INTERNATIONAL CONFERENCE ON PARALLEL PROCESSING
ICPP/2021/CHICAGO/USA
acm In-Cooperation
sighpc
AUGUST 9-12, 2021

FedCav: Contribution-aware Model Aggregation on Distributed Heterogeneous Data in Federated Learning

Hui Zeng[1], Tongqing Zhou[1], Yeting Guo[1], Zhiping Cai[1], Fang Liu[2]

[1] College of Computer, National University of Defense Technology

[2] School of Design, Hunan University

Outline

- Background
- Observation and Problem Statement
- Our Solution
- Experiments
- Conclusion

Background

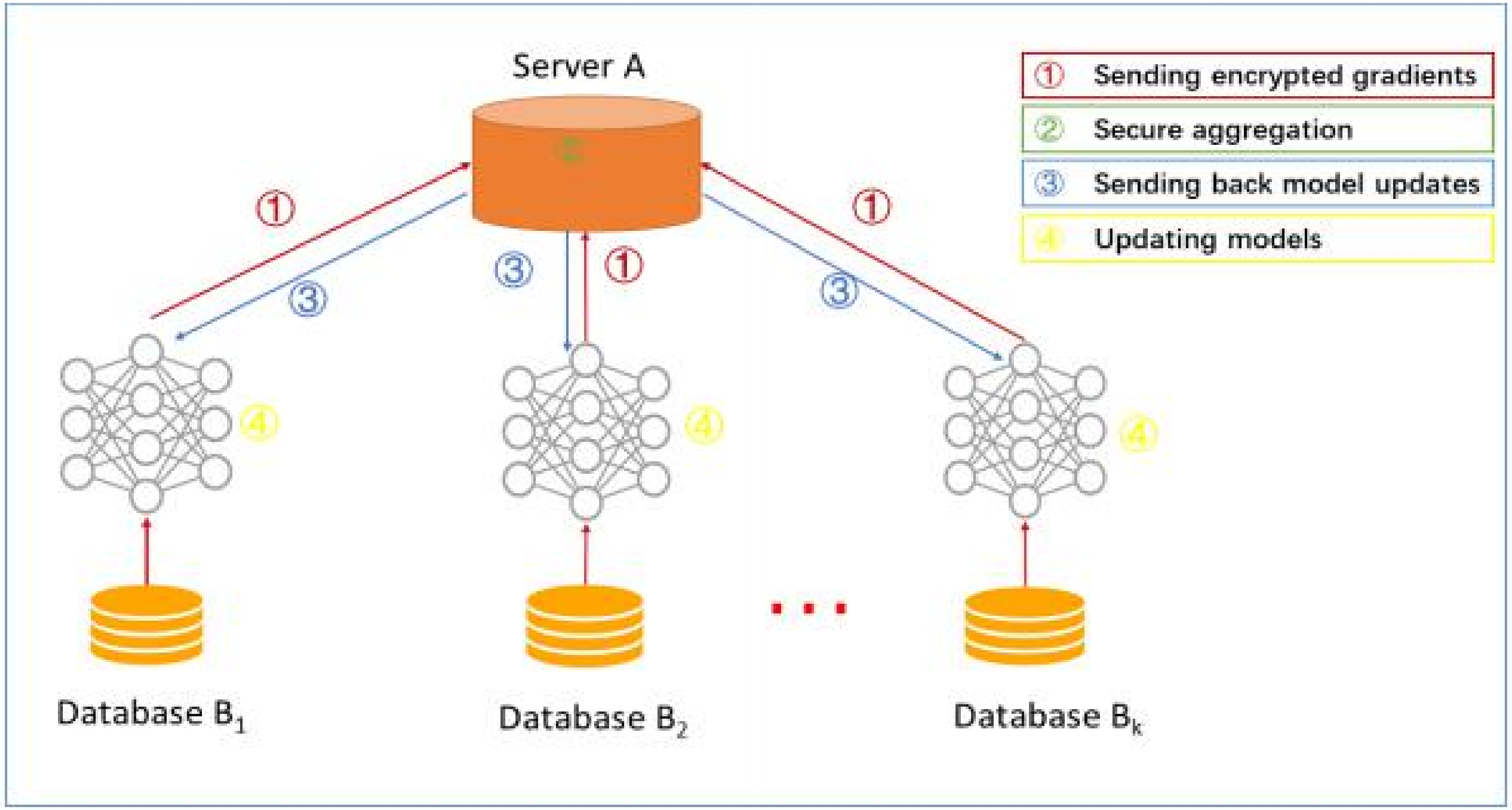
Two main problems in AI

data privacy

data island

