Stochastic Bandits Robust to Adversarial Attacks

Xuchuang Wang[1], Maoli Liu[2], Jinhang Zuo[3], Xutong Liu[4], John C.S. Lui[2], Mohammad Hajiesmaili[1]

1 University of Massachusetts Amherst 2 香港中文大學 The Chinese University of Hong Kong 3 CityU 香港城市大學 City University of Hong Kong 4 Carnegie Mellon University

Multi-armed bandits & Adversarial attacks

- K arms, each with a stochastic reward X_k with unknown mean μ_k
  - $\Delta_k \coloneqq \mu_{k^*} - \mu_k$ where $k^* \coloneqq \arg\max_k \mu_k$
- T sequential decision rounds.
- Regret: $R_T \coloneqq T\mu_{k^*} - \sum_{t=1}^{T} \mu_{I_t}$
- Attack budget: $C \coloneqq \sum_{t=1}^{T} |X_{k,t} - \tilde{X}_{k,t}|$

Attack (above) vs. Corruption (below)

Learner pull arm I_t → Stochastic reward $X_{I_t,t}$ drawn from I_t → The **adversary** chooses the attacked reward $\tilde{X}_{I_t,t}$ for arm I_t → **Learner** observes the attacked reward $\tilde{X}_{I_t,t}$

Stochastic reward $X_{k,t}$ drawn from all arms → The **adversary** chooses the corrupted reward $\tilde{X}_{k,t}$ for all arms → **Learner** pull arm I_t → **Learner** observes the corrupted reward $\tilde{X}_{I_t,t}$

**See our paper for a detailed separation of these two models.*

Robust algorithm designs & Regret bounds

Known Attack Budget

SE−WR: $O\left(\sum_{k\neq k^*} \frac{\log T}{\Delta_k} + KC\right)$ → Stop Cond. → SE−WR−STOP: $O(\sqrt{KT(\log T + C)})$

