HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

Privacy-aware machine learning and AI

Antti Honkela
Helsinki Institute for Information Technology HIIT
Department of Mathematics and Statistics &
Department of Public Health
University of Helsinki

Aalto University

UNIVERSITY OF HELSINKI

AI = learning from data

Interesting data are personal

Simple privacy protection is insufficient

Model Inversion Attacks that Exploit Confidence Information and Basic Countermeasures

Matt Fredrikson
Carnegie Mellon University

Somesh Jha
University of Wisconsin–Madison

Thomas Ristenpart
Cornell Tech

Figure 1: An image recovered using a new model inversion attack (left) and a training set image of the victim (right). The attacker is given only the person's name and access to a facial recognition system that returns a class confidence score.

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

Differential privacy (Dwork *et al.*, 2006)

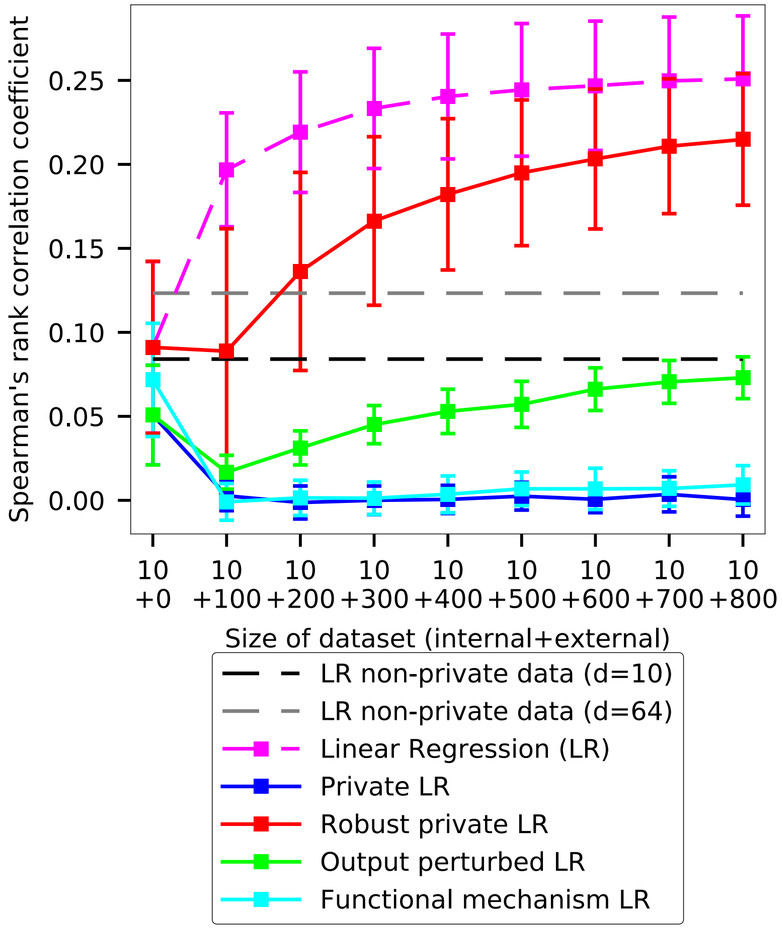
- The output of a *differentially private* algorithm must not change too much, even if one person's data are changed

$$\Pr(\mathcal{M}(\mathcal{D}) \in S) \leq e^{\epsilon} \Pr(\mathcal{M}(\mathcal{D}') \in S)$$

- New industry standard: Apple, Google, US Census 2020, ...
- Protection valid against adversaries with side information
- Degrades gracefully with repeated use, groups of related individuals, ...

Research highlight 1: Strong privacy with limited data

- Differential privacy challenge: strong privacy degrades modelling performance
- Case study: prediction of cancer drug efficacy using genomic data
- Our approach can provide reasonably accurate predictions already with <1000 samples (ϵ=2)

