Privacy
Experiments, History, Definitions, Metrics, Design Paradigms, and the Value of Information

Roy Dong
University of California, Berkeley
What’s at stake with privacy?
Helsinki privacy experiment

Figure 2. The Wi-Fi camera and an example positioning. The plan was redrawn from a subject’s original.
Figure 5. Subjects’ graphical depiction of how experience of surveillance changed over 0–6 months. Higher = worse, lower = better (* = interrupted at six months).
How do we define privacy?
The Right to Privacy (1890)

• *The Right to Privacy* (1890) by Warren and Brandeis.
  – **Conclusion**: Privacy is the right to be “left alone”.

![Portraits of Warren and Brandeis](image-url)
The first privacy tort
More privacy torts

  - Appropriating the plaintiff's identity for the defendant's benefit.
  - Placing the plaintiff in a false light in the public eye.
  - Publicly disclosing private facts about the plaintiff.
  - Unreasonably intruding upon the seclusion or solitude of the plaintiff.
Privacy as torts

  - **Conclusion:** Privacy was captured by 4 torts.
    - Appropriation
    - Slander & Libel
    - Public Disclosure
    - Intrusion upon Seclusion
Recent conceptions of privacy

- Privacy as Contextual Integrity (2004) by Helen Nissenbaum.
  - Spheres of life and norms of information flow.
Privacy in IoT

- Here’s a family on vacation June 9, 2012 to June 18, 2012:

[Graph showing power consumption with dates and power values]

[Makonin, Ellert, Bajić, Popowich 2016]
Privacy in IoT

• Here’s the same family watching TV on November 15, 2012:

[Image of a graph showing power usage over time with peaks at specific times.]
How can we quantify privacy?
Quantifying privacy

- Behavioral methods
- Non-statistical methods
- Differential privacy
- Inferential privacy
Behavioral methods

- **User** studies of privacy **decisions** and **perceptions** of privacy.

- *Example*: paradox of control and privacy. [Brandimarte, Acquisti, Loewenstein 2010]
### Study on Ethical Behavior

**IMPORTANT:** All answers are voluntary. By answering a question, you agree to give the researchers permission to publish your answer.

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Are you married?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Have you ever been fired by your employer?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Have you ever stolen anything (e.g.: from a shop, a person)?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Have you ever used drugs of any kind (e.g.: weed, heroin, crack)?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Have you ever lied about your age?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Have you ever had cosmetic surgery?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Have you ever done any kind of voluntary service?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Have you ever had sex in a public venue (e.g.: restroom of a club, airplane)?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Have you ever made a donation to a non-profit organization?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Do you have any permanent tattoos?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Brandimarte, Acquisti, Loewenstein 2010]
## Study on Ethical Behavior

**IMPORTANT:** All answers are voluntary. In order to give the researchers permission to publish your answers to the questions, please check the box below.

- [ ] Publication permission

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Are you married?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Have you ever been fired by your employer?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Have you ever stolen anything (e.g.: from a shop, a person)?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Have you ever used drugs of any kind (e.g.: weed, heroin, crack)?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Have you ever lied about your age?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Have you ever had cosmetic surgery?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Have you ever done any kind of voluntary service?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Have you ever had sex in a public venue (e.g.: restroom of a club, airplane)?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Have you ever made a donation to a non-profit organization?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Do you have any permanent tattoos?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Brandimarte, Acquisti, Loewenstein 2010]
### Study on Ethical Behavior

**IMPORTANT:** All answers are voluntary. In order to give the researchers permission to publish your answer to a question, please check the corresponding box.

<table>
<thead>
<tr>
<th>Question</th>
<th>Publication permission</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Are you married?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Have you ever been fired by your employer?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Have you ever stolen anything (e.g.: from a shop, a person)?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Have you ever used drugs of any kind (e.g.: weed, heroin, crack)?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Have you ever lied about your age?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Have you ever had cosmetic surgery?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Have you ever done any kind of voluntary service?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Have you ever had sex in a public venue (e.g.: restroom of a club, airplane)?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Have you ever made a donation to a non-profit organization?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Do you have any permanent tatoos?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Brandimarte, Acquisti, Loewenstein 2010]
Non-statistical methods

• Quantify privacy without explicit *randomness* or *probability* arguments.

