's inequality to get that with probability greater than $1 - \frac{1}{T\tilde{\Delta}_m^2}$,

$$\sum_{t=1}^{S_{m,j}} (Q_t - \mathbb{E}[Q_t | \mathcal{G}_{t-1}]) \leq \frac{2}{3} \log(T\tilde{\Delta}_m^2) + \sqrt{\frac{1}{12} n_m \log(T\tilde{\Delta}_m^2)}.$$

Bounding Term III.: For Term III., we again use Freedman's inequality (Theorem 10) but using Lemma 14 to show that $\{Y'_s\}_{s=0}^{\infty}$ with $Y'_s = \sum_{t=1}^{s} (\mathbb{E}[P_t | \mathcal{G}_{t-1}] - P_t)$ is a martingale with respect to $\{\mathcal{G}_s\}_{s=0}^{\infty}$ with increments $\{Z'_s\}_{s=0}^{\infty}$ satisfying $\mathbb{E}[Z'_s | \mathcal{G}_{s-1}] = 0$ and $Z'_s \leq 1$ for all s. Further, by Lemma 15, $\sum_{t=1}^{s} \mathbb{E}[Z_t^2 | \mathcal{G}_{t-1}] \leq m\mathbb{E}[\tau] \leq n_m/12$ with probability 1. Hence, with probability greater than $1 - \frac{1}{T\tilde{\Delta}_m^2}$,

$$\sum_{t=1}^{U_{m,j}} (\mathbb{E}[P_t | \mathcal{G}_{t-1}] - P_t) \leq \frac{2}{3} \log(T\tilde{\Delta}_m) + \sqrt{\frac{1}{12} n_m \log(T\tilde{\Delta}_m^2)}.$$

Bounding Term IV.: We bound term IV. using Lemma 16,

$$\begin{aligned}
&\sum_{t=1}^{S_{m,j}} \mathbb{E}[Q_t | \mathcal{G}_{t-1}] - \sum_{t=1}^{U_{m,j}} \mathbb{E}[P_t | \mathcal{G}_{t-1}] \\
&= \sum_{t=1}^{S_{m,j}} \mathbb{E}\left[\sum_{i=1}^{m} (A_{i,t} \mathbb{I}\{S_{i-1,j} \leq t \leq S_i - 1\} + B_{i,t} \mathbb{I}\{S_i \leq t \leq S_{i,j} - 1\}) \middle| \mathcal{G}_{t-1} \right] \\
&\qquad - \sum_{t=1}^{U_{m,j}} \mathbb{E}\left[\sum_{i=1}^{m} C_{i,t} \mathbb{I}\{S_{i,j} \leq t \leq U_{i,j}\} \middle| \mathcal{G}_{t-1} \right] \\
&= \sum_{i=1}^{m} \sum_{t=1}^{S_{m,j}} \mathbb{E}[A_{i,t} \mathbb{I}\{S_{i-1,j} \leq t \leq S_i - 1\} | \mathcal{G}_{t-1}] + \sum_{i=1}^{m} \sum_{t=1}^{S_{m,j}} \mathbb{E}[B_{i,t} \mathbb{I}\{S_i \leq t \leq S_{i,j} - 1\} | \mathcal{G}_{t-1}] \\
&\qquad - \sum_{i=1}^{m} \sum_{t=1}^{U_{m,j}} \mathbb{E}[C_{i,t} \mathbb{I}\{S_{i,j} \leq t \leq U_{i,j}\} | \mathcal{G}_{t-1}]
\end{aligned}$$

$$= \sum_{i=1}^{m} \left(\sum_{t=S_{i-1,j}}^{S_i - 1} \mathbb{E}[A_{i,t}|\mathcal{G}_{t-1}] + \sum_{t=S_i}^{S_{i,j}-1} \mathbb{E}[B_{i,t}|\mathcal{G}_{t-1}] - \sum_{t=S_{i,j}}^{U_{i,j}} \mathbb{E}[C_{i,t}|\mathcal{G}_{t-1}] \right)$$

$$\leq \sum_{i=1}^{m} \left(2\mathbb{E}[\tau] + \mu_{j_i'} \sum_{l=0}^{\nu_i} \mathbb{P}(\tau > l) - \mu_j \sum_{l=0}^{\nu_i} \mathbb{P}(\tau > l) \right) \leq 3m\mathbb{E}[\tau].$$

since $R_{t,j} \in [0,1]$.

Combining all terms: To get the final high probability bound, we sum the bounds for each term I.-IV.. Then, with probability greater than $1 - \frac{3}{T\tilde{\Delta}_m^2}$, either $j \notin \mathcal{A}_m$ or arm j is played n_m times by the end of phase m and

$$\frac{1}{n_m} \sum_{t \in T_j(m)} (X_t - \mu_j) \leq \frac{4\log(T\tilde{\Delta}_m^2)}{3n_m} + \left(\frac{2}{\sqrt{12}} + \frac{1}{\sqrt{2}} \right) \sqrt{\frac{\log(T\tilde{\Delta}_m^2)}{n_m}} + \frac{3m\mathbb{E}[\tau]}{n_m}$$

$$\leq \frac{4\log(T\tilde{\Delta}_m^2)}{3n_m} + \sqrt{\frac{2\log(T\tilde{\Delta}_m^2)}{n_m}} + \frac{3m\mathbb{E}[\tau]}{n_m} = w_m.$$

Defining $\mathbf{n_m}$: Setting

$$n_m = \left\lceil \frac{1}{\tilde{\Delta}_m^2} \left(\sqrt{2\log(T\tilde{\Delta}_m^2)} + \sqrt{2\log(T\tilde{\Delta}_m^2) + \frac{8}{3}\tilde{\Delta}_m \log(T\tilde{\Delta}_m^2) + 6\tilde{\Delta}_m m\mathbb{E}[\tau]} \right)^2 \right\rceil. \tag{14}$$

ensures that $w_m \leq \frac{\tilde{\Delta}_m}{2}$ which concludes the proof. □

B.2. Regret Bounds

Here we prove the regret bound in Theorem 2 under Assumption 1 and the choice of n_m given by (14). Under Assumption 1, the bridge period is not necessary so the results here hold for the version of Algorithm 1 with the bridge period omitted. Note that if we were to include the bridge period, we would be playing each arm at most $2n_m$ times by the end of phase m so our regret would simply increase by a factor of 2.

Theorem 2 *Under Assumption 1, the expected regret of Algorithm 1 is upper bounded as*

$$\mathbb{E}[\mathfrak{R}_T] \leq \sum_{\substack{j=1 \\ j \neq j^*}}^{K} O\left(\frac{\log(T\Delta_j^2)}{\Delta_j} + \log(1/\Delta_j)\mathbb{E}[\tau] \right). \tag{5}$$

Proof: Our proof is a restructuring of the proof of (Auer & Ortner, 2010). For any arm j, define M_j to be the random variable representing the phase when arm j is eliminated in. We set $M_j = \infty$ if the arm did not get eliminated before time step T. Note that if M_j is finite, $j \in \mathcal{A}_{M_j}$ (this also means that $\mathcal{A}_{M_j}$ is well-defined) and if $\mathcal{A}_{M_j+1}$ is also defined (M_j is not the last phase) then $j \notin \mathcal{A}_{M_j+1}$. We also let m_j denote the phase arm j *should* be eliminated in, that is $m_j = \min\{\, m \geq 1 \,:\, \tilde{\Delta}_m < \frac{\Delta_j}{2} \,\}$. From the definition of $\tilde{\Delta}_m$ in our algorithm, we get the relations

$$2^{m_j} = \frac{1}{\tilde{\Delta}_{m_j}} \leq \frac{4}{\Delta_j} < \frac{1}{\tilde{\Delta}_{m_j+1}} \quad \text{and} \quad \frac{\Delta_j}{4} \leq \tilde{\Delta}_{m_j} \leq \frac{\Delta_j}{2}\,. \tag{15}$$

Define $N_j = \sum_{t=1}^{T} \mathbb{I}\{J_t = j\}$ be the number of times arm j is used and let $\mathfrak{R}_T^{(j)} = N_j\Delta_j$ be the "pseudo"-regret contribution from each arm $1 \leq j \leq K$ so that $\mathbb{E}[\mathfrak{R}_T] = \mathbb{E}\left[\sum_{j=1}^{K} \mathfrak{R}_T^{(j)}\right]$. Let M^* be the round when the optimal arm j^* is eliminated. Hence,

$$\mathbb{E}[\mathfrak{R}_T] = \mathbb{E}\left[\sum_{j=1}^{K} \mathfrak{R}_T^{(j)}\right] = \underbrace{\mathbb{E}\left[\sum_{j=1}^{K} \mathfrak{R}_T^{(j)} \mathbb{I}\{M^* \geq m_j\}\right]}_{\text{Term I.}} + \underbrace{\mathbb{E}\left[\sum_{j=1}^{K} \mathfrak{R}_T^{(j)} \mathbb{I}\{M^* < m_j\}\right]}_{\text{Term II.}}.$$

We will bound the regret in each of these cases in turn. To do so, we need the following results which consider the probabilities of confidence bounds failing and arms being eliminated in the incorrect rounds.

Lemma 18 *For any suboptimal arm j,*

$$\mathbb{P}(M_j > m_j \textit{ and } M^* \geq m_j) \leq \frac{6}{T\tilde{\Delta}_{m_j}^2}\,.$$

Proof: Define

$$E = \{\bar{X}_{m_j,j} \leq \mu_j + w_{m_j}\} \quad \text{and} \quad H = \{\bar{X}_{m_j,j^*} > \mu^* - w_{m_j}\}\,.$$

If both E and F occur, it follows that,

$$\begin{aligned}
\bar{X}_{m_j,j} &\leq \mu_j + w_{m_j} \\
&= \mu_j^* - \Delta_j + w_{m_j} && \text{(since } \Delta_j = \mu_{j^*} - \mu_j\text{)} \\
&\leq \bar{X}_{m_j,j^*} + w_{m_j} - \Delta_j + w_{m_j} \\
&< \bar{X}_{m_j,j^*} - 2\tilde{\Delta}_{m_j} + 2w_{m_j} && \text{(by (15))} \\
&\leq \bar{X}_{m_j,j^*} - \tilde{\Delta}_{m_j} && \text{(since } n_m \text{ is such that } w_m \leq \tilde{\Delta}_m/2\text{)}
\end{aligned}$$

and arm j would be eliminated. Hence, on the event $M^* \geq m_j$, $M_j \leq m_j$. Thus, $M^* \geq m_j$ and $M_j > m_j$ imply that either E or H does not occur and so $\mathbb{P}(M_j > m_j \text{ and } M^* \geq m_j) \leq \mathbb{P}(\{E^c \cup H^c\} \cap \{j, j^* \in \mathcal{A}_{m_j}\}) \leq \mathbb{P}(E^c \cap j \in \mathcal{A}_{m_j}) + \mathbb{P}(H^c \cap j^* \in \mathcal{A}_{m_j})$. Using Lemma 1, we then get that,

$$\mathbb{P}(M_j \geq m_j \text{ and } M^* \geq m_j) \leq \frac{6}{T\tilde{\Delta}_{m_j}^2}.$$

□

Note that the random set $\mathcal{A}_m$ may not be defined for certain $\omega \in \Omega$. That is, $\mathcal{A}_m$ is a partially defined random element. For convenience, we modify the definition of $\mathcal{A}_m$ so that it is an emptyset for any ω when it is not defined by the previous definition. Define the event $F_j(m) = \{\bar{X}_{m,j^*} < \bar{X}_{m,j} - \tilde{\Delta}_m\} \cap \{j, j^* \in \mathcal{A}_m\}$ to be the event that arm j^* is eliminated by arm j in phase m (given our note on $\mathcal{A}_m$, this is well-defined). The probability of this occurring is bounded in the following lemma.

Lemma 19 *The probability that the optimal arm j^* is eliminated in round $m < \infty$ by the suboptimal arm j is bounded by*

$$\mathbb{P}(F_j(m)) \leq \frac{6}{T\tilde{\Delta}_m^2}\,.$$

Proof: First note that for a suboptimal arm j to eliminate arm j^* in round m, both j and j^* must be active in round m and $\bar{X}_{m,j} - w_m > \bar{X}_{m,j^*} + w_m$. Hence,

$$\mathbb{P}(F_j(m)) = \mathbb{P}(j, j^* \in \mathcal{A}_m \text{ and } \bar{X}_{m,j} - w_m > \bar{X}_{m,j^*} + w_m)$$

Then, observe that if

$$E = \{\bar{X}_{m,j} \leq \mu_j + w_m\} \quad \text{and} \quad H = \{\bar{X}_{m,j^*} > \mu^* - w_m\}$$

both hold in round m, it follows that,

$$\bar{X}_{m,j} - \tilde{\Delta}_m \leq \mu_j + w_m - \tilde{\Delta}_m \leq \mu_j - \frac{\tilde{\Delta}_m}{2} \leq \mu_{j^*} - \frac{\tilde{\Delta}_m}{2} \leq \bar{X}_{m,j^*} + w_m - \frac{\tilde{\Delta}_m}{2} \leq \bar{X}_{m,j^*}$$

so arm j^* will not be eliminated by arm j in round m. Hence, for arm j^* to be eliminated by arm j in round m, one of E or H must not occur and the probability of this is bounded by Lemma 1 as,

$$\mathbb{P}(F_j(m)) \leq \mathbb{P}((E^C \cup H^C) \cap (j, j^* \in \mathcal{A}_m)) \leq \mathbb{P}(E^C \cap (j \in \mathcal{A}_m)) + \mathbb{P}(H^C \cap (j^* \in \mathcal{A}_m)) \leq \frac{6}{T\tilde{\Delta}_m^2}.$$

□

We now return to bounding the expected regret in each of the two cases.

