Evaluating the Resilience of Decentralized Federated Learning to Model Poisoning Attacks

SAPIENZA
UNIVERSITÀ DI ROMA

Giovanni Pica

A.Y. 2022/2023

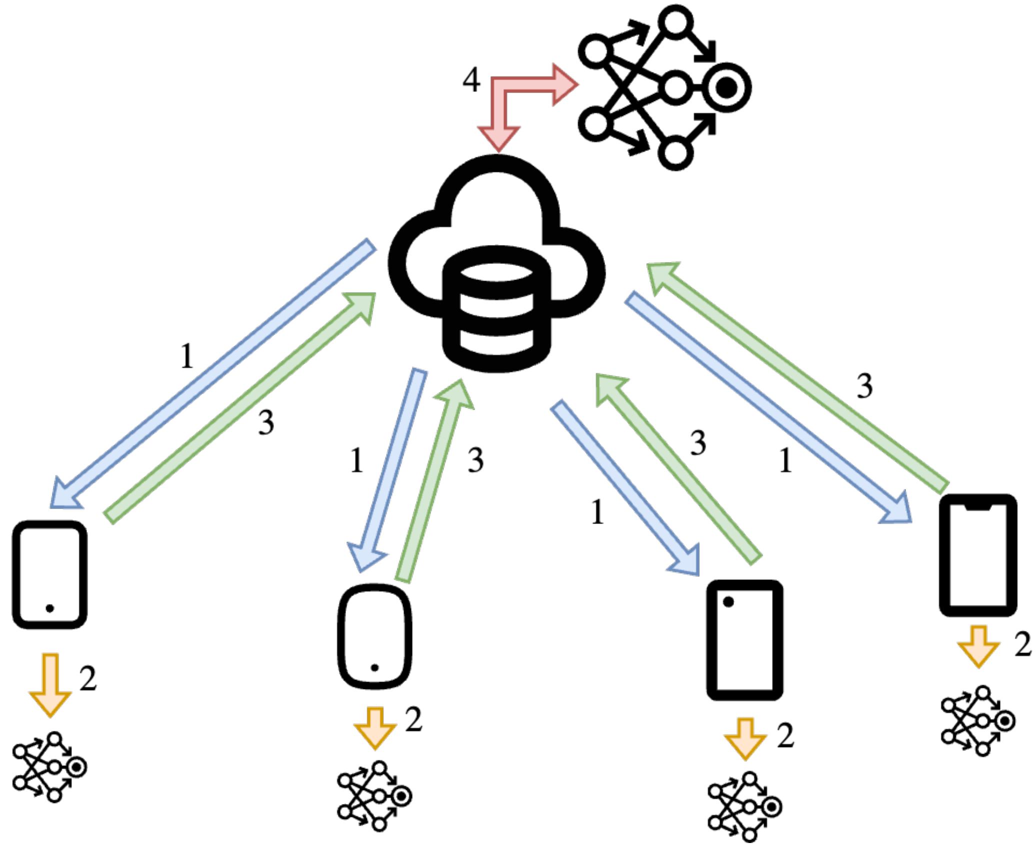
Advisor: Gabriele Tolomei

Federated Learning

Federated Learning is a technique to train a global ML model, shared across multiple clients. There is a central server that acts as an orchestrator, collecting and aggregating local models from clients until convergence.

Federated Learning

Federated Learning is a technique to train a global ML model, shared across multiple clients. There is a central server that acts as an orchestrator, collecting and aggregating local models from clients until convergence.



Federated Learning - Problems

- Security and Privacy
- Communication Efficiency
- Data and System Heterogeneity
- Incentive Mechanisms

Federated Learning - Problems

- Security and Privacy
- Communication Efficiency
- Data and System Heterogeneity
- Incentive Mechanisms
- **Centralized Orchestration**

Federated Learning - Problems

- Security and Privacy
- Communication Efficiency
- Data and System Heterogeneity
- Incentive Mechanisms
- **Centralized Orchestration**

How to overcome these limitations?

Federated Learning - Problems

- Security and Privacy
- Communication Efficiency
- Data and System Heterogeneity
- Incentive Mechanisms
- **Centralized Orchestration**

How to overcome these limitations?

