CASCADING BANDIT UNDER DIFFERENTIAL PRIVACY

Kun Wang[1], Jing Dong[2], Baoxiang Wang[3], Shuai Li[1]

[1]Shanghai Jiao Tong University,[2]University of Michigan
[3]The Chinese University of Hong Kong, Shenzhen

ABSTRACT

This paper studies *differential privacy (DP)* and *local differential privacy (LDP)* in cascading bandits. Under DP, we propose a UCB-based algorithm which guarantees ϵ-indistinguishability and a regret of $\mathcal{O}((\frac{\log T}{\epsilon})^{1+\xi})$ for an arbitrarily small ξ. This result significantly improves $O(\frac{\log^3 T}{\epsilon})$ in the previous work. Under (ϵ,δ)-LDP, we relax the K^2 dependence through the tradeoff between privacy budget ϵ and error probability δ, and obtain a regret of $\mathcal{O}(\frac{K\log(1/\delta)\log T}{\epsilon^2})$, where K is the size of the arm subset. This result holds for both Gaussian mechanism and Laplace mechanism by analyses on the composition. Extensive experiments corroborate our theoretic findings.

***Index Terms*—** Multi-armed bandit, Differential privacy, Online learning, Sequential decision making

1. INTRODUCTION

There exists a rich literature on multi-armed bandits (MAB) as it captures a most fundamental problem in sequential decision making - the exploration-exploitation dilemma [1, 2, 3, 4]. Despite its simpleness, the MAB is widely applied in a broad spectrum of applications such as online advertising and clinical trials. In the classic stochastic case of MAB, an agent is presented with K arms and is asked to pull an arm at each round through a finite-time horizon. Depending on the agent's choice, the agent will receive a reward and its goal is set to maximize the cumulative reward. The natural choice of the performance metric is thus the difference between the optimal reward possible and the actual reward received by the agent, which is termed *regret* in bandit literature.

The basic stochastic case of the MAB, though powerful, is insufficient to cope with the complexity of real applications. The cascading bandit problem is then proposed with the intention to better model complicated scenarios such as recommendation systems and search engines [5, 6, 7, 8, 9, 10]. This variant of MAB gives a realistic depiction of user click behavior in these applications. In each round, the agent recommends a list of items to the user. The user checks from the start of the list and stops at the first attractive item, which may be manifested by clicks in web recommendations. Then the agent receives the feedback in the form of user's click information.

However, the cascading bandit model raises concern of privacy despite its superiority in the depiction of user's click behavior. Many applications rely heavily on sensitive user data, which reveals the preference of a particular user. If additional measures were not taken, one can easily get this information by analyzing the algorithm's output [11]. To that end, *differential privacy (DP)* is proposed and becomes a standard in privacy notion because of its well-defined measure of privacy [11, 12, 13, 14]. A desirable algorithm that can protect differential privacy should guarantee the difference between outputs of this algorithm over two adjacent inputs is insignificant. Under the canonical definition of DP, bandit algorithms are known to enjoy an upper bound of $\mathcal{O}(\log^3 T)$ [15] for upper confidence bound (UCB)-based method and $\mathcal{O}(\log T)$ for successive elimination (SE)-based method [16]. Nevertheless, a number of problems still remain unsolved: First, these algorithms are merely applicable to basic MAB and lose efficacy in more complex scenarios such as recommender systems. Designing private algorithms for cascading bandit can effectively cope with this condition. Second, the $\mathcal{O}(\log T)$ regret [1, 17, 5] is well known for non-private UCB-based method, but it remains an open problem whether it is possible to achieve the same regret under DP. Third, compared to the definition of differential privacy, there is a much stricter definition of privacy guarantee known as *local differential privacy (LDP)*. It does not need a trusted center to guarantee privacy and gains attractions in an increasing number of scenarios. A natural problem is whether there exists an algorithm that can achieve LDP for cascading bandit.

