Robust Collaborative Inference with Vertically Split Data Over Dynamic Device Environments

PURDUE UNIVERSITY

Elmore Family School of Electrical and Computer Engineering

Surojit Ganguli, Zeyu Zhou, Christopher G. Brinton, David I. Inouye

Background

- Collaborative Learning (CL) over vertically split data is a scenario where features for the same entity are split across multiple devices
- Some studies exist on dealing with arbitrary joining and dropping of devices; no investigation under CL consider both test time faults and absence of a special node
- **Research Question:** Can we develop cross-device decentralized collaborative learning (CL) algorithms that are robust and maintains strong performance at test time even under near-catastrophic faults?

Use Cases

- Some real-world scenarios, where data is distributed and ability to handle dynamic changes in network is critical:
  - **Robustness:** Sensors observing different data modalities across an ocean or sea**,** may become unreliable given the harsh outdoor conditions. Thus, sensors may leave or join the network arbitrarily. System performance must be reliable even in such a scenario.
  - **Remote Monitoring:** Having a central server for surveilling a network of sensors in difficult to access harsh environments, such as the ocean, is challenging. Thus, there is a need for sensor network that operates without a central server.

Dynamic Network Context (DN-CL)

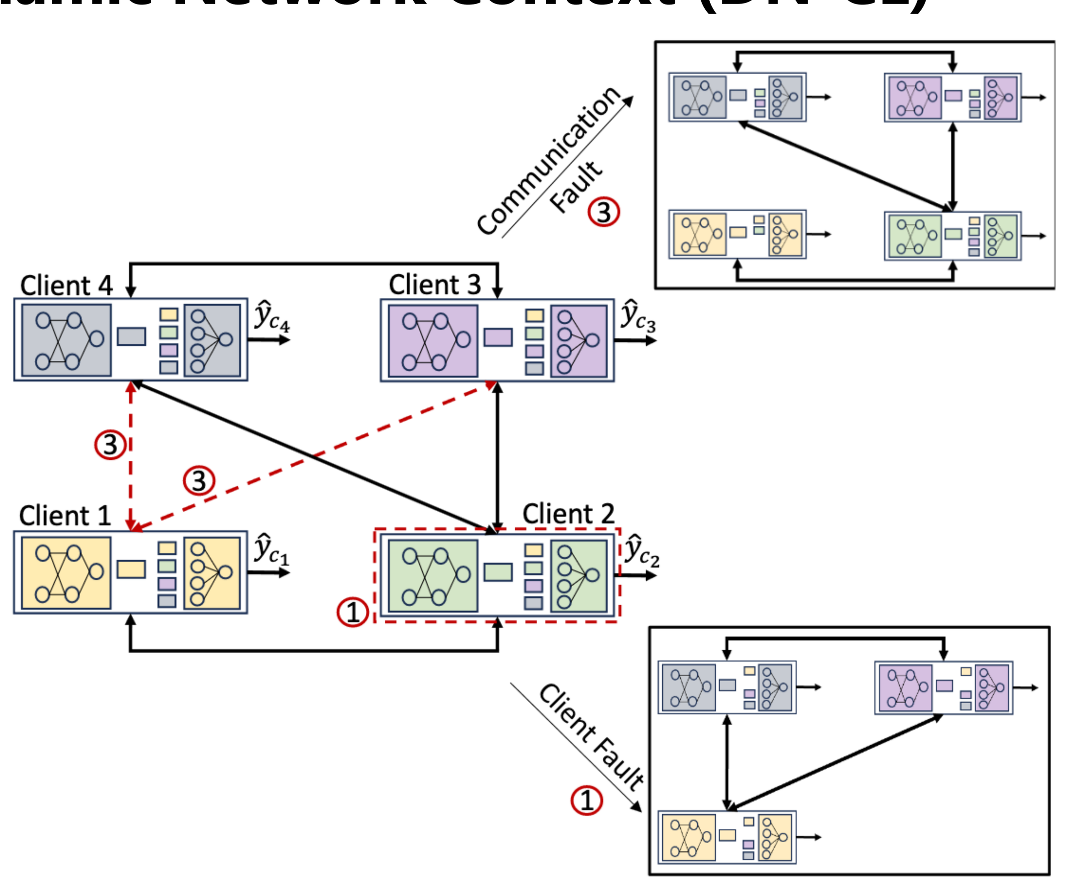
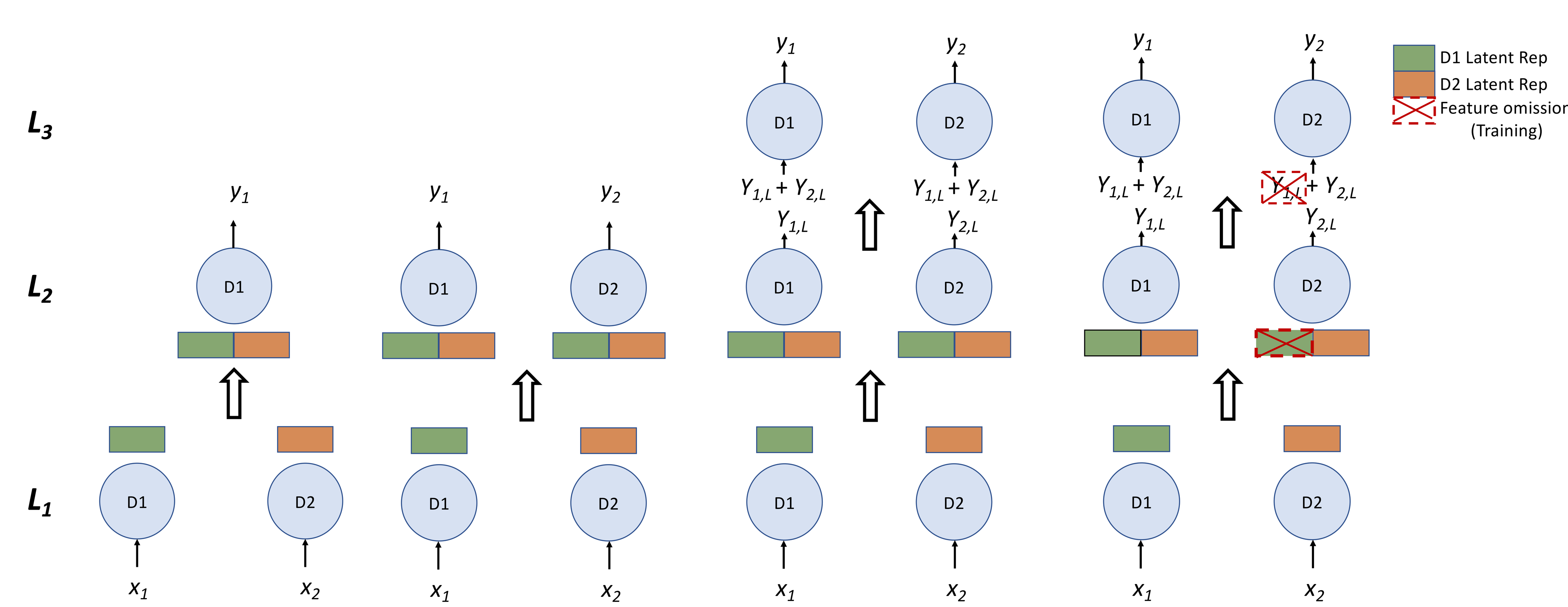


