Edge Analytics for Privacy & Performance

Hamed Haddadi

Queen Mary University of London

The Data Ecosystem

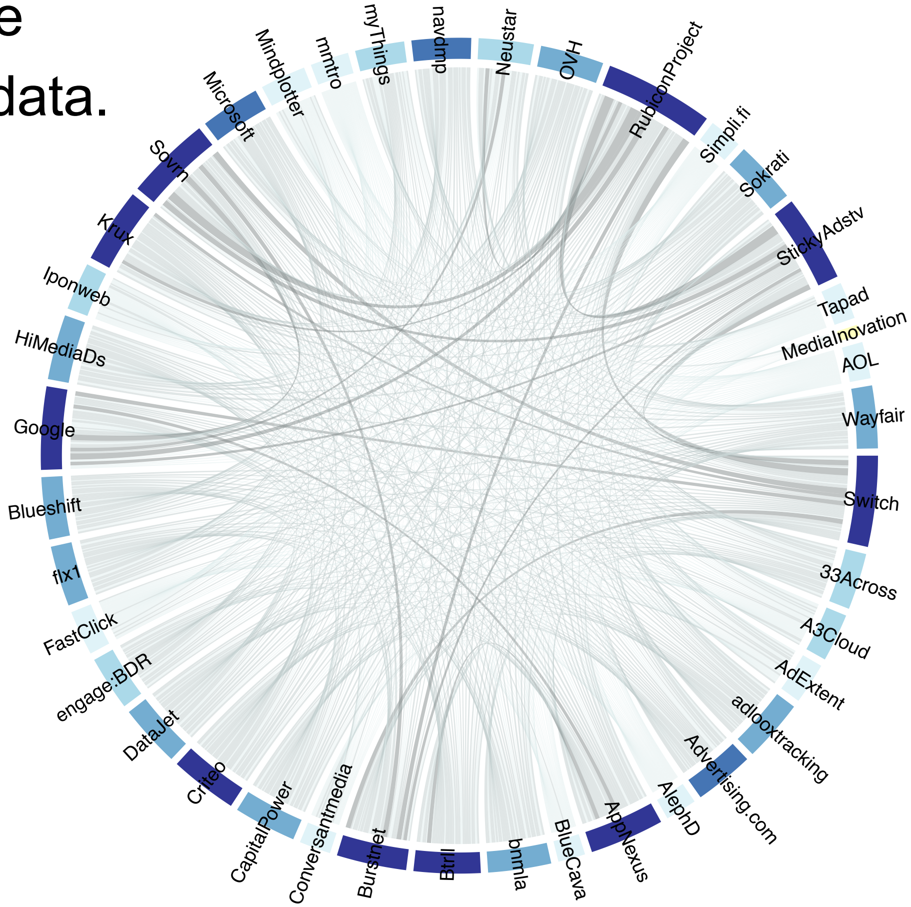
Data about us:

Data generated by us:

Data around us:

Data About Us

We found thousands of trackers across the world who follow our clicks and trade our data.



Our digital footprint include data we are not even aware of. Hence Provenance is a major issue.

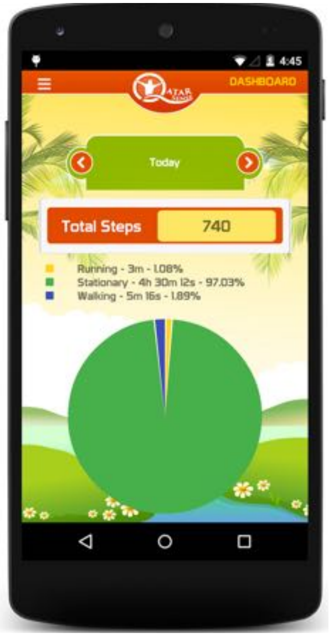
TMA 2014, PAM 2016 and “Anatomy of the Third-Party Web Tracking Ecosystem” on MIT TR 2014.

- Ad Blocking is not the long-term solution, see: “Ad-Blocking and Counter Blocking: A Slice of the Arms Race”, USENIX 2016.