• *Example*: k-anonymity. [Sweeney 1998]
Differential privacy

- Bound how much the observables change due to a small perturbation in private parameters.

- Let $\theta$ and $\theta'$ be two private parameters we wish to keep indistinguishable.

- Let $Y(\theta)$ and $Y(\theta')$ be the observables, respectively.

[Dwork 2006]
Differential privacy

- The observables are $\epsilon$ differentially private if, for any adjacent $\theta$ and $\theta'$, and for any measurable set $A$:

  $$\Pr(Y(\theta) \in A) \leq \exp(\epsilon) \Pr(Y(\theta') \in A)$$

- Oftentimes called the gold standard of privacy.
  - No adversary model required!
  - In practice, noise is often added so models of system uncertainty are not needed.
  - In practice, usually requires many users in the database to ‘hide’ amongst.

[Dwork 2006]
Inferential privacy

- Quantify the expected **loss** due to an adversary’s inference.
Setup

- Let $\theta$ be a random element in $\Theta$.
  - Interpret this as the *private parameter*.

- Let $Y$ be a random element in $\mathcal{Y}$.
  - Interpret this as the *public observable*.

- An adversary observes $Y$ and picks a probability distribution $q \in \Delta(\Theta)$.
  - Interpret this as the *posterior distribution* of $\theta$. 

[du pin Calmon and Fawaz 2012] [Jia, Dong, Spanos, Sastry]
Setup

- If the realized value of $\theta$ is $\phi \in \Theta$, and the adversary picks $q \in \Delta(\Theta)$, then the user experiences loss $-\ell(\phi, q)$.

- Without observing $Y$, the worst-case adversary attack is:
  \[ q^*_0 = \arg\min_q \mathbb{E}[\ell(\theta, q)] \]

- After observing $Y = y$, the worst-case adversary attack is:
  \[ q^*(y) = \arg\min_q \mathbb{E}[\ell(\theta, q) | Y = y] \]

[du pin Calmon and Fawaz 2012] [Jia, Dong, Spanos, Sastry]
Setup

• The inferential privacy loss of $\theta$ from $Y$ is:

$$\mathbb{E}[\ell(\theta, q_0^*) - \ell(\theta, q^*(Y))]$$

• This is the difference between the inference cost of the prior and the posterior.

• This is a general framework for modeling inference attacks in the literature:
  
  • $\ell(\theta, q) = -\log q(\theta)$ implies the inferential privacy loss is the mutual information between $\theta$ and $Y$.

  • $\ell(\theta, q) = \begin{cases} 1 \text{ when } \theta = \arg\max_{\phi\in\Theta} q(\phi) \\ 0 \text{ otherwise} \end{cases}$ implies the inferential privacy loss is determined by the probability of error in the MAP estimate.

[du pin Calmon and Fawaz 2012] [Jia, Dong, Spanos, Sastry]
Quantifying privacy

• What do we mean by privacy?
  – Behavioral methods
    • How users feel/respond?
  – Non-statistical methods
    • “Due diligence”
  – Differential privacy
    • How culpable am I for a privacy breach?
  – Inferential privacy
    • How vulnerable are my users?
How can we design systems in a privacy-aware fashion?
Design paradigms for privacy

• **Passive** privacy analysis
  – For a fixed system, *quantify* the privacy risk of users.

• **Active** privacy mechanisms
  – Fix a *parameterized* privacy-preserving scheme.
  – Pick the privacy parameter to best *trade-off* the utility of the collected data with the privacy of users.

• **Optimal** privacy-by-design
  – Fix *performance* metrics and *privacy* metrics.
  – Design a privacy-preserving mechanism that *maximizes* privacy, subject to performance constraints.
Direct Load Control

[Ruiz, Cobelo, and Oyarzabal 2009] [Moura, Bendtsen, and Ruiz 2013] [Mathieu, Koch, and Callaway 2013]
Privacy-Aware Direct Load Control

• Define a **controller** for different privacy settings.
• Model **private** variables and the **effect** of private variables on user **actions**.
• Analyze **closed loop** performance and how much is **revealed** by public observations.

