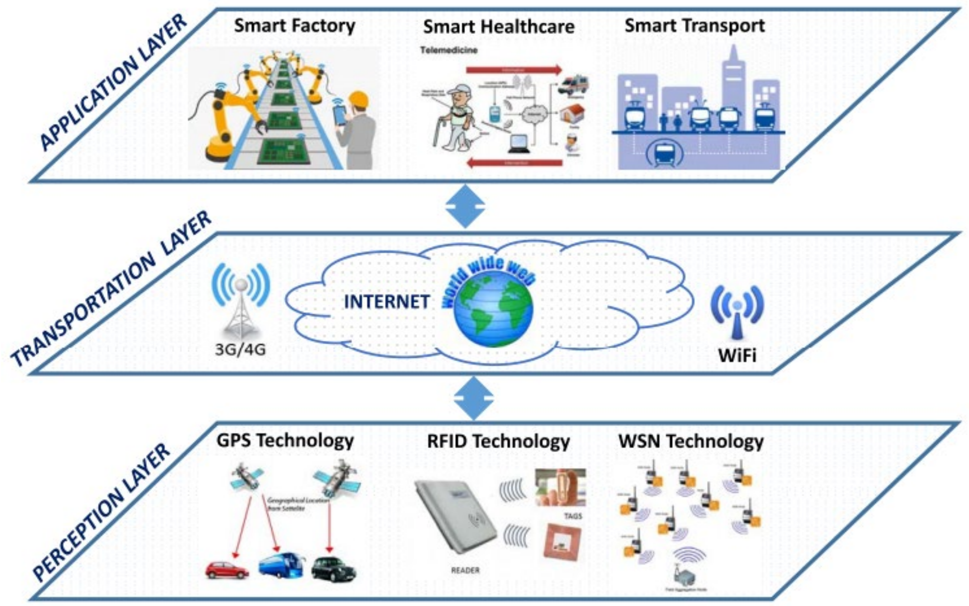
# Model Poison Attack in Federated Learning for IoT

# IoT Systems

**The Internet of Things (IoT) refers to the network of physical devices, vehicles, home appliances, and other items embedded with electronics, software, sensors, actuators, and connectivity.**

