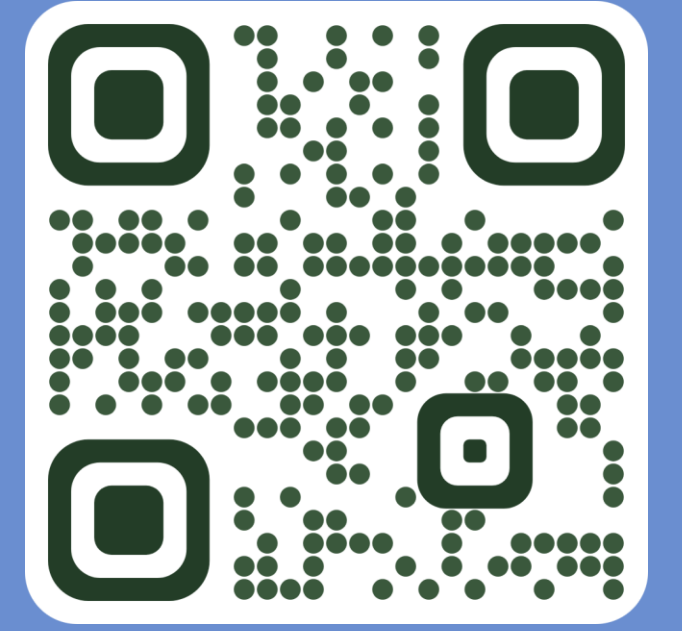


# Fairness of Exposure in Online Restless Multi-armed Bandits

Archit Sood<sup>1</sup>, Shweta Jain<sup>1</sup>, Sujit Gujar<sup>2</sup>

<sup>1</sup> IIT Ropar, India ; <sup>2</sup> IIIT Hyderabad, India

[2020mcb1230@iitrpr.ac.in](mailto:2020mcb1230@iitrpr.ac.in) [shwetajain@iitrpr.ac.in](mailto:shwetajain@iitrpr.ac.in) [sujit.gujar@iiit.ac.in](mailto:sujit.gujar@iiit.ac.in)



## Healthcare Intervention

Need to provide medical intervention to patients, but only have a limited budget to do so (number of doctors, number of available rooms, etc.). There are many things to consider for any potential solution.

- The patient's condition may change.
- The doctor may not have a good estimate of how a patient's condition might evolve.
- The limited budget.

