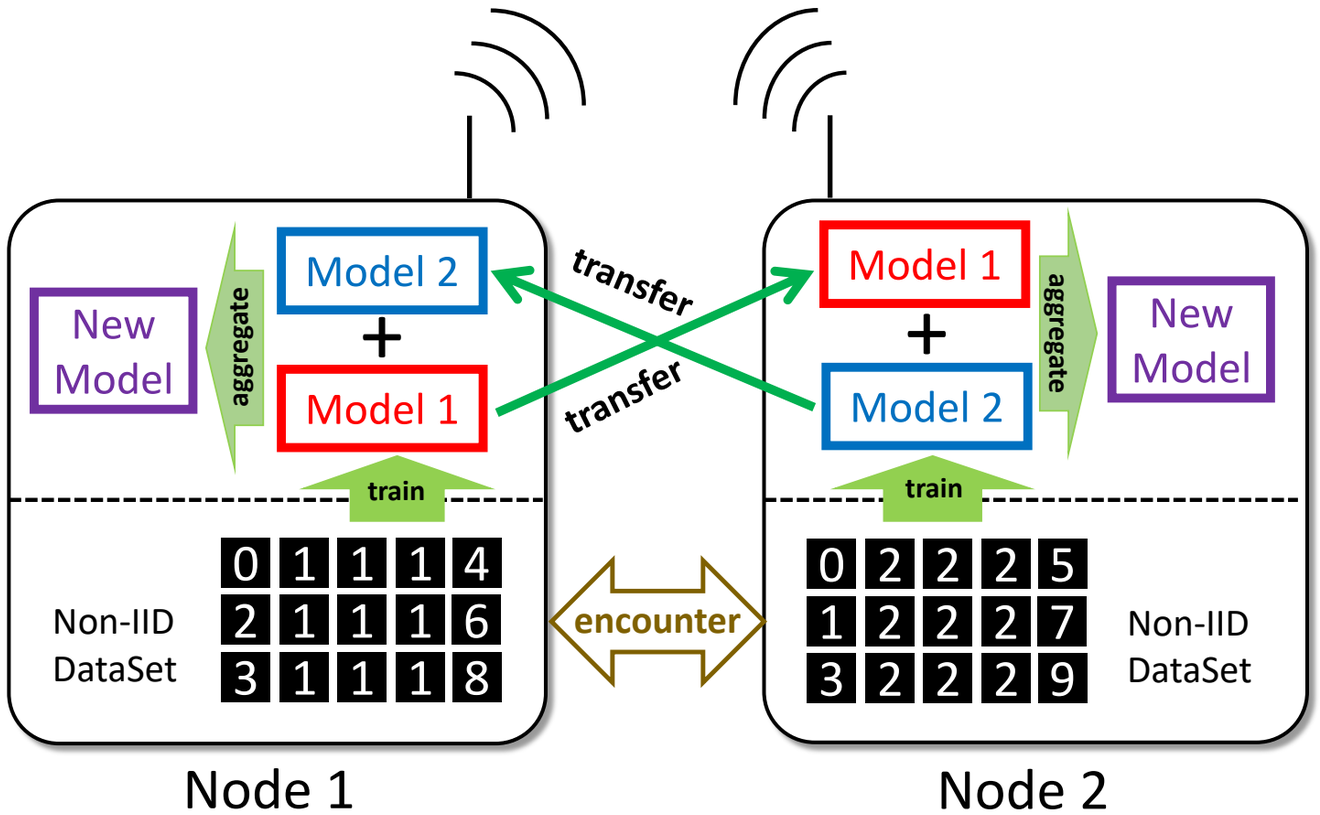
Wireless Ad Hoc Federated Learning: A Fully Distributed Cooperative Machine Learning



Associate Professor, Hideya Ochiai, Ph.D.,

The University of Tokyo, Japan

Table of Contents

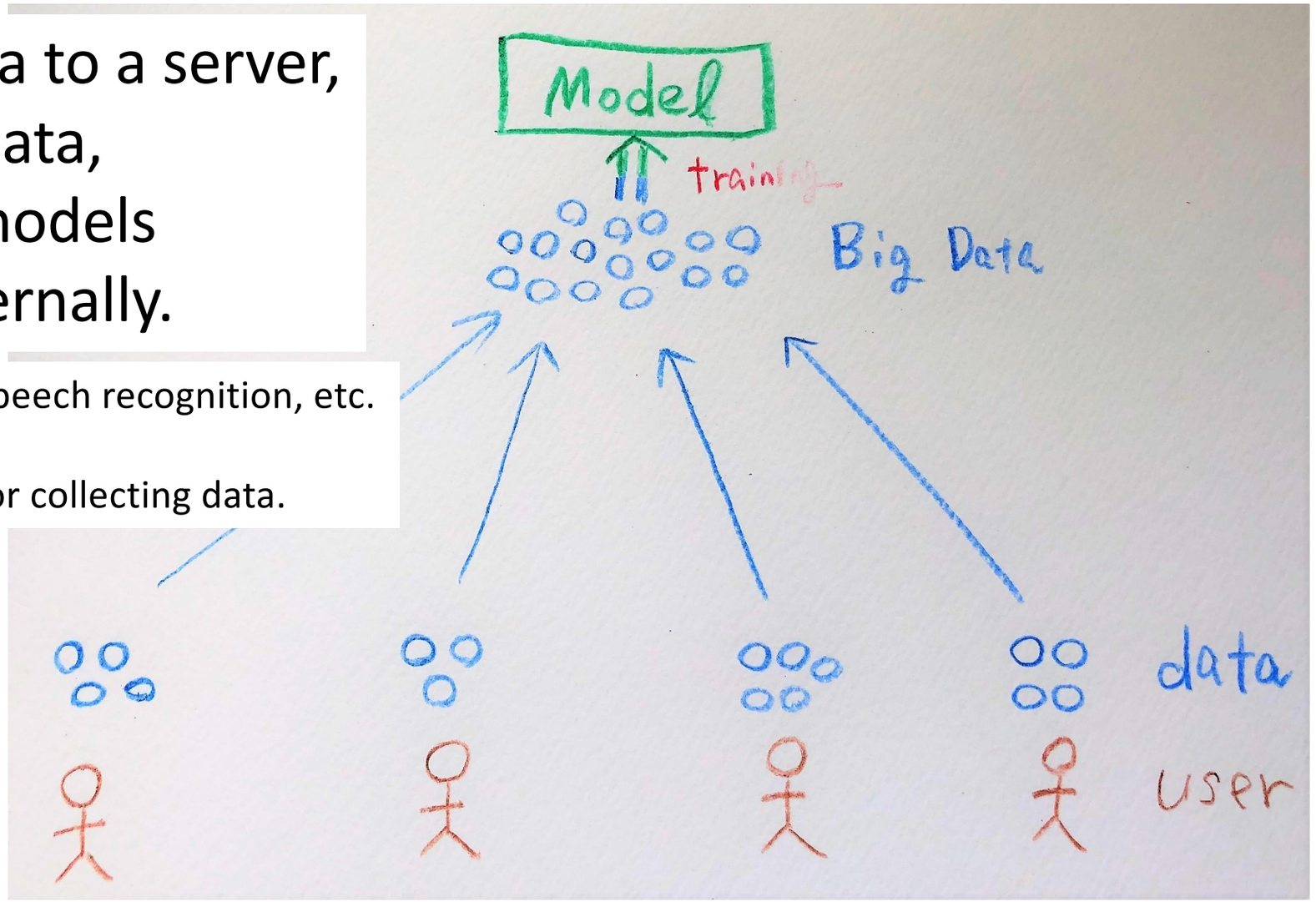
Table of Contents

<< Background >>

Machine Learning was Server-Centric and Data-Oriented.

1. They collected data to a server,
2. They formed Big Data,
3. They trained ML models
   in their server internally.

ML is good at Image processing, Speech recognition, etc.
Their interest was Data.
The Internet was just a platform for collecting data.



<< Background >>

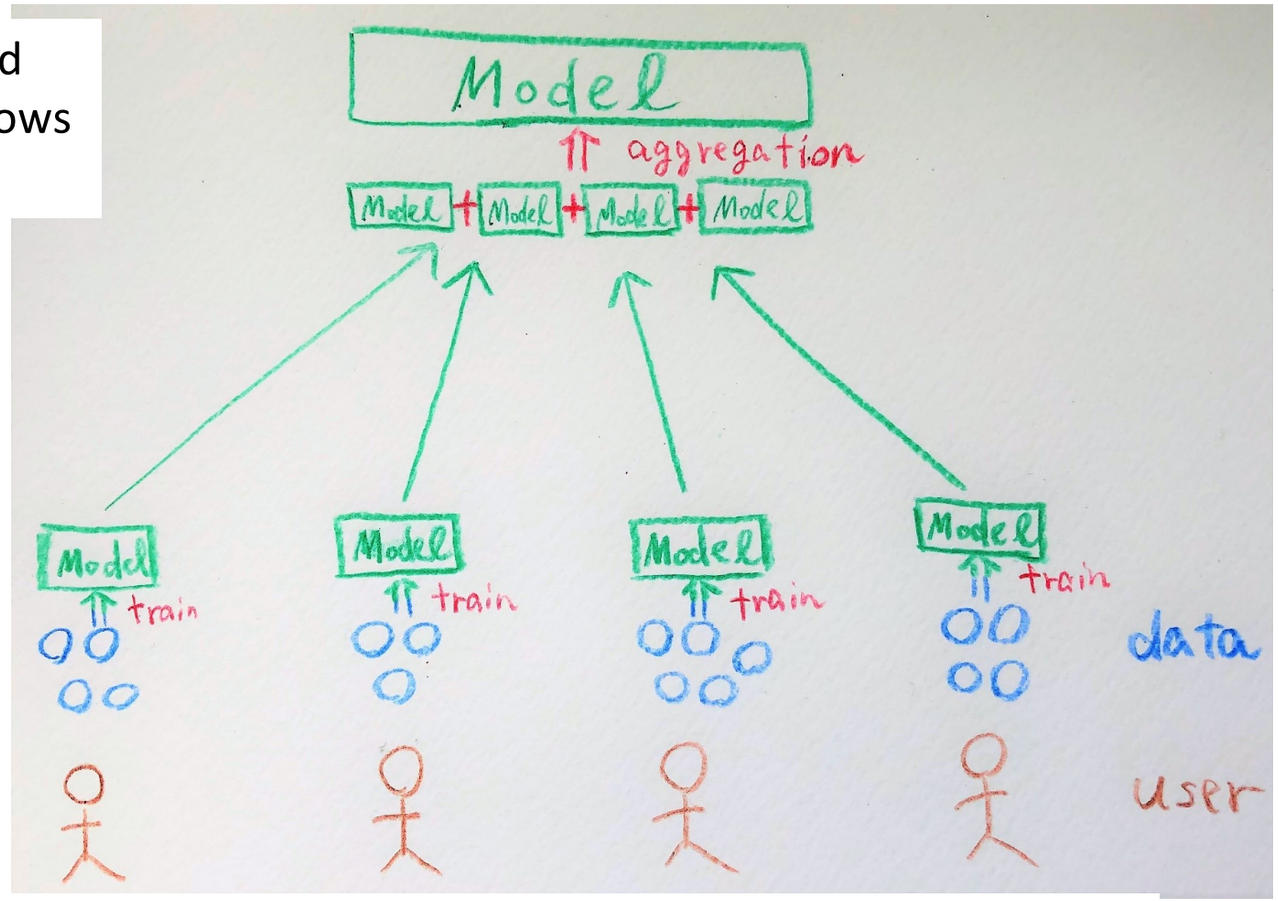
Federated Learning: Collaborative Machine Learning without Centralized Training Data
https://ai.googleblog.com/2017/04/federated-learning-collaborative.html

Emergence of Federated Learning (2017 – by Google, etc.)

Privacy Regulations (e.g., GDPR in Europe) motivated the emergence of Federated Learning (FL) which allows machine learning without collecting user data.

Point 2: FL can aggregate ML models

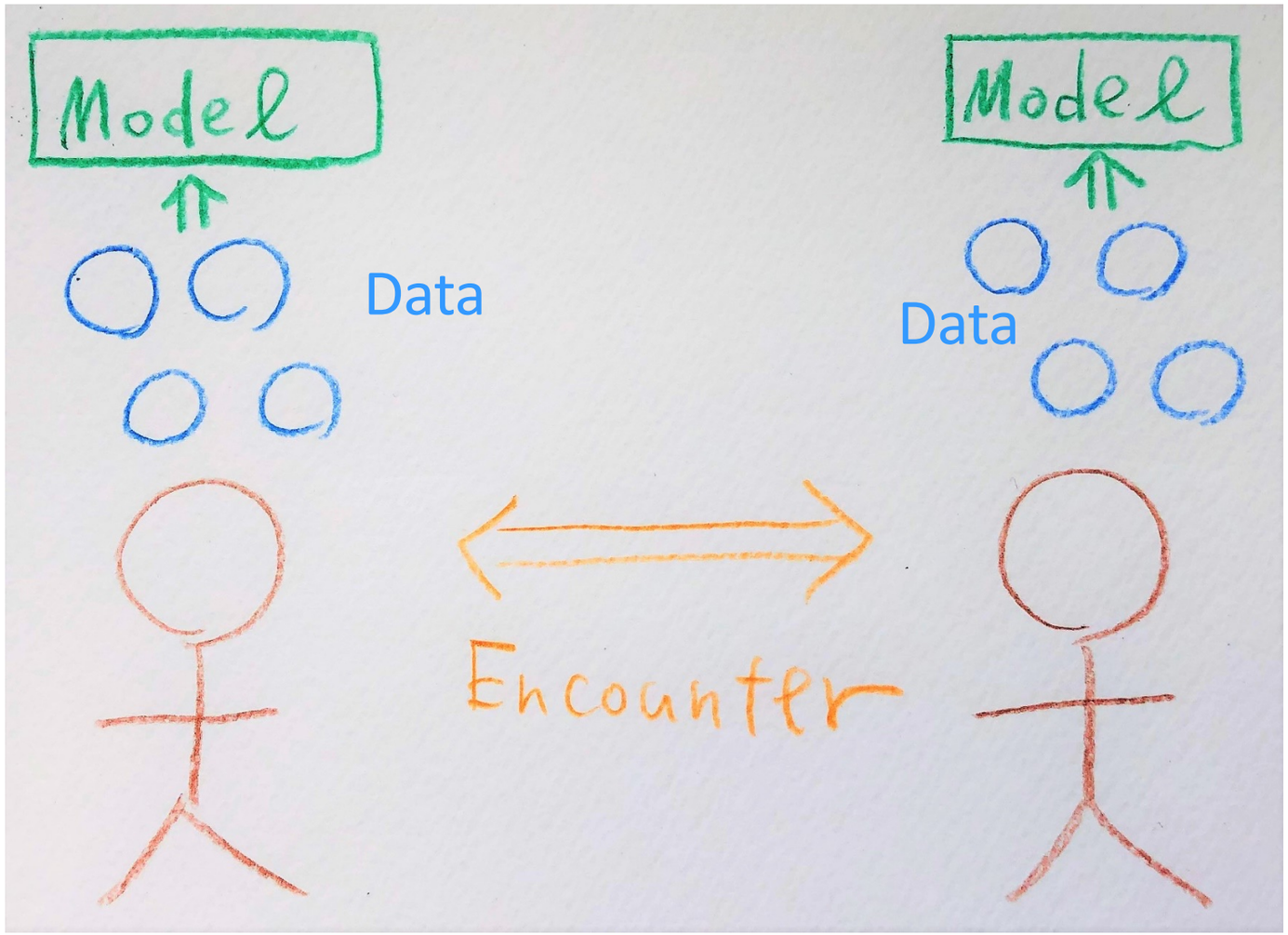
Point 1: without exchanging user data



However, this is still a Server-Client system!! (which is Centralized)

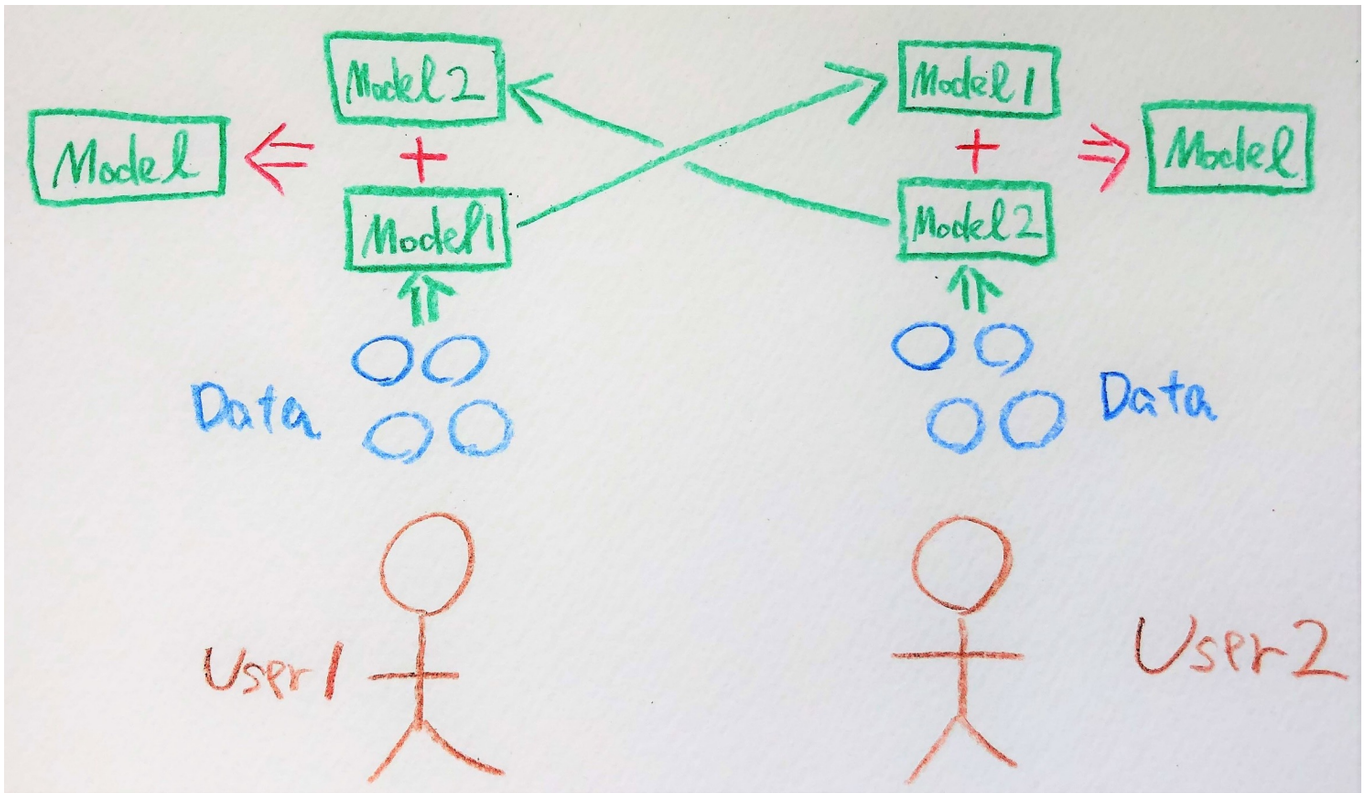
➔ Let's fully decentralize the system.

