Open Challenged in Decentralized (edge) AI

Sonia BEN MOKHTAR

The 88th Meeting of the IFIP WG 10.4 on Dependable Computing and Fault-Tolerance
Ischia, Italy

29/06/2025

Who am I?

- Head of the DRIM team @LIRIS lab Lyon
  - Distrubuted systems
    - Dependability
    - Privacy (e.g., location privacy, private web search, private recommender systems)
    - Performance
  - Information Retrieval

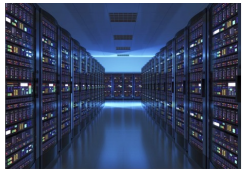
  - Increasing interest for Distributed Learning
    - Numerous challenges in terms of dependability, privacy & performance

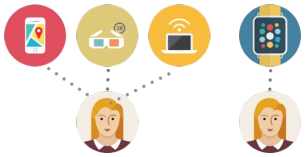
Ongoing projects

- Post-covid investments (PEPR national projects)
  - Co-Leading the Cybersecurity PEPR (65M€)
  - Carrying out research in
    - AI PEPR (resilient decentralized learning)
    - Cloud PEPR (confidential storage)
- Joint lab with iExec Blockchain-tech
  - Web 3.0 decentralized systems
  - TEEs

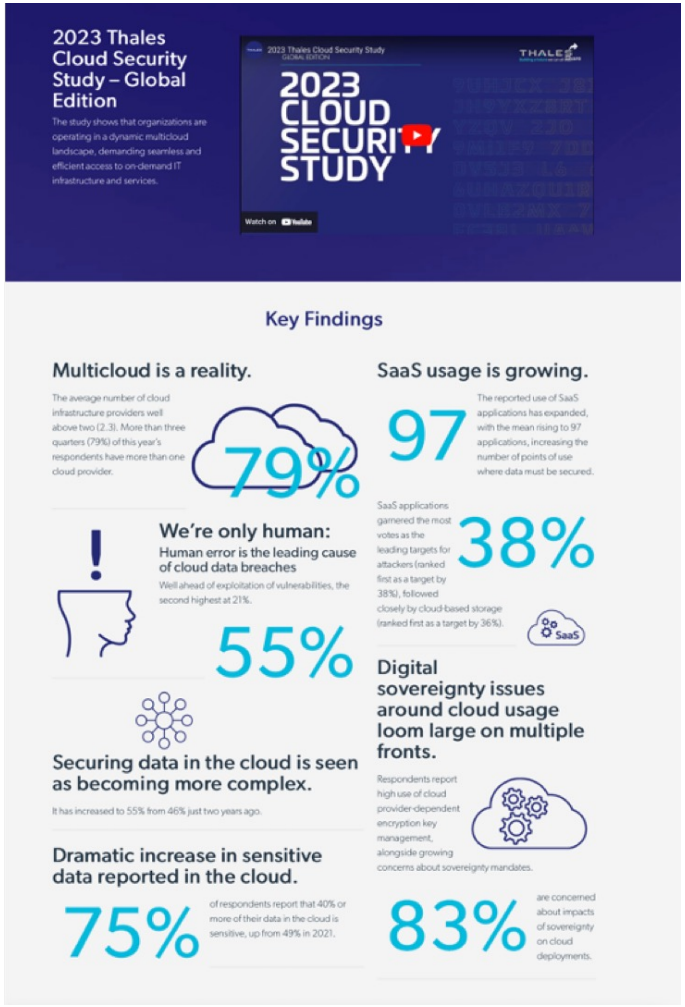
Today’s Online Services

- Heavily centralized (governance)
- Data-centric (data is the new oil)
- Open numerous threats



