Batch-Size Independent Regret Bounds for Combinatorial Semi-Bandits with Probabilistically Triggered Arms or Independent Arms

Xutong Liu
The Chinese University of Hong Kong
Hong Kong SAR, China
liuxt@cse.cuhk.edu.hk

Jinhang Zuo
Carnegie Mellon University
Pittsburgh, PA, USA
jzuo@andrew.cmu.edu

Siwei Wang
Tsinghua University
Beijing, China
wangsw2020@mail.tsinghua.edu.cn

Carlee Joe-Wong
Carnegie Mellon University
Pittsburgh, PA, USA
cjoewong@andrew.cmu.edu

John C.S. Lui
The Chinese University of Hong Kong
Hong Kong SAR, China
cslui@cse.cuhk.edu.hk

Wei Chen
Microsoft Research
Beijing, China
weic@microsoft.com

Abstract

In this paper, we study the combinatorial semi-bandits (CMAB) and focus on reducing the dependency of the batch-size K in the regret bound, where K is the total number of arms that can be pulled or triggered in each round. First, for the setting of CMAB with probabilistically triggered arms (CMAB-T), we discover a novel (directional) triggering probability and variance modulated (TPVM) condition that can replace the previously-used smoothness condition for various applications, such as cascading bandits, online network exploration and online influence maximization. Under this new condition, we propose a BCUCB-T algorithm with variance-aware confidence intervals and conduct regret analysis which reduces the $O(K)$ factor to $O(\log K)$ or $O(\log^2 K)$ in the regret bound, significantly improving the regret bounds for the above applications. Second, for the setting of non-triggering CMAB with independent arms, we propose a SESCB algorithm which leverages on the non-triggering version of the TPVM condition and completely removes the dependency on K in the leading regret. As a valuable by-product, the regret analysis used in this paper can improve several existing results by a factor of $O(\log K)$. Finally, experimental evaluations show our superior performance compared with benchmark algorithms in different applications.

1 Introduction

Stochastic multi-armed bandit (MAB) [26, 3, 4] is a classical model that has been extensively studied in online decision making. As an extension of MAB, combinatorial multi-armed bandits (CMAB) have drawn much attention recently, owing to its wide applications in marketing, network optimization and online advertising [13, 17, 7, 8, 29, 23]. In CMAB, the learning agent chooses a combinatorial action in each round, and this action would trigger a set of arms (or a super arm) to be pulled simultaneously, and the outcomes of these pulled arms are observed as feedback. Typically, such feedback is known as the semi-bandit feedback. The agent's goal is to minimize the expected *regret*, which is the difference in expectation for the overall rewards between always playing the best action

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

Table 1: Summary of the algorithms and results for CMAB with probabilistically triggered arms.

Algorithm	Smoothness	Independent Arms?	Computation	Regret
CUCB [29]	1-norm TPM, B_1	Not required	Efficient	$O(K \sum_{i\in[m]} \frac{B_1^2 \log T}{\Delta_i^{\min}})$
BOIM-CUCB [25, Section 4]*	1-norm TPM, B_1	Required	Hard	$O((\log K)^2 \sum_{i\in[m]} \frac{B_1^2 \log T}{\Delta_i^{\min}})$
BCUCB-T (Algorithm 1)	TPVM$_<$, B_v,† $\lambda > 1$	Not required	Efficient	$O(\log K \sum_{i\in[m]} \frac{B_v^2 \log T}{\Delta_i^{\min}})$
BCUCB-T (Algorithm 1)	TPVM$_<$, B_v,† $\lambda = 1$	Not required	Efficient	$O((\log K \log \frac{B_v K}{\Delta_{\min}} \sum_{i\in[m]} \frac{B_v^2 \log T}{\Delta_i^{\min}})$
BOIM-CUCB (Appendix C)‡	1-norm TPM, B_1	Required	Hard	$O(\log K \sum_{i\in[m]} \frac{B_1^2 \log T}{\Delta_i^{\min}})$

* This work is for a specific application, but we treat it as a general framework; † Generally, $B_v = O(B_1\sqrt{K})$, and the existing regret bound is improved when $B_v = o(B_1\sqrt{K})$; ‡ Using our new analysis.

(i.e., the action with highest expected reward) and playing according to the agent's own policy. For CMAB, an agent not only need to deal with the exploration-exploitation tradeoff: whether the agent should explore arms in search for a better action, or should the agent stick to the best action observed so far to gain rewards; but also need to handle the exponential explosion of all possible actions.

To model a wider range of application scenarios where action may trigger arms probabilistically, Chen et al. [8] first generalize CMAB to CMAB with probabilistically triggered arms (or CMAB-T for short), which successfully covers cascading bandit [9] (CB) and online influence maximization (OIM) bandit [31] problems. Later on, Wang and Chen [29] improve the regret bound of [8] by introducing a smoothness condition, called the triggering probability modulated (TPM) condition, which removes a factor of $1/p^*$ compared to [8], where p^* is the minimum positive probability that any arm can be triggered. However, in both studies, the regret bounds still depend on a factor of *batch-size* K, where K is the maximum number of arms that can be triggered, and this factor could be quite large, e.g., for OIM K can be as large as the number of edges in a large social network.

Our Contributions. In this paper, we reduce or remove the dependency on K in the regret bounds. For CMAB-T, we first discover a new *triggering probability and variance modulated (TPVM)* bounded smoothness condition, which is stronger than the TPM condition, yet still holds for several applications (such as CB and OIM) where only the TPM condition is known previously. We observe that for these applications, the previous TPM condition bounds the global speed of reward change regarding the parameter change, which will cause a large K coefficient due to the rapid change at the boundary regions (i.e., when an arm's mean μ_i is close to 0 or 1). Our TPVM condition utilizes this observation by raising up the regret contribution of those boundary regions, leading to a significant reduction on the dependency of K. Second, we propose a "variance-aware" BCUCB-T algorithm that adaptively changes the width of the confidence interval according to the (empirical) variance, cancelling out the large regret contribution raised by the TPVM condition at the boundary regions (where the variances are also very small). Combining these two techniques, we successfully reduce the batch-size dependence from $O(K)$ to $O(\log K)$ or $O(\log^2 K)$ for all CMAB-T problems satisfying the TPVM condition, leading to significant improvements of the regret bounds for applications like CB or OIM. As a by-product, we also give refined proofs that shall improve the regret for several existing works by a factor of $O(\log K)$, e.g., [11, 23], which may be of independent interests.

In addition to the general CMAB-T setting, we show how a non-triggering version of the TPVM condition (i.e., VM condition) can help to completely remove the batch-size K, under the additional independent arm assumption for non-triggering CMAB problems. In particular, we propose a novel Sub-Exponential Efficient Sampling for Combintorial Bandits Policy (SESCB) that produces tighter sub-exponential concentrated confidence intervals. In our analysis, we show that the total regret only depends on the arm that is observed least instead of all K arms, so that we can achieve a completely batch-size independent regret bound. Our empirical results demonstrate that our proposed algorithms can achieve around 20% lower regrets than previous ones for several applications. Due to the space limit, we will move the complete proofs and empirical results into the appendix.

Related Work. The stochastic CMAB has received much attention recently. From the modelling point of view, these CMAB works can be divided into two categories: CMAB with or without probabilistically triggered arms (i.e. CMAB-T setting or non-triggering CMAB). For CMAB-T, our work improves (a) the general framework in [8, 29], (b) the combinatorial cascading bandit [17], (c) the online multi-layered network exploration [21] problem, (d) the online influence maximization bandits [29, 25], by reducing or removing the batch-size dependent factor K in the regret bounds with our new TPVM condition and/or our refined analysis. We defer the detailed technical comparison to Section 3.1 and Section 5. For the algorithm, most CMAB-T studies use Combinatorial Upper Confidence Bound (CUCB) based on Chernoff concentration bounds [29], our BCUCB-T algorithm is different and uses the Bernstein concentration bound [2, 23] that considers variance of the arms.

Table 2: Summary of the algorithms and results for non-triggering CMAB problems.

Algorithm	Smoothness	Independent Arms?	Computation	Regret
CUCB [29]	1-norm, B_1	Not required	Efficient	$O(K\sum_{i\in[m]}\frac{B_1^2\log T}{\Delta_i^{\min}})$
CTS [30]*	1-norm, B_1	Required	Efficient	$O(K\sum_{i\in[m]}\frac{B_1^2\log T}{\Delta_i^{\min}})$
ESCB [9]	1-norm, B_1**	Required	Hard	$O((\log K)^2\sum_{i\in[m]}\frac{B_1^2\log T}{\Delta_i^{\min}})$
AESCB [10]	Linear	Required	Efficient	$O((\log K)^2\sum_{i\in[m]}\frac{\log T}{\Delta_i^{\min}})$
BC-UCB [23]†	VM, B_v ‡	Not required	Efficient	$O((\log K)^2\sum_{i\in[m]}\frac{B_v^2\log T}{\Delta_i^{\min}})$.
CTS [30]*	Linear	Required	Efficient	$O(\log K\sum_{i\in[m]}\frac{\log T}{\Delta_i^{\min}})$
SESCB (Algorithm 2)	VM, B_v ‡	Required	Efficient***	$O(\sum_{i\in[m]}\frac{B_v^2\log T}{\Delta_i^{\min}})$
BC-UCB (Appendix C)§	VM, B_v ‡	Not required	Efficient	$O(\log K\sum_{i\in[m]}\frac{B_v^2\log T}{\Delta_i^{\min}})$

* Requires exact offline oracle instead of (α,β)-approximate oracle; † This work gives sufficient smoothness condition with factor γ_g and translates to $B_v = 3\sqrt{2}\gamma_g$ in our setting; § Using our new analysis. ** This work is for the linear case, but can easily generalize to 1-norm B_1 case; ‡ Generally, $B_v = O(B_1\sqrt{K})$ and the existing regret bound is improved when $B_v = o(B_1\sqrt{K})$; *** Efficient when the reward function is submodular, otherwise the computation is hard;

For non-triggering CMAB, [13] is the first study on stochastic CMAB, and its regret has been improved by Kveton et al. [18], Combes et al. [9], Chen et al. [8], but they still have $O(K)$ factor in their regrets. When arms are mutually independent, Combes et al. [9] build a tighter ellipsoidal confidence region for exploration, and devise the Efficient Sampling for Combinatorial Bandit policy (ESCB), which reduces the dependence on $O(K)$ to $O(\log^2 K)$ at the cost of high computational complexity (since combinatorial optimization over the ellipsoidal region is NP-hard in general [1]). Later on, the computational complexity is improved by AESCB [10] in the linear CMAB problem. Recently, Merlis and Mannor [23] focus on the Probabilistic Maximum Coverage (PMC) bandit problem and propose the BC-UCB algorithm with the Gini-smoothness condition to achieve a similar improvement as ESCB/AESCB, but without the independent arm assumption. Our work is largely inspired by their work, however, our study generalizes theirs to the CMAB-T setting which can handle much broader application scenarios beyond the non-triggering CMAB (more detailed comparison is given in Section 3). In addition, we provide a refined analysis that can save a $O(\log K)$ factor for BC-UCB (or ESCB/AESCB) algorithm. Compared with other ESCB-type algorithms for independent arms, as far as we know, our SESCB algorithm are the first to completely remove the dependence of K in the leading regrets, owing to our non-triggering version of the TPVM condition. The detailed comparisons are summarized in Table 1 and Table 2.

The usage of variance-aware algorithms to give improved regret bounds can be dated back to [2]. Recently, there is a surge of interest to apply the variance-aware principle in bandit [23, 28] and reinforcement learning (RL) settings [33, 32]. It is notable that Vial et al. [28] share a similar variance-aware principle as ours but focus on the distribution-independent regret bounds for the cascading bandits [28]. Our work is more general and achieves the matching regret bound when translating to the distribution-independent regret bound. Compared with RL works, our paper studies a different setting as we do not consider the state transitions.

From the application's point of view, this paper covers the applications of PMC bandit [23], combinatorial cascading bandits [17, 19], network exploration [21], and online influence maximization [31, 29, 20]. Our proposed algorithms can significantly reduce the regret bounds of them, e.g., from $O(K)$ to $O(\log^2 K)$ for OIM where K can be hundreds of thousands in large social networks.

2 Problem Settings

We study the combinatorial multi-armed bandit problem with probabilistic triggering arms, which is denoted as CMAB-T for short. Following the setting from [29], a CMAB-T *problem instance* can be described by a tuple $([m], \mathcal{S}, \mathcal{D}, D_{\text{trig}}, R)$, where $[m] = \{1, 2, ..., m\}$ is the set of base arms; $\mathcal{S}$ is the set of eligible actions and $S \in \mathcal{S}$ is an action;[1] $\mathcal{D}$ is the set of possible distributions over the outcomes of base arms with bounded support $[0,1]^m$; D_{trig} is the probabilistic triggering function and R is the reward function, the definitions of which will be introduced shortly.

In CMAB-T, the learning agent interacts with the unknown environment in a sequential manner as follows. First, the environment chooses a distribution $D \in \mathcal{D}$ unknown to the agent. Then, at round

[1] In some cases $\mathcal{S}$ is a collection of subsets of $[m]$, in which case we often refer to $S \in \mathcal{S}$ as a super arm. In this paper we treat $\mathcal{S}$ as a general action space, same as in [29].

$t = 1, 2, ..., T$, the agent selects an action $S_t \in \mathcal{S}$ and the environment draws from the unknown distribution D a random outcome $\boldsymbol{X}_t = (X_{t,1}, ...X_{t,m}) \in [0, 1]^m$. Note that the outcome $\boldsymbol{X}_t$ is assumed to be independent from outcomes generated in previous rounds, but outcomes $X_{t,i}$ and $X_{t,j}$ in the same round could be correlated. Let $D_{\text{trig}}(S, \boldsymbol{X})$ be a distribution over all possible subsets of $[m]$, i.e. its support is $2^{[m]}$. When the action S_t is played on the outcome $\boldsymbol{X}_t$, base arms in a random set $\tau_t \sim D_{\text{trig}}(S_t, \boldsymbol{X}_t)$ are triggered, meaning that the outcomes of arms in τ_t, i.e. $(X_t)_{t\in\tau_t}$ are revealed as the feedback to the agent, and are involved in determining the reward of action S_t. Function D_{trig} is referred as the *probabilistic triggering function*. At the end of the round t, the agent will receive a non-negative reward $R(S_t, \boldsymbol{X}_t, \tau_t)$, determined by $S_t, \boldsymbol{X}_t$ and τ_t. CMAB-T significantly enhances the modeling power of CMAB [7, 18] and can model many applications such as cascading bandits and online influence maximization [29], which we will discuss in later sections.

The goal of CMAB-T is to accumulate as much reward as possible over T rounds, by learning distribution D or its parameters. Let $\boldsymbol{\mu} = (\mu_1, ..., \mu_m)$ denote the mean vector of base arms' outcomes. Following [29], we assume that the expected reward $\mathbb{E}[R(S, \boldsymbol{X}, \tau)]$ is a function of the unknown mean vector $\boldsymbol{\mu}$, where the expectation is taken over the randomness of $\boldsymbol{X} \sim D$ and $\tau \sim D_{\text{trig}}(S, \boldsymbol{X})$. In this context, we denote $r(S; \boldsymbol{\mu}) \triangleq \mathbb{E}[R(S, \boldsymbol{X}, \tau)]$ and it suffices to learn the unknown mean vector instead of the joint distribution D, based on the past observation.

The performance of an online learning algorithm A is measured by its *regret*, defined as the difference of the expected cumulative reward between always playing the best action $S^* \triangleq \operatorname{argmax}_{S\in\mathcal{S}} r(S; \boldsymbol{\mu})$ and playing actions chosen by algorithm A. For many reward functions, it is NP-hard to compute the exact S^* even when $\boldsymbol{\mu}$ is known, so similar to [29], we assume that the algorithm A has access to an offline (α, β)-approximation oracle, which for mean vector $\boldsymbol{\mu}$ outputs an action S such that $\Pr[r(S; \boldsymbol{\mu}) \geq \alpha \cdot r(S^*; \boldsymbol{\mu})] \geq \beta$. Formally, the T-round (α, β)-approximate regret is defined as

$$Reg(T; \alpha, \beta, \boldsymbol{\mu}) = T \cdot \alpha\beta \cdot r(S^*; \boldsymbol{\mu}) - \mathbb{E}\left[\sum_{t=1}^{T} r(S_t; \boldsymbol{\mu})\right], \quad (1)$$

where the expectation is taken over the randomness of outcomes $\boldsymbol{X}_1, ..., \boldsymbol{X}_T$, the triggered sets $\tau_1, ..., \tau_T$, as well as the randomness of algorithm A itself.

In the CMAB-T model, there are several quantities that are crucial to the subsequent study. We define *triggering probability* $p_i^{D,D_{\text{trig}},S}$ as the probability that base arm i is triggered when the action is S, the outcome distribution is D, and the probabilistic triggering function is D_{trig}. Since D_{trig} is always fixed in a given application context, we ignore it in the notation for simplicity, and use $p_i^{D,S}$ henceforth. Triggering probabilities $p_i^{D,S}$'s are crucial for the triggering probability modulated bounded smoothness conditions to be defined below. We define *batch size* K as the maximum number of arms that can be triggered, i.e., $K = \max_{S\in\mathcal{S}} |\{i \in [m] : p_i^{D,S} > 0\}|$. Our main contribution of this paper is to remove or reduce the regret dependency on batch size K, where K could be quite large, e.g., K can be hundreds of thousands in a large social network.

Owing to the nonlinearity and the combinatorial structure of the reward, it is essential to give some conditions for the reward function in order to achieve any meaningful regret bounds [7, 8, 29, 11, 23]. The following are two standard conditions originally proposed by Wang and Chen [29].

Condition 1 (Monotonicity)**.** *We say that a CMAB-T problem instance satisfies monotonicity condition, if for any action $S \in \mathcal{S}$, any two distributions $D, D' \in \mathcal{D}$ with mean vectors $\boldsymbol{\mu}, \boldsymbol{\mu}' \in [0, 1]^m$ such that $\mu_i \leq \mu'_i$ for all $i \in [m]$, we have $r(S; \boldsymbol{\mu}) \leq r(S; \boldsymbol{\mu}')$.*

Condition 2 (1-norm TPM Bounded Smoothness)**.** *We say that a CMAB-T problem instance satisfies the triggering probability modulated (TPM) B_1-bounded smoothness condition, if for any action $S \in \mathcal{S}$, any distribution $D, D' \in \mathcal{D}$ with mean vectors $\boldsymbol{\mu}, \boldsymbol{\mu}' \in [0, 1]^m$, we have $|r(S; \boldsymbol{\mu}') - r(S; \boldsymbol{\mu})| \leq B_1 \sum_{i\in[m]} p_i^{D,S} |\mu_i - \mu'_i|$.*

The first monotonicity condition indicates the reward is larger if the parameter vector $\boldsymbol{\mu}$ is larger. The second condition bounds the reward difference caused by the parameter change (from $\boldsymbol{\mu}$ to $\boldsymbol{\mu}'$). One key feature is that the parameter change in each base arm $i \in [m]$ is modulated by the triggering probability $p_i^{D,S}$. Intuitively, for base arm i that is unlikely to be triggered/observed (small $p_i^{D,S}$), Condition 2 ensures that a large change in μ_i only causes a small change (multiplied by $p_i^{D,S}$) in the reward, and thus one does not need to pay extra cost to observe such arms. Many applications satisfy Condition 1 and Condition 2, including linear combinatorial bandits [18], combinatorial cascading bandits [17], online influence maximization [29], etc. With the above two conditions, Wang and Chen [29] show that a CUCB algorithm achieves the distribution-dependent regret bound of $O(\sum_{i\in[m]} \frac{B_1^2 K \log T}{\Delta_i^{\min}})$, where $\Delta_i^{\min}$ is the distribution-dependent reward gap, to be formally defined in Definition 1. In the following sections, we will show how to remove or reduce the dependency on K in the above bounds under our new conditions.

3 Algorithm and Regret Analysis for CMAB-T

In this section, for the CMAB-T framework with probabilistic triggering, we improve the regret dependency on the batch size from $O(K)$ in [29] to $O(\log K)$ or $O(\log^2 K)$. Our main tool is a new condition called *triggering probability and variance modulated (TPVM) bounded smoothness condition*, replacing the TPM condition (Condition 2). We will define the TPVM condition, comparing it with the TPM condition and the gini-smoothness condition of [23], show our algorithm and regret analysis that utilize this condition. Later in Section 5, we will demonstrate how this condition is applied to applications such as cascading bandits and online influence maximization.

3.1 Triggering Probability and Variance Modulated (TPVM) Bounded Smoothness Condition

In this paper, we discover a new smoothness condition for many important applications as follows.

Condition 3 (Directional TPVM Bounded Smoothness)**.** *We say that a CMAB-T problem instance satisfies the directional TPVM (B_v, B_1, λ)-bounded smoothness condition ($B_v, B_1 \geq 0, \lambda \geq 1$), if for any action $S \in \mathcal{S}$, any distribution $D, D' \in \mathcal{D}$ with mean vector $\boldsymbol{\mu}, \boldsymbol{\mu}' \in (0,1)^m$, for any non-negative $\boldsymbol{\zeta}, \boldsymbol{\eta} \in [0,1]^m$ s.t. $\boldsymbol{\mu}' = \boldsymbol{\mu} + \boldsymbol{\zeta} + \boldsymbol{\eta}$, we have*

$$|r(S;\boldsymbol{\mu}') - r(S;\boldsymbol{\mu})| \leq B_v \sqrt{\sum_{i\in[m]} (p_i^{D,S})^{\lambda} \frac{\zeta_i^2}{(1-\mu_i)\mu_i}} + B_1 \sum_{i\in[m]} p_i^{D,S} \eta_i. \quad (2)$$

Remark 1 (Intuition for Condition 3). Looking at Eq. (2), if we ignore the $(1-\mu_i)\mu_i$ term in the denominator and set $\lambda = 2$, the RHS of Eq. (2) becomes $B_v\sqrt{\sum_{i\in[m]}(p_i^{D,S})^2\zeta_i^2} + B_1 \sum_{i\in[m]} p_i^{D,S}\eta_i$, which holds with $B_v = B_1\sqrt{K}$ by applying the Cauchy-Schwarz inequality to Condition 2. However, the regret upper bound following this modified Eq. (2) would not directly lead to the improvement in the regret due to the $\sqrt{K}$ factor in B_v. To deal with this issue, an important observation here is that for many applications, the reason B_v is large is because that the reward changes abruptly when parameters μ_i approaches 0 or 1. This motivates us to plug in the $1/(1-\mu_i)\mu_i$ term in Eq. (2) to enlarge the square root term when μ_i is close to 0 or 1, so that B_v can be as small as possible. On the other hand, notice that when μ_i approaches 0 or 1, the variance $V_i \leq (1-\mu_i)\mu_i$ is also very small, [2] so the estimation of μ_i should be quite accurate. Therefore, the gap ζ_i between our estimation and true value produces a variance-related term which cancels the $(1-\mu_i)\mu_i$ in the denominator. Since ζ_i in Eq. (2) is modulated by both triggering probability $p_i^{D,S}$ and inverse upper bound of the variance $1/(1-\mu_i)\mu_i$, we call Condition 3 the directional triggering probability and variance modulated (TPVM) condition for short, where the term "directional" is explained in the next remark. The exponent $\lambda \geq 1$ on the triggering probability gives flexibility to trade-off between the strength of the condition and the quantity of the regret bound: With a larger λ, we can obtain a smaller regret bound, while with a smaller λ, the condition is easier to satisfy and allows us to include more applications.

Remark 2 (On directional TPVM vs. undirectional TPVM). In the above definition, "directional" means that we have $\boldsymbol{\zeta}, \boldsymbol{\eta} \geq \boldsymbol{0}$ such that $\boldsymbol{\mu}' \geq \boldsymbol{\mu}$ in every dimension. This is weaker than the version of the undirectional TPVM condition, where $\boldsymbol{\zeta}, \boldsymbol{\eta} \in [-1,1]^m$, and the η_i in the right hand side of Eq.(2) is replaced with $|\eta_i|$. The reason we use the weaker version is that some of our applications considered in this paper only satisfy the weaker version. To differentiate, we use TPVM$_<$ when we refer to the directional TPVM condition.

Remark 3 (Relation between Conditions 2 and 3). First, when setting $\boldsymbol{\zeta}$ to $\boldsymbol{0}$, the directional TPVM condition degenerates to the directional TPM condition. However, Condition 2 is the undirectional TPM condition, which is typically stronger than its directional counterpart. Thus, in general Condition 3 does not imply Condition 2. Nevertheless, with some additional assumptions Condition 3 does imply

[2] For bounded random variable $X \in [0,1]$ with mean μ_i, variance $V_i = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \leq \mathbb{E}[X] - (\mathbb{E}[X])^2 \leq (1-\mu_i)\mu_i$, where the equality is achieved when X is a Bernoulli random variable.

Algorithm 1 BCUCB-T: Bernstein Combinatorial Upper Confidence Bound Algorithm for CMAB-T

1: **Input:** Base arms $[m]$, computation oracle ORACLE.
2: **Initialize:** For each arm i, $T_{0,i} \leftarrow 0, \hat{\mu}_{0,i} = 0, \hat{V}_{0,i} = 0$.
3: **for** $t = 1, ..., T$ **do**
4: For arm i, compute $\rho_{t,i}$ according to Eq. (3) and set UCB value $\bar{\mu}_{t,i} = \min\{\hat{\mu}_{t-1,i} + \rho_{t,i}, 1\}$.
5: S_t = ORACLE$(\bar{\mu}_{t,1}, ..., \bar{\mu}_{t,m})$.
6: Play S_t, which triggers arms $\tau_t \subseteq [m]$ with outcome $X_{t,i}$'s, for $i \in \tau_t$.
7: For every $i \in \tau_t$, update $T_{t,i} = T_{t-1,i} + 1$, $\hat{\mu}_{t,i} = \hat{\mu}_{t-1,i} + (X_{t,i} - \hat{\mu}_{t-1,i})/T_{t,i}$, $\hat{V}_{t,i} = \frac{T_{t-1,i}}{T_{t,i}}\left(\hat{V}_{t-1,i} + \frac{1}{T_{t,i}}(\hat{\mu}_{t-1,i} - X_{t,i})^2\right)$.
8: **end for**

Condition 2 with the same coefficient B_1 (See Appendix A for an example of such assumptions). Conversely, by applying the Cauchy-Schwartz inequality, one can verify that if a reward function is TPM B_1-bounded smooth, then it is (directional) TPVM $(B_1\sqrt{K}/2, B_1, \lambda)$-bounded smooth for any $\lambda \leq 2$. For applications considered in this paper, we are able to reduce their B_v coefficient from $B_1\sqrt{K}/2$ to a coefficient independent of K, leading to significant savings in the regret bound.

Remark 4 (Comparing with [23]). Merlis and Mannor [23] introduce a Gini-smoothness condition to reduce the batch-size dependency for CMAB problems, which largely inspires our TPVM$_<$ condition. Their condition is specified in a differential form of the reward function, with parameters γ_∞ and γ_g (See Appendix B for the exact definition). We emphasize that their original condition cannot handle the probabilistic triggering setting in CMAB-T. One natural extension is to incorporate triggering probability modulation into their differential form of Gini-smoothness. However, we found that the resulting TPM Gini-smoothness condition is not strong enough to guarantee desirable regret bounds (See Appendix B.1). This motivates us to provide a new condition directly on the difference form $|r(S;\boldsymbol{\mu}') - r(S;\boldsymbol{\mu})|$, similar to the TPM condition in [29]. Our TPVM$_<$ condition (Condition 3) can be viewed as extending Lemma 6 of [23] to incorporate triggering probabilities and bound the difference form $|r(S;\boldsymbol{\mu}') - r(S;\boldsymbol{\mu})|$. Intuitively, B_1 and B_v correspond to γ_∞ and γ_g, respectively, but since they are for different forms of definitions, their numerical values may not exactly match one another.

3.2 BCUCB-T Algorithm and Regret Analysis

Our proposed algorithm BCUCB-T is a generalization of the BC-UCB algorithm [23, Algorithm 1] which originally solves the non-triggering CMAB problem. Algorithm 1 maintains the empirical estimate $\hat{\mu}_{t,i}$ and $\hat{V}_{t,i}$ for the true mean and the true variance of the base arm outcomes. To select the action S_t, it feeds the upper confidence bound $\bar{\mu}_i$ into the offline oracle, where $\bar{\mu}_i$ optimistically estimates the μ_i by a confidence interval $\rho_{t,i}$. Compared with the CUCB algorithm [29, Algorithm 1] which uses confidence interval $\rho_{t,i} = \sqrt{\frac{3\log t}{2T_{t-1,i}}}$ for the CMAB-T problem, the novel part is the usage of empirical variance $\hat{V}_{t-1,i}$ to construct the following "variance-aware" confidence interval:

$$\rho_{t,i} = \sqrt{\frac{6\hat{V}_{t-1,i}\log t}{T_{t-1,i}}} + \frac{9\log t}{T_{t-1,i}} \tag{3}$$

This confidence interval leverages on the empirical Bernstein inequality instead of the Chernoff-Hoeffding inequality. As we will show in Appendix C.1, for the first term in Eq. (3), $\hat{V}_{t-1,i}$ is approximately equal to the true variance $V_i \leq (1-\mu_i)\mu_i$ and this indicates the estimation of μ_i is more accurate when μ_i is close to 0 or 1, which will cancel out the $(1-\mu_i)\mu_i$ coefficient of the B_v term in Condition 3 as we discussed before. The second term of Eq. (3) is to compensate the usage of the empirical variance $\hat{V}_{t-1,i}$, rather than the true variance V_i which is unknown to the learner.

To state the regret bound, we first give some definitions followed by our main result.

Definition 1 ((Approximation) Gap). *Fix a distribution $D \in \mathcal{D}$ and its mean vector $\boldsymbol{\mu}$, for each action $S \in \mathcal{S}$, we define the (approximation) gap as $\Delta_S = \max\{0, \alpha r(S^*;\boldsymbol{\mu}) - r(S;\boldsymbol{\mu})\}$. For each arm i, we define $\Delta_i^{\min} = \inf_{S\in\mathcal{S}:p_i^{D,S}>0,\,\Delta_S>0}\Delta_S$, $\Delta_i^{\max} = \sup_{S\in\mathcal{S}:p_i^{D,S}>0,\Delta_S>0}\Delta_S$. As a convention, if there is no action $S \in \mathcal{S}$ such that $p_i^{D,S} > 0$ and $\Delta_S > 0$, then $\Delta_i^{\min} = +\infty, \Delta_i^{\max} = 0$. We define $\Delta_{\min} = \min_{i\in[m]}\Delta_i^{\min}$ and $\Delta_{\max} = \max_{i\in[m]}\Delta_i^{\min}$.*

Theorem 1. *For a CMAB-T problem instance $([m], \mathcal{S}, \mathcal{D}, D_{trig}, R)$ that satisfies monotonicity (Condition 1), and TPVM$_<$ bounded smoothness (Condition 3) with coefficient (B_v, B_1, λ),*

(1) if $\lambda > 1$, BCUCB-T (Algorithm 1) with an (α, β)-approximation oracle achieves an (α, β)-approximate regret bounded by

$$O\left(\sum_{i\in[m]} \frac{B_v^2 \log K \log T}{\Delta_i^{\min}} + \sum_{i\in[m]} B_1 \log^2\left(\frac{B_1 K}{\Delta_i^{\min}}\right) \log T\right); \tag{4}$$

(2) if $\lambda = 1$, BCUCB-T (Algorithm 1) with an (α, β)-approximation oracle achieves an (α, β)-approximate regret bounded by

$$O\left(\sum_{i\in[m]} \log\left(\frac{B_v K}{\Delta_i^{\min}}\right) \frac{B_v^2 \log K \log T}{\Delta_i^{\min}} + \sum_{i\in[m]} B_1 \log^2\left(\frac{B_1 K}{\Delta_i^{\min}}\right) \log T\right). \tag{5}$$

Remark 5 (Discussion for Regret Bounds). Looking at the above regret bounds, for $\lambda > 1$ and $\lambda = 1$, the leading terms are $O(\sum_{i=1}^m \frac{B_v^2 \log K \log T}{\Delta_i^{\min}})$ and $O(\sum_{i=1}^m (\log \frac{B_v K}{\Delta_i^{\min}}) \frac{B_v^2 \log K \log T}{\Delta_i^{\min}})$. When $B_v \geq B_1$ (which typically holds, see Section 5) and gaps are small (i.e., $\Delta^i_{\min} \leq 1/\log^2 K$), the dependencies over K are $O(\log K)$ and $O(\log^2 K)$, respectively. For the setting of CMAB-T, [29] is the closest work to our paper, where the reward function satisfies Condition 1 and Condition 2 with coefficient B_1. As mentioned in Remark 3 in Section 3.1, their reward function trivially satisfies our Condition 3 with coefficient $(B_1\sqrt{K}/2, B_1, 2)$ so our work reproduces a bound of $O(\sum_{i\in[m]} \frac{B_1^2 K \log K \log T}{\Delta_i^{\min}})$, matching [29] up to a factor of $O(\log K)$. As will be shown in Section 5, for applications that satisfy TPVM (or TPVM$_<$) condition with non-trivial B_v, i.e., $B_v = o(B_1\sqrt{K})$, our work improves their regret bounds up to a factor of $O(K/\log K)$. As for the lower bound, according to the lower bound results by Merlis and Mannor [24], our regret bound is tight up to a factor of $O(\log^2 K)$ on the (degenerate) non-triggering CMAB case. We defer the details about the lower bound results and the distribution-independent regret bounds in the Appendix C.5.

