A Performance Evaluation of Federated Learning algorithms

Adrian Nilsson, Simon Smith, Gregor Ulm, Emil Gustavsson, Mats Jirstrand

Presented by Simon Smith at DIDL'18

2018-12-10

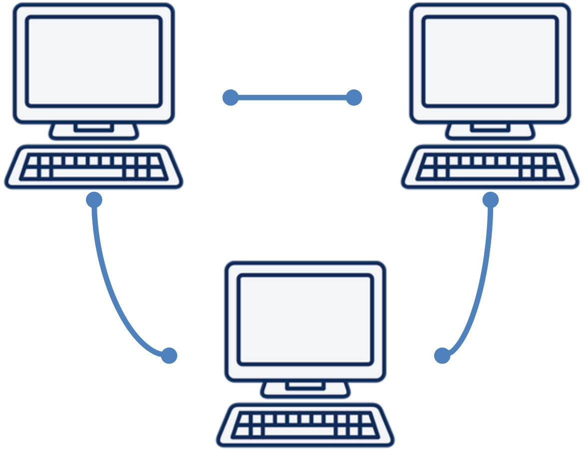
FRAUNHOFER CHALMERS
RESEARCH CENTRE FOR INDUSTRIAL MATHEMATICS

What is Federated Learning (FL)?

- Distributed machine learning
  - Communicate a model, not data.

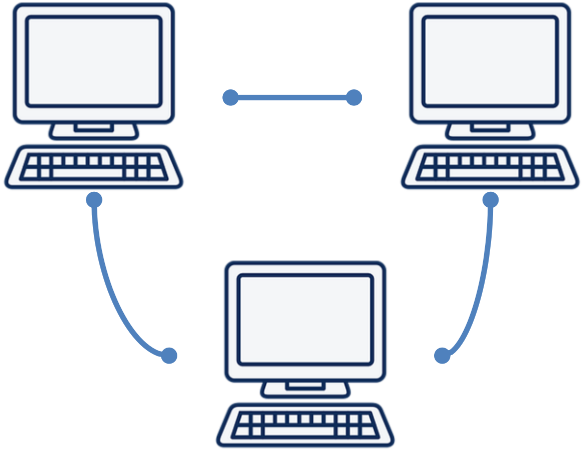
What is Federated Learning (FL)?

- Distributed machine learning
  - Communicate a model, not data.



What is Federated Learning (FL)?

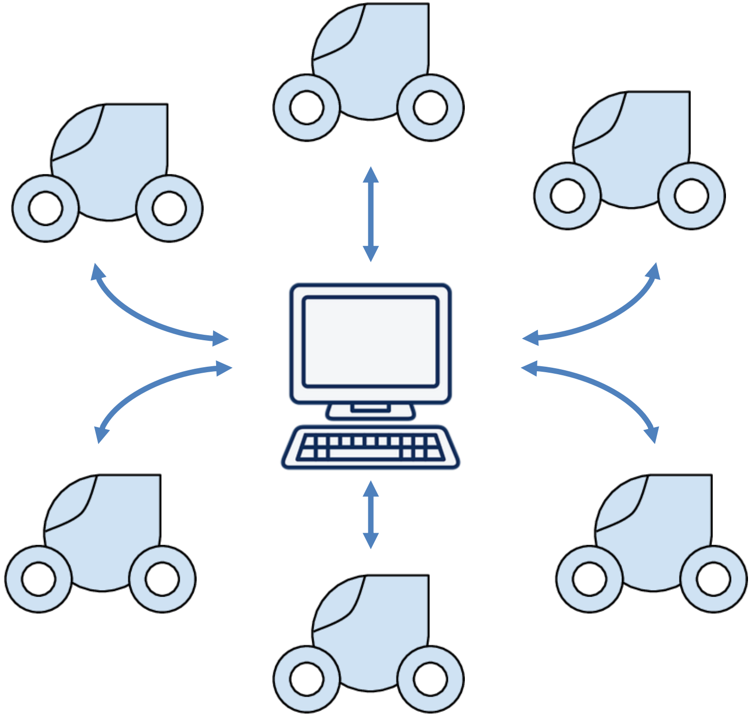
- Distributed machine learning
  - Communicate a model, not data.
- Massive number of clients
  - Slow, unreliable network
  - 250M connected vehicles by 2020[1]



[1] Hitachi, “The internet on wheels and Hitachi, Ltd.”, 2015

What is Federated Learning (FL)?

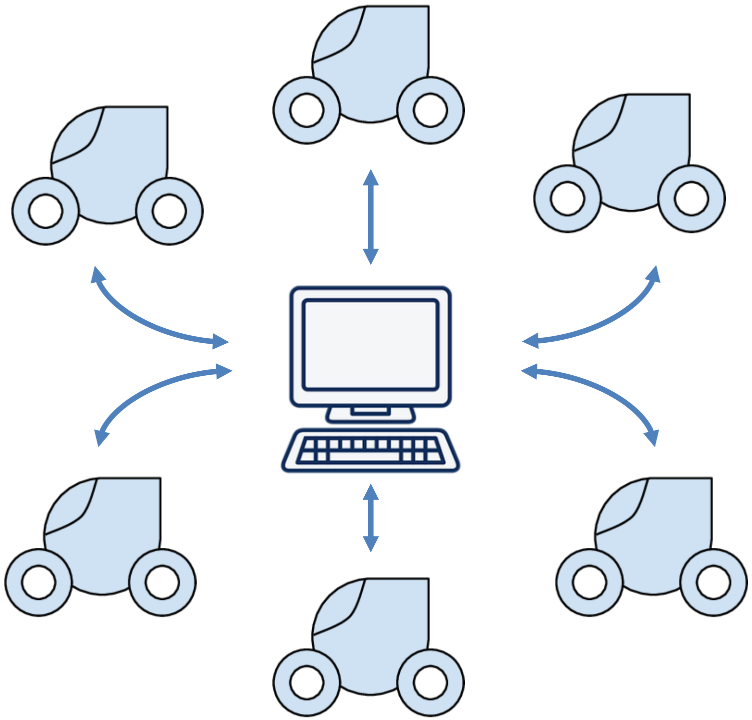
- Distributed machine learning
  - Communicate a model, not data.
- Massive number of clients
  - Slow, unreliable network
  - 250M connected vehicles by 2020[1]



[1] Hitachi, "The internet on wheels and Hitachi, Ltd.", 2015

What is Federated Learning (FL)?

- Distributed machine learning
  - Communicate a model, not data.
- Massive number of clients
  - Slow, unreliable network
  - 250M connected vehicles by 2020[1]
- GDPR



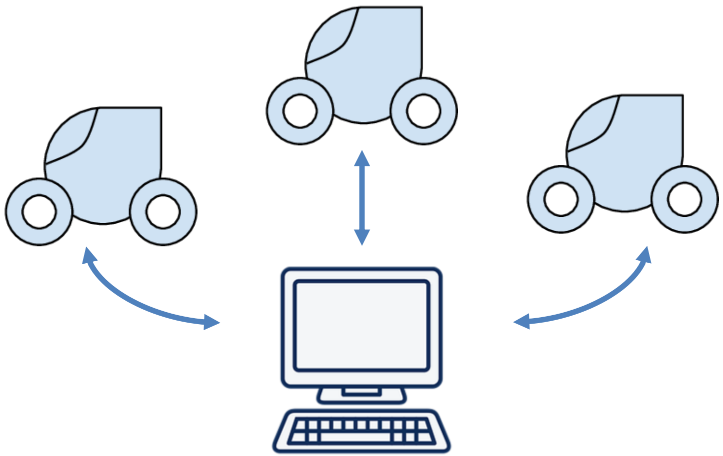
[1] Hitachi, "The internet on wheels and Hitachi, Ltd.", 2015

What have we done!?