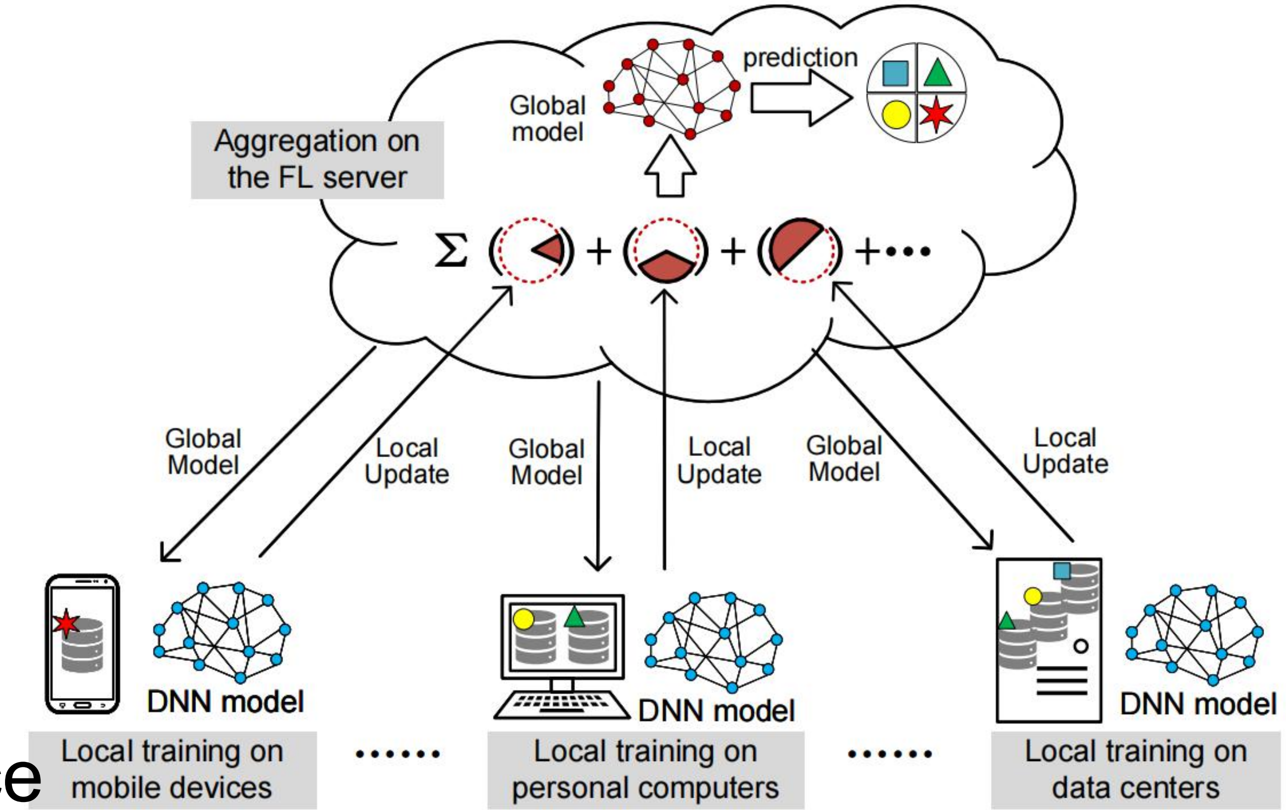
Federated Learning

- Federated Learning (FL) allows multiple distributed devices to cooperatively train models in parallel while preserving data privacy



Data Heterogeneity Problem in FedAvg

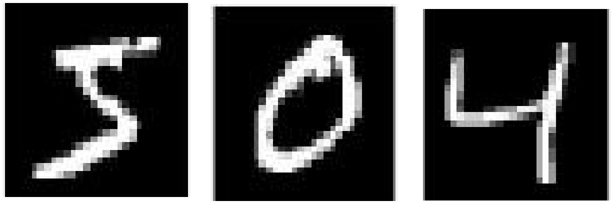
- Data Heterogeneity
  - Global non-IID
  - Local class imbalanced
- Problem
  - Slow convergence
  - Low training performance



Observations of FL with heterogeneous data

Setup -- Observation experiment

Dataset： MNIST



Training Model： LeNet-5

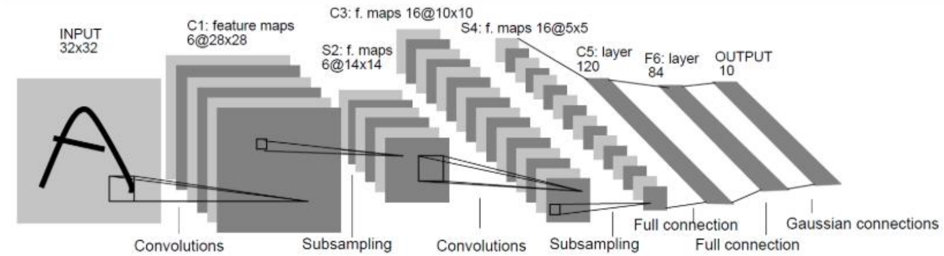
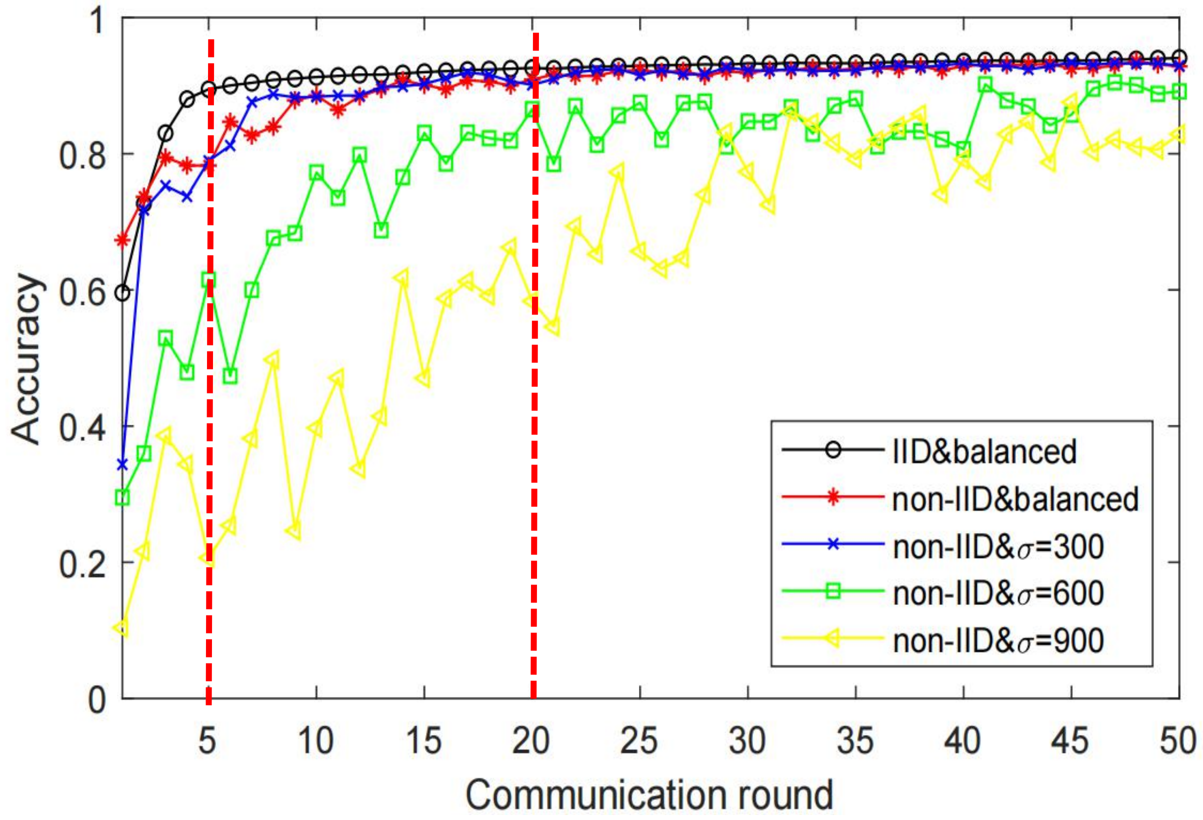


