The University of Texas at Austin

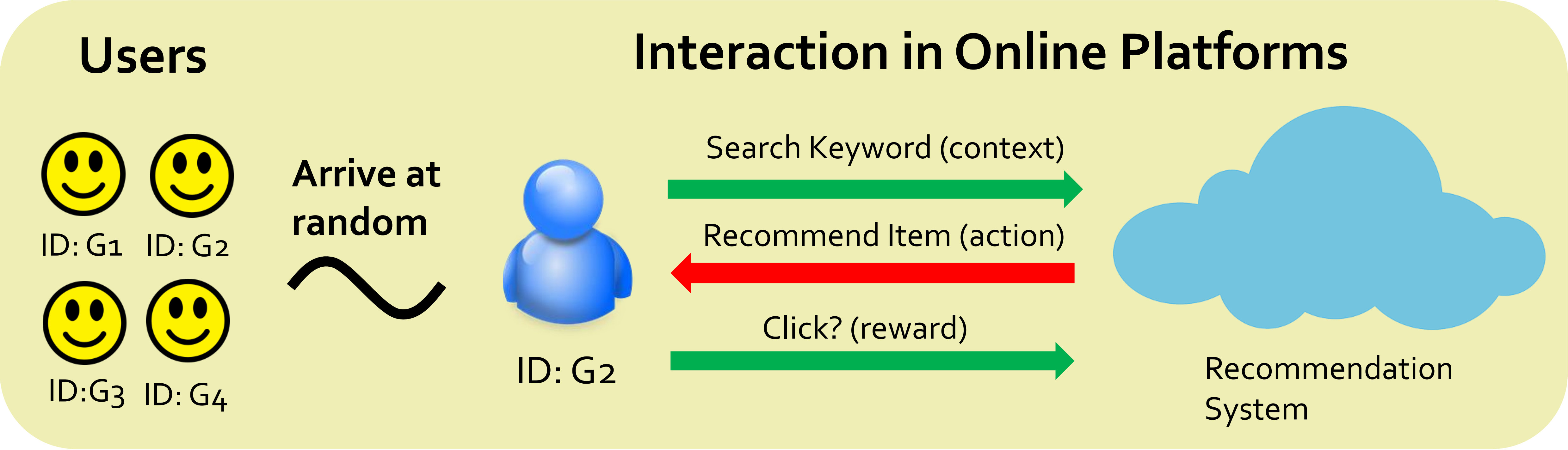
Coordinated Attacks against Contextual Bandits

Fundamental Limits and Defense Mechanisms

by **Jeongyeol Kwon***, Yonathan Efroni, Constantine Caramanis, Shie Mannor

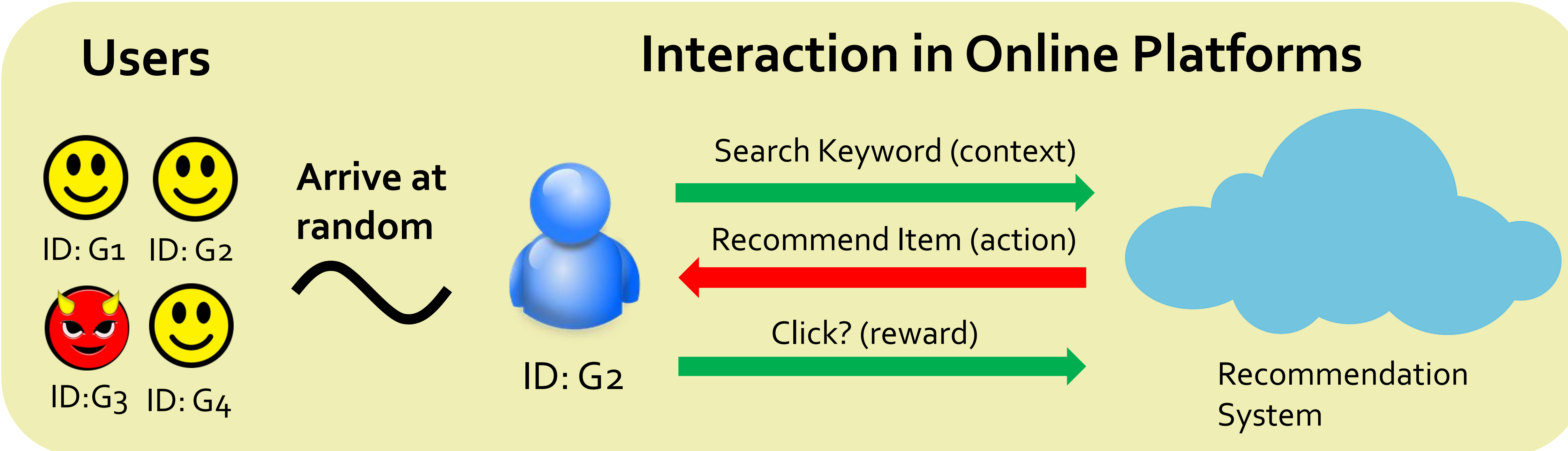
ICML 2022

Contextual Bandits with Adversaries



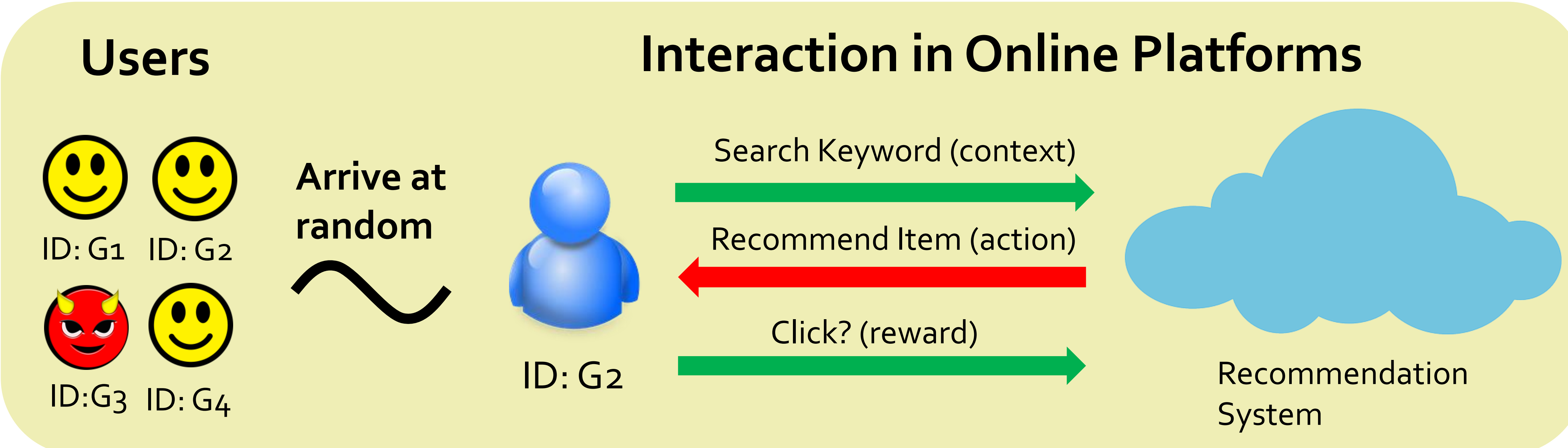
- Contextual Bandits
  - $\mathcal{B} \coloneqq (\mathcal{S}, \mathcal{A}, \mu)$: S contexts, A items, μ -- mean-rewards

Contextual Bandits with Adversaries



- Contextual Bandits
  - $\mathcal{B} \coloneqq (\mathcal{S}, \mathcal{A}, \mu)$: S contexts, A items, μ -- mean-rewards

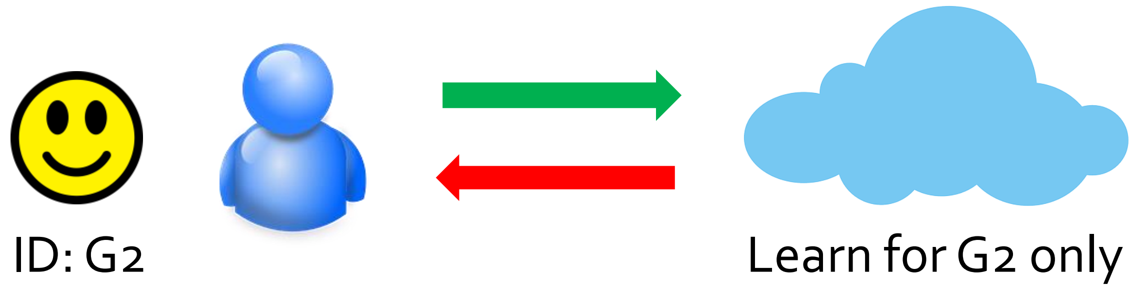
Contextual Bandits with Adversaries



- Contextual Bandits
  - $\mathcal{B} := (\mathcal{S}, \mathcal{A}, \mu)$: S contexts, A items, μ -- mean-rewards
- Multi-task(user) learning with adversaries
  - $1 - \alpha$ good users: $r(s, a) \sim \mathcal{B}$
  - α adversaries: $r(s, a) \in \mathbb{R}$, arbitrary – confuse the system
  - ***Unknown which users (with known IDs) are adversaries***