Toward Wireless Ad Hoc Federated Learning (WAFL)
Server-Client Architecture ➔ Peer-to-Peer Architecture



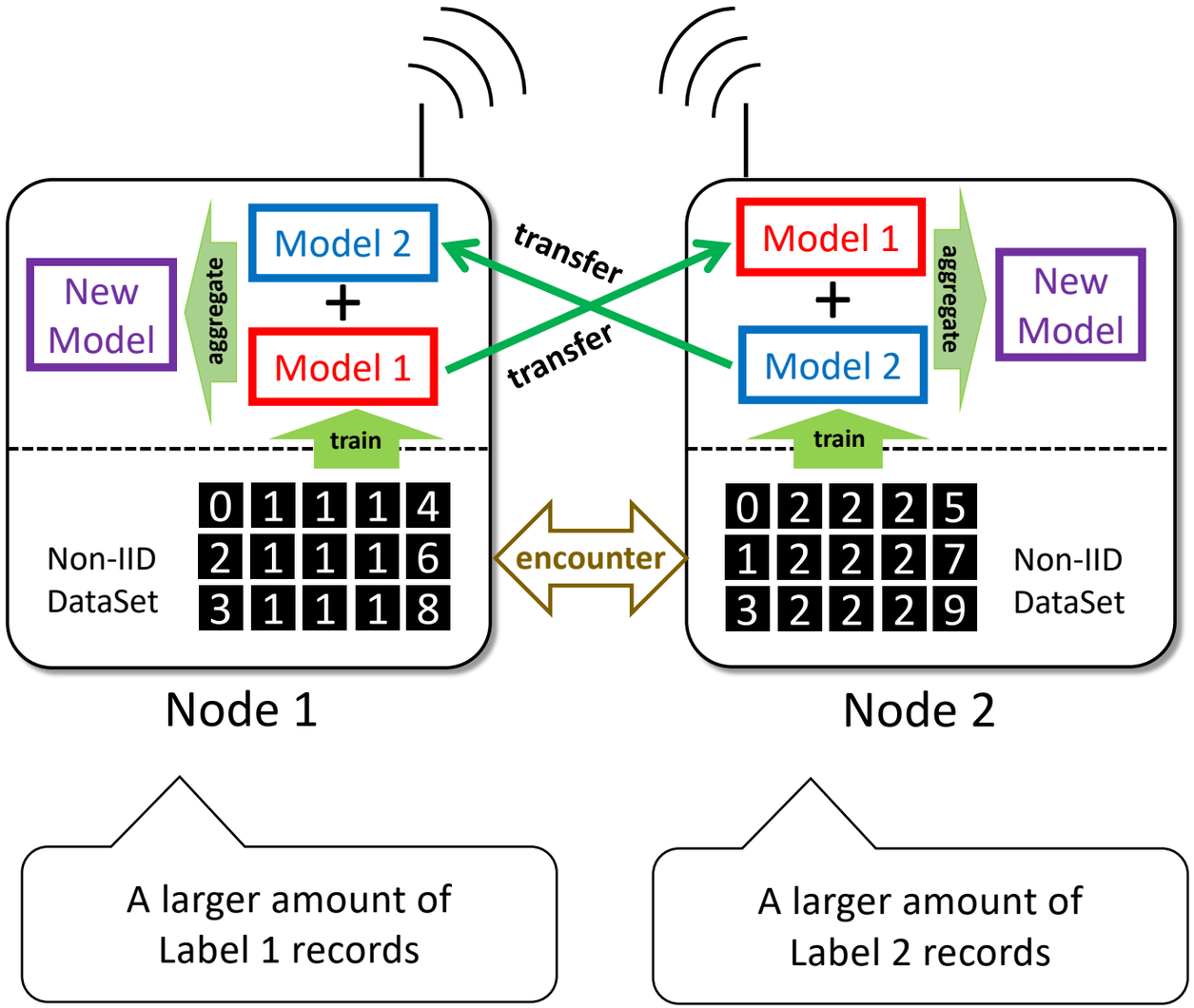
1. Each node individually trains its ML model using its local data.
2. Each node encounters the other.
3. They can communicate with local wireless communication media such as Wi-Fi Ad Hoc mode or Bluetooth

Toward Wireless Ad Hoc Federated Learning (WAFL)
Server-Client Architecture ➔ Peer-to-Peer Architecture



1. Each node individually trains its ML model using its local data.
2. Each node encounters the other.
3. They can communicate with local wireless communication media such as Wi-Fi Ad Hoc mode or Bluetooth
4. They exchange and aggregate the models to develop a new model.
5. This enables collaborative training.

Why nodes should collaborate?



1. Distribution of data on a node is not the same.
   e.g.,
   - Noodle lovers may have a lot of noodles photos.
   - Railway lovers may have a lot of railway photos.

   This is called Non-IID.
   (Not Independent and Identically Distributed)

2. If the training dataset is non-IID, the trained model will not be generalized for prediction.
   e.g., A noodle lovers device may recognize the photo of railway as noodle.

3. Mixture of the models will allow the generalization of model.

Hideya Ochiai et. al., “Wireless Ad Hoc Federated Learning – A Fully Distributed Cooperative Machine Learning”, arXiv, 2022.

Wireless Ad Hoc Federated Learning (WAFL)

Encountering many nodes leads to generalization of the model.

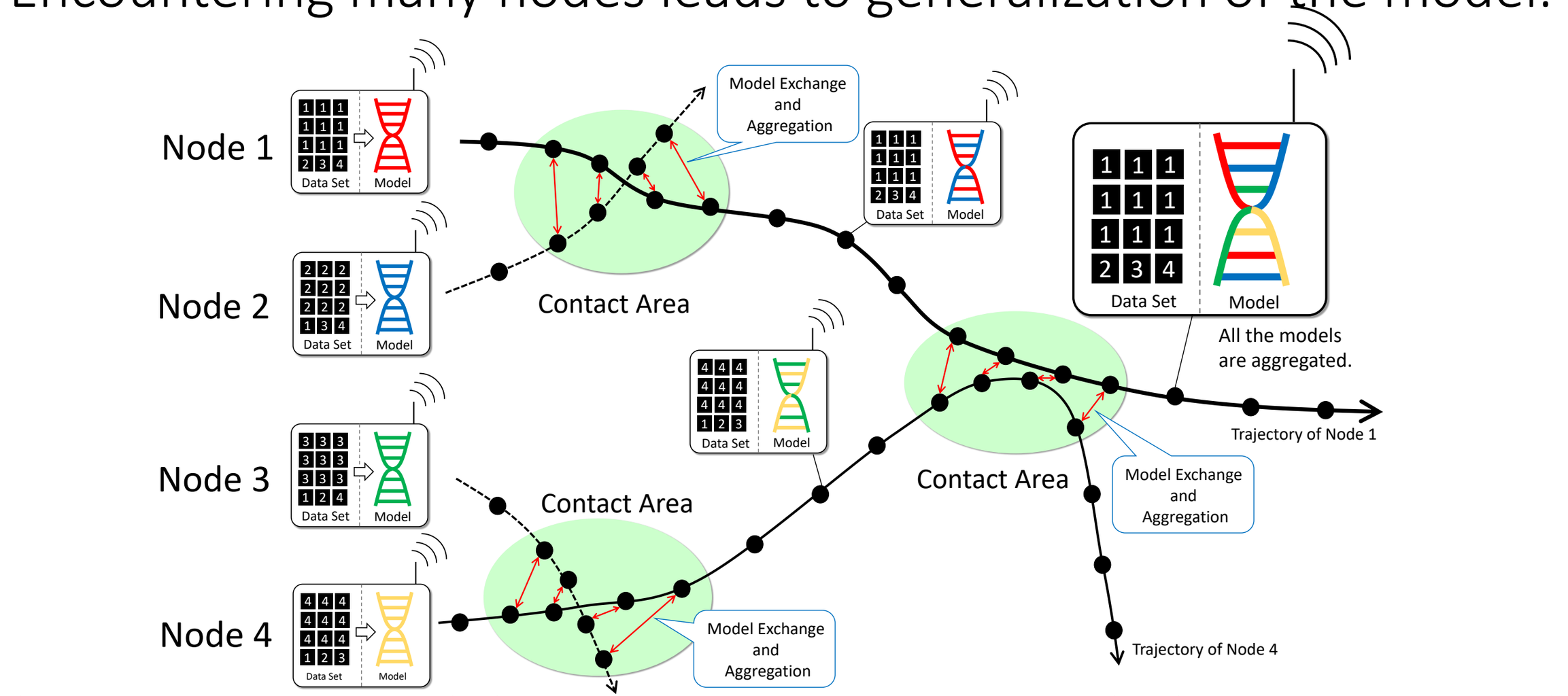


Table of Contents

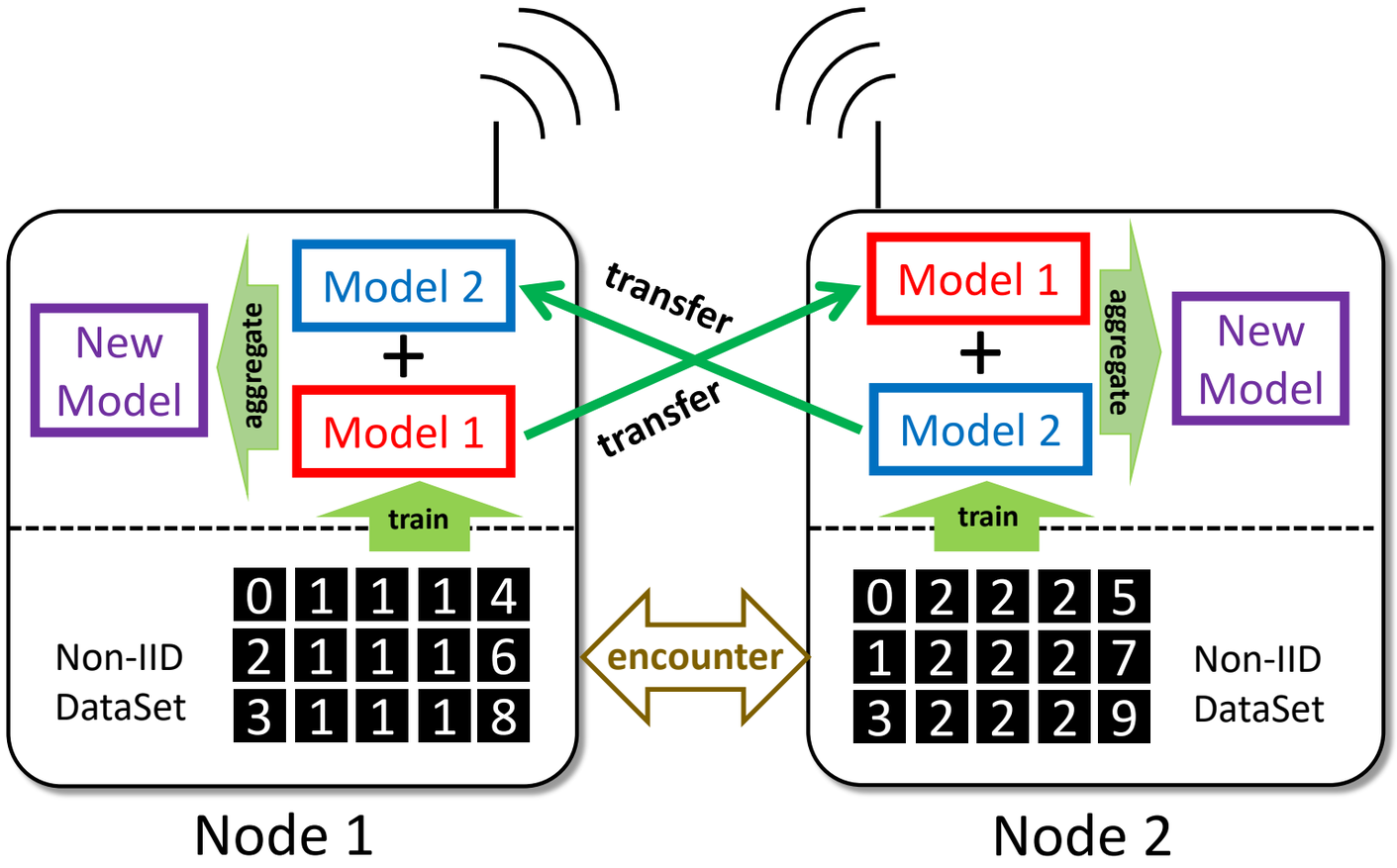
Characteristics of Wireless Ad Hoc Federated Learning (WAFL)

1. No third-party (or broker) mechanisms
   - Learning among peers without any third-party intervention.

2. No power structure
   - In Server-Client architecture, the service provider has the power.
   - Server-Client architecture can lead to a Master-Slave structure.
   - Every node is flat in Peer-to-Peer systems.

3. WAFL can realize multi-vendor scenarios
   - Anyone can join the system if collaboration protocols are defined.

Theoretical Aspects of Wireless Ad Hoc Federated Learning



Review of Machine Learning Mechanism in General

$(x, y) \in D$

x y

$\hat{y} = f(x, \theta)$ $\longleftrightarrow$ y

loss

$x \rightarrow \theta \rightarrow \hat{y}$ $\longleftrightarrow$ y

loss

ML adjusts the θ so that the loss is minimized for the entire D.

$$\theta \leftarrow \theta - \eta \nabla \left(\frac{\sum_{x,y \in D} loss(f(x, \theta), y)}{|D|} \right)$$

(*) θ denotes a set of model parameters

$$\bar{\theta} \cong \text{argmin}_{\theta} \frac{\sum_{x,y \in D} loss(f(x, \theta), y)}{|D|}$$

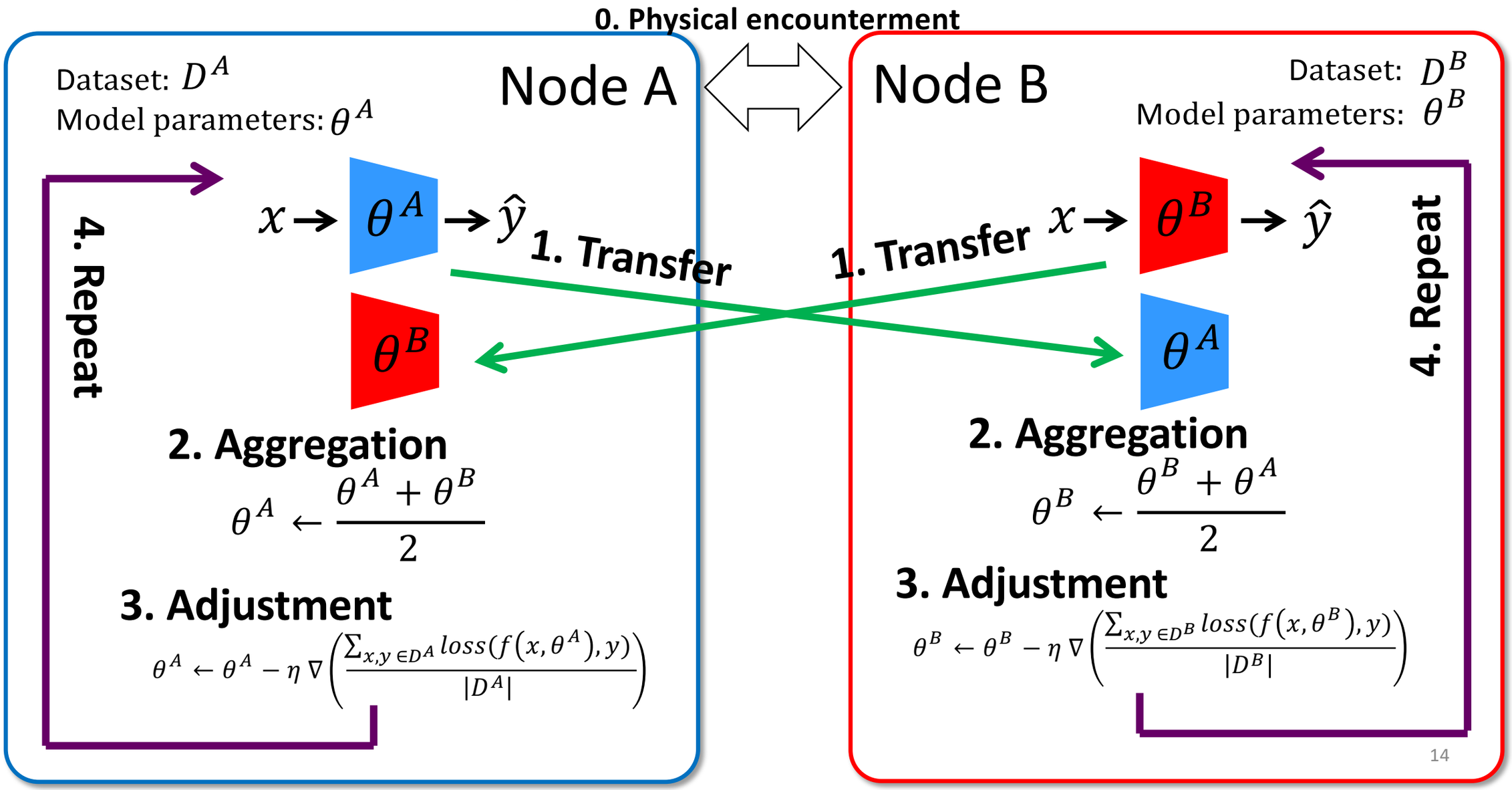
By using the optimized θ …

$x \rightarrow \bar{\theta} \rightarrow \hat{y}$

ML predicts proper y for x with enough accuracy.

