Asymptotic Optimality for Decentralised Bandits

Conor Newton, Ayalvadi Ganesh, Henry Reeve

ABSTRACT

We consider a large number of agents collaborating on a multi-armed bandit problem with a large number of arms. We present an algorithm which improves upon the *Gossip-Insert-Eliminate* method of Chawla et al. [3]. We provide a regret bound which shows that our algorithm is asymptotically optimal and present empirical results demonstrating lower regret on simulated data.

1. INTRODUCTION

The classical stochastic multi-armed bandit problem is specified by a collection of probability distributions $\{P_k\}_{k=1}^K$, commonly referred to as arms. There is a single agent which plays an arm I_t in $[K] := \{1, \ldots, K\}$ at each time step $t \in [T]$ and receives an associated reward $X_t \sim P_{I_t}$. The agent's goal is to minimise the expected regret $\mathbb{E}[R_T] = T\mu^* - \sum_{t=1}^T \mathbb{E}[X_t]$, where μ^* is the largest mean of the arms. The agent's decisions must be made using only the knowledge acquired from previous actions and observed rewards. A fundamental lower bound on the regret incurred by any bandit algorithm is proved in [7].

Motivated by applications in distributed computing, we consider a collection of agents collaborating on a multi-armed bandit problem [11, 3]. Agents may communicate with one another, and an agent's decision of which arm to play is made using information derived both from their own reward history, and from the sequence of messages received from other agents. However, communication between agents is tightly restricted as described in Section 2. Specifically, time is divided into growing phases and each agent may receive only one message per phase. Furthermore, a message is limited to recommending the id of a single arm; no additional information may be exchanged. We show in Theorem 3.1 that, even with these restrictions on communication, it is possible to asymptotically match the optimal total regret achievable with unlimited communication.

There has recently been growing interest in multi-agent bandits. A setting in which agents communicate with a central node is considered in [6], while [12, 2, 10, 4] consider settings where agents can communicate *rewards* (not just arm ids) with their neighbours. We follow the setting introduced in [11, 3] where agents may only communicate arm ids. In recent work, [1] introduced a method for achieving minimax optimal regret in this setting.

In this work, we consider a different direction and introduce an algorithm (Algorithm 1) which achieves asymptotically optimal instance-dependent regret in the distributed setting (Theorem 3.1). This leads to the surprising conclusion that, in the asymptotic regime, it is possible to perform on a par with algorithms with access to unlimited communication, whilst maintaining the communication constraints of [3, 11]. We also present empirical results demonstrating performance on simulated data (Figures 1, 2).

2. SETTING AND ALGORITHM

We now present our problem setting and algorithm. There are N agents, T time steps, and K arms. Let $X_{k,s}^n \in \{0, 1\}$ denote the reward that agent $n \in [N]$ receives by playing arm $k \in [K]$ for the s^{th} time. We assume that these are i.i.d. Bernoulli(μ_k) random variables. Let $k_* \in \operatorname{argmax} \mu_k$ and let $\mu_* := \mu_{k_*} = \max_{k \in [K]} \mu_k$. We assume throughout that there is a unique best arm, so k_* is uniquely defined.

Communication between agents is constrained by a strictly increasing sequence $(A_j)_{j \in \mathbb{N}}$ and an $N \times N$ probability matrix P as follows. The time horizon $[T]$ is partitioned into phases, with phase j consisting of time steps t for which $A_{j-1} < t \leq A_j$ where $A_0 := 0$. Communication between agents only occurs once a phase, on time steps A_j. On this time step agents request a message from their neighbours. The communicating agent is selected randomly according to P, with $P(n, q)$ denoting the probability that agent n will receive a message from agent q at the end of each phase j. We let $Q \equiv Q_j^n \sim P(n, \cdot)$ be the random variable corresponding to the agent who sends a message to agent n at the end of phase j. This message must take the form of an arm recommendation O_n^j, taking values in $[K]$.

Let $I_t^n \in [K]$ denote the random variable which specifies the index of the arm played by agent n in round t. This must be a measurable function of an agent's previous reward history and the previous messages they have received. We let $V_k^n(t) := \sum_{s=1}^t \mathbb{1}\{I_s^n = k\}$ denote the number of times agent n plays arm k in the first t rounds. Let $X^n(t) := X^n_{I_t^n, V_k^n(t)}$ denote the reward received by agent n in round t.

The goal of each agent $n \in [N]$ is to minimise their expected regret,

$$\mathcal{R}_T^n := T \cdot \mu^* - \sum_{t \in [T]} \mathbb{E}[X^n(t)].$$

Our algorithm (Algorithm 1) is based on the Gossip-Insert-Eliminate algorithm of [3]. Let $\{S_\circ^n\}_{n \in [N]}$ be an arbitrary partition of the set of arms $[K]$. We assume that each agent n is aware of $S_\circ^n$, its own set of arms within the partition, *a priori*. In each phase j, agent n will consider arms from a set $S_j^n \supseteq S_\circ^n$. Let $\hat{\mu}_{k,s}^n := \frac{1}{s}\sum_{i=1}^s X_{k,i}^n$. Denote by $\hat{\mu}_k^n(t) := \hat{\mu}_{k,V_k^n(t)}^n$ the mean reward obtained by agent n from arm k in the first t time steps.

