Privacy by design,
or,
How not to screw over your users

Lenny Foner

Outline

- Intro
- Why?
- What?
- How?
- Huh?

Who this is aimed at

- Handling data about people? Check.
- Trying to figure out your project’s architecture? Check.
- Not already a privacy expert? Check.
  - Experts can ask questions too, but hold until end.

A little about me

- MIT EECS, then Media Lab PhD
  - Yenta: Decentralized, crypto based social networking system implemented before the term had been invented
- A variety of industry experience
- Doing security, privacy, civil liberties since the early 90's
- Lots of policy stuff in the mix as well (CFP, WFPBD, NRC)
- Some unusual cases (more about these later)
  - AAAS Program in Science and Human Rights
  - National Network to End Domestic Violence

My agenda

- To change the world...

  ...by changing you
- We are privileged
  - We have enormous leverage
  - We are who will build tomorrow’s systems
  - The point of an MIT education is new ideas, not just coding
  - When others need help, they turn to us
  - You have a moral obligation not to waste that opportunity
  - Think of Hammurabi
- Maybe SIPB can influence the rest of MIT this way

What is privacy, anyway?

- The right to be left alone
  - Warren & Brandeis, SCOTUS, 1890
- A basic human right
  - OECD statement of principles
- A fundamental enabler of personal growth
  - People who feel without privacy are inhibited and timid
- The bogusness of "nothing to hide"
- PII is only part of the problem; metadata hurts badly

What kinds of privacy am I not covering today?

- HIPPA
- COUHES (IRB)
- PCI

What’s privacy’s opposite?

- The Panopticon (Bentham, late 18th century)
  - Every prisoner might be under surveillance
  - Impossible for any of them to tell
  - Random reinforcement equals maximum paranoia
- Inspiration for 1984
- Video surveillance is the modern Panopticon
- Big Data is its postmodern demon spawn

What’s a threat model?

- What are you trying to protect?
- How much are you willing to spend?
- What happens when you fail?
  - Note that’s not "if" you fail

Threats to privacy

- Outsiders
- Insiders
- Lawyers
- Murphy
- Evil

What is privacy by design?

- Building systems that are inherently privacy-protective
- Architecture and policies enforce the outcome
- Safety obvious even to outsiders
- The law cannot help us here
  - Subpoenas are an old thing (ask Fedex)
  - National Security Letters are a new thing

Why privacy by design is not the same as security engineering

- Security is only a delaying action
- Security failures often usually only cost money
- Privacy failures can cost lives
  - ...and let me give you some examples

Let’s talk about stalking

- "During a 12-month period an estimated 14 in every 1,000 (1.4%) persons age 18 or older were victims of stalking." [Bureau of Justice Statistics, US Department of Justice]
- Every reason to believe this is an underestimate
  - 18-25 year-olds are stalked at 30 per 1000, aka 3%
  - Some countries report much higher incidences (Australia, Korea, Iran...)
  - Many victims are totally off the map, can’t be found by/don’t trust surveys
  - Some estimates as high as 50% of all women will be stalked in their lifetime (NNEDV)
  - Most stalking unreported to police
  - Some stalk their victims more than 5 years
  - The most common motivations of stalkers center around anger, revenge, and control

More stalking

- Other demographics
  - 75% of victims know their stalkers (especially women)
  - 25% of stalking victims are male
  - Whites stalked more than others
  - Divorced/separated rates are 34 per 1000 and up (>3.4%)
  - Add in "harassment" and these rates go up another 2%
- What do stalkers do?
  - Threats (43%)
  - Property damage (16%)
  - Violence (12%)
  - Identity theft (10%)
- Run the numbers for MIT
  - ...and MITDIR

Threat models from hell

- NNEDV
  - Chronically underfunded
  - Victims typically flee with next to nothing
  - Granting agencies want to ensure no phantoms in shelters
  - Some stalkers have gotten jobs in shelters & state agencies!

Threat models from hell

- AAAS SHR
  - Victims of state-sponsored violence
  - Guatemala, Sri Lanka, South Africa, ...
  - Offenders have been tried and convicted at The Hague
  - "Who did what to whom"
  - Statistical approaches to uncovering the offenders
  - Interviewing tens of thousands of people, one at a time
  - If the database is stolen?

    ...thousands of people could be "disappeared"

How do we fix this?

