Commercial Privacy

Privacy Concerns are Ancient

- Physical privacy concerns go back more than 1800 years
- Credit reports date to the mid-19th century
- But modern technology has made the threat acute

Enter the Computer

- The Committee on Science and Law of the New York City Bar Association started its formal privacy study in 1962
  - This led to Alan Westin's 1967 book "Privacy and Freedom", a report on the committee's work
- The US Congress held hearings on technology and privacy throughout the 1960s
- Legal academics wrote extensively on the topic
- Major concern then: government databases

Notice and Consent

- **Westin (1967)**: “A central aspect of privacy is that individuals and organizations can determine for themselves which matters they want to keep private and which they are willing—or need—to reveal.”
- This has been the basis for virtually *all* privacy regulation since then

Notice and Consent

- Sites tell you what they'll collect, and what they'll do with it
- By using the site, you are deemed to have consented to this policy

Privacy Regulation

- **1973:** A US government committee came up with the “Fair Information Practices”
- **1974:** The US government passes the *Privacy Act of 1974*, implementing them—but only for the Federal government
- **1980:** The OECD guidelines suggested more or less the same thing, but for everyone
- **1994:** The EU’s *Data Protection Directive* is enacted
- **2012:** The EU’s GDPR is adopted
- From 10,000 meters, all of these are more or less the same: notice and consent

The HEW Committee

The HEW Advisory Committee

- In response to mounting privacy concerns, the then-cabinet department of Health, Education, and Welfare convened an advisory committee
- Its 1973 recommendations—the Fair Information Practices—still form the basis for privacy regulation around the world

The FIPs

- Basic rules for minimizing information collection, ensuring due process, protection against secret collection, provide security, ensure accountability
- Emphasize individual knowledge and consent
- Principles are broadly accepted, but individual principles not implemented uniformly

The Principles

- Collection limitation
- Data quality
- Purpose specification
- Use limitation
- Security
- Openness/notice
- Individual participation
- Accountability

Note: these revolve around PII (personally identifiable information)

The Web

Dawn of the Web

- In 1990, Tim Berners-Lee invented the web as a way to distribute documentation
  - Crucial notion: *hypertext,* a way for documents to contain links to other documents
  - (Hypertext in quasi-modern form also dates to the 1960s)
- Others wanted to enable e-commerce
- The original design couldn't quite accommodate this in a clean fashion

The Web: Design

- Two primary components, *HTML* and *HTTP*
  - **HTML:** HyperText Markup Language; describes how a page should be formatted
  - **HTTP:** HyperText Transport Protocol; used to transmit web pages over the Internet
- Stateless design
  - Open a connection, download a page, close the connection
  - No link between downloads—and hence no way to have a *session*

Sample HTTP

```
GET / HTTP/1.1
Host: greylock.cs.columbia.edu
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10.16; rv:85.0) Gecko/20100101 Firefox/85.0
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate
DNT: 1
Connection: keep-alive
Upgrade-Insecure-Requests: 1
```

Sample HTML

```
<html>
<title> weblog </title>
<body> <h1> I heard you say </h1> <font size+=4>
<pre>GET / HTTP/1.1
Host: greylock.cs.columbia.edu
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10.16; rv:85.0) Gecko/20100101 Firefox/85.0
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate
DNT: 1
Connection: keep-alive
Cookie: WhoYouAre=1804289383; ID-Age=1612720694; Last-Seen=1612721125; Flavor="Chocolate-chip"; Size="Large"
Upgrade-Insecure-Requests: 1
</font>
</pre>

</body>
</html>
```

Session Needs

- Ability to log in
- *No requirement for a login name*
- Persistent preferences, e.g., language
- Shopping cart
- The accepted answer: *cookies*

Cookies

- Cookies are arbitrary text strings sent to a browser by a web site
- They’re retained by the browser in non-volatile storage and returned when the site is next visited
- *Cookies can be persistent identifiers*
- They can hold anything else a site wants, too

Cookies

I heard you say

```
GET / HTTP/1.1
Host: greylock.cs.columbia.edu
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10.16; rv:85.0) Gecko/20100101 Firefox/85.0
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate
DNT: 1
Connection: keep-alive
Upgrade-Insecure-Requests: 1
```

from 24.194.9.206:47607

I just sent you, #1804289383, a cookie; reload this page to see it coming back to me.

