From Intelligent Coordination (etc.) to Human Data Interaction

Andy Crabtree

CHIST-ERA

20 June 2018

2. Keynote Brief

Topic 2. Intelligent Computation for Dynamic Networked Environments

Novel intelligent coordination strategies are needed in the new era of computing abstractions prompted by the generation of large data volumes and by the software needs of resource-constrained and mobile devices. This trend is foreseen to exacerbate in smart environments powered by post-5G networks. Processing will have to be delegated along the network path, including current fog nodes, edge servers or cloud datacentres. Hence, the need for ubiquitous solutions that dynamically orchestrate computing tasks among these planes. The aim is to provide full context-awareness in the distribution of resources while addressing potential security and service dynamicity issues.

3. A Social Scientist's View

What does it mean?

intelligent coordination … a new era of computing abstractions … processing delegated along the network path … ubiquitous solutions that dynamically orchestrate computing tasks among planes … provide full context-awareness in the distribution of resources … address potential security and service dynamicity issues

4. The Practical Need for Decentralisation [1]

A technical view on Internet of Things (IoT)

- A **massive increase** in network traffic

- Connected appliances and devices of all kinds will hit the network hard

- In layman's terms the IoT will 'clog up' the network and in more technical terms increase 'latency' and impact 'quality of service'

- Need to move beyond the current 'cloud' paradigm

- Move computation towards 'the edge' of the network

5. The Practical Need for Decentralisation [2]

There other mundane technical reasons for decentring computation including,

- **Resilience and local autonomy**

- **Security and opaque surveillance**
    - "We set up 19 different smart gadgets and investigated what data they collected. And we found a staggering level of corporate surveillance of your home. When we used a smart TV for just 15 minutes, it connected with a staggering 700 distinct addresses on the internet." Which? https://www.which.co.uk/news/2018/06/which-investigation-reveals-staggering-level-of-smart-home-surveillance/

- **Privacy and cloud 'honeypots'**

6. The Practical Need for Decentralisation [3]

Social issues

Widespread concern about privacy has (in part) motivated new data protection regulation in Europe

- Repairing the 'crisis in trust' by putting individuals in control of their data
  - Rapid technological developments … have brought new challenges for the protection of personal data. The scale of the collection and sharing of personal data has increased significantly. Technology allows both private companies and public authorities to make use of personal data on **an unprecedented scale** in order to pursue their activities. (Recital 6)
  - Those developments require a strong and more coherent data protection framework in the Union … Natural persons should have **control of their own personal data**. (Recital 7)

7. The Practical Need for Decentralisation [4]

GDPR also motivated by economic concerns

- " … it is time to build a stronger and more coherent data protection framework in the EU … that will **allow the digital economy to develop across the internal market** … (Recital 7)

- Need for privacy balanced by fundamental interest in utility and innovation

- However, GDPR is "technologically neutral" (Recital 15) and so tells us nothing about how to build a privacy-preserving, utilitarian personal data processing ecosystem

