Privacy-Aware Infrastructure for Managing Personal Data

Yousef Amar, Hamed Haddadi, Richard Mortier

http://www.databoxproject.uk/

Background

Personal data is proliferating due to:

- Our lives moving online more and more
- Through the rise of ubiquitous sensing via mobile devices
- Through the rise of ubiquitous sensing via IoT devices

Concerns over privacy, trust, and security are expressed more and more as different parties attempt to take advantage of this rich assortment of data.

We describe an architecture for the *Databox*, a personal networked device backed by cloud services, to:

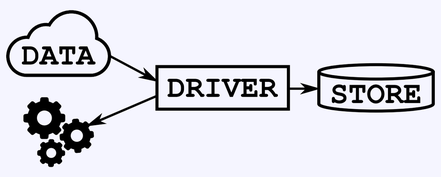
- Allow users to **collate**, **curate**, and **mediate** their data
- Enable multi-modal personal data analytics while providing **accountability** and **control** mechanisms
- Protect a user's privacy, allowing us to recover control of our online lives

Architecture

Databox components run locally on physical hardware, and remotely in the cloud, in a hybrid manner.

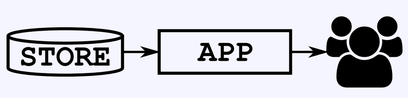
The main ones are:

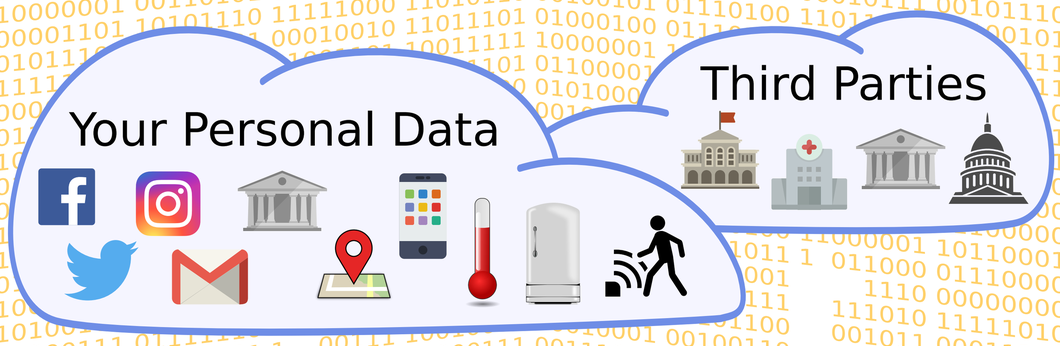
Drivers
Interface between the stores and the outside.