Cookies Reloaded

I heard you say

```
GET / HTTP/1.1
Host: greylock.cs.columbia.edu
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10.16; rv:85.0) Gecko/20100101 Firefox/85.0
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8
Accept-Language: en-US,en;q=0.5
Accept-Encoding: gzip, deflate
DNT: 1
Connection: keep-alive
Referer: http://greylock.cs.columbia.edu/
Cookie: Size="Large"; WhoYouAre=1804289383; ID-Age=1612724999; Last-Seen=1612725000
Upgrade-Insecure-Requests: 1
```

from 24.194.9.206:37450

I just sent you, #1804289383, a cookie; reload this page to see it coming back to me.
ID Age: Sun Feb 7 14:09:59 2021
Last visit: Sun Feb 7 14:10:00 2021

Third-Party Cookies

- Images and IFRAMES—embedded web pages—are loaded via separate URLs
- These URLs can point to a different site—and each such site can send and receive its own cookies
- HTTP requests for embedded content contain "`Referer`" lines that identify the parent page
- *Most ads are in IFRAMEs pointing to third-party ad brokers*
- *Consequence: third parties can track you around the web*

Internet Advertising

- Dominated by Google and Facebook
- They observe the content of the embedding pages to learn your interests
  - Other features—Facebook's embedded "like" buttons, Google Analytics, either's single sign-on—also contain embedded content and track you around the web
- Cookies are not PII per se—but many web sites know real names, etc.
  - Besides, gmail addresses and Facebook logins are PII

Privacy Regulation

How is Privacy Regulated?

- The EU: the General Data Protection Regulation (GDPR)
- The US
  - The Federal Trade Commission
  - The Federal government
  - Sector-specific (and state) regulations

GDPR

- An EU regulation, binding on member countries
  - Succeeds two earlier “directives”
- Strict privacy rules
- Enforced by government Data Protection Authorities
- Based on a design from a US Dept. of HEW advisory committee report in 1973: the FIPPs (Fair Information Practice Principles)

General Data Protection Regulation

- Applies within the EU
- Protects citizens of EU countries anywhere in the world
- Successor to the Data Protection Directive
- A regulation, not a directive—it's binding on member countries
- One of the strictest privacy laws anywhere
- Limited right of private action

Personal RIghts Under the FIPPs in the GDPR

- Access to data
- Accountability: companies must document how they comply
- Data processing only with consent, lawful obligation, or other valid reasons
- Transparency about data handling
- Security
- The right to be forgotten—we'll discuss more on this in a few weeks

The Federal Trade Commission

- With one exception, data about children, the FTC has no explicit statutory authority about privacy
  - In 1995, though, Congress urged it to get involved in privacy
  - The FTC (generally) does not issue regulations
- The FTC can act against “unfair or deceptive trade practices” that cause “harm”
- What does that mean?

The Federal Trade Commission

- The Federal Trade Commission Act bars “unfair or deceptive acts” if they cause or are likely to cause “substantial injury”
- The FTC has interpreted this authority very broadly

(1) Unfair methods of competition in or affecting commerce, and unfair or deceptive acts or practices in or affecting commerce, are hereby declared unlawful.

15 U.S.C. 45(a)(1)

The Commission shall have no authority under this section or section 57a of this title to declare unlawful an act or practice on the grounds that such act or practice is unfair unless the act or practice causes or is likely to cause substantial injury to

15 U.S.C. 45(n)

The FTC and Privacy Policies

- If a company violates its own privacy policies, that's obviously deceptive
- But—there's no requirement for a protective privacy policy
- And: what is "harm"?

