BASGD: Buffered Asynchronous SGD for Byzantine Learning

Yi-Rui Yang, Wu-Jun Li

ICML 2021

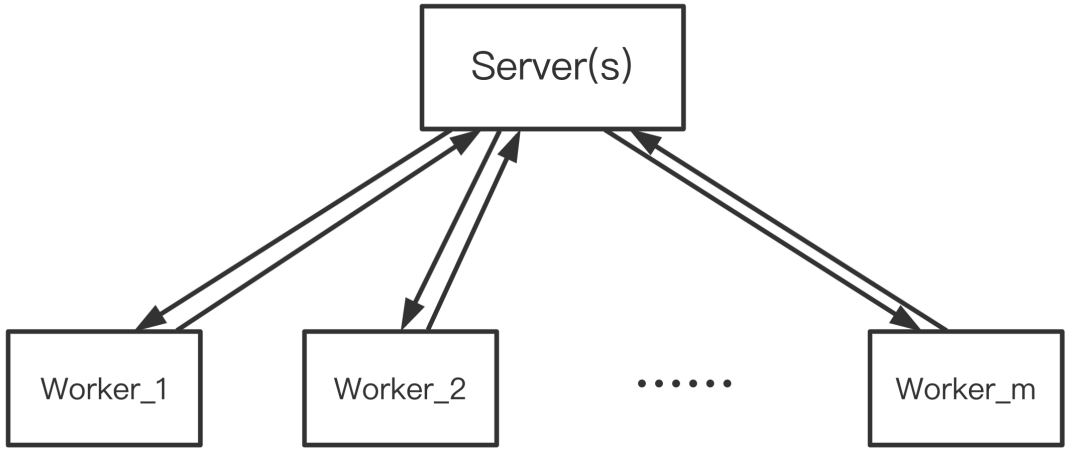
National Key Laboratory for Novel Software Technology, Department of Computer Science and Technology, Nanjing University, China.

Introduction

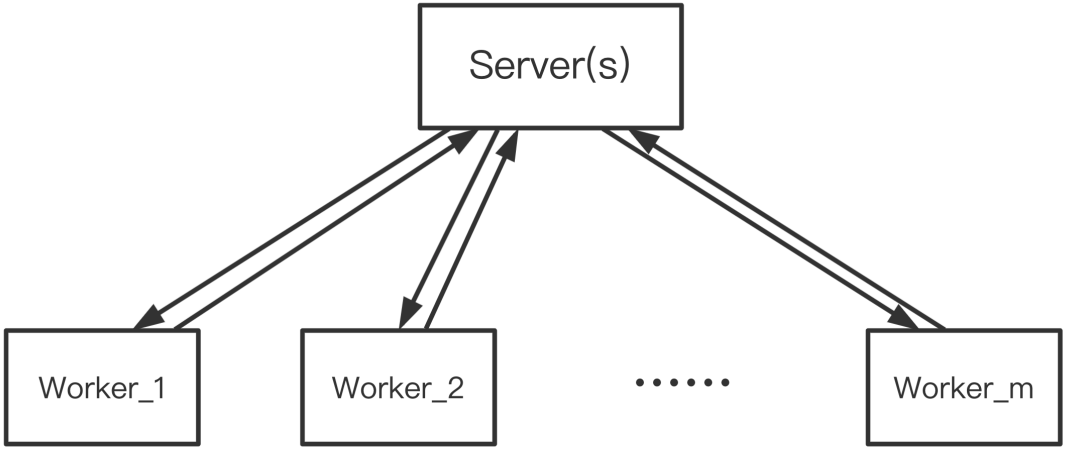
- Distributed machine learning (DML):

$$\min_{\mathbf{w}\in\mathbb{R}^d} F(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{w}; z_i),$$

- Parameter-server framework:

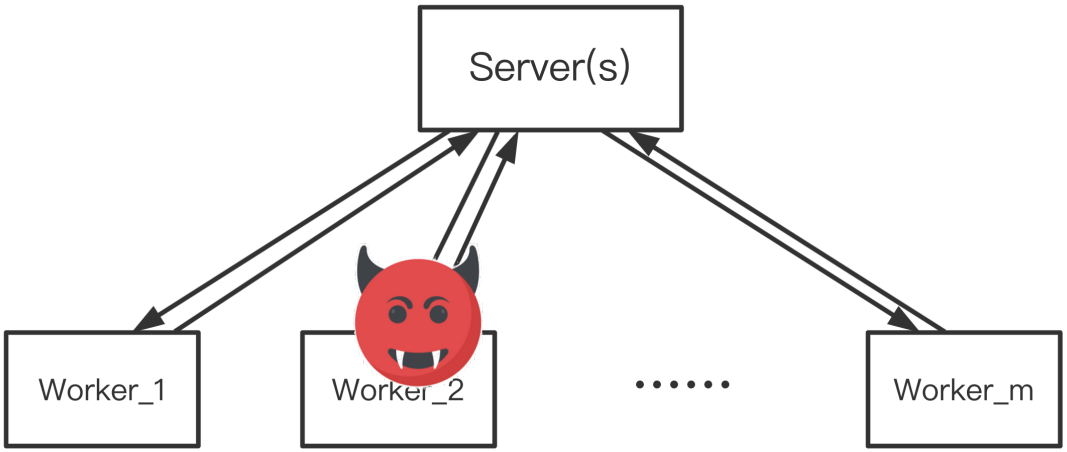


Introduction



- Traditional distributed learning methods typically assume no failure or attack.

