Optimal Privacy-Constrained Mechanisms

Ran Eilat, Kfir Eliaz, and Xiaosheng Mu (2019)

Presented by Silvio Ravaioli

August 1, 2019

Motivation

- Unprecedented amount of personal data stored and traded
- Concern about *privacy* [tastes, willingness to pay for products]
- Should government increase the regulation about collection and use of personal information?
- The paper explores the trade-off between profits and privacy
- We need to define:
  - *Privacy*: buyer's type (price elasticity)
  - *Privacy loss*: Bayesian measure (expected relative entropy) that takes into account the designer's initial information

Motivation

- Unprecedented amount of personal data stored and traded
- Concern about *privacy* [tastes, willingness to pay for products]
- Should government increase the regulation about collection and use of personal information?

- The paper explores the trade-off between profits and privacy
- We need to define:
  - *Privacy*: buyer's type (price elasticity)
  - *Privacy loss*: Bayesian measure (expected relative entropy) that takes into account the designer's initial information

Motivation

- Mussa-Rosen (1978) set-up: population of buyers with heterogeneous price elasticity (type), monopolist seller with increasing costs for producing a high quality product, designs an optimal menu of quality-price pairs
- After the transaction, the seller knows exactly the buyer's type (no privacy)
- Suppose the regulator wants to limit the amount of information that can be learned
- Additional constraint: the regulator decides how much the seller can learn about the buyer's type
- Bayesian approach to quantify the *privacy loss*

Motivation

- Mussa-Rosen (1978) set-up: population of buyers with heterogeneous price elasticity (type), monopolist seller with increasing costs for producing a high quality product, designs an optimal menu of quality-price pairs
- After the transaction, the seller knows exactly the buyer's type (no privacy)
- Suppose the regulator wants to limit the amount of information that can be learned
- Additional constraint: the regulator decides how much the seller can learn about the buyer's type
- Bayesian approach to quantify the *privacy loss*

Bayesian measure of privacy loss

- The seller initially has *some* information about buyers' type distribution (prior belief)
- Privacy loss is defined as the quantity of *additional* information that is released by the buyer by participating in the mechanism
- The difference between the designer's prior and posterior beliefs is calculated using the Kullback-Leibler divergence (expected relative entropy)
- The privacy constraint requires the privacy loss to be at most κ

Bayesian measure of privacy loss

- The seller initially has *some* information about buyers' type distribution (prior belief)
- Privacy loss is defined as the quantity of *additional* information that is released by the buyer by participating in the mechanism
- The difference between the designer's prior and posterior beliefs is calculated using the Kullback-Leibler divergence (expected relative entropy)
- The privacy constraint requires the privacy loss to be at most κ

Bayesian measure of privacy loss

- Most of the theoretical work on privacy is based on "differential privacy" [starting with Dwork et al. 2006], a concept from cryptography
- Differentially private algorithms are used to publish statistical aggregates while ensuring confidentiality of survey responses. The constraint lies in the ability of identifying individuals whose information may be in a database
- Other papers [including Agrawal and Aggarwal 2001 and Wang et al. 2016] discuss privacy-distortion and anonymization of databases, with different notions of privacy (including relative entropy), but without considering strategic interaction between privacy, mechanism, and agent behavior

Bayesian measure of privacy loss

- Most of the theoretical work on privacy is based on "differential privacy" [starting with Dwork et al. 2006], a concept from cryptography
- Differentially private algorithms are used to publish statistical aggregates while ensuring confidentiality of survey responses. The constraint lies in the ability of identifying individuals whose information may be in a database
- Other papers [including Agrawal and Aggarwal 2001 and Wang et al. 2016] discuss privacy-distortion and anonymization of databases, with different notions of privacy (including relative entropy), but without considering strategic interaction between privacy, mechanism, and agent behavior

Model Overview

- Mussa-Rosen set-up of monopolistic screening
- The seller designs a static mechanism that maps messages to quantity/price pairs
- **Incentive-compatibility**: the buyer reveals own type (in a coarse way)
- **Individual-rationality**: buyers participate in the mechanism
- **Privacy constraint**: the expected relative entropy is bounded
- Only some mechanisms are κ-feasible
- The profit maximizing ones are called κ-optimal mechanisms

Research Questions

- What are the key properties of the constrained-optimal mechanism?
- What information does each buyer type disclose?
- Do some buyer types disclose more information than others?
- What is the maximal amount of information that is revealed by any buyer type?
- Is the privacy constraint even binding?