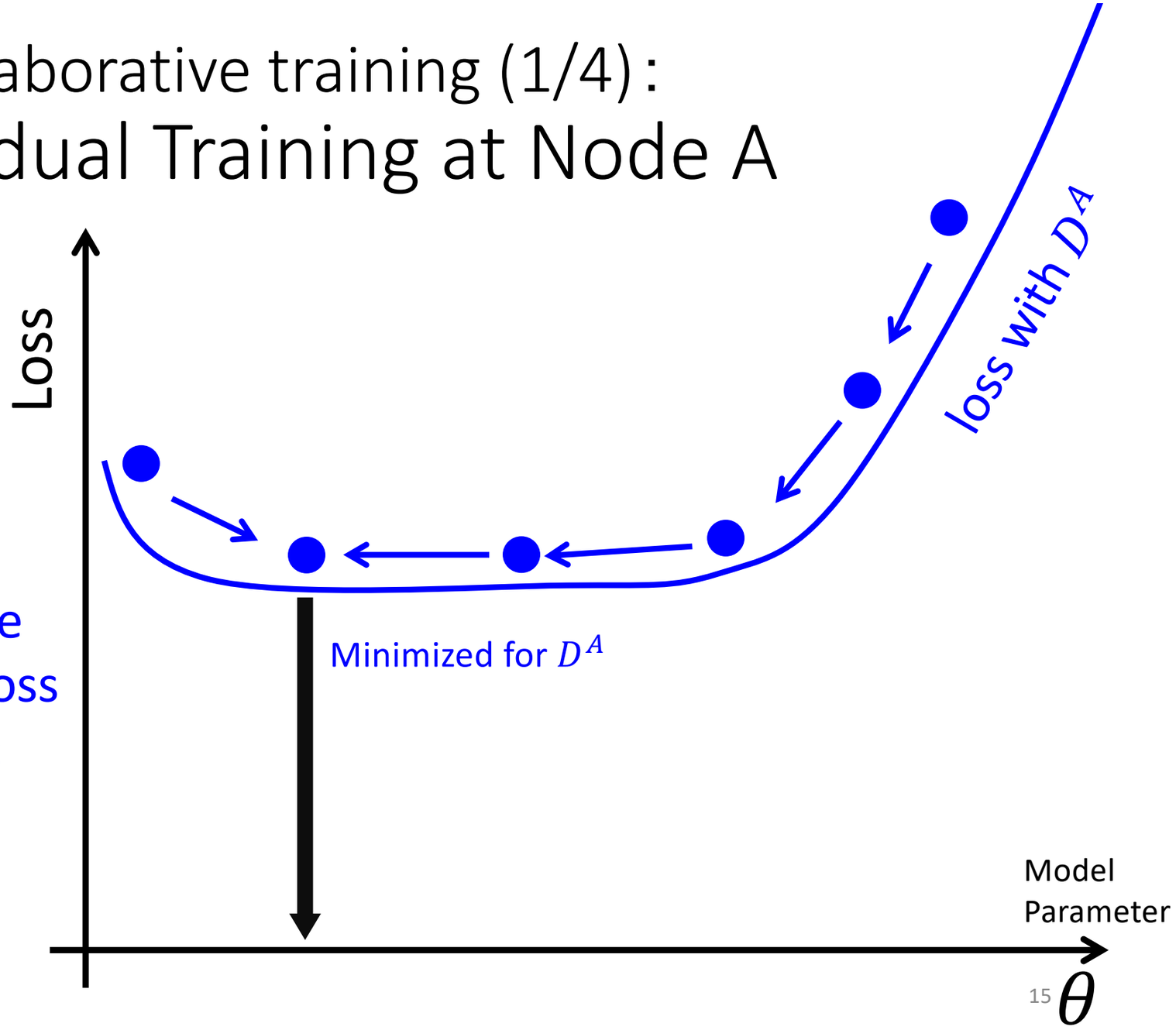
If it hits 100 out of 100, the accuracy is 100%.
If it hits 95 out of 100, the accuracy is 95%.

The learning algorithm of WAFL



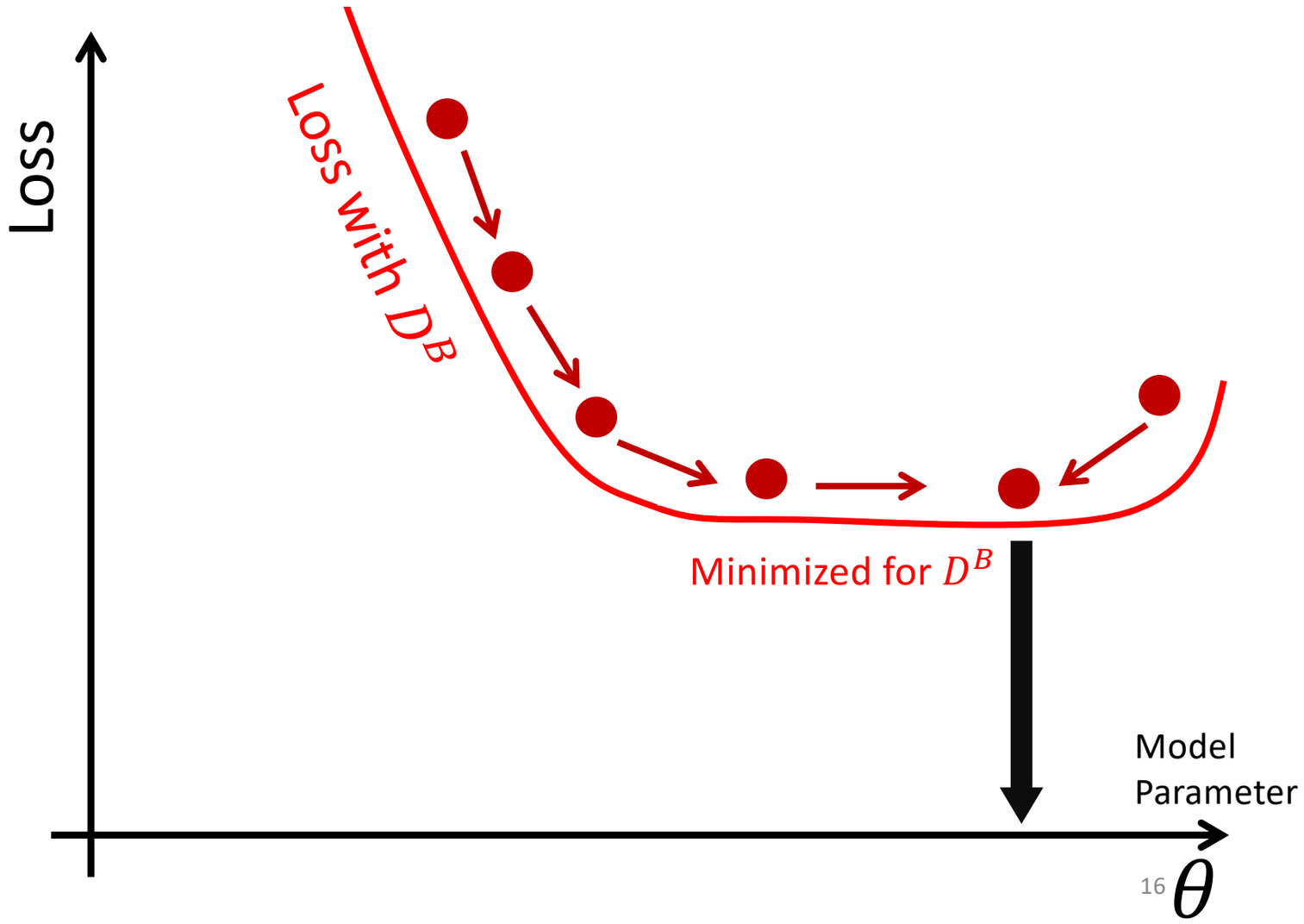
How WAFL allows collaborative training (1/4):
The Case of Individual Training at Node A

Model parameters will move to the θ that minimizes the loss in Valley & Basin structure



How WAFL allows collaborative training (2/4):
The Case of Individual Training at Node B

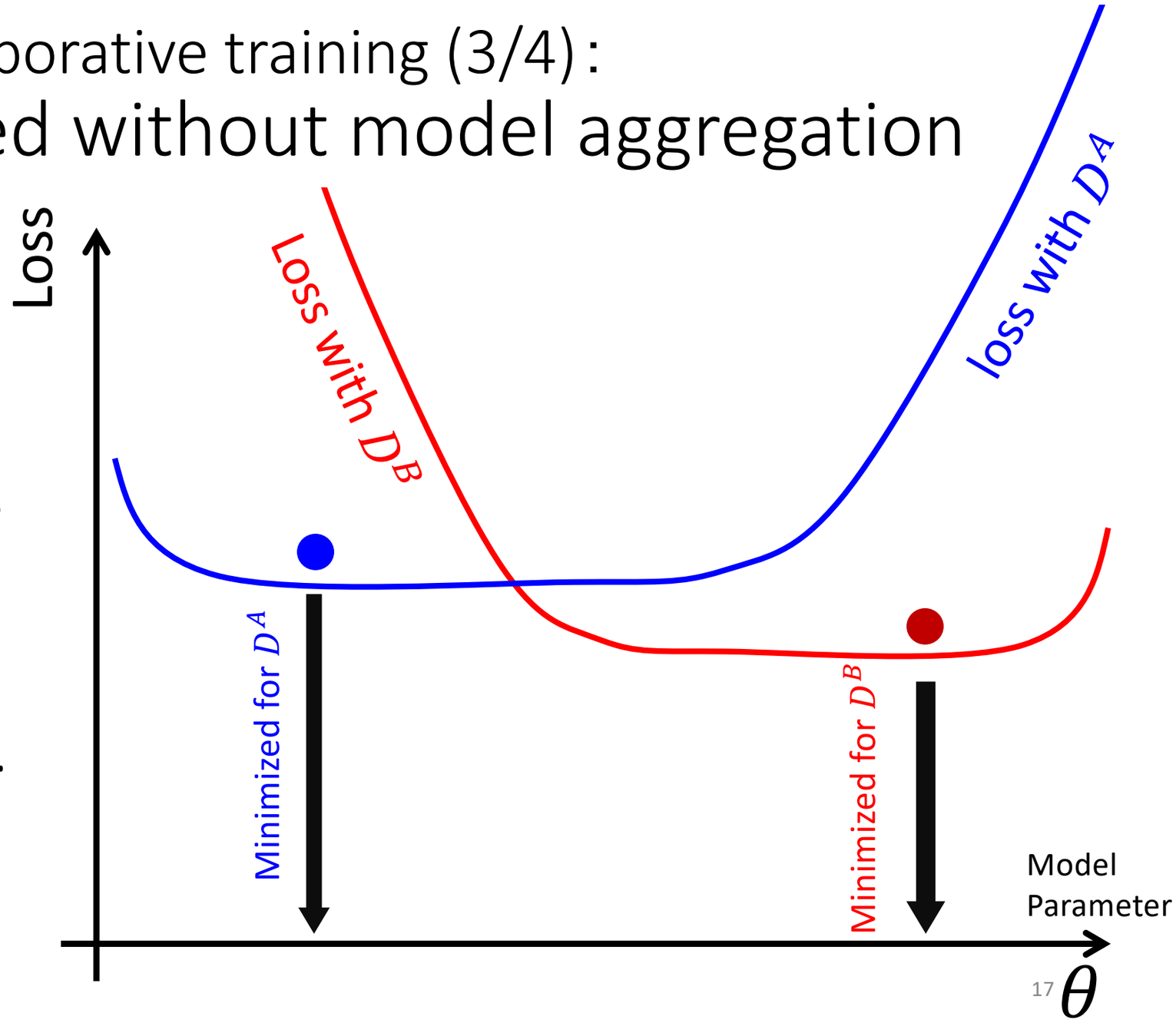
If the distribution of the dataset is different, the minimum location is different.



How WAFL allows collaborative training (3/4):
Not fully generalized without model aggregation

1. Minimum point for D^{A} is optimized for Node A, but not for Node B (D^{B}).

2. Minimum point for D^{B} is optimized for Node B, but not for Node A (D^{A}).



How WAFL allows collaborative training (4/4):
Aggregation finds an optimal model for Node A and B

The model aggregation,

$$\theta^A \leftarrow \frac{\theta^A + \theta^B}{2}$$

attracts encountered models with each other in the parameter space θ.

➔ This search for the optimal point for both nodes.

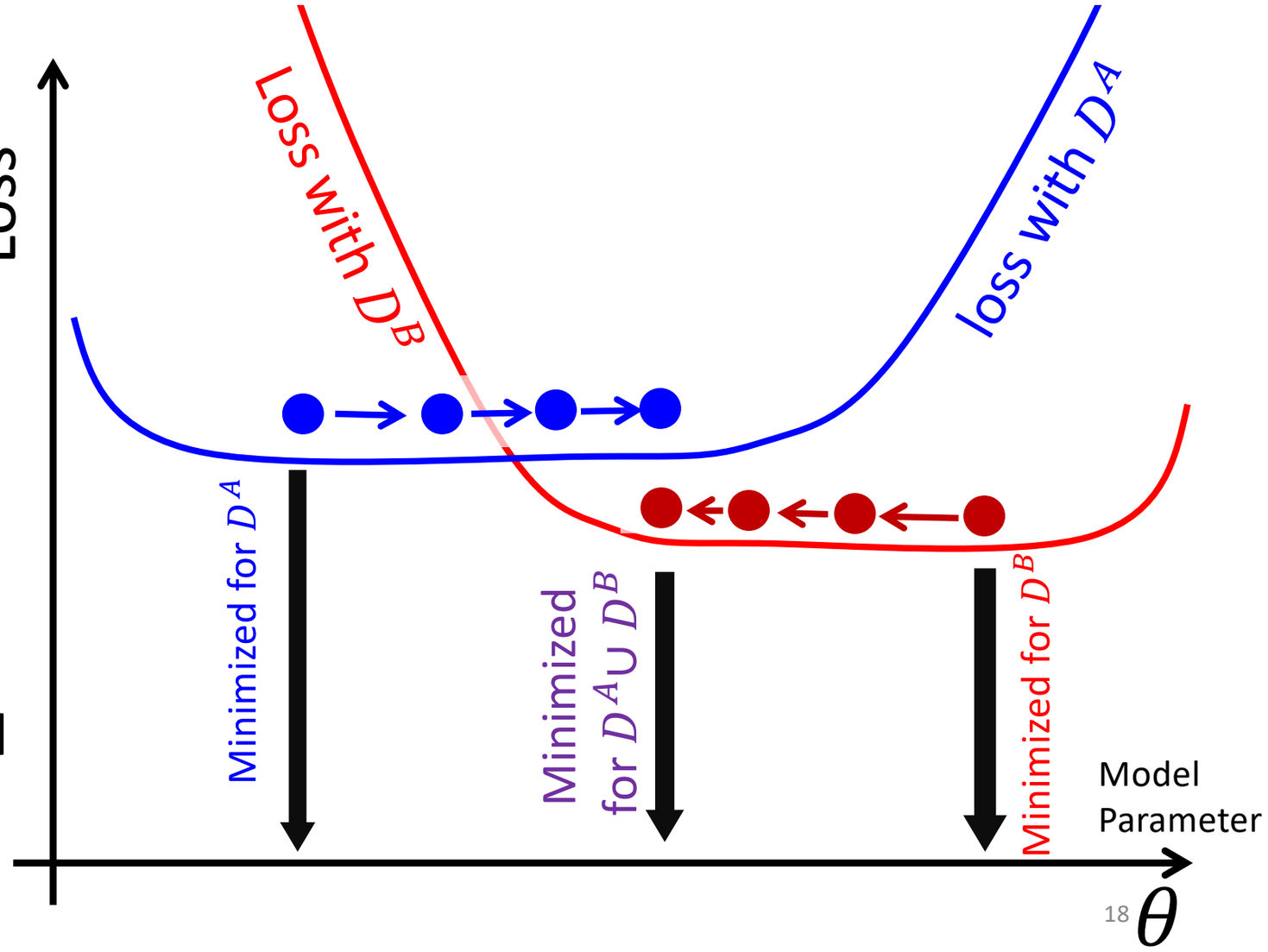
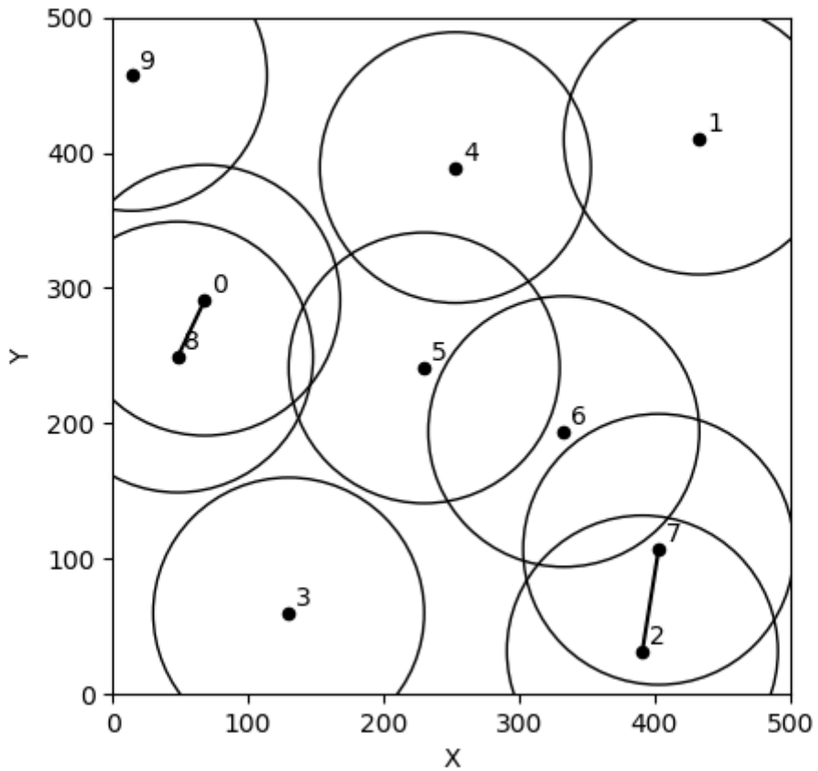
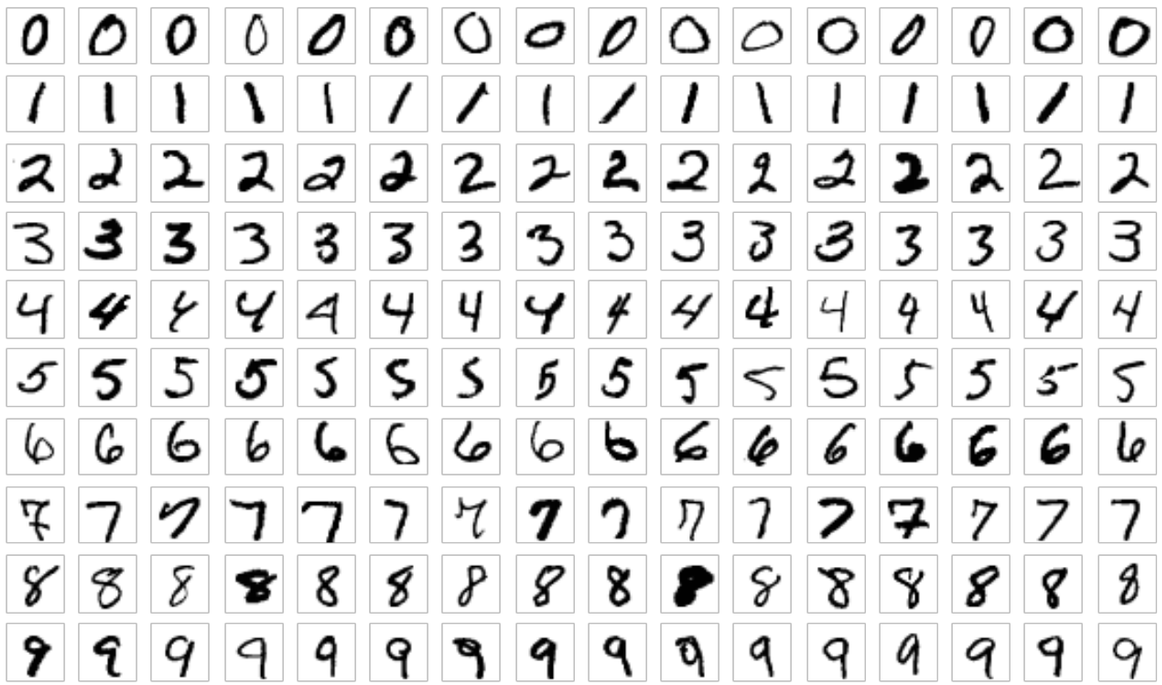


