HARVARD
School of Engineering
and Applied Sciences

CDC 2020

Online Residential Demand Response via Contextual Multi-Armed Bandit

Xin Chen, Yutong Nie, Na Li

School of Engineering and Applied Sciences, Harvard University

chen_xin@g.harvard.edu

Power Systems

Demand Response

Generation = **Load**

uncertainty ↑ intermittency

growing load peak

Large-scale renewable generation

State of emergency declared as California faces historic heat, possible power outages

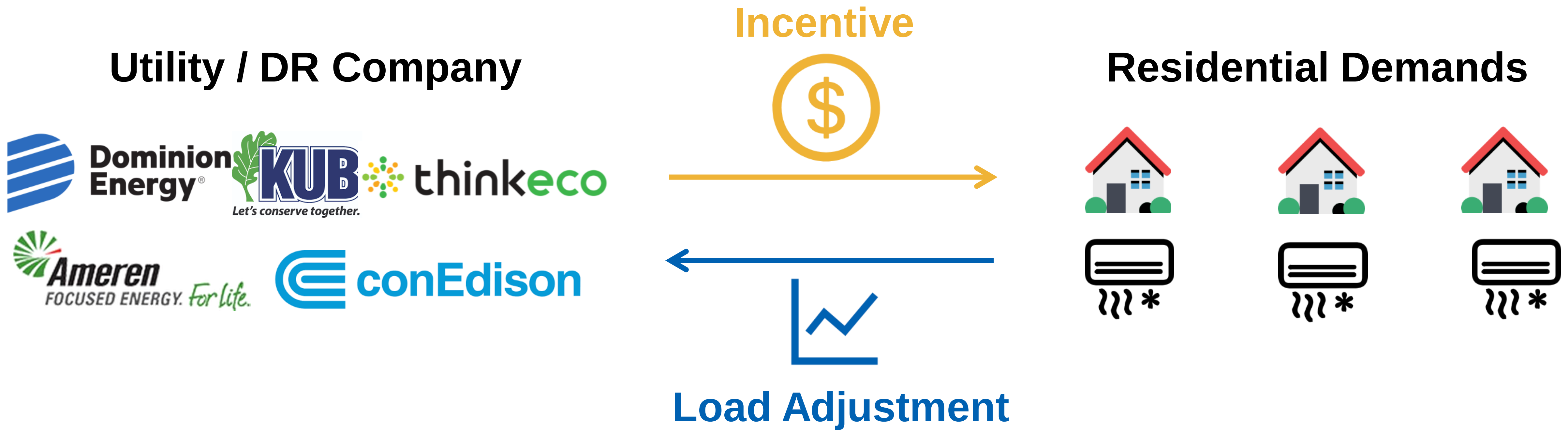
Los Angeles Times

Heat Wave Roasts Southern California With Record of 121 Degrees

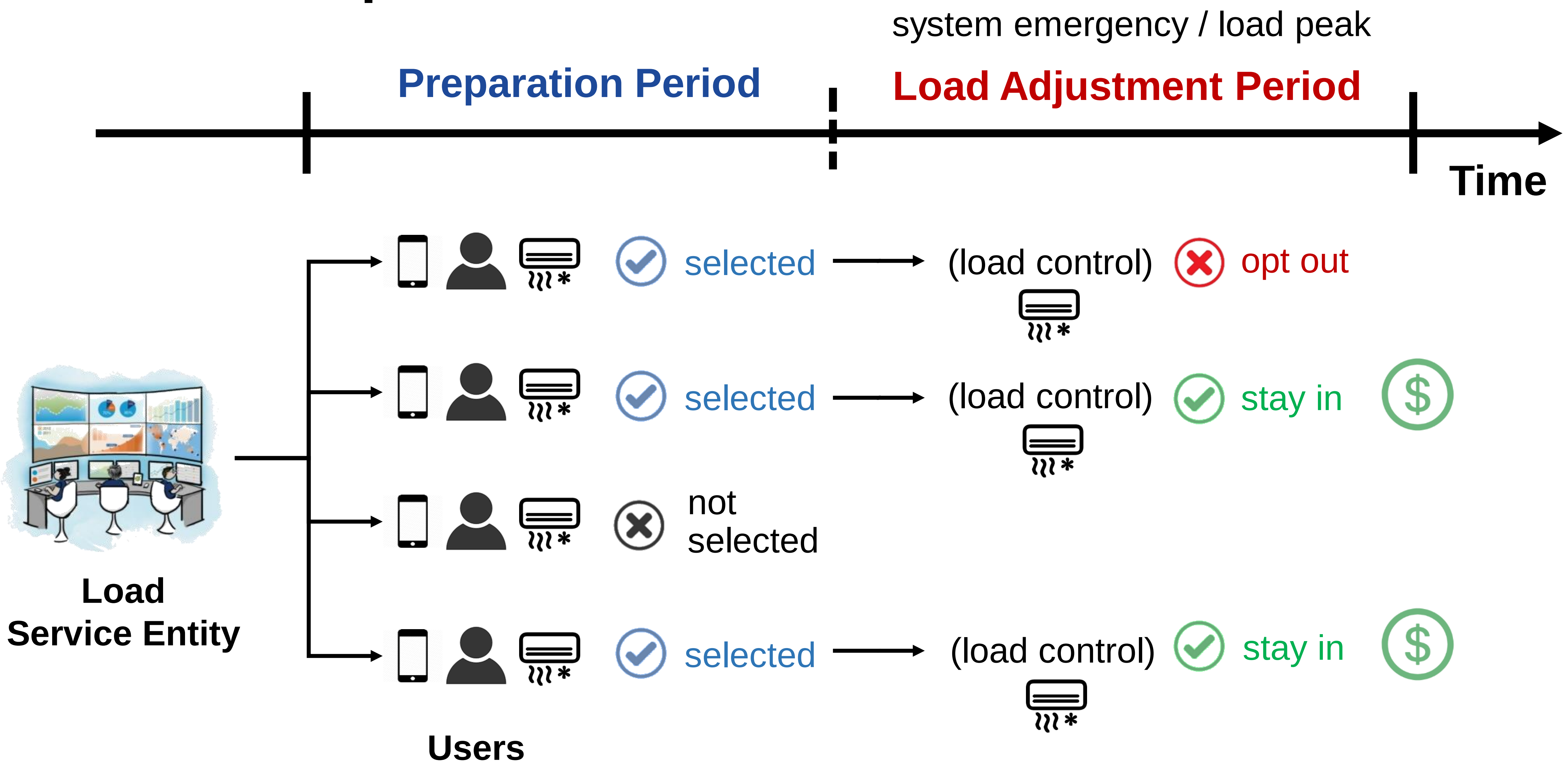
BRIEF

ERCOT calls 2 energy emergencies in one week, 3rd in 5 years