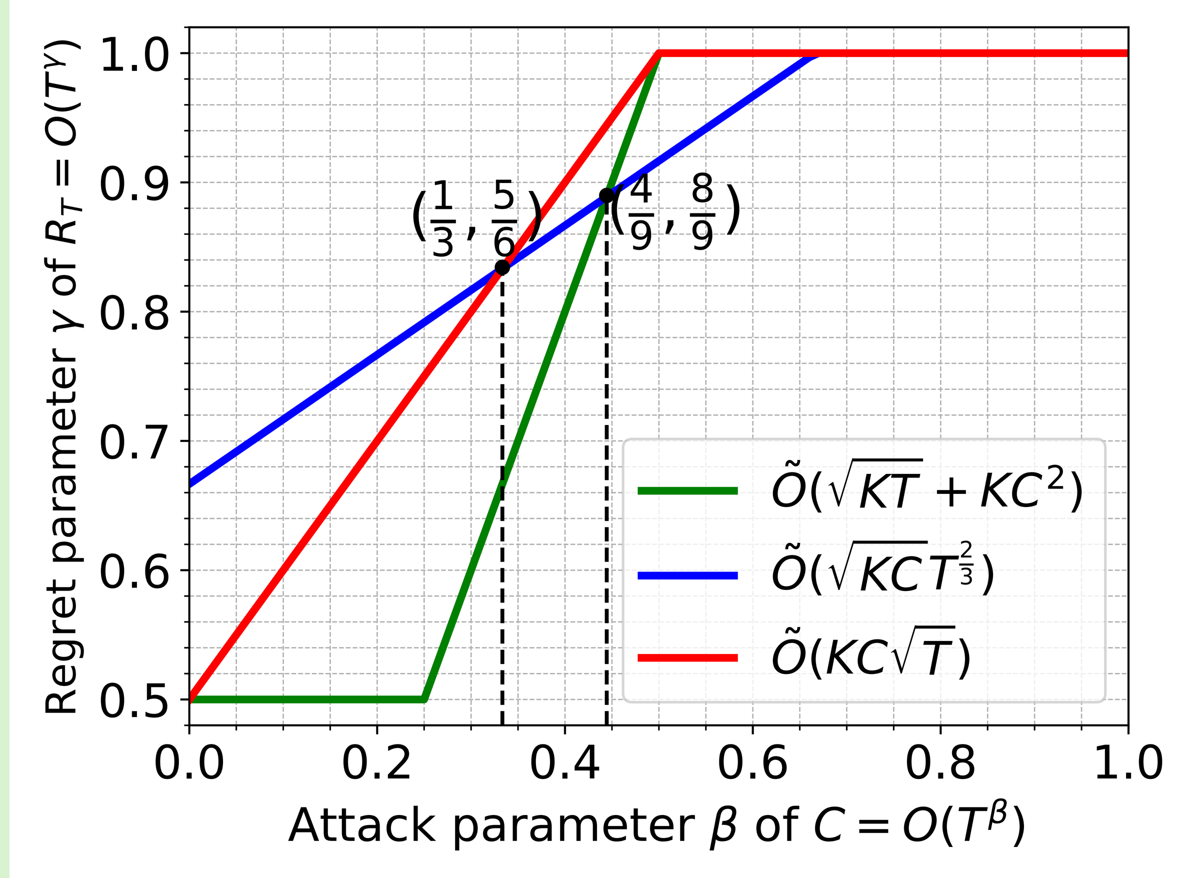
Multi-Phase ↓ Model Selection ↓

PE−WR: $\tilde{O}(\sqrt{KT} + KC^2)$

MS−SE−WR: $\tilde{O}(\sqrt{KCT^{\frac{2}{3}}})$, $\tilde{O}(KC\sqrt{T})$

Unknown Attack Budget

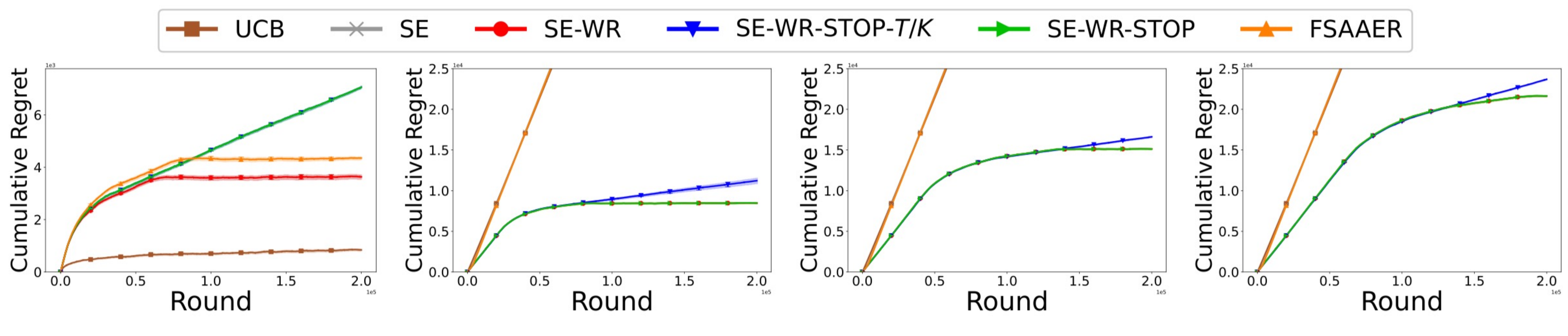
Additive vs. Multiplicative



**All upper bounds are tight to some logarithmic factors (see lower bounds in the paper).*

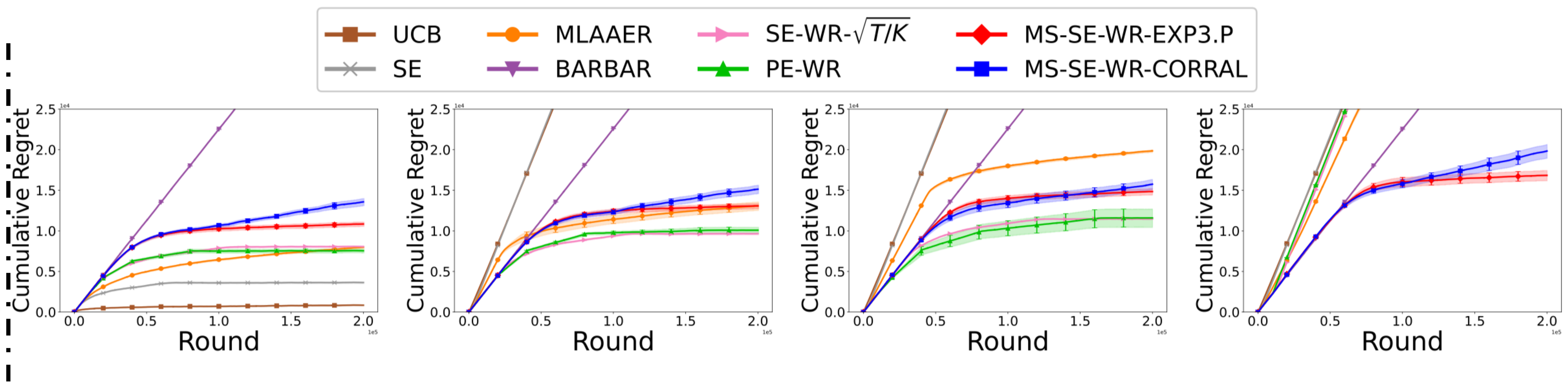
- Successive elimination with wide confidence radius (SE−WR) improves the $\sum_k C/\Delta_k$ of Lykouris et al. (2018, Theorem 1) to KC.
- SE−WR−STOP introduces a stopping condition to transfer the gap-dependent to independent (worse-case) bound.
- Phase Elimination (PE−WR) uses multiple phases, each dealing with one possible attack budget.
- Model Selection (MS−SE−WR) uses $\log_2 T$ base algorithms (SE−WR−STOP) with different attack budgets and pick instances as “arms” via either CORRAL or EXP3.P to defense adaptively.

Simulations: Regret comparison



(a) $C = 0$ (b) $C = 3000$ (c) $C = 6000$ (d) $C = 9000$

Algorithms with known attack budget C



(a) $C = 0$ (b) $C = 3000$ (c) $C = 6000$ (d) $C = 9000$

Algorithms with unknown attack budget C