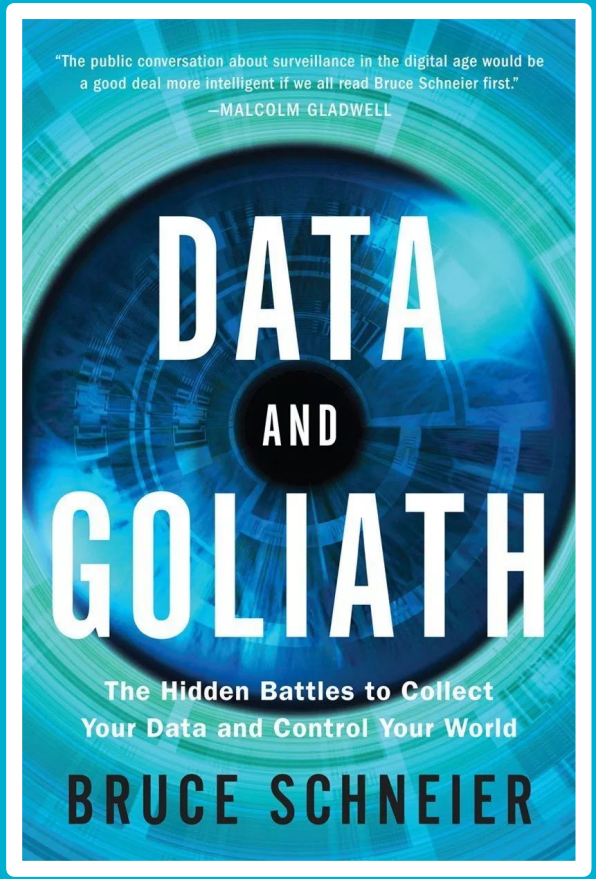
# Data and Goliath

## The Hidden Battles to Collect Your Data & Control Your World




Bruce Schneier

# Bruce Schneier

- Interesting guy… www.schneierfacts.com
- Fellow, lecturer at Berkman Center at Harvard Law
- Was CTO of Resilient Systems, Inc
  - Incident response management platform
  - Recently bought by IBM
- Leading security expert
- ***"You are under surveillance right now."***
  - *D&G* provides a comprehensive overview of modern surveillance and privacy.

# Data as a by-product of computing

- Data is a by-product of this high-tech socialization
  - These systems don't just transfer data; they also create data records of your interactions with others

- Metadata
  - "Metadata can be much more revealing than data, especially when collected in the aggregate. When you have one person under surveillance, the contents of conversations, text messages, and emails can be more important than the metadata. But when you have an entire population under surveillance, the metadata is far more meaningful, important, and useful."

- Implications of the cost of all aspects of computing continuous falling over the years
  - Amounts of data that were impractical to store and process a decade ago are easy to deal with today.

# Data as Surveillance

- We are under mass surveillance
  - Governments and corporations gather and analyze data about us
  - Make conclusions from the data they analyze

- “Google knows more about what I’m thinking of than I do, because Google remembers all of it perfectly and forever.”
  - Can look at google search history for any time you logged in
  - And even though Google lets you modify your ad preferences, you have **no rights to delete anything** you don’t want there

# Data as Surveillance

- Different Types of Surveillance
  - Cheaper Surveillance
  - Mass Surveillance
  - Hidden Surveillance
  - Automatic Surveillance
  - Ubiquitous Surveillance

# Big Data and Data Mining

- Government and private approach of “Save everything you can, and someday you’ll be able to figure out a use for it all”
- Modern surveillance vs. traditional surveillance:
  - You can look backwards at a target’s history rather than just track what they do next
- Companies and government agencies store data for years whether or not they have reason to believe it will be useful
- The NSA maps relationships by building networks up to 3-hops away from a target using phone call data
  - They also search the content of communications and match location data to find secret connections that don’t use direct lines of communication

