Leveraging Renewable Excess Energy in Federated Learning

Philipp Wiesner, Ramin Khalili, Dennis Grinwald, Pratik Agrawal, Lauritz Thamsen, Odej Kao

wiesner@tu-berlin.de

Low Carbon and Sustainable Computing Seminar on March 2, 2023
Agenda

1. What is excess energy?
2. Federated learning on excess energy: idea, use cases, challenges
3. *FedZero* protocol and client selection
4. Evaluation
5. Conclusion and next steps
Excess Energy

Quarterly wind and solar curtailments by the California ISO, who curtailed more than 27 TWh in 2022, around 7% of their entire solar production.
Idea: Federated Learning on Excess Energy

Federated Learning (FL) is an emerging machine learning technique that enables distributed model training across data silos or edge devices without sharing data.

From a scheduling perspective, we are dealing with a **iterative** execution of **distributed** batch jobs.
Use Cases

For our problem setting, we require FL clients with significant computing capabilities and electricity demand, for example:

- **On premise**
  - Health institutions training common models on confidential patient data
  - Financial institutions building credit score predictors

- **Edge computing**
  - Smart energy grids
  - Smart transportation services
  - Smart water distribution

- **Powerful edge devices**
  - Autonomous vehicles
Challenges

1. **Efficiency**: FedZero is designed with performance and energy efficiency in mind
2. **Common power budgets**: FedZero treats energy as a shared and limited resource during client selection and at runtime
3. **Fairness of participation**: FedZero ensures that all clients participate similarly, even if the availability of excess resources is imbalanced
4. **Robustness against forecasting errors**: FedZero remains functional if excess energy or load forecasts have a high error
5. **Scalability**: FedZero’s comes with a low overhead and runtime complexity
FedZero Protocol

Clients provide the following information:

- number of training samples
- maximum computational capacity (batches/timestep)
- energy efficiency (energy/batch)
- control plane addresses
FedZero Protocol

We require forecasts of

- Expected computational load
- Expected renewable excess energy
FedZero Protocol

1. Register clients
2. Collect forecasts
3. Select participating clients
4. Train locally on excess resources
5. Aggregate updates
Client Selection

Iterative algorithm for reduced optimization problem complexity

- Iteratively try out different maximum round durations
- Given a specific round duration, a mixed integer optimization problem (MIP) tries to select $n$ clients with sufficient computing capacity and energy

Heavy filtering of invalid solutions highly reduces the search space, e.g.

- Remove power domains without sufficient energy
- Remove clients without sufficient capacity and/or energy
- Remove clients that over-participated in the past

**Over-participation** is regulated by blocklisting clients after participation and removing them from the blacklist with a probability that corresponds to their statistical utility
Executing Training Rounds

1. Register clients
2. Collect forecasts
3. Select participating clients
4. Train locally on excess resources
5. Aggregate updates
Experimental Setup

- Evaluation is based on Flower (https://flower.dev) and simulated virtualized energy systems
- 100 clients of three sizes; load based on Alibaba GPU cluster traces
- Two energy scenarios
- Three datasets/models
  - CIFAR-10 (iid and non-iid) on ResNet-18
  - CIFAR-100 (iid and non-iid) on DenseNet-121
  - Shakespeare (non-iid) on a two-layer LSTM

We simulate training in 1-minute timesteps for up to 7 days. In each round we select 10% of all clients, who are supposed to perform between 1 and 5 local epochs.
Runtime and Energy Efficiency

- For **CIFAR-10** and **CIFAR-100**, FedZero
  - reaches the final accuracy of the random baseline around 30% faster
  - uses the same amount of energy
- For the **Shakespeare** dataset, FedZero improved the runtime by
  - factor 4 for the global scenario at 33.2% less energy
  - factor 3 for the co-located scenario at 46.6% less energy
Fairness Under Imbalanced Conditions

Global scenario

- Constrained
  - Participation per domain (%)
  - std=2.32

- FedZero
  - Participation per domain (%)
  - std=0.74

Power Domains
Fairness Under Imbalanced Conditions

```
Constrained participation per domain (%)  Global scenario  Global scenario (imbalanced)

<table>
<thead>
<tr>
<th></th>
<th>Power Domains</th>
<th>Power Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std.</td>
<td>std=2.32</td>
<td>std=3.44</td>
</tr>
<tr>
<td>FedZero participation per domain (%)</td>
<td>std=0.74</td>
<td>std=0.95</td>
</tr>
</tbody>
</table>
```

15
Robustness Against Forecasting Errors

![Accuracy vs. Training time](chart1)

- FedZero w/o error
- FedZero w/ error
- FedZero w/ error but no load forecast
- Unconstrained

![Percentage of rounds](chart2)

- Target accuracy
- Round durations (min)
Overhead

1440 timesteps correspond to 1 day in minutely resolution
Conclusion

Summary

- FedZero is a system design for fast, fair, and efficient training of FL models using only renewable excess energy and spare computational capacity
- It is robust against forecasting errors and highly scalable

Future work

- Integrate FedZero into existing client selection strategies
- Explicitly take energy storage and grid energy consumption into account
- Better understand the impact of periodic patterns in excess energy availability on training performance

Contact: Philipp Wiesner, wiesner@tu-berlin.de