We let M_j^n denote the most played arm by agent n in phase j so $M_j^n \in \operatorname{argmax}_{k\in[K]}\{V_k^n(A_j) - V_k^n(A_{j-1})\}$. Following [3], when an agent $q \in [N]$ is asked for an arm recommendation at the end of phase j, its recommendation will be its most played arm for that phase. Hence, when $Q \equiv Q_j^n \sim P(n,\cdot)$ communicates with agent $n \in [N]$ at the end of phase j, the recommendation will be $O_n^j = M_j^Q$.

Algorithm 1 makes two modifications to GoIE from [3]. Firstly, we use tighter KL based confidence intervals, following [5]. Secondly, we use a more efficient elimination scheme. To define our KL upper confidence bounds we first let $d : [0,1]^2 \to \mathbb{R}$ be the Kullback–Leibler divergence for two Bernoulli random variables and introduce a function $f_\alpha(t) = 1 + t^\alpha \log^2(t)$ indexed by α. The upper confidence bound for arm k at agent n at time t is defined by

$$\mathrm{U}_{k,\alpha}^n(t-1) := \max\left\{u \in [0,1] : d(\hat{\mu}_k^n(t-1), u) \le \frac{\log(f_\alpha(t))}{V_k^n(t-1)}\right\}$$

when $V_k^n(t-1) > 0$ and $\mathrm{U}_k^n(t-1) := \infty$ otherwise. When α is clear from context we suppress it for notational convenience.

Algorithm 1: Asymptotically Optimal Gossiping Bandits (AOGB)

$j \leftarrow 1$ and $S_1^n \leftarrow S_\circ^n$
for $t \in \mathbb{N}$ **do**
  $I_t^n \leftarrow \operatorname{argmax}_{k\in S_j^n} \mathrm{U}_{k,\alpha}^n(t-1)$
  if $t == A_j$ **then**
    $Q \leftarrow P(i,\cdot)$ and $O_j^n = M_j^Q$
    $S_{j+1}^n \leftarrow S_\circ^n \cup \{O_j^n, M_j^n\}$
    $j \leftarrow j+1$
  end
end

3. REGRET BOUND

We now present our asymptotically optimal regret bound for Algorithm 1.

THEOREM 3.1. *Suppose that P has a strongly connected graph and there exist $C \ge 1$, $\theta > 0$ such that $C^{-1}j^\theta \le A_j - A_{j-1} \le Cj^\theta$ for all $j \in \mathbb{N}$. Suppose that all agents select arms with Algorithm 1 with $\alpha = 1$. Then for each agent $n \in [N]$ we have the asymptotic bound*

$$\limsup_{T\to\infty} \frac{\mathcal{R}_T^n}{\log T} \le \sum_{k\in S_\circ^n\setminus[k_*]} \frac{\mu_* - \mu_k}{d(\mu_k, \mu_*)}.$$

Note that by summing over the regrets of the different agents the regret bound above matches the lower bound for the full communication setting implied by [7].

The proof of Theorem 3.1 hinges upon a random time $\hat{\tau}$ which corresponds to the phase after which all of the active sets S_j^n become fixed. After this random time all of the active sets become $S_\circ^n \cup \{k_*\}$, which leads to an asymptotic regret bound for agent n governed by the relationship between μ_k and μ_* for $k \in [K]$. The crucial difficulty then is to bound $\mathbb{E}[A_{\hat{\tau}}]$, the expected time until the end of phase $\hat{\tau}$. To bound $\mathbb{E}[A_{\hat{\tau}}]$ we show that, provided the phase lengths $A_j - A_{j-1}$ are sufficiently large in relationship to the gap, the probability of a sub-optimal arm being the most played, and subsequently being recommended decays exponentially.

4. NUMERICAL RESULTS

Here we will compare algorithm 1 and the GosInE algorithm on a range of synthetic data. We trial variants of both of these algorithms using sub-gaussian and KL upper confidence bounds. For GosInE, the sub-gaussian and KL variants are respectively labelled UCB-GIE and KLUCB-GIE and for algorithm 1, they are labelled GIE-FE (Gossip-Insert-Eliminate with Fast Elimination) and AOGB.

All the experiments are run with $N = 20$ nodes, $K = 50$ arms and phases growing cubically, i.e., $A_j = j^3$. Each experiment consists of 100 independent runs, and in each run the regret is averaged over the nodes. In each experiment, the algorithms encounter the same reward sequence. The first two experiments assume the agents are connected via a complete graph, while the third experiment compares different graphs. We compute the regret over a time horizon of $T = 100,000$ and plot the mean along with 95% confidence intervals.