[Dong, Ratliff, Cárdenas, Ohlsson, Sastry, “Quantifying the Utility-Privacy Tradeoff in the Internet of Things.”]
Privacy-Aware Direct Load Control

• Model \textit{private} variables and the \textit{effect} of private variables on user \textit{actions}.

[US Energy Information Administration 2009]

[Dong, Ratliff, Cárdenas, Ohlsson, Sastry, “Quantifying the Utility-Privacy Tradeoff in the Internet of Things.”]
Privacy-Aware Direct Load Control

- Analyze closed loop performance and how much is revealed by public observations.
Privacy-Aware Direct Load Control

- Quantify the utility-privacy tradeoff in data collection.

![Utility-privacy tradeoff in DLC of TCLs](image)
HVAC Control

• Heating, ventilation, and air conditioning control for buildings can be improved by incorporating occupancy estimates for different zones.
HVAC Control

HVAC Controller

Control signal

Occupancy Sensor

Building Manager

Occupancy

Zone: Z_1, Z_2, ..., Z_N
Time: 1, 2, 3, K

Occupancy:

Z_1: 0 → 1 → 1 → 0
Z_2: 0 → 1 → 2 → 1
Z_N: 1 → 0 → 1 → 1

Adversary

Auxiliary Knowledge

Office Directory
- Z_1: Alice
- Z_2: Chris
- Z_3: Bob
- Z_N: Pantry

Previous Location Traces

Privacy-Sensitive Information

Location Traces

Alice: Z_1 → Z_2 → Z_1
Bob: Z_2 → Z_5 → Z_3 → Z_3
Chris: Z_2 → Z_2 → Z_2

Time: 1, 2, 3, K
Privacy Metric

• We can use the mutual information as a privacy metric:

\[ I(\theta; y) = H(\theta) - H(\theta|y) \]

\[ H(\theta) = -\sum p(\theta) \ln(p(\theta)) \] is the entropy of \( \theta \).

\[ H(\theta|y) = -\sum p(y)\left[\sum p(\theta|y) \ln(p(\theta|y))\right] \] is the entropy of \( \theta \) conditioned on an observation of \( y \).

[Calmon and Fawaz 2012] [Sankar, Kar, Tandon, and Poor 2011]
Problem Setup

- The location traces of individual $m$: 
  \[ X_1^{(m)}, \ldots, X_k^{(m)} \]
- The **true** occupancy of zone $n$: 
  \[ Y_1^n, \ldots, Y_k^n \]
- How do we **modify observations** to balance good HVAC control performance with the privacy of individual’s locations?

[Jia, Dong, Sastry, Spanos, “Privacy-Enhanced Architecture for Occupancy-based HVAC Control.”]
Problem Setup

- **Distort** the occupancy with probability $p_{Z|Y}$:
  \[ Z_n^1, \ldots, Z_n^k \]

Minimizing privacy loss subject to a control performance constraint becomes a **channel design** problem!

\[
\min_{P_{Z|Y}} I \left( X_{1:M}^{1:K}, Z_{1:K}^{1:N} \right)
\]

subject to a performance constraint

[Jia, Dong, Sastry, Spanos, “Privacy-Enhanced Architecture for Occupancy-based HVAC Control.”]
Evaluating Control Performance

• For each true occupancy and reported occupancy, we evaluated the controller’s performance:

\[ J(Y_{1:K}^1, Z_{1:K}^1) \]

• Can approximate this via simulation of building dynamics and a model predictive controller (MPC), or historical data.

