Elastic Parameter Server: Accelerating ML Training With Scalable Resource Scheduling

谢礼安　2021.11.26

INTRODUCTION

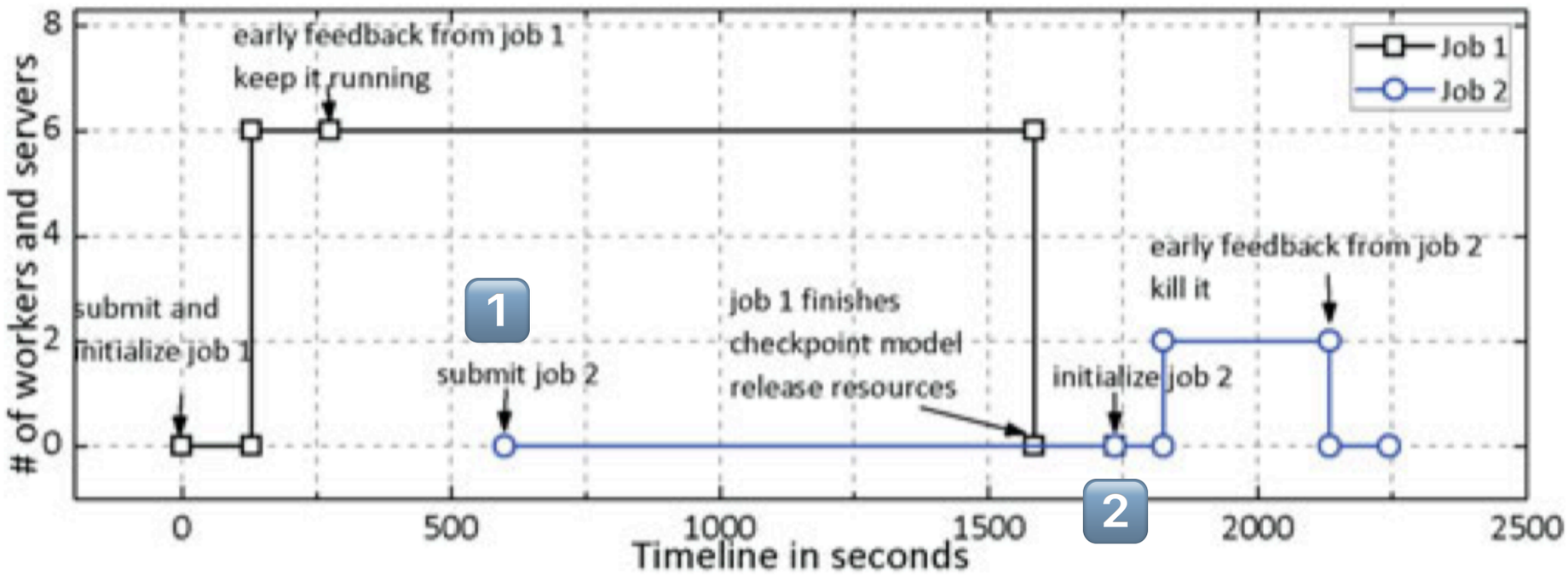
PS Cluster的限制

- 在job开始时进行资源分配，job进行时无法调整Worker和Server的数量
- 新job到来时，如果剩余资源无法满足训练的要求，则需要进入等待队列
- 排队会导致的job的early feedback延迟，无法快速地筛选出优质模型