Table 1. Three different types of data distribution

Notations	Description
IID & balanced	Global IID and Local class balanced
non-IID & balanced	Global non-IID and Local class balanced
non-IID & imbalanced	Global non-IID and Local class imbalanced, the variance of each class is σ

Observation Results

- Results
  - Slow convergence
    - balanced: 5~10
    - imbalanced: 20~35
  - Low training performance
    - balanced: ~95%
    - imbalanced: 80% ~ 93%



Problem Statement

- Observation analysis

Heterogeneous Data → (FedAvg) → Slow convergence, Low training performance

↓

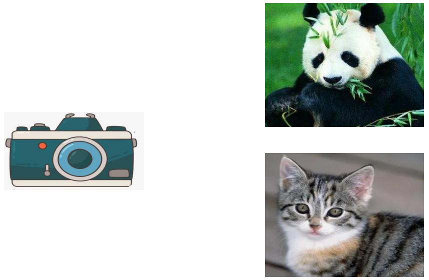
Data size = Contribution

- Reality

Data size ≠ Contribution

Which is more valuable?

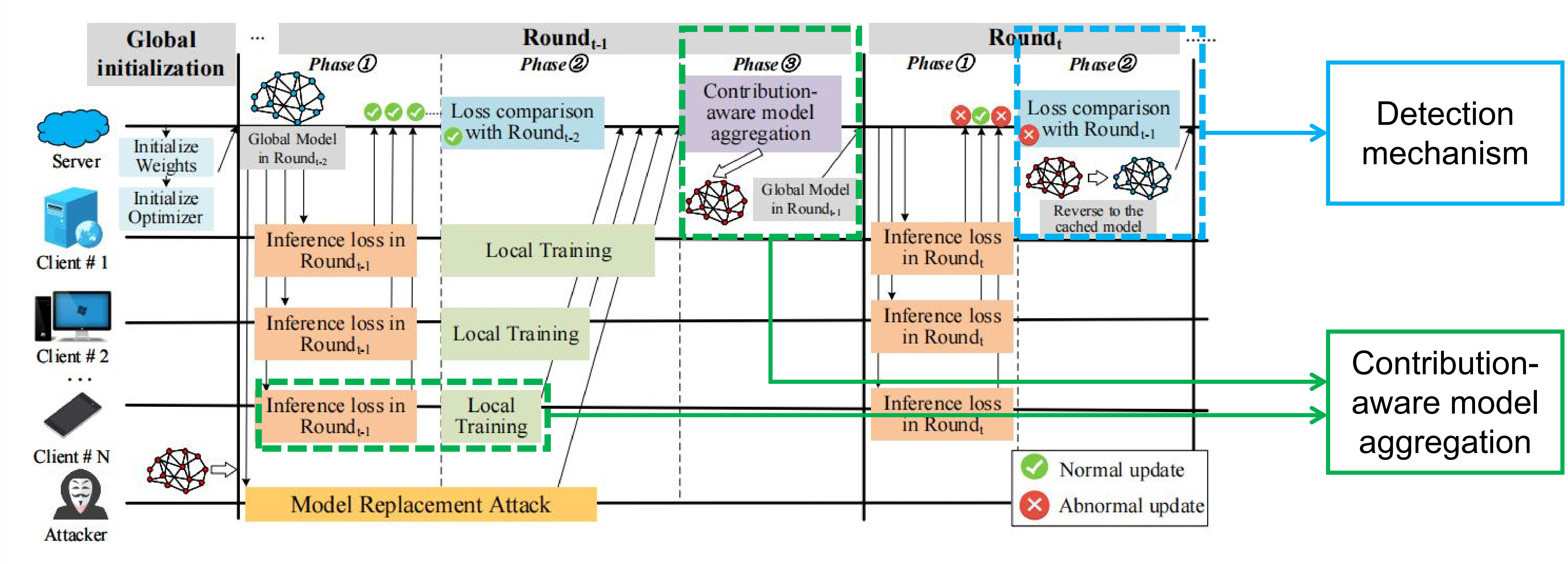
G-board



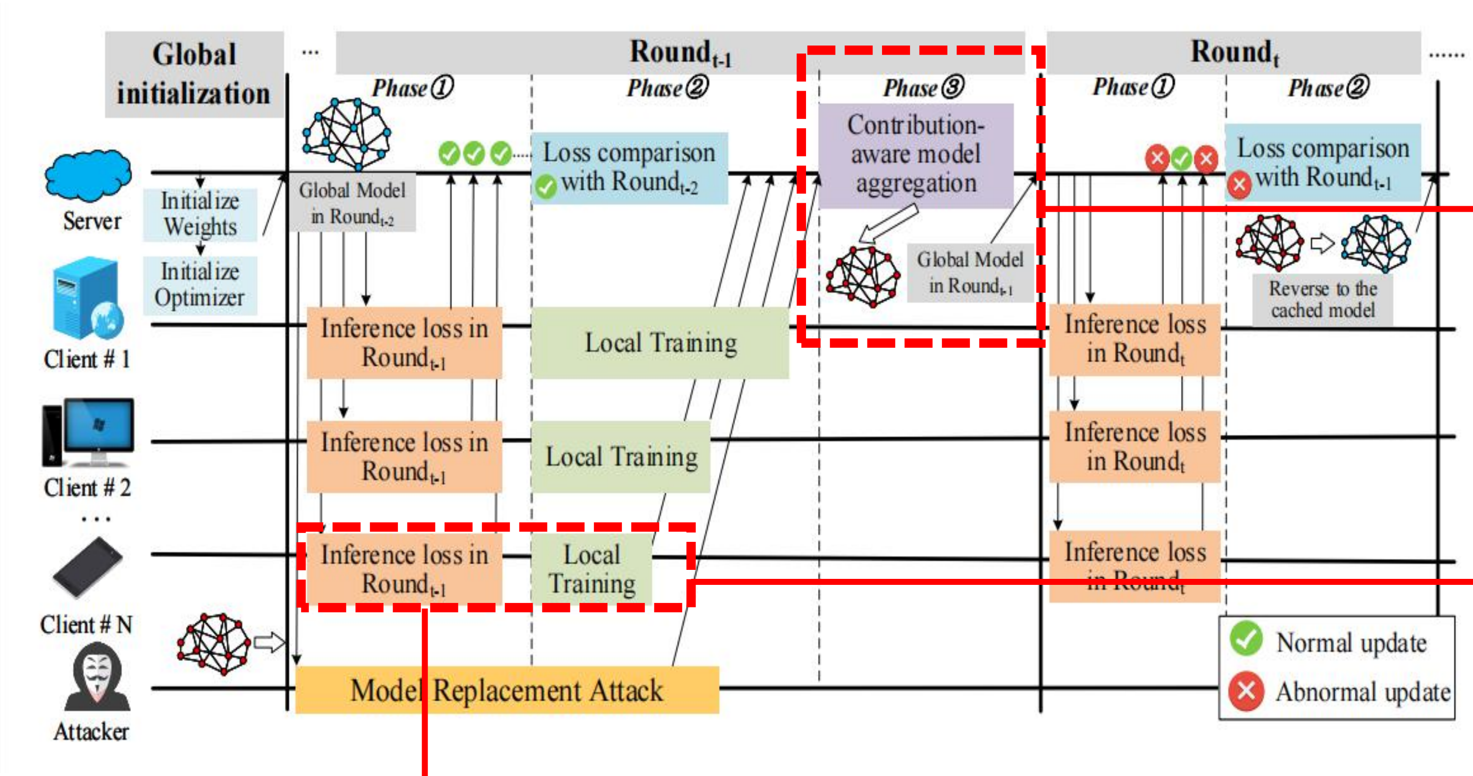
Animal pictures

Our Solution

FedCav workflow



Contribution-aware model aggregation



Algorithm 1: FedCav mdoel aggregation algorithm

Input : Number of clients n, sample ratio q.

Output : The expected optimization result of global model w_{opt}.

initialize global model w_0;
for *each communication round* $t = 1, 2, \ldots$ **do**
  $P_t \leftarrow$ client sets with random sampling ratio q;
  for *each client* $i \in P_t$ *in parallel* **do**
    $w^i_{t+1}, f_i(w_t) \leftarrow$ `LocalUpdate`(w_t) ;
  Clip the $f_i(w_t)$;
  $f_j(w_t) = min\{f_i(w_t), mean(f(w_t))\}$;
  $w_{t+1} \leftarrow \sum_i^{m_t} softmax\,[f_i(w_t)]\, w^i_{t+1}$;
$w_{opt} = w_{t+1}$;
return w_{opt}

$$w_{t+1} = \sum_i^n softmax[f_i(w_t)]\, w^i_{t+1}$$

Algorithm 2: Function **LocalUpdate**

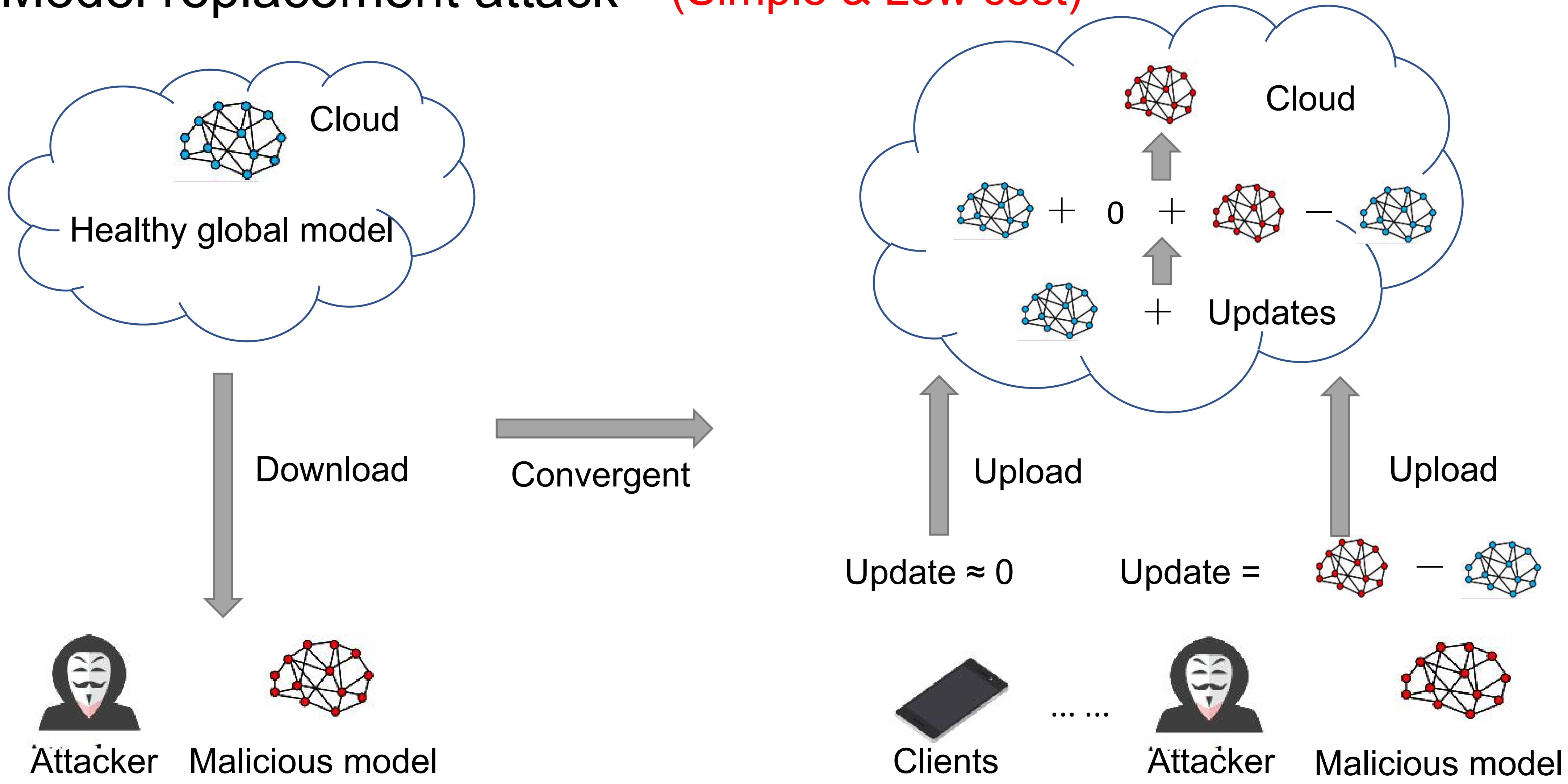
Input : Global model parameter w_t, local epochs E, local mini-batch size B.

Output : Local updated model w^i_{t+1}, inference loss $f_i(w_t)$.

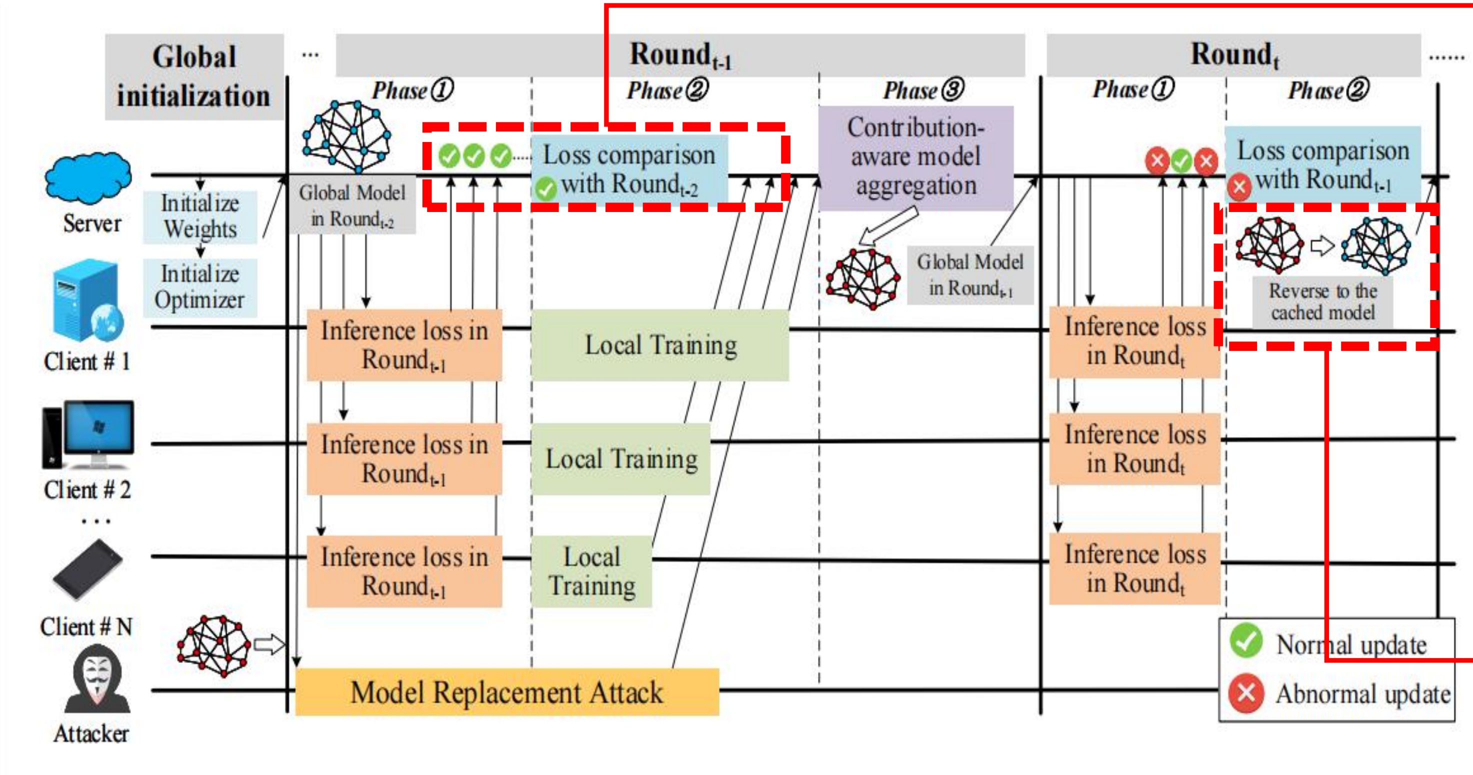
Compute the loss based on local data d_i;
$f_i(w_t, d_i) \leftarrow \ell(w_t, d_i)$;
initialize local model $w^i_t \leftarrow w_t$;
$\mathcal{B} \leftarrow$ (split d_i into batches of size B);
for *each local epoch* $e = 1, 2, \ldots, E$ **do**
  for *batch* $b \in \mathcal{B}$ **do**
    $w^i_{t+1} \leftarrow w^i_t - \eta \partial f_i(w^i_t, b)$;
return $w^i_{t+1}, f_i(w_t)$