Data Generated by Us

- Online Social media
- Wearable devices
  - Signals indicative of physical & mental health
  - Largely suffering from data isolation and poor user interaction (see publications: qmwearables.eecs.qmul.ac.uk)

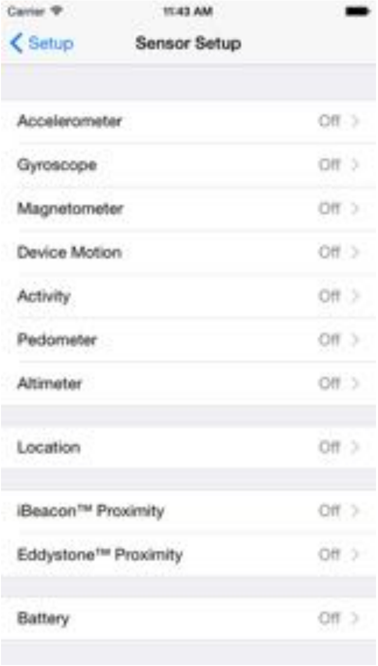
Emotional S... 8:15
?
very happy



Today

Total Steps 740

🟨 Running - 3m - 1.08%
🟩 Stationary - 4h 30m 12s - 97.03%
🟦 Walking - 5m 16s - 1.89%



Carrier 11:43 AM

< Setup Sensor Setup

Accelerometer Off >
Gyroscope Off >
Magnetometer Off >
Device Motion Off >
Activity Off >
Pedometer Off >
Altimeter Off >

Location Off >

iBeacon™ Proximity Off >
Eddystone™ Proximity Off >

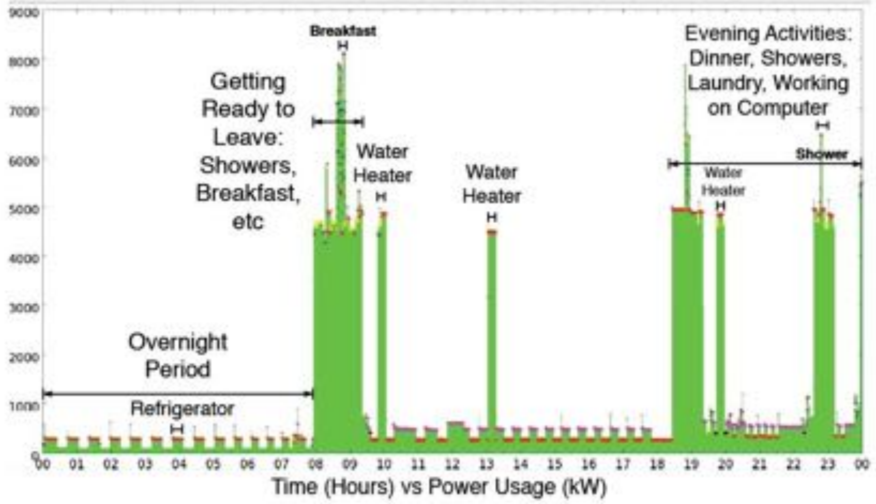
Battery Off >

SensingKit

CrowdSense
for iOS / Android

Data around us

- IoT devices
- Cyber Physical Systems



Without detailed knowledge of appliance signatures, intuitive observation with power consumption variations indicates human activity.

Credit: "Private Memoirs of a Smart Meter," Molina-Markham, et.al., 2nd ACM Workshop On Embedded Sensing Systems For Energy-Efficiency In Buildings (BuildSys 2010), Zurich, Switzerland, November 2, 2010.

www.connectedseeds.org/about/sensors

CPS Applications and Challenges

- Opportunities
  - Infrastructure monitoring
  - Understanding individuals' wellbeing & public health
  - Enabling personalised services

- Challenges
  - Real-time control & adaptation
  - Accountability & liability
  - Algorithmic bias, price discrimination, public exposure,...
  - Same with IoT/mobile data: see "Privacy Leakage in Mobile Computing: Tools, Methods, and Characteristics" 2014.