Proof ideas. Our proof uses a few events to filter the total regret and then bound these event-filtered regrets separately. As will be shown in the supplementary material, the event that contributes to the leading regret is $E_t = \{\Delta_{S_t} \leq e_t(S_t)\}$, where the error term $e_t(S_t) = O(B_v\sqrt{\sum_{i\in\tilde{S}_t}(\frac{\log t}{T_{t-1,i}})(p_i^{D,S_t})^\lambda} + B_1\sum_{i\in\tilde{S}_t}(\frac{\log t}{T_{t-1,i}})(p_i^{D,S_t}))$. To handle the probabilistic triggering, our key ingredient is to use the triggering probability group technique proposed by Wang and Chen [29] in the definition of above events. For the $\lambda = 1$ case, one new issue arises since the triggering probability group divides sub-optimal actions S into *infinite* geometrically separated bins $(1/2, 1], (1/4, 1/2]..., (2^{-j}, 2^{-j+1}), ...,$ over $p_i^{D,S}$, and the regret should be proportional to the number of bins (which are infinitely large). To handle this, we show that it suffices to consider the first $j \leq j_i^{\max} = O(\log \frac{B_v K}{\Delta_i^{\min}})$ bins (which is why Eq. (5) has this additional factor in the leading term) and the regret of other bins (with very small $p_i^{D,S}$) can be safely neglected. To bound the leading regret filtered by E_t as mentioned earlier, we use the reverse amortization trick from Wang and Chen [29, 30] and adaptively allocates each arm's regret contribution (according to thresholds on the number of times arm i is triggered). Note that these thresholds are carefully chosen for the error term $e_t(S_t)$, since trivially following the thresholds in Wang and Chen [29] would either yield no meaningful bound or suffer from additional $O(\log T)$ or $O(\log K)$ factors in the regret. As a by-product, one can also use our analysis to replace that of Merlis and Mannor [23] and Perrault et al. [25] (where similar error term $e_t(S_t)$ appears) to improve their bound by a factor of $O(\log K)$. For the detailed proofs, we defer them in the Appendix C. ■

4 Algorithm and Analysis For CMAB with Independent Arms

In this section, we aim to show that for the non-triggering CMAB, the assumption that all arms are independent, compounded with a non-triggering version of the above TPVM condition (named as VM condition below), together allow us to completely remove the $O(\log^2 K)$ or $O(\log K)$ dependence

Algorithm 2 SESCB: Sub-Exponential Sampling for Combinatorial Bandits with Independent Arms

1: **Input:** Base arms $[m]$, sub-Gaussian parameter C_1, VM smoothness coefficient B_v, (α, β)-approximation ORACLE $\bar{O}$.
2: **Initialize:** For each arm i, $T_{0,i} \leftarrow 0, \hat{\mu}_{0,i} = 0$.
3: **for** $t = 1, ..., T$ **do**
4:     For all $S \in \mathcal{S}$, define min-count $T^{\min}_{t-1,S} = \min_{i \in S} T_{t-1,i}$, let interval $\rho_t(S) = B_v\sqrt{\sum_{i\in S}\frac{C_1}{T_{t-1,i}} + \max\left\{8C_1\sqrt{\sum_{i\in S}\frac{\log(2|\mathcal{S}|T)}{T_{t-1,i}^2}}, \frac{8C_1\log(2|\mathcal{S}|T)}{T^{\min}_{t-1,S}}\right\}}$.
5:     For all $S \in \mathcal{S}$, define optimistic reward $\bar{r}_t(S) = r(S; \hat{\boldsymbol{\mu}}_{t-1}) + \rho_t(S)$.
6:     Play $S_t = \bar{O}(\hat{\boldsymbol{\mu}}_t, \boldsymbol{T}_t)$ s.t. $\Pr\left[\bar{r}_t(S) \geq \alpha \cdot \bar{r}_t(\bar{S}_t^*)\right] \geq \beta$,where $\bar{S}_t^* = \arg\max_{S\in\mathcal{S}} \bar{r}_t(S)$, and observe outcome $X_{t,i}$'s, for $i \in S_t$.
7:     For every $i \in S_t$, update $T_{t,i} = T_{t-1,i} + 1$, $\hat{\mu}_{t,i} = \hat{\mu}_{t-1,i} + (X_{t,i} - \hat{\mu}_{t-1,i})/T_{t,i}$.
8: **end for**

in the existing regret bounds. In particular, we focus on the a non-triggering CMAB problem instance $([m], \mathcal{S}, \mathcal{D}, R)$. Its setting is similar to CMAB-T, but here we assume that $\mathcal{S}$ are collections of subsets of $[m]$ and only arms pulled by action $S_t \in \mathcal{S}$ are revealed as feedback (i.e., $\tau_t = S_t$).

Condition 4 (VM Bounded Smoothness). *We say that a non-triggering CMAB problem instance $([m], \mathcal{S}, \mathcal{D}, R)$ satisfies the Variance Modulated (VM) $(B_v, B_1 \geq 0)$-bounded smoothness condition, if for any action $S \in \mathcal{S}$, any distribution $D, D' \in \mathcal{D}$ with mean vector $\boldsymbol{\mu}, \boldsymbol{\mu}' \in (0,1)^m$, for any $\boldsymbol{\zeta}, \boldsymbol{\eta} \in [-1,1]^m$ s.t. $\boldsymbol{\mu}' = \boldsymbol{\mu} + \boldsymbol{\zeta} + \boldsymbol{\eta}$, we have $|r(S;\boldsymbol{\mu}') - r(S;\boldsymbol{\mu})| \leq B_v\sqrt{\sum_{i\in S}\frac{\zeta_i^2}{(1-\mu_i)\mu_i}} + B_1\sum_{i\in[m]}|\eta_i|$.*

Condition 5 (Independent base arms). *We say that the base arms are independent, if for any $D \in \mathcal{D}$, the outcome vectors $\boldsymbol{X} \sim D$ are independent (across base arms), i.e., $D = \otimes_{i\in[m]} D_i$.*

Condition 6 ($C_1\mu_i(1-\mu_i)$ sub-Gaussian). *The outcome distribution D_i with mean μ_i is $C_1\mu_i(1-\mu_i)$ sub-Gaussian, where C_1 is a known coefficient.*

Remark 6 (Comparison with TPVM Condition and [23]). Condition 4 is the non-triggering version of TPVM, by setting $p_i^{D,S} = 1$ if $i \in S$ and 0 otherwise. As shown in Appendix B.2, Condition 4 can be implied by the original Gini-smoothness condition [23] with $(B_v, B_1) = (3\sqrt{2}\gamma_g, \gamma_\infty)$, so PMC application satisfies the VM condition (the fifth row in Table 3). But different from [23, Lemma 6] and TPVM$_<$, the VM condition is the undirectional version (i.e., we allow $\boldsymbol{\zeta}, \boldsymbol{\eta}$ to be negative). This is important for using empirical means in the algorithm (as we did in our SESCB policy), since they are not necessarily larger than the true means.

Remark 7 (Motivation and Feasibility for Condition 6). Condition 6 helps to cancel out the $(1-\mu_i)\mu_i$ effect in the VM condition without explicitly using the empirical variance that will bring in additional batch-size dependent errors. For Bernoulli arms with mean μ_i, we can compute the explicit value of C_1, i.e., $C_1 = \max_{i\in[m]}\frac{1-2\mu_i}{2\ln(\frac{1-\mu_i}{\mu_i})(1-\mu_i)(\mu_i)}$ by [22]. Notice that C_1 could be large when μ_i is approaching 0 or 1, but it is safe to consider μ_i over bounded supports that are not too close to 0 or 1, e.g., when $\mu_i \in [0.01, 0.99]$, $C_1 \approx 10.78$.

SESCB Algorithm. Our proposed algorithm is shown in Algorithm 2. Instead of maintaining one upper confidence bound for each base arm i, we maintain an upper confidence bound for each super arm S, based on the estimated reward of the empirical means and a confidence interval. In line 4, we compute the confidence interval $\rho_t(S)$ by taking the max of two tentative segments within the square root, which corresponds to two different segments of the concentration bound for the sub-exponential random variable [27]. Such a sub-exponential concentrated confidence interval comes from the VM condition by treating $(\zeta_i)_{i\in S}$ as $|S|$ *independent* sub-Gaussian random variables, whose summation produces a more concentrated sub-exponential random variable compared with considering them as $|S|$ possibly dependent variables. It is notable that for the second tentative interval, SESCB uses the min-counter $T^{\min}_{t-1,S}$ instead of all counters $T_{t-1,i}$ in S, which is the key ingredient that removes the $O(\log K)$ factor as to be shown in the analysis. After getting $\rho_t(S)$, the optimistic reward is defined in Line 5 and the learner selects S_t via the (α, β)-approximation oracle $\bar{O}$ and updates the corresponding statistics.

Regret Bound and Analysis. The following theorem summarizes the regret bound for Algorithm 2.

Theorem 2. *For a non-triggering CMAB problem instance $([m], \mathcal{S}, \mathcal{D}, R)$ that satisfies VM bounded smoothness (Condition 4) with coefficient (B_v, B_1), Condition 5 and Condition 6 with coefficient C_1,* *SESCB (Algorithm 2) with an (α, β)-approximation oralce achieves (α, β)-approximate regret that is bounded by $O\left(\sum_{i\in[m]} \frac{B_v^2 \log T}{\Delta_i^{\min}} + \frac{B_v^2 mK}{\Delta_{\min}} + m\Delta_{\max}\right)$.*

Looking at the above regret bound, the leading term totally removes the $O(\log K)$ dependency compared with Theorem 1. Compared with [23], our regret bounds improves theirs by $O(\log^2 K)$.

Proof Ideas. Similar to the proof of Theorem 1, we first identify an error term $e_t(S_t) = 2\rho_t(S_t)$ as Line 4 and consider the regret filtered by the event $\{\Delta_{S_t} \le e_t(S_t)\}$. The key ingredient is by following Condition 4 and Condition 6, and bound $|r(S;\hat{\boldsymbol{\mu}}) - r(S;\boldsymbol{\mu})| \le B_v\sqrt{\sum_{i\in S} u_{t,i}^2}$, where $u_{t,i}$ is a $(\frac{C_1}{T_{t-1,i}})$-sub-Gaussian random variable. Let $Y_{t,S} = \sum_{i\in S} u_{t,i}^2$. One can show $Y_{t,S}$ is a $(32C_1^2 \sum_{i\in S}\frac{1}{T_{t-1,i}^2}, 4C_1\frac{1}{T_{t-1,S}^{\min}})$-sub-Exponential random variable, so applying the concentration bounds on $Y_{t,S}$ [27] and one can obtain the above $e_t(S_t)$. Then we consider two cases based on the value of $\sum_{i\in S_t}\frac{1}{T_{t-1,i}}$. For both cases, we use the reverse amortization trick from [29] but different from Section 3.2, $e_t(S_t)$ ensures that we only need to consider regret contributions from the min-arm (which is least played in S_t) according to certain batch-size independent thresholds. This in turn gives batch-size independent regret bounds that totally removes $O(\log K)$ in the leading term. See Appendix D for more details. ■

Computational Efficiency. Notice that like other ESCB-type algorithms [9], for the general reward function $r(S;\boldsymbol{\mu})$, there may not exist efficient $\bar{O}$, so one needs to enumerate over all possible actions $S \in \mathcal{S}$ each round, where the time complexity could be as high as $O(|\mathcal{S}|T)$. However, when $r(S;\boldsymbol{\mu})$ is a monotone submodular function (e.g, the reward function of the PMC problem [8]), we can modify $\rho_t(S)$ so that the optimistic reward $\bar{r}_t(S)$ is also monotone submodular, which can be efficiently optimized with a greedy $(1-1/e, 1)$-approximation oracle. Observe that the current $\rho_t(S)$ is not submodular since the maximum of two submodular functions are not necessarily submodular, but we know the summation of two submodular functions are submodular. Based on this observation, we change $\rho_t(S)$ to $\rho_t'(S) = B_v\sqrt{\sum_{i\in S}\frac{C_1}{T_{t-1,i}} + 8C_1\sqrt{\sum_{i\in S}\frac{\log(2|\mathcal{S}|T)}{T_{t-1,i}^2}} + \frac{8C_1\log(2|\mathcal{S}|T)}{T_{t-1,S}^{\min}}}$, where $\max$ is replaced with a sum $(+)$, and we prove in Appendix D.3 that $\rho_t'(S)$ is a monotone submodular function. Now we can use the greedy oracle to maximize a new optimistic reward $r_t'(S) = r(S;\hat{\boldsymbol{\mu}}_{t-1}) + \rho_t'(S)$ in our SESCB algorithm. As for the final regret, using $\rho_t'(S)$ instead of $\rho_t(S)$ only worsens the final regret by a constant factor of two.

Now compared with [23] that achieves $(1-1/e, 1)$-approximate regret bound for PMC problem, our SESCB achieves the same $(1-1/e, 1)$-approximate regret bound but completely removes the $O(\log K)$ dependency. Moreover, our greedy oracle is efficient with computational complexity $O(TKL)$, where T is the total number of rounds, K is the number of source nodes to be selected in each round and L is the total number of source nodes, which is much faster than the enumeration method. For the regret analysis when using $r_t'(S)$, see Appendix D.3 for more details.

5 Applications

In this section, we show how various applications satisfy our new TPVM, TPVM$_<$ or VM smoothness condition and their corresponding (B_v, B_1, λ) coefficients with non-trivial B_v, i.e., $B_v = o(B_1\sqrt{K})$, which in turn improves the regret bounds over the batch-size dependence of K.

Theorem 3. *The combinatorial cascading bandits [17], the multi-layered network exploration [21], the influence maximization problems [29] and the probabilistic maximum coverage problem [23] satisfy the TPVM (TPVM$_<$ or VM) conditions with coefficients (B_v, B_1, λ), resulting regret bounds and improvements shown in Table 3.*

Note that the first four applications in Table 3 applies Theorem 1, while the last application applies Theorem 2. More specifically, the first two applications we consider are disjunctive and conjunctive cascading bandits [17], where m base arms represent web pages and routing edges in online advertising and network routing, respectively. Batch-size K is the maximum size of the ordered sequence $S \in \mathcal{S}$ to be selected in each round, which will trigger web pages/routing edge one by one until certain stopping condition is satisfied, i.e., a click or a routing edge being broken. The reward is 1

Table 3: Summary of the coefficients, regret bounds and improvements for various applications.

Application	Condition	(B_v, B_1, λ)	Regret	Improvement
Disjunctive Cascading Bandits [17]	TPVM$_<$	$(1, 1, 2)$	$O(\sum_{i\in[m]} \frac{\log K \log T}{\Delta_i^{\min}})$	$O(K/\log K)$
Conjunctive Cascading Bandits [17]	TPVM	$(1, 1, 1)$	$O(\sum_{i\in[m]} \log \frac{K}{\Delta_{\min}} \frac{\log K \log T}{\Delta_{i,\min}})$	$O(K/(\log K \log \frac{K}{\Delta_{\min}}))$
Multi-layered Network Exploration [21]	TPVM	$(\sqrt{1.25\lvert V\rvert}, 1, 2)$ †	$O(\sum_{i\in\mathcal{A}} \frac{\lvert V\rvert \log(n\lvert V\rvert)\log T}{\Delta_i^{\min}})$	$O(n/\log(n\lvert V\rvert))$
Influence Maximization on DAG [29]	TPVM$_<$	$(\sqrt{L}\lvert V\rvert, \lvert V\rvert, 1)$ †	$O(\sum_{i\in[m]} \log \frac{\lvert E\rvert}{\Delta_{\min}} \frac{L\lvert V\rvert^2 \log\lvert E\rvert \log T}{\Delta_i^{\min}})$	$O(\lvert E\rvert/(L\log\lvert E\rvert \log\frac{\lvert E\rvert}{\Delta_{\min}}))$
Probabilistic Maximum Coverage [23]*	VM	$(3\sqrt{2\lvert V\rvert}, 1, -)$	$O(\sum_{i\in[m]} \frac{\lvert V\rvert \log T}{\Delta_i^{\min}})$	$O(\log^2 k)$.

* This row is for the application in Section 4 and the rest of rows are for Section 3.1; † $|V|, |E|, n, k, L$ denotes the number of target nodes, the number of edges that can be triggered by the set of seed nodes, the number of layers, the number of seed nodes and the length of the longest directed path, respectively.

if *any* web page is clicked (or if *all* routing edges are live) and 0 otherwise. Compared with [29], we achieve an improvement $O(K/(\log K \log \frac{K}{\Delta_{\min}}))$ for the conjunctive case and an improvement $O(K/\log K)$ for the disjunctive case, due to the same $B_v, B_1 = 1$ but different orders λ.

The third application is the mutli-layered network exploration (MuLaNE) problem [21], and the MuLaNE task is to allocate B budgets into n layers to explore target nodes V. In MuLaNE, the base arms form a set $\mathcal{A} = \{(i, u, b) : i \in [n], u \in [V], b \in [B]\}$, the batch-size $K = (n+1)|V|$ and the reward is defined as the total reward give by the first visit of any target nodes. MuLaNE fits into our study, and compared with [21], the regret bound is improved by a factor of $O(n/\log K)$.

Our fourth application is the online influence maximization (OIM) problems direct acyclic graphs (DAG). For this application, the goal is to select at most k seed nodes to influence as many target nodes V as possible, where the influence process follows the independent cascade (IC) model [29] (see Appendix E for more details). The base arms are the edges with unknown edge probabilities and the batch-size K is the total number of edges that could be triggered by any set of k seed nodes, denoted as $|E|$. The improvements here are significant, improving the existing results [23] by a factor of $O(|E|/(L \log |E| \log \frac{|E|}{\Delta_{\min}}))$.

For the PMC problem [23], we consider a complete bipartite graph with L source nodes on the left and $|V|$ target nodes V on the right. The goal is to select k seed nodes from L nodes trying to influence as many as target nodes, so the edges E are independent base arms and the batch-size is $k|V|$. By using the computational efficient version of Algorithm 2 and applying Theorem 2, we achieve $O(\log^2 k)$ improvement compared with [23] while maintaining good computational efficiency.

Proof Ideas. For all above applications (except for the OIM on DAG), our proof involves the use of telescoping series to decompose the reward difference, together with a smart use of the Cauchy–Schwarz inequality aided by the variance terms. For disjunctive cascading bandits, for example, the reward difference $|r(S; \bar{\boldsymbol{\mu}}) - r(S; \boldsymbol{\mu})| = \prod_{i=1}^K (1-\mu_i) - \prod_{i=1}^K (1-\bar{\mu}_i)$ can be telescoped as $\sum_{i\in[K]} (\bar{\mu}_i - \mu_i) \left(\prod_{j=1}^{i-1} (1-\mu_j) \cdot \prod_{j=i+1}^K (1-\bar{\mu}_j) \right)$. After this decomposition, we replace certain terms with $p_i^{D,S}$ and bound above by $\sum_{i\in[K]} \zeta_i p_i^{D,S} \sqrt{\prod_{j=i+1}^K (1-\mu_j)} + \sum_{i\in[K]} \eta_i p_i^{D,S}$. Then we simultaneously multiply and divide the variance term $\sqrt{(1-\mu_i)\mu_i}$ on the first term and apply the Cauchy–Schwarz inequality to move the summation over K into the square root, concluding the satisfaction of Condition 3 with $B_v = \sqrt{\sum_{i\in[K]} (1-\mu_i)\mu_i \prod_{j=i+1}^K (1-\mu_j)} \le 1$. As for the OIM on DAG, since reward function have no closed-form solutions [5], the analysis is more involved with the need of advanced techniques such as the coupling technique [20], see Appendix E for details. ■

6 Conclusion and Future Direction

This paper studies the CMAB problem with probabilistically triggered arms or independent arms. We discover new TPVM and VM conditions, and propose BCUCB-T and SESCB algorithms to reduce and remove the batch-size K in the regret bounds, respectively. We also show that several important applications all satisfy our conditions to achieve improved regrets, both theoretically and empirically. There are many compelling directions for future study. For example, it would be interesting to study the setting of CMAB-T together with independent arms. One could also explore how to extend our application and consider general graphs in online influence maximization bandits.

Acknowledgments and Disclosure of Funding

The work of John C.S. Lui was supported in part by the HK RGC SRF2122-4202.

References

[1] Alper Atamtürk and Andrés Gómez. Maximizing a class of utility functions over the vertices of a polytope. *Operations Research*, 65(2):433–445, 2017.

[2] Jean-Yves Audibert, Rémi Munos, and Csaba Szepesvári. Exploration–exploitation tradeoff using variance estimates in multi-armed bandits. *Theoretical Computer Science*, 410(19): 1876–1902, 2009.

[3] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.

[4] Sébastien Bubeck, Nicolo Cesa-Bianchi, et al. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1):1–122, 2012.

[5] Wei Chen, Yajun Wang, and Siyu Yang. Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 199–208, 2009.

[6] Wei Chen, Laks VS Lakshmanan, and Carlos Castillo. Information and influence propagation in social networks. *Synthesis Lectures on Data Management*, 5(4):1–177, 2013.

[7] Wei Chen, Yajun Wang, and Yang Yuan. Combinatorial multi-armed bandit: General framework and applications. In *International Conference on Machine Learning*, pages 151–159. PMLR, 2013.

[8] Wei Chen, Yajun Wang, Yang Yuan, and Qinshi Wang. Combinatorial multi-armed bandit and its extension to probabilistically triggered arms. *The Journal of Machine Learning Research*, 17 (1):1746–1778, 2016.

[9] Richard Combes, Mohammad Sadegh Talebi Mazraeh Shahi, Alexandre Proutiere, et al. Combinatorial bandits revisited. *Advances in neural information processing systems*, 28, 2015.

[10] Thibaut Cuvelier, Richard Combes, and Eric Gourdin. Statistically efficient, polynomial-time algorithms for combinatorial semi-bandits. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 5(1):1–31, 2021.

[11] Rémy Degenne and Vianney Perchet. Combinatorial semi-bandit with known covariance. In *Advances in Neural Information Processing Systems*, pages 2972–2980, 2016.

[12] Devdatt P Dubhashi and Alessandro Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.

[13] Yi Gai, Bhaskar Krishnamachari, and Rahul Jain. Combinatorial network optimization with unknown variables: Multi-armed bandits with linear rewards and individual observations. *IEEE/ACM Transactions on Networking (TON)*, 20(5):1466–1478, 2012.

[14] Jean Honorio and Tommi Jaakkola. Tight bounds for the expected risk of linear classifiers and pac-bayes finite-sample guarantees. In *Artificial Intelligence and Statistics*, pages 384–392. PMLR, 2014.

[15] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146, 2003.

[16] Branislav Kveton, Csaba Szepesvari, Zheng Wen, and Azin Ashkan. Cascading bandits: Learning to rank in the cascade model. In *International Conference on Machine Learning*, pages 767–776. PMLR, 2015.

[17] Branislav Kveton, Zheng Wen, Azin Ashkan, and Csaba Szepesvári. Combinatorial cascading bandits. In *Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 1*, pages 1450–1458, 2015.

[18] Branislav Kveton, Zheng Wen, Azin Ashkan, and Csaba Szepesvari. Tight regret bounds for stochastic combinatorial semi-bandits. In *AISTATS*, 2015.

[19] Shuai Li, Baoxiang Wang, Shengyu Zhang, and Wei Chen. Contextual combinatorial cascading bandits. In *International conference on machine learning*, pages 1245–1253. PMLR, 2016.

[20] Shuai Li, Fang Kong, Kejie Tang, Qizhi Li, and Wei Chen. Online influence maximization under linear threshold model. *Advances in Neural Information Processing Systems*, 33:1192–1204, 2020.

[21] Xutong Liu, Jinhang Zuo, Xiaowei Chen, Wei Chen, and John CS Lui. Multi-layered network exploration via random walks: From offline optimization to online learning. In *International Conference on Machine Learning*, pages 7057–7066. PMLR, 2021.

[22] Olivier Marchal and Julyan Arbel. On the sub-gaussianity of the beta and dirichlet distributions. *Electronic Communications in Probability*, 22:1–14, 2017.

[23] Nadav Merlis and Shie Mannor. Batch-size independent regret bounds for the combinatorial multi-armed bandit problem. In *Conference on Learning Theory*, pages 2465–2489. PMLR, 2019.

[24] Nadav Merlis and Shie Mannor. Tight lower bounds for combinatorial multi-armed bandits. In *Conference on Learning Theory*, pages 2830–2857. PMLR, 2020.

[25] Pierre Perrault, Jennifer Healey, Zheng Wen, and Michal Valko. Budgeted online influence maximization. In *International Conference on Machine Learning*, pages 7620–7631. PMLR, 2020.

[26] Herbert Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535, 1952.

[27] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[28] Daniel Vial, Sujay Sanghavi, Sanjay Shakkottai, and R Srikant. Minimax regret for cascading bandits. *arXiv preprint arXiv:2203.12577*, 2022.

[29] Qinshi Wang and Wei Chen. Improving regret bounds for combinatorial semi-bandits with probabilistically triggered arms and its applications. In *Advances in Neural Information Processing Systems*, pages 1161–1171, 2017.

[30] Siwei Wang and Wei Chen. Thompson sampling for combinatorial semi-bandits. In *International Conference on Machine Learning*, pages 5114–5122, 2018.

[31] Zheng Wen, Branislav Kveton, Michal Valko, and Sharan Vaswani. Online influence maximization under independent cascade model with semi-bandit feedback. *Advances in neural information processing systems*, 30, 2017.

[32] Zihan Zhang, Jiaqi Yang, Xiangyang Ji, and Simon S Du. Improved variance-aware confidence sets for linear bandits and linear mixture mdp. *Advances in Neural Information Processing Systems*, 34:4342–4355, 2021.

[33] Dongruo Zhou, Quanquan Gu, and Csaba Szepesvari. Nearly minimax optimal reinforcement learning for linear mixture markov decision processes. In *Conference on Learning Theory*, pages 4532–4576. PMLR, 2021.

Appendix

The Appendix is organized as follows. We first compare the TPVM$_<$ Condition (Condition 3) with TPM Condition (Condition 2) in Appendix A. The comparisons between Gini-smoothness Condition with the TPVM (Condition 3) and VM (Condition 4) Condition are introduced in Appendix B. The detailed proofs for Theorem 1 together with some discussions are in Appendix C. The detailed proofs for Theorem 2 are in Appendix D. The application details and the proof details related to Theorem 3 are in Appendix E. The experiments for different applications are included in Appendix F.

A Comparing the TPVM$_<$ Condition (Condition 3) with the TPM Condition (Condition 2)

As discussed in Section 3.1, the TPVM$_<$ condition (Condition 3) does not imply the TPM condition (Condition 2) in general. In this section, we show that under some additional conditions, Condition 3 does imply Condition 2.

Lemma 1. *Assume that the followings are true: (a) Condition 1 holds; (b) there exist $D_\vee, D_\wedge \in \mathcal{D}$ with mean vectors $\boldsymbol{\mu}_\vee = \boldsymbol{\mu} \vee \boldsymbol{\mu}'$ and $\boldsymbol{\mu}_\wedge = \boldsymbol{\mu} \wedge \boldsymbol{\mu}'$ respectively, where operation $\vee$ and $\wedge$ means taking coordinate wise $\max$ and $\min$; and (c) $p_i^{D,S}$ increases as the mean vector of D increases. Then Condition 3 implies Condition 2 with the same B_1 coefficient.*

Proof. First, when setting $\boldsymbol{\zeta} = \mathbf{0}$ in Condition 3, we obtain the directional TPM condition. Then we prove with the following derivation that with the three assumptions stated in the lemma, directional TPM condition implies the undirectional TPM condition (Condition 2). For any $D, D' \in \mathcal{D}$ with mean vectors $\boldsymbol{\mu}, \boldsymbol{\mu}'$, without loss of generality, we assume that $r(S; \boldsymbol{\mu}') \geq r(S; \boldsymbol{\mu})$. we have

$$\begin{aligned}
&|r(S;\boldsymbol{\mu}') - r(S;\boldsymbol{\mu})| \\
&= r(S;\boldsymbol{\mu}') - r(S;\boldsymbol{\mu}) \\
&\leq r(S;\boldsymbol{\mu}_\vee) - r(S;\boldsymbol{\mu}_\wedge) && \text{by Assumptions (a) and (b)} \\
&\leq B_1 \sum_{i\in[m]} p_i^{D_\wedge,S}|\mu_i - \mu_i'| && \text{by Assumption (b) and the directional TPM condition} \\
&\leq B_1 \sum_{i\in[m]} p_i^{D,S}|\mu_i - \mu_i'|. && \text{By Assumption (c)}
\end{aligned}$$

Therefore, the undirectional TPM condition (Condition 2) holds. ■

It is not difficult to verify that for the online influence maximization application discussed in Section 5, all three assumptions in the lemma holds.

B Comparing the Gini-smoothness Condition [23] with the TPVM$_<$ Condition (Condition 3) and VM Condition (Condition 4)

Merlis and Mannor [23] define the following Gini-smoothness condition. In this section, we provide comparisons between this condition and our TPVM$_<$ condition (Condition 3) and VM condition (Condition 4).

Condition 7 (Gini-smoothness Condition, Restated, [23])**.** *Let $f(S;\boldsymbol{x}) : \mathcal{S} \times [0,1]^m \to \mathbb{R}$ be a differentiable function in $x \in (0,1)^m$ and continuous in $x \in [0,1]^m$, for any $S \in \mathcal{S}$. The function $f(S;x)$ is said to be monotonic Gini-smooth, with smoothness parameters (γ_g, γ_∞) if:*

1. For any $S \in \mathcal{S}$, the function $f(S;\boldsymbol{x})$ is monotonically increasing with bounded gradient, i.e., for any $i \in S$ and $x \in (0,1)^m$, $0 \leq \frac{\partial f(S;\boldsymbol{x})}{\partial x_i} \leq \gamma_\infty$. If $i \notin S$, then $\frac{\partial f(S;x)}{\partial x_i} = 0$ for all $x \in (0,1)^L$.

2. For any $S \in \mathcal{S}$ and $x \in (0,1)^m$, it holds that

$$\sqrt{\sum_{i\in S} x_i(1-x_i)(\frac{\partial f(S;x)}{\partial x_i})} \leq \gamma_g. \tag{6}$$

B.1 Triggering Probability Modulated Gini-smoothness Condition Does Not Imply TPVM$_<$ Condition

The original Gini-smoothness condition (Condition 7) does not work directly with probabilistically triggered arms. Thus, our first attempt is to add triggering probability modulation to the Gini-smoothness condition as given below, in hope that it would extend the result in [23] to the CMAB-T framework.

Condition 8 (TPM Gini-smoothness)**.** *For a CMAB-T problem instance $([m], \mathcal{S}, \mathcal{D}, D_{trig}, R)$, assume the reward function $r(S;\boldsymbol{\mu}) : \mathcal{S} \times [0,1]^m \to \mathbb{R}$ is a differentiable function in $\boldsymbol{\mu} \in (0,1)^m$ and continuous in $\boldsymbol{\mu} \in [0,1]^m$, for any $S \in \mathcal{S}$. The reward function $r(S;\boldsymbol{\mu})$ is said to be monotonic Triggering Probability Modulated (TPM) Gini-smooth, with smoothness parameters $(\gamma_g, \gamma_\infty, \lambda \geq 1)$ if:*

1. For any distribution $D \in \mathcal{D}$ with mean vector $\boldsymbol{\mu} \in (0,1)^m$ and any action $S \in \mathcal{S}$, the function $r(S;\boldsymbol{\mu})$ is monotonically increasing with bounded gradient: For any $i \in [m]$, if $p_i^{D,S} > 0$, then $0 \leq \frac{\partial r(S;\boldsymbol{\mu})}{\partial \mu_i}\frac{1}{p_i^{D,S}} \leq \gamma_\infty$; If $p_i^{D,S} = 0$, then $\frac{\partial r(S;\boldsymbol{\mu})}{\partial \mu_i} = 0$ for all $\boldsymbol{\mu} \in (0,1)^m$.