Bounding Term I. To bound the first term, we consider the cases where arm j is eliminated in or before the correct round ($M_j \le m_j$) and where arm j is eliminated late ($M_j > m_j$). Then, by Lemma 18,

$$\begin{aligned}
&\mathbb{E}\Bigg[\sum_{j=1}^K \mathfrak{R}_T^{(j)} \mathbb{I}\{M^* \ge m_j\}\Bigg] \\
&\quad = \mathbb{E}\Bigg[\sum_{j=1}^K \mathfrak{R}_T^{(j)} \mathbb{I}\{M^* \ge m_j\}\mathbb{I}\{M_j \le m_j\}\Bigg] + \mathbb{E}\Bigg[\sum_{j=1}^K \mathfrak{R}_T^{(j)} \mathbb{I}\{M^* \ge m_j\}\mathbb{I}\{M_j > m_j\}\Bigg] \\
&\quad \le \sum_{j=1}^K \mathbb{E}[\mathfrak{R}_T^{(j)} \mathbb{I}\{M_j \le m_j\}] + \sum_{j=1}^K \mathbb{E}[T\Delta_j \mathbb{I}\{M^* \ge m_j, M_j > m_j\}] \\
&\quad \le \sum_{j=1}^K \Delta_j n_{m_j} + \sum_{j=1}^K T\Delta_j \mathbb{P}(M_j > m_j \text{ and } M^* \ge m_j) \\
&\quad \le \sum_{j=1}^K \Delta_j n_{m_j} + \sum_{j=1}^K T\Delta_j \frac{6}{T\tilde{\Delta}_{m_j}^2} \\
&\quad \le \sum_{j=1}^K \left(\Delta_j n_{m_j} + \frac{24}{\tilde{\Delta}_{m_j}}\right) \le \sum_{j=1}^K \left(\frac{96}{\Delta_j} + \Delta_j n_{m_j}\right).
\end{aligned}$$

Bounding Term II For the second term, let $m_{\max} = \max_{j \neq j^*} m_j$. and recall that N_j is the total number of times arm j is played. Then,

$$\begin{aligned}
\mathbb{E}\Bigg[\sum_{j=1}^K \mathfrak{R}_T^{(j)} \mathbb{I}\left\{M^* < m_j\right\}\Bigg] &= \mathbb{E}\Bigg[\sum_{m=1}^{m_{\max}} \sum_{j:m<m_j} \mathfrak{R}_T^{(j)} \mathbb{I}\{M^* = m\}\Bigg] \\
&= \sum_{m=1}^{m_{\max}} \mathbb{E}\Bigg[\mathbb{I}\{M^* = m\} \sum_{j:m_j>m} \mathfrak{R}_T^{(j)}\Bigg] \\
&= \sum_{m=1}^{m_{\max}} \mathbb{E}\Bigg[\mathbb{I}\{M^* = m\} \sum_{j:m_j>m} N_j\Delta_j\Bigg] \\
&\le \sum_{m=1}^{m_{\max}} \mathbb{E}\Bigg[\mathbb{I}\{M^* = m\} T \max_{j:m_j>m} \Delta_j\Bigg] \\
&\le \sum_{m=1}^{m_{\max}} 4\mathbb{P}(M^* = m) T\tilde{\Delta}_m .
\end{aligned}$$

Now consider the probability that arm j^* is eliminated in round m. This includes the probability that it is eliminated by any suboptimal arm. For arm j^* to be eliminated in round m by a suboptimal arm with $m_j < m$, arm j must be active ($M_j > m_j$) and the optimal arm must also have been active in round m_j ($M^* \ge m_j$). Using this, it follows that

$$\begin{aligned}
\mathbb{P}(M^* = m) = \sum_{j=1}^K \mathbb{P}(F_j(m)) &= \sum_{j:m_j<m} \mathbb{P}(F_j(m)) + \sum_{j:m_j \ge m} \mathbb{P}(F_j(m)) \\
&\le \sum_{j:m_j<m} \mathbb{P}(M_j > m_j \text{ and } M^* \ge m_j) + \sum_{j:m_j \ge m} \mathbb{P}(F_j(m)).
\end{aligned}$$

Then, using Lemmas 18 and 19 and summing over all $m \le M$ gives,

$$\sum_{m=1}^{m_{\max}} \left(\sum_{j:m_j<m} 4\mathbb{P}(M_j > m_j \text{ and } M^* \ge m_j) T\tilde{\Delta}_m + \sum_{j:m_j \ge m} 4\mathbb{P}(F_j(m)) T\tilde{\Delta}_m \right)$$

$$
\begin{aligned}
&\leq \sum_{m=1}^{m_{\max}} \left(\sum_{j:m_j<m} 4\frac{6}{T\tilde{\Delta}_{m_j}^2} T \frac{\tilde{\Delta}_{m_j}}{2^{m-m_j}} + \sum_{j:m_j \geq m} \frac{24}{T\tilde{\Delta}_m^2} T\tilde{\Delta}_m \right) \\
&\leq \sum_{j=1}^{K} \frac{24}{\tilde{\Delta}_{m_j}} \sum_{m=m_j}^{m_{\max}} 2^{-(m-m_j)} + \sum_{j=1}^{K} \sum_{m=1}^{m_j} \frac{24}{2^{-m}} \\
&\leq \sum_{j=1}^{K} \frac{96 \cdot 2}{\Delta_j} + \sum_{j=1}^{K} 24 \cdot 2^{m_j+1} \\
&\leq \sum_{j=1}^{K} \frac{192}{\Delta_j} + \sum_{j=1}^{K} 48 \cdot \frac{4}{\Delta_j} = \sum_{j=1}^{K} \frac{384}{\Delta_j}.
\end{aligned}
$$

Combining the regret from terms I and II gives,

$$
\mathbb{E}[\mathfrak{R}_T] \leq \sum_{j=1}^{K} \left(\frac{480}{\Delta_j} + \Delta_j n_{m_j} \right).
$$

Hence, all that remains is to bound n_m in terms of Δ_j, T and d,

$$
\begin{aligned}
n_{m_j} &= \left\lceil \frac{1}{\tilde{\Delta}_{m_j}^2} \left(\sqrt{2\log(T\tilde{\Delta}_{m_j}^2)} + \sqrt{2\log(T\tilde{\Delta}_{m_j}^2) + \frac{8}{3}\tilde{\Delta}_{m_j} \log(T\tilde{\Delta}_{m_j}^2) + 6\tilde{\Delta}_{m_j} m_j \mathbb{E}[\tau]} \right)^2 \right\rceil \\
&\leq \left\lceil \frac{1}{\tilde{\Delta}_{m_j}^2} \left(8\log(T\tilde{\Delta}_{m_j}^2) + \frac{16}{3}\tilde{\Delta}_{m_j} \log(T\tilde{\Delta}_{m_j}^2) + 12\tilde{\Delta}_{m_j} m_j \mathbb{E}[\tau] \right) \right\rceil \\
&\leq 1 + \frac{8\log(T\Delta_j^2/4)}{\tilde{\Delta}_{m_j}^2} + \frac{16\log(T\Delta_j^2/4)}{3\tilde{\Delta}_{m_j}} + \frac{12\log_2(4/\Delta_j)\mathbb{E}[\tau]}{\tilde{\Delta}_{m_j}} \\
&\leq 1 + \frac{128\log(T\Delta_j^2)}{\Delta_j^2} + \frac{32\log(T\Delta_j^2)}{3\Delta_j} + \frac{96\log(4/\Delta_j)\mathbb{E}[\tau]}{\Delta_j},
\end{aligned}
$$

where we have used $(a+b)^2 \leq 2(a^2+b^2)$ for $a, b \geq 0$ and $\log_2(x) \leq 2\log(x)$ for $x > 0$.

Hence, the total expected regret from ODAAF with bounded delays can be bounded by,

$$
\mathbb{E}[\mathfrak{R}_t] \leq \sum_{j=1:j\neq j^*}^{K} \left(\frac{128\log(T\Delta_j^2)}{\Delta_j} + \frac{32}{3}\log(T\Delta_j^2) + 96\log(4/\Delta_j)\mathbb{E}[\tau] + \frac{480}{\Delta_j} + \Delta_j \right). \tag{16}
$$

□

We now prove the problem independent regret bound,

Corollary 3 *For any problem instance satisfying Assumption 1, the expected regret of Algorithm 1 satisfies*

$$
\mathbb{E}[\mathfrak{R}_T] \leq O(\sqrt{KT\log(K)} + K\mathbb{E}[\tau]\log(T)).
$$

Proof: Let

$$
\lambda = \sqrt{\frac{K\log(K)e^2}{T}}
$$

and note that for $\Delta > \lambda$, $\log(T\Delta^2)/\Delta$ is a decreasing function of Δ. Then, for some constants C_1, C_2, and using the previous theorem, we can bound the regret by,

$$
\mathbb{E}[\mathfrak{R}_T] \leq \sum_{j:\Delta_j \leq \lambda} \mathbb{E}[\mathfrak{R}_t^{(j)}] + \sum_{j:\Delta_j > \lambda} \mathbb{E}[\mathfrak{R}_T^{(j)}] \leq \frac{KC_1\log(T\lambda^2)}{\lambda} + KdC_2\log(1/\lambda) + T\lambda.
$$

Then, subsituting the above value of λ gives a worst case regret bound that scales with $O(\sqrt{KT\log(K)} + K\mathbb{E}[\tau]\log(T))$.

□

C. Results for Delays with Bounded Support

C.1. High Probability Bounds

Lemma 5 *Under Assumptions 1 of known expected delay and 2 of bounded delays, and choice of n_m given in (6), the estimates $\bar{X}_{m,j}$ obtained by Algorithm 1 satisfy the following: For any arm j and phase m, with probability at least $1-\frac{12}{T\tilde{\Delta}_m^2}$, either $j \notin \mathcal{A}_m$ or*

$$\bar{X}_{m,j} - \mu_j \leq \tilde{\Delta}_m/2.$$

Proof: Let

$$w_m = \frac{4\log(T\tilde{\Delta}_m^2)}{3n_m} + \sqrt{\frac{2\log(T\tilde{\Delta}_m^2)}{n_m}} + \frac{2\mathbb{E}[\tau]}{n_m}. \tag{17}$$

We show that with probability greater than $1-\frac{12}{T\tilde{\Delta}_m^2}$, either $j \notin \mathcal{A}_m$ or $\frac{1}{n_m}\sum_{t\in T_j(m)}(X_t - \mu_j) \leq w_m$. For now, assume that $n_m \geq md$.

For arm j and phase m, assume $j \in \mathcal{A}_m$ and define p_i to be the probability of the confidence bounds on arm j failing at the end of each phase $i \leq m$, ie. $p_i \doteq \mathbb{P}(\sum_{t\in T_j(i)}(X_t-\mu_j) \geq n_i w_i)$ with $p_0 = 0$. Again, let $B_{i,t} = R_t\mathbb{I}\{\tau_{t,J_t}+t \geq S_{i,j}\}$ and $C_{i,t} = R_t\mathbb{I}\{\tau_{t,J_t}+t > U_{i,j}\}$ (note that we don't need to consider $A_{i,t}$ since $\nu_i = n_i - n_{i-1} \geq d$ so all reward entering $[S_{i,j}, U_{i,j}]$ will be from the last $\nu_i \geq d$ plays) and for any event H, let $\mathbb{I}_i\{H\} := \mathbb{I}\{H \cap \{j \in \mathcal{A}_i\}\}$. Recall the filtration $\{\mathcal{G}_t\}_{t=0}^{\infty}$ from (12) where $\mathcal{G}_t = \sigma(X_1,\dots,X_t,J_1,\dots,J_t,\tau_{1,J_1},\dots,\tau_{t,J_t},R_{1,J_1},\dots,R_{t,J_t})$ and $\mathcal{G}_0 = \{\emptyset,\Omega\}$. Now, defining,

$$\begin{aligned}
Q_t &= \sum_{i=1}^{m} B_{i,t}\mathbb{I}\{S_{i,j}-d-1 \leq t \leq S_{i,j}-1\}),\\
P_t &= \sum_{i=1}^{m} C_{i,t}\mathbb{I}\{S_{i,j} \leq t \leq U_{i,j}\},
\end{aligned}$$

we use the decomposition

$$\begin{aligned}
\sum_{t\in T_j(m)}(X_t-\mu_j) &= \sum_{i=1}^{m}\sum_{t=S_{i,j}}^{U_{i,j}}(X_t-\mu_j)\\
&\leq \sum_{i=1}^{m}\left(\sum_{t=S_{i-1,j}}^{S_{i,j}-1} R_{t,J_t}\mathbb{I}_i\{\tau_{t,J_t}+t \geq S_{i,j}\} + \sum_{t=S_{i,j}}^{U_{i,j}}(R_{t,J_t}-\mu_j) - \sum_{t=S_{i,j}}^{U_{i,j}} R_{t,J_t}\mathbb{I}_i\{\tau_{t,J_t}+t > U_{i,j}\}\right)\\
&\leq \sum_{i=1}^{m}\left(\sum_{t=S_{i,j}-d}^{S_{i,j}-1} B_{i,t} + \sum_{t=S_{i,j}}^{U_{i,j}}(R_t-\mu_j) - \sum_{t=S_{i,j}}^{U_{i,j}} C_{i,t}\right)\\
&= \sum_{i=1}^{m}\sum_{t=S_{i,j}}^{U_{i,j}}(R_{t,J_t}-\mu_j) + \sum_{t=1}^{S_{m,j}} Q_t - \sum_{t=1}^{U_{m,j}} P_t\\
&= \underbrace{\sum_{i=1}^{m}\sum_{t=S_{i,j}}^{U_{i,j}}(R_{t,J_t}-\mu_j)}_{\text{Term I.}} + \underbrace{\sum_{t=1}^{S_{m,j}}(Q_t-\mathbb{E}[Q_t|\mathcal{G}_{t-1}])}_{\text{Term II.}} + \underbrace{\sum_{t=1}^{U_{m,j}}(\mathbb{E}[P_t|\mathcal{G}_{t-1}]-P_t)}_{\text{Term III.}}\\
&\qquad + \underbrace{\sum_{t=1}^{S_{m,j}}\mathbb{E}[Q_t|\mathcal{G}_{t-1}] - \sum_{t=1}^{U_{m,j}}\mathbb{E}[P_t|\mathcal{G}_{t-1}]}_{\text{Term IV.}}.
\end{aligned}$$

Outline of proof Again, the proof continues by bounding each term of this decomposition in turn. Note that we do not have the $A_{i,t}$ terms in this decomposition since there will be no reward from phase $i-1$ (before the bridge period) received in $[S_{i,j}, U_{i,j}]$. We bound each of these terms with high probability. For terms I. and III., this is the same as in the general case (see the proof of Lemma 1, Appendix B),. For term II. we need the following results to show that $Z_t = Q_t - \mathbb{E}[Q_s|\mathcal{G}_{t-1}]$ is a martingale difference (Lemma 20) and to bound its variance (Lemma 21) before we can apply Freedman's inequality. The bound for term IV. is also different due to the bridge period and boundedness of the delay. After bounding each term, we collect them together and recursively calculate the probability with which the bounds hold.

Lemma 20 *Let $Y_s = \sum_{t=1}^{s}(Q_t - \mathbb{E}[Q_t|\mathcal{G}_{t-1}])$ for all $s \geq 1$, and $Y_0 = 0$. Then $\{Y_s\}_{s=0}^{\infty}$ is a martingale with respect to the filtration $\{\mathcal{G}_s\}_{s=0}^{\infty}$ with increments $Z_s = Y_s - Y_{s-1} = Q_s - \mathbb{E}[Q_s|\mathcal{G}_{s-1}]$ satisfying $\mathbb{E}[Z_s|\mathcal{G}_{s-1}] = 0, |Z_s| \leq 1$ for all $s \geq 1$.*

Proof: To show $\{Y_s\}_{s=0}^{\infty}$ is a martingale we need to show that Y_s is $\mathcal{G}_s$-measurable for all s and $\mathbb{E}[Y_s|\mathcal{G}_{s-1}] = Y_{s-1}$.

Measurability: We show that $B_{i,s}\mathbb{I}\{S_{i,j} - d - 1 \leq s \leq S_{i,j} - 1\}$ is $\mathcal{G}_s$-measurable. This then suffices to show that Y_s is $\mathcal{G}_s$-measurable since the filtration $\mathcal{G}_s$ is non-decreasing in s.

First note that by definition of $\mathcal{G}_s$, τ_{t,J_t}, R_{t,J_t} are all $\mathcal{G}_s$-measurable for $t \leq s$. Hence, it is sufficient to show that $\mathbb{I}\{\tau_{s,J_s} + s \geq S_{i,j}, S_{i,j} - d - 1 \leq s \leq S_{i,j} - 1\}$ is $\mathcal{G}_s$-measurable since the product of measurable functions is measurable. For any $s' \in \mathbb{N} \cup \{\infty\}$, $\{S_{i,j} = s', s' - d - 1 \leq s\} \in \mathcal{G}_s$ for $s \geq S_i - \nu_{i-1}$ and so the union $\bigcup_{s' \in \mathbb{N}\cup\{\infty\}}\{\tau_{s,J_s} + s \geq s', s' - d - 1 \leq s \leq s' - 1, S_{i,j} = s'\} = \{\tau_{s,J_s} + s \geq S_{i,j}, S_{i,j} - d - 1 \leq s \leq S_{i,j} - 1\}$ is an element of $\mathcal{G}_s$.

Increments: Hence, $\{Y_s\}_{s=0}^{\infty}$ is a martingale with respect to the filtration $\{\mathcal{G}_s\}_{s=0}^{\infty}$ if the increments conditional on the past are zero. For any $s \geq 1$, we have that

$$Z_s = Y_s - Y_{s-1} = \sum_{t=1}^{s}(Q_t - \mathbb{E}[Q_t|\mathcal{G}_{t-1}]) - \sum_{t=1}^{s-1}(Q_t - \mathbb{E}[Q_t|\mathcal{G}_{t-1}]) = Q_s - \mathbb{E}[Q_s|\mathcal{G}_{s-1}].$$

Then,

$$\mathbb{E}[Z_s|\mathcal{G}_{s-1}] = \mathbb{E}[Q_s - \mathbb{E}[Q_s|\mathcal{G}_{s-1}]|\mathcal{G}_{s-1}] = \mathbb{E}[Q_s|\mathcal{G}_{s-1}] - \mathbb{E}[Q_s|\mathcal{G}_{s-1}] = 0$$

and so $\{Y_s\}_{s=0}^{\infty}$ is a martingale.