Introduction



- Traditional distributed learning methods typically assume no failure or attack.

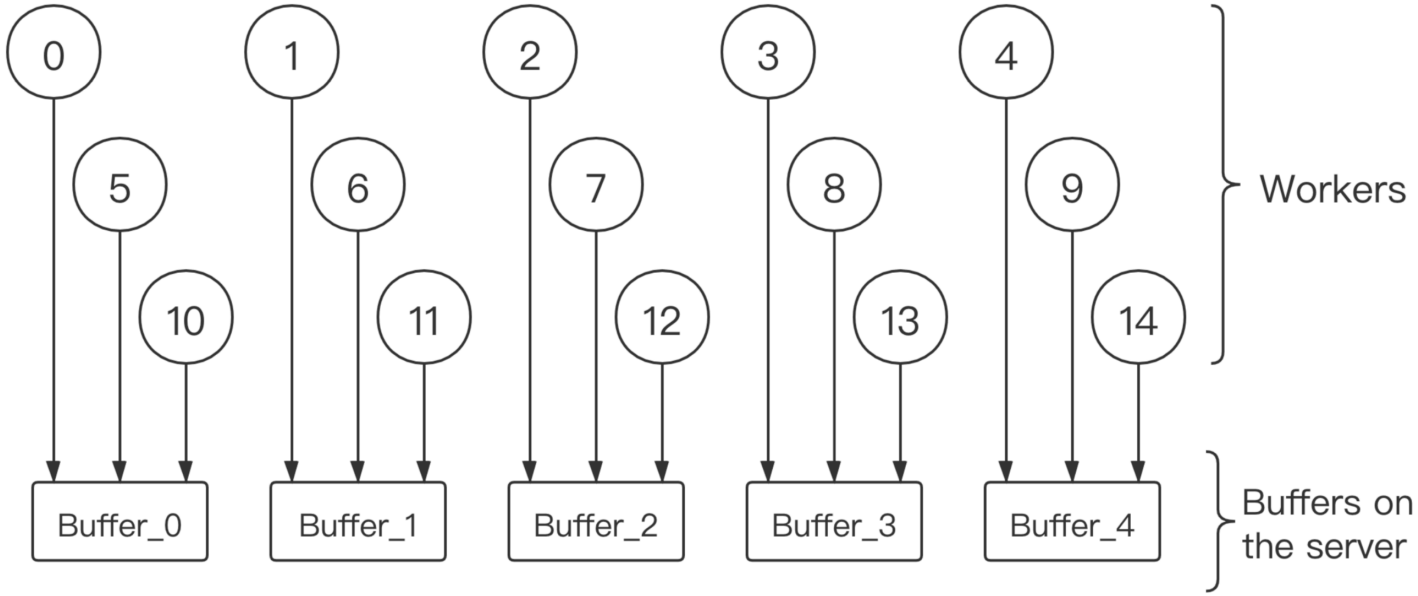
- In real applications, failure or attack may happen.

Buffered Asynchronous SGD

- The first asynchronous DML method that
  - can resist malicious attack
  - does not need to store instances on server
- Theoretical guarantee of convergence and robustness
- Significantly better empirical performance than baselines

