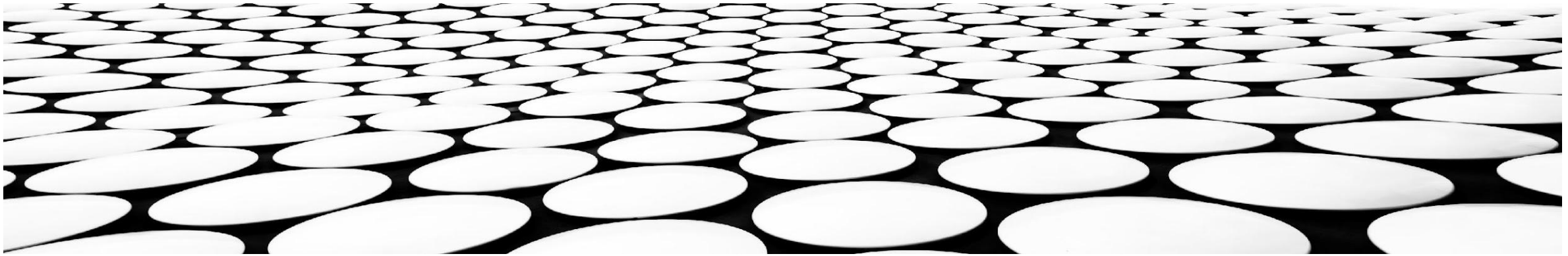
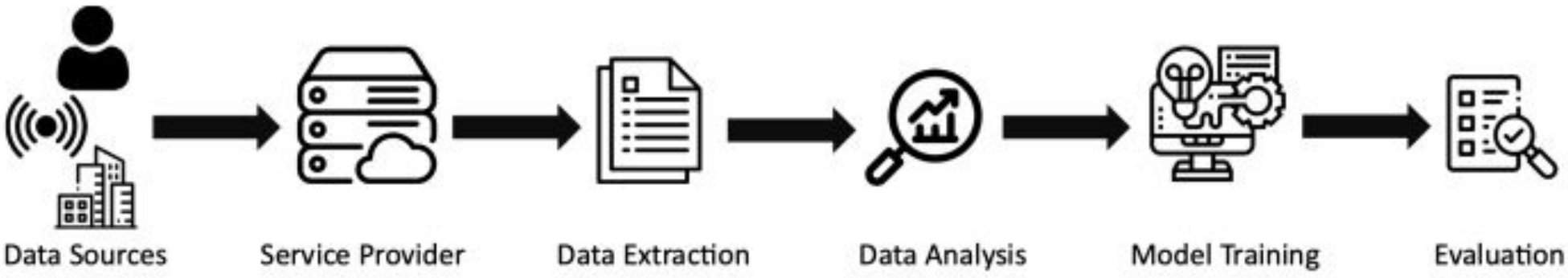
Ομοσπονδιακή Μάθηση
Federated Learning

Βασίλης Περηφάνης
03/04/2026

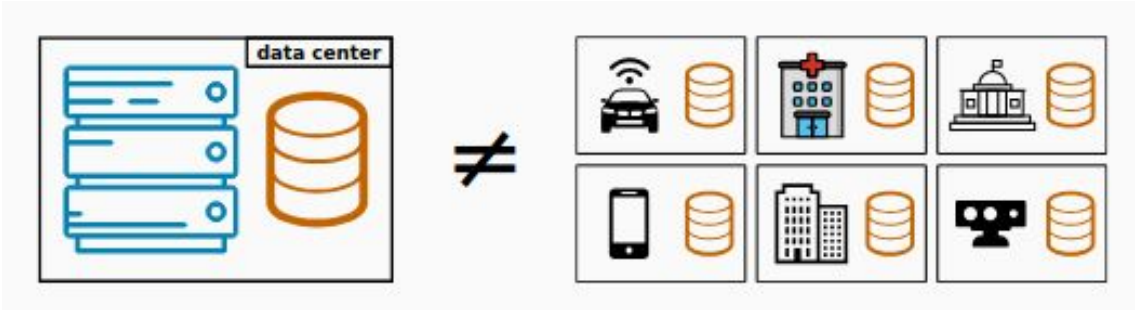


Περιεχόμενα

Μηχανική Μάθηση

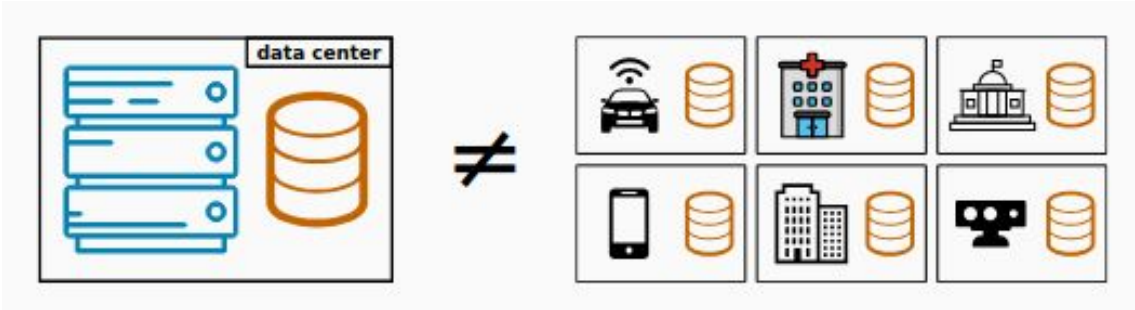


Γιατί FL



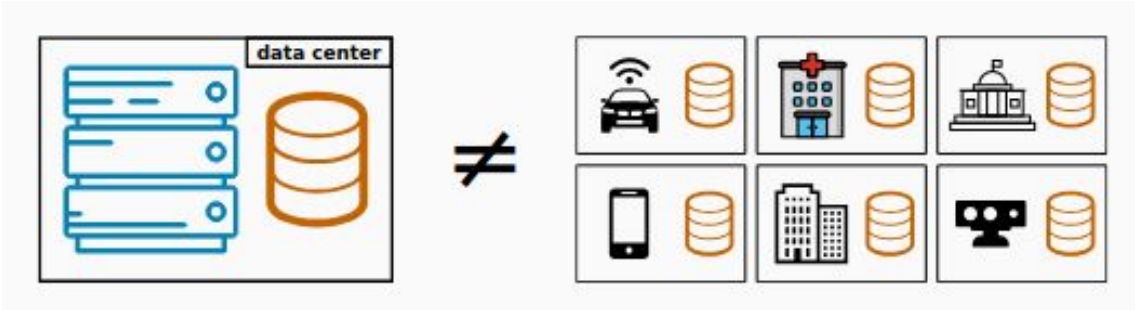
- Στον πραγματικό κόσμο, τα δεδομένα είναι κατανεμημένα σε πολλά μέρη
- Δε μπορούμε απλά να στείλουμε τα δεδομένα;

Γιατί FL



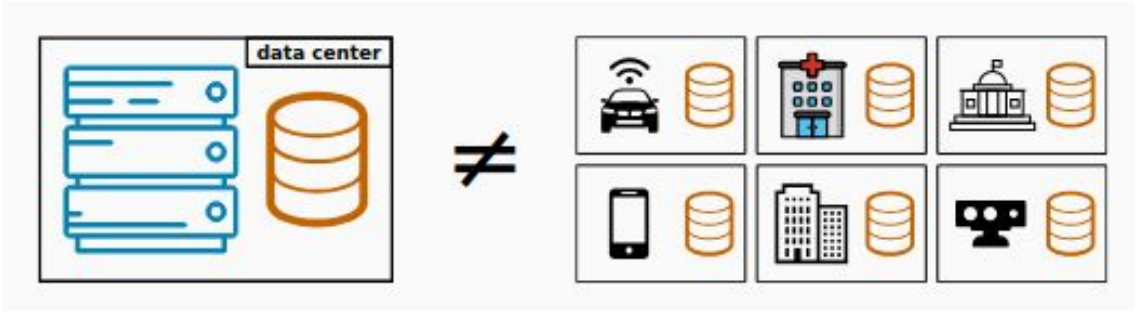
- Στον πραγματικό κόσμο, τα δεδομένα είναι κατανεμημένα σε πολλά μέρη
- Δε μπορούμε απλά να στείλουμε τα δεδομένα;
  - Πιθανώς να είναι αρκετά κοστοβόρο
  - Για παράδειγμα, μικρές συσκευές έχουν περιορισμένη μπαταρία ή μπορεί τα δεδομένα που παράγονται να είναι σε κλίμακα των TBs
- Ακόμη και να λύναμε το δικτυακό πρόβλημα, μπορούμε να στείλουμε ότι θέλουμε;

Γιατί FL



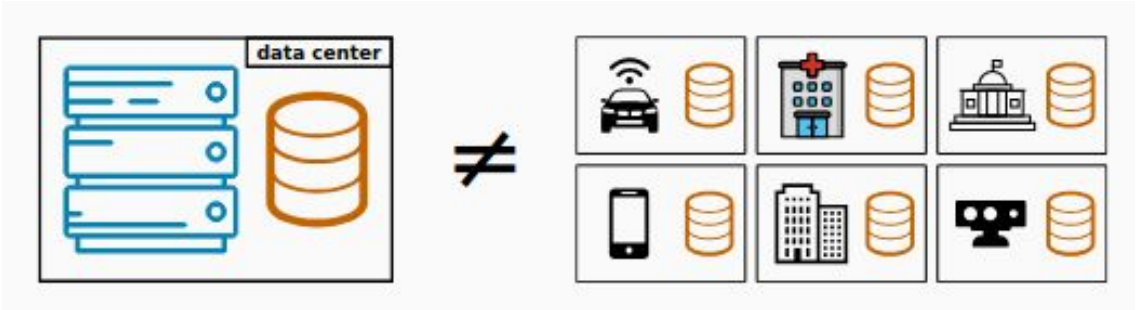
- Στον πραγματικό κόσμο, τα δεδομένα είναι κατανεμημένα σε πολλά μέρη
- Δε μπορούμε απλά να στείλουμε τα δεδομένα;
  - Πιθανώς να είναι αρκετά κοστοβόρο
  - Για παράδειγμα, μικρές συσκευές έχουν περιορισμένη μπαταρία ή μπορεί τα δεδομένα που παράγονται να είναι σε κλίμακα των TBs
- Ακόμη και να λύναμε το δικτυακό πρόβλημα, μπορούμε να στείλουμε ότι θέλουμε;
  - Σε πολλές περιπτώσεις, τα δεδομένα είναι ευαίσθητα
  - Π.χ., νοσοκομεία, εικόνες ανθρώπων, φορολογικά δεδομένα,...

Γιατί FL

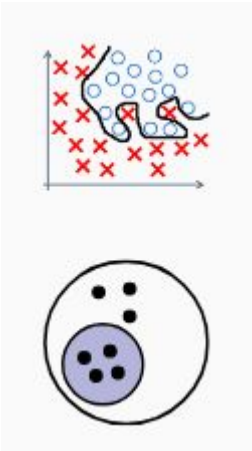


- Και γιατί κάθε ένας να μην εκτελεί αλγορίθμους μηχανικής μάθησης για τον εαυτό του;

Γιατί FL



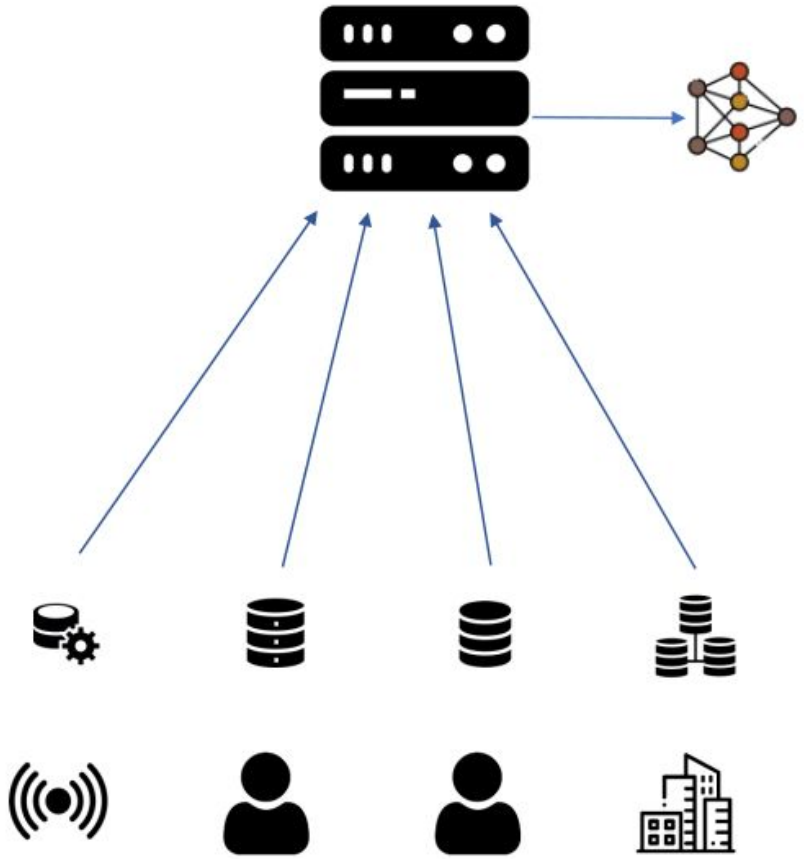
- Και γιατί κάθε ένας να μην εκτελεί αλγορίθμους μηχανικής μάθησης για τον εαυτό του;
  - Είναι αρκετά πιθανό τα τοπικά δεδομένα να είναι πολύ λίγα
    - Πολύ εύκολο overfitting
  - Είναι αρκετά πιθανό τα τοπικά δεδομένα να περιλαμβάνουν “διακρίσεις”
    - Να μην είναι αντιπροσωπευτικά μιας γενικής κατανομής