Residential Demand Response



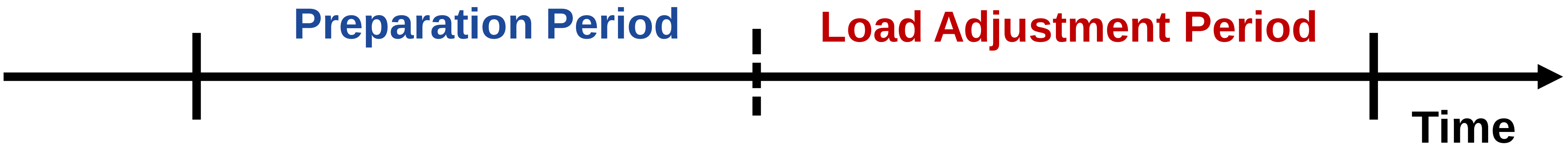
Demand Response Event



[1] **X. Chen**, Y. Li, J. Shimada, and N. Li, “**Online Learning and Distributed Control for Residential Demand Response**”, arXiv:2010.05153, 2020.

Q2- How to optimally control load in real-time?

A2: A follow-up paper [1].



Q1- How to select right users for DR?

A1: This talk.

Challenge: Uncertain and unknown user opt-out behaviors.

Individual Preference

- age
- education
- household size
- attitude to energy saving, etc.

Opt Out?

Environmental Factors

- indoor temperature
- offered reward
- electricity price
- weather conditions, etc.