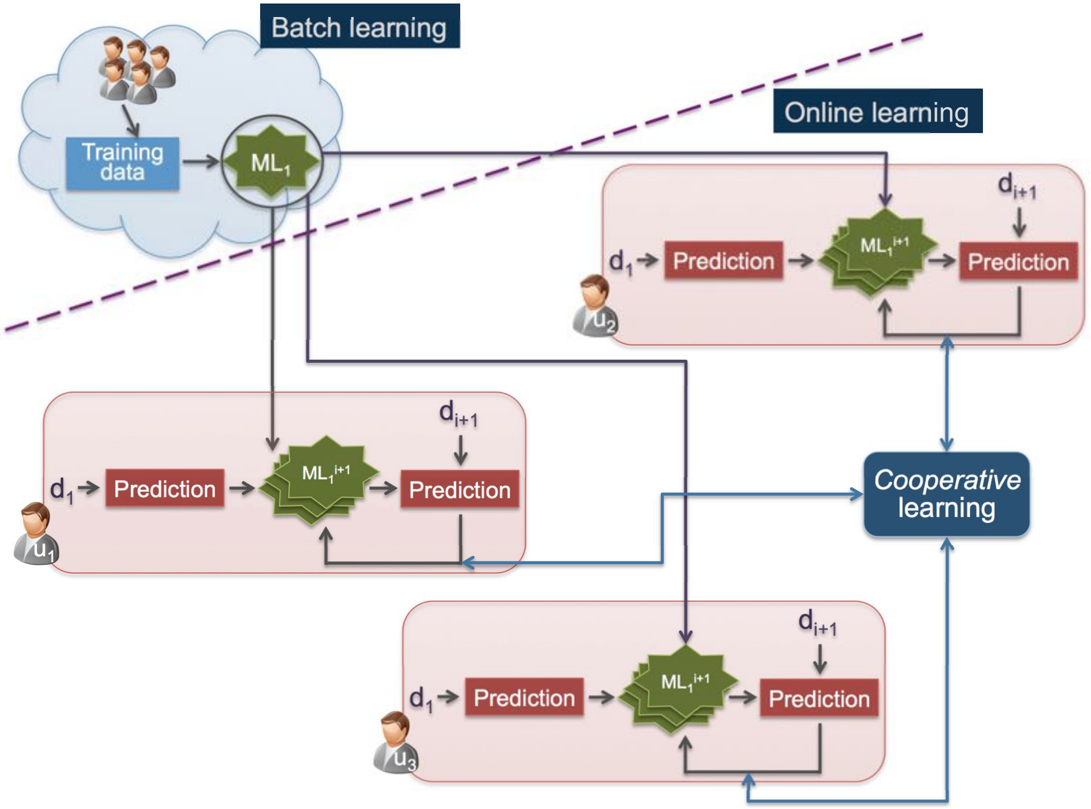
Can we do detailed, user-centric, contextual analytics without privacy disasters and legal challenges?

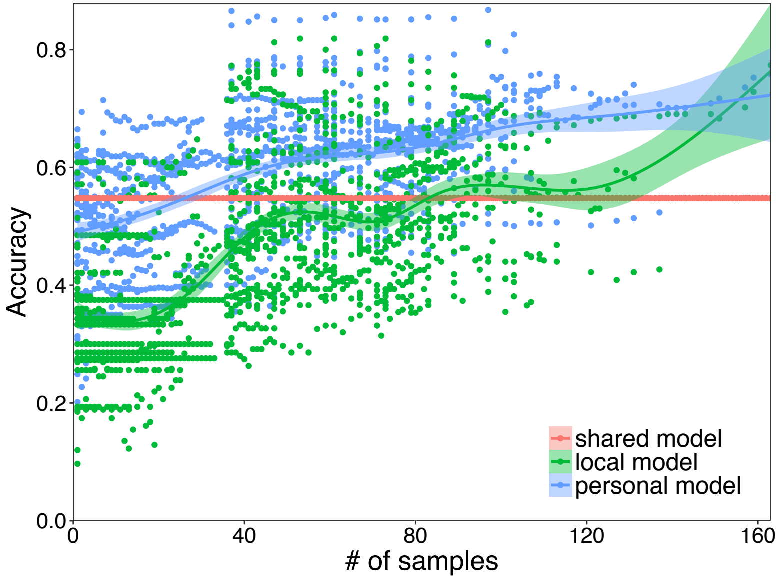
Outline

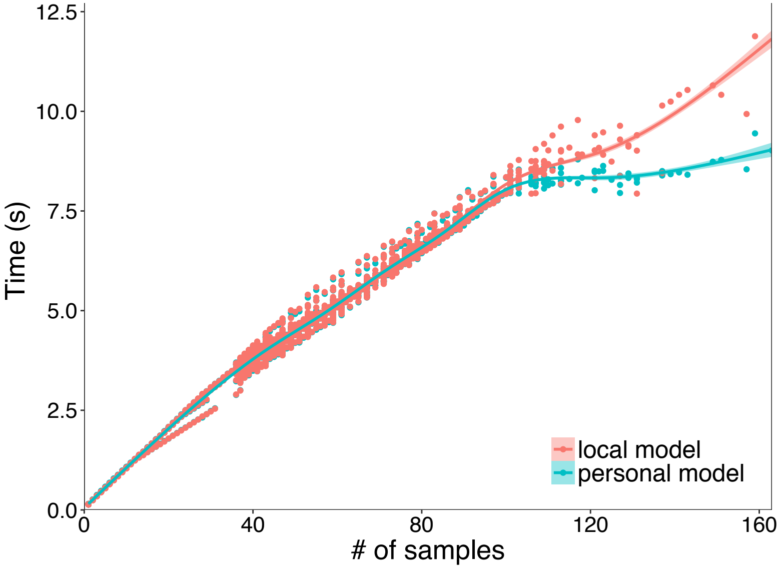
- CPS introduction & Motivations
- Privacy-preserving sensing & analytics
- The Databox platform

