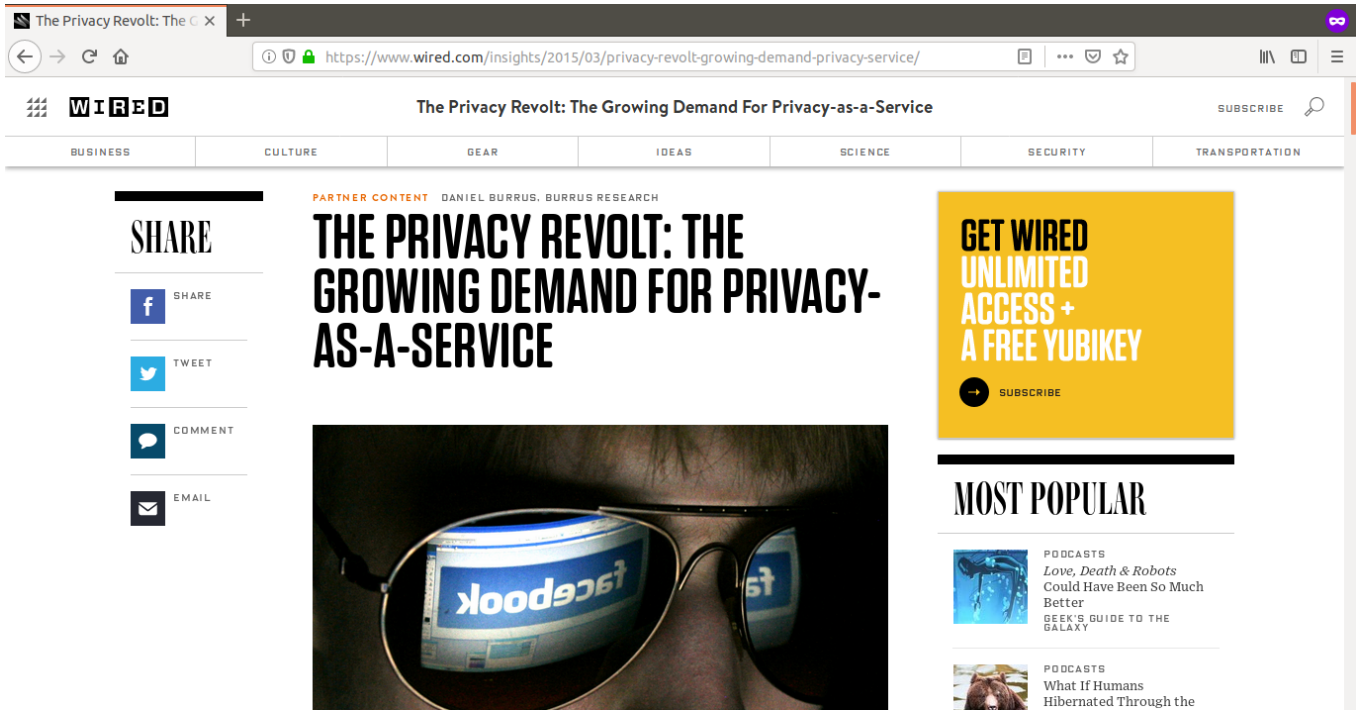
Privacy as a Service (slides)

Ashish Dandekar, Debabrota Basu, Poh Geong Sen, Jia Xu,
Stéphane Bressan

July 18, 2019

Motivation



What are you doing to provide privacy-as-a-service (PaaS)? Because this is a service that consumers increasingly view less as a bonus and more as an absolute necessity. How are you preparing for the coming privacy revolt? (Wired, March 2015).

NUS National University of Singapore

Overview of *Liánchéng*

- **Workflow-as-a-Service.** A data sharing cloud system that provides a graphical workflow language.
- **Privacy-as-a-Service.** A data sharing cloud system that provides operators to publish not only anonymised data but also models created by statistical machine learning with differential privacy guarantees.

Liánchéng is deployed on a hardware infrastructure consisting of 128 commodity servers!

NUS National University of Singapore

Liánchéng: A data sharing platform

- Liánchéng provides every user a private account that she uses to upload, download, organise and manage her data in the cloud. The internal sharing mechanism (user-to-user) relies on access control lists on directories.
- Liánchéng provides additional publishing mechanisms, such as public access through URLs, for files.
- Liánchéng provides both a web interface and a desktop computer synchronisation agent.