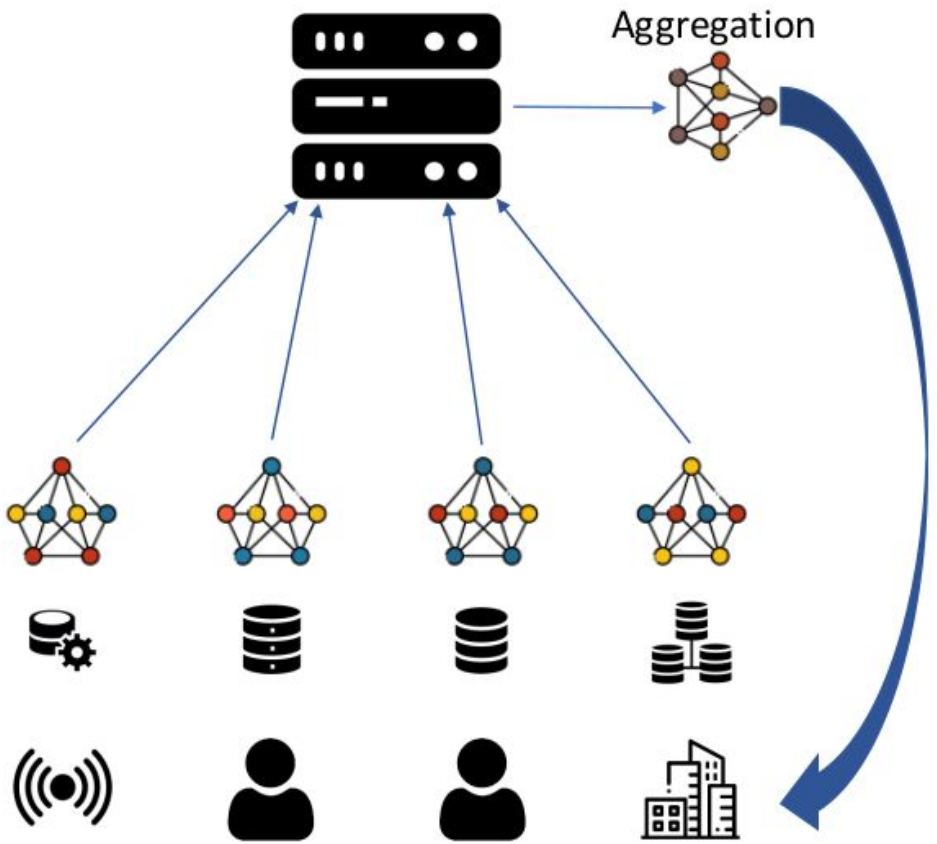


FL

- Η τεχνική Federated Learning είναι μια “ρύθμιση” κατά την διαδικασία της Μηχανικής Μάθησης όπου πολλά μέρη συνεργάζονται για την εκπαίδευση ενός μοντέλου, χωρίς να απαιτείται η αποστολή των δεδομένων τους
- Σκοπός είναι να δημιουργηθεί ένα μοντέλο με ακρίβεια παρόμοια με αυτή του μοντέλου που θα είχαμε, αν μαζέψουμε όλα τα δεδομένα σε ένα σημείο

Μηχανική Μάθηση vs FL



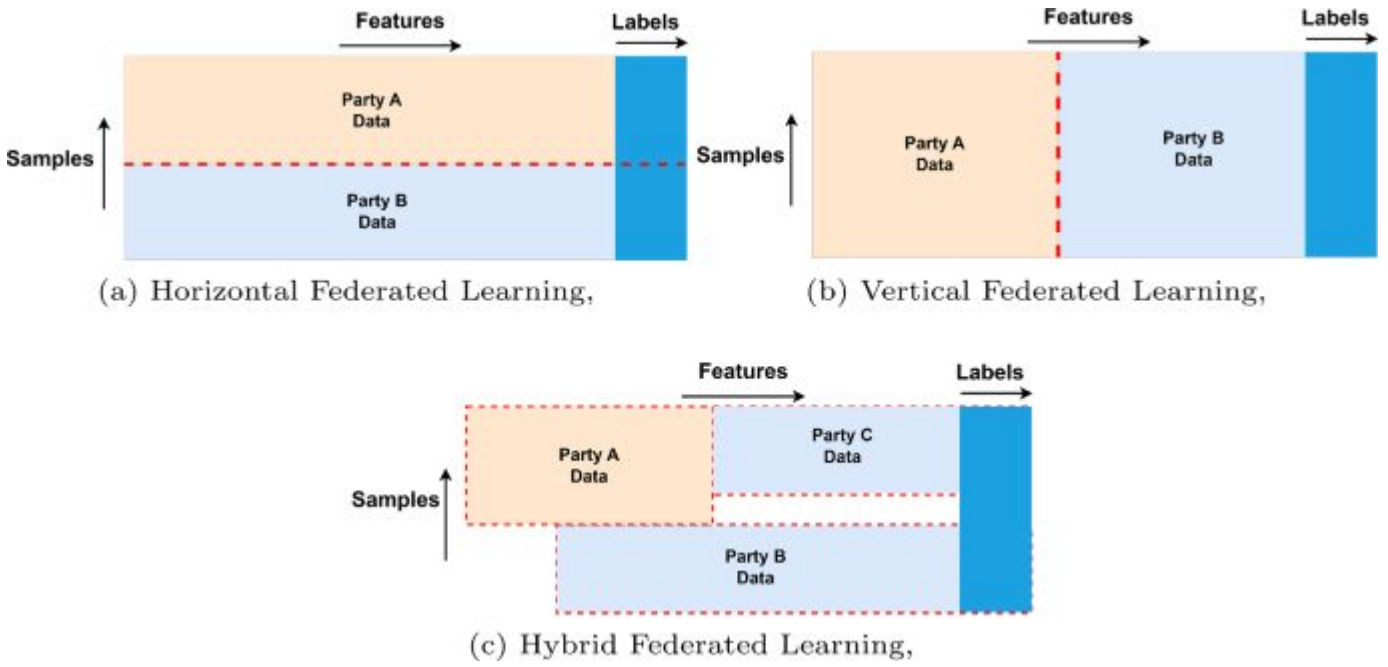


Διαφορές με την Κατανεμημένη Μηχανική Μάθηση

- Στην κατανεμημένη ΜΜ, τα δεδομένα είναι αποθηκευμένα κάπου κεντρικά
- Σκοπός είναι να γίνει γρηγορότερη εκπαίδευση
- Στο FL τα δεδομένα είναι απο τη φύση τους κατανεμημένα

Τύποι FL

Διαχωρισμός Δεδομένων FL



(a) Horizontal Federated Learning,

(b) Vertical Federated Learning,

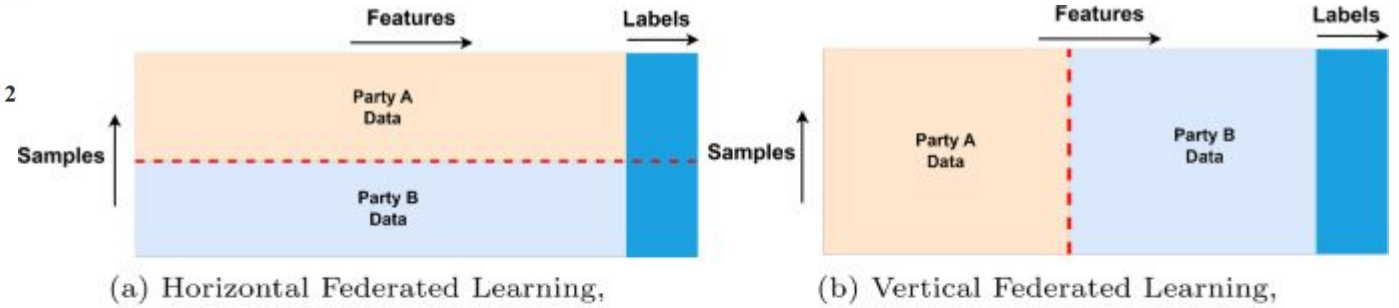
(c) Hybrid Federated Learning,

Διαχωρισμός Δεδομένων FL

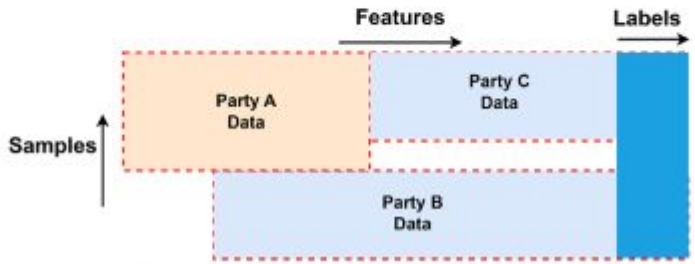
Features

Name	Age	Sex	Height	Weight	Label	
Person A	24	Male	178	78	1	Client 1
Person B	61	Female	165	64	0	
Person C	44	Male	182	89	1	
Person D	17	Female	159	52	0	Client 2
Person E	11	Male	137	36	1	
Person F	33	Female	171	60	0	

Samples



(a) Horizontal Federated Learning, (b) Vertical Federated Learning,



(c) Hybrid Federated Learning,

Διαχωρισμός Δεδομένων FL

Features

Samples

Name	Age	Sex	Height	Weight	Label	
Person A	24	Male	178	78	1	Client 1
Person B	61	Female	165	64	0	
Person C	44	Male	182	89	1	
Person D	17	Female	159	52	0	Client 2
Person E	11	Male	137	36	1	
Person F	33	Female	171	60	0	

Features

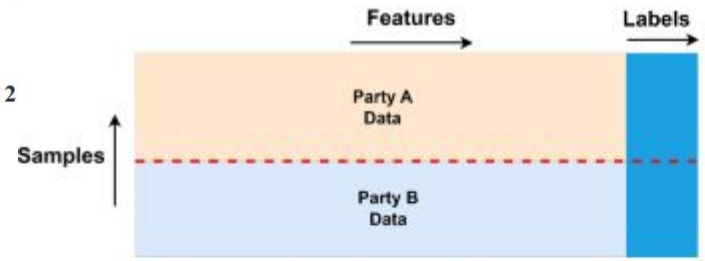
Samples

Name	Age	Height	Label
Person A	24	178	1
Person B	61	165	0
Person C	44	182	1
Person D	17	159	0
Person E	11	137	1
Person F	33	171	0

Client 1

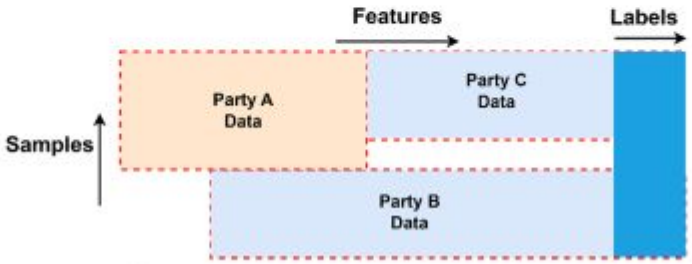
Name	Sex	Weight
Person A	Male	78
Person B	Female	64
Person C	Male	89
Person D	Female	52
Person E	Male	36
Person F	Female	60

Client 2



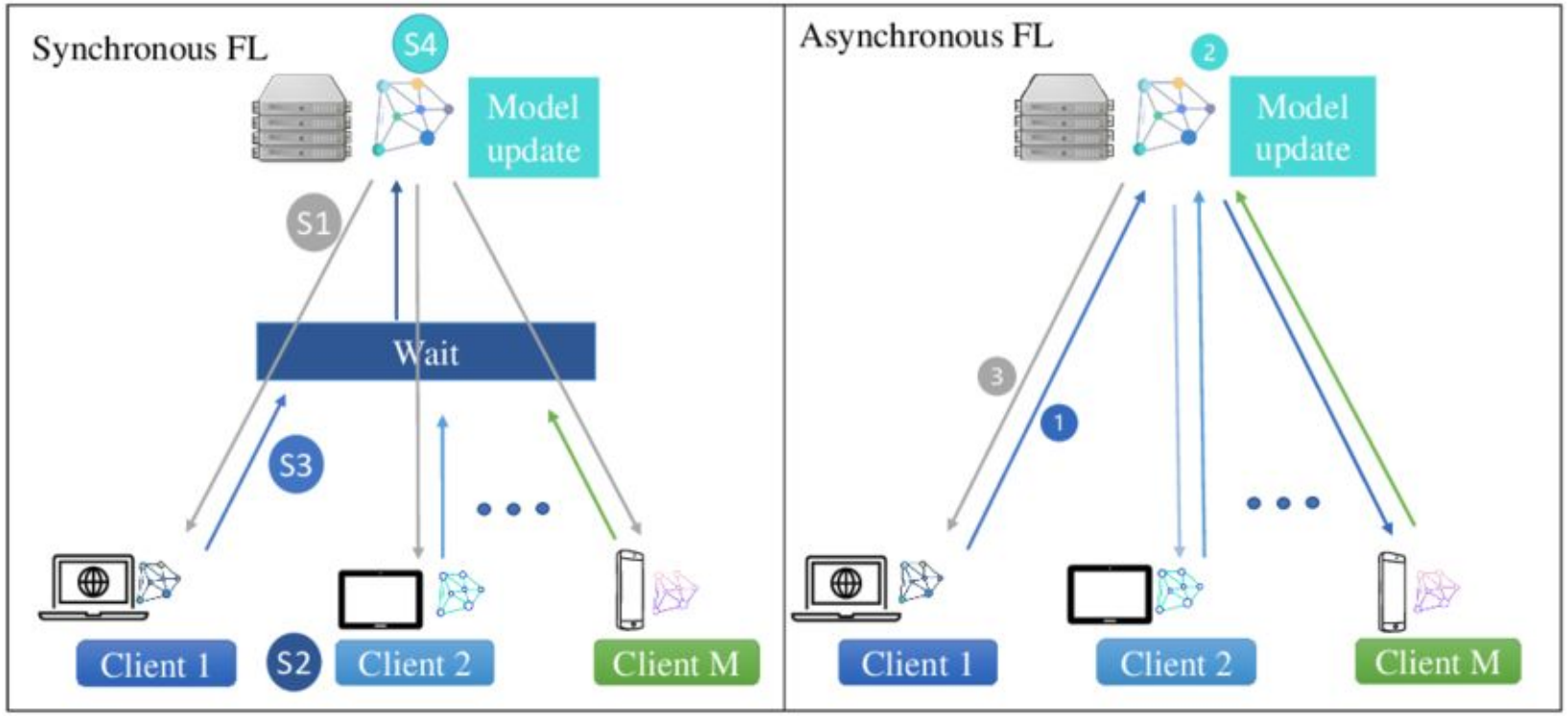
(a) Horizontal Federated Learning,

(b) Vertical Federated Learning,



(c) Hybrid Federated Learning,

Δικτύωση FL


Synchronous FL
S4
Model update
S1
Wait
S3
S2
Client 1
Client 2
Client M
Asynchronous FL
2
Model update
3
1
Client 1
Client 2
Client M

