NEARLY OPTIMAL BEST-OF-BOTH-WORLDS INDIVIDUAL REGRET IN COOPERATIVE BANDITS

Anonymous authors
Paper under double-blind review

ABSTRACT

We study the best-of-both-worlds problem of cooperative multi-armed bandits in the uninformed setting, where V agents interact with a common set of K arms over an arbitrary connected communication graph, and each agent initially knows only its local neighborhood. Over a horizon of T rounds, each agent selects an arm, observes only its local loss, and exchanges information with its neighbors. The environment can be either stochastic or adversarial, and the agents do not know the regime in advance. The objective is to simultaneously achieve small individual regret for every agent in both regimes while minimizing communication cost. We propose a single algorithm that achieves individual regret $\widetilde{\mathcal{O}}\big(\sqrt{K} + \min\{\sum_{i:\Delta_i>0} 1/(V\Delta_i), \sqrt{KT/V}\}\big)$ in the stochastic regime, where Δ_i is the suboptimality gap of arm i, and individual regret $\widetilde{\mathcal{O}}\big(\sqrt{KT/V} + K^{1/4}\sqrt{T}\big)$ in the adaptive adversarial regime. We establish matching worst-graph lower bounds up to logarithmic factors for both regimes. Our regret guarantees strictly improve the best known results in the stochastic regime. We further show that our algorithm requires only $\mathcal{O}(\log T)$ communication rounds per agent in the stochastic regime, while every packet has $\mathcal{O}(K \log V)$ bits in both regimes. Finally, we reveal a polynomial separation between adaptive and oblivious adversaries, showing that adaptivity can make cooperative bandit learning polynomially harder.

1 INTRODUCTION

Cooperative multi-armed bandits provide a natural model for sequential decision making among multiple agents that interact with a common set of actions, observe only local bandit feedback, and communicate over a network. This framework captures a wide range of distributed learning problems, including peer-to-peer and Internet of Things (IoT) networks (Szörényi et al., 2013; Chen et al., 2021), wireless channel selection and resource allocation (Avner & Mannor, 2016; Maghsudi & Stanczak, 2015), federated recommendation systems (Shi & Shen, 2021; Liu et al., 2022), and multi-robot or drone teams (Biyik et al., 2022; Wang et al., 2023). By pooling (i.e., aggregating) observations, cooperation can substantially improve individual learning efficiency, but it also incurs communication overhead, especially when agents are geographically distributed or connected only through local links (Madhushani & Leonard, 2021; Agarwal et al., 2022; Newton et al., 2023; Wang et al., 2023). A central goal is therefore to achieve small individual regret while keeping communication efficient.

Depending on how the losses are generated, cooperative bandits have been studied primarily under stochastic and adversarial regimes. In the stochastic model, every agent faces the same loss distribution for each arm, while the realized samples are independently generated across agents and rounds. Recent work obtains individual regret (i.e., regret per agent) over arbitrary connected graphs of order $\widetilde{\mathcal{O}}\big(\min\{\sum_{i:\Delta_i>0} 1/(V\Delta_i), \sqrt{KT/V}\} + K\big)$ (Barnea et al., 2025). The corresponding minimax individual-regret lower bound is $\Omega\big(\sqrt{KT/V} + \sqrt{K}\big)$ (Barnea et al., 2025), leaving a gap between the additive K and $\sqrt{K}$ dependencies. In the adversarial regime, all agents face the same loss vector at each round. Bar-On & Mansour (2019) first established an individual-regret guarantee of $\widetilde{\mathcal{O}}\big(K^2 + \sqrt{(1 + K/|\mathcal{N}(v)|)T}\big)$[1] for every agent v in the uninformed setting, where $\mathcal{N}(v)$ denotes the set containing agent v and its neighbors. Later, Yi & Vojnović (2023) prove a $\Omega\big(\sqrt{(1 + K/|\mathcal{N}(v)|)T}\big)$

[1†] Bar-On & Mansour (2019) assume $T \geq K^2 \log K$, which can be viewed as a burn-in requirement.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

lower bound under a restrictive communication model in which each agent transmits prescribed current-round information to its neighbors, rather than arbitrary aggregated or forwarded messages. Their lower bound therefore does not characterize the minimax regret under *unrestricted local communication*, i.e., communication in which a message may contain any information available to the sender, including information aggregated from earlier rounds, and may be forwarded across multiple hops subject only to the one-edge-per-round propagation constraint.

A recent work by Qiu et al. (2026) bridges the stochastic and adversarial regimes with a single distributed algorithm under a more general heterogeneous-loss model. Their framework allows different agents to have different arm means in the stochastic regime and different loss vectors in the adversarial regime, with performance measured against the global average loss across agents. To compare their guarantees with the cooperative setting considered here, let $\rho = (1 - \sigma_2(W))^{-1/2}$ denote the graph-dependent mixing quantity associated with their prescribed gossip matrix W, where $\sigma_2(W)$ is its second largest singular value. Under homogeneous losses, their best-of-both-worlds guarantee is $\widetilde{\mathcal{O}}\big((\rho/V + \rho^2)\sum_{i:\Delta_i>0} 1/\Delta_i + \rho^2\big)$ in the stochastic regime with a unique optimal arm and $\widetilde{\mathcal{O}}\big(\sqrt{\rho T} + \sqrt{KT/V} + \rho\big)$ in the oblivious adversarial regime. The quantity ρ captures the global mixing difficulty of the communication graph and can be large for poorly connected networks. Despite the greater generality of their loss model, three limitations remain for the setting considered here. (i) Their algorithm is informed and assumes that a common gossip matrix and its spectral information are available in advance, whereas in the *uninformed* setting each agent initially knows only its local neighborhood (Bar-On & Mansour, 2019; Barnea et al., 2025). (ii) Even after specializing their model to homogeneous cooperative bandits, the resulting regret bounds do not close the stochastic and adversarial individual-regret gaps described above. (iii) Their gossip protocol communicates in every round, resulting in $\mathcal{O}(T)$ communication rounds even when the environment is stochastic. Although best-of-both-worlds learning has been studied under a more general heterogeneous model, several fundamental questions remain unresolved even in the simpler homogeneous cooperative setting considered here.

1. *Can a single algorithm simultaneously achieve nearly optimal individual regret in stochastic and adversarial regimes without knowing the regime in advance?*
2. *When the environment is stochastic, can such an algorithm enjoy low communication cost?*
3. *What is the adversarial minimax individual-regret rate under unrestricted local communication?*

Contributions. We address these questions as follows. See Table 1 for an overview and comparison with prior work[2]. Our main contributions are summarized as follows.

- We develop a single algorithm for best-of-both-worlds cooperative bandits over arbitrary connected communication graphs under a general local communication model that allows aggregation and multihop forwarding. The algorithm operates in the uninformed setting, where each agent initially knows only its local neighborhood, and requires neither the loss gaps nor knowledge of whether the environment is stochastic or adversarial.
- We establish nearly optimal individual regret in both stochastic and adaptive adversarial regimes. In the stochastic regime, every agent achieves $\widetilde{\mathcal{O}}\big(\sqrt{K} + \min\{\sum_{i:\Delta_i>0} 1/(V\Delta_i), \sqrt{KT/V}\}\big)$ regret, improving the best known additive dependence from $\widetilde{\mathcal{O}}(K)$ to $\widetilde{\mathcal{O}}(\sqrt{K})$ (Barnea et al., 2025). We also strengthen the lower bound of Barnea et al. (2025) to a matching lower bound over inverse-gap budget classes up to logarithmic factors. In the adaptive adversarial regime, every agent achieves $\widetilde{\mathcal{O}}\big(\sqrt{KT/V} + K^{1/4}\sqrt{T}\big)$ individual regret, and we show a matching worst-graph minimax lower bound. Our algorithm is also communication efficient in the stochastic regime. With high probability, every agent communicates in only $\mathcal{O}(\log T)$ rounds. Every packet contains at most $\mathcal{O}(K \log V)$ bits in both stochastic and adaptive adversarial regimes.
- We reveal a polynomial separation between adaptive and oblivious adversaries. For oblivious adversaries, we obtain $\widetilde{\mathcal{O}}\big(\sqrt{KT/V}+\sqrt{T}+(KT)^{1/3}\big)$ individual regret and establish a worst-graph lower bound of $\Omega\big(\min\{T, \sqrt{KT/V}+\sqrt{T\log K}+\sqrt{K}\}\big)$ under unrestricted local communication. In particular, on a path with $K = V = T$, the oblivious upper bound is $\widetilde{\mathcal{O}}(T^{2/3})$, whereas our

[2]We also present (V, D)-worst-graph results in Appendix B for both upper and lower bounds. In the usual worst-graph case, the hard graph may choose its diameter as part of the worst case, whereas (V, D)-worst-graph results instead take the worst case among graphs with diameter fixed to D.

Table 1: Individual-regret guarantees on connected graphs. Gray rows are our results, and the cap $\min\{T, \cdot\}$ is omitted. Here $R^{\mathrm{sto}} = \min\{\sum_{i:\Delta_i>0} 1/(V\Delta_i), \sqrt{KT/V}\}$ and D is the graph diameter. For Qiu et al. (2026), $\rho = (1-\sigma_2(W))^{-1/2}$. Lower bounds are worst-graph bounds.

Reference	Uninformed	Individual regret	Comm. rounds	Msg. size
Stochastic Losses				
Theorem 4.2	–	$\Omega\big(R^{\mathrm{sto}} + \sqrt{K}\big)$	–	–
Wang et al. (2023)	✗	$\widetilde{\mathcal{O}}\big(R^{\mathrm{sto}} + D\sum_i \Delta_i + K\big)$	$\widetilde{\mathcal{O}}(D)$	$\widetilde{\mathcal{O}}(V)$
Barnea et al. (2025) Coop-SE	✓	$\widetilde{\mathcal{O}}\big(R^{\mathrm{sto}} + K\min\{K, D\}\big)$	$\mathcal{O}(T)$	$\widetilde{\mathcal{O}}(VK)$
Barnea et al. (2025) Coop-SE-CONGEST	✗	$\widetilde{\mathcal{O}}\big(R^{\mathrm{sto}} + K^2\big)$	$\mathcal{O}(T)$	$\widetilde{\mathcal{O}}(1)$
Barnea et al. (2025) Coop-SE-Comm-Cost	✗	$\widetilde{\mathcal{O}}\big(R^{\mathrm{sto}} + K\big)$	$\widetilde{\mathcal{O}}(1)$	$\widetilde{\mathcal{O}}(K)$
Qiu et al. (2026)	✗	$\widetilde{\mathcal{O}}\big((V\rho^2+\rho)R^{\mathrm{sto}} + \rho^2\big)$	$\mathcal{O}(T)$	$\mathcal{O}(K)$
Theorem 4.1	✓	$\widetilde{\mathcal{O}}\big(R^{\mathrm{sto}} + \sqrt{K}\big)$	$\widetilde{\mathcal{O}}(1)$	$\widetilde{\mathcal{O}}(K)$
Adaptive Adversarial Losses				
Theorem 4.2	–	$\Omega\big(\sqrt{KT/V} + K^{1/4}\sqrt{T}\big)$	–	–
Theorem 4.1	✓	$\widetilde{\mathcal{O}}\big(\sqrt{KT/V} + K^{1/4}\sqrt{T}\big)$	$\mathcal{O}(T)$	$\widetilde{\mathcal{O}}(K)$
Oblivious Adversarial Losses				
Bar-On & Mansour (2019)	✓	$\widetilde{\mathcal{O}}\big(K^2 + \sqrt{T} + \sqrt{KT/\lvert\mathcal{N}(v)\rvert}\big)$	$\mathcal{O}(T)$	$\widetilde{\mathcal{O}}(K)$
Qiu et al. (2026)	✗	$\widetilde{\mathcal{O}}\big(\sqrt{\rho T} + \sqrt{KT/V} + \rho\big)$	$\mathcal{O}(T)$	$\mathcal{O}(K)$
Theorem 4.4	–	$\widetilde{\Omega}\big(\sqrt{KT/V} + \sqrt{T} + \sqrt{K}\big)$	–	–
Theorem 4.3	✓	$\widetilde{\mathcal{O}}\big(\sqrt{KT/V} + \sqrt{T} + (KT)^{1/3}\big)$	$\mathcal{O}(T)$	$\widetilde{\mathcal{O}}(K)$

adaptive lower bound is $\Omega(T^{3/4})$. Thus, adversarial adaptivity can make cooperative bandit learning polynomially harder.

Related Work. Cooperative stochastic bandits aim to achieve small regret while keeping communication costs low. The problem has been studied under complete communication (Buccapatnam et al., 2015; Wang et al., 2023; Yang et al., 2026) and over general decentralized networks (Szörényi et al., 2013; Landgren et al., 2016; Kolla et al., 2018; Martínez-Rubio et al., 2019; Sankararaman et al., 2019). On complete graphs, near-optimal individual regret with low communication is now well established (Wang et al., 2023; Yang et al., 2026). On general graphs, consensus-based methods such as Coop-UCB2 (Landgren et al., 2021) and DDUCB (Martínez-Rubio et al., 2019) incur additional graph-dependent regret terms that can be large on poorly connected networks and require communication in every round. For individual regret, Wang et al. (2023) obtained a general-graph extension with an additive dependence on the graph diameter, while Barnea et al. (2025) removed this dependence over arbitrary connected graphs and further studied communication-efficient variants. Their best stochastic upper bound still contains an additive $\widetilde{\mathcal{O}}(K)$ term, compared with an additive $\Omega(\sqrt{K})$ term in the lower bound. Our result closes this gap up to logarithmic factors while requiring only $\mathcal{O}(\log T)$ communication rounds per agent with high probability.

Early work on cooperative adversarial bandits focused on group regret over communication networks (Cesa-Bianchi et al., 2019). For individual regret, Bar-On & Mansour (2019) obtain neighborhood-dependent upper bounds under an oblivious adversary, while Yi & Vojnović (2023) establish nearly matching lower bounds under a restricted communication model, also for oblivious adversaries. These results leave open both the minimax individual-regret rate under a general local communication model and whether adversarial adaptation changes that rate. Our results characterize the adaptive minimax rate and reveal a polynomial separation between adaptive and oblivious adversaries.

Best-of-both-worlds guarantees are well studied for single-agent bandits (Bubeck & Slivkins, 2012; Auer & Chiang, 2016; Zimmert & Seldin, 2021; Ito, 2021; Jin et al., 2023), but most cooperative bandit work treats stochastic and adversarial regimes separately. Recently, Qiu et al. (2026) gave a unified distributed algorithm under a more general heterogeneous-loss model, but their method

requires a prescribed gossip matrix and spectral information, incurs graph-mixing-dependent regret, and communicates in every round. In contrast, we study the uninformed setting and obtain nearly optimal individual regret in both regimes with only $\mathcal{O}(\log T)$ communication rounds per agent under stochastic losses.

2 PRELIMINARIES

We consider a cooperative multi-armed bandit problem over T synchronous rounds with a set of agents $[V] = \{1, \dots, V\}$ and a common action set $[K] = \{1, \dots, K\}$. The agents are connected by a fixed connected undirected graph $G = ([V], E)$, where E denotes the set of communication edges. We work in the *uninformed* setting, where each agent v initially knows only its neighborhood $\mathcal{N}(v) = \{u \in [V] : (u, v) \in E\} \cup \{v\}$, together with K, V, T, and has no knowledge of the global graph topology.[3] In each round t, every agent v independently draws an arm $I_t^v \in [K]$ from a distribution p_t^v and observes only its corresponding loss $X_t^v(I_t^v)$, where $X_t^v \in [0,1]^K$. We consider two regimes depending on how the loss vectors are generated, and the algorithm is not told which regime it is facing. Let $\mathcal{F}_{t-1}$ denote the full interaction history before round t, including all past actions, observations, and randomness.

Adversarial regime. For each round t, there exists an $\mathcal{F}_{t-1}$-measurable vector $\ell_t \in [0,1]^K$ such that $\mathbb{E}[X_t^v(i) \mid \mathcal{F}_{t-1}] = \ell_t(i), \forall v \in [V], \forall i \in [K]$. Conditionally on $\mathcal{F}_{t-1}$, the joint collection of loss vectors is independent of the agents' fresh action randomization. Its distribution may otherwise be arbitrary, including dependence across agents. This generalizes the setting where all agents face the same loss vector (Cesa-Bianchi et al., 2019; Bar-On & Mansour, 2019; Yi & Vojnović, 2023). Our adversarial analysis primarily focuses on the adaptive case. We also study the oblivious case and give the corresponding upper and lower bounds in Theorem 4.3 and Theorem 4.4.

Stochastic regime. The loss vectors are independent across agents and rounds with common fixed means $\mathbb{E}[X_t^v(i)] = \mu_i$ for every $v \in [V]$, $i \in [K]$, and $t \in [T]$. Let $\mu_* = \min_{i \in [K]} \mu_i$, $\Delta_i = \mu_i - \mu_*$.

Communication model. Agents communicate along the edges of G. At the end of each round, each agent v may send a message to and receive a message from every agent $u \in \mathcal{N}(v) \setminus \{v\}$. Each agent may send at most one packet to each neighbor per round, and information received in round t is available before the action in round $t+1$. Thus, a message can traverse at most one edge per round.

We measure communication at the individual agent level using the number of sending rounds and the packet size. A sending round for agent v is a round in which it sends at least one nonempty packet to a neighbor. Simultaneous sends to multiple neighbors count only once, while receiving alone does not count. The packet size is the maximum bit length of a packet transmitted from one agent to one neighbor in one round. Agents continue playing during setup, routing, and other coordination, and all associated communication is counted.

Objective. We seek a single algorithm that achieves small individual regret for every agent in both regimes without knowing the regime in advance. Following the standard pseudo-regret convention in the best of both worlds bandit literature (Bubeck & Slivkins, 2012; Auer & Chiang, 2016), we define the individual regret of agent v as $R^v(T) = \max_{i \in [K]} \mathbb{E}\left[\sum_{t=1}^T \left(X_t^v(I_t^v) - X_t^v(i)\right)\right]$.

Beyond regret, we also seek to reduce the communication cost. In the stochastic regime, our goal is to minimize both the number of sending rounds per agent and the packet size. In the adversarial regime, existing cooperative bandit algorithms typically exchange information in every round (Cesa-Bianchi et al., 2019; Bar-On & Mansour, 2019; Yi & Vojnović, 2023). Accordingly, we focus on minimizing the packet size rather than the number of sending rounds in this regime.

3 ALGORITHM

3.1 OVERVIEW AND HIGH-LEVEL IDEAS

Our algorithm follows the switching paradigm of the classical single-agent best-of-both-worlds algorithms of Bubeck & Slivkins (2012); Auer & Chiang (2016), which begin with a stochastic

[3] We follow the uninformed setting of Bar-On & Mansour (2019) and assume V is known for simplicity of presentation. Appendix A.4 covers the case where only a public upper bound on V is known.

learner and switch to an adversarial learner once the observed losses become inconsistent with stochastic behavior. Although FTRL-based approaches provide a natural alternative, our attempts with such approaches did not yield the same regret guarantees. Our algorithm therefore starts in stochastic mode and switches permanently to an adversarial learner once a discrepancy is detected. As long as the observations remain consistent with the stochastic model, agents pool observations to eliminate suboptimal arms while communicating only a constant number of times per phase. Eliminated arms continue to be tested, since they may later become competitive if the losses turn adversarial. When a test detects a discrepancy, the detecting agent raises an alarm. Each agent forwards the first alarm it observes and switches to the adversarial learner from its next action, without waiting for a graph-wide transition.

In our algorithm, cooperation operates at two levels. The first construction partitions the agents into fixed *initial trees*, whose membership, roots, and tree edges remain unchanged throughout all T rounds. These trees serve both as the basic units from which stochastic groups are formed and as the communication routes retained for adversarial learning. In phase 0, each initial tree is itself a group, whereas later phases may combine several initial trees into larger groups to pool more observations without ever splitting an initial tree. Each group maintains a shared state *(i.e., the common estimates and decisions followed by all of its members)* and a communication tree obtained by joining the fixed initial trees of its members with selected edges between them. This separates the fixed tree structure retained throughout the algorithm from the statistical groups that evolve across phases.

Each group runs phases with two windows of equal length W. The first collects observations for the next estimate. The second pools them, constructs the next groups when needed, and delivers their shared states. Agents continue playing arms and testing losses throughout both windows. Window lengths at least double, so each agent enters only logarithmically many phases. The first phase plays uniformly, and testing starts in phase 1. Algorithm 1 summarizes the procedure. We focus on the case where the initial tree construction does not cover the whole graph and therefore requires attachment. If the initial broadcast already reaches every agent, it directly produces a spanning initial tree, and no further grouping is needed. This tree is used as the single group in every phase. Due to space limitations, further implementation details are deferred to the appendix: the complete tree constructions are given in Appendix A.1, the phase sampling and message encoding and decoding in Appendix A.2, the complete protocol including the adversarial learner in Appendix A.3, and the communication-cost analysis in Appendix C.4.

3.2 Constructing Initial Trees and Groups

Initial trees. The first construction balances the number of observations available locally against the time needed to exchange them. It uses the exponential-shift decomposition of Miller et al. (2013), followed by attachment when necessary. While all agents play uniformly, agent 1 broadcasts its identifier (id) immediately, while every other agent schedules its own id after a random delay. An agent accepts the first arriving id, or its own scheduled release if that comes first, records the delivering neighbor as its parent, and forwards the id once. The resulting parent edges partition the agents into rooted trees. Each tree then aggregates its number of agents to the root. If one tree already contains all V agents, the construction ends with that spanning tree. Otherwise, trees with sufficiently many agents are retained, while agents in the remaining smaller trees attach toward the retained ones. We call the trees obtained after this attachment step the *initial trees*, distinguishing them from the temporary trees formed before attachment.

Let Q denote the set of agents in one initial tree, so that the sets Q partition $[V]$. Let $r_0 = \min\{\lceil 2\sqrt{K}\rceil, V\}$, $\Gamma = \Theta(\log(KVT/\delta))$, and $K_0 = \Theta(\Gamma)$, where $\delta \in (0,1)$ controls confidence. With high probability, every initial tree has at least $r = \lceil r_0/K_0\rceil$ agents and height $\widetilde{\mathcal{O}}(r_0)$. Each initial tree remains unchanged thereafter, and later phases only regroup whole initial trees.

Phase groups. For phase $s \geq 0$, the cooperation scale is $r_s = \min\{2^s\lceil 2\sqrt{K}\rceil, V\}$. Larger scales require groups with more observation sources and allow longer communication paths. In each reconstruction, the algorithm treats every initial tree as one super-node, weighted by its number of agents. Two such nodes are adjacent when an original communication edge joins their members. The first-arrival construction runs on this graph at scale $\bar{r}_s = \lceil r_s/r\rceil$. Tentative groups with at least r_s/K_0 agents are retained. Each initial tree outside them attaches as a whole to a nearby retained

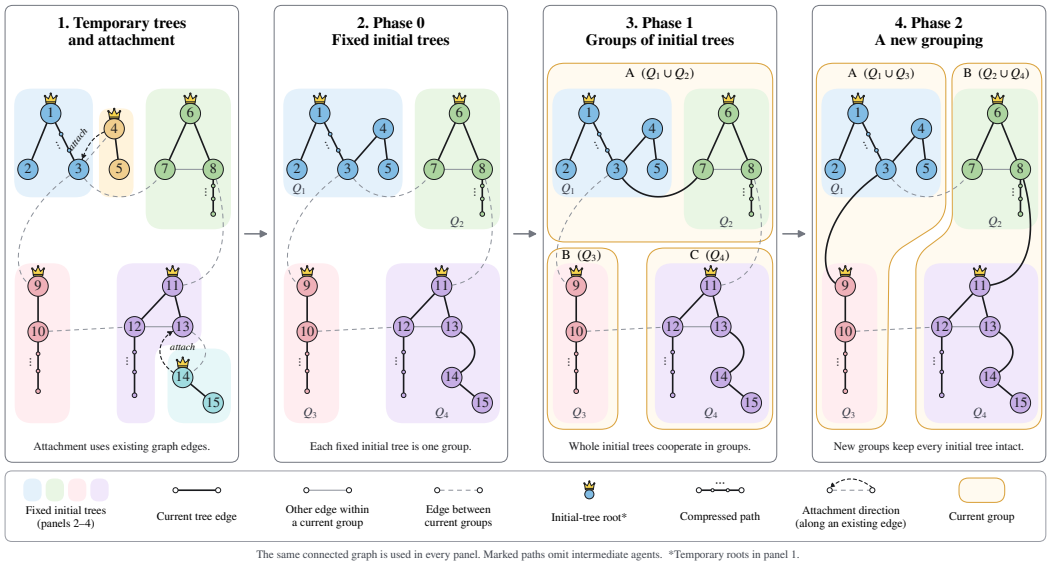


Figure 1: Illustration of the cooperation structure. Smaller trees first attach to retained trees to form the fixed initial trees. Each initial tree then forms one group in the first phase (i.e., phase 0). In later phases, whole initial trees are regrouped at increasing scales. The membership and internal communication structure of each initial tree remain fixed, while the groups may change across phases and need not be nested. All panels depict the same underlying connected graph, with intermediate agents along selected communication paths omitted for clarity.

group. The resulting group tree consists of the fixed initial-tree edges and one connecting edge for each edge of the selected tree between initial trees.

The reduced scale $\bar{r}_s$, rather than r_s, accounts for the fact that each initial tree already contributes at least r agents. With high probability, every group has at least r_s/K_0 members and tree height $\widetilde{\mathcal{O}}(r_s)$. Each construction still uses only $\mathcal{O}(1)$ sending rounds per agent. Lemma C.2 proves these claims and shows that the existing window lengths accommodate the extra logarithmic routing factors. New groups are prepared during the preceding second window until $r_s = V$, after which they remain fixed. A group may contain initial trees from several previous groups, so successive group memberships need not be nested. However, no initial tree is ever split or reconstructed. The current group's parent pointers are stored separately from the fixed rooted trees used for fallback.

3.3 Cooperative State Updates within a Phase

Shared state. All members of group g coordinate their learning in phase s using

$$(\widehat{\mu}_{s,g}, \lambda_{s,g}, \tau_s, W_{s,g}), \qquad \lambda_{s,g} = (\lambda_{s,g,1}, \ldots, \lambda_{s,g,K}), \qquad \tau_s = 2^{-s-1},$$

where $\widehat{\mu}_{s,g}$ estimates the best mean loss, $\lambda_{s,g,i}$ stores gap information for arm i, τ_s is the estimation precision, and $W_{s,g}$ is the length of each window. A candidate arm has stored gap zero, so $S_{s,g} = \{i \in [K] : \lambda_{s,g,i} = 0\}$. On elimination, its stored gap is roughly one quarter of the estimated gap (i.e., $\lambda_{s,g,i} \approx \Delta_i/4$ in the stochastic case, see Lemma 3.1). The initial state is $(0, \mathbf{0}, 1/2, W_{0,g})$, with all arms as candidates. The state stays fixed within the phase, while local statistics change with observations. To avoid clutter, the following phasewise formulas suppress (s, g).

