What Cryptography and TCS can do for Law

Thanks to K. Bamberger, R. Canetti, A. Cohen, F. Partnoy, Y. Shavit, M. Varia, P. Vasudevan, and D. Weitzner
My Journey into “Law For Algorithm” [Y]

• Law for Algorithms Fall 2018
  • Boston University, Harvard, Columbia, Berkeley
    (Canetti, Caruso, Dugan, Dwork, Minow, Williams, Goldwasser)

• Law for Algorithms Fall 2019
  • Boston University and Berkeley
    (Canetti, Dugan, Goldwasser, Partnoy)

• Danny Weitzner: Cohen, Frankle, Park, Shavit Fall 2017
Lessons Learned

Language Barrier

Knowledge Barrier

**Clear:** Need Interdisciplinary effort between law and computer science, to bridge the gap between legal standards and algorithms which stand to govern our life.
Time is Ripe

Law for Governing Algorithms

Algorithms for Implementing Law
The Power of Formal Definitions:
Mathematically (crypto style) define “The Right to be Forgotten: Discover How to be Compliant” [GGV2019, in submission]

Potential of using cryptographic protocols such as ZK and MPC for new and better legal procedures

AVAILABILITY OF DATA ABOUT INDIVIDUALS HAS ENORMOUS POTENTIAL TO BETTER OUR LIVES

- **Health:** disease control by trend prediction
- **Finance:** predictions for financial markets
- **Economic Growth:** intelligent consumer targeting
- **Infrastructure:** Traffic patterns and energy usage
- **Vision:** Facial and Image recognition
- **NLP:** Speech recognition, Machine Translation
- **Security:** Threat Prediction models, spam
- **Policing:** decide which neighborhood to police
- **Bail:** decide who is a flight risk
- **Credit Rating:** decide who gets a loan

Sudden Shift of Power
Privacy Regulations: Discussed, Codified, Made into Law

- Requires consent from the individual
- Wide definition of personal information including browser history, purchase behaviour, site/app interactions
- Allows for opt-out of collection/use
- Fines potentially in the millions of dollars
- Private right of action, class suits

- Has a legitimate interest component
- Defines PII and sensitive information
- Default to opt-in for collection/use
- Fines potentially in the millions of dollars
- Public complaints to an enforcement body to address

$10^6$ Q: AM I COMPLIANT?
Why “The Right to be Forgotten”? 

EU court backs 'right to be forgotten': Google must amend results on request

Individuals have right to control their data and can ask search engines to remove results, says European court

But How?

The test case privacy ruling by the European Union's court of justice against Google Spain was brought by a Spanish man, Mario Costeja González, after he failed to secure the deletion of an auction notice of his repossessed home dating from 1998 on the website of a mass circulation newspaper in Catalonia.

If You Want Life Insurance, Think Twice Before Getting A Genetic Test

As genetic testing expands, the health insurers are ramping up to deny coverage based on genetic factors.
To Determine Compliance Need to Formally Define “Data Deletion”

“The Right to be Forgotten: Discover How to be Compliant”
Joint work with Garg and Vasudevan
[in submission]

[NBWBGGOSV17] “Bridging the Gap between Computer Science and Legal Approaches to Privacy”, 2017

[CN19] “Towards Formalizing the GDPR Notion of Singling Out”, 2019
Data Privacy Laws

HIPAA, GDPR, CCPA, TITLE 13, FERPA...

Right to erasure (or to be forgotten)

"right to withdraw consent at any time. Right to object if their personal data is processed... by processors..."

Data Privacy Techniques

Encryption, Multi Party Computation, Obfuscation, Pseudonymization, K-anonymity, Differential Privacy...
1. **Extract** relevant text

4. **Solicit** legal opinions

---

Iterative Process

---

2. **Define** mathematically

3. **Analyze** taking into account other notions and attacks +

Design Mechanisms to Provably Achieve Definition
GDPR lists conditions when an entity may lawfully process personal data. “the data subject has given consent to the processing of his or her personal data for one or more specific purposes”.

[Article 6]

“The data subject shall have the right to withdraw his or her consent at any time”

[Article 7]

“The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay ... [under a number of conditions]”

[Article 17]
Defining Data Deletion: more than meets the eye

• Data Leaves May Traces

• Main Question: how does an honest data collector know it is doing its erasing as the law requires?