2. For any distribution $D \in \mathcal{D}$ with mean vector $\boldsymbol{\mu} \in (0,1)^m$ and any action $S \in \mathcal{S}$, it holds that

$$\sqrt{\sum_{i \in \tilde{S}} \mu_i(1-\mu_i)(\frac{\partial r(S;\boldsymbol{\mu})}{\partial \mu_i})^2 \frac{1}{(p_i^{D,S})^\lambda}} \leq \gamma_g. \tag{7}$$

However, using the above TPM Gini-smoothness condition, we cannot derive a desirable regret bound. In particular, the above condition only guarantees the following lemma (following the analysis of Lemma 6 in [23]), which is weaker than our TPVM$_<$ condition (Condition 3), leading to a weaker regret with an additional factor $\max_{S\in\mathcal{S}, i\in[m]} p_i^{\max,S}/p_i^{D,S}$. Such a factor could be exponentially large and undesirable in applications, similar to the factor being avoided in [8] by introducing the TPM condition.

Lemma 2. *Let $r(S;\boldsymbol{x})$ be a monotonic $(\gamma_g, \gamma_\infty, \lambda)$ TPM gini-smooth function. For any $\boldsymbol{\mu}, \boldsymbol{\mu}', \boldsymbol{\mu}'' \in [0,1]^m$, with $\boldsymbol{\zeta} = \boldsymbol{\mu}' - \boldsymbol{\mu}, \boldsymbol{\eta} = \boldsymbol{\mu}'' - \boldsymbol{\mu}'$, let $\mathcal{D}_{\boldsymbol{\mu},\boldsymbol{\mu}',\boldsymbol{\mu}''} = \{D \in \mathcal{D}$ with mean vector $\boldsymbol{x} : \forall i \in [m], \min\{\mu_i, \mu_i', \mu_i''\} \leq x_i \leq \max\{\mu_i, \mu_i', \mu_i''\}\}$ and $p_i^{\max,S} = \max_{D\in\mathcal{D}_{\boldsymbol{\mu},\boldsymbol{\mu}',\boldsymbol{\mu}''}} p_i^{D,S}$, it holds that*

$$|r(S;\boldsymbol{\mu}'') - r(S;\boldsymbol{\mu})| \leq 3\sqrt{2}\gamma_g\sqrt{\sum_{i\in S}\left(\frac{|\zeta_i|}{\sqrt{(1-\mu_i)\mu_i}}\right)^2 p_i^{\max,S}} + \sum_{i\in\tilde{S}} p_i^{\max,S}|\eta_i|, \tag{8}$$

Proof. **For the $|r(S;\boldsymbol{\mu}') - r(S;\boldsymbol{\mu})|$ term,**

First, we define two functions g, h, where

$$g(z) = \int_0^z \frac{dy}{\sqrt{y(1-y)}}, h(z) = \int_0^z \frac{dy}{\sqrt{y} \wedge \sqrt{1-y}}. \tag{9}$$

For $g(z)$, $g'(z) > 0$ so the inverse function g^{-1} is well defined. We also know that $h(z)$ has the following closed form,

$$h(z) = \begin{cases} 2\sqrt{z}, & \text{if } z \leq 1/2 \\ 2\sqrt{2} - 2\sqrt{1-z}, & \text{if } z \geq 1/2. \end{cases} \tag{10}$$

Note that these two functions are closely related: $h'(z) \leq g'(z) \leq \sqrt{2}h'(z)$ with $g(0) = h(0)$. Therefore, $h(z) \leq g(z) \leq \sqrt{2}h(z)$ and $g(z_2) - g(z_1) \leq \sqrt{2}(h(z_2) - h(z_1))$ for any $z_1 \leq z_2 \in [0,1]$. Now we set up a parameterization $z(t)$ for $t \in [0,1]$ such that $z_i(0) = \mu_i, z_i(1) = \mu_i'$. Specifically, we choose the parameterization to be

$$z_i(t) = g^{-1}([g(\mu_i') - g(\mu_i)]t + g(\mu_i)]), \tag{11}$$

Then its gradient is

$$z_i'(t) = \frac{g(\mu_i') - g(\mu_i)}{g'(z_i(t))} = (g(\mu_i') - g(\mu_i))\sqrt{z_i(t)(1-z_i(t))}. \tag{12}$$

Then we can use the gradient theorem to bound $|r(S;\boldsymbol{\mu}')-r(S;\boldsymbol{\mu})|$ as

$$
\begin{aligned}
|r(S;\boldsymbol{\mu}')-r(S;\boldsymbol{\mu})| &= \left|\int_{\boldsymbol{x}=\boldsymbol{\mu}}^{\boldsymbol{\mu}'} \nabla r(S;\boldsymbol{x})\cdot d\boldsymbol{x}\right| = \left|\int_{t=0}^{1}\sum_{i\in\tilde{S}}\frac{\partial r(S;z(t))}{\partial x_i}z_i'(t)dt\right| && (13)\\
&\le \int_0^1 \sqrt{\sum_{i\in\tilde{S}}(g(\mu_i')-g(\mu_i))^2(p_i^{z(t),S})^{\lambda}}\sqrt{\sum_{i\in\tilde{S}}(\frac{\partial r(S;z(t))}{\partial x_i})^2\frac{z_i(t)(1-z_i(t))}{(p_i^{z(t),S})^{\lambda}}}dt && (14)\\
&\le \int_0^1 \gamma_g\sqrt{\sum_{i\in\tilde{S}}(g(\mu_i')-g(\mu_i))^2(p_i^{z(t),S})^{\lambda}}dt && (15)\\
&\le \gamma_g\sqrt{\sum_{i\in\tilde{S}}(g(\mu_i')-g(\mu_i))^2(p_i^{\max,S})^{\lambda}}, && (16)
\end{aligned}
$$

Following the similar derivation for Eq. (28) in next subsection, we have

$$
|r(S;\boldsymbol{\mu}')-r(S;\boldsymbol{\mu})| \le 3\sqrt{2}\gamma_g\sqrt{\sum_{i\in\tilde{S}}\left(\frac{|\mu_i'-\mu_i|}{\sqrt{(1-\mu_i)\mu_i}}\right)^2(p_i^{\max,S})^{\lambda}}. \tag{17}
$$

For the $|r(S;\boldsymbol{\mu}'')-r(S;\boldsymbol{\mu}')|$ term,

We can use the gradient theorem to bound, let $\boldsymbol{x}(t)$ with $x_i(t)=t(\mu_i''-\mu_i')+\mu_i'$.

$$
\begin{aligned}
|r(S;\boldsymbol{\mu}'')-r(S;\boldsymbol{\mu}')| &= \left|\int_{t=0}^{1}\sum_{i\in\tilde{S}}\frac{\partial r(S;\boldsymbol{x}(t))}{\partial x_i}x_i'(t)dt\right|\\
&\le \left|\int_{t=0}^{1}\sum_{i\in\tilde{S}}\frac{\partial r(S;\boldsymbol{x}(t))}{\partial x_i p_i^{\boldsymbol{x}(t),S}}p_i^{\boldsymbol{x}(t),S}(\mu_i''-\mu_i')dt\right|\\
&\le \int_{t=0}^{1}\sum_{i\in\tilde{S}}\left|\frac{\partial r(S;\boldsymbol{x}(t))}{\partial x_i p_i^{\boldsymbol{x}(t),S}}\right|\left|p_i^{\boldsymbol{x}(t),S}(\mu_i''-\mu_i')\right|dt\\
&\le \int_{t=0}^{1}\sum_{i\in\tilde{S}}\gamma_\infty\left|p_i^{\boldsymbol{x}(t),S}(\mu_i''-\mu_i')\right|dt\\
&\le \int_{t=0}^{1}\sum_{i\in\tilde{S}}\gamma_\infty\left|p_i^{\max,S}(\mu_i''-\mu_i')\right|dt\\
&= \gamma_\infty\sum_{i\in\tilde{S}}p_i^{\max,S}|\eta_i|. && (18)
\end{aligned}
$$

Combining Eq. (17) and Eq. (18), we conclude the lemma. ■

The above lemma indicates that directly extending the Gini-smoothness condition may not be strong enough for the probabilistic triggering setting. This motivates us to define the new TPVM$_<$ condition not based on the differential form, but directly on the difference form $|r(S;\boldsymbol{\mu}')-r(S;\boldsymbol{\mu})|$. This can be viewed as incorporating triggering probability properly into the result of Lemma 6 in [23].

B.2 Gini-smoothness Condition Implies VM Condition

In this section, we show in the following lemma that the original Gini-smoothness condition (Condition 7) implies the VM condition (Condition 4), with $(B_v,B_1)=(3\sqrt{2}\gamma_g,\gamma_\infty)$. The proof of this lemma is similar to [23, Lemma 6], but we need to extend it to the undirectional case, where $\boldsymbol{\mu}''$ is not necessarily larger than $\boldsymbol{\mu}$ in all dimensions.

Lemma 3. *Let $r(S;\boldsymbol{\mu})$ be a monotonic (γ_g,γ_∞) gini-smooth function as given in Condition 7. For any $\boldsymbol{\mu},\boldsymbol{\mu}',\boldsymbol{\mu}''\in[0,1]^m$, with $\boldsymbol{\zeta}=\boldsymbol{\mu}'-\boldsymbol{\mu},\boldsymbol{\eta}=\boldsymbol{\mu}''-\boldsymbol{\mu}'$, it holds that*

$$|r(S;\boldsymbol{\mu}'')-r(S;\boldsymbol{\mu})|\le 3\sqrt{2}\gamma_g\sqrt{\sum_{i\in S}\left(\frac{|\zeta_i|}{\sqrt{(1-\mu_i)\mu_i}}\right)^2}+\gamma_\infty\sum_{i\in S}|\eta_i|\,. \tag{19}$$

Proof. We use $|r(S;\boldsymbol{\mu}'')-r(S;\boldsymbol{\mu})|\le|r(S;\boldsymbol{\mu}')-r(S;\boldsymbol{\mu})|+|r(S;\boldsymbol{\mu}'')-r(S;\boldsymbol{\mu}')|$ and separately bound two terms in the LHS.

For the $|r(S;\boldsymbol{\mu}')-r(S;\boldsymbol{\mu})|$ term,

We define two functions g,h, where

$$g(z)=\int_0^z\frac{dy}{\sqrt{y(1-y)}},h(z)=\int_0^z\frac{dy}{\sqrt{y}\wedge\sqrt{1-y}}. \tag{20}$$

For $g(z)$, $g'(z)>0$ so the inverse function g^{-1} is well defined. We also know that $h(z)$ has the following closed form,

$$h(z)=\begin{cases}2\sqrt{z}, & \text{if } z\le 1/2\\ 2\sqrt{2}-2\sqrt{1-z}, & \text{if } z\ge 1/2.\end{cases} \tag{21}$$

Note that these two functions are closely related: $h'(z)\le g'(z)\le\sqrt{2}h'(z)$ with $g(0)=h(0)$. Therefore, $h(z)\le g(z)\le\sqrt{2}h(z)$ and $g(z_2)-g(z_1)\le\sqrt{2}(h(z_2)-h(z_1))$ for any $z_1\le z_2\in[0,1]$. Now we set up a parameterization $z(t)$ for $t\in[0,1]$ such that $z_i(0)=\mu_i,z_i(1)=\mu_i'$. Specifically, we choose the parameterization to be

$$z_i(t)=g^{-1}([g(\mu_i')-g(\mu_i)]t+g(\mu_i)]), \tag{22}$$

Then its gradient is

$$z_i'(t)=\frac{g(\mu_i')-g(\mu_i)}{g'(z_i(t))}=(g(\mu_i')-g(\mu_i))\sqrt{z_i(t)(1-z_i(t))}. \tag{23}$$

Then we can use the gradient theorem to bound $r(S;\boldsymbol{\mu}')-r(S;\boldsymbol{\mu})$ as

$$\begin{aligned}|r(S;\boldsymbol{\mu}')-r(S;\boldsymbol{\mu})|&=\left|\int_{\boldsymbol{x}=\boldsymbol{\mu}}^{\boldsymbol{\mu}'}\nabla r(S;\boldsymbol{x})\cdot d\boldsymbol{x}\right|=\left|\int_{t=0}^1\sum_{i\in S}\frac{\partial r(S;z(t))}{\partial x_i}z_i'(t)dt\right| &(24)\\ &\le\int_{t=0}^1\sum_{i\in S}|g(\mu_i')-g(\mu_i)|\left|\frac{\partial r(S;z(t))}{\partial x_i}\right|\sqrt{z_i(t)(1-z_i(t))}dt &(25)\\ &\le\int_0^1\sqrt{\sum_{i\in S}(g(\mu_i')-g(\mu_i))^2}\sqrt{\sum_{i\in S}(\frac{\partial r(S;z(t))}{\partial x_i})^2 z_i(t)(1-z_i(t))dt} &(26)\\ &\le\int_0^1\gamma_g\sqrt{\sum_{i\in S}(g(\mu_i')-g(\mu_i))^2dt} &(27)\end{aligned}$$

To calculate the bound, we use the relation between g and h, and calculate the difference over h for the following cases:

Case 1: When $\mu_i\le\mu_i'\le 1/2$, then $|g(\mu_i')-g(\mu_i)|=g(\mu_i')-g(\mu_i)\le\sqrt{2}(h(\mu_i')-h(\mu_i))$.

$$\begin{aligned}h(\mu_i')-h(\mu_i)&=2\sqrt{\mu_i'}-2\sqrt{\mu_i}=2\sqrt{\mu_i}(\sqrt{1+\frac{|\mu_i'-\mu_i|}{\mu_i}}-1)\\ &\le 2\sqrt{\mu_i}\frac{|\mu_i'-\mu_i|}{2\mu_i}\le\frac{|\mu_i'-\mu_i|}{\sqrt{(1-\mu_i)\mu_i}},\end{aligned}$$

where the first inequality uses the fact that $\sqrt{1+x}\le 1+x/2$, for any $x>-1$. So $|g(\mu_i')-g(\mu_i)|\le\sqrt{2}\frac{|\mu_i'-\mu_i|}{\sqrt{(1-\mu_i)\mu_i}}$.

Case 2: When $\mu_i' \le \mu_i \le 1/2$, then $|g(\mu_i') - g(\mu_i)| = g(\mu_i) - g(\mu_i') \le \sqrt{2}(h(\mu_i) - h(\mu_i'))$.

$$h(\mu_i) - h(\mu_i') = 2\sqrt{\mu_i} - 2\sqrt{\mu_i'} = 2\sqrt{\mu_i}(1 - \sqrt{1 - \frac{|\mu_i' - \mu_i|}{\mu_i}})$$
$$\le 2\sqrt{\mu_i}\frac{|\mu_i' - \mu_i|}{\mu_i} \le \frac{2|\mu_i' - \mu_i|}{\sqrt{(1-\mu_i)\mu_i}}.$$

where the first inequality uses the fact that $1 - \sqrt{1-x} \le x$ for $x \in [0,1]$. So $|g(\mu_i') - g(\mu_i)| \le 2\sqrt{2}\frac{|\mu_i' - \mu_i|}{\sqrt{(1-\mu_i)\mu_i}}$.

Case 3: When $1/2 \le \mu_i \le \mu_i'$, then $|g(\mu_i') - g(\mu_i)| = g(\mu_i') - g(\mu_i) \le \sqrt{2}(h(\mu_i') - h(\mu_i))$.

$$h(\mu_i') - h(\mu_i) = 2\sqrt{1-\mu_i} - 2\sqrt{1-\mu_i'} = 2\sqrt{1-\mu_i}(1 - \sqrt{1 - \frac{|\mu_i' - \mu_i|}{1-\mu_i}})$$
$$\le 2\sqrt{1-\mu_i}\frac{|\mu_i' - \mu_i|}{(1-\mu_i)} \le \frac{2|\mu_i' - \mu_i|}{\sqrt{(1-\mu_i)\mu_i}},$$

So $|g(\mu_i') - g(\mu_i)| \le 2\sqrt{2}\frac{|\mu_i' - \mu_i|}{\sqrt{(1-\mu_i)\mu_i}}$.

Case 4: When $1/2 \le \mu_i' \le \mu_i$, then $|g(\mu_i') - g(\mu_i)| = g(\mu_i) - g(\mu_i') \le \sqrt{2}(h(\mu_i) - h(\mu_i'))$.

$$h(\mu_i) - h(\mu_i') = 2\sqrt{1-\mu_i'} - 2\sqrt{1-\mu_i} = 2\sqrt{1-\mu_i}(\sqrt{1 + \frac{|\mu_i' - \mu_i|}{1-\mu_i}} - 1)$$
$$\le \sqrt{1-\mu_i}\frac{|\mu_i' - \mu_i|}{(1-\mu_i)} \le \frac{|\mu_i' - \mu_i|}{\sqrt{(1-\mu_i)\mu_i}},$$

So $|g(\mu_i') - g(\mu_i)| \le \sqrt{2}\frac{|\mu_i' - \mu_i|}{\sqrt{(1-\mu_i)\mu_i}}$.

Case 5: When $1/2 \le \mu_i', \mu_i \le 1/2$, then $|g(\mu_i') - g(\mu_i)| = g(\mu_i') - g(\mu_i) \le \sqrt{2}(h(\mu_i') - h(\mu_i))$.

$$\begin{aligned} h(\mu_i') - h(\mu_i) &= (h(\mu_i') - h(1/2)) + (h(\mu_i) - h(1/2)) \\ &\le 2\frac{\mu_i' - 1/2}{\sqrt{1-1/2}} + \frac{1/2 - \mu_i}{\sqrt{\mu_i}} \\ &\le \frac{\mu_i' - \mu_i}{\sqrt{\mu_i}} + \frac{\mu_i' - \mu_i}{\sqrt{\mu_i}} \\ &\le 3\frac{|\mu_i' - \mu_i|}{\sqrt{\mu_i(1-\mu_i)}}, \end{aligned}$$

where the first inequality uses the results for $1/2 \le \mu_i \le \mu_i'$ and for $\mu_i \le \mu_i' \le 1/2$, the second inequality uses the relation that $\mu_i \le 1/2$ and $\mu_i' \ge 1/2$. So $|g(\mu_i') - g(\mu_i)| \le 3\sqrt{2}\frac{|\mu_i' - \mu_i|}{\sqrt{(1-\mu_i)\mu_i}}$.

Case 6: When $1/2 \le \mu_i, \mu_i' \le 1/2$, then $|g(\mu_i') - g(\mu_i)| = g(\mu_i) - g(\mu_i') \le \sqrt{2}(h(\mu_i) - h(\mu_i'))$.

$$\begin{aligned} h(\mu_i) - h(\mu_i') &= (h(\mu_i) - h(1/2)) + (h(\mu_i') - h(1/2)) \\ &\le \frac{\mu_i - 1/2}{\sqrt{1-\mu_i}} + 2\frac{1/2 - \mu_i'}{\sqrt{1/2}} \\ &\le \frac{\mu_i - \mu_i'}{\sqrt{1-\mu_i}} + 2\frac{\mu_i - \mu_i'}{\sqrt{1-\mu_i}} \\ &\le 3\frac{|\mu_i' - \mu_i|}{\sqrt{\mu_i(1-\mu_i)}}, \end{aligned}$$

where the first inequality uses the results for $1/2 \le \mu_i' \le \mu_i$ and for $\mu_i' \le \mu_i \le 1/2$, the second inequality uses the relation that $\mu_i' \ge 1/2$ and $1 - \mu_i \ge 1/2$. So $|g(\mu_i') - g(\mu_i)| \le 3\sqrt{2}\frac{|\mu_i' - \mu_i|}{\sqrt{(1-\mu_i)\mu_i}}$.

By above cases, we have $|g(\mu_i') - g(\mu_i)| \le 3\sqrt{2}\frac{|\mu_i' - \mu_i|}{\sqrt{(1-\mu_i)\mu_i}}$. Putting back this inequality into Eq. (27), we have

$$|r(S;\boldsymbol{\mu}') - r(S;\boldsymbol{\mu})| \le 3\sqrt{2}\gamma_g \sqrt{\sum_{i\in S}\left(\frac{|\mu_i' - \mu_i|}{\sqrt{(1-\mu_i)\mu_i}}\right)^2}. \tag{28}$$

For the $|r(S;\boldsymbol{\mu}'') - r(S;\boldsymbol{\mu}')|$ term,

We can use the gradient theorem to bound,

$$\begin{aligned}
|r(S;\boldsymbol{\mu}'') - r(S;\boldsymbol{\mu}')| &= \left|\int_{\boldsymbol{x}=\boldsymbol{\mu}'}^{\boldsymbol{\mu}''} \nabla r(S;\boldsymbol{x}) \cdot d\boldsymbol{x}\right| \\
&\le \sup_{\boldsymbol{x}} \|\nabla r(S;\boldsymbol{x})\|_\infty \sum_{i\in S} |\mu_i'' - \mu_i'| \\
&\le \gamma_\infty \sum_{i\in S} |\eta_i|. \qquad (29)
\end{aligned}$$

Combining Eq. (28) and Eq. (29), we conclude the lemma. ■

C Regret Analysis for CMAB-T with TPVM Bounded Smoothness (Proofs Related to Theorem 1)

In this section, we provide detailed proofs for Theorem 1 and give some discussions for the distribution-independent regret bounds as well as the lower bound results.

For the structure of this section, we first introduce some useful tools in Appendix C.1 that will be helpful for our analysis. Next we transform the total regret to the regret terms filtered by some events in Appendix C.2. Then we provide regret bounds for all these regret terms. For these regret terms, we give two different proofs for the leading regret term: the proof giving Theorem 1 that uses the reverse amortization trick (see Eq. (62) and Eq. (73)) are in Appendix C.3, while the proof Appendix C.4 directly follows [23]. Recall that former proof improves the latter by a factor of $O(\log K)$ and readers can skip the latter one if you are not interested. It is notable that this trick can be used to improve Degenne and Perchet [11], Merlis and Mannor [23], Perrault et al. [25] in a similar way, owing to the fact that their error terms have the similar form as ours shown in Eq. (48) (except without triggering probability modulation). Lastly, we summarize the detailed distribution-dependent regret, distribution-independent regret bounds and lower bounds in Appendix C.5.

C.1 Useful Concentration Bounds, Definitions and Inequalities

We use the following tail bound for the construction of the confidence radius and our analysis.

Lemma 4 (Empirical Bernstein Inequality [2])**.** *Let $(X_i)_{i\in[n]}$ be n i.i.d random variables with bounded support $[0,1]$ and mean $\mathbb{E}[X_i] = \mu$. Let $\hat{X}_n \triangleq \frac{1}{n}\sum_{i\in[n]} X_i$ and $\hat{V}_n \triangleq \frac{1}{n}\sum_{i\in[n]}(X_i - \hat{X}_n)^2$ be the empirical mean and empirical variance of $(X_i)_{i\in[n]}$. Then for any $n \in \mathbb{N}$ and $y > 0$, it holds that*

$$\Pr\left[|\hat{X}_n - \mu| \ge \sqrt{\frac{2\hat{V}_n y}{n}} + \frac{3y}{n}\right] \le 3e^{-y} \tag{30}$$

We use the following Bernstein Inequality to bound the difference between the empirical variance and the true variance.

Lemma 5 (Bernstein Inequality [12])**.** *Let $(X_i)_{i\in[n]}$ be n independent random variables in $[0,1]$ with mean $\mathbb{E}[X_i] = \mu$ and variance $Var[X_i] \triangleq \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2 = V$. Then with probability $1-\delta$:*

$$\frac{1}{n}\sum_{i\in[n]} X_i \le \mu + \frac{2\log 1/\delta}{3n} + \sqrt{\frac{2V\log 1/\delta}{n}}. \tag{31}$$

Similar to [29], we define the event-filtered regret, the triggering group, the counter, the nice triggering event and the nice sampling event to help our analysis.

Definition 2 (Event-Filtered Regret). *For any series of events $(\mathcal{E}_t)_{t\geq[T]}$ indexed by round number t, we define the $Reg^A_{\alpha,\boldsymbol{\mu}}(T,(\mathcal{E}_t)_{t\geq[T]})$ as the regret filtered by events $(\mathcal{E}_t)_{t\geq[T]}$, or the regret is only counted in t if $\mathcal{E}$ happens in t. Formally,*

$$Reg^A_{\alpha,\boldsymbol{\mu}}(T,(\mathcal{E}_t)_{t\geq[T]}) = \mathbb{E}\left[\sum_{t\in[T]} \mathbb{I}(\mathcal{E}_t)(\alpha\cdot r(S^*;\boldsymbol{\mu}) - r(S_t;\boldsymbol{\mu}))\right]. \tag{32}$$

For simplicity, we will omit $A,\alpha,\boldsymbol{\mu},T$ and rewrite $Reg^A_{\alpha,\boldsymbol{\mu}}(T,(\mathcal{E}_t)_{t\geq[T]})$ as $Reg(T,\mathcal{E}_t)$ when contexts are clear.

Definition 3 (Triggering Probability (TP) group). *For any arm i and index j, define the triggering probability (TP) group (of actions) as*

$$\mathcal{S}^D_{i,j} = \{S\in\mathcal{S} : 2^{-j} < p_i^{D,S} \leq 2^{-j+1}\}. \tag{33}$$

Notice $\{\mathcal{S}^D_{i,j}\}$ forms a partition of $\{S\in\mathcal{S} : p_i^{D,S}\}$.

Definition 4 (Counter). *For each TP group $S_{i,j}$, we define a counter $N_{i,j}$ which is initialized to 0. In each round t, if the action S_t is chosen, then we update $N_{i,j}$ to $N_{i,j}+1$ for (i,j) that $S_t\in S^D_{i,j}$. We also denote $N_{i,j}$ at the end of round t as $N_{t,i,j}$. Formally, we have the following recursive equation to define $N_{t,i,j}$ as follows:*

$$N_{t,i,j} = \begin{cases} 0, & \textit{if } t=0 \\ N_{t-1,i,j}+1, & \textit{if } t>0 \textit{ and } S_t\in S^D_{i,j} \\ N_{t-1,i,j}, & \textit{otherwise.} \end{cases} \tag{34}$$

Definition 5 (Nice triggering event $\mathcal{N}^t_t$). *Given a series integers $\{j_i^{\max}\}_{i\in[m]}$, we say that the triggering is nice at the beginning of round t, if for every triggered group identified by (i,j), as long as $\frac{6\ln t}{\frac{1}{3}N_{t-1,i,j}2^{-j}} \leq 1$, there is $T_{t-1,i} \geq \frac{1}{3}N_{t-1,i,j}\cdot 2^{-j}$. We denote this event as $\mathcal{N}^t_t$.*

Lemma 6 (Appendix B.1, Lemma 4 [29]). *For a series of integers $(j_i^{\max})_{i\in[m]}$, we have $\Pr[\neg\mathcal{N}^t_t] \leq \sum_{i\in[m]} j_i^{\max} t^{-2}$ for every round $t\in[T]$.*

Proof. We refer the readers to Lemma 4 in Appendix B.1 from Wang and Chen [29] for detailed proofs. ■

Definition 6. *We say that the sampling is nice at the beginning of round t if: (1) for every base arm $i\in[m]$, $|\hat{\mu}_{t-1,i}-\mu_i| \leq \rho_{t,i}$, where $\rho_{t,i} = \sqrt{\frac{6\hat{V}_{t-1,i}\log t}{T_{t-1,i}}} + \frac{9\log t}{T_{t-1,i}}$; (2) for every base arm $i\in[m]$, $\hat{V}_{t-1,i} \leq 2\mu_i(1-\mu_i) + \frac{3.5\log t}{T_{t-1,i}}$. We denote such event as $\mathcal{N}^s_t$.*

The following lemma bounds the probability that $\mathcal{N}^s_t$ does not happen.

Lemma 7. *For each round t, $\Pr[\neg\mathcal{N}^s_t] \leq 4mt^{-2}$.*

Proof. Let $\mathcal{N}^{s,1}_t, \mathcal{N}^{s,2}_t$ be the event (1) and event (2), where $\mathcal{N}^s_t = \mathcal{N}^{s,1}_t \bigcap \mathcal{N}^{s,2}_t$. We first bound the probability that $\mathcal{N}^{s,1}_t$ does not happen, we have

$$\Pr[\neg\mathcal{N}^{s,1}_t] = \Pr\left[\exists i\in[m] \text{ s.t. } |\hat{\mu}_{t-1,i}-\mu_i| > \sqrt{\frac{6\hat{V}_{t-1,i}\log t}{T_{t-1,i}}} + \frac{9\log t}{T_{t-1,i}}\right] \tag{35}$$

$$\leq \sum_{i\in[m]}\sum_{\tau\in[t]} \Pr\left[|\hat{\mu}_{t-1,i}-\mu_i| > \sqrt{\frac{6\hat{V}_{t-1,i}\log t}{\tau}} + \frac{9\log t}{\tau}, T_{t-1,i}=\tau\right] \tag{36}$$

$$\leq 3mt^{-2}, \tag{37}$$

where Eq. (36) is due to the union bound over i, τ, Eq. (37) is due to Lemma 4 by setting $y = 3\log t$ and when $T_{t-1,i} = \tau, \hat{\mu}_{t-1,i}$ and $\hat{V}_{t-1,i}$ are the empirical mean and empirical variance of τ i.i.d random variables with mean μ_i.

We then bound the probability that second event $\mathcal{N}_t^{s,2}$ does not happen using the similar proof of [23, Eq. (7)]. Fix $T_{t-1,i} = \tau$ and consider $(Y_i^1, ..., Y_i^\tau)$, where $Y_i^k = (X_i^k - \mu_i)^2 \in [0,1]$ and X_i^k is the random outcome of the k-th i.i.d trial. Since X_i^k are independent across k, Y_i^k are independent across k as well. In this case, one can verify that $\hat{V}_{t-1,i} = \frac{1}{\tau}\sum_{k=1}^{\tau}(X_i^k - \mu_i)^2 - (\frac{1}{\tau}\sum_{k=1}^{\tau} X_i^k - \mu_i)^2 \le \frac{1}{\tau}\sum_{k=1}^{\tau}(X_i^k - \mu_i)^2 = \frac{1}{\tau}\sum_{k=1}^{\tau} Y_i^k$; $\mathbb{E}[Y_i^k] = \mathbb{E}[(X_i^k)^2] - \mu_i^2 \le \mathbb{E}[X_i^k] \cdot 1 - \mu_i^2 = (1-\mu_i)\mu_i$; and $\text{Var}[Y_i] = \mathbb{E}[(Y_i^k)^2] - (\mathbb{E}[Y_i^k])^2 \le \mathbb{E}[(Y_i^k)^2] \le \mathbb{E}[Y_i^k] \le (1-\mu_i)\mu_i$. By Lemma 5 over τ i.i.d random variable $(Y_i^k)_{k\in\tau}$, it holds with probability at least $1 - t^{-3}$ that

$$\frac{1}{\tau}\sum_{k=1}^{\tau} Y_i^k \le \mathbb{E}[Y_i^k] + \frac{2\log t}{\tau} + \sqrt{\frac{6\text{Var}[Y_i^k]\log t}{\tau}} \tag{38}$$

This implies

$$\begin{aligned}
\hat{V}_{t-1,i} \le \frac{1}{\tau}\sum_{k=1}^{\tau} Y_i^k &\le \mathbb{E}[Y_i^k] + \frac{2\log t}{\tau} + \sqrt{\frac{6\text{Var}[Y_i^k]\log t}{\tau}} &(39)\\
&\le \mu_i(1-\mu_i) + \frac{2\log t}{\tau} + \sqrt{\frac{6(1-\mu_i)\mu_i \log t}{\tau}} &(40)\\
&\le \mu_i(1-\mu_i) + \frac{2\log t}{\tau} + \mu_i(1-\mu_i) + \frac{3\log t}{2\tau} &(41)\\
&= 2\mu_i(1-\mu_i) + \frac{3.5\log t}{\tau} &(42)
\end{aligned}$$

where Eq. (40) is using $2ab \le a^2 + b^2$ and $a = \sqrt{2\mu_i(1-\mu_i)}, b = \sqrt{\frac{3\log t}{n}}$.

Now by applying union bound over $i \in [m]$ and $\tau \in [t]$, we have $\Pr[\neg\mathcal{N}_t^{s,2}] \le mt^{-2}$. Lastly, applying union bound over $\mathcal{N}_t^{s,1}$ and $\mathcal{N}_t^{s,2}$, we have $\Pr[\neg\mathcal{N}_t^{s}] \le 4mt^{-2}$. ■

After setting up all above definitions, we can prove Lemma 8 about the confidence radius, which appears in the main content.

Lemma 8. *Fix every base arm i and every time t, with probability at least $1 - 4mt^{-3}$, it holds that*

$$\mu_i \le \bar{\mu}_{t,i} \le \min\{\mu_i + 2\rho_{t,i}, 1\} \le \min\left\{\mu_i + 4\sqrt{3}\sqrt{\frac{\mu_i(1-\mu_i)\log t}{T_{t-1,i}}} + \frac{28\log t}{T_{t-1,i}}, 1\right\}. \tag{43}$$

Proof. Recall that $\bar{\mu}_{t,i} = \min\{\hat{\mu}_{t-1,i} + \rho_{t,i}, 1\} = \min\{\hat{\mu}_{t-1,i} + \sqrt{\frac{6\hat{V}_{t-1,i}\log t}{T_{t-1,i}}} + \frac{9\log t}{T_{t-1,i}}, 1\}$. Under event $N_t^{s,1}$, we have $|\mu_i - \hat{\mu}_{t,i}| \le \rho_{t,i}$ by Lemma 7, hence the first and the second inequality in Lemma 8 holds.

For the last inequality, under event $N_t^{s,2}$, it holds that

$$\begin{aligned}
\mu_i + 2\rho_{t,i} &= \mu_i + 2\left(\sqrt{\frac{6\hat{V}_{t-1,i}\log t}{T_{t-1,i}}} + \frac{9\log t}{T_{t-1,i}}\right) &(44)\\
&\le \mu_i + 2\left(\sqrt{\frac{6\cdot(2\mu_i(1-\mu_i) + \frac{3.5\log t}{T_{t-1,i}})\log t}{T_{t-1,i}}} + \frac{9\log t}{T_{t-1,i}}\right) &(45)\\
&\le \mu_i + 4\sqrt{3}\sqrt{\frac{\mu_i(1-\mu_i)\log t}{T_{t-1,i}}} + 2\sqrt{21}\frac{\log t}{T_{t-1,i}} + \frac{18\log t}{T_{t-1,i}} &(46)
\end{aligned}$$

$$\leq \mu_{t-1,i} + 4\sqrt{3}\sqrt{\frac{\mu_i(1-\mu_i)\log t}{T_{t-1,i}}} + \frac{28\log t}{T_{t-1,i}}, \tag{47}$$

where Eq. (46) uses $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$.

Since $\mathcal{N}_t^s = \mathcal{N}_t^{s,1} \bigcap \mathcal{N}_t^{s,2}$ and by Lemma 7, Eq. (43) holds with probability at least $1 - 4mt^{-2}$. ■

C.2 Decompose the Total Regret to Event-Filtered Regrets

In this section, we decompose the regret $Reg(T, \{\}) = Reg(T, \mathcal{N}_t^s, \mathcal{N}_t^o) + Reg(T, \neg(\mathcal{N}_t^s \bigcap \mathcal{N}_t^o)) \leq Reg(T, \mathcal{N}_t^s, \mathcal{N}_t^o) + Reg(T, \neg\mathcal{N}_t^s) + Reg(T, \neg\mathcal{N}_t^o)$, where $\mathcal{N}_t^s$ is defined in Definition 6, $\mathcal{N}_t^o$ denotes the event where oracle successfully outputs an α-approximate solution (with probability at least β). We have the following lemma to do the decomposition.

Lemma 9. *[Leading Regret Term] Let $r(S;\boldsymbol{\mu})$ be TPVM smoothness with coefficients (B_v, B_1, λ), and define the error term*

$$e_t(S_t) = 4\sqrt{3}B_v\sqrt{\sum_{i\in\tilde{S}_t}(\frac{\log t}{T_{t-1,i}} \wedge \frac{1}{28})(p_i^{D,S_t})^\lambda} + 28B_1\sum_{i\in\tilde{S}_t}(\frac{\log t}{T_{t-1,i}} \wedge \frac{1}{28})(p_i^{D,S_t}) \tag{48}$$

and event $E_t = \mathbb{I}\{\Delta_{S_t} \leq e_t(S_t)\}$. The regret of Algorithm 1, when used with (α,β) approximation oracle is bounded by

$$Reg(T) \leq Reg(T, E_t) + \frac{2\pi^2}{3}m\Delta_{\max}. \tag{49}$$

Proof. Under event $\mathcal{N}_t^s, \mathcal{N}_t^o$, by Lemma 8, it is easily to check that

$$\begin{aligned}\bar{\mu}_{t,i} &\leq \min\{\mu_{t-1,i} + 4\sqrt{3}\sqrt{\frac{\mu_i(1-\mu_i)\log t}{T_{t-1,i}}} + \frac{28\log t}{T_{t-1,i}}, 1\} \\ &\leq \mu_{t-1,i} + 4\sqrt{3}\sqrt{\mu_i(1-\mu_i)(\frac{\log t}{T_{t-1,i}} \wedge \frac{1}{28})} + 28(\frac{\log t}{T_{t-1,i}} \wedge \frac{1}{28}) \end{aligned} \tag{50}$$

Therefore, it holds that

$$\alpha r(S^*;\boldsymbol{\mu}) \leq \alpha r(S^*;\bar{\boldsymbol{\mu}}_t) \leq r(S_t;\bar{\boldsymbol{\mu}}_t) \tag{51}$$

$$\leq r(S_t;\boldsymbol{\mu}) + 4\sqrt{3}B_v\sqrt{\sum_{i\in\tilde{S}_t}(\frac{\log t}{T_{t-1,i}} \wedge \frac{1}{28})(p_i^{D,S_t})^\lambda} + 28B_1\sum_{i\in\tilde{S}_t}(\frac{\log t}{T_{t-1,i}} \wedge \frac{1}{28})(p_i^{D,S_t}), \tag{52}$$

where the first inequality in Eq. (51) is due to monotonicity condition (**Condition 1**) and second inequality in Eq. (51) is due to event $\mathcal{N}_t^o$, Eq. (52) is because of Eq. (50) and the TPVM condition (**Condition 3**) by plugging in $\zeta_i = 4\sqrt{3}\sqrt{\mu_i(1-\mu_i)(\frac{\log t}{T_{t-1,i}} \wedge \frac{1}{28})}$ and $\eta_i = 28(\frac{\log t}{T_{t-1,i}} \wedge \frac{1}{28})$.

So $Reg(T, \mathcal{N}_t^s, \mathcal{N}_t^o) \leq Reg(T, E_t)$. Now for $Reg(T, \neg\mathcal{N}_t^s)$, by Lemma 7 it holds that

$$Reg(T, \neg\mathcal{N}_t^s) \leq \sum_{t=1}^T \Pr[\neg\mathcal{N}_t^s] \leq \sum_{t=1}^T 4mt^{-2} \leq \frac{2\pi^2}{3}m\Delta_{\max}. \tag{53}$$

Similarly by definition, it holds that

$$Reg(T, \neg\mathcal{N}_t^o) \leq (1-\beta)T\Delta_{\max}. \tag{54}$$

Therefore $Reg(T, \{\}) \leq Reg(T, E_t) + \frac{2\pi^2}{3}m\Delta_{\max} + (1-\beta)T\Delta_{\max}$. And we have $Reg(T) = Reg(T, \{\}) - (1-\beta)T\Delta\max \leq Reg(T, E_t) + \frac{2\pi^2}{3}\Delta_{\max}$, which concludes Lemma 9.

■

Recall that event $E_t = \{\Delta_{S_t} \le e_t(S_t)\}$, where $e_t(S_t) = 4\sqrt{3}B_v\sqrt{\sum_{i\in\tilde{S}_t}(\frac{\log t}{T_{t-1,i}} \wedge \frac{1}{28})(p_i^{D,S_t})^\lambda} + 28B_1\sum_{i\in\tilde{S}_t}(\frac{\log t}{T_{t-1,i}} \wedge \frac{1}{28})(p_i^{D,S_t})$. We will further decompose the event-filtered regret $Reg(T, E_t)$ into two event-filtered regret $Reg(T, E_{t,1})$ and $Reg(T, E_{t,2})$,

$$Reg(T, E_t) \le Reg(T, E_{t,1}) + Reg(T, E_{t,2}), \tag{55}$$

where $E_{t,1} = \{\Delta_{S_t} \le 2e_{t,1}(S_t)\}$, $E_{t,2} = \{\Delta_{S_t} \le 2e_{t,2}(S_t)\}$, $e_{t,1}(S_t) = 4\sqrt{3}B_v\sqrt{\sum_{i\in\tilde{S}_t}(\frac{\log t}{T_{t-1,i}} \wedge \frac{1}{28})(p_i^{D,S_t})^\lambda}$,$e_{t,2}(S_t) = 28B_1\sum_{i\in\tilde{S}_t}(\frac{\log t}{T_{t-1,i}} \wedge \frac{1}{28})(p_i^{D,S_t})$. The above inequality holds since the following facts: We can observe $e_{t,1}(S_t) + e_{t,2}(S_t) = e_t(S_t)$. From E_t, we know either $E_{t,1}$ holds or $E_{t,2}$ holds. So E_t implies that $1 \le \mathbb{I}\{E_{t,1}\} + \mathbb{I}\{E_{t,2}\}$, and thus $\Delta_{S_t}\mathbb{I}\{E_t\} \le \Delta_{S_t}\mathbb{I}\{E_{t,1}\} + \Delta_{S_t}\mathbb{I}\{E_{t,2}\}$, which concludes $Reg(T, E_t) \le Reg(T, E_{t,1}) + Reg(T, E_{t,2})$. The next two sections will provide two different proofs for $Reg(T, E_{t,1}), Reg(T, E_{t,2})$ separately, where the second improves the first by a factor of $O(\log K)$.

C.3 Our Improved Analysis Using the Reverse Amortized Trick

In this section, we are going to bound the $Reg(T, E_{t,1})$ and $Reg(T, E_{t,2})$ separately under the event $\mathcal{N}_t^t$, similar to Appendix C.4. The idea is to use a refined reverse amortization trick originated in [29] and to allocate the regret Δ_{S_t} to each base arm according to carefully designed thresholds. Note that it is highly non-trivial to derive the right thresholds and regret allocation strategy so that the K, T factors are as small as possible, which is our main contribution.

C.3.1 Upper bound for $Reg(T, E_{t,1})$

We first break $Reg(T, E_{t,1})$ into two parts and bound them separately: $Reg(T, E_{t,1} \bigcap \mathcal{N}_t^t)$ and $Reg(T, \neg\mathcal{N}_t^t)$.

For $Reg(T, E_{t,1} \bigcap \mathcal{N}_t^t)$, under the event $\mathcal{N}_t^t$, let $c_1 = 4\sqrt{3}$ and we set $j_i^{\max} = \frac{1}{\lambda}(\lceil \log_2 \frac{c_1^2 B_v^2 K}{(\Delta_i^{\min})^2}\rceil + 1)$. We first define a regret allocation function

$$\kappa_{i,j,T}(\ell) = \begin{cases} \frac{c_1^2 B_v^2 2^{(-j+1)(\lambda-1)}}{\Delta_i^{\min}}, & \text{if } \ell = 0 \text{ and } j \le j_i^{\max}, \\ 2\sqrt{\frac{24c_1^2 B_v^2 2^{(-j+1)(\lambda-1)}\log T}{\ell}}, & \text{if } 1 \le \ell \le L_{i,j,T,1} \text{ and } j \le j_i^{\max}, \\ \frac{48c_1^2 B_v^2 2^{(-j+1)(\lambda-1)}\log T}{\Delta_i^{\min}}\frac{1}{\ell}, & \text{if } L_{i,j,T,1} < \ell \le L_{i,j,T,2} \text{ and } j \le j_i^{\max}, \\ 0, & \text{if } \ell > L_{i,j,T,2} \text{ or } j > j_i^{\max}, \end{cases} \tag{56}$$

where $L_{i,j,T,1} = \frac{24c_1^2 B_v^2 2^{(-j+1)(\lambda-1)}\log T}{(\Delta_i^{\min})^2}$, $L_{i,j,T,2} = \frac{48c_1^2 B_v^2 2^{(-j+1)(\lambda-1)} K\log T}{(\Delta_i^{\min})^2}$.

Lemma 10. *For any time $t \in [T]$, if N_t^t and $E_{t,1}$ hold, we have*

$$\Delta_{S_t} \le \sum_{i\in\tilde{S}_t} \kappa_{i,j_i^{S_t},T}(N_{t-1,i,j_i^{S_t}}), \tag{57}$$

where $j_i^{S_t}$ is the index of the triggering group $S_{i,j}$ such that $2^{-j_i^{S_t}} < p_i^{D,S_t} \le 2^{-j_i^{S_t}+1}$.

Proof. By event $E_{t,1}$, which is defined in Eq. (48), we apply the reverse amortization (Eq. (62))

$$\Delta_{S_t} \le \sum_{i\in\tilde{S}_t} \frac{4c_1^2 B_v^2 (p_i^{D,S_t})^\lambda \min\{\frac{\log t}{T_{i,t-1}}, \frac{1}{28}\}}{\Delta_{S_t}} \tag{58}$$

$$\le -\Delta_{S_t} + 2\sum_{i\in\tilde{S}_t} \frac{4c_1^2 B_v^2 (p_i^{D,S_t})^\lambda \min\{\frac{\log t}{T_{i,t-1}}, \frac{1}{28}\}}{\Delta_{S_t}} \tag{59}$$

$$\le \sum_{i\in\tilde{S}_t}\left(\frac{8c_1^2 B_v^2 (p_i^{D,S_t})^\lambda \min\{\frac{\log t}{T_{i,t-1}}, \frac{1}{28}\}}{\Delta_{S_t}} - \frac{\Delta_{S_t}}{|\tilde{S}_t|}\right) \tag{60}$$

$$\le \sum_{i\in\tilde{S}_t}\left(\frac{8c_1^2 B_v^2 (p_i^{D,S_t})^\lambda \min\{\frac{\log t}{\frac{1}{3}N_{t-1,i,j_i^{S_t}} 2^{-j_i^{S_t}}}, \frac{1}{28}\}}{\Delta_{S_t}} - \frac{\Delta_{S_t}}{|\tilde{S}_t|}\right) \tag{61}$$

$$\leq \sum_{i \in \tilde{S}_t} \underbrace{\left(\frac{8c_1^2 B_v^2 (2^{-j_i^{S_t}+1})^\lambda \min\{\frac{\log t}{\frac{1}{3} N_{t-1,i,j_i^{S_t}} 2^{-j_i^{S_t}}}, \frac{1}{28}\}}{\Delta_{S_t}} - \frac{\Delta_{S_t}}{K} \right)}_{(62,i)}, \tag{62}$$

where Eq. (58) is by the definition of $E_{t,1}$ which says $\Delta_{S_t}^2 \leq \sum_{i \in \tilde{S}_t} 4c_1^2 B_v^2 (p_i^{D,S_t})^\lambda \min\{\frac{\log t}{T_{i,t-1}}, \frac{1}{28}\}$ and by dividing both sides by $\Delta_{S_t} > 0$, Eq. (59) is because we double the LHS and RHS of Eq. (58) at the same time and then put one into the RHS, Eq. (60) is by putting $-\Delta_{S_t}$ inside the summation, Eq. (61) is due to the same reason of Eq. (76) under event $\mathcal{N}_t^t$, Eq. (62) is due to $p_i^{D,S_t} \leq 2^{-j_i^{S_t}+1}$ given by the definition of $j_i^{S_t}$ and $|\tilde{S}_t| \leq K$.

Note that the Eq. (59) is called the reverse amortization trick, since we allocate two times of the total regret and then minus the Δ_{S_t} term to amortize the regret when $\ell > L_{i,j,T,2}$ or $j > j_i^{\max}$ in Eq. (57), which saves the analysis for arms that are sufficiently triggered. Now we bound (62, i) under different cases.

When $j > j_i^{\max}$,

we have $(62, i) \leq \frac{8c_1^2 B_v^2 (2^{-j_i^{S_t}+1})^\lambda}{\Delta_{S_t}} \cdot \frac{1}{28} - \frac{\Delta_{S_t}}{K} \leq \frac{8c_1^2 B_v^2}{\Delta_{S_t}} \frac{(\Delta_i^{\min})^2}{c_1^2 B_v^2 K} \cdot \frac{1}{28} - \frac{\Delta_{S_t}}{K} \leq \frac{\Delta_i^{\min}}{K} \cdot \frac{8}{28} - \frac{\Delta_{S_t}}{K} \leq 0 = \kappa_{i,j_i^{S_t},T}(N_{t-1,i,j_i^{S_t}})$.

When $N_{t-1,i,j_i^{S_t}} > L_{i,j_i^{S_t},T,2}$,

we have $(62, i) \leq \frac{8c_1^2 B_v^2 (2^{-j_i^{S_t}+1})^\lambda \log t}{\frac{1}{3} N_{t-1,i,j_i^{S_t}} \cdot 2^{-j_i^{S_t}} \Delta_{S_t}} - \frac{\Delta_{S_t}}{K} \leq \frac{48c_1^2 B_v^2 2^{(-j_i^{S_t}+1)(\lambda-1)} \log T}{\Delta_{S_t}} \frac{1}{N_{t-1,i,j_i^{S_t}}} - \frac{\Delta_{S_t}}{K} < \frac{(\Delta_i^{\min})^2}{K\Delta_{S_t}} - \frac{\Delta_{S_t}}{K} \leq 0 = \kappa_{i,j_i^{S_t},T}(N_{t-1,i,j_i^{S_t}})$.

When $L_{i,j_i^{S_t},T,1} < N_{t-1,i,j_i^{S_t}} \leq L_{i,j_i^{S_t},T,2}$ and $j \leq j_i^{\max}$,

We have $(62, i) \leq \frac{8c_1^2 B_v^2 (2^{-j_i^{S_t}+1})^\lambda \log t}{\frac{1}{3} N_{t-1,i,j_i^{S_t}} 2^{-j_i^{S_t}} \Delta_{S_t}} - \frac{\Delta_{S_t}}{K} \leq \frac{48c_1^2 B_v^2 2^{(-j_i^{S_t}+1)(\lambda-1)} \log T}{\Delta_{S_t}} \frac{1}{N_{t-1,i,j_i^{S_t}}} - \frac{\Delta_{S_t}}{K} < \frac{48c_1^2 B_v^2 2^{(-j_i^{S_t}+1)(\lambda-1)} \log T}{\Delta_{S_t}} \frac{1}{N_{t-1,i,j_i^{S_t}}} \leq \frac{48c_1^2 B_v^2 2^{(-j_i^{S_t}+1)(\lambda-1)} \log T}{\Delta_i^{\min}} \frac{1}{N_{t-1,i,j_i^{S_t}}} = \kappa_{i,j_i^{S_t},T}(N_{t-1,i,j_i^{S_t}})$.

When $N_{t-1,i,j_i^{S_t}} \leq L_{i,j_i^{S_t},T,1}$ and $j \leq j_i^{\max}$,

We further consider two different cases $N_{t-1,i,j_i^{S_t}} \leq \frac{24c_1^2 B_v^2 2^{(-j_i^{S_t}+1)(\lambda-1)} \log T}{(\Delta_{S_t})^2}$ or $\frac{24c_1^2 B_v^2 2^{(-j_i^{S_t}+1)(\lambda-1)} \log T}{(\Delta_{S_t})^2} < N_{t-1,i,j_i^{S_t}} \leq L_{i,j_i^{S_t},T,1} = \frac{24c_1^2 B_v^2 2^{(-j_i^{S_t}+1)(\lambda-1)} \log T}{(\Delta_i^{\min})^2}$.

For the former case, if there exists $i \in \tilde{S}_t$ so that $N_{t-1,i,j_i^{S_t}} \leq \frac{24c_1^2 B_v^2 2^{(-j_i^{S_t}+1)(\lambda-1)} \log T}{(\Delta_{S_t})^2}$, then we know $\sum_{q \in \tilde{S}_t} \kappa_{i,j_q^{S_t},T}(N_{t-1,q,j_q^{S_t}}) \geq \kappa_{i,j_i^{S_t},T}(N_{t-1,i,j_i^{S_t}}) = 2\sqrt{\frac{24c_1^2 B_v^2 2^{(-j_i^{S_t}+1)(\lambda-1)} \log T}{N_{t-1,i,j_i^{S_t}}}} \geq 2\Delta_{S_t} > \Delta_{S_t}$, which makes Eq. (57) holds no matter what. This means we do not need to consider this case for good.

For the later case, when $\frac{24c_1^2 B_v^2 2^{(-j_i^{S_t}+1)(\lambda-1)} \log T}{(\Delta_{S_t})^2} < N_{t-1,i,j_i^{S_t}}$, we know that $(62, i) \leq \frac{48c_1^2 B_v^2 2^{(-j_i^{S_t}+1)(\lambda-1)} \log T}{\Delta_{S_t}} \frac{1}{N_{t-1,i,j_i^{S_t}}} =$

$2\sqrt{\frac{24c_1^2B_v^22^{(-j_i^{S_t}+1)(\lambda-1)}\log T}{(\Delta_{S_t})^2}\frac{1}{N_{t-1,i,j_i^{S_t}}}}\sqrt{\frac{24c_1^2B_v^22^{(-j_i^{S_t}+1)(\lambda-1)}\log T}{N_{t-1,i,j_i^{S_t}}}} \leq$
$2\sqrt{\frac{24c_1^2B_v^22^{(-j_i^{S_t}+1)(\lambda-1)}\log T}{N_{t-1,i,j_i^{S_t}}}} = \kappa_{i,j_i^{S_t},T}(N_{t-1,i,j_i^{S_t}})$.

When $\ell = 0$ and $j \leq j_i^{\max}$,

We have $(62, i) \leq \frac{8c_1^2B_v^2(2^{-j_i^{S_t}+1})^\lambda}{\Delta_{S_t}} \cdot \frac{1}{28} - \frac{\Delta_{S_t}}{K} \leq \frac{c_1^2B_v^2(2^{-j_i^{S_t}+1})^\lambda}{\Delta_{S_t}} \leq \frac{c_1^2B_v^2(2^{-j_i^{S_t}+1})^\lambda}{\Delta_i^{\min}} = \kappa_{i,j_i^{S_t},T}(N_{t-1,i,j_i^{S_t}})$.

Combining all above cases, we have $\Delta_{S_t} \leq \sum_{i\in\tilde{S}_t}\kappa_{i,j_i^{S_t},T}(N_{t-1,i,j_i^{S_t}})$. ■

Since $N_{t,i,j_i^{S_t}}$ is increased if and only if $i \in \tilde{S}_t$ and consider all possible N_{t,i,j_i^S} where $\kappa_{i,j_i^S,T}(S, N_{t-1,i,j^S}) > 0$, we have

$$
\begin{aligned}
&Reg(T, E_{t,1} \bigcap N_t^t) \\
&\leq \sum_{t\in[T]}\sum_{i\in\tilde{S}_t}\kappa_{i,j_i^{S_t},T}(N_{t-1,i,j^{S_t}}) && (63)\\
&\leq \sum_{i\in[m]}\sum_{j=1}^{j_i^{\max}}\frac{c_1^2B_v^2(2^{-j+1})^\lambda}{\Delta_i^{\min}} + \sum_{i\in[m]}\sum_{j=1}^{j_i^{\max}}\sum_{\ell=1}^{L_{i,j,T,1}}2\sqrt{\frac{24c_1^2B_v^22^{(-j+1)(\lambda-1)}\log T}{\ell}} \\
&+ \sum_{i\in[m]}\sum_{j=1}^{j_i^{\max}}\sum_{L_{i,j,T,1}+1}^{L_{i,j,T,2}}\frac{48c_1^2B_v^22^{(-j+1)(\lambda-1)}\log T}{\Delta_i^{\min}}\frac{1}{\ell} && (64)\\
&\leq \sum_{i\in[m]}\sum_{j=1}^{j_i^{\max}}\frac{c_1^2B_v^2(2^{-j+1})^\lambda}{\Delta_i^{\min}} + \sum_{i\in[m]}\sum_{j=1}^{j_i^{\max}}\frac{96c_1^2B_v^22^{(-j+1)(\lambda-1)}\log T}{\Delta_i^{\min}} \\
&+ \sum_{i\in[m]}\sum_{j=1}^{j_i^{\max}}\frac{48c_1^2B_v^22^{(-j+1)(\lambda-1)}\log T}{\Delta_i^{\min}}(1+\log K) && (65)\\
&= \sum_{i\in[m]}\sum_{j=1}^{j_i^{\max}}\frac{c_1^2B_v^2(2^{-j+1})^\lambda}{\Delta_i^{\min}} + \sum_{i\in[m]}\sum_{j=1}^{j_i^{\max}}\frac{48c_1^2B_v^22^{(-j+1)(\lambda-1)}\log T}{\Delta_i^{\min}}(3+\log K) && (66)
\end{aligned}
$$

When $\lambda > 1$, we have $Reg(T, E_{t,1} \bigcap N_t^t) \leq \sum_{i\in[m]}\sum_{j=1}^{\infty}\frac{c_1^2B_v^22^{-j+1}}{\Delta_i^{\min}} + \sum_{i\in[m]}\sum_{j=1}^{\infty}\frac{48c_1^2B_v^22^{(-j+1)(\lambda-1)}\log T}{1-2^{(\lambda-1)}\Delta_i^{\min}}(3+\log K) = \sum_{i\in[m]}\frac{2c_1^2B_v^2}{\Delta_i^{\min}} + \sum_{i\in[m]}\frac{48c_1^2B_v^2\log T}{\Delta_i^{\min}}(3+\log K)$.

When $\lambda = 1$, we have $Reg(T, E_{t,1} \bigcap N_t^t) \leq \sum_{i\in[m]}\sum_{j=1}^{\infty}\frac{c_1^2B_v^22^{-j+1}}{\Delta_i^{\min}} + \sum_{i\in[m]} j_i^{\max}\frac{48c_1^2B_v^2\log T}{\Delta_i^{\min}}(3+\log K) = \sum_{i\in[m]}\frac{2c_1^2B_v^2}{\Delta_i^{\min}} + \sum_{i\in[m]}\log\frac{c_1^2B_v^2K}{(\Delta_i^{\min})^2}\frac{48c_1^2B_v^2\log T}{\Delta_i^{\min}}(3+\log K)$.

Similar to Eq. (96), with additional $Reg(T, \neg\mathcal{N}_t^t)$, we have the following inequality holds:

When $\lambda > 1$, we have $Reg(T, E_{t,1}) \leq \sum_{i\in[m]}\frac{2c_1^2B_v^2}{\Delta_i^{\min}} + \sum_{i\in[m]}\frac{48c_1^2B_v^2\log T}{\Delta_i^{\min}}(3+\log K) + \frac{m\pi^2}{6}\log_2\left(\frac{c_1^2B_v^2K}{\lambda(\Delta_{\min})^2}\right)\Delta_{\max}$.

When $\lambda = 1$, we have $Reg(T, E_{t,1}) \leq \sum_{i\in[m]}\frac{2c_1^2B_v^2}{\Delta_i^{\min}} + \sum_{i\in[m]}\log\frac{c_1^2B_v^2K}{(\Delta_i^{\min})^2}\frac{48c_1^2B_v^2\log T}{\Delta_i^{\min}}(3+\log K) + \frac{m\pi^2}{6}\log_2\left(\frac{c_1^2B_v^2K}{(\Delta_{\min})^2}\right)\Delta_{\max}$.

C.3.2 Upper bound for $Reg(T, E_{t,2})$

As usual, we first break $Reg(T, E_{t,2})$ into two parts and bound them separately: $Reg(T, E_{t,2} \bigcap \mathcal{N}_t^t)$ and $Reg(T, \neg \mathcal{N}_t^t)$.

For $Reg(T, E_{t,2} \bigcap \mathcal{N}_t^t)$, under the event $\mathcal{N}_t^t$, let $c_2 = 28$ be a constant and $K = \max_{S \in \mathcal{S}} |\tilde{S}|$. We set $j_i^{\max} = \lceil \log_2 \frac{4B_1 c_2 K}{\Delta_i^{\min}} \rceil + 1$. We first define a regret allocation function

$$\kappa_{i,j,T}(\ell) = \begin{cases} \Delta_i^{\max}, & \text{if } 0 \le \ell \le L_{i,j,T,1} \text{ and } j \le j_i^{\max} \\ \frac{24 c_2 B_1 \log T}{\ell}, & \text{if } L_{i,j,T,1} < \ell \le L_{i,j,T,2} \text{ and } j \le j_i^{\max} \\ 0, & \text{if } \ell > L_{i,j,T,2} + 1 \text{ or } j > j_i^{\max}, \end{cases} \tag{67}$$

where $L_{i,j,T,1} = \frac{24 c_2 B_1 \log T}{\Delta_i^{\max}}$, $L_{i,j,T,2} = \frac{24 c_2 B_1 K \log T}{\Delta_i^{\min}}$.

Lemma 11. *For any time $t \in [T]$, if N_t^t and $E_{t,2}$ hold, we have*

$$\Delta_{S_t} \le \sum_{i \in \tilde{S}_t} \kappa_{i, j_i^{S_t}, T}(N_{t-1,i,j_i^{S_t}}), \tag{68}$$

where $j_i^{S_t}$ is the index of the triggering group $S_{i,j}$ such that $2^{-j_i^{S_t}} < p_i^{D,S_t} \le 2^{-j_i^{S_t}+1}$.

Proof. By event $E_{t,2}$, we have

$$\begin{aligned}
\Delta_{S_t} &\le \sum_{i \in \tilde{S}_t} 2 c_2 B_1 p_i^{D,S_t} \min\{\frac{\log t}{T_{i,t-1}}, \frac{1}{28}\} &(69)\\
&\le -\Delta_{S_t} + 2 \sum_{i \in \tilde{S}_t} 2 c_2 B_1 p_i^{D,S_t} \min\{\frac{\log t}{T_{i,t-1}}, \frac{1}{28}\} &(70)\\
&\le \sum_{i \in \tilde{S}_t} \left(4 c_2 B_1 p_i^{D,S_t} \min\{\frac{\log t}{T_{i,t-1}}, \frac{1}{28}\} - \frac{\Delta_{S_t}}{|\tilde{S}_t|} \right) &(71)\\
&\le \sum_{i \in \tilde{S}_t} \left(4 c_2 B_1 p_i^{D,S_t} \min\{\frac{\log t}{\frac{1}{3} N_{t-1,i,j_i^{S_t}} 2^{-j_i^{S_t}}}, \frac{1}{28}\} - \frac{\Delta_{S_t}}{|\tilde{S}_t|} \right) &(72)\\
&\le \sum_{i \in \tilde{S}_t} \underbrace{\left(4 c_2 B_1 2^{-j_i^{S_t}+1} \min\{\frac{\log t}{\frac{1}{3} N_{t-1,i,j_i^{S_t}} 2^{-j_i^{S_t}}}, \frac{1}{28}\} - \frac{\Delta_{S_t}}{K} \right)}_{(73,i)}, &(73)
\end{aligned}$$

where Eq. (69) is by the definition of $E_{t,1}$ which says $\Delta_{S_t} \le \sum_{i \in \tilde{S}_t} 2 c_2 B_1 p_i^{D,S_t} \min\{\frac{\log t}{T_{i,t-1}}, \frac{1}{28}\}$ and by dividing both sides by $\Delta_{S_t} > 0$, Eq. (70) is because we double the LHS and RHS of Eq. (69) at the same time and then put one into the RHS, Eq. (71) is by putting $-\Delta_{S_t}$ inside the summation, Eq. (72) is due to the same reason of Eq. (76) under event $\mathcal{N}_t^t$, Eq. (73) is due to $p_i^{D,S_t} \le 2^{-j_i^{S_t}+1}$ given by the definition of $j_i^{S_t}$ and $|\tilde{S}| \le K$.

Similar to Eq. (73), Eq. (70) is called the reverse amortization. Now we bound (73, i) under different cases.

When $j > j_i^{\max}$,

we have $(73, i) \le 4 c_2 B_1 2^{-j_i^{S_t}+1} - \frac{\Delta_{S_t}}{K} \le 4 c_2 B_1 \frac{\Delta_i^{\min}}{c_2 B_1 K} - \frac{\Delta_{S_t}}{K} \le \frac{\Delta_i^{\min}}{K} \frac{4}{28} - \frac{\Delta_{S_t}}{K} \le 0 = \kappa_{i, j_i^{S_t}, T}(N_{t-1,i,j_i^{S_t}})$.

When $N_{t-1,i,j_i^{S_t}} > L_{i,j_i^{S_t},T,2}$,

we have $(73, i) \le 4 c_2 B_1 2^{-j_i^{S_t}+1} \frac{\log t}{\frac{1}{3} N_{t-1,i,j_i^{S_t}} 2^{-j_i^{S_t}}} - \frac{\Delta_{S_t}}{K} \le \frac{24 c_2 B_1 \log T}{N_{t-1,i,j_i^{S_t}}} - \frac{\Delta_{S_t}}{K} < \frac{\Delta_i^{\min}}{K} - \frac{\Delta_{S_t}}{K} \le 0 = \kappa_{i, j_i^{S_t}, T}(N_{t-1,i,j_i^{S_t}})$.