# Correlating Data Sets to Pierce our Anonymity

- Information that is exposed might not be harmful by itself, but by collecting data points over time and correlating different data sets, targets can be identified and tracked
  - “It takes much less data to identify us than we think”
  - “95% of Americans can be identified by name from just four time/date/location points”
- As in the Target-pregnancy scandal, “Data we’re willing to share can imply conclusions we don’t want to share.”
- Staying anonymous on the internet is almost impossible, as any slip up could be the thread that unravels everything
- Large data sets posted by AOL and  Netflix were de-anonymized by researchers

# Internet Surveillance

- Since people were unwilling to pay for services on the early internet, business models evolved around collecting data and advertising
  - Cookies are the primary mechanism -- anonymous, but easily correlated to real info including names
  - “It’s less Big Brother, and more hundreds of tattletale little brothers”
- Data Brokers combined the traditional streams of company records, direct marketing, credit bureaus, and government, allowing divisions into highly specific marketable or vulnerable groups
  - Uncanny Valley of personalized advertising
- The prevalence of advertising is making it less valuable: we are better at tuning it out, and there is only so much commerce to go around
  - There’s the potential for the advertising bubble to burst, as the value of each ad keeps falling

# The Feudal Lords of the Internet

- Modern technology platforms have positioned themselves as powerful middlemen
  - Based on their ubiquity and the speed at which they can pivot to capture new markets
- Feudal Metaphor: we are vassals of large internet corporations, spending our time on their platforms producing the data they sell.
  - They also store our data and ‘protect’ us (or not)
- It is not practical to opt out, only to choose which ‘feudal lord’ we give control of our data to

# Government Surveillance & Control (NSA)

- NSA Background
  - Formed by Truman in 1952 -- entirely for **foreign** intelligence gathering
  - Fall of communism 1980-90s -- **defense** (protect communication from others' spying)
  - After 9/11 -- **"Never again"**
- Most extensive surveillance network in the world
  - **$72B** annual intelligence budget across **17** agencies-- more than the rest of the world *combined*
  - Internet's physical wiring
  - Almost all large tech players based in the US and **subject to US laws**
- Is this Legal?
  - Executive order 12333 (Reagan), Patriot Act (2001), FISA (2008)
  - Many provisions **unconstitutional** (illegal search & seizure) -- debated in courts right now
- The US is not alone: **almost all countries**' intelligence agencies engage in similar acts
  - "US has way more legal controls and restrictions on data collection than any other country"

# Government Surveillance & Control

- Government on Population Surveillance -- everywhere
  - “Incidental” data collection
  - 3rd Party Doctrine → reduced privacy rights
  - Catching criminals  vs dissidents: Ideologically-driven surveillance
- Government on government Espionage -- follows geopolitical lines
  - Hacks: easiest way to eavesdrop (used to be to intercept communications)
    - Tailored Access Group (TAO)
    - China: Google, Canadian government, *NYT*, US military
  - Attacks: deliberate disruption of the target network
    - Infer attacker from list of victims
- Single Global Surveillance Network?
  - Since US has the most extensive network, incentive to form partnerships
    - → NSA has access to almost everything now
  - Single global network? Unlikely
    - Smaller countries likely to tap into such a network

# Consolidation of Institutional Control

- Tapped into corporate technologies to build surveillance network
  - PRISM, etc.: willing and forced ( secret courts, hacking) corporate cooperation
    - Feedback loop → Hard to get effective laws passed to curb corporate surveillance
  - UK (pays Vodaphone), France, China, Russia
    - ISPs required to keep data for few months in case government wants it
  - Some states (OH, TX, FL) sell license & voter registration data to private buyers
- Public-private surveillance partnership -- for- profit corporations
  - 70% of intelligence budget goes to private firms (1,931 firms)
  - Trade shows
  - Built for corporate use, used by developing countries
    - McAfee → Internet censoring in Tunisia and Iran
    - Tech is **value-neutral**