Choice of α: We begin by comparing algorithm 1 and GosInE for the two different types of upper confidence bounds by varying the exploration function $f(t) = 1 + t^\alpha \log^2(t)$ by choosing different values for α. From figure 1 we identify

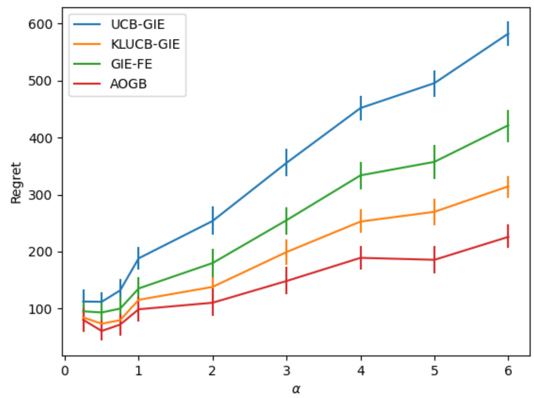

Figure 1: Regret for different choices of α with $\mu_* = 0.9$ and the rest of the arms divide the interval $[0.2, 0.8]$ uniformly.

that algorithm 1 and GosInE perform better when equipped with KL upper confidence bound. Additionally, algorithm 1 outperforms GosInE when they are both equipped with the same upper confidence bounds. Overall, performance is better for the smaller values of α and regret is minimised somewhere in the region $\alpha \le 1$. This implies that there may be more practical choices for $f_\alpha(t)$ than the asymptotically optimal choice at $\alpha = 1$.

$\Delta_{\min}$ vs Regret: Now we consider the affect of changing the sub-optimality gap $\Delta_{\min}$. This is the difference between the mean of the best arm and the second best arm. Figure 2 compares algorithm 1 and the GosInE algorithm for both types of confidence intervals. Similarly to the previous experiment, we observe that both algorithms perform better when equipped with the KL upper confidence bounds

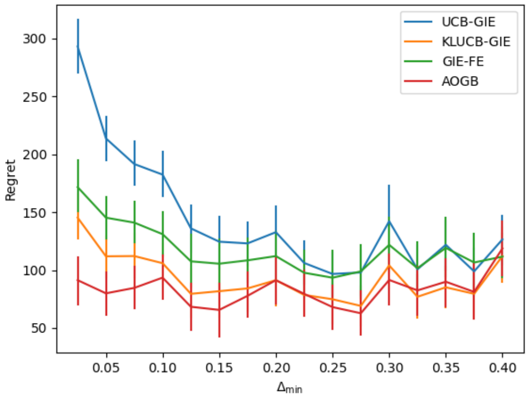

Figure 2: Regret for different choices of $\Delta_{\min}$ with $\alpha = 1$. The best arm has mean $\mu_\star = 0.9$ and the rest of the arms divide the interval $[0.9 - \Delta_{\min}, 0.2]$ uniformly.

and that algorithm 1 typically outperforms GosInE on average when they are equipped with the same upper confidence bounds.

Network Configurations: Here, we compare three different network configurations for agents implementing algorithm 1: a complete graph, a cycle graph and a star graph.

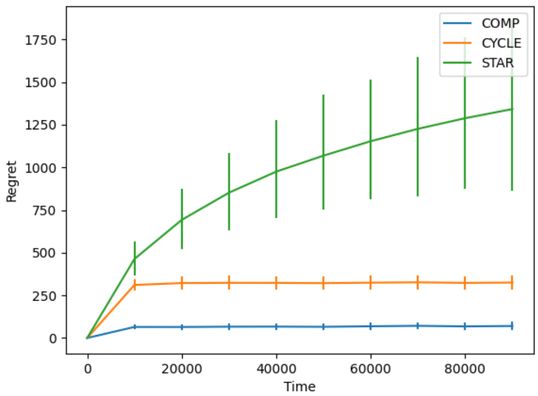

Figure 3: Regret over time for three different networks. Each in case we consider $\alpha = 1$, $\Delta_{\min} = 0.1$ and the means of the remaining arms divide the interval $[0.8, 0.2]$ uniformly.

The results in figure 3 show that the cycle graph performs slightly worse than the complete graph but the star graph struggles significantly along with a larger variance. In essence, this is because the best arm needs to spread to centre of the star before it can spread to all of the other nodes.

5. DISCUSSION

In this paper we presented an algorithm (Algorithm 1) for multi-agent bandits in a decentralised setting. Our algorithm builds upon the Gossip-Insert-Eliminate algorithm of [3] by making two modifications. First, we use tighter confidence intervals inspired by [5]. Second, we use a faster elimination scheme for reducing the number of arms that must be explored by an agent. Both modifications yield significant empirical improvement on simulated data (Figure 2). Finally, we prove a regret bound (Theorem 3.1) which demonstrates asymptotically optimal performance of our algorithm, matching the asymptotic performance of a collection of agents with unlimited communication.

There is substantial scope for future work in this direction. One challenge of great practical importance is the development of distributed algorithms which are robust to both malicious agents and faulty communication [9]. An interesting theoretical challenge is to develop a multi-agent bandit algorithm which is both asymptotically optimal and nearly minimax optimal with limited communication. In very recent work of [1] an algorithm has been proposed which is minimax optimal in the distributed setting, and it would be interesting to synthesise this with the insights provided in the current paper.

6. REFERENCES

[1] Mridul Agarwal, Vaneet Aggarwal, and Kamyar Azizzadenesheli. Multi-agent multi-armed bandits with limited communication. *arXiv preprint arXiv:2102.08462*, 2021.

[2] Nicolò Cesa-Bianchi, Tommaso Cesari, and Claire Monteleoni. Cooperative online learning: Keeping your neighbors updated. In *Algorithmic Learning Theory*, pages 234–250. PMLR, 2020.

