Correct Timings and Inspection of States for Federated Learning Simulations

Marko Putnik - m.putnik@student.tudelft.nl
Supervisor: Bart Cox
Responsible Professor: Dr. Jérémie Decouchant

TU Delft

1. Background

- **Federated Learning** is a machine learning approach where distributed devices collectively train a central global model [1]
- Popular FL simulators, such as Flower[2], are continuous simulators without a concept of simulation time
- This can incur problems, especially in asynchronous scenarios, such as global system inaccuracies and non-reproducibility of the simulated runs

2. Methodology

- **RQ**: How to ensure correct timings for a simulated FL system?
  - Introduce a **discrete-event simulator (DES)**
  - **Clock**: Moves from event time to event time
  - **Events List**:
    - INITIALIZE
    - START CLIENT TRAINING
    - END CLIENT TRAINING
    - START TRANSFER
    - END TRANSFER
    - SERVER AGGREGATE
  - **Random Variables**:
    - Latency - X
    - Throughput - Y
    - Computation - Z
  - **DAG**:

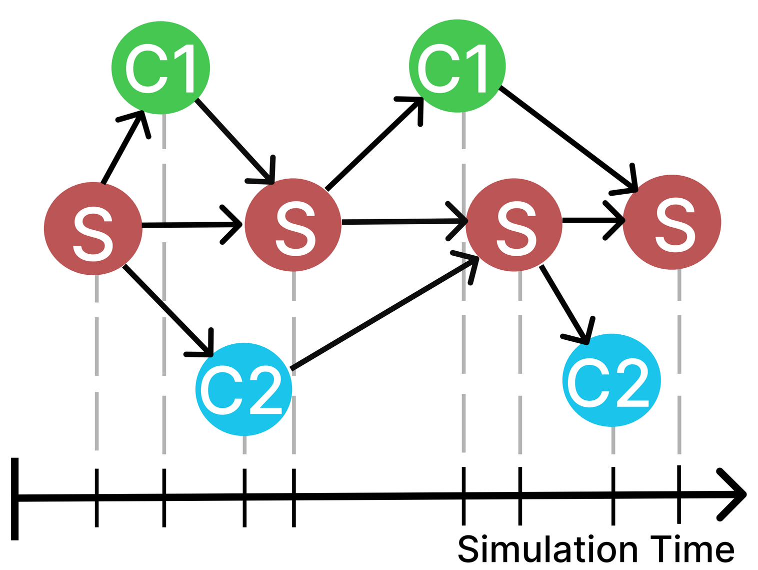

Figure 2: Resulting DAG as Output of the DES

- **Protocol:**
  - **Intialize Clients:**
    - Append **TRANSFER EVENT** to Each Client
  - **Event-Cycle Loop:**

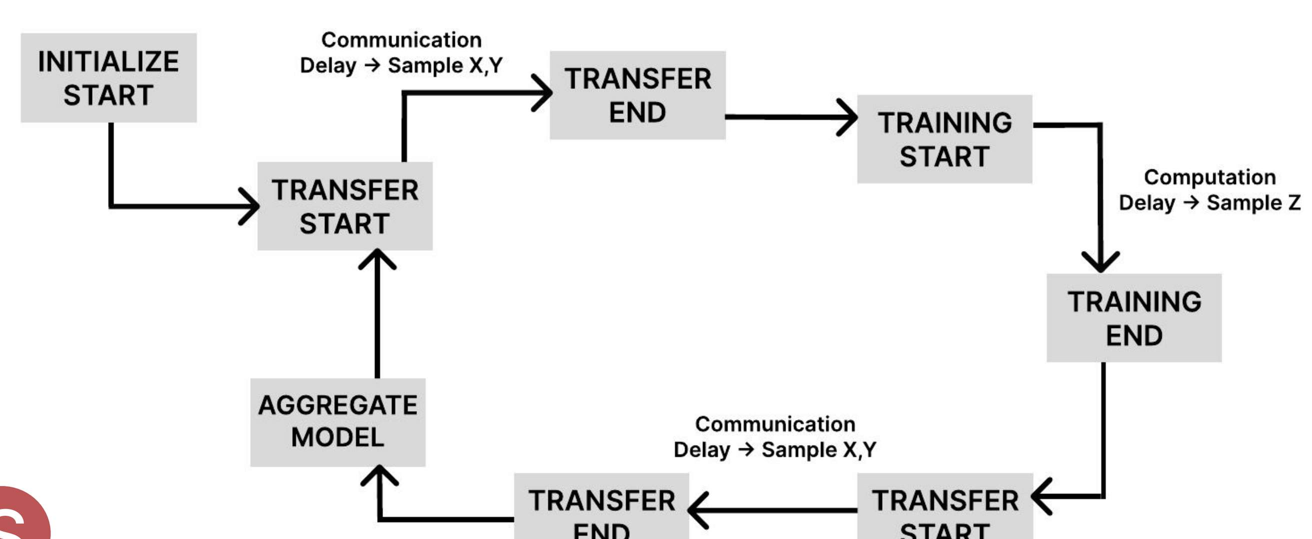

Figure 1: Simulation Event-Cycle for a Single Client-Server Pair

- **RQ**: How can we inspect the results of the system for any point in time?
  - Introduce a **log-structure**
  - Log-structure gathers simulation traces based on the global simulation time introduced by the **discrete-event simulator**

References

[1] H. B. McMahan, E. Moore, D. Ramage, and B. A.y Arcas, "Federated learning of deep networks using model averaging," arXiv preprint arXiv:1602.05629, 2016.
[2] D. J. Beutel, T. Topal, A. Mathur, X. Qiu, T. Parcollet, and N. D. Lane, "Flower: A friendly federated learning research framework," arXiv preprint arXiv:2007.14390, 2020.

3. Experiments

- Datasets: MNIST and CIFAR-10
- Number of Clients: 10
- Sampling:
  - Each client gets 2 labels for 80% of the data and 20% for the remaining labels

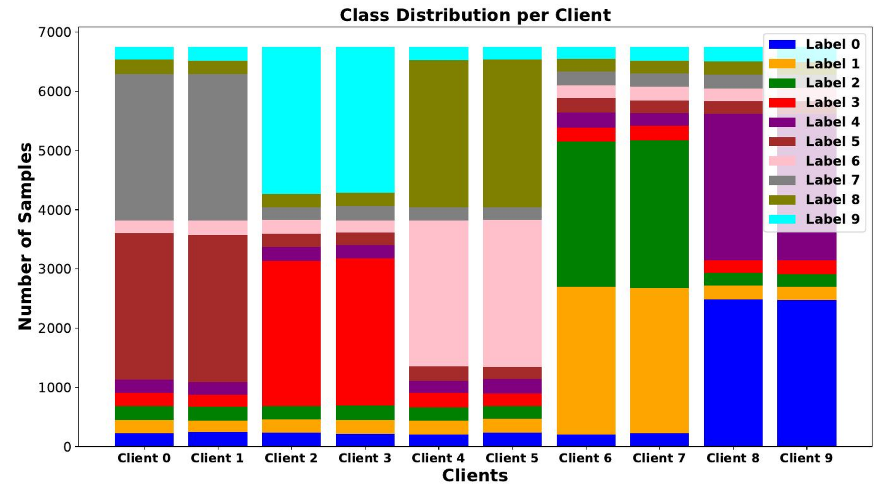

Figure 3: non-IID Sampling Strategy

- **Experiment 1 - Heterogeneous Clients**:
  - 'FASTER' VS 'SLOWER' Clients
  - Batch size of 10 vs 300
  - **Metrics:**
    - Global Model Accuracy
    - Standard Deviation Between Reruns
- **Hypothesis:**
  - Faster clients would bias the global model with their data distribution
  - With DES event ordering the bias would be reduced leading to improved accuracy