Table of Contents

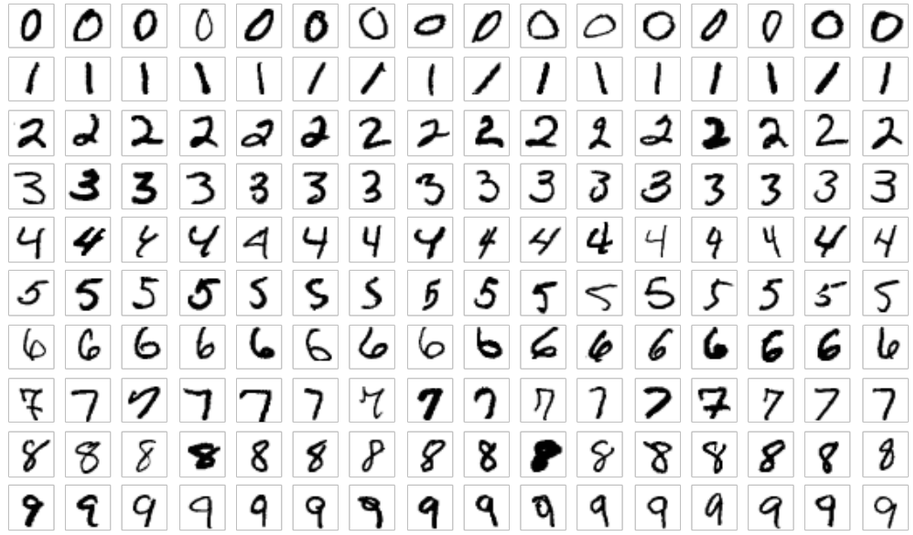
A Demonstration of
Wireless Ad Hoc Federated Learning with Benchmark Evaluation



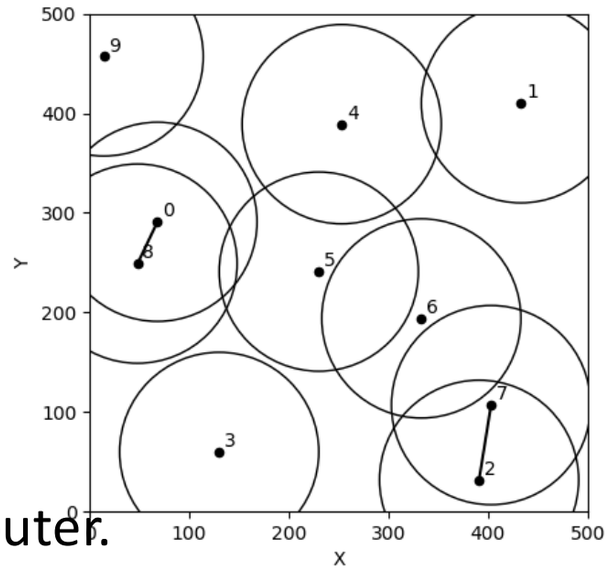


Benchmark Evaluation

- Why benchmark evaluation ?
  - For reproducibility, and standard
  - For understanding the technical characteristics
- ML model: Multilayer Perceptron (MLP)
- Dataset: MNIST
- Mobility Pattern
  - Static: 4 topologies
  - Dynamic Case 1: 3 Random Waypoint Mobility (RWP)
  - Dynamic Case 2: 3 Community-Structured Environment (CSE)
- Simulation
  - We carried out the experiment by simulation on a single computer.



Yann et.al., THE **MNIST** DATABASE of handwritten digits, 1998.
http://yann.lecun.com/exdb/mnist/



Random Waypoint Mobility

Benchmark Evaluation: Experiment Setting
90% Non-IID MNIST Dataset

■ Training Data Distribution

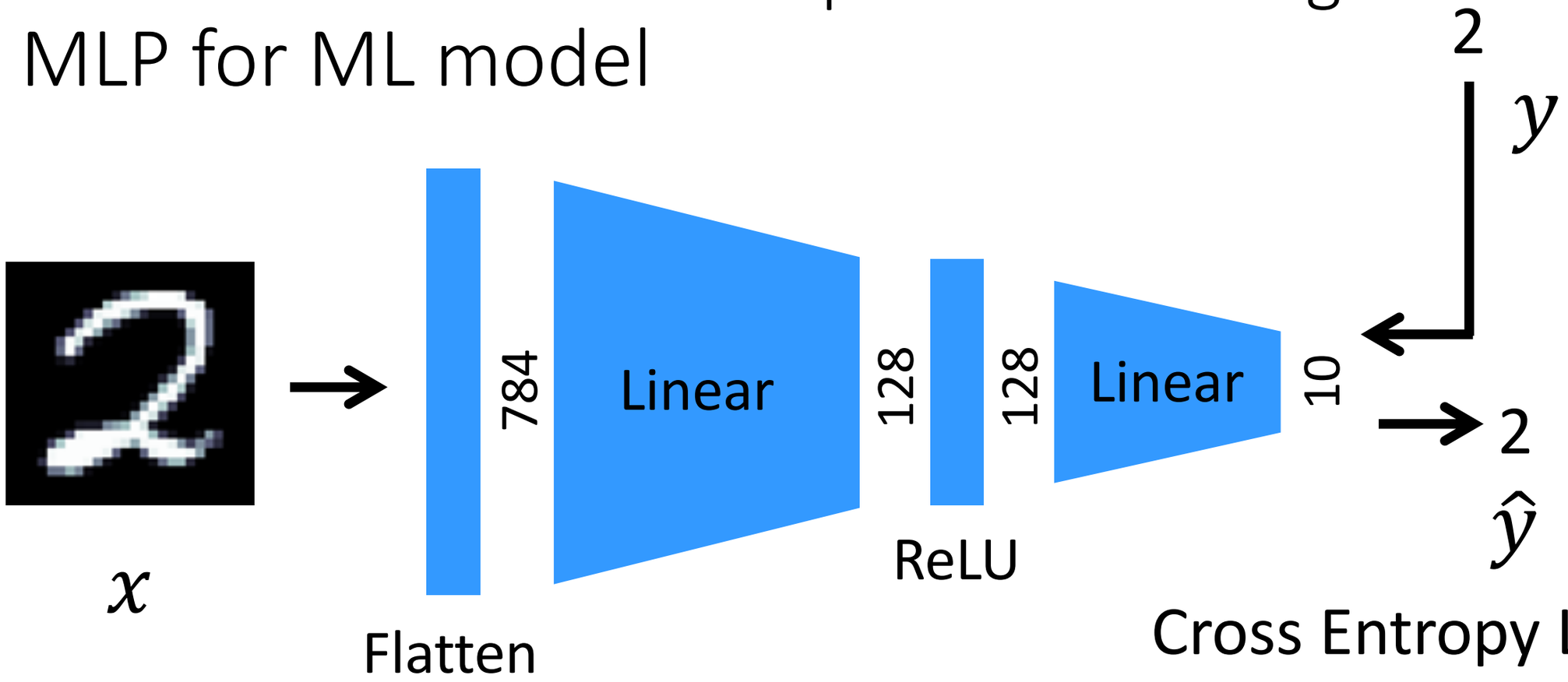
90% of Node *n's* samples are label *n* samples.

Node	L0	L1	L2	L3	L4	L5	L6	L7	L8	L9	Summary
0	5341	76	64	76	61	57	57	61	74	50	5917
1	79	6078	67	52	57	59	58	79	59	68	6656
2	58	67	5374	80	64	68	87	73	57	61	5989
3	68	74	73	5537	51	56	72	65	67	77	6140
4	73	67	80	67	5301	59	53	69	70	64	5903
5	60	66	57	74	68	4896	61	59	66	69	5476
6	52	78	53	56	58	66	5312	56	65	65	5861
7	67	90	66	74	55	61	65	5683	63	77	6301
8	59	80	59	54	63	48	88	67	5268	57	5843
9	66	66	65	61	64	51	65	53	62	5361	5914
Summary	5923	6742	5958	6131	5842	5421	5918	6265	5851	5949	60000

■ Testing Data Distribution

All the labels appear with equal probability – which is *independent and identically distributed* (IID)
We used the standard MNIST testing dataset.

Benchmark Evaluation: Experiment Setting
MLP for ML model



2-layer fully-connected neural networks (FC-NNs)

Cross Entropy Loss

Adam Optimizer

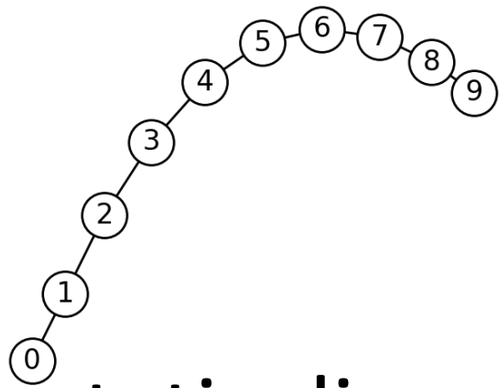
Learning rate 0.001

Coefficient of Aggregation 1.0

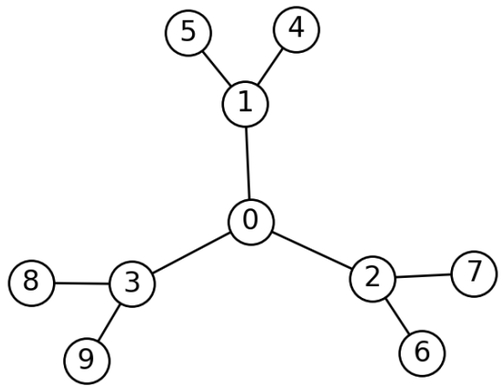
Benchmark Evaluation
1. Evaluation on Static Network Topologies

- If nodes are statically deployed, e.g., in wireless sensor network scenarios, the network topology becomes static.

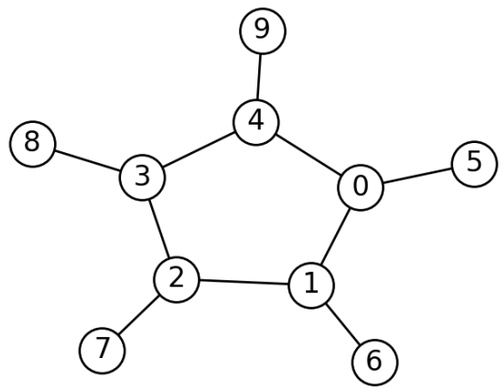
- We carried out WAFL assuming the four network topologies.



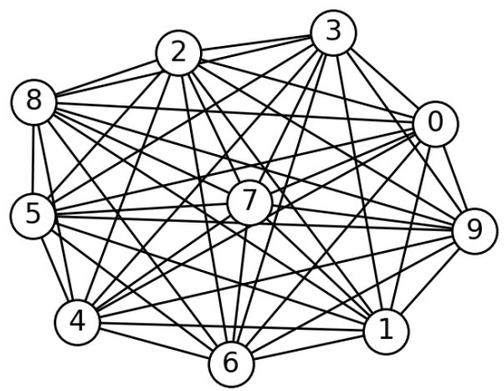
static_line



static_tree

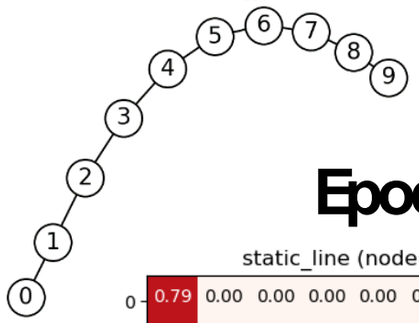


static_ringstar



static_dense

Benchmark Evaluation
Changes of Confusion Matrices to the Testing Data

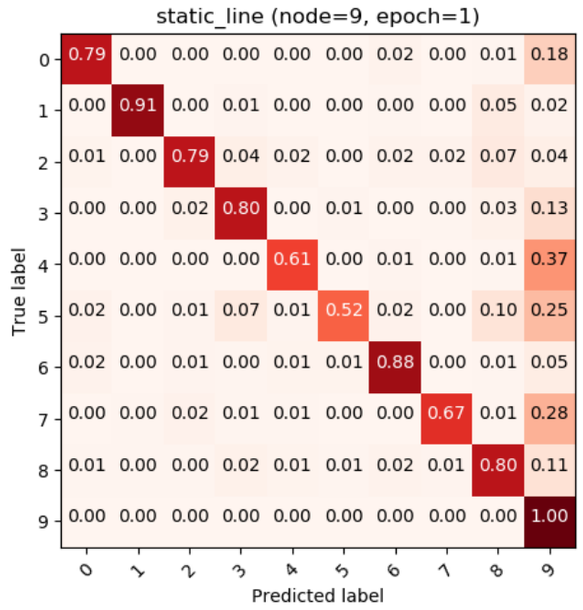


Case of static_line

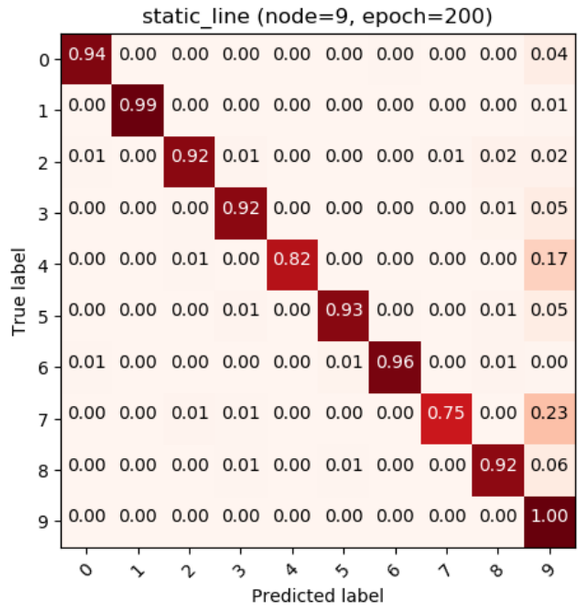
Example @Node 9

static_line

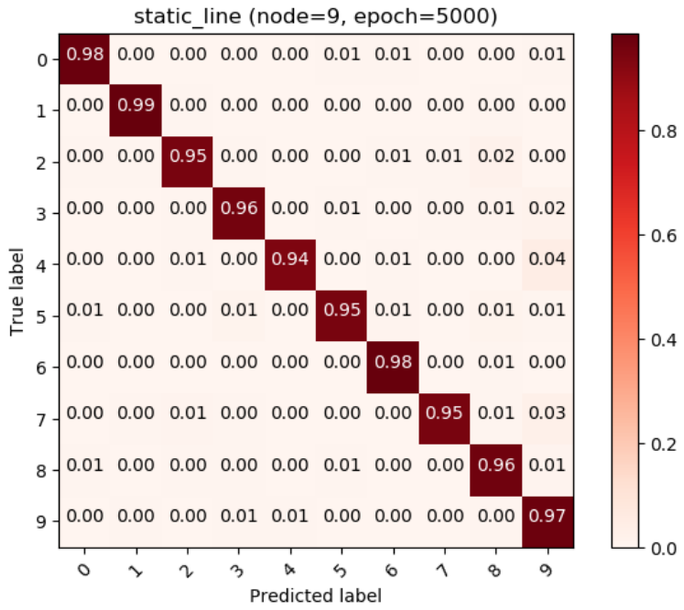
Epoch 1



Epoch 200



Epoch 5000



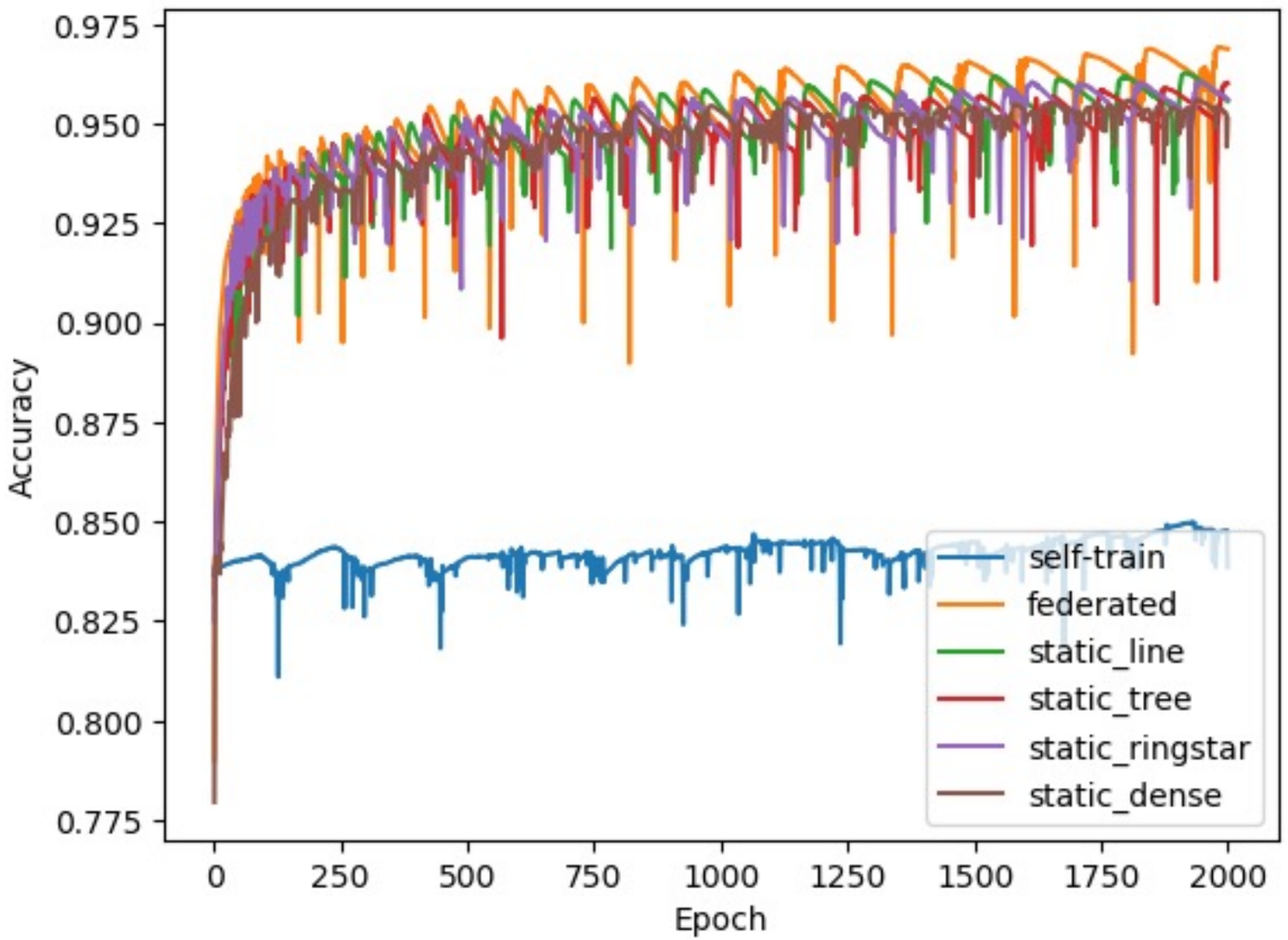
In the beginning, many testing records were misclassified into 9.