Εφαρμογές FL

- Google
  - Gboard
    - next word prediction, emoji prediction, new word prediction

Εφαρμογές FL

- Apple
  - Siri

MIT Technology Review

Sign in Subscribe

Artificial intelligence / Machine learning

How Apple personalizes Siri without hoovering up your data

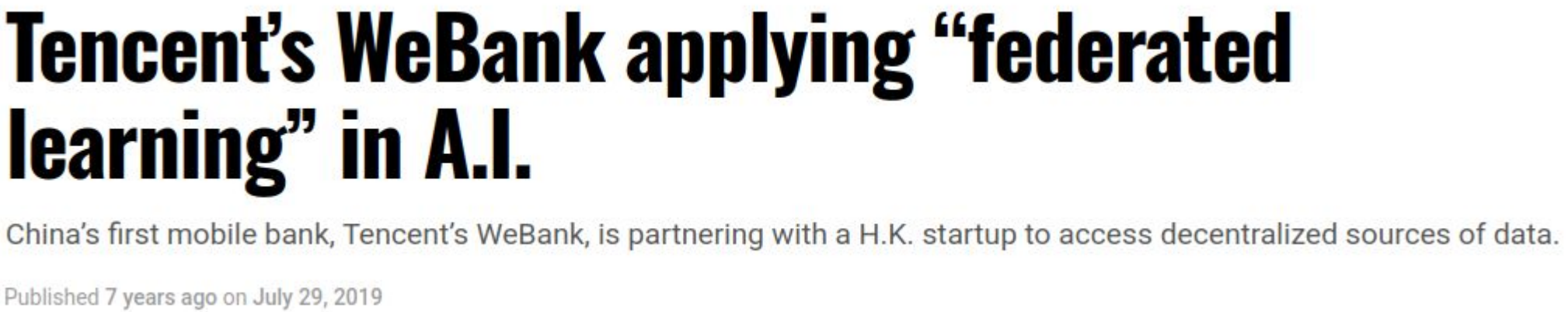
The tech giant is using privacy-preserving machine learning to improve its voice assistant while keeping your data on your phone.

by **Karen Hao** December 11, 2019

Εφαρμογές FL

- Brave
  - News recommendation

Εφαρμογές FL



Tencent's WeBank applying "federated learning" in A.I.

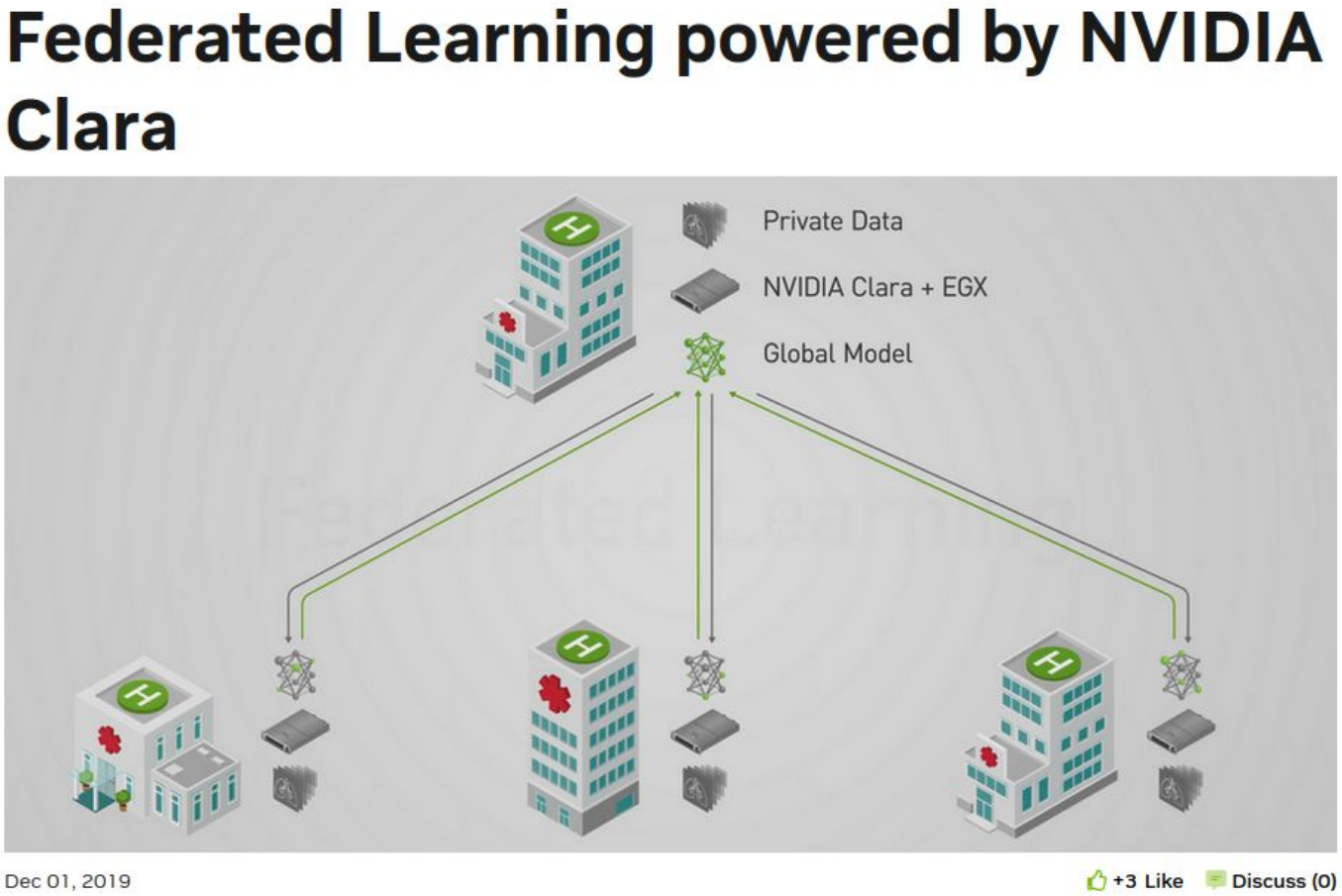
China's first mobile bank, Tencent's WeBank, is partnering with a H.K. startup to access decentralized sources of data.

Published 7 years ago on July 29, 2019

- WeBank
  - News recommendation

Εφαρμογές FL

- NVIDIA



Αλγόριθμος FL

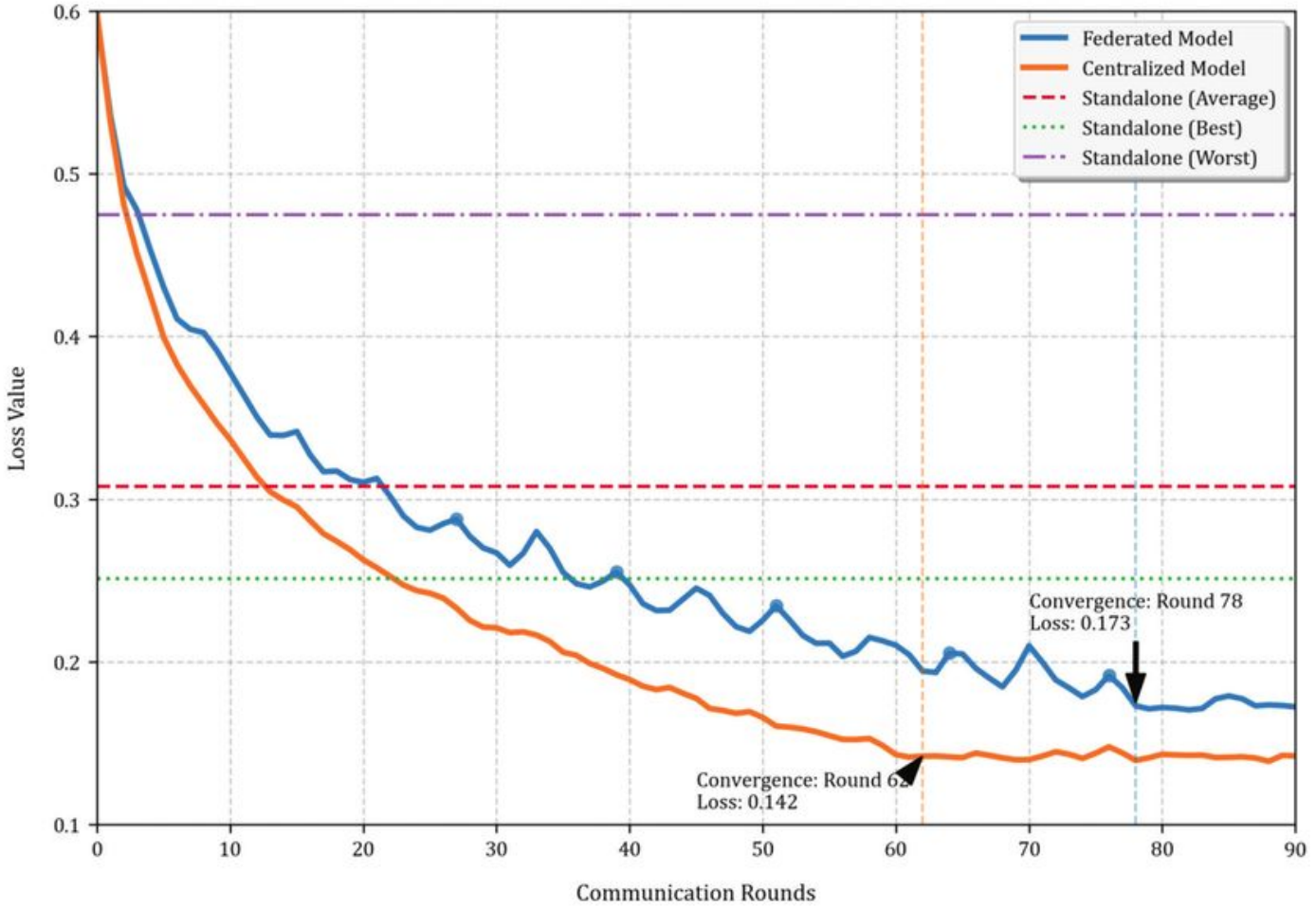
- Έστω K συμμετέχοντες
- Κάθε συμμετέχων έχει ένα τοπικό σετ δεδομένων μεγέθους nk
- Θεωρούμε ότι αν ενώσουμε όλα τα σετ δεδομένων, θα έχουμε λάβει ένα ενιαίο σετ δεδομένων
- Σκοπός είναι να εκπαιδεύσουμε ένα μοντέλο χωρίς να πάρουμε τα ενδιάμεσα δεδομένα και ταυτόχρονα να πετύχουμε ακρίβεια παρόμοια με το μοντέλο που θα εκπαιδεύαμε αν είχαμε πρόσβαση σε όλα τα δεδομένα.

Algorithm 1 `FederatedAveraging`. The K clients are indexed by k; B is the local minibatch size, E is the number of local epochs, and η is the learning rate.

