Transactional Privacy

Unknotting the Privacy Tussle w. Economics

Chris Riederer, Philippa Gill, Vijay Erramilli,
Pablo Rodriguez, Balachander Krishnamurthy, Dina Papagiannaki

A. Chaintreau (Columbia U.)

2/8/2013 @ DIMACS Workshop

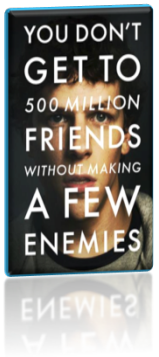
Acknowledgment

This is a joint work with Chris, Philippa, Vijay, Bala, Pablo & Dina!

Tech Bubbles: what they produce?

* Late 80s … cheap microprocessors, no applications
  – But had brought millions of pcs to business/home
* Late 90s … end of the dot-com boom
  – But the Internet infrastructure was built for most
* Early 2010s … peak of the social boom
  – Facebook 3rd "country",

,



MOBILE SOCIAL LAB
CS @CU

Today

What are we building for the next generation?

"The best mind of my generation are thinking about how to make people click ads." J. Hammerbacher

"This Tech Bubble Is Different."
A. Vance, Businessweek, 04/17/2011

Social Media & Computing

* The next generation could be the one with access to an unprecedented amount of behavioral data
* This can solve real problems

  … not just finding a movie or a restaurant!

  – ensuring energy efficiency
  – monitoring our environment
  – extend access to infrastructure
  – informing public decision

MOBILE SOCIAL LAB

CS @CU

“Data is web’s new oil”

And key to our society’s future!

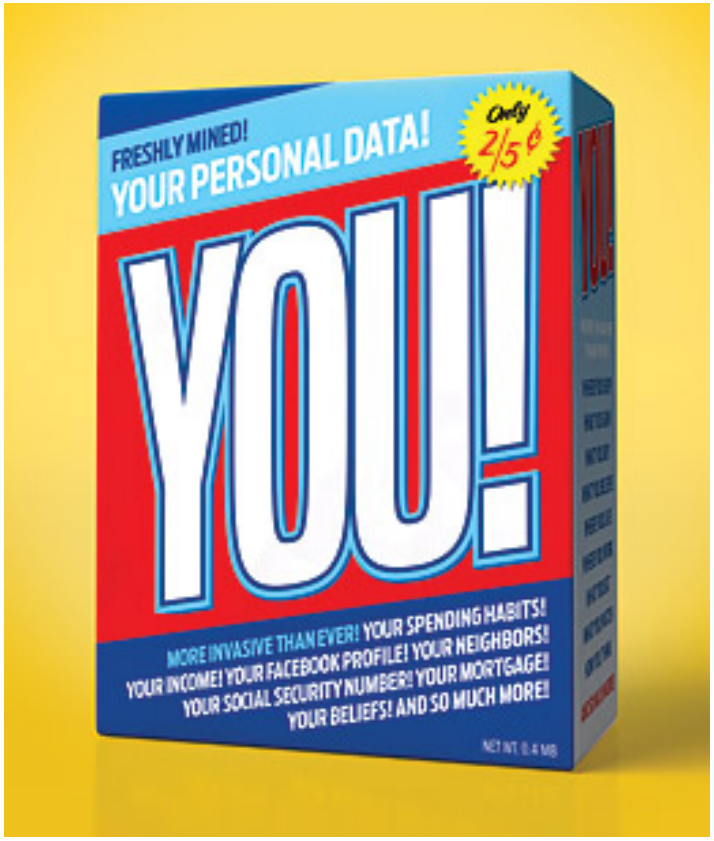
“You think you are Google’s customer?

De facto, you are Google’s product!”

S. Vaidhyanathan (2011)

Who produces this oil?

Who owns it? benefits from it?



MOBILE SOCIAL LAB

CS @CU

We have a problem …

"Privacy challenges **do not** and **must not** require us to **forego the benefits** of Network and Information Technology (NIT) in addressing national priorities.

Rather, we need a **practical science** of privacy protection, based on fundamental advances in NIT, to provide us with tools we can use to **reconcile privacy with progress**."