Objectives of a First Definition

• Generality making as few assumptions as possible about how the system used the data

• Simplicity

• Provides Protection As intended
Defining Data Deletion: Real-Ideal definition

Real World:

Data Collector X

Memory

Environment Z

Deletion Requester Y

\( \pi_1 \)

\( \pi_2 \)

\( del(\pi_1) \)

\( del(\pi_2) \)

Ideal World:

Data Collector X

Memory

Environment Z

Deletion Requester Y

\( Y_0 \)

Silent

(no communication)
Deletion Compliance

Real World:

\[ \text{Data Collector } X \rightarrow \text{Memory} \rightarrow \text{state}_{X}^{\text{Real}} \]

\[ \text{Environment} \]

\[ \text{Deletion Requester } Y \rightarrow \text{Memory} \rightarrow \text{state}_{X}^{\text{Real}} \]

\[ \text{Deletion} \]

\[ \text{view}_{Z}^{\text{Real}} \]

Ideal World:

\[ \text{Data Collector } X \rightarrow \text{Memory} \rightarrow \text{state}_{X}^{\text{Ideal}} \]

\[ \text{Environment} \]

\[ \text{Deletion Requester } Y \rightarrow \text{Memory} \rightarrow \text{state}_{X}^{\text{Ideal}} \]

\[ \text{Deletion} \]

\[ \text{view}_{Z}^{\text{Ideal}} \]

\[ X \text{ is deletion-compliant if for all } Z \text{ and } Y, \]

\[ (\text{state}_{X}^{\text{Real}}, \text{view}_{Z}^{\text{Real}}) \approx (\text{state}_{X}^{\text{Ideal}}, \text{view}_{Z}^{\text{Ideal}}) \]
Ex: Data Storage

• A data structure has
  • A “legitimate” interface: the set of operations allowed to be performed on it [Insert replace Delete]
  • A memory representation

• Techniques: History Oblivious data structures: randomized operations [Micciancio97, NaorTeague98]

• Definition: A data structure is history oblivious if any two sequences of operations $S_1$ and $S_2$ that yield the same content induce the same probability distribution on memory representation

• Theorem: Using history oblivious data structure + authentication can achieve deletion-compliance
Example: DP data release

- Sometimes, if output depends little "enough" on single inputs, don’t need to do anything else to delete.

Differential Privacy: A property of $M$ that limits how much the output $M(x)$ depends on any single $x$. 

![Diagram](image)
Example: ML Deletion

• Sometimes “history-independence” in other settings can be used to do deletion.

• Machine Learning Training Input Protection


• [VBE18] Algorithms that remember: Model inversion attacks and data protection law.
Remarks and Open Directions

• Our definition is very strong. Can a weaker definition provide a reasonable compromise at some settings (conditional compliance)?

• Which definition would be useful where?

• Composition of interacting compliant data collector subsystems?

• Better understanding of when it is required of a data collector to honor a given deletion request.
The Power of Formal Definitions: Mathematically (crypto style) define “The right to be Forgotten: Discover How to be Compliant” [GGV2019, in submission]

- Potential of using cryptographic Notions and Protocols for new and better legal procedures

1. Secure Multi Party Computation: enable detecting injustice (e.g. bias) without violating privacy

2. Zero Knowledge Proofs: enables accountability and compliance without learning the facts
Theorem: Distrustful parties can run arbitrary programs on the collective data without ever revealing individual data to each other [Under a variety of conditions]
Revealing the result, and nothing but the result

conditions: secret communication and/or good majority

Secure Multi-Party Computation
Compute safely on shared data

[Y86, GMW86, BGW87…]
Congressional definition

(12) SECURE MULTI-PARTY COMPUTATION.—The term “secure multi-party computation” means a computerized system that enables different participating entities in possession of private sets of data to link and aggregate their data sets for the exclusive purpose of performing a finite number of pre-approved computations without transferring or otherwise revealing any private data to each other or anyone else.

MPC can positively benefit Impact Policy and Law

Ex: Fusing Data Sets Data Sets for medical discovery and treatment
Detect and Address Discrimination without Finding Out Protected Attributes

Mulligan & Bamberger19: Enforcing “Privacy by Design” to reduce or eliminate the collection of data containing race and protected attributes, may make it impossible to determine “unequal treatment” via latent, redundant encoding of protected attributes and even normalize mass biases.