[3] Ronshee Chawla, Abishek Sankararaman, Ayalvadi Ganesh, and Sanjay Shakkottai. The gossiping insert-eliminate algorithm for multi-agent bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 3471–3481. PMLR, 2020.

[4] Abhimanyu Dubey and Alex 'Sandy' Pentland. Cooperative multi-agent bandits with heavy tails. In *International Conference on Machine Learning*, pages 2730–2739. PMLR, 2020.

[5] Aurélien Garivier and Olivier Cappé. The kl-ucb algorithm for bounded stochastic bandits and beyond. In *Proceedings of the 24th annual conference on learning theory*, pages 359–376. JMLR Workshop and Conference Proceedings, 2011.

[6] Varun Kanade, Zhenming Liu, and Bozidar Radunovic. Distributed non-stochastic experts. *Neurips*, 2012.

[7] T.L Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22, 1985.

[8] Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

[9] Nancy A Lynch. *Distributed algorithms*. Elsevier, 1996.

[10] David Martínez-Rubio, Varun Kanade, and Patrick Rebeschini. Decentralized cooperative stochastic bandits. *Neurips*, 2018.

[11] Abishek Sankararaman, Ayalvadi Ganesh, and Sanjay Shakkottai. Social learning in multi agent multi armed bandits. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 3(3):1–35, 2019.

[12] Balazs Szorenyi, Róbert Busa-Fekete, István Hegedus, Róbert Ormándi, Márk Jelasity, and Balázs Kégl. Gossip-based distributed stochastic bandit algorithms. In *International Conference on Machine Learning*, pages 19–27. PMLR, 2013.

7. APPENDIX I: PROOF OF THE ASYMPTOTIC REGRET BOUND

In this section we present the proof of Theorem 3.1. Throughout we restrict our attention to Algorithm 1 with $\alpha = 1$ and let $f(t) := 1 + t\log^2(t)$. For each $k \in [K]$ we let $\Delta_k := \mu_{k_*} - \mu_k$ and define $\Delta_{\min} := \min_{k \in [K]\setminus\{k_*\}} \Delta_k > 0$.

Next we recall some useful results from the bandit literature adapted to our setting. For each $\epsilon \in (0, \Delta_{\min})$ and $n \in [N]$ we define a random variable

$$\kappa_\epsilon^n := \min\left\{t \in \mathbb{N} : \max_{s \in [T]} \left(\underline{d}\left(\hat{\mu}_{k_*,s}^n, \mu_* - \epsilon\right) - \frac{\log(f(t))}{s}\right) \leq 0\right\},$$

where $\underline{d}(p,q) := d(p,q) \cdot \mathbb{1}\{p \leq q\}$.

Lemma 7.1. *For $\epsilon \in (0, \Delta_{\min})$, $\max_{n \in [N]} \mathbb{E}[\kappa_\epsilon^n] \leq 2/\epsilon^2$.*

Proof. See [8, Lemma 10.7]. □

Next we define for each $\epsilon \in (0, \Delta_{\min})$, $n \in [N]$, $k \in [K]\backslash[k_*]$,

$$\nu_{\epsilon,k}^n := \sum_{s=1}^{T} \mathbb{1}\left\{d(\hat{\mu}_{k,s}^n, \mu_* - \epsilon) \leq \frac{\log(f(T))}{s}\right\}.$$

Lemma 7.2. *For $\epsilon \in (0, \Delta_{\min})$, and $n \in [N]$ we have*

$$\mathbb{E}[\nu_{\epsilon,k}^n] \leq \inf_{\tilde{\epsilon} \in (0, \Delta_k - \epsilon)} \left(\frac{\log f(T)}{d(\mu_k + \tilde{\epsilon}, \mu_* - \epsilon)} + \frac{1}{2\tilde{\epsilon}^2}\right).$$

Proof. See [8, Lemma 10.8]. □

Let $\kappa_\circ^n := \kappa_{\Delta_{\min}/2}^n$ and for each agent $n \in [N]$ and phase j we define a Boolean random variable $\chi_j^n := \mathbb{1}\{k_* \in S_j^n,\ M_j^n \neq k_*, A_{j-1} \geq \kappa_\circ^n\}$. For each $n \in [N]$ we define

$$\begin{aligned}
\hat{\tau}_{\text{stab}}^n &:= \min\{j \in \mathbb{N}\ \ :\ A_{j-1} \geq \kappa_\circ^n, \forall j' \geq j,\ \chi_{j'}^n = 0\} \\
\hat{\tau}_{\text{stab}} &:= \max_{n \in [N]} \hat{\tau}_{\text{stab}}^n \\
\hat{\tau}_{\text{spr}}^n &:= \min\{j \geq \hat{\tau}_{\text{stab}}\ \ :\ k_* \in S_j^n\} - \hat{\tau}_{\text{stab}} \\
\hat{\tau}_{\text{spr}} &:= \max_{n \in [N]} \hat{\tau}_{\text{spr}}^n \\
\hat{\tau} &:= \hat{\tau}_{\text{stab}} + \hat{\tau}_{\text{spr}}.
\end{aligned}$$