User opt-out outcomes

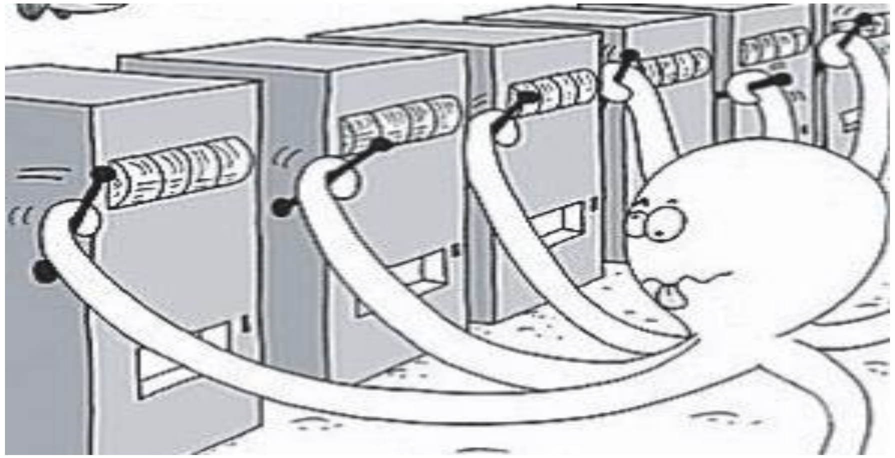
Selection Decision

Online Learning and Decision-making

Unknown and **uncertain** user behaviors.

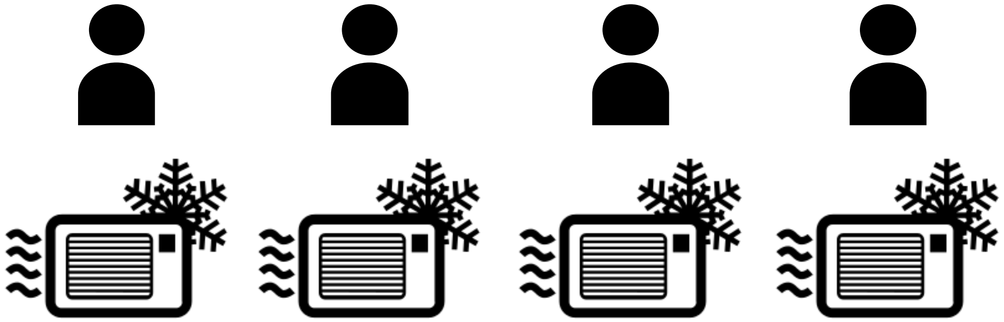
Contextual Multi-armed bandit (CMAB)

Slot Machine



- Select one arm to maximize the profits;
- Observe the reward of the selected arm;
- Improve play strategies from feedback.

Demand Response



- Select a subset of users for DR;
- Observe responses from selected users;
- Learn users' behaviors from responses.

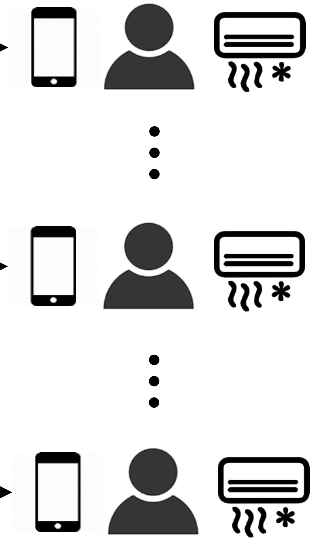
Problem Formulation

Consider a time horizon $[T] = \{1, 2, \cdots, T\}$

Each time $t \in [T]$ denotes a DR event.

Aggregator

budget $b_{i,t}$



user 1

user *i*

user *N*

- Binary random variable $z_{i,t} \in \mathrm{Bern}(p_{i,t})$.

$$z_{i,t} = \begin{cases} 1 & \text{: Stay-in} \\ 0 & \text{: Opt-out} \end{cases}$$

- Adjustable load capacity $c_{i,t}$.
- Incentive reward/ bidding price $r_{i,t}$.