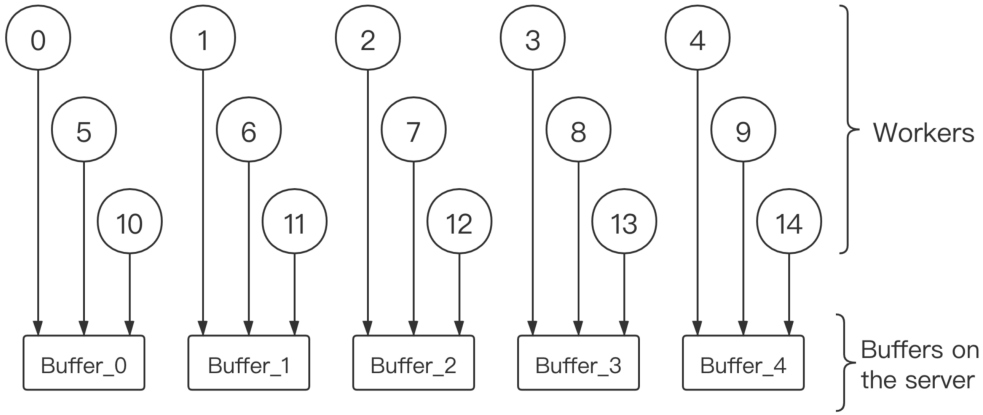
Buffered Asynchronous SGD

- Buffers: $h_1, h_2, \ldots, h_B$



Buffered Asynchronous SGD

- Buffers: $h_1, h_2, \ldots, h_B$

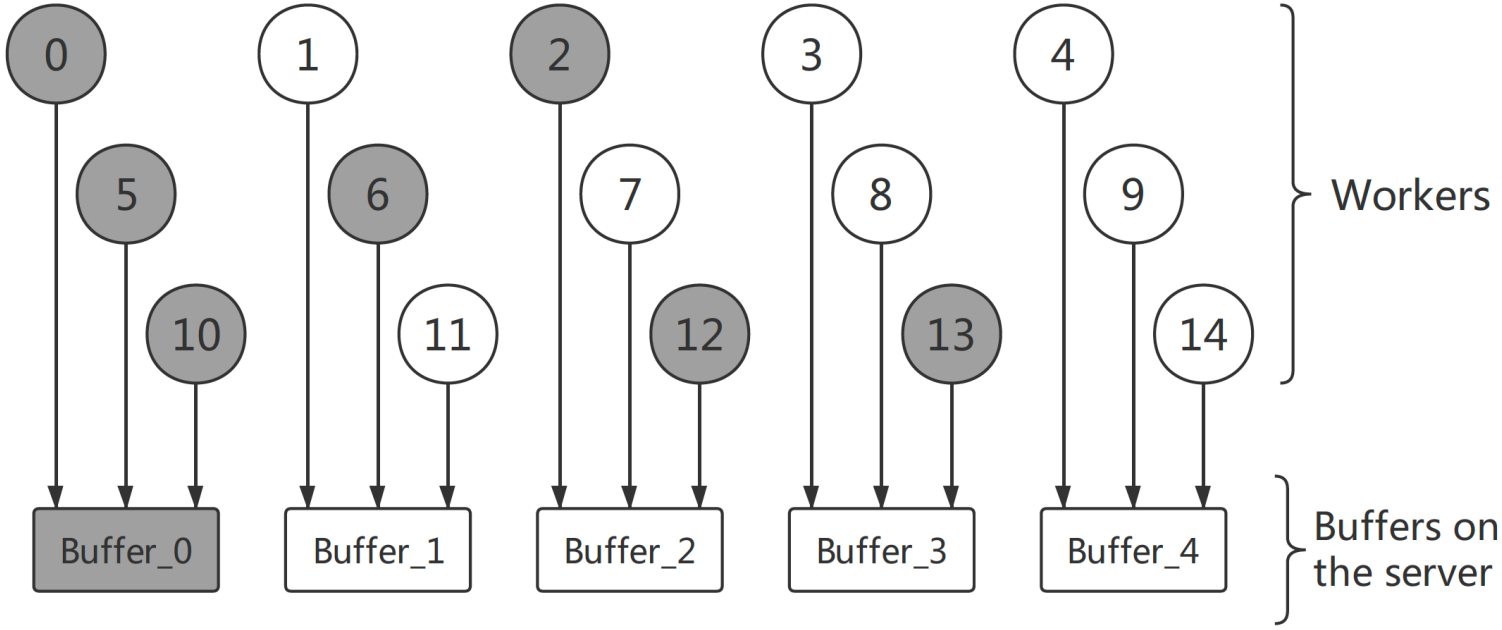


- Store the average value of gradients in buffers
- Update model parameters when all buffers are filled:
  - Aggregate: $\mathrm{G}^t = Aggr([h_1, h_2, \ldots, h_B])$
  - Execute SGD step: $w^{t+1} = w^t - \eta \cdot \mathrm{G}^t$
- Zero out all buffers after each SGD step

