



MAX-PLANCK-GESELLSCHAFT

A Systems-Theoretic Approach to Event-Based Distributed Learning

Guener Dilsad ER¹ Sebastian Trimpe² Michael Muehlebach¹¹ Max Planck Institute for Intelligent Systems ² RWTH Aachen UniversityMAX PLANCK INSTITUTE
FOR INTELLIGENT SYSTEMS

Summary

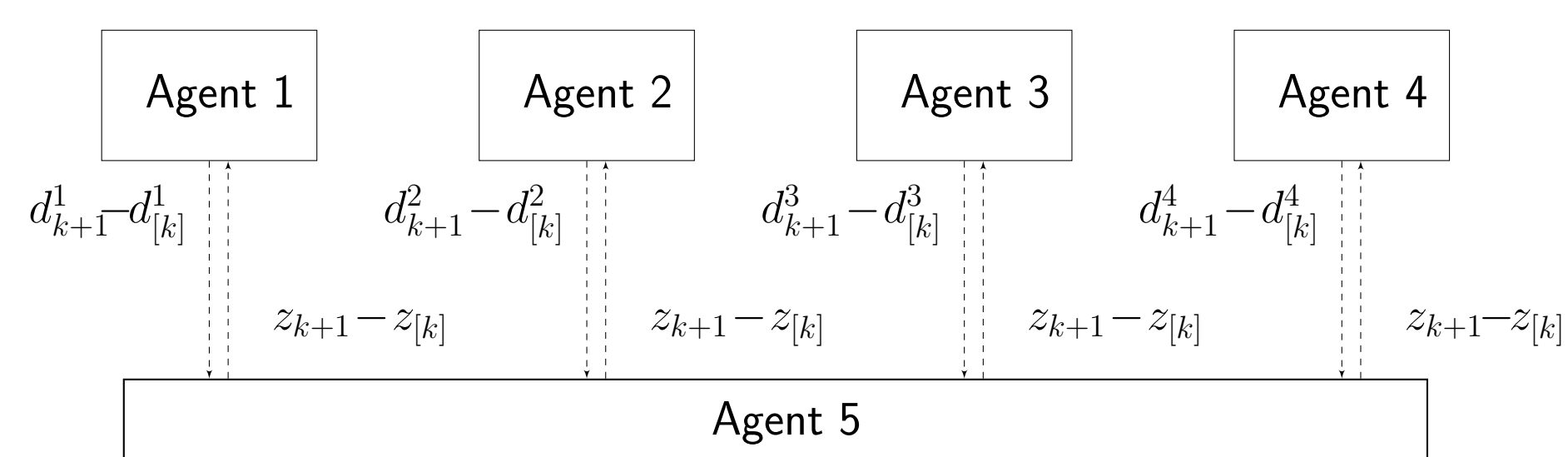
- **Objective:** Develop a communication-efficient distributed optimization algorithm that handles non-i.i.d. data distribution across agents.
- **Challenge:** High communication overhead and convergence with non-i.i.d. data.
- **Solution:** An event-based communication strategy that reduces unnecessary data exchanges while maintaining model accuracy.
Key mechanism: Agents communicate updates only when significant changes occur, based on predefined thresholds.
- **Theoretical Analysis:** Explores the trade-off between communication efficiency, model accuracy, and convergence speed.
- **Empirical Results:** Evaluated on MNIST and CIFAR-10 datasets.
Our method reduces communication costs by over 70% while preserving high classification accuracy.
- **Impact:** Provides a scalable, efficient solution for large-scale distributed optimization in federated learning.

We tackle a **distributed optimization problem** where multiple agents with local datasets collaborate to minimize a global objective function, represented as the sum of their individual objectives. The **non-independent and identically distributed (non-i.i.d.) data** across agents poses challenges for global optimality, requiring communication to ensure convergence.