In this paper, we study cascading bandit under both DP and LDP. Under DP, the post-processing lemma well known in DP and a more advanced hybrid mechanism with a tighter utility bound for private data release are used as our key tools. Building on this, we propose a UCB-based algorithm and give the $\mathcal{O}(\log^{1+\xi} T)$ upper bound, for an arbitrarily small ξ. This is not only the first result for cascading bandit, but also the best result for UCB-based method under DP. Under LDP, we relax the regret dependence on the size of the arm subset K through the balance between privacy budget ϵ and error probability δ. This holds for both Gaussian mechanism and Laplace mechanism by analyses on composition. To be specific, the main contributions of this paper are summarized as follows:

- For the cascading bandit problem under ϵ-DP, we propose

novel algorithms with regret bound of $\mathcal{O}((\frac{L\log T}{\epsilon})^{1+\xi})$, where L is the arm number.
- For the cascading bandit problem under (ϵ,δ)-LDP, we propose two approaches upper bounded by $\mathcal{O}(\frac{K\log 1/\delta\log T}{\epsilon^2})$.
- Experiments on synthetic data validate our theoretical results.

2. PROBLEM FORMULATION

2.1. Cascading bandit

A cascading bandit problem can be denoted by a tuple $B=(E,P,K)$, where $E=\{1,...,L\}$ is a set of L base items (also named arms), P is a probability distribution over $[0,1]^E$ and its expectation is $(\bar{w}(a_i))_{i=1}^L$. $K\leq L$ represents the number of recommended items each time. Through a time horizon of T, at each round $t\leq T$, weight $w_t\in\{0,1\}^E$ is instant Bernoulli reward drawn from P.

The problem proceeds iteratively, at each round t, the algorithm is asked to recommend a list of items $A_t=(a_1^t,a_2^t,...,a_K^t)\in\Pi_K(E)$, where $\Pi_K=\{(a_1,...,a_k):a_1,...,a_k\in E,a_i\neq a_j \text{ for any } i\neq j\}$. When the user receives the recommended list, he reviews the list from the top and stops at first attractive item C_t. The user's feedback to the algorithm is $(w_t(a_i))_{i=1}^{C_t}$, where $C_t=\arg\min\{1\leq k\leq K: w_t(a_k^t)=1\}$. The reward function is $f(A,w)=1-\prod_{k=1}^K(1-w(a_k))$. Note that, we can get $w_t(a_k^t)=\mathbb{1}\{C_t=k\}\quad k=1,...,\min\{C_t,K\}$. We say arm e is pulled if $w_t(e)=1$. Besides, we say item e is observed at time t if $e=a_k^t$ for some $1\leq k\leq\min\{C_t,K\}$. The cumulative regret $R(T)$ is defined as follows:

$$R(T)=\mathbb{E}\left[\sum_{t=1}^{T}R(A_t,w_t)\right],$$

where $R(A_t,w_t)=f(A^*,w_t)-f(A_t,w_t)$ is the instant stochastic regret at time t and $A^*=\arg\max_{A\in\Pi_K(E)}f(A,w_t)$ denote the optimal list at round t. Our goal is to minimize this cumulative regret to a sub-linear term of time horizon T. Next, we maintain a mild assumption that are common among cascading bandits literature.

Assumption 1. *The weights w are distributed as, $P(w)=\prod_{e\in E}P_e(w(e))$, where P_e is a Bernoulli distribution with mean $\bar{w}(e)$.*

Under this assumption, the expected reward for any list A, the probability that at least one item in A is attractive, can be expressed as $\mathbb{E}[f(A,w)]=f(A,\bar{w})$. At last, the optimal action is unique, i.e. $\Delta_{min}=\min_{e\in[L]\setminus A^*}\{\bar{w}^*-\bar{w}(e)\}>0$, where $\bar{w}^*$ is the maximum of $(\bar{w}(a_i))_{i=1}^L$.

2.2. (Locally) differential privacy

We study the problem of cascading bandits under both the classic (ϵ,δ)-differential privacy definition and (ϵ,δ)-local differential privacy. Firstly, we first give the rigorous definition of differential privacy.