The FTC and Security Breaches (or Issues)

- If a company skimps on security measures, that might be unfair competition
- If a company skimps on security measures *and* promises to keep your data secure, that's deceptive
  - But: what is the norm for security measures?
  - Most companies don't fight the FTC on this; Wyndham Hotels and LabMD did
- And again—what is "harm"?

Example: The FTC and Twitter

- Twice, hackers were able to gain administrative access to Twitter accounts, allowing them to send bogus tweets, read private DMs, etc.
- Twitter had bad security controls for admin accounts
- This was “deceptive” because “The privacy policy posted on Twitter’s website stated that “We employ administrative, physical, and electronic measures designed to protect your information from unauthorized access.”

Example: The FTC and Wyndham Hotels

- Wyndham Hotels had substandard security controls on its back-end systems
- They were hacked multiple times; consumer credit card numbers were taken
- The theft and abuse of these numbers was the “actual harm”
- It was “unfair” competition because other hotel chains went to the expense of securing their systems

What is “Harm”?

- Unclear!
- Often established by case law
- Easy case: financial loss to consumers
- Almost as easy: leaked health information
- Hard: other disclosure of personal data
- (Recent article: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3782222)

US Sector-Specific Privacy Laws

- The US has many sector-specific privacy laws
  - FERPA, for educational data
  - FCRA, for credit reports
  - HIPAA, for health data
  - Etc.
- They’re all different…
- And: some states are enacting their own, strict privacy laws
- Major issue for a possible comprehensive Federal law: should state laws be preempted?

State Privacy Laws

California Consumer Privacy Act (CCPA)

- Strongest privacy law in the US
- Origin was a ballot initiative, but the legislature acted first
- Applies to larger companies, but financial and health care institutions are generally excluded
- Protects only California residents
- Limited private right of action
- Violators can “cure” violations and escape penalties
- Again, based on the FIPs

Basic Principles

- Notice to data subjects
- Right of access
- (Limited) right to be forgotten
- Right to opt out of sale
- Right to receive services on equal terms
- Cannot discriminate against people who exercise CCPA rights
- Appropriate data security

Privacy Policies

- All of the usual provisions
- Must mention right of deletion
- Must include a “do not sell” link on web sites
- Must describe data sharing
- The California Attorney General believes that consumers have the right to see what businesses have inferred about them

California Privacy Rights Act

- Goes into effect in 2023
- Adds right to correction, right to opt out of AI systems, right to information about such systems
- Limits processing of sensitive data
- Mandatory risk assessments
- Creates a state privacy protection agency
- Requires data minimization, purpose limitation, and retention limitation
- Much more GDPR-like

Other State Privacy Laws

- Several other states (Virginia, Colorado, Utah, Connecticut) have their own privacy laws
- (See https://iapp.org/resources/article/us-state-privacy-legislation-tracker/ for up-to-date status)
- Some of these laws were effectively drafted by industry…

Daily Bird 2.0

Great blue heron on the ice in Morningside Park, December 24, 2019

CS@CU

Biometric Privacy Laws

What is a Biometric?

- “Biometrics are unique physical characteristics, such as fingerprints, that can be used for automated recognition” (Dept. of Homeland Security)
- Examples: fingerprints, faces, typing rhythm, voice print, retina or iris scan, gait, DNA, more
- Often used for matching or authentication

The Problems with Biometrics

- Biometrics aren’t (easily) changeable
- You can’t replace your fingerprints or your eyes
- Facial changes are rarely dramatic, and not frequently done
- Some biometrics, including facial images and fingerprints, can easily be captured by high-resolution photography
- They thus represent a considerable privacy risk

