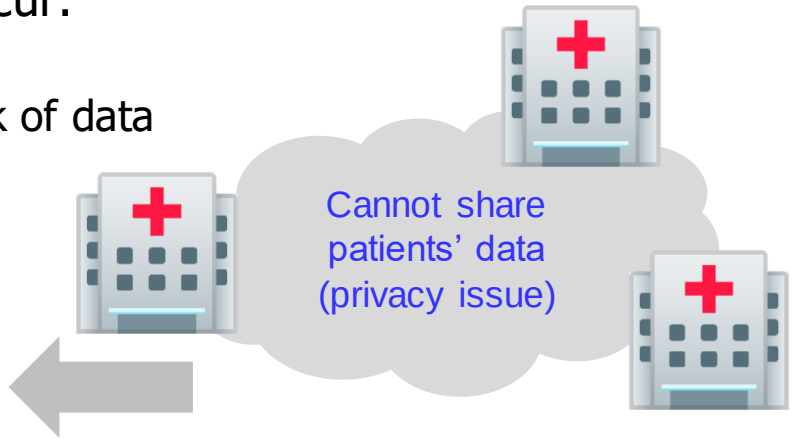
Privacy-Preserving Deep Learning Computation for Geo-Distributed Medical Big-Data Platforms

IEEE/IFIP Int'l Conf. on Dependable Systems and Networks (DSN) 2019

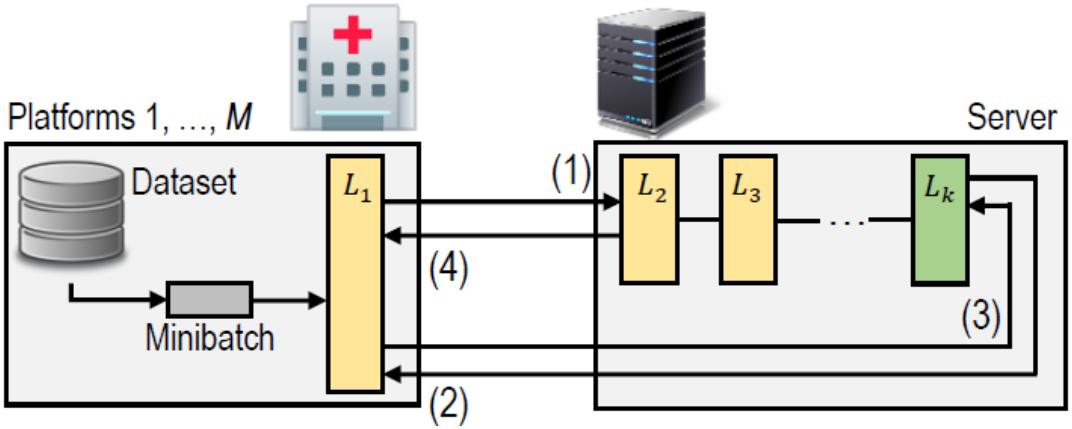
Joohyung Jeon, Junhui Kim, Joongheon Kim: Chung-Ang Univ., Seoul, Korea
Kwangsoo Kim: Seoul National Univ. Hospital, Seoul, Korea
Aziz Mohaisen: Univ. of Central Florida, Orlando, FL, USA
Jong-Kook Kim: Korea Univ., Seoul, Korea

- **Motivation**
  - It's not possible to gather all data in a single hospital/medical-cloud for deep learning computation (due to patients' privacy).
  - As a result, the following problems may occur:
    - **Overfitting** in each hospital
    - **Training Performance Degradation:** lack of data



