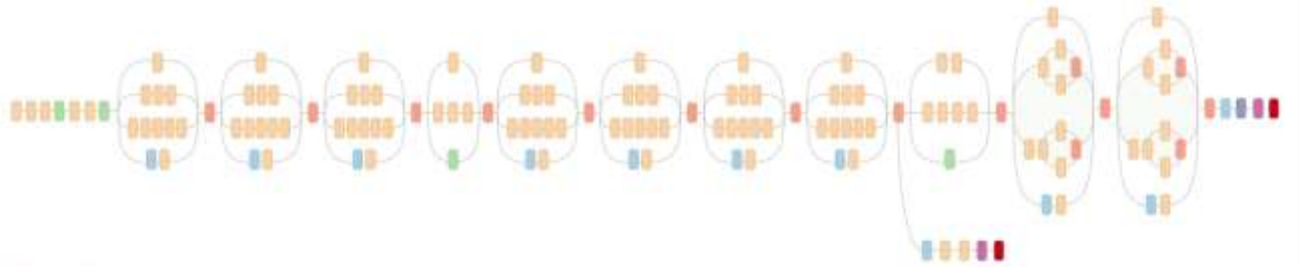
Deep Learning with Differential Privacy: Two Approaches

Ilya Mironov
Google Research

CNSTAT Privacy Workshop
June 6, 2019

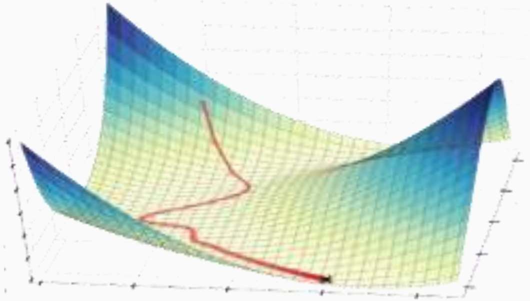
Deep Learning

- Non-convex optimization
- Large, deep models
- Diversity of input data
- Diversity of tasks and learning modalities



Stochastic Gradient Descent (SGD)

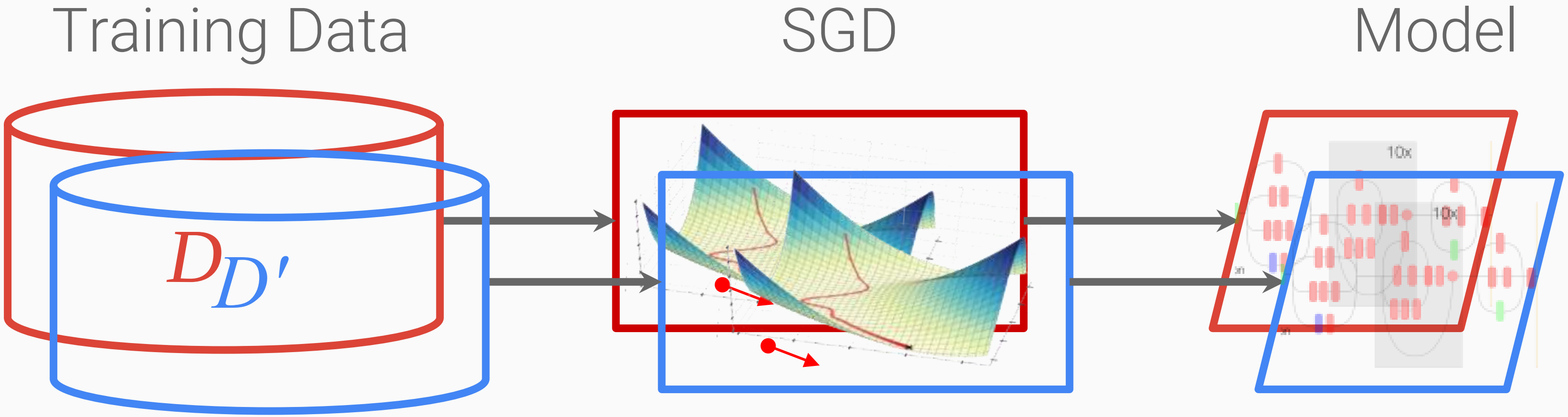
Compute $\nabla L(\theta_i)$ on a sample → $\theta_{i+1} := \theta_i - \eta \nabla L(\theta_i)$ → Compute $\nabla L(\theta_{i+1})$ on a sample → $\theta_{i+2} := \theta_{i+1} - \eta \nabla L(\theta_{i+1})$ →



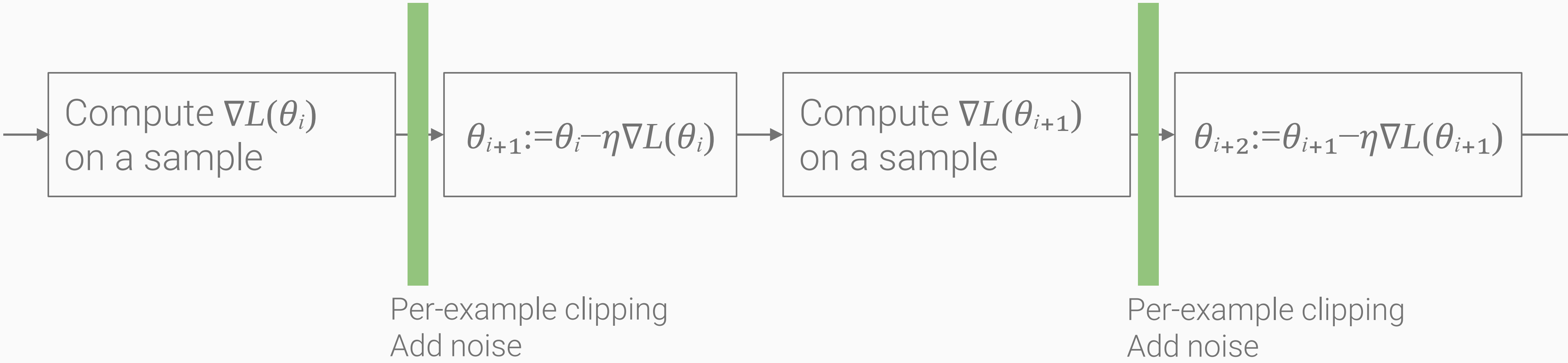
Differentially Private SGD

Abadi, Chu, Goodfellow, McMahan, Mironov, Talwar, Zhang, "Deep Learning with Differential Privacy", ACM CCS 2016

Differentially Private SGD



SGD with Differential Privacy



Naïve Privacy Analysis

1. Choose $\sigma = \frac{\sqrt{2 \log 1/\delta}}{\varepsilon}$ $= 4$
2. Each step is (ε, δ)-DP $(1.2, 10^{-5})$-DP
3. Number of steps T $10{,}000$
4. Composition: $(T\varepsilon, T\delta)$-DP $\boxed{(12{,}000, .1)\text{-DP}}$

Strong Composition Theorem

1. Choose $\sigma = \frac{\sqrt{2 \log 1/\delta}}{\varepsilon}$ $= 4$
2. Each step is (ε, δ)-DP $(1.2, 10^{-5})$-DP
3. Number of steps T $10{,}000$
4. Strong comp: $(\varepsilon\sqrt{T \log 1/\delta}, T\delta)$-DP $\boxed{(360, .1)\text{-DP}}$

Dwork, Rothblum, Vadhan, "Boosting and Differential Privacy", FOCS 2010
Dwork, Rothblum, "Concentrated Differential Privacy", https://arxiv.org/abs/1603.0188