Goal – Parallelization Gain

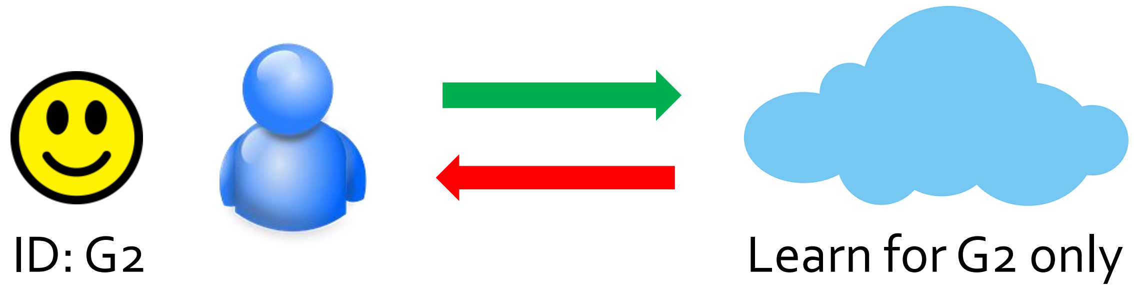
- Learn ϵ-optimal policy separately



Standard Contextual Bandits:
$O\left(\frac{SA}{\epsilon^2}\right)$ *per-user* samples

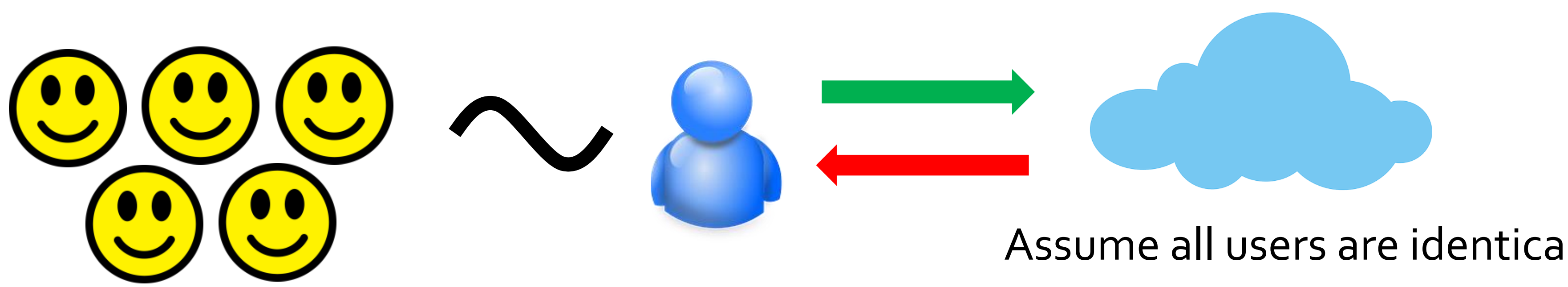
Goal – Parallelization Gain

- Learn ϵ-optimal policy separately



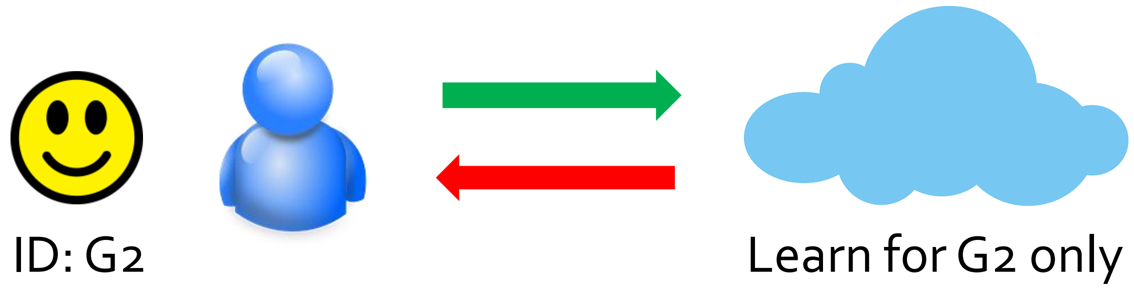
Standard Contextual Bandits:
$O\left(\frac{SA}{\epsilon^2}\right)$ *per-user* samples

- Exploit similarity between users
  - With L-good users: $O\left(\frac{\mathbf{1}}{\boldsymbol{L}} \cdot \frac{SA}{\epsilon^2}\right)$ *per-user* samples
  - $\frac{1}{L}$ - collaboration gain