Definition 1 (Differential Privacy). *Let $D=\langle x_1,x_2,\ldots,x_T\rangle$ be a sequence of data with domain X^T. Let $A(D)=Y$, where $Y=\langle y_1,y_2,\ldots,y_T\rangle\in Y^T$ be T outputs of the randomized algorithm A on input D. A is said to preserve (ϵ,δ)-differential privacy, if for any two data sequences D,D' that differ in at most one entry, and for any (measurable) subset $U\subset Y^T$, it holds that*

$$P(A(D)\in U)\leq e^{\epsilon}\cdot P(A(D')\in U)+\delta\,.$$

If $\delta=0$, then we say the algorithm satisfies ϵ-differential privacy.

The local differential privacy model, different from the differential privacy, requires masking data with noise before the accumulation of data in order to circumvent the need of a trusted center, which leads to a more promising privacy guarantee. Formally, the (ϵ,δ)-local differential privacy is defined as follows.

Definition 2 (Local Differential Privacy). *A mechanism $A: X\to Y$ is said to be (ϵ,δ)-local differentially private or (ϵ,δ)-LDP, if for any $x,x'\in X$, and any (measurable) subset $U\subset Y$, there is*

$$P(A(x)\in U)\leqslant e^{\epsilon}\cdot P(A(x')\in U)+\delta\,.$$

If $\delta=0$, then we say the algorithm satisfies ϵ-local differential privacy.

3. CASCADING BANDIT UNDER DIFFERENTIAL PRIVACY

In this section, we study cascading bandit under DP. The system operates as follows: Each round, the algorithm receives the noisy data manipulated by the database. Then the algorithm recommends a list of items to the user based on the noisy data. Finally, the user checks the lists and sends the feedback to the database.

The difficulty of this problem is governed by two parts: controlling the injected noise in the database and properly adjusting the confidence bound in the UCB-based algorithm. A tree-based mechanism from [18] has been used by [19, 15] to maintain the privacy of the stream data in the MAB setting. Tree-based mechanism solves this problem in an elegant way by adding only $\log t$ times of noises, which can control overall noises to an acceptable extent. However, the improper use of the tree-based mechanism over stream data in the database and adjusting confidence intervals based on the inappropriate utility bound for the tree-based mechanism in the algorithm leads to some sub-optimal regret bounds.

In this paper, we improve existing methods. Let $T_t(e)$ denote the pulled number of base arm e until time t. We find the regret of the algorithm is dominated by the maximal length of stream data imported into tree-based mechanism. Previous methods [19, 15] can only upper bound this length by a function of time horizon T as they use tree-based mechanism over the whole stream data. Directly upperbounding the stream data length with a function of time T leads to an undesirable $\log^3 T$ regret. In this work, first, we use a novel hybrid mechanism (instead of a tree mechanism). The hybrid mechanism transfers its dependence of the whole time horizon T to instant time moment t. Besides, we utilize the post-processing

lemma well-known in differential privacy, converting the privacy guarantee of whole algorithm's output to L items aimed at transforming this dependence on instant time t to the dependence on $T_t(e)$, for all e. Based on these two measures, we can control the pulled numbers of any sub-optimal arm to $\log^{1+\xi} t$, which is the dominant term in regret. Lacking any one of these two tools, the regret would degrade to $\mathcal{O}(\log^3 T)$.

We now describe our algorithm under DP. First, the algorithm receives noisy empirical mean $\hat{w}_{t-1}(e)$ from database (using hybrid mechanism to guarantee privacy) and calculates the upper confidence bound of the empirical mean for each arm e. Compared to classic UCB algorithm, we add an extra term due to the injected noise from the hybrid mechanism. The algorithm then recommends chosen K arms to users. Finally, the user checks this list of arms and sends the feedback to the database. Our method is depicted in Algorithm 1.

Algorithm 1: Cascading-UCB under DP.

