Argonne NATIONAL LABORATORY

KEEPING DATA PRIVATE WHILE LEARNING TOGETHER: ADVANCES IN PRIVACY-PRESERVING FEDERATED LEARNING FOR THE GRID

KIBAEK KIM
Computational Mathematician
Mathematics and Computer Science
Argonne National Laboratory

February 11, 2025

MOTIVATION AND CONTEXT

Why Privacy-Preserving Federated Learning for the Grid?

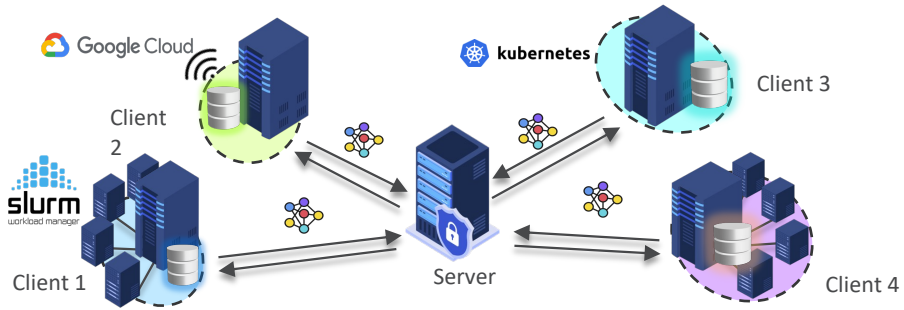
- **The Problem:** Grid operators, utilities, and researchers need to train models collaboratively while ensuring data privacy.
- **The Challenge:** Sharing raw data across multiple entities is not an option due to privacy concerns.
- **The Solution:** Federated Learning (FL) allows multiple stakeholders to train models without sharing their data.
- **Why Now?**
  - Increasing cybersecurity & privacy concerns
  - Advances in FL + AI enabling real-world deployment

❌ Traditional ML	✅ Federated Learning
Data sent to central server	Data stays local
One central dataset	Distributed datasets
Privacy risk: Raw data exposed	Privacy-preserving: No raw data transfer
High bandwidth usage	Low bandwidth usage

WHAT IS FEDERATED LEARNING?

Federated Learning: Enabling Collaborative AI Without Data Sharing

- **Federated Learning (FL) Basics:**
  - Data stays local (trained on edge devices or servers).
  - Only model updates are shared with a central aggregator.
- **Benefits:**
  - **Privacy:** Raw data never leaves local sites.
  - **Efficiency:** Reduces bandwidth and data movement.
  - **Scalability:** Works across multiple grid operators & industries.
- **Most widely used FL Algorithm:** FedAvg (introduced in 2016, still widely used today)
- **Our Contribution:** APPFL
  - Open-source FL framework for privacy-enhancing techniques.
  - Supports deployment across HPC, cloud, and edge devices.

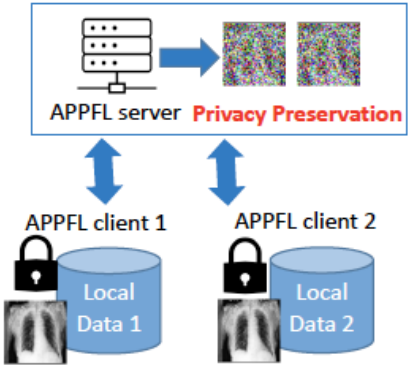


Argonne NATIONAL LABORATORY

PRIVACY RISKS & SOLUTIONS IN FL

Protecting Data Privacy in Federated Learning

- **The Risk:** Even without raw data, attackers can reconstruct data from gradients.
- **Key Privacy-Preserving FL (PPFL) Techniques:**
  - *Differential Privacy (DP):* Adds noise to model updates.
  - *Secure Multiparty Computation (SMPC):* Encrypts updates to prevent reconstruction.
  - *Homomorphic Encryption (HE):* Enables computations on encrypted data.