Optimal User Selection Model

$$\text{Obj.} \quad \max_{\mathcal{S}_t \subseteq [N]} \mathbb{E}\Big(\sum_{i \in \mathcal{S}_t} c_{i,t} z_{i,t}\Big) = \sum_{i \in \mathcal{S}_t} c_{i,t} \boxed{p_{i,t}}$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{S}_t} r_{i,t} \leq b_t$$

maximize expected total load reduction.

Unknown

User Behavior Learning with Contexts

- **Logistic model** to predict $p_{i,t} := p_i(t)$ for user i :

$$p_i(t) = g(\boldsymbol{\theta}_i^\top \boldsymbol{x}_i(t)) = \frac{1}{1 + \exp\left(-\boldsymbol{\theta}_i^\top \boldsymbol{x}_i(t)\right)}$$

Personal Weight

$$\boldsymbol{\theta}_i = (\underbrace{\theta_{i,1}}_{\text{individual preference}}, \underbrace{\theta_{i,2}, \cdots, \theta_{i,d}}_{\text{response to context}})$$

Context at time t

$$\boldsymbol{x}_i(t) = (1, \underbrace{x_{i,2}, \cdots, x_{i,d}}_{\text{environmental factors}})(t)$$

Goal : to learn parameter $\boldsymbol{\theta}_i$ for each user.

- **Thompson sampling** is used to balance ***exploration*** and ***exploitation***.

Thompson Sampling (Bayesian learning)

Assume unknown $\boldsymbol{\theta}_i$ be a random variable with Gaussian prior $\mathbb{P}_{\boldsymbol{\theta}_i} = \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$.

Step 1: Sample $\hat{\boldsymbol{\theta}}_i$ from its distribution $\mathbb{P}_{\boldsymbol{\theta}_i}$.

$$p_{i,t} = \frac{1}{1 + \exp\left(-\hat{\boldsymbol{\theta}}_i^\top \boldsymbol{x}_i(t)\right)}$$

Step 2: Select users by solving

$$\begin{aligned} &\text{Obj.} \quad \max_{\mathcal{S}_t \subseteq [N]} \mathbb{E}\Big(\sum_{i \in \mathcal{S}_t} c_{i,t} z_{i,t}\Big) = \sum_{i \in \mathcal{S}_t} c_{i,t} p_{i,t} \\ &\text{s.t.} \quad \sum_{i \in \mathcal{S}_t} r_{i,t} \leq b_t \end{aligned}$$

$$\begin{aligned} &\text{Obj.} \quad \max_{\alpha_{i,t} \in \{0,1\}} \sum_{i=1}^{N} c_{i,t} p_{i,t} \alpha_{i,t} \\ &\text{s.t.} \quad \sum^{N} r_{i,t} \alpha_{i,t} \leq b_t \end{aligned}$$

Binary Optimization

Step 3: Update posterior $\mathbb{P}_{\boldsymbol{\theta}_i} \leftarrow \mathbb{P}_{\boldsymbol{\theta}_i}(\cdot | \boldsymbol{x}_{i,t}, z_{i,t})$ with the observation $\boldsymbol{x}_{i,t}, z_{i,t}$.

variational Bayesian inference approach [2]

[2] T. S. Jaakkola and M. I. Jordan, “A variational approach to Bayesian logistic regression models and their extensions”. in Sixth International Workshop on Artificial Intelligence and Statistics, vol. 82, pp. 4, 1997.

Regret Analysis

- (Expected) Reward Function: $f_{\boldsymbol{\theta}}(\mathcal{S}_t, t) = \mathbb{E}(\sum_{i \in \mathcal{S}_t} c_{i,t} z_{i,t}) = \sum_{i \in \mathcal{S}_t} \frac{c_{i,t}}{1 + \exp(-\boldsymbol{x}_{i,t}^\top \boldsymbol{\theta}_i)}$
- T-time Regret: $\mathrm{Regret}(T, \boldsymbol{\theta}) = \sum_{t=1}^{T} \mathbb{E}\left[f_{\boldsymbol{\theta}}(\mathcal{S}_t^*, t) - f_{\boldsymbol{\theta}}(\mathcal{S}_t, t) \,|\, \boldsymbol{\theta}\right]$
- T-time Bayesian Regret: $\mathrm{BayesRegret}(T) = \mathbb{E}_{\boldsymbol{\theta} \sim P_0}\left[\mathrm{Regret}(T, \boldsymbol{\theta})\right]$