Lemma 7.3. *For all phases $j > \hat{\tau}$ and all $n \in [N]$ we have $S_j^n = S_\circ^n \cup \{k_*\}$.*

Proof. For each $n \in [N]$ we see by induction that for all $j \geq \hat{\tau}_{\text{spr}}^n + \hat{\tau}_{\text{stab}}$ we have $M_j^n = k_* \in S_j^n$. Moreover, since $S_{j+1}^n = S_\circ^n \cup \{M_j^n, M_j^Q\}$ for some Q in $[N]$ it follows that $S_{j+1}^n = S_\circ^n \cup \{k_*\}$, for all $j \geq \hat{\tau} = \hat{\tau}_{\text{stab}} + \hat{\tau}_{\text{spr}}$. □

Lemma 7.4. *For each $n \in [N]$ and $k \in [K]\backslash\{k_*\}$ we have*

$$\sum_{t=A_{\hat{\tau}}+1}^{T} \mathbb{1}\{I_t^n = k\} \leq \begin{cases} \inf_{\epsilon \in (0,\Delta_{\min})} \left\{\nu_{\epsilon,k}^n + \kappa_\epsilon^n\right\} & \text{if } k \in S_\circ^n \\ 0 & \text{if } k \notin S_\circ^n. \end{cases}$$

Proof. By Lemma 7.3 First note that by Lemma 7.3 we have $S_j^n = S_\circ^n\{k_*\}$ for all $j > \hat{\tau}$. In particular, this means that $I_t^n \notin S_\circ^n \cup \{k_*\}$ cannot occur for $t \geq A_{\hat{\tau}} + 1$. Now take $\epsilon \in (0, \Delta_{\min})$ and consider $k \in S_\circ^n\backslash[k_*]$. If $I_t^n = k$ for some $t \geq (A_{\hat{\tau}} + 1) \vee \kappa_\epsilon^n$ then we must have $\mathrm{U}_k^n(t-1) \geq \mathrm{U}_{k_*}^n(t-1) \geq \mu_* - \epsilon$, and hence,

$$d(\hat{\mu}_{k,V_k^n(t-1)}^n, \mu_* - \epsilon) \leq \frac{\log(f(t))}{V_k^n(t-1)} \leq \frac{\log(f(T))}{V_k^n(t-1)},$$

and $V_k^n(t) = V_k^n(t-1) + 1$. Consequently, $\sum_{t=(A_{\hat{\tau}}+1)\vee\kappa_\epsilon^n}^{T} \mathbb{1}\{I_t^n = k\} \leq \nu_{\epsilon,k}^n$. The result follows by taking an infimum over $\epsilon \in (0, \Delta_{\min})$. □

This leads to the following regret bound.

Corollary 7.5. *For each $n \in [N]$, we have $\mathbb{E}[R_T^n] \leq \mathbb{E}[A_{\hat{\tau}}] + \sum_{k \in S_\circ^n\backslash[k_*]} \Delta_k \inf_{\epsilon \in \left(0, \frac{\Delta_{\min}}{2}\right)} \left\{\frac{\log f(T)}{d(\mu_k+\epsilon, \mu_*-\epsilon)} + \frac{3}{\epsilon^2}\right\}$.*

Proof. This follows from Lemmas 7.1, 7.2 and 7.4. □

For the remainder of the proof we must show that $\mathbb{E}[A_{\hat{\tau}}]$ may be bounded independently of T.

Lemma 7.6. *For $j \in \mathbb{N}$ s.t. $A_j - A_{j-1} \geq \frac{8}{\Delta^2}\left(\frac{K}{N} + 3\right)\log f(A_j)$, we have $\mathbb{E}[\chi_j^n] \leq \frac{8K}{\Delta_{\min}^2} \exp\left(-\frac{\Delta_{\min}^2 (A_j - A_{j-1})}{16(K/N+3)}\right)$.*

PROOF. First observe that if $\chi_j^n = 1$ then $k_* \in S_j^n$, $A_{j-1} \geq \kappa_\circ^n$ and $M_j^n \neq k_*$. Since $M_j^n \neq k_*$ we deduce that for some $k \in [K]\backslash\{k_*\}$ we have

$$V_k^n(A_j) - V_k^n(A_{j-1}) \geq \frac{A_j - A_{j-1}}{|S_j^n|} \geq \frac{A_j - A_{j-1}}{K/N+3},$$

and so for some $A_{j-1} < t \leq A_j$ we have $s = V_k^n(t-1) \geq \frac{A_j - A_{j-1}}{K/N+3} - 1$ and $I_t^n = k$, so $\mathrm{U}_k^n(t-1) \geq \mathrm{U}_{k_*}^n(t-1)$ as $k_* \in S_j^n$. Since $t \geq A_{j-1} \geq \kappa_\circ^n$ we deduce that $\mathrm{U}_k^n(t-1) \geq \mathrm{U}_{k_*}^n(t-1) \geq \mu_* - \Delta_{\min}/2$. Hence, by Pinsker's inequality

$$2\left(\hat{\mu}_{k,s}^n - \mu_* + \frac{\Delta_{\min}}{2}\right)^2 = 2\left(\hat{\mu}_k^n(t-1) - \mu_* + \frac{\Delta_{\min}}{2}\right)^2 \leq d\left(\hat{\mu}_k^n(t-1), \mu_* - \frac{\Delta_{\min}}{2}\right) \leq \frac{\log(f_\alpha(t))}{V_k^n(t-1)} \leq \frac{\log f(A_j)}{s}.$$