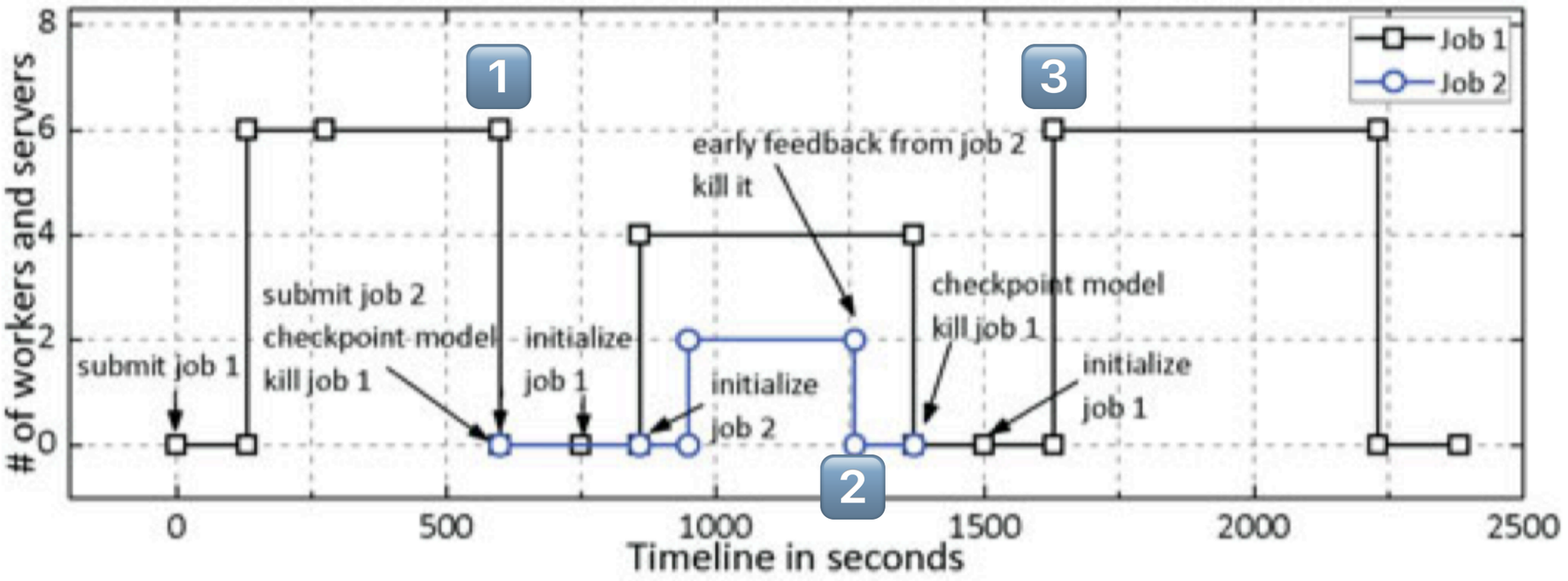
INTRODUCTION

PS Cluster的限制

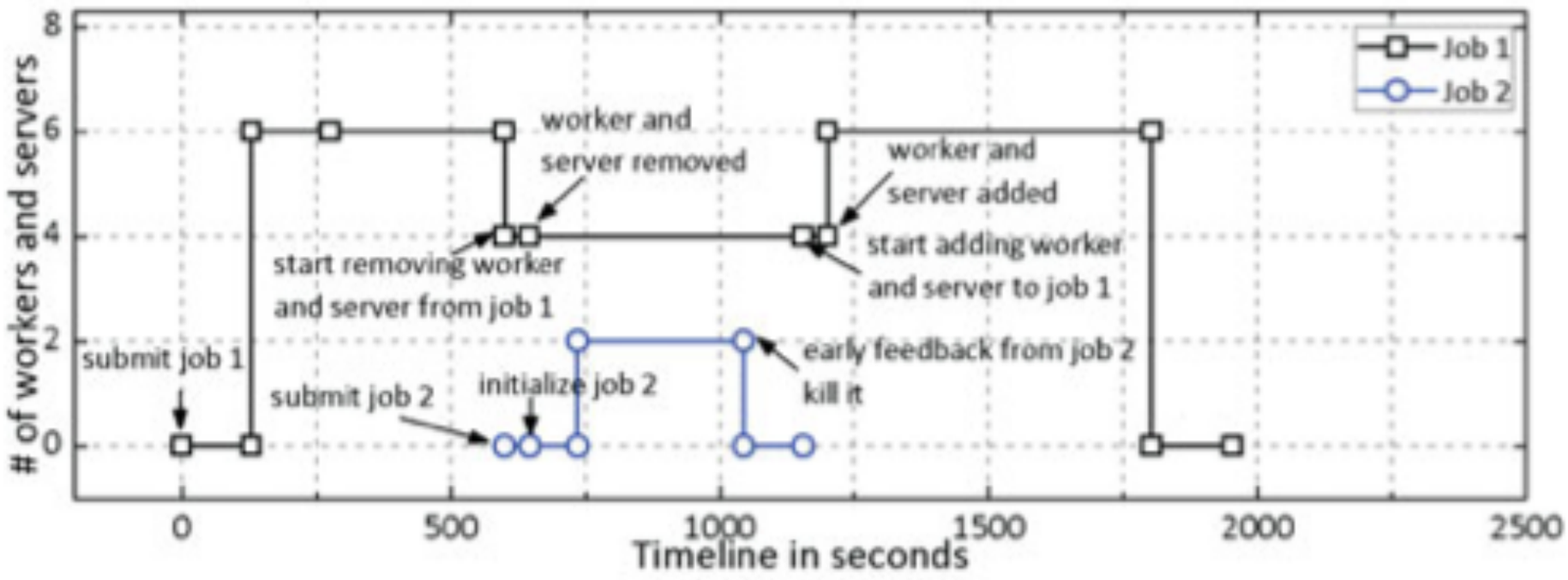
- 传统的PS：没有对jobs的调度能力，只能根据FIFO进行训练
- 基于Kill-based的PS：新jobs进入时，会存储当前所有训练jobs的状态，重新分配资源。
- Elastic PS：新jobs进入时，在2次iteration之间进行资源的释放



(a) Job execution due to the default PS.



(b) Job execution due to the kill-based approach.



(c) Job execution due to EPS.

