Evaluating Federated Learning Scenarios: Impact of Basic Parameters and Realistic Communication in Testbeds

Bruno H. Meyer[1], Aurora Pozo[1], Michele Nogueira[2], Wagner M. Nunan Zola[1]

[1]Departamento de Informática – Universidade Federal Paraná (UFPR)

[2]Depto. de Ciência da Computação – Universidade Federal de Minas Gerais (UFMG)

{bruno, aurora, wagner}@inf.ufpr.br, michele@dcc.ufmg.br

Abstract. *This paper evaluates Federated Learning in different scenarios to identify differences between simulations and hardware. The scenarios include the MENTORED Testbed with x86 nodes, an experimental cluster with aarch64 architectures, and a server simulating FL using Docker.*

1. Introduction

Federated Learning (FL) is a machine learning approach that enables training models in a distributed manner without sharing training data between nodes. FL can be applied in various applications, including edge devices like smartphones and IoT devices, which often have limited computational resources. Many studies have used simulations on a single server, which may not reflect real-world performance on edge devices [Liu et al. 2024, Božič et al. 2024]. Specifically, [Božič et al. 2024] introduced key concepts related to realistic FL experiments and testbeds. However, their work lacks a comprehensive exploration of FL experimentation leveraging computational resources widely distributed across diverse geographic locations, which serves as a primary motivation for the present study. Recent advances in FL introduced improvements and variations of the original approach (FedAvg) [Meyer et al. 2024], which involves training a global model using local updates from clients and aggregating these updates on a central server. Key hyperparameters such as Neural Network (NN) size, number of clients, and data size per client were evaluated in this work. Experiments were conducted using the MENTORED Testbed, a research infrastructure built upon the redeIpê network and the RNP Cluster, featuring computational resources distributed across multiple regions in Brazil. These resources are interconnected through Points of Presence (PoPs) located at various Brazilian universities, enabling geographically diverse experimentation. The RNP Cluster uses Kubernetes with x86 nodes with up to 48 cores and 150GB of RAM. Additionally, an experimental cluster with Raspberry Pi (aarch64 architectures) with 4 cores and 4GB of RAM was integrated with the MENTORED Testbed. Lastly, a unique server was used to simulate the FL topology using Docker technology. This work aims to identify differences between simulations and hardware used in FL experiments.

2. Results

The following results were obtained from experiments conducted in three scenarios, highlighting the impact of different model sizes using a two-layer neural network with varying neurons per layer (NPE), where model size equals NPE^2. Each experiment ran for 15 minutes, measuring F1-Score and communication time. In the Docker simulation (Figure 1), larger models achieved higher long-term F1-Scores at the cost of increased communication time. The IoT scenario (Figure 2) had fewer rounds due to limited computational resources, with communication time becoming a significant bottleneck (nearly 50 seconds per round, about 50%), likely caused by resource constraints and communication overhead on the aarch64 architecture. Lastly, the RNP Cluster scenario (Figure 3) showed highly non-monotonic behavior, with random peaks in communication time (e.g., at 450 seconds), especially for the largest model (NPE=4096), which exhibited stochastic oscillations between 10 and 20 seconds. These insights can guide future improvements in FL, such as dynamically adjusting model size or the number of clients based on network conditions. All methods and results are available at `https://github.com/BrunoMeyer/erad2025-fl-testbed`.