Input: ϵ, the differential privacy parameter
Instantiate L Hybrid Mechanisms with $\epsilon' = \frac{\epsilon}{L}$.
Observe w_0.
$\forall e \in E : T_0(e) \leftarrow 1, \hat{w}_1(e) \leftarrow w_0(e)$.
for $t = 1, 2, ..., T$ **do**
  Receive noisy empirical mean $\hat{w}_{t-1}(e)$ from hybrid mechanism.
  $\forall e \in E : U_t(e) =$
  $\hat{w}_{t-1}(e) + \sqrt{\frac{1.5 \log t}{T_{t-1}(e)}} + \frac{3c_1 L \log^{1.5} T_{t-1}(e) \log t}{\epsilon T_{t-1}(e)}$
  Let $a_1^t, ..., a_K^t$ be K items with largest private $U_t(e)$ in order.
  User play $A_t \leftarrow (a_1^t, ..., a_K^t)$ and observe the last click position $C_t \in \{1, ..., K, \infty\}$.
  $H \leftarrow \min\{C_t, K\}$.
  for $k = 1, ..., H$ **do**
    e $\leftarrow a_k^t$, $T_t(e) \leftarrow T_{t-1}(e) + 1$.
    $w_t(e) = \mathbb{1}\{H = C_t\}$.
    Insert $w_t(e)$ to hybrid mechanism for arm e.

Lemma 1 ([18], Utility Bound For Hybrid Mechanism)**.** *In the continual release procedure, the hybrid mechanism preserves ϵ-differential privacy and with probability $1 - \gamma$, the added noise $\mathcal{N}$ to the data at time n satisfies following inequality:*

$$|\mathcal{N}| \leq \frac{c_1 \log^{1.5} n \log 1/\gamma}{\epsilon},$$

where c_1 is a constant.

Lemma 2 (Post-Processing Lemma)**.** *If the sequence $(w_t(e))_{t=1}^{T_e}$ for all arm e is $\frac{\epsilon}{L}$-differentially private, then the Algorithm 1 is ϵ-differentially private.*

Based on Lemma 1, one can construct high probability events that empirical mean outside of this confidence interval happens with an arbitrary small probability. This is the basis of UCB algorithms. The Lemma 2 ensures Algorithm 1 is ϵ-differentially private.

Theorem 1. *Algorithm 1 guarantees ϵ-DP.*

Theorem 2. *The regret of Algorithm 1 is upper bounded by:*

$$R(T) \leq \sum_{e=K+1}^{L} \frac{192}{\Delta_{e,K}} \left(\frac{c_1 L}{\epsilon} \log T \right)^{1+\xi} + \frac{2\pi^2}{3} L + c_2 \,,$$

where ξ is an arbitrary small positive real value and c_1, c_2 are constants independent of the problem, $\Delta_{e,K}$ is the difference between $\bar{w}_e$ and $\bar{w}_K$.

To the best of our knowledge, this is not only the first result for cascading bandit, but also the first $\mathcal{O}(\log^{1+\xi} T)$ regret of UCB-based algorithm. We greatly improve the existing regret bound from $\mathcal{O}(\log^3 T)$ to $\mathcal{O}(\log^{1+\xi} T)$ by the post-processing lemma and the utility bound for hybrid mechanism. This matches the well known non-private lower bound up to an arbitrarily small poly-log factor. Besides, our method can extend to other bandits model under DP such as combinatorial semi-bandit and basic MAB.

4. CASCADING BANDIT UNDER LOCAL DIFFERENTIAL PRIVACY

Under local differential privacy, each round when the user browses through the list, he directly sends noisy feedback to the database rather than send data to the database and let the database add noises to guarantee privacy. This circumvents the need of a trusted center, so it has a much stronger privacy guarantee. Under this circumstance, we need to protect feedback at every time t with injected noises and let it satisfy (ϵ, δ)-local differential privacy. The database only plays a part in storage and integration. Besides, if we directly inject Laplace noise, which is most common in differential privacy, the regret of the algorithm shows a quadratic dependence on the size of the arm subset K. This is a notorious side-effect in real applications. For instance, in a personalized recommendation scenario, when a large amount of items is recommended at one time, the algorithm becomes impractical. To that end, we will provide two approaches to figure out the above two problems.