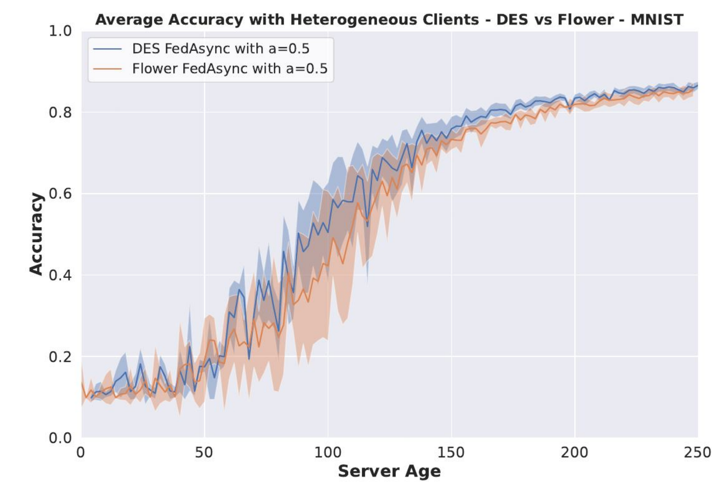

Figure 4: Average Accuracy - DES vs Flower on MNIST with Heterogeneous Clients

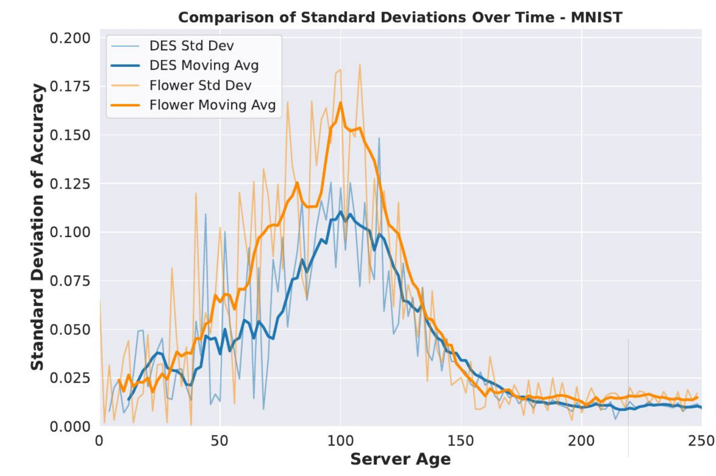

Figure 5: Standard Deviation Between Reruns - DES vs Flower on MNIST with Heterogeneous Clients

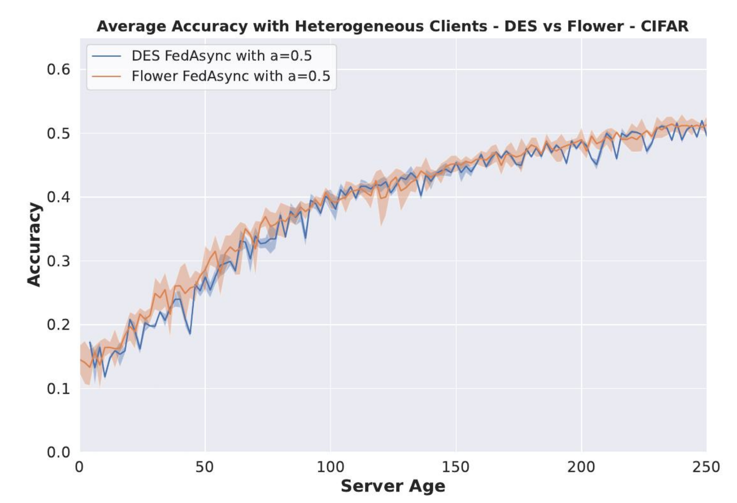

Figure 6: Average Accuracy - DES vs Flower on CIFAR-10 with Heterogeneous Clients

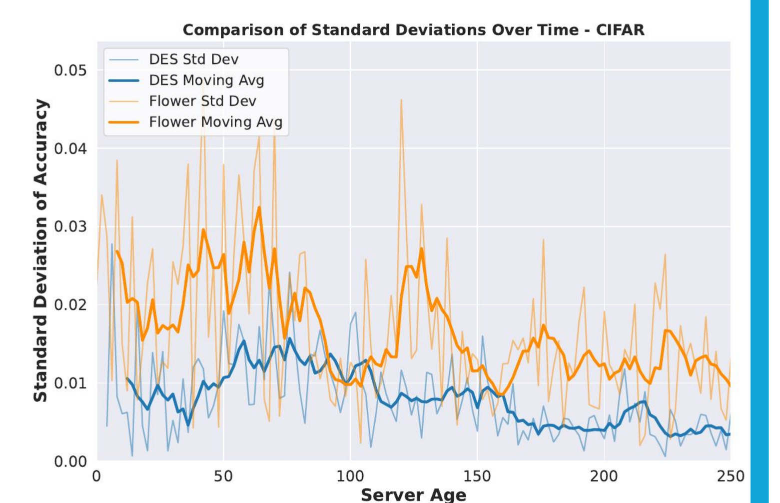

Figure 7: Standard Deviation Between Reruns - DES vs Flower on CIFAR-10 wtih Heterogeneous Clients