When $N_{t-1,i,j_i^{S_t}} \le L_{i,j_i^{S_t},T,2}$ and $j \le j_i^{\max}$,

We have $(73, i) \le 4c_2B_1 2^{-j_i^{S_t}+1} \frac{\log t}{\frac{1}{3}N_{t-1,i,j_i^{S_t}} 2^{-j_i^{S_t}}} - \frac{\Delta_{S_t}}{K} = \frac{24c_2B_1 \log T}{N_{t-1,i,j_i^{S_t}}} - \frac{\Delta_{S_t}}{K} < \frac{24c_2B_1 \log T}{N_{t-1,i,j_i^{S_t}}} = \kappa_{i,j_i^{S_t},T}(N_{t-1,i,j_i^{S_t}})$.

When $N_{t-1,i,j_i^{S_t}} \le L_{i,j_i^{S_t},T,1}$ and $j \le j_i^{\max}$,

If there exists $i \in \tilde{S}_t$ so that $N_{t-1,i,j_i^{S_t}} \le L_{i,j_i^{S_t},T,1} =$, then we know $\sum_{q\in\tilde{S}_t} \kappa_{i,j_q^{S_t},T}(N_{t-1,q,j_q^{S_t}}) \ge \kappa_{i,j_i^{S_t},T}(N_{t-1,i,j_i^{S_t}}) = \Delta_i^{\max} \ge \Delta_{S_t}$, which makes Eq. (68) holds no matter what. This means we do not need to consider this case for good.

Combining all above cases, we have $\Delta_{S_t} \le \sum_{i\in\tilde{S}_t} \kappa_{i,j_i^{S_t},T}(N_{t-1,i,j_i^{S_t}})$. ■

Since $N_{t,i,j_i^{S_t}}$ is increased if and only if $i \in \tilde{S}_t$ and consider all possible i, j_i^S and N_{t,i,j_i^S} where $\kappa_{i,j_i^S,T}(N_{t-1,i,j^S}) > 0$, we have

$$
\begin{aligned}
& Reg(T, E_{t,2} \bigcap N_t^t) \\
& \le \sum_{t\in[T]} \sum_{i\in\tilde{S}_t} \kappa_{i,j_i^{S_t},T}(N_{t-1,i,j^{S_t}}) \\
& \le \sum_{i\in[m]} \sum_{j=0}^{j_i^{\max}} \sum_{\ell=1}^{L_{i,j,T,1}} \Delta_i^{\max} + \sum_{i\in[m]} \sum_{j=1}^{j_i^{\max}} \sum_{L_{i,j,T,1}+1}^{L_{i,j,T,2}} \frac{24c_2B_1 \log T}{\ell} \\
& \le \sum_{i\in[m]} \sum_{j=1}^{j_i^{\max}} 24c_2B_1 \log T + \sum_{i\in[m]} \sum_{j=1}^{j_i^{\max}} 24c_2B_1 \log(\frac{K\Delta_i^{\max}}{\Delta_i^{\min}}) \log T \\
& = \sum_{i\in[m]} \sum_{j=1}^{j_i^{\max}} 24c_2B_1 \left(1 + \log(\frac{K\Delta_i^{\max}}{\Delta_i^{\min}})\right) \log T \\
& \le \sum_{i\in[m]} 24c_2B_1 \left(\log_2 \frac{B_1c_2K}{\Delta_i^{\min}}\right) \left(1 + \log(\frac{K\Delta_i^{\max}}{\Delta_i^{\min}})\right) \log T
\end{aligned}
$$

Similar to Eq. (96), with additional $Reg(T, \neg\mathcal{N}_t^t)$, we have

We have $Reg(T, E_{t,2}) \le \sum_{i\in[m]} 24c_2B_1 \left(\log_2 \frac{B_1c_2K}{\Delta_i^{\min}}\right) \left(1 + \log(\frac{K\Delta_i^{\max}}{\Delta_i^{\min}})\right) \log T + \frac{m\pi^2}{6} \log_2 \frac{4B_1c_2K}{\Delta_i^{\min}} \Delta_{\max}$.

C.4 The Proof Following [23] Using Infinitely Many Events With an Additional Factor of $O(\log K)$

Now we can separately bound these two event-filtered regrets. Recall that $\tilde{S}_t = \{i \in [m] : p_i^{D,S_t} > 0\}$ is the set of arms that could be triggered in round t. Let $K = \max_{t\in[T]} |\tilde{S}_t|$ be the maximum number of base arms that can be triggered in any rounds. In round t, given base arm i and action S_t, we denote $j_i^{S_t}$ to be the corresponding index of the triggering group $\mathcal{S}_{i,j}$ so that $2^{-j_i^{S_t}} < p_i^{D,S_t} \le 2^{-j_i^{S_t}+1}$. Our strategy is to find (perhaps infinitely many) events that must happen when $E_{t,1}$ (or $E_{t,2}$) happens. Then we can show that the number of times these events can happen are bounded or otherwise $E_{t,1}$ (or $E_{t,2}$) will not hold anymore.

C.4.1 Upper bound for $Reg(T, E_{t,1})$

To upper bound $Reg(T, E_{t,1})$, we bound it by $Reg(T, E_{t,1}) \le Reg(T, E_{t,1} \bigcap \mathcal{N}_t^t) + Reg(T, \neg\mathcal{N}_t^t)$. In the following, we will consider to first bound $Reg(T, E_{t,1} \bigcap N_t^t)$ and then $Reg(T, \neg\mathcal{N}_t^t)$.

Recall that $e_{t,1}(S_t) = 4\sqrt{3}B_v\sqrt{\sum_{i\in\tilde{S}_t}(\frac{\log t}{T_{t-1,i}} \wedge \frac{1}{28})(p_i^{D,S_t})^{\lambda}}$. Let $c_1 = 4\sqrt{3}$ be a constant and $O_t = \{i \in \tilde{S}_t : j_i^{S_t} \le j_i^{\max}\}$ be the set of base arms whose triggering probabilities are not too small, where the threshold $j_i^{\max} = \frac{1}{\lambda}(\lceil \log_2 \frac{c_1^2 B_v^2 K}{(\Delta_i^{\min})^2} \rceil + 1)$. Let $\alpha_1 > \alpha_2 > ... > \alpha_k > ... > \alpha_\infty$ and $1 = \beta_0 > \beta_1 > ... > \beta_k > ... > \beta_\infty$ be two infinite sequences of positive numbers that are decreasing and converge to 0, which will be used later to define specific set of base arms $(A_{t,k})_{k=1}^\infty$ and events $(\mathbb{G}_{t,k})_{k=1}^\infty$.

For positive integers k and t, we define $A_{t,k} = \{i \in \tilde{S}_t : N_{t-1,i,j_i^{S_t}} \leq \alpha_k \frac{g(K,\Delta_{S_t})f(t)}{\Delta_{S_t}^2}, j_i^{S_t} \leq j_i^{\max}\} = \{i \in \tilde{S}_t \cap O_{S_t} : N_{t-1,i,j_i^{S_t}} \leq \alpha_k \frac{g(K,\Delta_{S_t})f(t)}{\Delta_{S_t}^2}\}$, which is the set of arms in $\tilde{S}_t$ that are counted less that a threshold and whose triggering probabilities are not too small, where $g(K, \Delta_{S_t})$ and $f(t)$ are going to be tuned for later use. Moreover, we define the complementary set $\bar{A}_{t,k} = \{i \in \tilde{S}_t \cap O_t : N_{t-1,i,j_i^{S_t}} > \alpha_k \frac{g(K,\Delta_{S_t})f(t)}{\Delta_{S_t}^2}\}$.

Now we are ready to define the events $\mathbb{G}_{t,k} = \{|A_{t,k}| \geq \beta_k K; \forall h < k, |A_{t,h}| < \beta_h K\}$. Note that $\mathbb{G}_{t,k}$ is true when at least $\beta_k K$ arms triggered are in the set $A_{t,k}$ but less than $\beta_h K$ arms triggered are in the set $A_{t,h}$ for $h < k$. Let $\mathbb{G}_t = \bigcup_{k=1}^\infty \mathbb{G}_{t,k}$ and by definition its complementary $\overline{\mathbb{G}}_t = \{|A_{t,k}| < \beta_k K, \forall k \geq 1\}$. We first introduce a lemma saying that if there exists $k_0 > 0$ such that β_{k_0} is smaller than $1/K$, we can safely use finite many events to conclude infinitely many events.

Lemma 12. *If there exists k_0 such that $\beta_{k_0} \leq 1/K$, then $\mathbb{G}_t = \bigcup_{k=1}^{k_0} \mathbb{G}_{t,k}$ and $\overline{\mathbb{G}}_t = \{|A_{t,k}| < \beta_k K, \forall 1 \leq k \leq k_0\}$.*

Proof. Let k_0 such that $\beta_{k_0} \leq 1/K$. THen for all $k > k_0$, $\mathbb{G}_{t,k} = \{|A_{t,k}| \geq 1; \forall h < k_0, |A_{t,h}| \leq K\beta_k; \forall k_0 \leq h < k, |A_{t,h}| = 0\}$. But as the sequence of sets $A_{t,k}$ is decreasing, $\{|A_{t,k_0}| = 0\}$ and $\{|A_{t,k}| \geq 1\}$ cannot happen at the same time. Thus, $|A_{t,k}|$ cannot happen for $k > k_0$. ∎

Now we have the following lemma showing an upper bound of $e_{t,2}(S_t)$ when $\overline{\mathbb{G}}_t$ and N_t^t happens.

Lemma 13. *Under the event $\overline{\mathbb{G}}_t$ and N_t^t and if $\exists\, k_0$ such that $\beta_{k_0} \leq 1/K$, then*

$$(e_{t,1}(S_t))^2 < \frac{6c_1^2 B_v^2 2^{(-j_i^{S_t}+1)(\lambda-1)} \log t \Delta_{S_t}^2 K}{g(K,\Delta_{S_t})f(t)} (\sum_{k=1}^{k_0} \frac{\beta_{k-1}-\beta_k}{\alpha_k} + \frac{\beta_{k_0}}{\alpha_{k_0}}) + \frac{\Delta_{S_t}^2}{8} \tag{74}$$

Proof.

$$\begin{aligned}
(e_{t,1}(S_t))^2 &= \sum_{i \in \tilde{S}_t} c_1^2 B_v^2 (p_i^{D,S_t})^\lambda \min\{\frac{\log t}{T_{i,t-1}}, \frac{1}{28}\} && (75)\\
&\leq \sum_{i \in \tilde{S}_t} c_1^2 B_v^2 (p_i^{D,S_t})^\lambda \min\{\frac{\log t}{\frac{1}{3} N_{t-1,i,j_i^{S_t}} 2^{-j_i^{S_t}}}, \frac{1}{28}\} && (76)\\
&\leq \sum_{i \in \tilde{S}_t \cap O_{S_t}} c_1^2 B_v^2 (2^{-j_i^{S_t}+1})^\lambda \frac{\log t}{\frac{1}{3} N_{t-1,i,j_i^{S_t}} 2^{-j_i^{S_t}}} + \frac{1}{28} \sum_{i \in \tilde{S}_t \cap \bar{O}_{S_t}} c_1^2 B_v^2 (2^{-j_i^{\max}+1})^\lambda && (77)\\
&\leq \sum_{i \in \tilde{S}_t \cap O_{S_t}} \frac{6c_1^2 B_v^2 2^{(-j_i^{S_t}+1)(\lambda-1)} \log t}{N_{t-1,i,j_i^{S_t}}} + \sum_{i \in \tilde{S}_t \cap \bar{O}_{S_t}} B_v^2 c_1^2 \frac{(\Delta_i^{\min})^2}{8 B_v^2 c_1^2 K} && (78)\\
&\leq \sum_{k=1}^{k_0} \sum_{i \in \bar{A}_{t,k} \setminus \bar{A}_{t,k-1}} \frac{6c_1^2 B_v^2 2^{(-j_i^{S_t}+1)(\lambda-1)} \log t}{N_{t-1,i,j_i^{S_t}}} + \frac{\Delta_{S_t}^2}{8} && (79)\\
&< \sum_{k=1}^{k_0} \frac{6c_1^2 B_v^2 2^{(-j_i^{S_t}+1)(\lambda-1)} \log t \Delta_{S_t}^2 |\bar{A}_{t,k} \setminus \bar{A}_{t,k-1}|}{\alpha_k g(K,\Delta_{S_t}) f(t)} + \frac{\Delta_{S_t}^2}{8} && (80)\\
&< \frac{6c_1^2 B_v^2 2^{(-j_i^{S_t}+1)(\lambda-1)} \log t \Delta_{S_t}^2 K}{g(K,\Delta_{S_t}) f(t)} (\sum_{k=1}^{k_0} \frac{\beta_{k-1}-\beta_k}{\alpha_k} + \frac{\beta_{k_0}}{\alpha_{k_0}}) + \frac{\Delta_{S_t}^2}{8} && (81)
\end{aligned}$$

where Equation (75) is by definition, Equation (76) holds because if $\frac{\ln t}{\frac{1}{3}N_{i,j,t-1}2^{-j}} > 1/28$, then $\min\{\frac{\log t}{\frac{1}{3}N_{t-1,i,j_i^{S_t}}2^{-j_i^{S_t}}}, 1/28\} = 1/28$ and thus larger than $\min\{\frac{\log t}{T_{t-1,i}}, 1/28\}$, else we have $\frac{6\log t}{\frac{1}{3}N_{t-1,i,j_i^{S_t}}2^{-j_i^{S_t}}} < 6/28 < 1$ and by $\mathcal{N}_t^t$ we have $T_{t-1,i} \geq \frac{1}{3}N_{t-1,i,j^{S_t}} \cdot 2^{-j^{S_t}}$ and thus $\min\{\frac{\log t}{T_{t-1,i}}, 1/28\} < \frac{\log t}{\frac{1}{3}N_{t-1,i,j_i^{S_t}}2^{-j_i^{S_t}}}$, Eq. (77) is by considering O_t and $\bar{O}_t$, Equation (78) is due to definition of $j_i^{\max}$, Equation (79) is by setting k_0 be the largest number that $\beta_{k_0} \leq 1/K$, Equation (80) is by definition of $\bar{A}_{t,k}$, Equation (81) is due to the similar reason of Lemma 8 from [11]. ■

Now we set $g(K, \Delta_{S_t}) = 2^{(-j_i^{S_t}+1)(\lambda-1)}Kl$, where $l = \sum_{k=1}^{k_0}\frac{\beta_{k-1}-\beta_k}{\alpha_k} + \frac{\beta_{k_0}}{\alpha_{k_0}}$ and $f(t) = 48c_1^2B_v^2\log t$. By Lemma 13, we can show that $\Delta_{S_t} > 2e_{t,1}$ under event $\overline{\mathbb{G}}_t \bigcap \mathcal{N}_t^t$. In other words, under event $\mathcal{N}_t^t$, if $E_{t,1}$ holds, then $\mathbb{G}_t$ must hold.

For any arm i, let arm related event $\mathbb{G}_{t,k,i} = \mathbb{G}_{t,k} \bigcap \{i \in \tilde{S}_t, N_{t-1,i,j_i^{S_t}} \leq \alpha_k\frac{g(K,\Delta_{S_t})f(t)}{\Delta_{S_t}^2}, j_i^{S_t} \leq j_i^{\max}\}$. When $\mathbb{G}_{t,k}$ happens, we have $\mathbb{I}\{\mathbb{G}_{t,k}\} \leq \frac{1}{\beta_k K}\sum_{i\in[m]}\{\mathbb{G}_{t,k,i}\}$. We consider two cases when $\lambda > 1$ and when $\lambda = 1$.

Case 1: When $\lambda > 1$

The $Reg(T, E_{t,1}\bigcap\mathcal{N}_t^t)$ is bounded by,

$$Reg(T, E_{t,1}\bigcap\mathcal{N}_t^t) \leq \sum_{t=1}^{T}\sum_{k=1}^{k_0}\Delta_{S_t}\mathbb{I}\{\mathbb{G}_{t,k}\} \tag{82}$$

$$\leq \sum_{t=1}^{T}\sum_{k=1}^{k_0}\sum_{i=1}^{m}\frac{\Delta_{S_t}}{K\beta_k}\mathbb{I}\{\mathbb{G}_{t,k,i}\} \tag{83}$$

$$\leq \sum_{i=1}^{m}\sum_{k=1}^{k_0}\frac{1}{K\beta_k}\sum_{t=1}^{T}\Delta_{S_t}\mathbb{I}\{i \in \tilde{S}_t, N_{i,j_i^{S_t},t-1} \leq \frac{\theta_k 2^{(-j_i^{S_t}+1)(\lambda-1)}}{\Delta_{S_t}^2}, j_i^{S_t} \leq j_i^{\max}\}$$

(with $\theta_k = \alpha_k K l f(t)$)

$$\leq \sum_{i=1}^{m}\sum_{j=1}^{\infty}\sum_{k=1}^{k_0}\frac{1}{K\beta_k}\sum_{t=1}^{T}\Delta_{S_t}\mathbb{I}\{i \in \tilde{S}_t, N_{i,j_i^{S_t},t-1} \leq \frac{\theta_k 2^{(-j+1)(\lambda-1)}}{\Delta_{S_t}^2}, j_i^{S_t} = j\} \tag{84}$$

$$\leq \sum_{i=1}^{m}\sum_{j=1}^{\infty}\sum_{k=1}^{k_0}\frac{1}{K\beta_k}\sum_{t=1}^{T}\sum_{n=1}^{D_i}\Delta_{i,n}\mathbb{I}\{i \in \tilde{S}_t, N_{i,j_i^{S_t},t-1} \leq \frac{\theta_k 2^{(-j+1)(\lambda-1)}}{\Delta_{i,n}^2}, \Delta_{S_t} = \Delta_{i,n}, j_i^{S_t} = j\}$$

(with $\Delta_{i,1} \geq \Delta_{i,2} \geq ... \geq \Delta_{i,D_i}$) (85)

$$\leq \sum_{i=1}^{m}\sum_{j=1}^{\infty}\sum_{k=1}^{k_0}\frac{1}{K\beta_k}\sum_{t=1}^{T}\sum_{n=1}^{D_i}\sum_{p=1}^{n}\Delta_{i,p}\mathbb{I}\{i \in \tilde{S}_t, N_{i,j_i^{S_t},t-1} \in (\frac{\theta_k 2^{(-j+1)(\lambda-1)}}{\Delta_{i,p-1}^2}, \frac{\theta_k 2^{(-j+1)(\lambda-1)}}{\Delta_{i,p}^2}],$$

$$\Delta_{S_t} = \Delta_{i,n}, j_i^{S_t} = j\} \tag{86}$$

$$\leq \sum_{i=1}^{m}\sum_{j=1}^{\infty}\sum_{k=1}^{k_0}\frac{1}{K\beta_k}\sum_{t=1}^{T}\sum_{n=1}^{D_i}\sum_{p=1}^{D_i}\Delta_{i,p}\mathbb{I}\{i \in \tilde{S}_t, N_{i,j_i^{S_t},t-1} \in (\frac{\theta_k 2^{(-j+1)(\lambda-1)}}{\Delta_{i,p-1}^2}, \frac{\theta_k 2^{(-j+1)(\lambda-1)}}{\Delta_{i,p}^2}],$$

$$\Delta_{S_t} = \Delta_{i,n}, j_i^{S_t} = j\} \tag{87}$$

$$\leq \sum_{i=1}^{m}\sum_{j=1}^{\infty}\sum_{k=1}^{k_0}\frac{1}{K\beta_k}\sum_{t=1}^{T}\sum_{p=1}^{D_i}\Delta_{i,p}\mathbb{I}\{i \in \tilde{S}_t, N_{i,j_i^{S_t},t-1} \in (\frac{\theta_k 2^{(-j+1)(\lambda-1)}}{\Delta_{i,p-1}^2}, \frac{\theta_k 2^{(-j+1)(\lambda-1)}}{\Delta_{i,p}^2}]$$

$$, \Delta_{S_t} = \Delta_{i,n}, \Delta_{S_t} > 0, j_i^{S_t} = j\} \tag{88}$$

$$\leq \sum_{i=1}^{m}\sum_{j=1}^{\infty}\sum_{k=1}^{k_0}\frac{2^{(-j+1)(\lambda-1)}}{K\beta_k}(\frac{\theta_k}{\Delta_{i,1}}+\theta_k\sum_{p=2}^{D_i}\Delta_{i,p}(\frac{1}{\Delta_{i,p}^2}-\frac{1}{\Delta_{i,p-1}^2})) \tag{89}$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{\infty}\sum_{k=1}^{k_0}\frac{2^{(-j+1)(\lambda-1)}}{K\beta_k}(\frac{\theta_k}{\Delta_{i,D_i}}+\theta_k\sum_{p=1}^{D_i-1}\frac{\Delta_{i,p}-\Delta_{i,p+1}}{\Delta_{i,p}^2}) \tag{90}$$

$$\leq \sum_{i=1}^{m}\sum_{j=1}^{\infty}\sum_{k=1}^{k_0}\frac{2^{(-j+1)(\lambda-1)}}{K\beta_k}(\frac{\theta_k}{\Delta_{i,D_i}}+\theta_k\int_{\Delta_{i,D_i}}^{\Delta_{i,1}}x^{-2}dx) \tag{91}$$

$$\leq \sum_{i=1}^{m}\sum_{j=1}^{\infty}\sum_{k=1}^{k_0}\frac{2\theta_k 2^{(-j+1)(\lambda-1)}}{K\beta_k\Delta_{i,D_i}} \tag{92}$$

$$\leq \sum_{i=1}^{m}(\sum_{k=1}^{k_0}\frac{2}{1-2^{-(\lambda-1)}}\frac{1}{\Delta_i^{\min}}\frac{\alpha_k}{\beta_k}lf(T)) \tag{93}$$

$$\leq \sum_{i=1}^{m}\left(\frac{480c_1^2B_v^2}{(1-2^{-(\lambda-1)})}\right)\lceil\frac{\log K}{1.61}\rceil^2\frac{\log T}{\Delta_i^{\min}} \tag{94}$$

where Eq. (82) is because under event $\mathcal{N}_t^t$, if $E_{t,1}$ holds then $\mathbb{G}_t$ must hold, Eq. (83) is because $\mathbb{I}\{\mathbb{G}_{t,k}\} \leq \frac{1}{\beta_k K}\sum_{i\in[m]}\{\mathbb{G}_{t,k,i}\}$, Eq. (84) is by applying union bound over $j_i^{S_t} = 1,...,j_i^{\max}$, Eq. (85) is by considering D_i gaps for Δ_{S_t} and applying union bounds, Eq. (86) is by dividing $N_{i,j_i^{S_t},t-1} \leq \frac{\theta_k 2^{(-j+1)(\lambda-1)}}{\Delta_{i,n}^2}$ into non-overlapping sub-intervals, Eq. (87) is by extending summation over p to D_i, Eq. (88) is by replacing summation over $n = 1,...,D_i$ to $\Delta S_t > 0$, Eq. (89) is to bound the number of times the event happen to the length of interval, Eq. (90) to Eq. (93) are math calculation by replacing summation by integrals, Eq. (94) is similar to [11, Lemma 11, Appendix C] by setting $\alpha_k = \beta_k = 0.2^k$ and $\sum_{k=1}^{k0}\frac{\alpha_k}{\beta_k}l \leq 5\lceil\frac{\log K}{1.61}\rceil^2$.

Case 2: When $\lambda = 1$ The only difference is we have to sum over $j = 1$ to $j = j_{\max}^i$ in Eq. (92), instead of ∞, so that we replace $\frac{1}{(1-2^{-(\lambda-1)})}$ to $j_{\max}^i = \log_2\left(\frac{c_1^2B_v^2K}{(\Delta_i^{\min})^2}\right)$. we can bound the following inequality

$$Reg(T, E_{t,1}\wedge\mathcal{N}_t^t) \leq \sum_{i=1}^{m}\left(480c_1^2B_v^2\right)\log_2\left(\frac{c_1^2B_v^2K}{(\Delta_i^{\min})^2}\right)\lceil\frac{\log K}{1.61}\rceil^2\frac{\log T}{\Delta_i^{\min}} \tag{95}$$

Now for $Reg(T, \neg\mathcal{N}_t^t)$, by Lemma 6,

$$Reg(T, \neg\mathcal{N}_t^t) \leq \sum_{t=1}^{T}\sum_{i\in[m]}j_i^{\max}t^{-2}\Delta_{\max} \tag{96}$$

$$\leq \frac{m\pi^2}{6}\log_2\left(\frac{c_1^2B_v^2K}{(\Delta_{\min})^2}\right)\Delta_{\max} \tag{97}$$

So we have when $\lambda > 1$,

$$Reg(T, E_{t,1}) \leq \sum_{i=1}^{m}\left(\frac{480c_1^2B_v^2}{(1-2^{-(\lambda-1)})}\right)\lceil\frac{\log K}{1.61}\rceil^2\frac{\log T}{\Delta_i^{\min}}+\frac{m\pi^2}{6}\log_2\left(\frac{c_1^2B_v^2K}{\lambda(\Delta_{\min})^2}\right)\Delta_{\max}; \tag{98}$$

and when $\lambda = 1$,

$$Reg(T, E_{t,1}) \leq \sum_{i=1}^{m}\left(480c_1^2B_v^2\right)\log_2\left(\frac{c_1^2B_v^2K}{(\Delta_i^{\min})^2}\right)\lceil\frac{\log K}{1.61}\rceil^2\frac{\log T}{\Delta_i^{\min}}+\frac{m\pi^2}{6}\log_2\left(\frac{c_1^2B_v^2K}{(\Delta_{\min})^2}\right)\Delta_{\max} \tag{99}$$

C.4.2 Upper bound for $Reg(T, E_{t,2})$

Let $c_2 = 28$ be a constant and $O_t = \{i \in \tilde{S}_t : j_i^{S_t} \leq j_i^{\max}\}$ be the set of base arms whose triggering probabilities are not too small, where the threshold $j_i^{\max} = \lceil\log_2\frac{4B_1c_2K}{\Delta_i^{\min}}\rceil + 1$. Let

$\alpha_1 > \alpha_2 > ... > \alpha_k > ... > \alpha_\infty$ and $1 = \beta_0 > \beta_1 > ... > \beta_k > ... > \beta_\infty$ be two infinite sequences of positive numbers that are decreasing and converge to 0, which will be used later to define specific set of base arms $(A_{t,k})_{k=1}^{\infty}$ and events $(\mathbb{G}_{t,k})_{k=1}^{\infty}$.

For positive integers k and t, we define $A_{t,k} = \{i \in \tilde{S}_t : N_{t-1,i,j_i^{S_t}} \leq \alpha_k \frac{g(K,\Delta_{S_t})f(t)}{\Delta_{S_t}^2}, j_i^{S_t} \leq j_i^{\max}\} = \{i \in \tilde{S}_t \cap O_{S_t} : N_{t-1,i,j_i^{S_t}} \leq \alpha_k \frac{g(K,\Delta_{S_t})f(t)}{\Delta_{S_t}^2}\}$, which is the set of arms in $\tilde{S}_t$ that are counted less that a threshold and whose triggering probabilities are not too small, where $g(K,\Delta_{S_t})$ and $f(t)$ are going to be tuned for later use. Moreover, we define the complementary set $\bar{A}_{t,k} = \{i \in \tilde{S}_t \cap O_t : N_{t-1,i,j_i^{S_t}} > \alpha_k \frac{g(K,\Delta_{S_t})f(t)}{\Delta_{S_t}^2}\}$.

Now we are ready to define the events $\mathbb{G}_{t,k} = \{|A_{t,k}| \geq \beta_k K; \forall h < k, |A_{t,h}| < \beta_h K\}$. Note that $\mathbb{G}_{t,k}$ is true when at least $\beta_k K$ arms triggered are in the set $A_{t,k}$ but less than $\beta_h K$ arms triggered are in the set $A_{t,h}$ for $h < k$. Let $\mathbb{G}_t = \bigcup_{k=1}^{\infty} \mathbb{G}_{t,k}$ and by definition its complementary $\overline{\mathbb{G}}_t = \{|A_{t,k}| < \beta_k K, \forall k \geq 1\}$. We first introduce a lemma saying that if there exists $k_0 > 0$ such that β_{k_0} is smaller than $1/K$, we can safely use finite many events to conclude infinitely many events.

Lemma 14. *If there exists k_0 such that $\beta_{k_0} \leq 1/K$, then $\mathbb{G}_t = \bigcup_{k=1}^{k_0} \mathbb{G}_{t,k}$ and $\overline{\mathbb{G}}_t = \{|A_{t,k}| < \beta_k K, \forall 1 \leq k \leq k_0\}$.*

Proof. By the same argument as Lemma 12, the lemma is proved. ■

Now we have the following lemma showing an upper bound of $e_{t,2}(S_t)$ when $\overline{\mathbb{G}}_t$ and N_t^t happens.

Lemma 15. *Under the event $\overline{\mathbb{G}}_t$ and N_t^t and if $\exists\, k_0$ such that $\beta_{k_0} \leq 1/K$, then*

$$e_{t,2}(S_t) < \frac{6c_2 B_1 \log t \Delta_{S_t}^2 K}{g(K,\Delta_{S_t})f(t)} (\sum_{k=1}^{k_0} \frac{\beta_{k-1} - \beta_k}{\alpha_k} + \frac{\beta_{k_0}}{\alpha_{k_0}}) + \frac{\Delta_{S_t}}{4} \tag{100}$$

Proof.

$$\begin{aligned}
e_{t,2}(S_t) &= \sum_{i \in \tilde{S}_t} c_2 B_1 p_i^{D,S_t} \min\{\frac{\log t}{T_{t-1,i}}, \frac{1}{28}\} && (101)\\
&\leq \sum_{i \in \tilde{S}_t} c_2 B_1 p_i^{D,S_t} \min\{\frac{\log t}{\frac{1}{3} N_{t-1,i,j_i^{S_t}} 2^{-j_i^{S_t}}}, \frac{1}{28}\} && (102)\\
&\leq \sum_{i \in \tilde{S}_t \cap O_t} c_2 B_1 2^{-j_i^{S_t}+1} \frac{\log t}{\frac{1}{3} N_{t-1,i,j_i^{S_t}} 2^{-j_i^{S_t}}} + \frac{1}{28} \sum_{i \in \tilde{S}_t \cap \bar{O}_t} B_1 c_2 2^{-j_i^{\max}+1} && (103)\\
&\leq \sum_{i \in \tilde{S}_t \cap O_t} \frac{6c_2 B_1 \log t}{N_{t-1,i,j_i^{S_t}}} + \sum_{i \in \tilde{S}_t \cap \bar{O}_t} B_1 c_2 \frac{\Delta_i^{\min}}{4 B_1 c_2 K} && (104)\\
&\leq \sum_{k=1}^{k_0} \sum_{i \in \bar{A}_{t,k} \setminus \bar{A}_{t,k-1}} \frac{6c_2 B_1 \log t}{N_{t-1,i,j_i^{S_t}}} + \frac{\Delta_{S_t}}{4} && (105)\\
&< \sum_{k=1}^{k_0} \frac{6c_2 B_1 \log t \Delta_{S_t}^2 |\bar{A}_{t,k} \setminus \bar{A}_{t,k-1}|}{\alpha_k g(K,\Delta_{S_t}) f(t)} + \frac{\Delta_{S_t}}{4} && (106)\\
&< \frac{6c_2 B_1 \log t \Delta_{S_t}^2 K}{g(K,\Delta_{S_t})f(t)} (\sum_{k=1}^{k_0} \frac{\beta_{k-1} - \beta_k}{\alpha_k} + \frac{\beta_{k_0}}{\alpha_{k_0}}) + \frac{\Delta_{S_t}}{4} && (107)
\end{aligned}$$

where Equation (101) is by definition, Equation (102) holds because if $\frac{6 \ln t}{\frac{1}{3} N_{i,j,t-1} 2^{-j}} > \frac{1}{28}$, then $\min\{\frac{6 \log t}{\frac{1}{3} N_{t-1,i,j_i^{S_t}} 2^{-j_i^{S_t}}}, 1\} = \frac{1}{28}$ and thus larger than $\min\{\frac{\log t}{T_{t-1,i}}, 1\}$, else we have $\frac{6 \log t}{\frac{1}{3} N_{t-1,i,j_i^{S_t}} 2^{-j_i^{S_t}}} < 6/28 < 1$ and by $\mathcal{N}_t^t$ we have $T_{t-1,i} \geq \frac{1}{3} N_{t-1,i,j^{S_t}} \cdot 2^{-j^{S_t}}$ and thus

$\min\{\frac{\log t}{T_{t-1,i}}, 1/28\} < \frac{\log t}{\frac{1}{3}N_{t-1,i,j_i^{S_t}}2^{-j_i^{S_t}}}$, Eq. (103) is by considering O_t and $\bar{O}_t$, Equation (104) is due to definition of $j_i^{\max}$, Equation (105) is by setting k_0 be the largest number that $\beta_{k_0} \leq 1/K$, Equation (106) is by definition of $\bar{A}_{t,k}$, Equation (107) is due to the proof of Lemma 8 of [11]. ■

Now we set $g(K, \Delta_{S_t}) = K\Delta_{S_t}l$, where $l = \sum_{k=1}^{k_0} \frac{\beta_{k-1}-\beta_k}{\alpha_k} + \frac{\beta_{k_0}}{\alpha_{k_0}}$ and $f(t) = 24c_2B_1 \log T$. By Lemma 15, we can show that $\Delta_{S_t} > 2e_{t,2}$ under event $\overline{\mathbb{G}}_t \bigcap \mathcal{N}_t^t$. In other words, under event $\mathcal{N}_t^t$, if $E_{t,2}$ holds, then G_t must hold.