4.1. Gaussian mechanism under local differential privacy

Our first approach is based on Gaussian mechanism. It operates as follows: Each round, the algorithm receives the noisy data from the database and recommends a list of items based on the noisy data. Then the user receives the list of items and checks the list from the top and stops at the first attractive position. The user directly injects Gaussian noise into the data (feedback) and sends noisy data to the database for integration. The reason why we choose Gaussian noise is that Gaussian mechanism has superiority over Laplace mechanism[19], as it is sensitive to L_2-norm instead of L_1-norm. This key property leads to our design of a $\sqrt{K}$-dependent confidence interval for injected noises in the algorithm. This adjustment to the confidence interval, directly lessens the dominating effect on the regret bound and eliminate the K^2 dependency. Next, we give Lemma 3 to demonstrate this effect in detail.

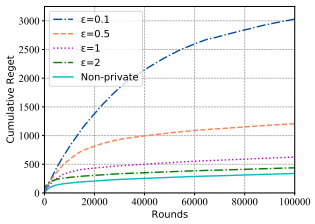

Fig. 1. DP, vary ϵ.

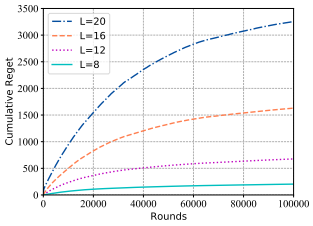

Fig. 2. DP, vary L.

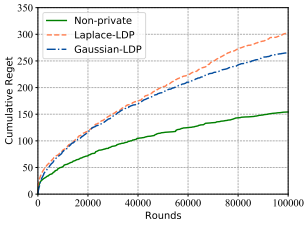

Fig. 3. LDP, $\epsilon = 0.2$.

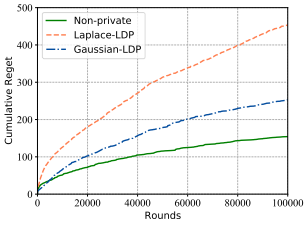

Fig. 4. LDP, $\epsilon = 0.5$.

Lemma 3 ([20]). *Let* $\Delta_f = \max_{D,D'} \|f(D) - f(D')\|_{L_2}$, *then* $\forall \delta \in (0,1)$, *and* $\sigma > \frac{\Delta_f}{\epsilon}\sqrt{2\log\frac{1.25}{\delta}}$, $M(D) = f(D) + \mathcal{N}(0,\sigma^2)$ *satisfied* (ϵ,δ)*-differential privacy.*

Lemma 3 gives the Gaussian mechanism privacy guarantee, if we inject $\mathcal{N}(0,\sigma^2)$ to the data of each item. Let u denote the expected mean of a Gaussian noise X, and $\bar{X}$ denotes the empirical mean of specified Gaussian noise. Based on the concentration bound for Gaussian distribution, we will get $u \in \left[\bar{X} - \sigma\sqrt{\frac{2\log 2/\gamma}{T_{t-1}(e)}}, \bar{X} + \sigma\sqrt{\frac{2\log 2/\gamma}{T_{t-1}(e)}}\right]$ with probability at least $1-\gamma$. In light of this, the upper confidence bound term in the algorithm can be enlarged accordingly to ensure the normal operation of the UCB algorithm. The new algorithm is denoted as Algorithm 2.

Theorem 3. *Algorithm 2 guarantees* (ϵ,δ)*-LDP.*

Theorem 4. *The regret of Algorithm 2 is upper bounded by:*

$$R(T) \leq \sum_{e=K+1}^{L} \frac{2\left(2\sqrt{1.5} + 8/\epsilon\sqrt{K\log\frac{1.25}{\delta}}\right)^2}{\Delta_{e,K}} \log T\,.$$

As Theorem 4 shows, the regret has the linear dependence on K, which is a great improvement compared to the Laplace mechanism with a K^2 dependence. Compared with the non-private theoretical guarantee, Theorem 4 implies that we can achieve optimal order in locally differentially private cascading bandit and with only additional K dependence for privacy protection, which is a bit surprising given the previous work about (locally) differentially private learning.