EPS DESIGN

Elastic Parameter Server

- 作者提出能够进行动态分配资源的Parameter Server的方案
- 在新的任务到来时，从正在运行的任务中释放一部分资源给新任务，更早开始训练新任务

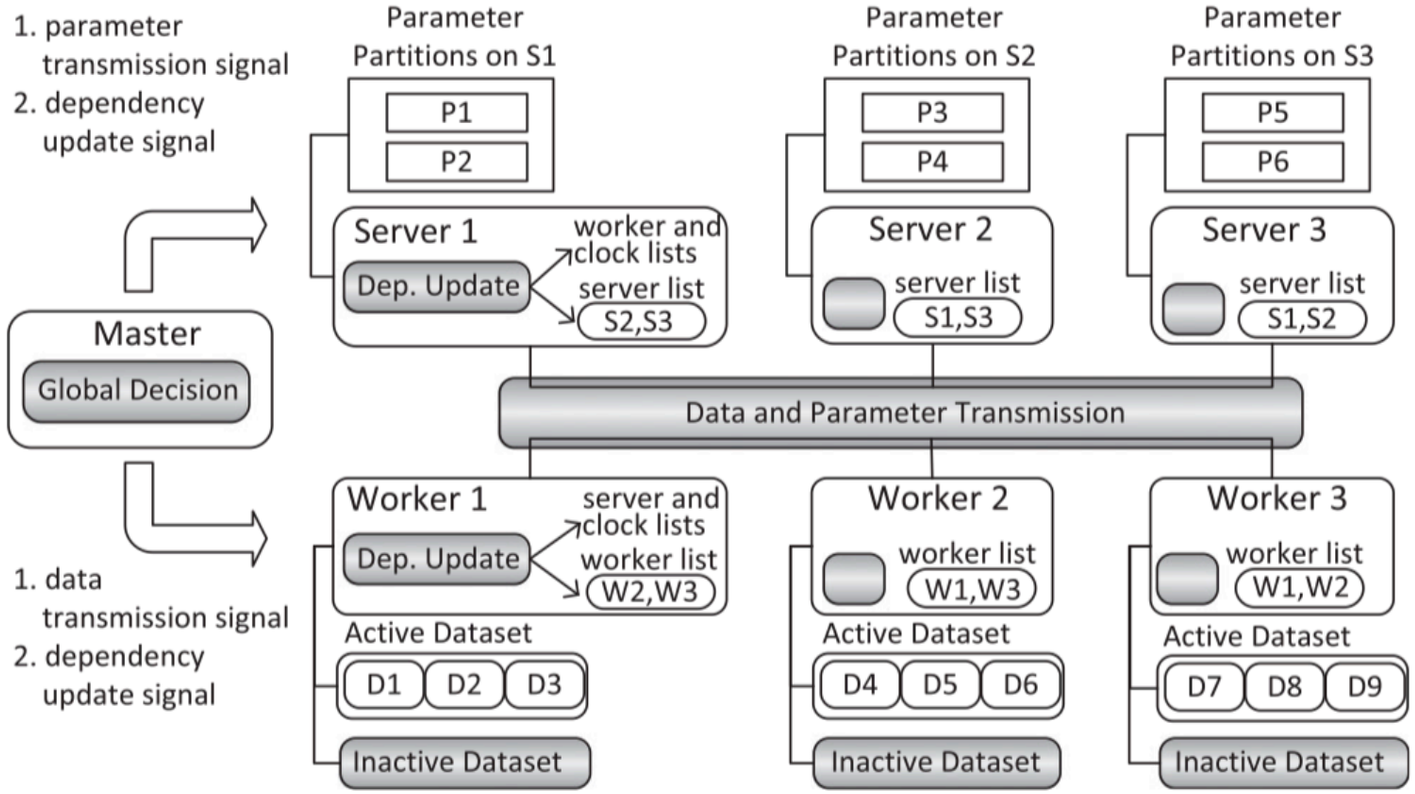


Fig. 4. The architecture of EPS.