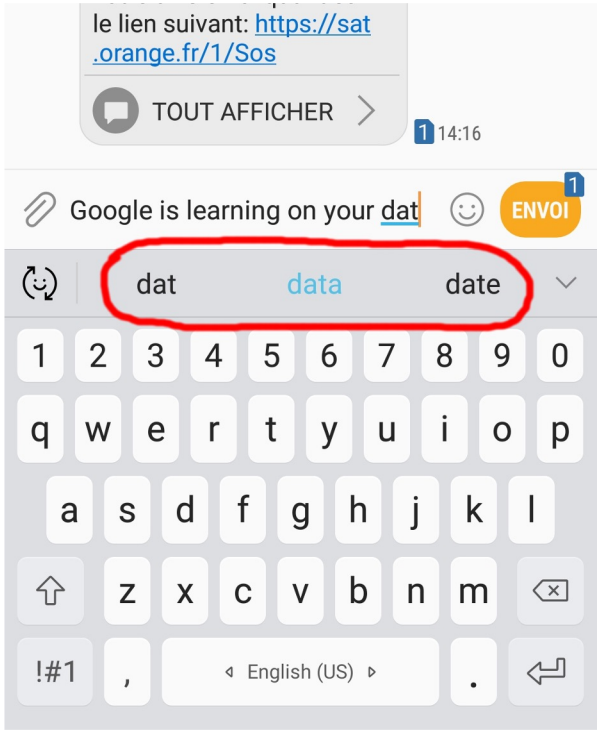


- Increased user awareness on privacy
- Legislator
  - GDPR, AI Act, …



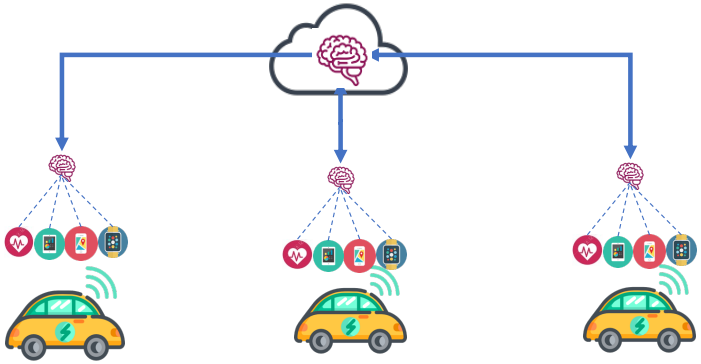
Federated Learning : a Natural Candidate for Preserving Data Confidentiality

- Federated learning (FL) aims at collaboratively train ML models while keeping the data decentralized
- 2016: Used by Google Research for training the Gboard (Google Android Keyboard)
- 2025: thousands of research papers published every year
- Interest coming from varius communities
  - AI/ML, optimization, distributed systems, networks, security, privacy, dependability, …
- Some real world deployments (e.g., hospitals)
- Libraries: PySyft, TensorFlow Federated, FATE, Flower, Substra...

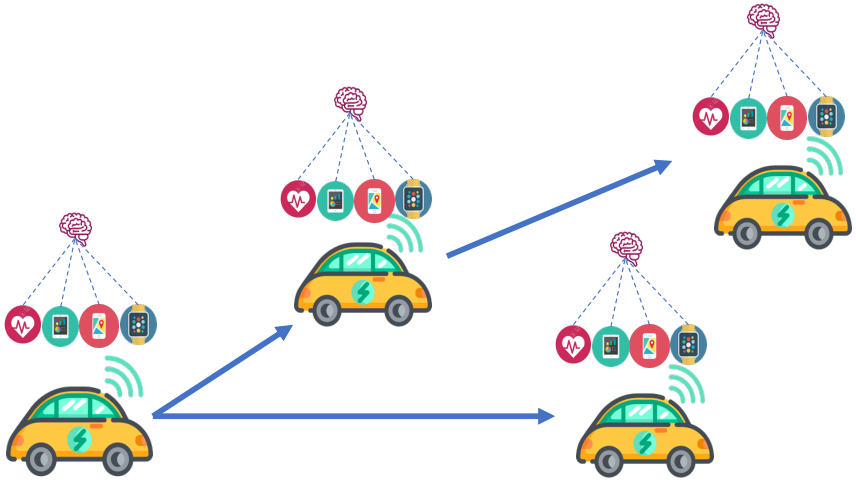


Server Orchestrated vs. Fully Decentralized

- Orchestrated
  - Server-client communication
  - Global coordination, global aggregation
  - Server is a single point of failure and may become a bottleneck