Sampling and test assignment. Agent v reserves mass $a_i^v(t)$ for testing arm i and samples from

$$p_t^v(i) = a_i^v(t) + \left(1 - \sum_j a_j^v(t)\right) \frac{\mathbf{1}\{i \in S\}}{|S|}, \qquad i \in [K]. \tag{1}$$

Algorithm 1: Cooperative best-of-both-worlds algorithm for agent v

Input $K, V, T, \delta, \mathcal{N}(v)$
Agent v plays uniformly while its fixed initial tree is constructed and retained
Set $s \leftarrow 0$, with g denoting the initial group $Q(v)$, $(\widehat{\mu}, \lambda, \tau, W) \leftarrow (0, \mathbf{0}, 1/2, W_{0,g})$ by Appendix A.1
for *each phase* (s, g) *before an alarm or round* T **do**
  Set $S \leftarrow \{i : \lambda_i = 0\}$, $Y_i^v \leftarrow 0$ for $i \in S$, $a_i^v \leftarrow 0$ for $i \in [K]$
  if $s \geq 1$ **then**
    Set $\xi_1^v \sim \mathrm{Unif}[0, 1)$, $\xi_i^v \leftarrow (\xi_1^v + (i-1)/K) \bmod 1$
    Set (a_i^v, b_i^v) by Appendix A.2, and $Z_i^v \leftarrow \widehat{\mu} + 4\lambda_i$ for assigned arms
  for $t = t_{s,g}, \ldots, \min\{t_{s,g} + 2W - 1, T\}$ **do**
    Play arm $I_t^v \sim p_t^v$ by Eq. (1), with observed loss $X_t^v(I_t^v)$
    if $a_{I_t^v}^v(t) > 0$ **then** Eq. (6) and Eq. (7) apply
    if $t < t_{s,g} + W$ and $I_t^v \in S$ **then** $Y_{I_t^v}^v$ changes by Eq. (3)
    During the second window, the group pools sums and prepares its successor by the procedure specified in Appendix A.2
    The decreasing rates follow Eq. (2)
  if $t_{s,g} + 2W \leq T$ **then** the prepared next state takes effect and $s \leftarrow s + 1$
on the first alarm stochastic work stops and the alarm is forwarded. From the next action, Algorithm 3 runs if the retained tree is ready, and otherwise play is uniform

The tests check whether the stored estimates still explain the losses. Having every agent repeatedly test every eliminated arm would forgo the cooperative reduction in exploration cost. Instead, agents share the testing burden, and a detected violation is communicated by an alarm. For each phase with $s \geq 1$, agent v independently draws ξ_1^v uniformly from $[0, 1)$ and sets $\xi_i^v = (\xi_1^v + (i-1)/K) \bmod 1$. Equal spacing limits the number of arms assigned to one agent, while independent shifts across agents provide nearby testers. The grids are drawn after the incoming state and group are fixed and remain unchanged only within that phase. Fixing the initial trees does not make these grids permanent.

Let M be the current group population. The fixed population floor r also determines the density of assigned testers, while H_i determines which of them use a decreasing rate,

$$r = \lceil r_0/K_0 \rceil, \qquad H_i = 1 + \frac{2M}{1 + M\lambda_i/\sqrt{K}}.$$

Agent v tests arm i when $r\xi_i^v \leq 48\Gamma$. Assigned testers with $H_i\xi_i^v \leq 48\Gamma$ use a decreasing rate, while the others use a constant rate. The former limits cumulative testing cost, while the latter maintains local checks as alarms propagate. All rates and statistics reset at the start of each phase. A decreasing-rate test uses

$$a_i^v(t_{s,g}) = \frac{1}{4096\Gamma^2(1 + K/H_i)}, \qquad \frac{1}{a_i^v(t+1)} = \frac{1}{a_i^v(t)} + \frac{\lambda_i^2}{c_1\Gamma^2}, \tag{2}$$

where $t_{s,g}$ is the first round of the phase and $c_1 \geq 2^{18}$. The reciprocal update is applied after every round in both windows, regardless of whether arm i is played, so a larger λ_i makes the rate decrease faster. Candidates have $\lambda_i = 0$ and therefore keep a fixed rate. Constant-rate tests use $a_i^v(t) = \gamma_i$ as specified in Appendix A.2, while unassigned arms have rate zero. Since the total testing mass is at most $1/16$, each candidate receives probability at least $15/(16|S|)$. Phase 0 sets all testing rates to zero and $S = [K]$, giving uniform play.

Pooling first-window observations. For each candidate $i \in S$, agent v initializes $Y_i^v = 0$. Whenever it plays arm i in the first window, it updates

$$Y_i^v \leftarrow Y_i^v + \frac{X_t^v(i) - \widehat{\mu}}{p_t^v(i)}. \tag{3}$$

The importance weight corrects for unequal sampling probabilities. Subtracting $\widehat{\mu}$ centers the expected statistic around $\mu_i - \widehat{\mu} = \mathcal{O}(\tau)$ for stochastic candidate arms. Without this centering, the encoded

values would grow as the precision decreases, thereby increasing message size. In the second window, the sums are added along the current group tree toward its root. The root broadcasts the total, so every member obtains the same estimate $\widetilde{\mu}_i$. Every member of every constituent initial tree contributes exactly once. The integer encoding and message checks are in Appendix A.2.

Preparing the next phase. During the second window, the same fixed initial trees are re-partitioned into groups at scale r_{s+1} when reconstruction is still needed. A new group g inherits one complete updated state from the previous group g' containing its new root. It does not average states from different previous groups. When the group is unchanged, $g' = g$. Previously positive stored gaps are retained by the update, and the remaining coordinates satisfy

$$\begin{aligned} \widehat{\mu}_{s+1,g} &= \min_{i \in S_{s,g'}} \widetilde{\mu}_{s,g',i}, \qquad S_{s+1,g} = \{i \in S_{s,g'} : \widetilde{\mu}_{s,g',i} - \widehat{\mu}_{s+1,g} \le 4\tau_s\}, \\ \lambda_{s+1,g,i} &= (\widetilde{\mu}_{s,g',i} - \widehat{\mu}_{s+1,g})/4, \qquad i \in S_{s,g'} \setminus S_{s+1,g}. \end{aligned} \tag{4}$$

The window lengths follow the unchanged recurrence

$$W_{0,g} = \lceil \Lambda r_0 \rceil, \qquad W_{s+1,g} = 2W_{s,g'} + \begin{cases} \left\lceil \Lambda \left(V + \dfrac{|S_{s+1,g}|}{V\tau_{s+1}^2} \right) \right\rceil, & r_{s+1} = V, \\ 0, & r_{s+1} < V, \end{cases} \tag{5}$$

where $\Lambda = \Theta(\Gamma^6)$. The first transition to scale V includes the fixed-group increment. The routing term covers all passes on the expanded group tree, while the sampling term provides enough observations for the next precision. The next state takes effect before round $t_{s+1,g} = t_{s,g'} + 2W_{s,g'}$. Absolute state values are sent only when the group tree changes. Once groups are fixed, members reconstruct later updates from the previous state and the same pooled integers.

3.4 Local Tests and Mode Switch

Local test. For each assigned arm i, agent v initializes a moving average $Z_i^v = \widehat{\mu} + 4\lambda_i$. Whenever it plays that arm, it updates

$$Z_i^v \leftarrow Z_i^v + \frac{(b_i^v)^2}{64\Gamma + (b_i^v)^2} \left(X_t^v(i) - Z_i^v \right). \tag{6}$$

The fixed phase tolerance b_i^v, specified in Eq. (13), determines both how quickly the average responds and which discrepancies are acceptable. The agent raises an alarm when

$$|Z_i^v - (\widehat{\mu} + 4\lambda_i)| > 3b_i^v/2. \tag{7}$$

The moving average of an assigned arm changes only when agent v plays that arm. The same test runs in both windows. The next lemma explains why estimation and testing remain consistent under stochastic losses.

Lemma 3.1 (Stochastic Behavior)**.** *In the stochastic regime, with high probability, simultaneously for every entered phase (s, g) with $s \ge 1$, the set $S_{s,g}$ contains every optimal arm and*

$$|\widehat{\mu}_{s,g} - \mu_*| \le \tau_s/256, \qquad \Delta_i < 10\tau_s \quad (i \in S_{s,g}), \qquad \Delta_i/5 \le \lambda_{s,g,i} \le \Delta_i/3 \quad (i \notin S_{s,g}).$$

Every assigned test passes, and no loss, range, or required-record alarm occurs.

Thus the algorithm stays in stochastic mode on this event. Appendix C proves the simultaneous statement using the new fixed-tree grouping lemma and the unchanged estimation and testing calculations.

Mode switch. On raising or receiving its first alarm, each agent stops stochastic learning permanently, forwards the alarm once to every graph neighbor, and starts the adversarial learner on its retained initial tree from its next action. Alarm forwarding uses all original graph edges and does not wait for scheduled group messages. Members can activate at different times, so no agent waits for a graph-wide switch. Later grouping messages cannot restart stochastic work. The fallback protocol and its conditional regret guarantee are in Appendix A.3.

4 Main Results

We first state the worst-graph guarantees of our best-of-both-worlds algorithm. We also consider (V, D)-worst-graph bounds in Appendix B, where the diameter is fixed rather than allowed to vary.

Theorem 4.1 (Best-of-Both-Worlds Regret and Communication Cost). *For every connected $G = ([V], E)$, $K \geq 2$, $V, T \geq 1$, and $\delta \in (0, 1)$, Algorithm 1 satisfies, for every agent v,*

$$R^v(T) \leq \widetilde{\mathcal{O}}\left(\sqrt{K} + \min\left\{\sum_{i:\Delta_i>0} \frac{1}{V\Delta_i}, \sqrt{\frac{KT}{V}}\right\}\right) \qquad \textit{under stochastic losses},$$

$$R^v(T) \leq \widetilde{\mathcal{O}}\left(\sqrt{\frac{KT}{V}} + K^{1/4}\sqrt{T}\right) \qquad \textit{under adaptive adversarial losses}.$$

Under stochastic losses, all agents simultaneously send in $\mathcal{O}(\log T)$ rounds with probability at least $1 - \delta$, and each agent has the same expected bound. Every packet has $\mathcal{O}(K \log V)$ bits on every execution in both regimes.

The stochastic guarantee improves the additive arm dependence from $\widetilde{\mathcal{O}}(K)$ in Barnea et al. (2025) to $\widetilde{\mathcal{O}}(\sqrt{K})$, and retains the $1/V$ speedup in leading term. The lower bounds below allow *unrestricted local communication*. Policies may know the entire graph and target initially and communicate every round using arbitrary packet contents and sizes, including aggregation and multihop forwarding. Information still travels at most one graph edge per round.

Theorem 4.2 (Stochastic and Adaptive Lower Bounds). *For every $K \geq 2$ and $V, T \geq 1$, the following hold under unrestricted local communication.*

(i) Stochastic losses. *For every finite $H \geq K - 1$, every cooperative policy on the V-vertex path, and every fixed agent v, there is a stochastic instance with a unique optimal arm and $H_\mu = \sum_{i:\Delta_i>0} \Delta_i^{-1} \leq H$ such that $R^v(T) = \Omega\Big(\min\big\{T, \sqrt{K} + \min\big\{H/V, \sqrt{KT/V}\big\}\big\}\Big)$.*

(ii) Adaptive adversarial losses. *For every cooperative policy on the V-vertex path and every fixed agent v, there is a deterministic adaptive common binary loss process with $R^v(T) = \Omega\Big(\min\big\{T, \sqrt{KT/V} + K^{1/4}\sqrt{T}\big\}\Big)$.*

The stochastic assertion applies to every inverse-gap budget class. The adaptive assertion is a worst-graph minimax lower bound without a diameter restriction. Together they match the bound given in Theorem 4.1 up to logarithmic factors in these respective senses. Their proofs are given in Appendix D.1 and Appendix D.2.

For oblivious losses, the loss process is chosen independently of the agents' random choices, with shared conditional means $\ell_1, \ldots, \ell_T$ fixed before interaction. The realized losses may still differ across agents, while the conditional-mean and fresh-action assumptions in Section 2 remain unchanged. It is noteworthy that the following lower bounds hold under the common-loss setting in which all agents face the same loss vector in every round, so the hardness does not arise from cross-agent differences in realized losses.

Theorem 4.3 (Oblivious Adversarial Upper Bound). *For every connected $G = ([V], E)$, $K \geq 2$, and $V, T \geq 1$, the configuration in Appendix E.1 satisfies, for every agent v, $R^v(T) = \widetilde{\mathcal{O}}\big(\sqrt{KT/V} + \sqrt{T} + (KT)^{1/3}\big)$. Every packet has at most $\mathcal{O}(K \log V)$ bits.*

Theorem 4.4 (Oblivious Adversarial Lower Bound). *For every $K \geq 2$ and $V, T \geq 1$, there are a fixed connected graph on $[V]$ and a fixed target v such that every cooperative policy under unrestricted local communication admits a deterministic common binary oblivious loss table with $R^v(T) = \Omega\big(\min\big\{T, \sqrt{KT/V} + \sqrt{T \log K} + \sqrt{K}\big\}\big)$.*

While the oblivious upper and lower bounds do not match, they nevertheless reveal a polynomial separation from adaptive adversaries under the same unrestricted local communication model. On a path with $K = V = T$, Theorem 4.3 gives $\widetilde{\mathcal{O}}(T^{2/3})$, whereas the adaptive lower bound in Theorem 4.2 is $\Omega(T^{3/4})$. The oblivious proofs appear in Appendix E.

Reproducibility statement

The problem setting and assumptions are stated in Section 2. Section 3 and Appendix A specify the algorithms, parameter choices, message formats, and communication schedules. The stochastic and adaptive upper bounds and communication guarantees are proved in Appendix C, the corresponding lower bounds in Appendix D, and the oblivious-adversary results in Appendix E. Appendix B states the fixed-diameter refinements. This theoretical work does not rely on datasets or empirical experiments.

AI use statement

In this work, we used generative AI tools to help develop theoretical models and algorithmic frameworks, formulate mathematical claims, propose and refine hypotheses, provide ingredients for proofs, assist in proof writing, give feedback on research methodology, and interpret theoretical results. We have not used generative AI tools to implement or run the proposed learning methods in empirical experiments, or to translate the manuscript. Synthetic dataset generation, dataset cleaning and reformatting, and qualitative or thematic data analysis are not applicable to this theoretical work. Additionally, we used generative AI tools to draft and edit text, summarize and identify relevant literature, create or modify scientific figures, format references, and create verification code and manuscript artifacts. We have reviewed the AI-assisted work by checking derivations against the stated assumptions, comparing related-work claims with the cited sources, and running finite checks of selected inequalities and protocol rules. These checks supplement, rather than replace, the mathematical arguments. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

References

Mridul Agarwal, Vaneet Aggarwal, and Kamyar Azizzadenesheli. Multi-agent multi-armed bandits with limited communication. *Journal of Machine Learning Research*, 23(212):1–24, 2022.

Peter Auer and Chao-Kai Chiang. An algorithm with nearly optimal pseudo-regret for both stochastic and adversarial bandits. In *29th Annual Conference on Learning Theory*, 2016.

Orly Avner and Shie Mannor. Multi-user lax communications: A multi-armed bandit approach. In *IEEE INFOCOM 2016*, 2016.

Yogev Bar-On and Yishay Mansour. Individual regret in cooperative nonstochastic multi-armed bandits. In *Advances in Neural Information Processing Systems*, 2019.

Idan Barnea, Tal Lancewicki, and Yishay Mansour. Individual regret in cooperative stochastic multi-armed bandits. In *Advances in Neural Information Processing Systems*, 2025.

Erdem Biyik, Anusha Lalitha, Rajarshi Saha, Andrea Goldsmith, and Dorsa Sadigh. Partner-aware algorithms in decentralized cooperative bandit teams. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022.

Sébastien Bubeck and Aleksandrs Slivkins. The best of both worlds: Stochastic and adversarial bandits. In *Proceedings of the 25th Annual Conference on Learning Theory*, 2012.

Swapna Buccapatnam, Jian Tan, and Li Zhang. Information sharing in distributed stochastic bandits. In *IEEE INFOCOM 2015*, 2015.

Nicolò Cesa-Bianchi, Claudio Gentile, and Yishay Mansour. Delay and cooperation in nonstochastic bandits. *Journal of Machine Learning Research*, 20(17):1–38, 2019.

Shuzhen Chen, Youming Tao, Dongxiao Yu, Feng Li, and Bei Gong. Distributed learning dynamics of multi-armed bandits for edge intelligence. *Journal of Systems Architecture*, 114:101919, 2021.

David A. Freedman. On tail probabilities for martingales. *The Annals of Probability*, 3(1):100–118, 1975.

Shinji Ito. Parameter-free multi-armed bandit algorithms with hybrid data-dependent regret bounds. In *Proceedings of the Thirty Fourth Conference on Learning Theory*, 2021.

Tiancheng Jin, Junyan Liu, and Haipeng Luo. Improved best-of-both-worlds guarantees for multi-armed bandits: FTRL with general regularizers and multiple optimal arms. In *Advances in Neural Information Processing Systems*, 2023.

Ravi Kumar Kolla, Krishna Jagannathan, and Aditya Gopalan. Collaborative learning of stochastic bandits over a social network. *IEEE/ACM Transactions on Networking*, 26(4):1782–1795, 2018.

Peter Landgren, Vaibhav Srivastava, and Naomi Ehrich Leonard. On distributed cooperative decision-making in multiarmed bandits. In *2016 European Control Conference*, 2016.

Peter Landgren, Vaibhav Srivastava, and Naomi Ehrich Leonard. Distributed cooperative decision making in multi-agent multi-armed bandits. *Automatica*, 125:109445, 2021.

Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.

Zengyan Liu, Linqi Song, and Christina Fragouli. Federated multi-armed bandits with vector rewards for aspect-based recommendations. In *IEEE Global Communications Conference*, 2022.

Udari Madhushani and Naomi Ehrich Leonard. When to call your neighbor? strategic communication in cooperative stochastic bandits. *CoRR*, abs/2110.04396, 2021.

Setareh Maghsudi and Slawomir Stanczak. Joint channel selection and power control in infrastructureless wireless networks: A multiplayer multiarmed bandit framework. *IEEE Transactions on Vehicular Technology*, 64(10):4565–4578, 2015.

David Martínez-Rubio, Varun Kanade, and Patrick Rebeschini. Decentralized cooperative stochastic bandits. In *Advances in Neural Information Processing Systems*, 2019.

Gary L. Miller, Richard Peng, and Shen Chen Xu. Parallel graph decompositions using random shifts. In *Proceedings of the 25th ACM Symposium on Parallelism in Algorithms and Architectures*, 2013.

Conor J. Newton, Ayalvadi Ganesh, and Henry W. J. Reeve. Asymptotic optimality for decentralised bandits. *Dynamic Games and Applications*, 13:307–325, 2023.

Hao Qiu, Mengxiao Zhang, and Nicolò Cesa-Bianchi. Near-optimal regret for distributed adversarial bandits: A black-box approach. In *Proceedings of the Thirty-Ninth Conference on Learning Theory*, 2026.

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.

Chengshuai Shi and Cong Shen. Federated multi-armed bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021.

Balázs Szörényi, Róbert Busa-Fekete, István Hegedüs, Róbert Ormándi, Márk Jelasity, and Balázs Kégl. Gossip-based distributed stochastic bandit algorithms. In *Proceedings of the 30th International Conference on Machine Learning*, 2013.

Joel A. Tropp. Freedman's inequality for matrix martingales. *Electronic Communications in Probability*, 16:262–270, 2011.

Xuchuang Wang, Lin Yang, Yu-Zhen Janice Chen, Xutong Liu, Mohammad Hajiesmaili, Don Towsley, and John C. S. Lui. Achieving near-optimal individual regret & low communications in multi-agent bandits. In *International Conference on Learning Representations*, 2023.

Lin Yang, Xuchuang Wang, Haoxu Chen, Mohammad H. Hajiesmaili, Lijun Zhang, John C. S. Lui, and Don Towsley. Cooperative bandit algorithms with optimal regret and communication costs. *IEEE/ACM Transactions on Networking*, 34:3609–3624, 2026.

Jialin Yi and Milan Vojnović. On regret-optimal cooperative nonstochastic multi-armed bandits. In *International Conference on Autonomous Agents and Multiagent Systems*, 2023.

Julian Zimmert and Yevgeny Seldin. Tsallis-INF: An optimal algorithm for stochastic and adversarial bandits. *Journal of Machine Learning Research*, 22(28):1–49, 2021.

Appendix

A Implementation Details 13
A.1 Initial Trees and Phase Groups . . . 13
A.2 Phase Sampling and Messages . . . 16
A.3 Complete Algorithms . . . 17
A.4 A Known Population Upper Bound . . . 19

B Worst-Graph Bounds at a Fixed Diameter 21

C Proof of Individual Regret in Theorem 4.1 22
C.1 Supporting Lemmas . . . 22
C.2 Analysis of Individual Regret under Stochastic Losses . . . 31
C.3 Analysis of Individual Regret under Adaptive Losses . . . 31
C.4 Communication Cost Analysis . . . 38

D Stochastic and Adaptive Lower Bounds 39
D.1 Stochastic Individual-Regret Lower Bound . . . 39
D.2 Adaptive Adversarial Individual-Regret Lower Bound . . . 42

E Oblivious Adversaries 47
E.1 Upper Bound . . . 47
E.2 The Lower Bound . . . 51

A Implementation Details

Table 2: Notation used in the algorithm and its analysis. Phase indices are omitted only while discussing one fixed phase.

Symbol	Meaning
K, V, T	Numbers of arms, agents, and interaction rounds
$G, \mathcal{N}(v), D$	Communication graph, closed neighborhood, and graph diameter
$X_t^v(i), \ell_t(i)$	Agent loss and its shared conditional mean
μ_i, Δ_i, H_μ	Stochastic mean, gap, and $\sum_{i:\Delta_i>0} 1/\Delta_i$
R^{sto}	$\min\{\sum_{i:\Delta_i>0} 1/(V\Delta_i), \sqrt{KT/V}\}$
$(s,g), M_{s,g}, t_{s,g}$	Phase and group, group population, and phase start
$\tau_s, W_{s,g}, r_s$	Phase precision, window length, and grouping scale
$Q(v), m_Q$	Fixed initial-tree member set and its population
$h_*, J, \bar{r}_s$	Initial-tree height bound, rounds per quotient step, and $\lceil r_s/r \rceil$
$\widehat{\mu}_{s,g}, \lambda_{s,g,i}, S_{s,g}$	Shared reference, stored quarter-gap, and candidate set
$\xi_i^v, a_i^v(t), b_i^v, Z_i^v$	Fresh grid, testing rate, tolerance, and moving average
$Y_i^v, \widetilde{\mu}_i$	First-window residual and decoded group estimate
Γ, Λ, K_0	Logarithmic confidence factor, window factor, and population factor
$r, H_i, B(x)$	Nearby testing radius, decreasing-test scale, and loss allowance
m, h, q	Retained-tree population, height, and block length $2h+3$
$w_b, U_{b,i}^v, C_{b,i}^v, G_b(i), \eta$	Common block distribution, retained zero and one counts, pooled loss estimate, and learning rate

Asymptotic regret statements written as $R = \mathcal{O}(f)$, $R = \widetilde{\mathcal{O}}(f)$, or $R = \Omega(f)$ denote one-sided upper or lower bounds. Explicit finite inequalities retain their usual meaning. Asymptotic regret rates are capped by T. All time indices refer to the interaction rounds, including setup and communication. The common horizon T is retained after group construction. For $T = 1$ or $K = 1$, the stated uniform action rule requires no communication. A singleton graph transmits no packets. The bounds involving $\log T$ and $\log V$ therefore concern $T \geq 2$ and $V \geq 2$, respectively. Otherwise fix

$$\Gamma = \log\left(\frac{2^{10}KVT^8}{\delta}\right), \qquad \varepsilon = \frac{\delta}{T^2}, \qquad \Lambda = c_2\Gamma^6, \qquad K_0 = \lceil c_3\Gamma \rceil.$$

Here $c_1 \geq 2^{18}$, $c_7 \geq 2^{24}$, and $c_3 \geq 64$ is sufficiently large for the partition tails. The constants are chosen in order, with c_6 sufficiently large for the estimation error below, then c_8 for the integer ranges, $c_5 \geq \max\{2^{20}, 128\sqrt{c_1}\}$, and finally c_2 sufficiently large for both statistical accuracy and all scheduled tree passes, including the fixed-tree quotient construction in Appendix A.1. Each is one universal numerical constant, independent of the instance. The requirements on c_6, c_8 do not depend on the final choice of c_2.

A.1 Initial Trees and Phase Groups

The first construction uses scale $r_0 \geq 1$ and confidence factor Γ. For standalone adaptive learning, $r_0 = \lceil 2\sqrt{K} \rceil$ and $\Gamma = \log(2^{11}KVT^8)$. The oblivious configuration uses $r_0 = \lceil (KT)^{1/3} \rceil$ with the same Γ. For best-of-both-worlds learning, the scales are $d_0 = \lceil 2\sqrt{K} \rceil$ and $r_s = \min\{d_0 2^s, V\}$, with the confidence factor above. The first construction uses scale r_0. In all three applications, $K_0 = \lceil c_3\Gamma \rceil$. The same initial construction serves all three configurations.

Put $R = \lceil c_3 r_0 \Gamma \rceil$ and $L = 8r_0 + R$. Each vertex independently draws $E_v \sim \mathrm{Exp}(1/r_0)$ and sets $\chi_v = \min\{R, \lfloor E_v \rfloor\}$. Vertex 1 sends its own label at the end of round 1. Every other vertex schedules its release at the end of round $1 + L - \chi_v$. A message received in one round is processed before the next round's action and can be forwarded only at its end. An unassigned vertex accepts the first arriving label or its own release, breaks simultaneous ties by root index, records its parent, and broadcasts its label, parent, and depth once. It suppresses later arrivals and its pending release. Thus one first-arrival wave both seeks a large group and supplies the smaller groups if the graph cannot be covered quickly.

A vertex waits for a declaration from every known neighbor. These declarations identify its children. After its children report, it sends its subtree population and maximum depth. A root broadcasts the

resulting population and height. A count equal to V certifies that this tree contains the entire graph. Its root also announces a common learning start at its report round plus $2h + 4$, where h is its height. All such announcements finish before the common slot $8(L + 2)$. No prior knowledge of paths, distances, or other neighborhoods is used.