- A fundamental mindset
- Some basic design principles
- Some case studies

Don’t be afraid to be the least popular person in the room

- "But we might need it!"
  - Physics envy
  - Big Data envy
- "But we don’t know how to do without it!"
  - What are you really trying to do?
  - Architecture matters
    - Examples for later: Yenta, NNEDV, SHR, space usage
- Mission creep
- Magical thinking

Scale matters

- If I drop one snowball on you from a rooftop

  ...you’ll be annoyed.
- If I drop 100,000,000 snowballs on you from a rooftop

  ...you’ll be dead.

Beware of slippery slopes.

- Automated license plate scanning
  - "It’s just like having a cop write down your plates."
  - No it’s not. We’re not stupid.
- This was seriously advanced by a cop at CFP 2015.

  (And I was disappointed I had to be the one to call him on it.)

Scaling effects, again

- Librarians believe your borrowing history is your own.
- Old-syle library cards
  - ...require manual traversal of every book in the library
  - ...which could take years
- A centralized computer database
  - ...makes lookup instantaneous

Which do you think enables fishing expeditions via subpoenas?

(We’ll talk later about how librarians solve this.)

Timeless advice: Never piss off a librarian.

A basic principle: Don’t collect it

- You can’t leak what you don’t have
- You can’t be subpoenaed for what you don’t have
- You don’t have to store what you don’t have
- You don’t have to back up what you don’t have
- Your users don’t have to trust you about what you don’t have

This is the single hardest thing to do in the world.

- You need force of will
- And a suitable architecture

A basic principle: Don’t keep it

- Design your system to completely flush old data
  - Bounds your liability for a breach
  - Subpoenas are slow

It’s almost impossible to get rid of data.

- Mission creep
- Backups
- Logfiles

A basic principle: Don’t own it

- Decentralize the system
  - Only peers have data
  - Removes single point of compromise
  - Increases jurisdictional barriers to lawyers
- Some things are inherently distributed
  - ...like your private keys
  - ...to your iPhone

If your business exists to extract rents from users,

this will be an unpopular stance.

  - ...but you probably aren’t renting keys

Yenta

So how do you get rid of data?

- Deletion is great, but backups are a problem
- Again: "What are you trying to do?"
- Use the structure of the problem to help
- If you’ve returned a checked out book...

  ...who cares if you previously checked it out?

Encrypted backups, keyed by date

- Destroy the key and the backup is gone
- Safeguarding a few recent keys is easy
  - ...certainly easier than safeguarding the tapes

NNEDV and AAAS SHR

- Some problems are hard to fix with technology
  - NNEDV check-in requirements met with pushback
  - Agencies need education in stalker APT
  - Sometimes requires legislative assist
- But some are easier
  - AAAS SHR trained caseworkers in crypto; laptops encrypted
  - Data immediately & automatically exported out of country
  - Theft loses only a few hours and exposes (almost) nothing
  - Workers also safer because less coercion threat
  - "This driver carries no money"

Bad ideas

- Blinding is hard
- Latanya Sweeney and medical records
  - Zipcode, birthday, and gender
  - >90% de-anonymized
  - Dropping gender doesn’t help much
- Why hashing won’t save you
  - There are only so many IPv4 addresses
  - Ditto gmail addresses
  - ...and common passwords, and birthdays, and...

Bad ideas

- Scenario:
  - Hundreds of shared artist studios and some workshops
  - Gym model: Flat-rate monthly pricing with membership types
  - Everyone must badge in; no tailgating; no badging out
  - Open 24x7 to those with membership cards
  - Minimal physical security
- What do you do about a member directory?
  - Do not allow creation of an automated directory
  - Use an opt-in wiki page instead
  - Only people who want to be listed are listed

Bad ideas

- Badge-in records
  - Would being closed certain hours inconvenience many?
- What’s wrong with this picture?

```
57  1449559848  Alyssa P. Hacker
58  1449559858  Ben Bitdiddle
```

- Hint #1: "The most common motivations of stalkers center around anger, revenge, and control."
- Hint #2: Who’s sleeping with whom?

Bad ideas

- Badge-in records
  - Would being closed certain hours inconvenience many?
- What’s wrong with this picture?
  - `57   1449559848   Alyssa P. Hacker`
  - `58   1449559858   Ben Bitdiddle`
  - Dither the timestamp?
    - By how much? Randomly?
    - But averaging noisy samples decreases variance (side-channel attacks)

Bad ideas

- Badge-in records
  - Would being closed certain hours inconvenience many?
- What’s still wrong with this picture?
  - `57  1449558901  Alyssa P. Hacker`
  - `58  1449557592  Ben Bitdiddle`

Bad ideas

- Badge-in records
  - Would being closed certain hours inconvenience many?
- What’s still wrong with this picture?
  - `57  1449558901  Alyssa P. Hacker`
  - `58  1449557592  Ben Bitdiddle`

Autoincrement will always show adjacencies, despite dither

- One solution is to use hour-wide histogram bins w/o autoincrement table
- ...even better, don’t keep the names associated with the histogram
- ...or maybe keep names a few days for auditing, then flush to histogram

What actually happened, and why

- Politics, broken promises, indefinite data retention, Big Data envy, ...

