Open Challenged in Decentralized (edge) AI

Sonia BEN MOKHTAR

The 87th Meeting of the IFIP WG 10.4 on Dependable Computing and Fault-Tolerance

Praia do forte
09/02/2025

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

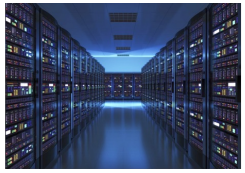
Ongoing projects

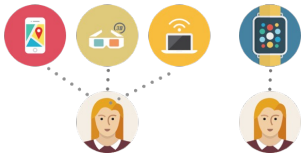
- Post-covid investments (PEPR national projects)
  - Co-Leading the Cybersecurity PEPR (65M€)
  - Carrying out research in
    - AI PEPR (resilient decentralized learning)
    - Cloud PEPR (confidential storage)
- Joint lab with iExec Blockchain-tech
  - Web 3.0 decentralized systems
  - TEEs

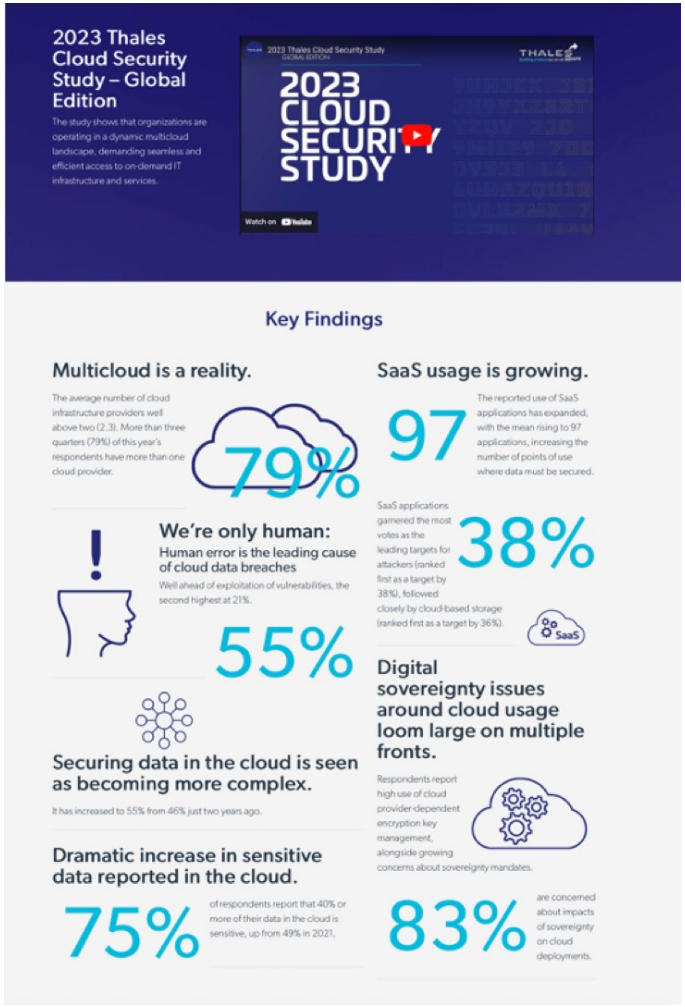
Today’s Online Services

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

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







Threats Illustrated

Angry Birds and 'leaky' phone apps targeted by NSA and GCHQ for user data

- US and UK spy agencies piggyback on commercial data
- Details can include age, location and sexual orientation
- Documents also reveal targeted tools against individual phones

Cybercriminals raid BBC pension database, steal records of over 25,000 people

This just in: We lost your personal info, but here's 2 years' worth of Experian

Connor Jones — Thu 30 May 2024 // 14:02 UTC

QUARTZ

EXCLUSIVE

Google collects Android users' locations even when location services are disabled