Thus, for some $k \in [K]\backslash\{k_*\}$ and $s \geq \frac{A_j - A_{j-1}}{K/N+3} - 1$,

$$\hat{\mu}_{k,s}^n \geq \mu_* - \frac{\Delta_{\min}}{2} - \sqrt{\frac{\log f(A_j)}{2s}} \geq \mu_k + \frac{\Delta_{\min}}{2} - \sqrt{\frac{\log f(A_j)}{2s}} \geq \mu_k + \frac{\Delta_{\min}}{4},$$

since $A_j - A_{j-1} \geq \frac{8}{\Delta^2}\left(\frac{K}{N} + 3\right)\log f(A_j)$. Thus, by Hoeffding's inequality we have

$$\begin{aligned}
\mathbb{E}[\chi_j^n] &\leq \sum_{k \in [K]\backslash\{k_*\}} \sum_{s \geq \frac{A_j - A_{j-1}}{K/N+3} - 1} \mathbb{P}\left[\hat{\mu}_{k,s}^n \geq \mu_k + \frac{\Delta_{\min}}{4}\right] \\
&\leq (K-1) \sum_{s \geq \frac{A_j - A_{j-1}}{K/N+3} - 1} \exp\left(-\frac{s\Delta_{\min}^2}{8}\right) \\
&\leq K \int_{\frac{A_j - A_{j-1}}{K/N+3} - 2}^{\infty} \exp\left(-\frac{s\Delta_{\min}^2}{8}\right) ds \\
&\leq \frac{8K}{\Delta_{\min}^2} \exp\left(-\frac{\Delta_{\min}^2(A_j - A_{j-1})}{16(K/N+3)}\right).
\end{aligned}$$

□

In what follows we let $p_{\min} := \min\left(\{P(i,j)\}_{(i,j)\in[N]^2}\backslash\{0\}\right)$ and $\mathrm{diam}(P)$ denote the maximum length of a directed path between two distinct nodes corresponding to the graph induced by P. We note that $\mathrm{diam}(P) < \infty$ if and only if P has a strongly connected graph.

LEMMA 7.7. *Suppose that P has a strongly connected graph. Then for $\xi \in \mathbb{N}$, $\mathbb{P}(\hat{\tau}_{\mathrm{spr}} \geq \xi) \leq N(1 - p_{\min}^{\mathrm{diam}(P)})^{\left\lfloor \frac{\xi}{2\mathrm{diam}(P)} - 1 \right\rfloor}$.*

PROOF. Fix $n \in [N]$ and choose a sequence $(\ell_i)_{i \in [q]\cup\{0\}} \in [K]^q$ with $q \leq \mathrm{diam}(P)$ and such that $\ell_0 = k_*$, $\ell_q = n$ and $P(\ell_i, \ell_{i-1}) > 0$ for each $i \in [q]$. Note that the definition of $\mathrm{diam}(P)$ entails the existence of at least one such a sequence. Recall that we let $Q_j^{\tilde{n}}$ denote the node which sends a message to agent $\tilde{n}$ and the end of phase j. Let $m = \lfloor \xi/(2q) - 1 \rfloor$ and observe that if for some $j_0 \in \{\hat{\tau}_{\mathrm{stab}}, \ldots, \hat{\tau}_{\mathrm{stab}} + 2mq\}$ we have $Q_j^{\ell_{j-j_0}} = \ell_{j-j_0-1}$ for $j \in \{j_0+1, \ldots, j_0+q\}$ then $\hat{\tau}_{\mathrm{spr}}^n + \hat{\tau}_{\mathrm{stab}} \leq j_0 + q < \xi + \hat{\tau}_{\mathrm{stab}}$, so . Hence, we have

$$\begin{aligned}
\mathbb{P}(\hat{\tau}_{\mathrm{spr}}^n \geq \xi) &\leq \mathbb{P}\left(\bigcap_{j_0 - \hat{\tau}_{\mathrm{stab}} \in \{0, 2q, \ldots, 2mq\}} \bigcup_{j \in \{j_0+1, \ldots, j_0+q\}} \left\{Q_j^{\ell_{j-j_0}} \neq \ell_{j-j_0-1}\right\}\right) \\
&= \prod_{j_0 - \hat{\tau}_{\mathrm{stab}} \in \{0, 2q, \ldots, 2mq\}} \mathbb{P}\left(\bigcup_{j \in \{j_0+1, \ldots, j_0+q\}} \left\{Q_j^{\ell_{j-j_0}} \neq \ell_{j-j_0-1}\right\}\right) \\
&= \prod_{j_0 - \hat{\tau}_{\mathrm{stab}} \in \{0, 2q, \ldots, 2mq\}} \left\{1 - \mathbb{P}\left(\bigcap_{j \in \{j_0+1, \ldots, j_0+q\}} \left\{Q_j^{\ell_{j-j_0}} = \ell_{j-j_0-1}\right\}\right)\right\} \\
&= \prod_{j_0 - \hat{\tau}_{\mathrm{stab}} \in \{0, 2q, \ldots, 2mq\}} \left\{1 - \prod_{j \in \{j_0+1, \ldots, j_0+q\}} \mathbb{P}\left(Q_j^{\ell_{j-j_0}} = \ell_{j-j_0-1}\right)\right\} \\
&\leq (1 - p_{\min}^q)^m \leq (1 - p_{\min}^{\mathrm{diam}(P)})^{\left\lfloor \frac{\xi}{2\mathrm{diam}(P)} - 1 \right\rfloor}.
\end{aligned}$$