Cooperative learning

Case study: smartphone activity recognition



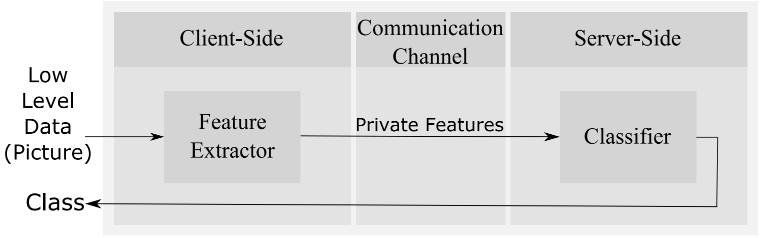


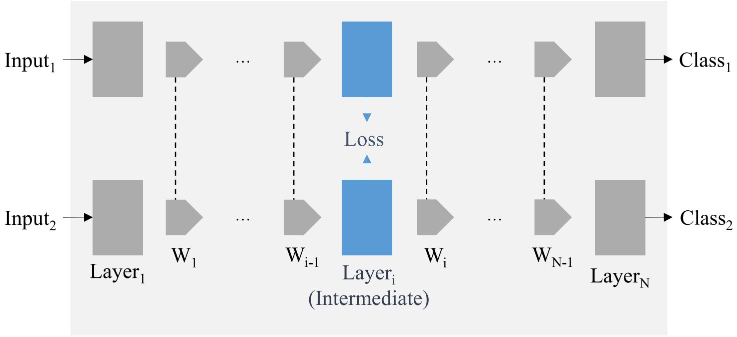


“Personal Model Training under Privacy Constraints”, on ArXiv 2017

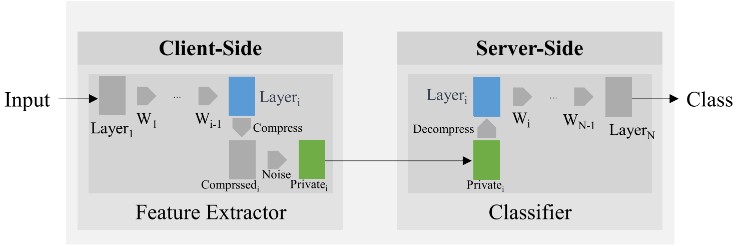
Edge computing paradigm

Case study: can we do gender detection without face recognition?



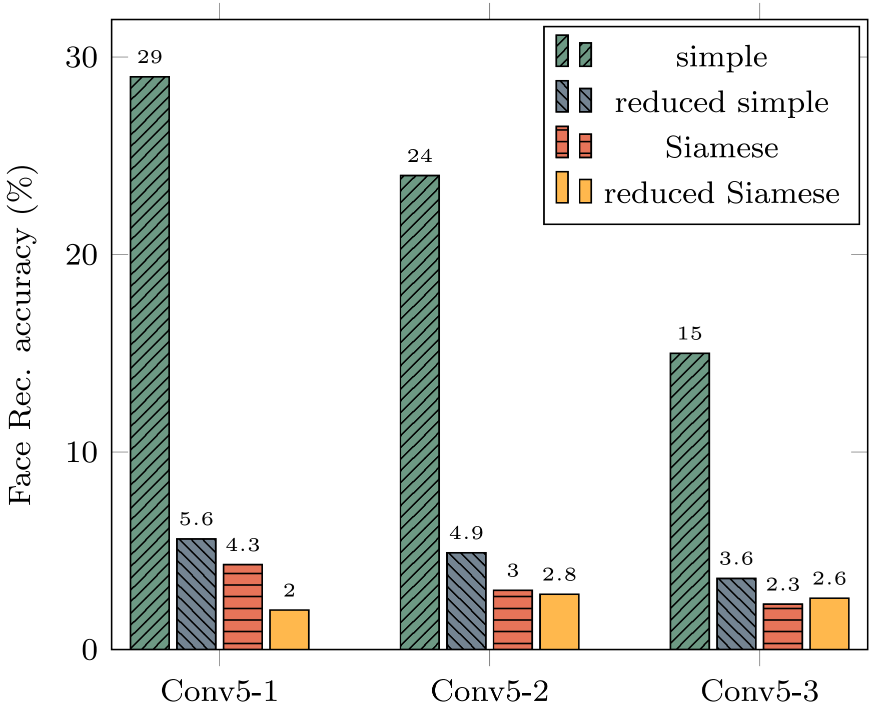


(a) Training advanced embedding with Siamese structure where we have identical network structure and weights connected by dashed lines are equal.



