Differentially Private Meta-Learning

Jeffrey Li[1], Mikhail Khodak[1],
Sebastian Caldas[1], Ameet Talwalkar[1,2]

[1]Carnegie Mellon University
[2]Determined AI

Why privacy in meta-learning?

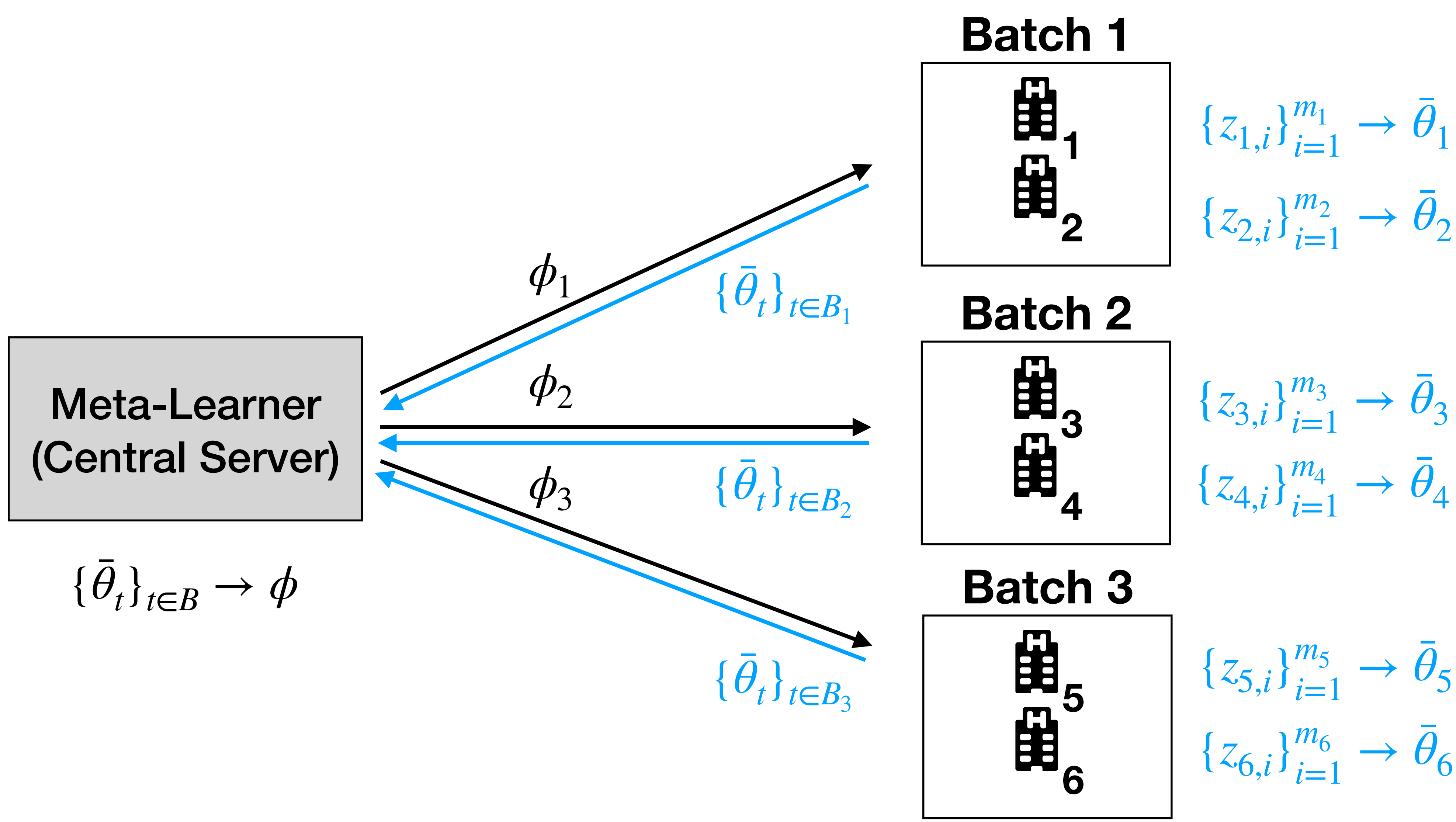
Problem: Meta-learning shares knowledge across tasks, leaving task-owners' (e.g. mobile users, hospitals) data vulnerable to inference