Measure the potential contribution

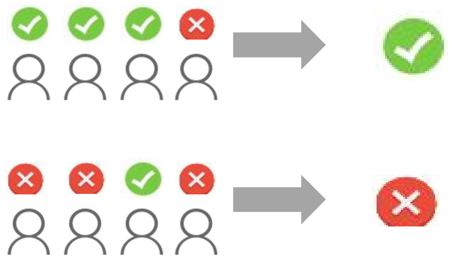
Model replacement attack (Simple & Low cost)



Detection mechanism



Comparing the inference loss with historical value.



The final decision is consistent with the majority

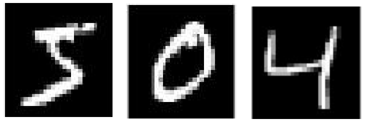
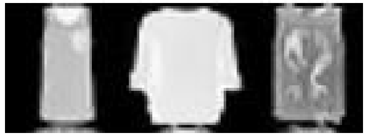
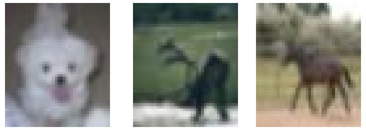
Reversing the current model to cached one.

INTERNATIONAL CONFERENCE ON PARALLEL PROCESSING

acm In-Cooperation sighpc

Experiment

Setup -- Dataset & Models

Dataset	Samples	Training Model
MNIST		LeNet-5[1]
FMNIST(Fashion-MNIST)		9-layers CNN
CIFAR-10		Resnet18[2]

[1]Yann Lecun, Leon Bottou, Y. Bengio, and Patrick Haffner. 1998. Gradient-Based Learning Applied to Document Recognition. Proc. IEEE (1998), 2278 – 2324.
[2]He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