Lastly, since for any t and $\omega \in \Omega$, there is at most one i where $\mathbb{I}\{S_{i,j} - d \leq t \leq S_{i,j} - 1\}(\omega) = 1$, and by definition of $R_{t,J_t}, B_{i,t} \leq 1$, it follows that $|Z_s| = |Q_s - \mathbb{E}[Q_s|\mathcal{G}_{s-1}]| \leq 1$ for all s. □

Lemma 21 *For any $t \geq 1$, let $Z_t = Q_t - \mathbb{E}[Q_t|\mathcal{G}_{t-1}]$, then*

$$\sum_{t=1}^{S_{m,j}-1} \mathbb{E}[Z_t^2|\mathcal{G}_{t-1}] \leq m\mathbb{E}[\tau].$$

Proof: Let us denote $S' \doteq S_{m,j} - 1$. Observe that

$$\sum_{t=1}^{S'}\mathbb{E}[Z_t^2|\mathcal{G}_{t-1}] = \sum_{t=1}^{S'}\mathbb{V}(Q_t|\mathcal{G}_{t-1}) \leq \sum_{t=1}^{S'}\mathbb{E}[Q_t^2|\mathcal{G}_{t-1}] = \sum_{t=1}^{S'}\mathbb{E}\left[\left(\sum_{i=1}^{m}(B_{i,t}\mathbb{I}\{S_{i,j} - d \leq t \leq S_{i,j} - 1\})\right)^2\middle|\mathcal{G}_{t-1}\right].$$

Then for all $i = 1, \ldots, m$, all indicator terms $\mathbb{I}\{S_{i,j} - d \leq t \leq S_{i,j} - 1\}$ are $\mathcal{G}_{t-1}$-measurable and only one can be non zero for any $\omega \in \Omega$. Hence, for any $i, i' \leq m$, $i \neq i'$,

$$B_{i,t} \times \mathbb{I}\{S_{i,j} - d - 1 \leq t \leq S_{i,j} - 1\} \times B_{i',t} \times \mathbb{I}\{S_{i',j} - d - 1 \leq t \leq S_{i',j} - 1\} = 0,$$

Using the above we see that

$$\sum_{t=1}^{S'}\mathbb{E}[Z_t^2|\mathcal{G}_{t-1}] \leq \sum_{t=1}^{S'}\mathbb{E}\left[\left(B_{i,t}\mathbb{I}\{S_{i,j} - d - 1 \leq t \leq S_{i,j} - 1\}\right)^2\middle|\mathcal{G}_{t-1}\right]$$
$$= \sum_{t=1}^{S'}\mathbb{E}\left[\sum_{i=1}^{m}B_{i,t}^2\mathbb{I}\{S_{i,j} - d - 1 \leq t \leq S_{i,j} - 1\}^2\middle|\mathcal{G}_{t-1}\right]$$

$$= \sum_{i=1}^{m} \sum_{t=1}^{S'} \mathbb{E}[B_{i,t}^2 \mathbb{I}\{S_{i,j} - d - 1 \leq t \leq S_{i,j} - 1\} | \mathcal{G}_{t-1}]$$

(using that the indicator is $\mathcal{G}_{t-1}$-measurable)

$$\leq \sum_{i=1}^{m} \sum_{t=S_{i,j}-d-1}^{S_{i,j}-1} \mathbb{E}[B_{i,t}^2 | \mathcal{G}_{t-1}].$$

Then, for any $i \geq 1$,

$$\begin{aligned}
\sum_{t=S_{i,j}-d-1}^{S_{i,j}-1} \mathbb{E}[B_{i,t}^2 | \mathcal{G}_{t-1}] &= \sum_{t=S_{i,j}-d-1}^{S_{i,j}-1} \mathbb{E}[R_{t,J_t}^2 \mathbb{I}\{\tau_{t,J_t} + t \geq S_{i,j}\} | \mathcal{G}_{t-1}] \\
&\leq \sum_{t=S_{i,j}-d-1}^{S_{i,j}-1} \mathbb{E}[\mathbb{I}\{\tau_{t,J_t} + t \geq S_{i,j}\} | \mathcal{G}_{t-1}] \\
&= \sum_{s=0}^{\infty} \mathbb{I}\{S_{i,j} = s\} \sum_{t=s-d-1}^{s-1} \mathbb{E}[\mathbb{I}\{\tau_{t,J_t} + t \geq S_{i,j}\} | \mathcal{G}_{t-1}] \\
&= \sum_{s=0}^{\infty} \sum_{t=s-d-1}^{s-1} \mathbb{E}[\mathbb{I}\{S_{i,j} = s, \tau_{t,J_t} + t \geq S_{i,j}\} | \mathcal{G}_{t-1}]
\end{aligned}$$

(Since $S_{i,j} \geq S_i$ and so, due to the bridge period, $\{S_{i,j} = s\} \in \mathcal{G}_{t-1}$ for any $t \geq s - d$)

$$\begin{aligned}
&= \sum_{s=0}^{\infty} \sum_{t=s-d-1}^{s-1} \mathbb{E}[\mathbb{I}\{S_{i,j} = s, \tau_{t,J_t} + t \geq s\} | \mathcal{G}_{t-1}] \\
&= \sum_{s=0}^{\infty} \mathbb{I}\{S_{i,j} = s\} \sum_{t=s-d-1}^{s-1} \mathbb{P}(\tau_{t,J_t} + t \geq s)
\end{aligned}$$

(Since $\{S_{i,j} = s\} \in \mathcal{G}_{t-1}$ for any $t \geq s - d$)

$$\begin{aligned}
&\leq \sum_{s=0}^{\infty} \mathbb{I}\{S_{i,j} = s\} \sum_{l=0}^{\infty} \mathbb{P}(\tau > l) \\
&\leq \mathbb{E}[\tau].
\end{aligned}$$

Combining all terms gives the result. □

We now return to bounding each term of the decomposition

Bounding Term I.: For term II., as in Lemma 1, we can use Lemma 17 to get that with probability greater than $1 - \frac{1}{T\tilde{\Delta}_m^2}$,

$$\sum_{i=1}^{m} \sum_{t=S_{i,j}}^{U_{i,j}} (R_{t,J_t} - \mu_j) \leq \sqrt{\frac{n_m \log(T\tilde{\Delta}_m^2)}{2}}.$$

Bounding Term II.: For Term II., we will use Freedmans inequality (Theorem 10). From Lemma 20, $\{Y_s\}_{s=0}^{\infty}$ with $Y_s = \sum_{t=1}^{s} (Q_t - \mathbb{E}[Q_t | \mathcal{G}_{t-1}])$ is a martingale with respect to $\{\mathcal{G}_s\}_{s=0}^{\infty}$ with increments $\{Z_s\}_{s=0}^{\infty}$ satisfying $\mathbb{E}[Z_s | \mathcal{G}_{s-1}] = 0$ and $Z_s \leq 1$ for all s. Further, by Lemma 21, $\sum_{t=1}^{s} \mathbb{E}[Z_t^2 | \mathcal{G}_{t-1}] \leq m\mathbb{E}[\tau] \leq \frac{4 \times 2^m \mathbb{E}[\tau]}{8} \leq n_m/8$ with probability 1. Hence we can apply Freedman's inequality to get that with probability greater than $1 - \frac{1}{T\tilde{\Delta}_m^2}$,

$$\sum_{t=1}^{S_{m,j}} (Q_t - \mathbb{E}[Q_t | \mathcal{G}_{t-1}]) \leq \frac{2}{3} \log(T\tilde{\Delta}_m^2) + \sqrt{\frac{1}{8} n_m \log(T\tilde{\Delta}_m^2)}.$$

Bounding Term III.: For Term III., we again use Freedman's inequality (Theorem 10). As in Lemma 1, we use Lemma 14 to show that $\{Y'_s\}_{s=0}^{\infty}$ with $Y'_s = \sum_{t=1}^{s}(\mathbb{E}[P_t|\mathcal{G}_{t-1}] - P_t)$ is a martingale with respect to $\{\mathcal{G}_s\}_{s=0}^{\infty}$ with increments $\{Z'_s\}_{s=0}^{\infty}$ satisfying $\mathbb{E}[Z'_s|\mathcal{G}_{s-1}] = 0$ and $Z'_s \leq 1$ for all s. Further, by Lemma 15, $\sum_{t=1}^{s}\mathbb{E}[Z_t^2|\mathcal{G}_{t-1}] \leq m\mathbb{E}[\tau] \leq n_m/8$ with probability 1. Hence, with probability greater than $1 - \frac{1}{T\tilde{\Delta}_m^2}$,

$$\sum_{t=1}^{U_{m,j}}(\mathbb{E}[P_t|\mathcal{G}_{t-1}] - P_t) \leq \frac{2}{3}\log(T\tilde{\Delta}_m^2) + \sqrt{\frac{1}{8}n_m\log(T\tilde{\Delta}_m^2)}.$$

Bounding Term IV.: For term IV., we consider the expected difference at each round $1 \leq i \leq m$ and exploit the independence of τ_{t,J_t} and R_{t,J_t}. Consider first $i \geq 2$ and let j'_i be the arm played just before arm j is played in the ith phase (allowing for j'_i to be the last arm played in phase $i-1$). Then, much in the same way as Lemma 21,

$$\begin{aligned}
\sum_{t=S_{i,j}-d-1}^{S_{i,j}-1}\mathbb{E}[B_{i,t}|\mathcal{G}_{t-1}] &= \sum_{t=S_{i,j}-d-1}^{S_{i,j}-1}\mathbb{E}[R_{t,J_t}\mathbb{I}\{\tau_{t,J_t}+t \geq S_{i,j}\}|\mathcal{G}_{t-1}] \\
&= \sum_{s'=d+1}^{\infty}\sum_{s=s'}^{\infty}\mathbb{I}\{S_i = s', S_{i,j} = s\}\sum_{t=s-d}^{s-1}\mathbb{E}[R_{t,J_t}\mathbb{I}\{\tau_{t,J_t}+t \geq S_{i,j}\}|\mathcal{G}_{t-1}] \\
&= \sum_{s'=d+1}^{\infty}\sum_{s=s'}^{\infty}\sum_{t=s-d}^{s-1}\sum_{k=1}^{K}\mathbb{E}[R_{t,J_t}\mathbb{I}\{S_i = s', S_{i,j} = s, \tau_{t,J_t}+t \geq S_{i,j}, J_t = k\}|\mathcal{G}_{t-1}] \\
&\qquad \text{(Due to the bridge period } \{S_i = s', S_{i,j} = s\} \in \mathcal{G}_{t-1} \text{ for } t \geq s-d \geq s'-d) \\
&= \sum_{s'=d+1}^{\infty}\sum_{s=s'}^{\infty}\sum_{t=s-d}^{s-1}\sum_{k=1}^{K}\mathbb{I}\{S_i = s', S_{i,j} = s, J_t = k\}\mathbb{E}[R_{t,k}\mathbb{I}\{\tau_{t,k}+t \geq s\}|\mathcal{G}_{t-1}] \\
&= \sum_{s'=d+1}^{\infty}\sum_{s=s'}^{\infty}\sum_{t=s-d}^{s-1}\sum_{k=1}^{K}\mu_k\mathbb{I}\{S_i = s', S_{i,j} = s, J_t = k\}\mathbb{P}(\tau \geq s-t) \\
&= \mu_{j'_i}\sum_{l=0}^{d-1}\mathbb{P}(\tau > l).
\end{aligned}$$

A similar argument works for $i = 1, j > 1$ with the simplification that $S_{i,j}$ is not a random quantity but known . Finally, for $i = 1, j = 1$ the sum is 0. Furthermore, using a similar argument, for all i, j,

$$\begin{aligned}
\sum_{t=S_{i,j}}^{U_{i,j}}\mathbb{E}[C_{i,t}|\mathcal{G}_{t-1}] &= \sum_{t=U_{i,j}-d+1}^{U_{i,j}}\mathbb{E}[C_{i,t}|\mathcal{G}_{t-1}] \\
&= \sum_{s'=d+1}^{\infty}\sum_{s=s'}^{\infty}\sum_{t=s-d}^{s}\mathbb{E}[R_{t,j}\mathbb{I}\{\tau_{t,j}+t > s\}\mathbb{I}\{U_{i,j} = s, S_i = s'\}|\mathcal{G}_{t-1}] \\
&= \mu_j\sum_{s=d+1}^{\infty}\mathbb{I}\{U_{i,j} = s, S_i = s'\}\sum_{t=s-d}^{s}\mathbb{P}(\tau + t > s) \\
&= \mu_j\sum_{l=0}^{d-1}\mathbb{P}(\tau > l).
\end{aligned}$$

Combining these we get the following bound for term IV for all $(i, j) \neq (1, 1)$,

$$\begin{aligned}
\sum_{t=S_{i,j}-d-1}^{S_{i,j}-1}\mathbb{E}[B_{i,t}|\mathcal{G}_{t-1}] - \sum_{t=S_{i,j}}^{U_{i,j}}\mathbb{E}[C_{i,t}|\mathcal{G}_{t-1}] &\leq \mu_{j'_i}\sum_{l=0}^{d-1}\mathbb{P}(\tau > l) - \mu_j\sum_{l=0}^{d-1}\mathbb{P}(\tau > l) \\
&\leq |\mu_{j'_i} - \mu_j|\mathbb{E}[\tau].
\end{aligned}$$

If $(i, j) = (1, 1)$ then we have the upper bounded by $\mu_1 \mathbb{E}[\tau] \leq \mathbb{E}[\tau] = \tilde{\Delta}_0 \mathbb{E}[\tau]$ since no pay-off seeps in and we define $\tilde{\Delta}_0 = 1$.

Let p_i be the probability that the confidence bounds for one arm hold in phase i and $p_0 = 0$. Then, the probability that either arm j_i' or j is active in phase i when it should have been eliminated in or before phase $i-1$ is less than $2p_{i-1}$. If neither arm should have been eliminated by phase i, this means that their mean rewards are within $\tilde{\Delta}_{i-1}$ of each other. Hence, with probability greater than $1 - 2p_{i-1}$,

$$\sum_{t=S_{i,j}-d-1}^{S_{i,j}-1} \mathbb{E}[B_{i,t}|\mathcal{G}_{t-1}] - \sum_{t=S_{i,j}}^{U_{i,j}} \mathbb{E}[C_{i,t}|\mathcal{G}_{t-1}] \leq \tilde{\Delta}_{i-1} \mathbb{E}[\tau].$$

Then, summing over all phases gives that with probability greater than $1 - 2\sum_{i=0}^{m-1} p_i$,

$$\sum_{i=1}^{m} \left(\sum_{t=S_{i,j}-d-1}^{S_{i,j}-1} \mathbb{E}[B_{i,t}|\mathcal{G}_{t-1}] - \sum_{t=S_{i,j}}^{U_{i,j}} \mathbb{E}[C_{i,t}|\mathcal{G}_{t-1}] \right) \leq \mathbb{E}[\tau] \sum_{i=1}^{m} \tilde{\Delta}_{i-1} = \mathbb{E}[\tau] \sum_{i=0}^{m-1} \frac{1}{2^i} \leq 2\mathbb{E}[\tau].$$

Combining all Terms: To get the final high probability bound, we sum the bounds for each term I.-IV.. Then, with probability greater than $1 - (\frac{3}{T\tilde{\Delta}_m^2} + 2\sum_{i=1}^{m-1} p_i)$ either $j \notin \mathcal{A}_m$ or arm j is played n_m times by the end of phase m and

$$\begin{aligned}
\frac{1}{n_m} \sum_{t \in T_j(m)} (X_t - \mu_j) &\leq \frac{4 \log(T\tilde{\Delta}_m^2)}{3n_m} + \left(\frac{2}{\sqrt{8}} + \frac{1}{\sqrt{2}} \right) \sqrt{\frac{\log(T\tilde{\Delta}_m^2)}{n_m}} + \frac{2\mathbb{E}[\tau]}{n_m} \\
&\leq \frac{4 \log(T\tilde{\Delta}_m^2)}{3n_m} + \sqrt{\frac{2\log(T\tilde{\Delta}_m^2)}{n_m}} + \frac{2\mathbb{E}[\tau]}{n_m} = w_m.
\end{aligned}$$

Using the fact that $p_0 = 0$ and substituting the other p_i's using the recursive relationship $p_i = \frac{3}{T\tilde{\Delta}_i^2} + 2\sum_{l=1}^{i-1} p_l$ gives,

$$\begin{aligned}
\frac{3}{T\tilde{\Delta}_m^2} + 2\sum_{i=0}^{m-1} p_i &= \frac{3}{T\tilde{\Delta}_m^2} + 2(\frac{3}{T\tilde{\Delta}_{m-1}^2} + 2(p_{m-2} + \cdots + p_1) + p_{m-2} + \cdots + p_1) \\
&= \frac{3}{T\tilde{\Delta}_m^2} + 2(\frac{3}{T\tilde{\Delta}_{m-1}^2} + 3(p_{m-2} + \cdots + p_1)) \\
&= \frac{3}{T\tilde{\Delta}_m^2} + 2(\frac{3}{T\tilde{\Delta}_{m-1}^2} + 3(\frac{3}{T\tilde{\Delta}_{m-2}^2} + 3(p_{m-3} + \cdots + p_1)) \\
&\leq \sum_{i=1}^{m} 3^{m-i} \frac{3}{T\tilde{\Delta}_i^2} \\
&= \frac{3}{T} \sum_{i=1}^{m} 3^{m-i} 2^{2i} \\
&= \frac{3}{T} \sum_{i=1}^{m} 3^{m-i} 4i \\
&= \frac{3}{T} \sum_{i=1}^{m} (\frac{3}{4})^{m-i} 4^{m-i} 4^i \\
&= \frac{3 \times 4^m}{T} \sum_{i=1}^{m} (\frac{3}{4})^{m-i} \\
&\leq \frac{12}{T\tilde{\Delta}_m^2}.
\end{aligned}$$

Hence, with probability greater than $1 - \frac{12}{T\tilde{\Delta}_m^2}$, either $j \notin \mathcal{A}_m$ or $\frac{1}{n_m} \sum_{t \in T_j(m)} (X_t - \mu_j) \leq w_m$.