Questions:

1. What are appropriate notions of privacy for meta-learning?
2. What are applications of these various notions?
3. Can we sufficiently privatize common methods while retaining utility?
4. How does our proposed approach work empirically?

Gradient-Based Meta-Learning

Algorithms alternate between **within-task queries** and **meta-level queries**



A task's data can potentially be inferred by any downstream agent.

What are appropriate notions of privacy in meta-learning?

Who are the adversaries?

less private → more private

What data is protected?

less private ↓ more private

	Other task-owners	Task-owners and the meta-learner
Single Records $z_{t,i}$	—	**Task-Global DP**
Entire Task Datasets $D_t = \{z_{t,i}\}_{i=1}^{m_t}$	Global DP (assumes trust in server)	Local DP (difficult to retain utility)

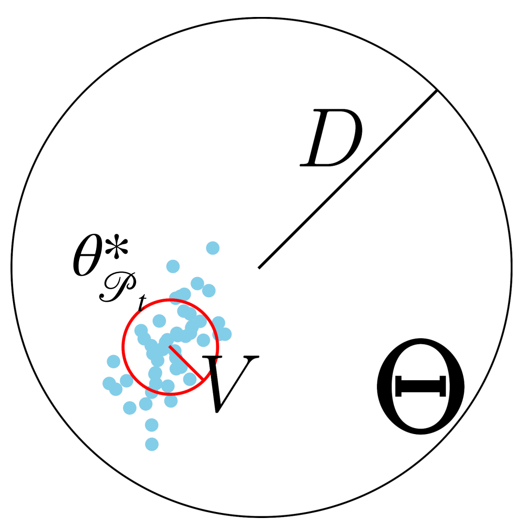
What does this mean practically?

	Mobile Users	Hospitals
Global	Whole SMS history private to only other users	Whole database private to only other hospitals
Local	Whole SMS history private to everyone	Whole database private to everyone
Task-Global	**Individual messages private to everyone**	**Each patient's record private to everyone**

Can we privatize Reptile[1] while still retaining the utility of meta-learning?

Results: Applying a noisy SGD procedure within-task, we can **guarantee both**

- Task-global DP in all settings
- Bound for the **transfer-risk** in convex settings



$$V^2 = \min_{\phi \in \Theta} \frac{1}{2} \mathbb{E}_{\mathcal{P} \sim \mathcal{Q}} \|\theta^*_{\mathcal{P}} - \phi\|_2^2$$