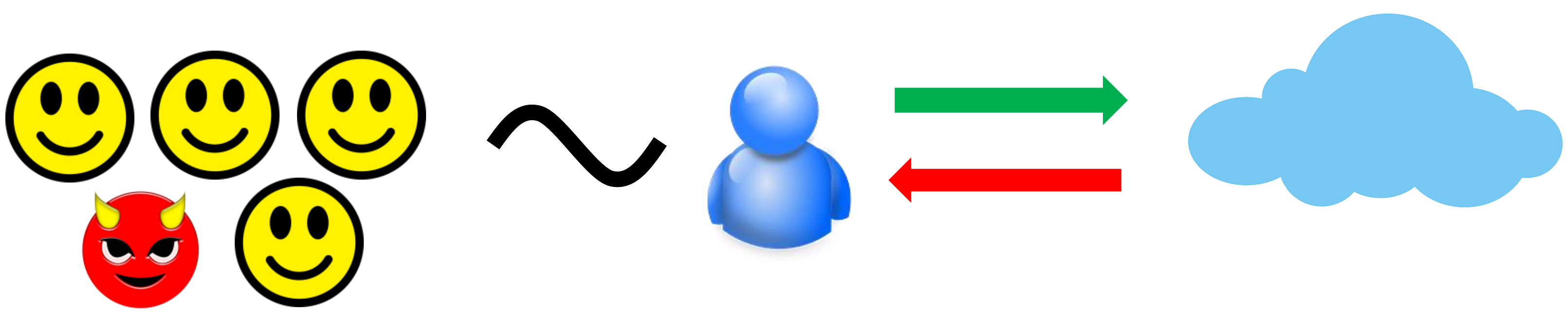
Goal – Parallelization Gain

- Learn ϵ-optimal policy separately



Standard Contextual Bandits:
$O\left(\frac{SA}{\epsilon^2}\right)$ *per-user* samples

- Exploit similarity between users
  - With L-good users: $O\left(\frac{\mathbf{1}}{\boldsymbol{L}} \cdot \frac{SA}{\epsilon^2}\right)$ *per-user* samples
  - $\frac{1}{L}$ - collaboration gain

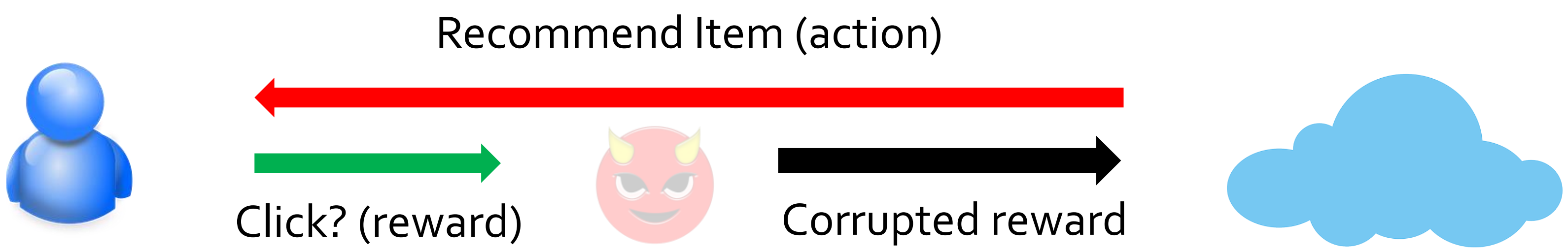


Q: *What is the maximum parallelization gain if α-fraction of users are adversarial? ($\alpha < 1/2$)*

Related Work

- Bandits with Adversarial Corruptions
  - *Single user*, rewards corrupted *at any time* with limited budget

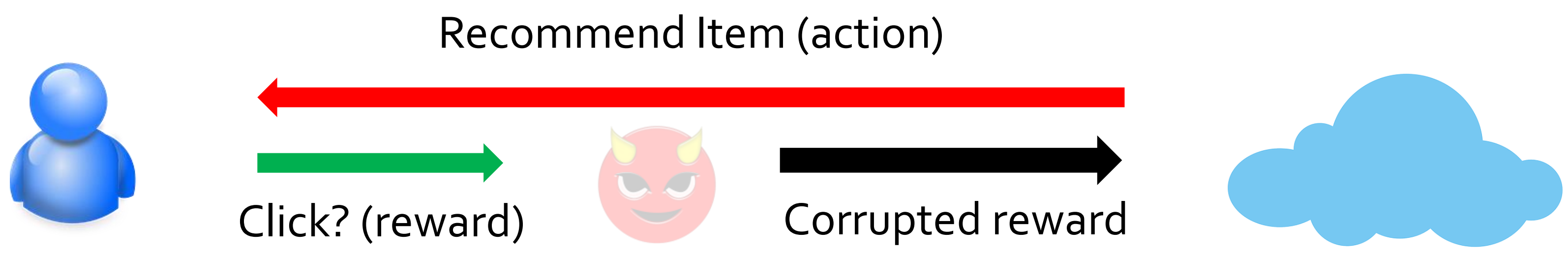
  [Gupta et al., 2019; Lykouris et al., 2018, 2021; Liu et al., 2021]