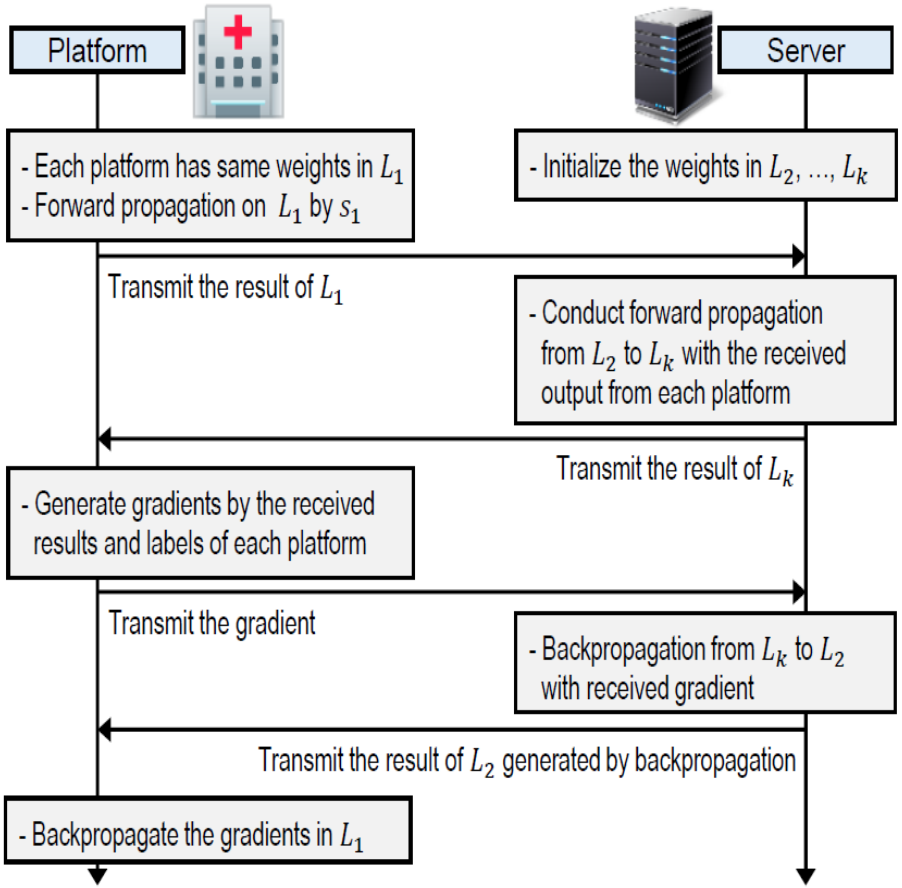
Goals
- Maintaining Deep Learning Computation Performance
- Eliminating Duplicated Patients' Data

Distributed Deep Learning Platform

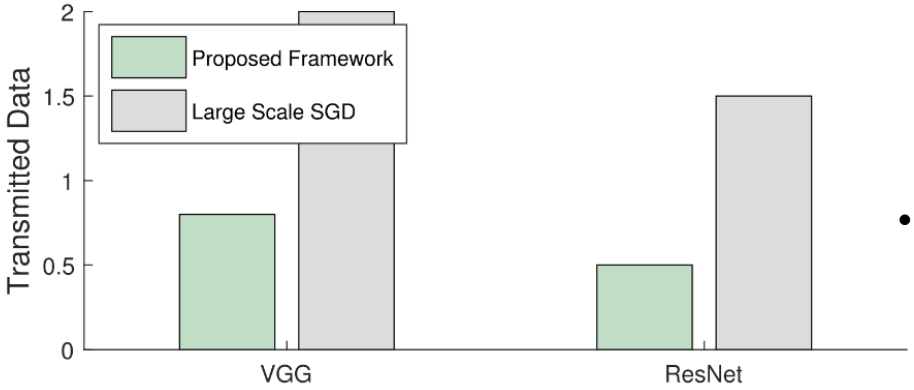


- **Privacy-Preserving Distributed Deep Learning Computation**
  - Each platform has the first hidden layer of deep learning model (L_1)
  - Server has the other hidden layers and the output layer (L_2, …, L_{k-1}, L_k)
  - During training process the data is shared in the form of the results of L_1

Distributed Deep Learning Platform (FlowChart)



Preliminary Result



- **Setup**
  - Data: CIFAR-10, CIFAR-100 data
  - Metric: communication overheads (lower is better).

- **Experimental Results**
  - Our proposed framework has low communication overheads compared to "Large-Scale SGD" (conventional distributed deep learning framework).
  - In addition,
    - Both of the proposed framework and Large-Scale SGD maintain same levels of accuracy, i.e., 95% in VGG, 75% in ResNet.
    - Our framework can preserve user-privacy in each medical platform during training

Summary and Future Work

- **Summary**
  - Distributed Deep learning Framework for **Privacy-Preserving** Computation.
  - How to? Based on the given deep neural network, the hidden layers are separated and then the first layer is left in each platform where the other layers are in a centralized server.
    - By doing this, the original/raw patients' data in each medical platform is not leaked during training, thus ensuring privacy.
    - Furthermore, utilizing the centralized server helps to improve learning performance by using all data from individual platforms during training.
- **Future Work**
  - Implementing this framework in geo-distributed hospitals (i.e., Seoul National University Hospitals) is anticipated.

Q&A

- More questions?
  - joongheon@gmail.com
  - joongheon@cau.ac.kr
  - mohaisen@ucf.edu