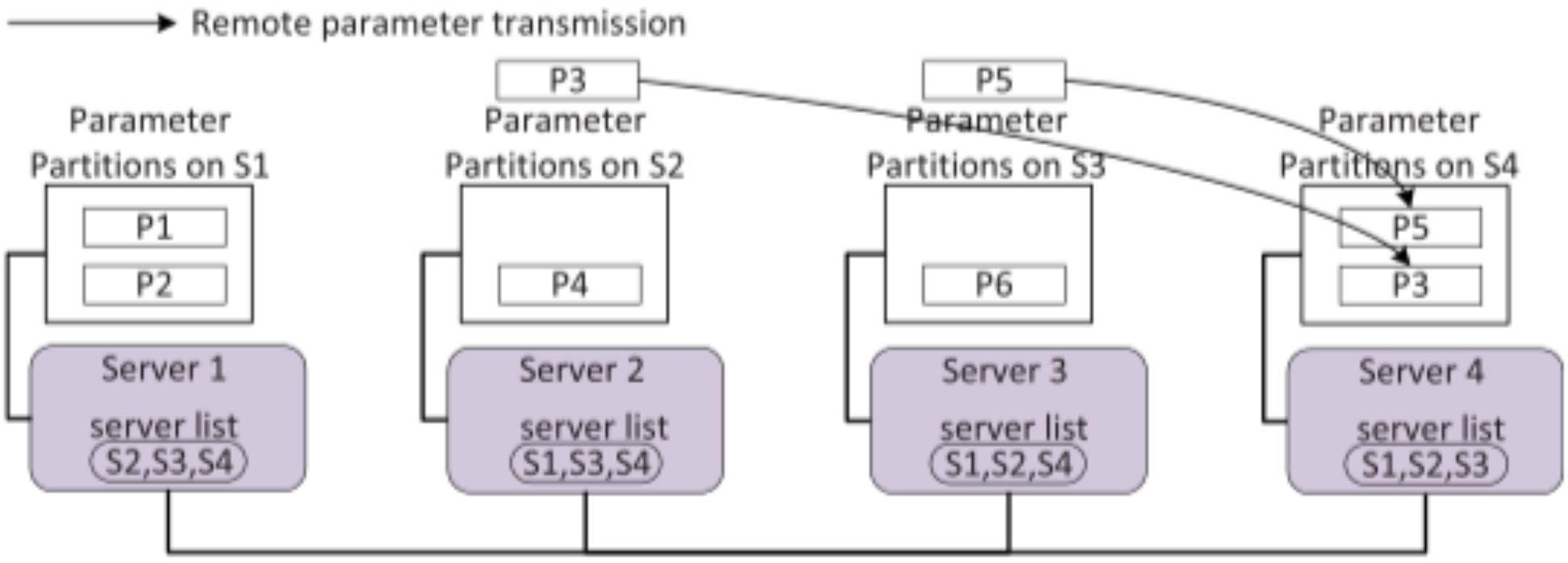
- Global Decision
- Two Datasets
- Update Function

EPS DESIGN

Adding Servers

- 查找已经early feedback的jobs，释放一部分这些jobs的Server，分配给新jobs

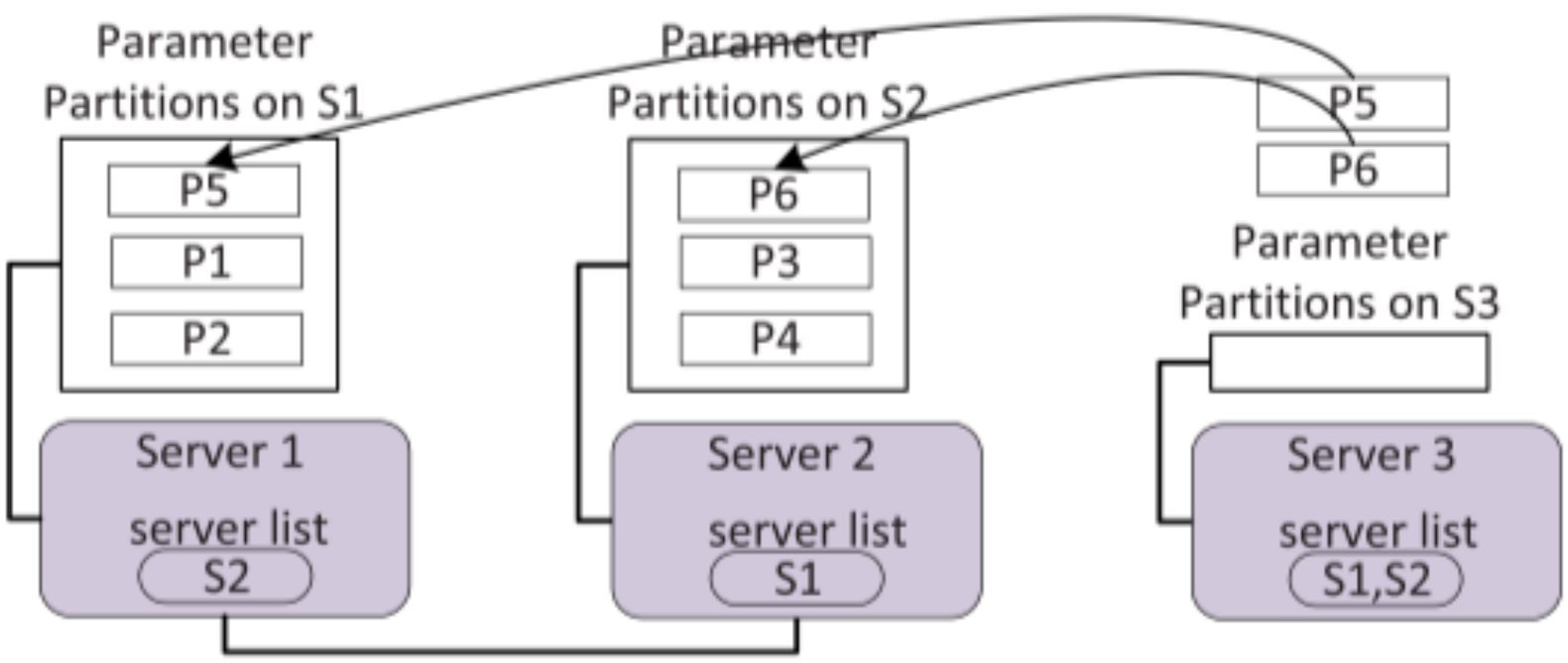
- 添加服务器，根据当前的服务器的部分参数传给新服务器并更新Server List



