BEST-OF-BOTH-WORLDS BANDITS UNDER GLOBAL DIFFERENTIAL PRIVACY

Anonymous authors
Paper under double-blind review

ABSTRACT

Best-of-both-worlds bandit algorithms exploit stochastic losses while retaining worst-case guarantees against adversarial losses, without knowing the environment in advance. Achieving this adaptivity under differential privacy requires balancing efficient learning and adversarial robustness within a fixed privacy budget. We introduce Private Detect-and-Switch (PrivDaS), which extends the detect-and-switch approach of Bubeck and Slivkins (COLT 2012) to private bandit learning. PrivDaS combines private successive elimination with tests for consistency with stochastic losses, continues to sample and monitor eliminated arms, and switches to a private adversarial algorithm when a test fails. Our construction controls the privacy cost across elimination, monitoring, and switching. For K arms and a horizon of T rounds, PrivDaS satisfies pure global ϵ-differential privacy and achieves expected regret $O\left((\log^2 T)\sqrt{\frac{KT}{\epsilon}} + \frac{1}{\epsilon}\right)$ against oblivious adversaries and $O\left(\sum_{i:\Delta_i>0} \frac{(\log T)\left(1+\log(1/\Delta_i)\right)}{\min\{\Delta_i,\epsilon\}}\right)$ in stochastic environments, where ϵ is the privacy parameter and Δ_i is the suboptimality gap of arm i. These bounds recover, up to additional logarithmic factors in the inverse gaps and the horizon, the gap-dependent stochastic guarantee of Azize and Basu (NeurIPS 2022) and the adversarial guarantee of Asi, Raman, and Talwar (ICML 2025), previously obtained by algorithms designed separately for the two environments. To our knowledge, PrivDaS is the first bandit algorithm to achieve best-of-both-worlds regret guarantees under global differential privacy.

1 INTRODUCTION

Multi-armed bandits model sequential decisions made with limited feedback: at each round, a learner selects an arm and observes only the loss of that arm. This restriction creates a tradeoff between gathering information and choosing actions that appear favorable. Whereas pure exploration focuses on identifying a desirable arm, regret minimization evaluates the cumulative cost of decisions made throughout the learning process (Lattimore & Szepesvári, 2020). The attainable regret depends on the assumptions imposed on the losses. Under stationary stochastic losses, repeated observations allow a learner to identify suboptimal arms and obtain logarithmic, gap-dependent regret. Against an oblivious adversary, these statistical regularities are absent, and the learner instead seeks a worst-case guarantee relative to the best fixed arm in hindsight. Best-of-both-worlds algorithms provide both types of guarantee without knowing which environment generates the observations (Bubeck & Slivkins, 2012; Seldin & Slivkins, 2014; Zimmert & Seldin, 2021; Orabona, 2019). This adaptivity is relevant when stationarity is a useful working assumption but cannot be guaranteed.

Privacy introduces a further constraint. Although a learner does not explicitly publish its observations, its subsequent actions depend on them and can reveal information about individual losses. Differential privacy limits this dependence by requiring similar distributions of observable outputs on neighboring inputs (Dwork et al., 2006; Dwork & Roth, 2014). We study pure global differential privacy: raw observations are available within a trusted agent, while the entire action sequence must protect the loss vector of any single round. This model permits aggregation before release and differs from local privacy, in which observations are privatized before reaching the learner.

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.

Private algorithms tailored to stochastic and adversarial bandits achieve substantially different regret guarantees. For bounded stochastic losses, Azize & Basu (2022) obtain a regret upper bound of order $O(\sum_{i:\Delta_i>0} \frac{\log T}{\min\{\Delta_i,\epsilon\}})$. For oblivious adversarial losses, Asi et al. (2025) give a private batching construction with an upper bound of order $O(\sqrt{\frac{KT}{\epsilon}} + \frac{1}{\epsilon})$ in the regime $\frac{1}{T} \leq \epsilon \leq 1$. Private best-of-both-worlds learning has also been studied through **local** privatization and heavy-tailed bandit methods (Zheng et al., 2020; Cheng et al., 2024). These developments motivate the following question: can a single **globally** private learner retain the favorable guarantees of algorithms designed separately for stochastic and adversarial environments?

We address this with Private Detect-and-Switch ($\mathrm{PrivDaS}$). The learner begins with a private elimination procedure and monitors whether the observations remain compatible with stochastic behavior. Active arms are checked using private estimates and cumulative-loss certificates. Eliminated arms still retain positive sampling probability and are monitored by private tests for revival relative to a reference. A failed active-arm certificate or sufficiently many eliminated-arm alarms triggers an irreversible switch to a private adversarial learner. The construction builds on the detect-and-switch approach (Bubeck & Slivkins, 2012) and the private adversarial conversion of Asi et al. (2025).

Two issues make this construction nontrivial. First, stage releases, monitoring queries, and the switching time all depend on previous private outputs. The privacy analysis must therefore cover their interleaving and the resulting adaptive action sequence. Second, eliminating an arm cannot justify discarding it permanently: an oblivious loss sequence can make an arm improve after an initially unfavorable period. Monitoring every eliminated arm too aggressively would, however, destroy the desired stochastic regret. Our construction uses aggregate releases and monitoring phases that trade sampling effort against evidence of improvement.

Contributions. To the best of our knowledge, we initiate the study of best-of-both-worlds bandits under global differential privacy. We propose $\mathrm{PrivDaS}$, which adapts the detect-and-switch framework (Bubeck & Slivkins, 2012; Auer & Chiang, 2016) through private successive elimination, stochasticity certificates, continued monitoring of eliminated arms, and a private adversarial fallback. We prove that its regret bounds recover, up to additional logarithmic factors, the best-known guarantees of private algorithms tailored to stochastic or adversarial environments. Experiments comparing $\mathrm{PrivDaS}$ with ADAP-UCB (Azize & Basu, 2022) and PRIVATECONVERSION (Asi et al., 2025), respectively, further demonstrate its ability to adapt to the unknown environment while remaining competitive with algorithms designed for each regime.

2 RELATED WORK

Best-of-both-worlds bandits. The best-of-both-worlds objective asks a single policy to adapt to stochastic losses while retaining a worst-case adversarial guarantee. Bubeck & Slivkins (2012) introduced a detect-and-switch approach, and Auer & Chiang (2016) refined it to obtain a stochastic pseudo-regret bound depending on the individual suboptimality gaps. A complementary approach establishes stochastic adaptivity directly for an adversarial algorithm. In particular, Tsallis-INF achieves adversarial expected regret of order $O(\sqrt{KT})$ and stochastic pseudo-regret of order $O(\sum_{i:\Delta_i>0} \frac{\log T}{\Delta_i})$ when the optimal arm is unique (Zimmert & Seldin, 2021). Our construction follows the detect-and-switch approach: elimination is accompanied by continued monitoring, and failed certificates trigger an adversarial fallback. Privacy requires both the estimates and these adaptive decisions to be analyzed jointly.

Differential privacy and interactive composition. Differential privacy controls how much an output distribution changes when one input record is replaced (Dwork et al., 2006; Dwork & Roth, 2014). In bandits, the protected input and observable output must be specified explicitly; local and global privacy impose different restrictions on the learner's access to feedback (Basu et al., 2019; Azize et al., 2025). Our analysis uses the Laplace mechanism, online AboveThreshold, and post-processing. Since private audits remain active while the controller interacts with other mechanisms, their composition is concurrent, which means we may interleave our interaction with multiple mechanisms, instead of following a sequential order. Vadhan & Wang (2021) establish composition guarantees for interleaved interactive mechanisms, including the pure-DP case used here. Subsequent

work extends interactive composition to further privacy notions and adaptive privacy parameters (Lyu, 2022; Haney et al., 2023). The additional task in our analysis is to account for the limited set of mechanisms that can access any changed record.

Private stochastic and adversarial bandits. For stochastic bandits, private stopping and elimination provide gap-dependent regret guarantees (Sajed & Sheffet, 2019). Azize & Basu (2022) develop private index policies using arm-dependent episodes and obtain the bounded-loss benchmark $O(\sum_{i:\Delta_i>0} \frac{\log T}{\min\{\Delta_i,\epsilon\}})$. Azize et al. (2025) sharpen the instance-dependent characterization and obtain asymptotically optimal guarantees for Bernoulli bandits. For oblivious adversarial bandits, Asi et al. (2025) privatize batch averages and combine them with a learner that tolerates unbounded noisy feedback, obtaining $O(\sqrt{\frac{KT}{\epsilon}} + \frac{1}{\epsilon})$ for $\epsilon \in [\frac{1}{T}, 1]$.

Private best-of-both-worlds guarantees. There are existing works on private best-of-both-worlds bandits under **local** differential privacy (LDP). Zheng et al. (2020) give a reduction yielding simultaneous stochastic and adversarial guarantees under approximate LDP, and Cheng et al. (2024) obtain pure-LDP guarantees with high probability through a heavy-tailed detect-and-switch construction. These results protect each feedback value before the learner receives it. Our work instead permits raw observations inside trusted mechanisms and protects the released output transcript under pure global DP. This distinction enables aggregation before noise addition and changes the attainable dependence on ϵ. The contribution is therefore the global-DP construction and its simultaneous regret guarantees; the broader private best-of-both-worlds objective and detect-and-switch methodology have precedents. Appendix A provides further comparisons.

3 Problem setting

Let $K \geq 2$ be the number of arms, let $T \geq K$ be a known horizon, and write $[K] = \{1, \ldots, K\}$. At round $t \in [T]$, a policy π selects $I_t \in [K]$ using its internal randomness and past bandit feedback, incurs loss $\ell_{t,I_t} \in [0,1]$, and observes only this selected loss. The policy is **not** informed whether the environment is stochastic or adversarial.

Stochastic environment. A stochastic instance is a collection $\nu = (P_1, \ldots, P_K)$ of distributions on $[0,1]$. Conditional on choosing arm i, the learner observes a fresh independent sample from P_i. Write $\mu_i = \mathbb{E}_{X\sim P_i}[X]$, $\mu_\star = \min_{i\in[K]} \mu_i$, and $\Delta_i = \mu_i - \mu_\star$. We assume a unique optimal arm. Let $N_i^{\mathrm{pull}}(T) = \sum_{t=1}^{T} \mathbf{1}\{I_t = i\}$. The expected stochastic pseudo-regret is

$$R_T^{\mathrm{sto}}(\pi, \nu) = \mathbb{E}_{\pi,\nu}\left[\sum_{t=1}^{T} \Delta_{I_t}\right] = \sum_{i:\Delta_i>0} \Delta_i \mathbb{E}_{\pi,\nu}[N_i^{\mathrm{pull}}(T)]. \tag{1}$$

The expectation is taken with respect to both the policy's randomness and the sampled losses.

Oblivious adversarial environment. The environment fixes a table $L = (\ell_1, \ldots, \ell_T) \in [0,1]^{T\times K}$ before the interaction begins. For a table, regret relative to the best arm in hindsight is

$$R_T(\pi; L) = \mathbb{E}_{\pi,L}\left[\sum_{t=1}^{T} \ell_{t,I_t}\right] - \min_{i\in[K]} \sum_{t=1}^{T} \ell_{t,i}, \qquad R_T^{\mathrm{adv}}(\pi) = \sup_{L} R_T(\pi; L). \tag{2}$$

Here the expectation is over the learner's randomness.

Pure global differential privacy. The definition of differential privacy is imposed on every deterministic loss table, independently of the stochastic or adversarial assumptions used in the analysis.

Definition 3.1 (Neighboring loss tables)**.** *Two tables $L, L' \in [0,1]^{T\times K}$ are neighboring, denoted as $L \sim L'$, if there exists $t_\star \in [T]$ such that $\ell_t = \ell'_t$ for every $t \neq t_\star$.*

For a policy π and table L, let $\mathbb{M}_{\pi,L}$ be the probability measure on action sequences $I_{1:T} \in [K]^T$ induced by the interaction between the learner π and the environment with loss table L.

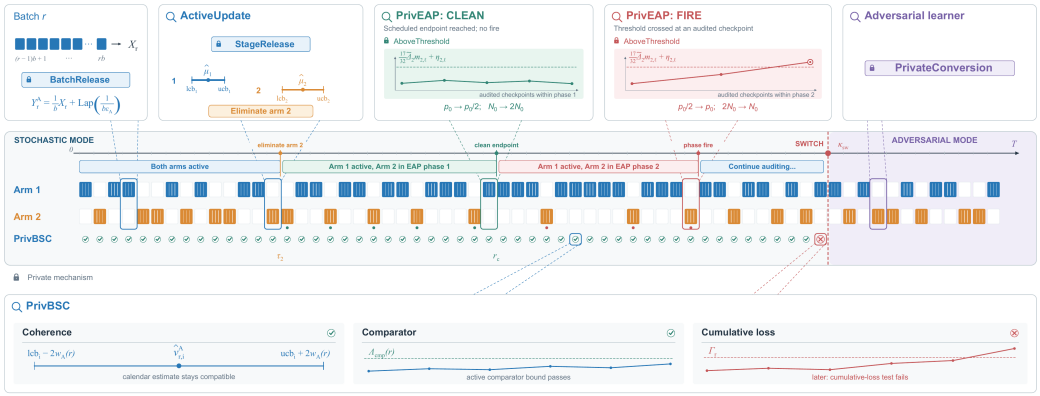


Figure 1: Two-arm simulation of PrivDaS, with arm 1 optimal; subroutine labels link to the corresponding pseudocode. Each batch comprises b consecutive pulls of one arm and releases private aggregates. Stage m ends when both active arms have received q_m batches, followed by private empirical-mean and confidence-bound updates. After arm 2 is eliminated, arm 1 observations support subsequent stages, while arm 2 observations enter eliminated-arm processing. Its two illustrated phases are clean and firing, respectively, with subsequent phase parameters adjusted accordingly. A PrivBSC failure for arm 1 or $\Theta(\log T)$ accumulated fires for arm 2 triggers a switch to adversarial mode using PRIVATECONVERSION; here, the switch follows a PrivBSC failure.

Definition 3.2 (Pure global differential privacy, Dwork & Roth (2014))**.** *A policy π is pure ϵ-globally differentially private if, for every pair $L \sim L'$ and every $E \subseteq [K]^T$,*

$$\mathbb{M}_{\pi,L}(E) \leq e^{\epsilon}\mathbb{M}_{\pi,L'}(E). \tag{3}$$

4 METHODOLOGY

We introduce Private Detect-and-Switch (PrivDaS), which initially uses a stochastic learning strategy and switches irreversibly to a private adversarial learner when a certificate fails. We state a sketch of PrivDaS in Algorithm 1 and the exact pseudocode in Algorithm 7. Appendix B specifies all subroutines and constants.

Batches and private feedback. Instead of making arm-pull decisions at each round, we divide the entire horizon of T rounds into batches of size $b = \lceil \frac{2}{\min\{\epsilon,1\}} \rceil$. We let $H = \lfloor \frac{T}{b} \rfloor$ denote the number of batches. For each complete batch $B_r = \{(r-1)b+1, \ldots, rb\}$, the learner selects one arm J_r and plays it throughout the batch. Let $X_r = \sum_{t \in B_r} \ell_{t,J_r}$. Let $\epsilon_A = \epsilon_S = \frac{1}{2}\min\{\epsilon, 1\}$. Every complete batch produces the private release $Y_r^A = \frac{X_r}{b} + \mathrm{Lap}(\frac{1}{b\epsilon_A})$. This stream, formally described in Algorithm 2, supports the stochasticity certificates before switching and the adversarial learner afterward. A second private channel (Algorithms 3 and 6) supplies active-stage aggregates and eliminated-arm monitoring responses during stochastic mode. For a completed active stage with batch-index set C, it releases $Y_C^S = \sum_{r \in C} X_r + \mathrm{Lap}(\frac{1}{\epsilon_S})$, which is used to compute empirical means and perform elimination of arms that are deemed sub-optimal. By adding Laplace noises to batch aggregates instead of to each individual observation, we reduce the frequency as well as scale of Laplace noise injection.

Private elimination. At batch r, the learner maintains active arms $\mathcal{A}$ and eliminated arms $\mathcal{B} = [K] \setminus \mathcal{A}$. At stage $m \geq 0$, the target confidence scale is $d_m = 2^{-m}$ and the stage quota is $q_m = \left\lceil 288 \left(\frac{L_H}{b d_m^2} + \frac{L_H}{b\epsilon_S d_m} \right) \right\rceil \vee 1$, where $L_H = \log(\frac{18KT^2}{\beta})$ and $\beta = \frac{1}{T}$ (in our analysis, it suffices that $\beta = O(\frac{1}{T})$). When every active arm has collected at least q_m batches in stage m, we declare that it is complete, and the learner forms a private empirical mean from the first q_m batches of each arm. For n observations, we define the confidence radius as $\rho(n) = \sqrt{\frac{L_H}{n}} + \frac{L_H}{\epsilon_S n}$. At completion of stage m, we

compute its confidence bounds $\mathrm{lcb}_i = [\hat{\mu}_{i,m} - \rho(bq_m)]_0^1, \quad \mathrm{ucb}_i = [\hat{\mu}_{i,m} + \rho(bq_m)]_0^1$. Write $u_m^{\mathrm{elim}} = \min_{j \in \mathcal{A}} \mathrm{ucb}_j$. An arm is eliminated if $\mathrm{lcb}_i > u_m^{\mathrm{elim}} + 2d_m$, indicating sub-optimality. A non-increasing benchmark u_r takes the minimum of the current pull-based upper bounds, importance-weighted upper bounds, and the previous benchmark (see (13)). If arm i is eliminated at batch τ_i, the algorithm stores $\widetilde{\ell}_i^{\mathrm{ref}} = \mathrm{lcb}_i$ and $\widetilde{\Delta}_i = \frac{1}{2}(\mathrm{lcb}_i - u_m^{\mathrm{elim}})$. These parameters are used to instantiate eliminated arm processing for arm i subsequently. The pseudocode for updating private statistics and performing eliminations is stated as Algorithm 4.

Basic stochasticity certificates. The private stochasticity check, with pseudocode stated as Algorithm 5, compares pull-based estimates with importance-weighted estimates for an arm $i \in \mathcal{A}$:

$$\widehat{L}_{r,i}^A = \sum_{s \le r} \frac{\mathbf{1}\{J_s = i\} Y_s^A}{p_{s,i}}, \qquad \widehat{\nu}_{r,i}^A = \frac{\widehat{L}_{r,i}^A}{r}, \qquad \widehat{G}_r = \sum_{s \le r} Y_s^A.$$

Here $p_{s,i}$ denotes the probability of selecting arm i in batch s. For an active arm, the importance-weighted increments estimate its batch losses in conditional expectation. The associated confidence width is $w_A(r) = 16(\sqrt{\frac{KL_H}{r}} + \frac{KL_H}{r})$. The check has three components. The coherence test (Algorithm 5) requires $\widehat{\nu}_{r,i}^A$ to lie within the active arm's confidence interval expanded by $2w_A(r)$. The active-comparator test (Algorithm 5) requires $ru_r - \widehat{L}_{r,i}^A \le 5rw_A(r)$ for every active arm. Finally, the cumulative-loss test (Algorithm 5) requires $\widehat{G}_r - ru_r \le \Gamma_r$, where Γ_r is defined in Appendix B. A failure of any test provides evidence of adversarial behavior of active arms and triggers the adversarial mode. In a stochastic instance, the probability of failure is small as shown in Lemma E.14.

Monitoring eliminated arms. An arm that appears suboptimal early in an oblivious loss sequence can later become favorable. Therefore, even after an arm is eliminated during stochastic mode, we still assign positive sampling probability to it. Subsequent samples of an eliminated arm are used by private eliminated-arm processing (Algorithm 6) to check for potential revivals. For an arm that is just eliminated, we initialize its sampling probability as $p_i = p_i^{\mathrm{init}} = \min\left\{\frac{1}{2}, \frac{1}{4K} + \frac{n_i^{\mathrm{elim}}}{4H}\right\}$, where n_i^{elim} is the total number of batches allocated to it before elimination. The sum of these probabilities is at most $1/2$, so the eliminated arms will never dominate the sampling procedure. Monitoring takes place in phases with sampling probability p_i and scheduled duration N_i, initialized as $N_i^{\mathrm{init}} = \lceil \frac{M_i}{p_i^{\mathrm{init}}} \rceil$. The expected number of audited batches is $M_i = \left\lceil 2048 \left(\frac{1}{b\widetilde{\Delta}_i^2} + \frac{1+\log(512/\widetilde{\Delta}_i)}{b\epsilon_S \widetilde{\Delta}_i} + \frac{1}{\widetilde{\Delta}_i} \right) \right\rceil \vee 1$.

Algorithm 1 PrivDaS: Algorithm sketch

Require: $K \ge 2, T \ge K, \epsilon > 0$.
Initialize parameters as in Appendix B.1.
for $r = 1, \ldots, H$ **do**
  if the algorithm is in adversarial mode **then**
    Run PRIVATECONVERSION (Alg. 1 of Asi et al. (2025)).
  else
    Draw $J_r \sim p_r$ defined in (23).
    Play J_r throughout B_r and compute Y_r^A.
    for $i \in \mathcal{A}$ **do**
      Update $\widehat{L}_{r,i}^A, \widehat{\nu}_{r,i}^A, \widehat{G}_r$.
    end for
    for $i \in \mathcal{B}$ **do**
      Run PrivEAP(i, r, J_r) (Algorithm 6).
    end for
    if an PrivEAP fire counter reaches E_0 **then**
      Switch to adversarial mode.
    else
      Run ActiveUpdate(r) (Algorithm 4).
      Run PrivBSC(r) (Algorithm 5).
      if PrivBSC(r) = fail **then**
        Switch to adversarial mode.
      end if
    end if
  end if
end for
Play arm 1 on any remaining rounds.

Within a phase, the learner accumulates $\widetilde{\ell}_i^{\mathrm{ref}} - \frac{X_r}{b}$ whenever arm i is sampled. At prescribed checkpoints, we apply the AboveThreshold mechanism to test this cumulative residual against $\theta_{i,\ell} = \frac{1}{4} p_i \widetilde{\Delta}_i N_i$, a threshold computed by parameters used to initialize the phase. The mechanism has sensitivity $1/b$ and privacy budget ϵ_S, and by Dwork & Roth (2014) is ϵ_S-DP. A positive response, or reaching a public cap on the number of audits, ends the phase with a *fire*. A phase that reaches its scheduled duration without a fire is *clean*. After a clean phase, p_i is halved and N_i is doubled. After a fire, we start a new phase where p_i doubles up to its initial value and N_i

halves down to its initial value. Intuitively, this process keeps comparing bandit feedback of an arm after elimination with a private reference frozen at elimination. A phase fire suggests that the arm is incurring loss that is too low as compared to when it is active, providing evidence of revival, while a clean phase indicates consistent performance before and after elimination. Thus, we increase monitoring upon a phase fire, while clean phases progressively reduce monitoring. A counter records fires for each arm; reaching a threshold $E_0 = \Theta(\log T)$ triggers the adversarial mode. By our design, in a stochastic instance, the probability of any single phase fire occurrence is no larger than $\frac{1}{256}$, proven in Lemma E.10, and Lemma E.11 establishes that the probability of a single arm accumulating $\Theta(\log T)$ phase fires is small. Together with the guarantee of low failure probability for basic stochasticity certificates in stochastic environments, this makes the probability of the learner falsely switching to adversarial mode small.

Sampling and adversarial mode. During stochastic mode, eliminated arms receive their current audit probabilities and the remaining probability is divided equally among active arms:

$$p_{r,i} = \begin{cases} p_i, & i \in \mathcal{B}, \\ (1 - \sum_{j \in \mathcal{B}} p_j)/|\mathcal{A}|, & i \in \mathcal{A}. \end{cases}$$

See (23). The total eliminated-arm probability remains at most $1/2$, so an active arm always has probability at least $1/(2K)$. Upon switching, the algorithm erases unfinished stochastic records and starts PRIVATECONVERSION (Huang et al., 2022; Asi et al., 2025). The adversarial learner is initialized with privacy budget ϵ_A and batch length b. The final incomplete batch plays an arbitrarily fixed arm and contributes less than b regret.

Theorem 4.1 (Privacy of PrivDaS)**.** *For every $K \geq 2$, $T \geq K$, and $\epsilon > 0$,* PrivDaS *satisfies pure ϵ-global differential privacy, under Definition 3.2.*

The proof of Theorem 4.1 is deferred to Appendix D. Figure 1 provides a simulation of PrivDaS on a two-arm instance.

5 MAIN RESULTS

We now state the two regret guarantees for PrivDaS. The proofs are deferred to Appendix F.