Decentralized Federated Learning

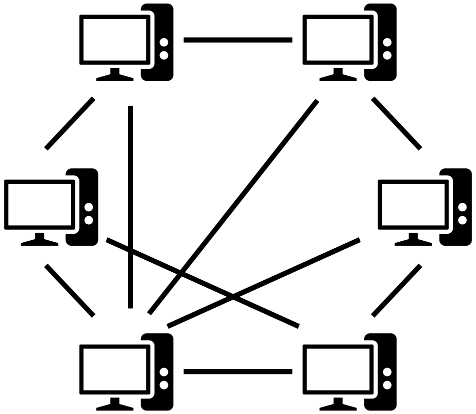
Decentralized Federated Learning

To obtain decentralization, in the literature we observed [1] two kind of approaches.

Decentralized Federated Learning

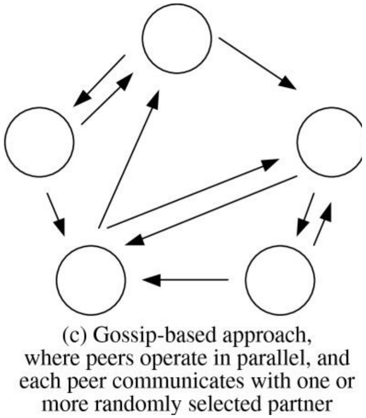
To obtain decentralization, in the literature we observed [1] two kind of approaches.

"Standard" Distributed Computing Techniques



P2P Networks

Image taken from wikipedia



Gossip communication

Image taken from https://haritibcoblog.wordpress.com/2018/11/01/what-is-a-gossip-protocol/

Decentralized Federated Learning

To obtain decentralization, in the literature we observed [1] two kind of approaches.

Blockchain-based

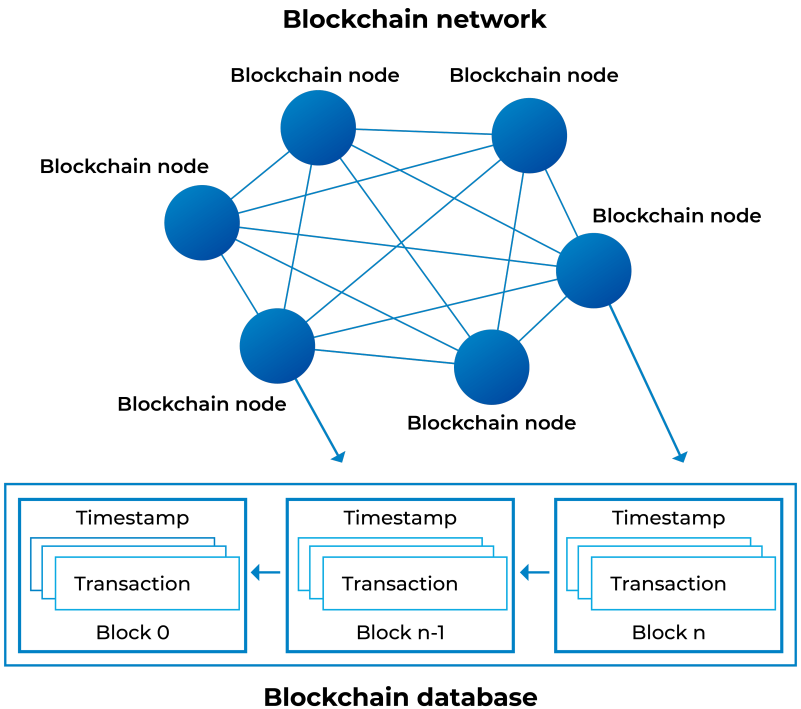


Image taken from https://unova.io/blockchain/

Decentralized Federated Learning - Problems?

- Centralized FL (and for extension decentralized FL) suffers **model poisoning** attacks.
- Model poisoning exploits the inherent feature of federated learning, allowing malicious participants to directly influence the collective model.

Decentralized Federated Learning - Problems?

- Centralized FL (and for extension decentralized FL) suffers **model poisoning** attacks.
- Model poisoning exploits the inherent feature of federated learning, allowing malicious participants to directly influence the collective model.
- The centralized FL literature offers various strategies to address model poisoning attacks, with the goal of filtering out potentially malicious local updates during the server-side aggregation process.