Defining $\mathbf{n_m}$: The above results rely on the assumption that $n_m \geq md$, so that only the previous arm can corrupt our observations. In practice, if d is too large then we will not want to play each active arm d times per phase because we will end up playing sub-optimal arms too many times. In this case, it is better to ignore the bound on the delay and use the results from Lemma 1 to set n_m as in (14). Formalizing this gives

$$n_m = \max\left\{ m\tilde{d}_m, \left\lceil \frac{1}{\tilde{\Delta}_m^2}\left(\sqrt{2\log(T\tilde{\Delta}_m^2)} + \sqrt{2\log(T\tilde{\Delta}_m^2) + \frac{8}{3}\tilde{\Delta}_m \log(T\tilde{\Delta}_m^2) + 4\tilde{\Delta}_m \mathbb{E}[\tau]}\right)^2 \right\rceil \right\} \tag{18}$$

where $\tilde{d}_m = \min\{d, \frac{(14)}{m}\}$. This ensures that if d is small, we play each active arm enough times to ensure that $w_m \leq \frac{\tilde{\Delta}_m}{2}$ for w_m in (17). Similarly, for large d, by Lemma 1, we know that n_m is suffiently large to guarantee $w_m \leq \frac{\tilde{\Delta}_m}{2}$ for w_m from (10). □

C.2. Regret Bounds

We now prove the regret bound given in Theorem 6. Note that for these results, it is necessary to use the bridge period of the algorithm.

Theorem 6 *Under Assumption 1 and bounded delay Assumption 2, the expected regret of Algorithm 1 satisfies*

$$\mathbb{E}[\mathfrak{R}_T] \leq \sum_{j=1; j\neq j^*}^{K} O\bigg(\frac{\log(T\Delta_j^2)}{\Delta_j} + \mathbb{E}[\tau] \\ + \min\left\{d, \frac{\log(T\Delta_j^2)}{\Delta_j} + \log(\frac{1}{\Delta_j})\mathbb{E}[\tau]\right\}\bigg).$$

Proof: For any sub-optimal arm j, define M_j to be the random variable representing the phase arm j is eliminated in and note that if M_j is finite, $j \in \mathcal{A}_{M_j}$ but $j \notin \mathcal{A}_{M_j+1}$. Then let m_j be the phase arm j *should* be eliminated in, that is $m_j = \min\{m | \tilde{\Delta}_m < \frac{\Delta_j}{2}\}$ and note that, from the definition of $\tilde{\Delta}_m$ in our algorithm, we get the relations

$$2^m = \frac{1}{\tilde{\Delta}_m}, \quad 2\tilde{\Delta}_{m_j} = \tilde{\Delta}_{m_j-1} \geq \frac{\Delta_j}{2} \quad \text{and so,} \quad \frac{\Delta_j}{4} \leq \tilde{\Delta}_{m_j} \leq \frac{\Delta_j}{2}. \tag{19}$$

Define $\mathfrak{R}_T^{(j)}$ to be the regret contribution from each arm $1 \leq j \leq K$ and let M^* be the round where the optimal arm $j^*$is eliminated. Hence,

$$\begin{aligned}\mathbb{E}[\mathfrak{R}_T] &= \mathbb{E}\bigg[\sum_{j=1}^{K} \mathfrak{R}_T^{(j)}\bigg] = \mathbb{E}\bigg[\sum_{m=0}^{\infty}\sum_{j=1}^{K} \mathfrak{R}_T^{(j)} \mathbb{I}\{M^* = m\}\bigg] \\ &= \mathbb{E}\bigg[\sum_{m=0}^{\infty}\sum_{j:m_j<m} \mathfrak{R}_T^{(j)} \mathbb{I}\{M^* = m\} + \sum_{j:m_j\geq m} \mathfrak{R}_T^{(j)} \mathbb{I}\{M^* = m\}\bigg] \\ &= \underbrace{\mathbb{E}\bigg[\sum_{m=0}^{\infty}\sum_{j:m_j<m} \mathfrak{R}_T^{(j)} \mathbb{I}\{M^* = m\}\bigg]}_{\text{I.}} + \underbrace{\mathbb{E}\bigg[\sum_{m=0}^{\infty}\sum_{j:m_j\geq m} \mathfrak{R}_T^{(j)} \mathbb{I}\{M^* = m\}\bigg]}_{\text{II.}}\end{aligned}$$

We will bound the regret in each of these cases in turn. First, however, we need the following results.

Lemma 22 *For any suboptimal arm j, if $j^* \in \mathcal{A}_{m_j}$, then the probability arm j is* not *eliminated by round m_j is,*

$$\mathbb{P}(M_j > m_j \textit{ and } M^* \geq m_j) \leq \frac{24}{T\tilde{\Delta}_{m_j}^2}$$

Proof: The proof is exactly that of Lemma 18 but using Lemma 5 to bound the probability of the confidence bounds on either arm j or j^* failing. □

Define the event $F_j(m) = \{\bar{X}_{m,j^*} < \bar{X}_{m,j} - \tilde{\Delta}_m\} \cap \{j, j^* \in \mathcal{A}_m\}$ to be the event that arm j^* is eliminated by arm j in phase m. The probability of this occurring is bounded in the following lemma.

Lemma 23 *The probability that the optimal arm j^* is eliminated in round $m < \infty$ by the suboptimal arm j is bounded by*

$$\mathbb{P}(F_j(m)) \leq \frac{24}{T\tilde{\Delta}_m^2}$$

Proof: Again, the proof follows from Lemma 19 but using Lemma 5 to bound the probability of the confidence bounds failing. □

We now return to bounding the expected regret in each of the two cases.

Bounding Term I. To bound the first term, we consider the cases where arm j is eliminated in or before the correct round ($M_j \leq m_j$) and where arm j is eliminated late ($M_j > m_j$). Then,

$$\begin{aligned}
\mathbb{E}\Big[\sum_{m=0}^{\infty}\sum_{j:m_j<m}\mathfrak{R}_T^{(j)}\mathbb{I}\{M^*=m\}\Big] &= \mathbb{E}\Big[\sum_{j=1}^{K}\mathfrak{R}_T^{(j)}\mathbb{I}\{M^*\geq m_j\}\Big] \\
&= \mathbb{E}\Big[\sum_{j=1}^{K}\mathfrak{R}_T^{(j)}\mathbb{I}\{M^*\geq m_j\}\mathbb{I}\{M_j\leq m_j\}\Big] + \mathbb{E}\Big[\sum_{j=1}^{K}\mathfrak{R}_T^{(j)}\mathbb{I}\{M^*\geq m_j\}\mathbb{I}\{M_j> m_j\}\Big] \\
&\leq \sum_{j=1}^{K}\mathbb{E}[\mathfrak{R}_T^{(j)}\mathbb{I}\{M_j\leq m_j\}] + \sum_{j=1}^{K}\mathbb{E}[T\Delta_j\mathbb{I}\{M^*\geq m_j, M_j>m_j\}] \\
&\leq \sum_{j=1}^{K}2\Delta_j n_{m_j,j} + \sum_{j=1}^{K}T\Delta_j\mathbb{P}(M_j>m_j \text{ and } M^*\geq m_j) \\
&\leq \sum_{j=1}^{K}2\Delta_j n_{m_j,j} + \sum_{j=1}^{K}T\Delta_j\frac{24}{T\tilde{\Delta}_{m_j}^2} \\
&\leq \sum_{j=1}^{K}\left(2\Delta_j n_{m_j,j} + \frac{384}{\Delta_j}\right),
\end{aligned}$$

where the extra factor of 2 comes from the fact that each arm will be played n_m times by the end of phase m to get the data for the estimated mean, then in the worst case, arm j is chosen as the arm to be played in the bridge period of each phase that it is active, and thus is played another n_m times.

Bounding Term II For the second term, we use the results from Theorem 2, but using Lemma 22 to bound the probability a suboptimal arm is eliminated in a later round and Lemma 23 to bound the probability j^* is eliminated by a suboptimal arm. Hence,

$$\mathbb{E}\Big[\sum_{m=0}^{\infty}\sum_{j:m_j\geq m}\mathfrak{R}_T^{(j)}\mathbb{I}\{M^*=m\}\Big] \leq \sum_{j=1}^{K}\frac{1536}{\Delta_j}.$$

Combining the regret from terms I and II gives,

$$\mathbb{E}[\mathfrak{R}_T] \leq \sum_{j=1}^{K}\left(\frac{1920}{\Delta_j} + 2\Delta_j n_{m_j,j}\right)$$

Hence, all that remains is to bound n_m in terms of Δ_j, T and d. Using $L_{m,T} = \log(T\tilde{\Delta}_m^2)$, we have that,

$$\begin{aligned}
n_{m_j,j} &= \max\left\{m_j\tilde{d}_{m_j}, \left\lceil \frac{1}{\tilde{\Delta}_m^2}\left(\sqrt{2\log(T\tilde{\Delta}_m)} + \sqrt{2\log(T\tilde{\Delta}_m) + \frac{8}{3}\tilde{\Delta}_m\log(T\tilde{\Delta}_m) + 4\tilde{\Delta}_m\mathbb{E}[\tau]}\right)^2\right\rceil\right\} \\
&\leq \max\left\{m_j\tilde{d}_{m_j}, \left\lceil \frac{1}{\tilde{\Delta}_{m_j}^2}\left(8L_{m_j,T} + \frac{16}{3}\tilde{\Delta}_{m_j}L_{m_j,T} + 8\tilde{\Delta}_{m_j}\mathbb{E}[\tau]\right)\right\rceil\right\}
\end{aligned}$$

$$\leq \max\left\{m_j\tilde{d}_{m_j}, 1 + \frac{8L_{m_j,T}}{\tilde{\Delta}^2_{m_j}} + \frac{8L_{m_j,T}}{3\tilde{\Delta}_{m_j}} + \frac{8\mathbb{E}[\tau]}{\tilde{\Delta}_{m_j}}\right\}$$

$$\leq \max\left\{m_j\tilde{d}_{m_j}, 1 + \frac{128L_{m_j,T}}{\Delta^2_j} + \frac{32L_{m_j,T}}{\Delta_j} + \frac{32\mathbb{E}[\tau]}{\Delta_j}.\right\}$$

where we have used $(a+b)^2 \leq 2(a^2+b^2)$ for $a, b \geq 0$.

Hence, using the definition of $\tilde{d}_m = \min\{d, \frac{(14)}{m}\}$ and the results from Theorem 2, the total expected regret from ODAAF with bounded delays can be bounded by,

$$\mathbb{E}[\mathfrak{R}_t] \leq \sum_{j=1;j\neq j^*}^{K} \max\left\{\min\{d,(16)\}, \left(\frac{256\log(T\Delta_j^2)}{\Delta_j} + 64\mathbb{E}[\tau] + \frac{1920}{\Delta_j} + 64\log(T\Delta_j^2) + 2\Delta_j\right)\right\}. \tag{20}$$

$$\leq \sum_{j=1;j\neq j^*}^{K} \left(\frac{256\log(T\Delta_j^2)}{\Delta_j} + 64\mathbb{E}[\tau] + \frac{1920}{\Delta_j} + 64\log(T\Delta_j^2) + 2\Delta_j\right.$$

$$\left. + \min\left\{d, \frac{128\log(T\Delta_j^2)}{\Delta_j} + 96\log(4/\Delta_j)\mathbb{E}[\tau]\right\}\right)$$

□

Note that the constants in these regret bounds can be improved by only requiring the confidence bounds in phase m to hold with probability $\frac{1}{T\tilde{\Delta}_m}$ rather than $\frac{1}{T\tilde{\Delta}_m^2}$. This comes at a cost of increasing the logarithmic term to $\log(T\Delta_j)$. We now prove the problem independent regret bound,

Corollary 7 *For any problem instance satisfying Assumptions 1 and 2 with* $d \leq \sqrt{\frac{T\log K}{K}} + \mathbb{E}[\tau]$*, the expected regret of Algorithm 1 satisfies*

$$\mathbb{E}[\mathfrak{R}_T] \leq O(\sqrt{KT\log(K)} + K\mathbb{E}[\tau]).$$

Proof: We consider the maximal value each part of the regret in (20) can take. From Corollary 3, the first term is bounded by

$$O(\min\{Kd, \sqrt{KT\log K} + K\log(T)\mathbb{E}[\tau]\}).$$

For the first term, we again set $\lambda = \sqrt{\frac{K\log(K)e^2}{T}}$. Then, as in corollary Corollary 3, for constants $C_1, C_2 > 0$, we bound the regret contribution by

$$\sum_{j:\Delta_j\leq\lambda} \mathbb{E}[\mathfrak{R}_t^{(j)}] + \sum_{j:\Delta_j>\lambda} \mathbb{E}[\mathfrak{R}_T^{(j)}] \leq \frac{KC_1\log(T\lambda^2)}{\lambda} + C_2K\mathbb{E}[\tau] + T\lambda.$$

Then, substituting in for λ implies that the second term of (20) is $O(\sqrt{KT\log K} + K\mathbb{E}[\tau])$.