If a full-population tree is certified, its members retain that tree and are ready at the announced round. The best-of-both-worlds algorithm begins the first phase then. Standalone adversarial learning begins at the first public block boundary at or after that round. Otherwise, every temporary tree with population at least r_0/K_0 is retained at slot $8(L + 2)$. All its vertices become seeds carrying its label. A second first-arrival wave, restricted to vertices outside these trees, attaches them toward the seeds for r_0 hops. A final parent exchange, census, and broadcast give the completed group metadata. An uncovered vertex or a final group below the required population fails this construction. In the best-of-both-worlds algorithm, a failed construction or missing required record raises an alarm. Standalone adversarial learners instead continue uniform play if construction fails. No membership lists are collected. At scale $r_0 \le V$, Lemma C.1 bounds the failure probability. If $r_0 > V$, the first census spans the graph deterministically by Lemma C.17, so this lemma is unnecessary. The routine uses the exponential-shift construction of Miller et al. (2013) followed by attachment to populated groups.

A nonspanning first construction is allocated $64(L + r_0 + 1)$ rounds. Its groups are ready at that deadline. The best-of-both-worlds algorithm starts the first phase there. Standalone adversarial learning starts on each tree at the next public block boundary, adding fewer than $q = 2h + 3$ rounds. A vertex without a required record interprets silence only at its deadline. Every vertex sends only a constant number of times during the construction. Counts, root and parent indices, and actual depths are bounded by V. A full-population start is also bounded by $\mathcal{O}(V)$, as shown in Lemma C.17. The initial actions are uniform, and every routine stops at round T.

Fixed Initial Trees For the best-of-both-worlds algorithm, let $\mathcal{Q}$ be the partition produced by the completed first construction, including its attachment step. Write $Q(v)$ for the part containing v and $m_Q = |Q|$. Phase 0 uses each Q as one group. Each agent permanently retains its initial tree, root, and original parent and child records. Subsequent phases never split a set Q or replace its tree edges. These are the *fixed initial trees* of Section 3.2. The temporary first-arrival clusters preceding attachment are not the fixed units.

If the first census certified a spanning tree before attachment, that group follows Eq. (10) throughout. Otherwise, define the public bounds

$$r = \lceil r_0/K_0 \rceil, \qquad h_* = 9r_0 + \lceil c_3 r_0 \Gamma \rceil, \qquad J = 8(h_* + 1). \tag{8}$$

Every valid completed initial tree has $m_Q \ge r$ and height at most h_*. The first construction already checks the population requirement. In particular,

$$\frac{J}{r} = \mathcal{O}(\Gamma^2),$$

since $r_0/r \le K_0 = \mathcal{O}(\Gamma)$. This ratio is why initial trees can be kept intact without an additional polynomial routing cost.

One exchange of the fixed initial-tree root ids across original graph edges identifies boundary neighbors. Define the quotient graph $G_{\mathcal{Q}}$ by treating each $Q \in \mathcal{Q}$ as a vertex of weight m_Q. Two quotient vertices are adjacent exactly when an original graph edge connects their members. No agent collects the whole quotient graph or a membership list. It needs only its retained tree records and the ids received from its original neighbors.

Regrouping the Fixed Initial Trees Only the best-of-both-worlds algorithm performs this step. For a new scale r_s with $s \ge 1$, put

$$\bar{r}_s = \left\lceil \frac{r_s}{r} \right\rceil, \qquad R_s = \lceil c_3 \bar{r}_s \Gamma \rceil. \tag{9}$$

Each communication step of this construction occupies J consecutive rounds of the original interaction. Each fixed root independently draws $E_Q \sim \mathrm{Exp}(1/\bar{r}_s)$ and sets $\chi_Q = \min\{R_s, \lfloor E_Q \rfloor\}$. Its id is scheduled at step $R_s - \chi_Q$, with no priority for agent 1. An unassigned initial tree accepts the first

candidate reaching any member, or its scheduled own release, and all its members accept the same choice. Ties use the root id followed by a fixed order on the realizing boundary edges. Earliest here refers to the first nonempty communication step, not the arrival order within that step. Its quotient parent is the initial tree from which the accepted candidate arrived. The entire initial tree then ignores later candidates and suppresses any pending own release.

Here is the sparse implementation of one communication step. Candidate ids from the preceding step cross original boundary edges in their allocated slots. Members gather their candidate records to the fixed root of the initial tree and broadcast its chosen record along that initial tree. A subtree with no candidate is silent, with absence interpreted only at its fixed depth deadline. An own release is an additional root candidate. Only the first nonempty candidate step of an unassigned Q invokes a nonempty gather and broadcast. Its choice reaches every member before the next communication step, when its boundary members forward the accepted id once. The allowance J covers this boundary exchange and a round trip within the initial tree. Thus empty communication steps consume time but not sending rounds. One flood uses only a constant number of sends per agent over its entire duration.

The first quotient flood produces tentative groups of whole initial trees. Its chosen boundary edges join the fixed internal edges of its constituent initial trees. A parent exchange, census, and broadcast give each tentative group's actual population and tree metadata. Every tentative group of population at least r_s/K_0 is retained, and its constituent initial trees carry that group's label. A second first-arrival flood then reaches initial trees outside the retained groups for $\bar{r}_s$ quotient hops. Each such initial tree attaches as one unit toward a seed. A tentative group may split at this step, but none of its constituent initial trees can split. A final parent exchange, census, and broadcast provide the completed group's records. Noncoverage, a missing required record, or a population below r_s/K_0 raises an alarm.

Each parent relation between initial trees selects one original edge joining their members. The group's communication tree contains all retained edges within its constituent initial trees and one selected edge for each such relation. Its population is

$$M_{s,g} = \sum_{Q \text{ in group } g} m_Q.$$

Every original member contributes one observation per round. The group root is the fixed root of the initial tree whose label won. Group messages follow the expanded tree toward or away from this root. The corresponding parent records are separate from the retained initial-tree records, which never change.

A reconstruction is allocated $64J(R_s + \bar{r}_s + 1)$ rounds, including its fixed number of passes. By Lemma C.2, this is $\mathcal{O}(r_s\Gamma^3)$. Before the scale reaches V, preparation fits into the preceding second window. The reconstruction at the first scale equal to V is the last one. Its groups remain fixed thereafter. The same schedule and testing rules remain in force if an attachment happens to produce only one group earlier. All operations stop at round T.

On a failed first construction, an affected agent continues uniform play if no valid retained tree is available and forwards any alarm. This exceptional behavior is charged by the horizon bound. On the geometry event, all retained trees are ready before testing begins. An alarm stops all later stochastic construction, but the original initial-tree records remain available for fallback.

Windows and Preparation When the first construction requires attachment, the windows follow Eq. (5) and $r_0 = d_0 < V$ by Lemma C.17. They double until the scale reaches V, including the fixed-group increment at that transition. All groups share their boundaries while scales increase. Each fixed group then uses its own windows. As in the main text, g' is the previous group containing the new root.

If the first broadcast and census certify a spanning tree, that single group remains fixed from the outset. For its height h, the parameters satisfy $q = 2h + 3$ and

$$\begin{aligned} W_{0,g} &= \lceil \Lambda(q + 4K/V) \rceil, \\ W_{s+1,g} &= 2W_{s,g} + \left\lceil \Lambda\left(q + \frac{|S_{s+1,g}|}{V\tau_{s+1}^2}\right) \right\rceil. \end{aligned} \tag{10}$$

Set $r = H_i = V$ in the testing rules below. No later construction is needed, and the tree is also retained for adversarial learning. The state, sampling, and alarm rules are the same as in the other branch.

The first phase (phase 0) is uniform and starts from the state $(\widehat{\mu}, \lambda, \tau, W) = (0, \mathbf{0}, 1/2, W_0)$. Only complete first windows prepare a successor. A window interrupted by the horizon supplies no successor and does not itself raise an alarm.

The second window contains a fixed number of passes for pooling, any reconstruction on $G_{\mathcal{Q}}$, and state delivery. The preceding group first pools its observations. The new grouping is then completed, and its root broadcasts the updated state it carries. Every initial tree moves to its new group as a whole. A change from g' to g takes effect immediately after the preceding second window, at $t_{s+1,g} = t_{s,g'} + 2W_{s,g'}$. Before a new phase's first action, the complete incoming state and window are available and validated, and only then are the new independent grids drawn. These preparation coins are included in the history preceding the phase's action randomization. The planned probabilities in this phase are then deterministic functions of its age. No grid is sent or inherited with a state.

A valid state satisfies

$$\widehat{\mu} \in [-1/8, 9/8], \qquad 0 \le \lambda_i \le 5/16, \qquad S = \{i : \lambda_i = 0\} \neq \varnothing, \qquad 0 < \tau \le 1. \tag{11}$$

Counts, precision, and window length must match their scheduled instructions. A missing instruction, invalid state, or failed required record raises an alarm. All numerical checks precede use of the state.

A.2 Phase Sampling and Messages

This subsection follows the sampling rules of Section 3.3 and the testing rules of Section 3.4. At the start of every phase with $s \ge 1$, each agent draws its independent grid ξ^v and initializes its assigned tests. The first phase has $a_i^v = 0$ for every arm and no tests. For a group formed by attachment or a successor of such a group, write $M = M_{s,g}$ and define

$$r = \lceil r_0/K_0 \rceil, \qquad H_i = 1 + \frac{2M}{1 + M\lambda_i/\sqrt{K}}, \qquad B(x) = c_5\Gamma^4\sqrt{(1 + K/x)T}. \tag{12}$$

Here M is the current group population and the fixed r allocates nearby tests. For these groups, integer populations satisfy $M \ge r$ because $M \ge r_0/K_0$. Completion of the first construction before testing gives $r \le r_0 < T$. Also $r^2 \le K$, since $r_0 \le \lceil 2\sqrt{K} \rceil$ and $K_0 \ge 4$. Finally $H_i \ge r$, since $2M/(1 + M\lambda_i/\sqrt{K}) \ge 2r/(1 + 5r/(16\sqrt{K})) > r$. For the spanning tree certified before attachment, $r = H_i = M = V$. This group uses exactly the same rate formula below. Only groups with $r < H_i$ can assign constant-rate tests. Their constant rate is

$$\gamma_i = \frac{2c_1\Gamma^2 T}{B(r)^2\left(1 + T\lambda_i/(32\sqrt{K})\right)}.$$

For $s \ge 1$, the complete initial rate is

$$a_i^v(t_{s,g}) = \begin{cases} [4096\Gamma^2(1 + K/H_i)]^{-1}, & H_i\xi_i^v \le 48\Gamma, \\ \gamma_i, & r\xi_i^v \le 48\Gamma < H_i\xi_i^v, \\ 0, & r\xi_i^v > 48\Gamma. \end{cases}$$

The first branch evolves by Eq. (2), equivalently

$$a_i^v(t) = \left[4096\Gamma^2(1 + K/H_i) + (t - t_{s,g})\lambda_i^2/(c_1\Gamma^2)\right]^{-1}.$$

The other two branches remain constant. When $r = H_i = V$, every assigned test uses the first branch. Its rate is constant when $\lambda_i = 0$. A decreasing-rate tester has the tolerance

$$b_i^v = \lambda_i + 16\tau + \frac{128\Gamma}{B(H_i)a_i^v(t_{s,g})}. \tag{13}$$

The constant-rate formula replaces $B(H_i)$ with $B(r)$. Each assigned average starts at $Z_i^v = \widehat{\mu} + 4\lambda_i$. In both windows, agent v samples by Eq. (1). When it plays an assigned arm, Eq. (6) and Eq. (7) apply. Every decreasing rate advances after every round by Eq. (2), even if its arm was not played. Only the current phase's test records are stored. Unassigned arms have no test record.

Pooling and State Updates During the first window, agents maintain the residual sums in Eq. (3). The second window delivers their sum and the next shared state. The following encoding keeps these messages short, without changing the update in Eq. (4).

For a complete first window, every member sends

$$k_{v,i} = \operatorname{round}\left(\frac{c_7 Y_i^v}{W\tau}\right), \qquad \widetilde{\mu}_i = \widehat{\mu} + \frac{\tau}{c_7 M}\sum_{v\in g} k_{v,i}, \qquad i \in S. \tag{14}$$

Nearest-integer ties are fixed. Intermediate nodes add exactly and the root broadcasts the integer sum. Every local integer and partial sum is checked against $[-c_8 M(M+1), c_8 M(M+1)]$. A failure sends an alarm, not a clipped or oversized number. All members use exactly the M required records, including zero records.

Only a changing tree needs an absolute state message. Temporary estimates are rounded to the grid τ/c_7 before Eq. (4). The new minimum lies on that grid and a newly positive gap on $\tau/(4c_7)$. Previously stored positive gaps lie on coarser nested grids and are sent exactly. The message contains the minimum and the ordered gap vector, including its zeros. There is no candidate mask. At any change of tree, the outgoing phase index satisfies $2^s < V/d_0$, so $1/\tau_s = 2^{s+1} < 2V/d_0 \le 2V$. Thus absolute indices need $\mathcal{O}(\log V)$ bits, including the last change that reaches scale V. On fixed trees, all agents update from the same preceding state and integer sum without sending absolute coordinates again.

For a tree of height h, let t_0 be the first round allocated to a gather. A nonroot at depth d and the root use the sending rounds

$$t_{\mathrm{up}} = t_0 + h - d, \qquad t_{\mathrm{root}} = t_0 + h. \tag{15}$$

A nonroot sends only after all required child records and its own record are available. The root broadcasts the complete sum, and each receiver forwards it in the following round. The pass finishes within $2(h+1)$ rounds and uses at most two sending rounds per agent. Other passes occupy successive allocated intervals. During first-arrival construction, a vertex instead sends when it first accepts a candidate.

An alarm takes priority over other records and is forwarded once to every graph neighbor in the first legal round. The agent permanently stops stochastic actions and messages, retains its first tree, and initializes the two fallback copies. Already sent records need not be recalled, but informed receivers ignore them. A canceled required statistic cannot affect an unaware receiver before the alarm can reach it. Lemma C.3 makes this causal rule precise. A pending incomplete fallback block is skipped, as in Appendix A.3.

A.3 Complete Algorithms

Algorithm 2 specifies the stochastic phases and the switch. All checks and messages use the preceding definitions. The tree learner in Algorithm 3 is shared by the adaptive and oblivious applications. An alarm known before an action determines that action's mode. An alarm generated by its observation changes the following action. This rule interrupts either algorithm's pending stochastic instructions immediately, including construction. Missing required instructions are checked before the action they would govern.

Learning on a Tree A retained tree has m members and height h. The horizon is divided into blocks of $q = 2h + 3$ rounds. Counts from block b are pooled and broadcast during block $b + 1$, then used before block $b + 2$. Two exponential-weights sequences, one for each parity, accommodate this delay. Both start uniformly. Their current distribution w_b is common to all members once activation is complete.

At the start of block b, each agent clears only the new counters $U^v_{b,i}, C^v_{b,i}$. The counts $C^v_{b-1,i}$ remain available until their scheduled transmission. Received feedback from block $b-2$ remains available until its parity update. The first counter records favorable retained outcomes and makes a small adjustment to the shared distribution. Before round t in block b, agent v samples using

$$p^v_t(i) = \frac{w_b(i) + 8\eta U^v_{b,i}}{1 + 8\eta \sum_{j=1}^{K} U^v_{b,j}}. \tag{16}$$

Algorithm 2: Complete best-of-both-worlds protocol for agent v

Input $K, V, T, \delta, \mathcal{N}(v)$
if $K = 1$ or $T = 1$ **then** play uniformly without messages through T and return
Set Γ, Λ, K_0 by Appendix A and $r_s = \min\{2^s\lceil 2\sqrt{K}\rceil, V\}$
Play uniformly during Appendix A.1 and retain the initial-tree records of $Q(v)$
Set $s \leftarrow 0$, $g \leftarrow Q(v)$, $(\widehat{\mu}, \lambda, \tau) \leftarrow (0, \mathbf{0}, 1/2)$, and $W \leftarrow W_{0,g}$ by Eq. (5) or Eq. (10)
for *each prepared phase* (s, g) *before an alarm or round* T **do**
  Set $M \leftarrow M_{s,g}$ and $t_{s,g}$ to its scheduled start
  if the state fails Eq. (11) or a required record is missing **then** raise an alarm
  Set $S \leftarrow \{i : \lambda_i = 0\}$, $Y_i^v \leftarrow 0$ for $i \in S$, and $a_i^v \leftarrow 0$ for $i \in [K]$
  if $s \geq 1$ **then**
    Set $\xi_1^v \sim \mathrm{Unif}[0, 1)$ and $\xi_i^v \leftarrow (\xi_1^v + (i-1)/K) \bmod 1$ for $i \in [K]$
    Set (a_i^v, b_i^v) by Appendix A.2 and $Z_i^v \leftarrow \widehat{\mu} + 4\lambda_i$ for assigned i
  for $t = t_{s,g}, \ldots, \min\{t_{s,g} + 2W - 1, T\}$ *while no alarm is known* **do**
    Play arm $I_t^v \sim p_t^v$ by Eq. (1) and observe $X_t^v(I_t^v)$
    if $a_{I_t^v}^v(t) > 0$ **then** update $Z_{I_t^v}^v$ and test by Eq. (6) and Eq. (7)
    if $t < t_{s,g} + W$ and $I_t^v \in S$ **then** update $Y_{I_t^v}^v$ by Eq. (3)
    In the second window, the scheduled passes pool by Eq. (14) and prepare the next groups and root-inherited state by Appendix A.1
    Set every decreasing rate for round $t + 1$ by Eq. (2)
  if $t_{s,g} + 2W \leq T$ and no alarm is known **then** set $(g, \widehat{\mu}, \lambda, \tau, W)$ to the prepared successor and $s \leftarrow s + 1$
on any failed check or first alarm stochastic work stops and the alarm is forwarded once to every neighbor
if the retained tree is ready **then** Algorithm 3 with Eq. (19) governs the next action onward, otherwise uniform play continues
All operations end at T, and an incomplete window supplies no successor

After observing loss x on arm i, it retains the observation with probability $w_b(i)/(4p_t^v(i))$ and independently draws a Bernoulli proxy with mean x. A retained zero increments $U_{b,i}^v$, and a retained one increments $C_{b,i}^v$. All auxiliary draws use fresh independent randomness. Since $\eta q \leq 1/8$, we have $p_t^v(i) \geq w_b(i)/2$, so retention is well defined.

Multiplying the sampling and retention probabilities gives $w_b(i)/4$, independently of the local correction. Therefore the integer counts can be pooled into

$$G_b(i) = \frac{4\sum_{z \text{ in the group}} C_{b,i}^z}{m w_b(i)}. \tag{17}$$

The common base then changes by

$$w_{b+2}(i) = \frac{w_b(i)\exp(-\eta G_b(i))}{\sum_{j=1}^K w_b(j)\exp(-\eta G_b(j))}. \tag{18}$$

Only integer counts are transmitted, not the probabilities. For adaptive losses, the learning rate is

$$\eta = \frac{1}{8q + \sqrt{2(q + K/m)T/\log K}}. \tag{19}$$

The conditional guarantee in Proposition C.12 remains valid when an alarm activates members at different times. The oblivious application instead uses Eq. (50), with the same remaining updates.

Activation and Complete Blocks We consider a retained rooted tree with m full members and height h. All members know their parent, children, depth, m, and h. The public blocks have $q = 2h + 3$ rounds and are

$$\mathcal{B}_b = \{bq + 1, \ldots, \min\{(b+1)q, T\}\}, \qquad b = 0, 1, \ldots. \tag{20}$$

The block index is retained in Algorithm 3 so that resetting current counters never erases the preceding block's counts. The tree uses the same two parity bases, local zero counters, and retained one counts

Algorithm 3: Tree learning for agent v

Input Retained root, parent, children, depth d, population m, height h, horizon T, rate η, and first active round σ_v

Set $q \leftarrow 2h+3$, $w^{(0)}, w^{(1)} \leftarrow \mathbf{1}/K$, with $\eta q \leq 1/8$, and all records to empty

for *each block b in Eq. (20) intersecting* $\{\sigma_v, \dots, T\}$ **do**

  if complete block-$(b-2)$ feedback is available **then** update $w^{(b \bmod 2)}$ by Eq. (17) and Eq. (18)

  Set $w_b \leftarrow w^{(b \bmod 2)}$ and $U^v_{b,i}, C^v_{b,i} \leftarrow 0$ for $i \in [K]$, with block eligibility $\mathbf{1}\{\sigma_v \leq bq+1\}$

  for $t = \max\{\sigma_v, bq+1\}, \dots, \min\{(b+1)q, T\}$ **do**

    Play arm $I^v_t \sim p^v_t$ by Eq. (16) and observe $x = X^v_t(I^v_t)$

    Set $i \leftarrow I^v_t$, $B \sim \mathrm{Bern}(w_b(i)/(4p^v_t(i)))$, and $Z \sim \mathrm{Bern}(x)$ independently

    Set $U^v_{b,i} \leftarrow U^v_{b,i} + B(1-Z)$ and $C^v_{b,i} \leftarrow C^v_{b,i} + BZ$

    if a gather or broadcast is due **then** send the complete block-$(b-1)$ record by Eq. (15) with $t_0 = bq+1$

An eligible local count and every child record, including zeros, are required for an aggregate

Preceding-block counts and feedback remain available until their scheduled use

as in Eq. (16) and Eq. (17). The adaptive application uses Eq. (19). In particular, $\eta q \leq 1/8$. Only this tuning differs from the fixed-mean application. The update rule and transmitted records are identical.

Both parity bases are initialized uniformly at activation. A member joining during a block uses this base and clears its counters, but does not contribute that incomplete block. The pooled count is $C_{b,i} = \sum_{z \text{ in the tree}} C^z_{b,i}$, with estimate from Eq. (17). Only complete counts update the matching parity base through Eq. (18). Stochastic observations are never included.

Sending Slots and Packet Size For source block b, Eq. (15) uses $t_0 = (b+1)q+1$. Thus its complete counts are gathered and broadcast in block $b+1$, before the first action of block $b+2$. A receiver forwards a complete root broadcast in the next round. A missing local or child record causes silence in its prescribed sending round. The shared deadline, not a count's numerical value, determines completeness. This also applies when activation occurs midway through a block. After its scheduled transmission, an agent may discard the preceding local count, but it keeps any root feedback until the corresponding parity update.

The parity base has not changed since the block whose counts it decodes. Thus no old probability vector is transmitted or stored separately. The bounds $0 \leq C_{b,i} \leq mq \leq 3m^2$, following from $h \leq m-1$, give $\mathcal{O}(K \log m)$ bits per count vector for $m \geq 2$. A singleton sends no messages. The local zero counters are not sent. A constant number of routing, mode, and alarm fields can be concatenated into the packet. The shared clock identifies its block without a transmitted timestamp. Feedback that would arrive after T is used only for virtual final updates in the proof.

A.4 A Known Population Upper Bound

Exact population knowledge is unnecessary. The extension also keeps the completed initial trees fixed.

Corollary A.1. *If only a public $\bar{V} \geq V$ is known, the implementation below retains the stochastic and adaptive regret bounds in Theorem 4.1 with the actual V, D and logarithmic factor $\bar{\Gamma} = \log(2^{10}K\bar{V}T^8/\delta)$. All stochastic sending counts are $\mathcal{O}(\log T)$ simultaneously with probability at least $1-\delta$ and in expectation per agent. Every packet has $\mathcal{O}(K \log \bar{V})$ bits on every execution.*

The extension uses $\bar{\Gamma}$ in place of Γ and $\bar{V}$ in the grouping scales and fixed-group window formulas until a census gives V. It starts at $r_0 = \min\{d_0, \bar{V}\}$ with the same one-time construction. At its first-arrival census before attachment, each vertex additionally records whether any original neighbor has a different root label. The subtree report includes the logical OR of these bits. A zero bit at the root certifies that no original edge leaves its tree. Connectedness then certifies a spanning tree, and its census gives the actual V. This adds one bit and no transmission. An early spanning outcome uses the actual-population global rules from the outset.

If a nonspanning first construction finishes before T, Lemma C.17 gives $D+1 > r_0$ and hence $V > r_0$. Its deadline also exceeds r_0, so $r_0 < T$. Consequently $r_0 = d_0 < V, T$. The resulting

partition $\mathcal{Q}$ is retained, with $m_Q \geq r = \lceil r_0/K_0 \rceil$ and Eq. (8) evaluated at $\bar{\Gamma}$. All later constructions use the quotient $G_{\mathcal{Q}}$ at quotient scale $b = \lceil r_s/r \rceil$ for the current desired population scale r_s. The initial trees are never reconstructed. Estimators and H_i always use the actual group population.

A Census When the Scale Exceeds the Population Seed noncoverage in a later quotient construction triggers a population census rather than an immediate loss alarm. On its ball-intersection event, a trigger at desired scale $r_s \leq \bar{V}$ implies $V < r_s$. Otherwise the weighted-ball proof of Lemma C.2 guarantees coverage. Since every quotient vertex has weight at least r,

$$|\mathcal{Q}| \leq V/r < r_s/r \leq b.$$

Any seed would therefore reach the entire quotient within the allowed b hops. Thus every initial tree is uncovered at the common deadline. If d is the preceding accepted population scale, then

$$d \leq r_s \leq 2d, \qquad V \geq d/K_0, \qquad r_s \leq 2K_0 V.$$

These statements concern desired population scales, not quotient hop counts.

To discover V, the fixed root of the initial tree containing agent 1 releases one label, using the same J-round communication steps. A first-arrival flood, boundary-label exchange, weighted census, and broadcast finish within $J(4b+8)$ interaction rounds. Indeed the quotient has fewer than b vertices. Expanding the resulting quotient tree as in Lemma C.2 gives a spanning tree that contains every initial tree unchanged. The zero boundary-label flag certifies coverage even without knowing V, and the census then supplies its exact value. A missing or inconsistent report at the deadline raises an alarm. The census is performed once at the common deadline.

At the trigger, each preceding group keeps its complete updated estimates but does not halve the preceding precision τ_{hold}, which satisfies $d\tau_{\text{hold}} = d_0/2$. It runs one additional two-window phase with $W_{\text{hold}} = \lceil \Lambda r_s \rceil$. Fresh grids and the same testing rules apply at its preparation. The updated estimates are more accurate than required at this unchanged precision. The census runs in its first window. Since $J/r = \mathcal{O}(\bar{\Gamma}^2)$ and $b \leq 2r_s/r$, its duration is $\mathcal{O}(r_s\bar{\Gamma}^2)$ and fits that window. During the second, agents pool observations on the preceding group trees and transfer the complete outgoing state of the group containing the new root. Their heights and the new tree's height fit the same padding. An alarm cancels this work permanently and still propagates over original edges, not quotient steps. The trigger cannot occur during a valid initialization, because a nonspanning first construction has $r_0 < V$.

The new fixed statistical group keeps all initial trees intact, retains the growing-branch testing floor r, and continues to use the original retained trees for fallback. Its initial precision is $\tau_1 = \tau_{\text{hold}}/2$ and its initial window is

$$W_1 = \left\lceil \Lambda \left(V + \frac{|S_1|}{V\tau_1^2} \right) \right\rceil.$$

The additional window is not inherited. Subsequent windows use the fixed-group rule in Eq. (5). This is a change of statistical grouping, not a new mode switch.