The lemma now follows by the union bound over $[N]$. □

LEMMA 7.8. *Suppose that there exist $C \geq 1$, $\theta > 0$ such that $C^{-1}j^\theta \leq A_j - A_{j-1} \leq Cj^\theta$ for all $j \in \mathbb{N}$. Then we have $C^{-1}j^{1+\theta} \leq A_j \leq C(1+j)^{1+\theta}$ for all $j \in \mathbb{N}$.*

Now define $\underline{j}(\Delta_{\min}) \in \mathbb{N}$ by

$$\underline{j}(\Delta_{\min}) := 1 + \max\left(\{0\} \cup \left\{ j \in \mathbb{N} \;:\; j^{\theta} < \frac{8C}{\Delta_{\min}^2}\left(\frac{K}{N}+3\right)\log f\left(C(1+j)^{\theta}\right)\right\}\right).$$

Note that $\underline{j}(\Delta_{\min})$ is always finite since $f(t) = O(\log t)$.

LEMMA 7.9. *Suppose that there exist $C \geq 1$, $\theta > 0$ such that $C^{-1}j^{\theta} \leq A_j - A_{j-1} \leq Cj^{\theta}$ for all $j \in \mathbb{N}$. Then for all $\xi \geq \underline{j}(\Delta_{\min})$ we have*

$$\mathbb{P}(\hat{\tau}_{\mathrm{stab}} \geq \xi) \leq \sum_{n\in[N]} \mathbb{P}(\kappa_{\circ}^n > C^{-1}(\xi-2)^{1+\theta}) + \frac{8KN}{\Delta_{\min}^2}\sum_{j\geq\xi}\exp\left(-\frac{\Delta_{\min}^2 j^{\theta}}{16C(K/N+3)}\right).$$

PROOF. Fix $n \in [N]$ and suppose that $\hat{\tau}_{\mathrm{stab}}^n \geq \xi$. Since $\hat{\tau}_{\mathrm{stab}}^n := \min\{j \in \mathbb{N} \;:\; A_{j-1} \geq \kappa_{\circ}^n, \forall j' \geq j,\ \chi_{j'}^n = 0\}$ it follows that either $A_{\xi-2} < \kappa_{\circ}^n$ or $\chi_j^n = 1$ for some $j \geq \xi$. Note also that by the upper bound in Lemma 7.8 for $j \geq \xi \geq \underline{j}(\Delta_{\min})$ we have

$$A_j - A_{j-1} \geq C^{-1}j^{\theta} \geq \frac{8}{\Delta_{\min}^2}\left(\frac{K}{N}+3\right)\log f\left(C(1+j)^{\theta}\right) \geq \frac{8}{\Delta^2}\left(\frac{K}{N}+3\right)\log f(A_j).$$

Hence, by Lemmas 7.6 and the lower bound in 7.8 we have

$$\begin{aligned}
\mathbb{P}(\hat{\tau}_{\mathrm{stab}}^n \geq \xi) &\leq \mathbb{P}(A_{\xi-2} < \kappa_{\circ}^n) + \sum_{j\geq\xi}\mathbb{E}[\chi_j^n] \\
&\leq \mathbb{P}(\kappa_{\circ}^n > C^{-1}(\xi-2)^{1+\theta}) + \frac{8K}{\Delta_{\min}^2}\sum_{j\geq\xi}\exp\left(-\frac{\Delta_{\min}^2(A_j - A_{j-1})}{16(K/N+3)}\right) \\
&\leq \mathbb{P}(\kappa_{\circ}^n > C^{-1}(\xi-2)^{1+\theta}) + \frac{8K}{\Delta_{\min}^2}\sum_{j\geq\xi}\exp\left(-\frac{\Delta_{\min}^2 j^{\theta}}{16C(K/N+3)}\right) \\
&\leq \mathbb{P}(\kappa_{\circ}^n > C^{-1}(\xi-2)^{1+\theta}) + \frac{8K}{\Delta_{\min}^2}\int_{z\geq\xi-1}\exp\left(-\frac{\Delta_{\min}^2 z^{\theta}}{16C(K/N+3)}\right)dz.
\end{aligned}$$

Once again conclusion of the lemma follows by union bounding over $n \in [N]$. $\square$

PROPOSITION 7.10. *Suppose that there exist $C \geq 1$, $\theta > 0$ such that $C^{-1}j^{\theta} \leq A_j - A_{j-1} \leq Cj^{\theta}$ for all $j \in \mathbb{N}$. Then there exists a constant $\phi \equiv \phi(\Delta_{\min}, C, \theta, N, K, p_{\min}, \mathrm{diam}(P))$ depending on $\Delta_{\min}, C, \theta, N, K, p_{\min}, \mathrm{diam}(P)$ but not T such that $\mathbb{E}[A_{\tau}] \leq \phi$.*