For $d \leq \sqrt{\frac{T\log K}{K}} + \mathbb{E}[\tau]$, $\min\{Kd, \sqrt{KT\log K} + K\log T\mathbb{E}[\tau]\} \leq \sqrt{KT\log K} + K\mathbb{E}[\tau]$. Hence the bound in (20) gives

$$\mathbb{E}[\mathfrak{R}_T] \leq O(\sqrt{KT\log K} + K\mathbb{E}[\tau] + \sqrt{KT\log K} + K\mathbb{E}[\tau]) = O(\sqrt{KT\log K} + K\mathbb{E}[\tau]).$$

□

D. Results for Delay with Known and Bounded Variance and Expectation

D.1. High Probability Bounds

Lemma 24 *Under Assumption 1 of known expected value and 3 of known (bound on) the expectation and variance of the delay, and choice of* n_m *given in (7), the estimates* $\bar{X}_{m,j}$ *obtained by Algorithm 1 satisfy the following: For any arm* j *and phase* m*, with probability at least* $1 - \frac{12}{T\tilde{\Delta}_m^2}$*, either* $j \notin \mathcal{A}_m$ *or*

$$\bar{X}_{m,j} - \mu_j \leq \tilde{\Delta}_m/2.$$

Proof: Let

$$w_m = \frac{4\log(T\tilde{\Delta}_m^2)}{3n_m} + \sqrt{\frac{2\log(T\tilde{\Delta}_m^2)}{n_m}} + \frac{2\mathbb{E}[\tau] + 4\mathbb{V}(\tau)}{n_m}. \tag{21}$$

We show that with probability greater than $1 - \frac{12}{T\tilde{\Delta}_m^2}$, $j \notin \mathcal{A}_m$ or $\frac{1}{n_m}\sum_{t\in T_j(m)}(X_t - \mu_j) \le w_m$.

For any arm j, phase i and time t, define,

$$A_{i,t} = R_{t,J_t}\mathbb{I}\{\tau_{t,J_t} + t \ge S_i\}, \quad B_{i,t} = R_{t,J_t}\mathbb{I}\{\tau_{t,J_t} + t \ge S_{i,j}\}, \quad C_{i,t} = R_{t,J_t}\mathbb{I}\{\tau_{t,J_t} + t > U_{i,j}\} \tag{22}$$

as in (11) and

$$\begin{aligned}
Q_t &= \sum_{i=1}^{m}(A_{i,t}\mathbb{I}\{S_{i-1,j} \le t \le S_i - \nu_{i-1} - 1\} + B_{i,t}\mathbb{I}\{S_i - \nu_{i-1} \le t \le S_{i,j} - 1\}),\\
P_t &= \sum_{i=1}^{m} C_{i,t}\mathbb{I}\{S_{i,j} \le t \le U_{i,j}\},
\end{aligned}$$

where $\nu_i = n_i - n_{i-1}$ is the number of times each active arm is played in phase $i \ge 1$ (assume $n_0 = 0$). Recall from the proof of Theorem 2, $\mathbb{I}_i\{H\} := \mathbb{I}\{H \cap \{j \in \mathcal{A}_i\}\} \le \mathbb{I}\{H\}$ and for all arms j and phases i, $\mathbb{I}_i\{\tau_{t,J_t} + t \ge S_{i,j}\} = \mathbb{I}\{\tau_{t,J_t} + t \ge S_{i,j}\}$ and $\mathbb{I}_i\{\tau_{t,J_t} + t > U_{i,j}\} = \mathbb{I}\{\tau_{t,J_t} + t > U_{i,j}\}$.

Then, using the convention $S_0 = S_{0,j} = 0$ for all arms j, we use the decomposition,

$$\begin{aligned}
\sum_{i=1}^{m}\sum_{t=S_{i,j}}^{U_{i,j}}(X_t - \mu_j) &\le \sum_{i=1}^{m}\left(\sum_{t=S_{i-1,j}}^{S_{i,j}-1} R_{t,J_t}\mathbb{I}_i\{\tau_{t,J_t} + t \ge S_{i,j}\} + \sum_{t=S_{i,j}}^{U_{i,j}}(R_{t,J_t} - \mu_j) - \sum_{t=S_{i,j}}^{U_{i,j}} R_{t,J_t}\mathbb{I}_i\{\tau_{t,J_t} + t > U_{i,j}\}\right)\\
&\le \sum_{i=1}^{m}\Bigg(\sum_{t=S_{i-1,j}}^{S_i-\nu_{i-1}-1} R_{t,J_t}\mathbb{I}\{\tau_{t,J_t} + t \ge S_i\} + \sum_{t=S_i-\nu_{i-1}}^{S_{i,j}-1} R_{t,J_t}\mathbb{I}\{\tau_{t,J_t} + t \ge S_{i,j}\}\\
&\qquad + \sum_{t=S_{i,j}}^{U_{i,j}}(R_{t,J_t} - \mu_j) - \sum_{t=S_{i,j}}^{U_{i,j}} R_{t,J_t}\mathbb{I}\{\tau_{t,J_t} + t > U_{i,j}\}\Bigg)\\
&= \sum_{i=1}^{m}\left(\sum_{t=S_{i-1,j}}^{S_i-\nu_{i-1}-1} A_{i,t} + \sum_{t=S_i-\nu_{i-1}}^{S_{i,j}-1} B_{i,t} + \sum_{t=S_{i,j}}^{U_{i,j}}(R_{t,J_t} - \mu_j) - \sum_{t=S_{i,j}}^{U_{i,j}} C_{i,t}\right)\\
&= \sum_{i=1}^{m}\sum_{t=S_{i,j}}^{U_{i,j}}(R_{t,J_t} - \mu_j) + \sum_{t=1}^{S_{m,j}} Q_t - \sum_{t=1}^{U_{m,j}} P_t\\
&= \underbrace{\sum_{i=1}^{m}\sum_{t=S_{i,j}}^{U_{i,j}}(R_{t,J_t} - \mu_j)}_{\text{Term I.}} + \underbrace{\sum_{t=1}^{S_{m,j}}(Q_t - \mathbb{E}[Q_t|\mathcal{G}_{t-1}])}_{\text{Term II.}} + \underbrace{\sum_{t=1}^{U_{m,j}}(\mathbb{E}[P_t|\mathcal{G}_{t-1}] - P_t)}_{\text{Term III.}} \qquad (23)\\
&\qquad + \underbrace{\sum_{t=1}^{S_{m,j}}\mathbb{E}[Q_t|\mathcal{G}_{t-1}] - \sum_{t=1}^{U_{m,j}}\mathbb{E}[P_t|\mathcal{G}_{t-1}]}_{\text{Term IV.}},
\end{aligned}$$

Recall that the filtration $\{\mathcal{G}_s\}_{s=0}^{\infty}$ is defined by $\mathcal{G}_0 = \{\Omega, \emptyset\}$ and

$$\mathcal{G}_t = \sigma(X_1, \dots, X_t, J_1, \dots, J_t, \tau_{1,J_1}, \dots, \tau_{t,J_t}, R_{1,J_1}, \dots R_{t,J_t}).$$

Furthermore, we have defined $S_{i,j} = \infty$ if arm j is eliminated before phase i and $S_i = \infty$ if the algorithm stops before reaching phase i.

Outline of proof: We will bound each term of the above decomposition in turn. We first show in Lemma 25 how the bounded second moment information can be incorporated using Chebychev's inequality. In Lemma 26, we show that $Z_t = Q_t - \mathbb{E}[Q_t|\mathcal{G}_{t-1}]$ is a martingale difference sequence and bound its variance in Lemma 27 before using Freedman's inequality. Then in Lemma 28, we provide alternative (tighter) bounds on $A_{i,t}, B_{i,t}, C_{i,t}$ which are used to bound term IV.. All these results are then combined to give a high probability bound on the entire decomposition.

Lemma 25 *For any $a > \lfloor \mathbb{E}[\tau] \rfloor + 1$, $a \in \mathbb{N}$,*

$$\sum_{l=a}^{\infty} \mathbb{P}(\tau \geq l) \leq \frac{\mathbb{V}(\tau)}{a - \lfloor \mathbb{E}[\tau] \rfloor - 1}.$$

Proof: For any $b > a$, $b \in \mathbb{N}$, and by denoting $\xi \doteq \lfloor \mathbb{E}(\tau) \rfloor$,

$$\begin{aligned}
\sum_{l=a}^{b} \mathbb{P}(\tau \geq l) = \sum_{l=a}^{b} \mathbb{P}(\tau - \xi \geq l - \xi) &= \sum_{l=a-\xi}^{b-\xi} \mathbb{P}(\tau - \xi \geq l) \\
&\leq \sum_{l=a-\xi}^{b-\xi} \frac{\mathbb{V}(\tau)}{l^2} \qquad \text{(by Chebychev's inequality since } l + \xi > \mathbb{E}[\tau] \text{ for } l \geq a - \xi) \\
&\leq \mathbb{V}(\tau) \sum_{l=a-\xi-1}^{b-\xi-1} \frac{1}{l(l+1)} \\
&= \mathbb{V}(\tau) \sum_{l=a-\xi-1}^{b-\xi-1} \left(\frac{1}{l} - \frac{1}{l+1} \right) \\
&= \mathbb{V}(\tau) \left(\frac{1}{a - \xi - 1} - \frac{1}{b - \xi} \right).
\end{aligned}$$

Hence, taking $b \to \infty$ gives

$$\sum_{l=a}^{\infty} \mathbb{P}(\tau \geq l) \leq \mathbb{V}(\tau) \frac{1}{a - \xi - 1}.$$

□

Lemma 26 *Let $Y_s = \sum_{t=1}^{s} (Q_t - \mathbb{E}[Q_t|\mathcal{G}_{t-1}])$ for all $s \geq 1$, and $Y_0 = 0$. Then $\{Y_s\}_{s=0}^{\infty}$ is a martingale with respect to the filtration $\{\mathcal{G}_s\}_{s=0}^{\infty}$ with increments $Z_s = Y_s - Y_{s-1} = Q_s - \mathbb{E}[Q_s|\mathcal{G}_{s-1}]$ satisfying $\mathbb{E}[Z_s|\mathcal{G}_{s-1}] = 0, |Z_s| \leq 1$ for all $s \geq 1$.*

Proof: To show $\{Y_s\}_{s=0}^{\infty}$ is a martingale we need to show that Y_s is $\mathcal{G}_s$-measurable for all s and $\mathbb{E}[Y_s|\mathcal{G}_{s-1}] = Y_{s-1}$.

Measurability: We show that $A_{i,s}\mathbb{I}\{S_{i-1,j} \leq s \leq S_i - \nu_{i-1}\} + B_{i,s}\mathbb{I}\{S_i - \nu_{i-1} + 1 \leq s \leq S_{i,j} - 1\}$ is $\mathcal{G}_s$-measurable for every $i \leq m$. This then suffices to show that Y_s is $\mathcal{G}_s$-measurable since each Q_t is a sum of such terms and the filtration $\mathcal{G}_s$ is non-decreasing in s.

First note that by definition of $\mathcal{G}_s$, τ_{t,J_t}, R_{t,J_t} are all $\mathcal{G}_s$-measurable for $t \leq s$. It is sufficient to show that $\mathbb{I}\{\tau_{s,J_s} + s \geq S_i, S_{i-1,j} \leq s \leq S_i - \nu_i\} + \mathbb{I}\{\tau_{s,J_s} + s \geq S_{i,j}, S_i - \nu_{i-1} + 1 \leq s \leq S_{i,j} - 1\}$ is $\mathcal{G}_s$-measurable since the product of measurable functions is measurable. The first summand is $\mathcal{G}_s$ measurable since $\{S_{i-1,j} \leq s\} \in \mathcal{G}_s$ and $\{S_i = s', S_{i-1,j} \leq s\} \in \mathcal{G}_s$ for all $s' \in \mathbb{N} \cup \{\infty\}$. So the union $\bigcup_{s' \in \mathbb{N} \cup \{\infty\}} \{\tau_{s,J_s} + s \geq s', S_{i-1,j} \leq s \leq s' - \nu_i, S_i = s'\} = \{\tau_{s,J_s} + s \geq S_i, S_{i-1,j} \leq s \leq S_i - \nu_{i-1}\}$ is an element of $\mathcal{G}_s$. The same argument works for the second summand since $\{S_{ij} = s', S_i - \nu_{i-1} \leq s\} \in \mathcal{G}_s$ for all $s' \in \mathbb{N} \cup \{\infty\}$

Increments: Hence, to show that $\{Y_s\}_{s=0}^{\infty}$ is a martingale with respect to the filtration $\{\mathcal{G}_s\}_{s=0}^{\infty}$ it just remains to show that the increments conditional on the past are zero. For any $s \geq 1$, we have that

$$Z_s = Y_s - Y_{s-1} = \sum_{t=1}^{s} (Q_t - \mathbb{E}[Q_t|\mathcal{G}_{t-1}]) - \sum_{t=1}^{s-1} (Q_t - \mathbb{E}[Q_t|\mathcal{G}_{t-1}]) = Q_s - \mathbb{E}[Q_s|\mathcal{G}_{s-1}].$$

Then,

$$\mathbb{E}[Z_s|\mathcal{G}_{s-1}] = \mathbb{E}[Q_s - \mathbb{E}[Q_s|\mathcal{G}_{s-1}]|\mathcal{G}_{s-1}] = \mathbb{E}[Q_s|\mathcal{G}_{s-1}] - \mathbb{E}[Q_s|\mathcal{G}_{s-1}] = 0$$

and so $\{Y_s\}_{s=0}^{\infty}$ is a martingale.

Lastly, since for any t and $\omega \in \Omega$, there is only one i where one of $\mathbb{I}\{S_{i-1,j} \le t \le S_i - \nu_{i-1}\}$ or $\mathbb{I}\{S_i - \nu_{i-1} + 1 \le t \le S_{i,j} - 1\}$ is equal to one (they cannot both be one), and by definition of R_{t,J_t}, $A_{i,t}, B_{i,t} \le 1$, it follows that $|Z_s| = |Q_s - \mathbb{E}[Q_s|\mathcal{G}_{s-1}]| \le 1$ for all s. □

Lemma 27 *For any $t \ge 1$, let $Z_t = Q_t - \mathbb{E}[Q_t|\mathcal{G}_{t-1}]$, then*

$$\sum_{t=1}^{S_{m,j}-1} \mathbb{E}[Z_t^2|\mathcal{G}_{t-1}] \le m\mathbb{E}[\tau] + m\mathbb{V}(\tau).$$

Proof: Let us denote $S' \doteq S_{m,j} - 1$. Observe that

$$\begin{aligned}\sum_{t=1}^{S'} \mathbb{E}[Z_t^2|\mathcal{G}_{t-1}] &= \sum_{t=1}^{S'} \mathbb{V}(Q_t|\mathcal{G}_{t-1}) \le \sum_{t=1}^{S'} \mathbb{E}[Q_t^2|\mathcal{G}_{t-1}] \\ &= \sum_{t=1}^{S'} \mathbb{E}\bigg[\bigg(\sum_{i=1}^{m}(A_{i,t}\mathbb{I}\{S_{i-1,j} \le t \le S_i - \nu_{i-1} - 1\} + B_{i,t}\mathbb{I}\{S_i - \nu_{i-1} \le t \le S_{i,j} - 1\})\bigg)^2\bigg|\mathcal{G}_{t-1}\bigg].\end{aligned}$$

Then all indicator terms $\mathbb{I}\{S_{i-1,j} \le t \le S_i - \nu_{i-1} - 1\}$ and $\mathbb{I}\{S_i - \nu_{i-1} \le t \le S_{i,j} - 1\}$ for all $i = 1, \dots, m$ are $\mathcal{G}_{t-1}$-measurable and only one can be non zero for any $\omega \in \Omega$. Hence, for any $\omega \in \Omega$, their product must be 0. Furthermore, for any $i, i' \le m$, $i \ne i'$,

$$\begin{aligned}A_{i,t}\mathbb{I}\{S_{i-1,j} \le t \le S_i - \nu_{i-1} - 1\} \times A_{i',t}\mathbb{I}\{S_{i'-1,j} \le t \le S_{i'} - \nu_{i'-1} - 1\} &= 0, \\ B_{i,t}\mathbb{I}\{S_i - \nu_{i-1} \le t \le S_{i,j} - 1\} \times B_{i',t}\mathbb{I}\{S_{i'} - \nu_{i'-1} \le t \le S_{i',j} - 1\} &= 0, \\ A_{i,t}\mathbb{I}\{S_{i-1,j} \le t \le S_i - \nu_{i-1} - 1\} \times B_{i',t}\mathbb{I}\{S_{i'} - \nu_{i'-1} \le t \le S_{i',j} - 1\} &= 0, \\ A_{i',t}\mathbb{I}\{S_{i'-1,j} \le t \le S_{i'} - \nu_{i'-1} - 1\} \times B_{i,t} \times \mathbb{I}\{S_i - \nu_{i-1} \le t \le S_{i,j} - 1\} &= 0.\end{aligned}$$