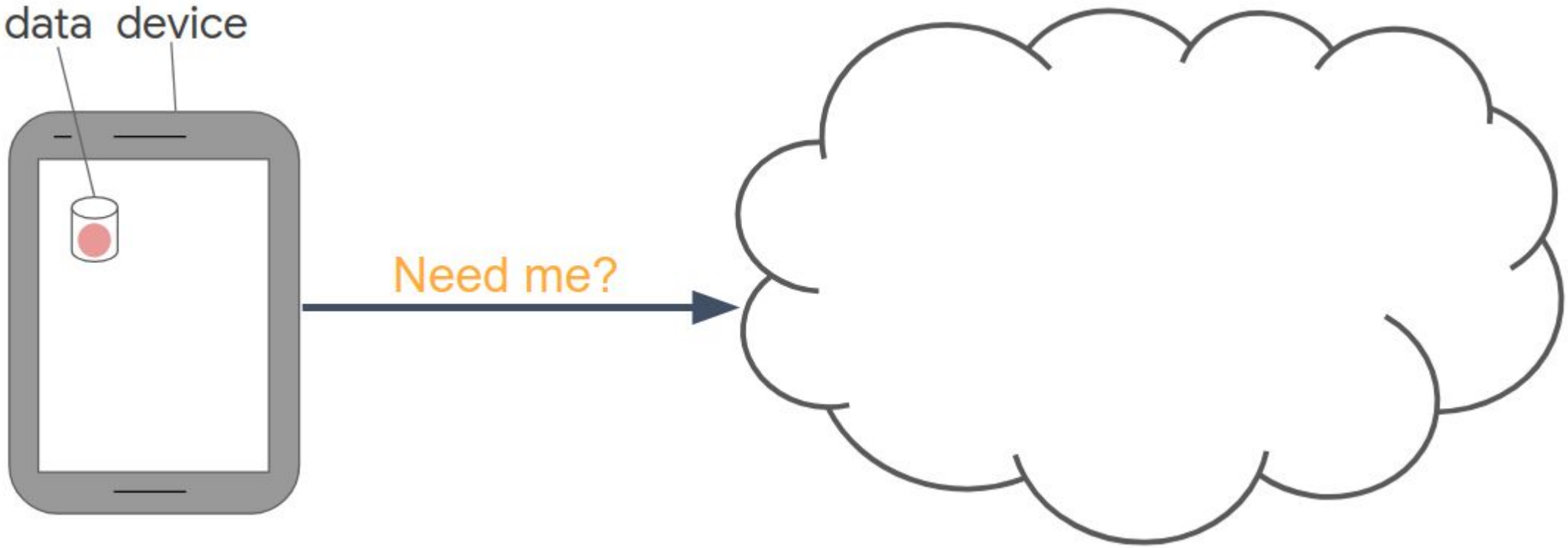
Server executes:
initialize w_0
for each round $t = 1, 2, \dots$ **do**
$m \leftarrow \max(C \cdot K, 1)$
$S_t \leftarrow$ (random set of m clients)
for each client $k \in S_t$ **in parallel do**
$w_{t+1}^k \leftarrow \text{ClientUpdate}(k, w_t)$
$w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$

ClientUpdate(k, w): *// Run on client k*
$\mathcal{B} \leftarrow$ (split $\mathcal{P}_k$ into batches of size B)
for each local epoch i from 1 to E **do**
for batch $b \in \mathcal{B}$ **do**
$w \leftarrow w - \eta \nabla \ell(w; b)$
return w to server

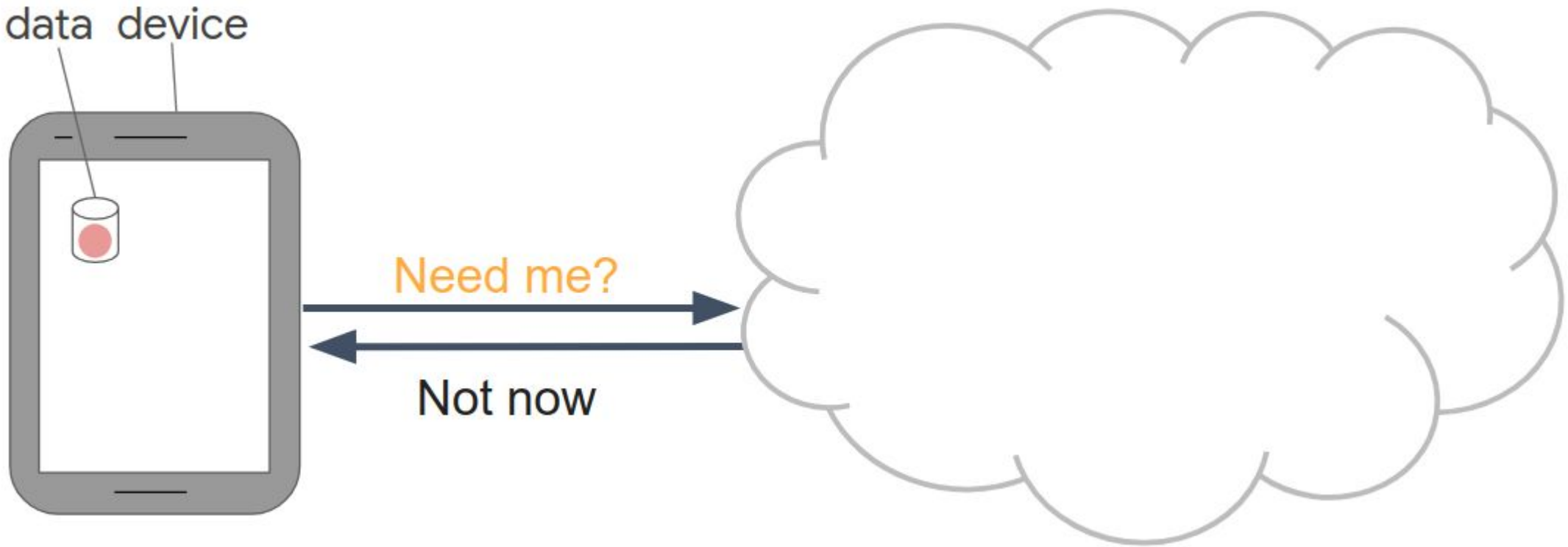
Σύγκλιση


Federated Model
Centralized Model
Standalone (Average)
Standalone (Best)
Standalone (Worst)
Loss Value
Communication Rounds
Convergence: Round 78
Loss: 0.173
Convergence: Round 62
Loss: 0.142
0.6
0.5
0.4
0.3
0.2
0.1
0
10
20
30
40
50
60
70
80
90