4.2. A new approach by composition theorem

In this part, we give another approach that also eliminates the K^2 dependence. Through detailed analysis of the regret of Gaussian mechanism, we get a deeper insight and identify a trade-off between privacy budget ϵ and error probability δ in regret: careful comparison of the difference between the regret of Gaussian mechanism and that of Laplace mechanism, we will find the dependence on privacy budget $\epsilon(\frac{\epsilon}{K})$ is exchanged by an additional multiplicative $\delta(\log\frac{1}{\delta})$ term. For this way about reducing dependence on K at the expense of a little δ, a natural question is whether can we extend the same idea into a more general situation. This enlightens us to give a new approach under LDP.

Lemma 4 ([21], Corollary 4.1). *For any* $\epsilon \in (0, 0.9]$ *and* $\delta \in (0,1]$, *if the database access mechanism satisfies* $(\sqrt{\epsilon^2/4k\log(e+\epsilon/\delta)}, \delta/2k)$*-local differential privacy on each query output, then it satisfies* (ϵ,δ)*-local differential privacy under k-fold composition.*

Building on Lemma 4, ensuring each item $\frac{\epsilon}{\sqrt{K}}$- indistinguishable instead of $\frac{\epsilon}{K}$-indistinguishable is enough to guarantee privacy of whole algorithm. Thus any ϵ-LDP mechanism can achieve half dependence on K at the cost of a $\log 1/\delta$ term in regret while simultaneously ensuring their (ϵ,δ)-LDP. We give an illustrating example to highlight the impact of the above Lemma. By Lemma 4, one can improve the regret of Laplace mechanism to $\mathcal{O}(K)$ dependence in regret, which is the same as Gaussian mechanism.

Theorem 5. *Cascading-UCB algorithm using Laplace mechanism with parameter* $\epsilon' = \frac{\epsilon}{\sqrt{4K\log(e+\epsilon/\delta)}}$ *and under K-fold composition each round achieves* $\mathcal{O}\left(\frac{K\log 1/\delta\log T}{\epsilon^2}\right)$ *regret while ensuring* (ϵ,δ)*-local differential privacy.*

By the definition of LDP, Laplace mechanism that attains $(\epsilon,0)$-LDP is also capable of offering (ϵ,δ)-DP protection. Using Lemma 4, every item observed in the list masked with a $Lap\left(\frac{\sqrt{4K\log(e+\epsilon/\delta)}}{\epsilon}\right)$ noise is enough to guarantee (ϵ,δ)-DP. An UCB algorithm with confidence interval of $\frac{4}{\epsilon}\sqrt{\frac{6K\log(e+\epsilon/\delta)\log t}{T_{t-1}(e)}} + \sqrt{\frac{3\log t}{2T_{t-1}(e)}}$ then suffers only $\mathcal{O}\left(\frac{K\log 1/\delta\log T}{\epsilon^2}\right)$ regret while ensuring privacy guarantee.

5. EMPIRICAL RESULTS

In this section, we provide extensive empirical results that corroborate our theoretical findings. We conduct experiments under DP and LDP for Gaussian and Laplace mechanism with varying values of ϵ. The Laplace mechanism uses the initial form. The experiments were proceeded with $L = 20, K = 4, \delta = 10^{-3}$ unless otherwise indicated. Without loss of generality, all noises are multiplied by 0.01 to ensure most of the samples distributed among $[0,1]$. Fig.(1) and Fig.(2) show regret varies with different privacy budget ϵ and arm number L under DP. Fig.(3) and Fig.(4) show our algorithm's advantage over previous work under LDP. They both validate our theoretical results with respect to ϵ and L.

6. CONCLUSION AND FUTURE WORK

In this paper, we study the differential privacy and local differential privacy in the cascading bandit setting. Under DP, we design a new UCB-based algorithm with state-of-the-art $\mathcal{O}(\log^{1+\xi} T)$ regret. Under LDP, we utilize the tradeoff between ϵ and δ, relaxing the $\mathcal{O}(K^2)$ dependence to $\mathcal{O}(K)$. We conjecture the algorithm under DP can be further improved. For example, distributing privacy budget based on arm's difference may improve regret order. We leave this direction for future work.