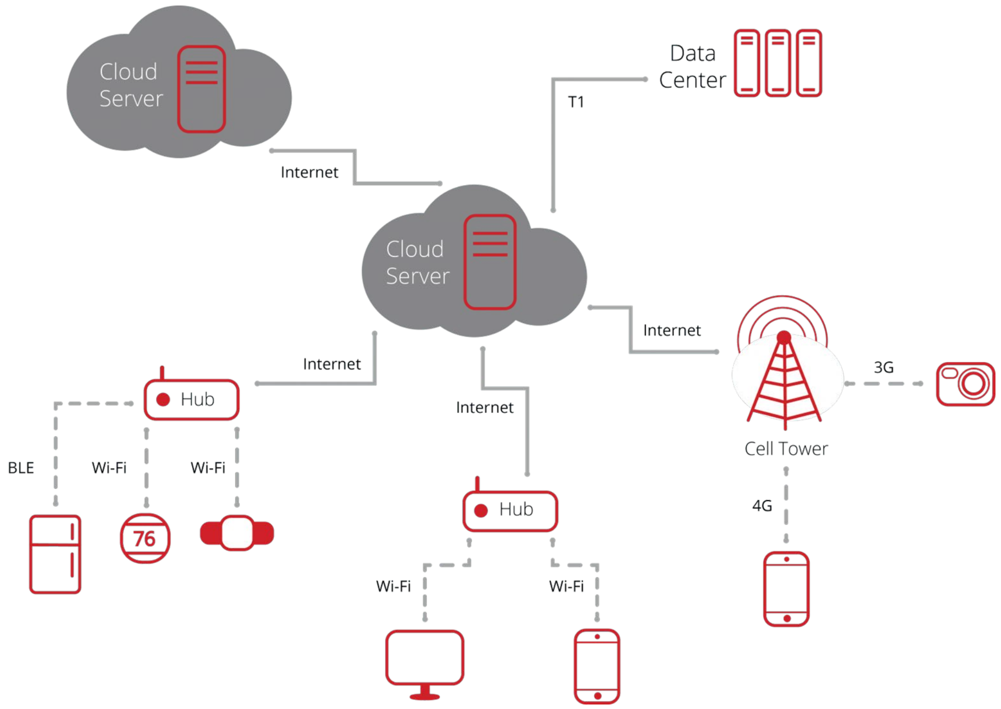



## Application Scenarios

- **Smart Cities:** Federated learning can be used in traffic management systems, optimizing traffic flow based on data from various sensors and cameras while keeping this data localized.
- **Healthcare Monitoring:** Wearable devices can collect health data and contribute to predictive health models without sharing sensitive personal health information.
- **Industrial IoT:** In factories, machines equipped with sensors can use federated learning to predict maintenance needs or optimize production processes, enhancing efficiency and safety.
- **Home Automation:** Smart home devices like thermostats and security cameras can learn user preferences and detect anomalies without sending sensitive data to the cloud.

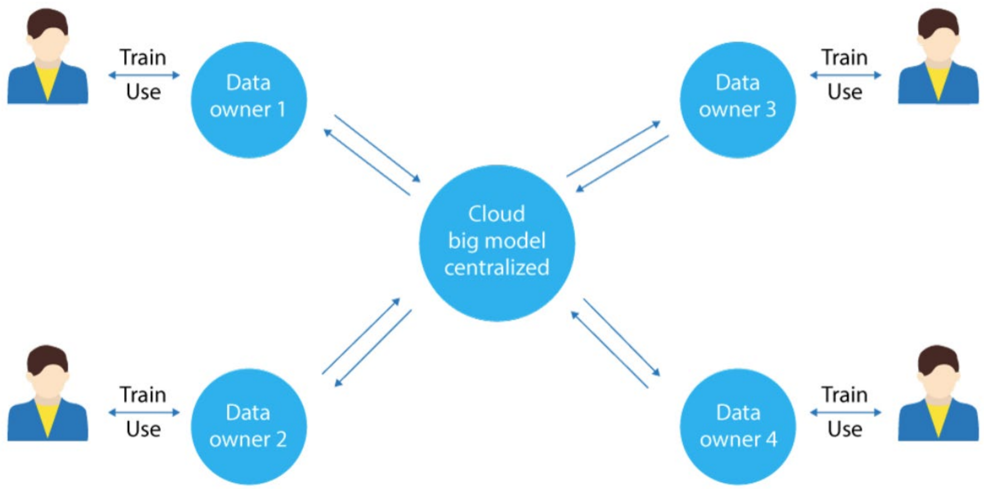
# Federated Learning in IoT Systems

- Federated learning addresses the critical challenge of data privacy in IoT.
- It allows for real-time analytics and decision-making at the edge of the network.
- Facilitates scalable machine learning models without the need for massive data centralization.



# Federated Learning

Federated Learning is the problem of training a shared global model under the coordination of a central server, from a federation of participation devices which maintain control of their own data.