Using the above we see that,

$$\begin{aligned}\sum_{t=1}^{S'} \mathbb{E}[Z_t^2|\mathcal{G}_{t-1}] &\le \sum_{t=1}^{S'} \mathbb{E}\bigg[\bigg(\sum_{i=1}^{m}(A_{i,t}\mathbb{I}\{S_{i-1,j} \le t \le S_i - \nu_{i-1} - 1\} + B_{i,t}\mathbb{I}\{S_i - \nu_{i-1} \le t \le S_{i,j} - 1\})\bigg)^2\bigg|\mathcal{G}_{t-1}\bigg] \\ &= \sum_{t=1}^{S'} \mathbb{E}\bigg[\sum_{i=1}^{m}(A_{i,t}^2\mathbb{I}\{S_{i-1,j} \le t \le S_i - \nu_{i-1} - 1\}^2 + B_{i,t}^2\mathbb{I}\{S_i - \nu_{i-1} \le t \le S_{i,j} - 1\}^2)\bigg|\mathcal{G}_{t-1}\bigg] \\ &= \sum_{i=2}^{m}\sum_{t=1}^{S'} \mathbb{E}[A_{i,t}^2\mathbb{I}\{S_{i-1,j} \le t \le S_i - \nu_{i-1} - 1\}|\mathcal{G}_{t-1}] \\ &\qquad + \sum_{i=1}^{m}\sum_{t=1}^{S'} \mathbb{E}[B_{i,t}^2\mathbb{I}\{S_i - \nu_i \le t \le S_{i,j} - 1\}|\mathcal{G}_{t-1}]\end{aligned}$$

(using that both indicators are $\mathcal{G}_{t-1}$-measurable)

$$\le \sum_{i=2}^{m}\sum_{t=S_{i-1,j}}^{S_i-\nu_{i-1}-1} \mathbb{E}[A_{i,t}^2|\mathcal{G}_{t-1}] + \sum_{i=1}^{m}\sum_{t=S_i-\nu_{i-1}}^{S_{i,j}-1} \mathbb{E}[B_{i,t}^2|\mathcal{G}_{t-1}].$$

Then, for any $i \ge 2$,

$$\sum_{t=S_{i-1,j}}^{S_i-\nu_{i-1}-1} \mathbb{E}[A_{i,t}^2|\mathcal{G}_{t-1}] = \sum_{t=S_{i-1,j}}^{S_i-\nu_{i-1}-1} \mathbb{E}[R_{t,J_t}^2\mathbb{I}\{\tau_{t,J_t} + t \ge S_i\}|\mathcal{G}_{t-1}]$$

$$
\begin{aligned}
&\le \sum_{t=S_{i-1,j}}^{S_i-\nu_{i-1}-1} \mathbb{E}[\mathbb{I}\{\tau_{t,J_t}+t\ge S_i\}|\mathcal{G}_{t-1}] \\
&= \sum_{s=0}^{\infty}\sum_{s'=s}^{\infty} \mathbb{I}\{S_{i-1,j}=s, S_i=s'\} \sum_{t=s}^{s'-\nu_{i-1}-1} \mathbb{E}[\mathbb{I}\{\tau_{t,J_t}+t\ge S_i\}|\mathcal{G}_{t-1}] \\
&= \sum_{s=0}^{\infty}\sum_{s'=s}^{\infty} \sum_{t=s}^{s'-\nu_{i-1}-1} \mathbb{E}[\mathbb{I}\{S_{i-1,j}=s, S_i=s', \tau_{t,J_t}+t\ge S_i\}|\mathcal{G}_{t-1}] \\
&\hspace{18em}\text{(Since } \{S_i=s', S_{i-1,j}=s\}\in\mathcal{G}_{t-1} \text{ for } t\ge s\text{)} \\
&= \sum_{s=0}^{\infty}\sum_{s'=s}^{\infty} \sum_{t=s}^{s'-\nu_{i-1}-1} \mathbb{E}[\mathbb{I}\{S_{i-1,j}=s, S_i=s', \tau_{t,J_t}+t\ge s'\}|\mathcal{G}_{t-1}] \\
&= \sum_{s=0}^{\infty}\sum_{s'=s}^{\infty} \mathbb{I}\{S_{i-1,j}=s, S_i=s'\} \sum_{t=s}^{s'-\nu_{i-1}-1} \mathbb{P}(\tau_{t,J_t}+t\ge s') \\
&\hspace{18em}\text{(Since } \{S_i=s', S_{i-1,j}=s\}\in\mathcal{G}_{t-1} \text{ for } t\ge s\text{)} \\
&\le \sum_{s=0}^{\infty}\sum_{s'=s}^{\infty} \mathbb{I}\{S_{i-1,j}=s, S_i=s'\} \sum_{l=\nu_{i-1}+1}^{\infty} \mathbb{P}(\tau>l) \\
&\le \mathbb{V}[\tau],
\end{aligned}
$$

by Lemma 25 since $\nu_i \ge \lfloor \mathbb{E}[\tau]\rfloor + 2$ for all i. Likewise, for any $i \ge 2$,

$$
\begin{aligned}
\sum_{t=S_i-\nu_{i-1}}^{S_{i,j}-1} \mathbb{E}[B_{i,t}^2|\mathcal{G}_{t-1}] &= \sum_{t=S_i-\nu_{i-1}}^{S_{i,j}-1} \mathbb{E}[R_{t,J_t}^2\mathbb{I}\{\tau_{t,J_t}+t\ge S_{i,j}\}|\mathcal{G}_{t-1}] \\
&\le \sum_{t=S_i-\nu_{i-1}}^{S_{i,j}-1} \mathbb{E}[\mathbb{I}\{\tau_{t,J_t}+t\ge S_{i,j}\}|\mathcal{G}_{t-1}] \\
&= \sum_{s=\nu_{i-1}+1}^{\infty}\sum_{s'=s}^{\infty} \mathbb{I}\{S_i=s, S_{i,j}=s'\} \sum_{t=s-\nu_{i-1}}^{s'-1} \mathbb{E}[\mathbb{I}\{\tau_{t,J_t}+t\ge S_{i,j}\}|\mathcal{G}_{t-1}] \\
&= \sum_{s=\nu_{i-1}+1}^{\infty}\sum_{s'=s}^{\infty} \sum_{t=s-\nu_{i-1}}^{s'-1} \mathbb{E}[\mathbb{I}\{S_i=s, S_{i,j}=s', \tau_{t,J_t}+t\ge s'\}|\mathcal{G}_{t-1}] \\
&\hspace{14em}\text{(Since } \{S_{i,j}=s', S_i=s\}\in\mathcal{G}_{t-1} \text{ for } t\ge s-\nu_i-1\text{)} \\
&= \sum_{s=\nu_{i-1}+1}^{\infty}\sum_{s'=s}^{\infty} \mathbb{I}\{S_i=s, S_{i,j}=s'\} \sum_{t=s-\nu_{i-1}}^{s'-1} \mathbb{P}(\tau_{t,J_t}+t\ge s') \\
&\le \sum_{s=\nu_{i-1}+1}^{\infty}\sum_{s'=s}^{\infty} \mathbb{I}\{S_i=s, S_{i,j}=s'\} \sum_{l=0}^{\infty} \mathbb{P}(\tau>l) \\
&\le \mathbb{E}[\tau]
\end{aligned}
$$

and for $i=1$ the derivation simplifies since we need to some over 1 to $S_{1,j}-1$ only. Combining all terms gives the result.
□

Lemma 28 *For $A_{i,t}, B_{i,t}$ and $C_{i,t}$ defined as in (22), let $\nu_i = n_i - n_{i-1}$ be the number of times each arm is played in phase i and j_i' be the arm played directly before arm j in phase i. Then, it holds that, for any arm j and phase $i\ge 1$,*

(i) $$\sum_{t=S_{i-1,j}}^{S_i-\nu_{i-1}-1} \mathbb{E}[A_{i,t}|\mathcal{G}_{t-1}] \le \sum_{l=\nu_{i-1}+1}^{\infty} \mathbb{P}(\tau\ge l).$$

(ii) $\displaystyle\sum_{t=S_i-\nu_{i-1}}^{S_{i,j}-1} \mathbb{E}[B_{i,t}|\mathcal{G}_{t-1}] \leq \sum_{l=\nu_{i-1}+1}^{\infty} \mathbb{P}(\tau \geq l) + \mu_{j'_i} \sum_{l=0}^{\nu_{i-1}} \mathbb{P}(\tau > l).$

(iii) $\displaystyle\sum_{t=S_{i,j}}^{U_{i,j}} \mathbb{E}[C_{i,t}|\mathcal{G}_{t-1}] = \mu_j \sum_{l=0}^{\nu_{i-1}} \mathbb{P}(\tau > l).$

Proof: The proof is very similar to that of Lemma 27. We prove each statement individually.

Statement (i): This is similar to the proof of Lemma 27,

$$\begin{aligned}
\sum_{t=S_{i-1,j}}^{S_i-\nu_{i-1}-1} \mathbb{E}[A_{i,t}|\mathcal{G}_{t-1}] &\leq \sum_{t=S_{i-1,j}}^{S_i-\nu_{i-1}-1} \mathbb{E}[\mathbb{I}\{\tau_{t,J_t}+t \geq S_i\}|\mathcal{G}_{t-1}] \\
&= \sum_{s=0}^{\infty}\sum_{s'=s}^{\infty} \mathbb{I}\{S_{i-1,j}=s, S_i=s'\} \sum_{t=s}^{s'-\nu_{i-1}-1} \mathbb{E}[\mathbb{I}\{\tau_{t,J_t}+t \geq S_i\}|\mathcal{G}_{t-1}] \\
&= \sum_{s=0}^{\infty}\sum_{s'=s}^{\infty} \sum_{t=s}^{s'-\nu_{i-1}-1} \mathbb{E}[\mathbb{I}\{S_{i-1,j}=s, S_i=s', \tau_{t,J_t}+t \geq s'\}|\mathcal{G}_{t-1}] \\
&\qquad\qquad (\text{Since } \{S_i=s', S_{i-1,j}=s\} \in \mathcal{G}_{t-1} \text{ for } t \geq s) \\
&= \sum_{s=0}^{\infty}\sum_{s'=s}^{\infty} \mathbb{I}\{S_{i-1,j}=s, S_i=s'\} \sum_{t=s}^{s'-\nu_{i-1}-1} \mathbb{P}(\tau_{t,J_t}+t \geq s') \\
&\leq \sum_{s=0}^{\infty}\sum_{s'=s}^{\infty} \mathbb{I}\{S_{i-1,j}=s, S_i=s'\} \sum_{l=\nu_{i-1}+1}^{\infty} \mathbb{P}(\tau > l) \\
&= \sum_{l=\nu_{i-1}+1}^{\infty} \mathbb{P}(\tau > l).
\end{aligned}$$

Statement (ii): For statement (ii), we have that for $(i,j) \neq (1,1)$,

$$\sum_{t=S_i-\nu_{i-1}}^{S_{i,j}-1} \mathbb{E}[B_{i,t}|\mathcal{G}_{t-1}] = \sum_{t=S_i-\nu_{i-1}}^{S_{i,j}-\nu_{i-1}-2} \mathbb{E}[B_{i,t}|\mathcal{G}_{t-1}] + \sum_{t=S_{i,j}-\nu_{i-1}-1}^{S_{i,j}-1} \mathbb{E}[B_{i,t}|\mathcal{G}_{t-1}].$$

Then, since$\{S_{i,j}=s'\} \cap \{S_i - \nu_{i-1} \leq t\} \in \mathcal{G}_{t-1}$ so we can use the same technique as for statement (i) to bound the first term. For the second term, since we will be playing only arm j'_i for $S_{i,j}-\nu_{i-1}-1, \dots, S_{i,j}-1$, so,

$$\begin{aligned}
\sum_{t=S_{i,j}-\nu_{i-1}-1}^{S_{i,j}-1} \mathbb{E}[B_{i,t}|\mathcal{G}_{t-1}] &= \sum_{t=S_{i,j}-\nu_{i-1}-1}^{S_{i,j}-1} \mathbb{E}[R_{t,J_t}\mathbb{I}\{\tau_{t,J_t}+t \geq S_{i,j}\}|\mathcal{G}_{t-1}] \\
&= \sum_{s=0}^{\infty} \mathbb{I}\{S_{i,j}=s\} \sum_{t=s-\nu_{i-1}-1}^{s-1} \mathbb{E}[R_{t,J_t}\mathbb{I}\{\tau_{t,J_t}+t \geq S_{i,j}\}|\mathcal{G}_{t-1}] \\
&= \sum_{s=0}^{\infty} \sum_{t=s-\nu_{i-1}-1}^{s-1} \mathbb{E}[R_{t,J_t}\mathbb{I}\{S_{i,j}=s, \tau_{t,J_t}+t \geq S_{i,j}\}|\mathcal{G}_{t-1}] \\
&\qquad\qquad (\text{Since } \{S_{i,j}=s', S_{i,j}-\nu_{i-1} \leq t\} \in \mathcal{G}_{t-1}\ ) \\
&= \sum_{s=0}^{\infty} \sum_{t=s-\nu_{i-1}-1}^{s-1} \mathbb{E}[R_{t,J_t}\mathbb{I}\{S_{i,j}=s, \tau_{t,J_t}+t \geq s\}|\mathcal{G}_{t-1}]
\end{aligned}$$

$$= \sum_{s=0}^{\infty} \mathbb{I}\{S_{i,j} = s\} \sum_{t=s-\nu_{i-1}-1}^{s-1} \mu_{j_i'} \mathbb{P}(\tau_{t,J_t} + t \geq s)$$

(Since $\{S_{i,j} = s\} \in \mathcal{G}_{t-1}$ for $t \geq s - \nu_{i-1} - 1$ and given $\mathcal{G}_{t-1}$, R_{t,J_t} and τ_{t,J_t} are independent)

$$= \sum_{s=0}^{\infty} \mathbb{I}\{S_{i,j} = s\} \mu_{j_i'} \sum_{l=0}^{\nu_{i-1}} \mathbb{P}(\tau > l)$$

$$= \mu_{j_i'} \sum_{l=0}^{\nu_{i-1}} \mathbb{P}(\tau > l).$$

Then, for $(i,j) = (1,1)$, the amount seeping in will be 0, so using $\nu_0 = 0, \mu_{1_1}' = 0$, the result trivially holds. Hence,

$$\sum_{t=S_i-\nu_{i-1}}^{S_{i,j}-1} \mathbb{E}[B_{i,t}|\mathcal{G}_{t-1}] \leq \sum_{l=\nu_{i-1}+1}^{\infty} \mathbb{P}(\tau \geq l) + \mu_{j_i'} \sum_{l=0}^{\nu_{i-1}} \mathbb{P}(\tau > l).$$

Statement (iii): This is the same as in Lemma 16. □

We now bound each term of the decomposition in (23).

Bounding Term I.: For Term I., we can again use Lemma 17 as in the proof of Lemma 1 to get that with probability greater than $1 - \frac{1}{T\tilde{\Delta}_m^2}$,

$$\sum_{i=1}^{m} \sum_{t=S_{i,j}}^{U_{i,j}} (R_{t,J_t} - \mu_j) \leq \sqrt{\frac{n_m \log(T\tilde{\Delta}_m^2)}{2}}.$$

Bounding Term II.: For Term II., we will use Freedmans inequality (Theorem 10). From Lemma 26, $\{Y_s\}_{s=0}^{\infty}$ with $Y_s = \sum_{t=1}^{s}(Q_t - \mathbb{E}[Q_t|\mathcal{G}_{t-1}])$ is a martingale with respect to $\{\mathcal{G}_s\}_{s=0}^{\infty}$ with increments $\{Z_s\}_{s=0}^{\infty}$ satisfying $\mathbb{E}[Z_s|\mathcal{G}_{s-1}] = 0$ and $Z_s \leq 1$ for all s. Further, by Lemma 27, $\sum_{t=1}^{s} \mathbb{E}[Z_t^2|\mathcal{G}_{t-1}] \leq m\mathbb{E}[\tau] + m\mathbb{V}(\tau) \leq \frac{4\times 2^m}{8}(\mathbb{E}[\tau] + \mathbb{V}(\tau)) \leq n_m/8$ with probability 1. Hence we can apply Freedman's inequality to get that with probability greater than $1 - \frac{1}{T\tilde{\Delta}_m^2}$,

$$\sum_{t=1}^{S_{m,j}} (Q_t - \mathbb{E}[Q_t|\mathcal{G}_{t-1}]) = \sum_{s=1}^{\infty} \mathbb{I}\{S_{m,j} = s\} \times Y_s \leq \frac{2}{3}\log(T\tilde{\Delta}_m^2) + \sqrt{\frac{1}{8} n_m \log(T\tilde{\Delta}_m^2)},$$

using that Freedman's inequality applies simultaneously to all $s \geq 1$.