8. The Practical Need for Decentralisation [5]

But we have some **key pointers**

- Move towards the edge of the network in order to …

- Reduce negative impacts of IoT on network latency and quality of service

- Build resilience and local autonomy into the IoT

- Address security challenges of connected devices and opaque device behaviours

- Circumvent privacy challenges created by the cloud-based paradigm

- Enable data utility to develop the digital economy

9. Databox (EPSRC EP/N028260/1)
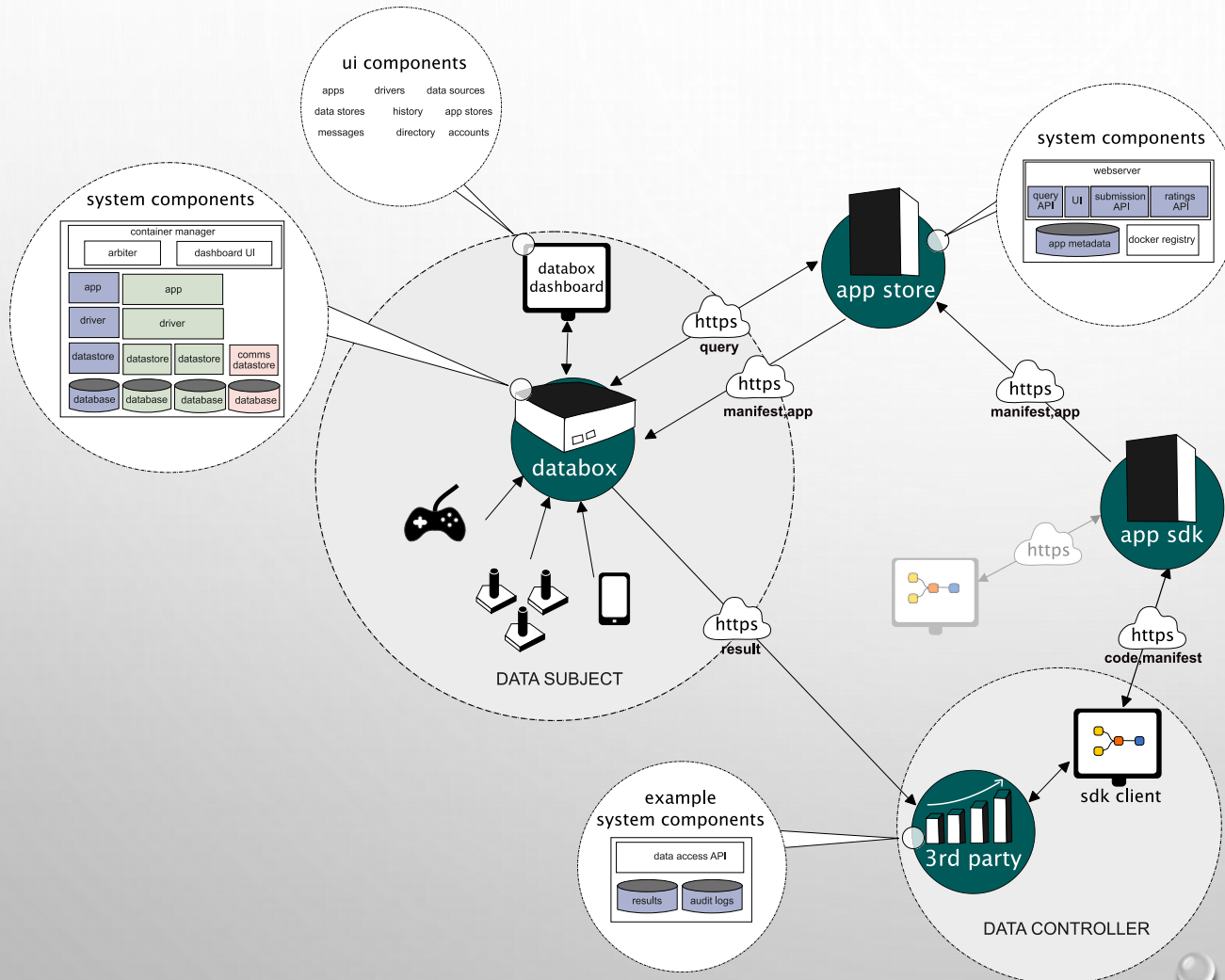
10. The IoT Databox Model [1]

The technical platform

- A **physical networked device** situated in the home

- Acts as a **gateway** to local and online 'data sources'

- **Drivers** enable data sources to be accessed, actuated and data to be stored on-the-box

- Data from sources is stored (encrypted at rest) in individual **'containers'** to reduce the potential attack surface

- Data processing is done by **sandboxed apps**

- Data distribution is limited to results of processing queries encoded in apps

- All processing operations are logged for **audit** and can be inspected or terminated at will by the user
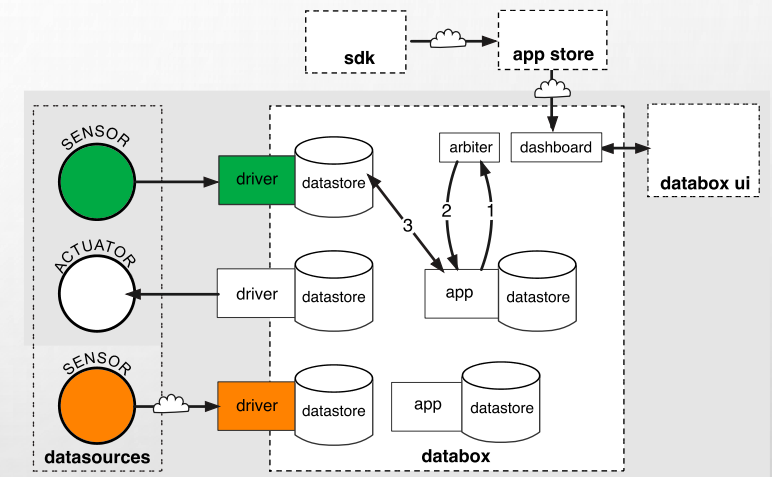
Building Accountability into the IoT,
DOI 10.1007/s40860-018-0054-5

11. The IoT Databox Model [2]

Enabling privacy-preserving app development

- The unintended impact of GDPR on developers

    - "Unfortunately the European Union's new GDPR … creates uncertainty and risk that I can't justify taking. GDPR threatens [developers] with fines of 4% of turnover or €20 million (whichever is higher) … Perversely, this new EU law hurts small and ethical startups, but helps reinforce the dominance of Facebook, Google and Twitter, who are able to prepare and defend themselves using established legal teams and cash reserves, and who now face less competition from startups." Streetland is no more :-(https://www.streetlend.com

- The Databox **SDK** enables rapid construction and publication to the app store of privacy-preserving apps