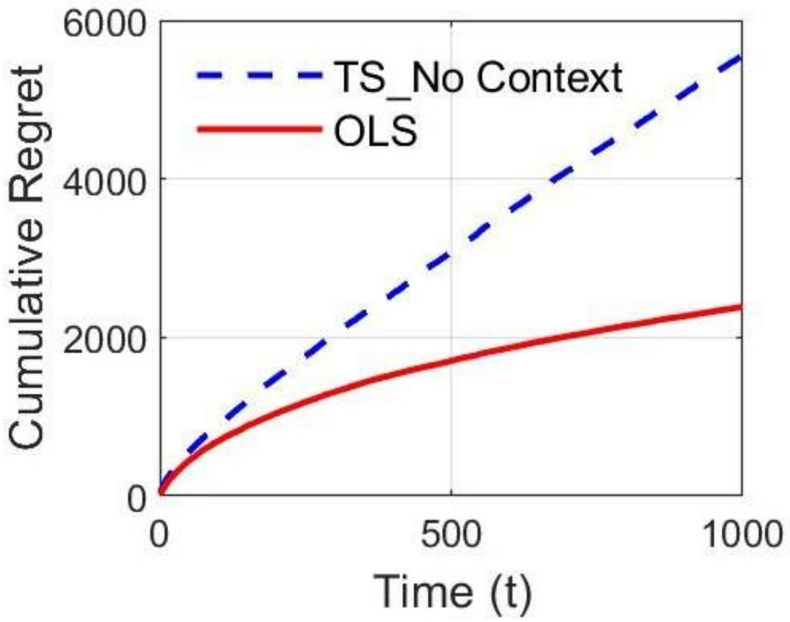
Theorem (informal): When T is sufficiently large, the Bayesian regret is

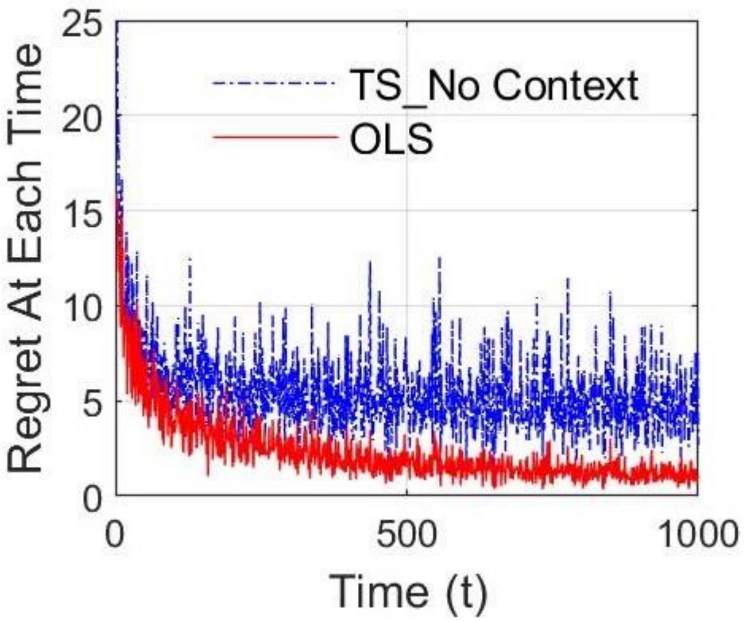
$$\mathrm{BayesRegret}(T) \leq O\left(N^2 \gamma^d \sqrt{T \log T (d + \log T)}\right) \sim O(\log(T)\sqrt{T})$$

where $\gamma = \exp(2 \sup_{i \in [N]} ||\boldsymbol{\theta}_i||_\infty)$ and d is the dimension of $\boldsymbol{\theta}_i$.

Simulation:

- N = 1000 users; m = 9 environmental factors.
- A Gaussian prior distribution $\mathcal{N}(\boldsymbol{\theta}_i^* + 0.3\boldsymbol{u}_i, 0.09\boldsymbol{I})$ for each user.
- Regret comparison between the proposed algorithm (OLS) and Thompson sampling without context information.





Thank you!