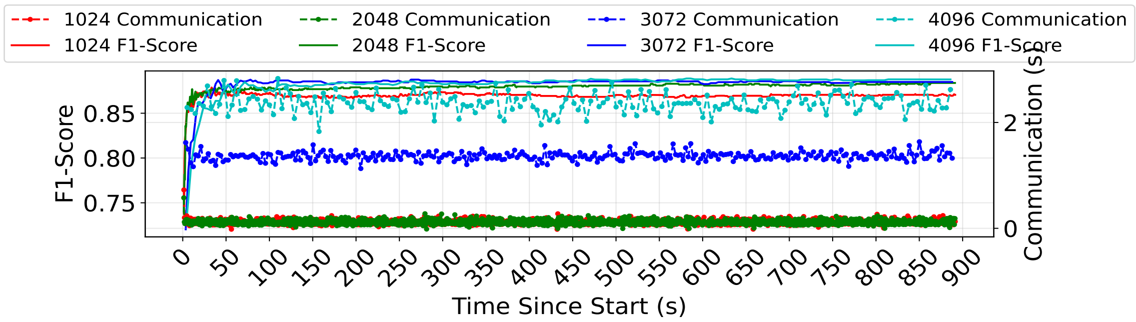

Fig. 1. Simulation. 6 nodes, 128 instances/client and different NN size.

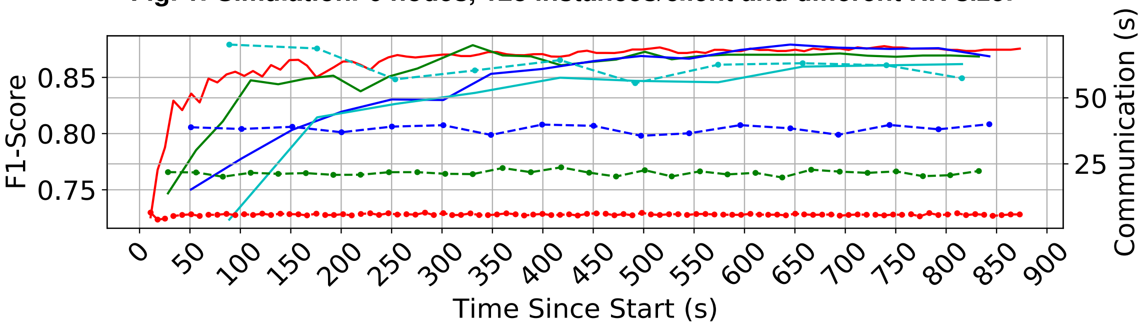

Fig. 2. IoT. 6 nodes, 128 instances/client and different NN size.

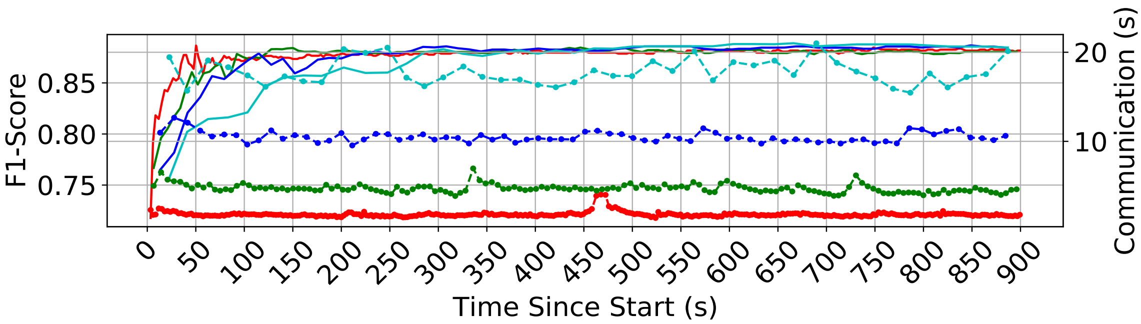

Fig. 3. RNP Cluster. 6 nodes, 128 instances/client and different NN size.

References

Božič, J., Faustino, A. R., Radovič, B., Canini, M., and Pejović, V. (2024). Where is the testbed for my federated learning research? In *2024 IEEE/ACM Symposium on Edge Computing (SEC)*, pages 249–264. IEEE.

Liu, B., Lv, N., Guo, Y., and Li, Y. (2024). Recent advances on federated learning: A systematic survey. *Neurocomputing*, page 128019.

Meyer, B. H., Pozo, et al. (2024). Aprendizado federado com autokeras e knowledge distillation. In *Anais da XXIV Escola Regional de Alto Desempenho da Região Sul*.