PCAST Report to the President and Congress,

Designing a Digital Future

MOBILE SOCIAL B

CS @CU

This talk

* Transactional Privacy, a primer
  - Need for alternative economic approach to privacy
* Highlights:
  - Can we practically build TP?
  - The real reasons why it may not work
  - Can it be incrementally deployed?
* Concluding remarks

The Privacy Tussle

Online Service Providers, Data Brokers, Aggregators

Users, Associations, Journalists, governments

More monetization of personal information

Stop the erosion of privacy? Regulate?

MOBILE SOCIAL LAB

CS @CU

What complicates the Tussle

* No limitation on 3rd party tracking
  - Permission ultimatum (Android, FB, Apple)
  - Aggregation (Re-targering, FB connect, quasi-logout)
  - Reselling (Rapleaf, bluekai, Google DDP)

* Privacy is difficult to perceive and to protect
  - Behavioral: Immediate gratification, illusion of control
  - Technical: inference (e.g. differential privacy)

Technical solutions

* Privacy preserving techniques
  - Anonymization: Tor, Obfuscation: TrackMeNot
  - Self-destructing data: Vanish
  - Monitoring: Dynamic Taint Analysis
  - Privacy-Preserving services: AdNostic, Privad, Repriv

* Not adopted, for 2 reasons:
  1. little user incentive, “privacy is not enough”
  2. Ignores data’s value, “really socially optimal?”

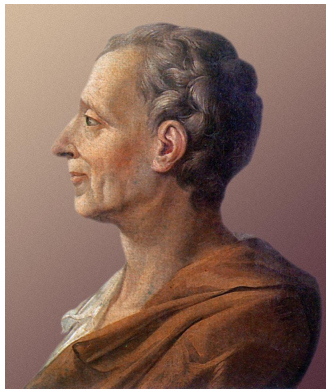
Fix the economy first!

MOBILE SOCIAL B

CS@CU

Transactional Privacy in a nutshell

* Principle 1: A relaxed definition of privacy
  - Is privacy the state of being free from observation? … or know and control who uses what about you?
  - We do not hide data, rather we enforce payment for their commercial use.
* Principle 2: A separation of powers
  - Who should decide what?
  - User "what is for sale?"
    market "what is it worth?"



MOBILE SOCIAL B

CS @CU

Privacy as usual vs. Transactional Pr.

Goal: free from observation

* Adversary:
  honest but curious

* Hard problem, requires
  - data through queries
  - Estimate privacy violation as negative externalities
* Many source of leakage
  - reselling
  - from price and bids

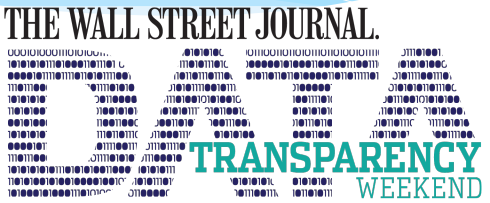
Goal: free from exploitation

* Adversary:
  malicious but rational

* Potentially easier
  - raw data
    works with any algorithm
  - simpler
* Inference is mostly useless
  - Brings no additional value

Economic solution to privacy

The price of free
https://github.com/ManConley/Price-of-Free/



1. Provide the right incentive to users
   - A perception of their data value
   - Information leakage = market arbitrage
2. Improve the new data economy
   - More transparent: give user a control
   - More democratic: let the best tech (not data) win!
   - More efficient? Avoid public campaigns, more data