- Implemented and compared three FL algorithms
- Compared with fully centralized approach

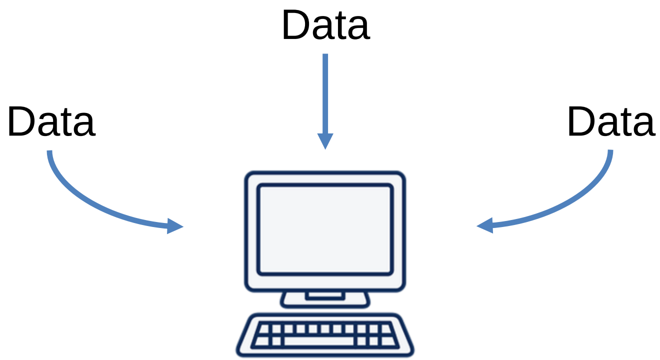
What have we done!?

- Implemented and compared three FL algorithms
- Compared with fully centralized approach



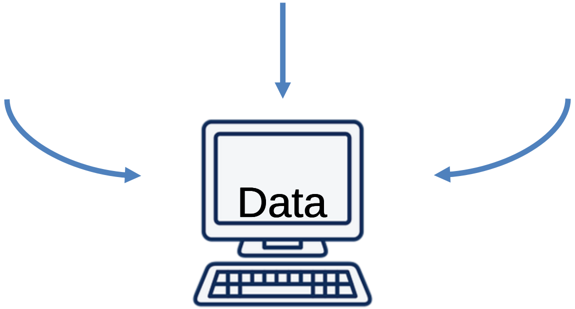
What have we done!?

- Implemented and compared three FL algorithms
- Compared with fully centralized approach



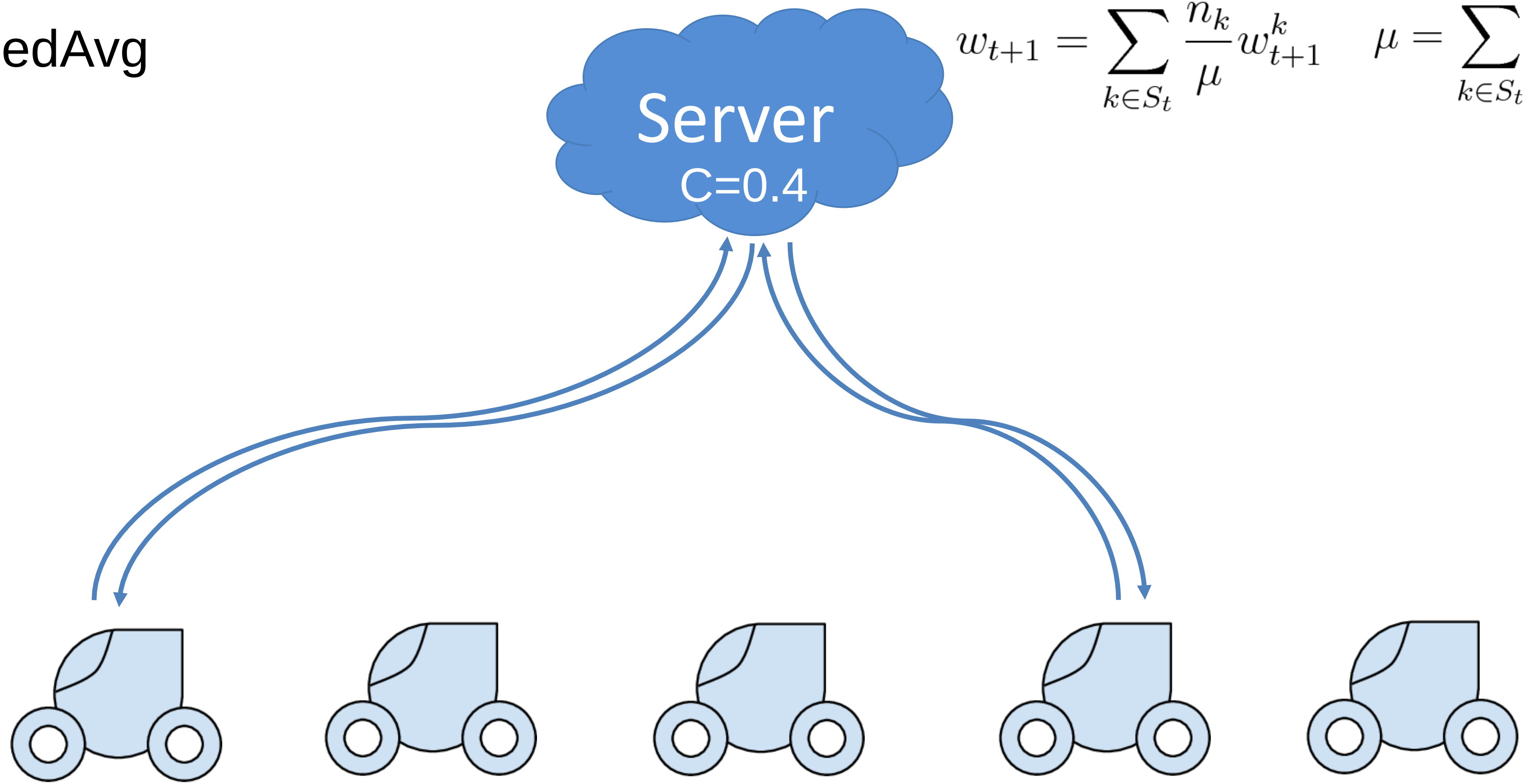
What have we done!?

- Implemented and compared three FL algorithms
- Compared with fully centralized approach