The Illinois Biometric Information Privacy Act

- Biometric information cannot be collected without advance written notice and written consent
- Notice must be given of what is collected, why, and how long it will be stored and used
- Biometric data cannot be sold or transferred without consent
- There is a private right of action
- Facebook had to pay $650 million to settle a class action suit for tagging people in photos via facial recognition

Texas Law

- Similar to Illinois', but no private right of action
- Only the Attorney General can file suit
- He has filed a large lawsuit against Meta

Washington; New York City; Others

- Washington's law is similar to Texas', but doesn't cover facial images
- New York City's law is primarily concerned with stores and walk-by facial recognition—signs must be prominently posted
- A number of other jurisdictions have banned facial recognition, especially by law enforcement
- Many other states are considering biometric privacy laws

Data Breach Notification Laws

The Problem

If a site is hacked, private data can be exposed

- Sometimes, the threat is potential—an attacker was known to be on the site—and sometimes it's actual
- Log files or deliberate data releases can show actual accesses
- Of late, ransomware perpetrators have leaked data to enhance the damage potential

When private data is released, people are harmed

The Laws

- Every state, the District of Columbia, and several territories have data breach notification laws
  - People whose data is taken have a right to know
- However, provisions differ, often wildly
  - Some include a right to credit monitoring

Issues

- Who is covered, businesses, data brokers, or government agencies?
- What is “personal information”?
- What is a breach?
- How and when must people be notified?
- Penalties?
- Exemptions, e.g., for encrypted data
- Remedies

Example: California

- Covers businesses, data brokers, and government agencies
- Breach must include name plus something else, or email address plus password
- Exempts encrypted data unless the key is taken
- People have to be notified promptly
  - A suggested notification form is provided
- A year of identity theft protection must be offered

Example: Mississippi

- Government isn’t covered
- Notification must be prompt
- Exempts encrypted data, but makes no mention of keys
- Violations are “unfair trade practices”, but there is no private right of action; only the state Attorney-General can act

In Other Words…

- Provisions vary widely
- There is no one national standard
- You may have to comply with the most restrictive laws

Jewish Sources

Ancient History of Privacy

- Privacy concerns are ancient
- A biblical text “How fair are your tents, O Jacob, Your dwellings, O Israel” (Numbers 24:5) was interpreted by ancient rabbis to mean that the tent doors were not aligned with each other, precisely to preserve privacy
- The Mishnah and Talmud extended that to “building code” requirements: windows and doors shouldn’t face into each other
- In 1890, Samuel Warren and Louis Brandeis published “The Right to Privacy” in the Harvard Law Review

Numbers 24:2, 5

וַיִּשָּׂא בִלְעָם אֶת-עֵינָיו , וַיַּרְא אֶת-יִשְׂרָאֵל, שֹׁכֵן, לִשְׁבָטָיו; וַתְּהִי עָלָיו, רוּחַ אֱלֹהִים.

…

מַה-טֹּבוּ אֹהָלֶיךָ, יַעֲקֹב; מִשְׁכְּנֹתֶיךָ, יִשְׂרָאֵל.

As Balaam looked up and saw Israel encamped tribe by tribe, the spirit of God came upon him… How fair are your tents, O Jacob, your dwellings, O Israel!

Mishnah Bava Batra 3

לֹא יִפְתַּח אָדָם לַחֲצַר הַשֻּׁתָּפִין פֶּתַח כְּנֶגֶד פֶּתַח וְחַלּוֹן כְּנֶגֶד חַלּוֹן

In a jointly held courtyard a man may not build a door directly opposite another’s door, or a window directly opposite another’s window.

Talmud Bava Batra 60a

אמר רבי יוחנן דאמר קרא וישא בלעם את עיניו וירא את ישראל שוכן לשבטיו מה ראה שאין פתחי אהליהם מכוונין זה לזה

Rabbi Yoḥanan says that verse (Numbers 24:2) states: “And Balaam lifted up his eyes, and he saw Israel dwelling tribe by tribe”; What did he see? He saw that the entrances of their tents were not aligned with each other.