The problem is formulated as a consensus problem:

$$\min_{x_1, \dots, x_N \in \mathbb{R}^n} \sum_{i=1}^N f^i(x^i) + g(z), \quad \text{subject to } x^i = z, \quad \forall i = 1, \dots, N,$$

where $g(z)$ represents nonsmooth part of the objective function, and z is the global consensus variable.



Algorithm 1 Event-Based Distributed Learning with Over-Relaxed ADMM

Require: Number of agents N , number of rounds $t_{\max}$, local objective functions f^i , penalty parameter ρ , communication thresholds Δ^d, Δ^z , reset period T

Require: Initialize $\hat{x}_0^i = x_0$, $\hat{z}_0 = \zeta_0 = x_0$, $\hat{u}_{-1}^i = u_0^i$

for $k = 0$ to $t_{\max}$ **do**

for $i = 1$ to N **do**

$\hat{z}_k^i \leftarrow$ event-based receive of $z_{k+1} - z_k$

Agent i

$u_k^i = u_{k-1}^i + \alpha x_{k-1}^i - \hat{z}_k^i + (1 - \alpha)z_{k-1}^i$

$x_{k+1}^i = \arg \min_{x^i} f^i(x^i) + \frac{\rho}{2} \|x^i - \hat{z}_k^i + u_k^i\|^2$

→ Local Training

 event-based send of $d_{k+1}^i - d_k^i$ where $d_{k+1}^i = \alpha x_{k+1}^i + u_k^i$

end for

$\hat{\zeta}_k \leftarrow$ event-based receive of $d_{k+1}^i - d_k^i$

Agent $N+1$

$z_{k+1} = \arg \min_z g(z) + \frac{N\rho}{2} \|z - \hat{\zeta}_k + (1 - \alpha)z_k\|^2$

→ Global Aggregation

if $\text{mod}(k+1, T) = 0$ **then**

 perform **reset** → $\hat{\zeta}_k = \zeta_k$, $\hat{z}_k = z_k$

end if

end for

A communication is triggered by the agent if the value d_{k+1}^i deviates by more than the predefined threshold $\Delta^d > 0$ from the last communicated value d_k^i . Specifically, communication occurs if $|d_{k+1}^i - d_k^i| > \Delta^d$, and the difference $d_{k+1}^i - d_k^i$ is then sent out. If $|z_{k+1}^i - z_k^i| > \Delta^z$, and the difference $z_{k+1}^i - z_k^i$ is sent out by the server.



References

- [1] G. Dilsad Er, Sebastian Trimpe, and Michael Muehlebach. "Distributed Event-Based Learning via ADMM". In: *arxiv:2405.10618* (2024).

Event-Based Communication in Federated Learning

Event-based communication reduces communication by triggering updates only when necessary and is robust to heterogeneous data-distributions among agents and communication failures.

Event-based communication is also effective in a distributed learning setting, where the aim is to minimize $\sum_{i=1}^N f^i(x)$, see [1] for more details.