Google's "Good to know" ~ 10m
Google Lobby +240% in 2012

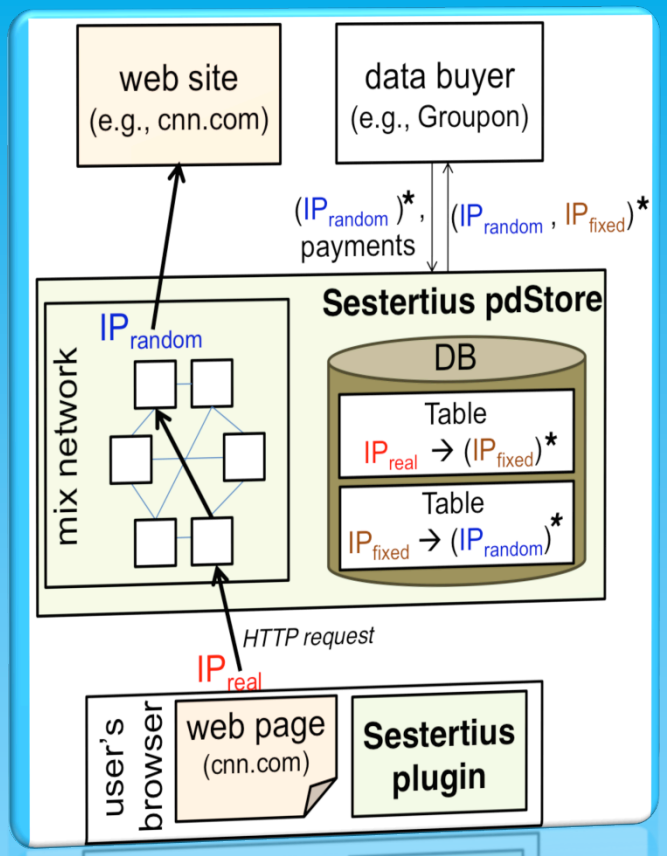
Nice but is it practical?

MOBILE SOCIAL B
CS @CU

This talk

* Transactional Privacy, a primer
  - Need for alternative economic approach to privacy
* Highlights:
  - Can we practically build TP?
  - The real reasons why it may not work
  - Can it be incrementally deployed?
* Concluding remarks

TP for web-browsing

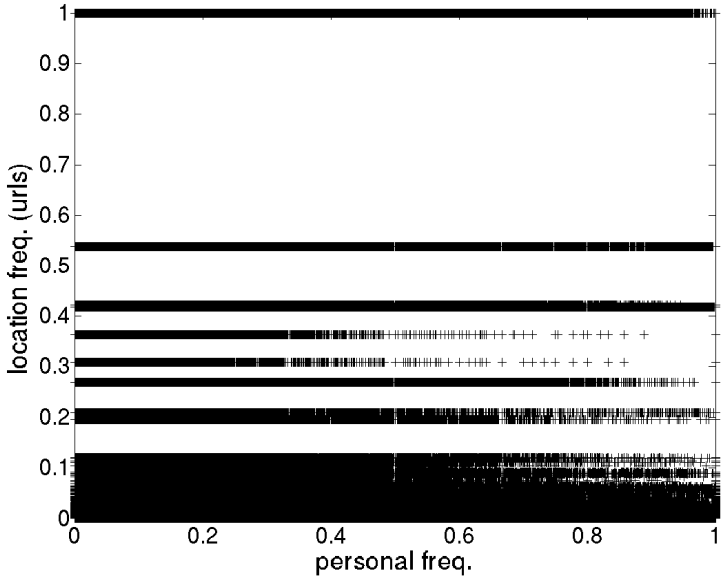


1. Data protection
   Mix network anonymize
   { IP address + cookies }
2. Data to sale+ Pricing
   unlim. supply auction
3. Revelation

Only those who paid can access the users identity during an impression

1. How to protect data?

* We don't protect to protect, we protect to sell later
  - Enough to make misbehavior economically inefficient
* What to sell? The really simple user Interface
  - How much do you value this bit? TOO HARD
  - Would you put this bit on the market? A BIT EASIER
  - Tune a simple scroll bar



2. How to Price Private Data?

1. As a function of User's loss?
   - Differential privacy + auctions [Ghosh-Roth11]
   - hard to put into practice: bid leaks, users' assessment
2. As a function of Provider's benefit?
   - Can be thought of as a coalition game [Kleinberg01]
   - Requires truthful revelation of value

* Run an auction (with unlimited supply)