- Two definitions of privacy loss: ex-ante and ex-post
- Here we focus on the former (more in the paper)

Research Questions

- What are the key properties of the constrained-optimal mechanism?
- What information does each buyer type disclose?
- Do some buyer types disclose more information than others?
- What is the maximal amount of information that is revealed by any buyer type?
- Is the privacy constraint even binding?

- Two definitions of privacy loss: ex-ante and ex-post
- Here we focus on the former (more in the paper)

Model Setup

- A monopolistic seller offers a menu of price/quantities to a single buyer and obtains a profit

$$\pi(p, q) = p - c(q)$$

based on realized price, quantity, and cost. She wants to maximize the expected profit wrt buyer's WTP

- A buyer with WTP $\theta \in \Theta = [\underline{\theta}, \bar{\theta}]$ and utility

$$u(p, q, \theta) = q \cdot \theta - p$$

- Buyers' WTP θ distribution is F, with support Θ and density f. Virtual valuation $v(\theta)$ is increasing, strictly positive, and continuously differentiable

Model Setup

- A monopolistic seller offers a menu of price/quantities to a single buyer and obtains a profit

$$\pi(p, q) = p - c(q)$$

based on realized price, quantity, and cost. She wants to maximize the expected profit wrt buyer's WTP

- A buyer with WTP $\theta \in \Theta = [\underline{\theta}, \bar{\theta}]$ and utility

$$u(p, q, \theta) = q \cdot \theta - p$$

- Buyers' WTP θ distribution is F, with support Θ and density f. Virtual valuation $v(\theta)$ is increasing, strictly positive, and continuously differentiable

Mechanism

- The seller designs a static mechanism $\mathbb{M} = \langle M, p, q \rangle$
- Set of messages M
- Quantity function $q : M \to \mathbb{R}^+$
- Price function $p : M \to \mathbb{R}^+$
- The buyer adopts a strategy $\sigma : \Theta \to \Delta M$
- Without privacy constraint, the optimal (revenue maximizing) mechanism is a direct revelation mechanism such that
  - buyers truthfully report own type $m = \theta$
  - quantities satisfy $v(\theta) = c'(q(\theta))$
  - prices satisfy $p(\theta) = q(\theta)\theta - \int_{\underline{\theta}}^{\theta} q(x)dx$

Mechanism

- The seller designs a static mechanism $\mathbb{M} = \langle M, p, q \rangle$
- Set of messages M
- Quantity function $q : M \to \mathbb{R}^+$
- Price function $p : M \to \mathbb{R}^+$
- The buyer adopts a strategy $\sigma : \Theta \to \Delta M$
- Without privacy constraint, the optimal (revenue maximizing) mechanism is a direct revelation mechanism such that
  - buyers truthfully report own type $m = \theta$
  - quantities satisfy $v(\theta) = c'(q(\theta))$
  - prices satisfy $p(\theta) = q(\theta)\theta - \int_{\underline{\theta}}^{\theta} q(x)dx$

Bayesian Privacy

- The seller knows the buyer's type distribution F (prior)
- After observing the message $m \in M$, the posterior distribution is $F(\cdot|m)$
- The relative entropy (Kullback-Leibler Divergence) from $F(\cdot|m)$ to F is

$$D_{KL}(F(\cdot|m) \mid\mid F) = \int_{\underline{\theta}}^{\bar{\theta}} f(\theta|m) \cdot log \frac{f(\theta|m)}{f(\theta)} d\theta$$

- If $F(\cdot|m)$ contains atoms, $D_{KL}(F(\cdot|m) \mid\mid F) = +\infty$

Ex-ante and Ex-post Privacy loss

- The ex-ante loss of privacy entailed by $\mathbb{M} = \langle M, p, q \rangle$ is

$$I(\mathbb{M}) = \mathbb{E}_m[D_{KL}(F(\cdot|m) \,||\, F)]$$

where $\mathbb{E}_m$ is evaluated according to the message probabilities in equilibrium