For any arm i, let arm related event $\mathbb{G}_{t,k,i} = \mathbb{G}_{t,k} \bigcap \{i \in \tilde{S}_t, N_{t-1,i,j_i^{S_t}} \leq \alpha_k \frac{g(K,\Delta_{S_t})f(t)}{\Delta_{S_t}^2}, j_i^{S_t} \leq j_i^{\max}\}$. When $\mathbb{G}_{t,k}$ happens, we have $\mathbb{I}\{\mathbb{G}_{t,k}\} \leq \frac{1}{\beta_k K}\sum_{i\in[m]}\{\mathbb{G}_{t,k,i}\}$. So the $Reg(T, E_{t,2} \bigcap \mathcal{N}_t^t)$ is bounded by,

$$Reg(T, E_{t,2} \bigcap \mathcal{N}_t^t) \leq \sum_{t=1}^{T}\sum_{k=1}^{k_0} \Delta_{S_t}\mathbb{I}\{\mathbb{G}_{t,k}\} \tag{108}$$

$$\leq \sum_{t=1}^{T}\sum_{k=1}^{k_0}\sum_{i=1}^{m} \frac{\Delta_{S_t}}{K\beta_k}\mathbb{I}\{G_{t,k,i}\} \tag{109}$$

$$\leq \sum_{i=1}^{m}\sum_{k=1}^{k_0} \frac{1}{K\beta_k}\sum_{t=1}^{T} \Delta_{S_t}\mathbb{I}\{i \in \tilde{S}_t, N_{t-1,i,j_i^{S_t}} \leq \frac{\theta_k}{\Delta_{S_t}}, j_i^{S_t} \leq j_i^{\max}\} \qquad (\text{with } \theta_k = \alpha_k K l f(t))$$

$$\leq \sum_{i=1}^{m}\sum_{j=1}^{j_i^{\max}}\sum_{k=1}^{k_0} \frac{1}{K\beta_k}\sum_{t=1}^{T} \Delta_{S_t}\mathbb{I}\{i \in \tilde{S}_t, N_{t-1,i,j_i^{S_t}} \leq \frac{\theta_k}{\Delta_{S_t}}, j_i^{S_t} = j\} \tag{110}$$

$$\leq \sum_{i=1}^{m}\sum_{j=1}^{j_i^{\max}}\sum_{k=1}^{k_0} \frac{1}{K\beta_k}\sum_{t=1}^{T}\sum_{n=1}^{D_i} \Delta_{i,n}\mathbb{I}\{i \in \tilde{S}_t, N_{t-1,i,j_i^{S_t}} \leq \frac{\theta_k}{\Delta_{i,n}}, \Delta_{S_t} = \Delta_{i,n}, j_i^{S_t} = j\}$$

$$(\text{with } \Delta_{i,1} \geq \Delta_{i,2} \geq ... \geq \Delta_{i,D_i}) \tag{111}$$

$$\leq \sum_{i=1}^{m}\sum_{j=1}^{j_i^{\max}}\sum_{k=1}^{k_0} \frac{1}{K\beta_k}\sum_{t=1}^{T}\sum_{n=1}^{D_i}\sum_{p=1}^{n} \Delta_{i,p}\mathbb{I}\{i \in \tilde{S}_t, N_{t-1,i,j_i^{S_t}} \in (\frac{\theta_k}{\Delta_{i,p-1}}, \frac{\theta_k}{\Delta_{i,p}}], \Delta_{S_t} = \Delta_{i,n}, j_i^{S_t} = j\} \tag{112}$$

$$\leq \sum_{i=1}^{m}\sum_{j=1}^{j_i^{\max}}\sum_{k=1}^{k_0} \frac{1}{K\beta_k}\sum_{t=1}^{T}\sum_{n=1}^{D_i}\sum_{p=1}^{D_i} \Delta_{i,p}\mathbb{I}\{i \in \tilde{S}_t, N_{t-1,i,j_i^{S_t}} \in (\frac{\theta_k}{\Delta_{i,p-1}}, \frac{\theta_k}{\Delta_{i,p}}], \Delta_{S_t} = \Delta_{i,n}, j_i^{S_t} = j\} \tag{113}$$

$$\leq \sum_{i=1}^{m}\sum_{j=1}^{j_i^{\max}}\sum_{k=1}^{k_0} \frac{1}{K\beta_k}\sum_{t=1}^{T}\sum_{p=1}^{D_i} \Delta_{i,p}\mathbb{I}\{i \in \tilde{S}_t, N_{t-1,i,j_i^{S_t}} \in (\frac{\theta_k}{\Delta_{i,p-1}}, \frac{\theta_k}{\Delta_{i,p}}], \Delta_{S_t} = \Delta_{i,n},$$

$$\Delta_{S_t} > 0, j_i^{S_t} = j\} \tag{114}$$

$$\leq \sum_{i=1}^{m}\sum_{j=1}^{j_i^{\max}}\sum_{k=1}^{k_0} \frac{1}{K\beta_k}(\theta_k + \theta_k\sum_{p=2}^{D_i} \Delta_{i,p}(\frac{1}{\Delta_{i,p}} - \frac{1}{\Delta_{i,p-1}})) \tag{115}$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{j_i^{\max}}\sum_{k=1}^{k_0} \frac{1}{K\beta_k}(\theta_k + \theta_k\sum_{p=1}^{D_i-1} \frac{\Delta_{i,p} - \Delta_{i,p+1}}{\Delta_{i,p}}) \tag{116}$$

$$\leq \sum_{i=1}^{m}\sum_{j=1}^{j_i^{\max}}\sum_{k=1}^{k_0} \frac{1}{K\beta_k}(\theta_k + \theta_k\int_{\Delta_{i,D_i}}^{\Delta_{i,1}} x^{-1}dx) \tag{117}$$

$$\leq \sum_{i=1}^{m}\sum_{j=1}^{j_i^{\max}}\sum_{k=1}^{k_0} \frac{\theta_k}{K\beta_k}(1 + \log\frac{\Delta_i^{\max}}{\Delta_i^{\min}}) \tag{118}$$

$$\leq \sum_{i=1}^{m}\sum_{j=1}^{j_i^{\max}}(\sum_{k=1}^{k_0}\frac{\alpha_k}{\beta_k}lf(T))(1+\log\frac{\Delta_i^{\max}}{\Delta_i^{\min}}) \tag{119}$$

$$\leq 120c_2B_1\sum_{i=1}^{m}\left(\log_2\frac{c_2B_1K}{\Delta_i^{\min}}\right)\left(1+\log\frac{\Delta_i^{\max}}{\Delta_i^{\min}}\right)\lceil\frac{\log K}{1.61}\rceil^2\log T \tag{120}$$

where Eq. (108) is because under event $\mathcal{N}_t^t$, if $E_{t,1}$ holds then $\mathbb{G}_t$ must hold, Eq. (109) is because $\mathbb{I}\{\mathbb{G}_{t,k}\} \leq \frac{1}{\beta_k K}\sum_{i\in[m]}\{\mathbb{G}_{t,k,i}\}$, Eq. (110) is by applying union bound over $j_i^{S_t} = 1,...,j_i^{\max}$, Eq. (111) is by considering D_i gaps for Δ_{S_t} and applying union bounds, Eq. (112) is by dividing $N_{i,j_i^{S_t},t-1} \leq \frac{\theta_k 2^{(-j+1)(\lambda-1)}}{\Delta_{i,n}^2}$ into non-overlapping sub-intervals, Eq. (113) is by extending summation over p to D_i, Eq. (114) is by replacing summation over $n=1,...,D_i$ to $\Delta S_t > 0$, Eq. (115) is to bound the number of times the event happen to the length of interval, Eq. (116) to Eq. (119) are math calculation by replacing summation by integrals, Eq. (120) is similar to [11, Lemma 11, Appendix C] by setting $\alpha_k = \beta_k = 0.2^k$ and $\sum_{k=1}^{k0}\frac{\alpha_k}{\beta_k}l \leq 5\lceil\frac{\log K}{1.61}\rceil^2$.

Similarly, consider $Reg(T, \neg\mathcal{N}_t^t) \leq \frac{m\pi^2}{6}\log_2\left(\frac{c_2B_1K}{\Delta_{\min}}\right)\Delta_{\max}$

We have

$$Reg(T, E_{t,2}) \leq 120c_2B_1\sum_{i=1}^{m}\left(\log_2\frac{c_2B_1K}{\Delta_i^{\min}}\right)\left(1+\log\frac{\Delta_i^{\max}}{\Delta_i^{\min}}\right)\lceil\frac{\log K}{1.61}\rceil^2\log T$$
$$+\frac{m\pi^2}{6}\log_2\left(\frac{c_2B_1K}{\Delta_{\min}}\right)\Delta_{\max} \tag{121}$$

C.5 Summary of Regret Upper Bounds and Discussions on Distribution-Independent Bounds and Lower Bounds

C.5.1 Analysis using the reverse amortization tricks (Appendix C.3).

When using the **improved** analysis in Appendix C.3, by Eq. (49), Appendix C.3.1, Appendix C.3.2, the total regret is bounded as follows

(1) if $\lambda > 1$,

$$Reg(T) \leq Reg(T, E_{t,1}) + Reg(T, E_{t,2}) + \frac{2\pi^2}{3}m\Delta_{\max}$$
$$\leq \sum_{i=1}^{m}\frac{48c_1^2B_v^2\log T}{\Delta_i^{\min}}(3+\log K) + \sum_{i\in[m]}24c_2B_1\left(\log_2\frac{B_1c_2K}{\Delta_i^{\min}}\right)\left(1+\log(\frac{K\Delta_i^{\max}}{\Delta_i^{\min}})\right)\log T$$
$$+\sum_{i\in[m]}\frac{2c_1^2B_v^2}{\Delta_i^{\min}} + \frac{m\pi^2}{6}\log_2\left(\frac{c_1^2B_v^2K}{\lambda(\Delta_{\min})^2}\right)\Delta_{\max} + \frac{m\pi^2}{6}\log_2\frac{4B_1c_2K}{\Delta_i^{\min}}\Delta_{\max} + \frac{2\pi^2}{3}m\Delta_{\max}+$$
$$\leq O\left(\sum_{i\in[m]}\frac{B_v^2\log K\log T}{\Delta_i^{\min}} + \sum_{i\in[m]}B_1\log^2\left(\frac{B_1K}{\Delta_i^{\min}}\right)\log T\right) \tag{122}$$

(2) if $\lambda = 1$,

$$Reg(T) \leq Reg(T, E_{t,1}) + Reg(T, E_{t,2}) + \frac{2\pi^2}{3}m\Delta_{\max}$$
$$\leq \sum_{i=1}^{m}\log\frac{c_1^2B_v^2K}{(\Delta_i^{\min})^2}\frac{48c_1^2B_v^2\log T}{\Delta_i^{\min}}(3+\log K)$$
$$+\sum_{i\in[m]}24c_2B_1\left(\log_2\frac{B_1c_2K}{\Delta_i^{\min}}\right)\left(1+\log(\frac{K\Delta_i^{\max}}{\Delta_i^{\min}})\right)\log T$$
$$+\sum_{i\in[m]}\frac{2c_1^2B_v^2}{\Delta_i^{\min}} + \frac{m\pi^2}{6}\log_2\left(\frac{c_1^2B_v^2K}{(\Delta_{\min})^2}\right)\Delta_{\max} + \frac{m\pi^2}{6}\log_2\frac{4B_1c_2K}{\Delta_i^{\min}}\Delta_{\max} + \frac{2\pi^2}{3}m\Delta_{\max}$$

$$\leq O\left(\sum_{i\in[m]}\frac{B_v^2\log\left(\frac{B_vK}{\Delta_i^{\min}}\right)\log K\log T}{\Delta_i^{\min}}+\sum_{i\in[m]}B_1\log^2\left(\frac{B_1K}{\Delta_i^{\min}}\right)\log T\right) \tag{123}$$

C.5.2 Regret Bound Using the Infinitely Many Events (Appendix C.4).

When using the analysis in Appendix C.4, by Eq. (49), Appendix C.4.1, Appendix C.4.2, the total regret is bounded as follows

(1) if $\lambda > 1$,

$$Reg(T)\leq Reg(T,E_{t,1})+Reg(T,E_{t,2})+\frac{2\pi^2}{3}m\Delta_{\max}$$
$$\leq\sum_{i=1}^{m}\left(\frac{480c_1^2B_v^2}{(1-2^{-(\lambda-1)})}\right)\lceil\frac{\log K}{1.61}\rceil^2\frac{\log T}{\Delta_i^{\min}}+120c_2B_1\sum_{i=1}^{m}\left(\log_2\frac{c_2B_1K}{\Delta_i^{\min}}\right)\left(1+\log\frac{\Delta_i^{\max}}{\Delta_i^{\min}}\right)$$
$$\lceil\frac{\log K}{1.61}\rceil^2\log T+\frac{m\pi^2}{6}\log_2\left(\frac{c_1^2B_v^2K}{\lambda(\Delta_{\min})^2}\right)\Delta_{\max}+\frac{m\pi^2}{6}\log_2\left(\frac{c_2B_1K}{\Delta_{\min}}\right)\Delta_{\max}+\frac{2\pi^2}{3}m\Delta_{\max}$$
$$\leq O\left(\sum_{i\in[m]}\frac{B_v^2\log^2K\log T}{\Delta_i^{\min}}+\sum_{i\in[m]}B_1\log^2\left(\frac{B_1K}{\Delta_i^{\min}}\right)\log^2K\log T\right) \tag{124}$$

(2) if $\lambda = 1$,

$$Reg(T)\leq Reg(T,E_{t,1})+Reg(T,E_{t,2})+\frac{2\pi^2}{3}m\Delta_{\max}$$
$$\leq\sum_{i=1}^{m}(480c_1^2B_v^2)\log_2\left(\frac{c_1^2B_v^2K}{(\Delta_i^{\min})^2}\right)\lceil\frac{\log K}{1.61}\rceil^2\frac{\log T}{\Delta_i^{\min}}$$
$$+120c_2B_1\sum_{i=1}^{m}\left(\log_2\frac{c_2B_1K}{\Delta_i^{\min}}\right)\left(1+\log\frac{\Delta_i^{\max}}{\Delta_i^{\min}}\right)\lceil\frac{\log K}{1.61}\rceil^2\log T$$
$$+\frac{m\pi^2}{6}\log_2\left(\frac{c_1^2B_v^2K}{(\Delta_{\min})^2}\right)\Delta_{\max}+\frac{m\pi^2}{6}\log_2\left(\frac{c_2B_1K}{\Delta_{\min}}\right)\Delta_{\max}+\frac{2\pi^2}{3}m\Delta_{\max}$$
$$\leq O\left(\sum_{i\in[m]}\frac{B_v^2\log\left(\frac{B_vK}{\Delta_i^{\min}}\right)\log^2K\log T}{\Delta_i^{\min}}+\sum_{i\in[m]}B_1\log^2\left(\frac{B_1K}{\Delta_i^{\min}}\right)\log^2K\log T\right) \tag{125}$$

C.5.3 Discussion on the Distribution-Independent Bounds

Similar to [29, Appendix B.3], for the distribution-independent regret bound, we fix a gap Δ to be decided later and we consider two events on Δ_{S_t}: $\{\Delta_{S_t}\leq\Delta\}$ and $\{\Delta_{S_t}>\Delta\}$.

For the former case, the regret is trivially $Reg(T,\{\Delta_{S_t}\leq\Delta\})\leq T\Delta$. For the later case, under $\{\Delta_{S_t}>\Delta\}$ it is also straight-forward to replace all $\Delta_i^{\min}$ with Δ in Appendix C.5.1 and derive $Reg(T,\{\Delta_{S_t}>\Delta\})\leq O\left(\frac{mB_v^2\log K\log T}{\Delta}+mB_1\log^2(\frac{B_1K}{\Delta})\log T\right)$ if $\lambda>1$ and $Reg(T,\{\Delta_{S_t}>\Delta\})\leq O\left(\frac{mB_v^2\log(\frac{B_vK}{\Delta})\log K\log T}{\Delta}+mB_1\log^2(\frac{B_1K}{\Delta})\log T\right)$ if $\lambda=1$.

Therefore, for $\lambda>1$, by selecting $\Delta=\Theta\left(\sqrt{\frac{mB_v^2\log T\log K}{T}}+\frac{B_1m\log K\log T}{T}\right)$, we have

$$Reg(T)\leq O\left(B_v\sqrt{m(\log K)T\log T}+B_1m\log^2(KT)\log T\right) \tag{126}$$

For $\lambda=1$, by selecting $\Delta=\sqrt{\frac{mB_v^2\log^2T\log K}{T}}+\frac{B_1m\log K\log T}{T}$, we have

$$Reg(T)\leq O\left(B_v\sqrt{m(\log K)T}\log(KT)+B_1m\log^2(KT)\log T\right) \tag{127}$$

C.6 Discussion on the Lower Bounds

We consider the degenerate case from the lower bound result [24], our regret bound is tight (up to polylogaritmic factors in K). More specifically, Merlis and Mannor [24] consider the special non-triggering CMAB (where $\Delta_{\min} = \Delta$ and $\boldsymbol{\mu}$ are not exponentially close to 0 or 1), and they prove $\Omega(\frac{m\gamma_g^2 \log T}{\Delta})$ and $\Omega(\frac{m\gamma_g^2 \log T}{\log K\Delta})$ regret lower bounds for non-monotone and monotone reward functions, respectively. In our paper, this setting is the same as letting $p_i^{D,S} = 1$ for $i \in S$ and $p_i^{D,S} = 0$ otherwise (i.e., TPVM condition degenerates to VM condition). According to the Remark 4 in Section 3.1, we know $B_v = 3\sqrt{2}\gamma_g$ and $\lambda = 2$ so this gives an $O(\frac{m\gamma_g^2 \log K \log T}{\Delta})$ bound, which is tight to the lower bound up to a $O(\log^2 K)$ factor.

D Regret Analysis for CMAB with Independent Arms (Proofs Related to Theorem 2)

D.1 Useful definitions and Inequalities

We first give the formal definition, the properties and the tail bounds for sub-Gaussian and sub-Exponential random variables, which helps our analysis.

Definition 7 (Sub-Gaussian Random Variable, [27])**.** *A random variable with mean $\mu = \mathbb{E}[X]$ is sub-Gaussian with parameter σ^2 if*

$$\mathbb{E}[e^{\lambda(X-\mu)}] \le e^{\frac{\lambda^2\sigma^2}{2}} \text{ for any } \lambda \in \mathbb{R}. \tag{128}$$

In this case, we write $X \in SG(\sigma^2)$.

Definition 8 (Sub-Exponential Random Variable, [27])**.** *A random variable with mean $\mu = \mathbb{E}[X]$ is sub-Exponential with parameter (ν^2, b) if*

$$\mathbb{E}[e^{\lambda(X-\mu)}] \le e^{\frac{\lambda^2\nu^2}{2}} \text{ for any } |\lambda| < \frac{1}{b}. \tag{129}$$

In this case, we write $X \in SE(\nu^2, b)$.

Lemma 16 (Tail bounds for sub-Exponential random variables, [27])**.** *Let $Y \in SE(\nu^2, b)$ with mean $\mu = \mathbb{E}[Y]$. Then*

$$\Pr[|Y-\mu| \ge \tau] \le \begin{cases} 2e^{-\tau^2/(2\nu^2)}, & \text{if } 0 < \tau \le \frac{\nu^2}{b} \\ 2e^{-\tau/(2b)}, & \text{if } \tau > \frac{\nu^2}{b}. \end{cases} \tag{130}$$

Lemma 17. *(Square of Sub-Gaussian Random Variable is Sub-Exponential [14, Appendix B]) For $X \in SG(\sigma^2)$ and let $Y = X^2$, then*

$$\mathbb{E}[e^{\lambda(Y-E[Y])}] \le 16\lambda^2\sigma^4, \text{ for any } |\lambda| \le \frac{1}{4\sigma^2}. \tag{131}$$

Thus, $X^2 \in SE(\nu^2, b)$ with $\nu = 4\sqrt{2}\sigma^2, b = 4\sigma^2$.

Lemma 18 (Composition of independent sub-Exponential random variables, [27])**.** *Let $Y_1, ..., Y_n$ be independent sub-Exponential random variables $Y_i \in SE(\nu_i^2, b_i)$ with $\mathbb{E}[Y_i] = \mu_i$. Then*

$$\sum_{i=1}^{n}(Y_i - \mu_i) \in SE\left(\sum_{i=1}^{n}\nu_i^2, \max_i b_i\right) \tag{132}$$

D.2 Proof of Theorem 2

Recall that at time t, $T_{t-1,i}$ is the number of times base arm i is observed and $\hat{\mu}_{t-1,i}$ is the empirical mean of arm i. Let $\delta_{t,i} = \hat{\mu}_{t-1,i} - \mu_i$ and by Condition 6, $\delta_{t,i}$ is a sub-Gaussian random variable SE($\frac{C_1(1-\mu_i)\mu_i}{T_{t-1,i}}$) with mean 0.

Let $u_{t,i} = \frac{\delta_{t,i}}{\sqrt{(1-\mu_i)\mu_i}}$, then $u_{t,i}$ is also sub-Gaussian SE($\frac{C_1}{T_{t-1,i}}$). By Condition 4, we have that $|r(S;\hat{\boldsymbol{\mu}}_{t-1}) - r(S;\boldsymbol{\mu})| \le B_v\sqrt{\sum_{i\in S}\left(\frac{|\hat{\mu}_{t,i}-\mu_i|}{\sqrt{(1-\mu_i)\mu_i}}\right)^2} = B_v\sqrt{\sum_{i\in S} u_{t,i}^2}$. Fix a super arm S, we will focus on random variable $Y_{t,S} \triangleq \sum_{i\in S} u_{t,i}^2$.

Since $u_{t,i} \in \text{SG}(\frac{C_1}{T_{t-1,i}})$ with $\mathbb{E}[u_{t,i}] = 0$, and $u_{t,i}$ are independent across $i \in [m]$, we know

$$Y_{t,S} \in \text{SE}(C_1^2 C_3 \sum_{i\in S} \frac{1}{T_{t-1,i}^2}, C_1 C_4 \frac{1}{T_{t-1,S}^{\min}}) \tag{133}$$

due to Lemma 17 and Lemma 18, where $T_{t-1,S}^{\min} = \min_{i\in S} T_{t-1,i}$, constants $C_3 = 32, C_4 = 4$.

For the mean of $Y_{t,S}$, we can also show that $\mathbb{E}[Y_{t,S}] = \sum_{i\in[S]} \mathbb{E}[u_{t,i}^2] = \sum_{i\in S}(\text{Var}[u_{t,i}] + (\mathbb{E}[u_{t,i}])^2) = \sum_{i\in S} \text{Var}[u_{t,i}] \le \sum_{i\in S} \frac{C_1}{T_{t-1,i}}$, where the last inequality is because the variance of any sub-Gaussian random variable $X \in \text{SE}(\sigma^2)$ is smaller than σ^2 [27].

For such a sub-Exponential random variable, we can give the confidence interval for $Y_{t,S}$ based on the tail bound Lemma 16. For any action $S \in \mathcal{S}$, any time $t \in [T]$, it holds with probability $1-\delta$,

$$Y_{t,S} \le \begin{cases} \mathbb{E}[Y_{t,S}] + \sqrt{2C_1^2 C_3 \log(\frac{2}{\delta}) \sum_{i\in S} \frac{1}{T_{t-1,i}^2}}, & \text{if } \frac{C_1 C_4}{T_{t-1,S}^{\min}}\sqrt{2\log(\frac{2}{\delta})} \le \sqrt{C_1^2 C_3 \sum_{i\in S}\frac{1}{T_{t-1,i}^2}} \\ \mathbb{E}[Y_{t,S}] + 2C_1 C_4 \log(\frac{2}{\delta})\frac{1}{T_{t-1,S}^{\min}}, & \text{otherwise.} \end{cases} \tag{134}$$

Equivalently, we can rewrite the above inequality by merging the above two segments as $Y_{t,S} \le C_1 \sum_{i\in S} \frac{1}{T_{t-1,i}} + \max\{\sqrt{2C_1^2 C_3 \log(\frac{2}{\delta}) \sum_{i\in S} \frac{1}{T_{t-1,i}^2}}, 2C_1C_4 \log(\frac{2}{\delta})\frac{1}{T_{t-1,S}^{\min}}\}$ and with probability at least $1-\delta$, it holds that

$$|r(S; \hat{\boldsymbol{\mu}}_{t-1}) - r(S; \boldsymbol{\mu})| \le \rho_t(S), \tag{135}$$

where $\rho_t(S) = B_v \sqrt{\sum_{i\in S} \frac{C_1}{T_{t-1,i}} + \max\left\{\sqrt{2C_1^2 C_3 \sum_{i\in S} \frac{\log(\frac{2}{\delta})}{T_{t-1,i}^2}}, \frac{2C_1C_4\log(\frac{2}{\delta})}{T_{t-1,S}^{\min}}\right\}}$. If S is selected as the action in any round t, then

$$\begin{aligned} \Delta_S &= \alpha r(S^*; \boldsymbol{\mu}) - r(S; \boldsymbol{\mu}) && (136)\\ &\le \alpha(r(S^*; \hat{\boldsymbol{\mu}}_{t-1}) + \rho_t(S^*)) - r(S; \boldsymbol{\mu}) && \text{by Eq. (135) over } S^* \quad (137)\\ &\le \bar{r}_t(S) - r(S; \boldsymbol{\mu}) && S \text{ is produced by } \bar{O} \text{ in line 6 of Algorithm 2} \quad (138)\\ &= r(S; \hat{\boldsymbol{\mu}}_{t-1}) + \rho_t(S) - r(S; \boldsymbol{\mu}) && (139)\\ &\le 2\rho_t(S) && \text{by Eq. (135) over } S, \quad (140) \end{aligned}$$

In other words, S can only be selected when $\rho_t(S) > \Delta_S/2$.

Now we consider two different cases based on the $\sum_{i\in S} \frac{1}{T_{t-1,i}}$.

Case 1: When $\sum_{i\in S} \frac{1}{T_{t-1,i}} < \frac{C_4 \Delta_S^2}{4C_1(C_3+C_4)(B_v)^2}$,

We first show by contraction that the confidence interval lies in the second part of Eq. (134), i.e. $\rho_t(S) = B_v \sqrt{C_1 \sum_{i\in S}\frac{1}{T_{t-1,i}} + 8C_1 \log(\frac{2}{\delta})\frac{1}{T_{t-1,S}^{\min}}}$. Based on Lemma 16, if the confidence interval lies in the first part, then

$$\begin{aligned} \tau &\le \frac{\nu^2}{b} = \frac{C_1^2 C_3 \sum_{i\in S}\frac{1}{T_{t-1,i}^2}}{C_1 C_4 \frac{1}{T_{t-1,S}^{\min}}} = \sum_{i\in S} \frac{C_1 C_3}{C_4} \frac{T_{t-1,S}^{\min}}{T_{t-1,i}^2} && (141)\\ &\le \sum_{i\in S} \frac{C_1C_3}{C_4}\frac{1}{T_{t-1,i}} && (142)\\ &\le \frac{C_1C_3}{C_4}\frac{C_4\Delta_S^2}{4C_1(C_3+C_4)(B_v)^2} = \frac{C_3\Delta_S^2}{4(C_3+C_4)(B_v)^2}. && (143) \end{aligned}$$

This indicates that

$$\begin{aligned} \rho_t(S) &= B_v\sqrt{\mathbb{E}[Y_{t,S}] + \tau} && (144)\\ &\le B_v\sqrt{\mathbb{E}[Y_{t,S}] + \frac{C_3\Delta_S^2}{4(C_3+C_4)(B_v)^2}} && (145) \end{aligned}$$

$$\leq B_v \sqrt{C_1 \sum_{i \in S} \frac{1}{T_{t-1,i}} + \frac{C_3 \Delta_S^2}{4(C_3+C_4)(B_v)^2}} \tag{146}$$

$$\leq B_v \sqrt{\frac{C_4 \Delta_S^2}{4(C_3+C_4)B_v^2} + \frac{C_3 \Delta_S^2}{4(C_3+C_4)(B_v)^2}} \tag{147}$$

$$= \Delta_S/2, \tag{148}$$

which contradicts the requirement $\rho_t(S) > \Delta_S/2$.

Therefore, the confidence interval lies in the second part of Eq. (134) with $\tau = \frac{2C_1C_4 \log(\frac{2}{\delta})}{T_{t-1,S}^{\min}}$. By the same analysis, we need $\tau \geq \frac{C_3\Delta_S^2}{4(C_3+C_4)(B_v)^2}$ (or otherwise we suffer from the same contradiction as shown in Eq. (148)), hence we require $\Delta_S \leq \sqrt{\frac{8(C_3+C_4)(B_v)^2C_1C_4\log(\frac{2}{\delta})}{C_3T_{t-1,S}^{\min}}}$.

Equivalently, we require $\Delta_S \leq 2\sqrt{\frac{8(C_3+C_4)(B_v)^2C_1C_4\log(\frac{2}{\delta})}{C_3T_{t-1,S}^{\min}}} - \Delta_S$ whenever S is selected, due to the use of the reverse amortization trick as in Eq. (60).

Now we can define the regret allocation, We first define a regret allocation function

$$\kappa_{i,\delta}(S,\ell) = \begin{cases} 2\sqrt{\frac{8(C_3+C_4)B_v^2C_1C_4\log(\frac{2}{\delta})}{C_3\ell}}, & \text{if } 1 \leq \ell \leq L_{i,\delta} \text{ and } i = \operatorname{argmin}_{j\in S} T_{t-1,j} \\ 0, & \text{otherwise,} \end{cases} \tag{149}$$

where $L_{i,\delta} = \frac{32(C_3+C_4)B_v^2C_1C_4\log(\frac{2}{\delta})}{C_3(\Delta_i^{\min})^2}$. Also note that if there are multiple arms that achieves the minimum, select the one with minimum index as the min-arm.

It can be easily shown that $\Delta_{S_t} \leq \sum_{i\in S_t} \kappa_{i,\delta}(S_t, T_{t-1,i})$ as follows.

