Federated Learning Local Model Poisoning Attack

Samuel Sibhatu and Qin, Hu(Ph.D.) ; Wang, Zhilin
Indiana University School of Informatics

Abstract

Federated Learning is distributed learning a global and local devices (e.g., PC, mobile, IOT) collaborate to produce a global model.

Having a partial knowledge of the training, the aim is to compromise the global model by manipulating local model.

Introduction

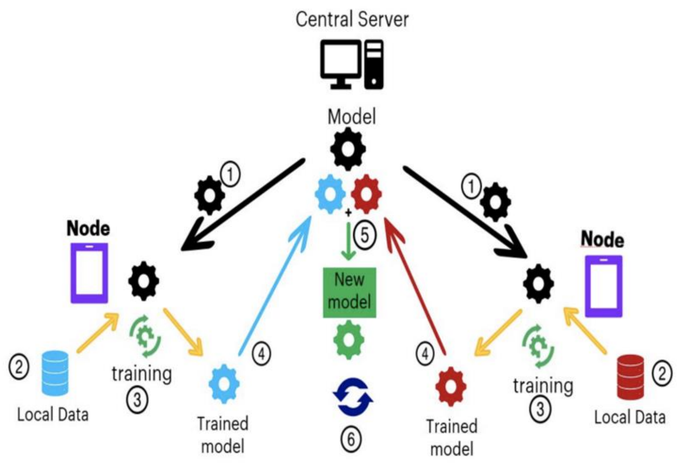

Figure 1: Federated Learning Steps

The method of learning is a great solution for industries interested in collaborating with out sharing their local datasets and data privacy.

However, an attacker can craft the local model parameters to deviate the accuracy of global model. The research proves an attacker can compromise the global model. Thus, alternative aggregation rule robust against failure needed.

Material and Methods

- 70000 Dermatological 28x 28 gray scale MNIST datasets
- 60000 training and 10000 testing examples
- Deviate local model parameter to compromise global model update

Results

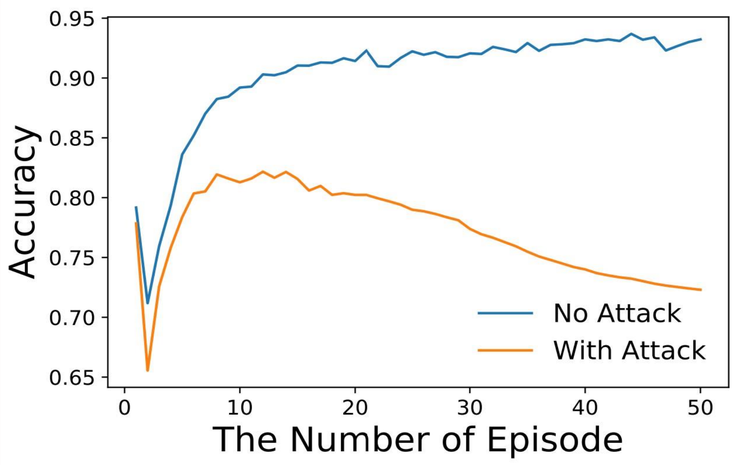

Figure 2: Before the attack

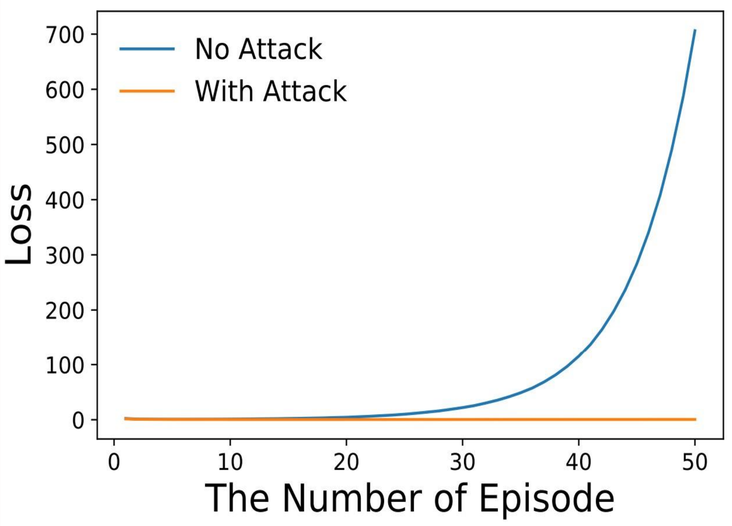

Figure 3: After the attack

Results continued

- The error rates are normal under non-adversarial situation (Figure 2).
- But the error rates increase after successful attack (Figure 3).

Conclusions

- Results indicate number of error rates increase as number of compromised worker devises increases.
- The survey yet need to experiment how local model poisoning attack perform in other aggregation rules “Krum, Bulyan trimmed, mean, and median agrégation aimed robust against byzantine failures of certain clients.” (Minghong Fang et al)

References

Zhilin Wang, Qiao Kang et al (Feb 13, 2022)” Strategies Toward Model Poisoning Attacks in Federated Learning: A Survey” https://arxiv.org/pdf/2202.06414.pdf accessed on 7/20/2022

Minghong Fang∗1, Xiaoyu Cao( ) “Local Model Poisoning Attacks to Byzantine-Robust Federated Learning” https://www.usenix.org/system/files/sec20summer_fang_prepub.pdf accessed on 7/22/2022

Acknowledgements

I thank my mentors and IUPUI LSAMP coordinators for encouraging and guiding me during the project.

This material is based upon work supported by the National Science Foundation under Grant No. 1618-408. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.