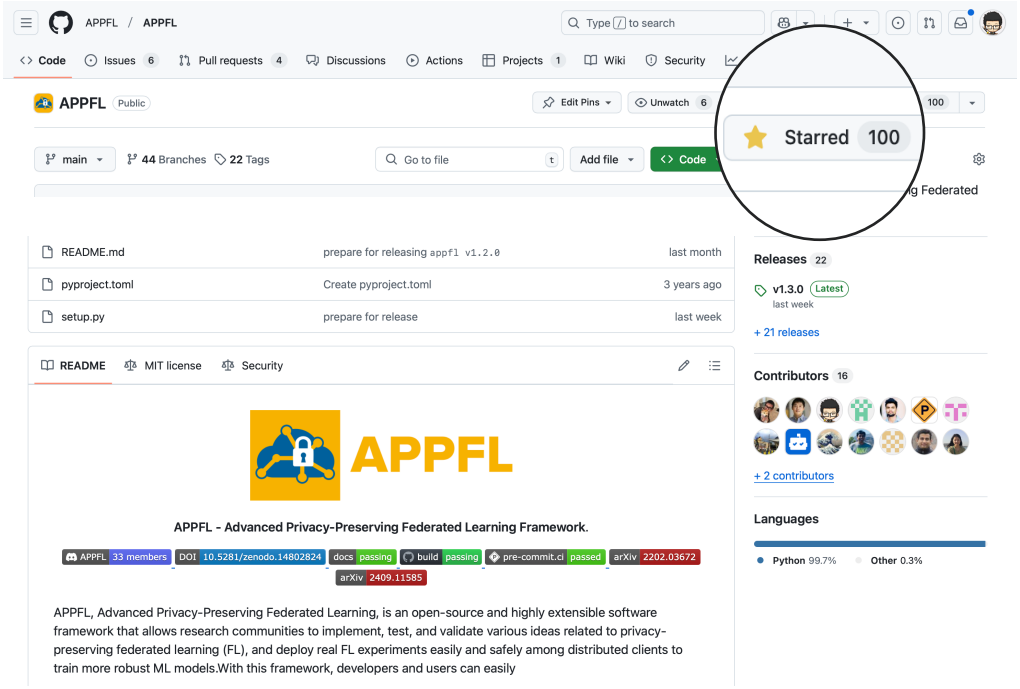
Argonne NATIONAL LABORATORY

https://appfl.ai

OPEN-SOURCE PPFL FRAMEWORK

Advanced Privacy-Preserving Federated Learning

- **APPFL v1.3.0:**
  - Open-source FL software for PPFL research & deployment
  - First release: Feb. 2022
  - Available in Github
- **Supports:**
  - Privacy (DP, HE, SMPC)
  - Heterogeneous computing (sync & async updates)
  - Scalable deployment (HPC, cloud, edge devices)
- **Extensively tested on:** DOE supercomputers (ALCF, OLCF, NERSC, ESnet FABRIC), Argonne's edge devices



Argonne NATIONAL LABORATORY

COMPARISON OF OPEN-SOURCE FL SOFTWARE

Key Capabilities Across FL Frameworks

TABLE I: Comparison of popular open-source federated learning frameworks. As of Aug, 2024

Framework	Data Hetero.	Sync. FL	Async. FL	Compression	Versatile Comm.	Privacy	Auth.	Real Deployment	FL Variants
LEAF [54]	✗	✓	✗	✗	✗	✗	✗	✗	✗
TFF [40]	✓	✓	✗	✗	✗	✓	✗	✗	✗
APPFL-V0 [22]	✓	✓	✗	✗	✗	✓	✗	✓	✗
FEDERATEDSCOPE [55]	✓	✓	✗	✗	✗	✓	✗	✓	*VFL*
FLARE [56]	✓	✓	✗	✗	✗	✓	✓	✓	*VFL*
OPENFL [57]	✓	✓	✗	✓	✗	✓	✓	✓	*VFL*
FEDSCALE [21]	✓	✓	✓	✓	✗	✓	✗	✓	✗
FLGO [58]	✓	✓	✓	✗	✗	✗	✗	✓	*VFL*
FEDLAB [59]	✓	✓	✓	✓	✗	✗	✗	✓	✗
FLOWER [19]	✓	✓	✗	✗	✓	✓	✓	✓	*VFL*
FEDML [20]	✓	✓	✗	✗	✓	✓	✓	✓	*VFL, HierFL, DFL*
APPFL (this work)	✓	✓	✓	✓	✓	✓	✓	✓	*VFL, HierFL, DFL*

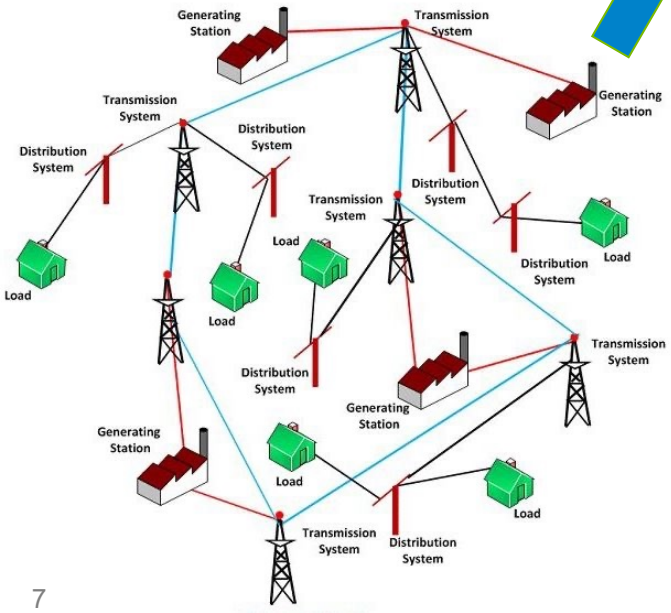
- *APPFL v1.x stands out with enhanced support for privacy, asynchronous algorithms, and versatile communication, advancing beyond APPFL v0 and other platforms.*

PROGRESS IN FOUNDATION MODELS FOR THE GRID

GNN Foundation Models for Electric Grid Operations

- **Why Graph Neural Networks (GNNs) for the grid?**
  - Grid operations depend on topological relationships
  - Traditional ML fails to generalize across grid configurations
- **Current Work:**
  - Training graph-based foundation models across different grid topologies
  - **Goal:** Integrate PPFL with these models

Training on graph relationship

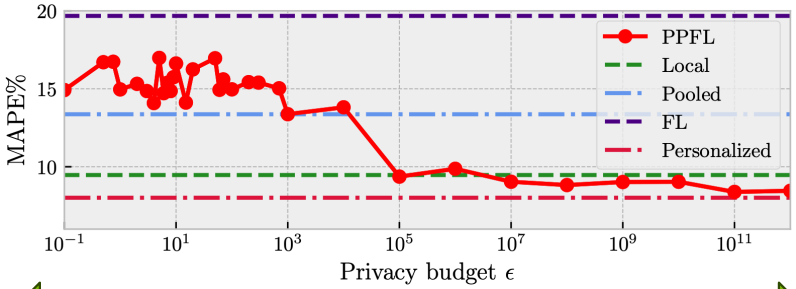


Argonne NATIONAL LABORATORY

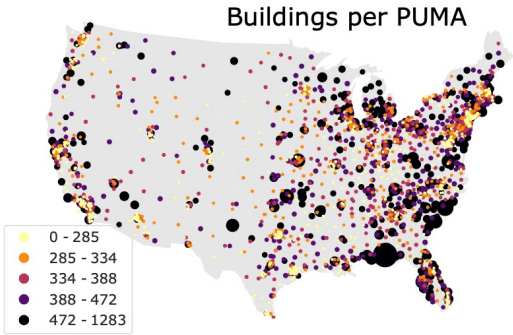
CASE STUDY:
TIME SERIES FL FOR BUILDING ENERGY

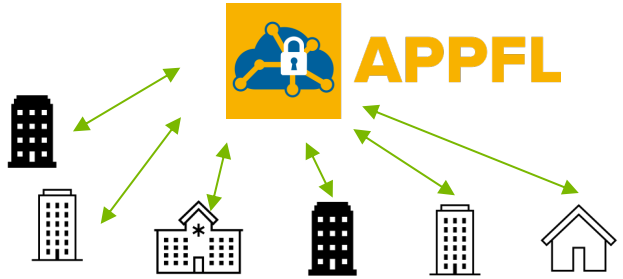
Federated Learning for Building Energy Forecasting

- **Data:** Electricity consumption from 42 buildings in CA, IL, NY.
- **Challenge:** Heterogeneous patterns across buildings.
- **Model:** Attention-based LSTM (long short-term memory) neural network architecture with personalized layers.
- **Results:**
  - **Personalized FL** achieves the lowest error.
  - **PPFL** successfully integrates to ensure data privacy.



Stronger privacy ⟷ *Weaker privacy*





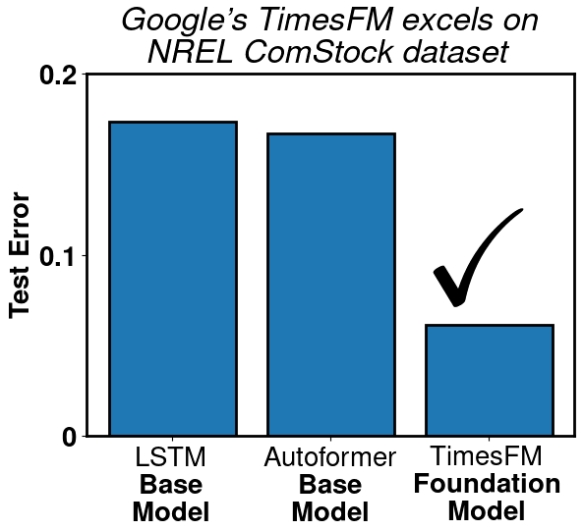
Training a forecast model without moving data + Privacy preserving technique