By Keith Collins · November 21, 2017

QUARTZ

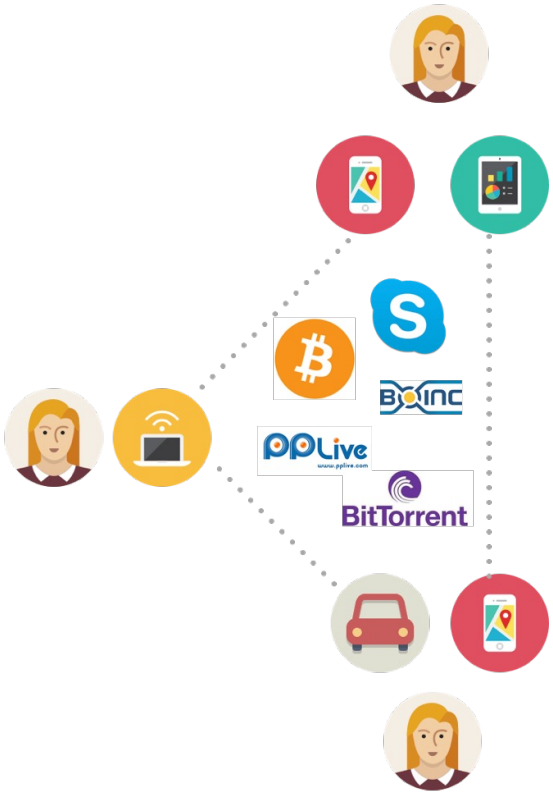
SWIPED

Dating app Tinder briefly exposed the physical location of its users

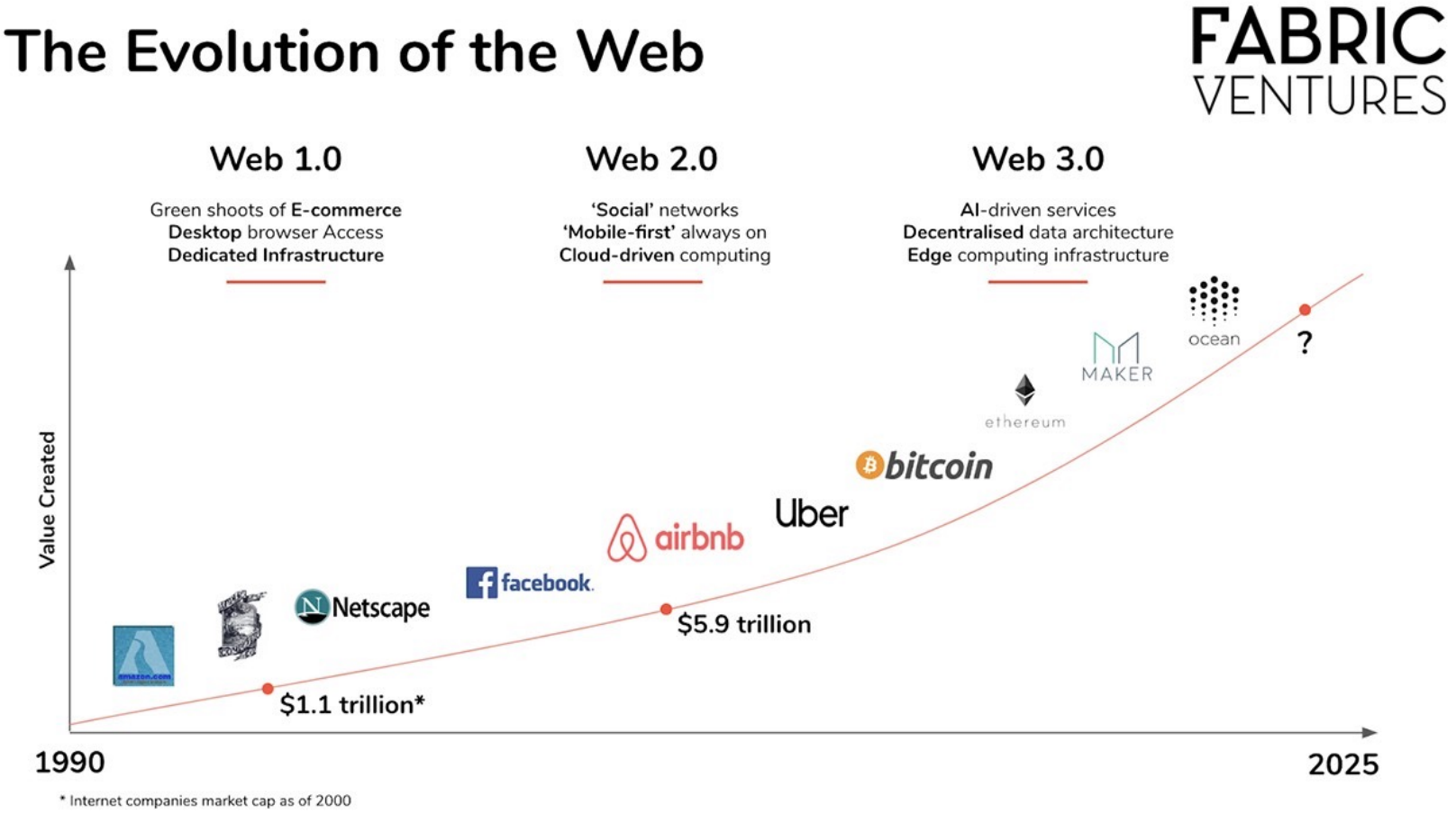
By Zachary M. Seward in California · July 23, 2013