Related Work

- Bandits with Adversarial Corruptions
  - *Single user*, rewards corrupted *at any time* with limited budget

  [Gupta et al., 2019; Lykouris et al., 2018, 2021; Liu et al., 2021]



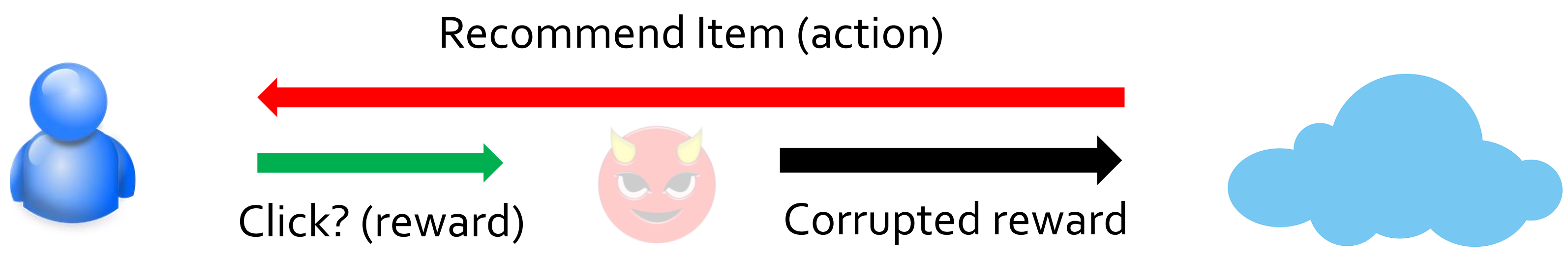
- Multitask Learning of Contextual Bandits
  - Multiple classes of users with separation

  [Mailard and Mannor, 2014; Gopalan et al., 2016; Sen et al., 2017; Gentile et al., 2014; Yang et al., 2020; Ghosh et al., 2021; Hu et al., 2021]

Related Work

- Bandits with Adversarial Corruptions
  - *Single user*, rewards corrupted *at any time* with limited budget

    [Gupta et al., 2019; Lykouris et al., 2018, 2021; Liu et al., 2021]



- Multitask Learning of Contextual Bandits
  - Multiple classes of users with separation

    [Mailard and Mannor, 2014; Gopalan et al., 2016; Sen et al., 2017; Gentile et al., 2014; Yang et al., 2020; Ghosh et al., 2021; Hu et al., 2021]

- **Our Work**: 1 good-user class & adversaries (fake-profiles)

Main Result: Fundamental Limits

- With polynomial number of users:
  - *i.e.*, $L = \mathrm{poly}\left(S, A, \frac{1}{\epsilon}\right)$
  - $\Omega\left(\mathbf{min}(\boldsymbol{S}, \boldsymbol{A}) \cdot \frac{\alpha^2}{\epsilon^2}\right)$ - per-user samples are necessary
  - Thus, parallelization gain can be at most $\frac{\alpha^2}{\max(S,A)}$

Main Result: Fundamental Limits

- With polynomial number of users:
  - *i.e.*, $L = \text{poly}\left(S, A, \frac{1}{\epsilon}\right)$
  - $\Omega\left(\mathbf{min}(\boldsymbol{S}, \boldsymbol{A}) \cdot \frac{\alpha^2}{\epsilon^2}\right)$ - per-user samples are necessary
  - Thus, parallelization gain can be at most $\frac{\alpha^2}{\max(S,A)}$

- Note: no "$\min(S, A)$" if $L \geq \exp\big(\omega(S)\big)$

Results Summary

- Multitask contextual bandits with adversarial users

- Lower Bound
  - $\Omega\left(\min(S, A) \cdot \frac{\alpha^2}{\epsilon^2}\right)$ - per-user samples are necessary

- Upper Bound
  - $O\left(\min(S, A) \cdot \frac{\alpha}{\epsilon^2}\right)$ - per-user samples are sufficient
  - Can be achieved with two robust estimators

- Some future directions
  - Investigate the gap on α
  - Extension to linear bandits / RL