    - Including in vivo articulation of personal data
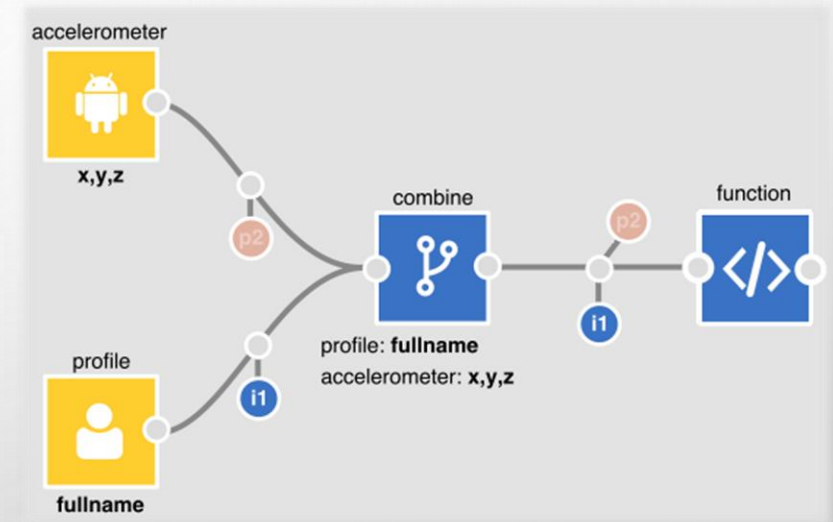
12. The IoT Databox Model [3]

Security

- In addition to container-based approach …

- Understand communication patterns of individual and clusters of IoT devices

- By **profiling** device and inter-device behaviors

- Developing **statistical** models to articulate normal and deviant behaviours

- Providing human-centred interaction mechanisms to make device behaviours **'accountable'** to ordinary people

13. The IoT Databox Model [4]

Distributed data analytics

- Core to the Databox model and an alternative to centralised processing

- Taking computing to the data

- Arguably **more flexible**

- A **better fit for modern regulation**

- And **more scalable**

14. The IoT Databox Model [5]

Fit with key pointers

- Moving towards the edge

- Reducing latency and impact on QoS

- Building-in resilience and local autonomy

- Addressing security, particularly opaque device behaviours

- Circumventing privacy challenges of cloud

- Enabling data utility to develop DE

- Edge-based gateway to local and cloud data sources

- Local data processing reduces latency and enhances QoS

- Local data processing also enables resilience to be built into IoT

- Containerisation reduces attack surface, and device profiling can surface hidden device behaviours

- Data stays on-the-box and distribution is 'minimised' to results of processing queries

- A familiar user-oriented app environment

15. Databox & HDI

Databox is an instantiation of the Human Data Interaction paradigm

https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/human-data-interaction

- HDI is distinct from HCI

  - HDI recognises that **data is a first class object in its own right**

  - An area of fundamental research

- Current investment largely focuses on data and the data processing ecosystem

- But what about the **H** and the **I** in HDI?

16. Human Data Interaction

Key challenges

- Need to consider how to build the **Human** into the emerging Data ecosystem and enable mundane **Interaction**

    - Within almost every discipline related to the digital economy, there are critical and emerging issues around humans and the data they generate … There is an urgent need to guide the realisation of system design principles that … fit with the ethics and values acceptable to wider society. Those who are expert in development of the systems, algorithms and analytics that raise such issues face challenging **culture gaps** … [especially] with regard to those who are inexpert in technology but who are increasingly impacted by it in their everyday lives. (Human Data Interaction: Legibility, Agency, Negotiability EPSRC EP/R045178/1)

17. Why We Need to Build the Emerging Data Ecosystem Around Human Interaction

A salutary tale

- Homework (EPRSC EP/F064276/1) and the practical impossibility of reengineering home networking
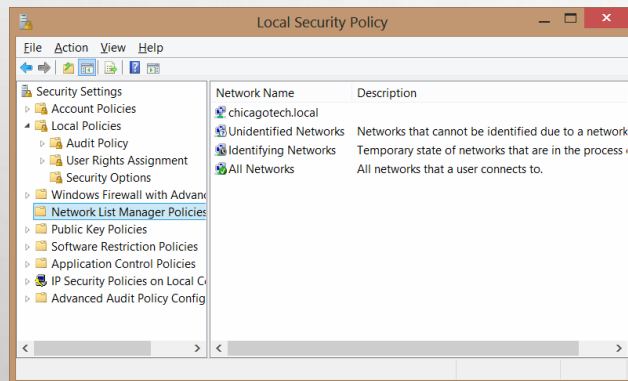
Home networking tools are designed for experts not ordinary people

If they were designed for ordinary people they would look very different

Human-centred networking
DOI 10.1007/s00779-014-0771-6

- Once it's built, it's built and we must live with its imperfections, so the moral of the story is build it around ordinary users in the first place

18. From Intelligent Coordination (etc.) to Human Data Interaction

Topic 2. Intelligent computation for dynamic networked environments

- If intelligent computation is to be effective it will have find a place **in everyday life**
  - And that requires that it is **accountable to** everyday life

- Fundamental challenges exist around human-data interaction in diverse social contexts
  - Databox type solutions will not address all

- At root we need to provide **ordinary people** with the tools they need to **make sense** of intelligent computation

19. Questions

Bearing in mind that I am a social scientist

- So please don't ask me any hard technical questions
  - Like how is Databox actually implemented or how does the arbiter enforce user choices

- I probably won't know the answer
  - "For a social scientist ignorance is more excusable than vagueness."
    Arthur L. Stinchcombe