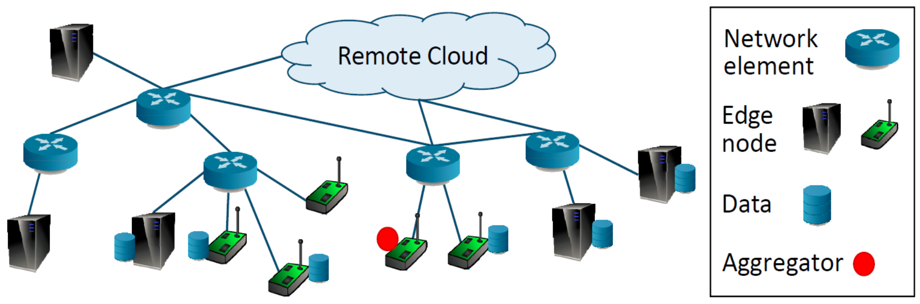
Distributed Machine Learning at Resource-Limited Tactical Edge:
Performance Bound and Control Algorithm

Shiqiang Wang (IBM US), **Tiffany Tuor** (Imperial College London), **Theodoros Salonidis** (IBM US),
Kin K. Leung (Imperial College London), **Christian Makaya** (IBM US), **Ting He** (PSU), **Kevin Chan** (ARL)

What is Distributed Learning?



- In tactical coalition applications, data is often collected and **distributed** at multiple nodes
- To analyze large amounts of data and obtain useful information for the detection, classification, and prediction of future events, **machine learning** techniques are usually applied
- Due to **bandwidth, storage**, and **privacy** constraints, it is impractical to send all the raw data to a centralized place

Distributed learning: Learning a model from data distributed at multiple nodes, without sending the raw data to a central place

The model has parameter vector $\mathbf{w}$
Each node i has a local loss function $F_i(\mathbf{w})$
The global loss function for *N* nodes is

$$F(\mathbf{w}) \triangleq \frac{\sum_{i=1}^{N} D_i F_i(\mathbf{w})}{D}$$

The learning problem is to find:

$$\mathbf{w}^* = \arg\min F(\mathbf{w})$$

Solution approach: Distributed gradient descent

- *Local update*: Update local parameter $\mathbf{w}_i$ using gradient descent on the local loss function $F_i(\mathbf{w})$ at each node i
- *Global aggregation*: Update the local parameter $\mathbf{w}_i$ at each node i to the weighted average of all nodes' parameters

Problem Formulation

Find the optimal trade-off between local update and global aggregation to minimize the learning loss function for a given resource budget

Formally:

- τ: # local updates between two aggregations
- T: total number of local updates
- R: total resource budget
- c: cost of a local update
- b: cost of a global aggregation

The problem is as follows:

$$\min_{\tau,T} \quad F(\mathbf{w}(T)) \quad \text{s.t.} \quad T\left(c+\frac{b}{\tau}\right) \leq R \qquad (1)$$

Challenge: How does τ and T impact $F(\mathbf{w}(T))$?

Solution

We prove the following upper bound:

$$F(\mathbf{w}(T)) - F(\mathbf{w}^*) \leq \frac{1}{T\left(\omega\eta\left(1-\frac{\beta\eta}{2}\right) - \frac{\rho h(\tau)}{\tau\varepsilon^2}\right)}$$

subject to some other constraints, where:

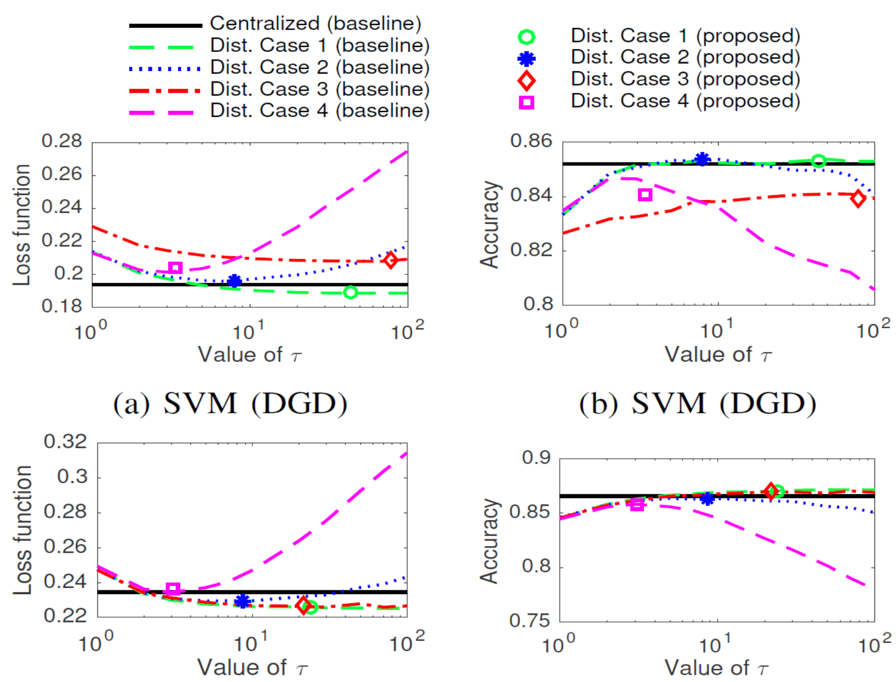
$$h(\tau) \triangleq \frac{\delta}{\beta}\left((\eta\beta+1)^\tau - 1\right) - \eta\delta\tau$$

δ: gradient deviation between local & global losses
β: smoothness parameter of loss function

- We use the above upper bound to approximate the obj. function in (1), and solve the problem with this approximation

The parameters c, b, δ, and β are estimated as part of a control algorithm in real time

Simulation Results



(a) SVM (DGD) (b) SVM (DGD)
(c) SVM (SGD) (d) SVM (SGD)