Figure: We study the vertically split data context such that the features are split across devices. We assume no centralized server node, and devices serve as data aggregators. Our goal is to obtain robust test time performance even under highly dynamic networks such as device faults ①, server faults ② and communication faults ③

Keys to **fault tolerant inference** over dynamic device network:

- *Training* with *simulated faults*
- *Replicating aggregators*
- *Ensembling predictions* via gossip

Methods



(a) CL (e.g., VFL) (b) MACL (c) MACL-G (d) CD-MACL-G

Figure (a) CL setup with D1 acting as a client as well as the aggregating server. (b) MACL arrangement has both the devices acting as data aggregators aside from being clients. (c) MACL-G is an extension of MACL wherein the output log probabilities are averaged for final prediction. (d) CD-MACL-G is a variant of MACL-G where during the training phase, representation from D2 to D1 is not communicated by design.

Theorem

Notation

r : fault rate , k : aggregator , $R(k)$: risk of the aggregator k , $\Psi()$: distributed inference algorithm , $R_{rnd}()$: risk of a random predictor , G : number of gossip rounds

Dynamic Risk Lower Bound

$$R_h(\theta;\mathcal{G}_r(t)) \geq (1-r^K)\cdot R_h(\theta;\mathcal{G}_{\text{base}}) + r^K \cdot R_{rnd}$$