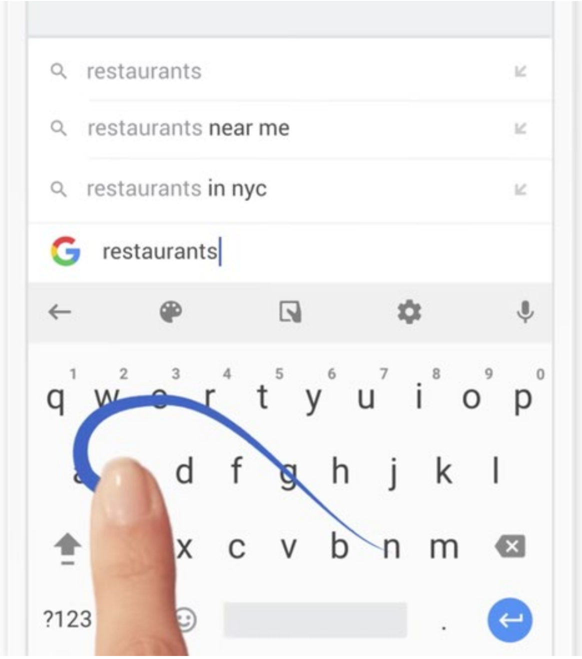
# Federated Learning

- Federated learning is a decentralized approach to training machine learning models.
- It doesn’t require an exchange of data from client devices to global servers.
- The raw data on edge devices is used to train the model locally, increasing data privacy.
- The final model is formed in a shared manner by aggregating the local updates.