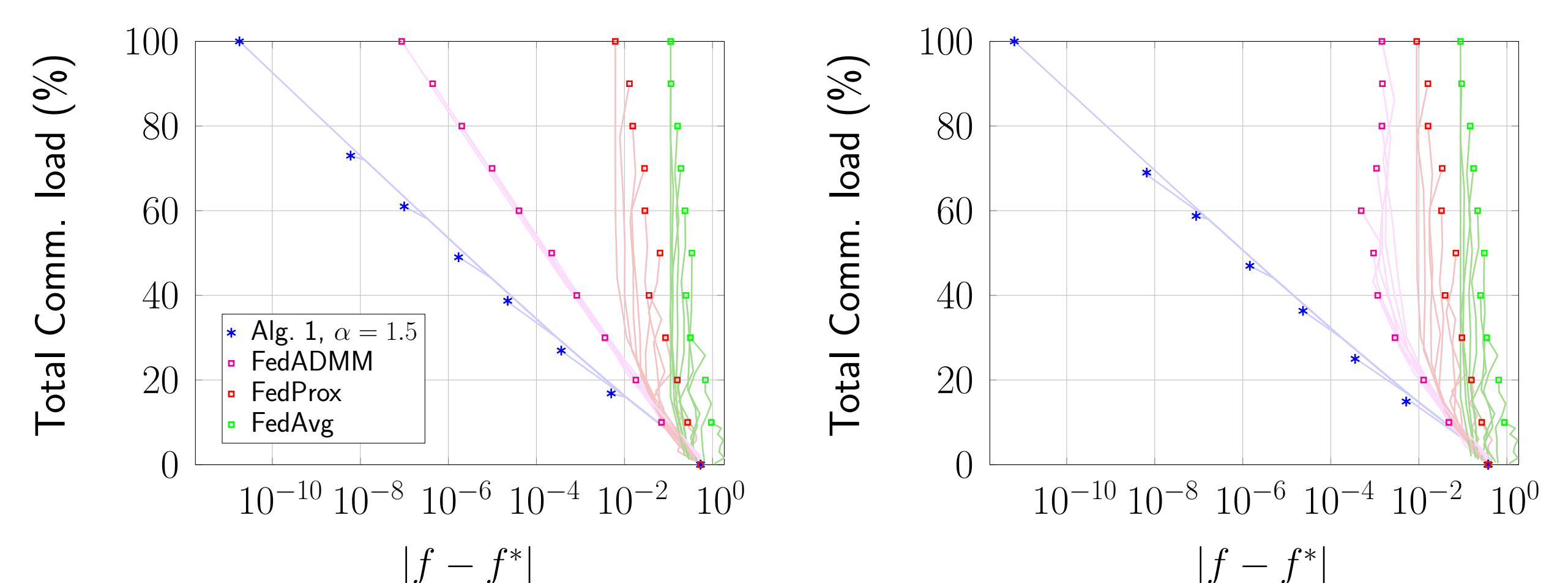
The event-based framework reduces communication by only transmitting updates when significant changes occur. We established the following theoretical result in a convex setting,

$$|\xi_k - \xi_*|^2 \leq \kappa_P |\xi_0 - \xi_*|^2 \left(1 - \frac{\alpha}{4\kappa^{\epsilon+\frac{1}{2}}}\right)^{2k} + \frac{60\kappa^{2+2\epsilon}}{\alpha(1-|\alpha-1|)} \Delta^2,$$