[Jia, Dong, Sastry, Spanos, “Privacy-Enhanced Architecture for Occupancy-based HVAC Control.”]
Optimal Privacy-Preserving Mechanisms

\[
\min_{P_{Z|Y}} I \left( X^{(1:M)}_1 : K, Z^{1:N}_1 : K \right) \\
\text{subject to } \mathbb{E} \left( J \left( Y^{1:N}_1 : K, Z^{1:N}_1 : K \right) \right) \leq \Delta
\]

- The expectation is taken across \( P_{Z|Y} \).
- This is a linear program and can easily be solved!

We are optimally designing our noising scheme to minimize privacy loss subject to a performance constraint!

[Jia, Dong, Sastry, Spanos, “Privacy-Enhanced Architecture for Occupancy-based HVAC Control.”]
Results

Visualization of Distortion Matrix

Temperature Evolutions for Different Distortion Levels

[Jia, Dong, Sastry, Spanos, “Privacy-Enhanced Architecture for Occupancy-based HVAC Control.”]
Results

Utility-Privacy Trade-off for Different Schemes

- Optimal Distortion
- Fixed Schedule
- Unperturbed Occupancy Data
- Uniform Distortion
- Multinomial Distortion

[Jia, Dong, Sastry, Spanos, “Privacy-Enhanced Architecture for Occupancy-based HVAC Control.”]
How do people decide how to share their data?
Privacy Contracts

• Thus far:
  – We can quantify how much the IoT company values data.
  – We can quantify how much private information the data contains.

• To go:
  – Can we account for the variability in privacy preferences among consumers?
  – Can we model how the market will behave when privacy-differentiated goods appear?

[Ratliff, Barreto, Dong, Ohlsson, Cárdenas, Sastry, “Effects of Risk on Privacy Contracts for Demand-Side Management.”]
Privacy Contracts

• Modeling privacy preferences:
  – The consumer’s type is \( \xi \in \{ \xi_1, \xi_2, ..., \xi_n \} \).
  - Here, type \( \xi_n \) values privacy the most and \( \xi_1 \) values it the least, with \( \xi_{i+1} \geq \xi_i \) for \( i = 1, 2, ..., n - 1 \).

• Modeling the privacy setting of the system:
  – The quality of the good is \( x \in \mathbb{R} \).
    - For each consumer type, we will offer a different quality of good: for type \( \xi_i \), assign quality \( x_i \).

• Modeling the market:
  – The IoT company wishes to design a menu of contracts \( (t_1, x_1), (t_2, x_2), ..., (t_n, x_n) \).
    - The good of quality \( x_i \) costs the consumer price \( t_i \), and is intended for consumer type \( \xi_i \).

[Ratliff, Barreto, Dong, Ohlsson, Cárdenas, Sastry, “Effects of Risk on Privacy Contracts for Demand-Side Management.”]
Privacy Contracts

• For the IoT company:
  – The cost of operations will depend on the quality: $g(x)$.
  – Revenue is generated by charging price $t$.
  – The IoT company wishes to maximize its utility:
    $$t - g(x)$$

• For the consumer:
  – The consumer’s utility function:
    $$U(x, \xi) - t$$
  – If the deal is unfair, the consumer will opt out.
    • Individual rationality constraint:
      $$U(x, \xi) - t \geq 0$$

[Ratliff, Barreto, Dong, Ohlsson, Cárdenas, Sastry, “Effects of Risk on Privacy Contracts for Demand-Side Management.”]
Privacy Contracts

• If the IoT company knew the consumer’s type was $\xi$, it would offer:
  \[
  \max\{t - g(x) : U(x, \xi) - t \geq 0\}
  \]

• This is known as the first best solution.

• However, the consumer may lie about her type.
  – Asymmetric information leads to adverse selection.