Theorem 5.1 (Adversarial regret)**.** *For every oblivious loss table $L \in [0,1]^{T \times K}$,*

$$R_T^{\mathrm{adv}}(\mathrm{PrivDaS}) \leq O\left((\log^2 T) \sqrt{\frac{KT}{\epsilon}} + \frac{1}{\epsilon} \right). \tag{4}$$

Theorem 5.2 (Stochastic regret)**.** *For every stochastic instance ν with a unique optimal arm,*

$$R_T^{\mathrm{sto}}(\mathrm{PrivDaS}, \nu) \leq O\left((\log T) \sum_{i:\Delta_i > 0} \frac{1 + \log(1/\Delta_i)}{\min\{\Delta_i, \epsilon\}} \right). \tag{5}$$

For $0 < \epsilon \leq 1$, the adversarial guarantee matches the $O(\sqrt{KT/\epsilon} + 1/\epsilon)$ upper bound of Asi et al. (2025), with an additional $O(\log^2 T)$ factor on the square-root term. In the stochastic setting, PrivDaS retains the $(\log T)/\Delta_i$ non-private contribution and incurs a privacy term $(\log T)(1 + \log(1/\Delta_i))/\epsilon$ per suboptimal arm. Relative to Azize & Basu (2022), the additional logarithmic dependence is in the gap, rather than an additional polynomial dependence on T or $1/\epsilon$.

5.1 PROOF SKETCHES

Appendix E supplies the concentration, monitoring, and regret arguments. In Appendix E.1, we define several good events and subsequently argue that they occur with probability $1 - \beta$, where β denotes the probability of failure and can be chosen by us. Since we incur at most regret T over T rounds by assumption of $[0,1]$-bounded loss, by choosing $\beta = \frac{1}{T}$, we can absorb the regret incurred upon failure of good events, and condition on the occurrence of the good events.

Stochastic regret. The stochastic core event controls private active-stage estimates, calendar-time estimates, realized losses, and adaptive batch counts. On this event, Lemma E.7 shows that the optimal arm remains active and an eliminated arm's frozen gap is comparable to its true gap: $\frac{7}{16}\Delta_i \leq \widetilde{\Delta}_i \leq \frac{1}{2}\Delta_i$. A monitoring phase need only have constant conditional reliability. Lemma E.10 bounds its fire probability by $\frac{1}{256}$ after conditioning on its start history and calibrated frozen parameters. The counter-amplification argument then controls the probability that any arm reaches E_0 fires. Together with the cumulative-loss certificate, this yields a stochastic safety event $\mathcal{G}^{\mathrm{sto}}_{\mathrm{safe}}$ of probability at least $1-\beta$ on which the adversarial fallback is never wrongly invoked (Theorem E.15).

For a suboptimal arm i, decompose its total pulls into the following:

$$N_i^{\mathrm{pull}}(T) = N_i^{\mathrm{act}}(T) + N_i^{\mathrm{audit}}(T) + N_i^{\mathrm{adv}}(T) + N_i^{\mathrm{rem}}(T),$$

where N_i^{act} denotes pulls of i when it is still active, N_i^{audit} denotes pulls of i after it is eliminated by the stochastic mode learner, N_i^{adv} denotes pulls of i during adversarial mode, and N_i^{rem} denotes pulls of i for the final incomplete batch (if any). On the safety event, $N_i^{\mathrm{adv}}(T) = 0$. Stage quotas and geometric elimination scales imply

$$N_i^{\mathrm{act}}(T) = O\left(L_H\left[\frac{1}{\Delta_i^2} + \frac{1}{\epsilon\Delta_i}\right]\right).$$

There are at most $P_H = O(L_H)$ activated monitoring phases per arm, and each phase processes at most $5(M_i+1)$ audits. Gap calibration therefore gives

$$\Delta_i N_i^{\mathrm{audit}}(T) = O\left(L_H\left[\frac{1}{\Delta_i} + \frac{1+\log(1/\Delta_i)}{\epsilon}\right]\right).$$

The incomplete-batch contribution is at most $b = O(1/\epsilon)$ and is absorbed by the arm sum. Summing the regret contributions and bounding the loss on the complement of the safety event by T proves Theorem 5.2.

Adversarial regret. Fix an oblivious table and a comparator arm i. Let κ_{sw} be the batch that declares a switch, with $\kappa_{\mathrm{sw}} = H+1$ if no switch occurs, and let $\tau_{\mathrm{fb}} = \min\{\kappa_{\mathrm{sw}}+1, H+1\}$. The stochastic mode period persists for batches $r < \tau_{\mathrm{fb}}$. Hence, we know that batches through $h = \kappa_{\mathrm{sw}} - 1$ have passed all preceding private stochasticity certificates. We mainly focus on the regret analysis before the switch, since after switching to the adversarial mode $\mathrm{PrivDaS}$ reduces to PRIVATECONVERSION, Algorithm 1 of Asi et al. (2025), and we can directly invoke their bounds for post-switch regret.

Write $G_h = \sum_{r\leq h} g_{r,J_r}$, where $g_{r,J_r} = b^{-1}\sum_{t\in B_r}\ell_{t,J_r}$. We define the pre-switch regret incurred by $\mathrm{PrivDaS}$ measured against a fixed comparator arm i as

$$R_{\text{pre-switch}}(i) = \sum_{r\leq h}(g_{r,J_r} - g_{r,i}).$$

We analyze by a decomposition using the following terms:

$$R_{\mathrm{learn}}(h) = G_h - hu_h, \qquad R_{\mathrm{act}}(h,i) = \min(\tau_i, h)u_{\min(\tau_i,h)} - \sum_{r\leq\min(\tau_i,h)} g_{r,i},$$

$$R_{\mathrm{eap}}(h,i) = \mathbf{1}\{\tau_i \leq h\}\sum_{r=\tau_i+1}^{h}(u_{\tau_i} - g_{r,i}).$$

Clearly, if $\tau_i > h$, we have $R_{\text{pre-switch}}(i) = R_{\mathrm{learn}}(h) + R_{\mathrm{act}}(h,i) + R_{\mathrm{eap}}(h,i)$. Otherwise, we have

$$\begin{aligned} R_{\text{pre-switch}}(i) &= R_{\mathrm{learn}}(h) + R_{\mathrm{act}}(h,i) + R_{\mathrm{eap}}(h,i) + h(u_h - u_{\tau_i}) \\ &\leq R_{\mathrm{learn}}(h) + R_{\mathrm{act}}(h,i) + R_{\mathrm{eap}}(h,i), \end{aligned} \tag{6}$$

where the last inequality uses the non-increasing property of u_r. Hence, in either case, the decomposition forms an upper bound of $R_{\text{pre-switch}}(i)$.

The remaining work is to control the decomposition terms. Since all private stochasticity certificates pass prior to the switch, Lemma E.16 shows that $R_{\mathrm{learn}}(h)$ is of order $O(\log^2 T\sqrt{KH})$ and

$R_{\mathrm{act}}(h,i)$ is of order $O(\log T\sqrt{KH})$. For an eliminated comparator, a stopped-martingale argument bounds each monitoring phase in terms of its probability of firing and its scheduled length. The doubling and halving updates yield a telescoping potential, established step by step via Lemmas E.17, E.18, E.19 and E.20, which bounds $R_{\mathrm{eap}}(h,i)$ by $O(\log T\sqrt{KH})$. Combining the above bounds $R_{\text{pre-switch}}(i)$ regret by $O(\log^2 T\sqrt{KH})$. We let $i=i^*$, the best arm in hindsight, to complete the analysis on pre-switch regret.

After switching to adversarial mode, PRIVATECONVERSION has batch regret at most $208\sqrt{K(H-\tau_{\mathrm{fb}}+1)}+2$ against any fixed comparator (Lemma F.3). Finally, multiplying batch regret by b and adding the incomplete-batch contribution proves Theorem 5.1.

Remark 5.3 (Source of extra multiplicative factors in our bounds)**.** *In the stochastic case, our bound matches Azize & Basu (2022) up to an extra factor of* $\log(\frac{1}{\Delta_i})$*, which depends on the inverse of the sub-optimality gap. This factor arises from privately monitoring eliminated arms. Each EAP phase allows* $O(1/\widetilde{\Delta}_i)$ $\mathrm{AboveThreshold}$ *queries. Although these queries share a single privacy budget, their noises create multiple opportunities for a false alarm. Controlling these opportunities with constant phase-level failure probability introduces a high-probability bound on the privacy noise perturbation of order* $O(\log(1/\widetilde{\Delta}_i))$*. The audit quota is calibrated so that the detection threshold, of order* $\widetilde{\Delta}_i M_i$*, dominates this perturbation. Consequently, using* $\widetilde{\Delta}_i=\Theta(\Delta_i)$ *on the stochastic good event and the fact that* $M_i=\Theta(\frac{\log(1/\widetilde{\Delta}_i)}{\Delta_i})$ *by its definition, summation of the audit costs over* $O(\log T)$ *phases gives the additional inverse-gap logarithm term in the bound.*

In the adversarial case, our bound matches Asi et al. (2025) up to an extra factor of $\log^2 T$*, a poly-logarithmic factor in* T*. This factor arises through the cumulative-loss certificate before switching. To avoid falsely switching on stochastic instances, this certificate must tolerate the cumulative costs of both active learning and eliminated-arm auditing. Its deterministic envelope* $\mathfrak{B}_H$*, defined in (10), includes confidence-calibrated costs accumulated over logarithmically many active stages, as well as* $O(\log T)$ *audit phases. The resulting uniform estimate is* $\mathfrak{B}_H=O(\log^2 T\sqrt{KH})$ *by Lemma E.12. This term controls* R_{learn} *in Lemma E.16, dominates the pre-switch regret and explains the* $\log^2 T$ *factor in the overall adversarial regret bound.*

6 EXPERIMENTS

We perform empirical comparisons of our algorithm $\mathrm{PrivDaS}$ with ADAP-UCB of Azize & Basu (2022) and PRIVATECONVERSION of Asi et al. (2025). For all experiments, we set $\epsilon=\frac{1}{512}$ and plot the average regret of 200 independent trials with ±1 standard error. We have provided four experiments with different settings.

For the first experiment, we run a two-arm stochastic instance. The best arm has loss following $\mathrm{Bernoulli}(0.99825)$, and the sub-optimal arm follows $\mathrm{Bernoulli}(0.999)$. Figure 2(a) gives the plot of pseudo-regret incurred by three algorithms. Up to $T=10^{11}$, $\mathrm{PrivDaS}$ incurs a pseudo-regret that is within 5 times that of ADAP-UCB, since it has to spend pulls to audit the eliminated arms. PRIVATECONVERSION, an adversarial learner, has incurred a pseudo-regret twice that of $\mathrm{PrivDaS}$.

For an adversarial instance, note that it is impractical to implement an oblivious adversary. For empirical demonstration, we consider the approach of Zimmert & Seldin (2021) to use a stochastically-constrained adversary for the second experiment. We let $K=8$ and construct two stochastic environments: In the first environment, arm 1 always outputs loss 0, while the rest follow $\mathrm{Bernoulli}(\frac{1}{8})$. In the second environment, arm 1 follows $\mathrm{Bernoulli}(\frac{7}{8})$, while the rest always output loss 1. We alternate between the two environments, with the alternating period growing exponentially with a power of 1.6. In this case, the best arm in hindsight is always arm 1. We note that ADAP-UCB has incurred significant regret, such that we have to draw a log-log plot, shown in Figure 2(b), as the regrets of the other two algorithms seem negligible otherwise. This is justified since ADAP-UCB is unable to detect the alternation of the environment and has early decided to commit to a sub-optimal arm for almost all rounds. On the other hand, $\mathrm{PrivDaS}$ is able to detect the alternation of the environment promptly and switches to the adversarial mode, in which it reduces to PRIVATECONVERSION. Up to $T=10^{11}$, $\mathrm{PrivDaS}$ incurs a regret that is within 7 times that of PRIVATECONVERSION.

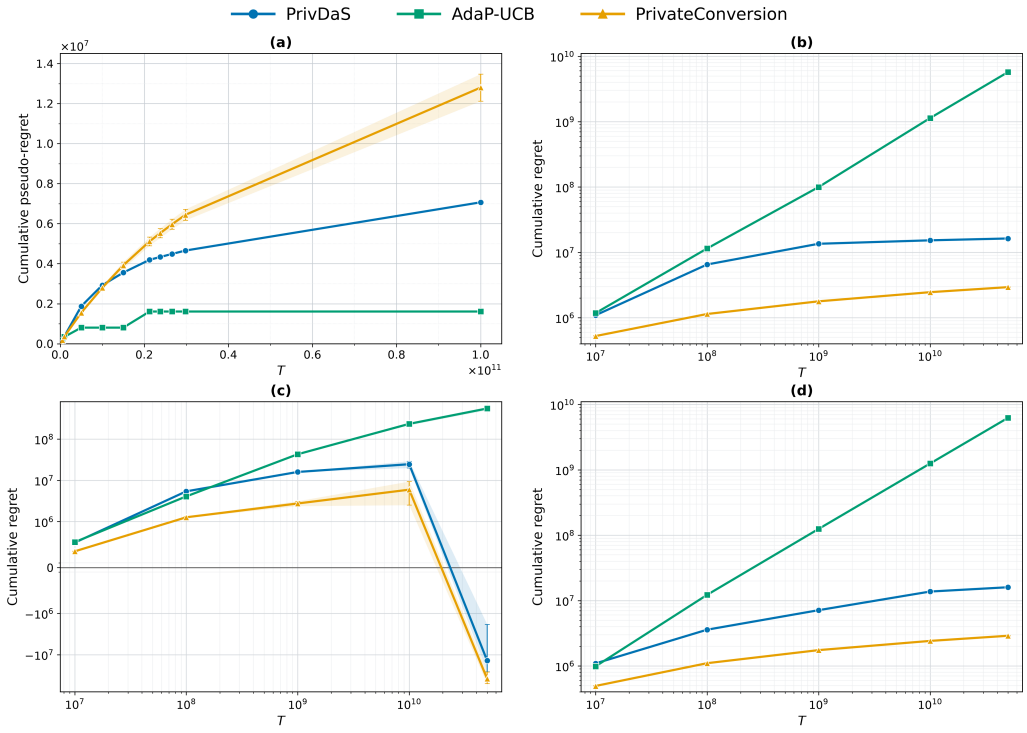


Figure 2: Plots of regret on stochastic and adversarial experiments.

Furthermore, we provide two more experiments on a stochastically-constrained adversary by modifying some of Zimmert & Seldin (2021)'s settings. In the third experiment, we alternate between two stochastic environments with different optimal arms: In the first environment, arm 1 always outputs loss 0, while the rest follow $\text{Bernoulli}(\frac{1}{8})$. In the second environment, arm 2 always outputs loss 0, while the rest follow $\text{Bernoulli}(\frac{1}{8})$. The identity of the optimal arm changes with the alternation in environments. The two environments still alternate with a period growing exponentially with a power of 1.6. Figure 2(c) plots the regret incurred by the three algorithms. Note that AdaP-UCB is still unable to detect the alternation of environments and mainly commits to sub-optimal arms. On the other hand, PrivDaS and PrivateConversion are prompt to adjust their commitments following the change in the environment. They incur negative regret in the long run as their flexibility allows them to outperform the constant strategy of always playing the best arm in hindsight.

In the last experiment, we still alternate between two stochastic environments with the same optimal arm, but now we make the alternation much more frequent by setting a constant period: The two environments are set the same as in the second experiment, but we now alternate with a period of $6,250,019$ rounds. Figure 2(d) plots the regret incurred by the three algorithms. The results are still consistent with all previous experiments: AdaP-UCB incurs significant regret as compared to the other two algorithms, and PrivDaS incurred a regret within twice that of PrivateConversion. These experiments empirically demonstrate that PrivDaS achieves best-of-both-worlds guarantees that are comparable to those of learners designed for a particular environment.

7 Discussion

We studied best-of-both-worlds bandits with global differential privacy. We have designed an algorithm that is globally differentially private and simultaneously matches the best-known bounds, up to logarithmic factors, achieved by private algorithms that are designed specifically for either adversarial or stochastic bandits. We conduct some empirical experiments demonstrating the effectiveness of our approach. Several interesting directions remain open, which we discuss in Appendix G.

AI USE STATEMENT

Generative AI was used to assist with drafting and editing this manuscript, organizing the mathematical notations, checking references, facilitating experiments, and preparing the LaTeX source. This disclosure describes the assistance used in preparing the present draft. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work studies privacy protection in sequential decision making under a trusted model. The guarantee limits the effect of changing one round's loss vector on the distribution of the released action sequence. It does not protect direct access to raw observations or provide user-level privacy when one individual contributes multiple rounds. Applications must align the adjacency relation with the intended unit of protection and restrict access to the learner's internal state. The empirical evaluation uses synthetic losses and does not require personal data.

REPRODUCIBILITY STATEMENT

Appendix B specifies the algorithm, numerical constants, private release mechanisms, update order, and stopping conventions. Appendices D and E provide the privacy and regret arguments, including the concentration events and the treatment of adaptive phase boundaries. Section 6 describes our experimental protocol, experiment results and interpretations. We have provided implementations of all algorithms in our experiments, execution code as well as plot-generating code as a self-contained package to be submitted as supplementary material.

REFERENCES

Hilal Asi, Vinod Raman, and Kunal Talwar. Faster rates for private adversarial bandits. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 1796–1832. PMLR, 2025. URL https://proceedings.mlr.press/v267/asi25b.html.

Peter Auer and Chao-Kai Chiang. An algorithm with nearly optimal pseudo-regret for both stochastic and adversarial bandits. In Vitaly Feldman, Alexander Rakhlin, and Ohad Shamir (eds.), *Proceedings of the 29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pp. 116–120. PMLR, 2016. URL https://proceedings.mlr.press/v49/auer16.html.

Achraf Azize and Debabrota Basu. When privacy meets partial information: A refined analysis of differentially private bandits. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 32199–32210. Curran Associates, Inc., 2022. doi: 10.52202/068431-2333. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/cffbaf4f47546ece96bb42c0edda40ee-Abstract-Conference.html.

Achraf Azize, Yulian Wu, Junya Honda, Francesco Orabona, Shinji Ito, and Debabrota Basu. Optimal regret of bandits under differential privacy. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, pp. 115153–115202. Curran Associates, Inc., 2025. doi: 10.52202/085713-3839. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/a70522d112322111cbf19c69e927eaf2-Abstract-Conference.html.

Debabrota Basu, Christos Dimitrakakis, and Aristide Tossou. Differential privacy for multi-armed bandits: What is it and what is its cost? *arXiv preprint arXiv:1905.12298*, 2019. URL https://arxiv.org/abs/1905.12298.

Sébastien Bubeck and Aleksandrs Slivkins. The best of both worlds: Stochastic and adversarial bandits. In Shie Mannor, Nathan Srebro, and Robert C. Williamson (eds.), *Proceedings of the 25th Annual Conference on Learning Theory*, volume 23 of *Proceedings of Machine Learning*

Research, pp. 42.1–42.23. PMLR, 2012. URL https://proceedings.mlr.press/v23/bubeck12b.html.

Duo Cheng, Xingyu Zhou, and Bo Ji. Taming heavy-tailed losses in adversarial bandits and the best-of-both-worlds setting. In *Advances in Neural Information Processing Systems*, volume 37. Curran Associates, Inc., 2024. doi: 10.52202/079017-1524. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/55f5aa8d997a8071d0651e0009ec49a2-Abstract-Conference.html.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4):211–407, 2014. doi: 10.1561/0400000042. URL https://www.cis.upenn.edu/~aaroth/privacybook.html.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006*, volume 3876 of *Lecture Notes in Computer Science*, pp. 265–284. Springer, 2006. doi: 10.1007/11681878_14. URL https://www.microsoft.com/en-us/research/publication/calibrating-noise-to-sensitivity-in-private-data-analysis/.

Samuel Haney, Michael Shoemate, Grace Tian, Salil Vadhan, Andrew Vyrros, Vicki Xu, and Wanrong Zhang. Concurrent composition for interactive differential privacy with adaptive privacy-loss parameters. In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1949–1963. Association for Computing Machinery, 2023. doi: 10.1145/3576915.3623128. URL https://doi.org/10.1145/3576915.3623128.

Jiatai Huang, Yan Dai, and Longbo Huang. Adaptive best-of-both-worlds algorithm for heavy-tailed multi-armed bandits. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 9173–9200. PMLR, 2022. URL https://proceedings.mlr.press/v162/huang22c.html.

Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020. doi: 10.1017/9781108571401. URL https://tor-lattimore.com/downloads/book/book.pdf.

Xin Lyu. Composition theorems for interactive differential privacy. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 9700–9712. Curran Associates, Inc., 2022. doi: 10.52202/068431-0705. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/3f52b555967a95ee850fcecbd29ee52d-Abstract-Conference.html.

Saeed Masoudian and Yevgeny Seldin. Improved analysis of the Tsallis-INF algorithm in stochastically constrained adversarial bandits and stochastic bandits with adversarial corruptions. In *Proceedings of the 34th Annual Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pp. 3330–3350. PMLR, 2021. URL https://proceedings.mlr.press/v134/masoudian21a.html.

Francesco Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*, 2019.

Wenbo Ren, Xingyu Zhou, Jia Liu, and Ness B. Shroff. Multi-armed bandits with local differential privacy. *arXiv preprint arXiv:2007.03121*, 2020. URL https://arxiv.org/abs/2007.03121.

Touqir Sajed and Or Sheffet. An optimal private stochastic-MAB algorithm based on optimal private stopping rule. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 5579–5588. PMLR, 2019. URL https://proceedings.mlr.press/v97/sajed19a.html.

Yevgeny Seldin and Aleksandrs Slivkins. One practical algorithm for both stochastic and adversarial bandits. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pp. 1287–1295. PMLR, 2014. URL https://proceedings.mlr.press/v32/seldinb14.html.

Aristide Tossou and Christos Dimitrakakis. Achieving privacy in the adversarial multi-armed bandit. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1):2653–2659, 2017. doi: 10.1609/aaai.v31i1.10896. URL https://ojs.aaai.org/index.php/AAAI/article/view/10896.

Salil Vadhan and Tianhao Wang. Concurrent composition of differential privacy. In Kobbi Nissim and Brent Waters (eds.), *Theory of Cryptography: 19th International Conference, TCC 2021, Proceedings, Part II*, volume 13043 of *Lecture Notes in Computer Science*, pp. 582–604. Springer, 2021. doi: 10.1007/978-3-030-90453-1_20. URL https://link.springer.com/chapter/10.1007/978-3-030-90453-1_20.

Kai Zheng, Tianle Cai, Weiran Huang, Zhenguo Li, and Liwei Wang. Locally differentially private (contextual) bandits learning. In *Advances in Neural Information Processing Systems*, volume 33, pp. 12300–12310. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/hash/908c9a564a86426585b29f5335b619bc-Abstract.html.

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. URL https://jmlr.org/papers/v22/19-753.html.

A Extended Related Work

A.1 Adaptivity across stochastic and adversarial environments

Stochastic and adversarial bandits differ in the structure available for learning. Stochastic observations permit estimation of fixed arm means, whereas an adversarial guarantee must remain valid when losses vary arbitrarily over time. Standard treatments distinguish cumulative regret minimization from pure exploration, where the objective is to identify a good arm rather than control the losses incurred while learning (Lattimore & Szepesvári, 2020). Our work concerns cumulative regret.

In the detect-and-switch framework, tests support stochastic learning only while their certificates remain valid. The algorithms of Bubeck & Slivkins (2012); Auer & Chiang (2016) illustrate why eliminating an apparently inferior arm does not justify permanently ignoring it: subsequent losses may make it a favorable comparator. Continued exploration and an adversarial fallback protect against this possibility. Our active-arm certificates and eliminated-arm audits serve these two roles under private feedback.

Tsallis-INF provides a different route by analyzing a regularized adversarial learner in both regimes (Zimmert & Seldin, 2021). Its stochastic analysis, and later refinements for self-bounding adversaries and corrupted stochastic environments (Masoudian & Seldin, 2021), show that a switch is not necessary for best-of-both-worlds adaptation. Our results address the stochastic and oblivious adversarial endpoints. They do not establish a corruption-dependent interpolation bound.

A.2 The stochastic global-DP benchmark

For bounded stochastic losses, the following expressions agree up to constant factors:

$$\sum_{i:\Delta_i>0} \frac{\log T}{\min\{\Delta_i, \epsilon\}} \quad \text{and} \quad \sum_{i:\Delta_i>0} \log T \left(\frac{1}{\Delta_i} + \frac{1}{\epsilon}\right).$$

Sajed & Sheffet (2019, Theorem 4.3) obtain this additive statistical and privacy dependence for sufficiently large horizons through private stopping and successive elimination. The AdaP-UCB guarantee of Azize & Basu (2022, Theorem 7) gives the first displayed form using adaptive episodes and fresh private estimates. This benchmark separates the sampling cost of resolving a small gap from the privacy cost associated with releasing an aggregate.

The stochastic bound proved here contains an additional factor $1+\log(\frac{1}{\Delta_i})$ in the privacy term. Thus, the comparison retains the dependence on individual gaps and separates the statistical and privacy contributions. The newer bounds of Azize et al. (2025) resolve a finer question: their privacy-dependent information quantity gives an asymptotic characterization for **Bernoulli** instances.