Accuracy, Regret, and Communication The accuracy budget during this census follows from $M \geq d/K_0$ and $r_s \geq d$,

$$\frac{MW_{\text{hold}}\tau_{\text{hold}}^2}{|S|} \geq \frac{\Lambda d r_s \tau_{\text{hold}}^2}{K_0 K} \geq c_6\bar{\Gamma}.$$

Keeping the precision fixed relaxes the incoming invariant. Candidate width is $\widetilde{\mathcal{O}}(r_s\tau_{\text{hold}}) = \widetilde{\mathcal{O}}(\sqrt{K})$. Conditional decreasing-test costs obey the same scale split as in a growing-group phase, because $N/M = \widetilde{\mathcal{O}}(r_s/d) = \widetilde{\mathcal{O}}(1)$. Constant-test cost is bounded by Eq. (27). This additional phase therefore costs $\widetilde{\mathcal{O}}(\sqrt{K})$ stochastically and one existing $B(r)$ allowance plus the same deterministic prices adversarially.

For its single state transfer, the reference and stored gaps are rounded at precisions τ_{hold}/c_7 and $\tau_{\text{hold}}/(4c_7)$. The reference error is below $\tau_{\text{hold}}/1024 + \tau_{\text{hold}}/(2c_7) < \tau_{\text{hold}}/512$, which fits the new $\tau_1/256$ invariant. Previously positive gaps lie on coarser nested grids and are exact. A newly

Table 3: Our (V, D)-worst-graph bounds. Each entry is capped by T. The stochastic upper bound uses the actual inverse-gap sum H_μ, whereas its lower bound is over the class $H_\mu \le H$ for any finite $H \ge K - 1$. The algorithms and communication guarantees are unchanged.

Regime	Upper bound $\widetilde{\mathcal{O}}(\cdot)$	Lower bound $\Omega(\cdot)$
Stochastic	$\min\{D, \sqrt{K}\} + \min\{H_\mu/V, \sqrt{KT/V}\}$	$\min\{D, \sqrt{K}\} + \min\{H/V, \sqrt{KT/V}\}$
Adaptive	$\sqrt{KT/V} + \sqrt{\min\{D, \sqrt{K}\}T}$	$\sqrt{KT/V} + \sqrt{\min\{D, \sqrt{K}\}T}$
Oblivious	$\sqrt{KT/V} + \sqrt{T} + \min\{D, (KT)^{1/3}\}$	$\sqrt{KT/V} + \sqrt{T \log K} + \min\{D, \sqrt{K}\}$

positive gap exceeds τ_{hold} and has error at most $\tau_{\text{hold}}/2048$ relative to its true quarter-gap, so the extra rounding cannot make it zero or violate $\Delta_i/5 \le \lambda_i \le \Delta_i/3$. Zero is represented exactly.

The expanded spanning tree has height at most $V - 1$. Also $V\tau_1 \le r_s\tau_{\text{hold}}/2 \le d_0/2 = \mathcal{O}(\sqrt{K})$. The fresh-restart calculation in Lemma C.9, with $q = V$ and no inherited window, therefore gives the same $\sqrt{K} + H_\mu/V$ and $\sqrt{K} + \sqrt{KT/V}$ stochastic rates. Conditional testing is unchanged. If the working cap is reached without discovery, accepted groups have $M \ge \bar{V}/K_0$, and the bounds with denominator $\bar{V}$ imply the ones with actual V.

For adaptive regret, the first retained trees still have population at least r_0/K_0 and height $\widetilde{\mathcal{O}}(r_0)$. Their conditional fallback regret and the unchanged floor r give $\widetilde{\mathcal{O}}(K^{1/4}\sqrt{T})$ on the nonspanning branch. Its first construction implies $D \ge \sqrt{K}$, so this is the order in Theorem 4.1. The one additional phase adds only one allowance of the same order. The expanded grouping never changes the agent-level alarm speed or the retained fallback routes. An early spanning outcome is still covered by Proposition C.16 and Lemma C.17 with the actual population.

The windows in Eq. (5) double. The extension adds one additional phase, one deterministic quotient census, and one state transfer. Each needs only $\mathcal{O}(1)$ extra sending rounds per agent, by the sparse communication-step implementation. Graph records are bounded counts and ids, and numerical packets retain their checked ranges. At the single transfer, $1/\tau_{\text{hold}} = 2d/d_0 \le 2\bar{V}$, so its indices use $\mathcal{O}(\log \bar{V})$ bits. Conditional geometry, concentration, and testing events use $\bar{\Gamma}$ and have total failure at most ε. Exceptional executions contribute at most $T\varepsilon \le 1$. This proves Corollary A.1 while preserving the fixed initial-tree partition.

B Worst-Graph Bounds at a Fixed Diameter

Worst-graph bounds fix the number of agents and allow the diameter to vary. Our (V, D)-worst-graph bounds also fix the diameter to D. They retain information about how far observations must travel that is omitted from the main results. A pair is feasible when $V = 1, D = 0$, or $V \ge 2$ and $1 \le D \le V - 1$.

Proposition B.1 (Bounds at a Fixed Diameter). *For every $K \ge 2$, $T \ge 1$, and feasible (V, D), the upper bounds in Table 3 hold for every connected graph on V agents of diameter D and every agent. The stochastic and adaptive bounds are achieved by Algorithm 1, and the oblivious bound uses the configuration in Appendix E.1. None requires prior knowledge of D.*

For the lower bounds, there are a fixed graph with these parameters and a fixed target such that every cooperative policy under unrestricted local communication admits a hard instance attaining the corresponding entry. The stochastic assertion holds for every finite budget $H \ge K - 1$ with a unique optimal arm and $H_\mu \le H$. The adaptive and oblivious assertions admit deterministic common binary loss processes of the respective types. The instance may depend on the policy and the regime, but the graph and target do not.

The upper bounds follow from the two first-construction outcomes in Appendix C and Appendix E.1. The lower bounds are proved in Appendix D.1, Appendix D.2, and Appendix E.2. They use a path ending in a clique, with the target at the other end. This graph has exactly the prescribed population and diameter. The lower bounds are worst-case over graphs of diameter D and target agents.

Relation to the Main Results A fixed-diameter upper bound applies to a smaller graph class and can be sharper on that class. Allowing all feasible diameters recovers a worst-graph upper bound. Conversely, a worst-graph upper bound applies to every fixed-diameter class, but need not capture its easier cases. A lower bound for one fixed diameter is also a worst-graph lower bound, whereas a worst-graph lower bound need not hold at every diameter. Thus the distinction is the graph class, not two incompatible notions of regret.

For our results, a path with $D = V - 1$ recovers the main lower bounds up to constants, including the horizon cap. The stochastic path proof is in Appendix D.1. The adaptive path calculation is in Appendix D.2, and the oblivious calculation is in Appendix E.2. The stochastic and adaptive bounds match up to logarithmic factors in their stated senses. The oblivious bounds match in some parameter ranges, but a gap remains in general, both with and without fixing the diameter.

Earlier Graph-Dependent Guarantees The prior upper bounds in Table 1 retain their diameter, neighborhood, or gossip-matrix dependence rather than maximizing over graphs.

For (Qiu et al., 2026), let $\rho = (1 - \sigma_2(W))^{-1/2}$ for the prescribed gossip matrix of Qiu et al. (2026). Specializing their Theorem 6 to $N = V$ and homogeneous losses makes their global gaps equal to Δ_i and their global-average pseudo-regret equal to each agent's local pseudo-regret. Their stochastic bound, which assumes a unique optimal arm, becomes $\widetilde{\mathcal{O}}((\rho/V + \rho^2)H_\mu + \rho^2)$, while their oblivious adversarial bound is $\widetilde{\mathcal{O}}(\sqrt{\rho T} + \sqrt{KT/V} + \rho)$. The additive ρ^2 and ρ terms come from $B^2 \log K$ and $B \log K$, respectively, where $B = \Theta(\rho \log(KVT))$ is their gossip block length. Their protocol exchanges K real-valued coordinates per neighbor per round, without a stated bit-precision guarantee. These bounds depend on the chosen matrix, not just (V, D), so they are not entries in our diameter-restricted table.

C Proof of Individual Regret in Theorem 4.1

We first establish the common ingredients, then the regret on growing groups. The full-population specialization and the first construction are analyzed in Proposition C.16 and Lemma C.17, respectively. Each adversarial argument fixes a comparator before taking expectations.

C.1 Supporting Lemmas

Trees and Complete Records

Lemma C.1 (First Arrivals and Populated Groups)**.** *At a scale $1 \le r_0 \le V$, the temporary trees in the one-time construction have height at most $L = 8r_0 + \lceil c_3 r_0 \Gamma \rceil$. With the allocated probability, its completed initial trees cover all vertices, have population at least r_0/K_0, and have height at most $L + r_0 = \mathcal{O}(r_0\Gamma)$. The construction takes $\mathcal{O}(r_0\Gamma)$ rounds, $\mathcal{O}(1)$ sending rounds per vertex, and $\mathcal{O}(\log V)$ bits per edge packet. A certified spanning first tree satisfies the stronger timing assertion in Lemma C.17. The first-arrival primitive without priority for agent 1 has height at most $R = \lceil c_3 r_0 \Gamma \rceil$ and is also used on the quotient graph below.*

Proof. Write the release time of vertex z as d_z. Its first-arrival label at v minimizes $(d_z + \mathrm{dist}(z, v), z)$ lexicographically. A winning label wins along a shortest path to its destination. Otherwise, a label that beats it at an intermediate vertex also beats it at the destination. Suppressing a displaced release therefore removes no winning path. Parents strictly precede their children and form trees. Comparing with v's own scheduled release bounds the root distance by L in the first construction. For the primitive $d_z = R - \chi_z$, the sharper comparison gives distance at most $\chi_z - \chi_v \le R$. All vertices accept by the largest scheduled release round plus one, and each broadcasts once.

The first step concerns the exponential-shift partition without advancing vertex 1. For the untruncated shifts, put $\beta = 1/r_0$ and $\phi_z = E_z - \mathrm{dist}(z, v)$. If a rounded-shift cluster meets a radius-b ball about v, its score ϕ_z is within $2b+1$ of the maximum. If at least j clusters meet that ball, the largest j scores are all within this amount. Conditional on the identity and value of the jth score and the indices above it, the higher scores remain independent with lower truncations. Exponential memorylessness bounds the chance that each lies within the next $2b + 1$ by $1 - e^{-\beta(2b+1)}$. A negative truncation only

reduces this probability. Thus

$$\mathbb{P}\{j \text{ or more clusters meet } B_G(v,b)\} \le (1-e^{-(2b+1)/r_0})^{j-1} + Ve^{-R/r_0}.$$

The last term pays for clipping. Giving vertex 1 an earlier release only enlarges its cluster and removes vertices from other clusters. Each other resulting cluster is a subset of its counterpart. Therefore any ball meets at most one more cluster than before this change.

A radius-r_0 ball contains at least r_0 vertices, or the entire graph with $V \ge r_0$. At $b = r_0$, $(2r_0 + 1)/r_0 \le 3$. Increasing the universal c_3 in $K_0 = \lceil c_3\Gamma \rceil$ if necessary, a union bound ensures that all these balls meet at most K_0 first-arrival clusters, including the advanced label. One of them has total population at least r_0/K_0. Its vertices are seeds, so every vertex is within r_0 hops of a seed. Keeping each seeded tree intact and attaching the remaining vertices gives height at most $L + r_0$ in the first construction and $R + r_0$ for the primitive. Attachment parents strictly decrease distance to the seed set, so cannot form a cycle. Every final group contains an intact sufficiently populated cluster. A spanning tree already contains all $V \ge r_0$ vertices and needs no attachment.

The first-arrival declarations identify every child once all neighbors have declared. Each census and broadcast therefore sends at most once at each vertex. In the first construction all declarations arrive by $L + 2$, the reports reach their roots by $2L + 4$, and metadata broadcasts finish by $3L + 5$. These passes, the attachment wave, and its final census fit the scheduled $\mathcal{O}(r_0\Gamma)$ rounds. All counts, indices, and actual depths have $\mathcal{O}(\log V)$ bits. For analysis, a canceled call's topology-only random tape is completed with independent unused shifts at stopped vertices. By Lemma C.3, cancellation cannot change an unaware receiver's intended tree before its alarm arrives. Conditional application at each scheduled call and a union over polynomially many vertices and calls fit $\varepsilon/32$ for the one-time construction. The completion supplies no loss observations and does not condition on future estimates or passing tests. □

Lemma C.2 (Regrouping Fixed Initial Trees)**.** *Suppose that a completed initial-tree partition $\mathcal{Q}$ has $m_{\mathcal{Q}} \ge r$ and height at most h_* for every initial tree. For every desired population scale $r_0 \le r_s \le V$, the construction in Appendix A.1, with quotient scale $b = \lceil r_s/r \rceil$, has the following properties on its allocated event. Every group is a union of whole initial trees, has population at least r_s/K_0, and has a communication tree of height*

$$\mathcal{O}\big(J(b\Gamma + 1)\big) = \mathcal{O}(r_s\Gamma^3).$$

It takes $\mathcal{O}(r_s\Gamma^3)$ interaction rounds, $\mathcal{O}(1)$ sending rounds per agent, and $\mathcal{O}(\log V)$ bits per construction packet. The fixed initial roots and tree edges are preserved. These properties hold simultaneously over all later constructions with total additional failure probability at most $\varepsilon/32$.

Proof. The quotient $G_{\mathcal{Q}}$ is connected because the original graph is connected and each contracted set is connected. Its vertex weights are at least r and sum to V. For any quotient vertex Q, its radius-b ball either is the entire quotient or contains at least $b + 1$ quotient vertices. In the latter case, a path to a vertex outside the ball supplies them. Consequently its total population is at least

$$\min\{rb, V\} \ge r_s.$$

The exponential-shift calculation in Lemma C.1 applies on this unweighted quotient with rate $1/b$ and clipping $R_b = \lceil c_3 b\Gamma \rceil$. The weights are used only for populations, not for the first-arrival distances. For each radius-b ball, the chance of meeting at least j tentative groups is at most

$$(1-e^{-(2b+1)/b})^{j-1} + |\mathcal{Q}|e^{-R_b/b} \le (1-e^{-3})^{j-1} + Ve^{-c_3\Gamma}.$$

Thus, with the allocated probability, every such ball meets at most K_0 tentative groups. Since its population is at least r_s, one of those groups has total population at least r_s/K_0. All initial trees in that group are seeds. Every quotient vertex is within b hops of a seed, so the second flood covers the whole quotient. Every completed group contains an intact sufficiently populated tentative group.

Each quotient vertex accepts only one parent. Within a retained tentative group its original quotient tree is unchanged. Outside the seed groups the new parents strictly decrease distance to the seed set, so they cannot create a cycle. The final quotient structure of each group is therefore a tree of height at

most $R_b + b$. Expanding a group with k initial trees adds all their fixed internal edges and exactly $k - 1$ boundary edges. It is connected and has

$$\sum_{Q \text{ in group } g} (m_Q - 1) + (k - 1) = M_{s,g} - 1$$

edges, and hence is a tree. A root-to-member route crosses at most $R_b + b + 1$ initial trees. Traversing one costs at most $2h_*$ internal edges and one boundary edge, giving height at most $(2h_* + 1)(R_b + b + 1)$. This is at most $\mathcal{O}(J(b\Gamma + 1))$. Independently, its actual height is always at most $M_{s,g} - 1 \le V - 1$.

Since $r_s \ge r_0 \ge r$, we have $b \le 2r_s/r$. Also $r_0/r \le K_0$ and $J = \mathcal{O}(r_0\Gamma)$. Therefore

$$J(b\Gamma + 1) \le C\frac{Jr_s\Gamma}{r} \le C' r_s \Gamma^3.$$

The rescaling in Eq. (9) is essential. Running the quotient flood at scale r_s instead of $\lceil r_s/r \rceil$ would not give this bound, because an initial-tree round trip can cost $\mathcal{O}(r_0\Gamma)$ interaction rounds.

For communication, consider the first nonempty candidate step in one initial tree. Each boundary vertex sends its chosen candidate toward the fixed root, each internal vertex combines any received candidates once, and the root's decision is forwarded once to every member. All these operations fit one communication step of J rounds. Before that step, empty subtrees send nothing and use fixed deadlines to interpret silence. Afterwards the entire initial tree is assigned and never forwards another candidate for that flood. The boundary forwarding to neighboring initial trees also occurs once. Thus the first flood and the attachment flood each use a constant number of sending rounds per agent, not per communication step. The parent exchanges, rerooting of the *group* routes, censuses, and broadcasts are a fixed number of additional passes. The retained initial-tree records are not overwritten. All these operations fit $64J(R_b + b + 1)$ interaction rounds.

A record contains a constant number of root ids, boundary endpoints, actual depths, population counts, and control bits, all bounded by V. The communication-step lengths and deadlines are public and need not be transmitted. A vertex with several neighboring initial trees sends separate bounded records along the corresponding original edges, and simultaneous sends count as one sending round. No message contains a membership list or a whole quotient-neighborhood list. Construction packets therefore have $\mathcal{O}(\log V)$ bits on every execution.

Conditional on the actual starting history, the quotient and its weights are fixed and this construction's shifts are fresh. The displayed tails can be unioned over at most polynomially many initial trees and started constructions before T. The universal choice of c_3 makes the total at most $\varepsilon/32$. If an alarm cancels a construction, independent unused topology randomness completes it only for analysis. The message-causality argument of Lemma C.3 applies to the gather and broadcast dependencies as well as to boundary edges, so cancellation cannot change an unaware agent's intended records before its alarm arrives. This completion supplies no observations and does not condition on future successful tests. □

Lemma C.3 (Alarm and Record Causality)**.** *Every accepted estimate uses all required generated observations. Canceling stochastic work cannot change an unaware vertex's required message transcript before an alarm reaches that vertex. On a retained tree of height h, a first alarm reaches every member within $2h + 2$ rounds.*

Proof. A leaf sends a statistic only after its full observation window. An internal vertex requires every required child record and adds its own. Induction on disjoint subtrees proves completeness. A zero vector is a record. A missing required vector at its deadline generates an alarm instead of a partial sum.

The sender of the first canceled transmission along a scheduled information path already knows an alarm. A locally generated alarm can replace that round's record, and a received alarm is forwarded in the following round, which is the earliest round in which the received information could be forwarded. Along every subsequent edge, alarm forwarding is no later than the canceled record could have been used. A decision from silence waits until the record's deadline. Induction covers both numerical and construction messages. An already complete state may arrive before an alarm generated later, but then it uses only observations that actually occurred.

This comparison concerns message paths on the actual interaction, not losses in a counterfactual adaptive environment. All concentration arguments below use actual stopped observations. Finally, two members of the retained tree are at distance at most $2h$, and the next-action convention adds at most two rounds. □

Sampling and Conditional Coverage The following quantities are used only in the proof. In a growing group put

$$n_i = \begin{cases} M, & \lambda_i = 0, \\ \min\{M, \lceil \sqrt{K}/\lambda_i \rceil\}, & \lambda_i > 0, \end{cases} \qquad L_i^{\text{test}} = \min\{N, 16n_i\},$$

where $N \le T$ is the horizon-clipped planned phase duration. Directly from Eq. (12),

$$n_i \le H_i \le 3n_i \le 3M, \qquad H_i\lambda_i \ge \sqrt{K} \text{ or } H_i \ge M, \qquad \lambda_i L_i^{\text{test}} \le 20\sqrt{K}. \tag{21}$$

For example, if $M\lambda_i \le \sqrt{K}$, then the fractional term in H_i is at least M. Otherwise it exceeds $\sqrt{K}/\lambda_i$, and the added one covers the ceiling. The upper comparison follows by the same two cases. It also follows that

$$\gamma_i \le \frac{2c_1\Gamma^2 T}{B(r)^2}, \qquad \lambda_i\gamma_i \le \frac{64c_1\Gamma^2\sqrt{K}}{B(r)^2}, \qquad \gamma_i \ge \frac{c_1\Gamma^2 L_i^{\text{test}}}{B(r)^2}. \tag{22}$$

The last inequality follows from $L_i^{\text{test}} \le T$ and $L_i^{\text{test}}(1 + T\lambda_i/(32\sqrt{K})) < 2T$.

Lemma C.4 (Feasibility and Nearby Tests)**.** *Every valid preparation and every realized grid satisfy* $\sum_i a_i^v(t) \le 1/16$. *On an event of failure probability at most* $\varepsilon/16$*, every prepared growing-group phase with* $s \ge 1$*, target* v*, and arm* i *has an assigned tester within tree distance* r *and a decreasing-rate tester within tree distance* n_i*. In a global group every arm has an assigned tester. An unavailable tester in these assertions is used only through its earlier alarm, never through ungenerated observations.*

Proof. Write $b_0 = 48\Gamma$ and $L = 4096\Gamma^2$ in this proof. A selected decreasing rate has initial value $H_i/[L(H_i + K)] \le b_0/[L(b_0 + K\xi_i^v)]$. The ordered grid is $\xi + j/K$, $0 \le \xi < 1/K$. Summing this decreasing upper envelope over $j = 0, \dots, K-1$ costs at most $(1 + b_0 \log K)/L$. By the calibration, $\gamma_i \le r/[L(r + K)]$. At most $1 + b_0K/r$ arms satisfy the second cutoff, so their total costs at most $(1 + b_0)/L$. The actual exclusive allocation is at most the sum of these two bounds, which is less than $1/16$. Rates do not increase with age. In a global group the direct bound is $(1 + b_0K/V)/[L(1 + K/V)] < 1/16$.

The coverage argument conditions on the history immediately before the new phase grids. Its tree, populations, state, and numerical radii are already fixed. Here the tree is the expanded group tree from Lemma C.2. Distances below count original graph edges, not quotient steps. Every initial-tree member draws its own independent phase grid, and alarms still travel one original edge per round. A tree ball of radius r contains at least r vertices because $M \ge r$. At each vertex the assigned cutoff has probability $\min\{1, b_0/r\}$, independently across vertices. Vacancy is at most e^{-b_0}. A radius-n_i ball similarly contains at least n_i vertices, and $H_i \le 3n_i$ makes its decreasing-test vacancy at most $e^{-16\Gamma}$. Global vacancy is at most $e^{-48\Gamma}$. The bounds remain true when a cutoff exceeds one, since vacancy is then impossible.

There are at most polynomially many actual group preparations, target vertices, and arms before T. Conditional bounds followed by the tower property and a union bound fit $\varepsilon/16$. No union over possible real-valued states is required. If an agent did not activate the scheduled phase because it already knew an alarm, we give it an unused independent grid for this coverage calculation. The coin only selects a vertex. Before any certified observation is used below, Lemma C.3 either ensures that it was generated in the shared phase, or delivers that vertex's earlier alarm. The auxiliary coin never supplies data. □

One Local Test and Two Loss Bounds All indices t below refer to the original interaction rounds. A prefix of length z in phase (s, g) is the interval $\{t_{s,g}, \dots, t_{s,g} + z - 1\}$. Endpoints described by their lengths are measured from $t_{s,g}$ without changing the loss or filtration indices.

Fresh-action independence gives, for every predictable p_t^v and fixed arm i,

$$\mathbb{E}\left[\sum_t A_t\big(X_t^v(I_t^v) - X_t^v(i)\big)\right] = \mathbb{E}\left[\sum_t A_t\big(\langle p_t^v, \ell_t\rangle - \ell_t(i)\big)\right] \tag{23}$$

for any predictable zero-one indicator A_t. An importance-weighted centered observation has mean $\ell_t(i) - \widehat{\mu}$ and second moment at most $(9/8)^2/p_t^v(i)$. These identities condition on the full preceding history, not on a current peer loss or a future passing event.

We use the conditional scalar Bernstein form of Freedman's inequality (Freedman, 1975; Tropp, 2011). If martingale differences have centered range bound R and predictable variance at most v, fixed at their actual preparation, then

$$\mathbb{P}\left\{\left|\sum_t D_t\right| > \sqrt{2vx} + 2Rx/3 \,\middle|\, \text{preparation}\right\} \le 2e^{-x}.$$

Predictably stopping both an increment and its conditional mean preserves the assertion. All ensuing intervals have integer endpoints. Conditional bounds are averaged over actual preparations, including their fresh grids, before the union bound.

For a single assigned test, suppress indices and write

$$c = \widehat{\mu} + 4\lambda, \quad e = \lambda + 16\tau, \quad \omega = 128\Gamma/(B_* a_0), \quad b = e + \omega,$$

where $a_0 > 0$ is its initial rate and, only in this local calculation, $B_* = B(H_i)$ for a decreasing test and $B_* = B(r)$ for a constant test. The two analysis baselines are

$$L_i = \widehat{\mu} + 2\lambda_i - 32\tau, \qquad U_i = \widehat{\mu} + 6\lambda_i + 32\tau.$$

An interval is called passing only when all its required observations were generated in this phase and its moving-average checks passed, including the endpoints. A crossing observation is not a passing endpoint.

Lemma C.5 (A Passing Moving Average)**.** *Except on an event of probability at most $\varepsilon/16$, every such test with $0 < b < 2$ and passing interval I satisfies both inequalities*

$$\sum_{t\in I}(\ell_t(i) - L_i), \quad \sum_{t\in I}(U_i - \ell_t(i)) \;\ge -\frac{512\Gamma}{b\min_{t\in I} p_t^v(i)} - 2\omega|I|. \tag{24}$$

No stochastic accuracy is required.

Proof. Let $R_0 = 3b/2$. On a selected round, $X_t^v(i) - c = (Z_{t-1} - c) + (64\Gamma + b^2)(Z_t - Z_{t-1})/b^2$. Consequently, over any passing interval, each of $\sum_t \mathbf{1}\{I_t^v = i\}(c - X_t^v(i) - R_0)$ and $\sum_t \mathbf{1}\{I_t^v = i\}(X_t^v(i) - c - R_0)$ is at most $2R_0(64\Gamma + b^2)/b^2 = 192\Gamma/b + 3b$. The lower margin is pointwise nonnegative when $L_i \le 0$, and the upper margin is so when $U_i \ge 1$. In either other case the selected increment, including zero, has range at most four and conditional variance at most $16p_t^v(i)$.

At the preparation the planned $p_t^v(i)$ is a fixed function of time. Freedman at 8Γ, with $V_I = \sum_{t\in I} p_t^v(i)$, bounds the centered sum by

$$16\sqrt{\Gamma V_I} + 64\Gamma/3 \le (b/2)V_I + 512\Gamma/(3b).$$

Combining with the moving-average identity shows that each corresponding margin f_t satisfies $\sum_{t\in I} p_t^v(i)(f_t + 2\omega) \ge -512\Gamma/b$. This holds simultaneously on all subintervals. If $p_t^v(i)$ decreases, summation by parts using suffixes divides this bound by the last probability. If it increases, the same argument with prefixes divides by the first. These are the two possible monotonicities under Eq. (1), giving Eq. (24).