- The ex-post loss of privacy entailed by $\mathbb{M} = \langle M, p, q \rangle$ is

$$I^{ep}(\mathbb{M}) = sup_m[D_{KL}(F(\cdot|m) \,||\, F)]$$

where the supremum is taken over messages that are sent with positive probability in equilibrium

Ex-ante and Ex-post Privacy loss

- The ex-ante loss of privacy entailed by $\mathbb{M} = \langle M, p, q \rangle$ is

$$I(\mathbb{M}) = \mathbb{E}_m[D_{KL}(F(\cdot|m) \; || \; F)]$$

  where $\mathbb{E}_m$ is evaluated according to the message probabilities in equilibrium

- The ex-post loss of privacy entailed by $\mathbb{M} = \langle M, p, q \rangle$ is

$$I^{ep}(\mathbb{M}) = sup_m[D_{KL}(F(\cdot|m) \; || \; F)]$$

  where the supremum is taken over messages that are sent with positive probability in equilibrium

Feasible and Optimal mechanisms

- The seller wants to design a mechanism $\mathbb{M} = \langle M, p, q \rangle$ and a strategy for the buyer σ such that she maximizes the expected profit subject to the three constraints

1. Incentive-compatibility: for all θ, $m \in supp(\sigma(\theta))$, $m' \in M$

$$u(p(m), q(m), \theta) \geq u(p(m'), q(m'), \theta)$$

2. Individual-rationality: for all θ and $m \in supp(\sigma(\theta))$

$$u(p(m)), q(m), \theta) \geq 0$$

3. Privacy constraint

$$I((M)) \leq \kappa$$

- κ-feasible mechanisms satisfy all the constraints
- κ-optimal m. maximize profits among all the κ-feasible m.

Feasible and Optimal mechanisms

- The seller wants to design a mechanism $\mathbb{M} = \langle M, p, q \rangle$ and a strategy for the buyer σ such that she maximizes the expected profit subject to the three constraints

1. Incentive-compatibility: for all θ, $m \in supp(\sigma(\theta))$, $m' \in M$

$$u(p(m), q(m), \theta) \geq u(p(m'), q(m'), \theta)$$

2. Individual-rationality: for all θ and $m \in supp(\sigma(\theta))$

$$u(p(m)), q(m), \theta) \geq 0$$

3. Privacy constraint

$$I((M)) \leq \kappa$$

- κ-feasible mechanisms satisfy all the constraints
- κ-optimal m. maximize profits among all the κ-feasible m.

Feasible and Optimal mechanisms

- The seller wants to design a mechanism $\mathbb{M} = \langle M, p, q \rangle$ and a strategy for the buyer σ such that she maximizes the expected profit subject to the three constraints

1. Incentive-compatibility: for all θ, $m \in supp(\sigma(\theta))$, $m' \in M$
$$u(p(m), q(m), \theta) \geq u(p(m'), q(m'), \theta)$$
2. Individual-rationality: for all θ and $m \in supp(\sigma(\theta))$
$$u(p(m)), q(m), \theta) \geq 0$$
3. Privacy constraint
$$I((M)) \leq \kappa$$

- κ-feasible mechanisms satisfy all the constraints
- κ-optimal m. maximize profits among all the κ-feasible m.

Coarse revelation principle

- Because of the privacy constraint, the seller obtains a *noisy* signal about the buyer's type
- *Coarse* revelation principle - focus on interval mechanisms
- **Lemma**. For any κ-feasible mechanism, there exists another κ-feasible mechanism with the same profit level, such that M consists of intervals that partition Θ and each type reports the message for which $\theta \in m$
- **Intuition of the proof**
- Mechanisms that rely on mixed strategies are "wasteful"
- We can transform $\mathbb{M}$ into an interval mechanism with the same expected profit and weakly lower privacy loss
- Remove duplicate messages s.t. $p(m) = p(m')$, $q(m) = q(m')$
- Single crossing property: convex sets of pooled types
- Define $\mu(m) = \{\theta \in \Theta | m \in supp(\sigma(\theta))\}$: interval or singleton