A.3 Batching and noisy adversarial feedback

Early private adversarial bandit algorithms illustrate the use of perturbations to protect online feedback (Tossou & Dimitrakakis, 2017). The conversion of Asi et al. (2025) repeats an action within a batch and perturbs the resulting average. Its privacy sensitivity is inversely proportional to the batch length. Choosing a batch length of order $\frac{1}{\epsilon}$ yields noisy batch observations with constant-scale noise, while reducing the number of decisions. Their HTINF instantiation gives the adversarial upper bound used in this paper; the base learner originates in the heavy-tailed bandit work of Huang et al. (2022).

A.4 Local privacy and earlier private best-of-both-worlds results

Local privacy protects observations before they reach the learner. Global privacy permits a trusted learner to access raw observations while restricting the distribution of the released outputs. In the context-free bandit model considered here, a locally private feedback protocol also protects actions obtained by post-processing those messages, but a global-DP policy need not satisfy the local requirement (Basu et al., 2019; Zheng et al., 2020).

The reduction of Zheng et al. (2020) instantiates a non-private best-of-both-worlds learner with Gaussian-perturbed feedback and analyzes approximate LDP. Cheng et al. (2024, Remark 6 and

Algorithm 2 $\mathrm{BatchRelease}(X_r)$

1: Draw $Z_r^A \sim \mathrm{Lap}(\frac{1}{b\epsilon_A})$.
2: Release $Y_r^A \leftarrow \frac{X_r}{b} + Z_r^A$.
3: **return** Y_r^A.

Appendix E) use Laplace-perturbed bounded losses within a heavy-tailed construction to obtain pure-LDP best-of-both-worlds guarantees with high probability. Our result uses the different trust model to aggregate observations inside private mechanisms and obtains the stated pure global-DP bounds. Consequently, the stronger access restriction has led to larger privacy penalties for locally differentially private models. For example, stochastic LDP bandits exhibit an ϵ^{-2} penalty in gap-dependent regret for $0 < \epsilon \leq 1$ (Ren et al., 2020). This differs from the additive ϵ^{-1} term in the global-DP stochastic benchmark.

A.5 Composition of the adaptive release streams

Post-processing protects decisions computed from a private transcript, but proving privacy of that transcript requires accounting for every mechanism that accesses sensitive data. The relevant mechanisms here include one-shot Laplace releases and persistent online $\mathrm{AboveThreshold}$ mechanisms (Dwork & Roth, 2014). Because the controller interleaves their queries and uses previous replies to choose subsequent actions, a sequential list of independent, completed releases does not describe the execution. Vadhan & Wang (2021) analyze concurrent composition of interactive mechanisms. For pure DP, their result supports composition without an additional penalty merely for interleaving the interactions. Lyu (2022) develops further interactive composition results, and Haney et al. (2023) study adaptive privacy-loss parameters. Our use of this literature is narrower: the budgets are fixed, and the proof tracks which mechanisms can inspect a changed record along a common interaction path.

B Complete algorithm specification

This appendix specifies the state, calibration, and subroutines of $\mathrm{PrivDaS}$. Algorithm 7 gives the complete order of execution underlying Algorithm 1. All logarithms are natural unless a base is indicated, and $[x]_0^1 = \min\{1, \max\{0, x\}\}$ denotes clipping to $[0, 1]$.

B.1 Calibration and batched observations

For a target privacy level $\epsilon > 0$, set

$$\epsilon_0 = \min\{\epsilon, 1\}, \qquad \epsilon_A = \epsilon_S = \epsilon_0/2, \qquad b = \left\lceil \frac{1}{\epsilon_A} \right\rceil, \qquad H = \left\lfloor \frac{T}{b} \right\rfloor, \qquad \beta = \frac{1}{T}.$$

Let

$$L_H = \log \frac{18KT^2}{\beta}, \qquad C_M = 2048, \qquad \rho_0 = \frac{1}{256}, \qquad E_0 = \lceil 2L_H \rceil.$$

Batch $r \in [H]$ consists of the rounds

$$B_r = \{(r-1)b + 1, \ldots, rb\}.$$

The learner chooses one arm J_r and plays it throughout B_r. Write $X_r = \sum_{t \in B_r} \ell_{t,J_r}$ and $g_{r,i} = b^{-1} \sum_{t \in B_r} \ell_{t,i}$.

B.2 Private loss release subroutines

The batch-release stream is used in both modes: it supplies the importance-weighted estimates for stochasticity certificates in stochastic mode and the bandit feedback in adversarial mode. Active-stage estimates use a separate private aggregate stream, released only after every active arm has met the current stage quota. Algorithms 2 and 3 specify these two mechanisms.

Algorithm 3 $\mathrm{StageRelease}(C)$

Require: Completed active-stage batch-index set C.
1: Compute $Y_C^S = \sum_{r\in C} X_r$.
2: Draw $Z_C^S \sim \mathrm{Lap}(1/\epsilon_S)$.
3: Set $Y_C^S \leftarrow Y_C^S + Z_C^S$.
4: **return** Y_C^S.

B.3 STOCHASTIC-MODE STATE

The stochastic mode maintains

$$\mathcal{A} = \text{active arms}, \qquad \mathcal{B} = [K] \setminus \mathcal{A},$$
$$m = \text{stage index}, \qquad d_m = 2^{-m}.$$

For each active arm i, the learner maintains a FIFO buffer Q_i of assigned batch records. The pair $(\mathrm{lcb}_i, \mathrm{ucb}_i)$ contains the monotone private confidence bounds, and n_i^{act} counts the batches in which arm i is played while active. Initially,

$$\mathcal{A} = [K], \qquad \mathcal{B} = \varnothing, \qquad m = 0, \qquad Q_i = \varnothing, \qquad n_i^{\mathrm{act}} = 0, \qquad \mathrm{lcb}_i = 0, \qquad \mathrm{ucb}_i = 1,$$
$$\widehat{G}_0 = 0, \qquad \widehat{L}_{0,i}^{A} = 0, \qquad \mathrm{ucb}_{0,i}^{A} = 1, \qquad u_0 = 1.$$

The active-stage quota is

$$q_m = \left\lceil 288 \left(\frac{L_H}{b d_m^2} + \frac{L_H}{b \epsilon_S d_m} \right) \right\rceil \vee 1. \tag{7}$$

A stage is eligible for release when at least two arms remain active and every active arm has at least q_m assigned batches. The estimate uses the first q_m batches of the current buffer; all remaining buffered records are discarded when the stage ends. Thus observations assigned to an active aggregate are never reused in a later active stage.

The cumulative-loss certificate uses a deterministic budget computed from the public parameters; no unknown stochastic gap enters this budget.

Let

$$a_{\max}(H) = 2 + \lceil \log_2(\max\{H, 2\}) \rceil, \qquad \mathcal{D}_H = \left\{ 2^{-j} : 0 \le j \le \lceil \log_2(\max\{H, 1\}) \rceil \right\},$$

and, for $\alpha \in \mathcal{D}_H$,

$$m(\alpha) = \left\lceil \log_2 \frac{8}{\alpha} \right\rceil, \qquad P_H = a_{\max}(H) + 2E_0 + 1, \qquad \Lambda_\alpha = 1 + \log \frac{8192}{7\alpha},$$

$$\mathfrak{A}_H(\alpha) = 32 \sum_{m=0}^{m(\alpha)} d_m q_m + 512 L_H, \tag{8}$$

$$\mathfrak{E}_H(\alpha) = 5 P_H \left[2 + C_M \left(\frac{256}{49 b \alpha} + \frac{16 \Lambda_\alpha}{7 b \epsilon_S} + \frac{16}{7} \right) \right], \tag{9}$$

$$\mathfrak{B}_H = \min \left\{ H, \min_{\alpha \in \mathcal{D}_H} \left[H\alpha + K\{\mathfrak{A}_H(\alpha) + \mathfrak{E}_H(\alpha)\} \right] \right\}. \tag{10}$$

For each eliminated arm, the state includes its frozen reference $\widetilde{\ell}_i^{\mathrm{ref}}$, certified scale $\widetilde{\Delta}_i$, sample target M_i, audit probability p_i, phase length N_i, phase index ℓ_i, calendar clock c_i, and fire counter E_i. The audit count is A_i^{eap}, the audit count at the latest checkpoint is A_i^{last}, and the number of submitted queries is j_i^{AT}.

B.4 ACTIVE-ARM UPDATE AND ELIMINATION

Algorithm 4 releases stage estimates only when the quota condition holds. An arm is eliminated when its lower confidence bound exceeds the smallest active upper bound by more than $2d_m$. The

Algorithm 4 ActiveUpdate(r)

if $|\mathcal{A}| \geq 2$ and $|Q_i| \geq q_m$ for every $i \in \mathcal{A}$ **then**
  for each $i \in \mathcal{A}$ **do**
    Choose the first q_m assigned batch indices and denote that index set by $C_{i,m}$.
    Release $\widehat{\mu}_{i,m} \leftarrow \frac{\mathrm{StageRelease}(C_{i,m})}{bq_m}$.
    $\mathrm{lcb}_i \leftarrow \max\{\mathrm{lcb}_i, [\widehat{\mu}_{i,m} - \rho(bq_m)]_0^1\}$.
    $\mathrm{ucb}_i \leftarrow \min\{\mathrm{ucb}_i, [\widehat{\mu}_{i,m} + \rho(bq_m)]_0^1\}$.
  end for
  Let j_m be the smallest index in $\arg\min_{j\in\mathcal{A}} \widehat{\mu}_{j,m}$ and define $u_m^{\mathrm{elim}} \leftarrow \min_{j\in\mathcal{A}} \mathrm{ucb}_j$.
  Define $\mathcal{E}_m = \left\{i \in \mathcal{A} \setminus \{j_m\} : \mathrm{lcb}_i > u_m^{\mathrm{elim}} + 2d_m\right\}$.
  for each $i \in \mathcal{E}_m$ **do**
    $\widetilde{\ell}_i^{\mathrm{ref}} \leftarrow \mathrm{lcb}_i$; $\widetilde{\Delta}_i \leftarrow (\mathrm{lcb}_i - u_m^{\mathrm{elim}})/2$.
    $n_i^{\mathrm{elim}} \leftarrow n_i^{\mathrm{act}}$.
    Clear Q_i and move i from $\mathcal{A}$ to $\mathcal{B}$.
  end for
  Update $u_r \leftarrow \min\left\{u_r, \min_{j\in\mathcal{A}} \mathrm{ucb}_j, \min_{j\in\mathcal{A}} \mathrm{ucb}_{r,j}^A\right\}$.
  for each $i \in \mathcal{E}_m$ **do**
    Initialize eliminated-arm processing using (17)–(20).
  end for
  Set $Q_j \leftarrow \varnothing$ for $j \in \mathcal{A}$.
  $m \leftarrow m + 1$, $d_m \leftarrow 2^{-m}$, and update q_m by (7).
end if

arm with the smallest current private stage estimate is protected from elimination, with ties broken by index, so the active set is always nonempty. For the analysis, let τ_i be arm i's elimination batch and define $\widetilde{v}_i = u_{\tau_i}$ using the benchmark after updating the active set. The private-aggregate confidence radius is

$$\rho(n) = \sqrt{\frac{L_H}{n}} + \frac{L_H}{\epsilon_S n}. \tag{11}$$

B.5 Basic stochasticity certificate

Algorithm 5 checks agreement between the two active-arm estimators, compares the running benchmark with each active comparator, and controls the learner's cumulative loss relative to that benchmark. All three tests use previously privatized quantities and require no additional access to raw losses. For an arm that is active at batch r, define

$$\widehat{L}_{r,i}^A = \sum_{s\leq r} \frac{\mathbf{1}\{J_s = i\} Y_s^A}{p_{s,i}}, \qquad \widehat{\nu}_{r,i}^A = \widehat{L}_{r,i}^A / r.$$

Set

$$w_A(r) = 16\left(\sqrt{\frac{KL_H}{r}} + \frac{KL_H}{r}\right), \qquad \widehat{G}_r = \sum_{s\leq r} Y_s^A.$$

Maintain batch-wise upper bounds

$$\mathrm{ucb}_{r,i}^A = \min\left\{\mathrm{ucb}_{r-1,i}^A, [\widehat{\nu}_{r,i}^A + 2w_A(r)]_0^1\right\}. \tag{12}$$

Before the active update, take the minimum of the preceding benchmark and the current active upper bounds:

$$u_r = \min\left\{u_{r-1}, \min_{j\in\mathcal{A}_r} \mathrm{ucb}_j, \min_{j\in\mathcal{A}_r} \mathrm{ucb}_{r,j}^A\right\}. \tag{13}$$

Consequently, (u_r) is non-increasing. Define

$$\Lambda_{\mathrm{cmp}}(r) = 5rw_A(r), \tag{14}$$

$$\Lambda_{\mathrm{noise}}(r) = 2\sqrt{2rL_H} + 2L_H + \sqrt{\frac{rL_H}{b}}, \tag{15}$$

$$\Gamma_r = \mathfrak{B}_H + \Lambda_{\mathrm{noise}}(r). \tag{16}$$

Algorithm 5 $\mathrm{PrivBSC}(r)$: private basic stochasticity check

Require: Active set $\mathcal{A}_r$; confidence intervals $(\mathrm{lcb}_i, \mathrm{ucb}_i)$; importance-weighted estimates $\widehat{\nu}^{A}_{r,i}$; benchmark u_r; cumulative learner loss $\widehat{G}_r$.

Active-arm coherence test:
if there exists $i \in \mathcal{A}_r$ such that $\widehat{\nu}^{A}_{r,i} \notin [\mathrm{lcb}_i - 2w_A(r), \mathrm{ucb}_i + 2w_A(r)]$ **then**
  return fail.
end if
Active-comparator certificate:
if there exists $i \in \mathcal{A}_r$ such that $ru_r - \widehat{L}^{A}_{r,i} > \Lambda_{\mathrm{cmp}}(r)$ **then**
  return fail.
end if
Cumulative-loss certificate:
if $\widehat{G}_r - ru_r > \Gamma_r$ **then**
  return fail.
end if
return pass.

B.6 Eliminated-arm processing

An eliminated arm remains under observation because its losses may subsequently decrease. Algorithm 6 compares later observations with the private reference frozen at elimination, using one online $\mathrm{AboveThreshold}$ (Dwork & Roth, 2014) instance per audit phase.

When arm i is eliminated, initialize

$$p_i \leftarrow \min\left\{\frac{1}{2}, \frac{1}{4K} + \frac{n_i^{\mathrm{elim}}}{4H}\right\}, \tag{17}$$

$$\Lambda_i = 1 + \log\frac{512}{\widetilde{\Delta}_i}, \qquad M_i = \left\lceil C_M\left(\frac{1}{b\widetilde{\Delta}_i^2} + \frac{\Lambda_i}{b\epsilon_S\widetilde{\Delta}_i} + \frac{1}{\widetilde{\Delta}_i}\right)\right\rceil \vee 1, \tag{18}$$

$$\begin{aligned} N_i \leftarrow \left\lceil \frac{M_i}{p_i} \right\rceil, \qquad \ell_i \leftarrow 1, \qquad c_i \leftarrow 0, \qquad E_i \leftarrow 0, \\ A_i^{\mathrm{eap}} \leftarrow 0, \qquad A_i^{\mathrm{last}} \leftarrow 0, \qquad j_i^{\mathrm{AT}} \leftarrow 0, \qquad S_i^{\mathrm{eap}} \leftarrow 0. \end{aligned} \tag{19}$$

Store $p_i^{\mathrm{init}} = p_i$ and $N_i^{\mathrm{init}} = N_i$. We use the pair (i, ℓ_i) to index an EAP phase, meaning that this is the ℓ_i-th phase of arm i. At the start of phase (i, ℓ_i), put

$$\begin{aligned} m_{i,\ell_i} = p_i N_i, \qquad q_{i,\ell_i} = \left\lceil \frac{\widetilde{\Delta}_i m_{i,\ell_i}}{64} \right\rceil, \\ U_{i,\ell_i} = \lceil 4m_{i,\ell_i} \rceil, \quad \theta_{i,\ell_i} = \frac{1}{4}\widetilde{\Delta}_i m_{i,\ell_i}. \end{aligned} \tag{20}$$

The initialization (17) implies

$$\sum_{i\in\mathcal{B}} p_i \le \frac{1}{4} + \frac{1}{4H}\sum_{i\in\mathcal{B}} n_i^{\mathrm{elim}} \le \frac{1}{2}. \tag{21}$$

The batches counted at the elimination of distinct arms are disjoint. Later updates only halve p_i or double it up to p_i^{init}, so (21) remains true throughout stochastic mode.

Private checkpoint queries. The query index j counts checkpoint submissions, whereas A_i^{eap} counts actual audits. At checkpoint j the query is the cumulative residual

$$D_{i,\ell_i}(j) = \sum_{r\in\mathcal{C}_{i,\ell_i}(j)} \left(\widetilde{\ell}_i^{\mathrm{ref}} - \frac{X_r}{b}\right), \tag{22}$$

where $\mathcal{C}_{i,\ell_i}(j)$ contains all audits of arm i in the phase through that checkpoint. One persistent $\mathrm{AboveThreshold}$ instance has sensitivity $1/b$, privacy budget ϵ_S, threshold θ_{i,ℓ_i}, and query cap

Algorithm 6 $\mathrm{PrivEAP}(i; r, J_r)$

if $c_i = 0$ **then**
  Set the phase parameters by (17)–(20).
  Initialize fresh persistent AT_{i,ℓ_i} with sensitivity $1/b$, budget ϵ_S, threshold θ_{i,ℓ_i}, and query cap $Q^{\mathrm{AT}}_{i,\ell_i}$.
end if
$c_i \leftarrow c_i + 1$.
if $J_r = i$ **then**
  $A_i^{\mathrm{eap}} \leftarrow A_i^{\mathrm{eap}} + 1$.
  $S_i^{\mathrm{eap}} \leftarrow S_i^{\mathrm{eap}} + \widetilde{\ell}_i^{\mathrm{ref}} - X_r/b$.
end if
if $A_i^{\mathrm{eap}} > A_i^{\mathrm{last}}$ and $(q_{i,\ell_i} \mid A_i^{\mathrm{eap}}$ or $A_i^{\mathrm{eap}} = U_{i,\ell_i}$ or $c_i = N_i)$ **then**
  $j_i^{\mathrm{AT}} \leftarrow j_i^{\mathrm{AT}} + 1$.
  Submit S_i^{eap} to AT_{i,ℓ_i} and receive $z \in \{\bot, \top\}$.
  $A_i^{\mathrm{last}} \leftarrow A_i^{\mathrm{eap}}$.
end if
if $z = \top$ or $A_i^{\mathrm{eap}} = U_{i,\ell_i}$ **then**
  $E_i \leftarrow E_i + 1$.
  if $E_i \geq E_0$ **then return** switch.
  end if
  $p_i \leftarrow \min\{p_i^{\mathrm{init}}, 2p_i\}$; $N_i \leftarrow \max\{N_i^{\mathrm{init}}, N_i/2\}$.
else if $c_i = N_i$ **then**
  $p_i \leftarrow p_i/2$; $N_i \leftarrow 2N_i$.
else
  return
end if
$\ell_i \leftarrow \ell_i + 1$; $(c_i, A_i^{\mathrm{eap}}, S_i^{\mathrm{eap}}) \leftarrow (0, 0, 0)$.

$Q^{\mathrm{AT}}_{i,\ell_i} = 2 + \lceil U_{i,\ell_i}/q_{i,\ell_i} \rceil$. It privately tests whether the residual $D_{i,\ell_i}(j)$ exceeds the threshold θ_{i,ℓ_i}. It returns a positive signal $\top$ if the threshold is exceeded, and a negative signal $\bot$ otherwise. A phase fires when its $\mathrm{AboveThreshold}$ instance returns $\top$ or its public audit count reaches the cap U_{i,ℓ_i}. After a fire, the next phase doubles the audit probability and halves the scheduled length, subject to their initial bounds. Clean completion makes the opposite update. All fires increment E_i, and reaching E_0 triggers the switch to adversarial mode. Completed phases are never activated again.

B.7 Sampling distribution in stochastic mode

At the beginning of a stochastic-mode batch, use the current EAP probabilities for eliminated arms and allocate the remaining probability uniformly among active arms:

$$p_{r,i} = p_i \quad (i \in \mathcal{B}), \qquad p_{r,i} = \frac{1 - \sum_{j \in \mathcal{B}} p_j}{|\mathcal{A}|} \quad (i \in \mathcal{A}). \tag{23}$$

By (21), the eliminated-arm mass is at most $1/2$, so every active arm receives probability at least $1/(2K)$.

B.8 Adversarial mode

When the algorithm switches, it discards all unfinished stochastic-mode records and initializes a fresh instance $\textsc{PrivateConversion}(\epsilon_A, b)$ of the private batching conversion of Asi et al. (2025), with

$$B = \mathrm{HTINF}(\alpha = 2, \sigma = \sqrt{6}), \qquad \tau = b = \left\lceil \frac{1}{\epsilon_A} \right\rceil, \qquad \eta = \epsilon_A. \tag{24}$$

In adversarial mode, the $\textsc{PrivateConversion}$ instance chooses J_r for batch r, the learner plays that arm throughout B_r, and the instance receives exactly the private batch aggregate Y_r^A returned by $\mathrm{BatchRelease}$.

B.9 Main algorithm

Algorithm 7 specifies the complete execution of $\mathrm{PrivDaS}$.

Algorithm 7 $\mathrm{PrivDaS}(K, T, \epsilon)$: Complete execution

Initialize the parameters in Appendix B.1 and Appendix B.3.
Compute $\mathfrak{B}_H$ by (10).
Set mode $\leftarrow$ STO.
for $r = 1, \ldots, H$ **do**
  if mode $=$ ADV **then**
    Obtain J_r from PRIVATECONVERSION and play it throughout B_r.
    Receive $Y_r^A \leftarrow \mathrm{BatchRelease}(r, J_r)$ and feed it to PRIVATECONVERSION.
    continue
  end if
  Draw $J_r \sim p_r$ by (23) and play it throughout B_r.
  Receive $Y_r^A \leftarrow \mathrm{BatchRelease}(r, J_r)$.
  $\widehat{G}_r \leftarrow \widehat{G}_{r-1} + Y_r^A$.
  for each $i \in \mathcal{A}$ **do**
    $\widehat{L}_{r,i}^A \leftarrow \widehat{L}_{r-1,i}^A + \mathbf{1}\{J_r = i\} Y_r^A / p_{r,i}$; $\widehat{\nu}_{r,i}^A \leftarrow \widehat{L}_{r,i}^A / r$.
    Update $\mathrm{ucb}_{r,i}^A$ by (12).
  end for
  Update u_r by (13).
  if $J_r \in \mathcal{A}$ **then**
    $Q_{J_r} \leftarrow Q_{J_r} \cup \{r\}$.
    $n_{J_r}^{\mathrm{act}} \leftarrow n_{J_r}^{\mathrm{act}} + 1$.
  end if
  for each $i \in \mathcal{B}$ **do**
    Call $\mathrm{PrivEAP}(i; r, J_r)$.
  end for
  if some PrivEAP call returned switch **then**
    Initialize PRIVATECONVERSION by (24).
    mode $\leftarrow$ ADV;
    continue
  end if
  Call $\mathrm{ActiveUpdate}(r)$.
  if $\mathrm{PrivBSC}(r) =$ fail **then**
    Initialize PRIVATECONVERSION by (24).
    mode $\leftarrow$ ADV.
  end if
end for
If $bH < T$, play arm 1 on rounds $bH + 1, \ldots, T$.

At the switch, erase every raw value in an active buffer Q_i, every EAP accumulator S_i^{eap}, and every unfinished online $\mathrm{AboveThreshold}$ instance. No discarded record in stochastic mode is used in a later release or decision.

For the regret analysis, define the *switch-declaration batch*

$$\kappa_{\mathrm{sw}} = \inf\Big\{ r \in [H] : \begin{array}{l}\text{PrivEAP returns switch, or} \\ \text{PrivBSC returns fail, at batch } r\end{array} \Big\}, \qquad \inf \varnothing = H + 1. \tag{25}$$

The first batch actually run by the adversarial mode is

$$\tau_{\mathrm{fb}} = \min\{\kappa_{\mathrm{sw}} + 1, H + 1\}. \tag{26}$$

C Auxiliary technical tools

C.1 Concurrent differential privacy

The Laplace mechanism and post-processing property can be dated back to Dwork & Roth (2014).

Theorem C.1 (Laplace mechanism, (Dwork & Roth, 2014)). *Let $q : X^n \to \mathbb{R}^d$ have ℓ_1-sensitivity*

$$\Delta_1(q) = \sup_{D \sim D'} \|q(D) - q(D')\|_1.$$

Then $q(D) + Z$, where the coordinates of Z are independent $\mathrm{Lap}(\Delta_1(q)/\epsilon)$ variables, is pure ϵ-DP.