Collection Limitation

- Only collect what you need
- Delete data when it is no longer needed
- *Data that doesn't exist can't be stolen or otherwise misused*

Data Quality

- Make sure the collected data is accurate
  - Using inaccurate data on people can itself cause harm
- Implication: people should have the right to see data about them to find inaccuracies
- Implication: people should have the right (and the ability!) to correct inaccurate data

Purpose Specification

- People must know why data about them is being collected
- (So must regulators!)
- Vagueness—“we may share with X and Y”—is unacceptable

Use Limitation

- A corollary to purpose specification—data cannot be used for other purposes
- Secondary use—reusing data collected for one purpose for another—is at the heart of the most serious privacy violations

Security

- If a system is insecure, data on it cannot be private
- If privacy is “the degree to which the entity is willing to share its personal information with others,” the subject has to know with whom the data is being shared
- If, due to insecurity, it is simply taken by another party, the subject’s privacy has been invaded by definition

Openness and Notice

- There should be no secret databases
- People should be aware of when their data is being collected
- People should be aware of what is being collected, and why

Individual Participation

- Individuals should have the right to consent to data collection
- Individuals should know what is being collected
- Individuals should have the right to obtain a copy of the records about them

Accountability

- Entities that collect or hold information about individuals must be accountable for compliance with these principles
- Record all accesses and transfers of personal data
- At such entities, particular individuals must be accountable for compliance, including for security
- Violations should be dealt with by civil and criminal penalties, as well as injunctive relief and personal lawsuits

Current Status

Problems with Notice and Consent

- Amount of data collected, and by whom
- Privacy policies
- Location data is collected, often without folks' knowledge
- Many governments

Where Are We Now?

- Notice and consent is still the norm, despite its many problems
  - No one has come up with a better paradigm
- However, details matter
- *Details vary by jurisdiction*

Analytics Platforms

“To those first-party profiles, Rubicon typically adds details from third-party data aggregators, like BlueKai or eXelate, such as users’ sex and age, interests, estimated income range and past purchases. Finally, Rubicon applies its own analytics to estimate the fair market value of site visitors and the ad spaces they are available to see.”

(New York Times)

Privacy Policies

- No one reads them
  - Cranor estimated the opportunity cost at US$3500/year to read them all
- They're deliberately vague and expansive
  - "We may collect personal information and other information about you from business partners, contractors and other third parties." (Reidenberg et al)
- "Only in some fantasy world do users actually read these notices and understand their implications before clicking to indicate their consent." (PCAST report)

Location Data

- Huge issue for mobile devices
  - Many apps collect and analyze such data
- IP geolocation also reveals a lot

Governments

- If data exists, it’s available to governments
- Some governments have a complex, restricted, and somewhat painful process required to gain access to data
- Other governments don’t care very much about such niceties
- Some governments collect data via espionage, technical and otherwise

Machine Learning

- Today's ML algorithms can infer things not directly observed, e.g., sexual orientation
- This is much harder to control: it is *not* based on data collection, which is usually what's regulated
- Even when some inputs are disallowed by law, there are often proxy variables that are strong correlates

Overcollection

- Data brokers—outside parties with whom consumers have no association, and to whom they have never consented—collect, buy, and sell a tremendous amount of data
- Websites track users
- Ads are from outside brokers, who use HTTP redirection to gather even more data
- Also: third-party “like” buttons (e.g., Facebook and Twitter),third-party analytic services, and third-party authentication (e.g., Facebook and Google)

Use Controls

- A different proposed paradigm: don't control collection, control how data is used
  - (Use limitations are in the FIPs)
- Supported by some academics and advisory groups; others think that large databases are *a priori* dangerous and shouldn't exist
- Under *Sorrell v. IMS Health Inc.*, 564 U.S. 552 (2011), such rules may not be constitutional in the US