Bad ideas

- But did any of this really solve the problem?
- Hint: Do people have to badge out?
- No, and you’ll never be able to force them to, either
  - ...so you have no idea how long they stayed
  - ...thus no idea about occupancy
  - ...and you don’t know which shop they were in
  - ...if they didn’t just spend the time in their studio anyway
- Record entrance to the second, but don’t record exit at all
- Measure with micrometer, mark with chalk, cut with axe
- I’ll discuss a real solution shortly

Bad ideas

- Blinding location data is especially hard
- This coincidence vulnerability affects the MBTA, too
  - Should the MBTA get to know who’s sleeping together?
  - Not to mention divorce lawyers
  - And how long does the MBTA keep this data, anyway?
  - Probably forever (because politics)
- And let’s not even talk about EZ-Pass
  - I predicted the divorce lawyers the day it was announced
- And GPS data, and cellular data, and...
- Then there’s the issue of cameras...

Video surveillance and magical thinking

- Video surveillance is an excellent Panopticon
- But the whole point of the Panopticon is security theater
- Most video surveillance is pointless, unless the point is fear
- Getting results w/o theater requires careful lifecycle analysis
  - ...which is rarely done
- Real-time monitoring can work
  - Casinos, which have bouncers and cops
  - Stores with shoplifters, which have cops
  - Most cameras are aimed at the cash register; guess why?

Video surveillance and magical thinking

- Stored video almost never useful
  - ID usually ineffective
  - How long do you store it?
  - Who broke it? Who knows?
  - Who cares about small items?
  - Where do you point the camera?
  - Access policies are a giant pile of snakes
    - Stalker heaven to an insider
    - Maybe even to an outsider
    - Better have good auditing
- A win: Occupancy via motion detection

Getting others on board: Short-term

- Implementation can fix things in stone
  - Get there first or live with the consequences
  - Provide an implementation that isn’t a civil-liberties disaster
  - Like security, privacy is very difficult to bolt on later
- Business methods & process are implementations, too
  - They run on people, memos, traditions, habits, and mindsets
  - Often even harder to fix than the code
  - A fish rots from the head

Getting others on board: Medium-term

- Find allies in your organization
  - If your designs are socially aware
    - ...you enable companies to do the right thing
    - ...but they can’t if the technology doesn’t exist
  - Beware empire-builders and evil people
    - Many organizations have both types
    - Fight them with organization, data, and alternatives
- Invoking the bean counters
  - Don’t-collect/don’t-keep/don’t-own can save resources
    - If you can avoid massive centralization of lots of private data...
    - ...fewer servers, backup, replication, security audits, ...
    - ...not to mention liability, lawsuits, compliance, paperwork, ...

Getting others on board: Long-term

- Teach
  - Those after you will build the systems you use
  - Computation and networking affect society
    - ...as much as the Manhattan Project did
    - ...just because it doesn’t go boom doesn’t mean lives aren’t affected worldwide
    - ...this is why ethics and worldviews matter
  - Join professional & lobbying organizations, interest groups, ...
- Create your own movements, organizations, and allies
  - This is part of why I’m speaking today
- Don’t despair: change is possible

"Just because you will not finish the job, you must still take the first step."

Some resources

- Books (not so obvious, but teach a useful mindset)
  - Between Silk and Cyanide (Leo Marks)
    - Tradecraft: Think like a spy
    - Know thy enemy: Make their lives difficult
  - Normal Accidents (Charles Perrow)
    - Close-coupled & fast vs loosely-coupled and slow
    - Fun with radar, or Murphy’s revenge
- Organizations
  - EFF
  - EPIC
  - ACLU

Summary

- Privacy matters, and you must help
- Politics are a problem; we need allies
- What are you really trying to do?
  - Solve the right problem
  - Don’t allow others to mushroom it
- Antipatterns
  - Big Data envy
  - Blinding is hard
- Patterns
  - Don’t collect it
  - Don’t keep it
  - Don’t own it

Suggestions for ongoing projects? Other questions?

- Want some advice for a project?
- Other questions?