# Governments Subverting Commercial Systems

- 1994: FBI lobbied Congress for CALEA (requires telcos to re-engineer digital switches so eavesdropping is built-in)
  - Currently lobbying to upgrade to cover all communication systems
    - Even video game chat windows
    - Easy to hard problem
      - Gmail: mail unencrypted on Google servers
      - Off the Record: encrypted on user's local device (nearly impossible)
      - Lavabit: gov demanded master encryption keys, but not to tell customers → company shutdown
      - Yahoo (2008): secretly threatened with $250k daily fines if it didn't join PRISM
- Lots of cooperation after 9/11: thought it was patriotic
- Other governments, too
  - China (Huawei)

# Political Liberty & Injustice

1. Accusation by Data
    - with enough data about someone, you can accuse them of *something*
2. Chilling Effect
    - inhibition of free expression and association that occurs as a result of the belief that your communication is being monitored
3. Censorship & Social Change
    - Most censorship is self-censorship
    - People become conformist & compliant when they think they're being observed
    - Due to this conformity, we lose our individuality and society stagnates
    - Flow of Social Change
        - Illegal and not okay → Illegal and not sure → Illegal and maybe okay → Legal
        - Censorship inhibits dissent, which disrupts the flow of social change

# Secrecy, Abuse, & Internet Freedom

- Secrecy is exerted to an extreme degree
  - Branches keep secrets from each other
  - Government deals severely with those who expose its secrets
  - Vast expansion of what is actually considered classified information
- Secrecy weakens checks and balances that oversee surveillance
- All surveillance systems are susceptible to abuse
  - Examples from the NSA & FBI
- Building the means for surveillance system can easily enable abuse
  - Systems can't be built for perfect usage
  - Must be built to withstand abuse
- Internet Freedom should be a human right, led by the US

# Harms of Commercial Surveillance

- Corporate surveillance enables government surveillance
- Redlining: practice of denying or charging more for services by using neighborhood as a proxy for race
  - Weblining
- Problem with price discrimination -- any time we're monitored and profiled, there's the potential for getting it wrong
- Chilling effects
- Surveillance data increases corporate profit at the expense of customers, and will continue as long as:
  - Sellers compete for our money
  - Software makes price discrimination easier
  - Discrimination can be hidden from customers

# Privacy & Business

- In 1993, the  Internet was new and the NSA tried to keep cryptography from foreign countries by limiting its export in U.S. products
- FBI and NSA were worried that they would not be able to eavesdrop on criminal conversations
    - Clipper Chip: allowed backdoor access to phone conversations
    - No sales due to foreign competition from secure products
- Companies are moving data out of the U.S. due to NSA surveillance
- New requirements by companies to contractually promise not to cooperate with government organizations

# Social Privacy

- If you do nothing wrong, you have nothing to hide
  - Privacy is an inherent human right, required for dignity
  - We can lose control of how we present ourselves without privacy
  - Psychological basis for privacy: surveillance is similar to being hunted -> we feel like prey
  - Especially true for minoritized groups (racial, religious, etc.)
- Information now lasts forever as data persists
  - Far reaching societal impact
  - Ex. Should criminal records last forever?
  - Forgetting is an important part of forgiving
- Algorithmic surveillance, as opposed to human surveillance
  - “Collection”: being looked at by a person, not an algorithm
  - NSA hoards all information, only a handful is “collected”
- Analogy: are you worried about a dog watching you?
  - Actions are based off of algorithm determination
  - No assurance that the computer deletes the information
  - Risk of exposure
  - Computer communicates with other people, dog cannot

# Security

- Fear gets in the way of smart security
  - We tend to fixate on rarer but wild occurrences
  - Ex. Getting killed by police is more likely than by terrorists
- Easy to tell story in hindsight
  - After terrorist event, NSA uses the connect the dots argument to advocate for further surveillance
  - Data mining works best when you search for a well defined profile and when there are many events per year
  - For terrorism, false positives overwhelm the system
  - No logic in collecting even more data to somehow find a pattern
  - Each terrorist attack is unique and hard to predict
- Failures of 9/11 were about bad analysis, not inadequate data
- Works in China to find dissidents because there is a well defined profile and many dissidents to track
- Attack vs. Defense: easier to hack security right now
  - Random vs. targeted attacks
  - Defense is inherently stronger mathematically, but the hardware/code/person behind add complications that make offense easier
  - Endpoint security is weak
  - Economics of development prioritize speed to market, not security