Let $j = \operatorname{argmin}_{i\in S_t} T_{t-1,i}$. If $T_{t-1,j} > L_{j,\delta}$, then $\Delta_{S_t} \leq 2\sqrt{\frac{8(C_3+C_4)B_v^2C_1C_4\log(\frac{2}{\delta})}{C_3T_{t-1,S}^{\min}}} - \Delta_{S_t} \leq 2\sqrt{\frac{8(C_3+C_4)B_v^2C_1C_4\log(\frac{2}{\delta})}{C_3T_{t-1,j}}} - \Delta_{S_t} \leq \Delta_j^{\min} - \Delta_{S_t} \leq 0 = \sum_{i\in S_t}\kappa_{i,\delta}(S_t,T_{t-1,i})$. If $T_{t-1,j} \leq L_{j,\delta}$, then $\Delta_{S_t} \leq 2\sqrt{\frac{8(C_3+C_4)B_v^2C_1C_4\log(\frac{2}{\delta})}{C_3T_{t-1,S}^{\min}}} - \Delta_{S_t} \leq 2\sqrt{\frac{8(C_3+C_4)B_v^2C_1C_4\log(\frac{2}{\delta})}{C_3T_{t-1,S}^{\min}}} = 2\sqrt{\frac{8(C_3+C_4)B_v^2C_1C_4\log(\frac{2}{\delta})}{C_3T_{t-1,j}}} + \sum_{i\in S_t\setminus\{j\}} 0 = \sum_{i\in S_t}\kappa_{i,\delta}(S_t,T_{t-1,i})$

Hence the regret for case 1 is upper bounded by

$$\sum_{t=1}^{T}\sum_{i\in S_t}\Delta_{S_t} \leq \sum_{t=1}^{T}\kappa_{i,\delta}(S_t,T_{t-1,i}) \tag{150}$$

$$\leq \sum_{i\in[m]}\sum_{\ell=1}^{L_{i,\delta}} 2\sqrt{\frac{8(C_3+C_4)B_v^2C_1C_4\log(\frac{2}{\delta})}{C_3\ell}} \tag{151}$$

$$\leq \sum_{i\in[m]} 2\sqrt{\frac{8L_{i,\delta}(C_3+C_4)B_v^2C_1C_4\log(\frac{2}{\delta})}{C_3}} \int_{x=0}^{L_{i,\delta}} \sqrt{\frac{1}{x}} \tag{152}$$

$$\leq \sum_{i\in[m]} \frac{64(C_3+C_4)B_v^2C_1C_4\log(\frac{2}{\delta})}{C_3\Delta_i^{\min}} \tag{153}$$

Case 2: When $\sum_{i\in S}\frac{1}{T_{t-1,i}} \geq \frac{C_4\Delta_S^2}{4C_1(C_3+C_4)B_v^2}$,

This case implies that $\frac{K}{T_{t-1,S}^{\min}} \geq \sum_{i\in S}\frac{1}{T_{t-1,i}} \geq \frac{C_4\Delta_S^2}{4C_1(C_3+C_4)B_v^2}$. Rewriting the inequality, we have $\Delta_S \leq \sqrt{\frac{4KC_1(C_3+C_4)B_v^2}{C_4T_{t-1,S}^{\min}}}$ and following the similar regret allocation and argument from case 1 (where we require $\Delta_S \le \sqrt{\frac{8(C_3+C_4)(B_v)^2 C_1 C_4 \log(\frac{2}{\delta})}{C_3 T_{t-1,S}^{\min}}}$), we have the case 2 contributes at most $\frac{32mKC_1(C_3+C_4)B_v^2}{C_4\Delta_{\min}}$, which is irrelevant of the time horizon T.

Now for the first m rounds so that each arm is observed at least once (i.e., counter $T_{t-1,i} \ge 1$ from $t \ge m+1$ rounds) and consider the bad events when there exists $t \in [T], S \in \mathcal{S}$ such that Eq. (134) does not hold, by setting $\delta = (|\mathcal{S}|T)^{-1}$ and using the union bound, the additional regret is upper bounded by

$$m\Delta_{\max} + \sum_{t\in[T]}\sum_{S\in\mathcal{S}} \frac{1}{|\mathcal{S}|T}\Delta_{\max} \le (m+1)\Delta_{\max}. \tag{154}$$

Therefore the total regret is upper bounded by (using the similar proof following Eq. (54) for the failure of oracle),

$$Reg(T) \le \sum_{i\in[m]} \frac{64(C_3+C_4)B_v^2 C_1 C_4 \log(2|\mathcal{S}|T)}{C_3\Delta_i^{\min}} + \frac{32mKC_1(C_3+C_4)B_v^2}{C_4\Delta_{\min}} + (m+1)\Delta_{\max} \tag{155}$$

$$\le \sum_{i\in[m]} \frac{64(C_3+C_4)B_v^2 C_1 C_4 \log(2T)}{C_3\Delta_i^{\min}}$$

$$+ \sum_{i\in[m]} \frac{64(C_3+C_4)B_v^2 C_1 C_4 \log(|\mathcal{S}|)}{C_3\Delta_i^{\min}} + \frac{32mKC_1(C_3+C_4)B_v^2}{C_4\Delta_{\min}} + (m+1)\Delta_{\max} \tag{156}$$

$$\le O\left(\sum_{i\in[m]} \frac{B_v^2 \log T}{\Delta_i^{\min}}\right) \tag{157}$$

For the distribution-independent regret, similar to Appendix C.5.3, $Reg(T) \le \frac{mB_v^2 \log T}{\Delta} + T\Delta$, when $T \to \infty$. By setting $\Delta = \Theta\left(\sqrt{\frac{mB_v^2\log T}{T}}\right)$, we have

$$Reg(T) \le O\left(B_v\sqrt{mT\log T}\right). \tag{158}$$

D.3 Computational Efficient Oracle for SESCB

Recall that $\rho'_t(S) = B_v\sqrt{\sum_{i\in S}\frac{C_1}{T_{t-1,i}} + 8C_1\sqrt{\sum_{i\in S}\frac{\log(2|\mathcal{S}|T)}{T_{t-1,i}^2}} + \frac{8C_1\log(2|\mathcal{S}|T)}{T_{t-1,S}^{\min}}}$ and $r'_t(S) = r(S;\hat{\boldsymbol{\mu}}_{t-1}) + \rho'_t(S)$. For the submodularity of $r'_t(S)$, it suffices to show $\rho'_t(S)$ is monotone submodular when $r(S;\boldsymbol{\mu})$ is monotone submodular. We know that $g(f(S))$ is submodular if $f(S)$ is submodular and g is a non-decreasing concave function, so it suffices to show three terms within the (non-decreasing concave) square root in $\rho'_t(S)$ are submodular. The first term is a modular function, the second term is the square root of a modular function, and the third term can be rewritten as $\max_{i\in S}\frac{8C_1\log(2|\mathcal{S}|T)}{T_{t-1,i}}$, which is also submodular.

For the regret bound when using $\rho'_t(S)$ instead of $\rho_t(S)$, it can be seen that $\rho_t(S) \le \rho'_t(S) \le \sqrt{2}\rho_t(S)$ for all $S \in \mathcal{S}$, since $\max\{a,b\} \le a+b \le 2\{a,b\}$ for any $a,b\in\mathbb{R}$. So we can equivalently use $B'_v = \sqrt{2}B_v$ to replace B_v and repeat the same proof in Theorem 2 with an additional factor of 2 in Eq. (157).

E Proof of TPVM Smoothness Conditions for Various Applications (Related to Theorem 3)

For convenience, we show our table again in this section.

E.1 Combinatorial cascading bandits

Combinatorial cascading bandits has two categories: conjunctive cascading bandits and disjunctive cascading bandits [17].

Table 4: Summary of the coefficients, regret bounds and improvements for various applications.

Application	Condition	(B_v, B_1, λ)	Regret	Improvement
Disjunctive Cascading Bandits [17]	TPVM$_<$	$(1, 1, 2)$	$O(\sum_{i\in[m]} \frac{\log K \log T}{\Delta_i^{\min}})$	$O(K/\log K)$
Conjunctive Cascading Bandits [17]	TPVM	$(1, 1, 1)$	$O(\sum_{i\in[m]} \log \frac{K}{\Delta_{\min}} \frac{\log K \log T}{\Delta_{i,\min}})$	$O(K/(\log K \log \frac{K}{\Delta_{\min}}))$
Multi-layered Network Exploration [21]	TPVM	$(\sqrt{1.25\lvert V\rvert}, 1, 2)$ †	$O(\sum_{i\in\mathcal{A}} \frac{\lvert V\rvert \log(n\lvert V\rvert) \log T}{\Delta_i^{\min}})$	$O(n/\log(n\lvert V\rvert))$
Influence Maximization on DAG [29]	TPVM$_<$	$(\sqrt{L}\lvert V\rvert, \lvert V\rvert, 1)$ †	$O(\sum_{i\in[m]} \log \frac{\lvert E\rvert}{\Delta_{\min}} \frac{L\lvert V\rvert^2 \log \lvert E\rvert \log T}{\Delta_i^{\min}})$	$O(\lvert E\rvert/(L \log \lvert E\rvert \log \frac{\lvert E\rvert}{\Delta_{\min}}))$
Probabilistic Maximum Coverage [23]*	VM	$(3\sqrt{2\lvert V\rvert}, 1, -)$	$O(\sum_{i\in[m]} \frac{\lvert V\rvert \log T}{\Delta_i^{\min}})$	$O(\log^2 k)$.

* This row is for the application in Section 4 and the rest of rows are for Section 3.1; † $|V|, |E|, n, k, L$ denotes the number of target nodes, the number of edges that can be triggered by the set of seed nodes, the number of layers, the number of seed nodes and the length of the longest directed path, respectively.

Disjunctive form For the disjunctive form, we want to select an ordered list S of K items from total L items, so as to maximize the probability that at least one of the outcomes of the selected items are 1. Each item is associated with a Bernoulli random variable with mean μ_i, indicating whether the user will be satisfied with the item if he scans the item. This setting models the movie recommendation system where the user sequentially scans a list of recommended items and the system is rewarded when the user is satisfied with any recommended item. After the user is satisfied with any item or scans all K items but is not satisfied with any of them, the user leaves the system. Due to this stopping rule, the agent can only observe the outcome of items until (including) the first item whose outcome is 1. If there are no satisfactory items, the outcomes must be all 0. In other words, the triggered set is the prefix set of items until the stopping condition holds. For this application, we have the following lemma.

Lemma 19. *Disjunctive conjunctive cascading bandit problem satisfies TPVM$_<$ bounded smoothness condition with coefficient $(B_v, B_1, \lambda) = (1, 1, 2)$.*

Proof. Without loss of generality, let the action be $\{1, ..., K\}$, then the reward function is $r(S; \boldsymbol{\mu}) = 1 - \prod_{j=1}^{K}(1-\mu_j)$ and the triggering probability is $p_i^{D,S} = \prod_{j=1}^{i-1}(1-\mu_j)$. Let $\bar{\boldsymbol{\mu}} = (\bar{\mu}_1, ..., \bar{\mu}_K)$ and $\boldsymbol{\mu} = (\mu_1, ..., \mu_K)$, where $\bar{\boldsymbol{\mu}} = \boldsymbol{\mu} + \boldsymbol{\zeta} + \boldsymbol{\eta}$ with $\bar{\boldsymbol{\mu}}, \boldsymbol{\mu} \in (0,1)^K, \boldsymbol{\zeta}, \boldsymbol{\eta} \in [0,1]^K$.

$$
\begin{aligned}
&|r(S; \bar{\boldsymbol{\mu}}) - r(S; \boldsymbol{\mu})| \\
&= \prod_{i=1}^{K}(1-\mu_i) - \prod_{i=1}^{K}(1-\bar{\mu}_i) && (159)\\
&= \sum_{i\in[K]} (\bar{\mu}_i - \mu_i)((1-\mu_1)...(1-\mu_{i-1})(1-\bar{\mu}_{i+1})...(1-\bar{\mu}_K)) && (160)\\
&= \sum_{i\in[K]} (\zeta_i + \eta_i)((1-\mu_1)...(1-\mu_{i-1})(1-\bar{\mu}_{i+1})...(1-\bar{\mu}_K)) && (161)\\
&\le \sum_{i\in[K]} (\zeta_i)((1-\mu_1)...(1-\mu_{i-1})(1-\mu_{i+1})...(1-\mu_K)) + \sum_{i\in[K]} (\eta_i)((1-\mu_1)...(1-\mu_{i-1})) && (162)\\
&\le \sum_{i\in[K]} \frac{\zeta_i p_i^{D,S}}{\sqrt{(1-\mu_i)\mu_i}} \sqrt{(1-\mu_{i+1})...(1-\mu_K)\mu_i} + \sum_{i\in[K]} \eta_i p_i^{D,S} && (163)\\
&\le \sqrt{\sum_{i\in[K]} \frac{\zeta_i^2}{(1-\mu_i)\mu_i} (p_i^{D,S})^2} \sqrt{\sum_{i\in[K]} (1-\mu_{i+1})...(1-\mu_K)\mu_i} + \sum_{i\in[K]} \eta_i p_i^{D,S} && (164)\\
&\le \sqrt{\sum_{i\in[K]} \frac{\zeta_i^2}{(1-\mu_i)\mu_i} (p_i^{D,S})^2} \sqrt{1-(1-\mu_1)...(1-\mu_K)} + \sum_{i\in[K]} \eta_i p_i^{D,S} && (165)\\
&\le \sqrt{\sum_{i\in[K]} \frac{\zeta_i^2}{(1-\mu_i)\mu_i} (p_i^{D,S})^2} + \sum_{i\in[K]} \eta_i p_i^{D,S}, && (166)
\end{aligned}
$$

where Eq. (159) uses $\mu_i \le \bar{\mu}_i, i \in [m]$, Eq. (160) is by telescoping the reward difference, Eq. (161) is by definition of $\boldsymbol{\zeta}, \boldsymbol{\eta}$, Eq. (162) is due to $\bar{\mu}_i \ge \mu_i, i \in [m]$, Eq. (163) is due to the definition of $p_i^{D,S}$ and we multiply the first term by $\sqrt{\mu_i}$ but divide it by $\sqrt{(1-\mu_i)\mu_i}$, Eq. (164) is due to the Cauchy–Schwarz inequality on the first term, is by math calculation. Hence, $(B_g, B_1, \lambda) = (1, 1, 2)$.

■

Conjunctive form. For the conjunctive form, the learning agent wants to select K paths from total L paths (i.e., base arms) so as to maximize the probability that the outcomes of the selected paths are all 1. Each item is associated with a Bernoulli random variable with mean μ_i, indicating whether the path will be live if the package will transmit via this path. Such a setting models the network routing problem [17], where the items are routing paths and the package is delivered when all paths are alive. The learning agent will observe the outcome of the first few paths till the first one that is down, since the transmission will stop if any of the path is down. In other words, the triggered set is the prefix set of paths until the stopping condition holds. We have the following lemma.

Lemma 20. *Conjunctive cascading bandit problem satisfies TPVM bounded smoothness condition with coefficient* $(B_v, B_1, \lambda) = (1, 1, 1)$.

Proof. Without loss of generality, suppose the selected base arms are $\{1, ..., K\}$, then the reward function is $r(S;\boldsymbol{\mu}) = \prod_{j=1}^{K}\mu_j$ and the triggering probability is $p_i^{\mu,S} = \prod_{j=1}^{i-1}\mu_j$. Let $\bar{\boldsymbol{\mu}} = (\bar{\mu}_1, ..., \bar{\mu}_K)$ and $\boldsymbol{\mu} = (\mu_1, ..., \mu_K)$, where $\bar{\boldsymbol{\mu}} = \boldsymbol{\mu} + \boldsymbol{\zeta} + \boldsymbol{\eta}$ with $\bar{\boldsymbol{\mu}}, \boldsymbol{\mu} \in (0,1)^K, \boldsymbol{\zeta}, \boldsymbol{\eta} \in [-1,1]^K$.

$$
\begin{aligned}
&|r(S;\bar{\boldsymbol{\mu}}) - r(S;\boldsymbol{\mu})| \\
&= \left|\prod_{i=1}^{K}\bar{\mu}_i - \prod_{i=1}^{K}\mu_i\right| \\
&\le \left|\sum_{i\in[K]}(\bar{\mu}_i - \mu_i)(\mu_1...\mu_{i-1}\bar{\mu}_{i+1}...\bar{\mu}_K)\right| \\
&= \sum_{i\in[K]}(|\zeta_i| + |\eta_i|)(\mu_1...\mu_{i-1}\bar{\mu}_{i+1}...\bar{\mu}_K) \\
&\le \sum_{i\in[K]}|\zeta_i|\,(\mu_1...\mu_{i-1}) + \sum_{i\in[K]}|\eta_i|\,(\mu_1...\mu_{i-1}) &(167)\\
&\le \sum_{i\in[K]}\frac{|\zeta_i|}{\sqrt{(1-\mu_i)\mu_i}}\sqrt{p_i^{D,S}}\sqrt{(\mu_1...\mu_{i-1})(1-\mu_i)} + \sum_{i\in[K]}|\eta_i|\,p_i^{D,S} &(168)\\
&\le \sqrt{\sum_{i\in[K]}\frac{\zeta_i^2}{(1-\mu_i)\mu_i}p_i^{D,S}}\sqrt{\sum_{i\in[K]}(\mu_1...\mu_{i-1})(1-\mu_i)} + \sum_{i\in[K]}|\eta_i|\,p_i^{D,S} &(169)\\
&\le \sqrt{\sum_{i\in[K]}\frac{\zeta_i^2}{(1-\mu_i)\mu_i}p_i^{D,S}}\sqrt{1-\mu_1...\mu_K} + \sum_{i\in[K]}|\eta_i|\,p_i^{D,S} &(170)\\
&\le \sqrt{\sum_{i\in[K]}\frac{\zeta_i^2}{(1-\mu_i)\mu_i}p_i^{D,S}} + \sum_{i\in[K]}|\eta_i|\,p_i^{D,S} &(171)
\end{aligned}
$$

where Eq. (167) uses $\bar{\mu}_i \in [0,1], i \in [m]$, Eq. (168) is due to the definition of $p_i^{D,S}$ and multiply the first term by $\sqrt{1-\mu_i}$ but divided it by $\sqrt{(1-\mu_i)\mu_i}$, Eq. (169) is due to the Cauchy–Schwarz inequality on the first term, Eq. (170) is by math calculation and Eq. (171) is due to $\mu_i \in [0,1], i \in [m]$. Hence, $(B_g, B_1, \lambda) = (1, 1, 1)$. ■

E.2 Multi-layered Network Exploration Problem (MuLaNE) [21]

We consider the MuLaNE problem with random node weights. After we apply the bipartite coverage graph, the corresponding graph is a tri-partite graph (n, V, R) (i.e., a 3-layered graph where the first layer and the second layer forms a bipartite graph, and the second and the third layer forms another bipartite graph), where the left nodes represent n random walkers; Middle nodes are $|V|$ possible targets V to be explored; Right nodes R are V nodes, each of which has only one edge connecting the middle edge. The MuLaNE task is to allocate B budgets into n layers to explore target nodes V and the base arms are $\mathcal{A} = \{(i,u,b) : i \in [n], u \in V, b \in [B]\}$.

With budget allocation $k_1, ..., k_L$, the (effective) base arms consists of two parts:

(1) $\{(i,j) : i \in [n], j \in V\}$, each of which is associated with visiting probability $x_{i,j} \in [0,1]$ indicating whether node j will be visited by explorer i given k_i budgets. All these base arms corresponds to budget $k_i, i \in [n]$ are triggered.

(2) $y_j \in [0,1]$ for $j \in V$ represents the random node weight. The triggering probability $p_j^{D,S} = 1 - \prod_{i \in [n]} (1 - x_{i,j})$.

We have the following lemma.

Lemma 21. *MuLaNE problem satisfies TPVM bounded smoothness condition with coefficient* $(B_v, B_1, \lambda) = (\sqrt{1.25|V|}, 1, 2)$*, where* $|V|$ *is the total number of vertices to be explored.*

Proof. Let effective base arms $\boldsymbol{\mu} = (\boldsymbol{x}, \boldsymbol{y}) \in (0,1)^{(n|V|+|V|)}, \bar{\boldsymbol{\mu}} = (\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) \in (0,1)^{(n|V|+|V|)}$, where $\bar{\boldsymbol{x}} = \boldsymbol{\zeta}_x + \boldsymbol{\eta}_x + \boldsymbol{x}, \bar{\boldsymbol{y}} = \boldsymbol{\zeta}_y + \boldsymbol{\eta}_y + \boldsymbol{y}$, for $\boldsymbol{\zeta}, \boldsymbol{\eta} \in [-1,1]^{(n|V|+|V|)}$. For the target node $j \in V$, the per-target reward function $r_j(S; \boldsymbol{x}, \boldsymbol{y}) = y_j(1 - \prod_{i \in [n]}(1 - x_{i,j}))$. Denote $\bar{p}_j^{D,S} = 1 - \prod_{i \in [n]} (1 - \bar{x}_{i,j})$.

$$
\begin{aligned}
&|r(S; \bar{\boldsymbol{\mu}}) - r(S; \boldsymbol{\mu})| \\
&= \left| \sum_{j \in V} r_j(S; \bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) - r_j(S; \boldsymbol{x}, \boldsymbol{y}) \right| \\
&= \left| \sum_{j \in V} \bar{y}_j \bar{p}_j^{D,S} - \bar{y}_j p_j^{D,S} + \bar{y}_j p_j^{D,S} - y_j p_j^{D,S} \right| \\
&= \left| \sum_{j \in V} \bar{y}_j \left(\prod_{i \in [n]} (1 - x_{i,j}) - \prod_{i \in [n]} (1 - \bar{x}_{i,j})) \right) + \sum_{j \in V} (\bar{y}_j - y_j) p_j^{D,S} \right| \qquad (172) \\
&\le \left| \sum_{j \in V} \sum_{i \in [n]} (\bar{x}_{i,j} - x_{i,j}) \left((1 - x_{1,j})...(1 - x_{i-1,j})(1 - \bar{x}_{i+1,j})...(1 - \bar{x}_{L,j}) \right) \bar{y}_j \right| \\
&+ \left| \sum_{j \in V} (\bar{y}_j - y_j) p_j^{D,S} \right| \\
&\le \sum_{j \in V} \sum_{i \in [n]} \left(|\zeta_{x,i,j}| + |\eta_{x,i,j}| \right) \left((1 - x_{1,j})...(1 - x_{i-1,j}) \right) \bar{y}_j + \sum_{j \in V} (|\zeta_{y,j}| + |\eta_{y,j}|) p_j^{D,S} \qquad (173) \\
&\le \sum_{j \in V} \sum_{i \in [n]} \left(\frac{|\zeta_{x,i,j}|}{\sqrt{(1 - x_{i,j}) x_{i,j}}} \right) \sqrt{\left((1 - x_{1,j})...(1 - x_{i-1,j}) \right) x_{i,j}} \bar{y}_j \\
&+ \sum_{j \in V} \left(\frac{|\zeta_{y,j}| \, p_j^{D,S}}{\sqrt{(1 - y_j) y_j}} \right) \sqrt{(1 - y_j) y_j} + \left(\sum_{j \in V} \sum_{i \in [n]} |\eta_{x,i,j}| + \sum_{j \in V} |\eta_{y,j}| \, p_j^{D,S} \right) \qquad (174) \\
&\le \sqrt{\sum_{j \in V} \sum_{i \in [n]} \left(\frac{\zeta_{x,i,j}^2}{(1 - x_{i,j}) x_{i,j}} \right) + \sum_{j \in V} \frac{\zeta_{y,j}^2 (p_j^{D,S})^2}{(1 - y_j) y_j}} \\
&\cdot \sqrt{\sum_{j \in V} \sum_{i \in [n]} \left((1 - x_{1,j})...(1 - x_{i-1,j}) \right) x_{i,j} \bar{y}_j^2 + \sum_{j \in V} (1 - y_j) y_j} \\
&+ \left(\sum_{j \in V} \sum_{i \in [n]} |\eta_{x,i,j}| + \sum_{j \in V} |\eta_{y,j}| \, p_j^{D,S} \right) \qquad (175)
\end{aligned}
$$

$$= \sqrt{\sum_{j\in V}\sum_{i\in[n]}(\frac{\zeta_{x,i,j}^2}{(1-x_{i,j})x_{i,j}}) + \sum_{j\in V}\frac{\zeta_{y,j}^2(p_j^{D,S})^2}{(1-y_j)y_j}}\sqrt{\sum_{j\in V}\left(1-\prod_{i\in[n]}(1-x_{i,j})\right)\bar{y}_j^2 + \sum_{j\in V}(1-y_j)y_j}$$

$$+ \left(\sum_{j\in V}\sum_{i\in[n]}|\eta_{x,i,j}| + \sum_{j\in V}|\eta_{y,j}|\, p_j^{D,S}\right) \tag{176}$$

$$\leq \sqrt{\sum_{j\in V}\sum_{i\in[n]}(\frac{\zeta_{x,i,j}^2}{(1-x_{i,j})x_{i,j}}) + \sum_{j\in V}\frac{\zeta_{y,j}^2(p_j^{D,S})^2}{(1-y_j)y_j}}\sqrt{\sum_{j\in V}\bar{y}_j^2 + (1-y_j)y_j} +$$

$$\left(\sum_{j\in V}\sum_{i\in[n]}|\eta_{x,i,j}| + \sum_{j\in V}|\eta_{y,j}|\, p_j^{D,S}\right) \tag{177}$$

$$\leq \sqrt{\sum_{j\in V}\sum_{i\in[n]}(\frac{\zeta_{x,i,j}^2}{(1-x_{i,j})x_{i,j}}) + \sum_{j\in V}\frac{\zeta_{y,j}^2(p_j^{D,S})^2}{(1-y_j)y_j}} \cdot \sqrt{1.25|V|}$$

$$+ \left(\sum_{j\in V}\sum_{i\in[n]}|\eta_{x,i,j}| + \sum_{j\in V}|\eta_{y,j}|\, p_j^{D,S}\right), \tag{178}$$

where Eq. (172) is by telescoping the difference of $\left(\prod_{i\in[n]}(1-x_{i,j}) - \prod_{i\in[n]}(1-\bar{x}_{i,j}))\right)$, Eq. (173) is due to $\bar{x}_{i,j} \in [0,1]$ for any i, j, Eq. (174) is because we multiply the first term by $\sqrt{x_{i,j}}$ but divide it by $\sqrt{x_{i,j}(1-x_{i,j})}$ and we multiply the second term by $\sqrt{y_j}$ but divide it by $\sqrt{y_j(1-y_j)}$, Eq. (175) is by applying the Cauchy-Schwarz inequality to the first term and the second term simultaneously, Eq. (176) is due to the math calculation on , Eq. (177) is due to the support of $x_{i,j}$ are bounded with $[0,1]$ for any i, j and Eq. (178) is due to the support of $y_j, \bar{y}_j$ are bounded with $[0,1]$ for any j. Hence $(B_g, B_1, \lambda) = (\sqrt{1.25|V|}, 1, 2)$. ■

E.3 Online Influence Maximization Bandit [29]

In this section, we first introduce the general problem setting of online influence maximization (OIM). Then we start with a easier problem instance, i.e., OIM on tri-partite graph (TPG) to get some intuition, where reward functions have closed form solutions. Finally, we give the results for OIM on the directed acyclic graph (DAG), which are much more involved since the reward function have no closed form solutions.

E.3.1 Setting of the Online Influence Maximization

Following the setting of [29, Section 2.1], we consider a weighted directed graph $G(V, E, p)$, where V is the set of vertices, E is the set of directed edges, and each edge $(u, v) \in E$ is associated with a probability $p(u, v) \in [0, 1]$. When the agent selects a set of seed nodes $S \subseteq V$, the influence propagates as follows: At time 0, the seed nodes S are activated; At time $t > 1$, a node u activated at time $t-1$ will have one chance to activate its inactive out-neighbor v with independent probability $p(u, v)$. The influence spread of S is denoted as $\sigma(S)$ and is defined as the expected number of activated nodes after the propagation process ends. The problem of Influence Maximization is to find seed nodes S with $|S| \leq k$ so that the influence spread $\sigma(S)$ is maximized.

For the problem of online influence maximization (OIM), we consider T rounds repeated influence maximization tasks and the edge probabilities $p(u, v)$ are assumed to be unknown initially. For each round $t \in [T]$, the agent selects k seed nodes as S_t, the influence propagation of S_t is observed and the reward is the number of nodes activated in round t. The agent's goal is to accumulate as much reward as possible in T rounds. The OIM fits into CMAB-T framework: the edges E are the set of base arms $[m]$, the (unknown) outcome distribution D is the joint of m independent Bernoulli random variables for the edge set E, the action S are any seed node sets with size k at most k. For the arm triggering, the triggered set τ_t is the set of edges (u, v) whose source node u is reachable from S_t. Let X_t be the outcomes of the edges E according to probability $p(u, v)$ and the live-edge graph $G_t^{\text{live}}(V, E^{\text{live}})$ be a induced graph with edges that are alive, i.e., $e \in E^{\text{live}}$ iff $X_{t,e} = 1$ for $e \in E$. The triggering probability distribution $D_{\text{trig}}(S_t, X_t)$ degenerates to a deterministic triggered set, i.e., τ_t is deterministically decided given S_t and X_t. The reward $R(S_t, X_t, \tau_t)$ equals to the number activated nodes at the end of t, i.e., the nodes that are reachable from S_t in the live-edge graph G_t^{live}. The offline oracle is a $(1 - 1/e - \varepsilon, 1/|V|)$-approximation algorithm given by the greedy algorithm from [15].

E.3.2 Online Influence Maximization Bandit on Tri-partite Graph (TPG)

We consider a OIM scenario where the underlying graph is a graph with three layers: the seed node layer, the intermediate layer and the target node layer. Specifically, the underlying graph is denoted as (L, M, R), where the first layer consists of L candidates for the seed node selection, the second layer consists of M intermediate nodes and the third layer are R target nodes. Such a setting is of significant interest since the edges connecting the target nodes can only be triggered and cannot be observed by seed-node selection, which requires the notion of triggering arms. Another favorable feature is that the reward function of this setting can be explicitly expressed, whereas for general IM, even calculating the explicit form of the reward function is NP-hard [5].

In this application, the base arms consists of two parts:

(1) $\{(i, j) : i \in [L], j \in [M]\}$, each of which is associated with probability $x_{i,j} \in [0, 1]$ indicating the probability whether edge (i, j) is alive when i is selected as seed nodes. Without loss of generality, we assume the seed nodes are $S = \{1, ..., K\}$.

(2)$\{(j, k) : j \in [M], k \in [R]\}$, each of which is associated with probability $y_{j,k} \in [0, 1]$ indicating whether edge (j, k) is live when j is triggered for the first time. The triggering probability $p_j^{D,S} = 1 - \prod_{i\in[K]} (1 - x_{i,j})$.

Let $\boldsymbol{\mu} = (\boldsymbol{x}, \boldsymbol{y}) \in (0, 1)^{(LM+MR)}$, $\bar{\boldsymbol{\mu}} = (\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) \in (0, 1)^{(LM+MR)}$, where $\bar{\boldsymbol{x}} = \boldsymbol{\zeta}_x + \boldsymbol{\eta}_x + \boldsymbol{x}, \bar{\boldsymbol{y}} = \boldsymbol{\zeta}_y + \boldsymbol{\eta}_y + \boldsymbol{y}$, for $\boldsymbol{\zeta}, \boldsymbol{\eta} \in [0, 1]^{(LM+MR)}$. Fix any target node k, the reward function $r_k(S; \boldsymbol{x}, \boldsymbol{y}) = 1 - \prod_{j\in[M]} \left(1 - y_{j,k}(1 - \prod_{i\in[K]}(1 - x_{i,j}))\right)$.

Lemma 22. *OIM-TPG problem satisfies TPVM$_<$ bounded smoothness condition with coefficient* $(B_v, B_1, \lambda) = (\sqrt{2}R, R, 1)$*, where R is the total number of target nodes.*

Proof. For notational brevity, we denote $p_j \triangleq p_j^{D,S} = 1 - \prod_{i\in[L]} (1 - x_{i,j})$, $\bar{p}_j = 1 - \prod_{i\in[K]}(1 - \bar{x}_{i,j})$, $g_j = 1 - \bar{y}_{j,k}p_j$, $\bar{g}_j = 1 - \bar{y}_{j,k}\bar{p}_j$.

The difference of $r_k(S; \bar{\boldsymbol{\mu}}), r_k(S; \boldsymbol{\mu})$ can be written as,

$$|r(S; \bar{\boldsymbol{\mu}}) - r(S; \boldsymbol{\mu})| = \sum_{k\in[R]} r_k(S; \bar{\boldsymbol{\mu}}) - r_k(S; \boldsymbol{\mu}) \tag{179}$$

$$= \sum_{k\in[R]} \underbrace{\left(1 - \prod_{j\in[M]} (1 - \bar{y}_{j,k}\bar{p}_j)\right) - \left(1 - \prod_{j\in[M]} (1 - \bar{y}_{j,k}p_j)\right)}_{\text{term (a)}}$$

$$+ \sum_{k\in[R]} \underbrace{\left(1 - \prod_{j\in[M]} (1 - \bar{y}_{j,k}p_j)\right) - \left(1 - \prod_{j\in[M]} (1 - y_{j,k}p_j)\right)}_{\text{term}(b)}, \tag{180}$$

where Eq. (180) is by adding and subtracting the same $\sum_{k\in[R]} \left(1 - \prod_{j\in[M]}(1 - \bar{y}_{j,k}p_j)\right)$.