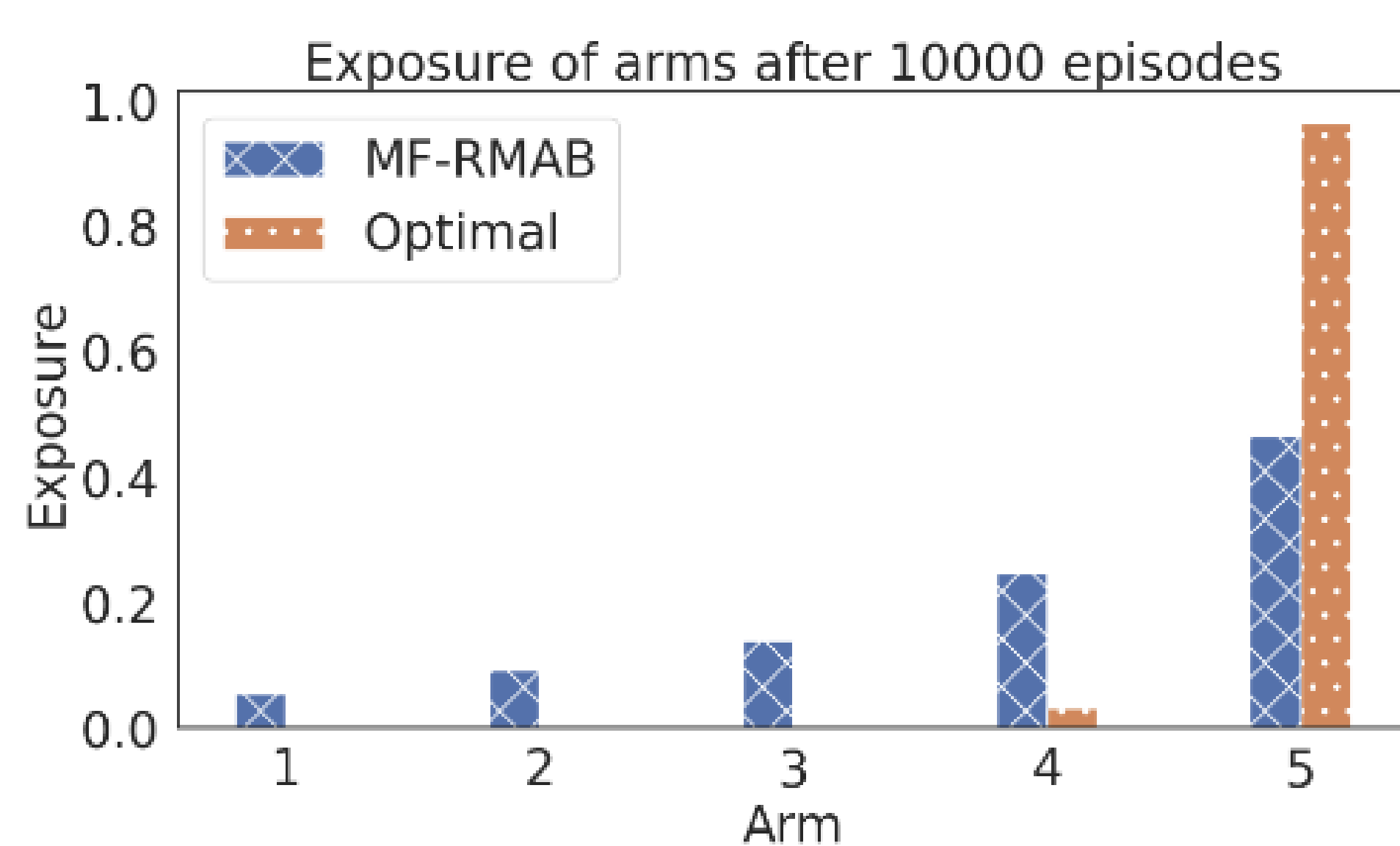
## Restless Multi-armed Bandit

- Each patient is modelled as an 'arm'. Total number of arms =  $N$
- Providing intervention to a patient is called 'pulling' that arm.
- Each arm has two states: 'good' and 'bad'.
- Arms transition from one state to another based on their 'transition probabilities'  $P$ .
- We run our algorithm (policy) for total  $T$  episodes.
- We can only pull  $K$  ( $< N$ ) arms at one-time step.
- Ideally, we want to pull the arm which will go from bad to good state due to our intervention.
- We can define the 'reward' of an arm as the benefit the arm receives from getting pulled.

The Optimal Policy need not be fair!