Decentralized Federated Learning - Problems?

- Centralized FL (and for extension decentralized FL) suffers **model poisoning** attacks.
- Model poisoning exploits the inherent feature of federated learning, allowing malicious participants to directly influence the collective model.
- The centralized FL literature offers various strategies to address model poisoning attacks, with the goal of filtering out potentially malicious local updates during the server-side aggregation process.
- **However**, there is a notable absence of experimental studies examining model poisoning attacks on decentralized FL systems in the literature.

Research Questions

- **RQ1**: Does decentralized FL exhibit resilience against significant model poisoning attacks?
- **RQ2**: Do the adapted aggregation methods perform effectively in decentralized FL environments?

Attacks

- We opted for a standard attack and two more advanced attacks.

Attacks

- We opted for a standard attack and two more advanced attacks.
- **Gaussian Attack [2]**

Attacks

- We opted for a standard attack and two more advanced attacks.
- **Gaussian Attack [2]**
- **“A Little Is Enough” Attack [3]**
- **“Fall of Empires” Attack [4]**

Adapted Aggregations

- Assumption: Each client in the network performing its own aggregation for its specific purposes, acting as a local server for itself.

Adapted Aggregations

- Assumption: Each client in the network performing its own aggregation for its specific purposes, acting as a local server for itself.
- **FedAvg [5]**

Adapted Aggregations

- Assumption: Each client in the network performing its own aggregation for its specific purposes, acting as a local server for itself.
- **FedAvg [5]**
- **Multi-Krum [6]**

Adapted Aggregations

- Assumption: Each client in the network performing its own aggregation for its specific purposes, acting as a local server for itself.
- **FedAvg [5]**
- **Multi-Krum [6]**
- **Median [7]**

Proposed Method - PENS

- We decided to choose PENS [8] as decentralized FL framework to perform our analysis.
- PENS is divided in two parts:
    - Identification of peers with similar data distributions.
    - Selection of peers who have been chosen more frequently than expected. Thanks to Gossip Learning techniques.

Proposed Method - Threat Model

- **Adversarial Model**
- **Attacker’s Knowledge**
- **Attacker’s Behaviour**

Experiments - Setup

- We work with three datasets: MNIST, Fashion-MNIST, Spambase.
- We have used two ML models, Logistic Regression and MLP.
- As evaluation metric we use the average accuracy of all models of the nodes.
- To simulate our environment we use Gossipy. We set the number of clients as 50, and two parameters of PENS defined as number of received model set to 5 and number of top-performing clients set to 2.

Experiments - Evaluation

- We compare the robustness of the decentralized framework employed with the three aggregation schemes FedAvg, Multi-Krum and Median.
- And with the three attacks Gaussian, A Little Is Enough and Fall of Empires.
- Finally, we decided to set the number of malicious nodes as b = {3, 5, 15, 25}.
- The approach exhibits resilience against all the proposed attacks, whether employing other aggregation schemes.