EPS DESIGN

Removing Servers

- 移除服务器，将需要移除的Server的参数分发给剩余的Server，并更新Server List



EPS DESIGN

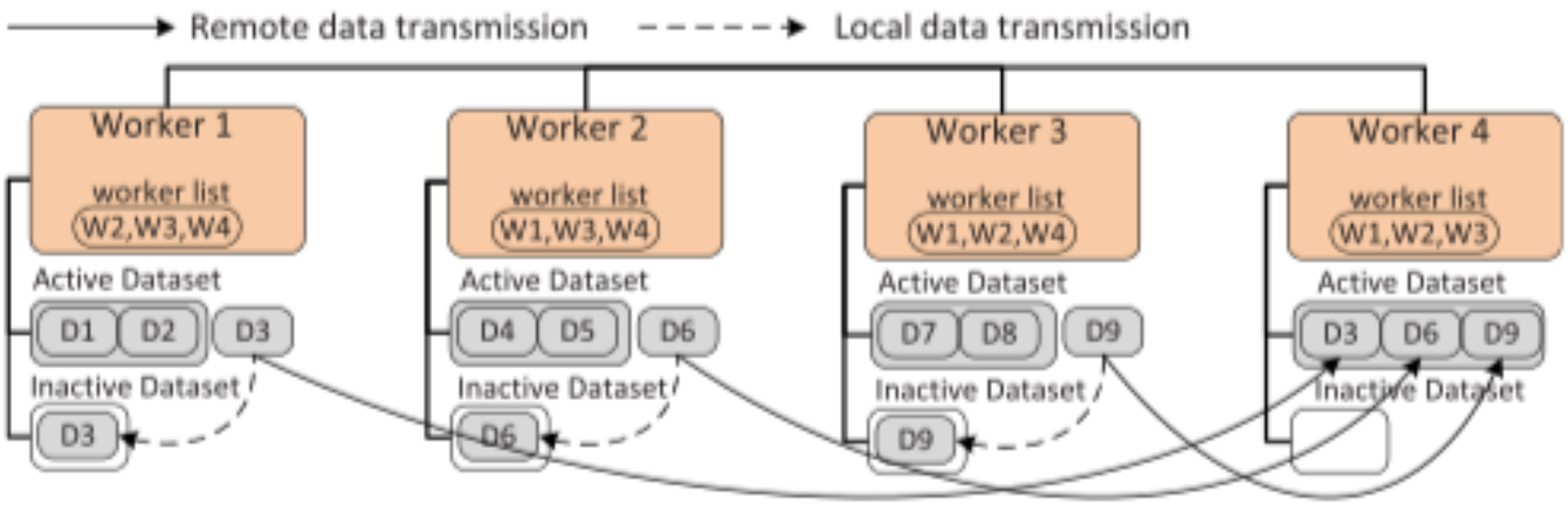
Adding Worker

- D：数据集
- N：Worker数量
- 每个Worker分到的数据集：D/N
- Worker对数据集进行进一步划分：D/N^2

- 传递Dataset：每个Worker传递一份Dataset给新Worker，剩余$(N-1)*(w/N^2)$，传递的Dataset设为Inactive

- 新Worker数据集：$N*(w/N^2)$

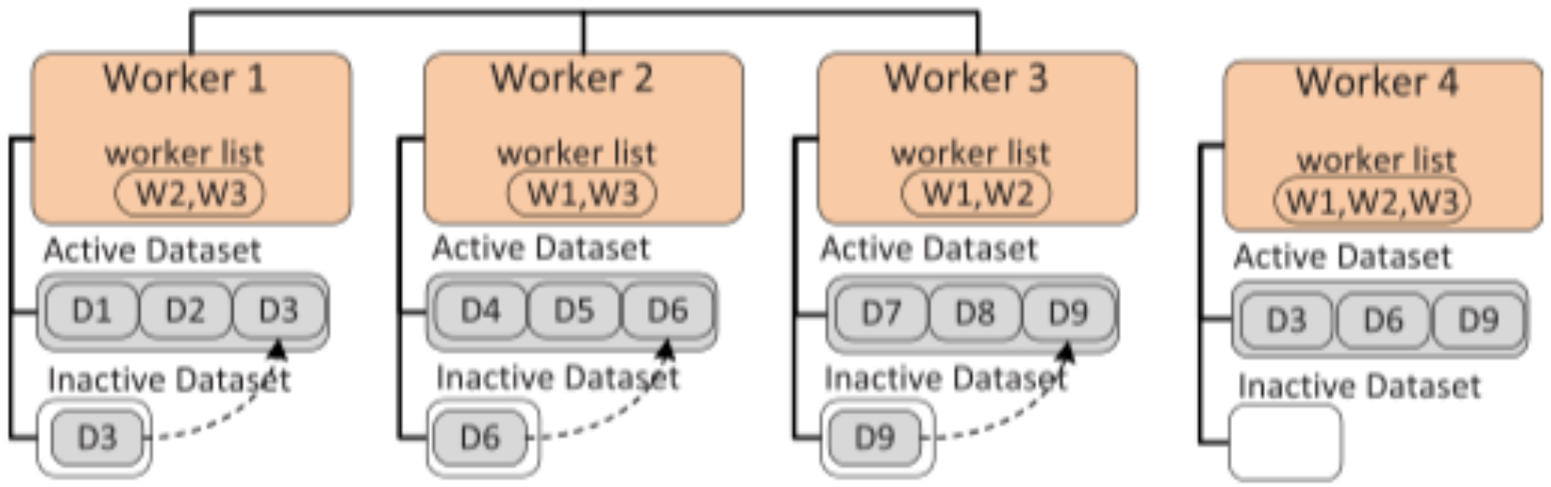
