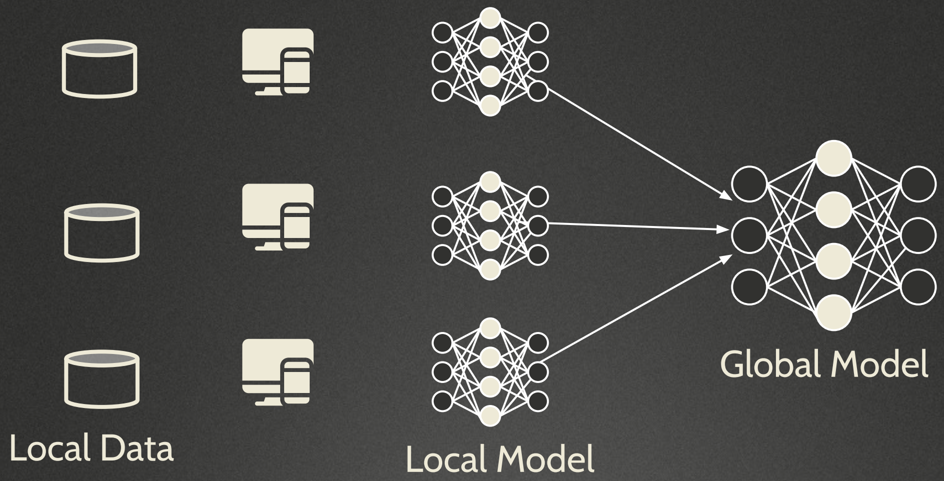
Better Together: Attaining the Triad of Byzantine-robust Federated Learning via Local Update Amplification

Liyue Shen, Yanjun Zhang, Jingwei Wang, Guangdong Bai

The University of Queensland, Deakin University

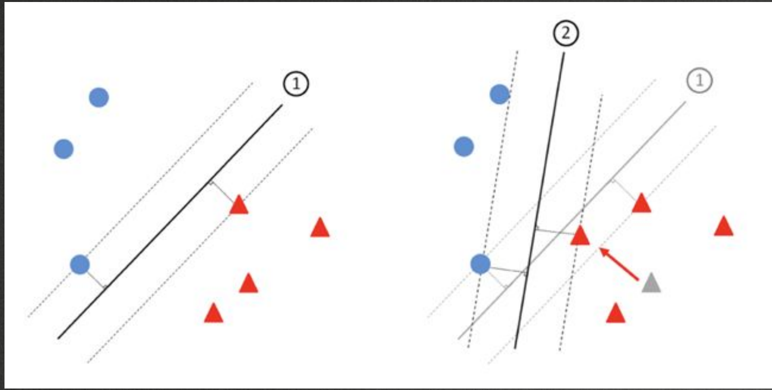
Federated Learning



Poisoning attacks in Federated Learning

Untargeted attack

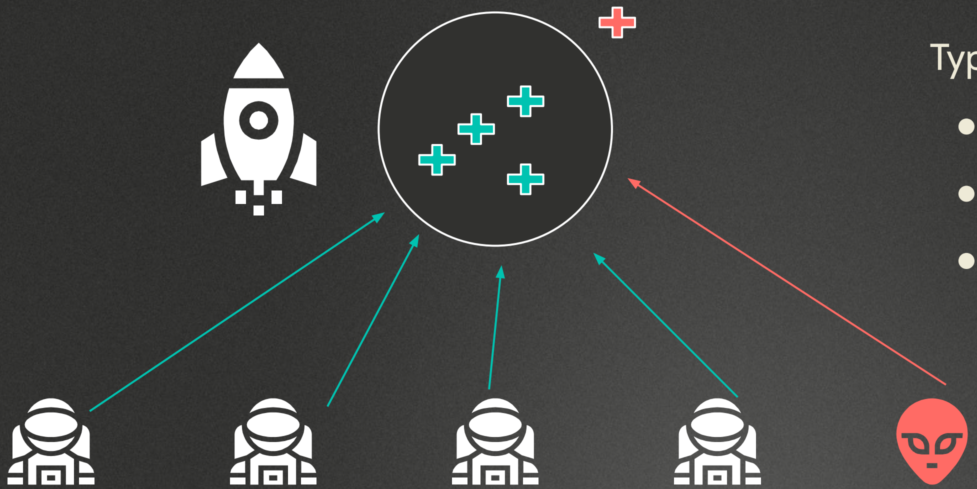
Targeted attack



An example: The decision boundary of the classifier is significantly impacted if just one training sample is changed [1]