Coarse revelation principle

- Because of the privacy constraint, the seller obtains a *noisy* signal about the buyer's type
- *Coarse* revelation principle - focus on interval mechanisms
- **Lemma**. For any κ-feasible mechanism, there exists another κ-feasible mechanism with the same profit level, such that M consists of intervals that partition Θ and each type reports the message for which $\theta \in m$
- **Intuition of the proof**
- Mechanisms that rely on mixed strategies are "wasteful"
- We can transform $\mathbb{M}$ into an interval mechanism with the same expected profit and weakly lower privacy loss
- Remove duplicate messages s.t. $p(m) = p(m')$, $q(m) = q(m')$
- Single crossing property: convex sets of pooled types
- Define $\mu(m) = \{\theta \in \Theta | m \in supp(\sigma(\theta))\}$: interval or singleton

Coarse revelation principle

- Because of the privacy constraint, the seller obtains a *noisy* signal about the buyer's type
- *Coarse* revelation principle - focus on interval mechanisms
- **Lemma**. For any κ-feasible mechanism, there exists another κ-feasible mechanism with the same profit level, such that M consists of intervals that partition Θ and each type reports the message for which $\theta \in m$
- **Intuition of the proof**
- Mechanisms that rely on mixed strategies are "wasteful"
- We can transform $\mathbb{M}$ into an interval mechanism with the same expected profit and weakly lower privacy loss
- Remove duplicate messages s.t. $p(m) = p(m')$, $q(m) = q(m')$
- Single crossing property: convex sets of pooled types
- Define $\mu(m) = \{\theta \in \Theta | m \in supp(\sigma(\theta))\}$: interval or singleton

Interval Mechanisms

- Interval mechanisms: M consists of intervals that partition $\Theta = [\underline{\theta}, \bar{\theta}]$
- We have a discrete distribution g_M over messages induced by the prior as $g_M(m) = F(\bar{m}) - F(\underline{m})$, the ex-ante privacy loss is

$$I(\mathbb{M}) = H(g_M) = -\sum_m [F(\bar{m}) - F(\underline{m})] \cdot log\,[F(\bar{m}) - F(\underline{m})]$$

- **Lemma 2**. The profit maximization problem is equivalent to finding a set of intervals that partition Θ, satisfy the privacy constraint, and such that expected profit is maximized. Quantity-price pairs are determined by the interval partition

Interval Mechanisms

- Interval mechanisms: M consists of intervals that partition $\Theta = [\underline{\theta}, \bar{\theta}]$
- We have a discrete distribution g_M over messages induced by the prior as $g_M(m) = F(\bar{m}) - F(\underline{m})$, the ex-ante privacy loss is

$$I(\mathbb{M}) = H(g_M) = -\sum_m [F(\bar{m}) - F(\underline{m})] \cdot log\,[F(\bar{m}) - F(\underline{m})]$$

- **Lemma 2**. The profit maximization problem is equivalent to finding a set of intervals that partition Θ, satisfy the privacy constraint, and such that expected profit is maximized. Quantity-price pairs are determined by the interval partition

Existence

- **Proposition**. There exists a κ-optimal mechanisms $\mathbb{M} = \langle M, p, q \rangle$ such that M consists of finitely many intervals that partition Θ, and each type $\theta \in \Theta$ reports the interval to which it belongs.
- **Intuition for the proof**
- Sequence of κ-feasible interval mechanisms $\mathbb{M}_j$ such as $\pi(\mathbb{M}_j)$ converges to π^*
- Suppose we can replace each $\mathbb{M}_j$ with $\tilde{\mathbb{M}}_j$, with a the new $\tilde{M}_j$ with at most N intervals (based on F and κ), then the limit partition $\tilde{M}_\infty$ would be optimal
- The replacement $\tilde{\mathbb{M}}_j$ is generated by merging two adjacent intervals, and dividing another interval. The profit is higher without violating the privacy constraint

Existence

- **Proposition**. There exists a κ-optimal mechanisms $\mathbb{M} = \langle M, p, q \rangle$ such that M consists of finitely many intervals that partition Θ, and each type $\theta \in \Theta$ reports the interval to which it belongs.