PROOF. Given $A_{\hat{\tau}} \geq \zeta \geq C(1+2\underline{j}(\Delta_{\min}))^{1+\theta} \vee C\cdot\{16\mathrm{diam}(P)\}^{1+\theta}$ then $\hat{\tau} \geq (\zeta/C)^{\frac{1}{1+\theta}} - 1$, so $\hat{\tau}_{\mathrm{spr}} \vee \hat{\tau}_{\mathrm{stab}} \geq \{(\zeta/C)^{\frac{1}{1+\theta}} - 1\}/2 \geq \underline{j}(\Delta_{\min})$. Hence, for $\zeta \geq \psi \equiv \psi(\Delta_{\min}, C, \theta) := C(1+2\underline{j}(\Delta_{\min}))^{1+\theta} \vee C\{16\mathrm{diam}(P)\}^{1+\theta}$,

$$\begin{aligned}
\mathbb{P}(A_{\hat{\tau}} \geq \zeta) &\leq \mathbb{P}\left(\hat{\tau}_{\mathrm{spr}} \geq \frac{1}{2}\{(\zeta/C)^{\frac{1}{1+\theta}} - 1\}\right) + \mathbb{P}\left(\hat{\tau}_{\mathrm{stab}} \geq \frac{1}{2}\{(\zeta/C)^{\frac{1}{1+\theta}} - 1\}\right) \\
&\leq N(1 - p_{\min}^{\mathrm{diam}(P)})^{\left\lfloor \frac{(\zeta/C)^{\frac{1}{1+\theta}}}{4\mathrm{diam}(P)} - 2\right\rfloor} + \frac{8KN}{\Delta_{\min}^2}\int_{z\geq(\zeta/C)^{\frac{1}{1+\theta}}/2-2}\exp\left(-\frac{\Delta_{\min}^2 z^{\theta}}{16C(K/N+3)}\right)dz \\
&\quad + \sum_{n\in[N]}\mathbb{P}(\kappa_{\circ}^n > \{(\zeta/C)^{\frac{1}{1+\theta}}/2 - 4\}^{1+\theta}/C) \\
&\leq N(1 - p_{\min}^{\mathrm{diam}(P)})^{\frac{(\zeta/C)^{\frac{1}{1+\theta}}}{2^4\mathrm{diam}(P)}} + \frac{8KN}{\Delta_{\min}^2}\int_{z\geq(\zeta/C)^{\frac{1}{1+\theta}}/2-2}\exp\left(-\frac{\Delta_{\min}^2 z^{\theta}}{16C(K/N+3)}\right)dz \\
&\quad + \sum_{n\in N}\mathbb{P}(\kappa_{\circ}^n > (2^{1+\theta}C)^{-2}\cdot\zeta).
\end{aligned}$$

Note also that by Lemma 7.1 we have

$$\sum_{n\in N}\sum_{\zeta\in\mathbb{N}}\mathbb{P}(\kappa_{\circ}^n > (2^{1+\theta}C)^{-2}\cdot\zeta) = \sum_{n\in N}\sum_{\zeta\in\mathbb{N}}\mathbb{P}((2^{1+\theta}C)^2\kappa_{\circ}^n > \cdot\zeta) = (2^{1+\theta}C)^2\sum_{n\in N}\mathbb{E}[k_{\circ}^n] \leq \frac{8N}{\Delta_{\min}^2}.$$

Hence, we have

$$\begin{aligned}\mathbb{E}[A_{\hat{\tau}}] &\le \psi + \sum_{\zeta>\psi} \mathbb{P}(A_{\hat{\tau}} \ge \zeta) \\ &\le \psi + \sum_{\zeta>\psi} \left\{ N(1-p_{\min}^{\mathrm{diam}(P)})^{\frac{(\zeta/C)^{\frac{1}{1+\theta}}}{2^4 \mathrm{diam}(P)}} + \frac{8K}{\Delta_{\min}^2} \int_{z \ge (\zeta/C)^{\frac{1}{1+\theta}}/2-2} \exp\left(-\frac{\Delta_{\min}^2 z^\theta}{16C(K/N+3)} \right) dz \right\} \\ &\quad + \sum_{n\in[N]} \sum_{\zeta\ge\psi} \mathbb{P}(\kappa_\circ^n > (2^{1+\theta}C)^{-2} \cdot \zeta) =: \phi(\Delta_{\min}, C, \theta, N, K, p_{\min}, \mathrm{diam}(P)) < \infty,\end{aligned}$$

for a finite constant $\phi \equiv \phi(\Delta_{\min}, C, \theta, N, K, p_{\min}, \mathrm{diam}(P))$ depending on $\Delta_{\min}, C, \theta, N, K, p_{\min}, \mathrm{diam}(P)$ but not T. $\square$

We can now complete the proof of the main result.

Proof of Theorem 3.1. The result follows from Corollary 7.5 combined with Proposition 7.10 by taking $\epsilon \to 0$. $\square$