7. REFERENCES

[1] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, vol. 47, no. 2, pp. 235–256, 2002.

[2] Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire, "The nonstochastic multiarmed bandit problem," *SIAM journal on computing*, vol. 32, no. 1, pp. 48–77, 2002.

[3] Sébastien Bubeck and Nicolò Cesa-Bianchi, "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," *Found. Trends Mach. Learn.*, vol. 5, no. 1, pp. 1–122, 2012.

[4] Omar Besbes, Yonatan Gur, and Assaf Zeevi, "Stochastic multi-armed-bandit problem with non-stationary rewards," *Advances in neural information processing systems*, 2014.

[5] Branislav Kveton, Csaba Szepesvari, Zheng Wen, and Azin Ashkan, "Cascading bandits: Learning to rank in the cascade model," in *International Conference on Machine Learning*, 2015.

[6] Branislav Kveton, Zheng Wen, Azin Ashkan, and Csaba Szepesvari, "Combinatorial cascading bandits," in *Advances in Neural Information Processing Systems*, 2015.

[7] Shi Zong, Hao Ni, Kenny Sung, Nan Rosemary Ke, Zheng Wen, and Branislav Kveton, "Cascading bandits for large-scale recommendation problems," in *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence*, 2016.

[8] Shuai Li, Baoxiang Wang, Shengyu Zhang, and Wei Chen, "Contextual combinatorial cascading bandits," in *International conference on machine learning*, 2016.

[9] Wang Chi Cheung, Vincent Tan, and Zixin Zhong, "A thompson sampling algorithm for cascading bandits," in *The 22nd International Conference on Artificial Intelligence and Statistics*, 2019.

[10] Kun Wang, Canzhe Zhao, Shuai Li, and Shuo Shao, "Conservative contextual combinatorial cascading bandit," *arXiv preprint arXiv:2104.08615*, 2021.

[11] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of cryptography conference*, 2006.

[12] Cynthia Dwork, Moni Naor, Toniann Pitassi, and Guy N Rothblum, "Differential privacy under continual observation," in *Proceedings of the forty-second ACM symposium on Theory of computing*, 2010.

[13] Cynthia Dwork, Aaron Roth, et al., "The algorithmic foundations of differential privacy.," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3-4, pp. 211–407, 2014.

[14] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016.

[15] Nikita Mishra and Abhradeep Thakurta, "(nearly) optimal differentially private stochastic multi-arm bandits," in *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, 2015.

[16] Touqir Sajed and Or Sheffet, "An optimal private stochastic-mab algorithm based on optimal private stopping rule," in *International Conference on Machine Learning*. PMLR, 2019, pp. 5579–5588.

[17] Wei Chen, Yajun Wang, and Yang Yuan, "Combinatorial multi-armed bandit: General framework and applications," in *International Conference on Machine Learning*, 2013.

[18] T-H Hubert Chan, Elaine Shi, and Dawn Song, "Private and continual release of statistics," *ACM Transactions on Information and System Security*, vol. 14, no. 3, pp. 1–24, 2011.

[19] Xiaoyu Chen, Kai Zheng, Zixin Zhou, Yunchang Yang, Wei Chen, and Liwei Wang, "(locally) differentially private combinatorial semi-bandits," in *International Conference on Machine Learning*, 2020.

[20] Jun Zhao, Teng Wang, Tao Bai, Kwok-Yan Lam, Zhiying Xu, Shuyu Shi, Xuebin Ren, Xinyu Yang, Yang Liu, and Han Yu, "Reviewing and improving the gaussian mechanism for differential privacy," *arXiv preprint arXiv:1911.12060*, 2019.

[21] Peter Kairouz, Sewoong Oh, and Pramod Viswanath, "The composition theorem for differential privacy," in *International conference on machine learning*, 2015.