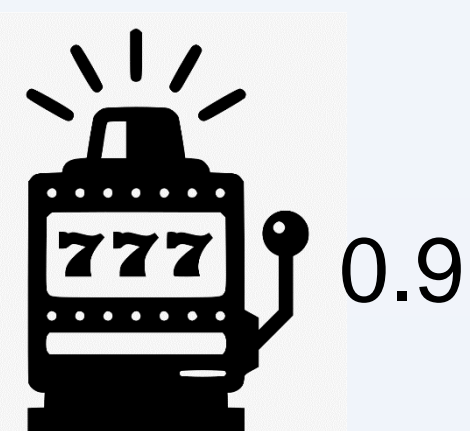
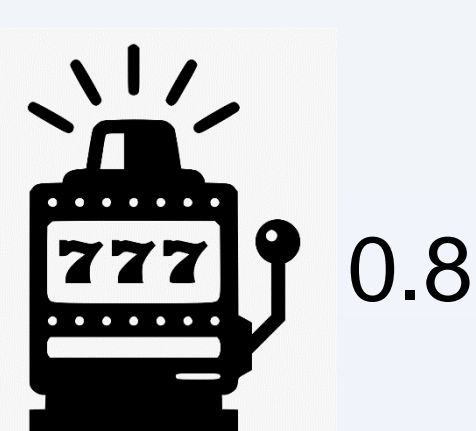
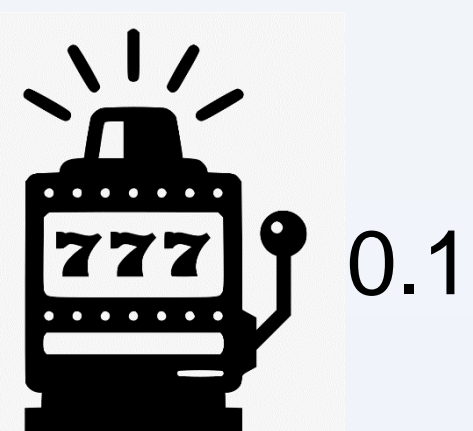
## Meritocratic Fairness

- Pulling arms with the highest reward leads to some arms getting starved of attention.
- In healthcare, this would imply that some patients barely receive medical help!
- There is a need for policies that provide fair exposure to each arm.



Pull each arm with probability proportional to their reward.

## Motivation



Optimal Allocation: Arm 3 with reward 0.9

Fair Allocation: Pulling probabilities of 0.055, 0.44, and 0.5 respectively

## Proposed Algorithm: MF-RMAB

For each episode  $t$ ,

1. Learn the transition probabilities  $P_i^t$  for each arm  $i$  via Upper Confidence Bound approach [1].
2. Find out steady state probability  $f_i(P_i^t, p_i)$  of arm being in 'good' state (when hypothetically pulled with probability  $p_i$ ) [2].
3. Estimate reward  $\mu_i^t = f_i(P_i^t, 1) - f_i(P_i^t, 0)$ .
4. Define meritocratic fair policy  $\pi$ , where  $\pi_i$  is the probability of arm  $i$  being pulled.  $\pi_i^t = \frac{g(\mu_i^t)}{\sum_j g(\mu_j^t)}$ , where  $g(\cdot)$  is a non-decreasing positive Lipschitz-continuous function [3].
5. Sample  $K$  arms from  $\pi^t$ .

## Fairness Regret

Suppose we already know the true transition probabilities  $P^*$  of all the arms. The subsequent policy according to MF-RMAB is denoted by  $\pi^*$ .