Bounding Term III.: For Term III., we again use Freedman's inequality (Theorem 10), using Lemma 14 to show that $\{Y_s'\}_{s=0}^{\infty}$ with $Y_s' = \sum_{t=1}^{s}(\mathbb{E}[P_t|\mathcal{G}_{t-1}] - P_t)$ is a martingale with respect to $\{\mathcal{G}_s\}_{s=0}^{\infty}$ with increments $\{Z_s'\}_{s=0}^{\infty}$ satisfying $\mathbb{E}[Z_s'|\mathcal{G}_{s-1}] = 0$ and $Z_s' \leq 1$ for all s. Further, by Lemma 15, $\sum_{t=1}^{s} \mathbb{E}[Z_t^2|\mathcal{G}_{t-1}] \leq m\mathbb{E}[\tau] \leq n_m/8$ with probability 1. Hence, with probability greater than $1 - \frac{1}{T\tilde{\Delta}_m^2}$,

$$\sum_{t=1}^{U_{m,j}} (\mathbb{E}[P_t|\mathcal{G}_{t-1}] - P_t) = \sum_{s=1}^{\infty} \mathbb{I}\{U_{m,j} = s\} \times Y_s' \leq \frac{2}{3}\log(T\tilde{\Delta}_m^2) + \sqrt{\frac{1}{8} n_m \log(T\tilde{\Delta}_m^2)}.$$

Bounding Term IV.: To begin with, observe that,

$$\sum_{t=1}^{S_{m,j}} \mathbb{E}[Q_t|\mathcal{G}_{t-1}] - \sum_{t=1}^{U_{m,j}} \mathbb{E}[P_t|\mathcal{G}_{t-1}]$$
$$= \sum_{t=1}^{S_{m,j}} \mathbb{E}\Bigg[\sum_{i=1}^{m} (A_{i,t} \times \mathbb{I}\{S_{i-1,j} \leq t \leq S_i - \nu_{i-1} - 1\} + B_{i,t} \times \mathbb{I}\{S_i - \nu_{i-1} \leq t \leq S_{i,j} - 1\}) \Bigg| \mathcal{G}_{t-1}\Bigg]$$

$$
\begin{aligned}
&\qquad - \sum_{t=1}^{U_{m,j}} \mathbb{E}\Bigg[\sum_{i=1}^{m} C_{i,t} \times \mathbb{I}\{S_{i,j} \le t \le U_{i,j}\}\Bigg|\mathcal{G}_{t-1}\Bigg] \\
&= \sum_{i=1}^{m}\sum_{t=1}^{S_{m,j}} \mathbb{E}[A_{i,t} \times \mathbb{I}\{S_{i-1,j} \le t \le S_i - \nu_{i-1} - 1\}|\mathcal{G}_{t-1}] \\
&\qquad + \sum_{i=1}^{m}\sum_{t=1}^{S_{m,j}} \mathbb{E}[B_{i,t} \times \mathbb{I}\{S_i - \nu_{i-1} \le t \le S_{i,j} - 1\}|\mathcal{G}_{t-1}] \\
&\qquad - \sum_{i=1}^{m}\sum_{t=1}^{U_{m,j}} \mathbb{E}[C_{i,t} \times \mathbb{I}\{S_{i,j} \le t \le U_{i,j}\}|\mathcal{G}_{t-1}] \\
&= \sum_{i=1}^{m}\Bigg(\sum_{t=S_{i-1,j}}^{S_i-\nu_{i-1}-1} \mathbb{E}[A_{i,t}|\mathcal{G}_{t-1}] + \sum_{t=S_i-\nu_{i-1}}^{S_{i,j}-1} \mathbb{E}[B_{i,t}|\mathcal{G}_{t-1}] - \sum_{t=S_{i,j}}^{U_{i,j}} \mathbb{E}[C_{i,t}|\mathcal{G}_{t-1}]\Bigg) \\
&\hfill \text{(using that the indicators are } \mathcal{G}_{t-1}\text{-measurable)} \\
&\le \sum_{i=1}^{m}\Bigg(\sum_{l=\nu_{i-1}+1}^{\infty} \mathbb{P}(\tau \ge l) + \mu_{j_i'}\sum_{l=0}^{\nu_{i-1}} \mathbb{P}(\tau > l) - \mu_j \sum_{l=0}^{\nu_i} \mathbb{P}(\tau > l)\Bigg), \\
&\le \sum_{i=1}^{m}\Bigg(\frac{2\mathbb{V}(\tau)}{\nu_{i-1} - \mathbb{E}[\tau]} + (\mu_{j_i'} - \mu_j)\sum_{l=0}^{\nu_i} \mathbb{P}(\tau > l)\Bigg), \\
&\le \sum_{i=1}^{m}\Bigg(\frac{2\mathbb{V}(\tau)}{2^{i-1}} + (\mu_{j_i'} - \mu_j)\sum_{l=0}^{\nu_i} \mathbb{P}(\tau > l)\Bigg), \qquad (24)
\end{aligned}
$$

by Lemma 28 and Lemma 25 where we have used the fact that since $n_m \le T$, the maximal number of rounds of the algorithm is $\frac{1}{2}\log_2(T/4)$ and for $m \le \frac{1}{2}\log_2(T/4)$, $\frac{\log(T\tilde{\Delta}_m^2)}{\tilde{\Delta}_m^2} \ge \frac{2\log(T\tilde{\Delta}_{m-1}^2)}{\tilde{\Delta}_{m-1}^2}$ so $n_m \ge 2n_{m-1}$ and $\nu_m \ge n_{m-1}$. Then for $\mathbb{E}[\tau] \ge 1$, $\nu_{i-1} - \mathbb{E}[\tau] \ge 2/\tilde{\Delta}_{i-1}\mathbb{E}[\tau] - \mathbb{E}[\tau] \ge (2 \times 2^{i-1} - 1)\mathbb{E}[\tau] \ge 2^{i-1}\mathbb{E}[\tau] \ge 2^{i-1}$ and for $\mathbb{E}[\tau] \le 1$, $\nu_{i-1} - \mathbb{E}[\tau] \ge \nu_{i-1} - 1 \ge 2\log(4)/\tilde{\Delta}_{i-1} - 1 \ge 2^{i-1}$ so $\nu_{i-1} - \mathbb{E}[\tau] \ge 2^{i-1}$. Then, the probability that either arm j_i' or j is active in phase i when it should have been eliminated in or before phase $i-1$ is less than $2p_{i-1}$, where p_i is the probability that the confidence bounds for one arm holds in phase i and $p_0 = 0$. If neither arm should have been eliminated by phase i, this means that their mean rewards are within $\tilde{\Delta}_{i-1}$ of each other. Hence, with probability greater than $1 - 2p_{i-1}$,

$$
\mu_{j_i'}\sum_{l=0}^{\nu_i} \mathbb{P}(\tau > l) - \mu_j \sum_{l=0}^{\nu_i} \mathbb{P}(\tau > l) \le \tilde{\Delta}_{i-1}\sum_{l=0}^{\nu_i} \mathbb{P}(\tau > l) \le \tilde{\Delta}_{i-1}\mathbb{E}[\tau].
$$

Then, summing over all phases gives that with probability greater than $1 - 2\sum_{i=0}^{m-1} p_i$,

$$
\begin{aligned}
\sum_{t=1}^{S_{m,j}} \mathbb{E}[Q_t|\mathcal{G}_{t-1}] - \sum_{t=1}^{U_{m,j}} \mathbb{E}[P_t|\mathcal{G}_{t-1}] &\le 2\mathbb{V}(\tau)\sum_{i=1}^{m}\frac{1}{2^{i-1}} + \mathbb{E}[\tau]\sum_{i=1}^{m}\tilde{\Delta}_{i-1} = (2\mathbb{V}(\tau) + \mathbb{E}[\tau])\sum_{i=0}^{m-1}\frac{1}{2^i} \\
&\le 4\mathbb{V}(\tau) + 2\mathbb{E}[\tau].
\end{aligned}
$$

Combining all terms: To get the final high probability bound, we sum the bounds for each term I.-IV.. Then, with probability greater than $1 - (\frac{3}{T\tilde{\Delta}_m^2} + 2\sum_{i=1}^{m-1} p_i)$, either $j \notin \mathcal{A}_m$ or arm j is played n_m times by the end of phase m and

$$
\begin{aligned}
\frac{1}{n_m}\sum_{t \in T_j(m)} (X_t - \mu_j) &\le \frac{4\log(T\tilde{\Delta}_m^2)}{3n_m} + \left(\frac{2}{\sqrt{8}} + \frac{1}{\sqrt{2}}\right)\sqrt{\frac{\log(T\tilde{\Delta}_m^2)}{n_m}} + \frac{2\mathbb{E}[\tau] + 4\mathbb{V}(\tau)}{n_m} \\
&\le \frac{4\log(T\tilde{\Delta}_m^2)}{3n_m} + \sqrt{\frac{2\log(T\tilde{\Delta}_m^2)}{n_m}} + \frac{2\mathbb{E}[\tau] + 4\mathbb{V}(\tau)}{n_m} = w_m.
\end{aligned}
$$

Using the fact that $p_0 = 0$ and substituting the other p_i's using the same recursive relationship $p_i = \frac{3}{T\tilde{\Delta}_i^2} + 2\sum_{l=1}^{i-1} p_l$ as in the case for bounded delays (see the proof of Lemma 5) gives, $p_m = \frac{12}{T\tilde{\Delta}_m^2}$ so the above bound holds with probability greater than $1 - \frac{12}{T\tilde{\Delta}_m^2}$.

Defining $\mathbf{n_m}$: Setting

$$n_m = \left\lceil \frac{1}{\tilde{\Delta}_m^2}\left(\sqrt{2\log(T\tilde{\Delta}_m^2)} + \sqrt{2\log(T\tilde{\Delta}_m^2) + \frac{8}{3}\tilde{\Delta}_m \log(T\tilde{\Delta}_m^2) + 4\tilde{\Delta}_m(\mathbb{E}[\tau] + 2\mathbb{V}(\tau))}\right)^2 \right\rceil. \tag{25}$$

ensures that $w_m \leq \frac{\tilde{\Delta}_m}{2}$ which concludes the proof. □

Remark: Note that if $\mathbb{E}[\tau] \geq 1$, then the confidence bounds can be tightened by replacing (24) with

$$\sum_{i=1}^{m}\left(\frac{2\mathbb{V}(\tau)}{2^{i-1}\mathbb{E}[\tau]} + (\mu_{j_i'} - \mu_j)\sum_{l=0}^{\nu_i}\mathbb{P}(\tau > l)\right)$$

This is obtained by noting that for $\mathbb{E}[\tau] \geq 1$. $\nu_{i-1} - \mathbb{E}[\tau] \geq 2/\tilde{\Delta}_{i-1}\mathbb{E}[\tau] - \mathbb{E}[\tau] \geq (2 \times 2^{i-1} - 1)\mathbb{E}[\tau] \geq 2^{i-1}\mathbb{E}[\tau]$. This leads to replacing the $\mathbb{V}(\tau)$ term in the definition of n_m by $\mathbb{V}(\tau)/\mathbb{E}[\tau]$.

D.2. Regret Bounds

Theorem 8 *Under Assumption 1 and Assumption 3 of known (bound on) the expectation and variance of the delay, and choice of n_m from (7), the expected regret of Algorithm 1 can be upper bounded by,*

$$\mathbb{E}[\mathfrak{R}_T] \leq \sum_{j=1:\mu_j\neq\mu^*}^{K} O\left(\frac{\log(T\Delta_j^2)}{\Delta_j} + \mathbb{E}[\tau] + \mathbb{V}(\tau)\right).$$

Proof: The proof is very similar to that of Theorem 2, however, for clarity, we repeat the main arguments here. For any sub-optimal arm j, define M_j to be the random variable representing the phase arm j is eliminated in and note that if M_j is finite, $j \in \mathcal{A}_{M_j}$ but $j \notin \mathcal{A}_{M_j+1}$. Then let m_j be the phase arm j *should* be eliminated in, that is $m_j = \min\{m | \tilde{\Delta}_m < \frac{\Delta_j}{2}\}$ and note that, from the new definition of $\tilde{\Delta}_m$ in our algorithm, we get the relations

$$2^m = \frac{1}{\tilde{\Delta}_m}, \quad 2\tilde{\Delta}_{m_j} = \tilde{\Delta}_{m_j-1} \geq \frac{\Delta_j}{2} \quad \text{and so,} \quad \frac{\Delta_j}{4} \leq \tilde{\Delta}_{m_j} \leq \frac{\Delta_j}{2}. \tag{26}$$

Define $\mathfrak{R}_T^{(j)}$ to be the regret contribution from each arm $1 \leq j \leq K$ and let M^* be the round where the optimal arm $j^*$is eliminated. Hence,

$$\begin{aligned}
\mathbb{E}[\mathfrak{R}_T] &= \mathbb{E}\left[\sum_{j=1}^{K}\mathfrak{R}_T^{(j)}\right] = \mathbb{E}\left[\sum_{m=0}^{\infty}\sum_{j=1}^{K}\mathfrak{R}_T^{(j)}\mathbb{I}\{M^* = m\}\right] \\
&= \mathbb{E}\left[\sum_{m=0}^{\infty}\sum_{j:m_j<m}\mathfrak{R}_T^{(j)}\mathbb{I}\{M^* = m\} + \sum_{j:m_j\geq m}\mathfrak{R}_T^{(j)}\mathbb{I}\{M^* = m\}\right] \\
&= \underbrace{\mathbb{E}\left[\sum_{m=0}^{\infty}\sum_{j:m_j<m}\mathfrak{R}_T^{(j)}\mathbb{I}\{M^* = m\}\right]}_{\text{I.}} + \underbrace{\mathbb{E}\left[\sum_{m=0}^{\infty}\sum_{j:m_j\geq m}\mathfrak{R}_T^{(j)}\mathbb{I}\{M^* = m\}\right]}_{\text{II.}}
\end{aligned}$$

We will bound the regret in each of these cases in turn. First, however, we need the following results.

Lemma 29 *For any suboptimal arm j, if $j^* \in \mathcal{A}_{m_j}$, then the probability arm j is* not *eliminated by round m_j is,*

$$\mathbb{P}(M_j > m_j \text{ and } M^* \geq m_j) \leq \frac{24}{T\tilde{\Delta}_{m_j}^2}$$

Proof: The proof is exactly that of Lemma 18 but using Lemma 24 to bound the probability of the confidence bounds on either arm j or j^* failing. □

Define the event $F_j(m) = \{\bar{X}_{m,j^*} < \bar{X}_{m,j} - \tilde{\Delta}_m\} \cap \{j, j^* \in \mathcal{A}_m\}$ to be the event that arm j^* is eliminated by arm j in phase m. The probability of this event is bounded in the following lemma.

Lemma 30 *The probability that the optimal arm j^* is eliminated in round $m < \infty$ by the suboptimal arm j is bounded by*

$$\mathbb{P}(F_j(m)) \leq \frac{24}{T\tilde{\Delta}_m^2}$$

Proof: Again, the proof follows from Lemma 19 but using Lemma 24 to bound the probability of the confidence bounds failing. □

We now return to bounding the expected regret in each of the two cases.