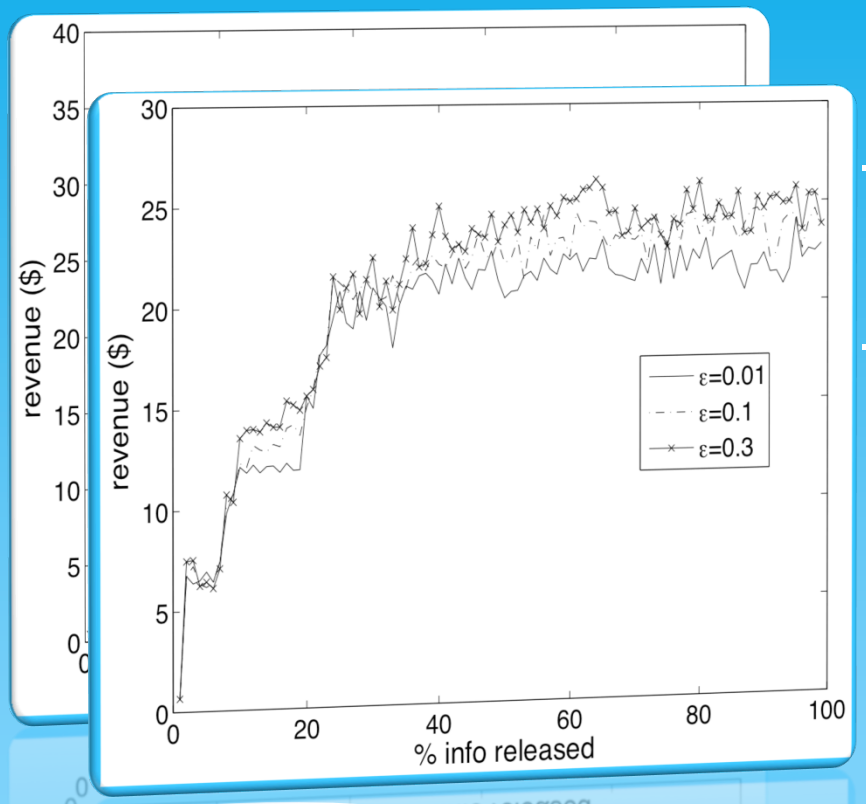
The personal data auction

* For sale: identifying your browsing in [t;t+1]
* Unlimited supply auctions
  - Sell your personal data to multiple purchasers
  - Every purchaser indicates a maximum price $p_{i,k}$
  - User's revenue $R\left((p_{i,k})_{k\in\mathcal{K}}, p\right) = \sum_{k\in\mathcal{K}} p \times \mathbb{I}_{\{p\leq p_{i,k}\}}$
  - Run exponential mechanism: $\frac{\exp\left(\varepsilon R\left((p_{i,k})_{k\in\mathcal{K}}, p\right)\right)\nu(p)}{\int_0^\infty \exp\left(\varepsilon R\left((p_{i,k})_{k\in\mathcal{K}}, s\right)\right)\nu(s)ds}$

3. Reveal

* Data obtained through de-anonymizer
  - The purchasers who won the auction are given the associating function IP-fake/IP-real for this user
  - Raw information: could be used for any algorithms
  - Real time: can be used for immediate action
* Re-run the bidding process periodically
  - Purchasers can infer users' profile from history
  - But they can't use it!

Case Study



- Mobile Web browsing large city, ~200k users
- Online Coupon Dealers crawl yipit.com
- Information released by decreasing popularity

Revenue vs. disclosure:

A sweet spot!

Confirms previous results on use of personal information to improve click-entropy (See [Krause-Horvitz 2008]).

This talk

* Transactional Privacy, a primer
  - Need for alternative economic approach to privacy
* Highlights:
  - Can we practically build TP?
  - The real reasons why it may not work
  - Can it be incrementally deployed?
* Concluding remarks

"This will not work because …

"I can resell your information to 1000 people"

"wait, I can even sell information about my friends!"

- BUT you can't sell access to info for commercial use!

"To bid, companies need information anyway"

- True, but for the same reason they can't monetize it

"You give away value of statistical information"

- Indeed, it becomes a public good. It's a feature!

"Price discrimination becomes unprofitable"

- Is that certain? Is that a bad thing?

“Still this will not work as …

“Tor is too slow anyway, and you can attack it”

- Something much lighter, since we only need to raise the bar. Companies care about reputation

“wouldn’t disclosing bulk of data scare users? today’s ecosystem relies on their ignorance”

- Aim at transparency; eventually users should know.

“wouldn’t it encourage users to over-expose.”

- Yes, which is why not all information can be traded

Why indeed it may not work

"What if users forge bogus data?"

"And get compensated for it, at the limit it means these signals are useless"

- still open problem: some data are verifiable

"What if there is there is not enough per user?"