- Decentralized
  - Device to device communication
  - No global coordination, local aggregation
  - Naturally scales to a large number of devices

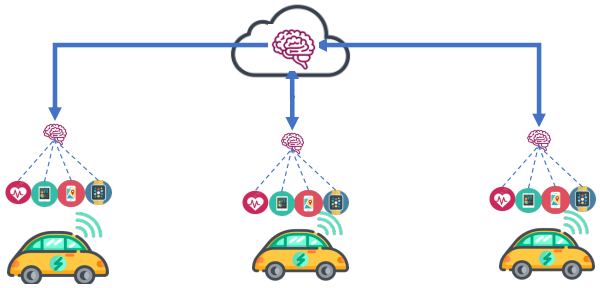


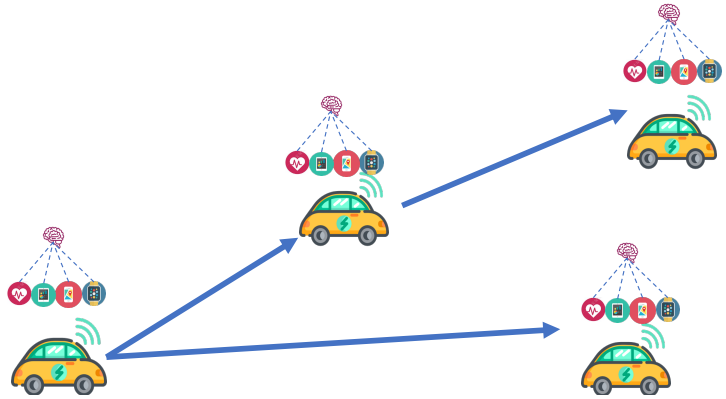
Orchestrated & Decentralized: threats

Adversary can:

- Run on the client side or on the server side vs be placed randomly in the communication graph
- Observe multiple snapshots of the model
- Reconstruct sensitive data (**Inversion attacks**)
- Infer sensitive properties about the participants (**Data property attacks**)
- Infer whether data samples have been used in training (**membership inference attacks**)
- Perform **data/model poisoning attacks**
- Inject **backdoors** into the model

Handling these threats all together is very challenging





Distributed/Decentralized Learning in Lyon

- Addressed challenges
  - Personalization
  - Privacy
  - Robustness (Byzantine Resilience)
- Ongoing work
  - [Personalization]
    - Decentralizing Recommender Systems with Gossip Learning [Ubicomp'22]
    - Personalized arrythmia detection in ECG signals
    - FL-based Location Privacy [Ubicomp'21][Middleware'20]
  - [Privacy]
    - Resilient FL with Trusted Execution Environments [Middleware'22]
    - Community detection attack in decentralized FL [ICDCS'25]
    - Differentially-private, decentralized mean estimation [arxiv]
    - Understanding the vulnerability of decentralized learning to membership attacks [arxiv]
  - [Robustness]
    - Private & Byz resilient decentralized ML
    - Byzantine resilient decentralized ML [arxiv]

Conclusion

- Today's online services are too centralized
- A new wave of decentralization is undergoing (Web 3.0)
- Revisiting decentralized/dependability/security algorithms (for decentralized ML) is needed
- Numerous challenges (ML, optimization, distributed systems/algorithms, security, privacy, networking…)
  - Understand the benefits/limits of decentralization
    - Does decentralization effectively improve personalization?
    - Does decentralization increase or reduce the attack surface?
    - Enforcing privacy & resilience to Byzantine nodes: compatible?