(b) Using advanced embedding (with PCA projection and noise addition in client side and reconstruction and classification in server side)

Accuracy on LFW			
	Conv5-1	Conv5-2	Conv5-3
simple	94%	94%	94%
reduced simple	89.7%	87%	94%
Siamese	92.7%	92.7%	93.5%
reduced Siamese	91.3%	92.9%	93.3%



“A Hybrid Deep Learning Architecture for Privacy-Preserving Mobile Analytics” on ArXiv 2017

Outline

- CPS introduction & Motivations
- Privacy-preserving sensing & analytics
- The Databox platform

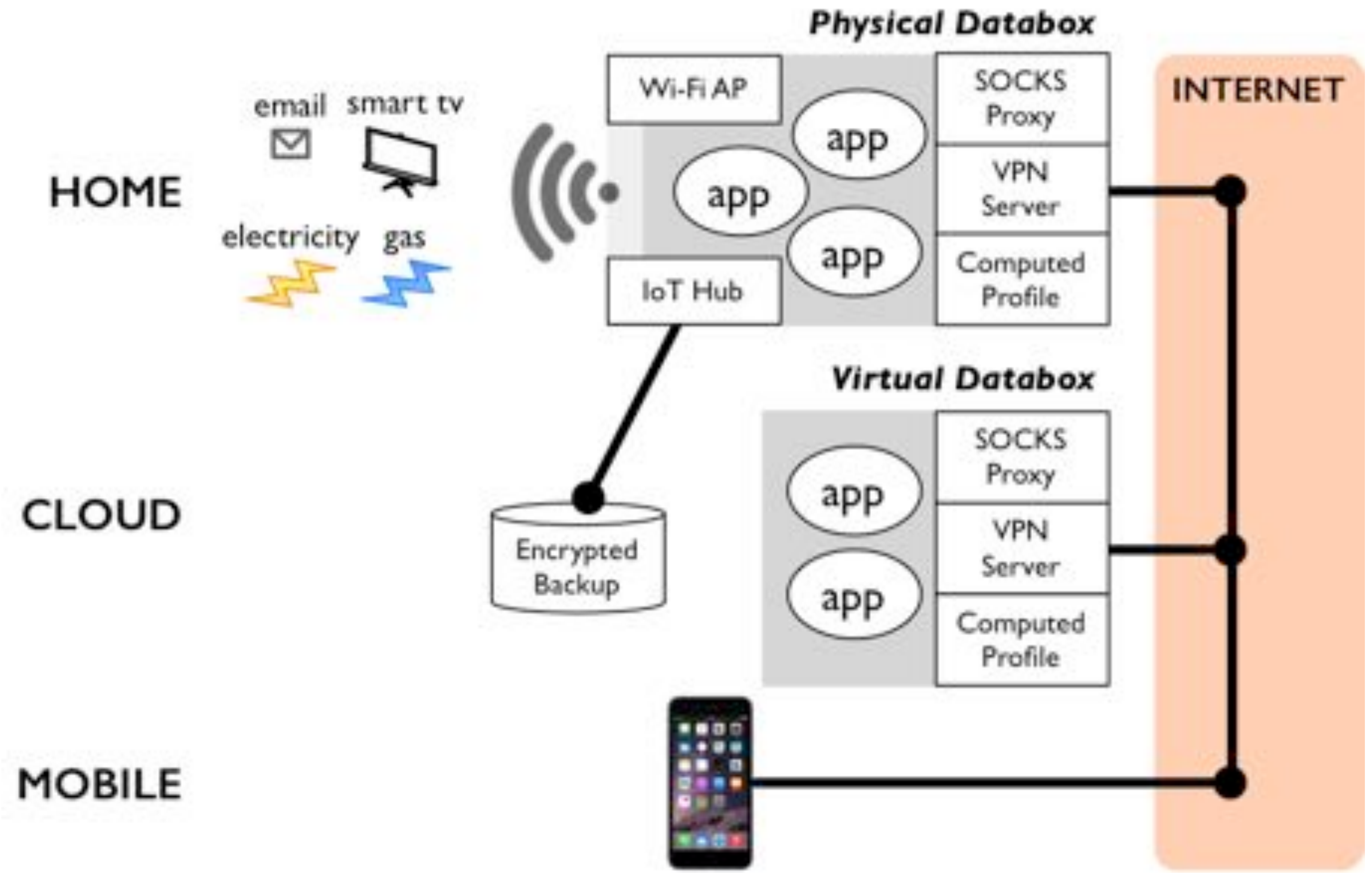
Databox vision

- An open-source personal networked system:
  - collates, curates, and mediates access to our personal data.
  - Enables interaction, sense-making, and privacy-preserving analytics on personal data, with potential wider societal benefits (Haddadi et al., CCR 2013)