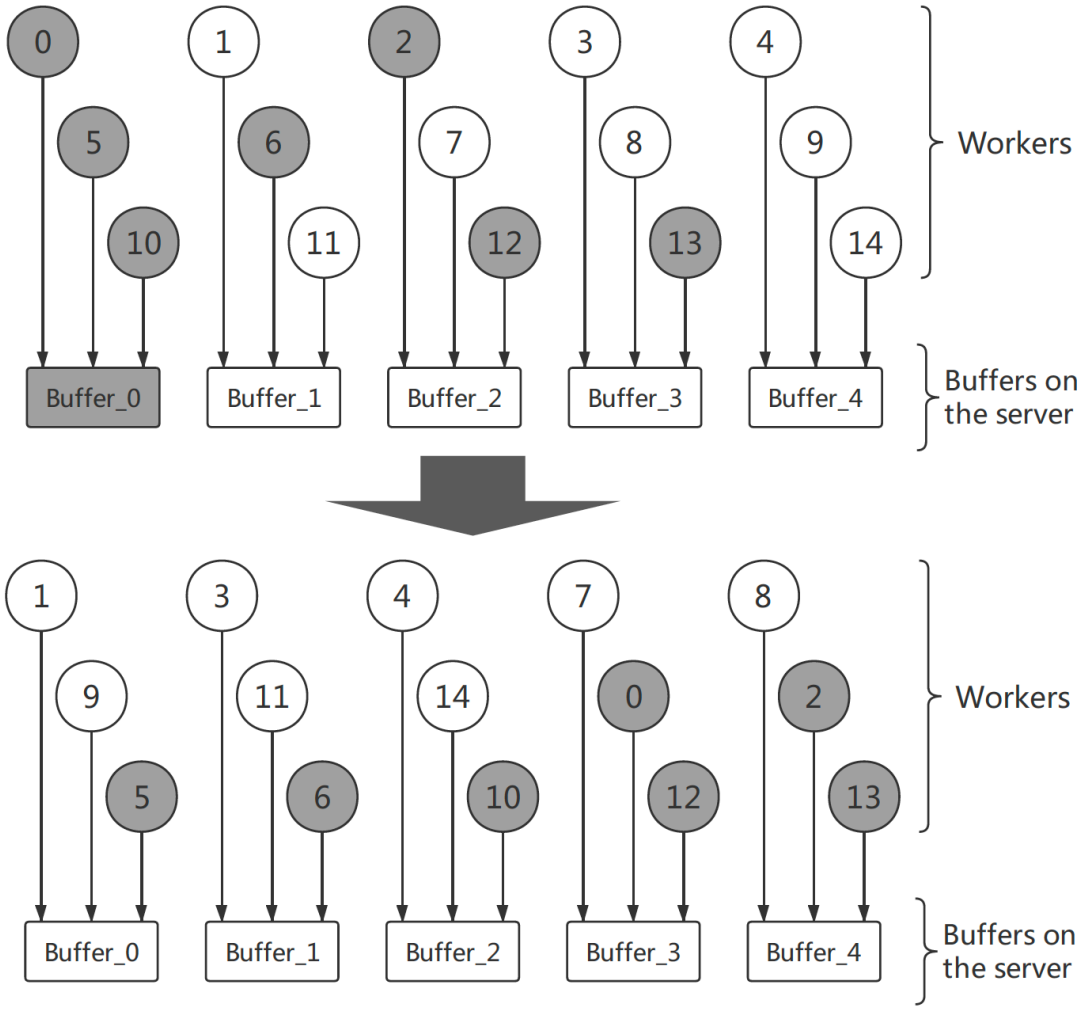
Buffered Asynchronous SGD

- Buffer reassignment:

  Improve the performance in extreme cases



Buffered Asynchronous SGD



Convergence

$$\frac{\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla F(\mathbf{w}^t)\|^2]}{T} \leq O\left(\frac{L^{\frac{1}{2}}[F(\mathbf{w}^0)-F^*]}{T^{\frac{1}{2}}}\right) + O\left(\frac{L^{\frac{1}{2}}\tau_{max}DA_2r^{\frac{1}{2}}}{T^{\frac{1}{2}}}\right) + O\left(\frac{L^{\frac{1}{2}}(A_2)^2}{T^{\frac{1}{2}}}\right) + O\left(\frac{L^{\frac{5}{2}}(A_2)^2\tau_{max}^2 r}{T^{\frac{3}{2}}}\right) + A_1.$$

New asynchronous methods can be obtained from synchronous ones when using BASGD.