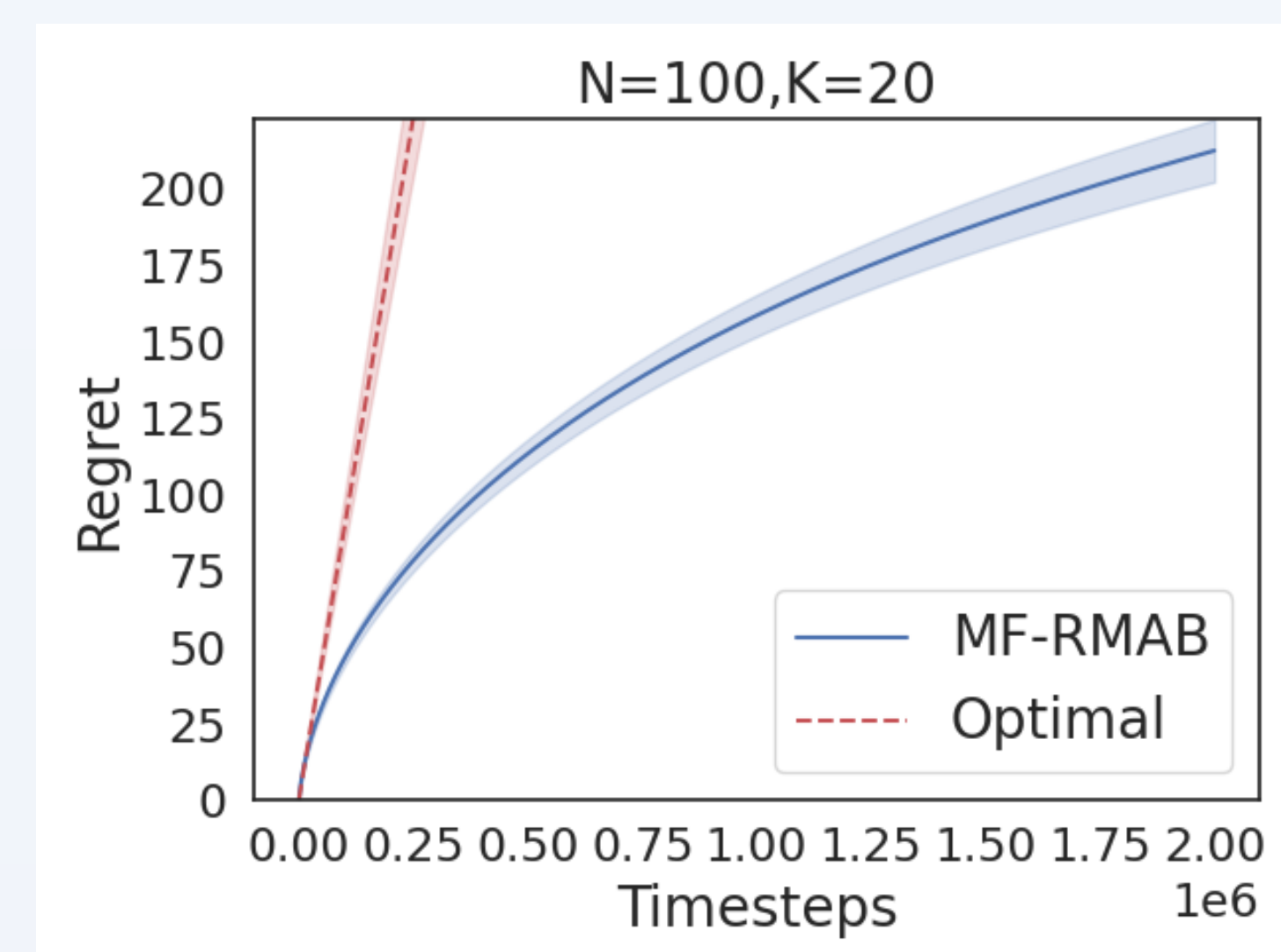
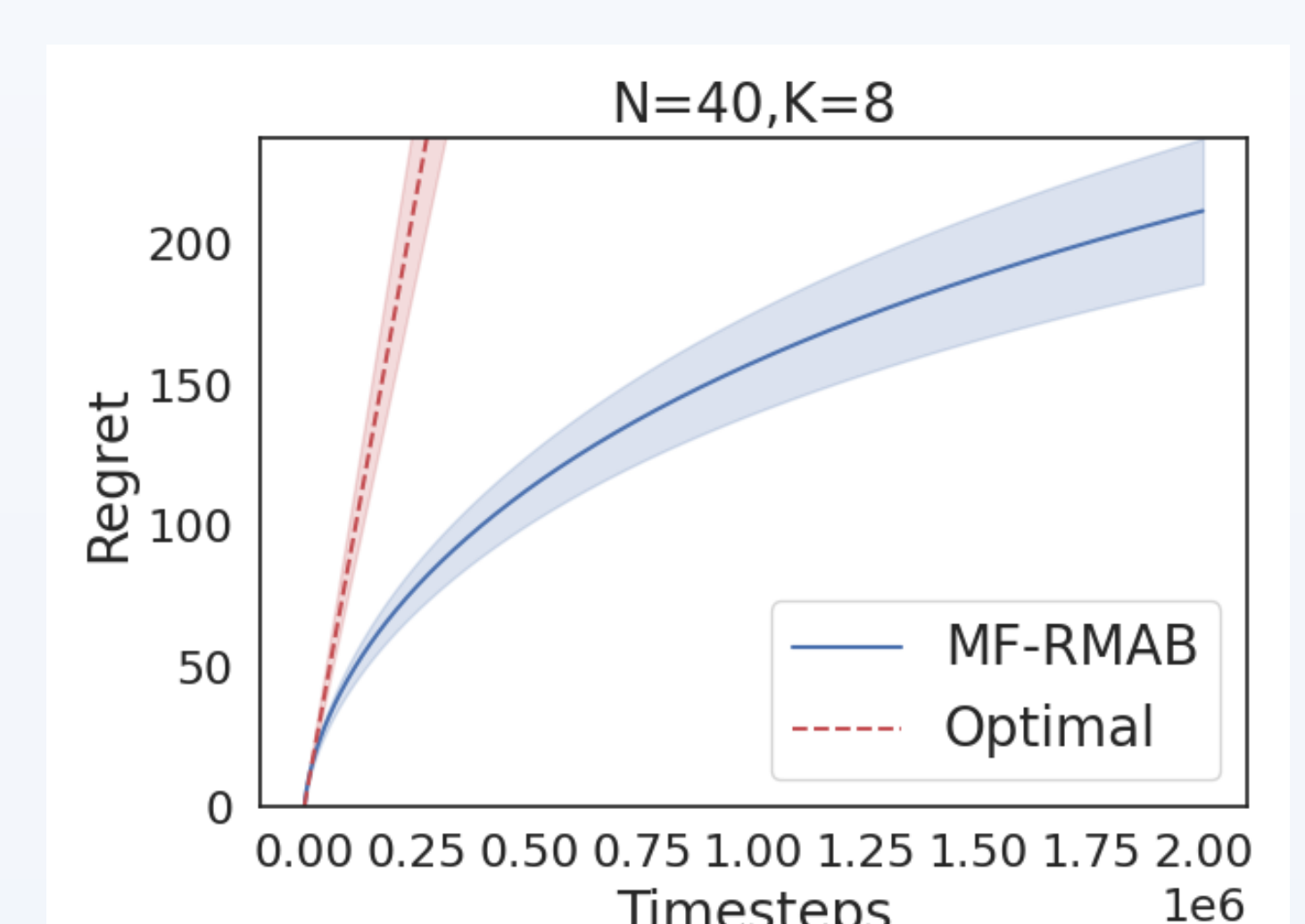
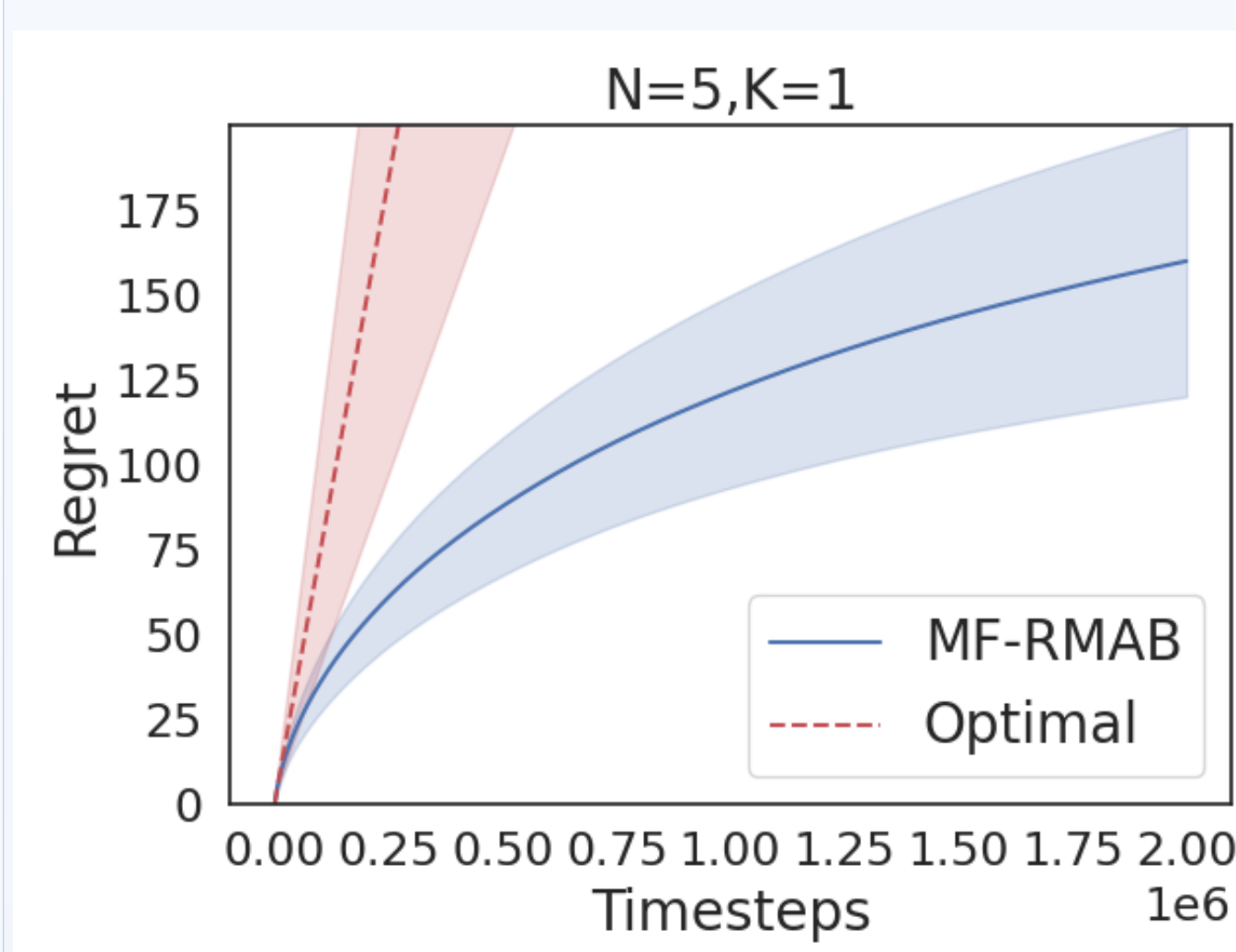
The Fairness Regret (FR) is defined as the difference between the policies when all information is known vs when we have to estimate the transition probabilities.

$$FR^T = \sum_{t=1}^T \sum_{i \in [N]} |\pi_i^* - \pi_i^t|$$

## Theoretical Results

**Theorem:** MF-RMAB incurs  $O(\sqrt{T \ln T})$  fairness regret for sufficiently large  $T$  and  $K = 1$ .

## Experimental Results



## References

- [1] Kai Wang, Lily Xu, Aparna Taneja, and Milind Tambe. Optimistic whittle index policy: Online learning for restless bandits. (AAAI 2023)
- [2] Christine Herlihy, Aviva Prins, Aravind Srinivasan, and John P Dickerson. Planning to fairly allocate: Probabilistic fairness in the restless bandit setting. (ACM SIGKDD 2023)
- [3] Lequn Wang, Yiwei Bai, Wen Sun, and Thorsten Joachims. Fairness of exposure in stochastic bandits. (ICML 2021)

## Acknowledgment:



GAME THEORY  
AND MACHINE  
LEARNING LAB