Theorem C.2 (Post-processing, (Dwork & Roth, 2014)). *If M is ϵ-DP and Φ is any randomized map whose conditional law, given $M(D)$, does not otherwise depend on D, then $\Phi(M(D))$ is ϵ-DP.*

Definition C.3 (Interactive protocol, (Vadhan & Wang, 2021)). *An interactive protocol is a pair $(\mathcal{A}, \mathcal{M})$ of randomized algorithms. On private inputs $x_{\mathcal{A}}$ and $x_{\mathcal{M}}$, and independent random coins $R_{\mathcal{A}}$ and $R_{\mathcal{M}}$, the parties alternately generate messages*

$$m_{2s} = \mathcal{A}\big(x_{\mathcal{A}}, m_1, m_3, \ldots, m_{2s-1}, R_{\mathcal{A}}\big),$$
$$m_{2s+1} = \mathcal{M}\big(x_{\mathcal{M}}, m_0, m_2, \ldots, m_{2s}, R_{\mathcal{M}}\big),$$

starting with $s = 0$ and terminating when either party sends the public symbol halt. *A party's next message may depend on its input, its random coins, and all messages it has received, but not on the other party's hidden input or coins.*

Definition C.4 (View of a party, (Vadhan & Wang, 2021)). *In the protocol of Definition C.3, the view of $\mathcal{A}$ is*

$$\mathrm{View}_{\mathcal{A}}\langle \mathcal{A}(x_{\mathcal{A}}; R_{\mathcal{A}}), \mathcal{M}(x_{\mathcal{M}}; R_{\mathcal{M}})\rangle = \big(R_{\mathcal{A}}, x_{\mathcal{A}}, m_1, m_3, \ldots\big).$$

Thus the view contains $\mathcal{A}$'s input and coins and all messages received by $\mathcal{A}$. Its sent messages are recoverable from this information and the code of $\mathcal{A}$.

The scheduler's sent messages are determined by its input, random coins, and received replies. Thus its complete command–response transcript can be recovered by replaying the protocol from its view using its random coins.

Definition C.5 (Interactive differential privacy, (Vadhan & Wang, 2021)). *A randomized interactive mechanism $\mathcal{M}$ is (ϵ, δ)-differentially private if, for every pair of neighboring datasets $D \sim D'$, every (possibly randomized) interacting party $\mathcal{A}$, and every measurable set G of views,*

$$\mathbb{P}\{\mathrm{View}\langle \mathcal{A}, \mathcal{M}(D)\rangle \in G\} \le e^{\epsilon}\mathbb{P}\{\mathrm{View}\langle \mathcal{A}, \mathcal{M}(D')\rangle \in G\} + \delta.$$

The probability is over the independent coins of both parties. The case $\delta = 0$ gives pure interactive differential privacy.

In our application, the party denoted $\mathcal{A}$ by Vadhan & Wang (2021) is the bandit controller. We call it the *scheduler*, to avoid confusion with the oblivious loss adversary. The scheduler observes mechanism replies and adaptively chooses the next tagged query.

Definition C.6 (Concurrent composition, (Vadhan & Wang, 2021)). *Let $\mathcal{M}_0, \ldots, \mathcal{M}_{k-1}$ be interactive mechanisms. Their concurrent composition*

$$\mathrm{ConComp}(\mathcal{M}_0, \ldots, \mathcal{M}_{k-1})$$

is the interactive mechanism with private input $(x_0, \ldots, x_{k-1})$ and independent mechanism coins $(R_0, \ldots, R_{k-1})$ that operates as follows:

Let $e_{<s} = \{(a_t, i_t, m_t)\}_{t<s}$ denote the interaction history before round s. The history is interpreted as follows: At round t, the scheduler $\mathcal{A}$ sends query a_t to mechanism $\mathcal{M}_{i_t}$, and the mechanism $\mathcal{M}_{i_t}$ provides response m_t. We assume that the queries (and the target mechanism to send queries to) from the scheduler $\mathcal{A}$ depend on all preceding interaction history, its private input $x_{\mathcal{A}}$ and random coins $R_{\mathcal{A}}$. The responses provided by any mechanism $\mathcal{M}_i$ depend on its local interaction history, the current query from $\mathcal{A}$, its private input $x_{\mathcal{M}_i}$ and random coins $R_{\mathcal{M}_i}$. Precisely,

$$(a_s, i_s) = \mathcal{A}(x_{\mathcal{A}}, e_{<s}, R_{\mathcal{A}})$$
$$m_s = \mathcal{M}_{i_s}(x_{\mathcal{M}_{i_s}}, [e_{<s}]_{\mathcal{M}_{i_s}}, (a_s, i_s), R_{\mathcal{M}_{i_s}})$$

where $[e_{<s}]_{\mathcal{M}_i} = \{(a_t, i_t, m_t) : i_t = i, t < s\}$ denotes the local history of mechanism $\mathcal{M}_i$ before round s.

Theorem C.7 (Concurrent composition of pure interactive DP, (Vadhan & Wang, 2021)). *If $\mathcal{M}_0, \ldots, \mathcal{M}_{k-1}$ are pure $\epsilon_0, \ldots, \epsilon_{k-1}$ interactive-DP mechanisms, respectively, then*

$$\mathrm{ConComp}(\mathcal{M}_0, \ldots, \mathcal{M}_{k-1})$$

is pure $\sum_{i=0}^{k-1} \epsilon_i$ interactive DP.

We provide the following refinement using the fact that, along a common transcript, only mechanisms that inspect the changed record can contribute to the likelihood ratio.

Corollary C.8 (Concurrent composition with transcript-wise bounded overlap)**.** *Let $\mathcal{M}_0, \dots, \mathcal{M}_{k-1}$ be a fixed finite family of interactive mechanisms, all operating on the same global dataset D, and let $\mathcal{A}$ be a scheduler interacting, under the protocol specified by Definition C.6, with*

$$\mathrm{ConComp}(\mathcal{M}_0(D), \dots, \mathcal{M}_{k-1}(D)).$$

Suppose that $\mathcal{M}_i$ is pure ϵ_i interactive DP for every i.

Assume that the scheduler has no access to the private dataset except through mechanism replies and that the mechanisms use independent random coins. For a scheduler view v, let $\mathrm{Read}_i(v)$ denote the set of global data indices inspected by mechanism $\mathcal{M}_i$.

For neighboring datasets $D \sim D'$ differing in the record with index $x^\star$, let

$$\mathrm{Own}_v(x^\star) = \{i : x^\star \in \mathrm{Read}_i(v)\}$$

denote the set of mechanisms that inspect the data $x^\star$ along v. Assume that

$$\sum_{i \in \mathrm{Own}_v(x^\star)} \epsilon_i \le \bar{\epsilon} \qquad \text{for all } v. \tag{27}$$

Let $V_D(\mathcal{A}) = \mathrm{View}_{\mathcal{A}}\langle \mathcal{A}, \mathrm{ConComp}(\mathcal{M}_0(D), \dots, \mathcal{M}_{k-1}(D))\rangle$. Then, for every measurable set G of scheduler views,

$$\mathbb{P}\{V_D \in G\} \le e^{\bar{\epsilon}} \mathbb{P}\{V_{D'} \in G\}. \tag{28}$$

Thus the resulting scheduler-view mechanism $D \mapsto V_D(\mathcal{A})$ is pure $\bar{\epsilon}$-DP.

Proof. We fix neighbouring datasets $D \sim D'$ differing at index $x^\star$ and a scheduler $\mathcal{A}$. For $Z \in \{D, D'\}$, let $\mathbb{P}_Z$ be the probability measure on the view of $\mathcal{A}$ induced by its interaction with $\mathrm{ConComp}(\mathcal{M}_0(Z), \dots, \mathcal{M}_{k-1}(Z))$. Fix a realized view v and let $M = M(v)$ denote the number of mechanism replies in v. Replaying $\mathcal{A}$ using its random coins $R_{\mathcal{A}}$ recovers the complete query–response sequence $e_{<M+1}$. Write $h_{<t} = (R_{\mathcal{A}}, e_{<t})$. For each mechanism $\mathcal{M}_i$ and $Z \in \{D, D'\}$, define $\mathbb{Q}_{i,Z}$ as the probability measure of $\mathrm{View}_{\mathcal{A}}\langle \mathcal{A}, \mathcal{M}_i(Z)\rangle$.

We now derive the likelihood identity explicitly. For $t \le M + 1$, let

$$\Sigma_{t,Z}(\cdot \mid h_{<t})$$

denote the conditional law of the t-th scheduler command from $\mathcal{A}$ on dataset Z. For $t \le M$, let

$$\mathsf{K}_{t,Z}(\cdot \mid h_{<t}, (a_t, i_t))$$

denote the conditional law of the corresponding mechanism reply from $\mathcal{M}_{i_t}$. The chain rule for Radon–Nikodym derivatives gives

$$\begin{aligned}\frac{d\mathbb{P}_D}{d\mathbb{P}_{D'}}(v) = &\left[\prod_{t=1}^{M+1} \frac{d\Sigma_{t,D}(\cdot \mid h_{<t})}{d\Sigma_{t,D'}(\cdot \mid h_{<t})}(a_t, i_t)\right] \\ &\times \prod_{t=1}^{M} \frac{d\mathsf{K}_{t,D}(\cdot \mid h_{<t}, (a_t, i_t))}{d\mathsf{K}_{t,D'}(\cdot \mid h_{<t}, (a_t, i_t))}(m_t).\end{aligned} \tag{29}$$

Since the scheduler $\mathcal{A}$ has no dataset input, its command is determined by the common history and its coins. Thus

$$\Sigma_{t,D}(\cdot \mid h_{<t}) = \Sigma_{t,D'}(\cdot \mid h_{<t}), \qquad t = 1, \dots, M+1,$$

so every command factor in (29) satisfies

$$\frac{d\Sigma_{t,D}(\cdot \mid h_{<t})}{d\Sigma_{t,D'}(\cdot \mid h_{<t})}(a_t, i_t) = 1. \tag{30}$$

Consequently,

$$\frac{d\mathbb{P}_D}{d\mathbb{P}_{D'}}(v) = \prod_{t=1}^{M} \frac{d\mathsf{K}_{t,D}(\cdot \mid h_{<t}, (a_t, i_t))}{d\mathsf{K}_{t,D'}(\cdot \mid h_{<t}, (a_t, i_t))}(m_t). \tag{31}$$

We define

$$K_{i,Z}(\cdot \mid [e_{<t}]_{\mathcal{M}_i}, (a_t, i))$$

to be the next-response kernel of $\mathcal{M}_i(Z)$ after local history $[e_{<t}]_{\mathcal{M}_i}$ and current local query (a_t, i). By Definition C.6, the mechanism random coin strings $R_0, \ldots, R_{k-1}$ are independent and each mechanism uses only its own local history. Therefore, for $Z \in \{D, D'\}$,

$$\mathsf{K}_{t,Z}\big(\cdot \mid h_{<t}, (a_t, i_t)\big) = K_{i_t,Z}\big(\cdot \mid [e_{<t}]_{\mathcal{M}_{i_t}}, (a_t, i_t)\big). \tag{32}$$

For mechanism $\mathcal{M}_i$, let $s_1 < s_2 < \cdots < s_{n_i}$ denote the rounds at which the scheduler addresses i. Apply the chain rule to the two reference interactions defining $\mathbb{Q}_{i,D}$ and $\mathbb{Q}_{i,D'}$. Every scheduler-command factor cancels as in (30), giving the local identity

$$\frac{d\mathbb{Q}_{i,D}}{d\mathbb{Q}_{i,D'}}(v) = \prod_{j=1}^{n_i} \frac{dK_{i,D}(\cdot \mid [e_{<s_j}]_{\mathcal{M}_i}, (a_{s_j}, i))}{dK_{i,D'}(\cdot \mid [e_{<s_j}]_{\mathcal{M}_i}, (a_{s_j}, i))}(m_{s_j}). \tag{33}$$

Substituting (32) into (31), and then reindexing the finite product by the pairs (i, j), yields

$$\begin{aligned}\frac{d\mathbb{P}_D}{d\mathbb{P}_{D'}}(v) &= \prod_{i=0}^{k-1} \prod_{j=1}^{n_i} \frac{dK_{i,D}(\cdot \mid [e_{<s_j}]_{\mathcal{M}_i}, (a_{s_j}, i))}{dK_{i,D'}(\cdot \mid [e_{<s_j}]_{\mathcal{M}_i}, (a_{s_j}, i))}(m_{s_j}) \\ &= \prod_{i=0}^{k-1} \frac{d\mathbb{Q}_{i,D}}{d\mathbb{Q}_{i,D'}}(v). \end{aligned} \tag{34}$$

The first equality uses the isolated mechanism kernels and independent mechanism random coins from Definition C.6; the second is exactly (33).

If $i \notin \mathrm{Own}_v(x^\star)$, then D and D' agree on every record inspected by $\mathcal{M}_i$ along the prescribed path. With access to identical data and dependency on identical local history, we have $\mathbb{Q}_{i,D}$ and $\mathbb{Q}_{i,D'}$ being identical probability measures.

If $i \in \mathrm{Own}_v(x^\star)$, pure ϵ_i-DP of $\mathcal{M}_i$ gives

$$\frac{d\mathbb{Q}_{i,D}}{d\mathbb{Q}_{i,D'}} \le e^{\epsilon_i} \qquad \mathbb{Q}_{i,D'}\text{-almost surely.}$$

Consequently, by substitution into (34),

$$\frac{d\mathbb{P}_D}{d\mathbb{P}_{D'}}(v) \le \exp\left(\sum_{i \in \mathrm{Own}_v(x^\star)} \epsilon_i\right) \le e^{\bar{\epsilon}}.$$

Integrating over G proves (28). □

C.2 Sparse vector / AboveThreshold

Theorem C.9 (Online AboveThreshold, Dwork & Roth (2014))**.** *Let D be a dataset and let $q_1, q_2, \ldots, q_M$ be an adaptively generated stream of query functions $q_m : \mathcal{D} \to \mathbb{R}$, each with sensitivity at most Δ_{sv}. The complete response stream from* AboveThreshold

$$(B(1), \ldots, B(\tau)), \qquad B(m) \in \{\bot, \top\},$$

where τ is the first fire time or the last submitted query, is pure ϵ_{sv}-DP.

One standard implementation samples

$$\eta \sim \mathrm{Lap}(2\Delta_{\mathrm{sv}}/\epsilon_{\mathrm{sv}}), \qquad \xi_m \overset{\mathrm{iid}}{\sim} \mathrm{Lap}(4\Delta_{\mathrm{sv}}/\epsilon_{\mathrm{sv}}),$$

sets the private threshold to $\theta + \eta$, and returns

$$B(m) = \top \quad \Longleftrightarrow \quad q_m(D) + \xi_m \ge \theta + \eta.$$

For every $\delta \in (0, 1)$, let

$$\alpha_{M,\delta}(\Delta_{\mathrm{sv}}) = 8\,\frac{\Delta_{\mathrm{sv}} \log(2M/\delta)}{\epsilon_{\mathrm{sv}}}.$$

On the event where

$$|\eta| \le \frac{\alpha_{M,\delta}(\Delta_{\mathrm{sv}})}{2}, \qquad \max_{m \le M} |\xi_m| \le \frac{\alpha_{M,\delta}(\Delta_{\mathrm{sv}})}{2}, \tag{35}$$

which has probability at least $1-\delta$, every submitted query satisfies

$$B(m) = \perp \Longrightarrow q_m(D) \le \theta + \alpha_{M,\delta}(\Delta_{\mathrm{sv}}), \tag{36}$$

$$B(m) = \top \Longrightarrow q_m(D) \ge \theta - \alpha_{M,\delta}(\Delta_{\mathrm{sv}}). \tag{37}$$

In particular, if a submitted query is at least $\theta + \alpha_{M,\delta}(\Delta_{\mathrm{sv}})$, the mechanism cannot answer $\perp$ to that query.

C.3 Adversarial mode learner

For a bandit algorithm B, let $\widetilde{R}_B(n, K, \lambda)$ denote the worst-case expected regret when B is run for n rounds on feedback of the form

$$\widetilde{\ell}_s(i) = \ell_s(i) + Z_s(i), \qquad Z_s(i) \overset{\text{iid}}{\sim} \mathrm{Lap}(\lambda),$$

while regret is measured against the underlying bounded losses $\ell_s(i) \in [0,1]$.

Theorem C.10 (Asi et al. (2025))**.** *Let B be any bandit algorithm, let $\tau \ge 1$ be an integer batch length, and let $\eta \in (0,1]$.* PrivateConversion *constructs a pure η-DP bandit algorithm A_τ satisfying*

$$R_{A_\tau}(n, K) \le \tau\, \widetilde{R}_B\left(\left\lfloor \frac{n}{\tau} \right\rfloor, K, \frac{1}{\eta\tau}\right) + \tau. \tag{38}$$

The mechanism plays one arm throughout each fixed batch of length τ, adds $\mathrm{Lap}(1/(\eta\tau))$ noise to the selected batch-average loss, and passes that scalar to B.

Corollary C.11 (Private HTINF conversion, Asi et al. (2025))**.** *Let $n \ge 1$, $K \ge 2$, and $\eta \in [1/n, 1]$. In Theorem C.10, choose*

$$B = \mathrm{HTINF}(\alpha = 2, \sigma = \sqrt{6}), \qquad \tau = \left\lceil \frac{1}{\eta} \right\rceil.$$

The resulting algorithm is pure η-DP and satisfies

$$R_{A_\tau}(n, K) \le 208\sqrt{\frac{Kn}{\eta}} + \frac{2}{\eta}. \tag{39}$$

In particular,

$$R_{A_\tau}(n, K) = O\left(\sqrt{\frac{Kn}{\eta}} + \frac{1}{\eta}\right).$$

D Privacy analysis

D.1 Concurrent composition of mechanisms

We index an EAP phase by $e = (i, \ell)$, where ℓ is its phase number after the elimination of arm i.

Definition D.1 (Batch-average loss table and chronological output)**.** *For a loss table $L = (\ell_{t,i})_{t \le T, i \le K}$, define*

$$g_{r,i}(L) = \frac{1}{b} \sum_{t \in B_r} \ell_{t,i}, \qquad r \in [H],\ i \in [K]. \tag{40}$$

At batch r, the chronological output is

$$\Xi_r = (Y_r^A, O_r^S), \qquad \mathsf{Tr}(L) = (\Xi_1, \ldots, \Xi_H), \tag{41}$$

where

$$Y_r^A = \mathrm{BatchRelease}(r, J_r),$$

and $O_r^S = (O_r^{\mathrm{act}}, O_r^{\mathrm{eap}})$ lists the stochastic private outputs emitted at batch r, defined as follows:

If stage m completes at batch r, let $\mathcal{A}$ denote the active set before that stage's eliminations and define $O_r^{\mathrm{act}} = ((i,m), Y_{i,m}^S)_{i\in\mathcal{A}}$, where

$$Y_{i,m}^S = \mathrm{StageRelease}(C_{i,m}). \tag{42}$$

Otherwise, $O_r^{\mathrm{act}} = \varnothing$.

Define $O_r^{\mathrm{eap}} = ((i,\ell), j, B_{i,\ell}(j))_{i\in\mathcal{B}'}$, where $\mathcal{B}' \subseteq \mathcal{B}$ denotes the set of eliminated arms that submit a query at batch r to the AboveThreshold instance in their ongoing EAP phase. $B_{i,\ell}(j) \in \{\bot, \top\}$ is the response emitted at the j-th audit checkpoint of the running phase $e = (i,\ell)$. If no query is submitted, $O_r^{\mathrm{eap}} = \varnothing$.

Algorithm 1 implements a scheduler interacting with the concurrent composition of the following finite family of potential mechanisms. Unused slots remain inactive.

Define the deterministic set of potential mechanism slots

$$\begin{aligned}\mathcal{Q}_0 = \{(\mathsf{Adv}_r, r) : r \in [H]\} \cup \{(\mathsf{Agg}_{i,m}, i, m) : i \in [K],\ m = 0, \dots, H\} \\ \cup \{(\mathsf{AT}_{i,\ell}, i, \ell) : i \in [K],\ \ell = 1, \dots, H\}.\end{aligned} \tag{43}$$

The slot mechanisms have the following interfaces:

P1. *Adversarial batch slot.* Slot (Adv_r, r) is a mechanism to report privatized loss values at batch r. It has a trivial initialization space. After activation, the scheduler submits the selected arm J_r as the mechanism's unique ordinary query. The slot then emits

$$\mathsf{Adv}_r(L; j) = \mathrm{BatchRelease}(r, j). \tag{44}$$

P2. *Active-aggregate slot.* Slot $(\mathsf{Agg}_{i,m}, i, m)$ is a mechanism to update private empirical means of active arm i for stage m. It is initialized by the transcript-measurable message

$$(i, m, q_m).$$

After activation, an ordinary query may be either a batch index r or a closing description $C_{i,m}$. On an index query r, the primitive reads the fixed arm-i batch sum

$$X_{r,i} = \sum_{t\in B_r} \ell_{t,i} = bg_{r,i}(L)$$

and stores it in a hidden accumulator. The scheduler routes a batch to this slot only when $J_r = i$. When the public queue counts show that the synchronized stage is complete, the scheduler submits the finite set $C_{i,m}$ of the first q_m routed indices as a closing query description. The mechanism then emits

$$\mathsf{Agg}_{i,m}(L) = \mathrm{StageRelease}(C_{i,m}). \tag{45}$$

If the queue is discarded before a closing query, the primitive erases its hidden accumulator and emits no aggregate.

P3. *EAP-phase slot.* Slot $(\mathsf{AT}_{i,\ell}, i, \ell)$ is the $\mathrm{AboveThreshold}$ mechanism contained in the ℓ-th phase of EAP of an eliminated arm i. It is initialized at phase start by

$$\big(i, \ell, N_e, p_e, \widetilde{\ell}_i^{\mathrm{ref}}, \widetilde{\Delta}_i, \theta_e, q_e, U_e, Q_e^{\mathrm{AT}}\big). \tag{46}$$

It emits $B_e(j) \in \{\bot, \top\}$ from the AboveThreshold mechanism according to Algorithm 6.

All action-sampling coins, batch-release noises, stage-release noises, and $\mathrm{AboveThreshold}$ noises are mutually independent. A potential slot is activated only when its corresponding batch, stage, or phase is reached, and an inactive slot emits no reply. The deterministic family $\mathcal{Q}_0$ contains every possible activation because there are at most H complete batches.

The scheduler sends arm identifiers, batch indices, aggregate-closing descriptions, and query-function descriptions. These descriptions and all data-access indices are recoverable from its view by replaying according to Algorithm 7.

D.2 PRIVACY OF MECHANISMS

Lemma D.2 (Uniform interactive privacy of the primitives)**.** *We have the following privacy guarantees:*

(i) Adv_r *is pure* ϵ_A*-DP;*

(ii) $\mathsf{Agg}_{i,m}$ *is pure* ϵ_S*-DP;*

(iii) $\mathsf{AT}_{i,\ell}$ *is pure* ϵ_S*-DP.*

Proof. Throughout the proof, let L and L' be neighbouring loss tables that differ at round $t^\star$. If $bH < t^\star \le T$, Algorithm 7 plays arm 1 regardless of the loss value observed, and L and L' agree on all loss values in the complete batches $r \in [H]$. All primitive output laws are then identical. Hence assume that $t^\star \in B_{r^\star}$ for a unique $r^\star \in [H]$.

For (i), fix an arbitrary batch slot r and an arbitrary arm j. We have

$$\begin{aligned}\big|g_{r,j}(L) - g_{r,j}(L')\big| &= \left|\frac{1}{b}\sum_{t\in B_r}\big(\ell_{t,j} - \ell'_{t,j}\big)\right| \\ &= \frac{\mathbf{1}\{r = r^\star\}}{b}\big|\ell_{t^\star,j} - \ell'_{t^\star,j}\big| \\ &\le \frac{1}{b}.\end{aligned}$$

Moreover,

$$\Delta_1(g_{r,j}) = \sup_{L\sim L'}\big|g_{r,j}(L) - g_{r,j}(L')\big| = \frac{1}{b}, \qquad \frac{\Delta_1(g_{r,j})}{\epsilon_A} = \frac{1}{b\epsilon_A}.$$

Hence (44) is pure ϵ_A-DP by Theorem C.1.

For (ii), fix an arbitrary stage m, an arbitrary active arm i, and an admissible routing and closing description $C_{i,m}$. A neighbouring change affects at most one accepted batch sum. Using the fact that each batch is accepted at most once,

$$\begin{aligned}\left|\sum_{r\in C_{i,m}} X_{r,i}(L) - \sum_{r\in C_{i,m}} X_{r,i}(L')\right| &= \left|\sum_{r\in C_{i,m}}\sum_{t\in B_r}\big(\ell_{t,i} - \ell'_{t,i}\big)\right| \\ &= \mathbf{1}\{r^\star \in C_{i,m}\}\big|\ell_{t^\star,i} - \ell'_{t^\star,i}\big| \\ &\le 1.\end{aligned}$$

The aggregate in (45) adds $\mathrm{Lap}(1/\epsilon_S)$ noise and is therefore pure ϵ_S-DP by Theorem C.1.