Experiments - Evaluation

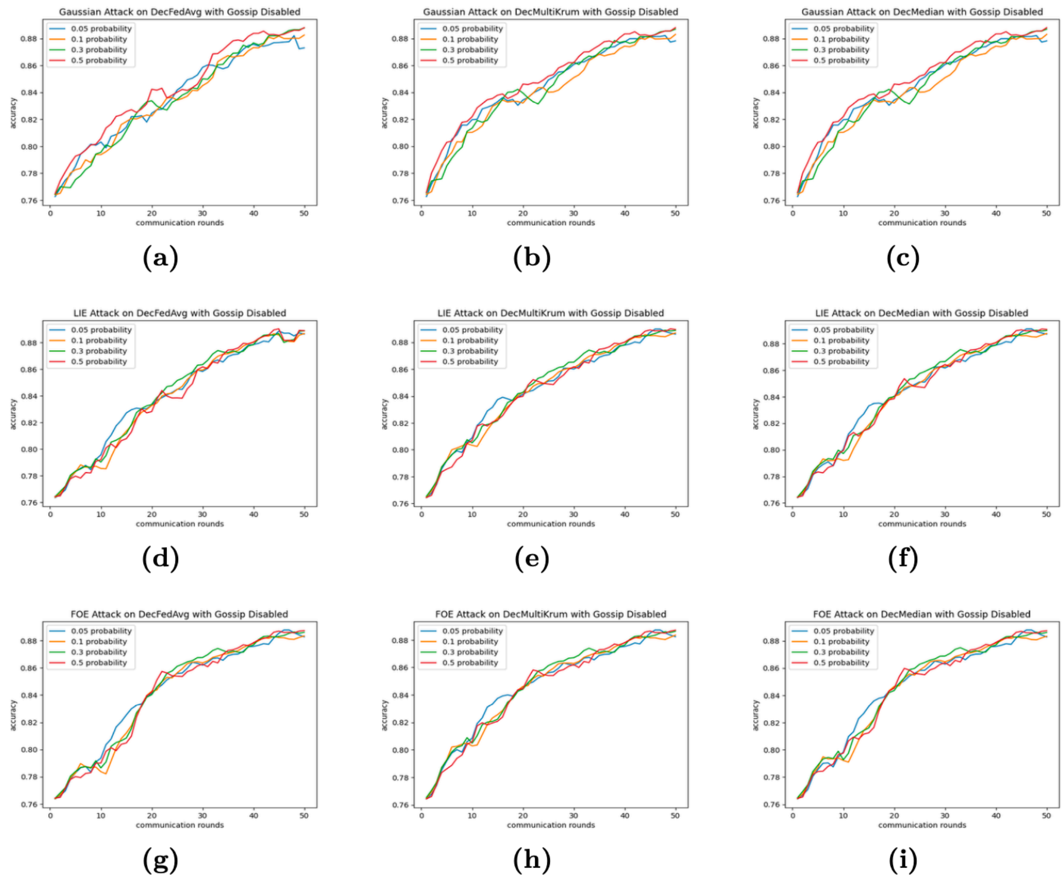


Figure 6.1. Comparing the robustness of PENS, with MNIST dataset. For each row we have the aggregation schemes, and for each column the attacks.

Experiments - Evaluation

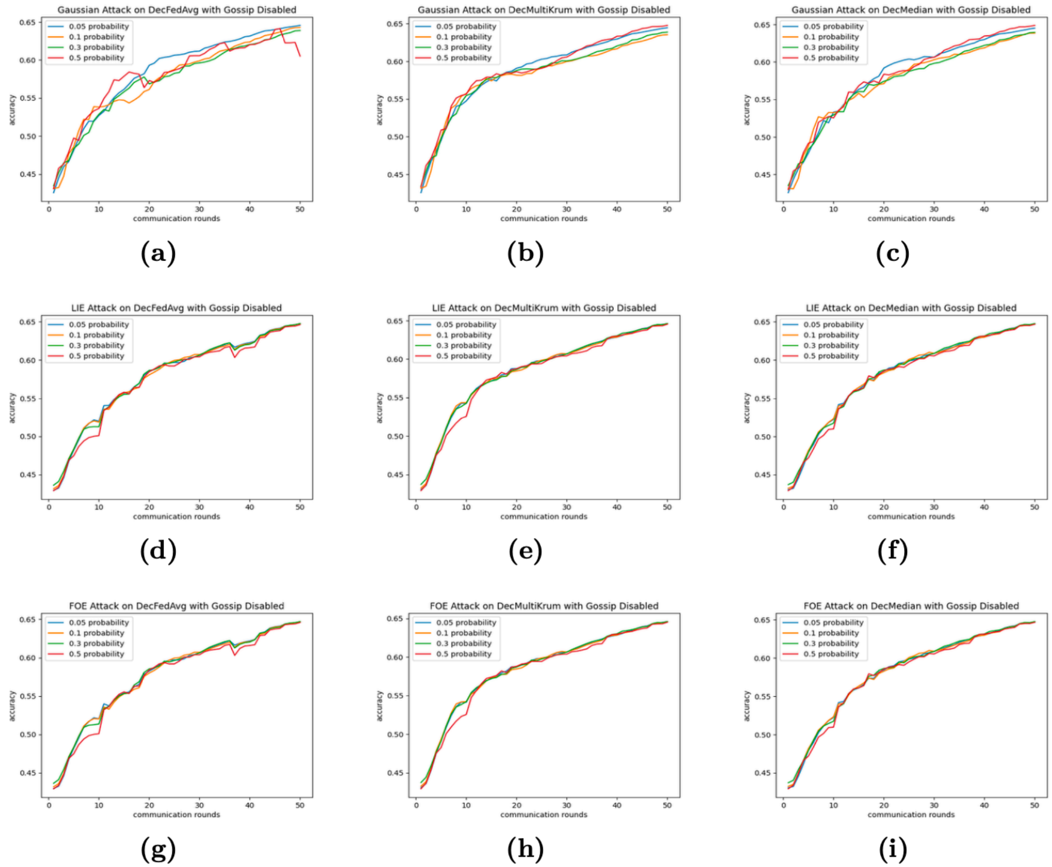


Figure 6.2. Comparing the robustness of PENS, with Fashion-MNIST dataset. For each row we have the aggregation schemes, and for each column the attacks.

Experiments - Evaluation

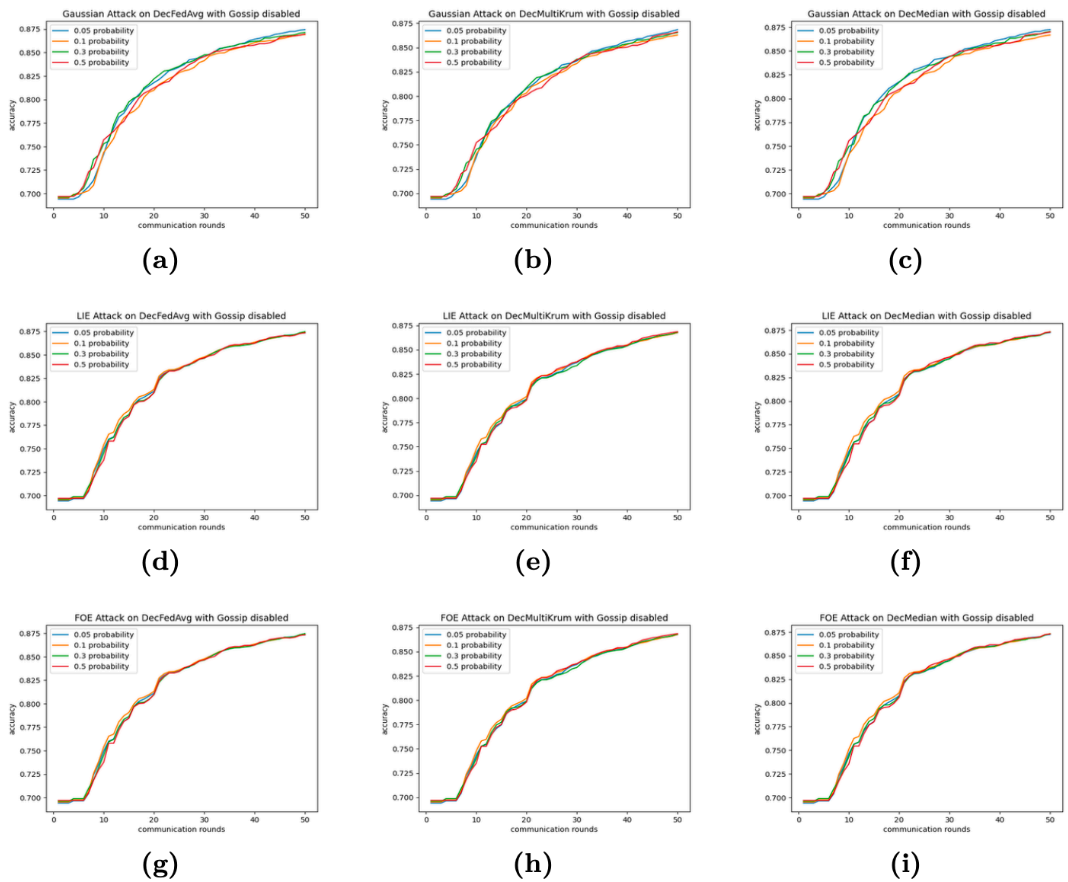


Figure 6.3. Comparing the robustness of PENS, with Spambase dataset. For each row we have the aggregation schemes, and for each column the attacks.

Experiments - Impact of Gossip Learning

- Notably, we have chosen to deactivate the second step of PENS, which involves gossip learning.
- This decision arises from the belief that, in the case of attacks, employing this algorithm would not yield significant benefits.

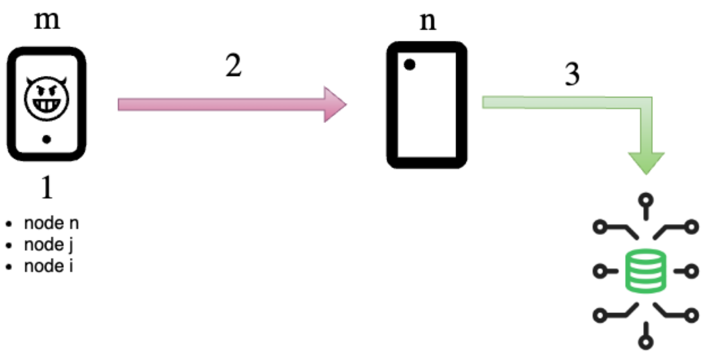
Experiments - Impact of Gossip Learning

- Notably, we have chosen to deactivate the second step of PENS, which involves gossip learning.
- This decision arises from the belief that, in the case of attacks, employing this algorithm would not yield significant benefits.

Algorithm 1 Gossip learning protocol

```
function MAIN
   while stopping criterion not met do
      WAIT(Δ)
      j ← RANDOMPEER() // select random peer
      SEND_{i→j}(w_i, j)
   end while
end function
function ONRECEIVEMODEL(w_i)
   SAVE(w_i)
   if no. of received models ≥ n_peers then
      w_j ← MERGE_SAVED_MODELS()
      w_j ← TRAIN(x; w_j) //update on local data x
   end if
end function
```