- **Experiment 2 - Homogeneous Clients**:
  - Same batch size
  - Effect of OS non-determinism: training process, OS scheduling, memory layouts of processes
  - **Metric:**
    - Standard Deviation Between Reruns
- **Hypothesis:**
  - Because of DES event ordering we would see a reduction in variability caused by the internals of the OS under which the simulation is ran

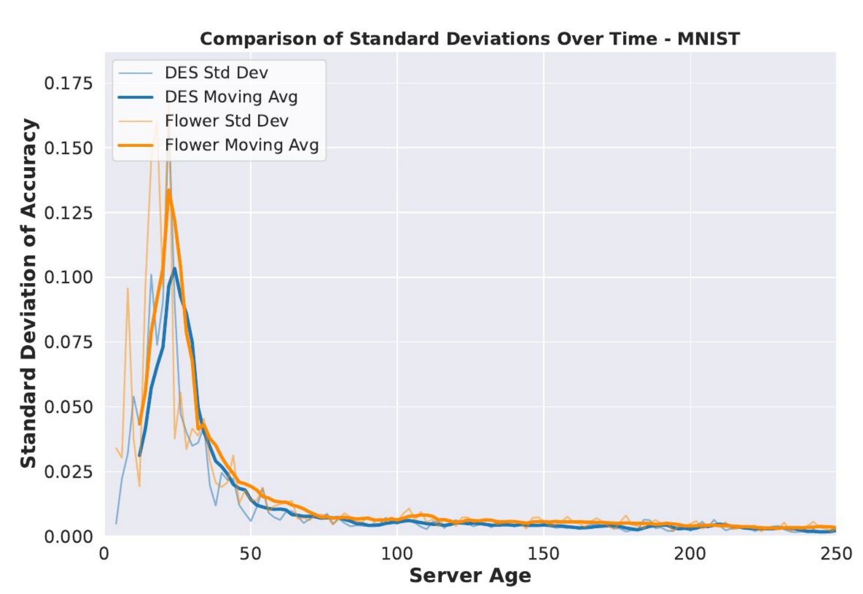

Figure 8: Standard Deviation Between Reruns - DES vs Flower on MNIST with Homogeneous Clients

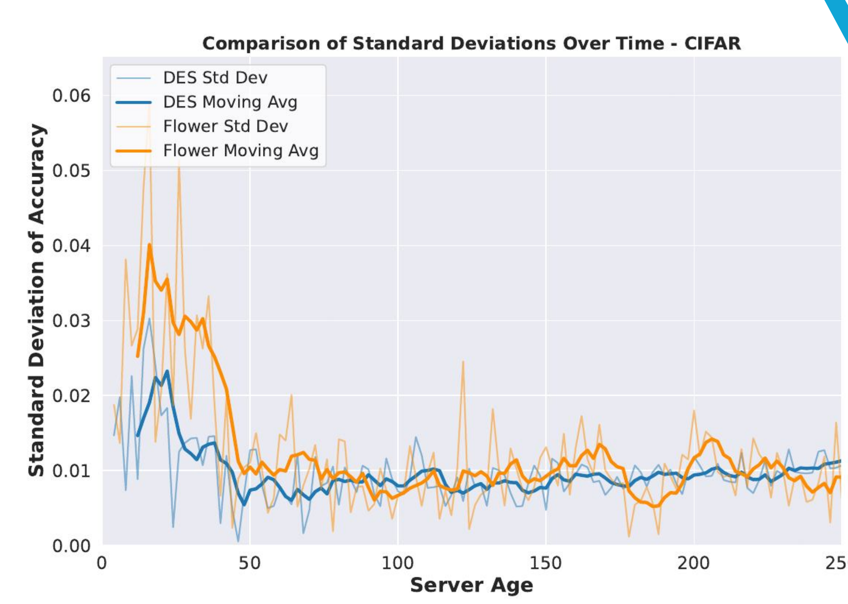

Figure 9: Standard Deviation Between Reruns - DES vs Flower on CIFAR-10 with Homogeneous Clients