- 数据划分完毕后，更新worker list
- $N=3$：



EPS DESIGN

Removing Worker

- 移除的Worker
- 其他Worker从Inactive Dataset中取回对应Dataset



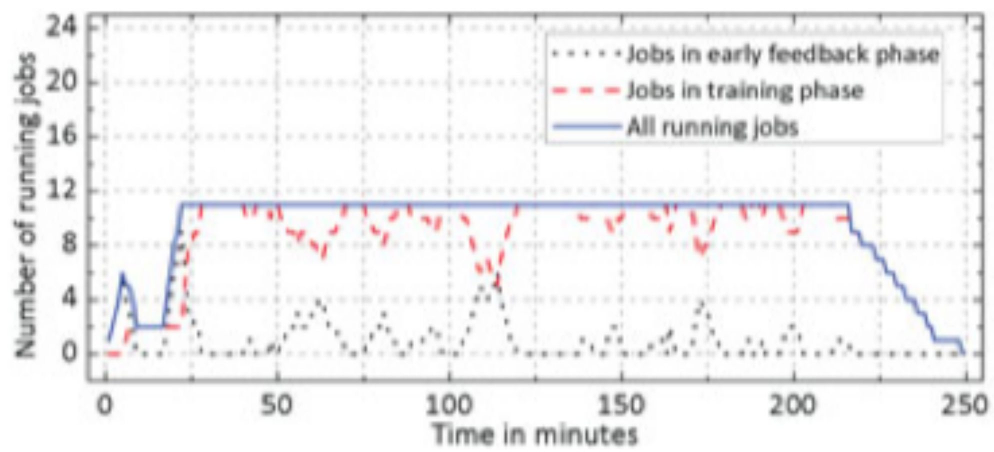
EPS DESIGN

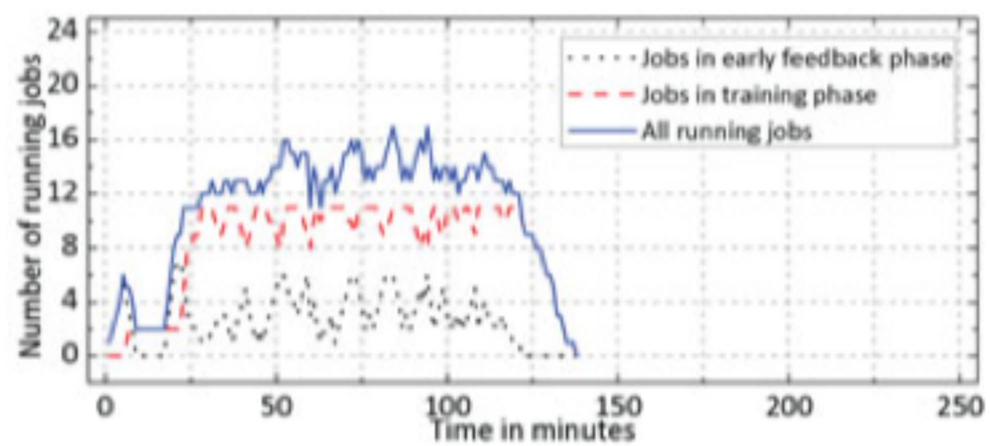
Job Scheduling

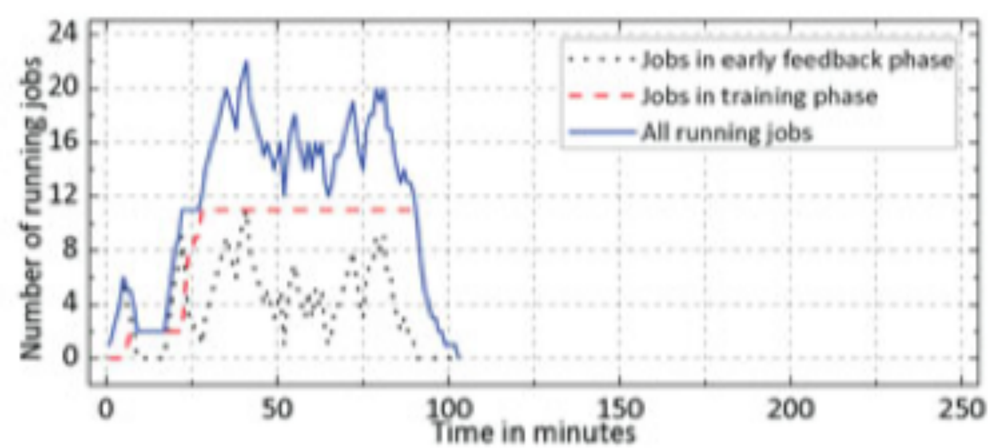
- 当集群中存在闲置资源时，EPS通过轮训，每轮对一个running job增加一个Server和Worker
- 监控每一个job，如果通信开销增大而导致训练速度降低，则需要rollback。