During shutdown, selected increments and means are multiplied by the predictable activity indicator, retaining the deterministic planned variance upper bound. On a generated interval the stopped and intended sums agree. The current crossing is included in concentration but never used as a pass. At most $\mathcal{O}(KVT^3)$ conditional events for preparations, arms, signs, and endpoints, each with probability $2e^{-8\Gamma}$, fit the stated confidence budget. □

Lemma C.6 (Prefix and Recent-Interval Certificates). *On the event of Lemma C.5, a decreasing-rate tester of scale H certifies each of the two margins in Eq. (24) on a passing prefix of length z by $-8B(H)-\lambda_i z$. If $\lambda_i = 0$, the allowance $-8B(H)$ holds on any passing subinterval. A constant-rate tester certifies any passing interval of length at most L_i^{test} by $-8B(r) - \lambda_i|I|$.*

Proof. For a decreasing test $B_* = B(H)$ and $a_0^{-1} = 4096\Gamma^2(1 + K/H)$. Its actual probability dominates its testing rate, so

$$\frac{1}{\min_{t_{s,g}\le t<t_{s,g}+z} p_t^v(i)} \le a_0^{-1} + \frac{\lambda_i^2(z-1)}{c_1\Gamma^2}, \qquad \frac{\omega T}{B_*} = \frac{2^{19}}{c_5^2\Gamma^5} < \frac{1}{8}.$$

Using $b \ge \omega$ and $b \ge \lambda_i$ in Eq. (24) gives $-4B_* - 512\lambda_i z/(c_1\Gamma) - 2\omega z \ge -8B_* - \lambda_i z$. When $\lambda_i = 0$ the same proof applies on any interval because $p_t^v(i) \ge a_0$.

For a constant test $B_* = B(r)$ and $a_0 = \gamma_i$. Eq. (22) implies $\omega L_i^{\mathrm{test}} \le 128B_*/(c_1\Gamma) < B_*/8$. The interval inequality then gives $-4B_* - 2\omega L_i^{\mathrm{test}} \ge -8B_*$, a stronger bound.

It remains to cover $b \ge 2$. A nontrivial lower margin has $L_i = c - 2e > 0$, so $e < c/2 \le 19/16$. A nontrivial upper margin has $U_i = c + 2e < 1$, so $e < (1-c)/2 \le 9/16$. Hence $\omega > 13/16$. The two preceding noise bounds respectively imply $T < 2B_*/13$ or $L_i^{\mathrm{test}} < 2B_*/13$. Since the one-round negative margins are at most $7/4$ and $9/8$, paying the entire interval directly costs less than B_*. Pointwise nonnegative margins require no payment. □

Lemma C.7 (Certificates at the Target). *On the coverage and local-test events, every growing-group phase prefix before the target learns an alarm satisfies*

$$\sum_{t=t_{s,g}}^{t_{s,g}+z-1} (\ell_t(i) - L_i) \ge -CB(r) - \lambda_i z, \qquad \sum_{t=t_{s,g}}^{t_{s,g}+z-1} (\ell_t(i) - U_i) \le CB(r) + \lambda_i z.$$

For candidate arms both inequalities hold on every subinterval, without the gap terms. In a global group the same assertions hold with $B(V) + q$ in place of $B(r)$, where $q = 2h + 3$.

Proof. For growing groups set $k = (z - 4r - 4)_+$. If a decreasing tester lies within distance r, its generated prefix through k is certified with allowance at most $8B(r)$, since $H_i \ge r$. Otherwise there is a constant tester within distance r and a decreasing tester within distance n_i. Put $j = (k - L_i^{\mathrm{test}})_+$. The latter certifies the first j observations and the former the following $k - j$ observations. If $j > 0$ then $L_i^{\mathrm{test}} < N$, so $L_i^{\mathrm{test}} = 16n_i$ covers the remote alarm route. If $L_i^{\mathrm{test}} = N$, the remote prefix is empty. Both allowances are bounded by $8B(r)$, and their gap charges total at most $\lambda_i k$.

Every claimed tester really passed these endpoints. An earlier crossing, known alarm, or missing phase instruction at the nearby or remote tester would already have reached the target by z, by Lemma C.3. The target and each still-active tester use the same prepared group state. No observations after a tester's stop are used. The final at most $4r + 4$ rounds are paid directly. They fit the allowance because $r^2 \le K$, $r \le T$, and thus

$$B(r) \ge c_5\Gamma^4\sqrt{KT/r} \ge c_5\Gamma^4 r.$$

The one-round margins are uniformly bounded on valid states. For candidates, the interval version of Lemma C.6 applies to the intersections with the same prefix, suffix, and final interval.

In the global group a tester for every arm is at tree distance at most $2h$. It must have generated and passed through $(z - q)_+$, or its alarm would already be known. The decreasing-test certificate applies at $H = V$, and the last q rounds are paid directly. The interval version treats candidates. □

Stochastic Accuracy and Test Silence

Lemma C.8 (Accurate Complete Windows). *Every valid scheduled window satisfies $MW\tau^2/|S| \ge c_6\Gamma$. With the allocated stochastic probability, simultaneously for all generated complete windows,*

$$|Y_i^v - W(\mu_i - \widehat{\mu})| \le C\big(\sqrt{|S|W\Gamma} + |S|\Gamma\big), \qquad |\widetilde{\mu}_i - \mu_i| \le \tau/1024.$$

On an accurate incoming state, no residual integer or partial sum overflows. Every transmitted numerical record has $\mathcal{O}(K \log V)$ bits on every execution.

Proof. Before groups freeze, $W_s = 2^s\lceil \Lambda d_0 \rceil$, $M \geq r_s/K_0$, and $r_s\tau_s = d_0/2 \geq \sqrt{K}$. The population bound now follows from Lemma C.1 for phase 0 and Lemma C.2 thereafter. Thus $MW_s\tau_s^2/|S_s| \geq \Lambda/K_0$. At and after the cap, the fresh window term is at least $\Lambda|S_s|/(V\tau_s^2)$ and $M \geq V/K_0$. In an early spanning group the same check uses $M = V$ and Eq. (10). The chosen c_2 therefore gives the required accuracy budget.

The second window first pools observations on the current group tree and broadcasts its update, then completes any quotient reconstruction, and finally delivers the state carried by the new root along the new group tree. In a growing phase these operations take at most $C(r_s + r_{s+1})\Gamma^3 \leq 3Cr_s\Gamma^3$ interaction rounds by Lemma C.2. The preceding second window has $W_s \geq \Lambda r_s = c_2\Gamma^6 r_s$ rounds, including when its successor is the first scale equal to V. Increasing only the universal c_2 covers all these operations. After this last construction, the actual group-tree height is at most $V - 1$, so the routing term ΛV covers every later pass. The early spanning case uses its own term Λq. Thus the communication fits the windows required by the statistical analysis.

Candidate probabilities are at least $15/(16|S|)$. A selected residual has centered range $\mathcal{O}(|S|)$ and conditional second moment $\mathcal{O}(|S|)$. Conditional Freedman gives the displayed individual error. For the pooled sum, stochastic observations and fresh action coins can be exposed consecutively across members in a round, giving range $\mathcal{O}(|S|)$ and variance $\mathcal{O}(MW|S|)$. Its normalized error is at most

$$C\sqrt{\frac{|S|\Gamma}{MW}} + C\frac{|S|\Gamma}{MW}.$$

The chosen c_6 makes this at most $\tau/4096$. Individual rounding contributes at most $\tau/(2c_7)$ to a pooled mean. If a tree changes, the additional absolute rounding contributes at most the same amount. Together they fit $\tau/1024$.

For the range claim, division of the individual error bound by $W\tau/c_7$ gives the normalized error. The budget bounds its two terms by $\mathcal{O}(\sqrt{M} + M\tau) \subseteq \mathcal{O}(M)$. On an accurate incoming candidate, $|\mu_i - \widehat{\mu}| < 11\tau$, so its centering contribution is bounded by a numerical constant. In the uniform first phase $\widehat{\mu} = 0$ and $\tau = 1/2$, giving the same bound. Each local integer is therefore $\mathcal{O}(M)$ and each partial sum $\mathcal{O}(M^2)$. A sufficiently large fixed c_8 excludes overflow. Outside this event, the explicit range check still prevents any oversized transmission.

Stopping is handled by multiplying an observation and its mean by the same predictable activity indicator. Concentration applies at its actual preparation before completeness is tested. On a fully generated window the stopped and intended sums agree. A union over the polynomial number of agents, arms, preparations, and endpoints gives total failure at most $\varepsilon/8$. The argument uses cross-agent loss independence only in this stochastic pooled calculation. □

Proof of Lemma 3.1. We use a first-failure induction. Suppose an incoming candidate set contains an optimal arm and its complete estimates have error at most $\tau/1024$. The estimated minimum differs from μ_* by at most $\tau/1024$, and each estimated gap differs from Δ_i by at most $2\tau/1024$. Every optimum passes the 4τ threshold. A survivor has $\Delta_i < 5\tau$, and a newly eliminated arm has $\Delta_i > 3\tau$ and $|\lambda_i - \Delta_i/4| \leq \tau/2048$. Thus it satisfies $\Delta_i/5 \leq \lambda_i \leq \Delta_i/3$. Previously stored positive coordinates are unchanged. Halving precision gives the stated incoming invariant. The new reference and gap vector lie inside their validity intervals. Copying one complete update to another tree preserves the same invariant at every arm, even though an individual agent's candidate sets need not be nested until its group is fixed.

Given such an incoming state and the current grid, the prediction bias of each assigned test is at most $e = \lambda_i + 16\tau \leq b_i^v$. Its successive selected observations are bounded and have conditional mean μ_i. Selection, including continued activity, is decided before that observation. For a deterministic selected count n and averaging coefficient $\eta = b^2/(64\Gamma + b^2)$,

$$Z_n - c = [1 - (1-\eta)^n](\mu_i - c) + \sum_{j=1}^{n} \eta(1-\eta)^{n-j}(X_j - \mu_i).$$

The squared weights sum to at most $\eta/(2-\eta) = b^2/(128\Gamma + b^2)$. Conditional Hoeffding bounds a centered fluctuation above $b/2$ by $2\exp(-64\Gamma - b^2/2) \leq 2e^{-64\Gamma}$. For analysis, independent

observations of mean μ_i are appended if selection terminates. The resulting process agrees with every actual selected observation, including a first crossing. A union bound over preparations, arms, and selected counts, always from accurate incoming histories, excludes a loss alarm before the first inaccurate state with failure at most $\varepsilon/16$.

The first uniform window has the required code range without an accurate reference. Thereafter the current invariant gives $|\mu_i - \widehat{\mu}| < 11\tau$ for candidates, so Lemma C.8 supplies range validity and the next accurate update. Scheduled delivery is complete before a first alarm by Lemma C.3. Therefore no range, numerical-state, or missing-record alarm can be the first failure either. The intersection of the geometry, window-concentration, and test events has total failure at most ε, giving accurate states and no alarms simultaneously. This proves both lemmas with the stronger internal probability $1 - \varepsilon$. □

A Common Window Calculation

Lemma C.9 (Doubled Windows)**.** *Every execution before shutdown has $\mathcal{O}(\log T)$ entered phases and fits all scheduled passes in its second windows. In the growing branch,*

$$\sum_s N_s \tau_s = \widetilde{\mathcal{O}}(\sqrt{K} + \sqrt{KT/V}).$$

On an accurate stochastic history, each fixed-group phase has

$$\frac{N_s}{|S_s|} \sum_{i \in S_s} \Delta_i = \widetilde{\mathcal{O}}(\sqrt{K} + H_\mu/V).$$

In the global branch replace these two right-hand sides by $\widetilde{\mathcal{O}}(q + \sqrt{KT/V})$ and $\widetilde{\mathcal{O}}(q + H_\mu/V)$, respectively. Here N_s is the planned duration clipped at the horizon. The actual generated duration can only be smaller.

Proof. Every successor window is at least twice its predecessor. During regrouping the predecessor length is common to all groups, so the same statement holds along each agent's history. Entering phase s requires completing phase $s-1$, proving the logarithmic count. The pass bound was verified in Lemma C.8. Before the cap, $W_s \tau_s = \mathcal{O}(\Lambda\sqrt{K})$, and candidate gaps on accurate histories with $s \geq 1$ are at most $10\tau_s$.

We give one calculation for either kind of fixed tree. Let $s_\star$ be its first fixed-phase index, let $q = V$ in Eq. (5) or $q = 2h + 3$ in Eq. (10), and let I be the inherited contribution to its first window. Thus $I = 2W_{s_\star - 1}$ on a transition into fixed groups, and $I = 0$ on a fresh fixed-tree start. With $k_s = |S_s|$, unrolling the recurrence gives

$$W_s \leq C \left(2^{s - s_\star}(I + \Lambda q) + \frac{\Lambda}{V \tau_s^2} \sum_{j=s_\star}^{s} 2^{j-s} k_j \right). \tag{25}$$

Ceilings are absorbed into Λq. For clarity, let A and B be the routing and sampling contributions on the right of Eq. (25), respectively. Then $N_s \leq 2(A + B)$ and $B\tau_s^2 \leq 2C\Lambda K/V$. If $N_s \leq 4A$, its width is at most $4A\tau_s$. Otherwise $N_s \leq 4B$, so $N_s \tau_s \leq 2\sqrt{N_s B \tau_s^2}$. In both cases,

$$N_s \tau_s \leq C'(I + \Lambda q)\tau_{s_\star} + C'\sqrt{\Lambda K N_s / V}.$$

Thus the routing and inherited contributions cost at most $C'(I + \Lambda q)\tau_{s_\star}$ per phase. This is $\mathcal{O}(\Lambda\sqrt{K})$ for growing groups, since $V\tau_{s_\star} \leq d_0/2$ and $W_{s_\star - 1}\tau_{s_\star} = \mathcal{O}(\Lambda\sqrt{K})$. It is $\mathcal{O}(\Lambda q)$ for the global branch, where $s_\star = 0$, $I = 0$, and $\tau_{s_\star} = 1/2$. Summation over the disjoint planned slots and Cauchy–Schwarz give the width bounds.

For the inverse-gap calculation, let $\overline{\Delta}_s = k_s^{-1} \sum_{i \in S_s} \Delta_i$. On a fixed tree the sets are nested. Writing $R_k = S_k \setminus S_{k+1}$ only in this proof,

$$\sum_{j=s_\star}^{s} 2^{j-s} k_j \leq 2k_s + \sum_{k=s_\star}^{s-1} 2^{k+1-s} |R_k|.$$

The routing and inherited terms in Eq. (25), multiplied by $\overline{\Delta}_s \le 10\tau_s$, have the same cost as above. Surviving positive-gap arms contribute at most

$$\frac{2k_s\overline{\Delta}_s}{\tau_s^2} \le 200 \sum_{i\in S_s:\Delta_i>0} \frac{1}{\Delta_i}.$$

Each arm removed at k contributes at most $2^{k+1-s}\overline{\Delta}_s/\tau_s^2 \le 20/\tau_k \le 200/\Delta_i$. These removal sets are disjoint and contain no optimal arm. Multiplication by $N_s/W_s \le 2$ proves the candidate bound. The first uniform phase satisfies $\Delta_i \le 1 < 10\tau_0$ as well. The same proof covers a fresh fixed-tree restart at any initial precision, with the routing contribution $\mathcal{O}(\Lambda q\tau_{s_\star})$ and no inherited window. □

Conditional Testing Costs The new grid is independent of the incoming state at its actual preparation. This permits the following direct conditional calculation.

Lemma C.10 (Testing in One Prepared Phase)**.** *Conditional on an accurate stochastic preparation before the phase grid is drawn, its S, λ, M, W, and planned duration $N \le T$ are fixed. Write $b_0 = 48\Gamma$, $L = 4096\Gamma^2$, and $F = 1 + c_1\Gamma^2 \log T$ for this calculation. For an eliminated arm of true gap Δ_i, its decreasing-test cost satisfies*

$$\mathbb{E}\left[\Delta_i \sum_{t=t_{s,g}}^{t_{s,g}+N-1} a^v_{i,\mathrm{dec}}(t) \,\middle|\, \textit{preparation}\right] \le \min\left\{\frac{b_0 N\Delta_i}{LK}, \frac{25b_0F}{H_i\Delta_i}\right\}. \tag{26}$$

Here the decreasing contribution is zero unless its cutoff is met. The global branch has $H_i = V$. The expected testing probability of each candidate in any round is at most $2b_0/(LK)$. In growing groups the constant-rate gap cost is, on each accurate execution, at most a logarithmic multiple of $\sqrt{K}\,N/T$.

Proof. Let $A = L(1 + K/H_i)$. Integration of $1/(A + x\lambda_i^2/(c_1\Gamma^2))$ over $0 \le x < N$ gives

$$\Delta_i \sum_{k=0}^{N-1} \frac{1}{A + k\lambda_i^2/(c_1\Gamma^2)} \le \min\left\{\frac{N\Delta_i H_i}{L(H_i+K)}, \frac{25F}{\Delta_i}\right\},$$

using $\lambda_i \ge \Delta_i/5$. The cutoff probability, conditional on this preparation, is $\min\{1, b_0/H_i\} \le b_0/H_i$. Multiplication proves Eq. (26).

For candidates $\lambda_i = 0$ and the numerical decreasing scale is fixed at $H_i = 1+2M$ in a growing group, or V globally. Its expected initial mass is at most $b_0/[L(H_i + K)] \le b_0/(LK)$. The constant-rate envelope has cutoff probability at most b_0/r and rate at most $r/[L(r+K)]$, giving the other $b_0/(LK)$. Counting the unused constant contribution at a decreasing tester only enlarges this bound.

Finally, there are at most $1 + b_0K/r$ arms under the nearby cutoff. By Eq. (22),

$$\sum_i \lambda_i N\gamma_i \mathbf{1}\{r\xi_i^v \le b_0\} \le (1 + b_0K/r)\frac{64c_1\Gamma^2\sqrt{K}\,N}{B(r)^2} = \widetilde{\mathcal{O}}(\sqrt{K}\,N/T). \tag{27}$$

On an accurate eliminated state $\Delta_i \le 5\lambda_i$. This proves the asserted stochastic constant-rate cost pathwise. □

The conditional expectation in Eq. (26) is taken before the fresh grid, not before a future test outcome. If an alarm interrupts a phase, its planned-prefix cost bounds its nonnegative stochastic cost. On the simultaneous good event these dominate all actual contributions. For an entered phase with an accurate past, this domination has a fixed incoming state, so one may drop the future-good-event indicator before applying the conditional bound. Earlier inaccurate histories are covered by the total failure contribution $T\varepsilon$.

C.2 Analysis of Individual Regret under Stochastic Losses

Let $\mathcal{E}$ be the joint stochastic event established above. Its complement has probability at most ε. On $\mathcal{E}$, the predictable regret increments are nonnegative and all states are accurate. Outside it, the total predictable regret is at most T.

The group construction has the properties needed for the following calculations. Growing groups have $M \geq r_s/K_0$, the windows satisfy $N = \widetilde{\mathcal{O}}(r_s)$, and fixed groups have $M \geq V/K_0$. Every member contributes exactly once, and the new grids are drawn after the actual group and state are fixed. Lemma C.2 and Lemma C.8 justify the routing schedule. No later argument requires an initial tree to be split or requires successive group memberships to be nested.

Candidate Actions For a nonempty candidate set S, write $\overline{\Delta}_S = |S|^{-1} \sum_{i \in S} \Delta_i$. In a prepared phase the uniform-candidate part of Eq. (1) costs at most $N\overline{\Delta}_S$. By Lemma C.10, the additional testing part has conditional expected cost at most

$$\frac{2b_0 N}{LK} \sum_{i \in S} \Delta_i \leq \frac{2b_0}{L} N \overline{\Delta}_S.$$

The phase window and candidate set were fixed before these grid draws. In a growing phase, $N\overline{\Delta}_S \leq 10N\tau = \widetilde{\mathcal{O}}(\sqrt{K})$. On a fixed group, Lemma C.9 instead bounds it by $\widetilde{\mathcal{O}}(\sqrt{K} + H_\mu/V)$. Summing the logarithmically many conditional phase bounds proves the inverse-gap candidate cost. For the gap-free cost, all candidate probability mass has gap below 10τ, so the pathwise width bound already controls it by $\widetilde{\mathcal{O}}(\sqrt{K} + \sqrt{KT/V})$.

Eliminated Arms If $H_i\lambda_i \geq \sqrt{K}$, then $H_i\Delta_i \geq 3\sqrt{K}$. The inverse bound in Eq. (26) costs $\widetilde{\mathcal{O}}(1/\sqrt{K})$ per arm and $\widetilde{\mathcal{O}}(\sqrt{K})$ per phase. In the other case, Eq. (21) gives $H_i \geq M$ and $\Delta_i \leq 5\sqrt{K}/M$. Before the groups freeze, the linear bound sums over all arms to $\widetilde{\mathcal{O}}(N\sqrt{K}/M) = \widetilde{\mathcal{O}}(\sqrt{K})$ because $N = \widetilde{\mathcal{O}}(r_s)$ and $M \geq r_s/K_0$. After they freeze, the inverse bound sums to $\widetilde{\mathcal{O}}(H_\mu/V)$ since $M \geq V/K_0$. The constant tests cost $\widetilde{\mathcal{O}}(\sqrt{K})$ over all disjoint slots by Eq. (27).

For the gap-free bound, all actions on arms of true gap at most ζ cost at most ζT. The remaining fixed-group inverse bounds total $\widetilde{\mathcal{O}}(K/(V\zeta))$. Choosing a balancing $\zeta \leq 1$, or using the trivial horizon cap when the balance exceeds one, yields $\widetilde{\mathcal{O}}(\sqrt{KT/V})$. Growing phases and the first scale case still cost $\widetilde{\mathcal{O}}(\sqrt{K})$.

Initialization and the Actual Population A first construction requiring attachment has $r_0 = d_0 < V$. Its construction and first uniform phase both cost $\widetilde{\mathcal{O}}(\sqrt{K})$. Adding $T\varepsilon \leq 1$ gives, for every agent,

$$R^v(T) = \widetilde{\mathcal{O}}\left(\min\left\{T, \sqrt{K} + \min\{H_\mu/V, \sqrt{KT/V}\}\right\}\right). \tag{28}$$

This is the stochastic bound in Theorem 4.1. Every member supplies an observation to every complete group estimate, including when $V > T$. If all gaps vanish, the actual stochastic regret is zero.

C.3 Analysis of Individual Regret under Adaptive Losses

Lemma C.11 (Weighted Certificates and Testing Price)**.** *For a fixed comparator i, every growing-group phase prefix of length $z \leq N$ with $s \geq 1$ before the target's alarm has predictable regret at most*

$$64\tau z + 7 \sum_{t=t_{s,g}}^{t_{s,g}+z-1} \sum_j \lambda_j a_j^v(t) + CB(r) \tag{29}$$

on the coverage and concentration events. For every valid state and every grid, the decreasing part of the testing sum is $\widetilde{\mathcal{O}}(\sqrt{K} + \sqrt{z} + \sqrt{Kz/M})$ and the constant part is $\widetilde{\mathcal{O}}(\sqrt{K}\, z/T)$. In a global group the allowance in Eq. (29) is $C(B(V) + q)$ and its testing sum is $\widetilde{\mathcal{O}}(\sqrt{z} + \sqrt{Kz/V})$.

Proof. For an eliminated arm, the prefix sums of $f_t = \ell_t(j) - U_j - \lambda_j$ are at most a common allowance $CB(r)$. Its probability is nonincreasing, so summation by parts bounds $\sum_t p_t^v(j) f_t$ by this allowance times $p^v_{t_{s,g}}(j)$. For a candidate, $f_t = \ell_t(j) - U_j$ is controlled by the suffix certificates. Its probability is nondecreasing, so the charge is the allowance times $p^v_{t_{s,g}+z-1}(j)$. The sum of endpoint weights is

$$\sum_{j \notin S} p^v_{t_{s,g}}(j) + \sum_{j \in S} p^v_{t_{s,g}+z-1}(j) \le 1 + \sum_j a_j^v(t_{s,g}) \le 17/16.$$

Thus there is no factor K. Also $\sum_j p_t^v(j) U_j = \widehat{\mu} + 32\tau + 6\sum_j \lambda_j a_j^v(t)$. Subtracting the fixed comparator's lower certificate uses the baseline $L_i = \widehat{\mu} + 2\lambda_i - 32\tau$. Its $\lambda_i z$ allowance consumes only half the negative $2\lambda_i z$ baseline difference. Dropping the remainder yields Eq. (29). The global argument is identical using its allowance from Lemma C.7.

For the pathwise testing bound put $b_0 = 48\Gamma$, $L = 4096\Gamma^2$, and $F = 1 + c_1\Gamma^2 \log T$ within the proof. One decreasing arm with $A_i = L(1 + K/H_i)$ has stored-gap price

$$\lambda_i \sum_{k=0}^{z-1} \frac{1}{A_i + k\lambda_i^2/(c_1\Gamma^2)} \le \min\{z\lambda_i/A_i, F/\lambda_i\} \le \sqrt{Fz/A_i}.$$

Zero gaps have zero price. The realized grid is ordered only for this calculation. The first point costs at most $\sqrt{Fz/L}$. At a remaining position $j \ge 1$, selection implies $H_i \le b_0 K/j$. If $H_i\lambda_i \ge \sqrt{K}$, its price is at most $Fb_0\sqrt{K}/j$. Summing costs $\widetilde{\mathcal{O}}(\sqrt{K})$. In the other case $M \le H_i \le 3M$, so at most b_0K/M remaining positions are selected, each with price at most $\sqrt{3FzM/[L(M+K)]}$. Their total is $\widetilde{\mathcal{O}}(\sqrt{Kz/M})$. Eq. (27) supplies the constant part. These are deterministic bounds, including when the adaptive loss law knows the new grid.

Globally $H_i = V$ for every assigned arm. Multiplying $\sqrt{Fz/[L(1+K/V)]}$ by the grid count $1 + b_0K/V$ gives the stated $\widetilde{\mathcal{O}}(\sqrt{z} + \sqrt{Kz/V})$ bound. □

The identity Eq. (23) applies with the indicator that the target still uses stochastic mode, before the good event is separated. Summing Eq. (29) over its actual phase sequence gives the total contribution. The associated horizon-clipped planned slots are disjoint, even when a final slot is interrupted. Since $z \le N$, each nonnegative prefix price is bounded by its value at the planned duration N. By Lemma C.9, Lemma C.11, and $M \ge r$, the nonnegative prices total

$$\widetilde{\mathcal{O}}\big(\sqrt{K} + \sqrt{KT/V} + \sqrt{KT/r}\big) = \widetilde{\mathcal{O}}(B(r)).$$

Here $r \le V$ and $r \le T$ allow $B(r)$ to absorb the other terms. The terminal crossing and the last uncertified communication intervals were already paid in Lemma C.7. The bad event contributes at most $T\varepsilon$. First-tree construction and uniform initialization cost $\widetilde{\mathcal{O}}(\min\{T, r_0 + K/V\})$ and fit the same final bound.