[1] Miller, David J., Zhen Xiang, and George Kesidis. "Adversarial learning targeting deep neural network classification: A comprehensive review of defenses against attacks." Proceedings of the IEEE 108.3 (2020): 402-433.

Byzantine-robust Methods against Poisoning Attacks



Typical AGR strategies:

- Distance Based [1, 2]
- Prediction Based [3]
- Trust Bootstrapping Based [4]

[1] Peva Blanchard, El Mahdi El Mhamdi, Rachid Guerraoui, and Julien Stainer. 2017. Machine learning with adversaries: Byzantine tolerant gradient descent. In Proceedings of the 31st International Conference on Neural Information Processing Systems. 118–128.
[2] El Mahdi El Mhamdi, Rachid Guerraoui, and Sébastien Rouault. 2018. The Hidden Vulnerability of Distributed Learning in Byzantium. (2018).
[3] Minghong Fang, Xiaoyu Cao, Jinyuan Jia, and Neil Gong. 2020. Local model poisoning attacks to byzantine-robust federated learning. In 29th {USENIX} Security Symposium ({USENIX} Security 20). 1605–1622.
[4] Xiaoyu Cao, Minghong Fang, Jia Liu, and Neil Zhenqiang Gong. 2020. FLTrust: Byzantine-robust Federated Learning via Trust Bootstrapping. arXiv preprint arXiv:2012.13995 (2020).

Triad of Byzantine-robust Federated Learning [1]

Robustness

The method shall minimize the decrease of the global model’s test accuracy caused by malicious updates.

Efficiency

The method shall retain the computation efficiency and scale in large-scale FL training

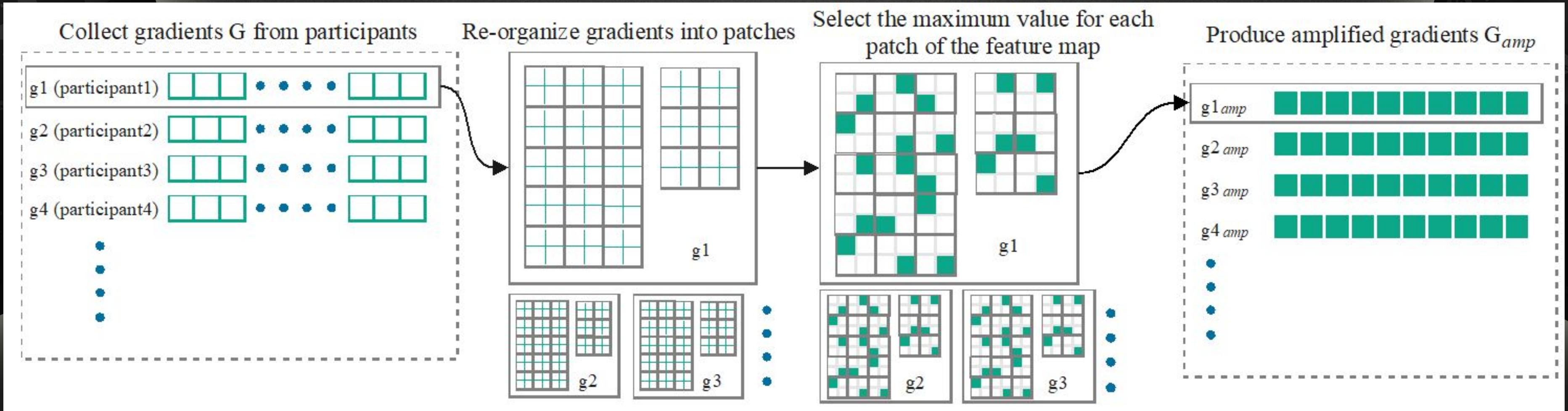
Fidelity

The method shall not harm the performance of the global model when there is no malicious updates.