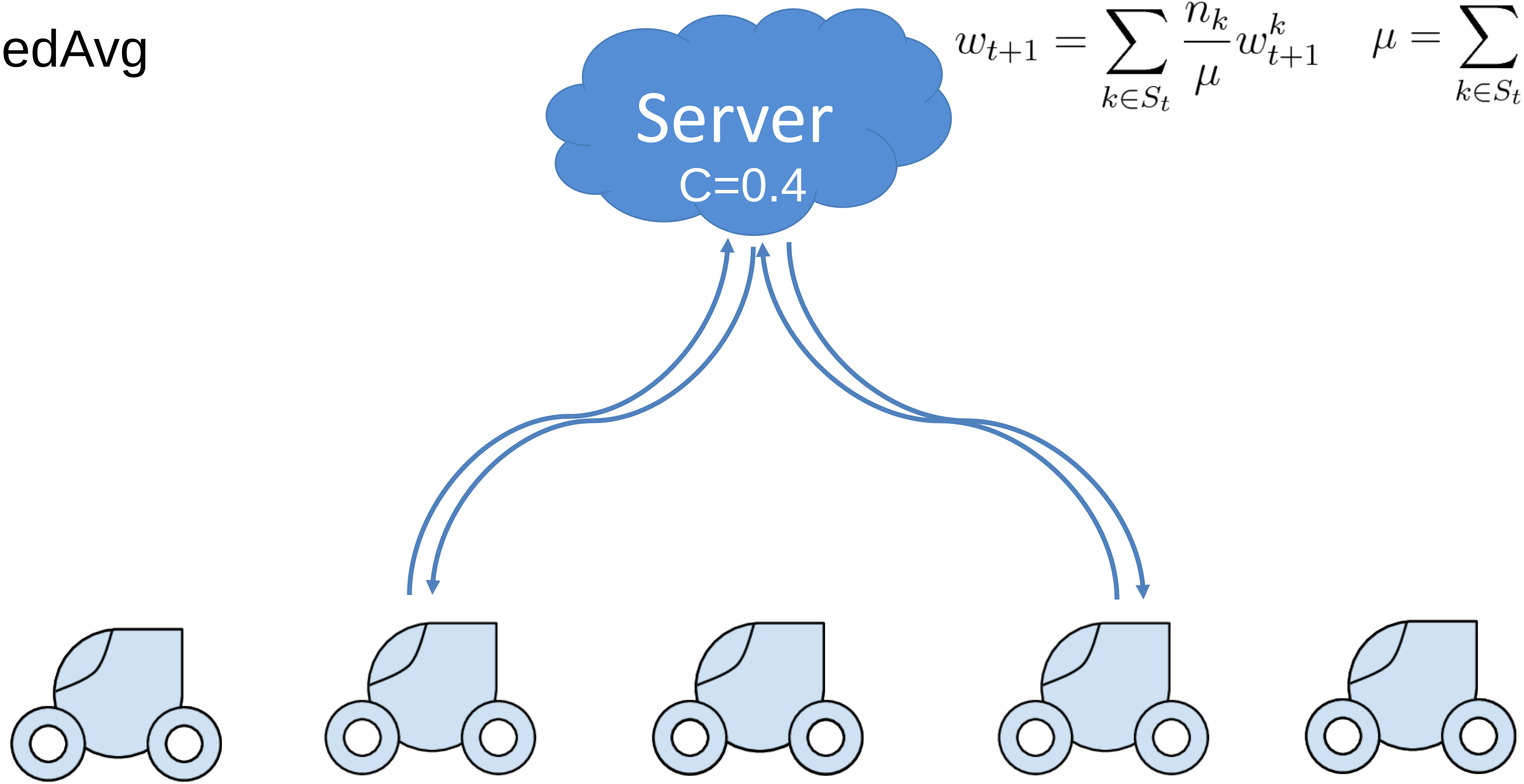
FedAvg

$$w_{t+1} = \sum_{k \in S_t} \frac{n_k}{\mu} w_{t+1}^k \qquad \mu = \sum_{k \in S_t} n_k$$