On the first-partition event each retained tree has $m \ge r_0/K_0$ and $h = \mathcal{O}(r_0\Gamma)$. All its full members use Algorithm 3 after their local alarms. These are the original fixed trees, not the larger statistical trees. Regrouping has changed neither their populations nor their heights. The certificate radii are still measured on the expanded agent-level group trees, so the J rounds allocated to each quotient step do not enter the alarm allowance $B(r)$. Proposition C.12 applies conditionally at the target's actual activation, against the same comparator. As $r \ge r_0/(2K_0)$, the prefix and continuation together give

$$R^v(T) = \widetilde{\mathcal{O}}\left(\min\{T, \sqrt{(r_0 + K/r_0)T}\}\right).$$

The horizon cap applies to the total regret, after setup, the pre-alarm prefix, and the conditional continuation are combined. In particular, a setup cost bounded by $C\Gamma^c r_0$ contributes at most $\min\{T, C\Gamma^c r_0\} \le \sqrt{C\Gamma^c r_0 T}$. The incomplete blocks and alarm interval are already included in Proposition C.12, not added again. The following inequalities hold for every horizon

$$\min\{T, x\} \le \sqrt{xT}, \qquad \min\{T, x + \sqrt{(1+x)T}\} \le 2\sqrt{(1+x)T}, \qquad x \ge 0. \tag{30}$$

If $T \geq d_0$, $r_0 = \min\{d_0, V\}$ gives $r_0 + K/r_0 \leq C(\sqrt{K} + K/V)$. If $T < d_0 \leq 3\sqrt{K}$, then $T \leq \sqrt{3\sqrt{K}\,T}$. Consequently

$$R^v(T) = \widetilde{\mathcal{O}}\left(\min\{T, \sqrt{KT/V} + K^{1/4}\sqrt{T}\}\right). \tag{31}$$

Every conditional continuation is analyzed on its actual retained tree, not conditional on later successful observations. Failed initial geometry is charged by its probability and the horizon bound. Only after combining expectations is the maximum over comparator arms taken.

Proposition C.12 (Conditional Tree Regret)**.** *Let $\sigma_v \leq T$ be a member's first action of the tree learner on a ready tree. Suppose either all members activate together, or activation is triggered by an alarm reaching every member within $2h + 2$ rounds. All activated members continue this learner. Conditionally on the actual full history before σ_v, for every fixed arm i, Algorithm 3 with Eq. (19) satisfies*

$$\mathbb{E}\left[\sum_{t=\sigma_v}^{T}\left(X_t^v(I_t^v) - X_t^v(i)\right)\,\middle|\,\mathcal{F}_{\sigma_v-1}\right] \leq 32\min\{T, \sqrt{(K/m+q)T\log K}\}.$$

The assertion is uniform over all admissible adaptive continuations, including correlated current losses. Every packet obeys the stated size bound on every execution.

Lemma C.13 (Complete Blocks Align the Bases)**.** *Let t_* be the first public block start at or after every member has activated, or $T + 1$ if none occurs by the horizon. Before t_*, no successful update occurs and every active parity base is uniform. From t_* onward, every block is complete and all members have the same two parity bases. Their action distributions can differ through their local zero counters. The interval from any member's activation to t_*, or to the horizon, has length at most $3q$.*

Proof. Induction on disjoint subtrees shows that a root aggregate is complete exactly when every member was active in every round of its source block. This condition depends on activation, not on a loss or proxy value. Every block starting before t_* has a missing member-round and supplies no update. Each newly joining member initializes uniform bases. From t_* onward, all counts are complete, and the fixed slots deliver the same aggregate before the next update of its parity. The two bases consequently remain identical, even though local sampling corrections need not coincide.

The alarm transit takes at most $2h + 2$ rounds, the next action at most one more round, and the next block boundary fewer than q additional rounds. Their sum is less than $3q$. Whether all members are active is measurable before a block's first action in the full filtration. Therefore t_* is a stopping time for the block analysis. We condition at that actual start and never condition on a later favorable count or loss. If no complete block starts by T, the same bound covers the entire remaining interval. □

Lemma C.14 (Adaptive Block Estimates and Local Correction)**.** *We consider a complete block of length $n \leq q$ with base w fixed at its start. Under the shared predictable-mean assumption, the estimate in Eq. (17) satisfies*

$$\mathbb{E}\left[\sum_i w(i)G_b(i)^2\,\middle|\,\textit{block-start history}\right] \leq n(4K/m + n). \tag{32}$$

For each member, its additional loss relative to playing the base satisfies, on every execution,

$$\sum_{t \text{ in the block}} \langle p_t^v - w, \ell_t\rangle \leq 4\eta n(n-1).$$

Proof. The number of retained zero proxies before a round is at most the number of preceding rounds of its block. Thus Eq. (16) and $\eta q \leq 1/8$ give $p_t^v(i) \geq w(i)/2$. Retention is therefore a valid probability. Conditional on the past and the current loss array, the probability of selecting and retaining arm i is exactly $w(i)/4$.

Let $Y_{t,i}^z$ indicate a retained one proxy and define, for analysis,

$$\widehat{\ell}_{t,i} = \frac{4}{mw(i)}\sum_{z=1}^{m} Y_{t,i}^z.$$

Fresh action and auxiliary coins give

$$\mathbb{E}_{t-1}[\widehat{\ell}_{t,i}] = \ell_t(i), \qquad \mathbb{E}_{t-1}[\widehat{\ell}^2_{t,i}] \le \frac{4}{mw(i)} + 1.$$

Indeed, a diagonal indicator has conditional expectation at most $w(i)/4$. For distinct members, conditional independence of their fresh coins given the loss array gives

$$\mathbb{E}_{t-1}[Y^z_{t,i} Y^{z'}_{t,i}] = \frac{w(i)^2}{16} \mathbb{E}_{t-1}[X^z_t(i) X^{z'}_t(i)] \le \frac{w(i)^2}{16}.$$

No independence of current losses across agents is used. For two rounds $s < t$ of the block, nonnegativity and $\ell_t(i) \le 1$ imply

$$\begin{aligned}\mathbb{E}[\widehat{\ell}_{s,i}\widehat{\ell}_{t,i} \mid \text{block-start history}] &= \mathbb{E}[\widehat{\ell}_{s,i}\ell_t(i) \mid \text{block-start history}] \\ &\le \mathbb{E}[\widehat{\ell}_{s,i} \mid \text{block-start history}] \le 1.\end{aligned}$$

This argument allows ℓ_t to depend on all earlier actions and observations. Expanding $G_b(i) = \sum_t \widehat{\ell}_{t,i}$ and summing with weights $w(i)$ yields a diagonal contribution at most $n(4K/m+1)$ and a cross-time contribution at most $n(n-1)$, proving Eq. (32).

For the local correction, let $z = \sum_i U^v_i$ just before a round. Since $U^v_i \ge 0$ and $\ell_t(i) \in [0,1]$,

$$\langle p^v_t - w, \ell_t \rangle = \frac{8\eta\left(\langle U^v, \ell_t\rangle - z\langle w, \ell_t\rangle\right)}{1 + 8\eta z} \le 8\eta z.$$

At the jth round of the block, $z \le j-1$. Summing proves the second assertion. In contrast to the sharper fixed-mean argument, this bound is pathwise and requires no restriction on how the conditional means change. □

Proof of Proposition C.12. The argument first conditions on the actual history at $t_* \le T$ from Lemma C.13. For a positive base w, nonnegative estimate G, and update $w^+_i \propto w_i e^{-\eta G_i}$,

$$\langle w, G\rangle - G_i \le \frac{-\log w_i + \log w^+_i}{\eta} + \frac{\eta}{2}\sum_j w_j G^2_j.$$

This follows from $e^{-x} \le 1 - x + x^2/2$ for $x \ge 0$ and $\log x \le x - 1$. It does not require bounded importance-weighted estimates. The logarithms telescope separately for the two bases, starting uniformly, and contribute at most $2\log K/\eta$.

The conditional mean of $\widehat{\ell}_{t,i}$ identifies the expected comparator difference of the base with the corresponding estimated difference. Lemma C.14 adds the loss of the local correction. Since the total number of rounds is at most T and $n \le q$ in each block, their sum is at most

$$\frac{2\log K}{\eta} + \eta\left(\frac{9}{2}q + \frac{2K}{m}\right)T \le \frac{2\log K}{\eta} + \frac{9}{2}\eta(q + K/m)T.$$

The identity Eq. (23) converts these conditional-mean differences into each member's expected individual loss difference. Updates scheduled after T are virtual and use only counts generated by the horizon. Substituting Eq. (19) gives

$$16q\log K + \frac{17}{2\sqrt{2}}\sqrt{(q+K/m)T\log K} \le 16q\log K + 7\sqrt{(q+K/m)T\log K}.$$

The initial interval adds at most $3q$, and the direct bound by T also applies. If $T \ge (q+K/m)\log K$, then both $q\log K$ and $q\log 2$ are bounded by constant multiples of the displayed square-root term. More precisely, $3q \le (3/\log 2)q\log K$, so the total is at most $28\sqrt{(q+K/m)T\log K}$. If $T < (q+K/m)\log K$, the square-root term exceeds T and the horizon bound suffices. The constant 32 covers both cases. Finally, the tower property returns the conditioning to the history before σ_v. Every step retains the same fixed comparator. This proves the proposition. □

The Standalone Adaptive Upper Bound The standalone adversarial algorithm uses only the first construction in Appendix A.1 and the tree learner in Algorithm 3. It does not use stochastic phases or loss tests. The construction scale and its logarithmic factor are

$$r_0 = \lceil 2\sqrt{K} \rceil, \qquad \Gamma = \log(2^{11} KVT^8).$$

The constant in the construction is chosen sufficiently large that the simultaneous geometry event in Lemma C.1 fails with probability at most $1/(2T^2)$. This is the original confidence schedule with $\delta = 1/2$ fixed. If $r_0 \geq V$, then $D + 1 \leq r_0$ and Lemma C.17 gives the spanning outcome deterministically, so no population lemma is needed. Otherwise, the ball-count estimate in Lemma C.1 bounds the failure probability for this one call by

$$V \exp\left(-e^{-3}(K_0 - 1)\right) + V^2 \exp(-R/r_0), \qquad K_0 = \lceil c_3 \Gamma \rceil, \quad R = \lceil c_3 r_0 \Gamma \rceil.$$

The extra possible cluster from advancing vertex 1 is included in this bound. Taking $c_3 \geq 64$, or increasing the fixed construction constant if other requirements demand it, makes the display at most $1/(2T^2)$. All initial shifts are drawn before the construction. This event concerns only those shifts and the fixed graph.

We first state the guarantee when the adversarial learner is used from the beginning, without stochastic phases or tests.

Theorem C.15 (Adaptive Adversarial Upper Bound)**.** *Under adaptive losses, for every connected $G = ([V], E)$, $K \geq 2$, and $V, T \geq 1$, one call of the construction in Appendix A.1 at scale $\lceil 2\sqrt{K} \rceil$, followed by Algorithm 3 with Eq. (19), gives every agent v*

$$R^v(T) = \widetilde{\mathcal{O}}\left(\sqrt{\frac{KT}{V}} + \sqrt{\min\{D, \sqrt{K}\}T}\right).$$

Every packet contains $\mathcal{O}(K \log V)$ bits on every execution. A singleton sends no packets.

The bound includes group construction and all actions during communication, so no additional setup term is needed for any horizon. This standalone result also identifies the rate of the continuation used by the best-of-both-worlds algorithm.

Proof of Theorem C.15. The agent v and comparator arm i are fixed before taking expectations. The tree lemma Proposition C.12 applies whenever all members activate together, without requiring an alarm. Indeed, its complete-block alignment proof then needs only the next public block boundary, and the conditional estimate and potential arguments are unchanged. Thus, on a ready tree with population m and height h, the continuation has conditional expected regret at most

$$32 \min\{T, \sqrt{(K/m + q)T \log K}\}, \qquad q = 2h + 3, \qquad \eta = \frac{1}{8q + \sqrt{2(q + K/m)T/\log K}}.$$

This is a bound at the actual starting history, uniformly over every admissible adaptive continuation. In particular, the losses after construction may depend on all preceding actions and messages. All members start at the same public block boundary, so the standalone algorithm has no incomplete initial learning block. The conditional bound also permits the asynchronous activation used later by the best-of-both-worlds algorithm.

For $V = 1$, the tree has $m = 1$, $h = 0$, and $q = 3$. Setup takes $\mathcal{O}(1)$ rounds, so the displayed bound gives $\widetilde{\mathcal{O}}(\sqrt{KT})$, including horizons ending during setup. This is the asserted rate because $D = 0$. The remaining argument assumes $V \geq 2$, hence $D \geq 1$.

Population and Height If the first-arrival census certifies the whole graph before attachment, then $m = V$. By Lemma C.17, its root is 1, its height satisfies $h \leq \min\{D, 8r_0 + R\}$, and its construction time is $\mathcal{O}(h)$. Waiting for the common block boundary adds fewer than q rounds. Since $r_0 \leq 3\sqrt{K}$, both its total setup length and q are at most

$$\mathcal{O}\left(\Gamma \min\{D, \sqrt{K}\}\right).$$

If attachment is required, Lemma C.17 implies $D+1 > r_0$. Together with $D+1 \le V$, this gives $r_0 < V$ and $D \ge \sqrt{K}$. On the geometry event, every completed tree has $m \ge r_0/K_0$ and $h = \mathcal{O}(r_0\Gamma)$. Its construction and block-alignment costs are also $\mathcal{O}(r_0\Gamma)$. Consequently, in both outcomes,

$$\frac{K}{m} + q = \mathcal{O}\left(\Gamma\left(\frac{K}{V} + \min\{D, \sqrt{K}\}\right)\right). \tag{33}$$

This distinction refers to the first-arrival census. If attachment later creates a tree containing all V agents, that tree still uses its actual height and the attachment bound. It is not treated as an early spanning-tree outcome. Every completed tree uses the same learner and the same rate formula.

Setup and Every Horizon The intended construction can be completed for analysis using its already drawn topology randomness, even if its scheduled end lies beyond T. This generates no additional loss observations. The preceding bounds show that the number of uniform setup actions is at most

$$\min\left\{T, \mathcal{O}\left(\Gamma \min\{D, \sqrt{K}\}\right)\right\}.$$

For any $x \ge 0$, $\min\{T, x\} \le \sqrt{xT}$. Therefore the setup contribution is absorbed by $\widetilde{\mathcal{O}}\left(\sqrt{\min\{D, \sqrt{K}\}T}\right)$, even when no learning block starts by the horizon. There is no horizon restriction and no remaining additive setup term.

At each actual learning start, the conditional tree bound combines with Eq. (33) and is averaged over the construction history. The geometry event is determined by the initial topology randomness, so this averaging does not condition the learning bound on future observations. Its complement contributes at most $T/(2T^2) \le 1$, which is absorbed by $\sqrt{T}$. Including the setup contribution gives, for the fixed comparator,

$$\mathbb{E}\left[\sum_{t=1}^{T}\left(X_t^v(I_t^v) - X_t^v(i)\right)\right] = \widetilde{\mathcal{O}}\left(\sqrt{KT/V} + \sqrt{\min\{D, \sqrt{K}\}T}\right).$$

The argument is uniform in i. Taking the maximum after expectation proves the individual-regret assertion.

Communication and Initial Knowledge The construction uses only local neighborhoods and received messages, with the bounded records established in Appendix A.1. During learning, each pooled count is at most $mq \le 3m^2$. Thus Appendix A.3 gives $\mathcal{O}(K \log V)$ bits per packet on every execution, and a singleton sends no packets. Both construction and learning remain uninformed and respect one-edge-per-round propagation. □

The Full-Population Group

Proposition C.16 (Learning on a Fixed Global Tree)**.** *On a spanning tree of height h with a common starting round, set $q = 2h+3$. With all V members, Eq. (10), and $r = H_i = V$ in Eq. (2), every agent has*

$$R^v(T) = \widetilde{\mathcal{O}}\left(\min\{T, \sqrt{(K/V+q)T}\}\right) \qquad \textit{adversarially}, \tag{34}$$

$$R^v(T) = \widetilde{\mathcal{O}}\left(\min\{T, q + \min\{H_\mu/V, \sqrt{KT/V}\}\}\right) \qquad \textit{stochastically}. \tag{35}$$

The adversarial bound holds conditionally after any actual starting history against each fixed comparator, using the original public horizon T. Stochastic sending rounds are $\mathcal{O}(\log T)$ simultaneously with probability at least $1-\delta$ and in expectation per member. Every packet has $\mathcal{O}(K \log V)$ bits on every execution.

Proof. All local state, estimation, coding, and test rules are exactly those in Appendix A, with actual population $M = V$, decreasing scale $H_i = V$, and no constant-rate tests. Grids are freshly drawn

after each incoming state. Thus Lemma C.8 and the first-failure induction prove the same accuracy and silence assertions. The root never changes, so no absolute state is transmitted after initialization.

For stochastic regret, the global cases of Lemma C.9 and Lemma C.10 apply. The uniform and testing candidate contributions in one phase are bounded in conditional expectation by a constant multiple of $N\overline{\Delta}_S$, giving $\widetilde{\mathcal{O}}(q + H_\mu/V)$ in total. Each eliminated arm contributes at most $\widetilde{\mathcal{O}}(1/(V\Delta_i))$ per phase. For the gap-free bound, candidate actions cost at most $10\sum_s N_s\tau_s$, and eliminated arms split at a fixed true gap ζ cost at most $\zeta T + \widetilde{\mathcal{O}}(K/(V\zeta))$. Balancing gives $\widetilde{\mathcal{O}}(q + \sqrt{KT/V})$. This proves Eq. (35), including the uniform first phase and the failure contribution $T\varepsilon$.

For adversarial regret, the global certificates and weighted argument give a phase allowance $C(B(V)+q)$ plus its width and testing prices. Their sums are bounded by $\widetilde{\mathcal{O}}(q + \sqrt{T} + \sqrt{KT/V})$ using the global cases of Lemma C.9 and Lemma C.11. The uniform first phase has regret at most its length, which is twice its width price. The predictable pre-switch indicator is retained before splitting the good event, and Proposition C.12 supplies the continuation on this same tree for the same fixed comparator. The total bound is capped by T after these contributions are added. When $q \le T$, $q \le \sqrt{qT}$ absorbs the linear routing cost. When $q > T$, $T \le \sqrt{qT}$ does so instead. The same argument applies when the first window or the first complete fallback block does not finish by T. Maximizing after expectation proves Eq. (34).

There are logarithmically many doubled windows and a constant number of passes per phase. No fallback communication occurs on the stochastic event. On its complement at most T sending rounds add $T\varepsilon$ to expectation. The residual and block-count integer ranges give the packet bound deterministically. □

The First Construction and the Main Bound

Lemma C.17 (Sparse Tree Certification)**.** *The first construction in Appendix A.1 uses only each vertex's own index, its known neighbors, and received messages. Its sending count is $\mathcal{O}(1)$ per vertex and its packet size is $\mathcal{O}(\log V)$ bits, deterministically. A first-arrival census equal to V certifies a spanning BFS tree rooted at 1, with height $h \le \min\{D, L\}$. Its common learning start occurs after $\mathcal{O}(q)$ rounds, where $q = 2h + 3$. If $D + 1 \le r_0$, this spanning outcome occurs deterministically. If the first-arrival census is nonspanning, its trees proceed to attachment without restarting the construction.*

Proof. After all known neighbors have declared their fixed labels and parents, a vertex knows its children. Subtree reports count disjoint sets of vertices. Their sum equals V exactly when this first-arrival tree spans the graph. The extra bit described in Appendix A.4 gives the same certificate when only a population upper bound is known.

Vertex 1 sends its own label in round 1 and never changes it. Thus any spanning first-arrival tree has root 1. In that event no other label is accepted anywhere, so its depths are shortest-path distances from 1. A vertex at depth d receives every neighbor declaration by round $d + 3$, because adjacent BFS depths differ by at most one. An induction from leaves then gives the complete root report by round $2h + 4$. If its report arrives at round s, the broadcast of a common start at $s + 2h + 4$ reaches every member before that start. Every vertex accepts a label by its own scheduled release, which is no later than $L + 1$. On a spanning outcome this gives $h \le L$ for every scale $r_0 \ge 1$, including $r_0 > V$. Both the announcement and the start therefore precede $8(L + 2)$. All vertices consequently agree whether the fixed attachment slot is needed. The start field is bounded by $4V + 4$, even if the public L is larger than V.

Every other release is at least $1 + 8r_0$. If $D + 1 \le r_0$, label 1 reaches every vertex before any other label can be released. The preceding census therefore certifies the whole graph and finishes in $\mathcal{O}(q)$ rounds. This is $\mathcal{O}(D)$ when $V \ge 2$ and $\mathcal{O}(1)$ for a singleton. If this does not happen, the declarations and censuses still provide all inputs needed by the attachment step. There is no second first-arrival partition. Each vertex sends at most a declaration, a census report, and a forwarded metadata message before attachment. All tree records contain only a constant number of indices, bounded counts, depths, and bits. These facts prove the communication and uninformed-setting assertions. □

Proof of Theorem 4.1. We prove the upper bounds in Table 3, which imply the main worst-graph bounds after allowing the diameter to vary. For $T = 1$, the uniform action has regret at most one,

which is no larger than the claimed rate, and no messages are sent. Hence assume $T \geq 2$. For $V = 1$, the first tree has $h = 0$, $q = 3$, and constant setup time. Proposition C.16 gives the claimed stochastic rate and adaptive rate $\widetilde{\mathcal{O}}(\sqrt{KT})$, including a horizon ending during setup. Thus the remaining proof concerns $V \geq 2$, where $D \geq 1$. For the stochastic fixed-diameter bound, any positive gap gives $H_\mu \geq 1$, which absorbs constant costs when $V = 1$. If all gaps vanish, the regret is zero. The adaptive fixed-diameter bound likewise needs no additive unit inside its propagation term. For $V \geq 2$, $\min\{D, \sqrt{K}\} \geq 1$ absorbs that unit, while for $V = 1$ its $\sqrt{T}$ contribution is absorbed by $\sqrt{KT/V}$.

The first construction uses no losses. On a spanning first-arrival outcome, its setup length and $q = 2h + 3$ are at most

$$\mathcal{O}\left(\min\{D, L\}\right) = \widetilde{\mathcal{O}}\left(\min\{D, \sqrt{K}\}\right),$$

since $r_0 \leq d_0 \leq 3\sqrt{K}$. Proposition C.16 applies at the actual common start. The stochastic setup and routing cost is $\widetilde{\mathcal{O}}(\min\{D, \sqrt{K}\})$, and Eq. (35) retains the same statistical term. Adversarially, its setup cost is absorbed by the square-root term using Eq. (30), so the theorem has no additional additive term.

Suppose a nonspanning first construction finishes before T. Its completion time exceeds r_0, so $r_0 < T$. By Lemma C.17, $D + 1 > r_0$. Since $D + 1 \leq V$, also $r_0 < V$. Consequently $r_0 = d_0$ and $D \geq \sqrt{K}$. The growing-group bounds Eq. (28) and Eq. (31) have exactly the required orders, including the cost of this first construction.

It remains to cover an initialization cut short by the horizon. If its completed first-arrival outcome would be spanning, the intended common start is $\mathcal{O}(\min\{D, L\})$. If it would be nonspanning, then $D + 1 > r_0$ and the intended completion is $\mathcal{O}(L + r_0)$. In either case the uniform prefix is at most $\widetilde{\mathcal{O}}(\min\{D, \sqrt{K}\})$ and at most T. Its stochastic cost is $\widetilde{\mathcal{O}}(\min\{D, \sqrt{K}\})$, while Eq. (30) absorbs its adversarial cost in $\widetilde{\mathcal{O}}\left(\sqrt{\min\{D, \sqrt{K}\}T}\right)$.

Conditional concentration is applied at the actual preparations, with the allocated failure probabilities. Adding the initial constant number of sends preserves the logarithmic stochastic count. Exceptional paths add at most $T\varepsilon \leq 1$ in expectation. All concatenated packets obey the deterministic integer range bounds. The fixed-comparator prefix and continuation are combined before taking the maximum over comparators, completing the proof. □

Combining the two adaptive branches yields a universal constant C and a fixed exponent c such that

$$R^v(T) \leq C\Gamma^c \min\left\{T, \sqrt{KT/V} + \sqrt{\min\{D, \sqrt{K}\}T}\right\}$$

for every $K \geq 2$ and $V, T \geq 1$. For $V \geq 2$, $D \geq 1$, so a square-root term absorbs the failure contribution $T\varepsilon \leq 1$ and any constant rounding cost. For $V = 1$, the term $\sqrt{KT/V}$ does so instead. Initial waiting is bounded by its actual duration clipped at T, and Eq. (30) covers horizons ending before construction or a complete learning block.

C.4 Communication Cost Analysis

There is one initial-tree construction. Every later reconstruction on $G_{\mathcal{Q}}$ uses $\mathcal{O}(1)$ sending rounds per agent over the entire reconstruction by Lemma C.2, even though it contains many silent communication steps. Each phase also uses a fixed number of sparse tree passes for statistics and state delivery. A vertex belongs to one fixed initial tree and one current statistical group, and may prepare one successor group. These roles add only a constant number of record streams. Drawing a fresh phase grid has no communication cost. On the joint stochastic event there are no alarms or fallback packets, so Lemma C.9 gives $\mathcal{O}(\log T)$ sending rounds simultaneously for all agents. On its complement the count is at most T, adding at most $T\varepsilon \leq 1$ in expectation.

Residual and state records obey Appendix A.2's checked integer bounds. Graph fields use $\mathcal{O}(\log V)$ bits, and the fallback bound follows from Proposition C.12. Concatenating the constant number of records that can share an edge in a round changes packet length by a constant factor. Sending zeros,

forwarding alarms, constructing trees, and all actions during these operations are included. This completes the growing-branch guarantees.

D Stochastic and Adaptive Lower Bounds

D.1 Stochastic Individual-Regret Lower Bound

Write $H_\mu = \sum_{i:\Delta_i>0} \Delta_i^{-1}$. We prove the stochastic assertion of Theorem 4.2, with the inverse-gap budget fixed before choosing the hard instance. For $K \geq 3$, the construction can moreover be calibrated so that $H_\mu = H$ exactly; the lower bound remains a statement over the budget class.

The cooperative policy on the V-vertex path and target agent v are fixed throughout the proof. Every hard instance below has independent losses across agents, rounds, and arms, so it belongs to the stochastic model in Section 2. The path propagation argument uses the information constraint of Barnea et al. (2025), and the likelihood calculation is the standard bandit change of measure (Lattimore & Szepesvári, 2020). The allowed messages remain unrestricted.

