Computing over Distributed Sensitive Data

PIs: Stephen Chong, James Honaker, Kobbi Nissim, Salil Vadhan (Harvard), Marco Gaboardi (University at Buffalo)

http://privacytools.seas.harvard.edu/computing-over-distributed-sensitive-data

The Research Challenge:

A vast array of different organizations collect similar data or data about similar populations. Sharing this data can bring benefits in social, scientific, business, and security domains. When the data is sensitive, can we unlock these benefits while avoiding the need to share the data between the different entities?

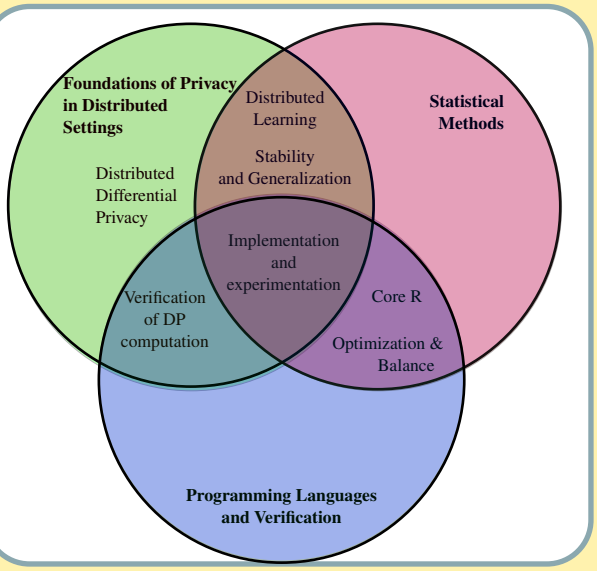
Our Vision

Enabling combined analyses across multiple datasets, leaving each dataset in the hands of its owner, who may choose to opt-in or opt-out of the analysis.

Sharing the computation (and not the data) across multiple entities.

Computing over similar data, or data about similar populations, stored at different organizations as a single, massive, joint dataset.

Re-design the computation to alleviate the privacy concerns of different entities.



Our framework:

Differential privacy: A formal mathematical framework for measuring and enforcing the privacy guarantees provided by statistical computations.

Our approach:

Distributing differential privacy computations: Using algorithm design, cryptographic, statistical, and programming languages tools.

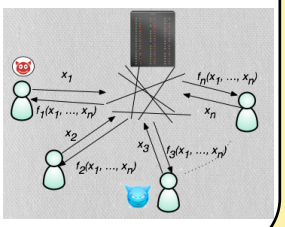
Foundation of Privacy in a Distributed Setting

Goals:

Theory of distributed differential privacy.

Protocols for distributed summation and aggregation.

Algorithms for distributed learning and statistical inference



Programming Languages and Verification

Goals:

Trusted interpreter for a core R language

Symbolic execution for differential privacy

Protocols for remote attestation for privacy

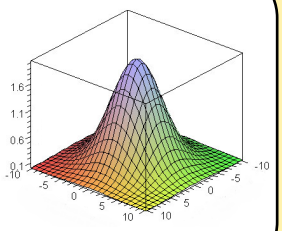
Statistical Methods

Goals:

Verified statistical estimation and balancing

Parametric optimizations over distributed sensitive data

Balance and matching for causa inference over distributed sensitive data



Preliminary results:

Generic Attacks on Secure Outsourced Databases. Kellaris G, Kollios G, Nissim K, O'Neill A. CCS 2016.

Differentially Private Bayesian Programming. G. Barthe, G. P. Farina, M. Gaboardi, E. J. Gallego Arias, A. D. Gordon, J. Hsu, P.-Y. Strub. CCS 2016.

Advanced Probabilistic Couplings for Differential Privacy. G. Barthe, N. Fong, M. Gaboardi, B. Grégoire, J. Hsu, and P.-Y. Strub. CCS 2016.

Interested in meeting the PIs? Attach post-it note below!

National Science Foundation
WHERE DISCOVERIES BEGIN

The 3rd NSF Secure and Trustworthy Cyberspace Principal Investigator Meeting
January 9-11, 2017
Arlington, Virginia