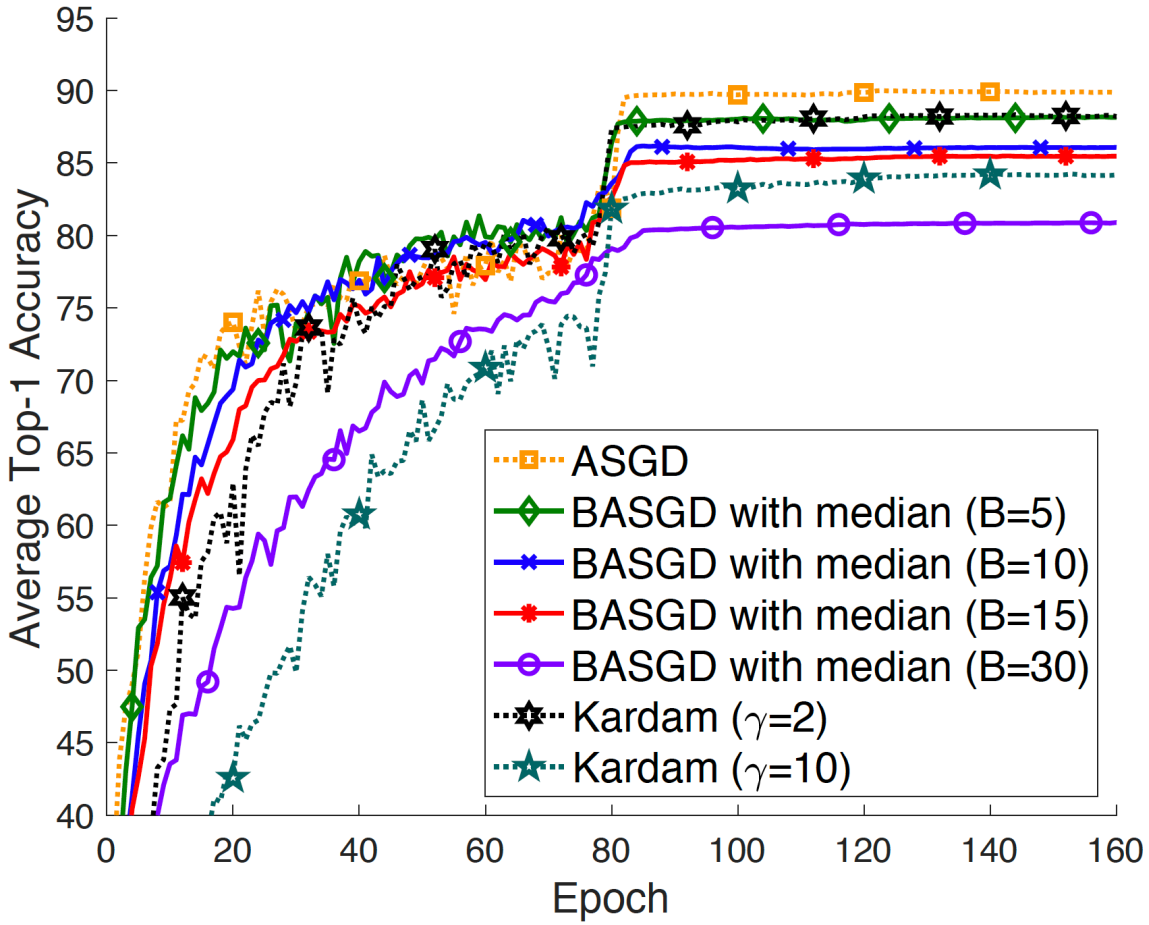
Experiment

Settings:

- Baselines: Asynchronous SGD & Kardam
- Attacks:
  - No attack
  - Negative-gradient (NG) attack
    (a typical kind of malicious attack)
  - Random-disturbance (RD) attack
    (can be seen as accidental failure)

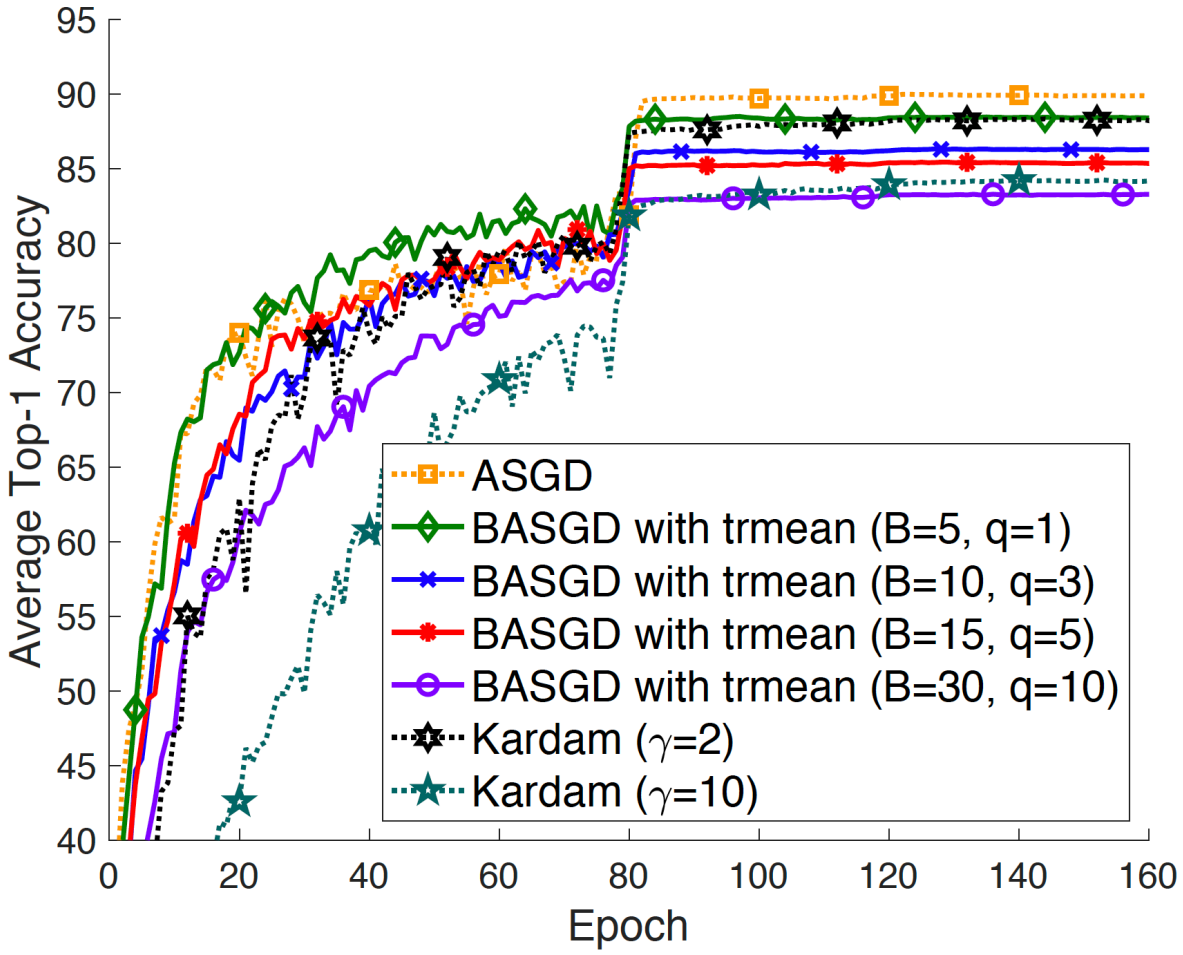
Experiment

CIFAR-10, ResNet-20 (no attack):