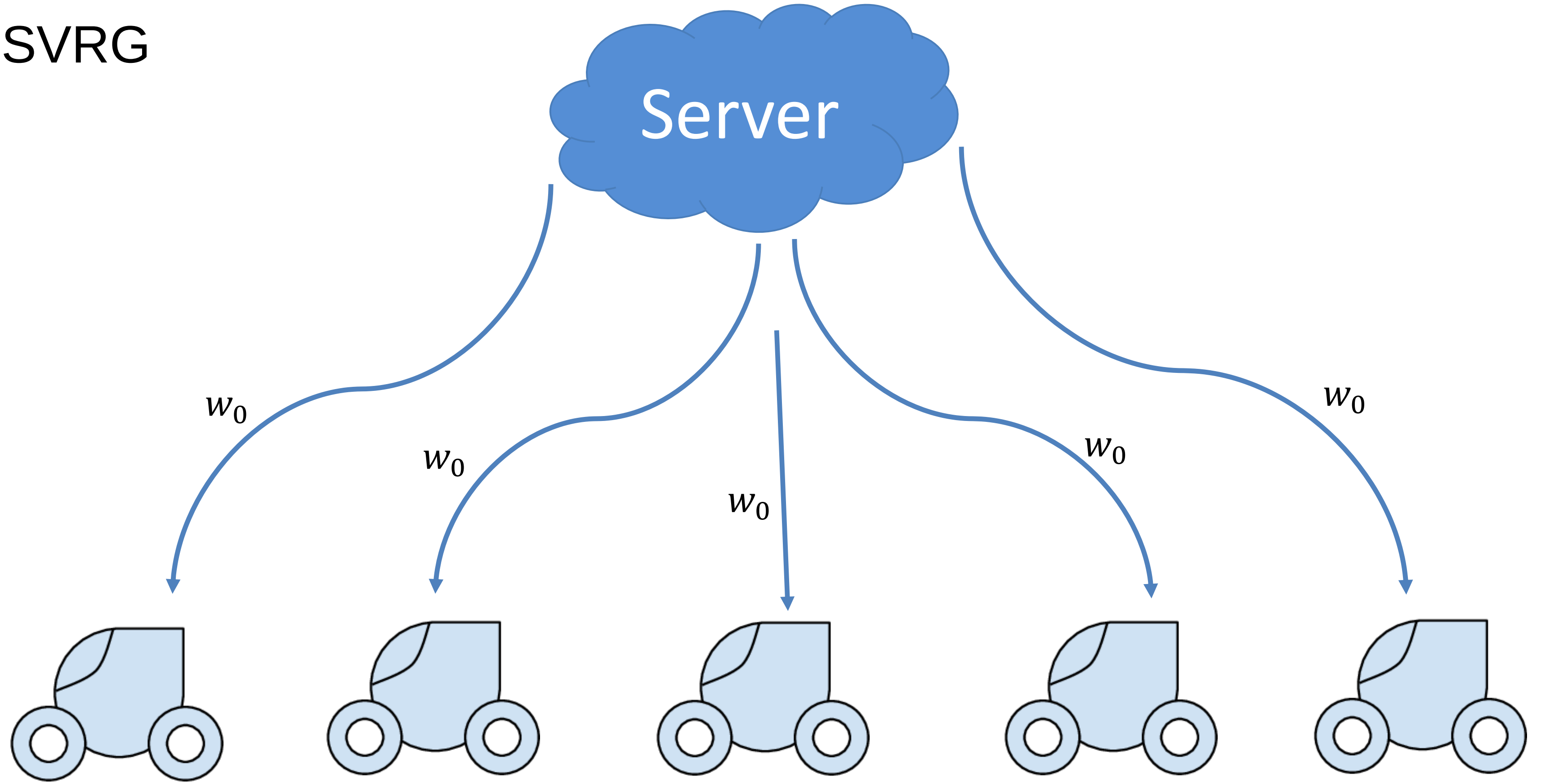


FedAvg

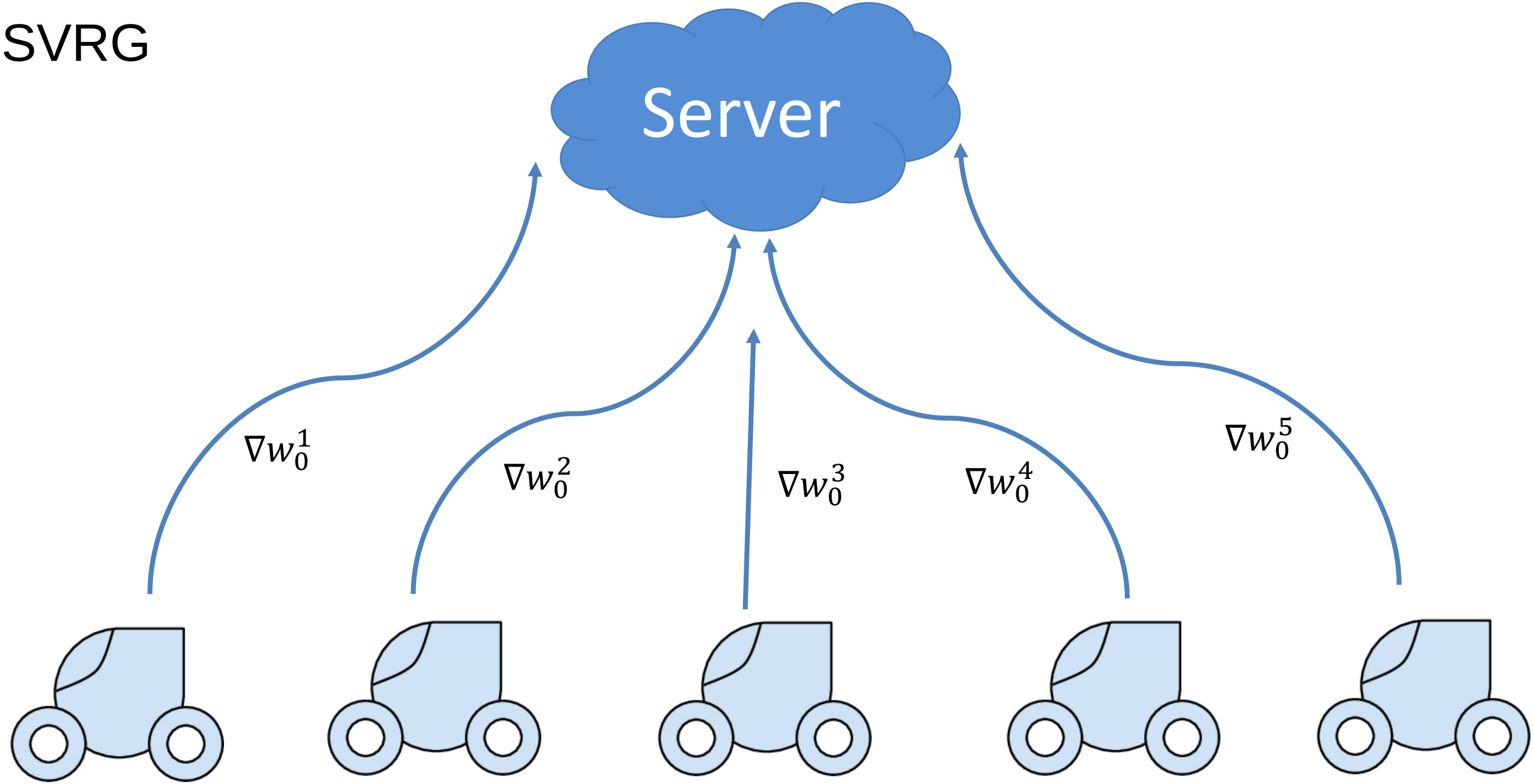
$$w_{t+1} = \sum_{k \in S_t} \frac{n_k}{\mu} w_{t+1}^k \qquad \mu = \sum_{k \in S_t} n_k$$