Proposal: Instead run MPC on collective data to gather information about “bias” and “fairness”, without learning protected attributes of individuals. Has the potentially relieve tension between Disparate Treatment vs. Disparate Outcome

Legal Q: Can you argue that an MPC based protocol in which you know there is a decision based on protected attribute but does not reveal protected attribute of individual does not constitute disparate treatment
The Carpenter Case

Following tip, law enforcement compelled two cell phone companies to provide continuous location information (12,898 location points) of cell-phones of several suspects (Carpenter included), over a period of 127 days.

- Carpenter moved to suppress evidence due to 4th Amendment violation. Motion denied by federal court.
- In a remarkable decision, the supreme court decided that the acquisition of data by law enforcement was in violation of Carpenter’s 4th amendment rights.
- **Proposal**: Use MPC to determine if there is sufficient cause to allow surveillance
Its Been Done!
Evaluate Wage (dis)parity by BU researchers

BOSTON closing the WAGE GAP
Becoming the Best City in America for Working Women

100% TALENT
The Boston Women’s Compact

SIMMONS COLLEGE
BOSTON, MASSACHUSETTS

STATE STREET

Raytheon

MassMutual

EMC2

SUFFOLK

STAPLES

Putnam

Suffolk

Eastern Bank

EVERSOURCE

nationalgrid

Associated Industries of Massachusetts
Goal: Benchmark Wage (dis)parity

Employers agree to contribute data to a report compiled by a third party on the Compact’s success to date. Employer-level data could not be identified in the report.
Machine Learning Governance

Ensure Algorithms for Bail, Credit, Policing are consistent with specs

verifying data analysis

Text from CA compass case

Blackbox v versus checking
Potential legal aspects of “secure computation”

Scenarios:
1. Private entities engage in a secure computation on the union of their proprietary data.
2. Same, but data is protected by regulation (HIPAA, FERPA, FCRA, SRTLBYGA…)
3. Government performing SC on collected [potentially personal] data
4. Same, but SC performed by private entities
5. Adjudication (administrative or in courts) regarding private information
6. Same, but with non-US entities (state and private)

Each scenario has a different set of concerns... let’s try to list.
Mathematical Proofs

\[ \sqrt{a^2+b^2} \]
Efficiently **Verifiable** Proofs

Prover Algorithm  \(\text{Claim}\)  Verifier Algorithm

Harder Job \(\text{proof}\)  Easier Job

(may know something which the verifier does not)

**Completeness:** if claim true, there exist a prover algorithm that will make the **Verifier algorithm** Accept

**Soundness:** if claim false, no prover algorithm can make the **Verifier algorithm** accept
Carl Friedrich Gauss
1777-1855

“The dignity of science itself seems to require that every possible means be explored for the solution of a problem so elegant.”

Factoring: find the prime factors of a number

<table>
<thead>
<tr>
<th>Number</th>
<th>Prime factors?</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>Trivial: 5 * 7</td>
</tr>
<tr>
<td>221</td>
<td>Quick: 13 * 17</td>
</tr>
<tr>
<td>562137</td>
<td>Long time for some humans</td>
</tr>
<tr>
<td>“1000 digit number”</td>
<td>1000 years fastest computer</td>
</tr>
</tbody>
</table>

Unless the computer is a quantum-computer
Claim: \( N=pq \) where \( p \) and \( q \) are primes

**Prover Algorithm**

Find \( p \) and \( q \)

**Verifier Algorithm**

Checks that \( p \) and \( q \) are primes and \( N=pq \)

*Efficient*

**Completeness** If \( N=pq \), there exist a prover who can make Verifer accept

**Soundness** If \( N\neq pq \), such prover doesn’t exist
Claim: $N$ is a product of 2 primes

After interaction, Verifier knows:
1) $N$ is product of 2 primes
2) $p$ and $q$
Zero Knowledge Proofs
[GMicaliRackoff85, GoldrMicaliWig87]

Proofs that reveal nothing but their validity

Main Idea:
Prove that I could prove it if I felt like it and therefore a proof exists
Why: Passwords = (Hard to find) Proofs of Claims

Verifier can be broken into, leaked from, have Malicious Insiders

P and Q s.t. \( N = PQ \)

Can Alice prove to Bob she knows P, Q without ever giving them to Bob
Other Proposed Uses

Passwords

Nuclear Disarmaments

DNA identity testing

DNA innocence proving

Validating Transactions on Blockchains preserving privacy and anonymity
“Public can verify votes of an electronic election were tallied correctly”

input = Encrypted(Vote). Keeping votes secret

“Public can verify tax return form is compliant with the rules”

without knowing which taxes paid to whom

“Can Verify Whether Documents are Privileged Attorney-Client”

without being exposed to the documents
Interactive and Probabilistically Verifiable Proofs [GMR85]

Completeness: if Claim is true, there exists a prover algorithm which will make the verifier accept with probability $>1$-negligible, regardless of prover

Soundness: if claim is false, verifier will reject with probability $>1$-negligible, regardless of prover

This is what a proof ultimately is!
What is Zero Knowledge?