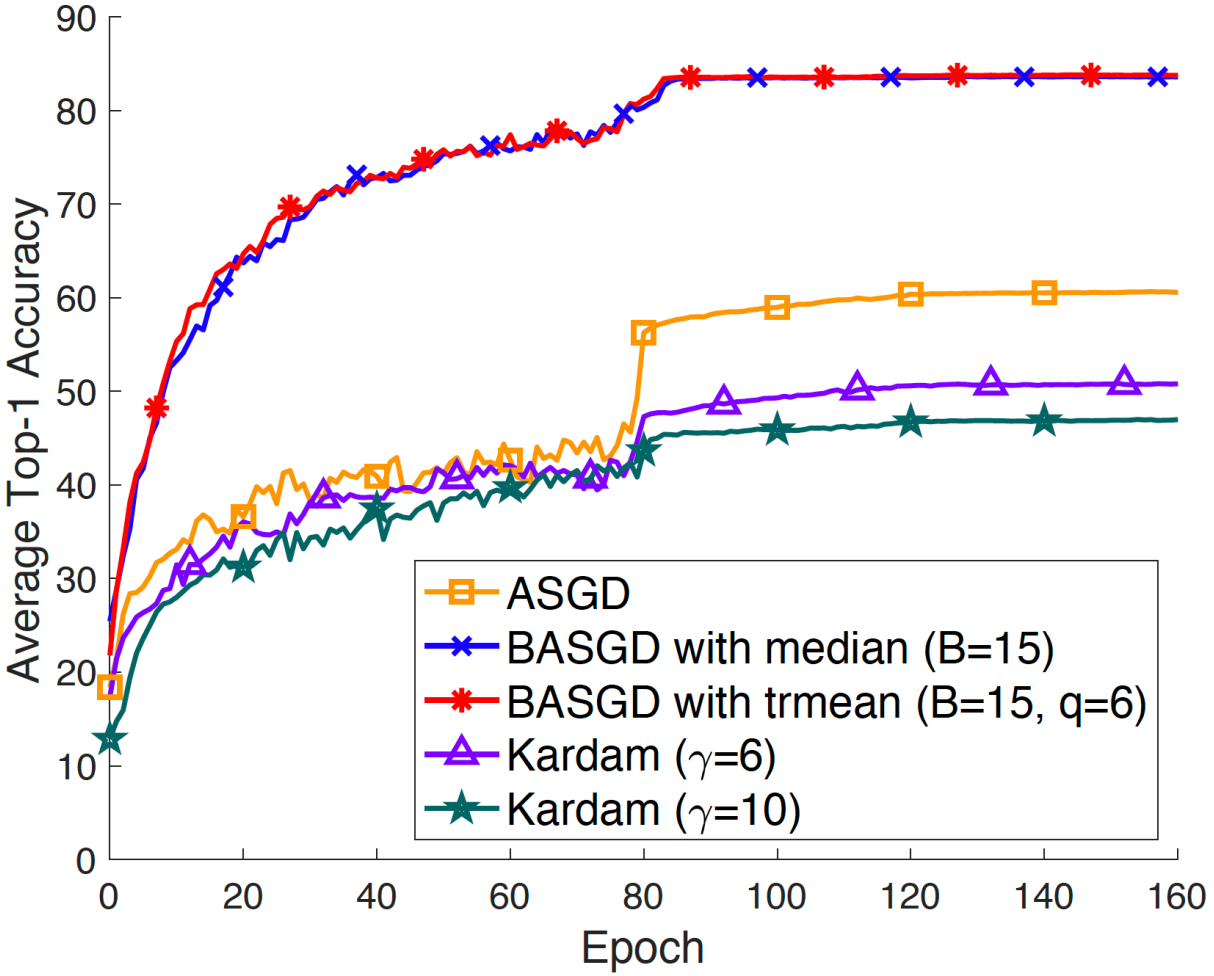
Experiment

CIFAR-10, ResNet-20 (no attack):