E.g., True 4 was misclassified into 9. True 7 was 9.

Finally, misclassifications were drastically reduced.

Benchmark Evaluation

Accuracy Trend (Static Network Topologies)



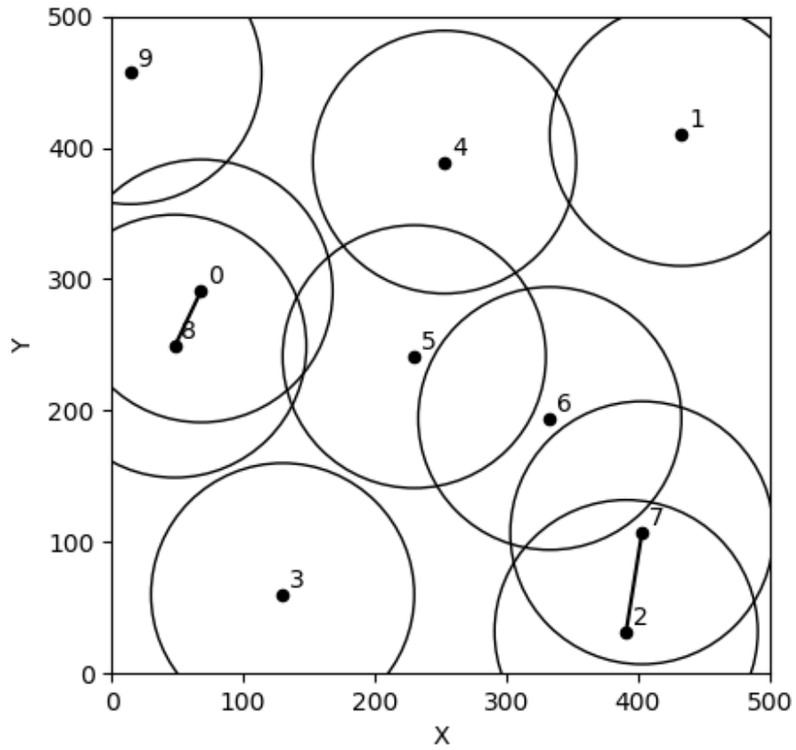
WAFL has achieved almost the same performance as conventional federated learning.

10.0-11.5% Improvement

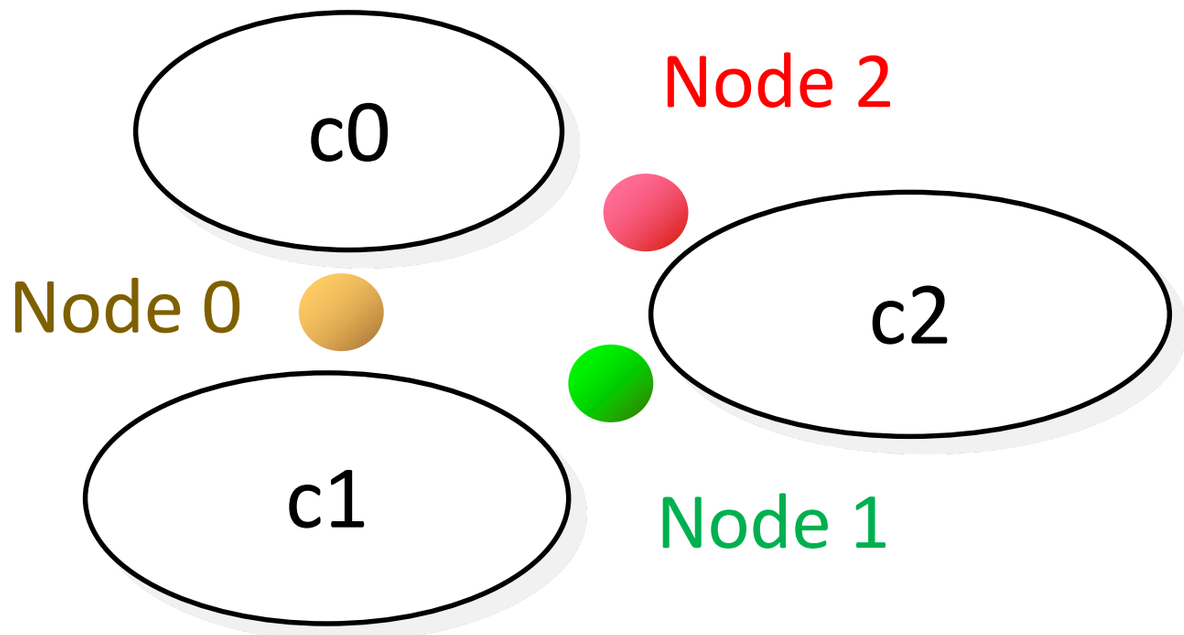
Accuracy of Self-Train

(*) Fluctuations were caused by “Optimizer Adam”

Benchmark Evaluation
2. Evaluation on Dynamic Network Topologies



Random Waypoint Mobility (RWP)

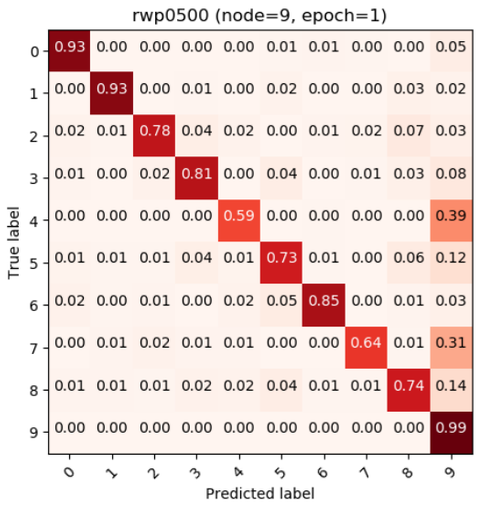


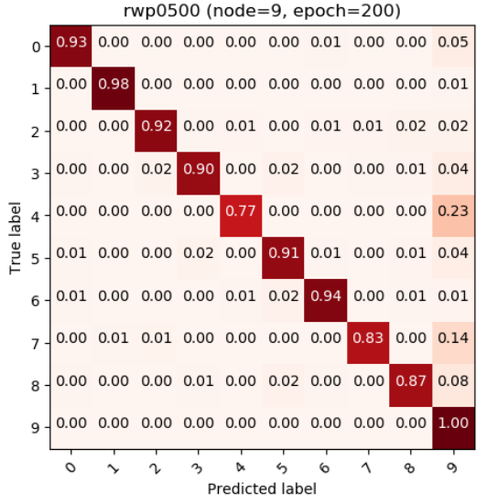
Community-Structured Environment (CSE)

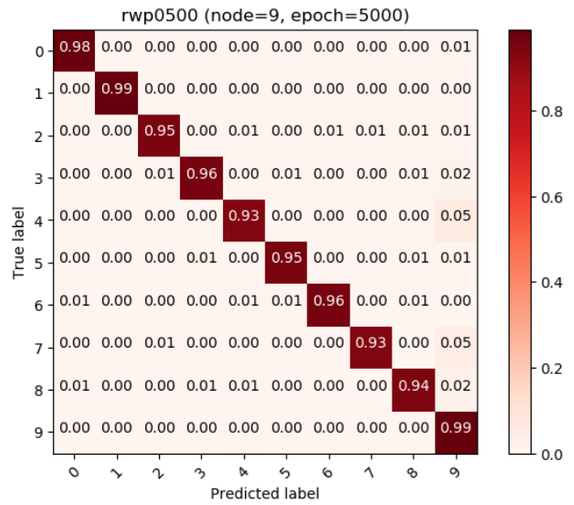
Benchmark Evaluation
Change of Confusion Matrices to the Testing Data

Epoch 1 | **Epoch 200** | **Epoch 5000**

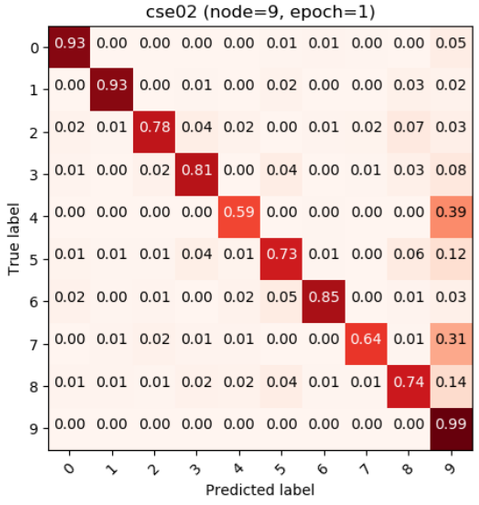
rwp0500

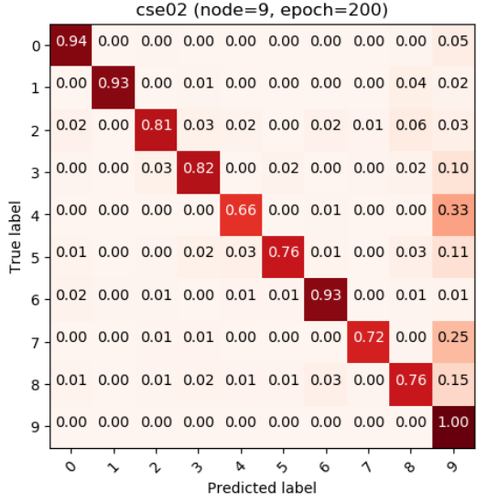


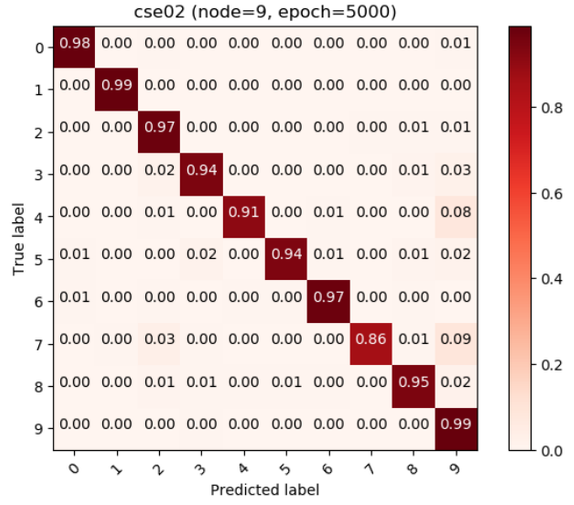




cse02





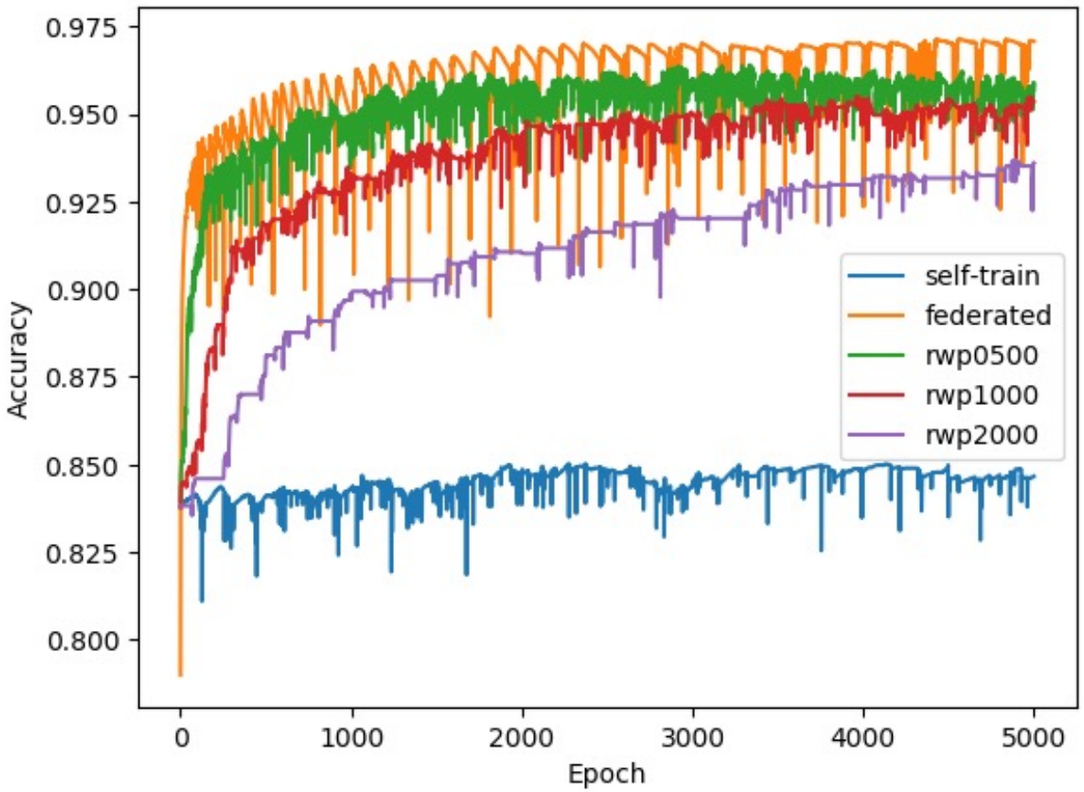


RWP, CSE @Node 9

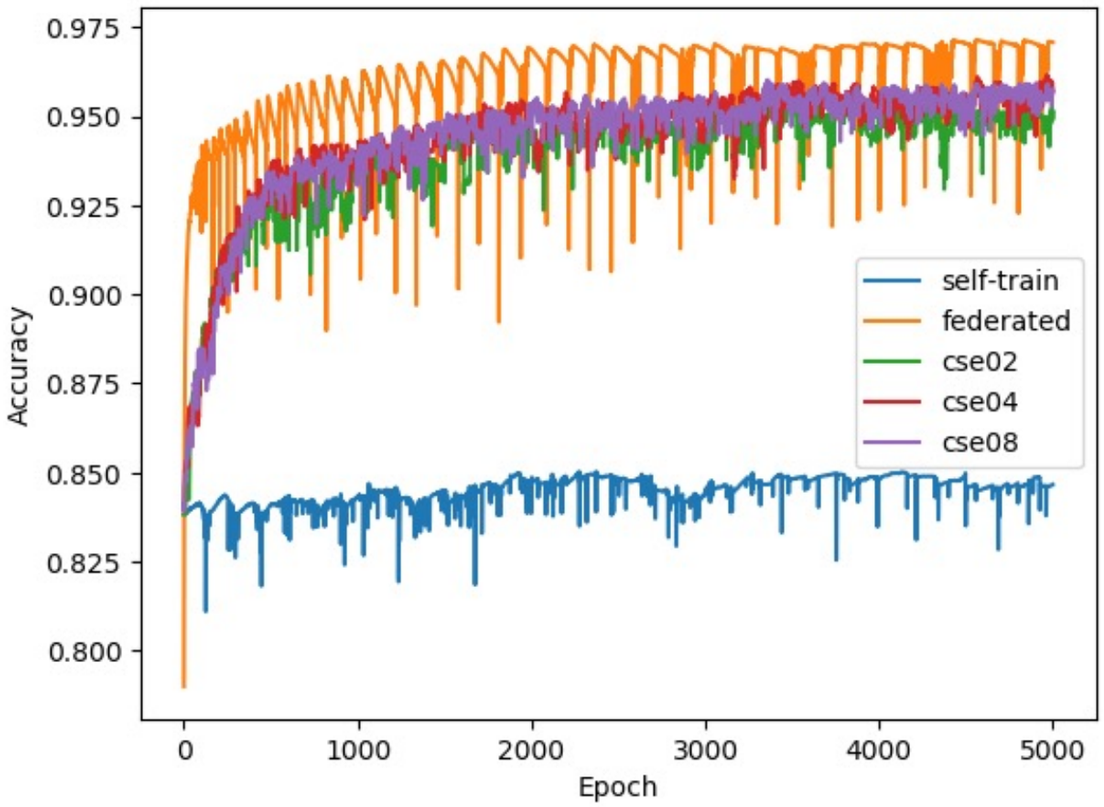
Misclassifications were drastically reduced.