Image taken from PENS paper



Experiments - Impact of Gossip Learning

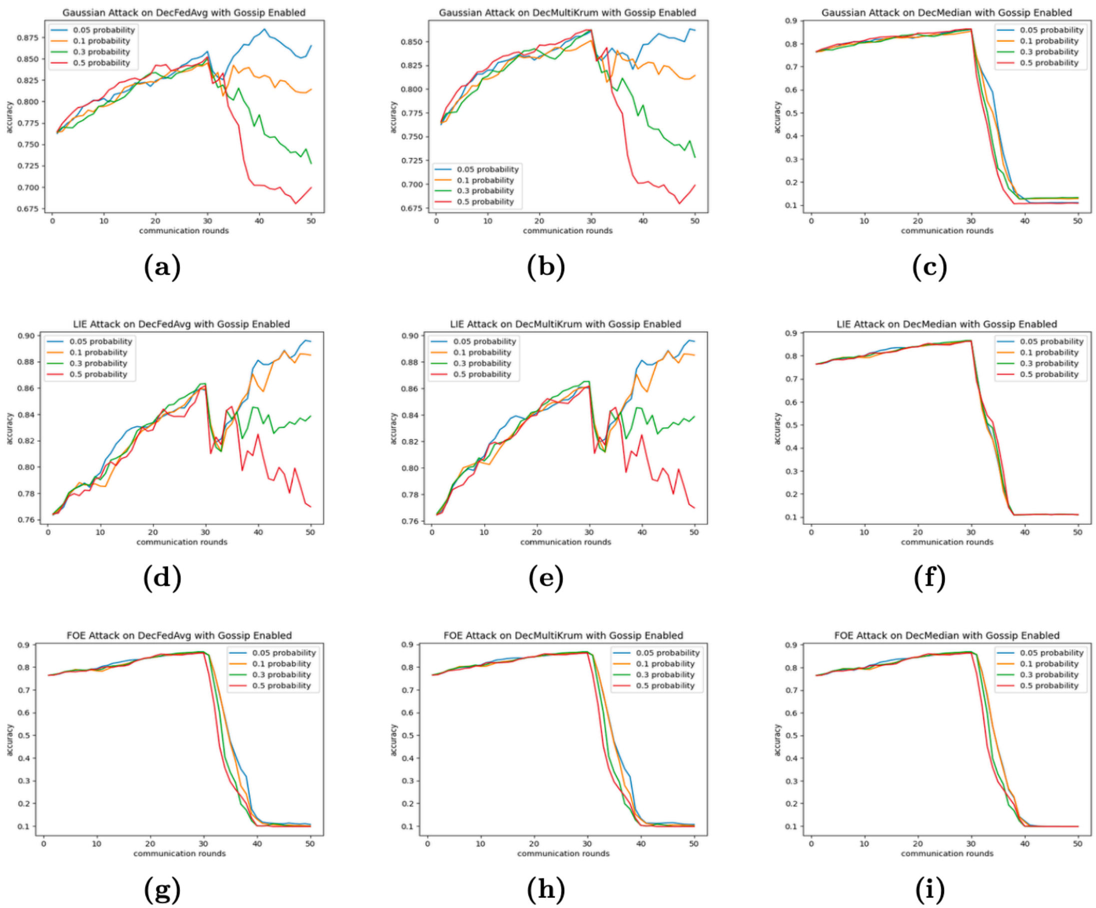


Figure 6.4. Comparing the robustness of PENS, with MNIST dataset and gossip learning enabled. For each row we have the aggregation schemes, and for each column the attacks.

Conclusions

- In this research, we have presented a novel study investigating the resilience of decentralized approaches.
- Notably, we have demonstrated and answered to RQ1 that the decentralized approach exhibits robustness against three distinct types of attacks even when we are in the standard case with FedAvg.
- Also, we answered to RQ2 through the implementation of the other aggregation schemes, and the approach demonstrates its robustness also with them.
- Our future objectives include verifying the resilience of these approaches against newly developed attacks specifically tailored for this approach.
- Additionally, we aim to extend our experimental findings to envelop other datasets and various neural network architectures.
- Finally, we plan to expand our experiments to incorporate other decentralized FL approaches that do not have the requirement for gossip learning or incorporate a more generalized variant of it, and may not necessarily rely on neighbor selection.

Thank You!

For Your Attention

References

1. Gabrielli, E., Pica, G., and Tolomei, G. A survey on decentralized federated learning (2023). http://arxiv.org/abs/2308.04604
2. Fang, M., Cao, X., Jia, J., and Gong, N. Z. Local model poisoning attacks to byzantine-robust federated learning (2021). http://arxiv.org/abs/1911.11815
3. Baruch, G., Baruch, M., and Goldberg, Y. A Little Is Enough: Circumventing Defenses for Distributed Learning. In *Proc. of NeurIPS '19*, pp. 8632–8642 (2019). https://proceedings.neurips.cc/paper/2019/hash/ec1c59141046cd1866bbbcdfb6ae31d4-Abstract.html
4. Xie, C., Koyejo, S., and Gupta, I. Fall of empires: Breaking byzantine-tolerant sgd by inner product manipulation (2019). http://arxiv.org/abs/1903.03936
5. McMahan, B., Moore, E., Ramage, D., Hampson, S., and Arcas, B. A. y. Communication-Efficient Learning of Deep Networks from Decentralized Data. In *Proc. of AISTATS '17'* (edited by A. Singh and J. Zhu), vol. 54, pp. 1273–1282. PMLR (2017). https://proceedings.mlr.press/v54/mcmahan17a.html
6. Blanchard, P., El Mhamdi, E. M., Guerraoui, R., and Stainer, J. Machine learning with adversaries: Byzantine tolerant gradient descent. In *Adv. in NeurIPS* (edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett), vol. 30. Curran Associates, Inc. (2017). https://proceedings.neurips.cc/paper_files/paper/2017/file/f4b9ec30ad9f68f89b29639786cb62ef-Paper.pdf
7. Yin, D., Chen, Y., Kannan, R., and Bartlett, P. Byzantine-robust distributed learning: Towards optimal statistical rates. In *Proceedings of the 35th International Conference on Machine Learning* (edited by J. Dy and A. Krause), vol. 80 of *Proceedings of Machine Learning Research*, pp. 5650– 5659. PMLR (2018). https://proceedings.mlr.press/v80/yin18a.html
8. Onoszko, N., Karlsson, G., Mogren, O., and Zec, E. L. Decentralized federated learning of deep neural networks on non-iid data (2021). http://arxiv.org/abs/2107.08517