"and they won't bother for 2c a month"

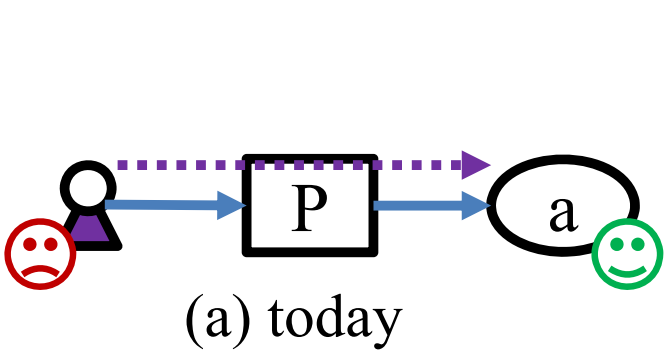
- still open problem: (1) we still have to make the math as the pie may grows, (2) we could make it more attractive: lottery, pay with services

This talk

* Transactional Privacy, a primer
  - Need for alternative economic approach to privacy
* Highlights:
  - Can we practically build TP?
  - The real reasons why it may not work
  - Can it be incrementally deployed?
* Concluding remarks

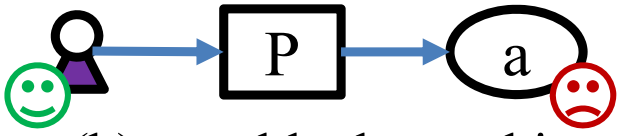
"Why Johnny can't opt-out"

* Current choice:

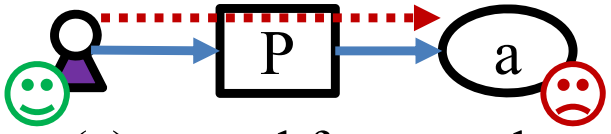


(a) today

or



(b) user blocks tracking

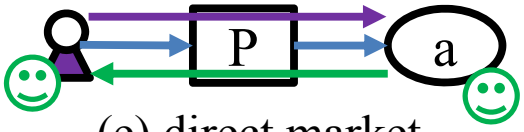


(c) user obfuscates data

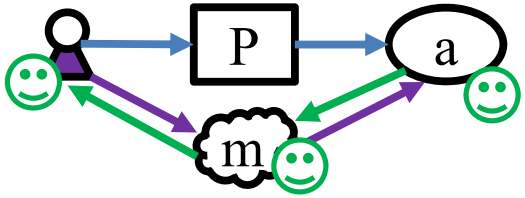
impression inferred data explicit data
obfuscated data revenue

MOBILE SOCIAL LAB CS @CU

Tomorrow possible's vision



(e) direct market



(f) mediated market

* But this creates initially some revenue loss
  - is there a deployment that is incentive compatible?

impression inferred data explicit data
obfuscated data revenue

MOBILE SOCIAL B CS @CU

A closer view at today's ads

* Using multiple traces (Residential, Mobile, Campus)
  - And a simple model of Cost-Per-Mille

$$CPM(u,p,a) = RON_a \times TQM_p \times \mathcal{I}_a(u)$$

  - RON is base price, TQM quality of site
* I is the "Intent" of user u as seen by aggregator a

$$\mathcal{I}_a(u) = \begin{cases} II_a(u) & \text{u do nothing} \\ EI(u) & \text{u sells data to a} \\ 1 & \text{u block tracking} \end{cases}$$

  - Estimated using categories and browsing + adwords

Characterizing Deployment

* Deployment under two scenarios:
  - Let $r_{u,a} = \frac{EI(u) - 1}{II_a(u) - 1}$ "consented tracking ratio"
  - r>1 because explicit intent is larger than implicit
  - relates intuitively to user's bargaining power
* Market deployment as a coalitional game
  - Prop: In a direct market, distributing revenue according to Shapley value (i.e. under fairness axioms) is incentive compatible iff r>2
  - Prop: In a mediated market, it is iff r>3/2

Distinguishing 1st and 3rd party

Publishers

* Make impressions

Publisher	Frac. Rev.	Frac. Users	Category
facebook.com	0.09	0.15	society
google.co.uk	0.04	0.11	computers
bbc.co.uk	0.03	0.07	arts
fbcdn.net	0.03	0.13	society
twitter.com	0.03	0.04	computers
yahoo.com	0.03	0.04	computers
google.com	0.02	0.18	computers
skysports.com	0.02	0.04	regional
premierleague.com	0.01	0.01	regional
ebay.com	0.01	0.02	shopping