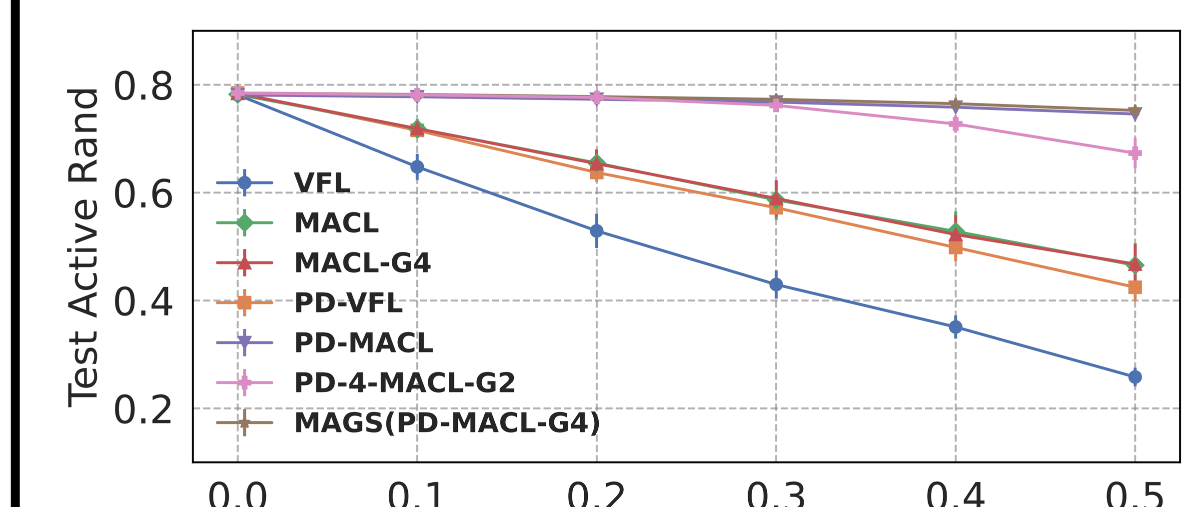
Equation 1: Multiple aggregators, K, reduce the chance of catastrophic failure of the system, even for high failure rate, r

Dynamic Risk Upper Bound

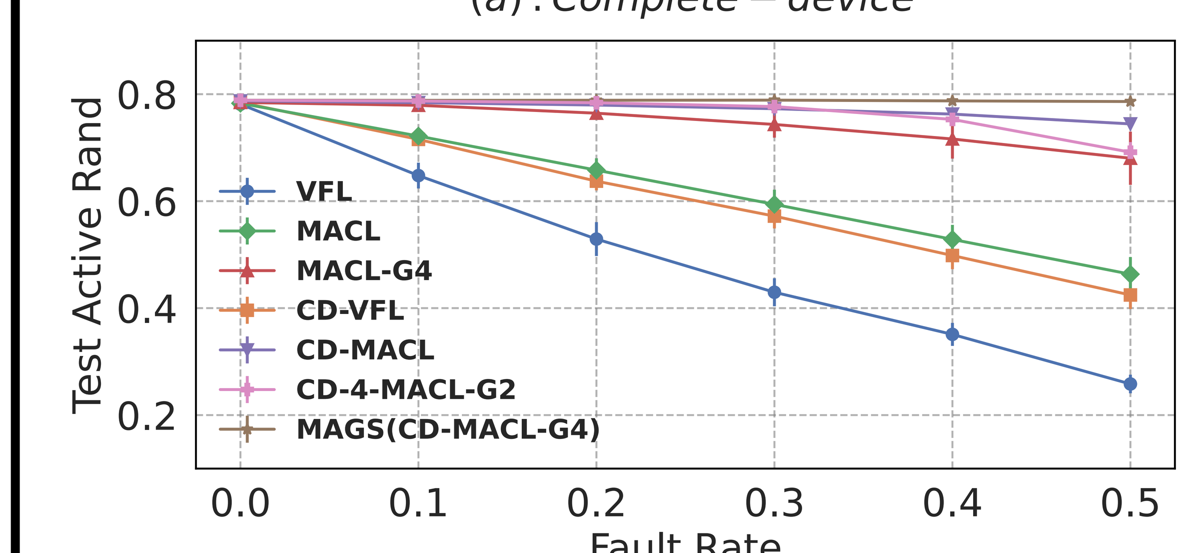
$$R_h(\theta;\mathcal{G}_r(t)) \leq r^K \cdot R_{rnd} + (1-r^K)\left(\tfrac{1}{K}\sum_k R(k) - \tfrac{1}{K}\sum_k \mathbb{E}[\ell(\Psi_k,\Psi^{ens})]\right) + (1-r^K)\sqrt{2K}\lambda^G \mathbb{E}\left[\max_{j,j'\in\mathcal{K}} \|\Psi_j - \Psi_{j'}\|_2\right]$$

Equation 2: Gossiping, to simulate model ensembling, leads to reducing the dynamic risk and variability among devices.

Results



$(a): Complete-device$



$(b): Complete-communication$

Figure: Test accuracy with and without comm. (CD-) dropout method. Across different fault rates, MACL-G4 (MAGS) trained with feature omissions has the highest average performance.

Acknowledgment

This work supported in part by ONR (N00014-23-C-1016), NSF (IIS-2212097) and ARL (W911NF-2020-221)