Information Available to a Fixed Agent Before round t, the target is given all instance-independent policy random seeds and the actual actions and selected losses at positions

$$\mathcal{C}_t = \{(w,s) : 1 \leq s < t,\ s + \mathrm{dist}(w,v) \leq t\}, \qquad n_t = |\mathcal{C}_t|.$$

This enlarged transcript determines I_t^v. Indeed, if an observation at (z,r) influences an action at $(w,s) \in \mathcal{C}_t$, then $r < s$ and $r + \mathrm{dist}(z,w) \leq s$, and hence

$$r + \mathrm{dist}(z,v) \leq r + \mathrm{dist}(z,w) + \mathrm{dist}(w,v) \leq t.$$

It too lies in the transcript. Chronological induction reconstructs all necessary actions, messages, forwarding, and decisions based on silence. No unplayed loss is revealed. All actions taken during setup are included in the same transcript and horizon. On a path, the number of vertices within distance j is at most $2j+1$. Therefore

$$n_t \leq \min\{V(t-1), 3(t-1)^2\}, \qquad n_1 = 0. \tag{36}$$

For $t \geq 2$, the second estimate follows from $\sum_{j=1}^{t-1}(2j+1) = (t-1)(t+1) \leq 3(t-1)^2$.

A Family with Two Distinguished Arms We first assume that $K \geq 3$ and choose $0 < \Delta_0 \leq \Delta \leq 1/4$. For $j \in [K]$, let j^+ be its cyclic successor and define the Bernoulli instance $\mathcal{I}_j$ by

$$\mu_i^{(j)} = \begin{cases} 1/2 - \Delta, & i = j, \\ 1/2 - \Delta + \Delta_0, & i = j^+, \\ 1/2, & i \notin \{j, j^+\}. \end{cases} \tag{37}$$

It has a unique optimum, one gap Δ_0, and $K-2$ gaps Δ. Thus $H_{\mu^{(j)}} = 1/\Delta_0 + (K-2)/\Delta$ is the same for every j.

Let $\mathcal{I}_0$ have mean $1/2$ on every arm. For a fixed t, write P_j for the enlarged transcript's law, and let $N_i(t)$ count its arm-i observations under the reference. Exposing fresh observations in time order, conditional on the independent policy seeds, gives

$$\begin{aligned} \mathrm{KL}(P_0 \| P_j) &= \mathrm{kl}(1/2, 1/2-\Delta)\mathbb{E}_0[N_j(t)] + \mathrm{kl}(1/2, 1/2-\Delta+\Delta_0)\mathbb{E}_0[N_{j^+}(t)] \\ &\leq 4\Delta^2 \mathbb{E}_0[N_j(t) + N_{j^+}(t)]. \end{aligned}$$

Here $\mathrm{kl}(1/2, 1/2-x) = -\frac{1}{2}\log(1-4x^2) \leq 4x^2$ for $0 \leq x \leq 1/4$. Each chosen arm is determined by earlier exposed observations. Messages and seeds introduce no additional likelihood factor.

Pinsker's inequality applied to $\{I_t^v \in \{j, j^+\}\}$, followed by averaging and Cauchy–Schwarz, yields

$$\frac{1}{K}\sum_j \mathbb{P}_j(I_t^v \in \{j, j^+\}) \leq \frac{2}{K} + 2\Delta\sqrt{n_t/K}.$$

Every arm belongs to exactly two cyclic pairs under the same reference law, and $\sum_i N_i(t) = n_t$ on every execution. No symmetry of the policy is assumed.

Set, only for the proof,

$$T_\Delta = \max\{K/(V\Delta^2), \sqrt{K}/\Delta\}, \qquad L = \min\{T, \lceil T_\Delta/128 \rceil\}.$$

For $t \le L$, $t - 1 < T_\Delta/128$. The first bound in Eq. (36) applies when $T_\Delta = K/(V\Delta^2)$, and the second applies otherwise. In either case $n_t < K/(128\Delta^2)$. Since $K \ge 3$, the average probability of playing outside the low-loss pair is at least

$$1 - \frac{2}{3} - \frac{2}{\sqrt{128}} \ge \frac{1}{8}.$$

Such an action has gap Δ. The gap on the second arm is nonnegative and can be dropped. Since stochastic regret increments are nonnegative and $L \ge \min\{T, T_\Delta\}/128$, we obtain

$$\frac{1}{K}\sum_j R^v_{\pi,\mathcal{I}_j}(T) \ge 2^{-11}\min\{T\Delta, K/(V\Delta) + \sqrt{K}\}. \tag{38}$$

Before replacing the maximum by a sum, the constant is 2^{-10} and the second argument is $\max\{K/(V\Delta), \sqrt{K}\}$. The displayed bound follows from $\max\{x, y\} \ge (x + y)/2$.

Calibrating a Fixed Inverse-Gap Sum Suppose $K \ge 3$ and $H \ge 4(K - 1)$. Let r be the positive solution of $Tr = K/(Vr) + \sqrt{K}$, and choose

$$\begin{aligned} r &= \frac{\sqrt{K} + \sqrt{K + 4KT/V}}{2T}, \\ \Delta &= \max\left\{\frac{K-1}{H}, \min\{1/4, r\}\right\}, \qquad \Delta_0 = \left(H - \frac{K-2}{\Delta}\right)^{-1}. \end{aligned} \tag{39}$$

The assumptions imply $0 < \Delta_0 \le \Delta \le 1/4$. In particular, $H \ge (K - 1)/\Delta$ makes the denominator of Δ_0 at least $1/\Delta$. Every member of Eq. (37) has $H_\mu = H$ exactly.

We compare Eq. (38) with the right-hand expression in Theorem 4.2(i) without its numerical constant. If $r \ge 1/4$, then $\Delta = 1/4$ and monotonicity of $Tx - K/(Vx) - \sqrt{K}$ gives $T/4 \le 4K/V + \sqrt{K}$. Thus the minimum in Eq. (38) is $T/4$, at least one quarter of the desired expression. If $r < 1/4$ and $(K - 1)/H \le r$, then $\Delta = r$ and

$$T\Delta = K/(V\Delta) + \sqrt{K} = \frac{\sqrt{K} + \sqrt{K + 4KT/V}}{2} \ge \frac{\sqrt{K} + \sqrt{KT/V}}{2}.$$

If $r < 1/4$ and $(K - 1)/H > r$, then $\Delta = (K - 1)/H$ and the minimum equals

$$\frac{K}{K-1}\frac{H}{V} + \sqrt{K},$$

which is at least the desired expression. These cases show that Eq. (38) is at least 2^{-13} times the benchmark in Theorem 4.2(i), which is stronger than the stated constant.

Small Complexity and Two Arms For any $K \ge 2$ and $K - 1 \le H < 4(K - 1)$, set $\Delta = (K - 1)/H \in (1/4, 1]$. In instance $\mathcal{I}_j$, arm j has deterministic loss zero and every other arm has deterministic loss Δ. The inverse-gap sum is exactly H.

Couple all instances to the deterministic reference with loss Δ on every arm using the same policy seeds. If arm j never appears in the reference transcript $\mathcal{C}_t$, the transcript is unchanged under $\mathcal{I}_j$ and the target takes the same action. Thus among all choices of j, the target can choose the optimum in at most $n_t + 1$ cases. Averaging over the seeds,

$$\frac{1}{K}\sum_j \mathbb{P}_{\mathcal{I}_j}(I^v_t = j) \le \frac{n_t + 1}{K}.$$

Put $T_0 = \max\{K/V, \sqrt{K}\}$ and $L_0 = \min\{T, \lceil T_0/16 \rceil\}$. The same two cases in Eq. (36) give $n_t < K/16$ for $t \le L_0$, so the average error probability is at least $1 - 1/K - 1/16 \ge 1/4$. Hence

$$\frac{1}{K}\sum_j R^v_{\pi,\mathcal{I}_j}(T) \ge 2^{-8}\min\{T, T_0\}$$
$$\ge 2^{-11}\min\left\{T, \sqrt{K} + \min\{H/V, \sqrt{KT/V}\}\right\}.$$

For the last comparison use $H < 4K$ and $\min\{T, \sqrt{K} + 4K/V\} \le 4\min\{T, \sqrt{K} + K/V\}$.

Finally let $K = 2$ and $H \ge 4$. With r and Δ as in Eq. (39), the instance has a single low-loss arm with mean $1/2 - \Delta$ and one arm of mean $1/2$. Then $H_\mu = 1/\Delta \le H$. The one-arm likelihood calculation gives

$$\frac{1}{K}\sum_j \mathbb{P}_j(I^v_t = j) \le \frac{1}{K} + \Delta\sqrt{2n_t/K}.$$

The same T_Δ, L used above give error probability at least $1/8$ and therefore Eq. (38). The three calibration cases do not use the second low-loss arm, so the budget lower bound follows for $K = 2$ as well.

In every case, a single hard family member is selected after summing over all rounds. The comparator is that member's fixed optimal arm, and the construction is independent across agents and times. This proves the stochastic assertion of Theorem 4.2. The equality assertion is restricted to $K \ge 3$ because fixing $H_\mu = H$ with two arms fixes the only gap to $1/H$, for which regret is at most T/H.

A Fixed Diameter We now prove the stochastic lower bound in Proposition B.1. For $V \ge 2$, the graph is a path $v_0, \dots, v_{D-1}$ ending in a clique containing v_{D-1} and $V - D$ additional vertices. The target is v_0. For $V = 1$, the graph is a singleton. This choice depends only on (V, D).

The statistical part works on any fixed graph and target. A central observer receives all $V(t-1)$ previous selected losses and all policy seeds. In the independent Bernoulli family with mean $1/2 - \Delta$ on arm j and $1/2$ on every other arm, the same change-of-measure calculation as above gives

$$\frac{1}{K}\sum_{j=1}^{K} \mathbb{P}_j(I^v_t = j) \le \frac{1}{K} + \Delta\sqrt{2V(t-1)/K}.$$

For the first $\min\{T, \lceil K/(128V\Delta^2)\rceil\}$ rounds, the average error probability is at least $1/4$. Every error has gap Δ, and later stochastic regret is nonnegative. Consequently, for $0 < \Delta \le 1/4$,

$$\frac{1}{K}\sum_j R^v_{\pi,j}(T) \ge c\min\{T\Delta, K/(V\Delta)\}.$$

When $H \ge 4(K-1)$, the choice

$$\Delta = \max\left\{\frac{K-1}{H}, \min\left\{\frac{1}{4}, \sqrt{\frac{K}{VT}}\right\}\right\}$$

ensures $H_\mu = (K-1)/\Delta \le H$. If the square-root term is at least $1/4$, the preceding minimum is $T/4$. If it lies between $(K-1)/H$ and $1/4$, the minimum is $\sqrt{KT/V}$. In the remaining case the minimum is $KH/(V(K-1)) \ge H/V$. Thus the statistical lower bound is

$$c'\min\{T, H/V, \sqrt{KT/V}\}.$$

For $K - 1 \le H < 4(K-1)$, one arm has deterministic loss zero and all other arms have loss $(K-1)/H > 1/4$. Coupling to the constant-loss reference bounds the average success probability by $(V(t-1)+1)/K$. Summation through $\min\{T, \lceil K/(16V)\rceil\}$ gives $c\min\{T, K/V\}$, which implies the same statistical bound since $H < 4K$.

For propagation, one arm has deterministic loss zero and every other arm has loss one. This is a valid stochastic instance with $H_\mu = K - 1 \le H$. Its constants are independent across agents and rounds, even though their numerical values coincide. The reference-cone argument in Appendix E.2 gives $c \min\{T, D, \sqrt{K}\}$ on the same graph and target. The stochastic statistical instance instead uses losses that are independent across agents and rounds.

For any fixed policy, the harder of the statistical and propagation instances yields

$$R^v(T) \ge c'' \min\left\{T, \min\{D, \sqrt{K}\} + \min\{H/V, \sqrt{KT/V}\}\right\}.$$

Their maximum is at least half the capped sum. The same graph and target work for every H, completing the stochastic assertion of Proposition B.1. Choosing a sufficiently large finite H also gives its minimax version with statistical term $\sqrt{KT/V}$.

D.2 Adaptive Adversarial Individual-Regret Lower Bound

We prove the adaptive assertion of Proposition B.1 and its worst-graph consequence in Theorem 4.2. The graph and target are fixed before the policy is chosen, for each feasible population and diameter. Messages may have arbitrary contents and sizes. The construction below is adaptive. Together with the upper bound in Table 3, it yields nearly optimal (V, D)-worst-graph regret, where the lower bound is worst-case over graphs and target agents with the prescribed parameters.

The proof establishes the following explicit, horizon-capped fixed-diameter lower bound, with the graph and target quantifiers of Proposition B.1,

$$R^v(T) \ge \frac{1}{128} \min\left\{T, \sqrt{\frac{KT}{V}} + \sqrt{\min\{D, \sqrt{K}\}T}\right\}. \tag{40}$$

The construction first yields a local lower bound for every fixed graph and target, from which the path and fixed-diameter statements follow. Its loss vector is common to every agent and binary, and its only adaptivity is to earlier chosen arms. No packet-size or feedback-volume restriction is imposed. Actions taken during setup are included in the horizon and in the argument below.

An Adaptive Hard Family The graph and target vertex v are fixed throughout this section. Define

$$d_v(z) = \max\{1, \mathrm{dist}(z, v)\}.$$

The maximum with one accounts for an agent using its own round-s observation only in a later action. An observation at (z, s) could first influence the target's action at round $s + d_v(z)$.

For a gap $0 < \Delta \le 1/16$, instance $j \in [K]$ has

$$\theta_i = \tfrac{1}{2} - \Delta \mathbf{1}\{i = j\}.$$

The independent arrays $\{\psi_{i,k} : i \in [K], k \in [T]\}$ have entries $\psi_{i,k} \sim \mathrm{Bernoulli}(\theta_i)$. They are independent of every primitive policy random variable. The reference instance, denoted 0, uses $\theta_i = 1/2$ for every arm.

Each arm has a current version index k_i, a birth time τ_i, and an expiration time e_i. Its loss is shared by every agent and remains equal to ψ_{i,k_i} until that version expires. Initialization is $k_i = 1$, $\tau_i = 1$, and $e_i = \infty$.

Old pending arrivals are discarded on a refresh. Thus an old-version observation cannot refresh a newer version. Equivalently, an arm with current birth time τ_i is refreshed at the first round t for which some query of this version satisfies

$$\tau_i \le s < t, \qquad I_s^z = i, \qquad s + d_v(z) \le t. \tag{41}$$

At the beginning of that round, increment k_i, set $\tau_i = t$ and $e_i = \infty$, and use ψ_{i,k_i} for the new value. After the actions in round s, a current-version query by z updates $e_i \leftarrow \min\{e_i, s + d_v(z)\}$. Thus all loss values are selected before the current actions.

The expiration uses the *first possible* information-arrival time, not the actual time that the policy forwards the observation. Suppressing a packet, choosing an aggregate instead, or keeping silent does not postpone expiration. The construction does not censor any feedback. Agents still immediately observe their selected local loss and may forward any information they possess.

Admissibility At the start of round t, every refresh decision uses only earlier actions. The loss vector is chosen before the agents' fresh current action draws. The array is common across agents, so the common-conditional-mean assumption holds even when a version's value is already known elsewhere in the network. Under the randomized family, define $\ell_t(i) = \mathbb{E}[\psi_{i,k_i(t)} \mid \mathcal{F}_{t-1}]$. This is the required common predictable mean. It need not equal θ_i. After fixing all Bernoulli arrays, the loss vector is a deterministic function of the past action history. At most one refresh per arm per round is possible because $d_v(z) \geq 1$, so the finite arrays of length T suffice.

The Target's Transcript and Freshness Let $\mathcal{U}$ denote the policy's entire instance-independent primitive random tape. For lower-bound analysis, we enlarge the target's information before round t to

$$\mathcal{C}_{v,t} = \{(z,s) : 1 \leq s < t,\ s + d_v(z) \leq t\}, \qquad \mathcal{H}_{v,t} = \left(\mathcal{U}, \left(I_s^z, X_s^z(I_s^z)\right)_{(z,s)\in\mathcal{C}_{v,t}}\right). \tag{42}$$

The cone is a deterministic set of source–time positions. The primitive tape is included only for analysis, not as an assumption on the algorithm. Every observation is the actual selected-coordinate observation, not a full loss vector.

Lemma D.1 (Causal Replay and Unqueried Current Versions)**.** *For each t and each fixed instance j, the following hold.*

1. *The transcript $\mathcal{H}_{v,t}$ determines I_t^v and every current version index $k_i(t)$.*
2. *Conditional on $\mathcal{U}$, the transcript can be generated by adaptively querying the independent arrays $\psi_{i,k}$. Query addresses depend only on previously exposed array values. A repeated address exposes no new random variable.*
3. *For each arm i, exactly the labels $1, \ldots, k_i(t) - 1$ have been exposed in this transcript. In particular,*
$$\mathbb{E}_j[X_t^v(i) \mid \mathcal{H}_{v,t}] = \theta_i. \tag{43}$$

Proof. We separately check the policy's communication dependencies and the additional dependencies created by the adaptive adversary.

Communication closure. If an observation (w,r) can affect an action at $(z,s) \in \mathcal{C}_{v,t}$, then $r < s$ and $r + \mathrm{dist}(w,z) \leq s$. Since

$$d_v(w) \leq \mathrm{dist}(w,z) + d_v(z),$$

we have $r + d_v(w) \leq s + d_v(z) \leq t$. Thus (w,r) is also in $\mathcal{C}_{v,t}$. The same argument applies to messages that determine that action, including a message's content, the decision whether to send, and any decision based on silence after a known deadline. With the primitive tape fixed, all such dependencies can be replayed from earlier cone observations. In particular, I_t^v is determined by (42).

Closure of the refresh rule. For any round $s \leq t$, the version indices at its start depend only on earlier refresh decisions and on actions (w,r) with $r + d_v(w) \leq s$. These positions belong to $\mathcal{C}_{v,s} \subseteq \mathcal{C}_{v,t}$, and have time index strictly below s. This follows directly from (41). Actions whose possible arrival is later than s cannot affect a version used at s and may be omitted when reconstructing it. In particular, one does not need to know future expiration candidates created by queries outside the cone.

The positions in $\mathcal{C}_{v,t}$ can now be replayed in chronological order, conditional on $\mathcal{U}$. At each round s, the preceding paragraph determines the version labels from earlier cone actions. Communication closure determines the current actions of nodes whose positions occur in the cone. All these actions are fixed before exposing any loss from round s. Their observations are therefore obtained by querying the corresponding addresses $(i, k_i(s))$. A new address is an independent Bernoulli variable of mean θ_i, and all other uses of the address repeat its previously exposed value. This is an adaptive-query

representation on independent arrays. It also proves that the version indices at round t are computable from the final transcript. This verifies causal closure for both the policy and the loss-generation rule.

Every version completed by round t has a witness query (z, s) of that version with $s + d_v(z)$ equal to its expiration time, which is at most t. That query is in the cone, so its label has been exposed. Conversely, if the current label $k_i(t)$ had been exposed in the cone, the corresponding query would satisfy (41) by round t, forcing another refresh. This is a contradiction. A later label cannot yet have been observed. Hence the labels exposed for arm i are precisely $1, \ldots, k_i(t) - 1$.

In an adaptive querying of independent arrays, conditional on the query transcript, every unqueried entry retains its original distribution. The random address $(i, k_i(t))$ is measurable from the transcript and is unqueried. Its value therefore has conditional mean θ_i, proving (43). This is true even if agents outside the target's cone have already seen the same value. Their observations do not occur in the replay. □

The identity (43) is local to the target's information. It does not assert $\ell_t(i) = \theta_i$ conditional on the full global history. In particular, the hard family is not an instance of the stochastic regime.

Corollary D.2 (One Fixed Comparator Throughout)**.** *In each fixed instance j, the target's regret is exactly*

$$R^v_{\pi,j}(T) = \Delta \sum_{t=1}^{T} \mathbb{P}_j(I^v_t \neq j). \tag{44}$$

Proof. The target action is measurable in $\mathcal{H}_{v,t}$. By (43), its expected played loss is $\sum_i \theta_i \mathbb{P}_j(I^v_t = i)$, whereas each fixed arm i has expected loss θ_i in that round. Arm j minimizes the latter quantity at every round. Substituting in the regret definition in Section 2 gives (44). The comparator is the same fixed arm j over all rounds. Comparator losses are evaluated on the actual interaction history, as in Section 2, not on a counterfactual rerun in which the comparator was always played. □

A Graph-Local Information Bound Write $B_G(v, r) = \{z : \mathrm{dist}(z, v) \leq r\}$, $b_v(r) = |B_G(v, r)|$, and define

$$\kappa_v(G, K) = \min_{r \in \mathbb{Z}_{\geq 0}} \left\{ r + 1 + \frac{K}{b_v(r)} \right\}.$$

The minimum can equivalently be restricted to radii at most the eccentricity of v, since the ball is full beyond that point. Let

$$N_i(t) = k_i(t) - 1$$

be the number of completed versions of arm i before the target's action at round t. By Lemma D.1, it is also the number of independent arm-i Bernoulli entries exposed in the target's enlarged transcript.

Lemma D.3 (Counting Completed Versions)**.** *On every execution, every $t \in [T]$, and every integer $r \geq 1$,*

$$\sum_{i=1}^{K} N_i(t) \leq (t-1) \left(b_v(r-1) + \frac{K}{r} \right). \tag{45}$$

Consequently,

$$\sum_{i=1}^{K} N_i(t) \leq \frac{8K(t-1)}{\kappa_v(G, K)}. \tag{46}$$

Proof. A completed version has a birth time τ, expiration time $e \leq t$, and lifetime $L = e - \tau \geq 1$. A witness query (z, s) satisfies $e = s + d_v(z)$ for this version, with ties broken deterministically. Since $s \geq \tau$, we have $L \geq d_v(z)$.

For versions with $L < r$, their witness nodes lie in $B_G(v, r-1)$. Different completed versions have different witness source–time pairs, because a single action queries only one arm and one version. Thus there are at most $b_v(r-1)(t-1)$ such short versions. For versions with $L \geq r$, the completed lifetimes of each arm are disjoint intervals within the first $t-1$ rounds. There are at most $(t-1)/r$ long versions for that arm, and at most $K(t-1)/r$ in total. Adding the two counts gives (45).

Let $R \geq 1$ be the smallest integer such that

$$(R+1)b_v(R) \geq K.$$

Such an integer exists. Since $K \geq 2$ and $b_v(0) = 1$, minimality yields $Rb_v(R-1) < K$. Moreover,

$$\kappa_v(G,K) \leq R + 1 + \frac{K}{b_v(R)} \leq 2(R+1) \leq 4R.$$

With $r = R$, the right-hand side of (45) is at most $2K(t-1)/R$, which is at most $8K(t-1)/\kappa_v(G,K)$. This also covers $t = 1$, when both counts are zero. The balancing radius R is analytical and need not be used by the policy or the adversary. □

The Lower-Bound Theorem

Proposition D.4 (Graph-Local Adaptive Lower Bound)**.** *For every connected graph G, $K \geq 2$, $T \geq 1$, every cooperative policy π, and every fixed target agent v, there is a deterministic nonanticipating adaptive common binary loss process such that*

$$R^v_{\pi,\mathcal{I}}(T) \geq \frac{1}{64}\min\{T, \sqrt{\kappa_v(G,K)T}\}.$$

Arbitrary forwarding, aggregation, message contents, and message sizes are allowed. Full topology and the target identity may be given to the policy initially. The hard loss process may depend on the target and the policy.

Proof. Set $\kappa = \kappa_v(G,K)$ and choose

$$\Delta = \frac{1}{16}\min\left\{1, \sqrt{\frac{\kappa}{T}}\right\}. \tag{47}$$

Let P^t_j be the law of the enlarged transcript in instance j, and P^t_0 its reference law. Lemma D.1 gives an adaptive sequence of independent-entry queries. Conditional on the instance-independent tape, the chain rule for relative entropy counts a likelihood-ratio contribution only at the first exposure of an entry. Actions, version labels, repeats, messages, and silence are determined by the preceding queried values and introduce no additional factor. Therefore

$$\mathrm{KL}(P^t_0 \| P^t_j) = \mathrm{kl}(1/2, 1/2-\Delta)\mathbb{E}_0[N_j(t)] \leq 4\Delta^2\mathbb{E}_0[N_j(t)].$$

The equality also holds after averaging over the tape, whose marginal law is identical across instances. The Bernoulli bound follows from $\mathrm{kl}(1/2, 1/2-\Delta) = -\frac{1}{2}\log(1-4\Delta^2)$ and $\Delta \leq 1/16$.

Pinsker's inequality applied to the transcript-measurable event $I^v_t = j$, followed by averaging over j, Cauchy–Schwarz, and (46), gives

$$\begin{aligned}\frac{1}{K}\sum_{j=1}^{K}\mathbb{P}_j(I^v_t = j) &\leq \frac{1}{K} + \frac{\Delta}{K}\sum_{j=1}^{K}\sqrt{2\mathbb{E}_0[N_j(t)]} \\ &\leq \frac{1}{K} + \Delta\sqrt{\frac{2\sum_j \mathbb{E}_0[N_j(t)]}{K}} \\ &\leq \frac{1}{2} + 4\Delta\sqrt{\frac{t-1}{\kappa}} \leq \frac{3}{4}.\end{aligned}$$

The last inequality follows from (47) for every $t \leq T$, including when $T < \kappa$. By the exact fixed-instance identity (44),

$$\frac{1}{K}\sum_{j=1}^{K} R^v_{\pi,j}(T) \geq \frac{\Delta T}{4} = \frac{1}{64}\min\{T, \sqrt{\kappa T}\}. \tag{48}$$

A single j attains this average. This choice is made after summing the fixed-instance regrets, not inside any comparator maximum.

A deterministic adaptive witness. For the selected j, the fixed-comparator difference against arm j has expectation at least (48), where the expectation includes the finite Bernoulli arrays and the policy tape. Averaging first over the finite arrays yields one fixed array table attaining at least this expectation. The refresh rule in Eq. (41) then uses that fixed table. Its current loss vector is now a deterministic function of past actions, selected before the next action randomization. The expected difference against the same fixed arm j remains at least the bound, and the maximum in the regret definition is no smaller. Its refresh times depend only on past actions, so the fixed table defines a deterministic nonanticipating adaptive witness. □

Information in Arbitrary Messages The transcript includes every raw observation in the entire causal cone, not merely direct-neighbor feedback. An aggregate can transmit all information present in its inputs, and the proof grants the target those inputs directly whenever they could causally arrive. What limits information is the number of distinct completed loss versions, not the number of coordinates in a packet. Multiple observations of the same version are the same Bernoulli entry, regardless of how many agents observed it.