Bounding Term I. As in the proof of Theorem 2, to bound the first term, we consider the cases where arm j is eliminated in or before the correct round ($M_j \leq m_j$) and where arm j is eliminated late ($M_j > m_j$). Then, using Lemma 22,

$$\mathbb{E}\Bigg[\sum_{m=0}^{\infty} \sum_{j:m_j<m} \mathfrak{R}_T^{(j)} \mathbb{I}\{M^* = m\}\Bigg] \leq \sum_{j=1}^{K} \left(2\Delta_j n_{m_j,j} + \frac{384}{\Delta_j}\right)$$

Bounding Term II For the second term, we again use the results from Theorem 2, but using Lemma 29 to bound the probability a suboptimal arm is eliminated in a later round and Lemma 30 to bound the probability j^* is eliminated by a suboptimal arm. Hence,

$$\mathbb{E}\Bigg[\sum_{m=0}^{\infty} \sum_{j:m_j\geq m} \mathfrak{R}_T^{(j)} \mathbb{I}\{M^* = m\}\Bigg] \leq \sum_{j=1}^{K} \frac{1920}{\Delta_j}.$$

Combining the regret from terms I and II gives,

$$\mathbb{E}[\mathfrak{R}_T] \leq \sum_{j=1}^{K} \left(\frac{1920}{\Delta_j} + 2\Delta_j n_{m_j,j}\right)$$

Hence, all that remains is to bound n_m in terms of Δ_j, T and $\mathbb{E}[\tau], \mathbb{V}(\tau)$. Using $L_{m,T} = \log(T\tilde{\Delta}_m^2)$, we have that,

$$\begin{aligned}
n_{m_j,j} &= \left\lceil \frac{1}{\tilde{\Delta}_m^2}\left(\sqrt{2\log(T\tilde{\Delta}_m^2)} + \sqrt{2\log(T\tilde{\Delta}_m^2) + \frac{8}{3}\tilde{\Delta}_m \log(T\tilde{\Delta}_m) + 4\tilde{\Delta}_m(\mathbb{E}[\tau] + 2\mathbb{V}(\tau))}\right)^2 \right\rceil \\
&\leq \left\lceil \frac{1}{\tilde{\Delta}_{m_j}^2}\left(8L_{m_j,T} + \frac{16}{3}\tilde{\Delta}_{m_j} L_{m_j,T} + 8\tilde{\Delta}_{m_j}\mathbb{E}[\tau] + 16\tilde{\Delta}_{m_j}\mathbb{V}(\tau)\right)\right\rceil \\
&\leq 1 + \frac{8L_{m_j,T}}{\tilde{\Delta}_{m_j}^2} + \frac{16L_{m_j,T}}{3\tilde{\Delta}_{m_j}} + \frac{8\mathbb{E}[\tau]}{\tilde{\Delta}_{m_j}} + \frac{16\mathbb{V}(\tau)}{\tilde{\Delta}_{m_j}} \\
&\leq 1 + \frac{128L_{m_j,T}}{\Delta_j^2} + \frac{32L_{m_j,T}}{\Delta_j} + \frac{32\mathbb{E}[\tau]}{\Delta_j} + \frac{64\mathbb{V}(\tau)}{\Delta_j}.
\end{aligned}$$

where we have used $(a+b)^2 \leq 2(a^2+b^2)$ for $a, b \geq 0$.

Hence, the total expected regret from ODAAF with bounded delays can be bounded by,

$$\mathbb{E}[\mathfrak{R}_t] \leq \sum_{j=1}^{K}\left(\frac{256\log(T\Delta_j^2)}{\Delta_j} + 64\mathbb{E}[\tau] + 128\mathbb{V}(\tau) + \frac{1920}{\Delta_j} + 64\log(T) + 2\Delta_j\right).$$

□

Note that again, these constants can be improved at a cost of increasing $\log(T\Delta_j^2)$ to $\log(T\Delta_j)$. We now prove the problem independent regret bound.

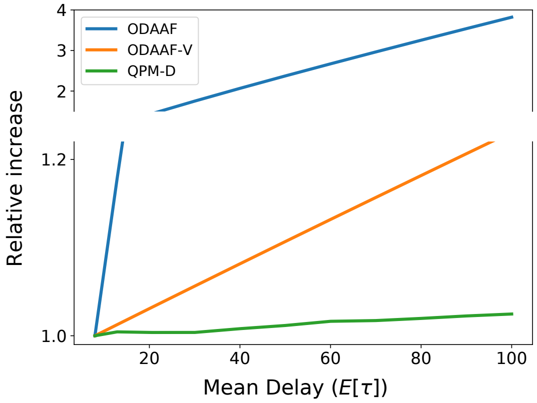


Figure 5: The relative increase in regret at horizon $T = 250000$ for increasing mean delay when the delay is $\mathcal{N}_+$ with variance 100.

Corollary 9 *For any problem instance satisfying Assumptions 1 and 3, the expected regret of Algorithm 1 satisfies*

$$\mathbb{E}[\mathfrak{R}_T] \leq O(\sqrt{KT\log(K)} + K\mathbb{E}[\tau] + K\mathbb{V}(\tau)).$$

Proof: Let $\lambda = \sqrt{\frac{K\log(K)e^2}{T}}$ and note that for $\Delta > \lambda$, $\log(T\Delta^2)/\Delta$ is decreasing in Δ. Then, for constants $C_1, C_2 > 0$ we can bound the regret in the previous theorem by

$$\mathbb{E}[\mathfrak{R}_T] \leq \sum_{j:\Delta_j \leq \lambda} \mathbb{E}[\mathfrak{R}_t^{(j)}] + \sum_{j:\Delta_j > \lambda} \mathbb{E}[\mathfrak{R}_T^{(j)}] \leq \frac{KC_1\log(T\lambda^2)}{\lambda} + KC_2(\mathbb{E}[\tau] + \mathbb{V}(\tau)) + T\lambda.$$

substituting in the above value of λ gives a worst case regret bound that scales with $O(\sqrt{KT\log(K)} + K(\mathbb{E}[\tau] + \mathbb{V}(\tau)))$. □

Remark: If $\mathbb{E}[\tau] \geq 1$, we can replace the $\mathbb{V}(\tau)$ terms in the regret bounds with $\mathbb{V}(\tau)/\mathbb{E}[\tau]$. This follows by using the alternative definition of n_m suggested in the remark at the end of Section D.1.

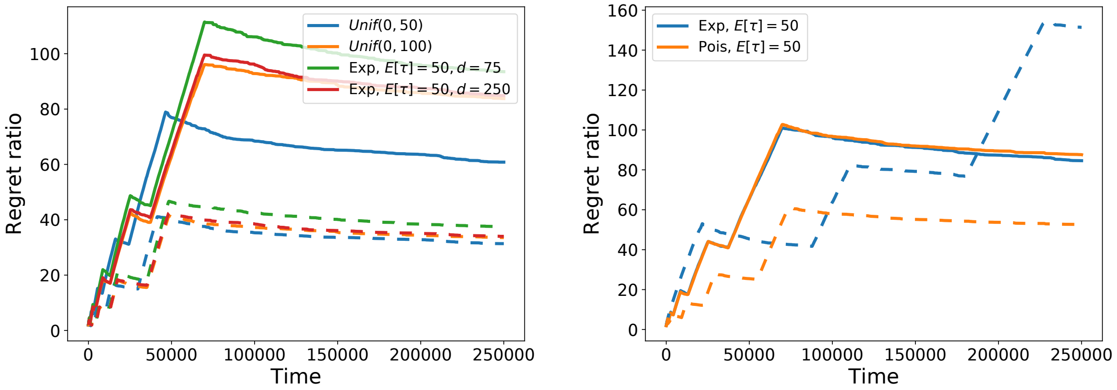
E. Additional Experimental Results

E.1. Increasing the Expected Delay

Here we investigate the effect of increasing the mean delay on both our algorithm and QPM-D (Joulani et al., 2013) and demonstrate that the regret of both algorithms increases linearly with $\mathbb{E}[\tau]$, as indicated by our theoretical results. We use the same experimental set up as described in Section 5. In Figure 5, we are interested in the impact of the mean delay on the regret so we kept the delay distribution family the same, using a $\mathcal{N}_+(\mu, 100)$ (Normal distribution with mean μ, variance 100, truncated at 0) as the delay distribution. We then ran the algorithms for increasing mean delays and plotted the ratio of the regret at T to the regret of the same algorithm when the delay distribution was $\mathcal{N}_+(0, 100)$. In this case, the regret was averaged over 1000 replications for ODAAF and ODAAF-V, and 5000 for QPM-D (this was necessary since the variance of the regret of QPM-D was significant). Here, it can be seen that increasing the mean delay causes the regret of all three algorithms to increase linearly. This is in accordance with the regret bounds which all include a linear factor of $\mathbb{E}[\tau]$ (since here $\log(T)$ is kept constant). It can also be seen that ODAAF-V scales better with $\mathbb{E}[\tau]$ than ODAAF (for constant variance). Particularly, at $\mathbb{E}[\tau] = 100$, the relative increase in ODAAF-V is only 1.2 whereas that of ODAAF is 4 (QPM-D has the best relative increase of 1.05).

E.2. Comparison with Vernade et al. (2017)

Here we compare our algorithms, ODAAF, ODAAF-B and ODAAF-V, to the (non-censored) DUCB algorithm of Vernade et al. (2017). We use the same experimental setup as described in Section 5. As in the comparison to QPM-D, in Figure 6



(a) Bounded delays. Ratios of regret of ODAAF (solid lines) and ODAAF-B (dotted lines) to that of DUCB.

(b) Unbounded delays. Ratios of regret of ODAAF (solid lines) and ODAAF-V (dotted lines) to that of DUCB.

Figure 6: The ratios of regret of variants of our algorithm to that of DUCB for different delay distributions.

we plot the ratios of the cumulative regret of our algorithms to that of DUCB for different delay distributions. In Figure 6a, we consider bounded delay distributions and in Figure 6b, we consider unbounded delay distributions. From these plots, we observe that, as in the comparison to QPM-D in Figure 3, the regret ratios all converge to a constant. Thus we can conclude that the order of regret of our algorithms match that of DUCB, even though the DUCB algorithm of Vernade et al. (2017) has considerably more information about the delay distribution. In particular, along with knowledge on the individual rewards of each play (non-anonymous observations), DUCB also uses complete knowledge of the cdf of the delay distribution to re-weigh the average reward for each arm. Thus, our algorithms are able to match the rate of regret of Vernade et al. (2017) and QPM-D of Joulani et al. (2013) while just receiving aggregated, anonymous observations and using only knowledge of the expected delay rather than the entire cdf.

We ran the DUCB algorithm with parameter $\epsilon = 0$. As pointed out in Vernade et al. (2017), the computational bottleneck in the DUCB algorithm is evaluating the cdf at all past plays of the arms in every round. For bounded delay distributions, this can be avoided using the fact that the cdf will be 1 for plays more than d steps ago. In the case of unbounded distributions, in order to make our experiments computationally feasible, we used the approximation $\mathbb{P}(\tau \leq d) = 1$ for $d \geq 200$. Another nuance of the DUCB algorithm is due to the fact that in the early stages, the upper confidence bounds are dominated by the uncertainty terms, which themselves involve dividing by the cdf of the delay distributions. The arm that is played last in the initialization period will have the highest cdf and so it's confidence bound will be largest and DUCB will play this arm at time $K + 1$ (and possibly in subsequent rounds unless the cdf increases quickly enough). In order to overcome this, we randomize the order that we play the arms in during the initialization period in each replication of the experiment. Note that we did not run DUCB with half normal delays as DUCB divides by the cdf of the delay distribution and in this case the cdf would be 0 at some points.

F. Naive Approach for Bounded Delays

In this section we describe a naive approach to defining the confidence intervals when the delay is bounded by some $d \geq 0$ and show that this leads to sub-optimal regret. Let

$$w_m = \sqrt{\frac{\log(T\tilde{\Delta}_m^2)}{2n_m}} + \frac{md}{n_m}\,.$$

denote the width of the confidence intervals used in phase m for any arm j. We start by showing that the confidence bounds hold with high probability:

Lemma 31 *For any phase m and arm, j,*

$$\mathbb{P}(|\bar{X}_{m,j} - \mu_j| > w_m) \leq \frac{2}{T\tilde{\Delta}_m^2}\,.$$

Proof: First note that since the delay is bounded by d, at most d rewards from other arms can seep into phase i of playing arm j and at most d rewards from arm j can be lost. Defining $S_{i,j}$ and $U_{i,j}$ as the start and end points of playing arm j in phase i, respectively, we have

$$\left| \sum_{t=S_{i,j}}^{U_{i,j}} R_{j,t} - \sum_{t=S_{i,j}}^{U_{i,j}} X_t \right| \leq d\,, \tag{27}$$

because we can pair up some of the missing and extra rewards, and in each pair the difference is at most one. Then, since $T_j(m) = \cup_{i=1}^m \{S_{i,j}, S_{i,j}+1, \dots, U_{i,j}\}$ and using (27) we get

$$\frac{1}{n_m} \left| \sum_{t \in T_j(m)} R_{j,t} - \sum_{t \in T_j(m)} X_t \right| \leq \frac{md}{n_m}\,.$$

Define $\bar{R}_{m,j} = \frac{1}{|T_j(m)|} \sum_{t \in T_j(m)} R_{j,t}$ and recall that $\bar{X}_{m,j} = \frac{1}{|T_j(m)|} \sum_{t \in T_j(m)} X_t$. For any $a > \frac{md}{n_m}$,

$$\begin{aligned} \mathbb{P}\left(|\bar{X}_{m,j} - \mu_j| > a\right) &\leq \mathbb{P}\left(|\bar{X}_{m,j} - \bar{R}_{m,j}| + |\bar{R}_{m,j} - \mu_j| > a\right) \leq \mathbb{P}\left(|\bar{R}_{m,j} - \mu_j| > a - \frac{md}{n_m}\right) \\ &\leq 2\exp\left\{-2n_m\left(a - \frac{md}{n_m}\right)^2\right\}\,, \end{aligned}$$

where the first inequality is from the triangle inequality and the last from Hoeffding's inequality since $R_{j,t} \in [0,1]$ are independent samples from ν_j, the reward distribution of arm j. In particular, taking $a = \sqrt{\frac{\log(T\tilde{\Delta}_m^2)}{2n_m}} + \frac{md}{n_m}$ guarantees that $\mathbb{P}\left(|\bar{X}_j - \mu_j| > a\right) \leq \frac{2}{T\tilde{\Delta}_m^2}$, finishing the proof. □

Observe that setting

$$n_m = \left\lceil \frac{1}{2\tilde{\Delta}_m^2}\left(\sqrt{\log(T\tilde{\Delta}_m^2)} + \sqrt{\log(T\tilde{\Delta}_m^2) + 4\tilde{\Delta}_m md}\right)^2 \right\rceil. \tag{28}$$

ensures that $w_m \leq \frac{\tilde{\Delta}_m}{2}$. Using this, we can substitute this value of n_m into Improved UCB and use the analysis from (Auer & Ortner, 2010) to get the following bound on the regret.

Theorem 32 *Assume there exists a bound $d \geq 0$ on the delay. Then for all $\lambda > 0$, the expected regret of the Improved UCB algorithm run with n_m defined as in (28) can be upper bounded by*

$$\sum_{\substack{j \in A \\ \Delta_j > \lambda}} \left(\Delta_j + \frac{64\log(T\Delta_j^2)}{\Delta_j} + 64\log(2/\Delta_j)d + \frac{96}{\Delta_j}\right) + \sum_{\substack{j \in A \\ 0 < \Delta_j < \lambda}} \frac{64}{\lambda} + T\max_{\substack{j \in A \\ \Delta_j \leq \lambda}} \Delta_j$$

Proof: The result follows from the proof of Theorem 3.1 of (Auer & Ortner, 2010) using the above definition of n_m. □

In particular, optimizing with respect to λ gives worst case regret of $O(\sqrt{KT\log K} + Kd\log T)$. This is a suboptimal dependence on the delay, particularly when $d >> \mathbb{E}[\tau]$.