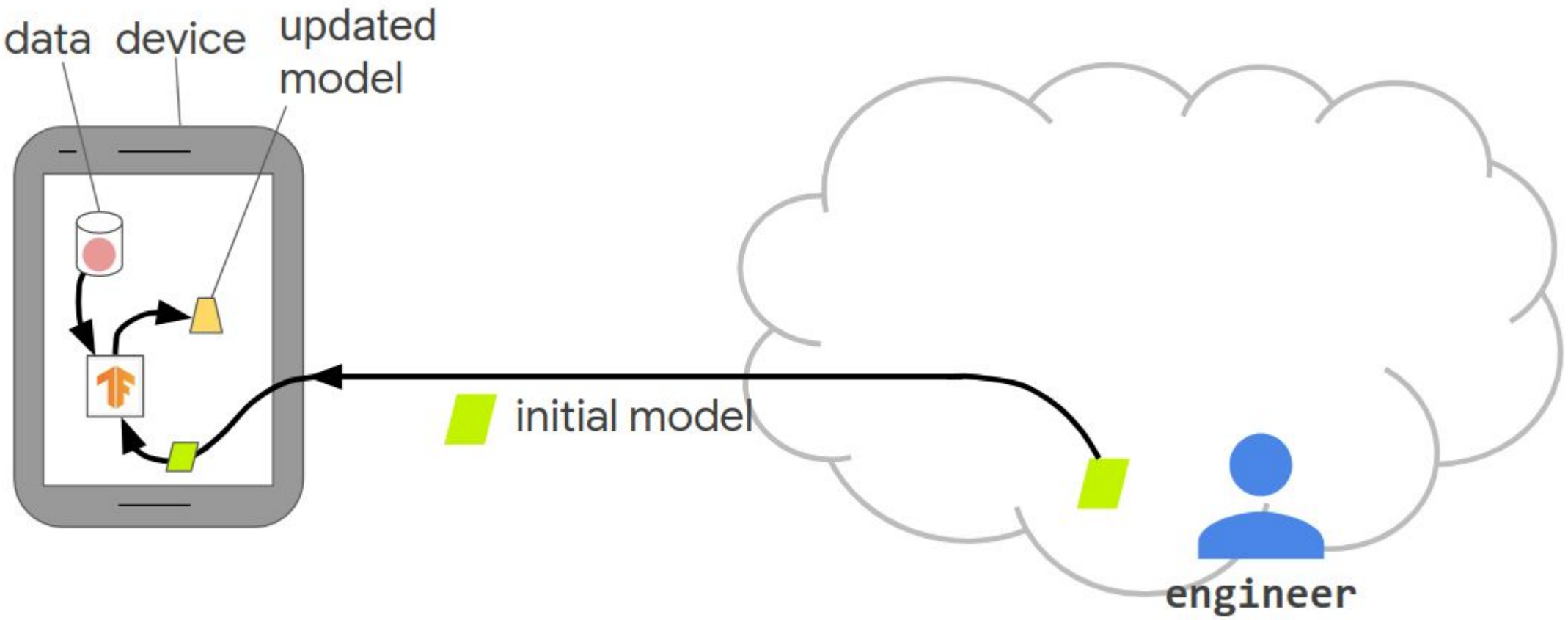
FL Βήμα-προς-Βήμα



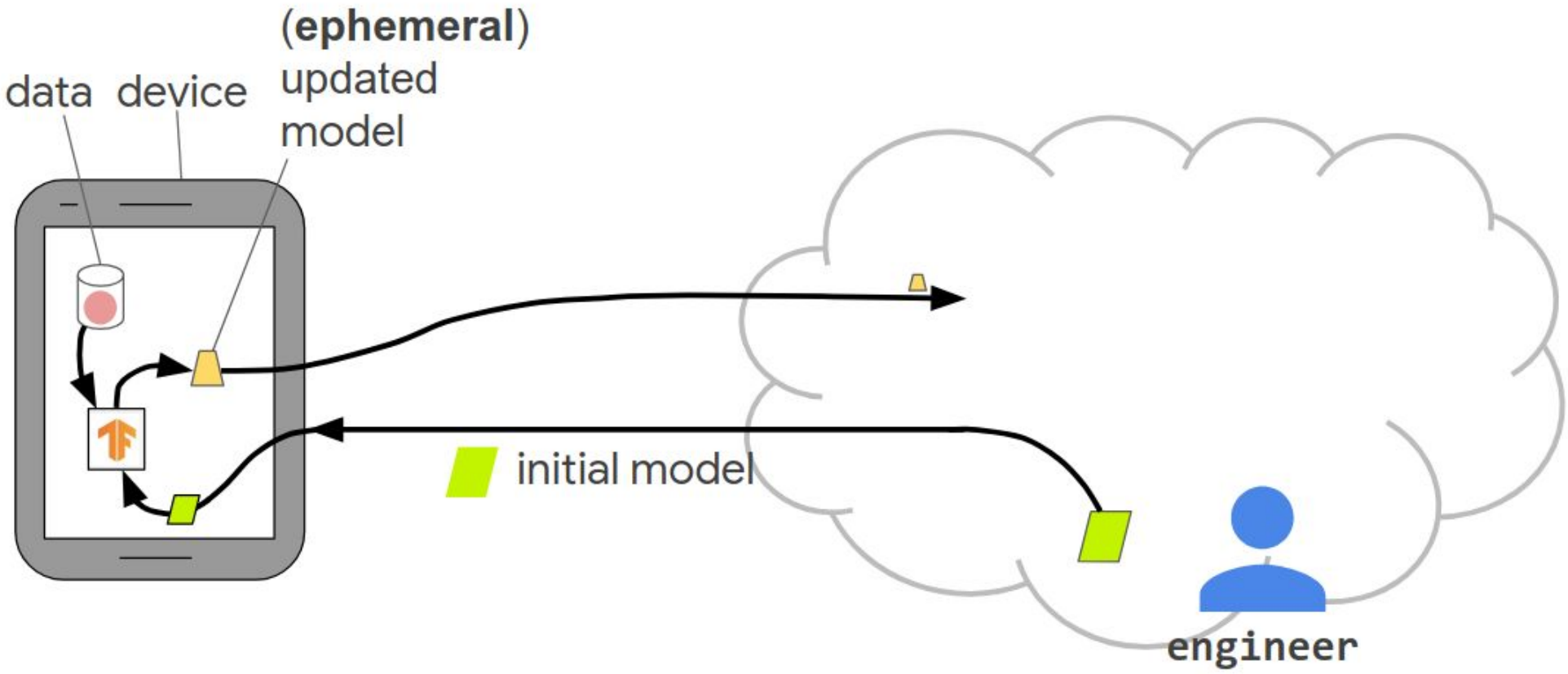
FL Βήμα-προς-Βήμα



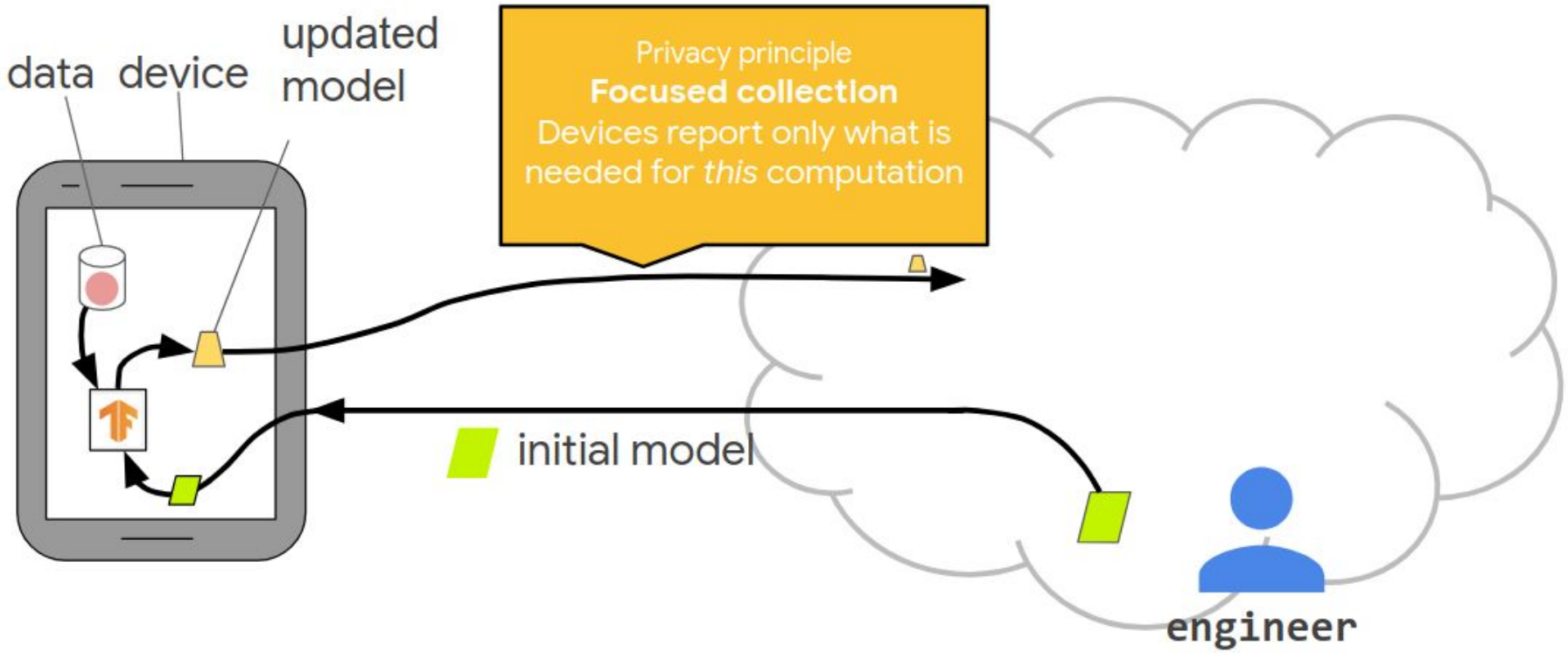
FL Βήμα-προς-Βήμα



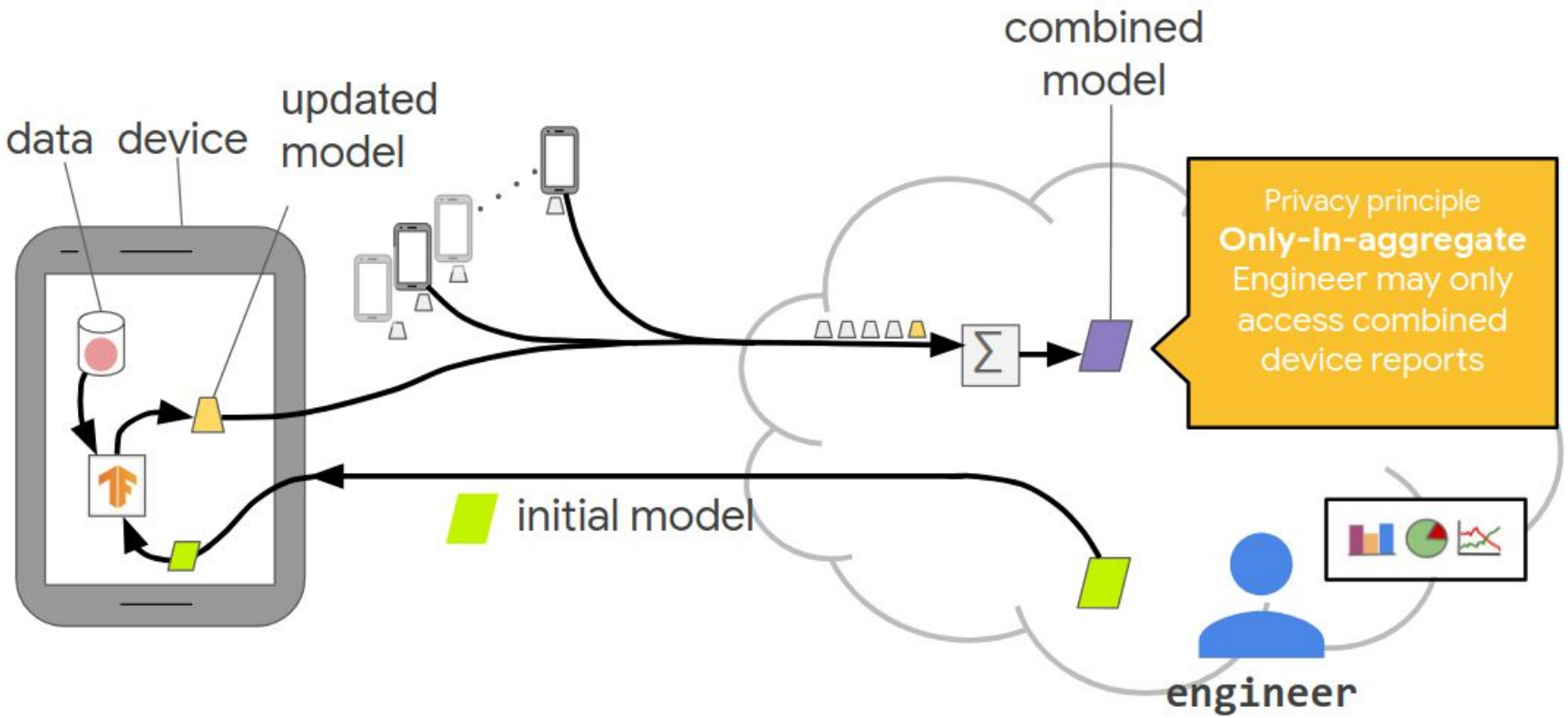
FL Βήμα-προς-Βήμα



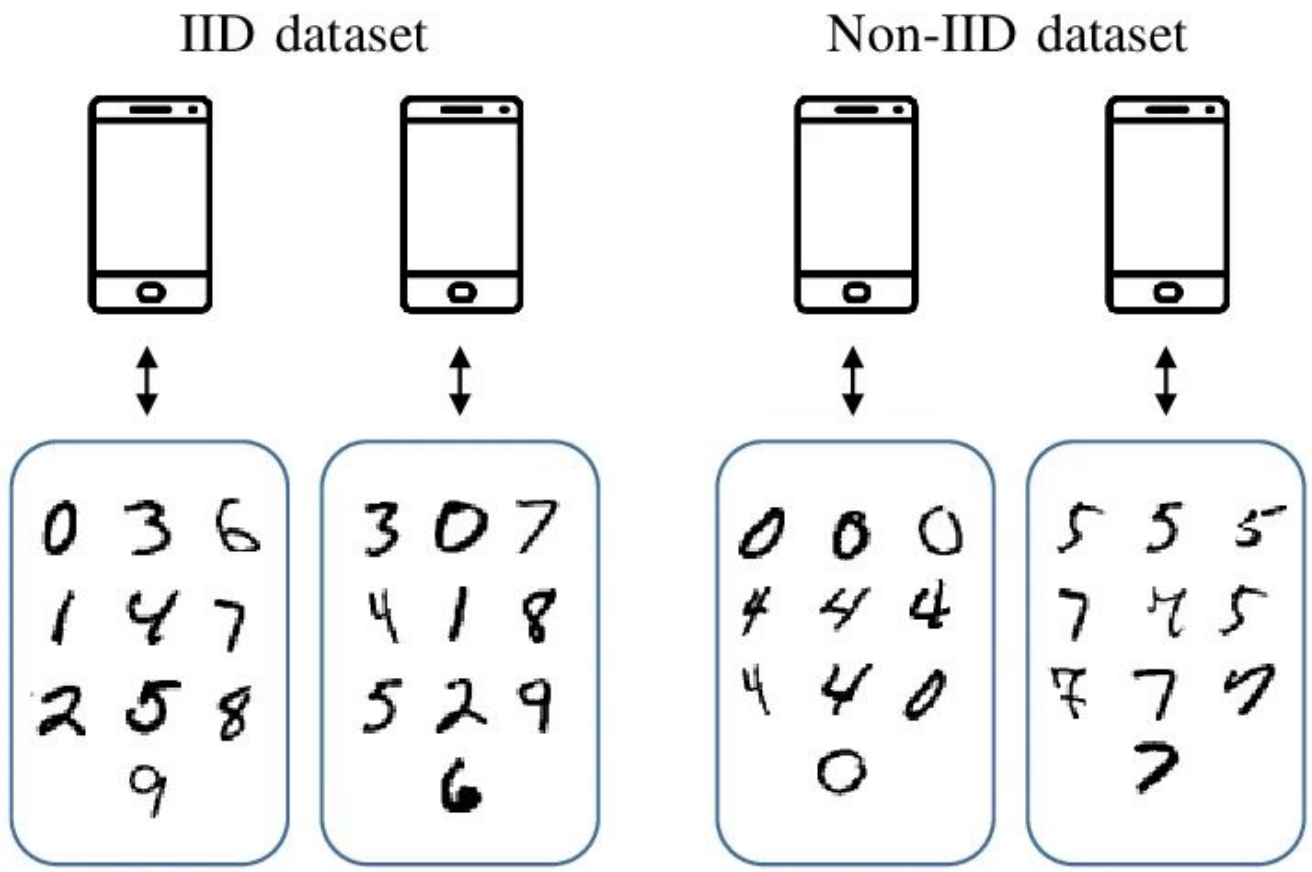
FL Βήμα-προς-Βήμα



FL Βήμα-προς-Βήμα



Προβλήματα στο FL: NON-I.I.D Data

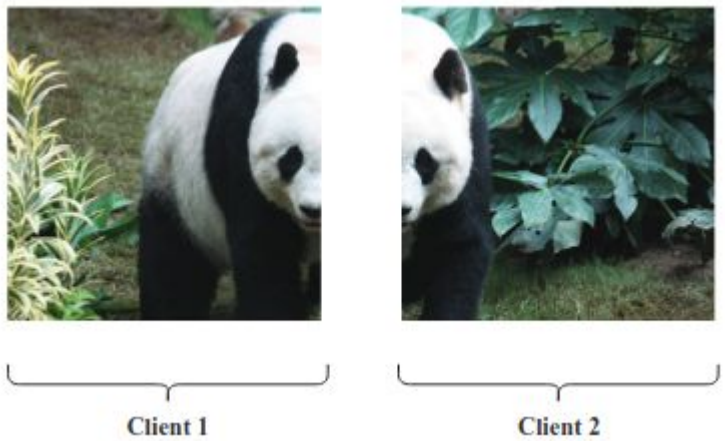


Προβλήματα στο FL: NON-I.I.D Data

- Πού μπορεί να υπάρξουν non-iid data?
  - Παντού

Προβλήματα στο FL: NON-I.I.D Data

- Πού μπορεί να υπάρξουν non-iid data?
  - Κατανομή εισόδου



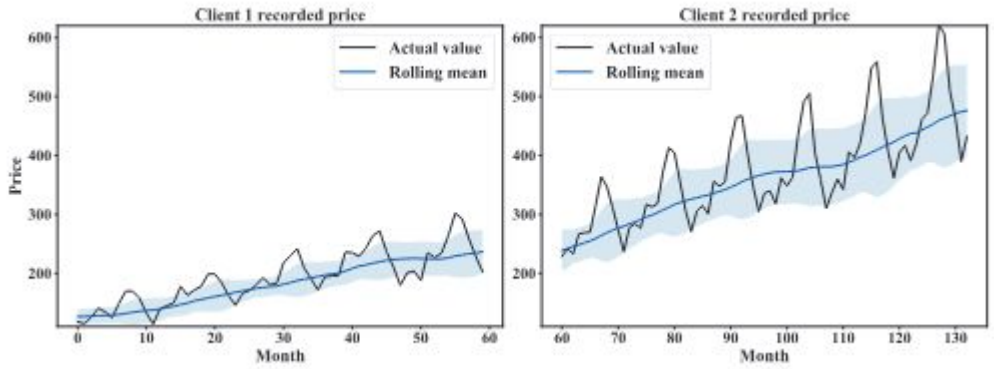
Προβλήματα στο FL: NON-I.I.D Data

- Πού μπορεί να υπάρξουν non-iid data?
  - Κατανομή εξόδου

	bird	deer	frog	ship
Client 1				
Client 2				

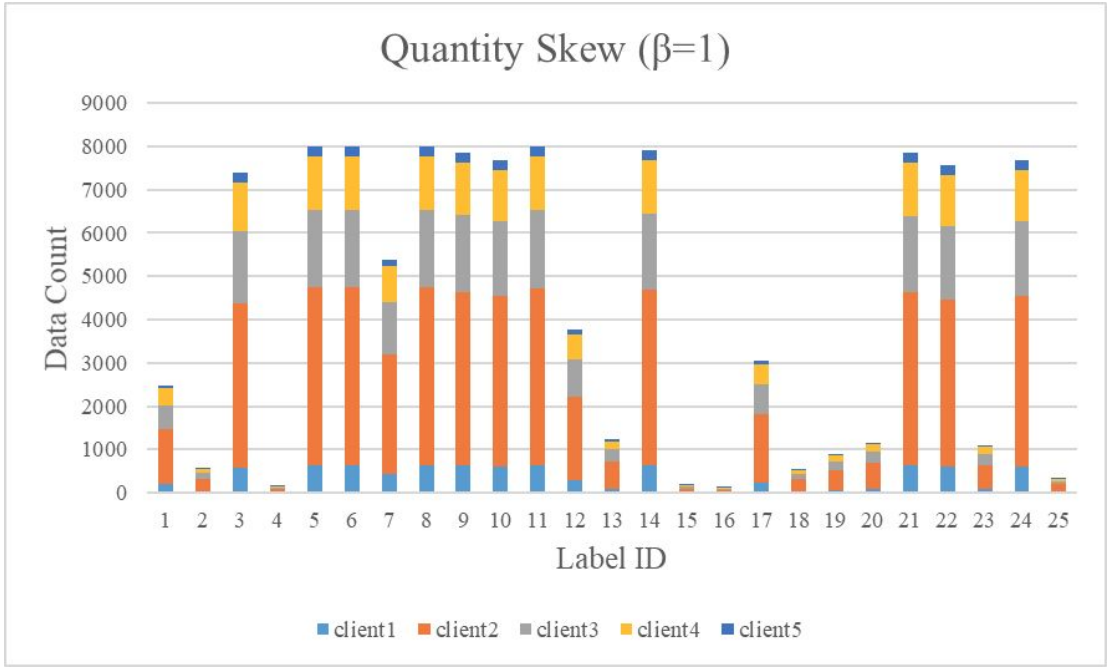
Προβλήματα στο FL: NON-I.I.D Data

- Πού μπορεί να υπάρξουν non-iid data?
  - Χρονικότητα

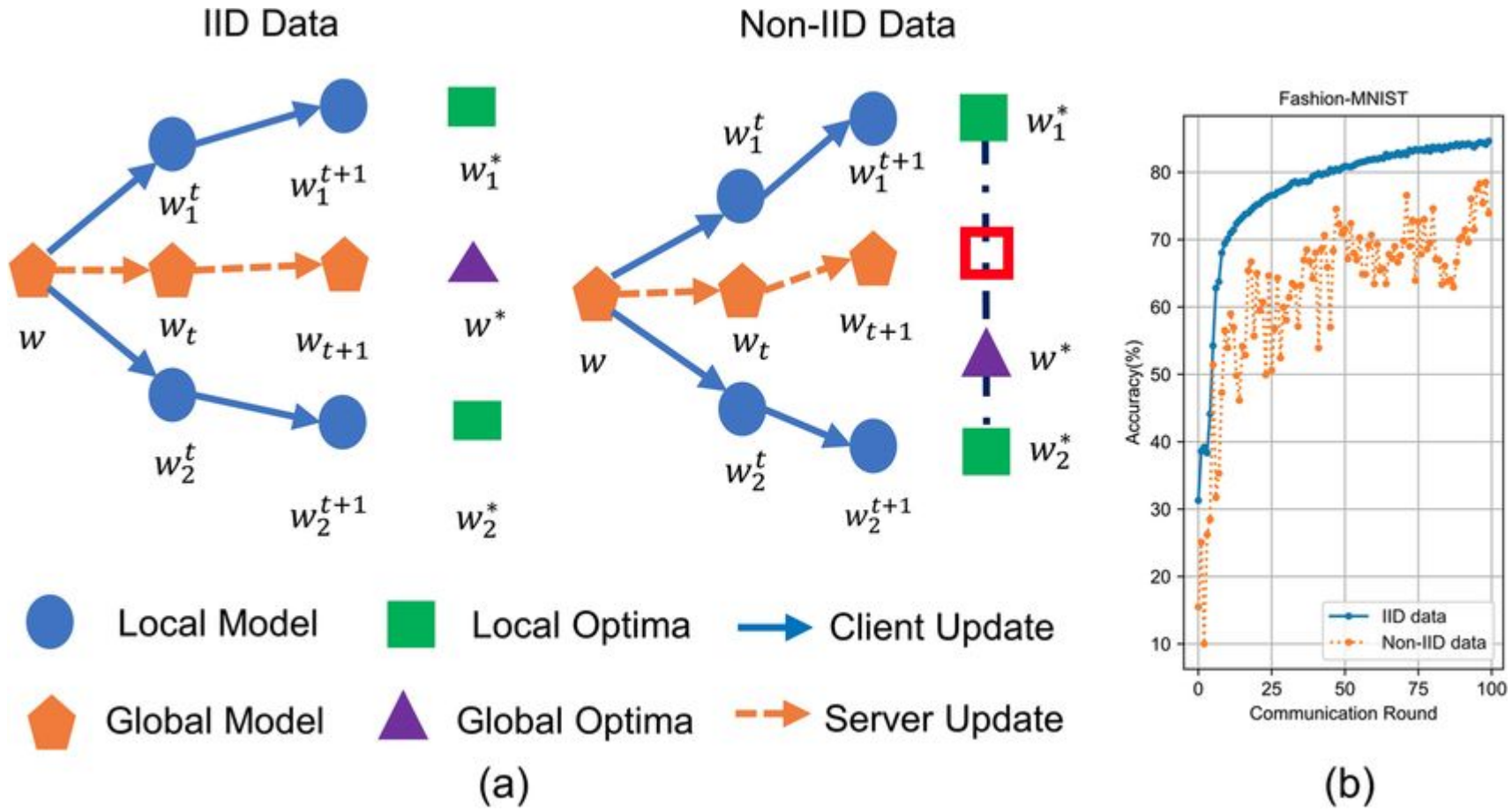


Προβλήματα στο FL: NON-I.I.D Data

- Πού μπορεί να υπάρξουν non-iid data?
  - Αριθμός δειγμάτων

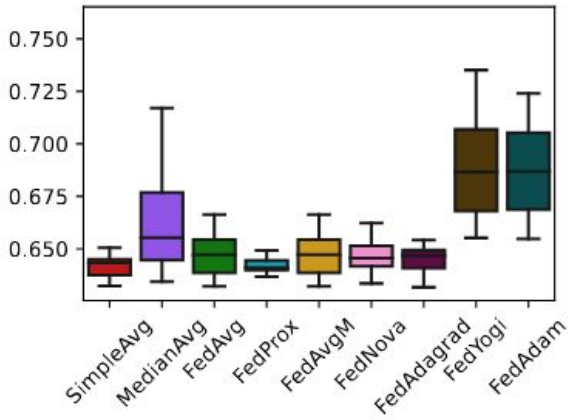


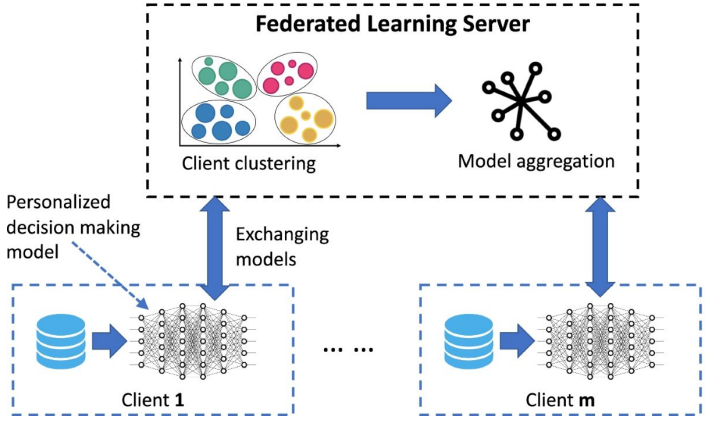
Προβλήματα στο FL: NON-I.I.D Data



Προβλήματα στο FL: NON-I.I.D Data - Λύσεις