Argonne NATIONAL LABORATORY

FOUNDATION MODEL FOR BUILDING LOAD FORECASTING

TimesFM: A Foundation Model for Time Series Outperforms State-of-the-Art Methods

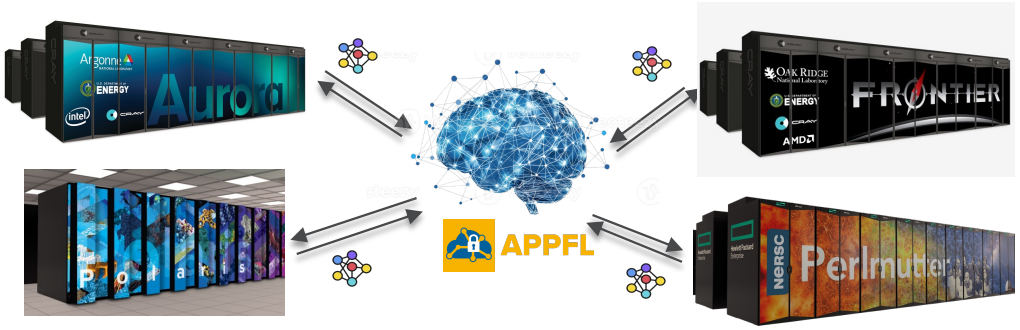
- **What are Foundation Models?**
  - Trained on vast, general-purpose data before being fine-tuned on task-specific datasets.
  - Pre-training on diverse data leads to significant performance gains in downstream tasks.
- **Applying TimesFM for Load Forecasting:**
  - Federated fine-tuning of foundation models (e.g., Google TimesFM) is a promising approach for building-level load forecasting.

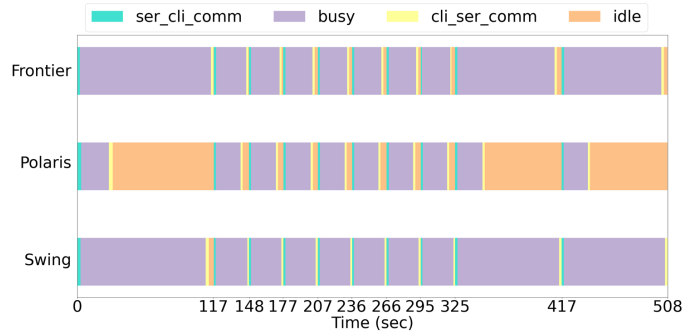


COMPUTING AT SCALE: RUNNING FL ON DOE SUPERCOMPUTERS

Scaling PPFL on High-Performance Computing (HPC) Infrastructure

- **Where We Run Our FL Models:**
  - ALCF Polaris
  - OLCF Frontier
  - NERSC Perlmutter
  - ESnet FABRIC Testbed
  - NCSA Delta
  - Many other clusters and clouds
- **Lessons Learned:**
  - FL scalability challenges in large systems
  - Need for adaptive scheduling & asynchronous updates





NEXT STEPS

Bringing PPFL to Grid Foundation Models

- **Immediate Goal:**
  - Run PPFL on GNN-based grid foundation models
- **Challenges:**
  - Adapting PPFL techniques to large-scale AI models
  - Interoperability between FL systems and grid operators
- **Call for Collaboration:**
  - Interested in testing PPFL in industry and national lab settings?

Near term	Mid-term	Long-term
APPFL on DOE HPC Baseline Privacy Techniques	Advanced Privacy Technique PPFL for Real-time Operations	Industry-wide Adoption Deployment-Ready PPFL Models
GNN-based FMs	Federated Fine-Tuning of FMs across Operators	Continual Adaptation of FMs
Scaling FL on HPC/Cloud	Cross-Institution PPFL	Self-Improving FL Ecosystem Policy/Regulatory Alignment

DISCUSSION / Q&A

Open Discussion & Collaboration Opportunities

- How can federated learning benefit your work?
- Are there specific technical, regulatory, or adoption challenges that need to be addressed?
- What would make utilities or grid operators more willing to adopt PPFL?
- What privacy concerns do you see in grid applications?
- Are there additional industry/national lab partners interested in PPFL testing?

ACKNOWLEDGEMENTS

- DOE ASCR Early Career Research Program (2019 - 2024)
- DOE ASCR PALISADE-X Project (2022 – 2024)
- DOE ASCR EXPRESS (2023 – 2024)
- DOE ASCR Resilient Distributed Systems (2024 – 2028)
- DOE ASCR AI4S (2025 – 2027)

U.S. DEPARTMENT OF ENERGY

- **Collaborators**:

Argonne NATIONAL LABORATORY
Brookhaven National Laboratory
OAK RIDGE National Laboratory
ASU
RUTGERS UNIVERSITY
THE UNIVERSITY OF CHICAGO
UNIVERSITY OF CALIFORNIA SANTA CRUZ
CLEMSON
ILLINOIS
BROAD INSTITUTE

THANK YOU