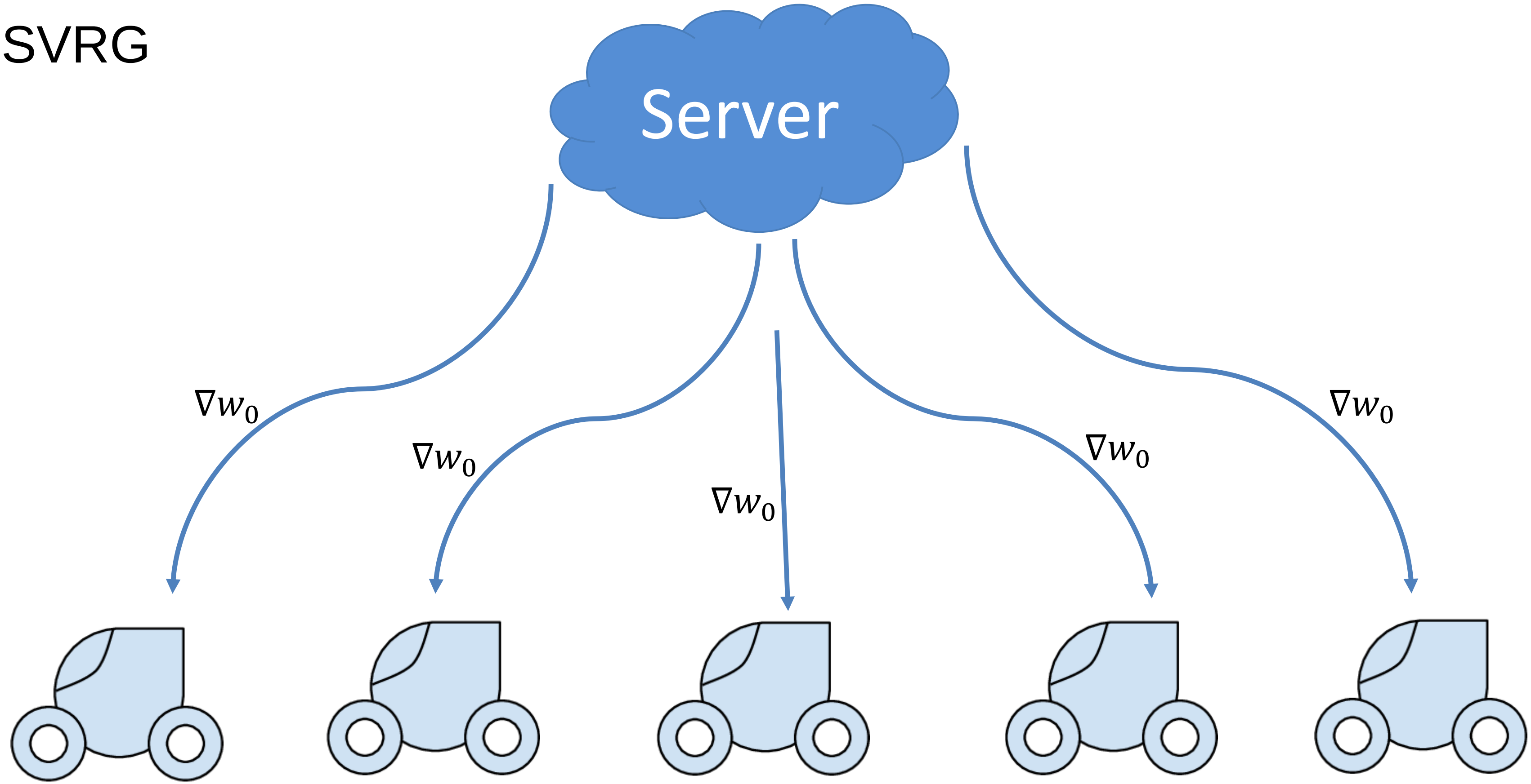
FSVRG



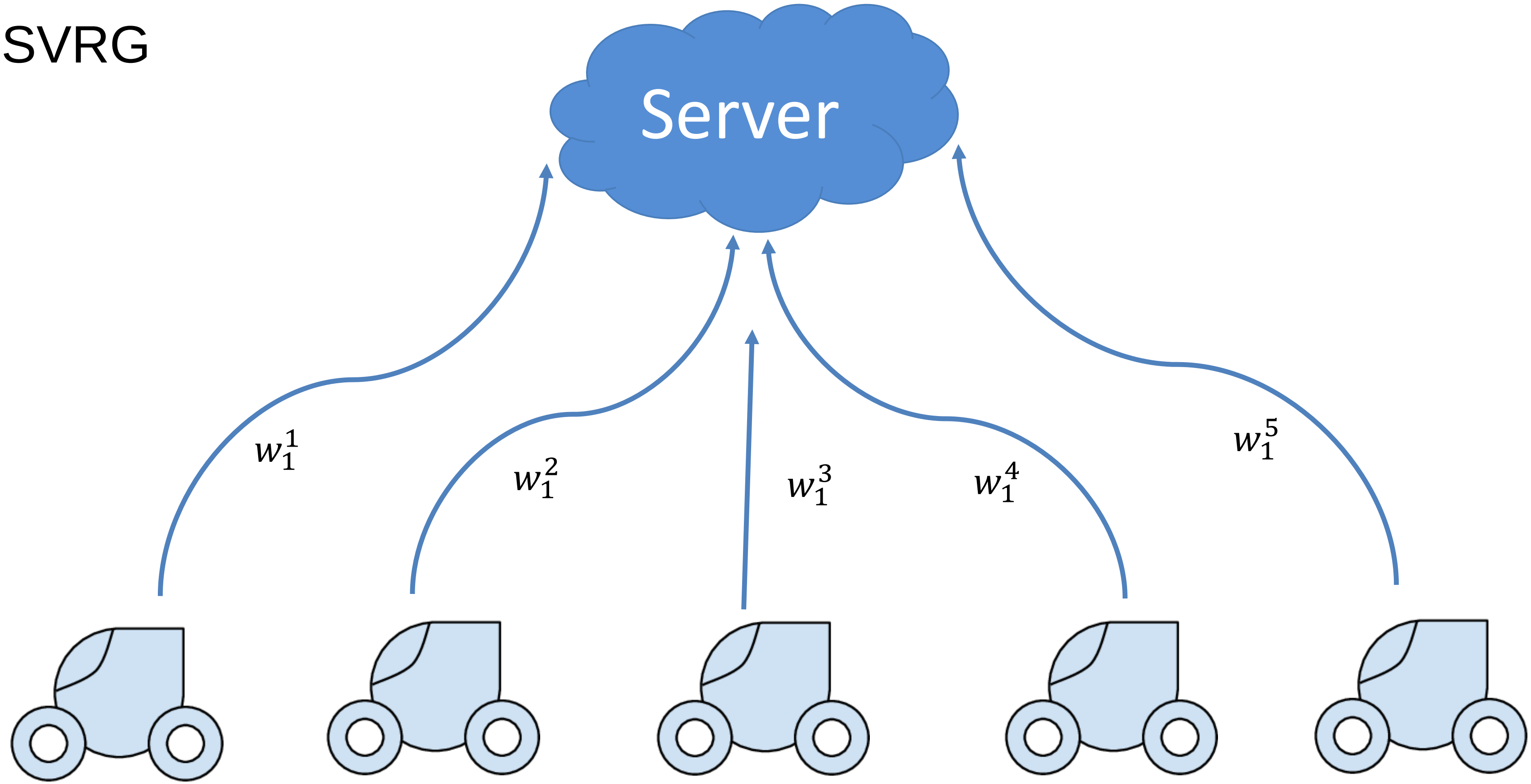
FSVRG



FSVRG

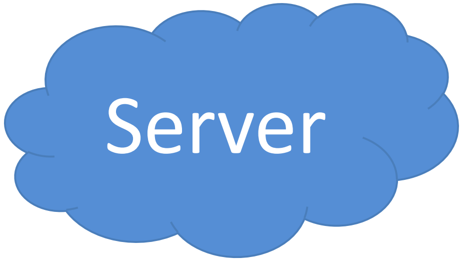


FSVRG

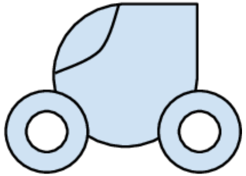


CO-OP

Age filters: b_l, b_u



$$b_l \leq a - a_k \leq b_u$$



Server age: a

Client age: a_k

Algorithms

	Synchronous	Opt. Algorithm	New hyperparam.	Note
FedAvg	✓	SGD	C, E	C – fraction of clients E – epoch before upload
FSVRG	✓	SVRG	h	
CO-OP		SGD	b_l, b_u	"Age filters" - Mitigate staleness

Evaluation approach

MNIST digit recognition

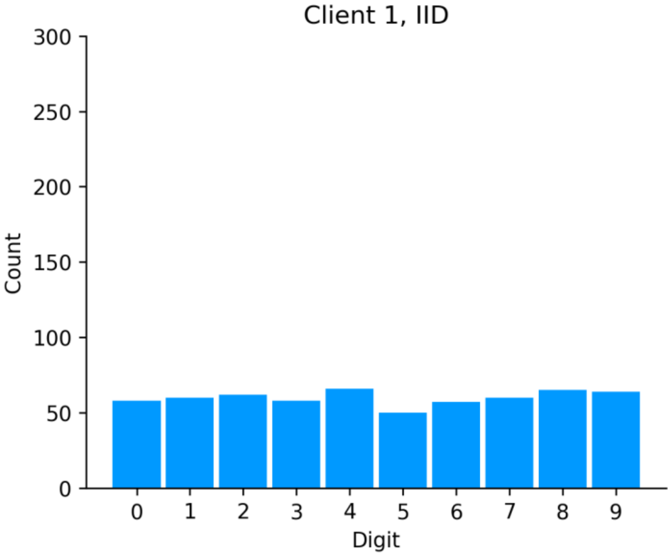
Feed-forward ANN with 2 hidden layers

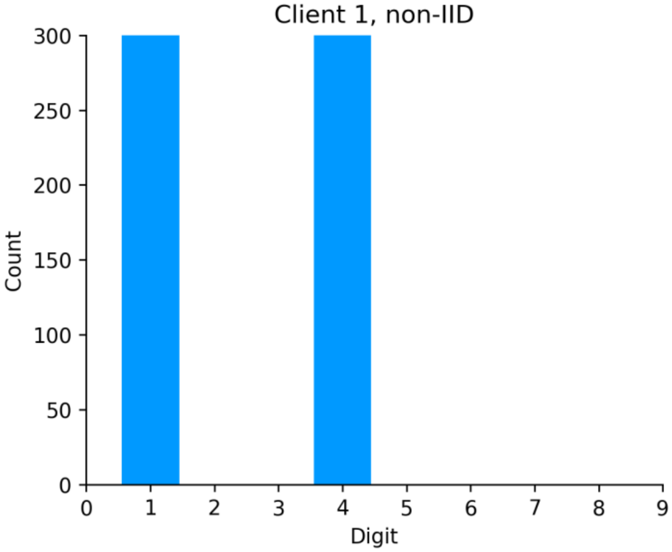
100 clients

Image by Josef Steppan [CC BY-SA 4.0], from Wikimedia Commons

Evaluation approach

IID & non-IID partitionings



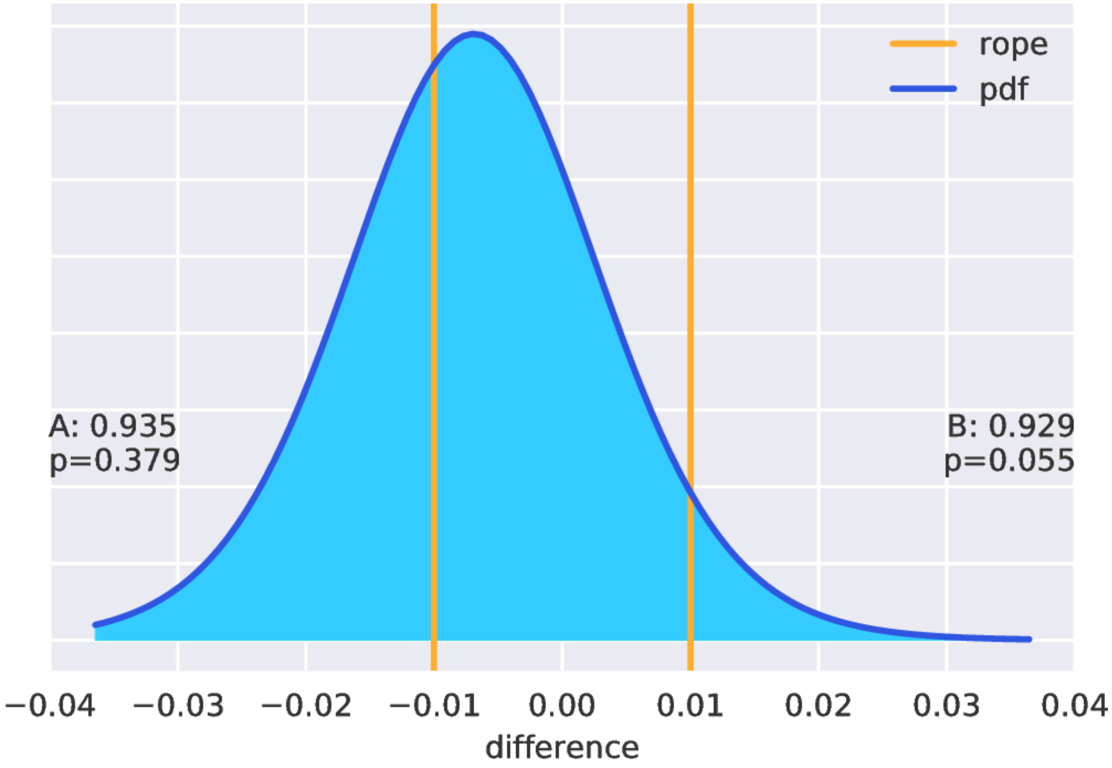


Evaluation approach

- Hyperparameter search
  - Learning rate, decay, epochs, batch size, global step size, age filter
- Cross-validation
  - We allow 10,000 uploads from 100 simulated clients