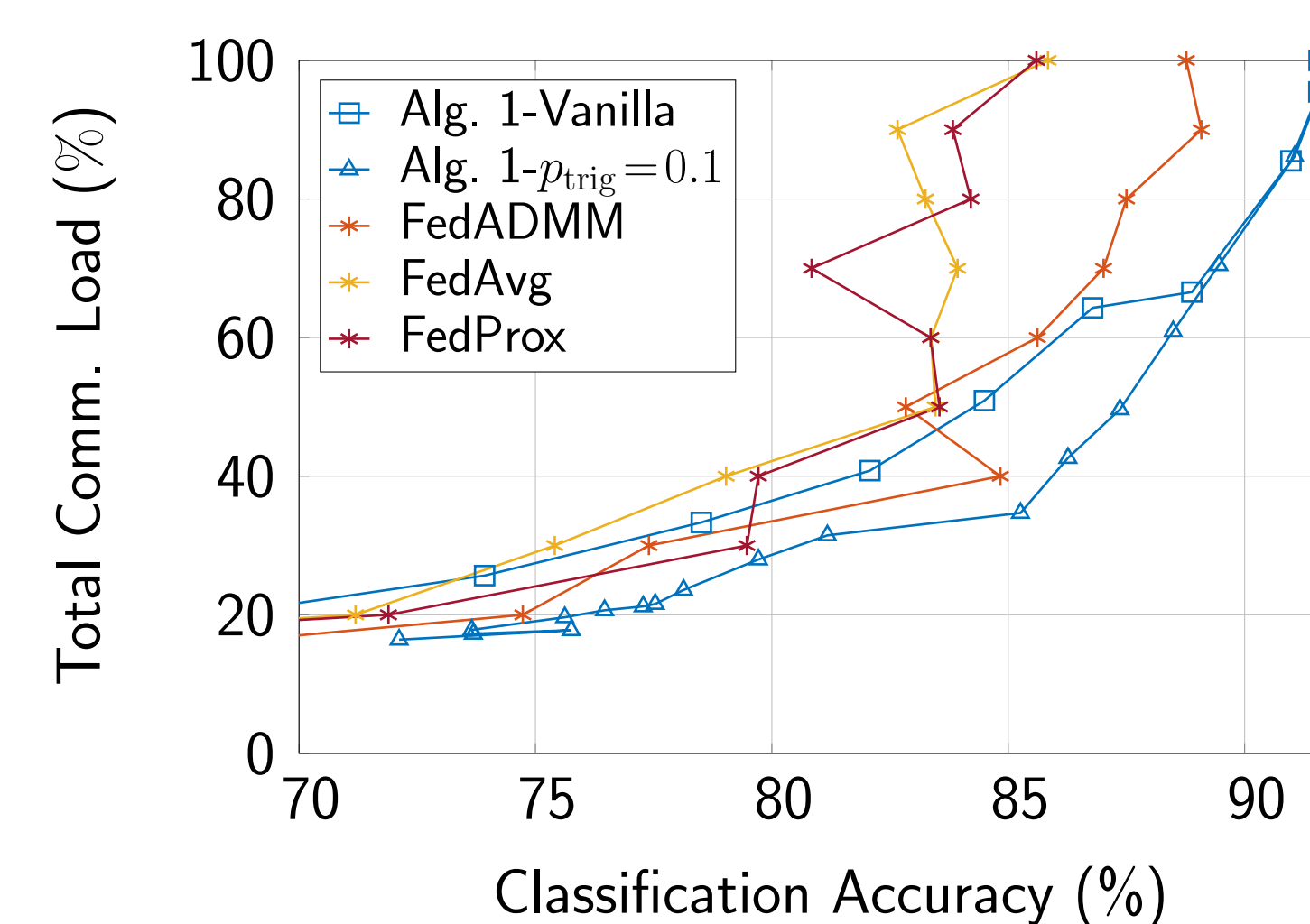
where ξ_k is the server variable and ξ_* is the optimizer of $\sum_{i=1}^N f^i(x)$. Furthermore, Δ represents the error arising from the event-based communication which scales by the number of agents, κ_P is a function of the relaxation parameter α and the condition number κ of the objective function $\sum_{i=1}^N f^i(x)$, and ϵ is a parameter that scales the step-size of the algorithm.

Empirical Evaluation

Event-based communication reduces communication cost and result in a better trade-off compared to well-known baselines FedAvg, FedProx and FedADMM.