[Ratliff, Barreto, Dong, Ohlsson, Cárdenas, Sastry, “Effects of Risk on Privacy Contracts for Demand-Side Management.”]
Privacy Contracts

• To handle adverse selection, we must design a menu of contracts \((t_i, x_i)\) such that they are incentive compatible:

\[
U(x_i, \xi_i) - t_i \geq U(x_j, \xi_i) - t_j \quad \forall \ i, j
\]

This introduces the screening problem:

maximize: \(\sum_i P(\xi = \xi_i)[t_i - g(x_i)]\)

subject to: \(U(x_i, \xi_i) - t_i \geq 0 \quad \forall \ i\)

\[
U(x_i, \xi_i) - t_i \geq U(x_j, \xi_i) - t_j \quad \forall \ i, j
\]

[Ratliff, Barreto, Dong, Ohlsson, Cárdenas, Sastry, “Effects of Risk on Privacy Contracts for Demand-Side Management.”]
Privacy Contracts

• Finding the **first best solution**:

\[
\max \{ t - g(x) : U(x, \xi) - t \geq 0 \} \\
= \max_x U(x, \xi) - g(x)
\]

• The first best solution is the "**social optimum**".
  – Social benefit:

\[
U(x, \xi) - t + t - g(x) = U(x, \xi) - g(x)
\]

[Ratliff, Barreto, Dong, Ohlsson, Cárdenas, Sastry, “Effects of Risk on Privacy Contracts for Demand-Side Management.”]
Privacy Contracts

• Finding the second best solution:
• Assume Spence-Mirrlees single-crossing condition:
  \[ U(\cdot, \xi_{i+1}) - U(\cdot, \xi_i) \text{ is increasing for each } i. \]
• Then:
  \[
  \text{maximize: } \sum_i P(\xi = \xi_i)[t_i - g(x_i)] \\
  \text{subject to: } U(x_i, \xi_i) - t_i \geq 0 \quad \forall i \\
  \quad U(x_i, \xi_i) - t_i \geq U(x_j, \xi_i) - t_j \quad \forall i, j \\
  = \text{maximize: } \sum_i P(\xi = \xi_i)[t_i - g(x_i)] \\
  \text{subject to: } U(x_1, \xi_1) - t_1 = 0 \\
  \quad U(x_i, \xi_i) - t_i = U(x_{i-1}, \xi_i) - t_{i-1} \quad \forall i \\
  \quad x_i \geq x_{i-1} \quad \forall i
  \]
Privacy Contracts

• So what changes from the first best solution to the second best solution?

– The quality **does not change** for the highest type \( \xi_n \).

\[
x_{n}^{fb} = x_{n}^{sb}
\]

– Other types receive **lower** quality.

\[
x_{i}^{fb} \geq x_{i}^{sb}
\]

– The lowest type \( \xi_1 \) receives **zero** surplus.

\[
U(x_{1}^{sb}, \xi_1) - t_{1}^{sb} = 0
\]

– The other types receive **information rent**.

• Thus, the other types **free ride**.

\[
U(x_{i}^{sb}, \xi_i) - t_{i}^{sb} \geq 0
\]

[Ratliff, Barreto, **Dong**, Ohlsson, Cárdenas, Sastry, “Effects of Risk on Privacy Contracts for Demand-Side Management.”]
Privacy Contracts in Practice

These privacy-differentiated goods already exist!

How AT&T Wants You to Pay For Your Privacy

Victor Luckerson @VLuck | Feb. 18, 2015

ISP can track your web history and searches

The privilege of not having your every click tracked, saved and regurgitated in the form of targeted ads will only cost you $20 per month on AT&T’s super-fast Internet service.

The company, which just announced it’s bringing its 1-gigabit-per-second service to Kansas City, touts a price tag of $70 per month for the high-speed connection.

Telematics explained

Telematics is not a new invention and has been used for over a decade in commercial vehicles, by the emergency services and Formula One teams. The technology is now widely being used in road cars.

The ingenie box is a self-contained unit the size of a smartphone that includes a:

- GPS unit which captures when and where the car is driven
- high frequency motion sensor which captures how the car is driven
- SIM card which is used to transmit the data

The cost of the ingenie box and the fitting of the box is included in the price of your insurance.
Closing Remarks

• Privacy is an integral facet of new IoT technologies.

• To consider privacy, we can:
  – passively quantify privacy in cyber-physical systems,
  – actively consider privacy-utility tradeoff in the design,
  – optimally tradeoff between privacy and utility.

• We need to consider privacy in the context of the emerging data market.