Amplification by Sampling

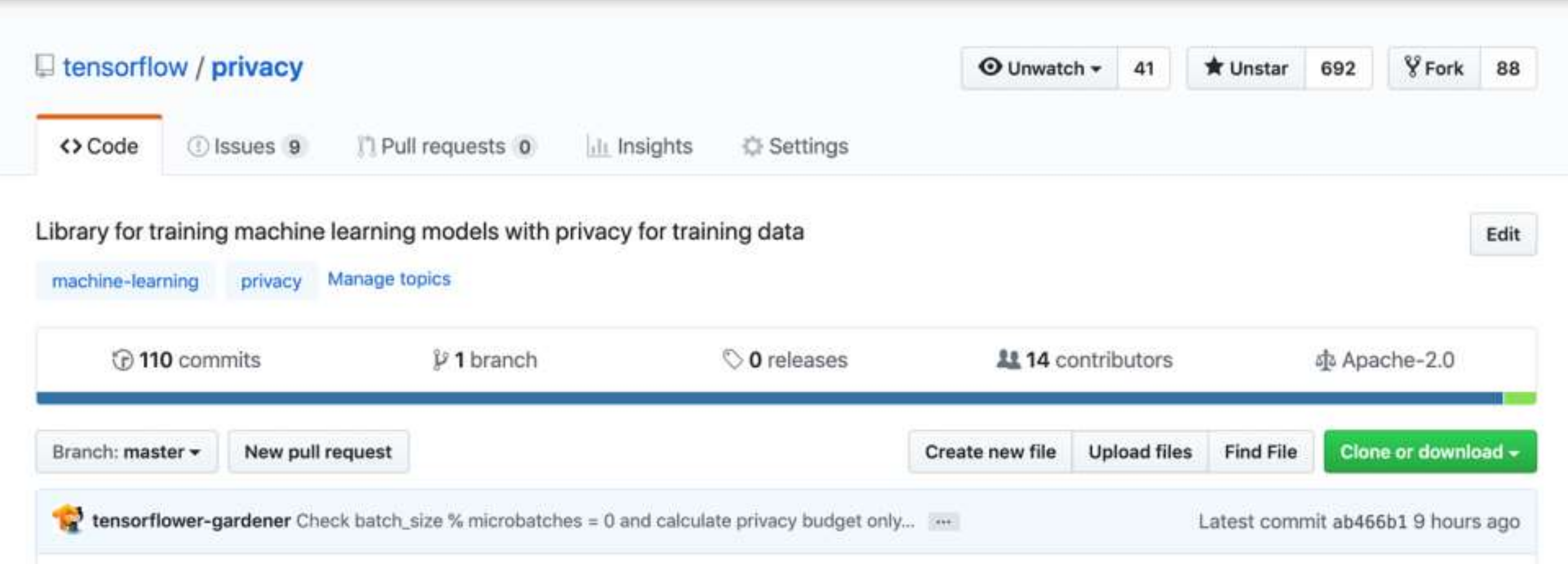
1. Choose $\sigma = \frac{\sqrt{2\log 1/\delta}}{\varepsilon}$ $= 4$
2. Each batch is q fraction of data 1%
3. Each step is $(2q\varepsilon, q\delta)$-DP $(.024, 10^{-7})$-DP
4. Number of steps T 10,000
5. Strong comp: $(2q\varepsilon\sqrt{T\log 1/\delta}, qT\delta)$-DP $\boxed{(10, .001)\text{-DP}}$

S. Kasiviswanathan, H. Lee, K. Nissim, S. Raskhodnikova, A. Smith, "What Can We Learn Privately?", SIAM J. Comp, 2011

Moments Accountant (Rényi Differential Privacy)

1. Choose $\sigma = \frac{\sqrt{2 \log 1/\delta}}{\varepsilon}$ $= 4$
2. Each batch is q fraction of data 1%
3. Keeping track of privacy loss's **moments**
4. Number of steps T $10{,}000$
5. Moments: $(2q\varepsilon\sqrt{T}, \delta)$-DP $\boxed{(1.25, 10^{-5})\text{-DP}}$