# Security and Privacy

- False trade-off
- Privacy is fundamental to security
- Often phrased with leading questions
- Security vs. surveillance
  - Systems can be built for one or the other
  - Ex. Infrastructure, Tor, Airports

# Transparency

- You should know what data is being collected about you
- Transparency of Algorithms
- Half-life of secrets are decreasing
  - Increasing effect
- Cultural Changes
- Oversight and Accountability
- Whistleblowers

# NSA

- Less Secrecy
  - NSA vs Police
- FISA Court and Amendments Act
- Wiretap laws vs Patriot Act
- NSA oversight
- Overstepping bounds
  - Executive Order 12333
  - Section 215 of Patriot Act
  - Section 702 of Fisa Amendments Act

# Solutions for the Government

- Targeted surveillance vs data gathering
  - “The front door governed by law; the backdoor governed by game theory.”
- Information online can still be private
- Fixing Vulnerabilities
- Don’t subvert procedures
  - U.S. has a disproportionate amount of power over the internet
- Espionage vs Surveillance
- Fight Cyber Sovereignty
- Provide for Commons

# OECD Privacy Framework (1980)

Principle of:

- **Collection limitation**
- **Data quality**
- **Purpose specification**
- **Use limitation**
- **Security safeguards**
- **Openness**
- **Individual participation**
- **Accountability**

*Developed by the Organization for Economic Cooperation and Development in 1980 to formalize a structure for how firms might reduce data privacy risks.*

# Solutions for firms

***Stay in line with framework of maintaining security, privacy, and transparency (OECD principles)***

- Hold corporations accountable for breaches in privacy
- Regulate data *use* and data *collection*
- Make do with less data! (or store for shorter periods of time)
  - Waze doesn't need everyone's surveillance data to infer traffic flow
  - Waze only needs your data in real time!
- Make data collection and privacy salient! (in line with transparency)
  - Establish information fiduciaries (Middle Ages Catholic Church)
- Give people rights to their data (standardized in 2012 by White House, see next slide…)

# US Consumer Privacy Bill of Rights (2012)

- Individual control
- Transparency
- Respect for context
- Security
- Access & Accuracy
- Focused collection
- Accountability

# Solutions for we, the users

- Defend against surveillance
  - Avoid
  - Block
  - Distort
  - Break
- Set a personal privacy/convenience threshold
  - “I don’t use Gmail, and I never access my e-mail via the web. I don’t have a personal Facebook account [...] But I do carry a cell phone pretty much everywhere I go, and I collect frequent flier miles whenever possible [...] ***find your own sweet spot***.”
- Work to be an activist for change

# Beating the surveillance society

- People don’t care that sensitive data is being collected and used (still overly scared of terrorism)
  - We underestimate how much surveillance occurs
- We must “modify our feelings” to beat a surveillance society
  - The fear of terrorism dominates the fear of tyranny. Must recalibrate our fear.
  - Recalibrate our privacy. Never before were we able to Google/Facebook stalk. To probe. To create and submit revenge porn. To send nudes.
  - “Don’t wait.”

# Discussion questions...

- Imagine if the data-advertising bubble ‘burst’ -- would you be willing to pay for services that are currently free, in exchange for guaranteed security of your data? Would this only be feasible if your data footprint is ‘wiped clean’ first?
- We discussed the notion of “legal whistleblowing”... Should Snowden have a fair trial?
- What are some challenges to constructing channels of legal whistleblowing?
- How much of the burden should be put on the user to be careful about the data they provide (vs. on the firm-side)?
- Are you for or against firms like 23&Me performing personalized marketing based on user data?
- Is it our responsibility to lobby *more* in order to stop it?