We thank the reviewers for their insightful comments. The numbered citations refer to references in the submitted paper. The remaining additional citations are listed at the end of the page.

Reviewers 3 and 4: motivation. We admit the comment of **Reviewer 4** about the lack of motivation on *the combined scenario of nonstochastic bandits with corruption*. The isolated motivation about bandits with corruption is mentioned in 2nd paragraph of the introduction (L25-35) with extensive recent results [8,9,11,14,15,17,18,25]. The motivation about nonstochastic bandits is given in the 3rd paragraph (L36-43). There exist application scenarios where either 'nonstochastic bandits' or 'bandits with corruption' in insolation fail to fully characterize the underlying model. To response *the combined motivation* and strengthen the motiviation, we will add one such example, described below, to the paper. This will also address **Reviewer 3**'s comments on *the real example of targeted attacks* and **Reviewer 4**'s comments on attacks to the *actual vs. observation of the reward.*

A concrete example of nonstochastic bandits with corruptions is the Online Shortest Path Routing (OSPR) problem under the denial of service (DoS) attacks. OSPR is a classic example of MAB problems [20]. And there is also extensive research on routing under DoS attacks, including the recent work [Zhou et al., 2019] focusing on bandit modeling of this scenario. OSPR could be reasonably modeled as nonstochastic bandits when the delays on the links change dynamically [György et al. 2007], or once is it difficult to characterize the combined distribution of a path including multiple links [20]. In this nonstochastic scenario, the DoS attack could be modeled by our bandit with targeted corruptions. Specifically, the DoS attacker can be aware of the selected paths by detecting the transmitted packets over the path and manipulate the latency of the selected path by flooding the path with dummy packets. Also, the budget of the attacker is simply the available resources for the DoS attacker to keep her undetectable. Arguably, none of 'nonstochastic bandits' and 'bandits with corruption' alone would suffice to fully characterize the underlying model here. We believe presenting this example in the introduction would prove useful to connect the dots between *nonstochastic bandits* and *bandits with corruption*.

Reviewer 2: experiments. We will add the following experiment to the supplementary material. The goal here is to compare `ExpRb` with `Exp3`. We constructed a simple scenario where the attacker follows a simple policy that attacks the optimal arms (see L172-181 of the paper) with 1 high-reward and $K-1$ low-reward arms. In Fig. 1, we report the average regret of 100 independent runs, with $\Phi = O(\sqrt{T})$. The results show that the regret of `Exp3` is largely degraded with the attack, while `ExpRb` achieves a sublinear regret. This experiment is not meant to be exhaustive, rather it is intended to validate the theoretical results and illustrate the potential of our approach.

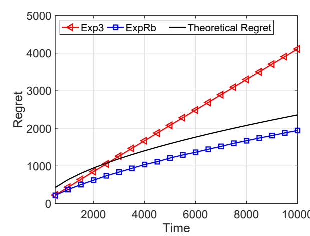

Figure 1: `ExpRb` vs. `Exp3`

Reviewers 3 and 4: lower bound. R3/Q1: Different bandit algorithms tolerate different levels of corruption. Hence, finding a more refined bound for the attacker's budget is highly algorithm-specific and we were not able to generalize our result in Corollary 2 for algorithms beyond `Exp3`. **R3/Q2:** Without substantial change in the proofs, the lower bound in Theorem 1 can be applied to the algorithms that are aware of the existence of corruptions, but do not know the budget. We will add a remark after Theorem 1 to involve this case. **R4 (the impossibility Theorem 1):** The impossibility result is a generic result for any attack-agnostic algorithms; one can have more refined versions for specific algorithms such as our result for `Exp3` (see **R1/Q1**). **R4: the algorithm needs to know the attacker's budget)**: Yes, it is a negative result on the attack-agnostic model. `ExpRb` algorithm is parameterized by a robustness parameter γ. Our further results which will appear in the future version can characterize the regret as a function demonstrating the regret reduction with improper γ. As future work, it is promising to dig out more interesting results on the attack-agnostic case.

Additional comments: Reviewer 1: Both questions are valid and due to a mistake when defining $\mathcal{T}_i$, $t_i(n)$ and N_i in the proof of Lemma 8. Thanks for your careful reading. $\mathcal{T}_i$ should be referred to (Do you simply mean: "$\mathcal{T}_i$ denotes") the set of time slots that the i-th arm is selected and the selection probability for arm i is lower than the previous one (only in this case, $\delta(t)$ will be larger than 0). And accordingly, $t_i(n)$, $n = 1, 2, \ldots, N_i$ are the indices for those time slots. By redefining those variables, we hope we can clarify the reviewer's concern as follows. **R1/Q1:** We only consider the time slots that the selection probability for the i-th arm is smaller than the previous, so the algorithm falls into the case in Line (5) of algorithm 1, but not case (1) in line 522. By checking the conditions for Equation (8), we have $\tilde{p}_i(t_i(n)) = p_i(t_i(n))$. **R1/Q2:** Yes, we do require that condition. This inequality holds, since we only consider the time slots that the selection probability for the i-th arm is smaller than the previous one.

Additional References

[Zhou et al., 2019] Zhou, P., Xu, J., Wang, W., et al. (2019). Toward Optimal Adaptive Online Shortest Path Routing With Acceleration Under Jamming Attack. *IEEE/ACM Transactions on Networking*, 27(5), 1815-1829.

[György et al. 2007] György, A., Linder, T., Lugosi, G., and Ottucsák, G. (2007). The on-line shortest path problem under partial monitoring. *Journal of Machine Learning Research*, 8(Oct), 2369-2403.