Transfer-risk =

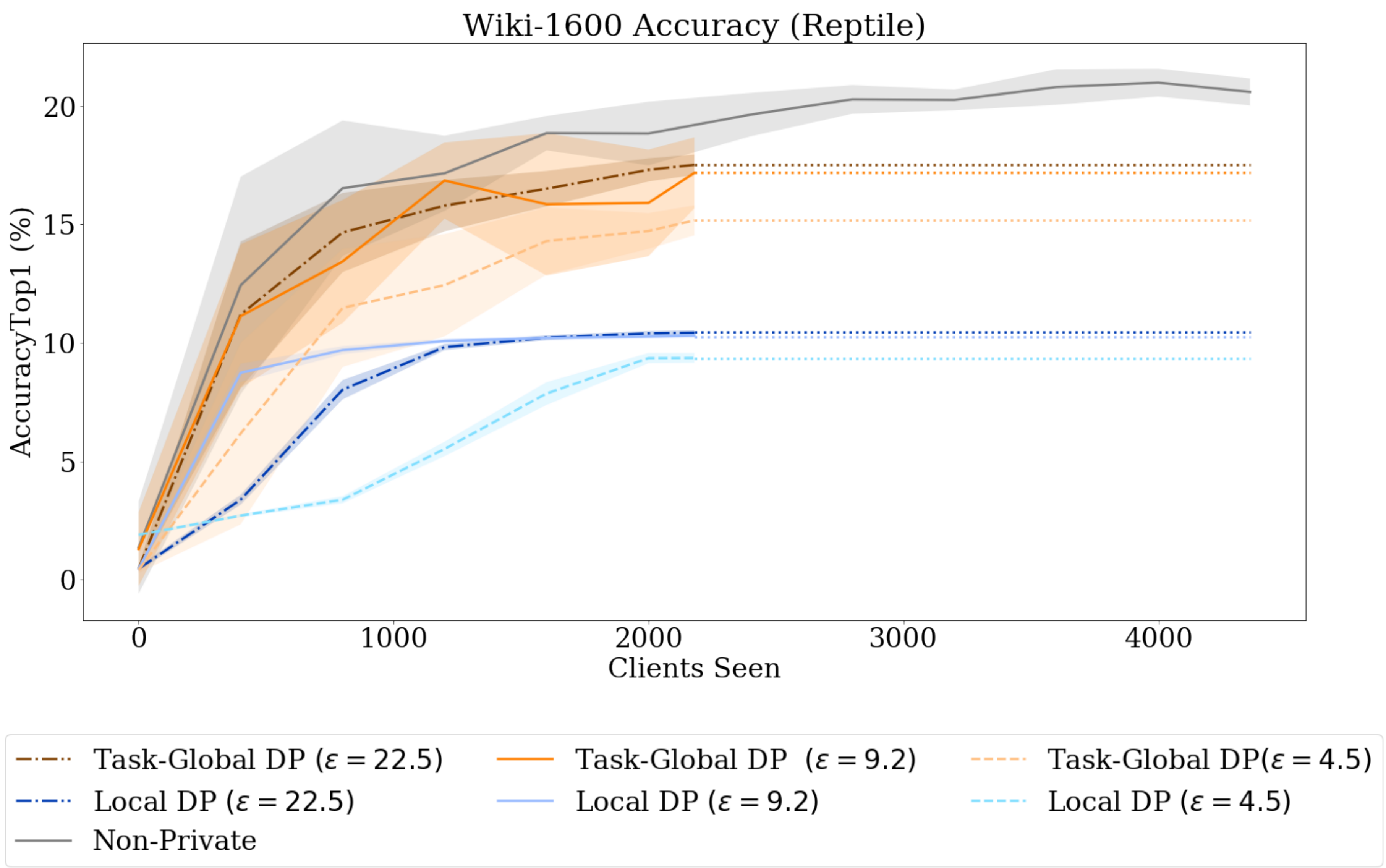
$$\tilde{\mathcal{O}}\left(\frac{V}{\sqrt{m}} + \frac{\alpha D^2}{T} + \frac{1}{\alpha} \max\left(\frac{d \log \frac{1}{\delta}}{\varepsilon^2 m^2}, \frac{1}{m}\right)\right)$$

Standard $\sqrt{m}$ term scales with task similarity

Terms incurred by DP

[1](Nichol et al., 2018)

Federated Language Modeling



Non-private learning

Task-Global DP (Ours)

Local DP

Our Contributions

1. What are appropriate notions of privacy in meta-learning?

   *Formalized **task-global DP** as useful relaxation of local DP*

2. What are applications of these various notions?

   We show natural applications to personalized federated learning

3. Can we sufficiently privatize common methods while retaining utility?

   Reptile-like method with both privacy and learning guarantees

4. How does our approach work empirically?

   *Showed usefulness of **task-global DP** in non-convex experiments*

Find out more!

- Full paper: https://openreview.net/forum?id=rJgqMRVYvr
- Contact me: jwl3@andrew.cmu.edu