Benchmark Evaluation

Accuracy Trend (Dynamic Network Topologies)



(a) RWP



(b) CSE

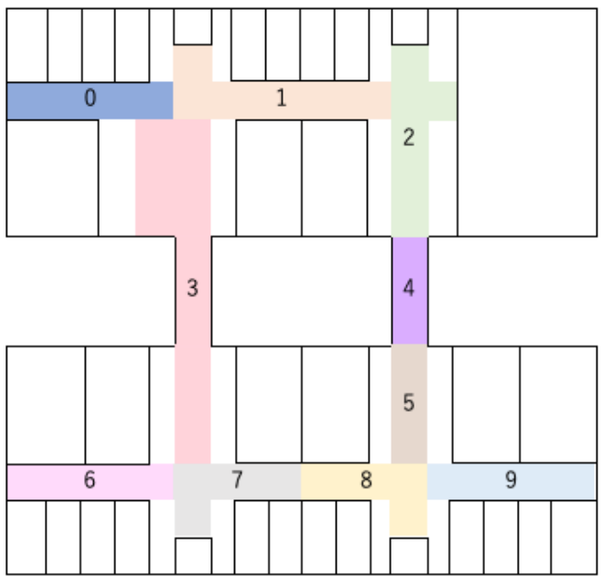
(*) Fluctuations were caused by “Optimizer Adam”

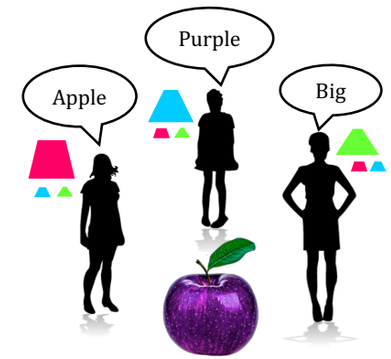
Benchmark Evaluation: Summary of Accuracy

Training Method	Accuracy [%]	Diff(FL) [%]	Diff(Self-Train) [%]
WAFL (static_line)	**96.337±0.595**	**-0.419**	**11.674**
WAFL (static_tree)	95.906±1.196	-0.850	11.243
WAFL (static_ringstar)	96.110±0.745	-0.646	11.447
WAFL (static_dense)	95.643±1.036	-1.113	10.98
WAFL (dynamic_rwp0500)	**95.389±1.218**	**-1.367**	**10.726**
WAFL (dynamic_rwp1000)	95.230±0.981	-1.526	10.567
WAFL (dynamic_rwp2000)	93.659±0.948	-3.097	8.996
WAFL (dynamic_cse02)	94.840±0.833	-1.916	10.177
WAFL (dynamic_cse04)	95.454±1.288	-1.302	10.791
WAFL (dynamic_cse08)	**95.560±0.946**	**-1.196**	**10.897**
Federated Learning	96.756±0.928	-	-
Self-Training	84.663±1.285	-	-

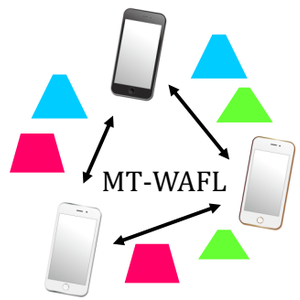
Table of Contents

Applications of
Wireless Ad Hoc Federated Learning

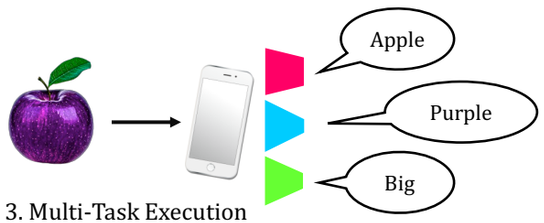




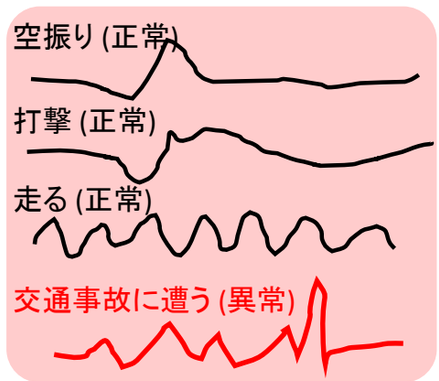
1. Labeling by their own perspective

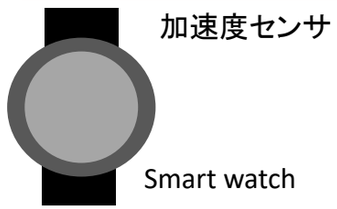


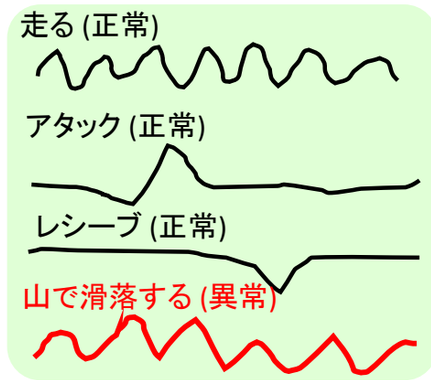
2. Model mixture by MT-WAFL

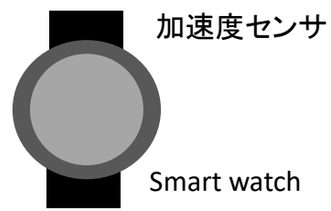


3. Multi-Task Execution









(1) Self-Localization

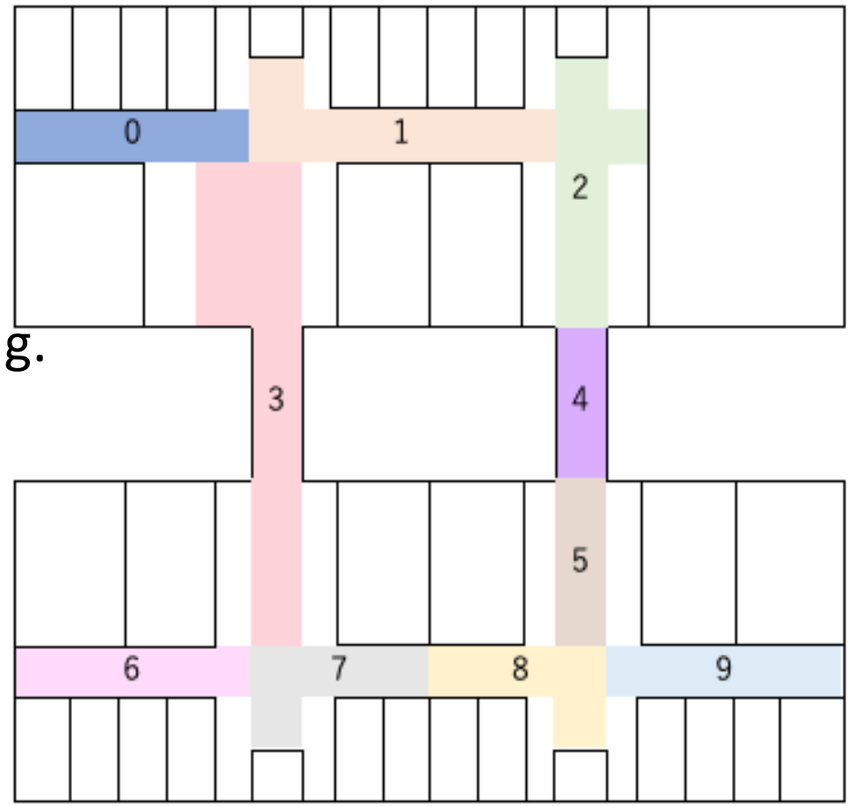
(2) Brainstormer

(3) Anomaly Detection

Collaborative Training for Self-Localization with WAFL

1. Self-localization with observed Wi-Fi AP's RSSI

SSID	BSSID	RSSI	CHANNEL	HT	CC	SECURITY (auth/unicast/group)
sanshiro	74:88:bb:02:d7:c9	-90	100,+1	Y	JP	WPA2(PSK/AES/AES)
91A2-A-WPA3	90:96:f3:c4:15:9a	-86	100	Y	JP	WPA2(SAE/AES/AES)
91A2-A	90:96:f3:c4:15:99	-85	100	Y	JP	WPA2(PSK/AES/AES)
elab-internal	04:fe:7f:a1:2d:43	-83	52	Y	--	WPA2(PSK/AES/AES)
Buffalo-A-A4C0	c6:36:c0:00:a4:ca	-82	64	Y	JP	WPA2(PSK/AES/AES)



2. Problems in the previous studies
   We had to collect RSSI list at all the locations for training.

3. WAFL allows model development with
   (a) a node training around area 0,
   (b) another node training around area 9,
   (c) aggregating the developed models among them.

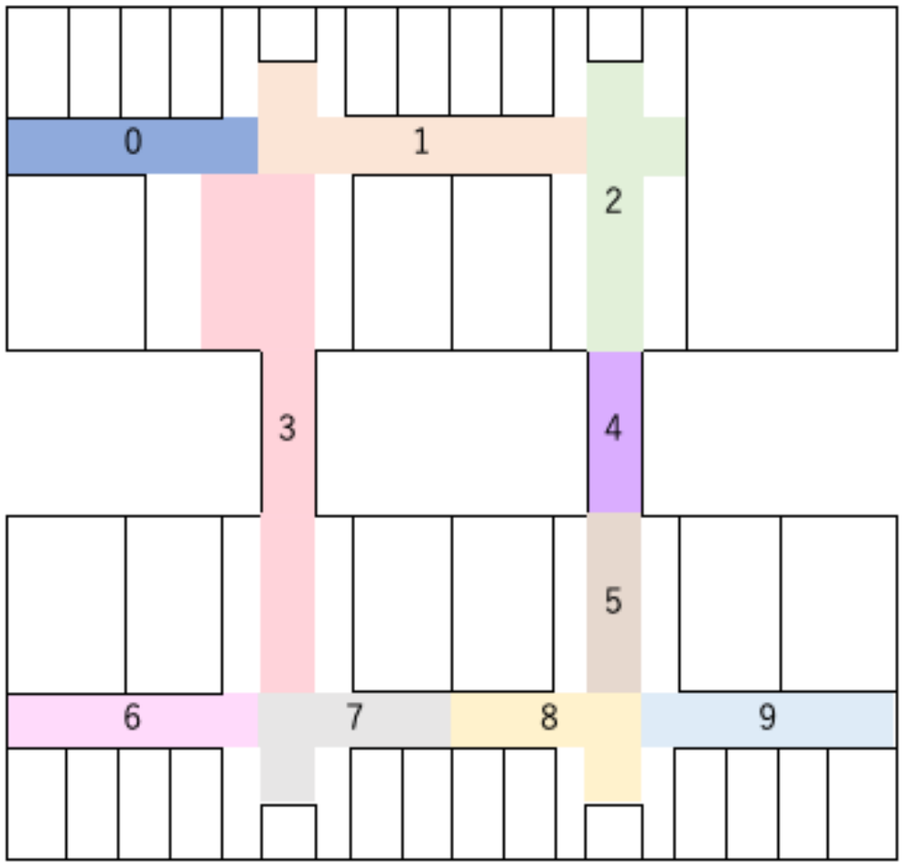
4. The aggregate model can predict the location at anywhere.

Sugizaki, Ochiai, and Esaki . "Wi-Fi の信号強度を用いた自律分散協調学習による広域自己位置推定." IEICE-Tech Report (MSS) 2022.17 (2022): 1-4.

Collaborative Training for Self-Localization with WAFL

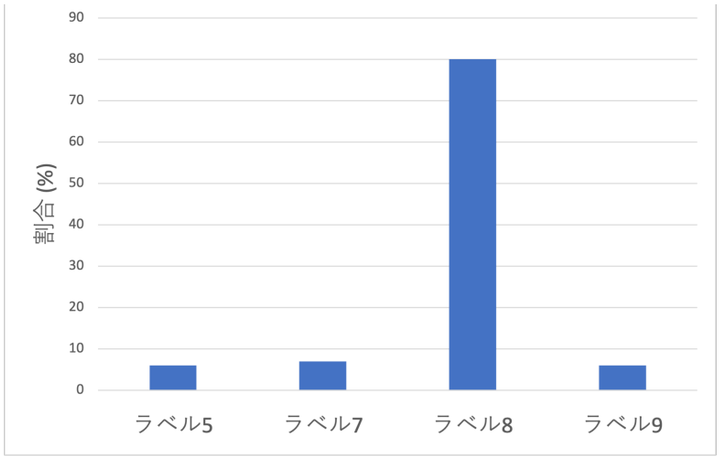
Experiment Setting

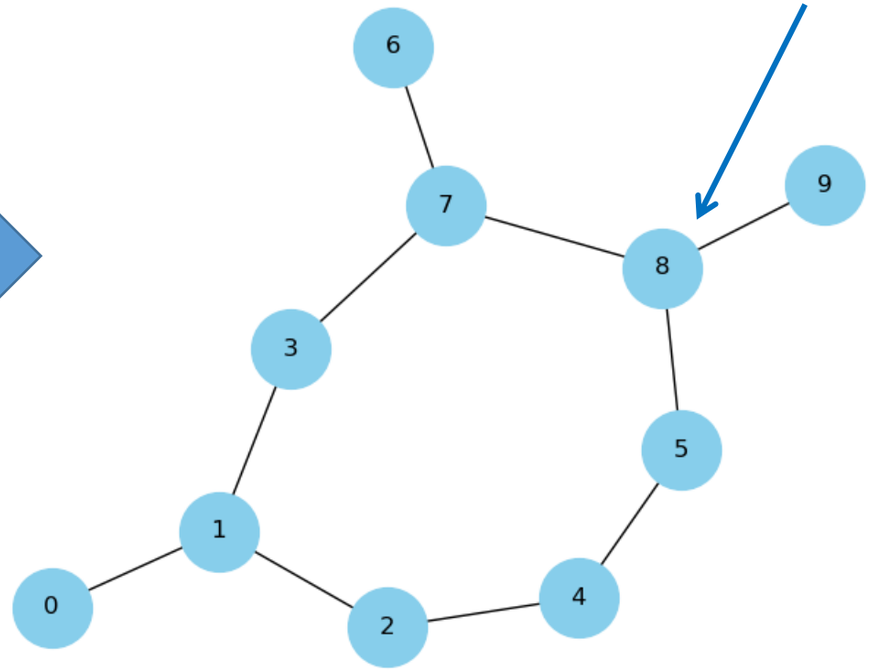
* Assumption
Single node in each area.
Each node communicate only with neighbors.



Area Labels

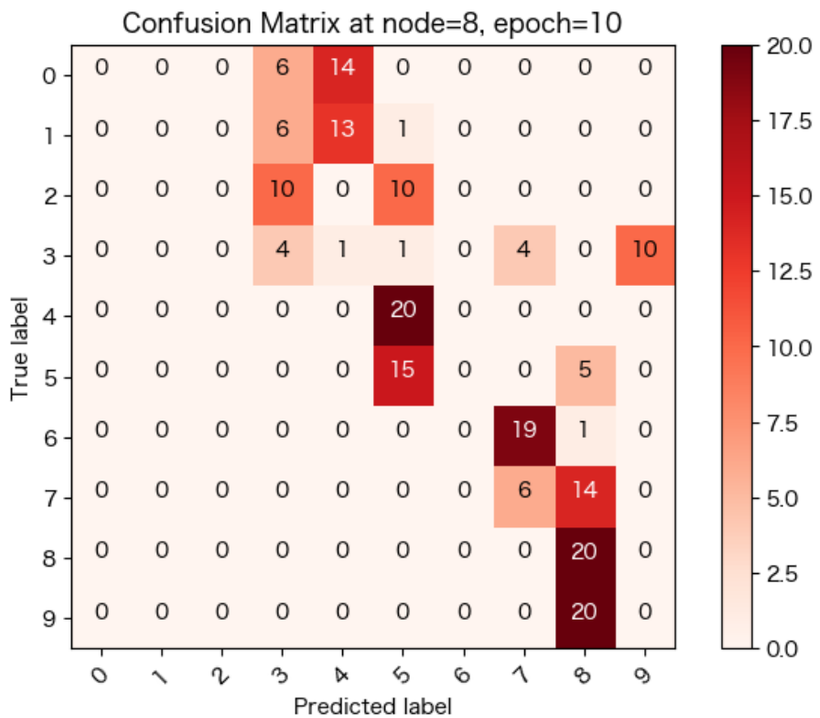
Distribution of the training data @Node 8



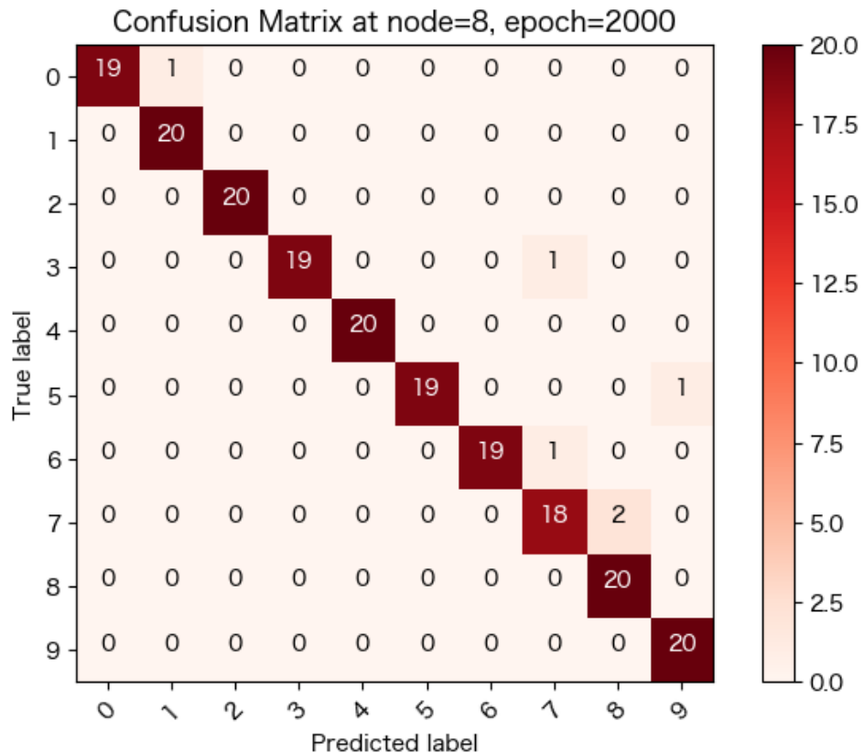


Network Topology

Node 8's Location Prediction Example



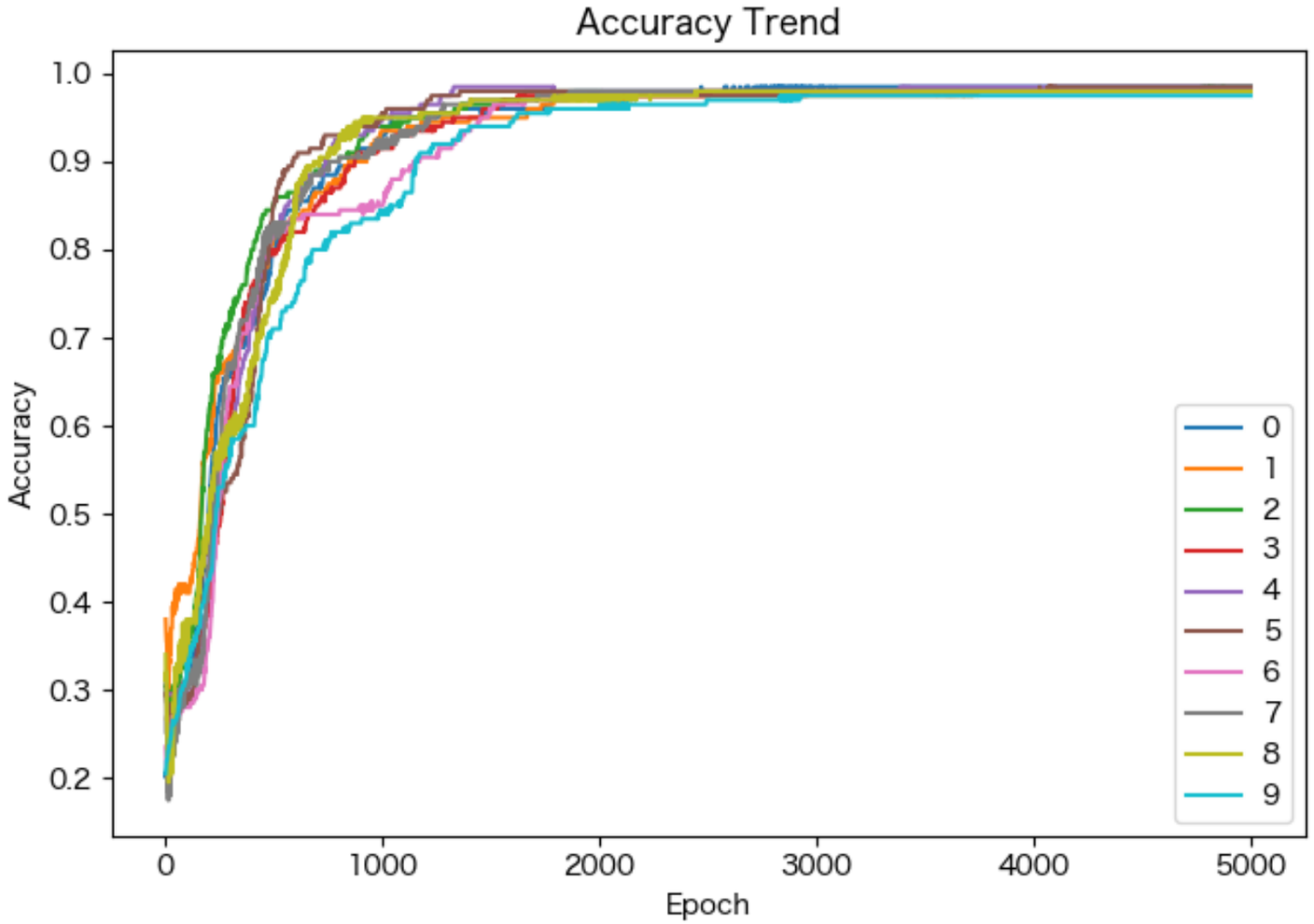
Epoch 10



Epoch 2000

At the beginning of training, the model predicted wrong location from observed RSSIs, but finally, it could predict the location precisely.

Testing Accuracy of all the Nodes

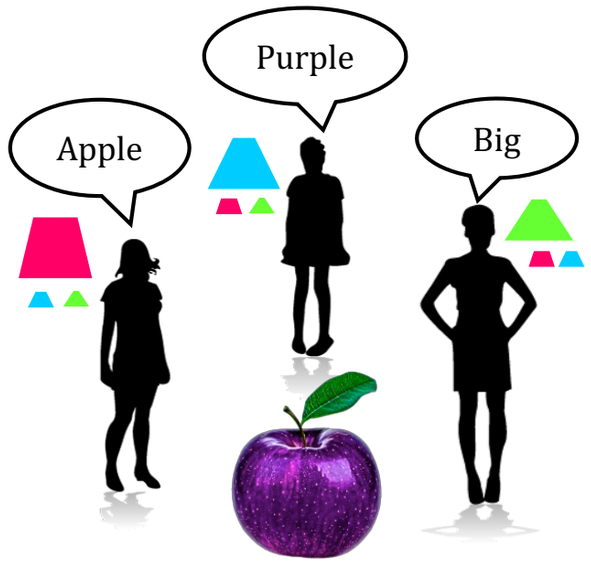


All the nodes could achieve more than 95% accuracies for the testing data.

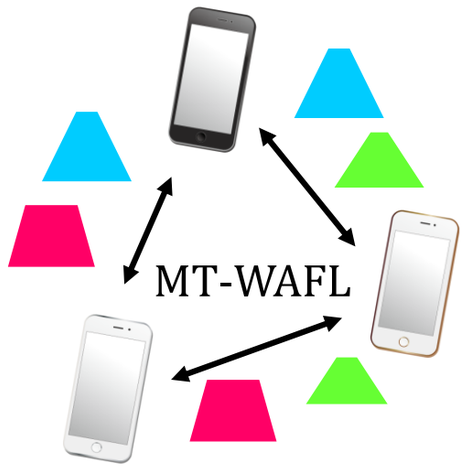
Integration of Multiple Perspectives by WAFL

1. People in a team react in different ways to a presented object, because they have different perspectives in their idea.

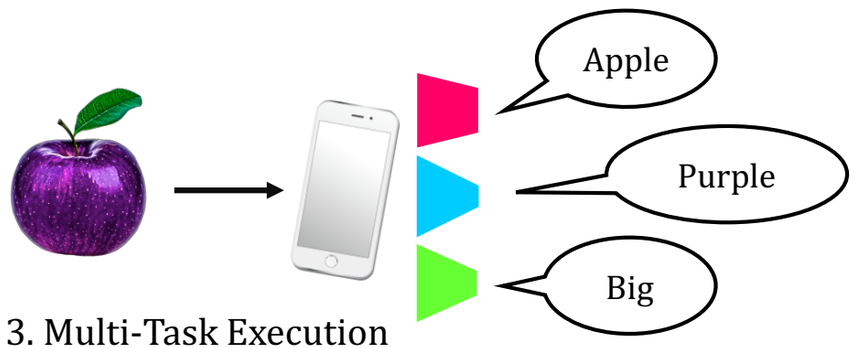
2. Integration of these perspectives of team members will develop the reaction model for the team.



1. Labeling by their own perspective



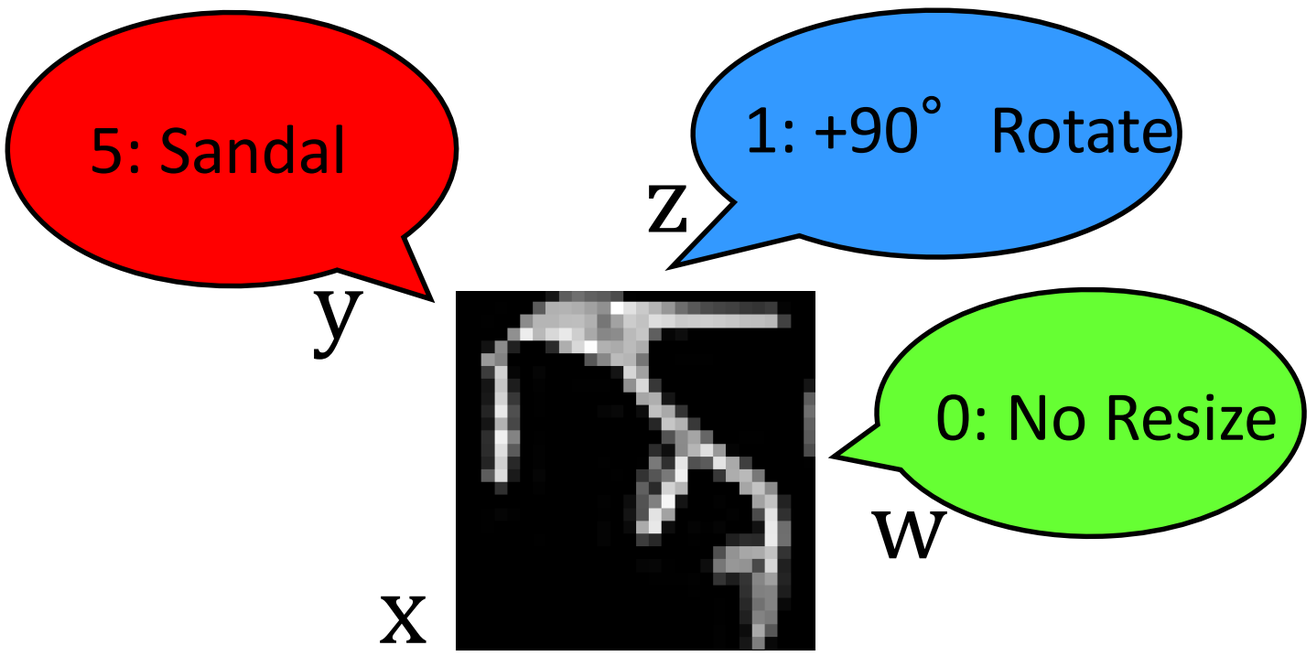
2. Model mixture by MT-WAFL



3. Multi-Task Execution

Experiment Setting

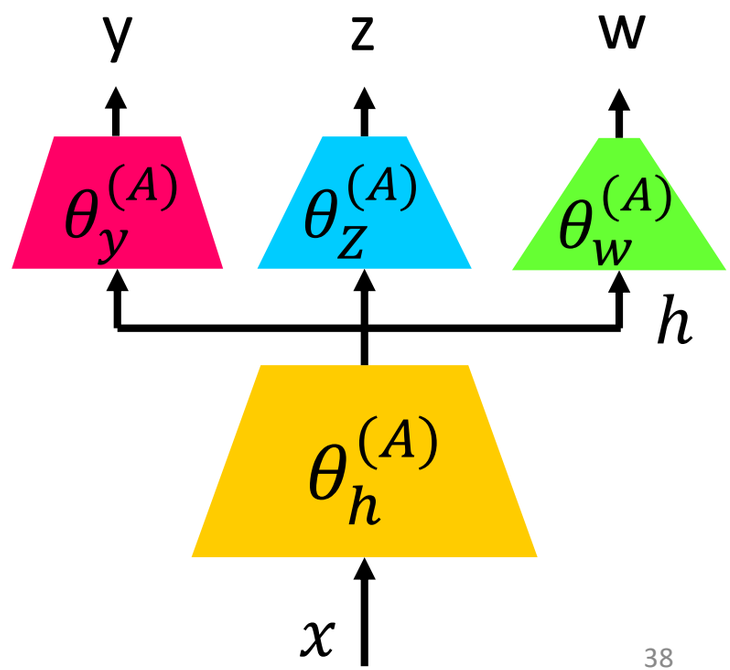
X: Rotated and Resized Fashion-MNIST Image
Y: Class of the object (0, 1, …, 9)
Z: Rotation of the object (0, 1, 2, 3)
W: Size of the object (0, 1)



Mobility: rwp0500

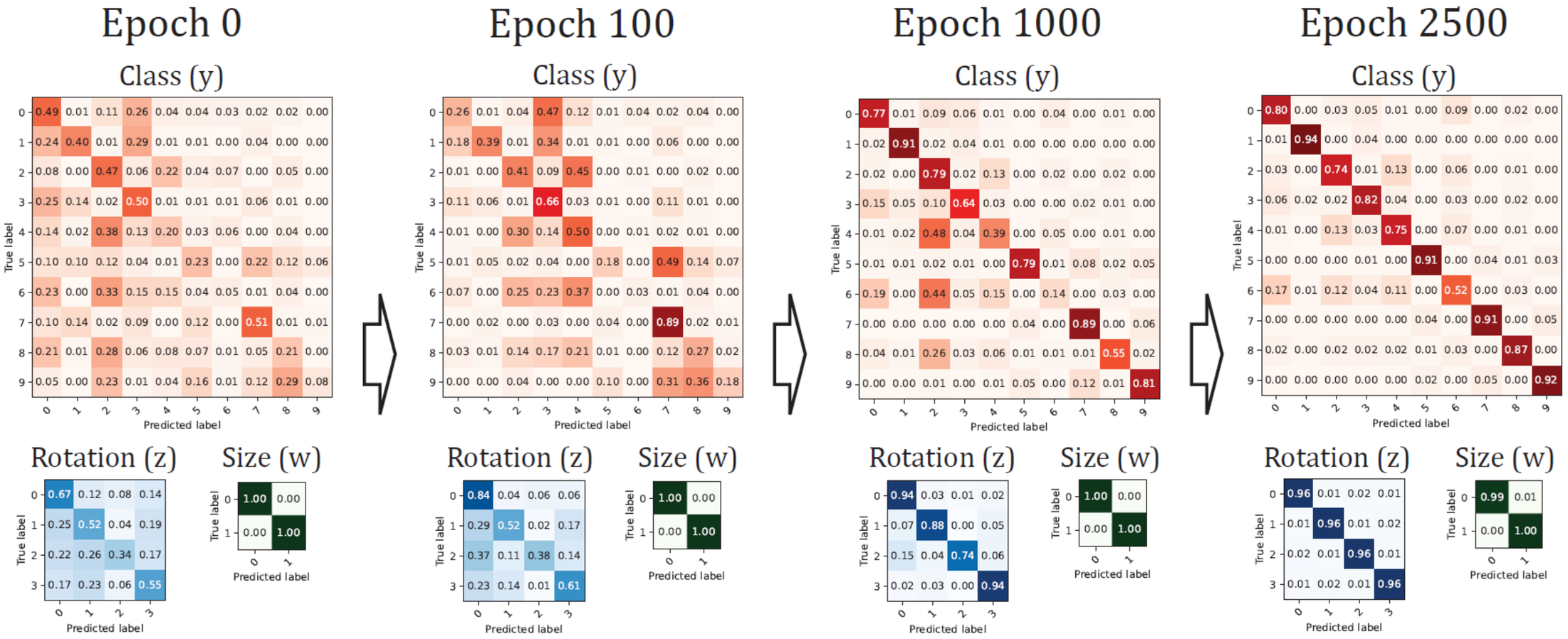
Distribution of the Label Existence

	Label Y	Label Z	Label W
Node 0-4	99%	1%	1%
Node 5-7	1%	99%	1%
Node 8-9	1%	1%	99%



Integration of Multiple Perspectives by WAFL

Node 8's Predictions for Object Class, Rotation, and Size

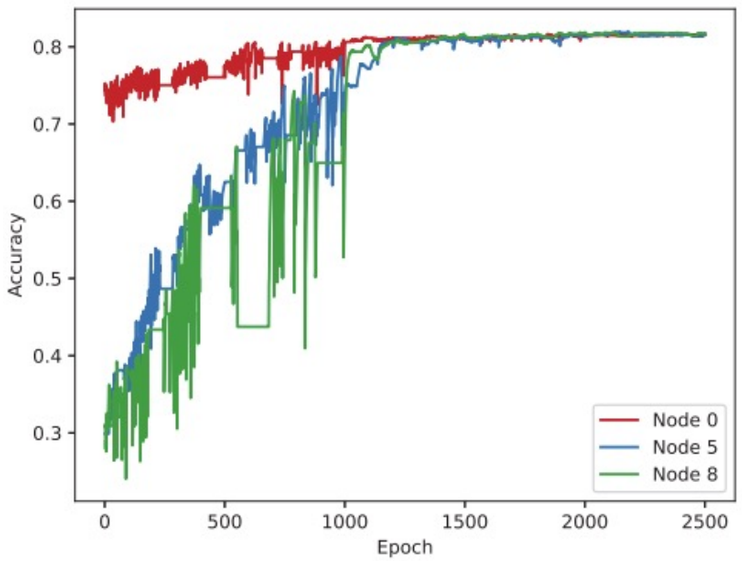


Node 8 originally has many labels in Size(w) perspective.

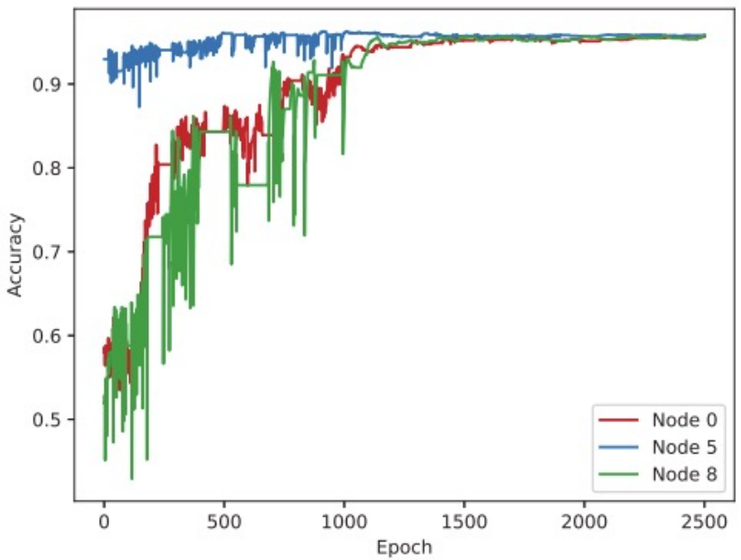
As the training proceeds, misclassifications in Object Class and Rotation predictions have improved.

Integration of Multiple Perspectives by WAFL

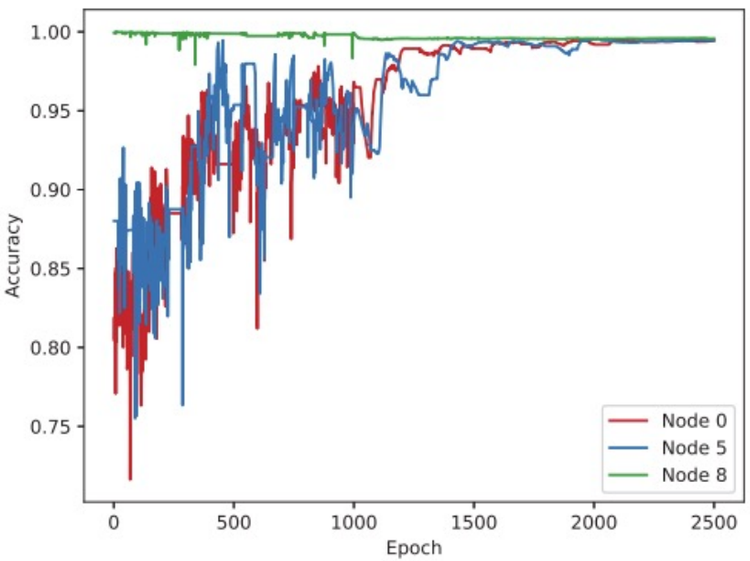
Accuracy of Class, Rotation, and Size @Node (0, 5, 8)



(a) Test Accuracy for Class (Y)



(b) Test Accuracy for Rotation (Z)



(c) Test Accuracy for Size (W)

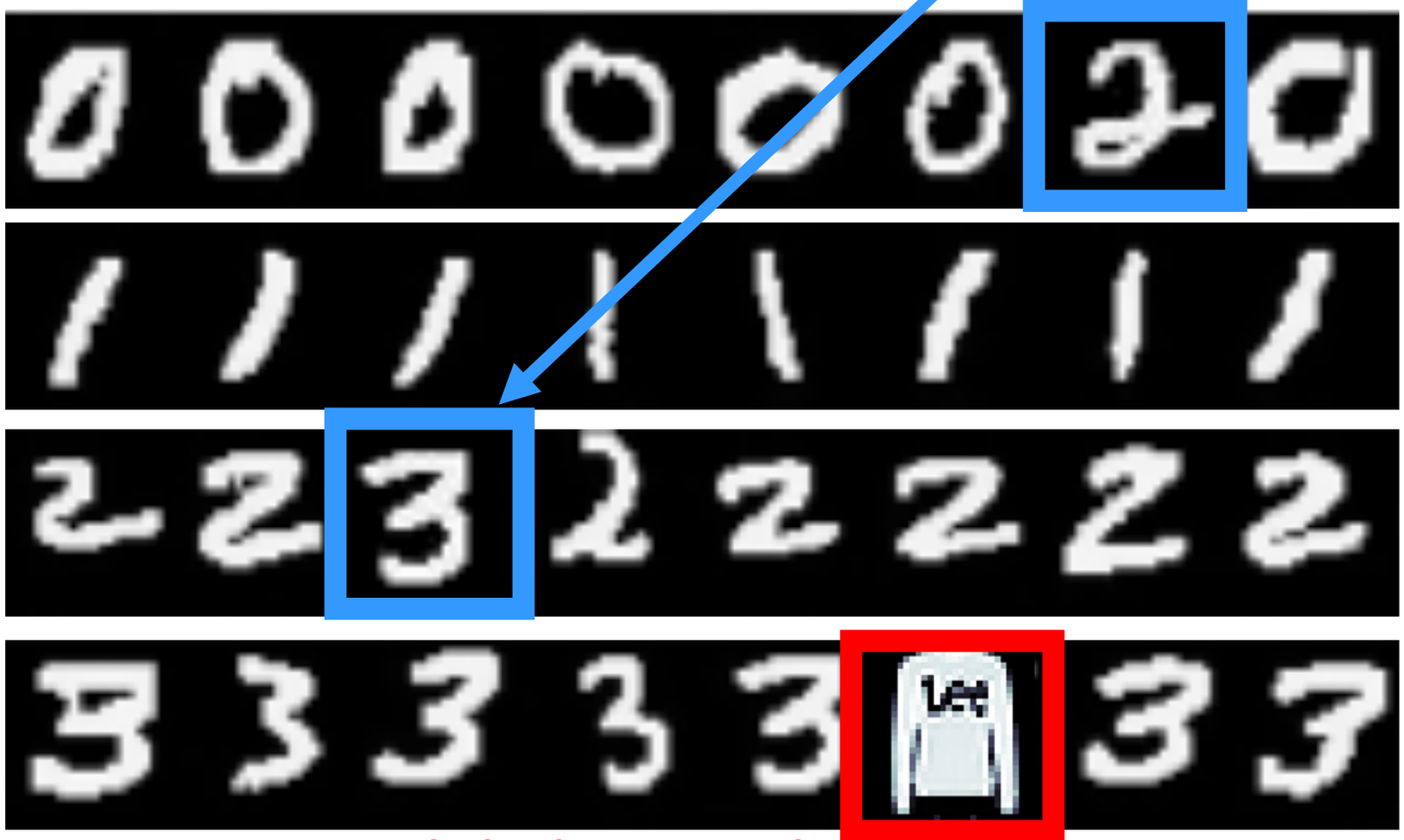
Node 0 has Class labels.