Paths at Every Fixed Agent On the V-vertex path, $b_v(r) \le \min\{V, 2r+1\}$ for every target. With $x = r+1$,

$$x + \frac{K}{b_v(r)} \ge x + \frac{K}{2x} \ge \sqrt{2K}, \qquad x + \frac{K}{b_v(r)} \ge 1 + \frac{K}{V}.$$

Thus κ_v is bounded below by a constant multiple of $\sqrt{K} + K/V$. Conversely, in every connected graph $b_v(r) \ge \min\{V, r+1\}$. If $V \le \sqrt{K}$, the choice $r = V-1$ gives $V + K/V \le 2K/V$. Otherwise $r = \lceil \sqrt{K} \rceil - 1$ gives the required bound. This proves

$$\kappa_v(G_{\text{path}}, K) = \Theta\left(\sqrt{K} + \frac{K}{V}\right)$$

uniformly over path vertices and all $V \ge 1$.

Together with Theorem 4.1, this proves matching bounds on the path, at each fixed target, of order

$$\min\left\{T, \sqrt{KT/V} + K^{1/4}\sqrt{T}\right\},$$

up to logarithmic factors. The lower-bound loss process may differ between targets. This proves the adaptive assertion of Theorem 4.2.

An Explicit Hard Graph for Each Diameter For $V \ge 2$ and $1 \le D \le V-1$, form a path on vertices $v_0, \dots, v_{D-1}$. Together with $V - D$ additional vertices, let v_{D-1} form a clique. This graph has V vertices and diameter exactly D. The target is $v = v_0$, whose balls satisfy

$$b_v(r) = r+1 \quad (0 \le r < D), \qquad b_v(r) = V \quad (r \ge D).$$

Consequently,

$$\kappa_v(G, K) = \min\left\{\min_{1 \le s \le D}\left(s + \frac{K}{s}\right),\ D + 1 + \frac{K}{V}\right\}. \tag{49}$$

Put $A = 1 + K/V + \min\{D, \sqrt{K}\}$. The global option in (49) is at least A. Each local option is at least $2\sqrt{K} \ge 1 + \min\{D, \sqrt{K}\}$ and at least K/V, hence at least $A/2$. For the reverse comparison, the global option gives the required upper bound when $D \le \sqrt{K}$. Otherwise $s = \lceil \sqrt{K} \rceil \le D$ gives a local option at most $2\sqrt{K} + 1 \le 2A$. Thus

$$\frac{A}{2} \le \kappa_v(G, K) \le 2A.$$

Since $\sqrt{x} + \sqrt{y} \le \sqrt{2(x+y)}$ for nonnegative x, y, we have

$$\sqrt{\kappa_v T} \ge \frac{1}{2}\left(\sqrt{KT/V} + \sqrt{\min\{D, \sqrt{K}\}T}\right).$$

Taking the minimum with T preserves this inequality up to the same factor. The graph-local bound proves Eq. (40) with constant $1/128$. If $V = 1$ and $D = 0$, then $\kappa_v = 1 + K$ and the same conclusion follows directly. This proves the adaptive lower bound in Proposition B.1.

The graph and target in this construction depend only on (V, D), not on the policy. The policy-dependent loss process may use arbitrary earlier actions but never current random draws. Combining this lower bound with the adaptive upper bound in Table 3 gives the worst-case order over graphs with these fixed parameters. The graph-local lower bound also holds separately at every fixed (G, v), while the diameter-only expression characterizes the worst case over graphs and target agents with the prescribed parameters.

E Oblivious Adversaries

Here the shared conditional-mean sequence $\ell_1, \ldots, \ell_T$ is fixed before interaction. The joint loss laws are chosen independently of the policy's randomization and satisfy

$$\mathbb{E}\left[X_t^v(i) \mid \mathcal{F}_{t-1}\right] = \ell_t(i) \qquad \text{for every } t, v, i.$$

Current losses remain conditionally independent of fresh action and auxiliary coins, but may be correlated across agents.

The standalone adaptive and oblivious configurations use the same one-time initial-tree construction in Appendix A.1 and the same tree-learning algorithm in Appendix A.3. Neither performs the later quotient reconstructions of the best-of-both-worlds algorithm. Only the construction scale and learning rate are retuned,

$$r_0 = \lceil (KT)^{1/3} \rceil, \qquad \eta = \frac{1}{8q + \sqrt{2(1 + K/m)T/\log K}}. \tag{50}$$

The confidence factor is unchanged. There is one construction followed by learning on fixed trees, with no stochastic phases or tests. Within the tree learner, only η differs from the adaptive setting. All sampling, counter updates, and messages are identical.

The oblivious worst-graph upper bound is stated separately in Theorem 4.3, and its fixed-diameter refinement is in Table 3. The learning rate changes the tree regret, and the construction scale balances that regret against setup time. Thus the displayed graph-wide rate uses both tunings in Eq. (50). Neither the sampling rule nor the message format changes.

Proposition B.1 states the fixed-diameter lower bound, and Theorem 4.4 gives its worst-graph consequence under unrestricted local communication. We prove both in Appendix E.2. These are separate from the adaptive assertion of Theorem 4.2.

The two lower-bound statements have different graph quantifiers. Stochastic losses are a special case of oblivious losses, so their worst-graph $\sqrt{K}$ term remains necessary. Thus the diameter-unrestricted statement retains the $\sqrt{K}$ term, while the fixed-diameter statement keeps the sharper diameter dependence. The appendix statement keeps that diameter dependence, while the main theorem uses the same worst-graph comparison as the stochastic row of Table 1. The proof below derives the main statement from the fixed-diameter result, including short horizons.

The diameter contributes additively to the upper bound. For every fixed graph, the leading term is $\widetilde{\mathcal{O}}\left(\sqrt{KT/V} + \sqrt{T}\right)$. The upper bound and the fixed-diameter lower bound agree up to logarithms when $D \le \sqrt{K}$, and when the additive upper-bound term is absorbed by the square-root terms. A gap remains for some intermediate parameter ranges.

The following subsections prove these statements under the conditional-mean assumptions above.

E.1 Upper Bound

The algorithm uses one call of the construction in Appendix A.1 at scale $r_0 = \lceil (KT)^{1/3} \rceil$, with $\Gamma = \log(2^{11} KVT^8)$. This is the same calibration as $\delta = 1/2$. Agents play uniformly until the call completes, retain its trees, and begin Algorithm 3 at the first public block boundary at or after their

tree's announced ready round. The wait is shorter than $q = 2h + 3$, where h is the completed tree's height. The construction is not repeated, and no stochastic phase or test is run.

When $D + 1 \le r_0$, Lemma C.17 guarantees a spanning first-arrival tree, with population V, height at most D, and ready time $\mathcal{O}(q)$. Waiting for the block boundary preserves this order. It is $\mathcal{O}(D)$ for $V \ge 2$ and $\mathcal{O}(1)$ for a singleton. This conclusion also holds when $r_0 > V$, since agent 1 reaches every vertex before any competing release at or after round $8r_0$. Thus no attachment or population lemma is needed for scales exceeding V.

For $r_0 \le V$, Lemma C.1 gives an event of probability at least $1 - 1/(2T^2)$ on which the call covers every vertex, every completed group has at least r_0/K_0 members, and its tree has height $\mathcal{O}(r_0\Gamma)$. The call and the wait for a block boundary together take $\mathcal{O}(r_0\Gamma)$ rounds. This is the only geometric estimate needed here. If a required construction record is missing at its deadline or a vertex remains uncovered, that vertex continues uniform play through the horizon. These outcomes are included in the same failure event. All routines stop at round T.

Construction packets contain the same bounded indices, depths, and counts as in Appendix A.1. Thereafter the shared protocol in Appendix A.3 aggregates retained one counts and broadcasts the complete sum in the following block. All members start together, so every generated source block is complete. Each arm count is at most $mq \le 3m^2$, and every learning packet has $\mathcal{O}(K \log V)$ bits on every execution. A singleton sends no messages. The construction uses only local neighborhoods and received messages, and remains uninformed.

Learning on a Fixed Tree

Lemma E.1. *From a common start on a ready tree of population m and height h, the expected loss difference of any member to any fixed arm is at most*

$$\frac{2\log K}{\eta} + 4\eta(1 + K/m)T \le 16q\log K + 4\sqrt{2(1 + K/m)T\log K}.$$

The bound also holds with the actual number of learning rounds as an additional upper bound.

Proof. There are at most q retained observations in a block at one agent. Hence $\eta q \le 1/8$ implies $p_t^v(i) \ge w_b(i)/2$. Conditional on the full past and the current loss array, the probability of selecting and retaining arm i is exactly $w_b(i)/4$. This identity holds because the retention coin is independent of the loss value.

Conditionally on a block-start history, let w be its base, t_0 its first round, and $n \le q$ its length. Position j in this block is the original round $t_0 + j - 1$. For this proof, $g_j = \mathbf{1} - \ell_{t_0+j-1}$ and $Q_j = \sum_{k=1}^{j} g_k$. These vectors are deterministic under the conditioning. Define

$$\bar{p}_j(i) = \frac{w_i\,(1 + 2\eta Q_{j-1}(i))}{1 + 2\eta\langle w, Q_{j-1}\rangle}, \qquad \Phi_w(x) = -\frac{1}{\eta}\log\sum_{i=1}^{K} w_i \exp(-\eta x_i).$$

The proof compares the actual sampling rule with $\bar{p}_j$, then compares the true block loss with its pooled estimate.

The Mean Sampling Rule Put $v_j = \mathrm{Var}_{i\sim w}(Q_j(i))$, and $d_j = \mathrm{Var}_{i\sim w}(g_j(i))$. Expanding the variance of $Q_{j-1} + g_j$ gives

$$\langle \bar{p}_j, g_j\rangle = \langle w, g_j\rangle + \frac{\eta}{1 + 2\eta\langle w, Q_{j-1}\rangle}\,(v_j - v_{j-1} - d_j)\,.$$

The coefficient of the variance increment is nonincreasing in j. Since $v_j \ge 0$, $v_0 = 0$, and $d_j \le 1$, summation by parts yields

$$\sum_{j=1}^{n}\langle \bar{p}_j, g_j\rangle \ge \langle w, Q_n\rangle + \frac{\eta v_n}{1 + 2\eta\langle w, Q_n\rangle} - \eta n.$$

For the centered variable $Q_n(i) - \langle w, Q_n \rangle$, whose absolute value is at most n, the exponential Taylor remainder gives

$$\frac{1}{\eta} \log \sum_i w_i \exp\left(\eta Q_n(i)\right) \le \langle w, Q_n \rangle + \frac{\eta \exp(\eta n)}{2} v_n.$$

Here we used $\log(1+x) \le x$ after taking the weighted expectation of the Taylor expansion. Because $\eta n \le 1/8$,

$$\frac{1}{1 + 2\eta \langle w, Q_n \rangle} \ge \frac{4}{5} > \frac{\exp(1/8)}{2}.$$

Writing $L = \sum_{j=1}^{n} \ell_{t_0+j-1}$, we conclude that

$$\sum_{j=1}^{n} \langle \bar{p}_j, \ell_{t_0+j-1} \rangle \le \Phi_w(L) + \eta n. \tag{51}$$

Local Sampling Noise Let $Y^{v,0}_{t_0+k-1,i}$ indicate that agent v retains a zero proxy on arm i at round $t_0 + k - 1$. Fix j and put

$$S_i = \sum_{k<j} \frac{4Y^{v,0}_{t_0+k-1,i}}{w_i}, \quad A = \langle w, S \rangle, \quad B = \langle w \odot g_j, S \rangle, \quad c = \langle w, g_j \rangle.$$

The symbol $\odot$ denotes coordinatewise multiplication. The actual complementary loss is

$$\langle p^v_{t_0+j-1}, g_j \rangle = \frac{c + 2\eta B}{1 + 2\eta A}.$$

Every source-round increment of A and B lies in $[0, 4]$ and has a deterministic conditional mean in $[0, 1]$. Its variance is at most four. Centered increments from different source rounds are orthogonal by the martingale difference property. Thus

$$\mathrm{Var}(A) \le 4(j-1), \qquad \mathrm{Var}(B) \le 4(j-1).$$

These variances and expectations remain conditional on the fixed block-start history.

Let $(A(z), B(z))$ be the linear interpolation between $(\mathbb{E}[A], \mathbb{E}[B])$ and (A, B), with displayed ratio $f(z)$. With $\delta A = A - \mathbb{E}[A]$ and $\delta B = B - \mathbb{E}[B]$, direct differentiation gives

$$f''(z) = -\frac{8\eta^2 \delta A \left(\delta B - f(z)\delta A\right)}{(1 + 2\eta A(z))^2}.$$

Since $0 \le B(z) \le A(z)$ and $0 \le c \le 1$, we have $0 \le f(z) \le 1$. Taylor's integral remainder and the zero expectation of the linear term therefore give the two-sided bound

$$\begin{aligned} \left|\mathbb{E}[\langle p^v_{t_0+j-1} - \bar{p}_j, g_j \rangle]\right| &\le 4\eta^2 \left(\sqrt{\mathrm{Var}(A)\,\mathrm{Var}(B)} + \mathrm{Var}(A)\right) \\ &\le 32\eta^2(j-1). \end{aligned}$$

The two-sided form matters because an upper bound on expected loss requires a lower bound on expected complementary loss. Summing the last display costs at most $16\eta^2 n(n-1) \le 2\eta n$. Together with Eq. (51), this proves

$$\mathbb{E}\left[\sum_{j=1}^{n} \langle p^v_{t_0+j-1}, \ell_{t_0+j-1} \rangle\right] \le \Phi_w(L) + 3\eta n. \tag{52}$$

Pooled Sampling Noise Let $Y_{j,i}^{z,1}$ indicate a retained one proxy at round $t_0 + j - 1$. For this proof, define the per-round pooled estimator

$$\widehat{\ell}_{t_0+j-1}(i) = \frac{4\sum_{z=1}^{m} Y_{j,i}^{z,1}}{mw_i}.$$

Conditioning on the full past gives $\mathbb{E}_{t_0+j-2}[\widehat{\ell}_{t_0+j-1}(i)] = \ell_{t_0+j-1}(i)$. Its second moment is bounded by conditioning further on the current loss array. Fresh action, retention, and proxy coins are independent across agents under this conditioning. A diagonal indicator has expectation at most $w_i/4$, while the product for two distinct agents has expectation at most $w_i^2/16$. Consequently,

$$\mathbb{E}_{t_0+j-2}[\widehat{\ell}_{t_0+j-1}(i)^2] \le \frac{4}{mw_i} + 1.$$

This argument allows arbitrary correlations within the current loss array.

The block estimate in Eq. (17) is $G = \sum_{j=1}^{n} \widehat{\ell}_{t_0+j-1}$. Martingale orthogonality gives

$$\mathbb{E}[G] = L, \qquad \sum_i w_i \mathbb{E}[(G(i) - L(i))^2] \le n(4K/m + 1).$$

Moreover, $G \ge 0$ and $\langle w, G\rangle \le 4n$ on every execution. On the segment joining L and G, the tilted distribution associated with Φ_w obeys

$$P_i(x) = \frac{w_i \exp(-\eta x_i)}{\sum_j w_j \exp(-\eta x_j)} \le \exp\left(\eta\langle w, x\rangle\right) w_i \le \exp(1/2) w_i < 2w_i.$$

The first inequality follows from Jensen's inequality and $x \ge 0$. The Hessian is $\nabla^2\Phi_w(x) = -\eta\left(\operatorname{diag} P(x) - P(x)P(x)^\top\right)$. Taylor expansion at L, followed by unbiasedness, therefore yields

$$\Phi_w(L) \le \mathbb{E}[\Phi_w(G)] + \eta n(4K/m + 1).$$

Combining this with Eq. (52) establishes

$$\mathbb{E}\left[\sum_{t\in\mathcal{B}_b} \langle p_t^v, \ell_t\rangle\right] \le \mathbb{E}[\Phi_{w_b}(G_b)] + 4\eta(1 + K/m)|\mathcal{B}_b|. \tag{53}$$

The Two Parity Bases For each fixed arm j, the update in Eq. (18) gives the exact identity

$$\Phi_{w_b}(G_b) - G_b(j) = \frac{\log w_{b+2}(j) - \log w_b(j)}{\eta}.$$

Summing along each parity telescopes. Each base starts uniformly, and its final coordinate is at most one, so the two sums together are at most $2\log K/\eta$. Conditional application of Eq. (53) at each block start uses $\mathbb{E}[G_b(j)] = \sum_{t\in\mathcal{B}_b} \ell_t(j)$. A final update scheduled after the horizon is used only in the proof and depends solely on observations generated before T. This proves

$$\mathbb{E}\left[\sum_t \left(X_t^v(I_t^v) - X_t^v(j)\right)\right] \le \frac{2\log K}{\eta} + 4\eta(1 + K/m)T.$$

Finally,

$$\frac{1}{\eta} = 8q + \sqrt{\frac{2(1 + K/m)T}{\log K}}, \qquad \eta \le \sqrt{\frac{\log K}{2(1 + K/m)T}},$$

which give the asserted bound. Bounded losses also cap the expected difference by the number of learning rounds. □

from Trees to the Full Graph

Proof of the oblivious upper bounds. We prove the upper bound in Table 3, which implies Theorem 4.3. The agent and comparator arm are fixed. The construction and its common starting times depend only on topology and independent construction coins. Conditioning on the ready trees and setup history therefore leaves the deterministic conditional-mean assumption valid for the subsequent interaction. The learner begins on a public block boundary with uniform parity bases, so Lemma E.1 applies.

If $V = 1$, then $m = 1$, $q = 3$, and setup takes $\mathcal{O}(1)$ rounds. The fixed-tree bound is $\widetilde{\mathcal{O}}(\sqrt{KT})$, which proves the claim, including short horizons. The rest of the proof assumes $V \geq 2$, so $D \geq 1$.

If $D + 1 \leq r_0$, Lemma C.17 gives a spanning tree of height at most D and initialization cost $\mathcal{O}(D)$, including the wait for the first learning block. The fixed-tree bound gives

$$\widetilde{\mathcal{O}}\left(D + \sqrt{T} + \sqrt{KT/V}\right).$$

If the horizon ends during initialization, its regret is at most both T and the allocated initialization length, so the same bound holds. This argument includes every $r_0 \geq V$, because $D + 1 \leq V$. Since D is an integer and $r_0 = \lceil (KT)^{1/3} \rceil$, it also gives $D \leq r_0 - 1 < (KT)^{1/3}$. The diameter is consequently the smaller term in the claimed bound.

In the remaining case $D \geq r_0$, connectedness implies $r_0 \leq D < V$. The scale is therefore within the range of Lemma C.1. With probability at least $1 - 1/(2T^2)$, the initialization cost and q are $\mathcal{O}(r_0\Gamma)$, while the target's tree population satisfies $m \geq r_0/K_0$. Conditional application of Lemma E.1 gives

$$\mathcal{O}\left(r_0\Gamma \log K + \sqrt{T \log K} + \sqrt{KK_0T \log K/r_0}\right) = \widetilde{\mathcal{O}}\left(r_0 + \sqrt{T} + \sqrt{KT/r_0}\right).$$

The good geometry event depends only on the construction coins, so this averaging does not condition on a favorable future loss. On its complement the entire regret is at most T, contributing at most $1/(2T) \leq 1$. A horizon ending before construction completes costs at most its elapsed setup length and is covered by the same bound. We conclude

$$R^v(T) = \widetilde{\mathcal{O}}\left(r_0 + \sqrt{T} + \sqrt{KT/r_0}\right).$$

Now $r_0 = \lceil (KT)^{1/3} \rceil$ makes both r_0-dependent terms $\mathcal{O}((KT)^{1/3})$, and $D \geq r_0 \geq (KT)^{1/3}$. This proves the required minimum with the diameter. The two cases analyze the same construction and do not require knowledge of D. The additive one is absorbed by $\sqrt{T}$. Maximizing over the fixed comparator after taking expectations and using Appendix E.1 for the deterministic packet bound completes the proof. □

E.2 The Lower Bound

Proof of the oblivious lower bounds. We first prove the fixed-diameter assertion of Proposition B.1. For $V \geq 2$, we use a path $v_0, \ldots, v_{D-1}$ and a clique consisting of v_{D-1} and the remaining $V - D$ vertices. Its diameter is D. We fix target v_0. For $V = 1$, we use the singleton graph. The first two arguments below work on every graph, while the last uses this particular graph.

Population For a fixed hidden arm j, the common binary losses are independent over arms and rounds, with

$$X_t^z(i) = Z_t(i), \qquad Z_t(i) \sim \mathrm{Bern}\left(\tfrac{1}{2} - \Delta \mathbf{1}\{i = j\}\right) \quad \text{for every } z.$$

Let P_j be this law and let P_0 be the all-$1/2$ law. A central observer is given all policy seeds and every previously queried arm–round coordinate. A shared coordinate queried by several agents is revealed only once. If $N_j(t)$ counts distinct previous rounds in which arm j was queried, the likelihood-ratio chain rule gives

$$\mathrm{KL}(P_0^t \| P_j^t) \leq 4\Delta^2 \mathbb{E}_0[N_j(t)], \qquad \sum_j N_j(t) \leq V(t-1).$$

The one-coordinate divergence satisfies $-\frac{1}{2}\log(1-4\Delta^2) \le 4\Delta^2$ when $\Delta \le 1/8$. Conditional on a transcript, unqueried coordinates retain their independent Bernoulli laws, so adaptive querying does not change the chain-rule argument. Policy seeds have the same distribution under all instances and contribute zero divergence.

Pinsker's inequality applied to the event $\{I_t^v = j\}$, followed by Cauchy–Schwarz, gives

$$\frac{1}{K}\sum_j P_j(I_t^v = j) \le \frac{1}{K} + \Delta\sqrt{\frac{2V(t-1)}{K}}.$$

One can obtain the form of Pinsker used here by applying the log-sum inequality to an event and its complement and then using the binary divergence bound $\mathrm{kl}(a,b) \ge 2(a-b)^2$. Choose

$$\Delta = \frac{1}{8}\min\left\{1, \sqrt{K/(VT)}\right\}.$$

The average error probability is at least $1/4$ in every round. Against comparator j, the expected loss difference is $\Delta\sum_t P_j(I_t^v \ne j)$. Averaging over j therefore proves a lower bound of order $\min\left\{T, \sqrt{KT/V}\right\}$.

Full Information The same family is used, now with

$$\Delta = \frac{1}{8}\min\left\{1, \sqrt{\log K/T}\right\},$$

and every agent is given each complete past loss vector. For a uniform hidden index J, the mutual information with this enriched history $\mathcal{H}_t$, including independent policy seeds, satisfies

$$I(J, \mathcal{H}_t) \le \frac{1}{K}\sum_j \mathrm{KL}(P_j^t \| P_0^t) \le 4\Delta^2(t-1) \le \frac{\log K}{16}.$$

The first inequality follows by inserting the mixture law of the histories into the average divergence. The excess over mutual information is the nonnegative divergence of that mixture from P_0^t. The one-coordinate inequality $\mathrm{kl}(1/2-\Delta, 1/2) \le 4\Delta^2$ follows from $\log x \le x-1$.

For $K \ge 4$, let e_t be the error probability of the action as an estimate of J. Conditioning on whether the estimate is correct yields

$$H(J \mid \mathcal{H}_t) \le \log 2 + e_t \log(K-1).$$

Since $H(J) = \log K$, the preceding mutual-information bound implies $e_t \ge 7/16$. This is the required finite-alphabet form of Fano's inequality. The average regret against the hidden arm is consequently of order $\min\left\{T, \sqrt{T\log K}\right\}$. For $K = 2, 3$, the Pinsker calculation above with all coordinates revealed and $\Delta = 1/(8\sqrt{T})$ gives the same order.

Propagation Set one unknown arm j to zero and every other arm to one at every agent and time. We fix all policy seeds and compare the resulting execution with the all-one reference. Before the target acts at time t, only observations in

$$\mathcal{C}_t = \{(z,s) : s < t,\ s + \mathrm{dist}(z, v_0) \le t\}$$

can influence it. This set is closed under all earlier information needed to choose its recorded actions. Indeed, if an observation at (w,a) influences an action at $(z,s) \in \mathcal{C}_t$, then

$$a + \mathrm{dist}(w, v_0) \le a + \mathrm{dist}(w,z) + \mathrm{dist}(z, v_0) \le s + \mathrm{dist}(z, v_0) \le t.$$

The argument includes information encoded in arbitrary messages or silence.

For $t \le D$, the graph within this cone is just the path, and

$$|\mathcal{C}_t| = \sum_{k=1}^{t-1}(k+1) = \frac{(t-1)(t+2)}{2}.$$

If arm j is absent from all queries in this reference cone, changing its losses leaves the target's transcript unchanged. Including its current reference action, the average success probability over the hidden index is therefore at most

$$\frac{|\mathcal{C}_t| + 1}{K} \le \frac{t^2}{K}.$$

For $D \ge 1$, the first $\min\left\{T, D, \max\left\{1, \lfloor\sqrt{K}/2\rfloor\right\}\right\}$ rounds thus have error probability at least a fixed positive constant. For $K = 2, 3$, the first-round error is at least $1/2$, which covers the same statement. Later losses against the zero-loss arm are nonnegative. Averaging over the policy seeds proves a lower bound of order $\min\left\{T, D, \sqrt{K}\right\}$. For $D = 0$, the claimed propagation term is zero.

Deterministic Witnesses and the Combined Bound In each Bernoulli construction, first choose a fixed j whose expected loss difference is at least the average over hidden indices. Averaging that same comparator difference over deterministic binary loss tables selects a table with at least the stated risk. Regret on that table is no smaller, since it maximizes over fixed comparators. Thus random instances are used only to prove existence of a deterministic oblivious witness. The propagation construction is already deterministic once j is selected.

All three constructions use the same graph and target. The supremum over deterministic loss tables is at least the maximum of their three lower bounds. For nonnegative a, b, c, the maximum of $\min\{T, a\}$, $\min\{T, b\}$, and $\min\{T, c\}$ is at least $\min\{T, a + b + c\}/3$. Combining the three terms proves the oblivious lower bound in Proposition B.1. The hard table may depend on the policy, and for each fixed target, the hard table may depend on the policy.

Recovering the Stochastic Arm Dependence For the diameter-unrestricted statement, choose the path with $D = V - 1$ and the same endpoint target, or the singleton when $V = 1$. If $T \ge V$, then $\sqrt{KT/V} \ge \sqrt{K}$. If $T < V$, integer horizons give $V - 1 \ge T$, and hence

$$\min\{V - 1, \sqrt{K}\} \ge \min\{T, \sqrt{K}\}.$$

Thus, in both cases,

$$\begin{aligned}&\min\left\{T, \sqrt{KT/V} + \sqrt{T \log K} + \sqrt{K}\right\} \\ &\quad \le 2 \min\left\{T, \sqrt{KT/V} + \sqrt{T \log K} + \min\{V - 1, \sqrt{K}\}\right\}.\end{aligned}$$

The fixed-diameter result therefore proves Theorem 4.4 with the additive $\sqrt{K}$ term. Thus the diameter-unrestricted statement recovers the additive $\sqrt{K}$ term, while the fixed-diameter result retains the sharper $\min\{D, \sqrt{K}\}$ dependence. □