Differential Privacy in TensorFlow

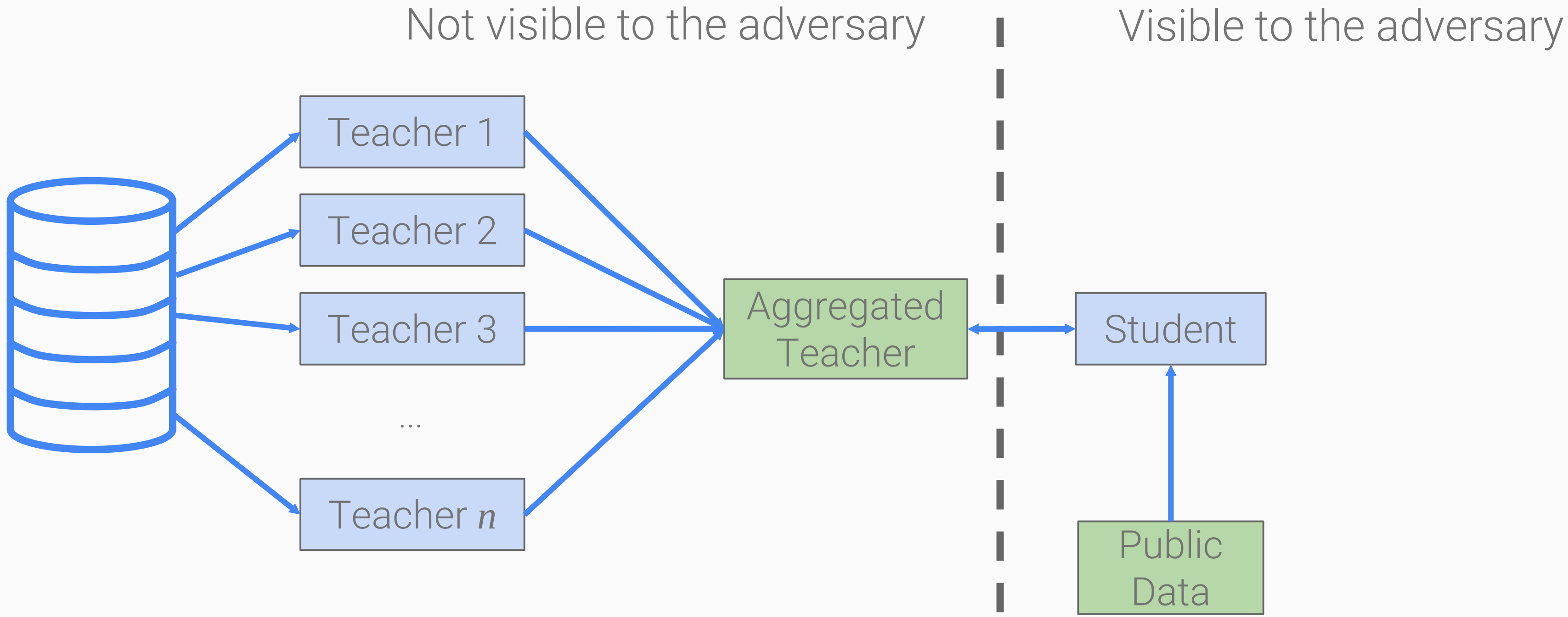


Private Aggregation of Teacher Ensembles: PATE

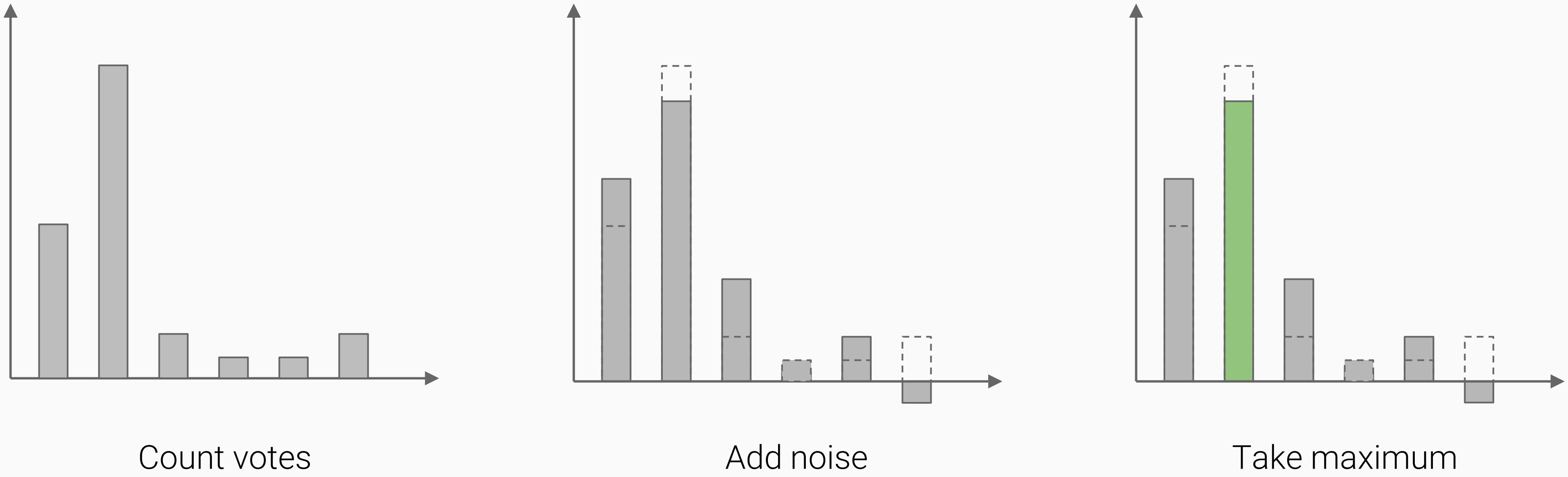
Papernot, Abadi, Goodfellow, Erlingsson, Talwar, "Semi-supervised Knowledge Transfer for Deep Learning from Private Training Data", ICLR 2017