[1] Xiaoyu Cao, Minghong Fang, Jia Liu, and Neil Zhenqiang Gong. 2020. FLTrust: Byzantine-robust Federated Learning via Trust Bootstrapping. arXiv preprint arXiv:2012.13995 (2020).

Better Together: Attaining the Triad of Byzantine-robust Federated Learning via Local Update Amplification

AgrAmplifier



Achieving Robustness

Since the most activated features are extracted, the local updates become more distinguishable after the amplification

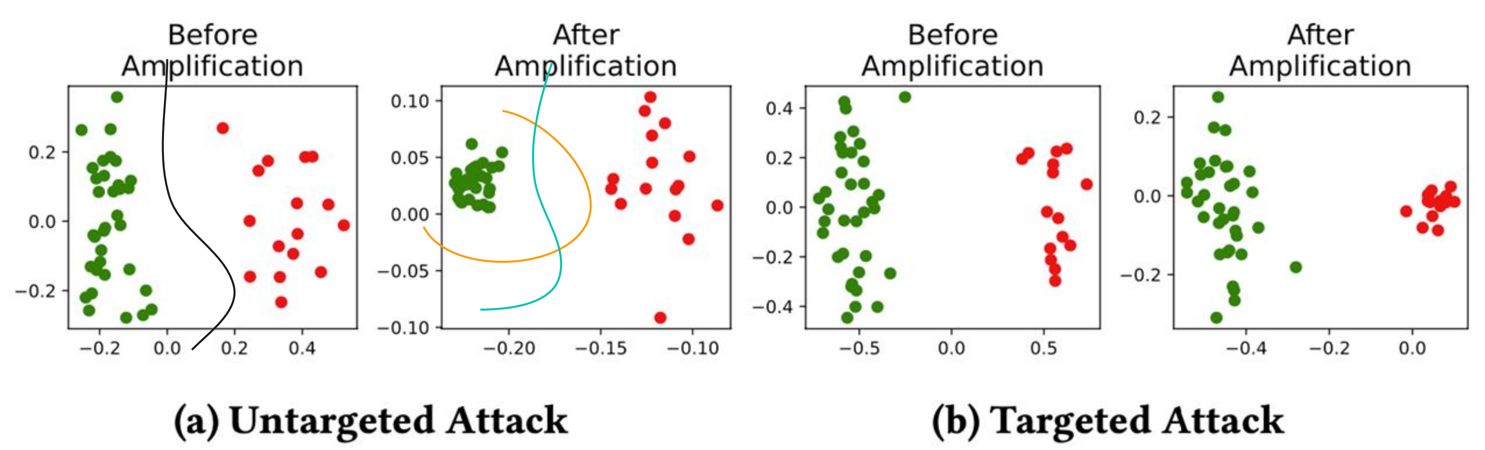
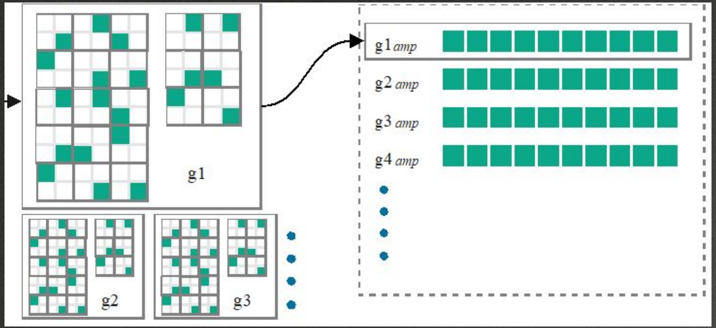


Figure: Gradients projected to 2-D surface using PCA. We plot 50 local updates at the 70th epoch of the training on the LOCATION30 dataset. Red dots (16/50) are malicious updates and green ones are benign. The untargeted attack is implemented via label flipping, and the targeted attack is the Scaling attack proposed by Cao et al.

Achieving Fidelity and Efficiency



Fidelity

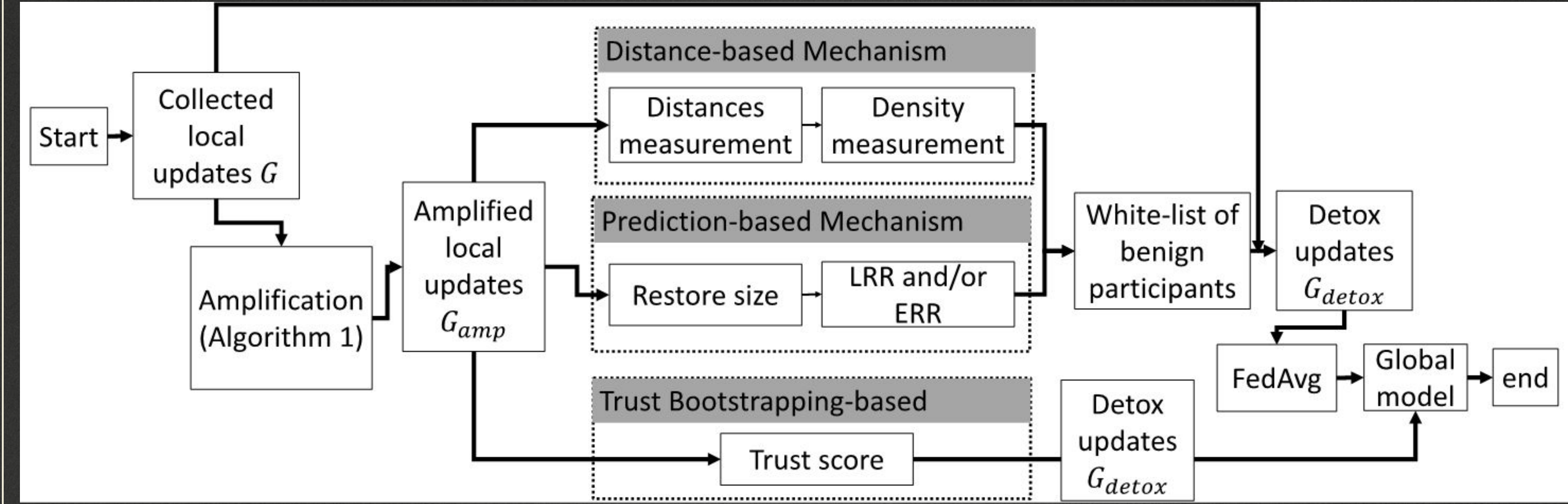
AgrAmplifier provides invariance to distortion. Similar to a max pooling layer in typical CNNs, it can ignore the small changes. This makes AgrAmplifier a noise canceller when no attack is present.

Efficiency

The significant dimension reduction in the feature space can greatly benefit the efficiency of AGR.

Equip AgrAmplifier

Universally compatible with existing AGRs regardless of the underlying aggregation rules.

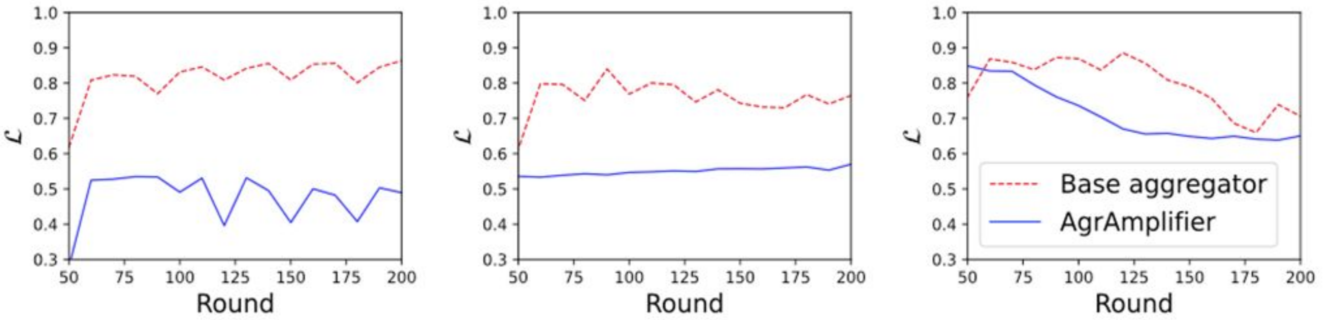


Evaluate AgrAmplifier

- Datasets
  - CIFAR-10, MNIST, Location30, Purchase100 (non-iid), Texas100 (non-iid) [1]
- Evaluated Attacks
  - Data poisoning via label-flipping (L-flip)
  - Model poisoning via gradient manipulation (G-asc)
  - <L-flip>+<G-asc>
  - Optimized and adaptive attack (S&H Attack) [2]
  - Scaling attack, Targeted (T-Scal) [3]
- Evaluated Defence
  - Distance-based: Cosine similarity / Euclidean distance / Merged distance combined with the Density measurement
  - Prediction-based: Fang defence [4]
  - Trust Bootstrapping-based: FLTrust [3]

[1] Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. 2017. Membership inference attacks against machine learning models. In 2017 IEEE Symposium on Security and Privacy (SP). IEEE, 3–18.
[2] Virat Shejwalkar and Amir Houmansadr. 2021. Manipulating the Byzantine: Optimizing Model Poisoning Attacks and Defenses for Federated Learning. Internet Society (2021), 18.
[3] Xiaoyu Cao, Minghong Fang, Jia Liu, and Neil Zhenqiang Gong. 2020. FLTrust: Byzantine-robust Federated Learning via Trust Bootstrapping. arXiv preprint arXiv:2012.13995 (2020).
[4] Minghong Fang, Xiaoyu Cao, Jinyuan Jia, and Neil Gong. 2020. Local model poisoning attacks to byzantine-robust federated learning. In 29th {USENIX} Security Symposium ({USENIX} Security 20). 1605–1622.

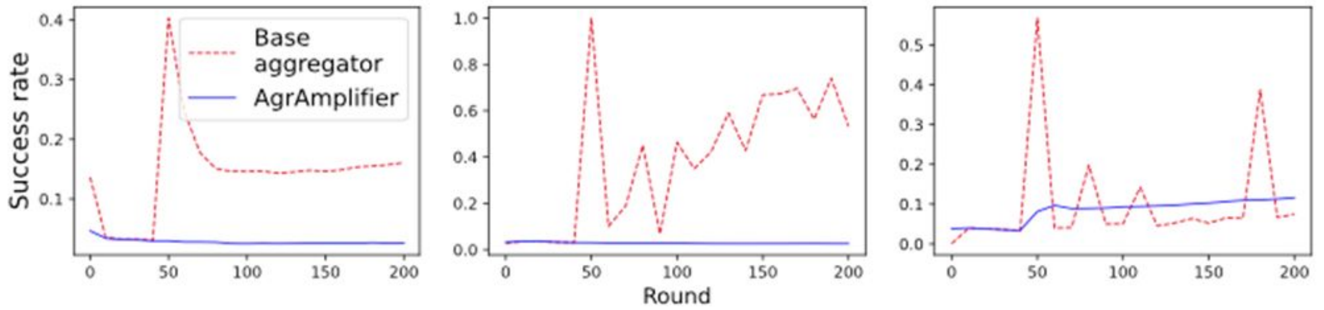
Robustness Performance against Untargeted Attack



(a) Distance-based **(b) Prediction-based** **(c) Trust bootstrapping-based**

L = [Non-attacked Test Accuracy] - [Attacked Test Accuracy]