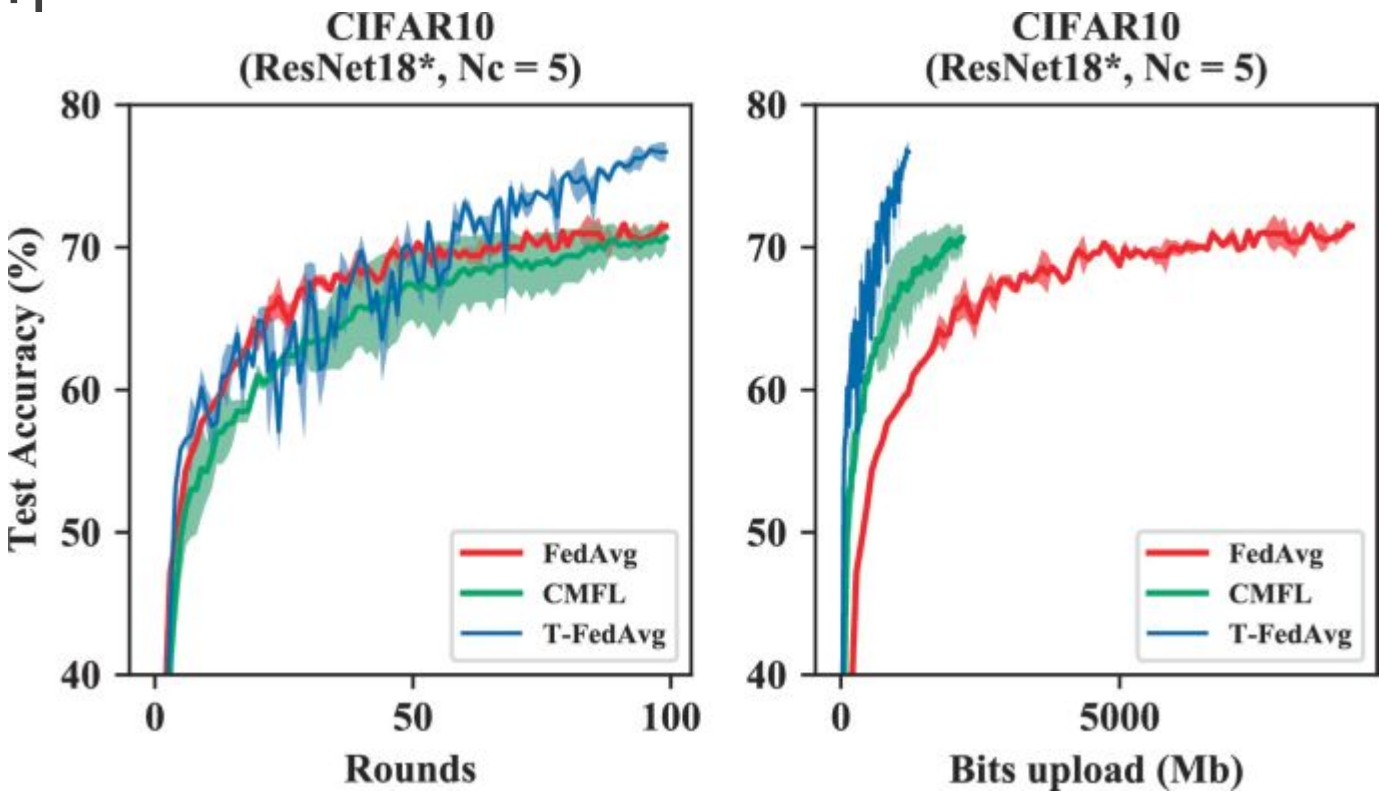
- Διαμοιρασμός (κάποιων) δεδομένων
  - Καταπατείται ωστόσο μια απο τις βασικές αρχές του FL
- Αλγοριθμικές μέθοδοι βελτιστοποίησης
  - Χρήση άλλων aggregation συναρτήσεων
  - Τοπικό fine-tuning
  - Χρήση “στρωμάτων” personalization
- Συστεμικές Μέθοδοι
  - Ομαδοποίηση χρηστών και εκτέλεση FL σε κάθε ομάδα





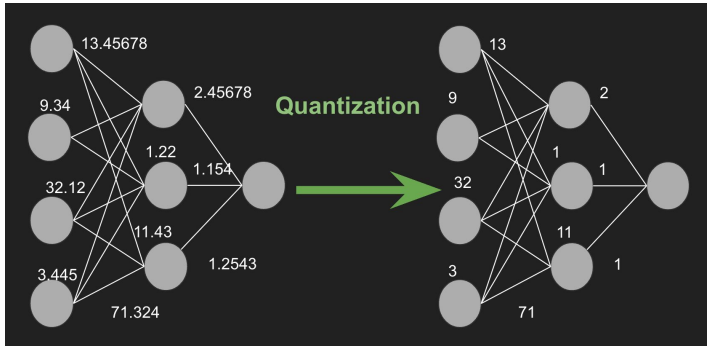
Προβλήματα στο FL: Δικτύωση

- Μεγάλα μοντέλα ⇒ πολλά MB/GB δεδομένων
- Πολλές συσκευές (κινητά, IoT) έχουν χαμηλή ταχύτητα σύνδεσης
- Συμφόρηση δικτύου όταν συμμετέχουν πολλοί clients
- Καθυστέρηση στην αποστολή/λήψη ενημερώσεων
- Διαφορετικές ταχύτητες σύνδεσης ανά client
- Μεγάλη χρήση δεδομένων ⇒ κατανάλωση μπαταρίας

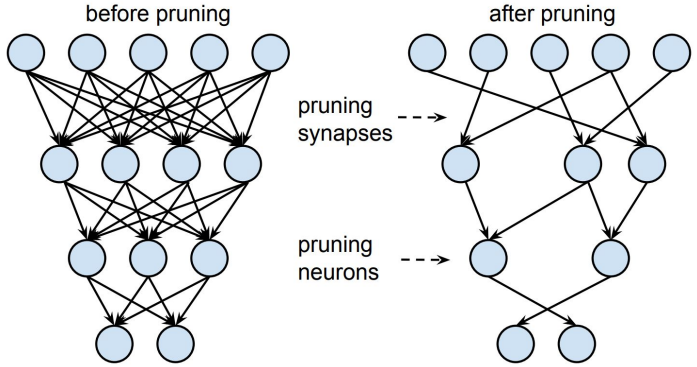


Προβλήματα στο FL: Δικτύωση - Λύσεις

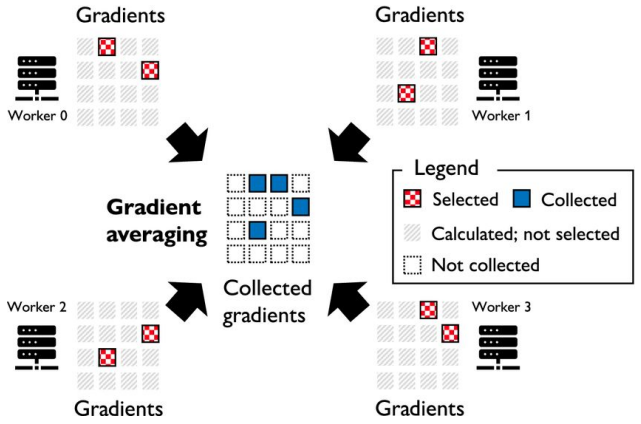
- Quantization



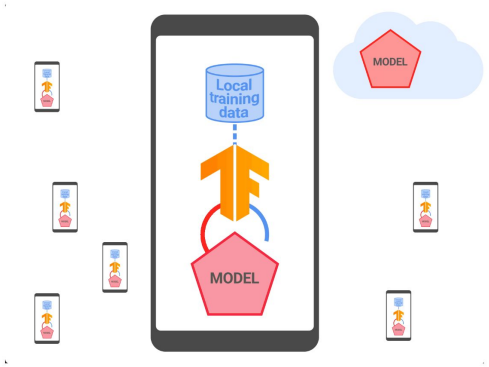
- Pruning



- Top-k gradients

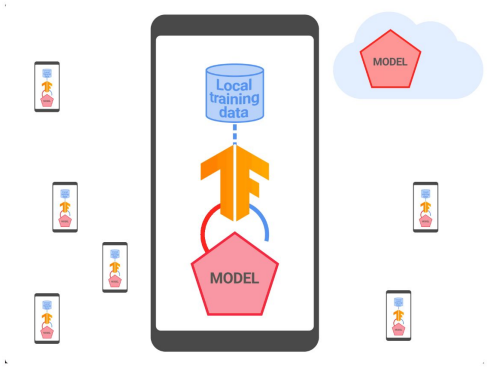


Προβλήματα στο FL: Ιδιωτικότητα



- Ωραία, με το Federated Learning κρατάμε τα δεδομένα τοπικά. Άρα είμαστε εντάξει σε σχέση με την αποστολή δεδομένων και την συμμόρφωση με κανονισμούς προστασίας;

Προβλήματα στο FL: Ιδιωτικότητα



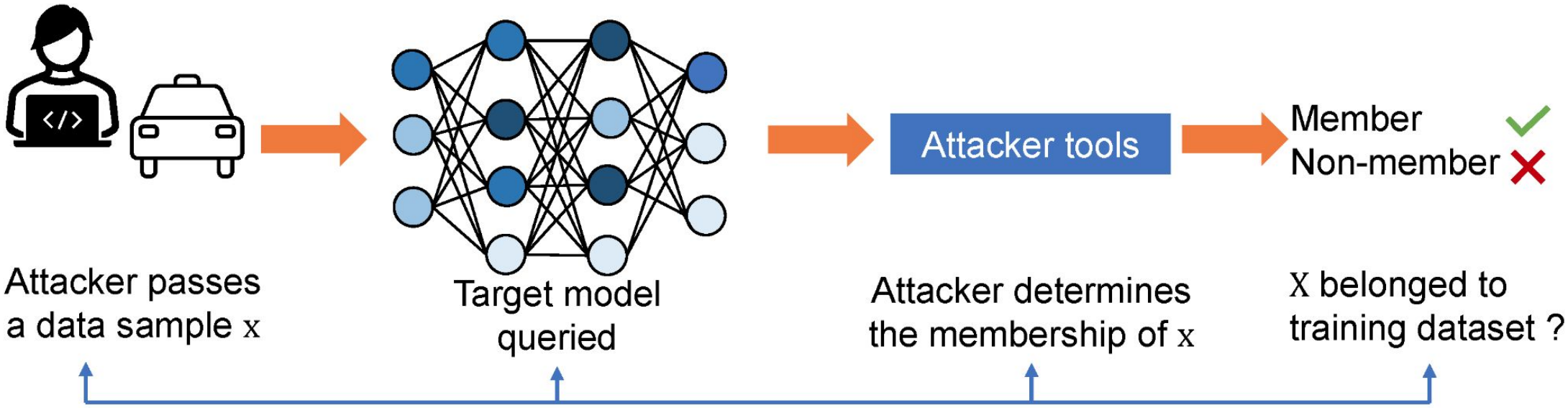
- Ωραία, με το Federated Learning κρατάμε τα δεδομένα τοπικά. Άρα είμαστε εντάξει σε σχέση με την αποστολή δεδομένων και την συμμόρφωση με κανονισμούς προστασίας;
  - Όχι
  - Ναι μεν δεδομένα δεν στέλνονται
  - Αλλά, στέλνονται gradients, weight updates

Προβλήματα στο FL: Ιδιωτικότητα

- Μα πώς γίνεται να διαρρεύσουν δεδομένα από τα μοντέλα;
  - Επιθέσεις τύπου Gradient Leakage

Προβλήματα στο FL: Ιδιωτικότητα

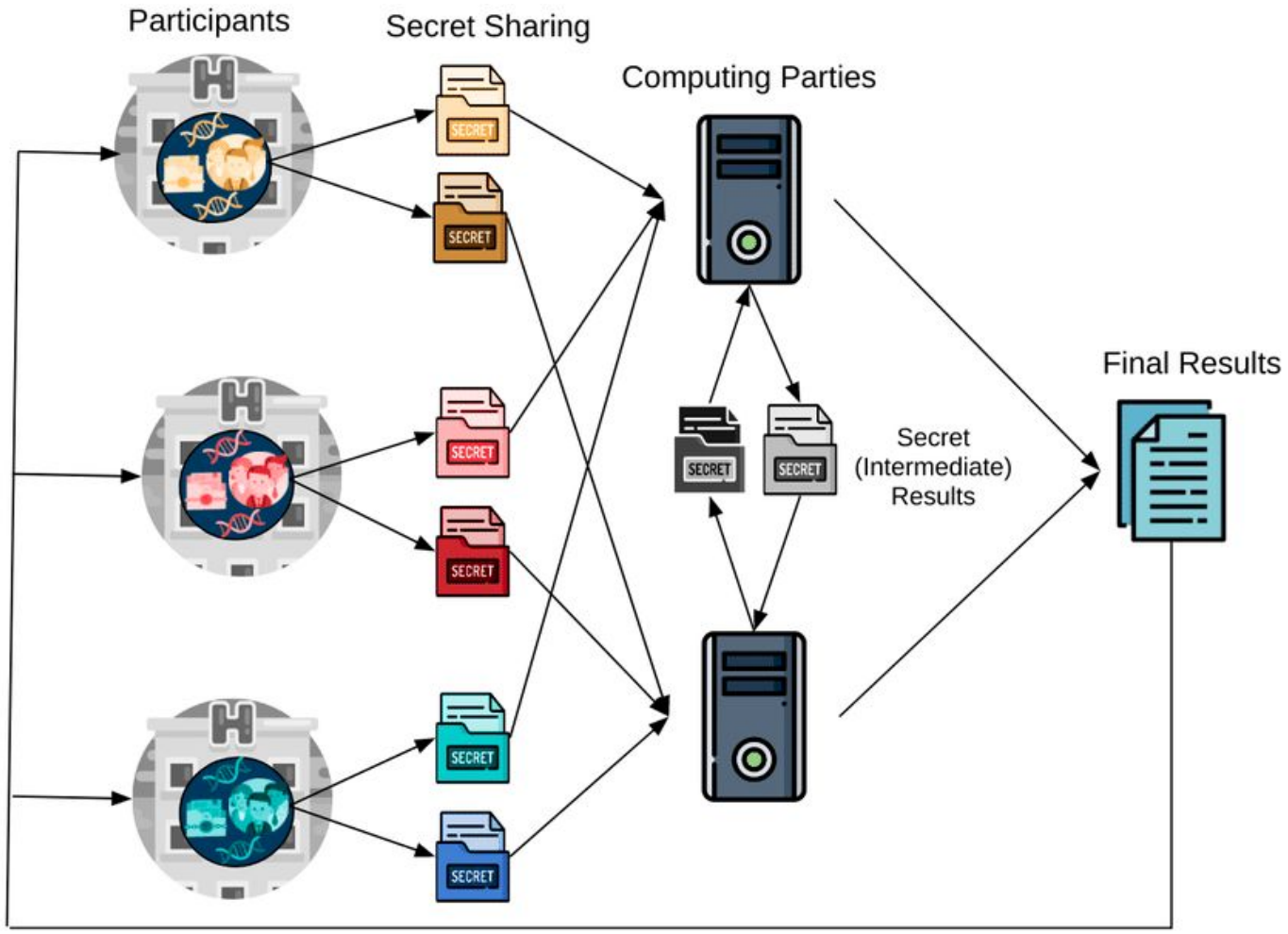
- Μα πώς γίνεται να διαρρεύσουν δεδομένα από τα μοντέλα;
  - Επιθέσεις τύπου Membership Inference



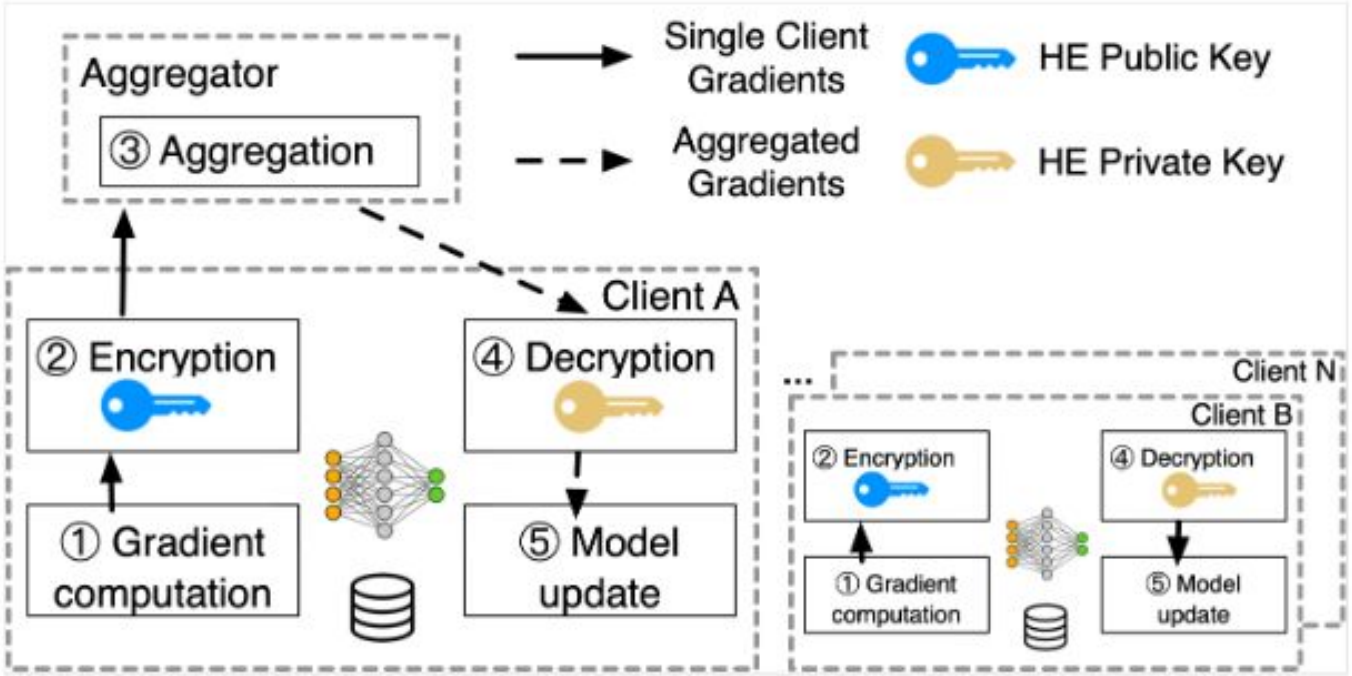
Προβλήματα στο FL: Ιδιωτικότητα - Λύσεις

- Κρυπτογραφία
  - Secure MultiParty Computation, (Fully) Homomorphic Encryption
- Differential Privacy

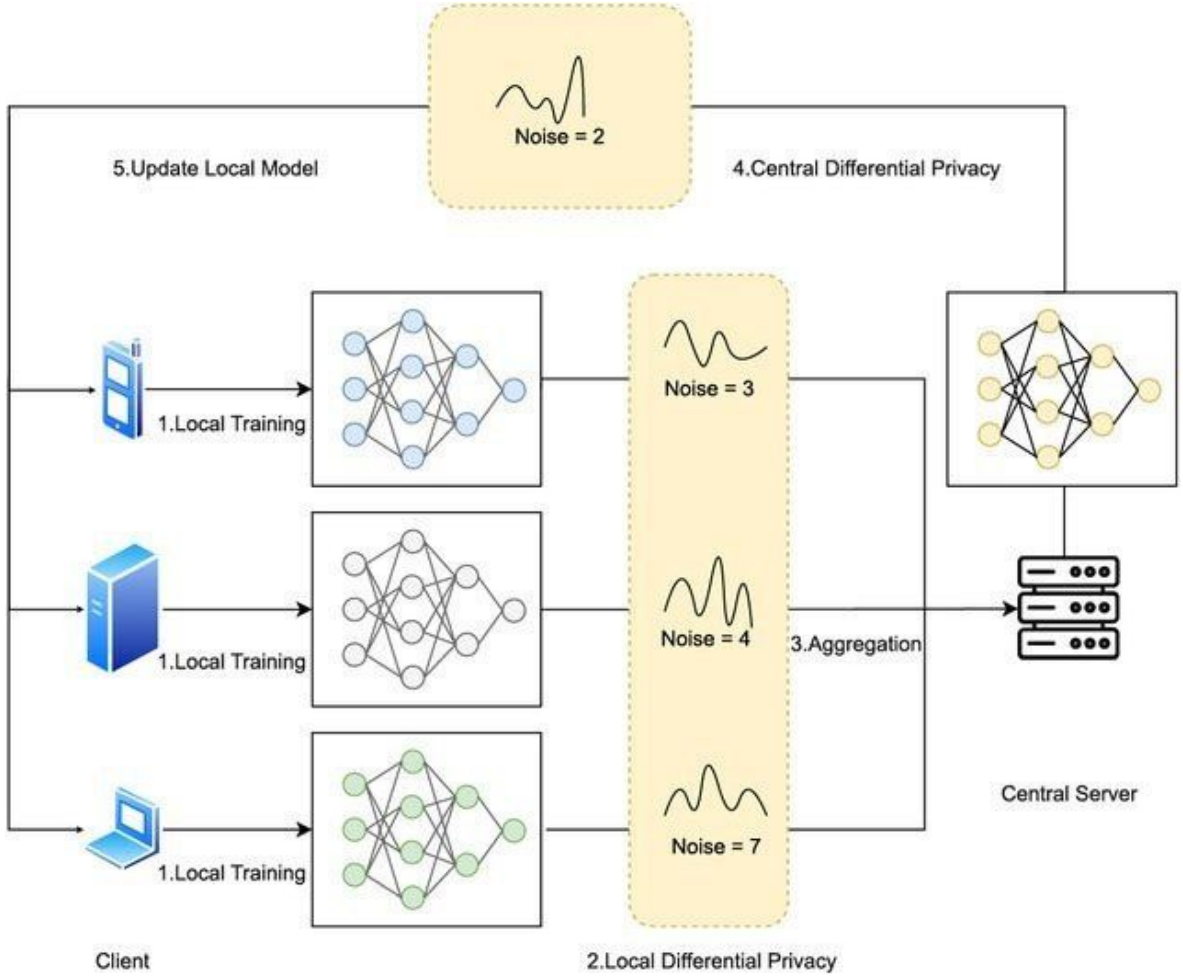
Προβλήματα στο FL: Ιδιωτικότητα - SMPC



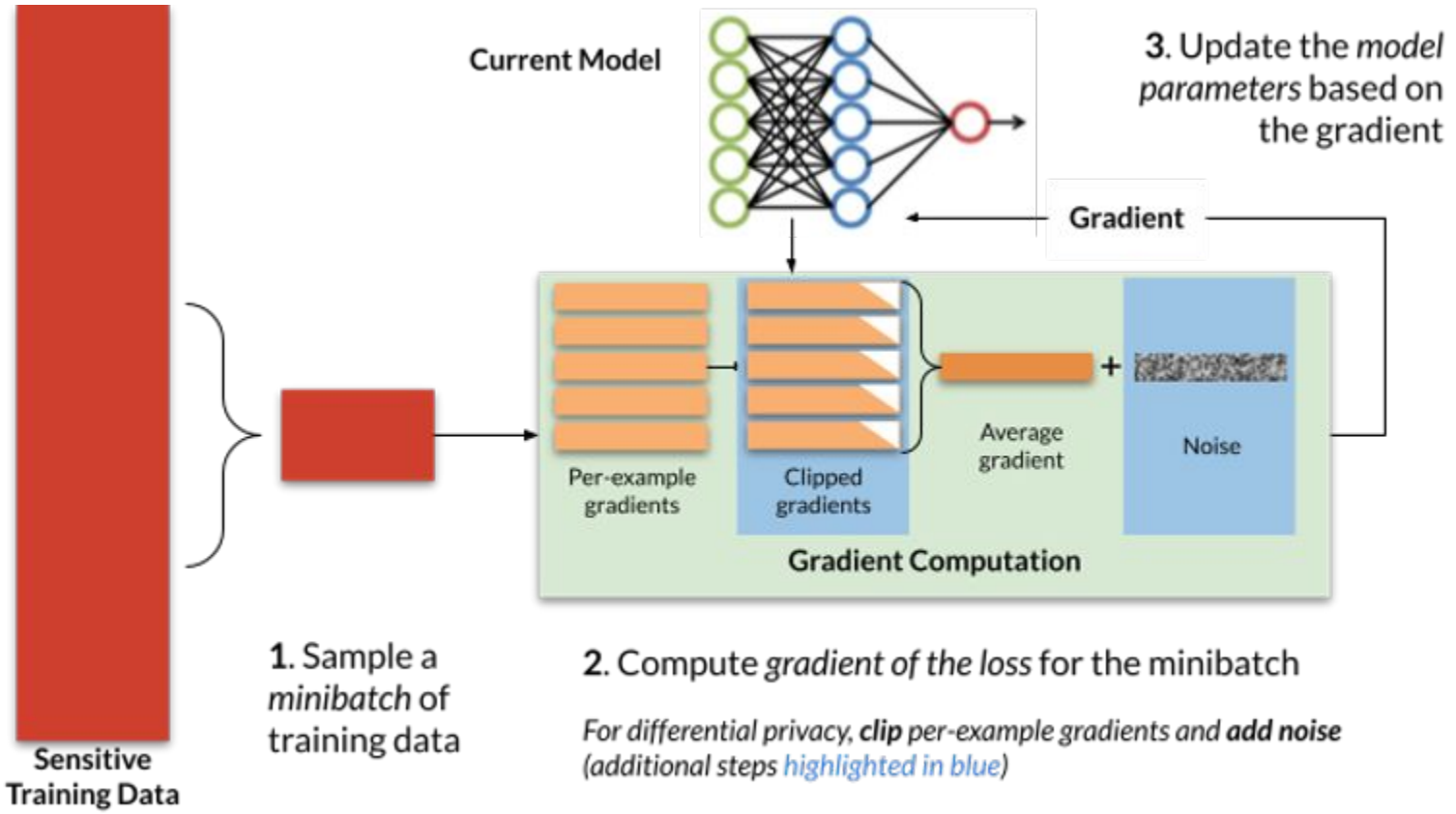
Προβλήματα στο FL: Ιδιωτικότητα - Homomorphic Encryption



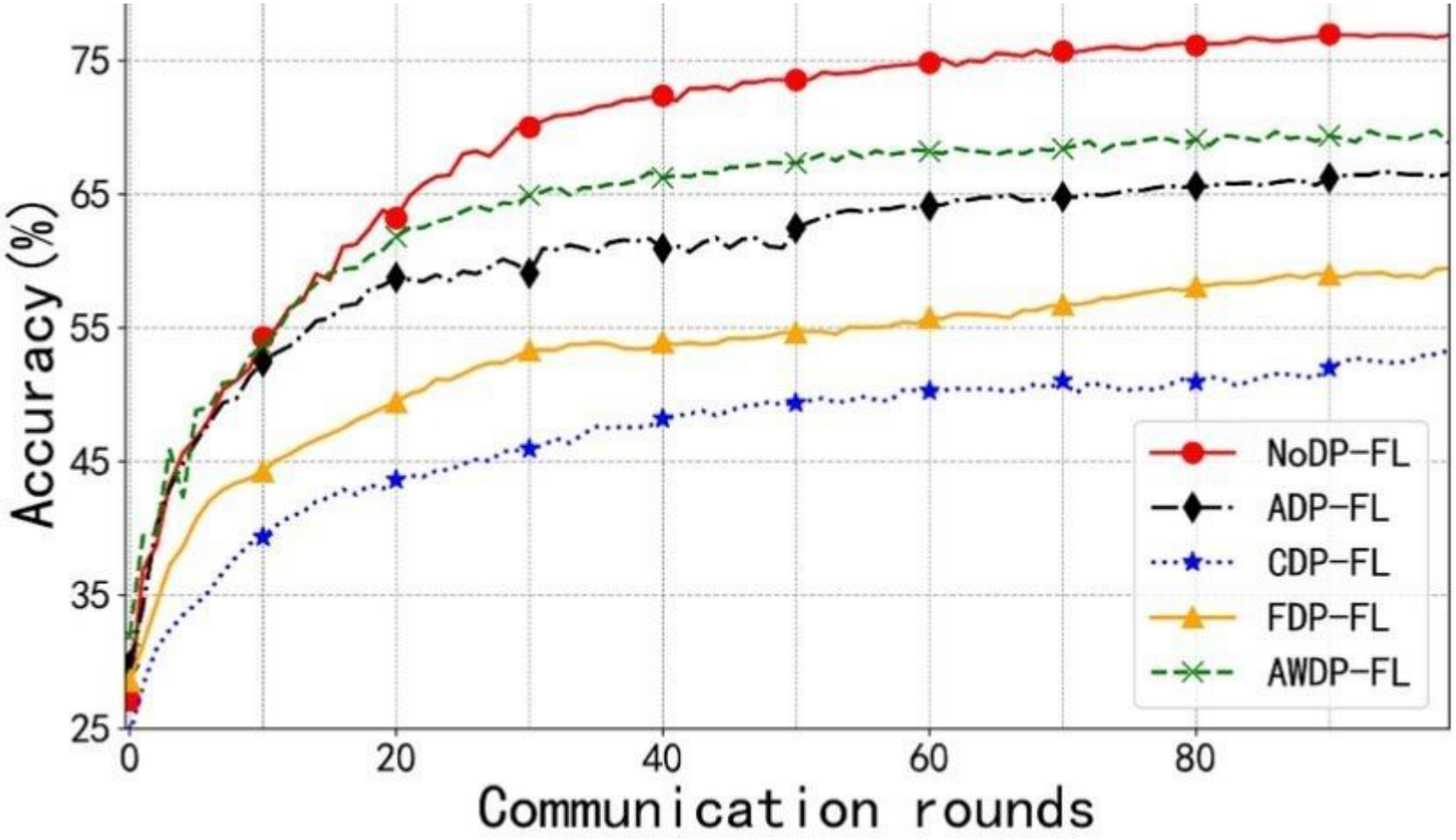
Προβλήματα στο FL: Ιδιωτικότητα - Differential Privacy



Προβλήματα στο FL: Ιδιωτικότητα - Differential Privacy

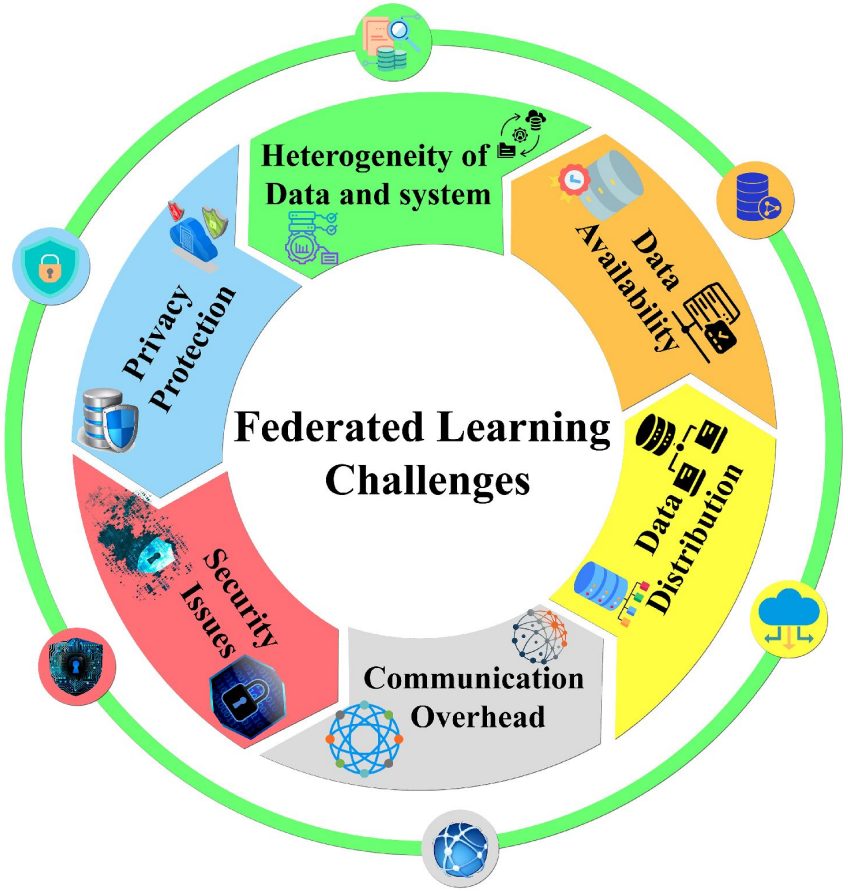


Προβλήματα στο FL: Ιδιωτικότητα - Differential Privacy



Federated Learning - Σύνοψη

- Τα δεδομένα μένουν τοπικά, ενώ ένα μοντέλο είναι δημόσιο
  - Η εκπαίδευση συμβαίνει στις συσκευές και όχι σε έναν server
- Η ιδιωτικότητα είναι επαυξημένη, αλλά όχι εγγυημένη
  - Απαιτούνται τεχνικές όπως secure aggregation και differential privacy
- Η επικοινωνία είναι ένα από τα σημαντικότερα κωλύματα
  - Απαιτούνται αποδοτικοί αλγόριθμοι μεταφοράς
- Τα δεδομένα είναι αποκεντρωμένα και ετερογενή
  - non-IID data
- Υπάρχουν trade-off παντού
  - Ιδιωτικότητα - Αποδοτικότητα - Ακρίβεια - Ενέργεια
- Το FL αλλάζει την μηχανική μάθηση
  - Από data-centric σε model-centric



Βιβλιογραφία

- McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017, April). Communication-efficient learning of deep networks from decentralized data. In Artificial intelligence and statistics (pp. 1273-1282). Pmlr.
- Lu, Z., Pan, H., Dai, Y., Si, X., & Zhang, Y. (2024). Federated learning with non-iid data: A survey. IEEE Internet of Things Journal, 11(11), 19188-19209.
- Kairouz, P., & McMahan, H. B. (2021). Advances and open problems in federated learning. Foundations and trends in machine learning, 14(1-2), 1-210.
- Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., ... & Seth, K. (2016). Practical secure aggregation for federated learning on user-held data. arXiv preprint arXiv:1611.04482.
- El Ouadrhiri, A., & Abdelhadi, A. (2022). Differential privacy for deep and federated learning: A survey. IEEE access, 10, 22359-22380.
- Zhang, C., Li, S., Xia, J., Wang, W., Yan, F., & Liu, Y. (2020). {BatchCrypt}: Efficient homomorphic encryption for {Cross-Silo} federated learning. In 2020 USENIX annual technical conference (USENIX ATC 20) (pp. 493-506).
- Khan, N., Nisar, S., Khan, M. A., Rehman, Y. A. U., Noor, F., & Barb, G. (2025). Optimizing federated learning with aggregation strategies: A comprehensive survey. IEEE Open Journal of the Computer Society.
- Perifanis, V., Pavlidis, N., Nikolaidou, F., Kampouri, D., & Efraimidis, P. S. (2025, June). Bridging the Gap: Challenges and Limitations of Federated Learning in Real-World Applications. In 2025 10th International Conference on Smart and Sustainable Technologies (SpliTech) (pp. 1-6). IEEE.
- Perifanis, V., Pavlidis, N., Koutsiamanis, R. A., & Efraimidis, P. S. (2023). Federated learning for 5G base station traffic forecasting. Computer Networks, 235, 109950.