Decentralized Systems: not a new concept

- Peer-to-Peer systems (as opposed to client-server architectures)
- 1999: Napster file sharing system
  - Followed: Gnutella, G2, eDonkey, BitTorrent, PPlive, ToR…
- Tim Berners-Lee's vision for the World Wide Web was close to a P2P*": each user of the web would be an active editor and contributor, creating and linking content to form an interlinked "web" of links".*



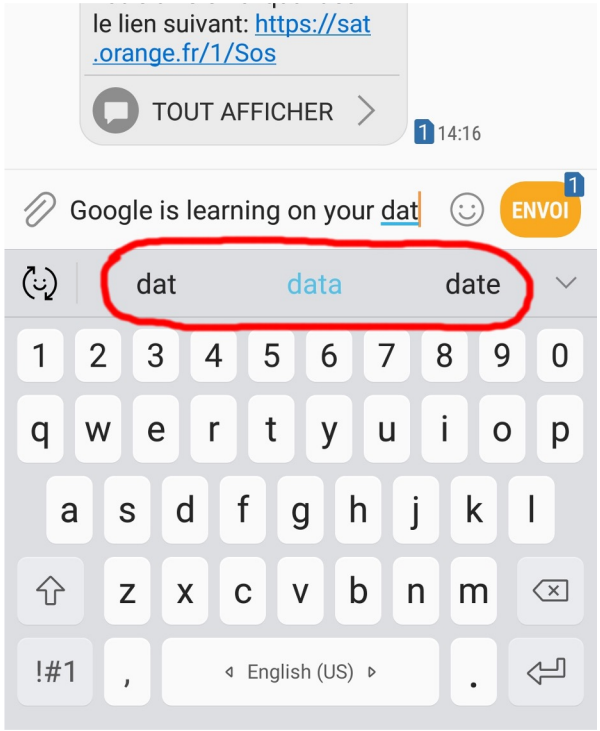
Web 3.0: a new wave of Web Decentralization



There will be no *decentralized services* without *decentralized learning*

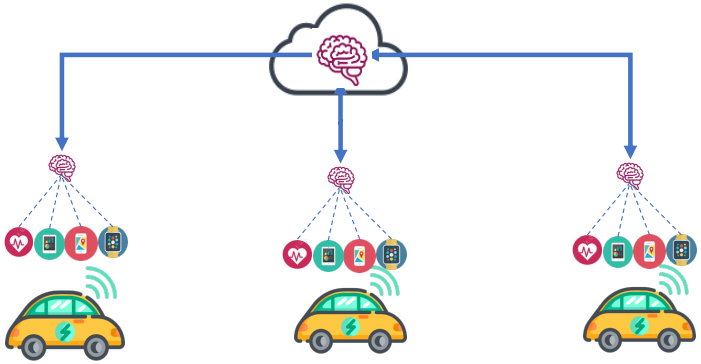
Federated Learning : a Natural Candidate

- 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)
- 2024: 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...