Papernot, Song, Mironov, Raghunathan, Talwar, Erlingsson, "Scalable Private Learning with PATE", ICLR 2018

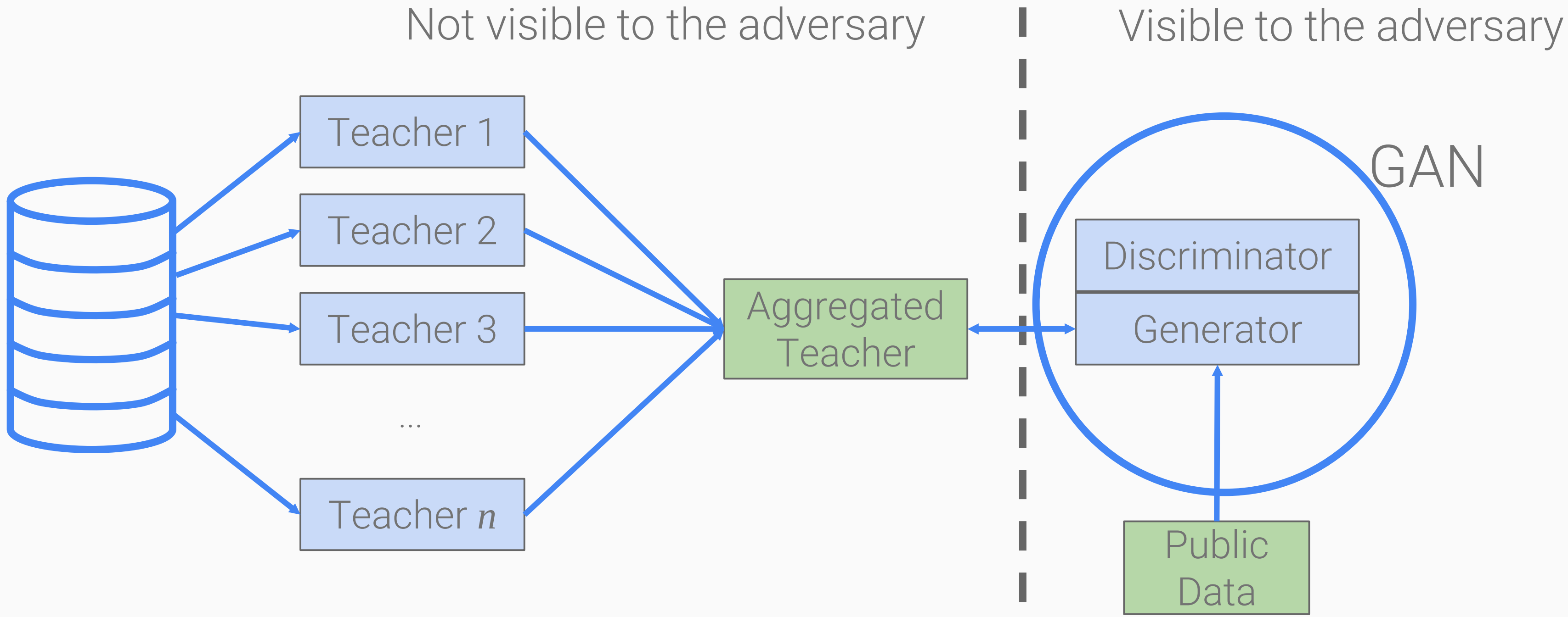
PATE at a Glance: Sample-and-Aggregate



Differentially Private Aggregation



Semi-Supervised Setting: PATE-G



References

DP-SGD

- Abadi et al., “Deep Learning with Differential Privacy”, ACM CCS 2016
- https://github.com/tensorflow/privacy
- Blog post: Radebaugh and Erlingsson, “Introducing TensorFlow Privacy: Learning with Differential Privacy for Training Data”, 2019

PATE:

- Papernot et al., Semi-supervised Knowledge Transfer for Deep Learning from Private Training Data”, ICLR 2017
- Papernot et al., “Scalable Private Learning with PATE”, ICLR 2018