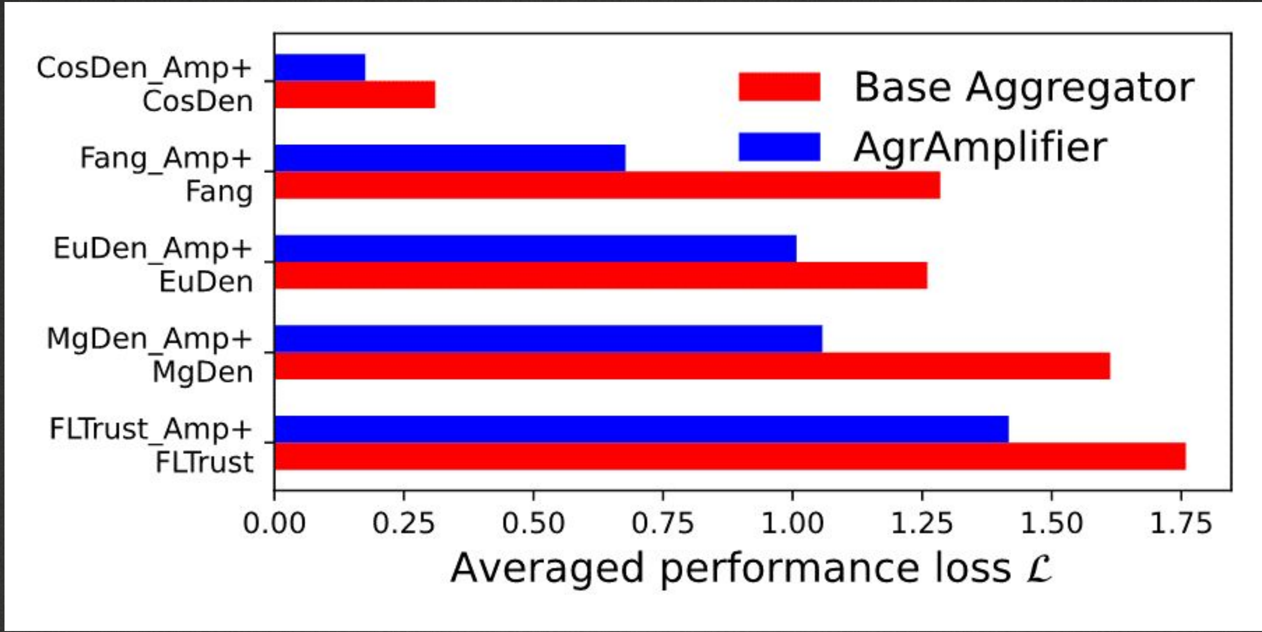
Robustness Performance against Targeted Attack



(a) Distance-based (b) Prediction-based (c) Trust bootstrapping-based

A = [attacked samples predicted as the attacker wish] - [other predictions]

Fidelity Performance



Efficiency Performance

Time consumption (in second) of aggregation

Hidden Layer	C	C+	F	F+	T	T+
512	3110.7	3038.6	3916.8	3822.6	3334.1	3283.6
1024	6722.9	6479.4	8612.4	7588.0	6672.6	6339.5
2048	12984.6	10367.1	17039.5	15883.0	13610.2	11084.4

† In this table, we use C to stand for CosDen, C+ for CosDen_Amp+, F for Fang, F+ for Fang_Amp+; T for FLTrust, T+ for FLTrust_Amp+

The difference is more obvious when the neural network architecture is larger, which benefits Byzantine-robust mechanisms towards learning on large/deep models.

Conclusion

- AgrAmplifier benefits Byzantine-robust mechanisms on all three properties
- We recommend COSDEN_Amp+ (distance based) as it achieves a desirable trade-off among the three. It performs the best in term of fidelity, and the joint best in robustness and efficiency

Thanks!

Does anyone have any questions?

liyue.shen@uqconnect.edu.au
UQ-TrustLab