EVALUATION

Training Time







EVALUATION

Early Feedback

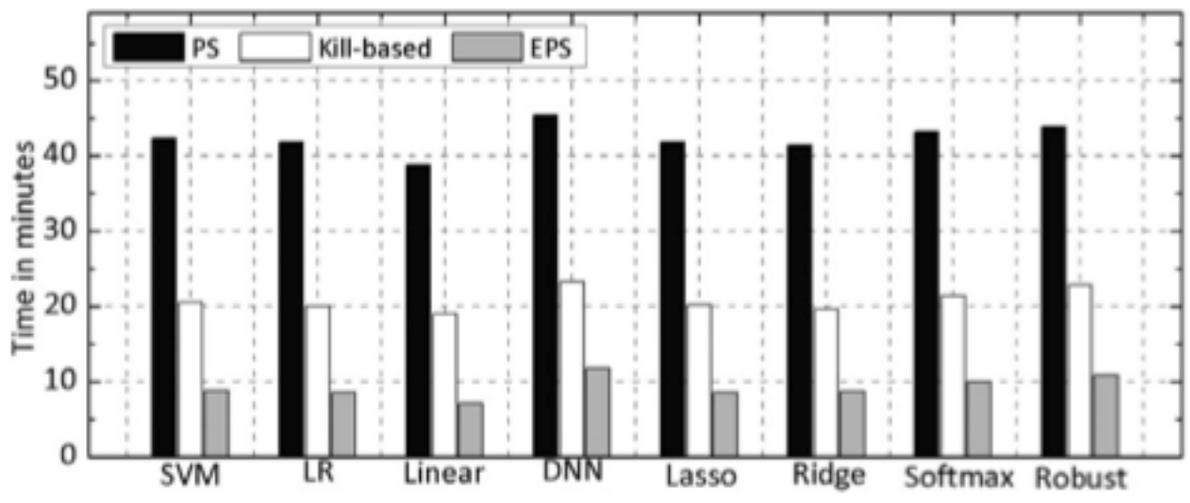


Fig. 17. The average early feedback latency.