For (iii), the desired conclusion follows directly from Theorem C.9. □

D.3 PRIVACY OF PrivDaS

Theorem D.3 (Pure global DP)**.** *Algorithm 1 is pure global* ϵ_0*-DP, where* $\epsilon_0 = \min\{\epsilon, 1\}$*, and hence is pure global* ϵ*-DP.*

Proof. Let $L \sim L'$ differ only at original round $t^\star$. If $t^\star > bH$, then the changed record lies in the final incomplete batch. Algorithm 7 plays the fixed arm 1 on that batch and never submits its loss to a private primitive, so the action laws under L and L' are identical. Hence assume $t^\star \le bH$ and let $r^\star$ be the unique batch with $t^\star \in B_{r^\star}$.

Let R_{PrivDaS} denote the random coins of PrivDaS. Denote the complete view of PrivDaS by

$$V(L) = \mathrm{View}\langle \mathrm{PrivDaS}, \mathrm{ConComp}(\mathcal{Q}_0)\rangle = (R_{\mathrm{PrivDaS}}, \mathsf{Tr}(L)). \tag{47}$$

It remains to verify the hypotheses of Corollary C.8 for this scheduler and mechanism family. We note that every mechanism $q \in \mathcal{Q}_0$ is either ϵ_A-DP or ϵ_S-DP by Lemma D.2. The random coins are independent by the implementation of Algorithm 7. Fix a realization of view v. Replaying PrivDaS using the random coins R_{PrivDaS} determines the selected arm $i = J_{r^\star}$. At batch $r^\star$, we have two cases:

1. The algorithm has already switched to adversarial mode. In this case, the selected loss value $\ell_{t^\star,i}$ is only accessed by the mechanism $(\mathsf{Adv}_{r^\star}, r^\star)$, because no mechanism for stochastic mode remains active in adversarial mode. Therefore,
$$\sum_{q\in \mathrm{Own}_v(t^\star)} \epsilon_q = \epsilon_A.$$

2. The algorithm remains in stochastic mode. In this case, the selected arm i is either active or eliminated, not both. Therefore, its loss value $\ell_{t^\star,i}$ is accessed by $(\mathsf{Adv}_{r^\star}, r^\star)$ and at most one of the following:
   (a) $(\mathsf{Agg}_{i,m}, i, m)$, if i is active and at stage m;
   (b) $(\mathsf{AT}_{i,\ell}, i, \ell)$, if i is eliminated and at phase ℓ.

   Active queues discard unused batches at stage boundaries, and EAP phases use disjoint calendar intervals. Thus the batch is not reused by a later stochastic slot. In either sub-case,
$$\sum_{q\in \mathrm{Own}_v(t^\star)} \epsilon_q \le \epsilon_A + \epsilon_S.$$

Hence, in both cases,
$$\sum_{q\in \mathrm{Own}_v(t^\star)} \epsilon_q \le \epsilon_A + \epsilon_S = \epsilon_0.$$

Therefore, by Corollary C.8, the complete view $V(L)$ defined in (47) is pure ϵ_0-DP.

Recovering the chronological transcript and the actions. We finally describe the post-processing explicitly. From a realization v of $V(L)$, replay PrivDaS using the scheduler coins and received replies contained in v. This recovers the arms J_r sampled for each batch $r \in [H]$.

The action sequence is thus reconstructed by the deterministic map
$$I_t(v) = \begin{cases} J_r(v), & t \in B_r,\ r \in [H], \\ 1, & bH < t \le T. \end{cases} \tag{48}$$

Applying Theorem C.2 to (48) proves pure global ϵ_0-DP. Since $\epsilon_0 \le \epsilon$, PrivDaS is also pure global ϵ-DP. □

E Proofs of lemmas

We now give proofs of the lemmas that support the proofs of our main results. The complete regret bounds are proved after these lemmas.

E.1 Good events

Let $\mathscr{F}_r$ be the sigma-field generated by the full primitive history through the end of batch r: actions, raw batch losses, all private releases and checkpointed AboveThreshold responses, the private coins already drawn by activated mechanisms, and all algorithmic parameters. The fresh action-sampling coin at batch r remains independent conditional on $\mathscr{F}_{r-1}$, so
$$\Pr\{J_r = i \mid \mathscr{F}_{r-1}\} = p_{r,i}. \tag{49}$$

Let $\mathcal{R}_{\mathrm{act}}$ be the random set of completed active-aggregate releases. For $a = (i, m) \in \mathcal{R}_{\mathrm{act}}$, let $C_a \subseteq [H]$ be its batch indices and $n_a = b|C_a|$.

For a fixed batch-average loss table, define
$$\bar{g}_{r,i} = \frac{1}{r}\sum_{s\le r} g_{s,i}, \qquad G_r = \sum_{s\le r} g_{s,J_s}, \qquad \Lambda_A(r) = 2\sqrt{2rL_H} + 2L_H.$$

Define

$$\mathcal{G}_{\mathrm{bsc}} = \left\{\forall r < \tau_{\mathrm{fb}},\ \forall i \in \mathcal{A}_r : |\widehat{\nu}^A_{r,i} - \bar{g}_{r,i}| \le w_A(r)\right\} \cap \left\{\forall r < \tau_{\mathrm{fb}} : |\widehat{G}_r - G_r| \le \Lambda_A(r)\right\}. \quad (50)$$

For a stochastic instance, define the active-aggregate concentration event

$$\mathcal{G}^{\mathrm{sto}}_{\mathrm{act}} = \bigcap_{a=(i,m)\in\mathcal{R}_{\mathrm{act}}} \left\{\left|\frac{Y^S_a}{n_a} - \mu_i\right| \le \rho(n_a)\right\}, \quad (51)$$

and the raw-loss concentration event

$$\mathcal{G}^{\mathrm{sto}}_{\mathrm{raw}} = \bigcap_{r\le H,\, i\in[K]} \left\{\left|\sum_{s=1}^{r}(g_{s,i} - \mu_i)\right| \le \sqrt{\frac{rL_H}{b}}\right\} \cap \bigcap_{r\le H} \left\{\left|\sum_{s=1}^{r}(g_{s,J_s} - \mu_{J_s})\right| \le \sqrt{\frac{rL_H}{b}}\right\}. \quad (52)$$

Let

$$p_{r,i} = \Pr(J_r = i \mid \mathscr{F}_{r-1}),$$

and, for every interval $I = [s,t] \subseteq [H]$, define

$$N^{\mathrm{pull}}_i(I) = \sum_{r=s}^{t} \mathbf{1}\{J_r = i\}, \qquad \Pi_i(I) = \sum_{r=s}^{t} p_{r,i}.$$

We then define the batch-count event

$$\mathcal{G}^{\mathrm{sto}}_{\mathrm{count}} = \bigcap_{i\in[K]} \bigcap_{1\le s\le t\le H} \left\{\frac{1}{2}\Pi_i([s,t]) - 6L_H \le N^{\mathrm{pull}}_i([s,t]) \le 2\Pi_i([s,t]) + 6L_H\right\}. \quad (53)$$

Let $F^{\mathrm{eap}}_i(H)$ be the number of PrivEAP fires accumulated by arm i throughout the horizon, and define the fire-count event

$$\mathcal{G}^{\mathrm{sto}}_{\mathrm{ctr}} = \bigcap_{i=1}^{K} \{F^{\mathrm{eap}}_i(H) < E_0\}. \quad (54)$$

Finally, set

$$\mathcal{G}^{\mathrm{sto}} = \mathcal{G}^{\mathrm{sto}}_{\mathrm{act}} \cap \mathcal{G}_{\mathrm{bsc}} \cap \mathcal{G}^{\mathrm{sto}}_{\mathrm{raw}} \cap \mathcal{G}^{\mathrm{sto}}_{\mathrm{count}}, \qquad \mathcal{G}^{\mathrm{adv}} = \mathcal{G}_{\mathrm{bsc}}, \quad (55)$$

$$\mathcal{G}^{\mathrm{sto}}_{\mathrm{safe}} = \mathcal{G}^{\mathrm{sto}} \cap \mathcal{G}^{\mathrm{sto}}_{\mathrm{ctr}}. \quad (56)$$

E.2 Lemmas

Lemma E.1. *Let $U_1, \dots, U_n \in [0,1]$ be independent with common mean μ, and $Z \sim \mathrm{Lap}(1/\epsilon_S)$. For all $u > 0$,*

$$\mathbb{P}\left\{\left|\frac{\sum_{s=1}^{n} U_s + Z}{n} - \mu\right| > \sqrt{\frac{u}{2n}} + \frac{u}{\epsilon_S n}\right\} \le 3e^{-u}.$$

Proof. Let $\bar{U}_n = n^{-1}\sum_s U_s$. Hoeffding gives

$$\mathbb{P}\{|\bar{U}_n - \mu| > a\} \le 2e^{-2na^2}.$$

With $a = \sqrt{u/(2n)}$,

$$\mathbb{P}\{|\bar{U}_n - \mu| > \sqrt{u/(2n)}\} \le 2e^{-u}.$$

For $Z \sim \mathrm{Lap}(1/\epsilon_S)$,

$$\mathbb{P}\{|Z| > u/\epsilon_S\} = e^{-u}.$$

Hence, by

$$\left|\frac{\sum_s U_s + Z}{n} - \mu\right| \le |\bar{U}_n - \mu| + |Z|/n,$$

the claim follows. □

Lemma E.2. *Fix a stochastic instance and a potential active-aggregate slot $a = (i, m)$ with quota q_m. Let σ_a be the batch at which its active stage starts, and let*

$$\tau_{a,1} < \cdots < \tau_{a,q_m}$$

be the successive batches in that stage at which arm i is pulled, on the event that the stage completes. If the slot releases

$$Y_a^S = b \sum_{k=1}^{q_m} g_{\tau_{a,k},i} + Z_a, \qquad Z_a \sim \mathrm{Lap}(1/\epsilon_S),$$

then, for every $u > 0$,

$$\Pr_\nu\left(a \text{ completes and } \left| \frac{Y_a^S}{bq_m} - \mu_i \right| > \sqrt{\frac{u}{2bq_m}} + \frac{u}{\epsilon_S bq_m} \right) \le 3e^{-u}. \tag{57}$$

Proof. For each batch r, let

$$\mathscr{P}_r = \sigma(\mathscr{F}_{r-1}, J_r)$$

be the pre-batch sigma-field: it contains the complete history through batch $r-1$ and the current action, but not the losses in the fresh batch B_r. For the fixed slot $a = (i, m)$, let $Q_{a,r-1}$ be the number of batch indices assigned to a before batch r, with $Q_{a,r-1} = 0$ before activation. Define

$$\begin{aligned} A_{a,r} = \mathbf{1}\{&a \text{ is the current active slot before batch } r\} \\ &\cdot \mathbf{1}\{Q_{a,r-1} < q_m\}\mathbf{1}\{J_r = i\}. \end{aligned}$$

Thus $A_{a,r}$ is $\mathscr{P}_r$-measurable, vanishes before activation and after the slot closes, and selects exactly the first q_m arm-i batches assigned to a; later queue overshoots, if any, have indicator zero.

Under the stochastic model, the potential losses in B_r are independent of $\mathscr{P}_r$, and the b variables $(\ell_{t,i})_{t\in B_r}$ are independent, $[0,1]$-valued, and have common mean μ_i. Hence Hoeffding's lemma gives, for every $\lambda \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E}_\nu\left[e^{\lambda(g_{r,i}-\mu_i)} \middle| \mathscr{P}_r \right] &= \prod_{t\in B_r} \mathbb{E}_\nu\left[e^{(\lambda/b)(\ell_{t,i}-\mu_i)} \right] \\ &\le \prod_{t\in B_r} e^{\lambda^2/(8b^2)} = e^{\lambda^2/(8b)}. \end{aligned} \tag{58}$$

Because $A_{a,r} \in \{0,1\}$ is $\mathscr{P}_r$-measurable, (58) implies

$$\mathbb{E}_\nu\left[\exp\left\{ \lambda A_{a,r}(g_{r,i} - \mu_i) - \frac{\lambda^2}{8b} A_{a,r} \right\} \middle| \mathscr{P}_r \right] \le 1. \tag{59}$$

Define

$$S_{a,n} = \sum_{r=1}^{n} A_{a,r}(g_{r,i} - \mu_i), \qquad N_{a,n} = \sum_{r=1}^{n} A_{a,r},$$

and, for fixed $\lambda \in \mathbb{R}$,

$$M_{a,n}(\lambda) = \exp\left\{ \lambda S_{a,n} - \frac{\lambda^2}{8b} N_{a,n} \right\}, \qquad M_{a,0}(\lambda) = 1.$$

Using the tower property and (59),

$$\begin{aligned} \mathbb{E}_\nu[M_{a,n}(\lambda) \mid \mathscr{F}_{n-1}] &= M_{a,n-1}(\lambda)\mathbb{E}_\nu\left[\mathbb{E}_\nu\left[e^{\lambda A_{a,n}(g_{n,i}-\mu_i) - \lambda^2 A_{a,n}/(8b)} \middle| \mathscr{P}_n \right] \middle| \mathscr{F}_{n-1} \right] \\ &\le M_{a,n-1}(\lambda). \end{aligned}$$

Therefore $(M_{a,n}(\lambda))_{n=0}^H$ is a nonnegative supermartingale and $\mathbb{E}_\nu M_{a,H}(\lambda) \le 1$.

On the event that a completes,

$$N_{a,H} = q_m, \qquad S_{a,H} = \sum_{k=1}^{q_m} (g_{\tau_{a,k},i} - \mu_i).$$

Consequently, for every $x > 0$ and $\lambda > 0$, Markov's inequality gives

$$\Pr_{\nu}\left(a \text{ completes and } \sum_{k=1}^{q_m}(g_{\tau_{a,k},i} - \mu_i) \geq x\right) \leq \exp\left\{-\lambda x + \frac{q_m\lambda^2}{8b}\right\}.$$

Optimizing at $\lambda = 4bx/q_m$ yields

$$\Pr_{\nu}\left(a \text{ completes and } \sum_{k=1}^{q_m}(g_{\tau_{a,k},i} - \mu_i) \geq x\right) \leq \exp\left\{-\frac{2bx^2}{q_m}\right\}.$$

Applying the same argument to $-S_{a,H}$ and taking $x = \sqrt{q_m u/(2b)}$ gives

$$\Pr_{\nu}\left(a \text{ completes and } \left|\frac{1}{q_m}\sum_{k=1}^{q_m} g_{\tau_{a,k},i} - \mu_i\right| > \sqrt{\frac{u}{2bq_m}}\right) \leq 2e^{-u}. \tag{60}$$

Finally, the slot-specific noise is independent of the loss process and satisfies

$$\Pr\{|Z_a| > u/\epsilon_S\} = e^{-u}.$$

Since

$$\left|\frac{Y_a^S}{bq_m} - \mu_i\right| \leq \left|\frac{1}{q_m}\sum_{k=1}^{q_m} g_{\tau_{a,k},i} - \mu_i\right| + \frac{|Z_a|}{bq_m}$$

on the completion event, a union bound with (60) proves (57). □

Lemma E.3 (Online AboveThreshold accuracy event)**.** *Consider one EAP phase with at most M sequential prefix queries $D(1), \ldots, D(M)$, each of sensitivity at most Δ_{sv}. Implement the online mechanism of Theorem C.9 with threshold noise η and query noises $(\xi_m)_{m\leq M}$. Define*

$$\mathcal{E}^{\mathrm{svt}}_{M,\delta} = \left\{|\eta| \leq \frac{\alpha_{M,\delta}(\Delta_{\mathrm{sv}})}{2}, \quad \max_{m\leq M}|\xi_m| \leq \frac{\alpha_{M,\delta}(\Delta_{\mathrm{sv}})}{2}\right\}, \tag{61}$$

where

$$\alpha_{M,\delta}(\Delta_{\mathrm{sv}}) = 8\frac{\Delta_{\mathrm{sv}}\log(2M/\delta)}{\epsilon_S}.$$

Then

$$\Pr\{\mathcal{E}^{\mathrm{svt}}_{M,\delta}\} \geq 1 - \delta.$$

On this event, for every query actually submitted before the phase terminates,

$$B(m) = \bot \Longrightarrow D(m) \leq \theta + \alpha_{M,\delta}(\Delta_{\mathrm{sv}}), \tag{62}$$

$$B(m) = \top \Longrightarrow D(m) \geq \theta - \alpha_{M,\delta}(\Delta_{\mathrm{sv}}). \tag{63}$$

Proof. The two Laplace tails and a union bound give

$$\Pr\{|\eta| > \alpha/2\} = \exp\left(-\frac{\alpha\epsilon_S}{4\Delta_{\mathrm{sv}}}\right) \leq \delta/2,$$

$$\Pr\left\{\max_{m\leq M}|\xi_m| > \alpha/2\right\} \leq M\exp\left(-\frac{\alpha\epsilon_S}{8\Delta_{\mathrm{sv}}}\right) = \delta/2,$$

where $\alpha = 8\Delta_{\mathrm{sv}}\log(2M/\delta)/\epsilon_S$. Hence the noise event has probability at least $1 - \delta$.

If $B(m) = \bot$, then $D(m) + \xi_m < \theta + \eta$, so on the noise event

$$D(m) < \theta + \eta - \xi_m \leq \theta + |\eta| + |\xi_m| \leq \theta + \alpha.$$

If $B(m) = \top$, the inequality is reversed and

$$D(m) \geq \theta + \eta - \xi_m \geq \theta - |\eta| - |\xi_m| \geq \theta - \alpha.$$

This proves (62)–(63). □

Lemma E.4. *Let $(M_u, \mathcal{F}_u)_{u=0}^t$ be a martingale with*

$$M_0 = 0, \qquad X_s = M_s - M_{s-1}, \quad s = 1, \dots, t.$$

Suppose that there are nonnegative predictable quantities v_s, c_s for $s = 1, \dots, t$ such that, for every integer $q \geq 2$,

$$\mathbb{E}[|X_s|^q \mid \mathcal{F}_{s-1}] \leq \frac{q!}{2} v_s c_s^{q-2}, \qquad s = 1, \dots, t. \tag{64}$$

Let

$$\bar{V}_t \geq \sum_{s=1}^{t} v_s, \qquad \bar{C}_t \geq \max_{1 \leq s \leq t} c_s \tag{65}$$

be deterministic bounds. Then, for every $x > 0$,

$$\mathbb{P}\big\{|M_t| > 2\sqrt{\bar{V}_t x} + 2\bar{C}_t x\big\} \leq 2e^{-x}. \tag{66}$$

In addition, for any finite index set $\mathfrak{I}$ and martingales $(M_j)_{j \in \mathfrak{I}}$ with thresholds $B_j(x) = 2\sqrt{\bar{V}_j x} + 2\bar{C}_j x$, we have

$$\mathbb{P}\big\{\exists j \in \mathfrak{I} : |M_j| > B_j(x)\big\} \leq 2|\mathfrak{I}|e^{-x}. \tag{67}$$

In particular, choosing $x = \log(2|\mathfrak{I}|/\beta)$ makes the failure probability at most β.

Proof. Fix s and let $|\theta|c_s < 1$. The exponential power series is absolutely integrable under (64): indeed,

$$\sum_{q=2}^{\infty} \frac{|\theta|^q}{q!} \mathbb{E}[|X_s|^q \mid \mathcal{F}_{s-1}] \leq \frac{\theta^2 v_s}{2} \sum_{q=2}^{\infty} (|\theta|c_s)^{q-2} < \infty.$$

Since $\mathbb{E}[X_s \mid \mathcal{F}_{s-1}] = 0$,

$$\begin{aligned}
\mathbb{E}[e^{\theta X_s} \mid \mathcal{F}_{s-1}] &= 1 + \sum_{q=2}^{\infty} \frac{\theta^q}{q!} \mathbb{E}[X_s^q \mid \mathcal{F}_{s-1}] \\
&\leq 1 + \sum_{q=2}^{\infty} \frac{|\theta|^q}{q!} \mathbb{E}[|X_s|^q \mid \mathcal{F}_{s-1}] \\
&\leq 1 + \frac{\theta^2 v_s}{2} \sum_{q=2}^{\infty} (|\theta|c_s)^{q-2} \\
&= 1 + \frac{\theta^2 v_s}{2(1 - |\theta|c_s)} \\
&\leq \exp\left(\frac{\theta^2 v_s}{2(1 - |\theta|c_s)} \right).
\end{aligned}$$

Take $0 < \theta < 1/\bar{C}_t$. Since $c_s \leq \bar{C}_t$ for $s \leq t$,

$$\mathbb{E}[e^{\theta X_s} \mid \mathcal{F}_{s-1}] \leq \exp\left(\frac{\theta^2 v_s}{2(1 - \theta\bar{C}_t)} \right). \tag{68}$$

Define, for $0 \leq u \leq t$,

$$W_u = \exp\left(\theta M_u - \frac{\theta^2 \sum_{s=1}^{u} v_s}{2(1 - \theta\bar{C}_t)} \right). \tag{69}$$

For $u = 0$, the convention gives

$$W_0 = \exp(0) = 1.$$

Moreover, by (68),

$$\mathbb{E}[W_u \mid \mathcal{F}_{u-1}] = W_{u-1} \exp\left(-\frac{\theta^2 v_u}{2(1 - \theta\bar{C}_t)} \right) \mathbb{E}[e^{\theta X_u} \mid \mathcal{F}_{u-1}] \leq W_{u-1}.$$

Thus $(W_u)_{u \leq t}$ is a nonnegative supermartingale. Taking expectations and iterating,

$$\mathbb{E}W_t \leq \mathbb{E}W_{t-1} \leq \cdots \leq \mathbb{E}W_0 = 1.$$

Markov's inequality gives, for every $a > 0$,

$$\mathbb{P}\{M_t \geq a\} \leq \exp\left(-\theta a + \frac{\theta^2 \bar{V}_t}{2(1-\theta \bar{C}_t)}\right). \tag{70}$$

Set

$$a = 2\sqrt{\bar{V}_t x} + 2\bar{C}_t x. \tag{71}$$

If $\bar{V}_t = 0$, then $v_s = 0$ for all $s \leq t$, and the $q = 2$ case of (64) gives $X_s = 0$ almost surely for all $s \leq t$; the tail bound is then trivial. Assume henceforth that $\bar{V}_t > 0$ and take

$$\theta = \frac{\sqrt{x}}{\sqrt{\bar{V}_t} + \bar{C}_t\sqrt{x}}.$$

Then $0 < \theta < 1/\bar{C}_t$ when $\bar{C}_t > 0$; the case $\bar{C}_t = 0$ follows by taking $\theta = \sqrt{x/\bar{V}_t}$. For $\bar{C}_t > 0$,

$$\begin{aligned}
\theta a &= \frac{\sqrt{x}}{\sqrt{\bar{V}_t} + \bar{C}_t\sqrt{x}}\left(2\sqrt{\bar{V}_t x} + 2\bar{C}_t x\right) = 2x, \\
1 - \theta\bar{C}_t &= \frac{\sqrt{\bar{V}_t}}{\sqrt{\bar{V}_t} + \bar{C}_t\sqrt{x}}, \\
\frac{\theta^2 \bar{V}_t}{2(1-\theta\bar{C}_t)} &= \frac{x\bar{V}_t}{2(\sqrt{\bar{V}_t} + \bar{C}_t\sqrt{x})^2}\,\frac{\sqrt{\bar{V}_t} + \bar{C}_t\sqrt{x}}{\sqrt{\bar{V}_t}} \\
&= \frac{x\sqrt{\bar{V}_t}}{2(\sqrt{\bar{V}_t} + \bar{C}_t\sqrt{x})} \leq \frac{x}{2}.
\end{aligned}$$

Substituting in (70) gives

$$\mathbb{P}\{M_t \geq 2\sqrt{\bar{V}_t x} + 2\bar{C}_t x\} \leq e^{-3x/2} \leq e^{-x}. \tag{72}$$

The same argument applied to $-M_t$ gives the lower-tail bound, because (64) is unchanged when X_s is replaced by $-X_s$. Therefore

$$\mathbb{P}\big\{|M_t| > 2\sqrt{\bar{V}_t x} + 2\bar{C}_t x\big\} \leq 2e^{-x}. \tag{73}$$

Finally, for a finite index set $\mathfrak{I}$,

$$\mathbb{P}\big\{\exists j \in \mathfrak{I} : |M_j| > B_j(x)\big\} \leq \sum_{j\in\mathfrak{I}} \mathbb{P}\{|M_j| > B_j(x)\} \leq 2|\mathfrak{I}|e^{-x}. \tag{74}$$

Choosing $x = \log(2|\mathfrak{I}|/\beta)$ yields (67). □

Lemma E.5. *Let $(S_n, \mathcal{F}_n)_{n=0}^N$ be a martingale with $S_0 = 0$ and increments $X_n = S_n - S_{n-1}$ satisfying $|X_n| \leq 1$. Define the predictable quadratic-variation process by*

$$V_0 = 0, \qquad V_n = \sum_{j=1}^{n} \mathbb{E}[X_j^2 \mid \mathcal{F}_{j-1}], \qquad n = 1, \ldots, N.$$