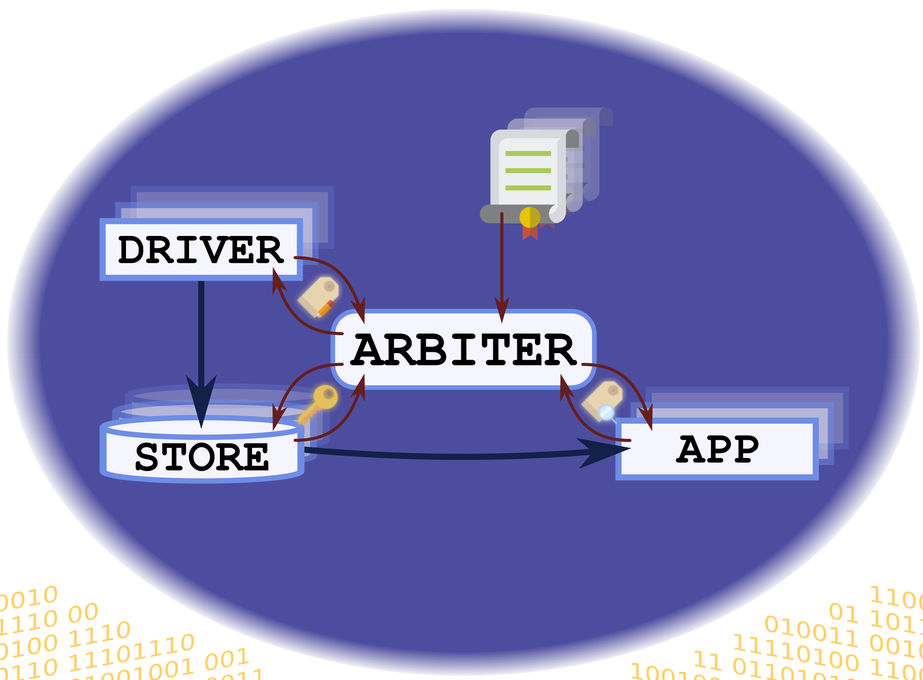


Stores
Make data available on presentation of appropriate credentials, and log access.

Applications
Process data and can connect to external parties to emit results of analytics, all to the extent of app permissions.

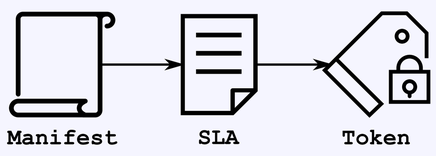






The Arbiter

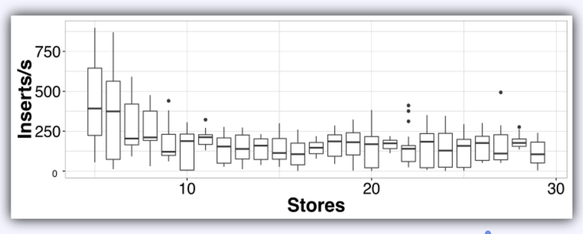
The arbiter manages interactions internally between components and external parties by minting refinable bearer tokens based on app manifests that list the extent of permissions an app may require.



Through a dashboard interface, a user negotiates what permissions they are prepared to grant based on their individual risk propensity and perceived benefits, generating a Service Level Agreement, that the arbiter encodes into signed bearer tokens.

Stores can then validate these tokens and grant read and/or write access to themselves within the constraints of approved permissions, keeping access to your personal data limited and secure at all times.

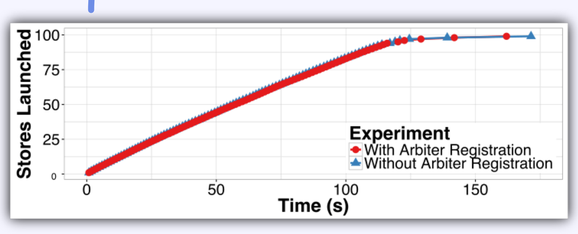
Preliminary Evaluation



The rate at which data is written to stores is not noticeably affected by the number of stores being written to simultaneously. Indeed – as supplementary tests further indicate – when it comes to scaling up, the bottleneck is system memory.

The launch rates with and without arbiter interaction match almost exactly.

The rate at which new stores are launched slows down as memory runs out at just below one hundred stores.



Conclusion and Next Steps

We have presented the architecture, components, and early evaluations of the Databox; a networked system for managing access to personal data while enabling privacy, negotiability, auditing, and control.

Databox provides APIs supporting a range of uses ranging from privacy-preserving detailed analytics (e.g. on mental/physical health) to aggregate population surveying and statistics.

For the ecosystem to be widely deployable, next steps to be taken include:

- Understanding user interactions
- Characterising performance bounds
- Limiting security and privacy risks

The Databox project is open source; contribute at:

https://github.com/me-box

References

Hamed Haddadi, Heidi Howard, Amir Chaudhry, Jon Crowcroft, Anil Madhavapeddy, Derek McAuley, Richard Mortier, "Personal Data: Thinking Inside the Box", *The 5th decennial Aarhus conference* (Aarhus 2015), August 2015, available on arXiv

Hamed Haddadi, Richard Mortier, Derek McAuley, Jon Crowcroft, "Human-data interaction", *University of Cambridge* (2013)

Queen Mary University of London

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

UNIVERSITY OF CAMBRIDGE