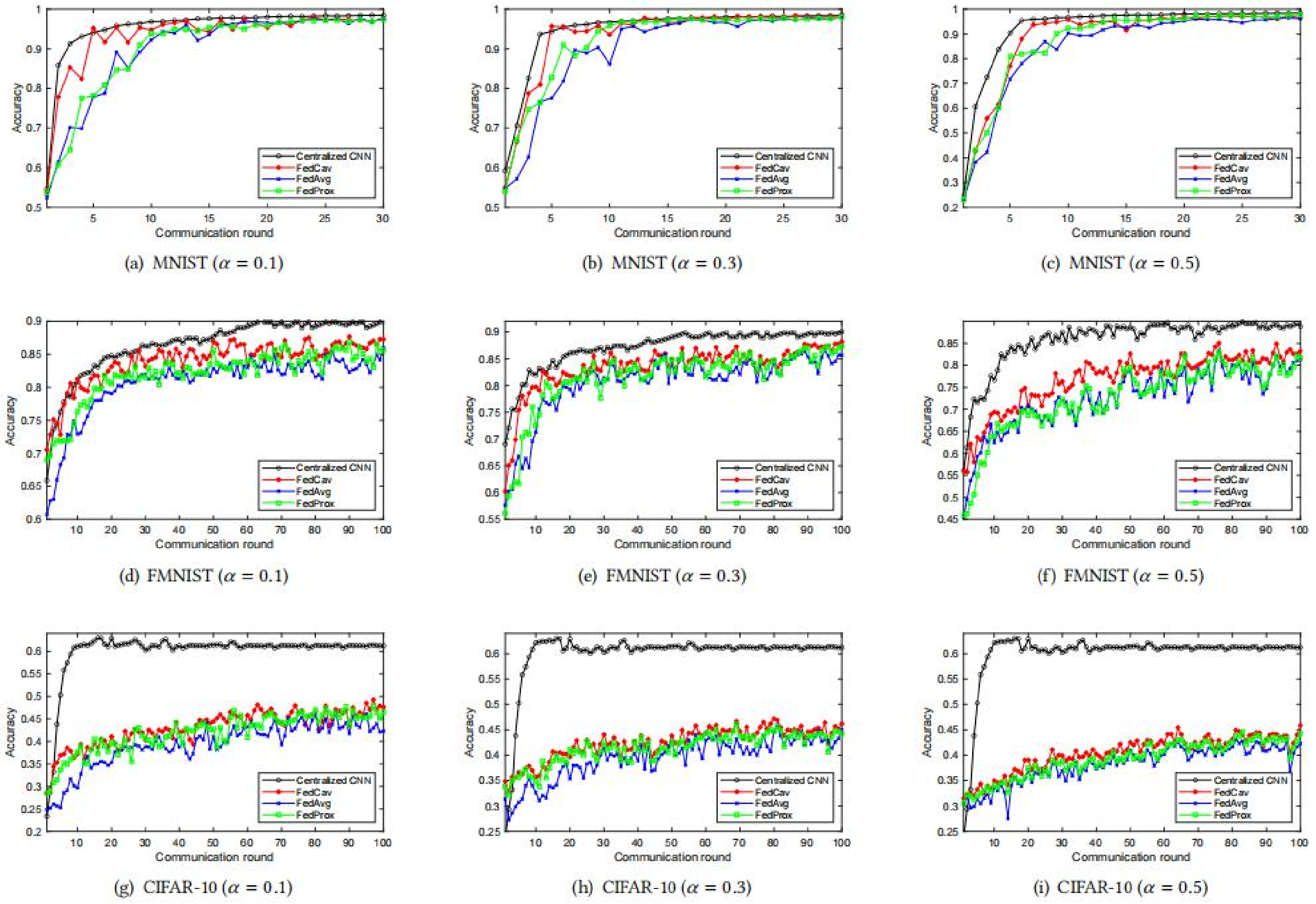
Results -- Classification performance

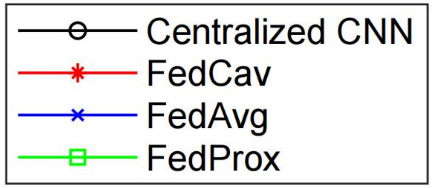
Table 4: Average classification accuracy under different levels of data heterogeneity (varying σ) on three datasets. Here we list the accuracy performance of different methods after the learning process gets convergence.

	$\sigma = 300$			$\sigma = 600$			$\sigma = 900$		
	FedAvg	FedProx	FedCav	FedAvg	FedProx	FedCav	FedAvg	FedProx	FedCav
MNIST	0.9333	**0.9391**	0.9365	0.9175	**0.9200**	**0.9200**	0.8467	0.8498	**0.8623**
FMNIST	0.8447	0.8459	**0.8621**	0.8111	0.8236	**0.8349**	0.7397	0.7716	**0.7913**
CIFAR-10	0.4612	0.4644	**0.4686**	0.4239	0.4254	**0.4287**	0.4003	0.424	**0.4387**

FedProx: Tian Li, Anit Kumar Sahu, Manzil Zaheer, et al. 2018. Federated optimization in heterogeneous networks. arXiv preprint arXiv:1812.06127 (2018).