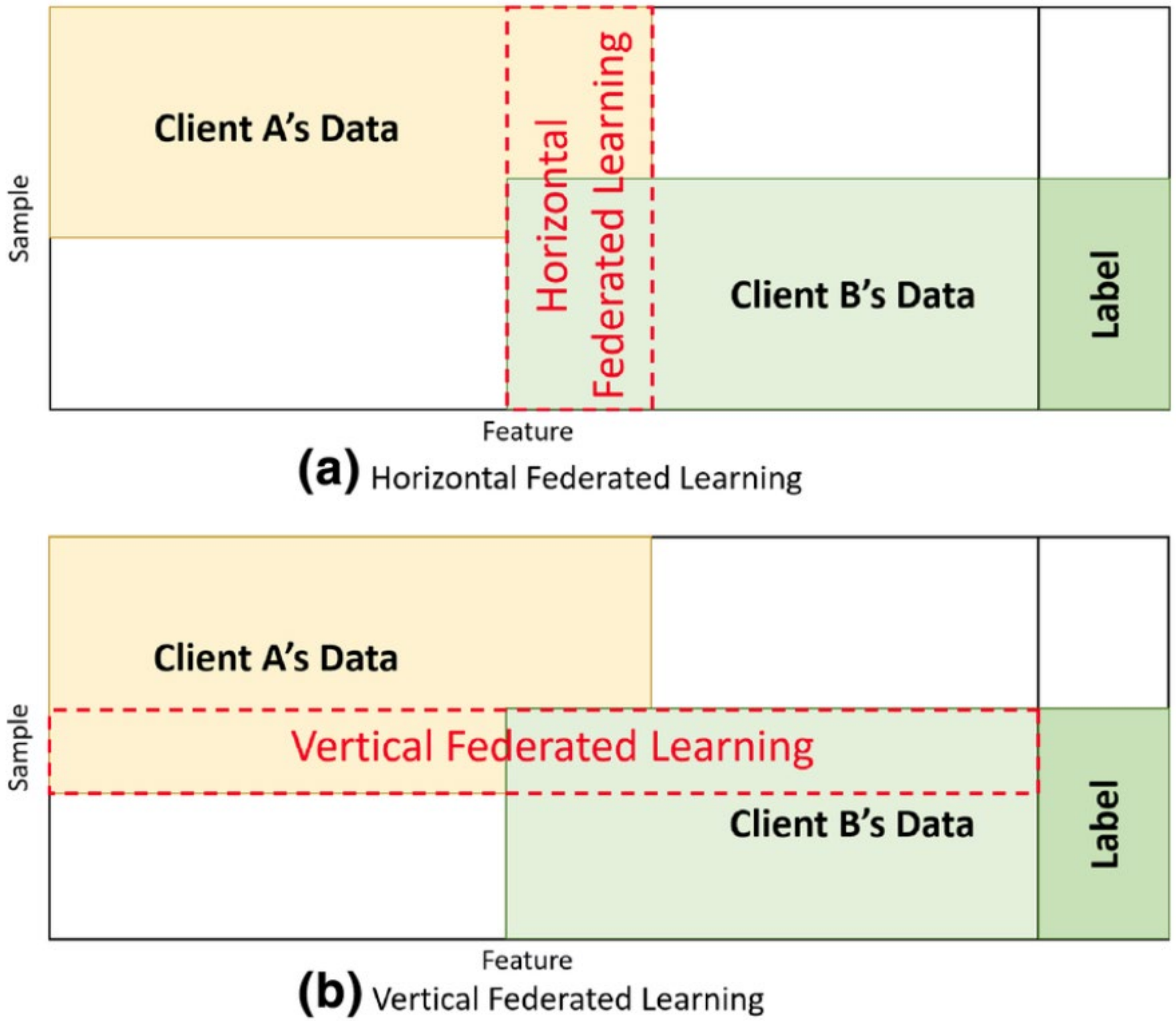
# Federated Learning Applications

The Federated learning is first put forward by Google to predict user's next-word prediction from Gboard on Android devices.

- Smart Phone
- Organization
- IoT
- Healthcare
- Advertising
- Autonomous vehicles
- Federated learning in the field of financial fraud
- Federated learning in the field of insurance



# Horizontal FL & Vertical FL




- **Horizontal FL**

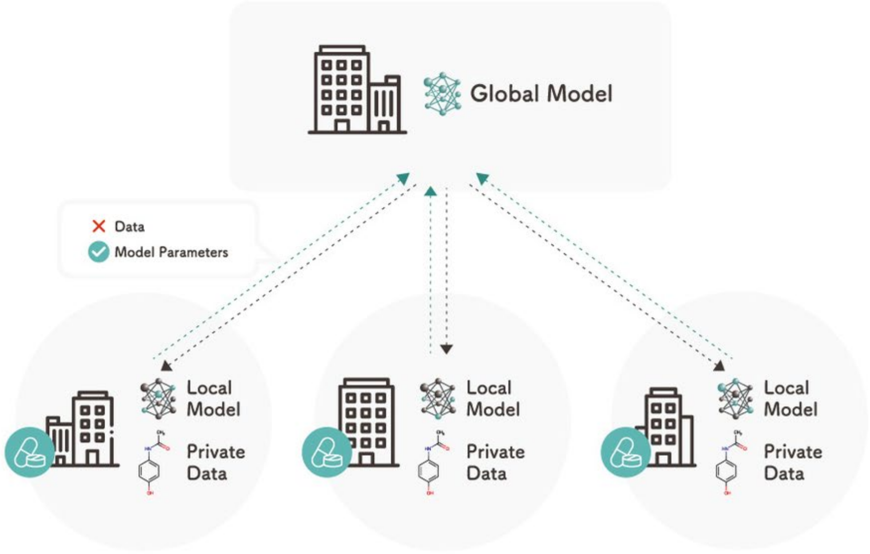
  each participant share similar features but concern different samples

- **Vertical FL**

  participants have overlaps in the sample space but differ in the feature space

# Federated Learning

- The server chooses an initialized model and broadcasts the current global model to (a subset of) the clients
- Each client fine-tunes the global model using its local training data and reports its model update to the server
- The server aggregates the clients' model updates following some aggregation rule and uses the aggregated model update to update the global model.



# Federated Learning

local model updates:

$$\boldsymbol{g}_t^i = \frac{\partial \mathcal{L}_i(\boldsymbol{w}_t)}{\partial \boldsymbol{w}_t}$$

global model updates:

$$\boldsymbol{g}^t = \mathcal{A}(\boldsymbol{g}_1^t, \boldsymbol{g}_2^t, \cdots, \boldsymbol{g}_n^t)$$

global model:

$$\boldsymbol{w}^{t+1} \leftarrow \boldsymbol{w}^t + \eta \boldsymbol{g}^t$$

| | |
|---|---|
| $\boldsymbol{w}^t$ | global model in the *t*-th round |
| $\mathcal{L}$ | loss function |
| $\boldsymbol{g}_t^i$ | model update of *i*-th client |
| $\mathcal{A}$ | aggregation rule |
| $\boldsymbol{g}^t$ | global model updates |
| $\eta$ | learning rate |

# Aggregation Rules

- **FedAvg**
  Average value of the local model updates as the parameter of global model update.
- **Median**
  Median value of the local model updates as the parameter of global model update.
- **Trimmed-mean**
  Removes the largest and smallest k values from its sorted values, and then computes the average of the remaining values as the parameter of global model update.