Suppose that

$$V_N \leq v \quad \textit{almost surely},$$

where $v \geq 0$ is deterministic. Then, for every $x > 0$,

$$\Pr\left\{\max_{n\leq N} S_n \geq x\right\} \leq \exp\left(-\frac{x^2}{2(v + x/3)}\right). \tag{75}$$

Proof. If $v = 0$, then $V_N = 0$ almost surely. Since every summand in V_N is nonnegative, $\mathbb{E}[X_n^2 \mid \mathcal{F}_{n-1}] = 0$ for every n, so $X_n = 0$ almost surely and the claim is immediate. Hence assume $v > 0$. For $0 < \lambda < 3$, the elementary inequality

$$e^{\lambda z} \leq 1 + \lambda z + \frac{\lambda^2 z^2}{2(1-\lambda/3)}, \qquad |z| \leq 1,$$

implies

$$\mathbb{E}[e^{\lambda X_n} \mid \mathcal{F}_{n-1}] \le \exp\left(\frac{\lambda^2 \mathbb{E}[X_n^2 \mid \mathcal{F}_{n-1}]}{2(1-\lambda/3)}\right).$$

Consequently,

$$W_n = \exp\left(\lambda S_n - \frac{\lambda^2}{2(1-\lambda/3)} V_n\right), \qquad n = 0, \dots, N,$$

is a nonnegative supermartingale with $W_0 = 1$. Let

$$\tau = \inf\{n \le N : S_n \ge x\}, \qquad \inf \varnothing = N + 1.$$

On $\{\tau \le N\}$, we have $S_\tau \ge x$ and, because (V_n) is nondecreasing, $V_\tau \le V_N \le v$. Moreover, $\tau \wedge N \le N$ is a bounded stopping time. Optional stopping for the nonnegative supermartingale (W_n) therefore gives

$$\begin{aligned} 1 = W_0 &\ge \mathbb{E}[W_{\tau \wedge N}] \\ &\ge \mathbb{E}[W_\tau \mathbf{1}\{\tau \le N\}] \\ &\ge \exp\left(\lambda x - \frac{\lambda^2 v}{2(1-\lambda/3)}\right) \Pr\{\tau \le N\}. \end{aligned}$$

Rearranging and using $\{\tau \le N\} = \{\max_{n \le N} S_n \ge x\}$ yields

$$\Pr\left\{\max_{n \le N} S_n \ge x\right\} \le \exp\left(-\lambda x + \frac{\lambda^2 v}{2(1-\lambda/3)}\right).$$

Finally, choose $\lambda = x/(v + x/3) \in (0, 3)$. Then

$$-\lambda x + \frac{\lambda^2 v}{2(1-\lambda/3)} = -\frac{x^2}{2(v + x/3)},$$

which proves (75). □

Lemma E.6 (Probability of the core good events)**.** *We have*

$$\Pr_\nu((\mathcal{G}^{\mathrm{sto}})^c) \le \beta/2, \qquad \Pr_L((\mathcal{G}^{\mathrm{adv}})^c) \le \beta/2. \tag{76}$$

Proof. For active aggregates, enumerate the deterministic set of potential slots (i, m) with $i \in [K]$ and $m = 0, \dots, H$. Lemma E.2 with $u = L_H$ gives, for each slot,

$$\Pr_\nu\big((i,m) \text{ completes and violates } \mathcal{G}^{\mathrm{sto}}_{\mathrm{act}}\big) \le 3e^{-L_H}.$$

Therefore a union bound over at most $K(H+1)$ potential slots gives

$$\Pr_\nu((\mathcal{G}^{\mathrm{sto}}_{\mathrm{act}})^c) \le 3K(H+1)e^{-L_H}.$$

For the calendar estimator, let $\mathcal{A}_s^-$ denote the active set immediately before the action is drawn in batch s. This set is $\mathscr{F}_{s-1}$-measurable. Fix a deterministic pair (r, i) and define the stopped martingale

$$M_{r,i}^{\mathrm{stop}} = \sum_{s=1}^{r} \mathbf{1}\{s < \tau_{\mathrm{fb}},\ i \in \mathcal{A}_s^-\} \left(\frac{\mathbf{1}\{J_s = i\} Y_s^A}{p_{s,i}} - g_{s,i}\right).$$

The activity indicator is $\mathscr{F}_{s-1}$-measurable. Conditional on the past, the bracketed increment has mean zero; when the indicator is one, its conditional moments satisfy (64) with

$$v_s = 8/p_{s,i}, \qquad c_s = 4/p_{s,i},$$

and $p_{s,i} \ge 1/(2K)$. When the indicator is zero, take $v_s = c_s = 0$. Hence Lemma E.4 applies with deterministic bounds $\bar{V}_r = 16Kr$ and $\bar{C}_r = 8K$. On the event $r < \tau_{\mathrm{fb}}$ and $i \in \mathcal{A}_r$, the arm has been active throughout all earlier controller batches, so

$$M_{r,i}^{\mathrm{stop}} = \widehat{L}_{r,i}^A - \sum_{s \le r} g_{s,i}.$$

A union bound over the KH deterministic pairs (r, i) therefore yields

$$\left| \widehat{L}^A_{r,i} - \sum_{s \le r} g_{s,i} \right| \le 8\sqrt{KrL_H} + 16KL_H \le rw_A(r) \tag{77}$$

for all relevant (r, i), with probability at least $1 - 2KHe^{-L_H}$. Similarly, define

$$M^{G,\mathrm{stop}}_r = \sum_{s=1}^{r} \mathbf{1}\{s < \tau_{\mathrm{fb}}\} Z^A_s.$$

This is a stopped martingale; Lemma E.4 applies with $v_s = 2, c_s = 1$ when the indicator is one and zero otherwise. On $r < \tau_{\mathrm{fb}}$ it equals $\widehat{G}_r - G_r$. A union bound over $r \le H$ gives

$$|\widehat{G}_r - G_r| \le 2\sqrt{2rL_H} + 2L_H = \Lambda_A(r) \tag{78}$$

with probability at least $1 - 2He^{-L_H}$. Hence

$$\Pr_L(\mathcal{G}^c_{\mathrm{bsc}}) \le 2KHe^{-L_H} + 2He^{-L_H}.$$

For the raw stochastic losses, one batch average satisfies, conditionally on the past,

$$\mathbb{E}\Big[e^{\lambda(g_{r,i} - \mu_i)} \mid \mathscr{F}_{r-1}\Big] \le e^{\lambda^2/(8b)}.$$

Define

$$S_{n,i} = \sum_{s=1}^{n} (g_{s,i} - \mu_i), \qquad S_{0,i} = 0.$$

Iterating the one-batch bound gives, for every $\lambda \in \mathbb{R}$,

$$\mathbb{E}[e^{\lambda S_{n,i}}] \le e^{n\lambda^2/(8b)}.$$

Consequently, for every $x > 0$ and $\lambda > 0$, Markov's inequality gives

$$\Pr_\nu\{S_{r,i} \ge x\} \le \exp\left\{-\lambda x + \frac{r\lambda^2}{8b}\right\}.$$

Optimizing at $\lambda = 4bx/r$ yields

$$\Pr_\nu\{S_{r,i} \ge x\} \le \exp\left\{-\frac{2bx^2}{r}\right\}.$$

Applying the same calculation to $-S_{r,i}$, taking a union bound over the two tails, and substituting $x = \sqrt{rL_H/b}$ gives

$$\Pr_\nu\left\{\left|\sum_{s=1}^{r}(g_{s,i} - \mu_i)\right| > \sqrt{\frac{rL_H}{b}}\right\} \le 2e^{-2L_H},$$

for each fixed pair (r, i). The same one-batch moment-generating-function bound holds with $i = J_r$ after conditioning additionally on J_r, because J_r is selected before batch r is generated; the identical argument therefore applies to every played-loss prefix. Union-bounding over $KH+H$ prefixes gives

$$\Pr_\nu((\mathcal{G}^{\mathrm{sto}}_{\mathrm{raw}})^c) \le 2(KH + H)e^{-2L_H}.$$

For the count event, fix an arm i and an interval $I = [a, c] \subseteq [H]$. Write

$$X_n = \mathbf{1}\{J_n = i\}, \qquad \pi_n = \mathbb{E}[X_n \mid \mathscr{F}_{n-1}],$$

and, for $n = a - 1, \ldots, c$, define

$$N_n = \sum_{u=a}^{n} X_u, \qquad P_n = \sum_{u=a}^{n} \pi_u, \qquad M_n(\lambda) = \exp\{\lambda N_n - (e^\lambda - 1)P_n\},$$

with empty sums at $n = a-1$, so $M_{a-1}(\lambda) = 1$. Conditional on $\mathscr{F}_{n-1}$, the variable X_n is Bernoulli with mean π_n, and hence

$$\mathbb{E}[e^{\lambda X_n} \mid \mathscr{F}_{n-1}] = 1 + \pi_n(e^\lambda - 1) \le \exp\{\pi_n(e^\lambda - 1)\}.$$

The last inequality is $1 + x \le e^x$ with $x = \pi_n(e^\lambda - 1)$. It follows that

$$\begin{aligned}\mathbb{E}[M_n(\lambda) \mid \mathscr{F}_{n-1}] &= M_{n-1}(\lambda)e^{-(e^\lambda-1)\pi_n}\mathbb{E}[e^{\lambda X_n} \mid \mathscr{F}_{n-1}] \\ &\le M_{n-1}(\lambda).\end{aligned}$$

Thus $(M_n(\lambda))_{n=a-1}^c$ is a nonnegative supermartingale and, by iteration, $\mathbb{E}M_c(\lambda) \le M_{a-1}(\lambda) = 1$. At the right endpoint, abbreviate $N = N_c = N_i^{\mathrm{pull}}(I)$ and $P = P_c = \Pi_i(I)$. With $\lambda = \log 2$,

$$M_c(\log 2) = \exp\{(\log 2)N - P\}.$$

On the event $N \ge 2P + 6L_H$,

$$(\log 2)N - P \ge (2\log 2 - 1)P + 6(\log 2)L_H \ge 6(\log 2)L_H,$$

because $2\log 2 - 1 > 0$. Markov's inequality therefore gives

$$\Pr\{N \ge 2P + 6L_H\} \le e^{-6(\log 2)L_H}\mathbb{E}M_c(\log 2) \le e^{-6(\log 2)L_H} \le e^{-4L_H}.$$

With $\lambda = -\log 2$, one has

$$M_c(-\log 2) = \exp\{-(\log 2)N + P/2\}.$$

On $N \le P/2 - 6L_H$, the exponent is at least

$$\frac{1 - \log 2}{2}P + 6(\log 2)L_H \ge 6(\log 2)L_H;$$

if $P/2 - 6L_H < 0$, the event is empty. The same Markov argument gives

$$\Pr\{N \le P/2 - 6L_H\} \le e^{-4L_H}.$$

There are exactly $KH(H+1)/2$ arm–interval pairs. Accounting for both tails and using $H+1 \le 2H$ when $H \ge 1$ gives

$$\Pr_\nu((\mathcal{G}_{\mathrm{count}}^{\mathrm{sto}})^c) \le KH(H+1)e^{-4L_H} \le 2KH^2e^{-4L_H}.$$

For $H = 0$, the interval family is empty and the same conclusion is immediate.

Substituting $e^{-L_H} = \beta/(18KT^2)$ and $H \le T$ into the displayed failure bounds gives

$$\begin{aligned}\Pr_\nu((\mathcal{G}^{\mathrm{sto}})^c) &\le 3K(H+1)e^{-L_H} + 2KHe^{-L_H} + 2He^{-L_H} \\ &\quad + 2(KH+H)e^{-2L_H} + 2KH^2e^{-4L_H} \le \beta/2,\end{aligned}$$

and

$$\Pr_L((\mathcal{G}^{\mathrm{adv}})^c) \le 2KHe^{-L_H} + 2He^{-L_H} \le \beta/2.$$

□

Lemma E.7. *Assume $\mathcal{G}_{\mathrm{act}}^{\mathrm{sto}}$ and let i be suboptimal with gap $\Delta_i > 0$. If i is eliminated at active stage m_i, then*

$$\frac{\Delta_i}{8} < d_{m_i} < \frac{\Delta_i}{2}, \tag{79}$$

and the stored EAP gap satisfies

$$\frac{7\Delta_i}{16} < \widetilde{\Delta}_i \le \frac{\Delta_i}{2}. \tag{80}$$

The optimal arm is never eliminated on $\mathcal{G}_{\mathrm{act}}^{\mathrm{sto}}$.

Proof. The quota definition gives

$$\rho(bq_m) \le d_m \left(288^{-1/2} + 288^{-1}\right) \le d_m/16.$$

Hence, after the stage-m release,

$$\mathrm{lcb}_j \le \mu_j \le \mathrm{ucb}_j, \qquad \mathrm{lcb}_j \ge \mu_j - d_m/8, \qquad \mathrm{ucb}_j \le \mu_j + d_m/8 \tag{81}$$

for every active arm j. Clipping preserves the first containment because it is monotone and fixes every $\mu_j \in [0,1]$.

Let $i^\star$ be optimal. If it is active, then

$$u_m^{\mathrm{elim}} = \min_{j \in \mathcal{A}} \mathrm{ucb}_j \ge \mu_\star, \qquad \mathrm{lcb}_{i^\star} \le \mu_\star,$$

so it cannot satisfy the elimination inequality. Induction over stages proves that $i^\star$ is never eliminated.

If $d_m \le \Delta_i/4$, then (81) gives

$$\widehat{\mu}_{i,m} - \widehat{\mu}_{i^\star,m} \ge \Delta_i - d_m/8 > 0,$$

so $i \ne j_m$. Moreover,

$$\begin{aligned} \mathrm{lcb}_i - u_m^{\mathrm{elim}} &\ge (\mu_i - d_m/8) - (\mu_\star + d_m/8) \\ &= \Delta_i - d_m/4 \ge 15\Delta_i/16 > 2d_m. \end{aligned}$$

Thus i is eliminated when the first stage with $d_m \le \Delta_i/4$ completes.

At the actual elimination stage,

$$2d_{m_i} < \mathrm{lcb}_i - u_{m_i}^{\mathrm{elim}} \le \mu_i - \mu_\star = \Delta_i,$$

so $d_{m_i} < \Delta_i/2$. Elimination cannot occur at stage 0, because $d_0 = 1$ and all gaps are at most one. Since i survived the preceding completed stage, $d_{m_i-1} > \Delta_i/4$, and therefore $d_{m_i} > \Delta_i/8$. Finally,

$$2\widetilde{\Delta}_i = \mathrm{lcb}_i - u_{m_i}^{\mathrm{elim}} \ge \Delta_i - d_{m_i}/4 > 7\Delta_i/8,$$

while the preceding upper bound gives $2\widetilde{\Delta}_i \le \Delta_i$. This proves (80). □

Lemma E.8. *Before switching to adversarial mode or reaching the horizon, the following statements hold.*

(i) *On $\mathcal{G}_{\mathrm{count}}^{\mathrm{sto}}$, let s_m be the first batch of an active stage m. For every prefix $[s_m, r]$ of that stage, including an unfinished prefix, and every arm i active throughout that prefix,*

$$N_i^{\mathrm{pull}}([s_m, r]) \le 4q_m + 30L_H. \tag{82}$$

(ii) *Every activated EAP phase e of arm i contains at most*

$$N_e^{\mathrm{audit}} \le U_e \le 5(M_i + 1) \tag{83}$$

audited batches, deterministically.

(iii) *Let $F_i(r)$ be the number of fires of arm i before a prefix ending at r, and let $C_i(r)$ be the number of phases activated by that prefix. Then*

$$C_i(r) \le a_{\max}(H) + 2F_i(r) + 1. \tag{84}$$

Consequently, before the E_0-th fire its index is at most $a_{\max}(H) + 2E_0$, and on $\mathcal{G}_{\mathrm{ctr}}^{\mathrm{sto}}$,

$$C_i(H) \le P_H = a_{\max}(H) + 2E_0 + 1. \tag{85}$$

Proof. Every active stage starts with empty queues because the unused overshoot is discarded at the preceding stage boundary. During a fixed stage all active arms have the same conditional sampling probability. For part (i), the stage contains at least two active arms, so its quota-completion rule remains enabled. Fix any prefix $[s_m, r]$ of stage m. If the stage has not completed by r, at least one active arm has fewer than q_m batches on this prefix; if r is the completion batch, the last deficient arm has exactly q_m batches. Let $P_m(r)$ be the common predictable mass of an active arm on $[s_m, r]$. The lower side of (53) applied to that deficient arm gives

$$P_m(r)/2 - 6L_H \le q_m, \qquad \text{hence} \qquad P_m(r) \le 2q_m + 12L_H.$$

The upper side of (53), now applied to any active arm i, yields

$$N_i^{\mathrm{pull}}([s_m, r]) \le 2P_m(r) + 6L_H \le 4q_m + 30L_H.$$

This proves (i).

For (ii), the algorithm forces a fire as soon as the audit counter reaches U_e. Since $m_e = p_e N_e \le M_i + 1$,

$$U_e = \lceil 4m_e \rceil \le 4(M_i + 1) + 1 \le 5(M_i + 1).$$

For (iii), write a_ℓ for the level of the ℓ-th phase, so that $N_\ell = 2^{a_\ell} N_i^{\mathrm{init}}$. Let $C_{\ell-1}$ be the number of preceding clean phases and let $F'_{\ell-1}$ be the number of preceding fires strictly above the floor. The update rule gives

$$a_\ell = C_{\ell-1} - F'_{\ell-1}.$$

Before level a can first be activated, a clean phase must have completed at every level $0, \dots, a-1$. These disjoint phases consume at least

$$N_i^{\mathrm{init}} \sum_{j=0}^{a-1} 2^j \ge 2^a - 1$$

calendar batches, so $a_\ell \le a_{\max}(H)$. If $F_{\ell-1}$ counts all preceding fires, including floor fires, then

$$\begin{aligned} \ell - 1 &= C_{\ell-1} + F_{\ell-1} \\ &= a_\ell + F'_{\ell-1} + F_{\ell-1} \\ &\le a_{\max}(H) + 2F_{\ell-1}. \end{aligned}$$

This proves both phase-count conclusions. □

Definition E.9. *Fix a potential slot $e = (i, \ell)$. Let A_e be the event that, at the beginning of a stochastic-controller batch, slot e becomes the current phase of arm i and receives its first calendar batch. On A_e, let $\sigma_e \le H$ be that batch and set*

$$\mathcal{H}_e^- := \mathscr{F}_{\sigma_e - 1}, \qquad \bar{N}_e := \min\{N_e, H - \sigma_e + 1\}.$$

Define

$$Q_e^{\mathrm{AT}} := 2 + \left\lceil \frac{U_e}{q_e} \right\rceil.$$

At phase initialization draw fresh independent noises

$$\eta_e \sim \mathrm{Lap}\left(\frac{2}{b\epsilon_S}\right), \qquad \xi_{e,1}, \dots, \xi_{e,Q_e^{\mathrm{AT}}} \overset{\mathrm{iid}}{\sim} \mathrm{Lap}\left(\frac{4}{b\epsilon_S}\right),$$

and define

$$\mathcal{H}_e^+ := \mathcal{H}_e^- \vee \sigma(\eta_e, \xi_{e,1}, \dots, \xi_{e,Q_e^{\mathrm{AT}}}), \qquad \mathcal{N}_e^{\mathrm{AT}} := |\eta_e| + \max_{j \le Q_e^{\mathrm{AT}}} |\xi_{e,j}|.$$

For $1 \le \kappa \le \bar{N}_e$, let

$$r_{e,\kappa} = \sigma_e + \kappa - 1, \qquad \mathcal{F}_{e,0}^+ = \mathcal{H}_e^+, \qquad \mathcal{F}_{e,\kappa}^+ = \mathcal{H}_e^+ \vee \mathscr{F}_{r_{e,\kappa}}.$$

Let ρ_e be the first local batch at which the phase returns $\top$ or reaches the overflow audit count U_e. Let ζ_e be the first external termination time, with value ∞ if none occurs. Set

$$\tau_e = \min\{\rho_e, \bar{N}_e, \zeta_e\}. \tag{86}$$

On A_e^c, use the convention $\sigma_e = H + 1$, $\mathcal{H}_e^- = \mathcal{H}_e^+ = \mathscr{F}_H$, and set all slot variables, including $N_e, \bar{N}_e, m_e, p_e, q_e, U_e, Q_e^{\mathrm{AT}}, \tau_e$, and $\mathcal{N}_e^{\mathrm{AT}}$, to zero.

Lemma E.10. *Fix a potential EAP phase $e = (i, \ell)$. Conditional on $\mathcal{H}_e^-$, suppose that the phase is activated and its frozen reference satisfies*

$$\widetilde{\ell}_i^{\mathrm{ref}} \le \mu_i.$$

Then

$$\Pr_\nu\{e \textit{ fires} \mid \mathcal{H}_e^-\} \le \rho_0 = \frac{1}{256}. \tag{87}$$

Proof. Condition on $\mathcal{H}_e^-$ and abbreviate

$$a = \widetilde{\ell}_i^{\mathrm{ref}}, \qquad \delta = \widetilde{\Delta}_i, \qquad m = p_e N_e,$$

$$q = \left\lceil \frac{\delta m}{64} \right\rceil, \qquad U = \lceil 4m \rceil, \qquad \theta = \frac{\delta m}{4}.$$

The invariant and quota definition imply

$$b\delta^2 m \ge C_M, \qquad b\epsilon_S \delta m \ge C_M \Lambda_i, \qquad m \ge \frac{C_M}{\delta}, \qquad \Lambda_i = 1 + \log \frac{512}{\delta}.$$

In particular,

$$q \le \frac{\delta m}{32}, \qquad U \le 5m, \qquad Q_e^{\mathrm{AT}} \le \frac{512}{\delta} \le e^{\Lambda_i - 1}.$$

For every audit that actually occurs, let r_j be its batch and put $Z_j = a - g_{r_j,i}$. Since $a \le \mu_i$, by Hoeffding's lemma,

$$\mathbb{E}_\nu[e^{\lambda Z_j} \mid \mathscr{F}_{r_j - 1}] \le e^{\lambda^2/(8b)}, \qquad \lambda > 0.$$

The associated exponential supermartingale and Ville's inequality give, with $x = \delta m/8$,

$$\begin{aligned} \Pr_\nu\left\{ \max_{n \le U} \sum_{j=1}^{n} Z_j \ge x \,\middle|\, \mathcal{H}_e^- \right\} &\le \inf_{\lambda > 0} \exp\left(-\lambda x + \frac{U\lambda^2}{8b} \right) \\ &= \exp\left(-\frac{2bx^2}{U} \right) \le \exp\left(-\frac{b\delta^2 m}{160} \right). \end{aligned}$$

The AboveThreshold noises satisfy, by Lemma E.3,

$$\Pr\left\{ \eta_e < -\frac{\delta m}{32} \right\} \le \frac{1}{2} e^{-b\epsilon_S \delta m/64},$$

and

$$\Pr\left\{ \max_{j \le Q_e^{\mathrm{AT}}} \xi_{e,j} > \frac{\delta m}{32} \right\} \le \frac{Q_e^{\mathrm{AT}}}{2} e^{-b\epsilon_S \delta m/128}.$$

Outside these noise events and the preceding raw-residual event, every submitted query is at most $5\delta m/32$, whereas the noisy threshold is at least $7\delta m/32$; hence AboveThreshold cannot emit $\top$.

For overflow, extend the audit indicators after phase termination by fresh independent Bernoulli(p_e) variables up to the scheduled length N_e, and let S be their total. While the phase remains active, its audit probability conditional on the full preceding history is exactly p_e; the appended indicators have the same property. Thus, conditional on $\mathcal{H}_e^-$, one has $S \sim \mathrm{Binom}(N_e, p_e)$ with mean $m = p_e N_e$. Moreover, $N_e^{\mathrm{audit}} \le S$ pathwise. For any $\lambda > 0$, the exponential Markov inequality gives

$$\begin{aligned} \Pr_\nu\{N_e^{\mathrm{audit}} \ge U \mid \mathcal{H}_e^-\} &\le \Pr_\nu\{S \ge U \mid \mathcal{H}_e^-\} \\ &\le e^{-\lambda U} \mathbb{E}_\nu[e^{\lambda S} \mid \mathcal{H}_e^-] \\ &= e^{-\lambda U}(1 - p_e + p_e e^{\lambda})^{N_e} \\ &\le \exp\big(-\lambda U + m(e^{\lambda} - 1)\big), \end{aligned}$$

where the last inequality uses $1 + y \le e^y$. Taking $\lambda = \log 4$ and using $U = \lceil 4m \rceil \ge 4m$ yields

$$\begin{aligned}\Pr_\nu\{N_e^{\mathrm{audit}} \ge U \mid \mathcal{H}_e^-\} &\le \exp(-U \log 4 + 3m) \\ &\le \exp(-(4 \log 4 - 3)m) \\ &\le e^{-m},\end{aligned}$$

since $4 \log 4 - 3 > 1$. Consequently,

$$\Pr_\nu\{e \text{ fires} \mid \mathcal{H}_e^-\} \le e^{-b\delta^2 m/160} + \frac{1}{2} e^{-b\epsilon_S \delta m/64} + \frac{Q_e^{\mathrm{AT}}}{2} e^{-b\epsilon_S \delta m/128} + e^{-m}.$$

Using $C_M = 2048$ and $Q_e^{\mathrm{AT}} \le e^{\Lambda_i - 1}$ makes the right-hand side at most $1/256$. □

Lemma E.11. *With $E_0 = \lceil 2L_H \rceil$ and $\rho_0 = 1/256$,*

$$\Pr_\nu\big((\mathcal{G}_{\mathrm{ctr}}^{\mathrm{sto}})^c \cap \mathcal{G}_{\mathrm{act}}^{\mathrm{sto}}\big) \le K(3e\rho_0)^{E_0} \le \frac{\beta}{2}. \tag{88}$$

Proof. For an eliminated arm i, let

$$\mathcal{B}_i^{\mathrm{ref}} = \{\widetilde{\ell}_i^{\mathrm{ref}} \le \mu_i\}.$$

This event is measurable at elimination, and $\mathcal{G}_{\mathrm{act}}^{\mathrm{sto}}$ implies $\mathcal{B}_i^{\mathrm{ref}}$ whenever arm i is eliminated.

By Lemma E.8, if the E_0-th fire occurs, it occurs among the first

$$n_0 = a_{\max}(H) + 2E_0 \le 3E_0$$

activated phase slots. Here $a_{\max}(H) \le E_0$ follows directly from $H \le T$, $K \ge 2$, $T \ge K$, $\beta = T^{-1}$, and the definitions of $a_{\max}(H), L_H, E_0$. Put the fire indicator equal to zero for an inactive slot. Lemma E.10 and repeated conditioning show that, for every fixed set of E_0 slots, the probability that all those slots fire, jointly with $\mathcal{B}_i^{\mathrm{ref}}$, is at most $\rho_0^{E_0}$. Therefore

$$\Pr_\nu\{E_i \ge E_0, \mathcal{B}_i^{\mathrm{ref}}\} \le \binom{n_0}{E_0} \rho_0^{E_0} \le \left(\frac{e n_0 \rho_0}{E_0}\right)^{E_0} \le (3e\rho_0)^{E_0}.$$

A union bound over arms proves the first inequality. Since $3e/256 < e^{-1}$ and $E_0 \ge 2L_H$,

$$K(3e\rho_0)^{E_0} \le K e^{-2L_H} \le \frac{\beta}{2}.$$

□

Lemma E.12. *The quantity $\mathfrak{B}_H$ in (10) satisfies*

$$\mathfrak{B}_H = O(L_H^2 \sqrt{KH}). \tag{89}$$

Proof. From (7),

$$d_m q_m \le d_m + 288 L_H \left(\frac{1}{b d_m} + \frac{1}{b\epsilon_S}\right).$$

For $m \le m(\alpha)$,

$$\sum_m d_m \le 2, \qquad \sum_m d_m^{-1} = O(\alpha^{-1}), \qquad m(\alpha) + 1 = O(L_H),$$

because $\log(1/\alpha) = O(\log(\max\{H, 2\})) = O(L_H)$ on the grid and $b\epsilon_S \ge 1$. Therefore

$$\mathfrak{A}_H(\alpha) = O\left(\frac{L_H}{b\alpha} + L_H^2\right).$$

Since $P_H = O(L_H)$, $\Lambda_\alpha = O(L_H)$ on $\mathcal{D}_H$, and $b\epsilon_S \ge 1$,

$$\mathfrak{E}_H(\alpha) = O\left(\frac{L_H}{b\alpha} + \frac{L_H^2}{b\epsilon_S} + L_H\right) = O\left(\frac{L_H}{b\alpha} + L_H^2\right).$$

Choose a grid point within a factor of two of $\min\{1, \sqrt{K/H}\}$. Then

$$\mathfrak{B}_H \le O\left(L_H^2 \sqrt{KH} + K L_H^2\right).$$

If $H \ge K$, the second term is at most $L_H^2 \sqrt{KH}$; if $H < K$, the defining truncation gives $\mathfrak{B}_H \le H \le \sqrt{KH}$. This proves (89). □

Lemma E.13. *On*

$$\mathcal{G}_{\mathrm{act}}^{\mathrm{sto}} \cap \mathcal{G}_{\mathrm{count}}^{\mathrm{sto}} \cap \mathcal{G}_{\mathrm{ctr}}^{\mathrm{sto}},$$

for every batch r before switching to adversarial mode,

$$\sum_{s=1}^{r} \Delta_{J_s} \le \mathfrak{B}_H. \tag{90}$$

Proof. Fix $\alpha \in \mathcal{D}_H$. Batches allocated to arms with $\Delta_i \le \alpha$ contribute at most $H\alpha$.

Consider an arm with $\Delta_i > \alpha$. An optimal arm remains active, so every active stage containing i has at least two active arms. By Lemma E.7, every active stage in which it is pulled has $\Delta_i < 8d_m$. Lemma E.8 therefore gives

$$\begin{aligned}\Delta_i N_i^{\mathrm{act}}(r) &\le \sum_{m \le m(\alpha)} 8d_m(4q_m + 30L_H) \\ &\le \mathfrak{A}_H(\alpha).\end{aligned}$$

After elimination, (80) gives $\widetilde{\Delta}_i > 7\Delta_i/16$. On $\mathcal{G}_{\mathrm{ctr}}^{\mathrm{sto}}$, Lemma E.8 and the hard audit cap give

$$N_i^{\mathrm{audit}}(r) \le 5P_H(M_i + 1).$$

For $\Delta_i > \alpha$,

$$\begin{aligned}\Delta_i(M_i + 1) &\le 2 + C_M\left(\frac{\Delta_i}{b\widetilde{\Delta}_i^2} + \frac{\Delta_i\Lambda_i}{b\epsilon_S\widetilde{\Delta}_i} + \frac{\Delta_i}{\widetilde{\Delta}_i}\right) \\ &\le 2 + C_M\left(\frac{256}{49b\alpha} + \frac{16\Lambda_\alpha}{7b\epsilon_S} + \frac{16}{7}\right).\end{aligned}$$

Consequently,

$$\Delta_i N_i^{\mathrm{audit}}(r) \le \mathfrak{E}_H(\alpha).$$

Summing the active and audit contributions over at most K large-gap arms yields

$$\sum_{s=1}^{r} \Delta_{J_s} \le H\alpha + K\{\mathfrak{A}_H(\alpha) + \mathfrak{E}_H(\alpha)\}.$$

Taking the minimum over $\alpha \in \mathcal{D}_H$ and using the trivial bound by H proves (90). □

Lemma E.14. *On*

$$\mathcal{G}_{\mathrm{act}}^{\mathrm{sto}} \cap \mathcal{G}_{\mathrm{bsc}} \cap \mathcal{G}_{\mathrm{raw}}^{\mathrm{sto}} \cap \mathcal{G}_{\mathrm{count}}^{\mathrm{sto}} \cap \mathcal{G}_{\mathrm{ctr}}^{\mathrm{sto}},$$

the running benchmark satisfies, for every pre-switch batch r,

$$\mu_\star \le u_r \le \mu_\star + \frac{49}{16}w_A(r), \tag{91}$$

and all three tests in $\mathrm{PrivBSC}(r)$ *pass.*

Proof. For every active arm,

$$|\widehat{\nu}_{r,i}^A - \mu_i| \le w_A(r) + \sqrt{\frac{L_H}{br}} \le \frac{17}{16}w_A(r).$$

The pull-based confidence interval contains μ_i. Moreover,

$$\widehat{\nu}_{r,i}^A + 2w_A(r) \ge \mu_i + \frac{15}{16}w_A(r),$$

so every batch-wise UCB candidate, and hence every running batch-wise UCB, is at least μ_i. Since every historical pull-based or batch-wise candidate used in the running minimum is at least the mean of its arm, $u_r \ge \mu_\star$. The optimal arm remains active by Lemma E.7, and its current calendar candidate gives

$$u_r \le \widehat{\nu}_{r,i^\star}^A + 2w_A(r) \le \mu_\star + \frac{49}{16}w_A(r),$$

which proves (91).

The coherence test passes because

$$\widehat{\nu}^A_{r,i} \in [\mu_i - 17w_A(r)/16, \mu_i + 17w_A(r)/16] \subseteq [\mathrm{lcb}_i - 2w_A(r), \mathrm{ucb}_i + 2w_A(r)].$$

For the active-comparator test,

$$ru_r - \widehat{L}^A_{r,i} = r(u_r - \widehat{\nu}^A_{r,i}) \le \frac{66}{16} r w_A(r) < \Lambda_{\mathrm{cmp}}(r).$$

Finally,

$$\begin{aligned}\widehat{G}_r - ru_r &= (\widehat{G}_r - G_r) + \sum_{s=1}^{r}(g_{s,J_s} - \mu_{J_s}) + \sum_{s=1}^{r}\Delta_{J_s} + r(\mu_\star - u_r) \\ &\le \Lambda_A(r) + \sqrt{\frac{rL_H}{b}} + \mathfrak{B}_H = \Gamma_r,\end{aligned}$$

where Lemma E.13 controls the third term and $u_r \ge \mu_\star$ makes the last term nonpositive. Thus the cumulative test also passes. □

Theorem E.15 (Low probability of false switch in stochastic environment)**.** *For the stochastic safety event in (56),*

$$\Pr_\nu(\mathcal{G}^{\mathrm{sto}}_{\mathrm{safe}}) \ge 1 - \beta, \tag{92}$$

and on $\mathcal{G}^{\mathrm{sto}}_{\mathrm{safe}}$ the algorithm never enters adversarial mode:

$$\kappa_{\mathrm{sw}} = H + 1. \tag{93}$$

Proof. Lemma E.6 and Lemma E.11 give

$$\begin{aligned}\Pr_\nu((\mathcal{G}^{\mathrm{sto}}_{\mathrm{safe}})^c) &\le \Pr_\nu((\mathcal{G}^{\mathrm{sto}})^c) + \Pr_\nu\big((\mathcal{G}^{\mathrm{sto}}_{\mathrm{ctr}})^c \cap \mathcal{G}^{\mathrm{sto}}_{\mathrm{act}}\big) \\ &\le \beta.\end{aligned}$$

On $\mathcal{G}^{\mathrm{sto}}_{\mathrm{safe}}$, suppose that a first switching batch exists. It cannot be an EAP switch, by $\mathcal{G}^{\mathrm{sto}}_{\mathrm{ctr}}$. Immediately before a possible BSC switch, all EAP counters are below E_0, so Lemmas E.13 and E.14 imply that the BSC call returns pass. This is a contradiction. Hence $\kappa_{\mathrm{sw}} = H + 1$. □

Lemma E.16. *On $\mathcal{G}_{\mathrm{bsc}}$, if $\mathrm{PrivBSC}(r)$ returns pass, then for every $i \in \mathcal{A}_r$,*

$$\bar{g}_{r,i} \in [\mathrm{lcb}_i - 3w_A(r), \mathrm{ucb}_i + 3w_A(r)], \tag{94}$$

$$ru_r - \sum_{s \le r} g_{s,i} \le 6r w_A(r), \tag{95}$$

$$G_r - ru_r \le \mathfrak{B}_H + 2\Lambda_{\mathrm{noise}}(r). \tag{96}$$

Proof. The coherence claim follows by adding the realized importance-weighted estimation error $w_A(r)$ to the private interval test. The active-comparator test and (77) give

$$ru_r - \sum_{s \le r} g_{s,i} \le 5r w_A(r) + r w_A(r).$$

The cumulative test gives $\widehat{G}_r - ru_r \le \mathfrak{B}_H + \Lambda_{\mathrm{noise}}(r)$, and $|\widehat{G}_r - G_r| \le \Lambda_A(r) \le \Lambda_{\mathrm{noise}}(r)$, proving (96). □

Lemma E.17. *When an arm i is eliminated by Algorithm 4,*

$$\widetilde{\ell}^{\mathrm{ref}}_i - \widetilde{v}_i \ge \frac{9}{8}\widetilde{\Delta}_i. \tag{97}$$

Proof. The algorithm stores

$$2\widetilde{\Delta}_i = \mathrm{lcb}_i - u_m^{\mathrm{elim}}, \qquad \widetilde{\ell}_i^{\mathrm{ref}} = \mathrm{lcb}_i\,.$$

After removing the eliminated arms it updates the running benchmark by a minimum, so $\widetilde{v}_i = u_r \le u_m^{\mathrm{elim}}$. Therefore

$$\widetilde{\ell}_i^{\mathrm{ref}} - \widetilde{v}_i \ge \mathrm{lcb}_i - u_m^{\mathrm{elim}} = 2\widetilde{\Delta}_i \ge \frac{9}{8}\widetilde{\Delta}_i.$$

□

Lemma E.18. *Fix an activated EAP phase $e = (i, \ell)$ and a fixed oblivious batch-average loss table, and abbreviate*

$$p = p_e, \quad N = N_e, \quad \bar{N} = \bar{N}_e, \quad m = pN, \quad \Delta = \widetilde{\Delta}_i, \quad a = \widetilde{\ell}_i^{\mathrm{ref}}, \quad v = \widetilde{v}_i, \quad \tau = \tau_e.$$

Suppose

$$a - v \ge \frac{9}{8}\Delta. \tag{98}$$

Then we have

$$\mathbb{E}[\mathcal{N}_e^{\mathrm{AT}} \mid \mathcal{H}_e^-] + q_e \le \frac{1}{8}\Delta m. \tag{99}$$

Define

$$R_e = \sum_{\kappa=1}^{\tau} \big(v - g_{r_{e,\kappa},i}\big). \tag{100}$$

It satisfies

$$\mathbb{E}[R_e \mid \mathcal{H}_e^-] \le \frac{3}{8}\Delta N - \frac{9}{8}\Delta\,\mathbb{E}[\tau \mid \mathcal{H}_e^-]. \tag{101}$$

Proof. For local time κ, define

$$A_\kappa = \mathbf{1}\{J_{r_{e,\kappa}} = i\}, \qquad h_\kappa = a - g_{r_{e,\kappa},i}, \qquad Q_n = \sum_{\kappa=1}^{n} A_\kappa h_\kappa, \qquad 0 \le n \le \bar{N}.$$

Because the loss table is oblivious, h_κ is predictable once the phase start is fixed. Also $a, g_{r,i} \in [0,1]$, so $|h_\kappa| \le 1$.

Since $m \ge M_i \ge C_M/\Delta$, one has $m \ge 1$ and

$$Q_e^{\mathrm{AT}} \le 3 + \frac{4m+1}{\Delta m/64} \le \frac{512}{\Delta}.$$

For $Z \sim \mathrm{Lap}(s)$, $\mathbb{E}|Z| = s$ and $\mathbb{E}\max_{j\le Q}|Z_j| \le s(1+\log(2Q))$. Consequently, with $\Lambda_e = 1 + \log(512/\Delta)$,

$$\mathbb{E}[\mathcal{N}_e^{\mathrm{AT}} \mid \mathcal{H}_e^-] \le \frac{8\Lambda_e}{b\epsilon_S}. \tag{102}$$

Moreover,

$$q_e \le \frac{\Delta m}{64} + 1, \qquad \Delta m \ge C_M, \qquad \Delta m \ge \frac{C_M \Lambda_e}{b\epsilon_S}.$$

For $C_M = 2048$ these inequalities prove (99).

We next control the stopped prefix pathwise. After a $\perp$ checkpoint, the raw accumulator is at most $\theta_e + \mathcal{N}_e^{\mathrm{AT}}$. Before the next checkpoint fewer than q_e new audits occur, and a first checkpoint contains at most q_e audits. Therefore

$$Q_\tau \le \theta_e + \mathcal{N}_e^{\mathrm{AT}} + q_e. \tag{103}$$

Now define the stopped-difference process, for $n = 0, \dots, \bar{N}$,

$$M_n = \sum_{\kappa=1}^{n} \mathbf{1}\{\tau \ge \kappa\}(A_\kappa - p)h_\kappa, \qquad M_0 = 0. \tag{104}$$

By Definition E.9, on $\{\tau \geq \kappa\}$ the phase is active and

$$\mathbb{E}[A_\kappa \mid \mathcal{F}^+_{e,\kappa-1}] = p.$$

Consequently,

$$\begin{aligned}\mathbb{E}[M_\kappa - M_{\kappa-1} \mid \mathcal{F}^+_{e,\kappa-1}] &= \mathbf{1}\{\tau \geq \kappa\} h_\kappa \mathbb{E}[A_\kappa - p \mid \mathcal{F}^+_{e,\kappa-1}] \\ &= 0.\end{aligned}$$

Thus $(M_n)_{n=0}^{\bar{N}}$ is a bounded martingale. Conditional on $\mathcal{H}^+_e$, the available horizon $\bar{N}$ is deterministic. At that terminal index,

$$M_{\bar{N}} = Q_\tau - p \sum_{\kappa=1}^{\tau} h_\kappa.$$

Taking conditional expectation given $\mathcal{H}^+_e$ gives

$$\mathbb{E}[Q_\tau \mid \mathcal{H}^+_e] = p\, \mathbb{E}\left[\sum_{\kappa=1}^{\tau} (a - g_{r_{e,\kappa},i}) \,\middle|\, \mathcal{H}^+_e \right]. \tag{105}$$

This is the finite-horizon stopped-martingale identity. The martingale is stopped at the actual phase duration, while the threshold and phase-size condition continue to use the scheduled length N; a shorter available horizon can only reduce the number of possible increments.

Average (105) over the phase noises and combine it with (103) and (99):

$$\mathbb{E}\left[\sum_{\kappa=1}^{\tau} (a - g_{r_{e,\kappa},i}) \,\middle|\, \mathcal{H}^-_e \right] \leq \frac{\theta_e + \mathbb{E}[\mathcal{N}^{\mathrm{AT}}_e \mid \mathcal{H}^-_e] + q_e}{p} \leq \frac{3}{8}\Delta N.$$

Finally,

$$R_e = \sum_{\kappa=1}^{\tau} (a - g_{r_{e,\kappa},i}) + (v - a)\tau,$$

and (98) gives $v - a \leq -9\Delta/8$. This proves (101). □

Lemma E.19. *Fix an arm i. For each deterministic index $\ell \in [H]$, let $e_\ell = (i, \ell)$ be its potential EAP slot. Let*

$$C_i = \sum_{\ell=1}^{H} \mathbf{1}_{A_{e_\ell}}$$

be the number of activated phases. For an activated phase, let R_ℓ be defined in (100), N_ℓ its scheduled length, and τ_ℓ its actual duration; for an inactive slot set $R_\ell = N_\ell = \tau_\ell = 0$. Suppose (98) holds. Let N_i^{init} be the floor phase length and $\Delta = \widetilde{\Delta}_i$, with both set to zero if arm i is never eliminated. Then

$$\mathbb{E}\left[\sum_{\ell=1}^{C_i} R_\ell \right] \leq \frac{9}{8} E_0 \mathbb{E}\big[\Delta N_i^{\mathrm{init}} \mathbf{1}\{C_i \geq 1\} \big]. \tag{106}$$

In particular,

$$\mathbb{E}\left[\sum_{\ell=1}^{C_i} R_\ell \right] \leq 3 E_0 \mathbb{E}\big[\Delta N_i^{\mathrm{init}} \mathbf{1}\{C_i \geq 1\} \big]. \tag{107}$$

Proof. The contribution vanishes on paths with $C_i = 0$. For every deterministic $\ell \leq H$, write

$$\mathcal{H}^-_\ell = \mathcal{H}^-_{e_\ell}, \qquad \Phi_\ell = \frac{3}{4}\Delta N_\ell,$$

where the inactive-slot values are those of Definition E.9; in particular $R_\ell = N_\ell = \tau_\ell = \Phi_\ell = 0$ when A_{e_ℓ} fails. Define the clean, fire, floor-fire, and terminal events to be empty for an inactive slot. Since $A_{e_\ell} = \{\ell \leq C_i\}$ and $A_{e_\ell} \in \mathcal{H}^-_{e_\ell}$ by Definition E.9, the indicator $\mathbf{1}\{\ell \leq C_i\}$ is $\mathcal{H}^-_\ell$-measurable. For each activated phase, Lemma E.18 gives

$$\mathbb{E}[R_\ell \mid \mathcal{H}^-_\ell] \leq \frac{3}{8}\Delta N_\ell - \frac{9}{8}\Delta\, \mathbb{E}[\tau_\ell \mid \mathcal{H}^-_\ell]. \tag{108}$$

Classify a phase by whether a subsequent EAP phase for arm i receives at least one batch. Let C_ℓ be the event that the phase completes cleanly and the subsequent phase starts, let F_ℓ be the event that the phase fires and the subsequent phase starts, and let T_ℓ be the complementary terminal event. Thus T_ℓ includes horizon termination, a switch triggered by another certificate, a fire that raises the error counter to E_0, and every tie in which no next phase starts. Fire has precedence over clean at a scheduled endpoint. Hence C_ℓ, F_ℓ, T_ℓ are disjoint and form a partition. Let

$$F_\ell^0 = F_\ell \cap \{N_\ell = N_i^{\mathrm{init}}\}.$$

On C_ℓ, $N_{\ell+1} = 2N_\ell$; on a nonterminal fire above the floor, $N_{\ell+1} = N_\ell/2$; on a nonterminal fire at the floor, $N_{\ell+1} = N_\ell$; and on T_ℓ set $\Phi_{\ell+1} = 0$. Therefore

$$\begin{aligned}\mathbb{E}[\Phi_{\ell+1} - \Phi_\ell \mid \mathcal{H}_\ell^-] = \frac{3}{4}\Delta N_\ell \Pr(C_\ell \mid \mathcal{H}_\ell^-) - \frac{3}{8}\Delta N_\ell \Pr(F_\ell \setminus F_\ell^0 \mid \mathcal{H}_\ell^-) \\ - \frac{3}{4}\Delta N_\ell \Pr(T_\ell \mid \mathcal{H}_\ell^-).\end{aligned} \tag{109}$$

Moreover,

$$\mathbb{E}[\tau_\ell \mid \mathcal{H}_\ell^-] \geq N_\ell \Pr(C_\ell \mid \mathcal{H}_\ell^-).$$

Using $\Pr(C_\ell) + \Pr(F_\ell \setminus F_\ell^0) + \Pr(F_\ell^0) + \Pr(T_\ell) = 1$ in (108)–(109) yields

$$\begin{aligned}&\mathbb{E}[R_\ell + \Phi_{\ell+1} - \Phi_\ell \mid \mathcal{H}_\ell^-] \\ &\leq \frac{3}{8}\Delta N_i^{\mathrm{init}} \Pr(F_\ell^0 \mid \mathcal{H}_\ell^-) - \frac{3}{8}\Delta N_\ell \Pr(T_\ell \mid \mathcal{H}_\ell^-) \\ &\leq \frac{3}{8}\Delta N_i^{\mathrm{init}} \Pr(F_\ell^0 \mid \mathcal{H}_\ell^-).\end{aligned} \tag{110}$$

Using these deterministic potential-slot variables, take expectations in (110) and sum over the deterministic range $\ell = 1, \ldots, H$. The tower property gives

$$\begin{aligned}&\mathbb{E}\left[\sum_{\ell=1}^{C_i}(R_\ell + \Phi_{\ell+1} - \Phi_\ell)\right] \\ &\qquad \leq \frac{3}{8}\mathbb{E}\left[\Delta N_i^{\mathrm{init}} \sum_{\ell=1}^{C_i} \mathbf{1}_{F_\ell^0}\right].\end{aligned}$$

The potentials telescope:

$$\sum_{\ell=1}^{C_i}(\Phi_{\ell+1} - \Phi_\ell) = \Phi_{C_i+1} - \Phi_1 = -\Phi_1, \qquad \Phi_{C_i+1} = 0.$$

Every floor fire increments the EAP error counter, and a switch occurs after at most E_0 fires, so

$$\sum_{\ell=1}^{C_i} \mathbf{1}_{F_\ell^0} \leq E_0.$$

Since $\Phi_1 = (3/4)\Delta N_i^{\mathrm{init}} \mathbf{1}\{C_i \geq 1\}$ and $E_0 \geq 1$,

$$\begin{aligned}\mathbb{E}\left[\sum_{\ell=1}^{C_i} R_\ell\right] &\leq \mathbb{E}[\Phi_1] + \frac{3}{8}E_0\mathbb{E}\big[\Delta N_i^{\mathrm{init}} \mathbf{1}\{C_i \geq 1\}\big] \\ &\leq \frac{9}{8}E_0\mathbb{E}\big[\Delta N_i^{\mathrm{init}} \mathbf{1}\{C_i \geq 1\}\big].\end{aligned}$$

This proves (106), and the looser (107) follows immediately. □

Lemma E.20. *If $\mathcal{G}_{\mathrm{act}}^{\mathrm{sto}}$ holds and a suboptimal arm i is eliminated in a stochastic instance, then*

$$b\Delta_i(M_i + 1) \leq O\left(\frac{1}{\Delta_i} + \frac{1 + \log(1/\Delta_i)}{\epsilon_0}\right). \tag{111}$$

For arbitrary oblivious losses with $H \geq K$,

$$\widetilde{\Delta}_i N_i^{\mathrm{init}} \leq O\left(\sqrt{KH}\right). \tag{112}$$

Proof. In the stochastic case, Lemma E.7 gives $7\Delta_i/16 < \widetilde{\Delta}_i \le \Delta_i/2$. Moreover,

$$M_i \le 1 + C_M \left(\frac{1}{b\widetilde{\Delta}_i^2} + \frac{\Lambda_i}{b\epsilon_S \widetilde{\Delta}_i} + \frac{1}{\widetilde{\Delta}_i} \right). \tag{113}$$

The calibration also gives

$$\Lambda_i = 1 + \log \frac{512}{\widetilde{\Delta}_i} \le 1 + \log \frac{8192}{7\Delta_i} = O(1 + \log(1/\Delta_i)).$$

Using $\epsilon_S = \epsilon_0/2$ and $b \le 3/\epsilon_0$,

$$\begin{aligned} b\Delta_i(M_i + 1) &\le O(b\Delta_i) + C_M \left(\frac{\Delta_i}{\widetilde{\Delta}_i^2} + \frac{\Lambda_i \Delta_i}{\epsilon_S \widetilde{\Delta}_i} + \frac{b\Delta_i}{\widetilde{\Delta}_i} \right) \\ &\le O\left(\frac{1}{\Delta_i} + \frac{1 + \log(1/\Delta_i)}{\epsilon_0} \right). \end{aligned}$$

This proves (111).

Write $\delta = \widetilde{\Delta}_i$ and

$$\begin{aligned} M_i^{(0)} &= C_M \left(\frac{1}{b\delta^2} + \frac{\Lambda_i}{b\epsilon_S \delta} \right), \\ M_i^{(1)} &= \frac{C_M}{\delta} + 1, \qquad M_i \le M_i^{(0)} + M_i^{(1)}. \end{aligned}$$

At an actual elimination, the eliminating stage completed, so $q_{m_i} \le H$. (7) gives

$$d_{m_i}^2 \ge \frac{288 L_H}{bH} \ge \frac{288 L_H}{T}.$$

Since $\delta > d_{m_i}$, it follows that $\Lambda_i = O(\log T) = O(L_H)$. Comparing the two terms in (7) with those in $M_i^{(0)}$ therefore gives an absolute constant $c_0 > 0$ such that

$$n_i^{\mathrm{elim}} \ge q_{m_i} \ge c_0 M_i^{(0)}. \tag{114}$$

Furthermore, $b\epsilon_S \ge 1$ and $\sup_{0<\delta\le 1} \delta[1 + \log(512/\delta)] < 8$, so

$$\delta^2 M_i^{(0)} = C_M \left(\frac{1}{b} + \frac{\delta \Lambda_i}{b\epsilon_S} \right) = O(1), \qquad \delta \sqrt{M_i^{(0)}} = O(1). \tag{115}$$

The initialization satisfies

$$p_i^{\mathrm{init}} \ge \frac{1}{2} \sqrt{\frac{n_i^{\mathrm{elim}}}{KH}}, \qquad p_i^{\mathrm{init}} \ge \frac{1}{4K}. \tag{116}$$

Using $N_i^{\mathrm{init}} \le 1 + M_i/p_i^{\mathrm{init}}$ and (114),

$$\begin{aligned} \delta N_i^{\mathrm{init}} &\le 1 + 2\delta M_i^{(0)} \sqrt{\frac{KH}{n_i^{\mathrm{elim}}}} + 4K\delta M_i^{(1)} \\ &\le 1 + \frac{2}{\sqrt{c_0}} \delta \sqrt{M_i^{(0)}} \sqrt{KH} + O(K) \\ &= O(\sqrt{KH}), \end{aligned}$$

where (115) and $H \ge K$ were used. If $H < K$, the total batch-level regret is at most $H \le \sqrt{KH}$, so this scale lemma is unnecessary. This proves (112). □

F Proofs of main results

F.1 Stochastic regret

Theorem F.1 (Stochastic regret)**.** *For every stochastic instance with a unique optimal arm,*

$$R_T^{\mathrm{sto}}(\mathrm{PrivDaS}, \nu) \le O\left(L_H \sum_{i:\Delta_i > 0} \left[\frac{1}{\Delta_i} + \frac{1 + \log(1/\Delta_i)}{\epsilon_0} \right] \right). \tag{117}$$

In particular,

$$R_T^{\mathrm{sto}}(\mathrm{PrivDaS}, \nu) \le O\left(L_H \sum_{i:\Delta_i>0} \frac{1+\log(1/\Delta_i)}{\min\{\Delta_i, \epsilon_0\}} \right). \tag{118}$$

Proof. Let κ_{sw} be the switch-declaration batch in (25). For every suboptimal arm i, decompose its round-wise pull count as

$$N_i^{\mathrm{pull}}(T) = N_i^{\mathrm{act}} + N_i^{\mathrm{audit}} + N_i^{\mathrm{adv}} + N_i^{\mathrm{rem}}, \tag{119}$$

where N_i^{act} and N_i^{audit} count original-round pulls contained in complete pre-switch batches in which i is respectively active or eliminated, N_i^{adv} counts pulls in complete adversarial mode batches, and N_i^{rem} counts pulls in the final incomplete batch.

On $\mathcal{G}_{\mathrm{safe}}^{\mathrm{sto}}$, Theorem E.15 gives $\kappa_{\mathrm{sw}} = H+1$, and therefore

$$N_i^{\mathrm{adv}} = 0. \tag{120}$$

We next bound the first two terms of (119) on this event.

For the active part, an arm with no active pull has zero contribution and needs no bound. Otherwise, define $\bar{m}_i$ to be the largest active-stage index in which arm i is pulled before the horizon. Lemma E.7 shows that a completed stage with $d_m \le \Delta_i/4$ would eliminate i, hence

$$d_{\bar{m}_i-1} > \Delta_i/4, \qquad d_{\bar{m}_i} > \Delta_i/8.$$

Therefore every stage in which i is active and pulled satisfies $d_m > \Delta_i/8$, and

$$\bar{m}_i + 1 = O\left(\log \frac{1}{\Delta_i}\right).$$

On the safety event, the optimal arm is always active, so every stage in which a suboptimal arm is pulled has at least two active arms. The prefix form of Lemma E.8 therefore applies also to its final unfinished stage, and gives at most $4q_m + 30L_H$ batches allocated to it in every such stage. Hence

$$\begin{aligned} N_i^{\mathrm{act}} &\le b \sum_{m=0}^{\bar{m}_i} (4q_m + 30L_H) \\ &\le O\big(L_H \left[\Delta_i^{-2} + \epsilon_S^{-1}\Delta_i^{-1}\right]\big). \end{aligned} \tag{121}$$

Consequently,

$$\Delta_i N_i^{\mathrm{act}} \le O\big(L_H \left[\Delta_i^{-1} + \epsilon_0^{-1}\right]\big). \tag{122}$$

For the audit part, EAP fires may occur on the stochastic safety event, but every counter remains below E_0. Lemma E.8 gives at most $P_H = O(L_H)$ activated phases for arm i. Every phase (i,ℓ) satisfies

$$U_{i,\ell} \le 5(M_i+1).$$

Thus

$$\begin{aligned} \Delta_i N_i^{\mathrm{audit}} &\le 5b\Delta_i P_H (M_i+1) \\ &\le O\left(L_H \left[\frac{1}{\Delta_i} + \frac{1+\log(1/\Delta_i)}{\epsilon_0}\right]\right), \end{aligned} \tag{123}$$

where Lemma E.20 controls $b\Delta_i(M_i+1)$.

The final incomplete original-round batch contributes at most b total pseudo-regret. This term is absorbed by the sum in (117), since each suboptimal-arm summand is at least L_H/ϵ_0 and $b \le 3/\epsilon_0$. Indeed, uniqueness of the optimal arm and $K \ge 2$ ensure at least one suboptimal arm, and for every such arm, since $0 < \Delta_i \le 1$,

$$\frac{1}{\min\{\Delta_i, \epsilon_0\}} \ge 1, \qquad \frac{1}{\min\{\Delta_i, \epsilon_0\}} \ge \frac{1}{\epsilon_0}.$$

Therefore

$$b = \left\lceil \frac{2}{\epsilon_0} \right\rceil \le 1 + \frac{2}{\epsilon_0} \le \frac{3}{\min\{\Delta_i, \epsilon_0\}}.$$

Combining (120), (122), and (123),

$$\sum_{i:\Delta_i>0} \Delta_i N_i^{\mathrm{pull}}(T) \le O\left(L_H \sum_{i:\Delta_i>0} \left[\frac{1}{\Delta_i} + \frac{1+\log(1/\Delta_i)}{\epsilon_0} \right] \right) \quad \text{on } \mathcal{G}_{\mathrm{safe}}^{\mathrm{sto}}. \tag{124}$$

On the complement of the safety event, total pseudo-regret is at most T. Theorem E.15 gives $\Pr_\nu((\mathcal{G}_{\mathrm{safe}}^{\mathrm{sto}})^c) \le \beta$. Thus (117) follows after taking expectations in (124). Finally,

$$\frac{1}{x} + \frac{1+\log(1/x)}{y} \le \frac{2[1+\log(2/x)]}{\min\{x,y\}}, \qquad 0 < x, y \le 1,$$

with $(x, y) = (\Delta_i, \epsilon_0)$ proves (118). □

F.2 Adversarial regret

Lemma F.2. *Let R_H^{batch} be the expected batch-level regret on the induced losses $g_{r,i}$. Then*

$$R_T(\mathrm{PrivDaS}; L) \le b R_H^{\mathrm{batch}} + b.$$

Proof. For any fixed comparator i and complete batch B_r,

$$\sum_{t \in B_r} (\ell_{t,J_r} - \ell_{t,i}) = b(g_{r,J_r} - g_{r,i}).$$

Summing over complete batches gives equality on the first bH rounds. The remaining $T - bH < b$ rounds contribute at most b. □

Lemma F.3 (Post-switch regret guarantee)**.** *Condition on any history $\mathscr{F}_{\tau_{\mathrm{fb}}-1}$ immediately before* PrivateConversion *starts. If the remaining complete suffix contains h batches, then for every fixed comparator arm $i \in [K]$,*

$$\mathbb{E}_L \left[\sum_{s=1}^{h} \big(g_{\tau_{\mathrm{fb}}+s-1,J_s} - g_{\tau_{\mathrm{fb}}+s-1,i} \big) \,\middle|\, \mathscr{F}_{\tau_{\mathrm{fb}}-1} \right] \le 208\sqrt{Kh} + 2. \tag{125}$$

The bound is understood as zero when $h = 0$.

Proof. Fix the stated history. Because the adversary is oblivious, the remaining loss vectors are then a deterministic suffix, and PrivateConversion uses fresh randomness. If $h = 0$ there is nothing to prove. Suppose $h \ge 1$ and set

$$n = bh.$$

PrivateConversion in the adversarial mode of PrivDaS is exactly Algorithm 1 of Asi et al. (2025) with privacy parameter $\eta = \epsilon_A$, batch length

$$\tau = b = \left\lceil \frac{1}{\epsilon_A} \right\rceil,$$

and base learner $\mathrm{HTINF}(2, \sqrt{6})$. Moreover, $\epsilon_A \le 1$ and

$$n = bh \ge b \ge \frac{1}{\epsilon_A},$$

so $\epsilon_A \in [1/n, 1]$ and Corollary C.11 applies conditionally. Its regret against the best fixed suffix arm is at most

$$208\sqrt{\frac{Kbh}{\epsilon_A}} + \frac{2}{\epsilon_A}. \tag{126}$$

For any fixed arm i, regret against arm i is no larger than regret against the best fixed suffix arm. Since every complete adversarial mode batch repeats one action for b rounds,

$$\sum_{t \in B_{\tau_{\mathrm{fb}}+s-1}} (\ell_{t,J_s} - \ell_{t,i}) = b\big(g_{\tau_{\mathrm{fb}}+s-1,J_s} - g_{\tau_{\mathrm{fb}}+s-1,i}\big).$$

Dividing (126) by b and using $b\epsilon_A \geq 1$ gives

$$\mathbb{E}_L\left[\sum_{s=1}^{h}(g_{\tau_{\rm fb}+s-1,J_s} - g_{\tau_{\rm fb}+s-1,i})\middle|\mathscr{F}_{\tau_{\rm fb}-1}\right] \leq 208\sqrt{\frac{Kh}{b\epsilon_A}} + \frac{2}{b\epsilon_A}$$
$$\leq 208\sqrt{Kh} + 2.$$

This proves (125). □

Definition F.4 (Pre-switch regret decomposition)**.** *Let h be a pre-switch batch and let i be a comparator. If i is never eliminated by time h, set $\tau_i = h+1$. If it is eliminated, let τ_i be the elimination batch and let $\widetilde{v}_i$ be its frozen benchmark. Define*

$$R_{\rm learn}(h) = G_h - hu_h, \tag{127}$$

$$R_{\rm act}(h,i) = \begin{cases} hu_h - \sum_{r\leq h} g_{r,i}, & \tau_i > h, \\ \tau_i u_{\tau_i} - \sum_{r\leq\tau_i} g_{r,i}, & \tau_i \leq h, \end{cases} \tag{128}$$

$$R_{\rm eap}(h,i) = \mathbf{1}\{\tau_i \leq h\}\sum_{r=\tau_i+1}^{h}(\widetilde{v}_i - g_{r,i}). \tag{129}$$

Lemma F.5. *For every fixed oblivious batch-average loss table L and every comparator arm i,*

$$\mathbb{E}_L\left[\mathbf{1}_{\mathcal{G}^{\rm adv}}\sum_{r<\tau_{\rm fb}}(g_{r,J_r} - g_{r,i})\right] \leq O\Big(L_H^2\sqrt{KH}\Big) + 1. \tag{130}$$

Proof. Let

$$h_{\rm cert} = \min\{\kappa_{\rm sw} - 1, H\}, \qquad \chi_{\rm trig} = \mathbf{1}\{\kappa_{\rm sw} \leq H\}.$$

All BSC calls through $h_{\rm cert}$ have returned pass. If a switch is declared, batch $\kappa_{\rm sw}$ is the switch-triggering batch and no passed certificate is asserted there. Hence, pathwise,

$$\sum_{r<\tau_{\rm fb}}(g_{r,J_r} - g_{r,i}) = \sum_{r=1}^{h_{\rm cert}}(g_{r,J_r} - g_{r,i}) + \chi_{\rm trig}(g_{\kappa_{\rm sw},J_{\kappa_{\rm sw}}} - g_{\kappa_{\rm sw},i}), \tag{131}$$

and the last term is at most one because all batch-average losses lie in $[0,1]$.

If $h_{\rm cert} = 0$, only this trigger charge remains. Assume $h_{\rm cert} \geq 1$. On $\mathcal{G}^{\rm adv}$, Lemma E.16 applied at the last passed batch $h_{\rm cert}$ gives

$$G_{h_{\rm cert}} - h_{\rm cert}u_{h_{\rm cert}} \leq \mathfrak{B}_H + 2\Lambda_{\rm noise}(h_{\rm cert}). \tag{132}$$

We now bound the comparator part

$$h_{\rm cert}u_{h_{\rm cert}} - \sum_{r=1}^{h_{\rm cert}} g_{r,i}.$$

If arm i is active at $h_{\rm cert}$, then the active-comparator certificate passed at that batch and

$$h_{\rm cert}u_{h_{\rm cert}} - \sum_{r=1}^{h_{\rm cert}} g_{r,i} \leq 6h_{\rm cert}w_A(h_{\rm cert}). \tag{133}$$

Suppose instead that i was eliminated at batch $\tau_i \leq h_{\rm cert}$. The last batch at which i was certainly active and the BSC certificate passed is $\tau_i - 1$. Since (u_r) is non-increasing and $u_r, g_{r,i} \in [0,1]$, for $\tau_i \geq 2$,

$$\tau_i u_{\tau_i} - \sum_{r=1}^{\tau_i} g_{r,i} = (\tau_i - 1)u_{\tau_i} - \sum_{r<\tau_i} g_{r,i} + (u_{\tau_i} - g_{\tau_i,i})$$
$$\leq (\tau_i - 1)u_{\tau_i - 1} - \sum_{r<\tau_i} g_{r,i} + 1$$
$$\leq 6(\tau_i - 1)w_A(\tau_i - 1) + 1. \tag{134}$$

For $\tau_i = 1$, the left side is at most one, so the same conclusion holds with the first term omitted.

By definition, $\widetilde{v}_i = u_{\tau_i}$. Monotonicity gives $u_{h_{\mathrm{cert}}} \le \widetilde{v}_i$, and therefore

$$h_{\mathrm{cert}} u_{h_{\mathrm{cert}}} - \sum_{r=1}^{h_{\mathrm{cert}}} g_{r,i} \le \tau_i u_{\tau_i} - \sum_{r=1}^{\tau_i} g_{r,i} + \sum_{r=\tau_i+1}^{h_{\mathrm{cert}}} (\widetilde{v}_i - g_{r,i}). \tag{135}$$

The actual EAP phase intervals cover every batch in $\{\tau_i + 1, \dots, h_{\mathrm{cert}}\}$. On a switching path, the last terminal phase may additionally contain the switch-triggering batch $\kappa_{\mathrm{sw}} = h_{\mathrm{cert}} + 1$. Since each summand belongs to $[-1, 1]$,

$$\sum_{r=\tau_i+1}^{h_{\mathrm{cert}}} (\widetilde{v}_i - g_{r,i}) \le \sum_{\ell=1}^{C_i} R_\ell + 1. \tag{136}$$

Lemmas E.17, E.18, E.19, and E.20 therefore imply

$$\mathbb{E}_L\left[\sum_{r=\tau_i+1}^{h_{\mathrm{cert}}} (\widetilde{v}_i - g_{r,i})\right] \le O\Big(L_H\sqrt{KH}\Big) + 1.$$

The sum is bounded below by $-H$, so

$$\mathbb{E}_L\left[\mathbf{1}_{\mathcal{G}^{\mathrm{adv}}} \sum_{r=\tau_i+1}^{h_{\mathrm{cert}}} (\widetilde{v}_i - g_{r,i})\right] \le O\Big(L_H\sqrt{KH}\Big) + 1 + H \Pr_L((\mathcal{G}^{\mathrm{adv}})^c).$$

Combining (131)–(135), using

$$\mathfrak{B}_H = O(L_H^2\sqrt{KH}), \qquad \Lambda_{\mathrm{noise}}(H) + H w_A(H) = O(L_H\sqrt{KH})$$

when $H \ge K$ and the trivial bound $H \le \sqrt{KH}$ when $H < K$, proves (130). The failure-probability terms are absorbed because Lemma E.6 gives probability at most $\beta/2$ and the algorithm fixes $\beta = T^{-1}$. □

Theorem F.6 (Adversarial regret)**.** *For every oblivious loss table L, let $\epsilon_0 = \min\{\epsilon, 1\}$. Then*

$$R_T(\mathrm{PrivDaS}; L) \le O\left(L_H^2\sqrt{\frac{KT}{\epsilon_0}} + \frac{1}{\epsilon_0}\right).$$

Consequently the same upper bound holds for $R_T^{\mathrm{adv}}(\mathrm{PrivDaS})$. In the usual privacy regime $0 < \epsilon \le 1$, this is $O(L_H^2\sqrt{KT/\epsilon} + 1/\epsilon)$.

Proof. Let $i^\star \in \arg\min_{i\in[K]} \sum_{t=1}^T \ell_{t,i}$. The stochastic mode period consists of batches $r < \tau_{\mathrm{fb}}$, and the adversarial mode consists of batches $r \ge \tau_{\mathrm{fb}}$. Thus

$$R_H^{\mathrm{batch}} = \mathbb{E}\left[\sum_{r<\tau_{\mathrm{fb}}} (g_{r,J_r} - g_{r,i^\star})\right] + \mathbb{E}\left[\sum_{r=\tau_{\mathrm{fb}}}^{H} (g_{r,J_r} - g_{r,i^\star})\right]. \tag{137}$$

For the first term, split according to the good event $\mathcal{G}^{\mathrm{adv}}$:

$$\begin{aligned}
&\mathbb{E}\left[\sum_{r<\tau_{\mathrm{fb}}} (g_{r,J_r} - g_{r,i^\star})\right] \\
&= \mathbb{E}\left[\mathbf{1}_{\mathcal{G}^{\mathrm{adv}}} \sum_{r<\tau_{\mathrm{fb}}} (g_{r,J_r} - g_{r,i^\star})\right] + \mathbb{E}\left[\mathbf{1}_{(\mathcal{G}^{\mathrm{adv}})^c} \sum_{r<\tau_{\mathrm{fb}}} (g_{r,J_r} - g_{r,i^\star})\right] \\
&\le O\Big(L_H^2\sqrt{KH}\Big) + 1 + H \Pr_L((\mathcal{G}^{\mathrm{adv}})^c) \\
&\le O\Big(L_H^2\sqrt{KH}\Big) + 1,
\end{aligned} \tag{138}$$

where Lemma F.5 controls the good-event term, the stochastic mode batch regret is trivially at most H on every path, and Lemma E.6 gives $\Pr_L((\mathcal{G}^{\mathrm{adv}})^c) \le \beta/2$.

Conditional on the history $\mathscr{F}_{\tau_{\mathrm{fb}}-1}$, its random length

$$h = H - \tau_{\mathrm{fb}} + 1$$

is known, and $i^\star$ is a fixed comparator determined by the oblivious loss table. Lemma F.3 gives

$$\mathbb{E}_L\left[\sum_{r=\tau_{\mathrm{fb}}}^{H}(g_{r,J_r} - g_{r,i^\star})\middle|\mathscr{F}_{\tau_{\mathrm{fb}}-1}\right] \le 208\sqrt{Kh} + 2.$$

Taking expectation and using $h \le H$ yields

$$\mathbb{E}_L\left[\sum_{r=\tau_{\mathrm{fb}}}^{H}(g_{r,J_r} - g_{r,i^\star})\right] \le 208\sqrt{KH} + 2. \tag{139}$$

Combining (137), (138), and (139),

$$R_H^{\mathrm{batch}} \le O\Big(L_H^2\sqrt{KH}\Big) + 1.$$

Lemma F.2 gives

$$R_T(\mathrm{PrivDaS}; L) \le bR_H^{\mathrm{batch}} + b \le O\Big(L_H^2 b\sqrt{KH} + b\Big).$$

Since $H \le T/b$ and

$$b = \left\lceil \frac{2}{\epsilon_0} \right\rceil \le 1 + \frac{2}{\epsilon_0} \le \frac{3}{\epsilon_0},$$

we have

$$b\sqrt{KH} \le \sqrt{bKT} \le \sqrt{\frac{3KT}{\epsilon_0}}, \qquad b \le \frac{3}{\epsilon_0}.$$

This proves the theorem. □

G FUTURE WORK

In this section, we discuss several aspects of interesting future work to be explored further.

Adaptation to intermediate environments. Stochastic bandits with adversarial corruptions provide a natural setting between the two regimes considered in our guarantees. Non-private best-of-both-worlds analyses can exploit limited corruption to obtain intermediate regret bounds (Zimmert & Seldin, 2021). Our adversarial guarantee already applies to corrupted loss sequences that remain bounded and oblivious, but it does not quantify an improvement when the corruption is small. An interesting future direction is therefore a private regret guarantee that adapts to the amount of corruption, without requiring this quantity as an input. A difficulty for the present framework is that a temporary departure from stochastic behavior can trigger an irreversible switch, after which the present analysis provides only the adversarial guarantee. Corruption-tolerant certificates or carefully controlled restarts may address this issue, provided that their privacy costs and regret are analyzed under an explicit corruption model and comparator.

Heavy-tailed losses. Extending the framework beyond bounded losses is another direction, motivated by existing work on heavy-tailed best-of-both-worlds bandits (Huang et al., 2022; Cheng et al., 2024). Under suitable moment assumptions, one could investigate private robust estimators or clipped aggregates together with stochasticity certificates that account for estimation error and truncation bias. This extension requires both statistical and privacy control: a finite-moment assumption alone does not bound the worst-case sensitivity of the raw sums used by our Laplace mechanisms. Moreover, unbounded losses invalidate the deterministic bound of T on regret over exceptional events, so that part of the expected-regret analysis would also need to be replaced. A useful goal is to characterize how the attainable regret depends jointly on the tail parameters, suboptimality gaps, and privacy level.

More responsive updates and practical calibration. The time required to complete an active stage or an auditing phase can be substantial even when the asymptotic regret guarantees are favorable. This motivates investigating arm-specific update schedules, adaptive auditing rules, and tighter finite-sample calibration of the release and detection mechanisms. For example, allowing an arm's statistics to be updated without waiting for every active arm to complete a common stage could reduce the delay before elimination. Such changes require new justification: update and stopping decisions must respect the privacy analysis, and the resulting sampling process must preserve both stochastic safety and adversarial detection guarantees.