EVALUATION

Resource Utilization

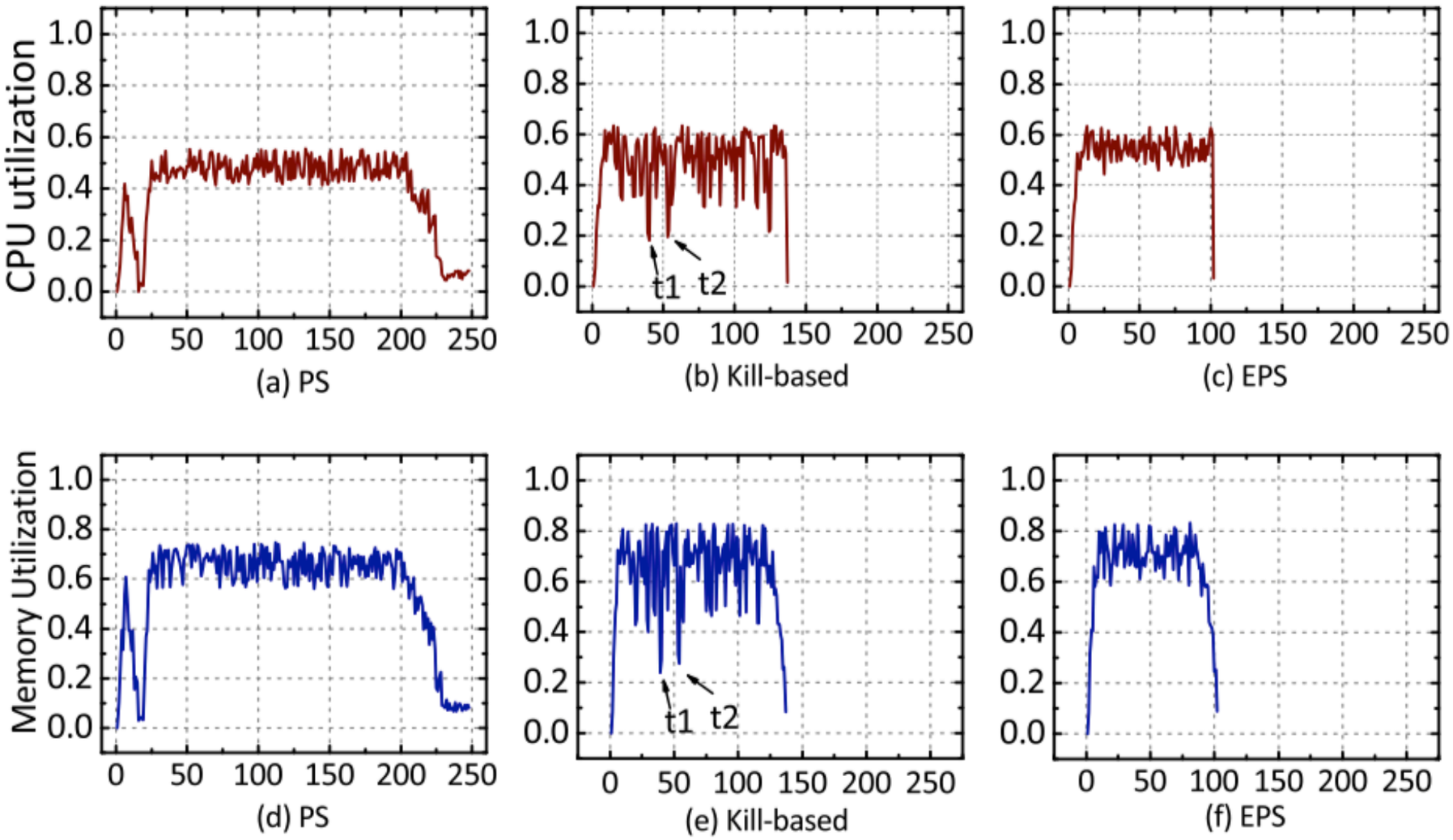
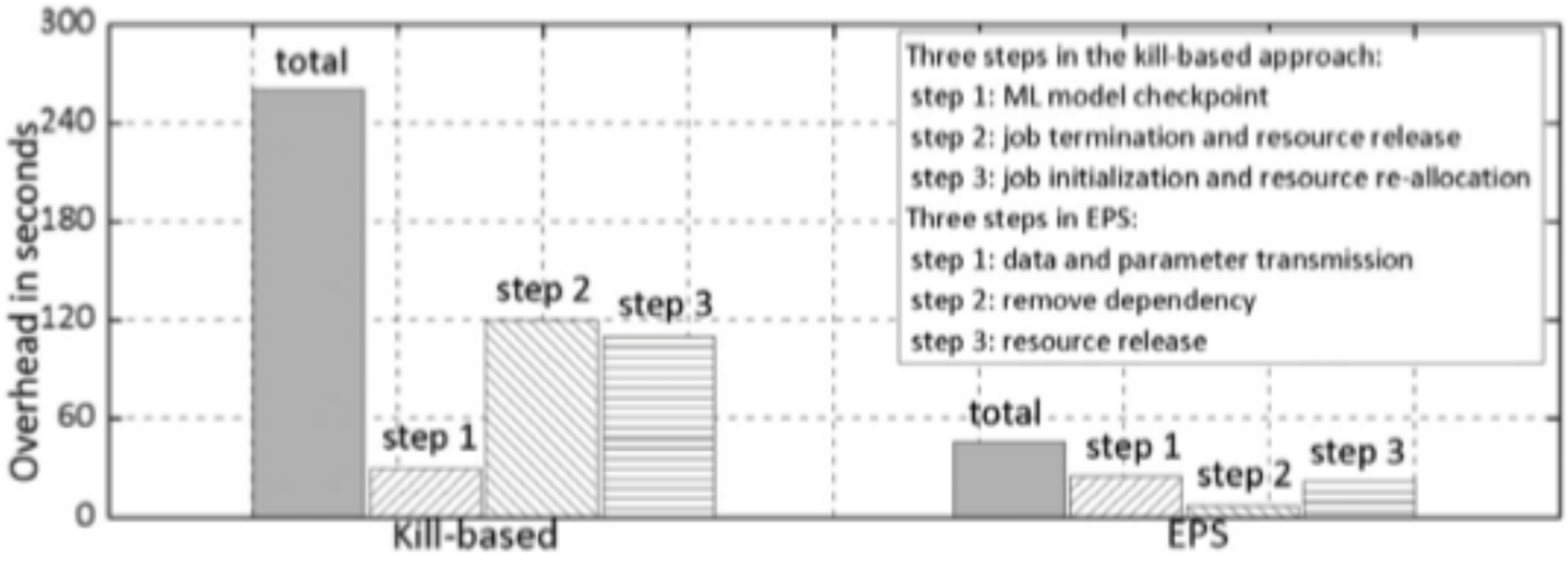


Fig. 19. CPU and memory utilizations during LR training.

EVALUATION

Removing Overhead



THINKING

- 优点

1. 通过每个Worker对dataset的备份，降低了在分配资源时的数据拷贝开销
2. 通过给jobs增加闲置资源，加快训练速度

- 缺点

1. Master结点需要实时获得Server和Worker结点的状态，需要较高的时延要求