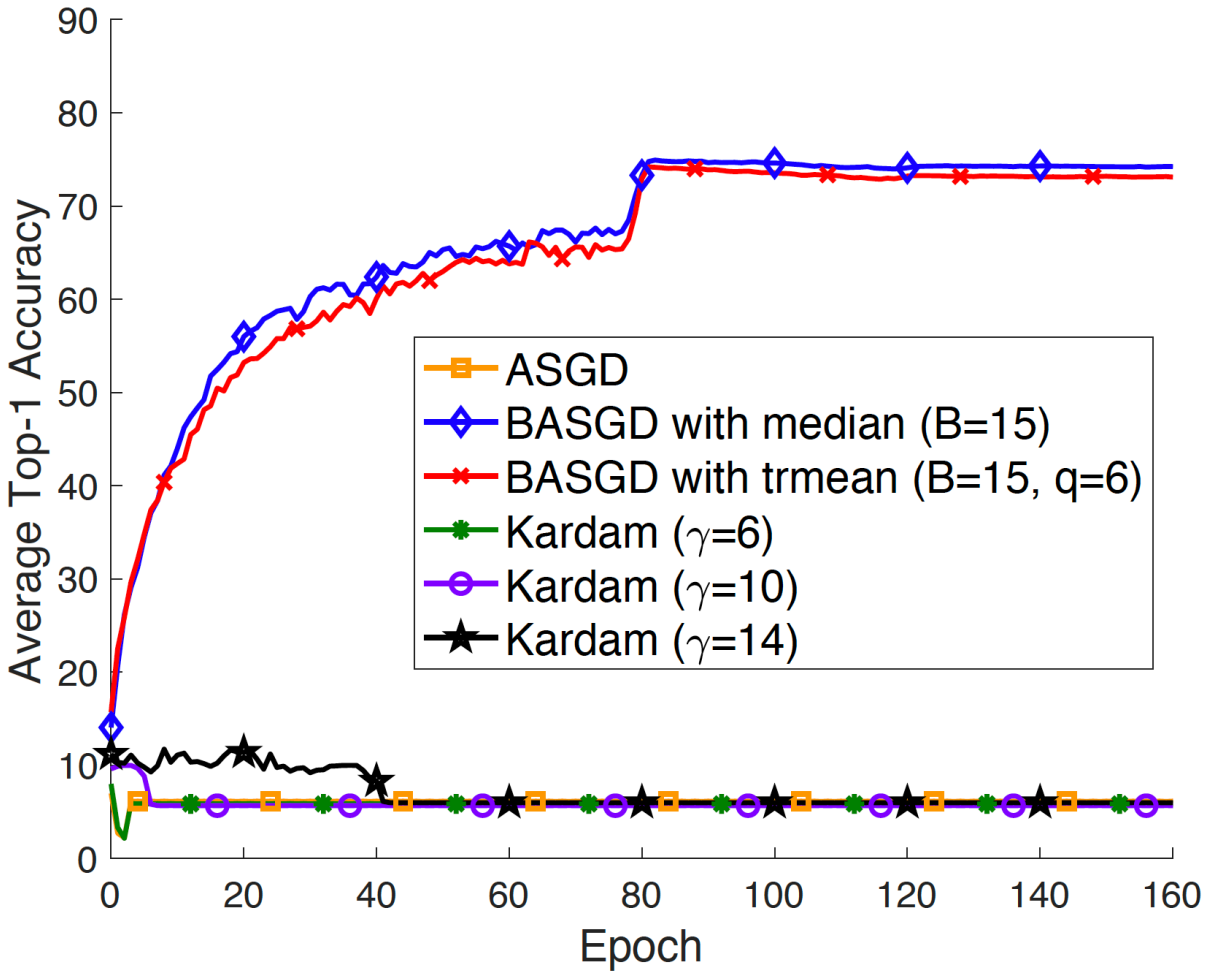
Experiment

CIFAR-10, ResNet-20 (6 malicious workers, RD-attack):



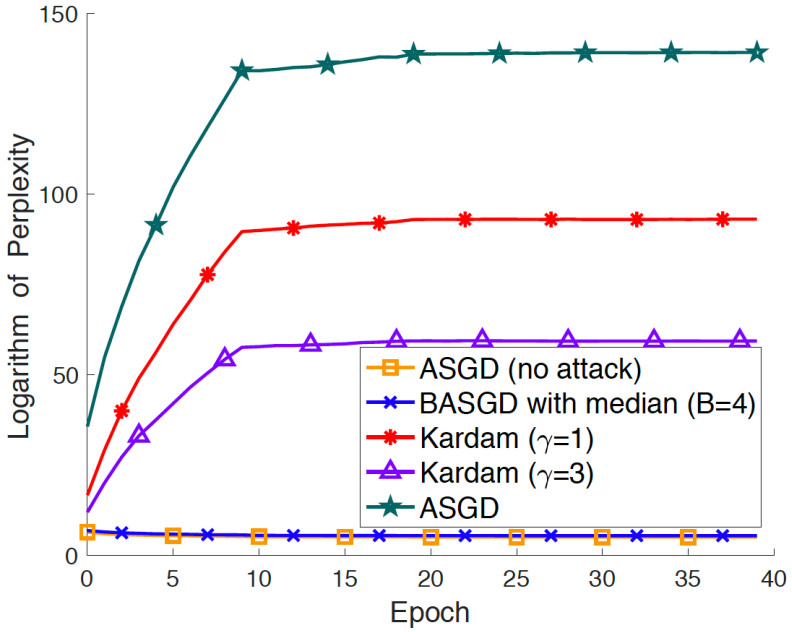
Experiment

CIFAR-10, ResNet-20 (6 malicious workers, NG-attack):