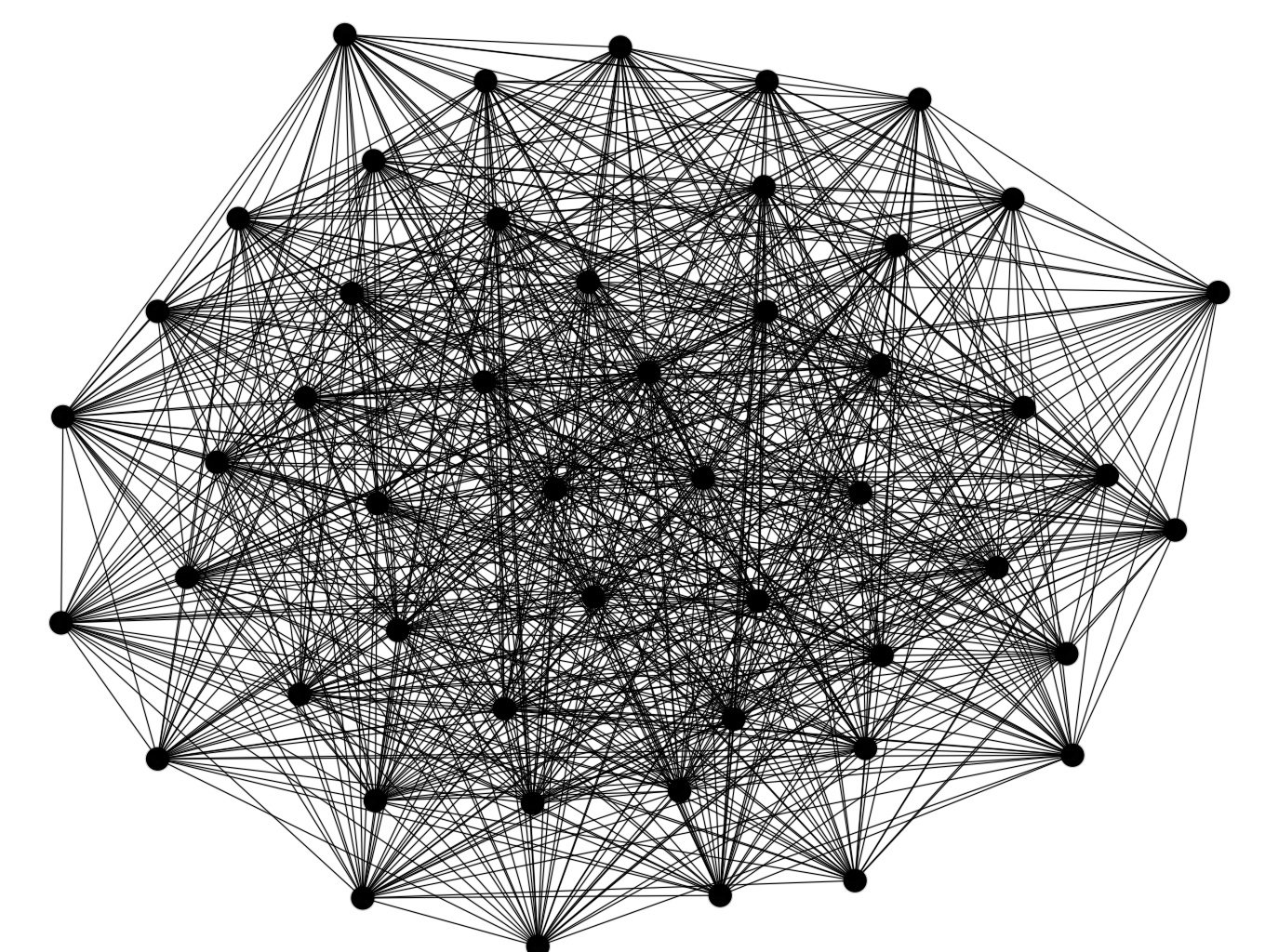
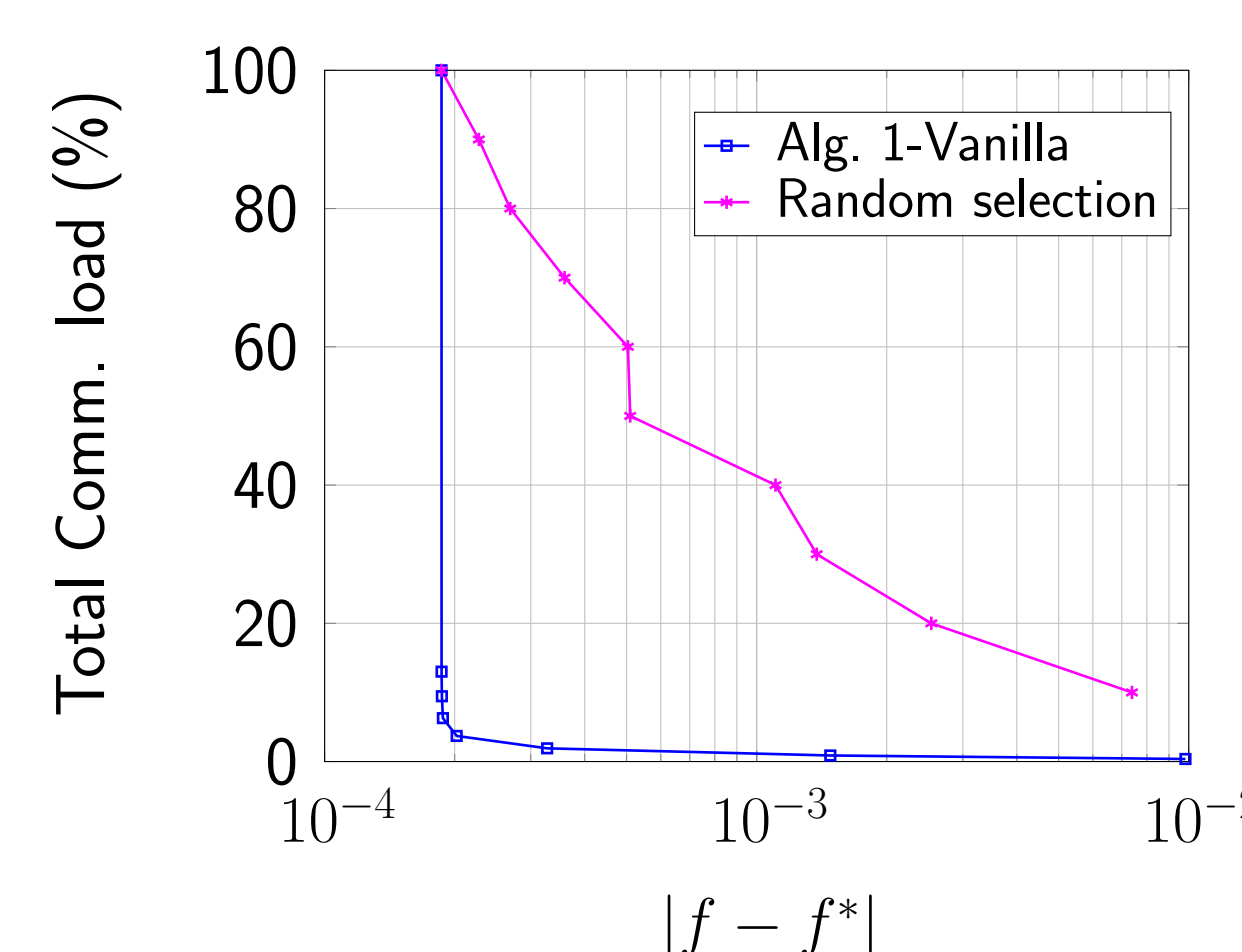
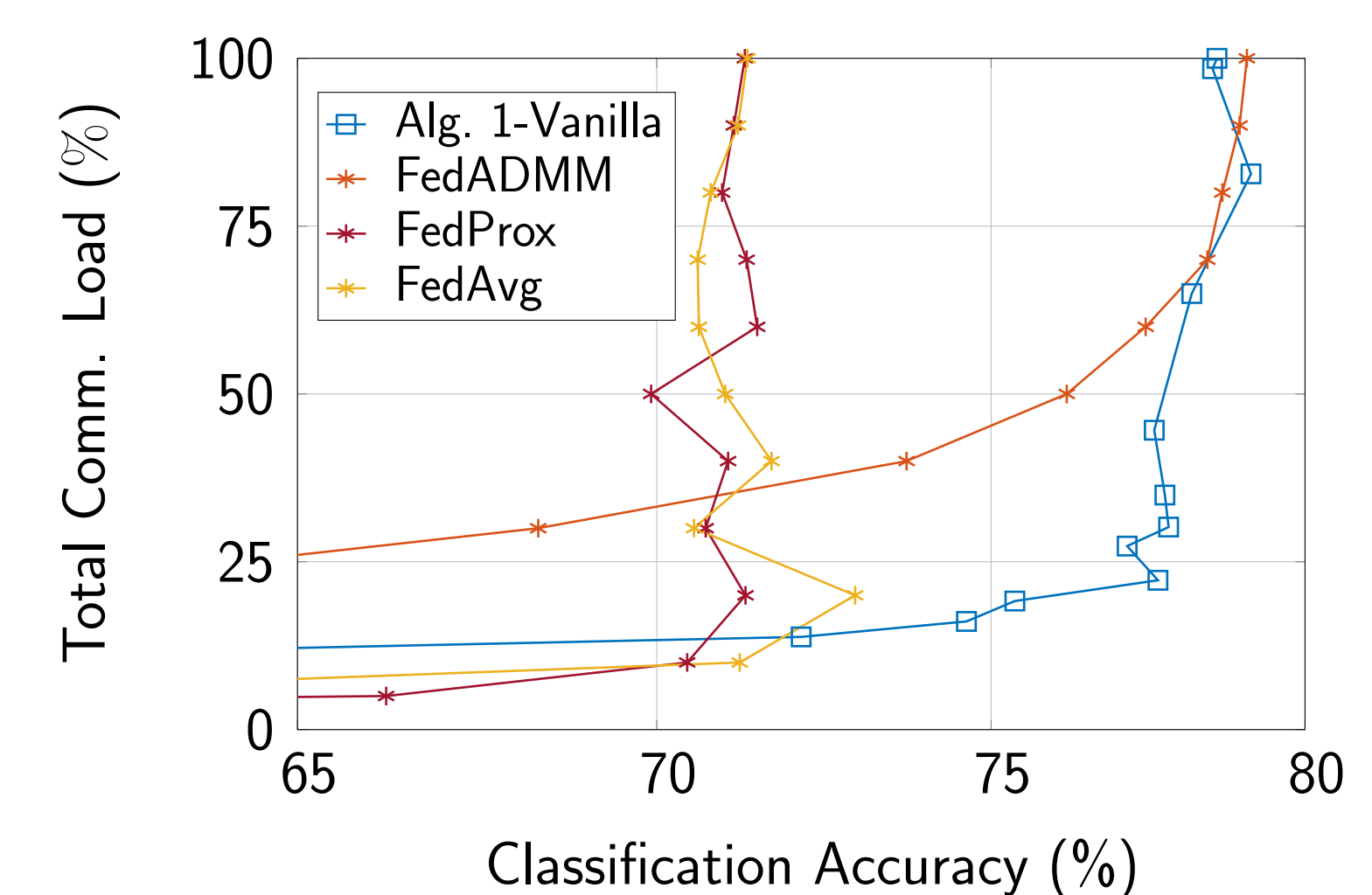


The figure shows the communication load versus accuracy trade-off for different methods applied to two distinct problems: linear regression (left panel), and LASSO (right panel).



Comparison of different federated learning methods with respect to the resulting trade-off between total communication load and classification accuracy on the entire **MNIST** dataset. Setup consists of $N = 10$ training agents, where each agent i possesses only images of digit- i .

Comparison of different federated learning methods on the **CIFAR-10** dataset with respect to the resulting trade-off between total communication load and classification accuracy on the test set. The setup includes $N = 100$ agents with a non-i.i.d. data distribution.



The left panel compares the communication load versus solution accuracy for different communication methods applied to the linear regression problem. The right panel visualizes the agent network with 50 agents connected by 1762 edges.