- **Intuition for the proof**

- Sequence of κ-feasible interval mechanisms $\mathbb{M}_j$ such as $\pi(\mathbb{M}_j)$ converges to π^*

- Suppose we can replace each $\mathbb{M}_j$ with $\tilde{\mathbb{M}}_j$, with a the new $\tilde{M}_j$ with at most N intervals (based on F and κ), then the limit partition $\tilde{M}_\infty$ would be optimal

- The replacement $\tilde{\mathbb{M}}_j$ is generated by merging two adjacent intervals, and dividing another interval. The profit is higher without violating the privacy constraint

Further Properties

- **Proposition 2**. Under the ex-ante privacy measure, the privacy constraint is exhausted in any κ-optimal mechanism.
- **Proposition 3**. There exists $\underline{\kappa} > 0$ such that in any κ-optimal interval mechanism with $0 < \kappa \leq \underline{\kappa}$, the message set M consists of exactly two intervals
- **Proposition 4**. Suppose $c(q)$ has non-negative third derivative, and $v(\theta)$ is strictly less convex than $F(\theta)$. Then any κ-optimal mechanism consists of intervals that are ordered in increasing mass from left to right.
  Symmetrically, the intervals in the optimal mechanism would be ordered in decreasing mass if $c''' \leq 0$ and $v(\theta)$ were strictly more convex than $F(\theta)$

Uniform-Quadratic Case

- $F(\theta)$ is uniform $U[\underline{\theta}, \bar{\theta}]$, $c(q)$ is quadratic $c(q) = \frac{q^2}{2}$
- $v(\theta)$ is linear - therefore as convex as $F(\theta)$, $c''' = 0$
- From Proposition 4: the ordering of intervals does not matter, focus on the lengths
- **Lemma**. In the uniform-quadratic-case, given any $n \geq 1$ and $\kappa > 0$, the $(n - \kappa)$-optimal mechanism is such that:
  1. if $log\ n \leq \kappa$ then M consists of n intervals of equal length
  2. if $log\ n > \kappa$ then exactly one interval has length l_s, and the remaining $n - 1$ intervals have length $l_b > l_s$, with the interval lengths determined by the privacy constraint
- Intuition: 1) intervals' order does not matter, 2) FOC: intervals can have at most two lengths, 3) SOC: $n - 1$ intervals have the same lengths, the last one is weakly shorter

Uniform-Quadratic Case

- $F(\theta)$ is uniform $U[\underline{\theta}, \bar{\theta}]$, $c(q)$ is quadratic $c(q) = \frac{q^2}{2}$
- $v(\theta)$ is linear - therefore as convex as $F(\theta)$, $c''' = 0$
- From Proposition 4: the ordering of intervals does not matter, focus on the lengths
- **Lemma**. In the uniform-quadratic-case, given any $n \geq 1$ and $\kappa > 0$, the $(n - \kappa)$-optimal mechanism is such that:
  1. if $log\ n \leq \kappa$ then M consists of n intervals of equal length
  2. if $log\ n > \kappa$ then exactly one interval has length l_s, and the remaining $n - 1$ intervals have length $l_b > l_s$, with the interval lengths determined by the privacy constraint
- Intuition: 1) intervals' order does not matter, 2) FOC: intervals can have at most two lengths, 3) SOC: $n - 1$ intervals have the same lengths, the last one is weakly shorter

Uniform-Quadratic Case: privacy-profit trade-off

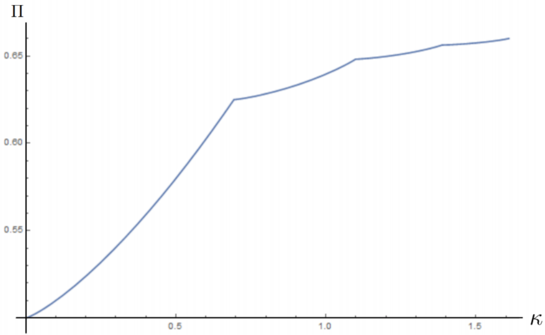

Figure 1. The privacy-profit frontier in the uniform-quadratic case

- Expected profit as a function of κ
- Kinks indicate that n increases
- Diminishing returns when n increases
- Increasing returns when κ increases (and n does not)

Optimal Privacy-Constrained Mechanisms

Ran Eilat, Kfir Eliaz, and Xiaosheng Mu (2019)

Presented by Silvio Ravaioli

August 1, 2019