Results -- Classification accuracy with dynamic environment

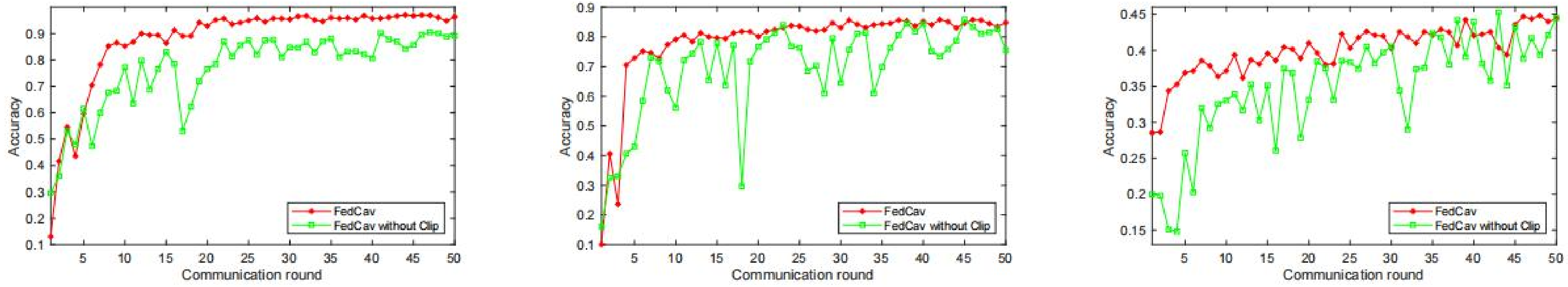




Fewer rounds to achieve a certain accuracy

Figure 4: Classification accuracy with dynamic data distribution adjustment controlled by factor α on three dataset. Results on different datasets are painted, wherein FedCav shows a generally stable and superior performance.

Results -- Impact of Clip strategy



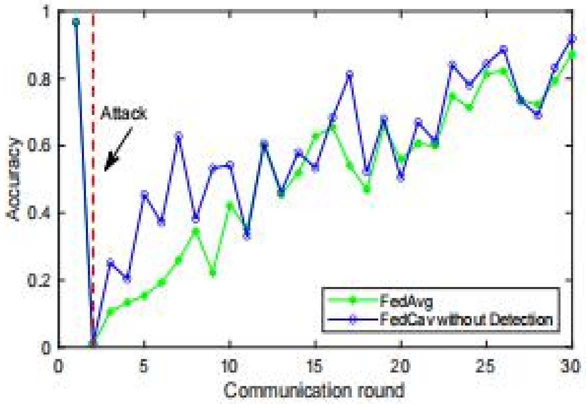
(a) Clip versus non-Clip on MNIST.

(b) Clip versus non-Clip on FMNIST.

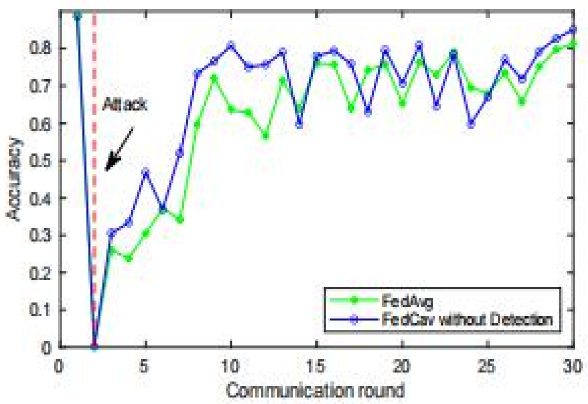
(c) Clip versus non-Clip on CIFAR-10.

Figure 5: Training process with four different algorithms on three datasets. Comparing the difference of whether it is necessar to use the Clip strategy. The curve shows that FedCav without Clip occurs great up-and-down oscillation.