4. Results

Experiment 1: Heterogeneous Clients

- **Global Model Accuracy:**
  - In MNIST we see an improvement of 3.3% on average with DES compared to Flower
  - In CIFAR-10 we see worse accuracy: -0.85% on average with DES compared to Flower
  - In CIFAR-10 we do not see better accuracy
    - An issue could be our proposed sampling strategy as it does not properly encapsulate heterogeneity with faster and slower clients
- **Standard Deviation Between Reruns:**
  - In MNIST the mean standard deviation between reruns for DES is 0.039 and for Flower is 0.052
  - In CIFAR-10 the mean standard deviation between reruns for DES is 0.007 and for Flower is 0.016
  - Results align with our expectations due to the event ordering of the DES

Experiment 2: Homogeneous Clients

- **Standard Deviation Between Reruns:**
  - In MNIST the mean standard deviation between reruns for DES is 0.012 and for Flower is 0.015
  - In CIFAR-10 the mean standard deviation between reruns for DES is 0.009 and for Flower is 0.012
  - Results align with our expectations due to the event ordering of the DES

Inspecting Client state

- From the **log-structure** we can inspect client state for any point in time because client training times occur at a predefined time intervals, visualized in Figure 10

5. Conclusions

- **Improved accuracy** for **3.3%** on average in the heterogeneous scenario on MNIST
- **Reduced standard deviation** in all cases for around **31%** on average
- **Inspecting client models** possible based on time due to the predefined event timings because of the DES

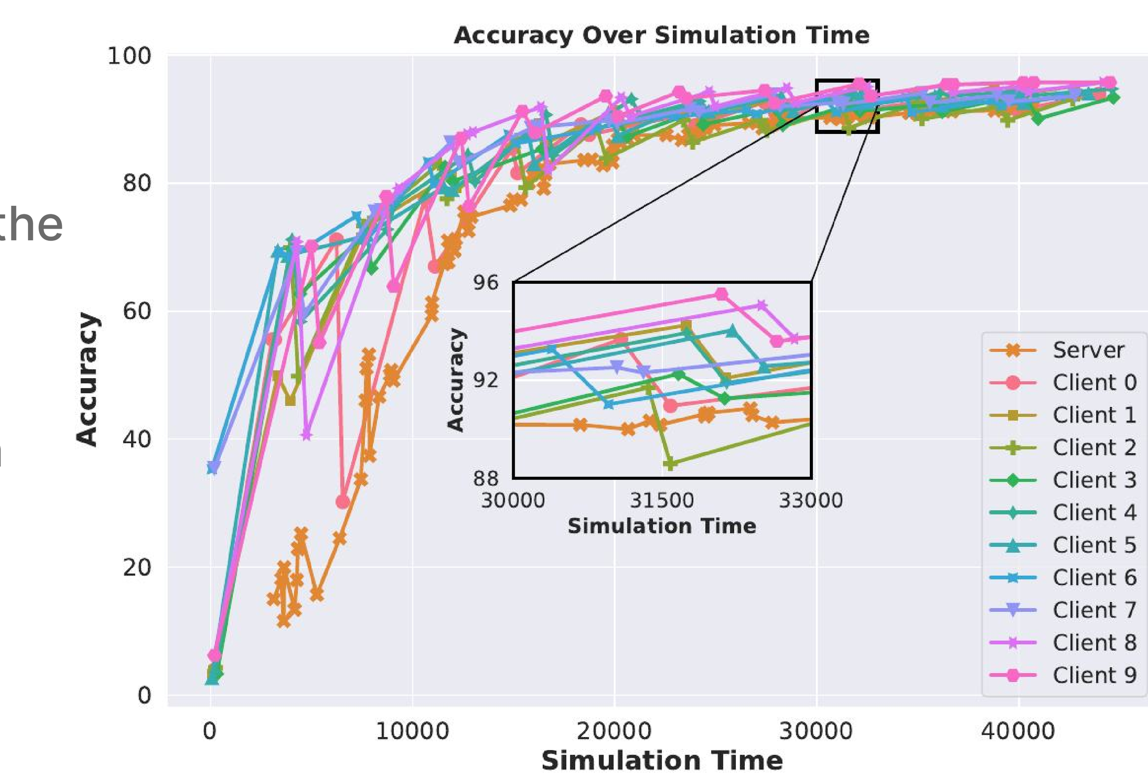

Figure 10: Accuracy of Clients and Server over Global Simulation Time on MNIST

6. Future Work

- Improve **random variables** by gathering real-life FL deployment data
- Propose a **better sampling strategy** taking heterogeneity better into account and test DES against Flower
- Incorporate multi-server scenarios into the simulator as they can introduce new conflicting variables when it comes to timing