Privacy Tech work at the Turing

Jon Crowcroft,

http://www.cl.cam.ac.uk/~jac22

.

Privacy Enhancing Technologies

- we have law (e.g. GDPR):

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2865811

- We have technologies:

https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/privacy-enhancing-technologies-report.pdf

- then we have people:

https://www.rsaconference.com/writable/presentations/file_upload/pdac-f02-why-johnny-still-can-t-encrypt.pdf

Why PETs?

- Discourage under-sharing of data by encouraging PETs
- Discourage over-sharing of data with inappropriate PETs
- Figure out best use case & practice for each PET…

technologies

- Don't keep data (extreme data minimisation☺
- Encrypt data at rest and in transit
- Keys, roles, id management challenges…

- Snazzier technologies, possibly allowing use of public cloud
- Diffpriv, enclaves, MPC, SHC, edge cloud
- Turing has projects on all these....

Differential privacy

- Available, works, understood
- Limited number of users/queries
- Unknown known data can do triangulation attacks
- Probably limited use for e.g. NHS cases
- But other cases, or in combination
- Lots of related tech (fuzzing, k-anon etc)

Secure enclaves

- Trusted Exec Environments – de-risk of OS/hypervisor vulnerabilities
- Trustzone, other tech ok, but not for BIG data science
- Ok for modest scale stuff though
- SGX side channel pb…
- mitigations or combined with other tech may be ok
- Have a report if people interested…

Secure multiparty computation

- Clever – hard to explain to people
- High latency, but potentially high throughput
- Good fit for inherently federated data use cases as well
- (NHS or edge cloud…)

Homomorphic encryption

- Excellent privacy tech, but low throughput
- Some interesting use cases including
- Homomorphic crypto deep learning/training
- At decent speed – several groups have
- Promising – industry has tools and use with diffpriv (in NHS)
- Turing has benchmarking toolkit for comparing solutions…

Data safe havens

- Requirement from Turing work with NHS, FCA, NCSC et al
- Private data centers
- But still internally shared
- Defense in depth
- Define 4 levels of security – design will be published….
- Have used for several large data studies....
- A bit clunky, but confidentiality and ethics/legal solutions

Edge cloud

- Everyone keeps own data
- Fully decentralized machine learning
- Some attacks (see Emiliano’s work)
- But mitigatable (same as federated, and lower risk inherently)
- Several projects/products/big company interest too

Challenges

- Explainability, usability, performance
- Key management for dummies ☺
- Identity management (Canada has recent cool announcements)
- Verification&robustness
- Long term fun with quantum computing, sensing
  - Actually pretty avoidable (QC-resistant encryption algo exist)
  - QKD quite interesting, but narrow applications to be honest…
- Add yours here...
- “Every decoding is another encoding”