Bayesian comparisons

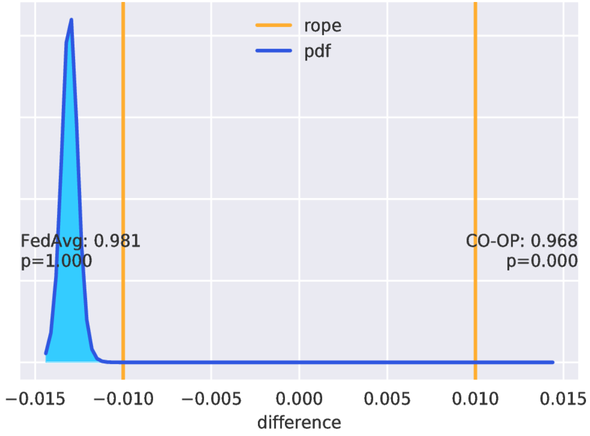
- x-axis shows *mean difference in accuracy* between **A** and **B**
- Region of practical equivalence (**rope**)
- Area is interpreted as a probability.



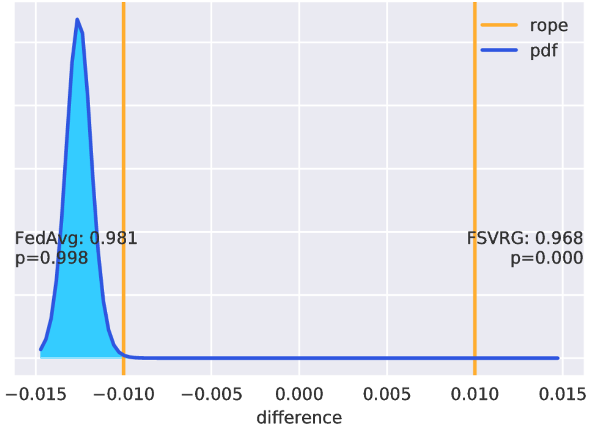
Benchmarking on IID data

Adrian Nilsson & Simon Smith

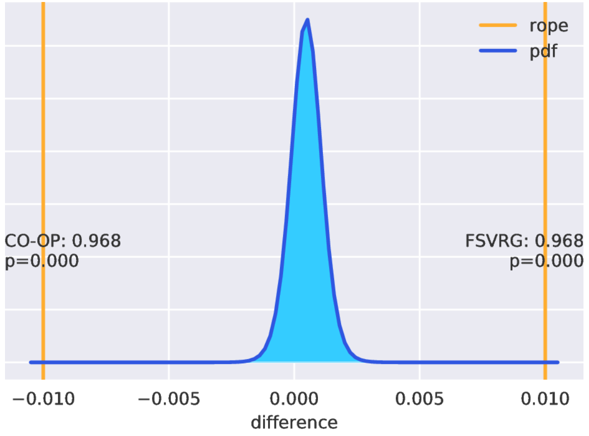
Results – Federated Learning IID



FedAvg vs CO-OP

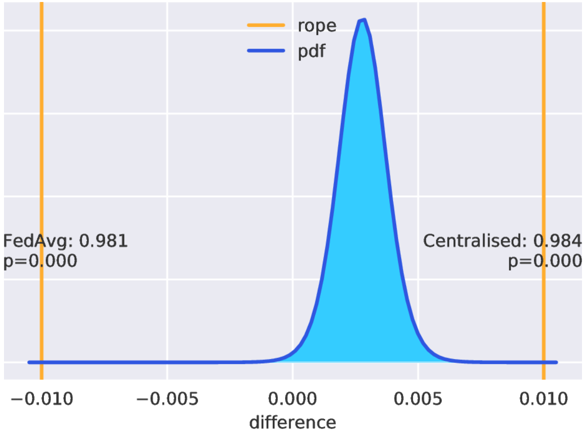


FedAvg vs FSVRG

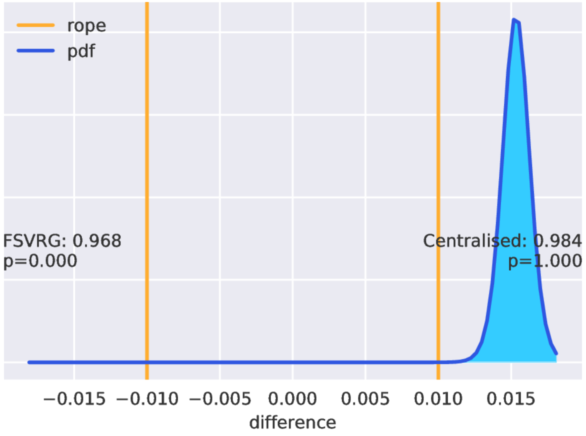


CO-OP vs FSVRG

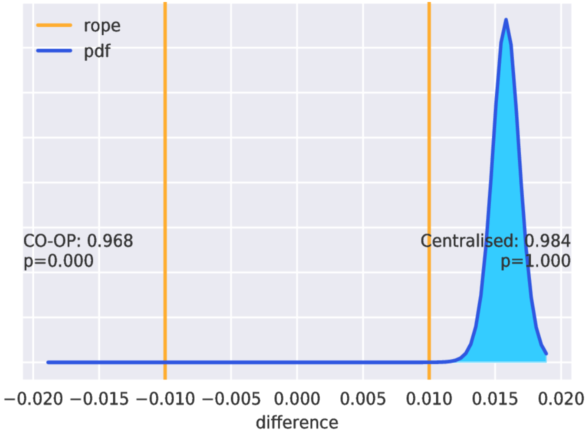
Results – Federated Learning IID vs Centralized Learning



FedAvg



FSVRG

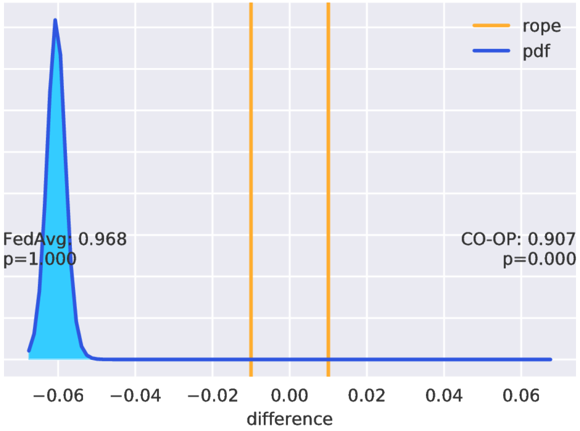


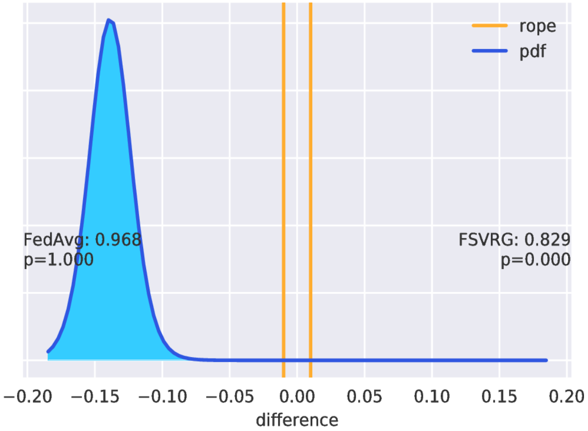
CO-OP

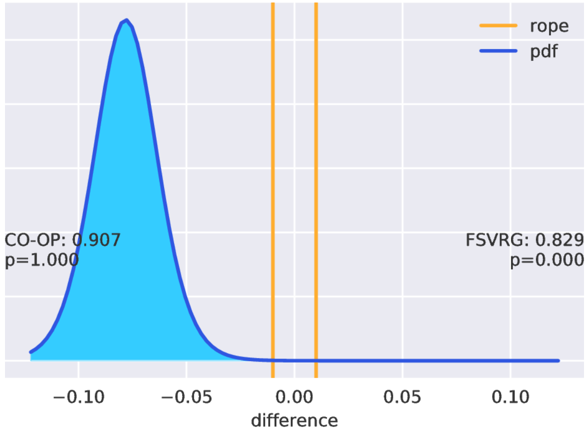
Benchmarking on non-IID data

Adrian Nilsson & Simon Smith

Results – FL non-IID





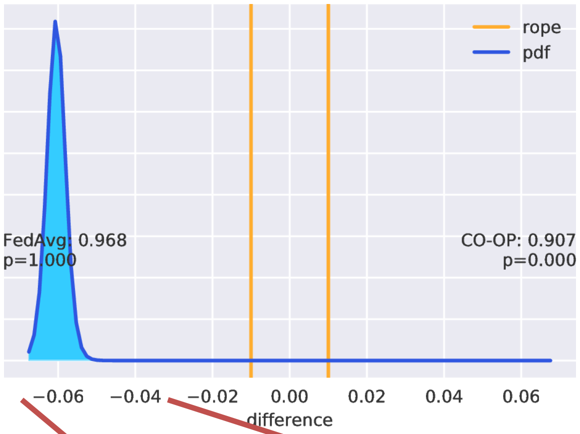


FedAvg vs CO-OP

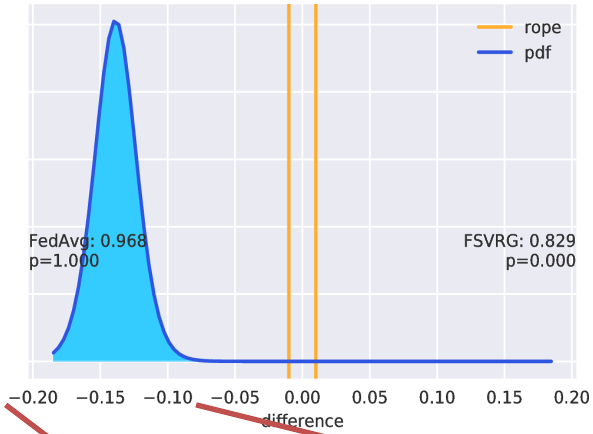
FedAvg vs FSVRG

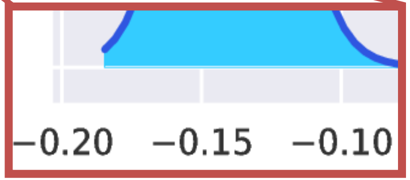
CO-OP vs FSVRG

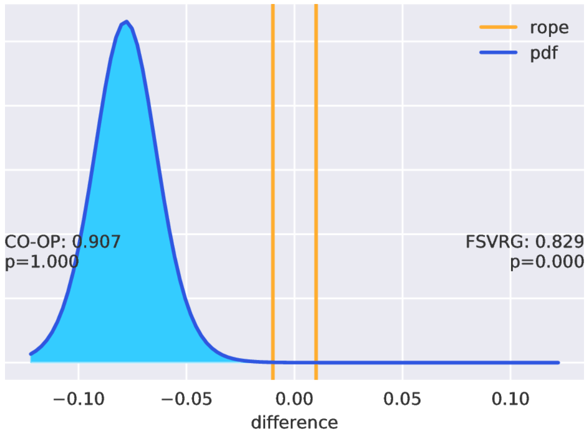
Results – FL non-IID



OP

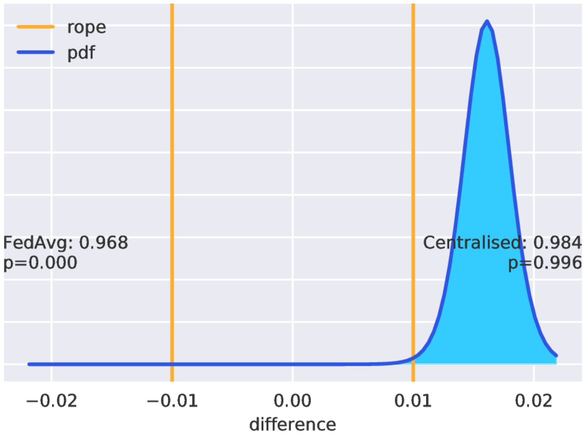




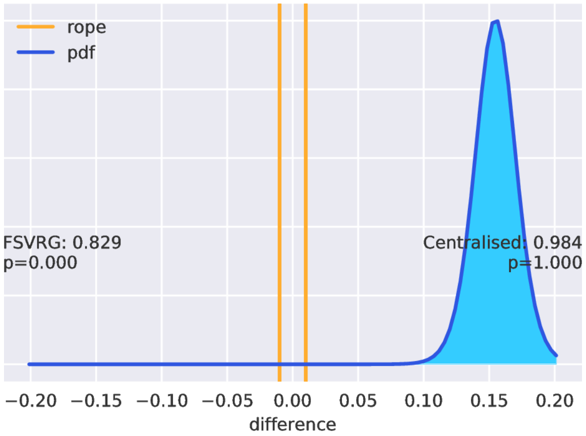


CO-OP vs FSVRG

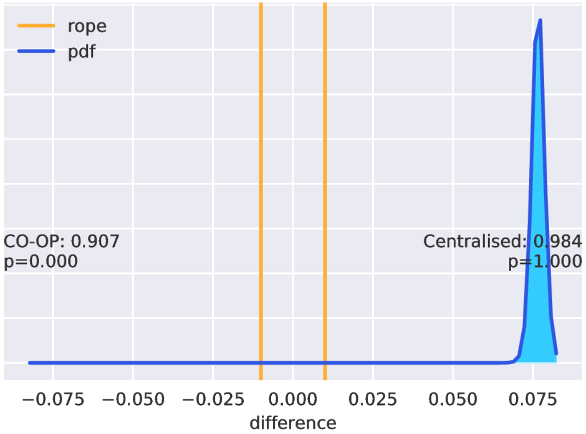
Results – FL non-IID vs Centralized Learning



FedAvg



FSVRG



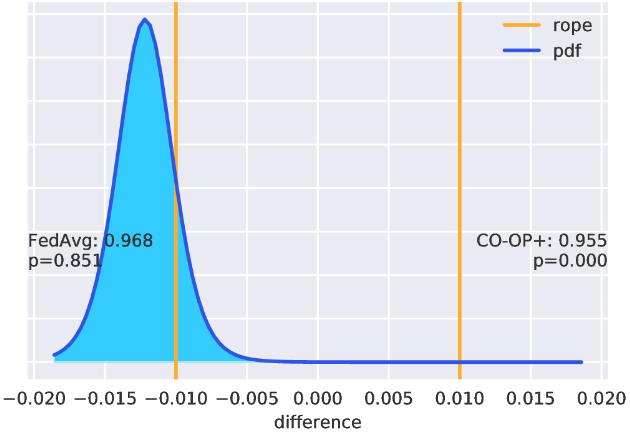
CO-OP

Allowing more communication for FSVRG and CO-OP

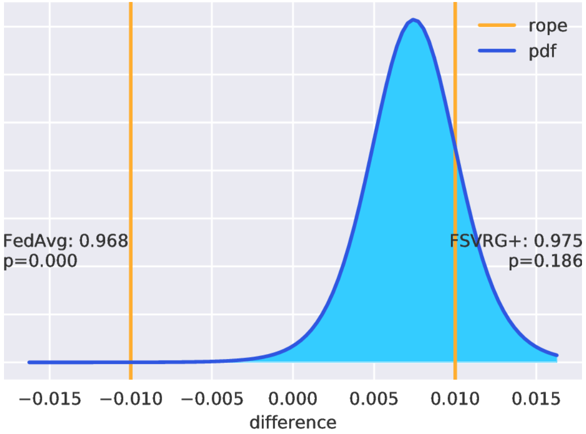
Adrian Nilsson & Simon Smith

FRAUNHOFER CHALMERS
RESEARCH CENTRE FOR INDUSTRIAL MATHEMATICS

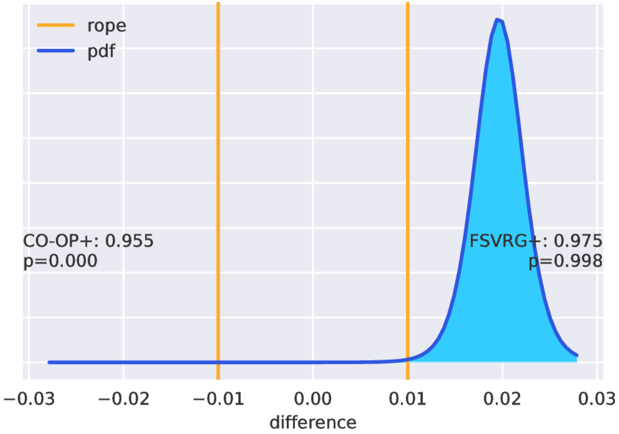
Results – FL non-IID; more uploads



FedAvg vs CO-OP$^{(\times 5)}$

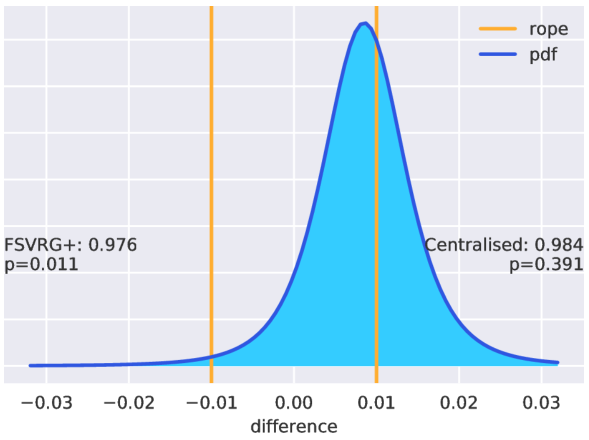


FedAvg vs FSVRG $^{(\times 10)}$

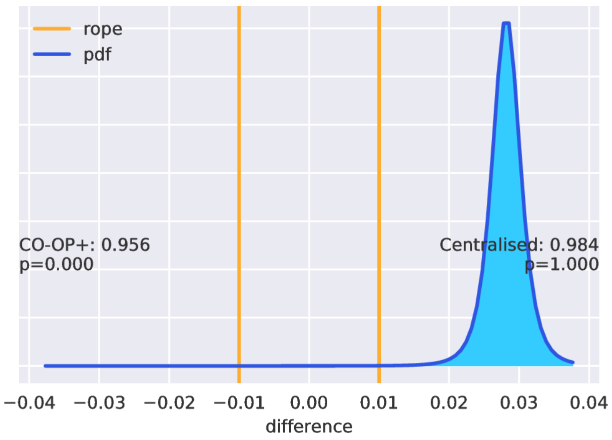


CO-OP $^{(\times 5)}$ vs FSVRG $^{(\times 10)}$

Results – FL non-IID; more uploads vs Centralized Learning



FSVRG $^{(\times 10)}$



CO-OP $^{(\times 5)}$

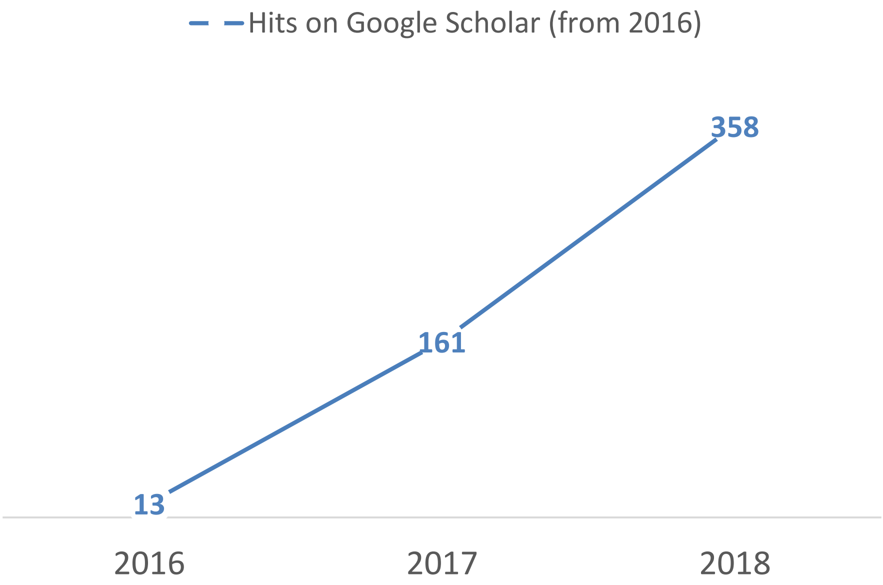
Practical considerations

- FSVRG
  - Requires more communication per global update
- CO-OP
  - Age filters are difficult to tune

Future work

- Evaluate on multiple datasets
- Examine unevenly distributed data
  - i.e. a few cars hold most of the data
- New algorithms

Algorithm + ”Federated Learning”



What has happened since?

- Non-IID is still an issue [1,2]
- More privacy [3,4]
- New Algorithms
  - Asynchronous FL [5]
  - Dynamic Averaging Protocol [6]
  - Federated Kernelized Multi-task Learning [7]

Thank you

References

(Non-IID)
[1] **Federated Learning with Non-IID Data,** Zhao et al., 2018 (arXiv:1806.00582)
[2] **Communication-Efficient On-Device Machine Learning: Federated Distillation and Augmentation under Non-IID Private Data**, Jeong et al., 2018 (arXiv:1811.11479)

(Privacy)
[3] **Biscotti: A Ledger for Private and Secure Peer-to-Peer Machine Learning,** Shayan et al., 2018 (arXiv:1811.09904)
[4] **cpSGD: Communication-efficient and differentially-private distributed SGD,** Agarwal et al., 2018 (arXiv:1805.10559)

(Algorithms)
[5] **Asynchronous Federated Learning for Geospatial Applications,** Sprague et al., DMLE'18, *https://dmle.iais.fraunhofer.de/papers/sprague2018asynchronous.pdf*
[6] **Efficient Decentralized Deep Learning by Dynamic Model Averaging,** Kamp et al., 2018 (arXiv:1807.03210)
[7] **Federated Kernelized Multi-Task Learning,** Caldas et al., Poster at SysML 2018, *https://www.sysml.cc/doc/30.pdf*