For term (a), it holds that

$$\text{term}(a) = \sum_{j\in[M]} (g_j - \bar{g}_j)(g_1...g_{j-1}\bar{g}_{j+1}...\bar{g}_M) \tag{181}$$

$$= \sum_{j\in[M]} \bar{y}_{j,k}(\bar{p}_j - p_j)(g_1...g_{j-1}\bar{g}_{j+1}...\bar{g}_M)$$

$$
\begin{aligned}
&= \sum_{j\in[M]} \bar{y}_{j,k}(g_1...g_{j-1}\bar{g}_{j+1}...\bar{g}_M)\left(\left(1-\prod_{i\in[K]}(1-\bar{x}_{i,j})\right)-\left(1-\prod_{i\in[K]}(1-x_{i,j})\right)\right)\\
&= \sum_{j\in[M]} \bar{y}_{j,k}(g_1...g_{j-1}\bar{g}_{j+1}...\bar{g}_M)\sum_{i\in[K]}(\bar{x}_{i,j}-x_{i,j})\left((1-x_{1,j})...(1-x_{i-1,j})(1-x_{i+1,j})...(1-x_{K,j})\right) && (182)\\
&= \sum_{i\in[K]}\sum_{j\in[M]}(\zeta_{x,i,j}+\eta_{x,i,j})\bar{y}_{j,k}(g_1...g_{j-1}\bar{g}_{j+1}...\bar{g}_M)\left((1-x_{1,j})...(1-x_{i-1,j})(1-x_{i+1,j})...(1-x_{K,j})\right)\\
&\le \underbrace{\sum_{i\in[K]}\sum_{j\in[M]}\left(\frac{\zeta_{x,i,j}}{\sqrt{(1-x_{i,j})x_{i,j}}}\right)\sqrt{x_{i,j}(1-x_{1,j})...(1-x_{i-1,j})\bar{y}_{j,k}(g_1...g_{j-1})}}_{\text{term}(c)}\\
&+ \underbrace{\sum_{i\in[K]}\sum_{j\in[M]}\eta_{x,i,j}}_{\text{term}(d)}, && (183)
\end{aligned}
$$

where Eq. (181) is by telescoping the term (b), Eq. (182) is by telescoping $\bar{p}_j - p_j$, Eq. (183) is because $\bar{g}_j \in [0,1]$ for all j and multiplying the first term by $\sqrt{x_{i,j}}$ but dividing it by $\sqrt{(1-x_{i,j})x_{i,j}}$.

For term (b), it holds that

$$
\begin{aligned}
\text{term}(b) &= \left(1-\prod_{j\in[M]}(1-\bar{y}_{j,k}p_j)\right)-\left(1-\prod_{j\in[M]}(1-y_{j,k}p_j)\right)\\
&= \sum_{j\in[M]} p_j(\bar{y}_{j,k}-y_{j,k})(1-y_{1,k}p_1)...(1-y_{j,k}p_j)(1-\bar{y}_{j+1,k}p_{j+1})...(1-\bar{y}_{M,k}p_M) && (184)\\
&\le \underbrace{\sum_{j\in[M]}\left(\frac{\zeta_{y,j,k}\sqrt{p_j}}{\sqrt{(1-y_{j,k})y_{j,k}}}\right)\sqrt{y_{j,k}p_j(1-y_{1,k}p_1)...(1-y_{j-1,k}p_{j-1})}}_{\text{term}(e)} + \underbrace{\sum_{j\in[M]} p_j\eta_{y,j,k}}_{\text{term}(f)}, && (185)
\end{aligned}
$$

where Eq. (184) is by telescoping on the term (b), Eq. (185) is by multiplying the first term by $\sqrt{y_{i,j}}$ but dividing it by $\sqrt{(1-y_{i,j})y_{i,j}}$.

Next, we apply Cauchy–Schwarz inequality to term (c) and term (e) simultaneously,

$$
\begin{aligned}
&\sum_{k\in[R]}(\text{term}(c)+\text{term}(e))\\
&\le \sqrt{\left(\sum_{i\in[K],j\in[M],k\in[R]}\frac{\zeta_{x,i,j}^2}{(1-x_{i,j})x_{i,j}}\right)+\left(\sum_{j\in[M]}\sum_{k\in[R]}\frac{\zeta_{y,j,k}^2 p_j}{(1-y_{j,k})y_{j,k}}\right)}\\
&\cdot\left[\left(\sum_{i\in[K]}\sum_{j\in[M]}\sum_{k\in[R]}x_{i,j}(1-x_{1,j})...(1-x_{i-1,j})\bar{y}_{j,k}(g_1...g_{j-1})\right)\right.\\
&\left.+\left(\sum_{j\in[M]}\sum_{k\in[R]}y_{j,k}p_j(1-y_{1,k}p_1)...(1-y_{j-1,k}p_{j-1})\right)\right]^{1/2}\\
&\le \sqrt{\left(\sum_{i\in[K],j\in[M],k\in[R]}\frac{\zeta_{x,i,j}^2}{(1-x_{i,j})x_{i,j}}\right)+\left(\sum_{j\in[M]}\sum_{k\in[R]}\frac{\zeta_{y,j,k}^2 p_j}{(1-y_{j,k})y_{j,k}}\right)}
\end{aligned}
$$

$$\cdot \sqrt{\left(\sum_{j\in[M]}\sum_{k\in[R]}(1-\prod_{i\in[K]}(1-x_{i,j}))\bar{y}_{j,k}(g_1...g_{j-1})\right)+\sum_{k\in[R]}\left(1-\prod_{j\in[M]}(1-y_{j,k}p_j)\right)} \tag{186}$$

$$= \sqrt{\left(\sum_{i\in[K],j\in[M],k\in[R]}\frac{\zeta_{x,i,j}^2}{(1-x_{i,j})x_{i,j}}\right)+\left(\sum_{j\in[M]}\sum_{k\in[R]}\frac{\zeta_{y,j,k}^2 p_j}{(1-y_{j,k})y_{j,k}}\right)}$$

$$\cdot \sqrt{\sum_{k\in[R]}\left(\sum_{j\in[M]}(1-g_j)(g_1...g_{j-1})\right)+\sum_{k\in[R]}\left(1-\prod_{j\in[M]}(1-y_{j,k}p_j)\right)} \tag{187}$$

$$= \sqrt{\left(\sum_{i\in[K],j\in[M],k\in[R]}\frac{\zeta_{x,i,j}^2}{(1-x_{i,j})x_{i,j}}\right)+\left(\sum_{j\in[M]}\sum_{k\in[R]}\frac{\zeta_{y,j,k}^2 p_j}{(1-y_{j,k})y_{j,k}}\right)}$$

$$\cdot \sqrt{\sum_{k\in[R]}\left(1-\prod_{j\in[M]}(1-\bar{y}_{j,k}p_j)\right)+\sum_{k\in[R]}\left(1-\prod_{j\in[M]}(1-y_{j,k}p_j)\right)} \tag{188}$$

$$\leq \sqrt{R}\sqrt{\left(\sum_{i\in[K]}\sum_{j\in[M]}\frac{\zeta_{x,i,j}^2}{(1-x_{i,j})x_{i,j}}\right)+\left(\sum_{j\in[M]}\sum_{k\in[R]}\frac{\zeta_{y,j,k}^2 p_j}{(1-y_{j,k})y_{j,k}}\right)}$$

$$\cdot \sqrt{\sum_{k\in[R]}\left(1-\prod_{j\in[M]}(1-\bar{y}_{j,k}p_j)\right)+\sum_{k\in[R]}\left(1-\prod_{j\in[M]}(1-y_{j,k}p_j)\right)} \tag{189}$$

$$\leq \sqrt{\left(\sum_{i\in[K],j\in[M],k\in[R]}\frac{\zeta_{x,i,j}^2}{(1-x_{i,j})x_{i,j}}\right)+\left(\sum_{j\in[M]}\sum_{k\in[R]}\frac{\zeta_{y,j,k}^2 p_j}{(1-y_{j,k})y_{j,k}}\right)}\cdot\sqrt{2}R, \tag{190}$$

where Eq. (186) is due to the math calculation on the summation over $i \in [K]$, Eq. (187) is due to the definition of g_j, Eq. (188) is due to the math calculation on the summation over $j \in [M]$, Eq. (189) is because we take one summation over $k \in [R]$ out for the first square root term and Eq. (190) is because $y_{j,k}, \bar{y}_{j,k}, p_j$ are all bounded with support $[0, 1]$.

Combined with $\sum_{k\in[R]}(\text{term}(d) + \text{term}(f))$, we can derive that

$$\begin{aligned}|r(S;\bar{\boldsymbol{\mu}})-r(S;\boldsymbol{\mu})| &= \sqrt{\left(\sum_{i\in[K]}\sum_{j\in[M]}\frac{\zeta_{x,i,j}^2}{(1-x_{i,j})x_{i,j}}\right)+\left(\sum_{j\in[M]}\sum_{k\in[R]}\frac{\zeta_{y,j,k}^2 p_j}{(1-y_{j,k})y_{j,k}}\right)}\cdot\sqrt{2}R \\ &\quad+\sum_{i\in[K]}\sum_{j\in[M]}\sum_{k\in[R]}\eta_{x,i,j}+\sum_{j\in[M]}\sum_{k\in[R]}p_j\eta_{y,j,k} \\ &\leq \sqrt{\left(\sum_{i\in[K]}\sum_{j\in[M]}\frac{\zeta_{x,i,j}^2}{(1-x_{i,j})x_{i,j}}\right)+\left(\sum_{j\in[M]}\sum_{k\in[R]}\frac{\zeta_{y,j,k}^2 p_j}{(1-y_{j,k})y_{j,k}}\right)}\cdot\sqrt{2}R \\ &\quad+R\left(\sum_{i\in[K]}\sum_{j\in[M]}\eta_{x,i,j}+\sum_{j\in[M]}\sum_{k\in[R]}p_j\eta_{y,j,k}\right),\end{aligned} \tag{191}$$

where Eq. (191) is because the summation over $k \in [R]$ in the second term.

Hence we have $(B_g, B_1, \lambda) = (\sqrt{2}R, R, 1)$. ■

E.3.3 Influence Maximization on Directed Acyclic Graphs (DAG)

In this section, we introduce how to derive (B_g, B_1, λ) for DAGs. Note that TBG (or even multi-layered bipartite graphs) are special cases of DAG. However, the main challenge is that for general DAGs, there are no closed-form solutions for the reward function since it is NP-Hard to compute their influence spread [5]. To deal with this challenge, we will use the live-edge graph and edge coupling technique as [20] and prove the following lemma.

Lemma 23. *OIM-DAG problem satisfies TPVM$_<$ bounded smoothness condition with coefficient $(B_v, B_1, \lambda) = (\sqrt{L}|V|, |V|, 1)$, where L is the length of the longest directed path.*

Proof. Denote the DAG graph as $G(V, E)$ where V are nodes and E are the edges. With a little abuse of the notation, we will denote $u(e), v(e)$ as the starting node and ending node of the edge e. Inspired by previous applications with closed-form solutions, we will partition edges E (i.e., base arms) into L groups, where L is the length of the longest directed path of G. More specifically, we apply the topological sort on G and label each node with $l(v) \in \{0, 1, ..., L\}$, where $l(v)$ is the length of the longest path that starts from node v. For simplicity, we say node v in layer $l(v)$. This sorting and labelling procedure can be done in $O(V + E)$ time complexity. Given this labelling $l(v)$ for $v \in V$, we partition edges E into L disjoint edge sets so that $E = \bigcup_{s\in[L]} E_s$, where E_s contains edges that point from node in layer s to node in lower layers, i.e., $E_s = \{(u, v) : l(u) = s, l(v) < s\}$. One critical property we will use later is that for any two edges $e, e' \in E_s$ in the same layer s, there are no directed path p from any node $v \in V$ so that both e, e' are in the path p, or otherwise $l(u(e)) \neq l(u(e'))$ which contradicts the assumption that starting nodes $u(e)$ and $u(e')$ belong to the same layer s.

Let $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_L)$ be the true mean vector for the partitions $(E_1, ..., E_L)$ mentioned above, where each partition contains $n_s = |E_s|$ base arms (with $\boldsymbol{\mu}_s \in (0, 1)^{n_s}$), for $s \in [L]$. Similarly, we denote $\bar{\boldsymbol{\mu}} = (\bar{\boldsymbol{\mu}}_1, ..., \bar{\boldsymbol{\mu}}_L), \boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, ..., \boldsymbol{\zeta}_L), \boldsymbol{\eta} = (\boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_L)$ such that $\bar{\boldsymbol{\mu}} = \boldsymbol{\mu} + \boldsymbol{\zeta} + \boldsymbol{\eta}$ and $\bar{\boldsymbol{\mu}}_s \in (0, 1)^{n_s}, \boldsymbol{\zeta}, \boldsymbol{\eta} \in [0, 1]^{n_s}$. Now for the reward, we focus on each node t as target, so that $r(S; \boldsymbol{\mu}) \triangleq \sum_{t\in V} r_t(S; \boldsymbol{\mu})$ and $r(S; \bar{\boldsymbol{\mu}}) \triangleq \sum_{t\in V} r_t(S; \bar{\boldsymbol{\mu}})$. Now fix any target node $t \in V$, we can telescope the reward by gradually changing one partition from $\bar{\boldsymbol{\mu}}_s$ to $\boldsymbol{\mu}_s$ as follows:

$$
\begin{aligned}
&|r_t(S; \bar{\boldsymbol{\mu}}) - r_t(S; \boldsymbol{\mu})| \\
&= r_t(S; \bar{\boldsymbol{\mu}}_1, ..., \bar{\boldsymbol{\mu}}_L) - r_t(S; \boldsymbol{\mu}_1, \bar{\boldsymbol{\mu}}_2, ..., \bar{\boldsymbol{\mu}}_L) \\
&= r_t(S; \boldsymbol{\mu}_1, \bar{\boldsymbol{\mu}}_2, ..., \bar{\boldsymbol{\mu}}_L) - r_t(S; \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \bar{\boldsymbol{\mu}}_3, ..., \bar{\boldsymbol{\mu}}_L) \\
&+ ... \\
&+ \underbrace{r_t(S; \boldsymbol{\mu}_1, ...\boldsymbol{\mu}_{s-1}, \bar{\boldsymbol{\mu}}_s, ..., \bar{\boldsymbol{\mu}}_L) - r_t(S; \boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_s, \bar{\boldsymbol{\mu}}_{s+1}, ..., \bar{\boldsymbol{\mu}}_L)}_{\text{s-th partition}} \\
&+ ... \\
&+ r_t(S; \boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_{L-1}, \bar{\boldsymbol{\mu}}_L) - r_t(S; \boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_L)
\end{aligned} \tag{192}
$$

For the s-th partition, since the first term and the second term differs only in $\boldsymbol{\mu}_s$ and $\bar{\boldsymbol{\mu}}_s$, we denote the first $s - 1$ partitions as $\cup_{\ell=1}^{s-1}\boldsymbol{\mu}_\ell$ and the last $L - s$ partitions as $\cup_{\ell=s+1}^{L}\bar{\boldsymbol{\mu}}_\ell$. We can further telescope the reward difference by gradually changing one parameter from $\bar{\mu}_{s,i}$ to $\mu_{s,i}$ as follows:

$$
\begin{aligned}
&\text{s-th partition} \\
&= r_t(S; \cup_{\ell=1}^{s-1}\boldsymbol{\mu}_\ell, \bar{\mu}_{s,1}, ..., \bar{\mu}_{s,n_s}, \cup_{\ell=s+1}^{L}\bar{\boldsymbol{\mu}}_\ell) - r_t(S; \cup_{\ell=1}^{s-1}\boldsymbol{\mu}_\ell, \mu_{s,1}, \bar{\mu}_{s,2}, ..., \bar{\mu}_{s,n_s}, \cup_{\ell=s+1}^{L}\bar{\boldsymbol{\mu}}_\ell) \\
&= r_t(S; \cup_{\ell=1}^{s-1}\boldsymbol{\mu}_\ell, \mu_{s,1}, \bar{\mu}_{s,2}, ..., \bar{\mu}_{s,n_s}, \cup_{\ell=s+1}^{L}\bar{\boldsymbol{\mu}}_\ell) - r_t(S; \cup_{\ell=1}^{s-1}\boldsymbol{\mu}_\ell, \mu_{s,1}, \mu_{s,2}, \bar{\mu}_{s,3}, ..., \bar{\mu}_{s,n_s}, \cup_{\ell=s+1}^{L}\bar{\boldsymbol{\mu}}_\ell) \\
&+ ... \\
&+ \underbrace{r_t(S; \cup_{\ell=1}^{s-1}\boldsymbol{\mu}_\ell, \cup_{\ell=1}^{i-1}\mu_{s,\ell}, \cup_{\ell=i}^{n_s}\bar{\mu}_{s,\ell}, \cup_{\ell=s+1}^{L}\bar{\boldsymbol{\mu}}_\ell) - r_t(S; \cup_{\ell=1}^{s-1}\boldsymbol{\mu}_\ell, \cup_{\ell=1}^{i}\mu_{s,\ell}, \cup_{\ell=i+1}^{n_s}\bar{\mu}_{s,\ell}, \cup_{\ell=s+1}^{L}\bar{\boldsymbol{\mu}}_\ell))}_{\text{i-th change in s-th partition}} \\
&+ ... \\
&+ r_t(S; \cup_{\ell=1}^{s-1}\boldsymbol{\mu}_\ell, \mu_{s,1}, ..., \mu_{s,n_s-1}, \bar{\mu}_{s,n_s}, \cup_{\ell=s+1}^{L}\bar{\boldsymbol{\mu}}_\ell) - r_t(S; \cup_{\ell=1}^{s-1}\boldsymbol{\mu}_\ell, \mu_{s,1}, ..., \mu_{s,n_s}, \cup_{\ell=s+1}^{L}\bar{\boldsymbol{\mu}}_\ell)
\end{aligned} \tag{193}
$$

For the i-th change in s-th partition, we use the live edge graph [6] technique (which links the probability of a node is activated to the probability this node is reachable in the random live-edge graph where each edge i is independently selected as live-edge with probability μ_i), in order to transform the reward difference into the probability of some events happens. Note that by using "bypass of edge e", we mean the path that connects seed nodes S and target node t but does not contains edge e in this path. Let $e_{s,i} = (u(e_{s,i}), v(e_{s,i}))$ denote the i-th edge with staring node $u(e_{s,i})$ and ending node $v(e_{s,i})$ in partition E_s, for $i \in [n_s]$. We use $\Pr[A|\boldsymbol{\mu}]$ to denote the probability of all live-edge graphs under parameter $\boldsymbol{\mu}$ so that event A happens. Consider the live-edge graphs without $e_{s,i}$,

$$
\begin{aligned}
&\text{i-th change in s-th partition}\\
&= (\bar{\mu}_{s,i} - \mu_{s,i}) \Pr[S \to u(e_{s,i}), v(e_{s,i}) \to t, \text{ no bypass of } e_{s,i} \mid \cup_{\ell=1}^{s-1} \boldsymbol{\mu}_\ell,\\
&\mu_{s,1}, ...\mu_{s,i-1}, \bar{\mu}_{s,i+1}, ..., \bar{\mu}_{s,n_s}, \cup_{\ell=s+1}^{L} \bar{\boldsymbol{\mu}}_\ell] \quad (194)\\
&\le (\bar{\mu}_{s,i} - \mu_{s,i}) \Pr[S \to u(e_{s,i}), v(e_{s,i}) \to t, \text{ no bypass of } e_{s,i} \mid \cup_{\ell=1}^{s-1} \boldsymbol{\mu}_\ell, \boldsymbol{\mu}_s \backslash \mu_{s,i}, \cup_{\ell=s+1}^{L} \bar{\boldsymbol{\mu}}_\ell] \quad (195)\\
&\le \zeta_{s,i} \Pr[S \to u(e_{s,i}), v(e_{s,i}) \to t, \text{ no bypass of } e_{s,i} \mid \cup_{\ell=1}^{s-1} \boldsymbol{\mu}_\ell, \boldsymbol{\mu}_s \backslash \mu_{s,i}, \cup_{\ell=s+1}^{L} \bar{\boldsymbol{\mu}}_\ell]\\
&+ \eta_{s,i} \Pr[S \to u(e_{s,i}) \mid \cup_{\ell=1}^{s-1} \boldsymbol{\mu}_\ell] \quad (196)\\
&= \sqrt{\frac{\zeta_{s,i}^2}{\mu_{s,i}}} \sqrt{\mu_{s,i} (\Pr[S \to u(e_{s,i}), v(e_{s,i}) \to t, \text{ no bypass of } e_{s,i} \mid \cup_{\ell=1}^{s-1} \boldsymbol{\mu}_\ell, \boldsymbol{\mu}_s \backslash \mu_{s,i}, \cup_{\ell=s+1}^{L} \bar{\boldsymbol{\mu}}_\ell])^2}\\
&+ \eta_{s,i} \Pr[S \to u(e_{s,i}) \mid \cup_{\ell=1}^{s-1} \boldsymbol{\mu}_\ell] \quad (197)\\
&\le \sqrt{\frac{\zeta_{s,i}^2 \Pr[S \to u(e_{s,i}) \mid \cup_{\ell=1}^{s-1} \boldsymbol{\mu}_\ell]}{(1-\mu_{s,i})\mu_{s,i}}}\\
&\cdot \sqrt{\Pr[S \to u(e_{s,i}), e_{s,i} \text{ is live}, v(e_{s,i}) \to t, \text{ no bypass of } e_{s,i} \mid \cup_{\ell=1}^{s} \boldsymbol{\mu}_\ell, \cup_{\ell=s+1}^{L} \bar{\boldsymbol{\mu}}_\ell]}\\
&+ \eta_{s,i} \Pr[S \to u(e_{s,i}) \mid \cup_{\ell=1}^{s-1} \boldsymbol{\mu}_\ell], \quad (198)
\end{aligned}
$$

where Eq. (194) is due to the coupling technique [20] and the contribution of live-edge graph G' to the reward difference is $\bar{\mu}_{s,i} - \mu_{s,i}$ if $u(e_{s,i})$ is reachable from S, the target node t is reachable from $v(e_{s,i})$ but t is not reachable from any other paths in G' that bypasses edge $e_{s,i}$ or otherwise the contribution of G' is 0; Eq. (195) is due to $e_{s,i}$ and $e_{s,j}$ for $j \in [n_s] \backslash \{i\}$ are in the same s-th partition, so that the parameter change from $\bar{\mu}_{s,j}$ to $\mu_{s,j}$ for $j \in [n_s] \backslash \{i\}$ neither affects the probability of $u(e_{s,i})$ is reachable from S nor affects the probability of the target node t is reachable from $v(e_{s,i})$, but only increases the probability of there is no bypass of $e_{s,i}$ (since there is less possibility of a path connects S and t by reducing $\bar{\mu}_{s,j}$ to $\mu_{s,j}$); Eq. (196) is because for the second term, we only require $u(e_{s,i})$ is reachable from S; Eq. (197) is because we multiply and divide the first term by $\sqrt{\mu_{s,i}}$ at the same time; Eq. (198) is because we divide the first term by $\sqrt{1-\mu_{s,i}}$ which is within $(0, 1)$.

Now we can summation over all $(s, i)_{s \in [L], i \in [n_s]}$, we have

$$
\begin{aligned}
&|r_t(S; \bar{\boldsymbol{\mu}}) - r_t(S; \boldsymbol{\mu})|\\
&\le \sum_{s,i} \sqrt{\frac{\zeta_{s,i}^2 \Pr[S \to u(e_{s,i}) \mid \cup_{\ell=1}^{s-1} \boldsymbol{\mu}_\ell]}{(1-\mu_{s,i})\mu_{s,i}}}\\
&\cdot \sqrt{\Pr[S \to u(e_{s,i}), e_{s,i} \text{ is live}, v(e_{s,i}) \to t, \text{ no bypass of } e_{s,i} \mid \cup_{\ell=1}^{s} \boldsymbol{\mu}_\ell, \cup_{\ell=s+1}^{L} \bar{\boldsymbol{\mu}}_\ell]}\\
&+ \sum_{s \in [L]} \sum_{i \in [n_s]} \eta_{s,i} \Pr[S \to u(e_{s,i}) \mid \cup_{\ell=1}^{s-1} \boldsymbol{\mu}_\ell]\\
&\le \sqrt{\sum_{i \in [m]} \frac{\zeta_i^2 p_i^{D,S}}{\mu_i(1-\mu_i)} \left(\sum_{s \in [L]} \sum_{i \in [n_s]} \Pr[S \to u(e_{s,i}), e_{s,i} \text{ is live}, v(e_{s,i}) \to t, \text{ no bypass of } e_{s,i} \mid \cup_{\ell=1}^{s} \boldsymbol{\mu}_\ell, \cup_{\ell=s+1}^{L} \bar{\boldsymbol{\mu}}_\ell] \right)}
\end{aligned}
$$

$$+\sum_{i\in[m]}\eta_i p_i^{D,S} \tag{199}$$

$$\leq \sqrt{\sum_{i\in[m]}\frac{\zeta_i^2 p_i^{D,S}}{\mu_i(1-\mu_i)}}\sqrt{L}+\sum_{i\in[m]}\eta_i p_i^{D,S}, \tag{200}$$

where Eq. (199) uses the Cauchy-Schwatz Inequality; Eq. (200) uses the critical property that for $e_{s,i}$ and $e_{s,j}$ with $i \neq j$, there does not exist any path that can contain both $e_{s,i}$ and $e_{s,j}$ as mentioned earlier, which means under the same parameter $(\cup_{\ell=1}^{s}\boldsymbol{\mu}_\ell, \cup_{\ell=s+1}^{L}\bar{\boldsymbol{\mu}}_\ell)$, the live edge graphs that satisfies $\{S \to u(e_{s,i}), e_{s,i} \text{ is live}, v(e_{s,i}) \to t, \text{ no bypass of } e_{s,i}\}$ and the live edge graphs that satisfies $\{S \to u(e_{s,j}), e_{s,j} \text{ is live}, v(e_{s,j}) \to t, \text{ no bypass of } e_{s,j}\}$ are disjoint (or otherwise contradicts the fact that there is no bypass of $e_{s,i}$ or $e_{s,j}$), so it holds that $\sum_{i\in[n_s]}\Pr[S \to u(e_{s,i}), e_{s,i} \text{ is live}, v(e_{s,i}) \to t, \text{ no bypass of } e_{s,i} \mid \cup_{\ell=1}^{s}\boldsymbol{\mu}_\ell, \cup_{\ell=s+1}^{L}\bar{\boldsymbol{\mu}}_\ell] \leq 1$.

Considering all the target nodes $t \in [V]$, we have $(B_v, B_1, \lambda) = (|V|\sqrt{L}, |V|, 1)$. ■

E.4 Probabilistic Maximum Coverage Bandit [8, 23]

In this section, we consider the probabilistic maximum coverage (PMC) problem. PMC is modeled by a weighted bipartite graph $G = (L, V, E)$, where L are the source nodes, V is the target nodes and each edge $(u, v) \in E$ is associated with a probability $p(u, v)$. The task of PMC is to select a set $S \subseteq L$ of size k so as to maximize the expected number of nodes activated in V, where a node $v \in V$ can be activated by a node $u \in S$ with an independent probability $p(u, v)$. PMC can naturally models the advertisement placement application, where L are candidate web-pages, V are the set of users, and $p(u, v)$ is the probability that a user v will click on web-page u.

PMC fits into the non-triggering CMAB framework: each edge $(u, v) \in E$ corresponds to a base arm, the action is the set of edges that are incident to the set $S \subseteq L$, the unknown mean vectors $\boldsymbol{\mu} \in (0, 1)^E$ with $\mu_{u,v} = p(u, v)$ and we assume they are independent across all base arms. In this context, the reward function $r(S; \boldsymbol{\mu}) = \sum_{v\in V}(1 - \prod_{u\in S}(1 - \mu_{u,v}))$.

Lemma 24. *PMC problem satisfies VM bounded smoothness condition (Condition 4) with coefficient* $(B_v, B_1) = (3\sqrt{2|V|}, 1)$.

Proof. We prove PMC satisfies VM condition by the definition of Gini-smoothness condition (Condition 7) and Lemma 3. First, we know $\frac{\partial r(S;\boldsymbol{\mu})}{\partial \mu_{u,v}} = \prod_{i\in S, i\neq u}(1-\mu_{i,j}) \leq 1$, thus $\gamma_\infty = 1$. Also $\sqrt{\sum_{u\in S, v\in V}\mu_{u,v}(1-\mu_{u,v}\frac{\partial(r(S;\boldsymbol{\mu})}{\partial \mu_{u,v}})^2} = \sqrt{\sum_{u\in S, v\in V}\mu_{u,v}(1-\mu_{u,v})\prod_{i\in S, i\neq u}(1-\mu_{i,v})^2} = \sqrt{\sum_{v\in V}\sum_{u\in S}\left(\mu_{u,v}\prod_{i\in S, i\neq u}(1-\mu_{i,v})\right)\left(\prod_{i\in S}(1-\mu_{i,v})\right)} \leq \sqrt{\sum_{v\in V} 1/4} = \sqrt{|V|/4}$, where the second last inequality uses the fact consider S coins and the i-th coin is up with prob. $\mu_{i,v}$, the first term $\left(\mu_{u,v}\prod_{i\in S, i\neq u}(1-\mu_{i,v})\right)$ corresponds to probability P_1 that only one coin is up and the second term $\left(\prod_{i\in S}(1-\mu_{i,v})\right)$ corresponds to the probability P_2 that all coins are down, thus $P_1 * P_2 \leq P_1(1-P_1) \leq 1/4$. Hence $\gamma_g = \sqrt{V/4}$. By Lemma 3, we have $(B_v, B_1) = (3\sqrt{2}\gamma_g, \gamma_\infty) = (3\sqrt{|V|/2}, 1)$. ■

F Experiments

In this section, we conduct experiments to validate our proposed algorithms. All experiments were performed on a desktop with i7-9700K CPU and 32 GB RAM.

F.1 Cascading Bandits

We consider cascading bandits as the application for CMAB with probabilistically triggered arms. More specifically, we choose the disjunctive cascading bandit problem to compare the performance of CUCB and BCUCB-T, where the reward function is $r(S; \boldsymbol{\mu}) = 1 - \prod_{i\in S}(1-\mu_i)$. Similar to the experimental setup in [16], we set batch-size $K = 10$ and generate 30 base arms with means randomly sampled from the uniform distribution $U(0, 0.1)$. Figure 1 shows the cumulative regrets of CUCB and BCUCB-T for $200,000$ rounds. We repeat each experiment 20 times and show the

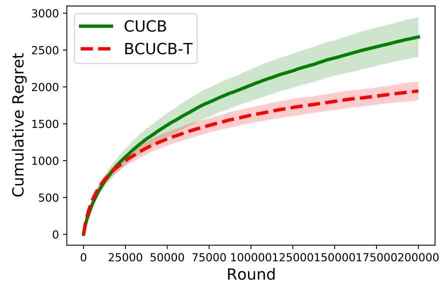

Figure 1: Cumulative regrets of CUCB and BCUCB-T for disjunctive cascading bandit.

average regret with shaded standard deviation. The average running time of CUCB and BCUCB-T for $200,000$ rounds are 13s and 19s, respectively. As shown in the figure, BCUCB-T achieves around 20% less regret than CUCB, owing to its variance-aware confidence radius control.

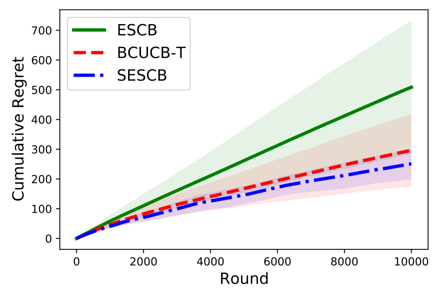

Figure 2: Cumulative regrets of ESCB, BCUCB-T and SESCB for the PMC problem.

F.2 Probabilistic Maximum Coverage

We consider Probabilistic Maximum Coverage (PMC) as the application for non-triggering CMAB. We generate a complete bipartite graph with 10 source nodes on the left and 20 target nodes on the right. The goal is to select 5 seed nodes from source nodes to influence as many as target nodes. The edge probabilities are randomly sampled from the uniform distribution $U(0.05, 0.06)$. We run ESCB [9], BCUCB-T and SESCB on this graph. For SESCB, we set sub-Gaussian parameter $C_1 = 3$ (according to Remark 7) and VM smoothness coefficient $B_v = 3\sqrt{2 \cdot 20}/2$. Since the number of base arms to be learned in PMC is large, we shrink the confidence intervals of all algorithms by $\alpha_\rho = 0.01$ to speed up the learning, e.g., for SESCB, $\bar{r}_t(S) = r(S; \hat{\boldsymbol{\mu}}_{t-1}) + \alpha_\rho \cdot \rho_t(S)$. We repeat each experiment 10 times and show the average regret. The average running time of ESCB, BCUCB-T and SESCB for $10,000$ rounds are 80s, 7s and 115s, respectively. Figure 2 shows the cumulative regrets with shaded standard deviations. SESCB achieves around 15% less regret than BCUCB-T, since it utilizes the independence of base arms while BCUCB-T; it also outperforms ESCB as ESCB is mainly designed for the linear reward case.