For true claims, for every a verifier

what the verifier can compute after the interaction =
what the verifier could compute before interaction
More Generally

Can Decouple

“Verifying Correctness”

from

“Knowledge of the proof”
How Would Law Treat

Decoupling

“Verifying Claims”

from

“Knowledge of the facts”
Decisions for Mergers and Acquisitions

Using Zero Knowledge to verify “business facts” to:

• Mitigate the risks in disclosure of information for “due diligence” when one doesn’t know the motives of the enquirer

• Mitigate the costs of disclosure such as suing for expected damages after contract breaches, which might dissuade vindication of legal rights
Outline

The Power of Formal Definitions:
Mathematically (crypto style) define “The Right to be Forgotten: Discover How to be Compliant” [GGV2019, in submission]

Potential of using cryptographic protocols such as ZK and MPC for new and better legal procedures

Since 1979, [the FISA] court has processed over 28,000 warrant applications and renewals, a rate of nearly one thousand secret cases a year.

Today discuss the federal docket that handles tens of thousands of “electronic surveillance warrants” cases every year.
Disclosure of a surveillance order during the course of a criminal investigation would...be self-defeating.

Sealing a case – preventing court records from becoming accessible to the general public.

Gag order – preventing parties involved with the case from mentioning it to others.
Kudzu in the Courthouse, Judgments Made in the Shade (2009)

Gagged, Sealed, & Delivered: Reforming ECPA’s Secret Docket (2013)
Problems

2. Busy magistrate judges issue hundreds of...orders with literally no appellate guidance

3. The only affected party which might have incentive to object...is never given prior notice of the order.

4. Neither congress nor the public can accurately access the breadth and depth of current electronic surveillance activity.
A “cover sheet” containing meta data about sealed cases should always be made public

Obvious Tension: Privacy vs. Accountability
Propose a System:
Offers greater accountability and privacy is achievable when modern cryptography is brought in: MPC, zero-knowledge proofs, encryption, public ledger

• Demonstrate to the public that each participant performs its role properly and lawfully

• Make public Aware of the extent of Electronic Surveillance
Threat Model

• Judges are honest but forgetful
• Law enforcement agencies might be malicious.
• Companies might be malicious.
• The public might be malicious
My hope is that...court administrators will embrace the possibility of enhancing public oversight while preserving necessary secrecy.

Lessons learned here will... smooth the way towards greater accountability for a broader class of secret information processes, which are a hallmark of our digital age.
<table>
<thead>
<tr>
<th>Algorithmic Task</th>
<th>Societal Value</th>
<th>Math Formulation</th>
<th>Legal Doctrine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extract Utility from Data</td>
<td>Privacy</td>
<td>Differential Privacy Multi Party Computation</td>
<td>The right to Privacy</td>
</tr>
<tr>
<td>Verify Compliance and Accountability</td>
<td>Privacy</td>
<td>Zero Knowledge Proof</td>
<td>The right to Privacy</td>
</tr>
<tr>
<td>Make accurate Predictions and decisions</td>
<td>Fairness</td>
<td>Algorithmic Fairness- several definitions which cannot co-exist</td>
<td>Anti Discrimination</td>
</tr>
<tr>
<td>Drive a Vehicle</td>
<td>Safety</td>
<td>Formal Models proposed</td>
<td>Negligence, duty of Care</td>
</tr>
<tr>
<td>Moderate online content</td>
<td>Freedom of Speech</td>
<td>Attribute Based Encryption</td>
<td>Copy Right, Fair Use</td>
</tr>
</tbody>
</table>

Quest is **Not** Restricted to Achieving Privacy nor to using Cryptography

*From “Algorithms and Law” by Elkin-Koren, Feldman, Goldwasser, Talgram-Cohen*