# Model Poisoning Attack to FL

- **Targeted Model Poisoning Attacks:**
  Aim to force the global model to output attacker chosen target labels for attacker-chosen target input.
- **Untargeted Model Poisoning Attacks:**
  Aim to decrease the test accuracy of the global model.

- Compromised genuine clients compute the genuine local model updates based on their genuine local training data.
- They perturb their genuine local model updates such that the poisoned global model updates will substantially deviate from the genuine ones.

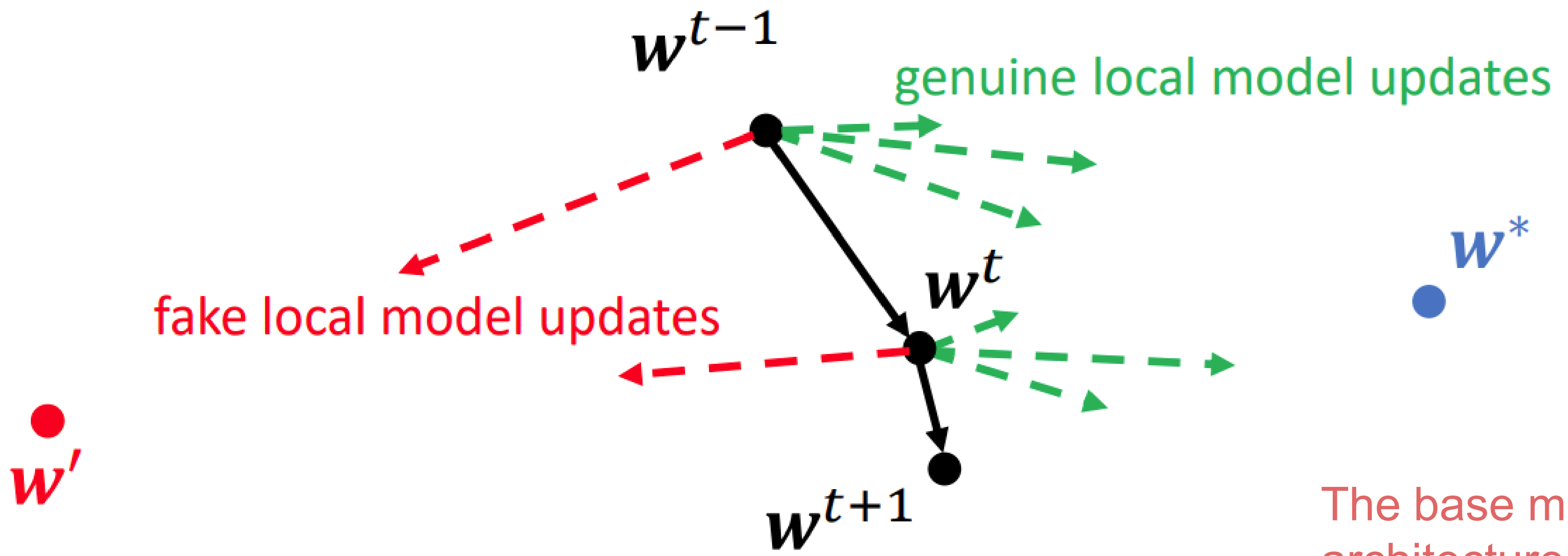
# Threat Model

- **Attacker's goal**
  Decrease the test accuracy of global model.

- **Attacker's capability**
  Inject fake clients and control fake clients to send fake local model updates.

- **Attacker's knowledge**
  Received global model during training.

# Model Poisoning Attack

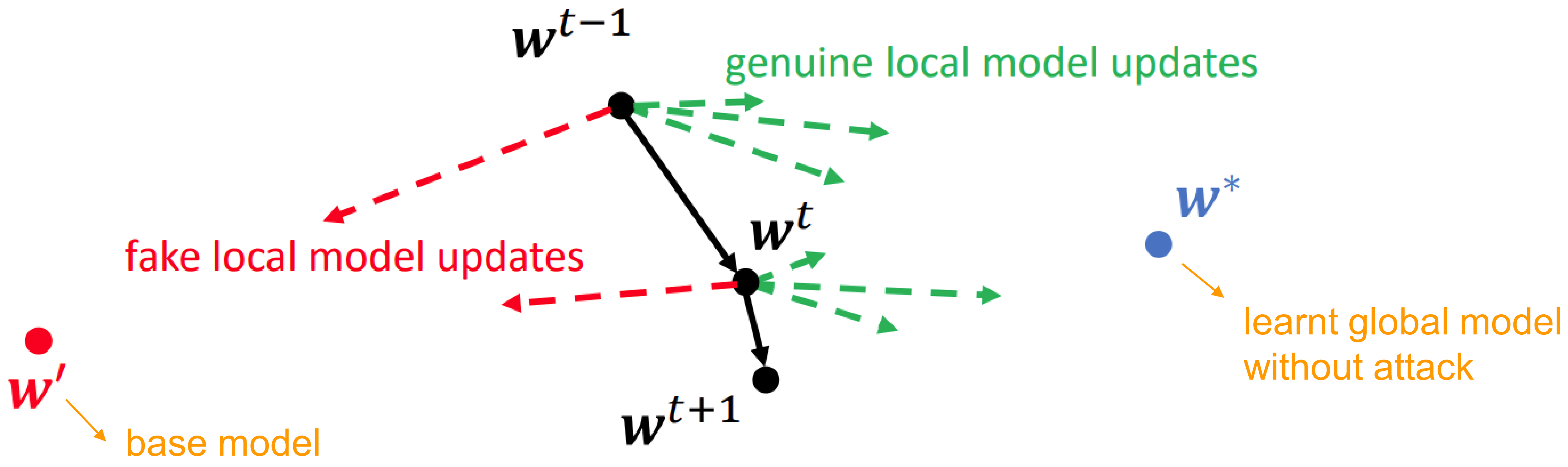



The base model has the same architecture as the global model and has low test accuracy.

$w'$ attacker-chosen base model.

$w^*$ learnt global model without attack.

# Model Poisoning Attack

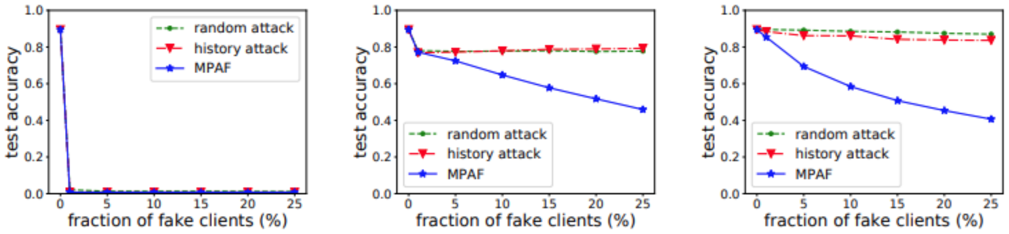



$$\min_{g_i^t, i \in [n+1, n+m], t \in [0, T-1]} \|w^T - w'\|$$

$$g_i^t = \lambda(w' - w^t)$$

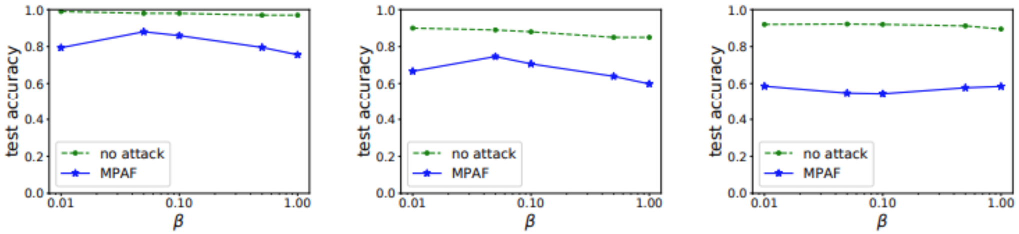
# Evaluation

Different fraction of fake clients:



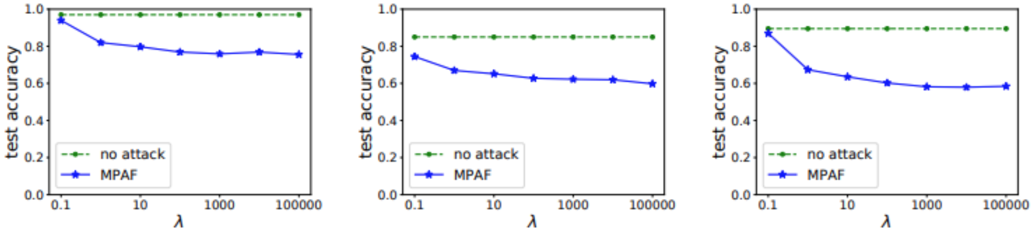

(a) FedAvg (b) Median (c) Trimmed-mean

Different sample rate β:




(a) MNIST (b) Fashion-MNIST (c) Purchase

Different scaling factor λ:

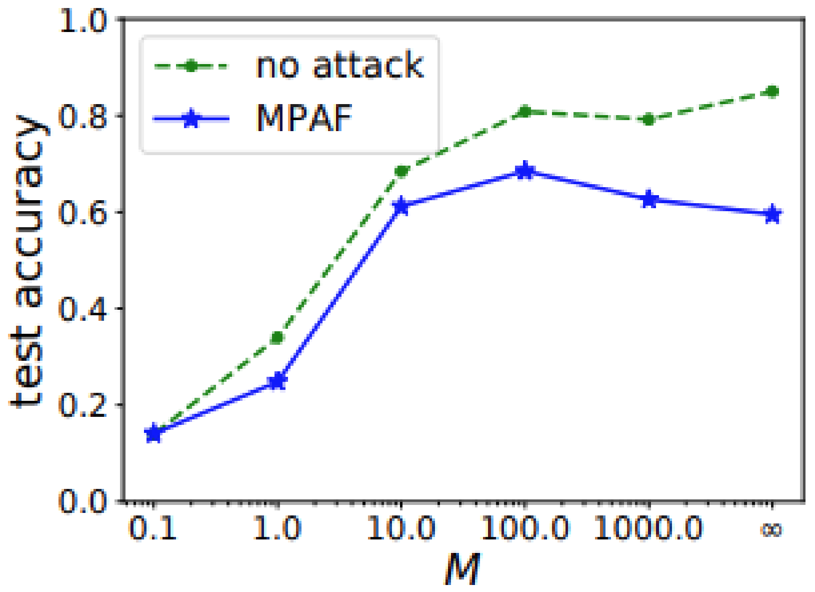



(a) MNIST (b) Fashion-MNIST (c) Purchase

A lower test accuracy indicates a stronger attack

# Evaluation

## Performance under Norm Clipping




$$\boldsymbol{g} \longrightarrow \frac{\boldsymbol{g}}{\max(1, \|\boldsymbol{g}\|_2 / M)}$$

$M \longrightarrow 0$ Test accuracy under attack/no attack decrease

$M \longrightarrow \infty$ No norm clipping

# Conclusion

- **Addressing Privacy and Security Concerns:** Federated Learning stands at the forefront of addressing the critical privacy and security concerns in IoT by keeping data localized and minimizing the risk of data breaches.
- **Enabling Real-Time, Efficient IoT Applications:** By facilitating on-device data processing, Federated Learning enables IoT systems to make real-time decisions, greatly reducing latency and enhancing the efficiency of IoT applications.
- **Meeting the Challenges of Scalability and Diversity:** Given the diverse and expansive nature of IoT devices, Federated Learning's scalable and adaptable framework is essential for managing the complexity and variety inherent in IoT ecosystems.