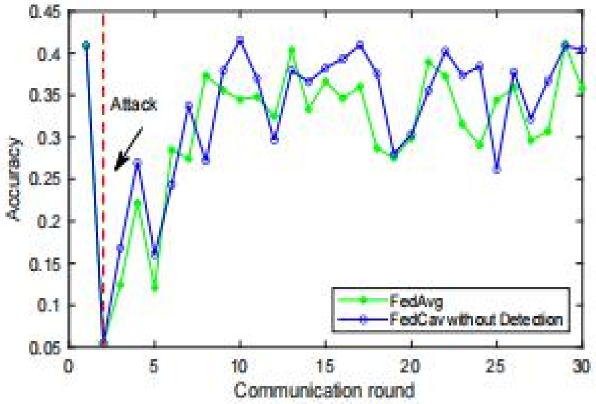
Results -- Impact of model replacement attack



(a) Model replacement attack on MNIST.



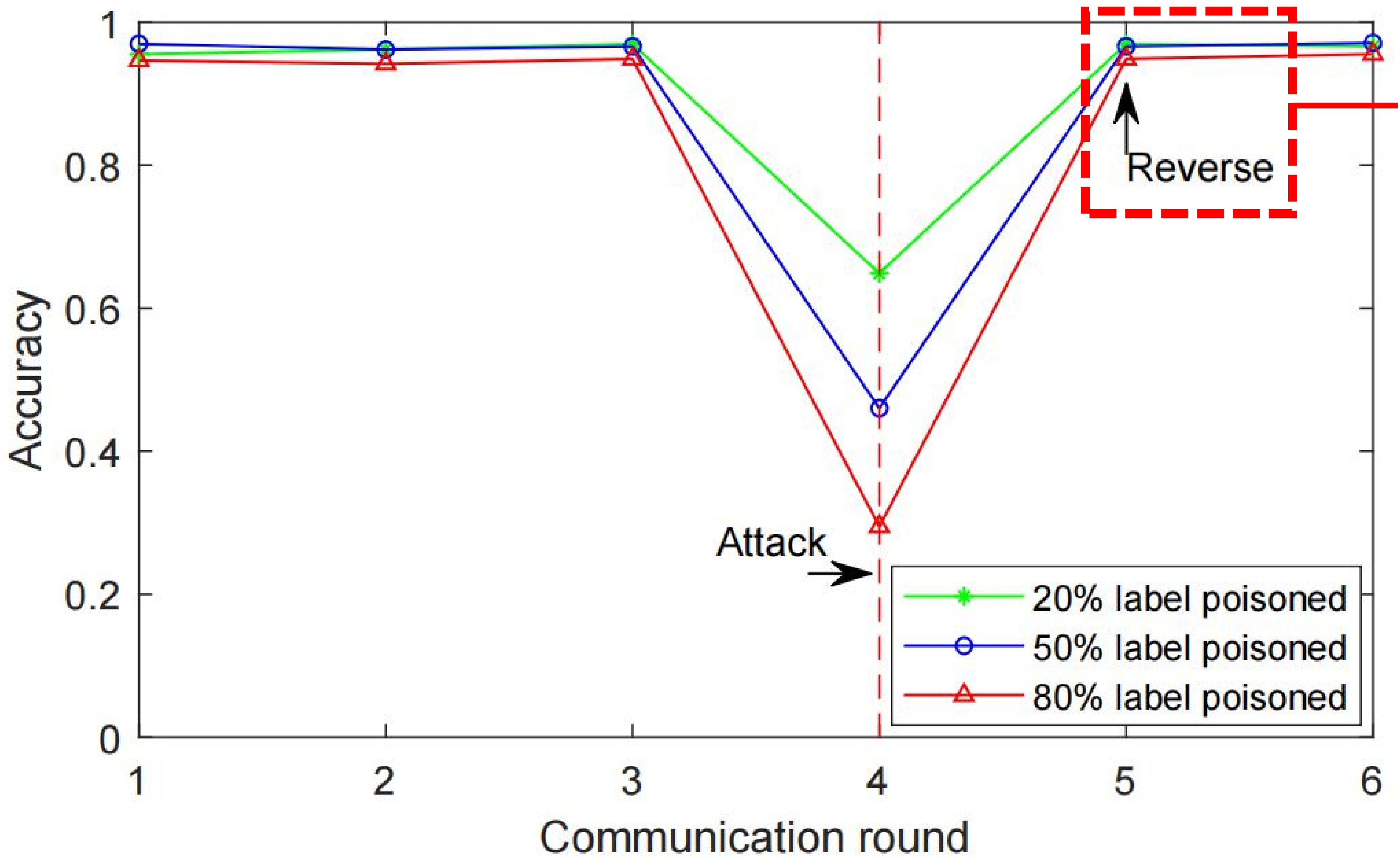
(b) Model replacement attack on FMNIST.



(c) Model replacement attack on CIFAR-10.

Figure 6: Part of training process of FedCav without detection and FedAvg after the model replacement attack on three datasets.

Results -- Reverse to cached one



Reverse the current model to cached one

Three different strengths of attacks

Conclusion

- Problem: Address the data heterogeneous problem in FL
- Observation: Heterogeneous data cause the slow convergent and low training accuracy.
- Key idea: Contribution-aware model aggregation. Aggregate with the contribution not the data size. Also consider malicious attack.
- Evalution: Higher classification accuracy (~2.4%) than baselines on average, fewer communication rounds (~34%) to achieve convergence

INTERNATIONAL CONFERENCE ON PARALLEL PROCESSING

ICPP/2021/CHICAGO/USA

acm In-Cooperation

sighpc

AUGUST 9-12, 2021

Thank you!

Hui Zeng[1], Tongqing Zhou[1], Yeting Guo[1], Zhiping Cai[1], Fang Liu[2]

[1] College of Computer, National University of Defense Technology

[2] School of Design, Hunan University

Email: zenghui116@nudt.edu.cn