* Largest ≠ more profitable

Aggregators

* Track users and play ads

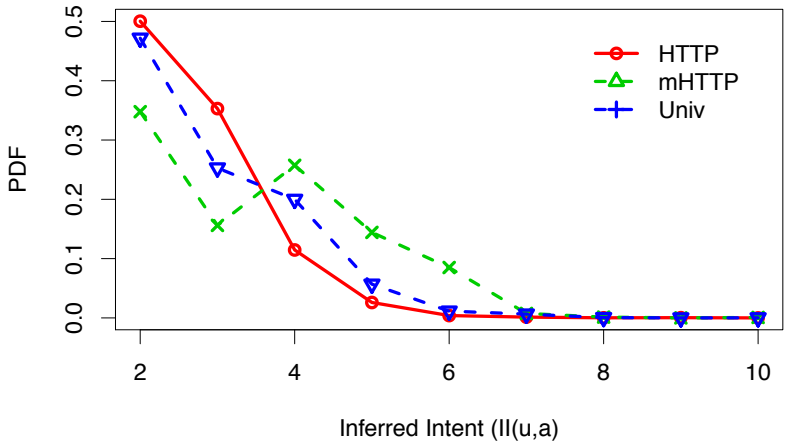
Aggregator	Frac. Rev.	Frac. Users	Frac. Pubs.
Google	0.18	0.17	0.80
Facebook	0.06	0.09	0.23
GlobalCrossing (AdMob)	0.04	0.11	0.19
AOL	0.03	0.04	0.07
Microsoft	0.03	0.04	0.17
Omniture	0.03	0.05	0.07
Yahoo! (AS42173)	0.03	0.04	0.07
Internap (RevSci)	0.02	0.03	0.01
Quantcast	0.02	0.03	0.09
Yahoo! (AS43428)	0.01	0.03	0.11

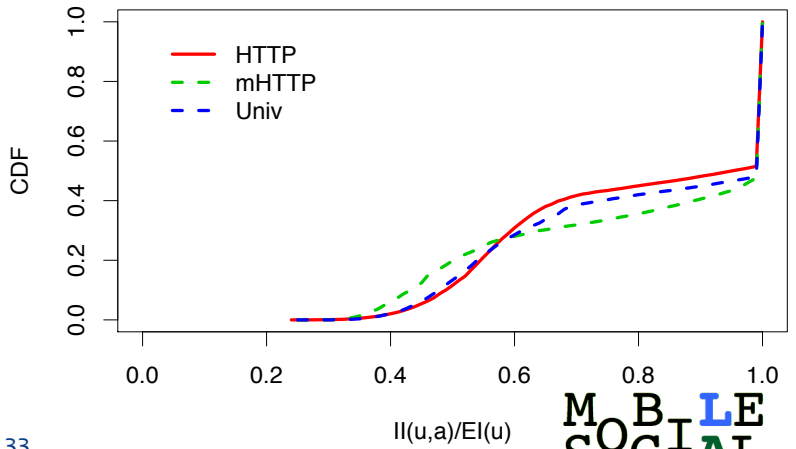
* Revenue more skewed

MOBILE SOCIAL B CS@CU

The power of large aggregators

* Largest aggregators have specific advantage
  - Implicit intent: based on what aggregator can infer
  - Explicit intent ≠ implicit intent
* But implicit intent is still not perfectly accurate
  - Leaving users some bargaining power





Concluding remarks

* We need to explore alternative approaches to privacy with an economic angle
  - Transactional privacy focuses on keeping data in control of which data is used and how
* Encouraging observations
  - Revenue vs. disclosures exhibits a sweet spot
  - Data revelation can exhibit mutual benefits
* Not shown today: adoption, location privacy

Thank you!

Riederer, C., Erramilli, V., Chaintreau, A., Krishnamurty, B., & Rodriguez, P. (2011).
For sale : Your Data By : You. *Proceedings of ACM SIGCOMM HotNets*

Gill, P., Erramilli, V., Chaintreau, A., Krishnamurty, B., Papagiannaki, D. & Rodriguez, P. (2013).
Money for nothing and click for free. *Working paper*

Riederer, C., Erramilli, V., Chaintreau, A., & Krishnamurty, B. (2013).
The price (and the place) is right: an economic solution to location privacy. *Working paper*