We thank the reviewers for their insightful comments, which have helped us improve the paper. We now discuss the issues raised by the reviewers and our proposed changes in the final version to resolve these. We omit discussion for each suggested change which we will directly implement in the final version.

Simplified Regret Upper Bound: We define $D_{min} := \min_i D_i$, $D_{max} := \max_i D_i$, and $\Delta_{i,i+1} := (\mu_i - \mu_{i+1})$. The regret admits the bound $R_T^{(1-1/e)} \leq O(\frac{1}{\epsilon}\log(\frac{1}{\epsilon})) + \frac{32K_g(K-K^*_\epsilon)}{\min_{i\in[K^*_\epsilon,\dots,K_g]}\Delta_{i,i+1}}\log(T)$ for any $\epsilon > 0$. Notably $\Delta_{i,i+1}$ for $i=1$ to $(K^*_\epsilon - 1)$ do not appear in the bound. Furthermore, $(1-\epsilon)D_{min} \leq K^*_\epsilon \leq \min(K, (1-\epsilon)D_{max} + 1)$ for any $\epsilon \geq 0$, and $D_{min} \leq K_g \leq \min(K, D_{max})$. Using these bounds we can simplify the regret upper bound.

Reviewer #4: • We will add the above simplified regret upper bound in the final version of the paper.

• The problem resolves to semi bandit only if *all the delays are equal*. When all the delays are less than K (number of arms) but distinct we do not have reduction to the semi-bandit case. Nonetheless, we simulate the results for systems where $\max_i D_i > K$ ($K = 20$ and $\max_i D_i = 40$) and observe similar behavior as reported in the current version of the paper. We will report the results of the new experiments in the final version of the paper.

Reviewer #5 • The upper bound for the K^*-set instance (the instance used in the lower bound proof), is $O(1) + \frac{K^*(K-K^*)\log(T)}{\Delta}$. This follows by choosing any $0 < \epsilon < 1/K^*$, using $K_g = K^*$, and observing that for all $i \in \{1,\dots,K^*\}$ and $j \in \{K^*+1,\dots,K\}$, $\Delta_{ij} = \Delta$.

• The lower bound is $\Omega((K-K^*)/\Delta(K,K^*))$ as proved in the paper. Therefore, the dependence on $\Delta(K,K^*)$ is statistically optimal. However, the lower bound is smaller than the upper bound by a factor of $K_g \leq \min(K, D_{max})$. Closing this gap is left as future work.

• We would like to emphasize here that a key message of our paper is that in the offline setting the simple greedy algorithm is (1-1/e) optimal. The goal of the next result is to show that the UCB greedy algorithm has low regret *w.r.t the greedy algorithm*, and the delay information is not required for this result. This does not rule out the possibility of improved algorithms that use the delay information, both in the offline and online settings.

• A gap-independent regret upper bound of $2H(4)/(D_{min})^4 + \sqrt{32K_gKT\log(T)}$ holds. We will add the result and its proof in the supplementary material for the sake of completeness.

• The use of ϵ in Prop. 3.4 is purely technical to avoid tie breaking in favor of arm 3. Indeed, if arm 3 has reward 1, same as arm 1 and arm 2. The greedy algorithm will play 13231323… which gives the optimal reward. However, for any $\epsilon > 0$ greedy plays suboptimally as 12341234… Therefore, using $\epsilon > 0$ is required for the proof. We will make the statement of Prop 3.4 rigorous by swapping the order of "there exists" and mentioning $\epsilon < 1$.

• Prop 3.5 states suboptimality of an algorithm called the greedy-per-round which plays the available arm with the highest "reward/delay". Here, we have $(1+\epsilon)/(K+1) > 1/(K+1)$, not $(1+\epsilon)/(K+1) > 1$. We will explain the greedy-per-round algorithm clearly and explicitly mention the inequality $(1+\epsilon)/(K+1) > 1/(K+1)$.

• The UCB-greedy algorithm may enter suboptimal cycles but as high reward arms become available, UCB strategy ensures they are played with high probability. We will consider the high probability regret upper bound as a future work.

• For Fig. 1.b, where $K^* = 20$, all delays are equal. Here we highlight the constant regret behavior when $K^* = K$.

Reviewer #6 • The hardness result stated in the paper is with respect to the unary representation of the number of arms, i.e. time complexity is measured w.r.t. the number of arms, K. Specifically, the number of bits required to express the total number of timeslots T can be expressed in $\log(\Pi_i D_i)$ bits which is polynomial in the number of arms. Now recall that the PINWHEEL scheduling under consideration is dense. Therefore, by construction, for the MAXREWARD instance the delays will satisfy $\sum_{i=1}^K 1/D_i = 1$. Now the decision problem for the MAXREWARD becomes "$OPT = T\sum_{i=1}^K a_i/D_i$?" Thus the decision problem is expressible in polynomial number of bits. Therefore, the reduction is valid. We will add the above explanation in the proof.

• We thank the reviewer for pointing out this key typographical error. We will replace $\min_{t'>=1}$ with $\max_{t'<=t}$.

• Our focus in terms of the offline problem is studying the complexity of the problem both in negative and positive direction. Specifically, we prove the hardness of the optimization problem (in number of arms), and in the positive side we show that the greedy heuristics under full information is a (1-1/e) approximation. Indeed, as pointed out by the reviewer, we leave open the design of an algorithm with approximation larger than (1-1/e) [or proving approximating beyond (1-1/e) is hard]. We will highlight this message clearly and specify the future works.

• We have checked the existing literature very carefully and we believe that the paper introduces and studies a new problem. Furthermore, the other reviews also acknowledge the novelty of the problem. Therefore, we request the reviewer to re-evaluate the work, as mentioned at the end of the review.