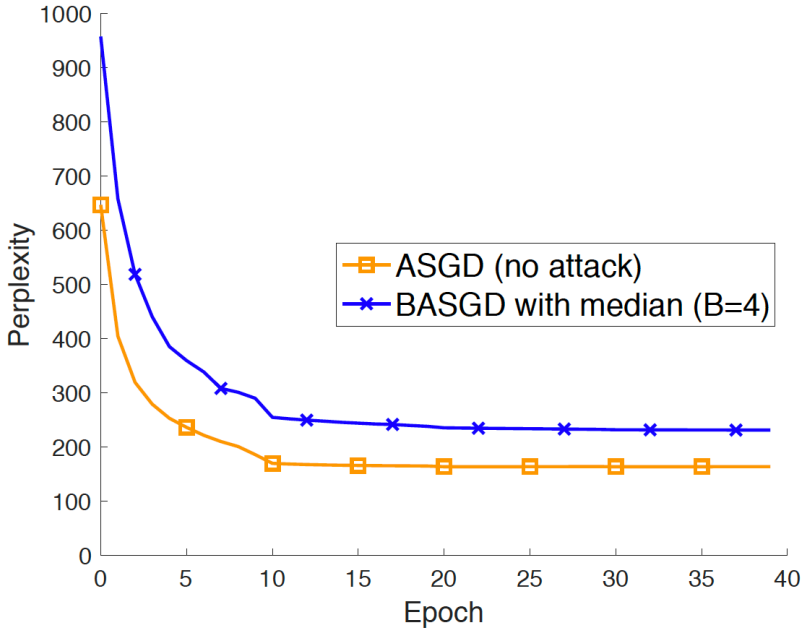


Experiment

WikiText-2, LSTM (1 malicious worker, RD-attack):



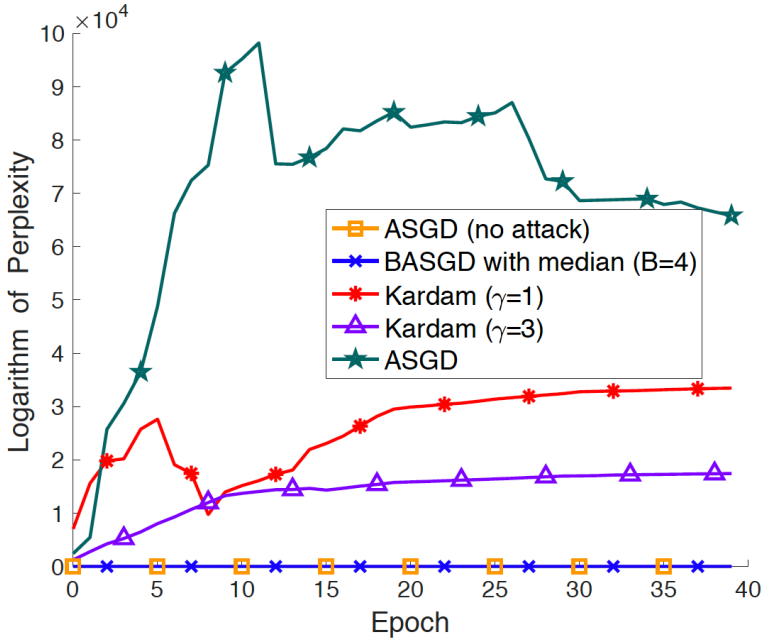
(a) RD-attack



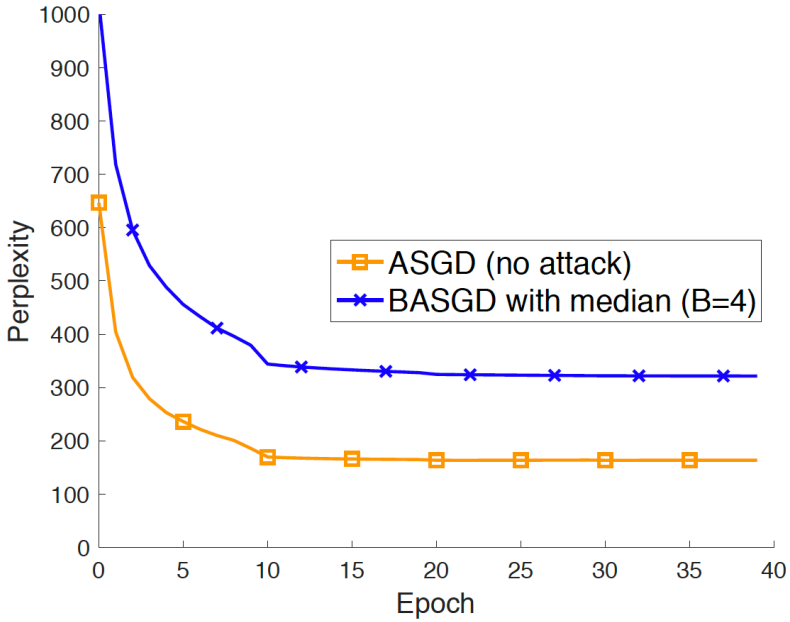
(b) RD-attack (magnified)

Experiment

WikiText-2, LSTM (1 malicious worker, NG-attack):



(c) NG-attack



(d) NG-attack (magnified)

Conclusion

Buffered Asynchronous SGD:

- The first asynchronous DML method that
  - can resist malicious attack
  - does not need to store instances on server
- Theoretical guarantee of convergence and robustness
- Significantly better empirical performance than baselines


LAMDA
Learning And Mining from DatA
http://lamda.nju.edu.cn
南京大学
NANJING UNIVERSITY
1902

Thank you !