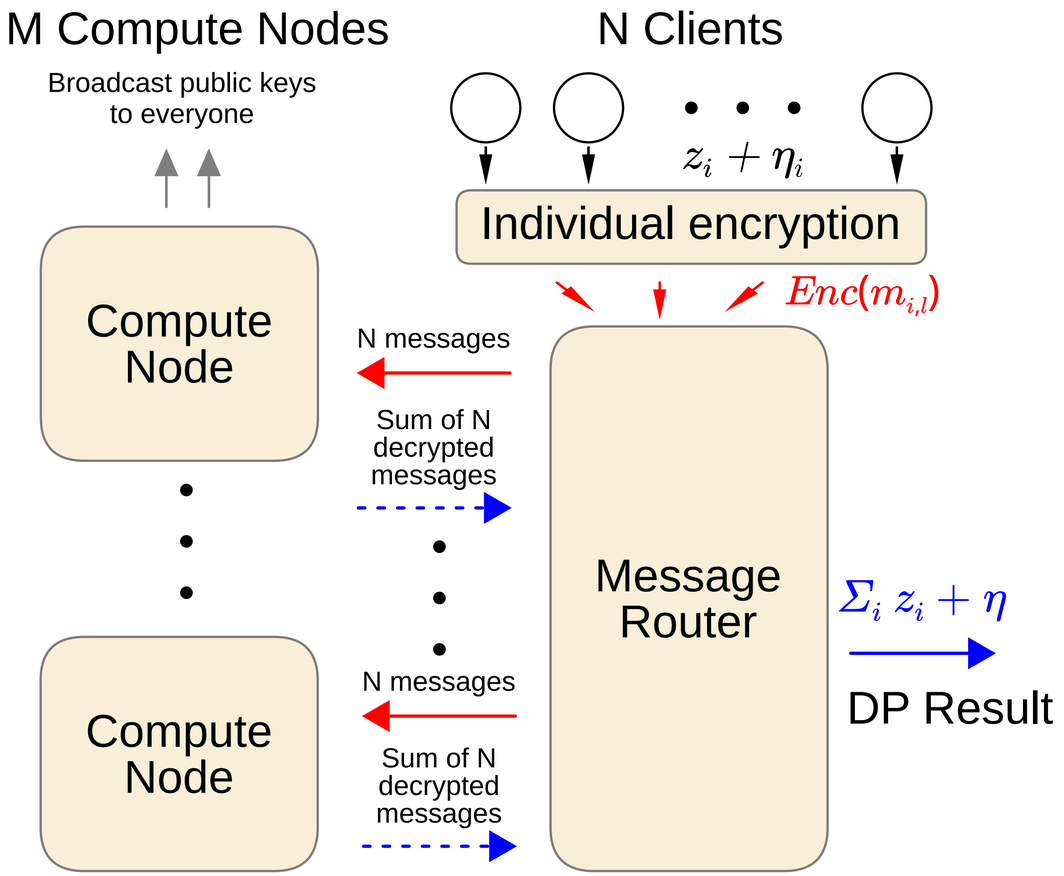


Joint work with: Mrinal Das, Arttu Nieminen, Onur Dikmen and Samuel Kaski
arXiv:1606.02109

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

Research highlight 2: Private learning on distributed data

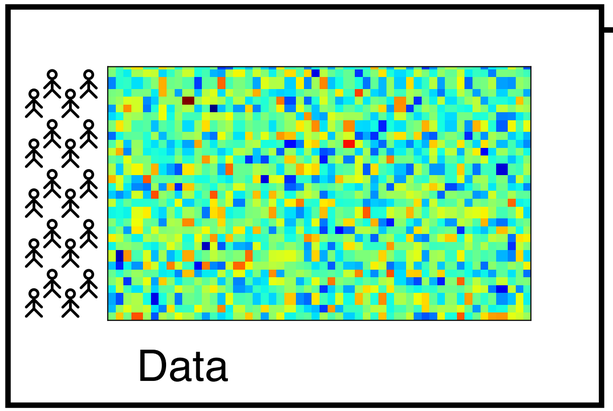
- Collecting the data to a single location is a risk in itself: better to work with data held by individual users
- Combining DP with Secure Multi-party Computation: data processed in encrypted form until DP noise is added
- Quick and scalable, communication is the main bottleneck



Joint work with Mikko Heikkilä, Eemil Lagerspetz, Samuel Kaski, Kana Shimizu and Sasu Tarkoma.
NIPS 2017, arXiv:1703.01106

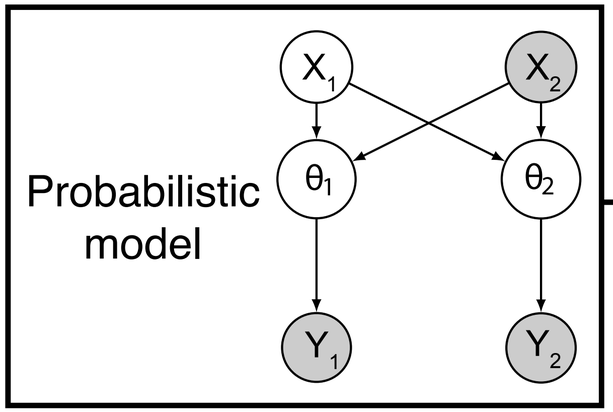
HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

Towards general privacy-aware probabilistic modelling and programming



Privacy wall

$$\left|\log \frac{\Pr(\mathcal{M}(\mathcal{D} \in S))}{\Pr(\mathcal{M}(\mathcal{D}' \in S))}\right| \leq \epsilon$$



Variational inference

$$\max \mathcal{L}(q_\xi) = \mathrm{E}_{q_\xi(\theta)}\left[\log \frac{p(\mathcal{D}, \theta)}{q_\xi(\theta)}\right]$$

Differentially private variational inference (Jälkö, Dikmen & Honkela, UAI 2017)

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

Contact details and acknowledgements

Antti Honkela
email: antti.honkela@helsinki.fi

Collaborators:
Mrinal Das, Onur Dikmen, Mikko Heikkilä, Joonas Jälkö, Eemil Lagerspetz, Arttu Nieminen
Samuel Kaski, Sasu Tarkoma, Kana Shimizu

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY