Strategic Goals and Mechanisms of the GDPR and CCPA

Chris Hoofnagle

Beyond Differential Privacy

Simons Institute

May 6-10, 2019

Abstract

The General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) are the most consequential developments in information policy in a generation. In this talk, I will explain their strategic goals and the most interesting mechanisms these instruments use to create incentives and disincentives, to structurally strengthen some relationships while disadvantaging others, and to create privacy markets.

- My high level goal: fair balancing of these provisions
- 3 CCPA examples
- 5 GDPR examples – most are about risk

Information regulation dynamics

- Regulating information is different from ordinary products
  - Info is abstract
  - Unlike a normal product, information can be reshaped
  - One can't see information uses; violations can be hidden
  - We are solipsistic
- Privacy law is drifting toward prescription, high detail
  - Info companies go beyond econ motivations; ideologically motivated
  - Info industries lie by omission (they learned from Radio Shack)
    - Eric Schmidt's "hiding strategy"
    - Larry Page's opposition to Zeitgeist, search term billboard

The human rights tensions

- US law often regulates marketing behaviors, not privacy
- US law fundamentally treats privacy as an economic issue
- EU has elevated privacy to a *fundamental* human right
  - Holocaust, Stasi, Soviet Union
  - Privacy as group interest
  - Yet, GDPR could be seen as market-creating for privacy & risk-based
- Regulatory dynamics requires principles-level language which U.S. lawyers find wanting

California Consumer Privacy Act 1

- Businesses may pay consumers for collecting/selling personal data.
  - Ambiguous legislative language result of compromise
- Businesses may also offer different prices/level of service "if that price or difference is directly related to the value provided to the [business] by the consumer's data."
  - Notice is required
  - "A business shall not use financial incentive practices that are unjust, unreasonable, coercive, or usurious in nature."
    - ∴ CCPA imposes a ceiling
    - But what would floors look like? What if the value to the business is psychological (lock-in), or some other platform value?

CCPA 2

- Definition of personal data includes “probabilistic identifier:” identification of a consumer or a device to a degree of certainty of **more probable than not** based on any categories of personal information included in, or similar to, the categories enumerated in the definition of personal information.

Towards Formalizing the GDPR's Notion of Singling Out

Aloni Cohen* Kobbi Nissim†

April 15, 2019

Abstract

There is a significant conceptual gap between legal and mathematical thinking around data privacy. The effect is uncertainty as to which technical offerings adequately match expectations expressed in legal standards. The uncertainty is exacerbated by a litany of successful privacy attacks, demonstrating that traditional statistical disclosure limitation techniques often fall short of the sort of privacy envisioned by legal standards.

We define *predicate singling out*, a new type of privacy attack intended to capture the concept of singling out appearing in the General Data Protection Regulation (GDPR). Informally, an adversary predicate singles out a dataset $\mathbf{x}$ using the output of a data release mechanism $M(\mathbf{x})$ if it manages to find a predicate p matching exactly one row $x \in \mathbf{x}$ with probability much better than a statistical baseline. A data release mechanism that precludes such attacks is *secure against predicate singling out* (*PSO secure*).

We argue that PSO security is a mathematical concept with legal consequences. Any data release mechanism that purports to “render anonymous” personal data under the GDPR must be secure against singling out, and hence must be PSO secure. We then analyze PSO security, showing that it fails to self-compose. Namely, a combination of $\omega(\log n)$ exact counts, each individually PSO secure, enables an attacker to predicate single out. In fact, the composition of just two PSO-secure mechanisms can fail to provide PSO security.

Finally, we ask whether differential privacy and k-anonymity are PSO secure. Leveraging a connection to statistical generalization, we show that differential privacy implies PSO security. However, k-anonymity does not: there exists a simple and general predicate singling out attack under mild assumptions on the k-anonymizer and the data distribution.

CCPA 3

- Consumers can sue for security breaches if:
  - There is unauthorized access and exfiltration, theft, or disclosure
  - As a result of a failure to implement reasonable security procedures
  - The consumer notifies the business & give 30 days for response
  - The business fails to “cure” the violation

GDPR strategic goals 1

- GDPR's structure deters opportunism, guile
- Processing is illegal unless justified
  - "Legitimate interests"–when the interests of the controller outweigh the data subject
    - "necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such **interests are overridden** by the interests or fundamental rights and freedoms of the data subject which require protection of personal data"
      - Thus, there is an opportunity to object + a balancing test
    - To prevent surprise and opportunism, all legitimate interests must be disclosed
  - Consent (next slide)

GDPR strategic goals 2

- Deterring “consent” by burdening it with many requirements:
  - Freely given
  - Explicit
  - Informed
  - Specific (Omnibus consents presumptively invalid)
  - Unambiguous
  - Right to withdraw (and stop processing)
  - Burden on controller to prove validity
- CNIL (2019): €50 million fine for Google’s failure to obtain informed, specific, and unambiguous consent for ad personalization

GDPR strategic goals 3

- GDPR advantages first parties, is really tough on 3rd parties & unforeseen uses of data
- Human-in-the-loop data analysis
  - The GDPR is a pre-ML regulation and in fact it appears to ban it.
  - Companies like Palantir (I am on their board) are cleaning up in Europe
- New data uses must be “compatible”
  - What is the link between the old and new purpose?
  - What is the context in which the data were collected?
  - The nature of the data (sensitive data presumably more limited for reuse)
  - Possible consequences of the processing
  - Existence of safeguards

GDPR strategic goals 4

- Deterring “high risk” data uses
- Must do “Data Protection Impact Assessments” (DPIA)
  - If processing could infringe a person’s natural rights or freedoms, it may be “high risk”
    - E.g. activities that may cause discrimination, fraud, or financial loss
  - Controllers must complete a privacy impact assessment (PIA) that analyzes the need for and proportionality of the processing
  - If risk cannot be mitigated, must inform a supervisory authority

GDPR strategic goals 5

- Higher level of security, risk based
- Technological safeguards keyed to nature, scope, context, and purposes of the processing as well as the risks to the rights and freedoms of individuals.
- The GDPR conception of information security incorporates confidentiality, integrity, availability, as well as an interest in system resilience. Yet what measures are required is unclear, as the GDPR both signals a “state of the art” standard, but tempers the standard with a consideration of the costs involved.

GDPR security breach

- Incidents are “a breach of security leading to…[unauthorized] destruction, loss, alteration, disclosure of, access to, personal data”
- Must report to supervisory authority within 72 hours
- Must report to data subjects without undue delay, unless
  - Breach unlikely to result in a high risk for rights/freedoms of data subjects
  - Data are encrypted (or other technical measures protect the data)
  - Notice would require disproportionate efforts