Liánchéng: Workflow-as-a-Service

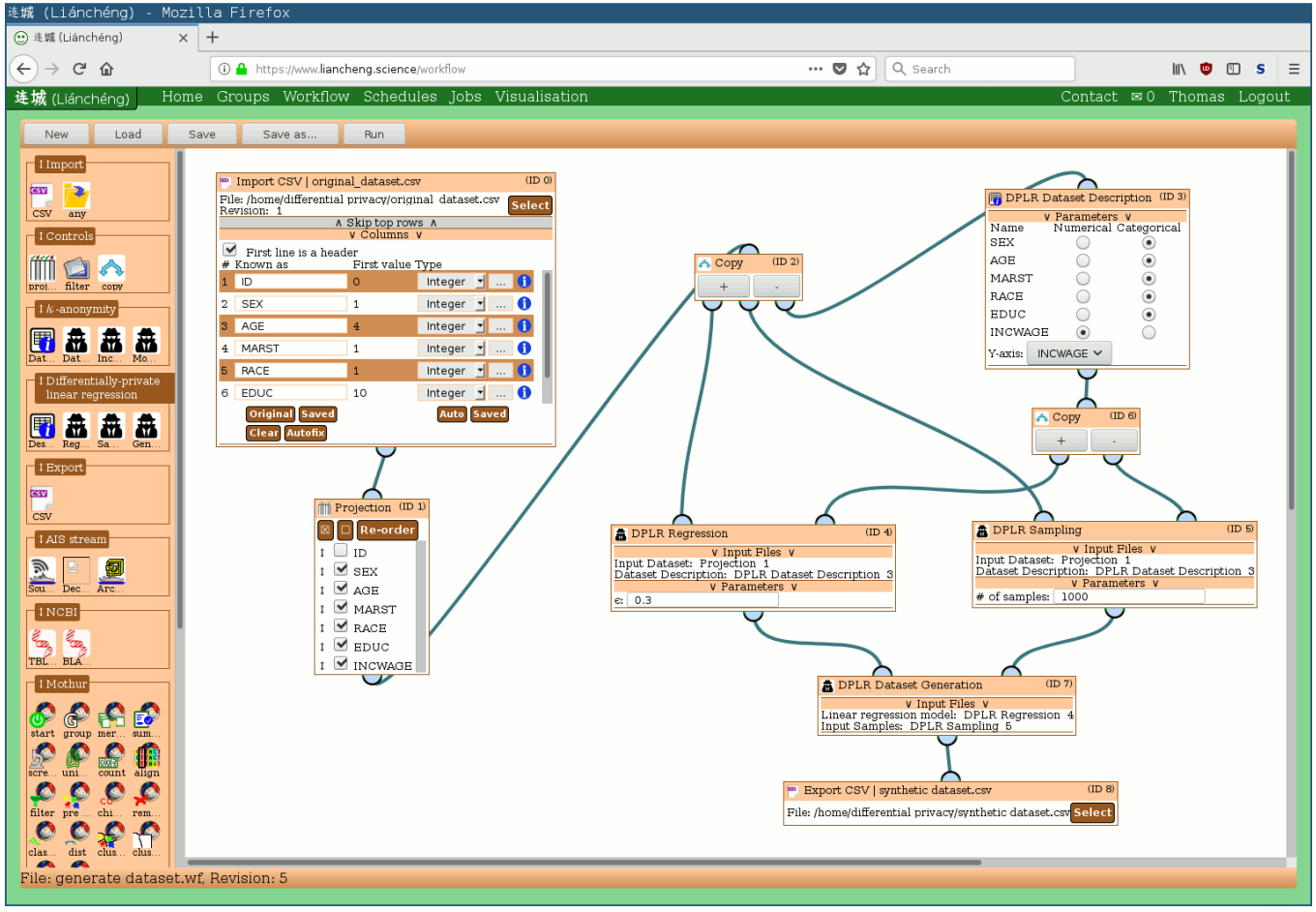
- Liánchéng offers an interactive GUI-based workflow language and with domain specific operators that work on the data in the cloud.
- Liánchéng workflow is a directed acyclic graph whose vertices represent operators and whose edges represent data flow. An operator can have an arbitrary number of parameters and has at least one input or output interface.

NUS National University of Singapore

Liánchéng: Privacy-as-a-Service

- **Publishing data.** Liánchéng provides traditional anonymisation operators such as *k-anonymity [Sweeney, 2002]*, *l-diversity [Machanavajjhala et al., 2006] and t-closeness [Li et al., 2007]*. Alternatively, it also provides operators to synthetically generate datasets using differentially private machine learning model trained on private datasets.
- **Publishing models.** Liánchéng provides operators to publish parameters of parametric models using Functional mechanism [Zhang et al., 2012] and publishing non-parametric models using functional perturbation [Hall et al., 2012].

Liánchéng: Screenshot



Demo

References I

Hall, R., Rinaldo, A., and Wasserman, L. (2012).
Random differential privacy.
Journal of Privacy and Confidentiality, 4(2):43–59.

Li, N., Li, T., and Venkatasubramanian, S. (2007).
t-closeness: Privacy beyond k-anonymity and l-diversity.
In *Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on*, pages 106–115. IEEE.

Machanavajjhala, A., Gehrke, J., Kifer, D., and Venkitasubramaniam, M. (2006).
\ell-diversity: Privacy beyond\kappa-anonymity.
In *null*, page 24. IEEE.

Sweeney, L. (2002).
k-anonymity: A model for protecting privacy.
International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(05):557–570.

Zhang, J., Zhang, Z., Xiao, X., Yang, Y., and Winslett, M. (2012).
Functional mechanism: regression analysis under differential privacy.
Proceedings of the VLDB Endowment, 5(11):1364–1375.

NUS National University of Singapore