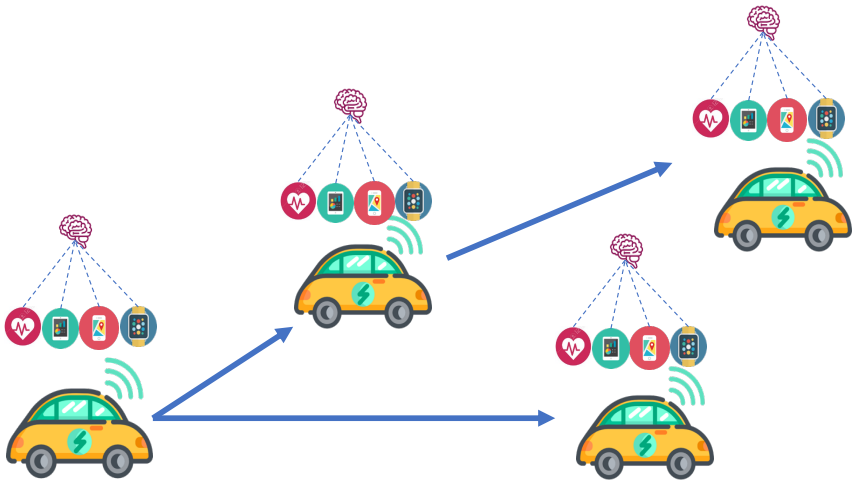


Server Orchestrated vs. Fully Decentralized

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



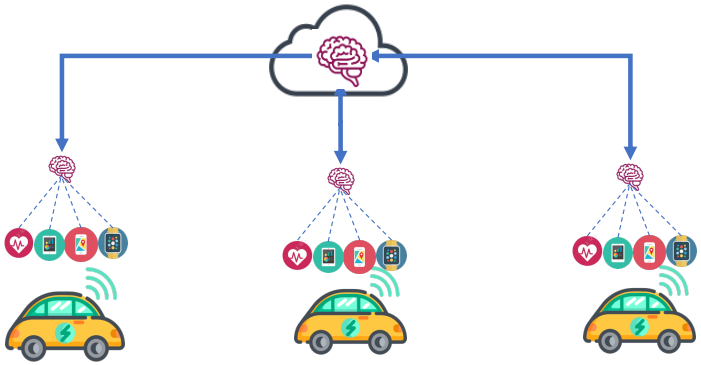
- Decentralized
  - Device to device communication
  - No global coordination, local aggregation
  - Naturally scales to a large number of devices

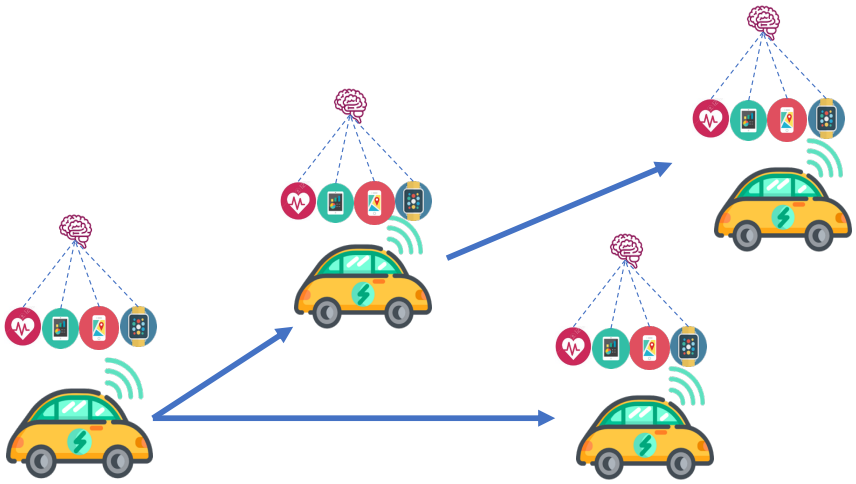


Orchestrated & Decentralized: threats

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





Distributed/Decentralized Learning in Lyon

- Addressed challenges (🐱&🐶)
  - Personalization
  - Privacy
  - Robustness (Byzantine Resilience)
- Ongoing work
  - [Personalisation] Decentralizing Recommender Systems with Gossip Learning
    - PhD Yacine Bellal [Ubicomp'22]
    - PhD Julien Nicolas -> with Mark Coates, McGill (Canada)
  - [Personalisation] FL-based Location Privacy
    - PhD Besma Khalfoun [Ubicomp'21][Middleware'20]
  - [Personalisation] Anomaly detection in ECG signals
    - PhD Joey Bekkink
  - [Privacy] Resilient FL with Trusted Execution Environments
    - PhD Aghiles Ait Messaoud [Middleware'22]
  - [Robustness] Private & Byz resilient decentralized ML
    - PhD Ousmane Touat

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?