Node 5 has rotation (Z) labels.

Node 8 has size (W) labels.

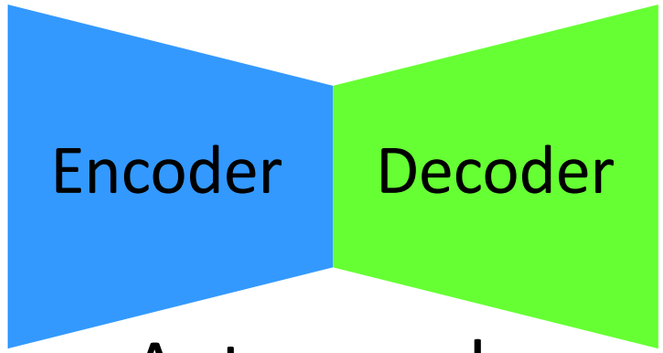
Accuracy has been improved in all the perspectives.

Anomaly Detection in Non-IID Scenario



To Detect a Global Anomaly

Legitimate Input

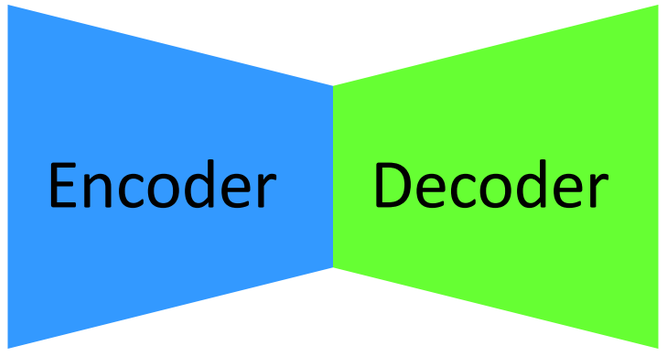


Autoencoder

Output (Reconstruction)

"To precisely reconstruct the legitimate inputs", WAFL-Autoencoder should learn the legitimate features.

Anomaly Input



Autoencoder

Output (Reconstruction)

WAFL-Autoencoder should fail in reconstructing a Global Anomaly input.

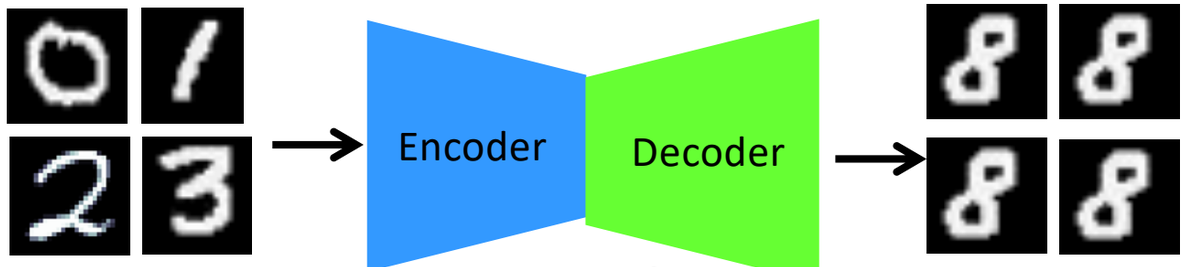
Feature Learning and Anomaly Detection with WAFL

Experiment Setting

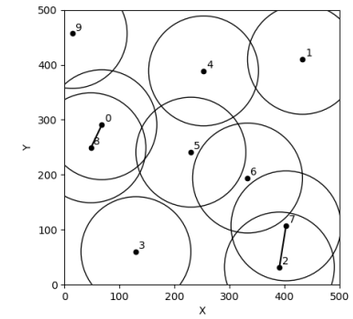
Distribution of Training Data (99.95% Non-IID)

Node	L0	L1	L2	L3	L4	L5	L6	L7	L8	L9	Total
Node 0	4736	0	0	0	0	0	0	0	0	0	4736
Node 1	0	5418	0	0	2	1	0	0	0	1	5422
Node 2	0	0	4779	0	0	0	0	0	0	0	4479
Node 3	1	0	0	4911	0	0	1	1	0	0	4914
Node 4	0	0	0	1	4733	0	0	0	0	0	4734
Node 5	0	0	0	0	0	4343	1	0	0	0	4344
Node 6	0	0	0	1	0	0	4712	0	0	0	4713
Node 7	1	0	1	0	1	0	2	5046	0	0	5051
Node 8	1	0	0	0	0	0	1	0	4714	0	4716
Node 9	0	0	0	0	0	0	0	0	1	4751	4752

For example, we expect the Autoencoder at Node 8 will give outputs as follows (if no WAFL).



Legitimate Input Autoencoder Output

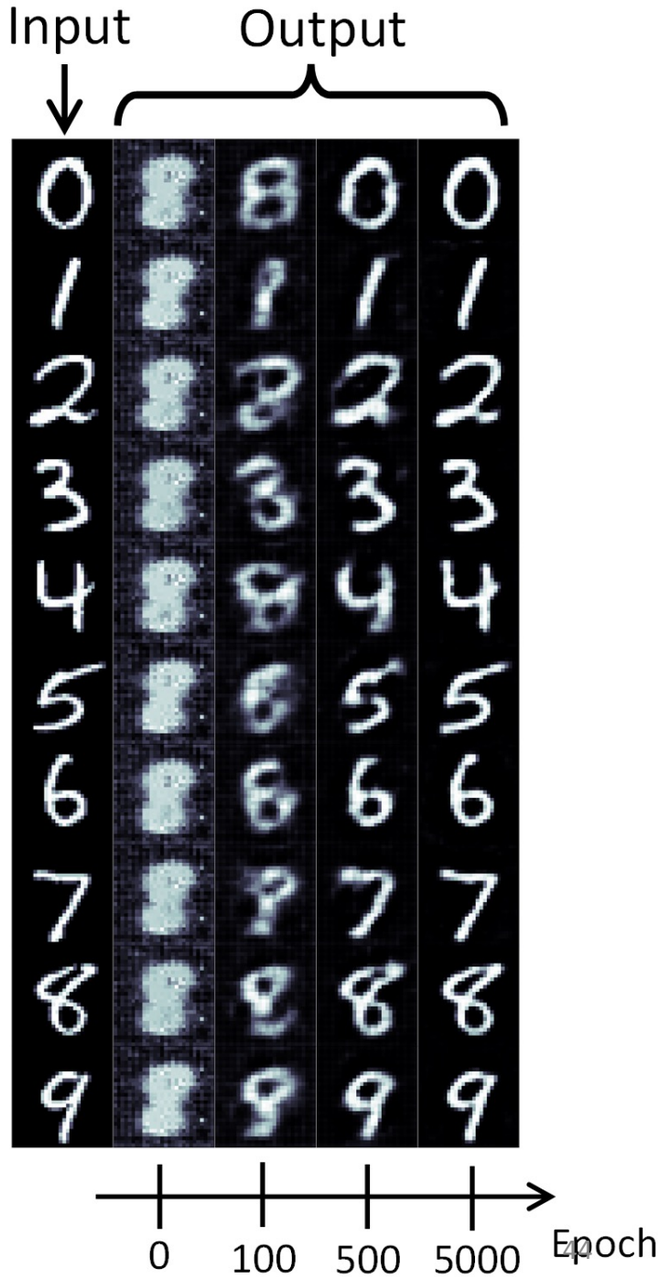


Mobility Model: rwp0500

Reconstructions at Node 8 with WAFL In case of Legitimate Inputs

1. WAFL (epoch 0 – after Self-Training)
   the autoencoder gave 8 to any inputs 0 ∼ 9.

2. As the WAFL training proceeds, WAFL's model aggregation allowed the precise reconstruction of all the legitimate samples (0 ∼ 9).



Reconstructions at Node 3 with WAFL In case of Global Anomaly Inputs.

1. WAFL (epoch 0 – after Self-Training)
   the autoencoder gave 3 to any global anomalies.

2. Even though the WAFL training proceeded,
   the autoencoder did not reconstruct the
   global anomaly input (which is succeeded).

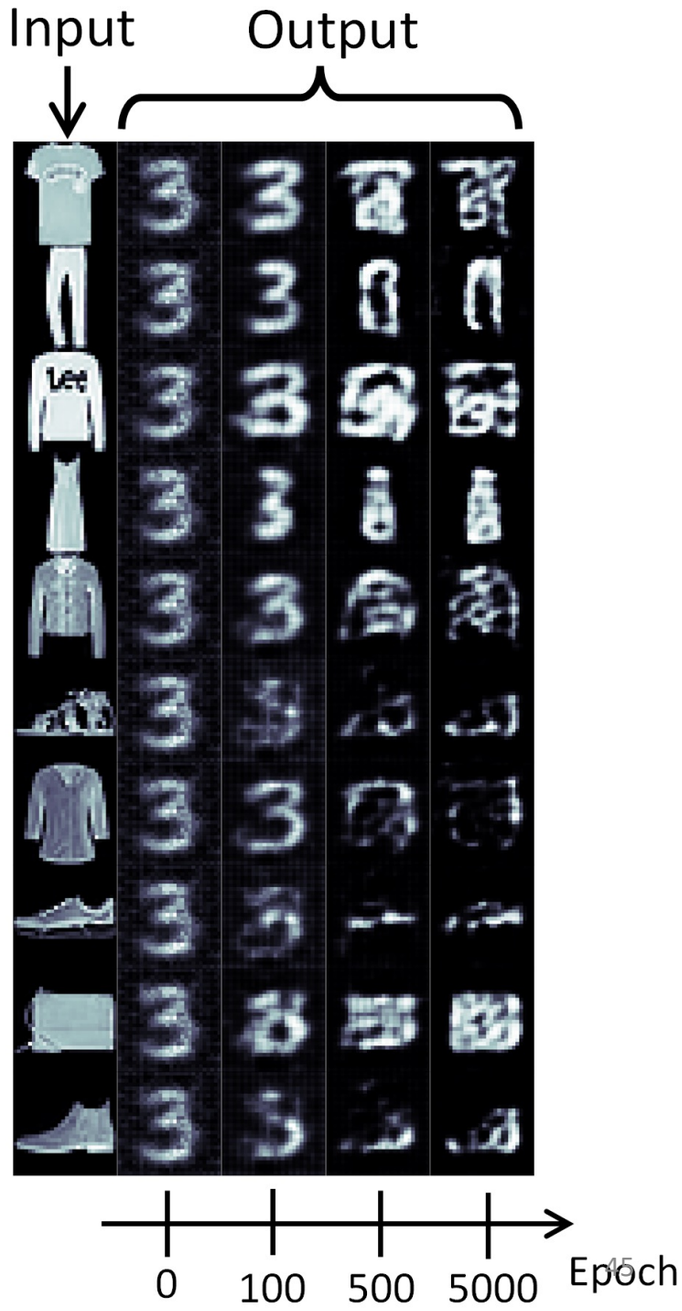


Table of Contents

Open Research Questions (Open Issues)

- MNIST is just a simple dataset ….
  - How about general images or videos?
  - How about text?
  - How about audio?
  - How about IoT or Industrial data?
    - Can we combine with sensors?
  - How about the logs of computers?
    - Can we use for network security?
- How about Generative models (e.g., GAN)?
- How about Multi-Domain Adaptation (Out-of-Distribution) Issues?
- How about the security of WAFL?
- How about the implementation?
- How about the protocol for discovery and model exchange?
- How can we operate WAFL as a learning system?

Conclusion

- A Fully Autonomous and Distributed Collaborative Machine Learning
  -- Wireless Ad Hoc Federated Learning (WAFL)

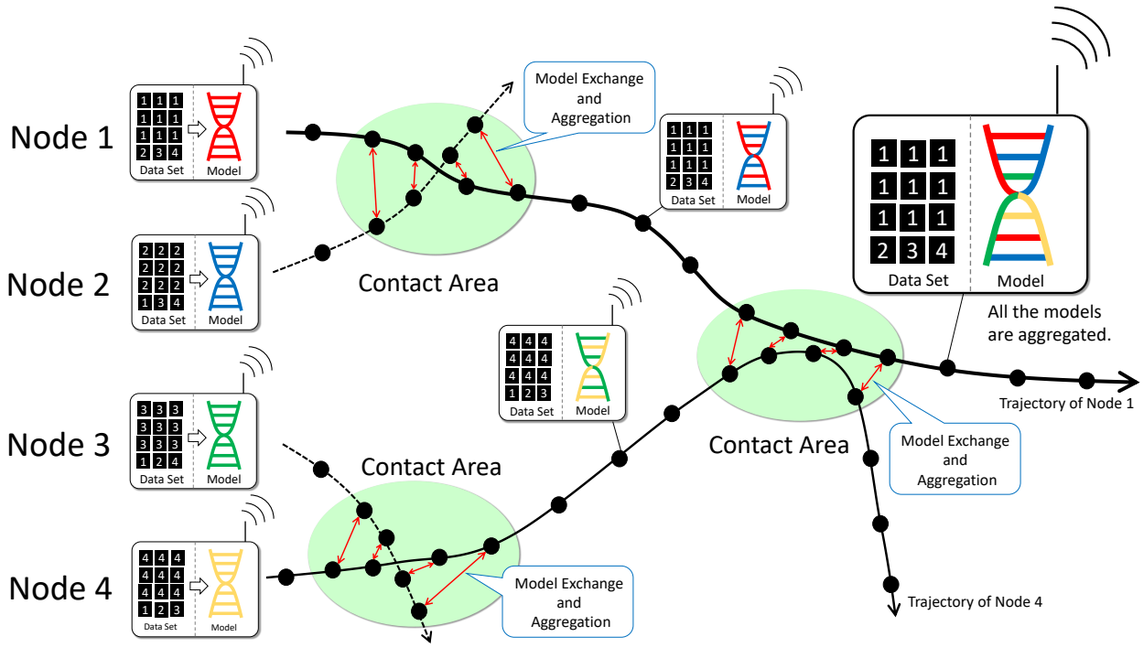
- Characteristics of WAFL
  - All the nodes are even – no centralized power mechanism.
  - WAFL allows multi-vendor system if protocol is defined.

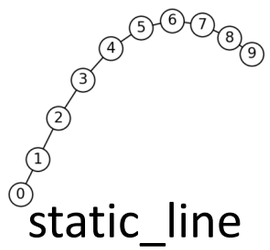
- Current Stage of Research
  - Benchmark-based evaluation in basic and application-oriented scenarios

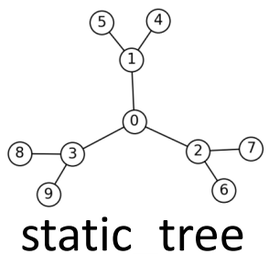
- Future Research Directions
  - Expansion into various applications
  - Implementation, operation, protocol design

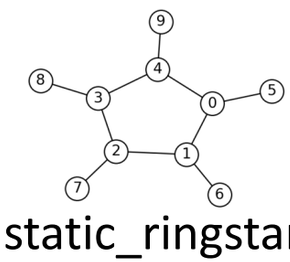
} We look for research partners.

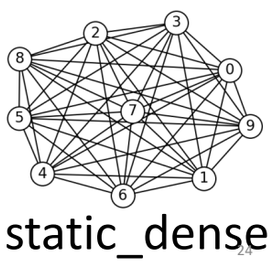
Thank you very much

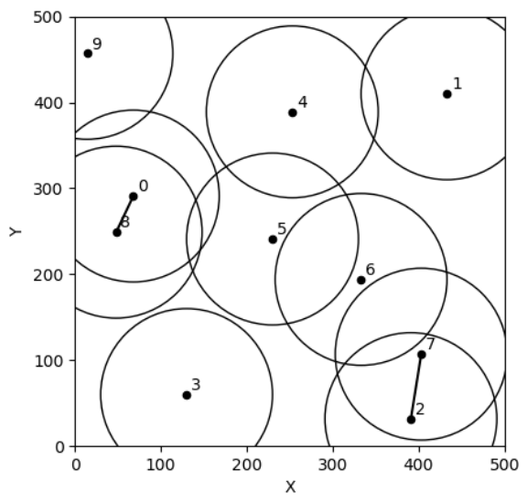


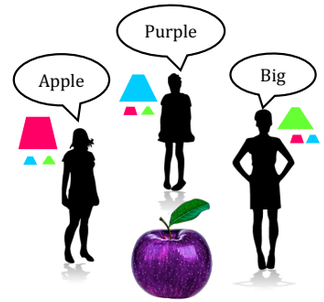




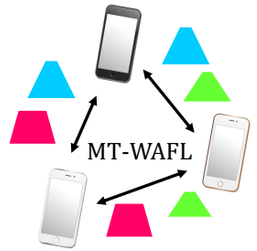




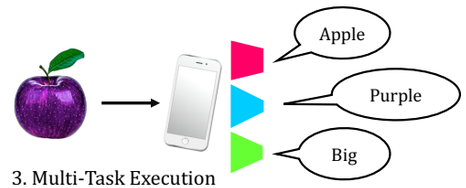




1. Labeling by their own perspective



2. Model mixture by MT-WAFL



3. Multi-Task Execution