- **Not** yet another data silo:
  - cooperative design approach, involving engagement with **all** stakeholders (sources, collectors, processors, organisations, and subjects)

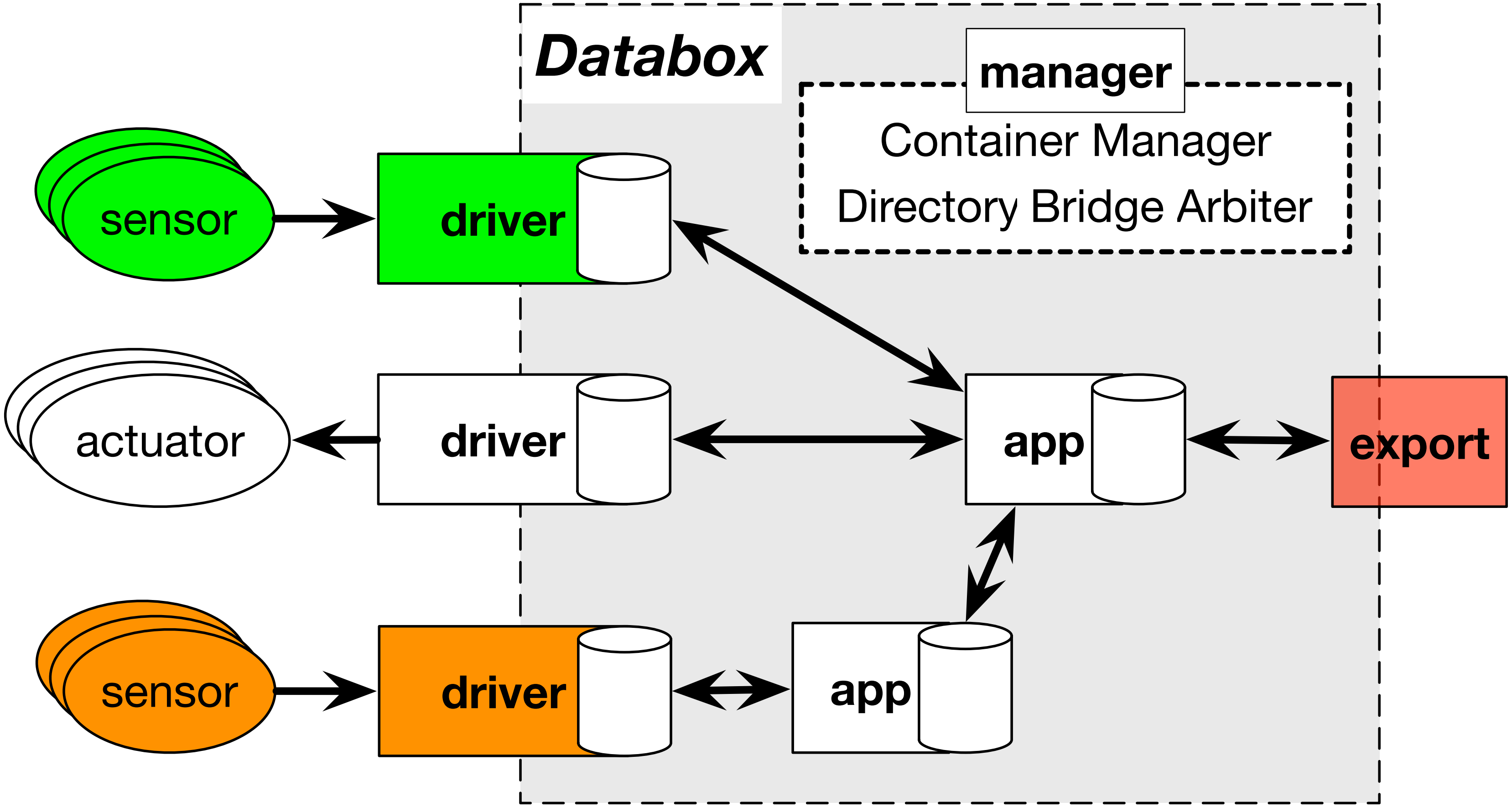
See Haddadi et al., "Personal Data: Thinking Inside the Box”, (MIT-TR, Aarhus 2015)

Privacy-Aware Personal Data Platform



EPSRC Databox: Privacy-Aware Infrastructure for Managing Personal Data
3-years, started October 2016: www.databoxproject.uk

System architecture



Databox

manager

Container Manager

Directory Bridge Arbiter

sensor

driver

actuator

driver

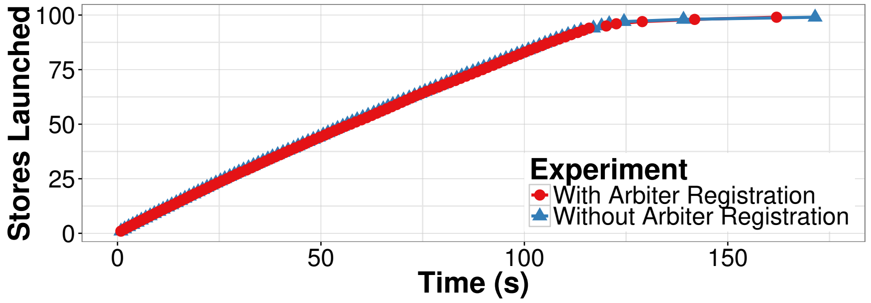
app

export

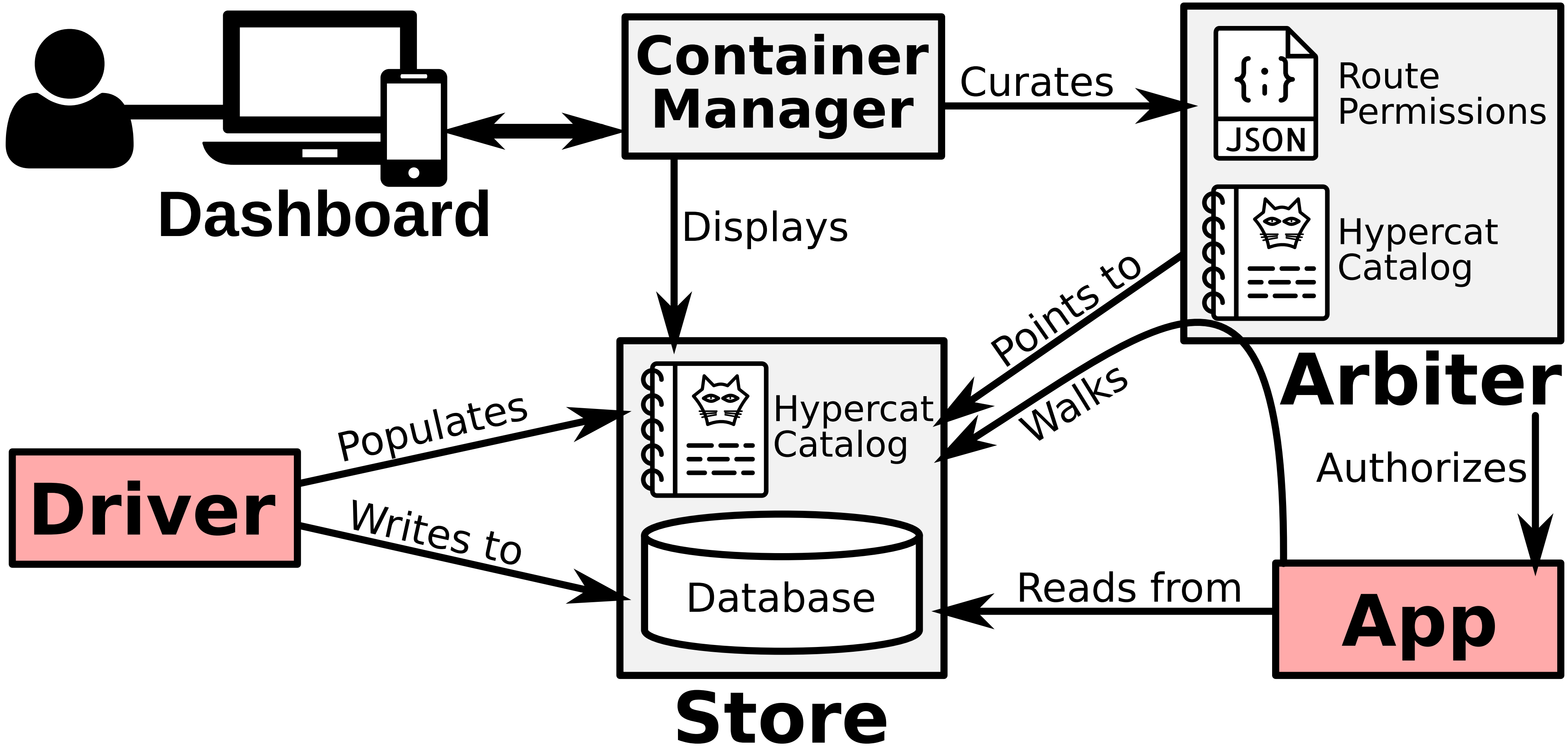
sensor

driver

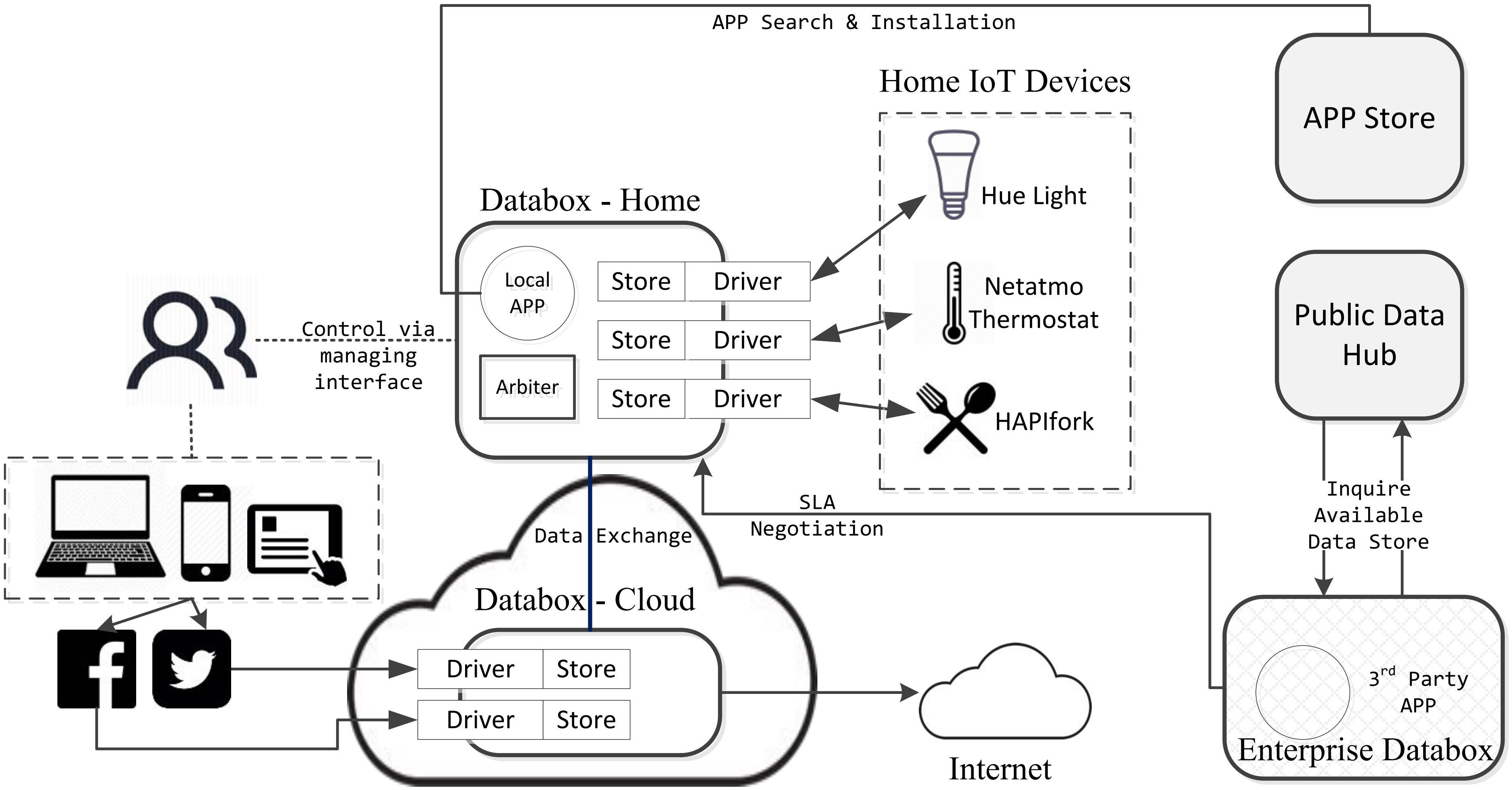
app



Interaction between the components



Databox and apps ecosystem



Code available on https://github.com/me-box/

Developer Community Engagement

DATABOX

01000100 01100001 01110100 01100001 01100010 01111000

mozilla Festival

www.databoxproject.uk

BBC R&D

Conclusions

- Personal Data analytics face complex challenges and we need new approaches for data utilisation.

- Databox, edge-computing, and user-centric processing methods are timely enablers in this direction

- Interesting new approaches for personal data, ambient sensing, actuation, and HDI

For more information, software, and papers:

http://www.eecs.qmul.ac.uk/~hamed/