Privacy, Ethics, and Big (Smartphone) Data

ICISSP 2016
February 20, 2016
Smartphones are Pervasive

- Over 1B smartphones sold last year
- 75% penetration in the US as of late 2014, about 25% worldwide
- Over 100B apps downloaded on each of Android and iOS
Fun Facts about Millennials

83% sleep with phones
Fun Facts about Millennials

83% sleep with phones
90% check first thing in morning
Fun Facts about Millennials

83% sleep with phones
90% check first thing in morning
1 in 3 use in bathroom
Smartphone Data is Intimate

Who we know (contacts + call log)

Where we go (gps, photos)

Sensors (accel, sound, light)
The Opportunity

- We are creating a worldwide sensor network with our smart devices
- We can now capture and analyze human behavior at unprecedented fidelity and scale

http://www.flickr.com/photos/robby_van_moor/478725670/
Talk Overview

• What might this brave new world be like?
  – What kinds of amazing opportunities?

• What kinds of privacy challenges will we face?
  – What kinds of new approaches for privacy?
  – Today’s notice and choice won’t work
Talk Overview

• Story 1: Livehoods Urban Analytics
  – How can we use geotagged data to understand cities?
  – Privacy issues about inferences and uses of data

• Story 2: PrivacyGrade
  – How can we help people understand what their apps are really doing?
  – Why are developers collecting so much data?

• Reflections on improving the privacy ecosystem
Story 1: The Challenge of Getting Data About Our Cities

- Today’s methods for getting city data slow, costly, limited
  - Ex. Travel Behavioral Inventory
  - US Census 2010 cost $13b
  - Quality of life surveys
- Some approaches today:
  - Call Data Records
  - Deploy a custom app
The Vision: Urban Analytics

- How can we use smartphones + social media + machine learning to offer new and useful insights about the dynamics of cities?
Understanding Urban Areas

• AT&T Work on Human Mobility

Median distance traveled per day
Eric Fischer’s Maps of Tourists vs Locals
Livehoods
Our First Urban Analytics Tool

- The character of an urban area is defined not just by the types of places found there, but also by the people that make it part of their daily life.

Cranshaw et al, The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City, ICWSM 2012.
Two Perspectives on Cities

“Politically constructed”

Neighborhoods have fixed borders defined by the city government

“Socially constructed”

Neighborhoods are organic, cultural artifacts. Borders are blurry, imprecise, and may be different to different people
Two Perspectives on Cities

Can we discover automated ways of identifying the “organic” boundaries of the city?

Can we extract local cultural knowledge from social media?

Can we build a collective cognitive map from data?

“Socially constructed”

Neighborhoods are organic, cultural artifacts. Borders are blurry, imprecise, and may be different to different people.
Livehoods Data Source

• Crawled 18m check-ins from foursquare
  – Claims 20m users
  – People who linked their foursquare accts to Twitter
• Spectral clustering based on geographic and social proximity
If you watch check-ins over time, you’ll notice that groups of like-minded people tend to stay in the same areas.
We can aggregate these patterns to compute relationships between check-in venues.
These relationships can then be used to identify natural borders in the urban landscape.
We call the discovered clusters “Livehoods” reflecting their dynamic character.
Welcome to Livehoods!

Each dot on the map (●) represents a check-in location. Groups of nearby dots of the same color form a Livehood.

The shapes of Livehoods are determined by the patterns of people that check-in to them. If many of the same people check-in to two nearby locations, then these locations will likely be part of the same Livehood.

Livehoods reveal how the people and places of a city come together to form the dynamic character of local urban areas.

Click on a location to learn about its Livehood.

Try it out at livehoods.org
South Side Pittsburgh
“Whenever I was living down on 15th Street [LH7] I had to worry about drunk people following me home, but on 23rd [LH8] I need to worry about people trying to mug you... so it’s different. It’s not something I had anticipated, but there is a distinct difference between the two areas of the South Side.”
There is this interesting mix of people there I don’t see walking around the neighborhood. I think they are coming to the Giant Eagle [grocery store] from lower income neighborhoods... I always assumed they came from up the hill.”
South Side Pittsburgh

“I always assumed they came from up the hill.”
South Side Pittsburgh
Other Urban Analytics

- Whole Foods Market
  Number Users: 4.6K
  Repeat proportion: 0.8

- Disneyland Park
  Number Users: 58.9K
  Repeat proportion: 0.6

- San Francisco International Airport (SFO)
  Number Users: 145K
  Repeat proportion: 0.7

- Yankee Stadium
  Number Users: 42.9K
  Repeat proportion: 0.6

- Philadelphia International Airport (PHL)
  Number Users: 79.6K
  Repeat proportion: 0.6

- The White House
  Number Users: 23K
  Repeat proportion: 0.2
Topic Modeling (LDA)

- Tourists – far away point is airport
- The “music” cluster
- The GayBar cluster
Geotagged Tweets in Pittsburgh
Geotagged Tweets in Pittsburgh
Geotagged Tweets in Pittsburgh

jssaylor: Picking up a sweeper I found on Craigslist. Might get murdered. (@greenfield in Pittsburgh, PA)
https://t.co/TwsMdTbpwk

tully_412: So the cops showed up at the house party last night. Guess we were too loud, oops. Sorry greenfield 🙁
tully_412: It’s kinda sad that I can’t walk down the street to the mail box in shorts without getting honked at. :/ Show some class, greenfield. #Pgh

aofesh: Skype Fred is better than no Fred at all. #Fredro #catsofinstagram #cats @greenfield, Pittsburgh…
https://t.co/IdRqa9FB2h

cvciking82: Happy Easter with the fam. @greenfield https://t.co/ORFqghwRo
BigSnuke: Anyone else in greenfield lose their @comcast the last two nights right around 8:30PM? Or is it just our lucky few blocks?
BigSnuke: @brianmcnulty11 not tryin to put it on front street but your honesty is real refreshing. I’m from greenfield so I’ve seen it in action

NickyBelke: Our first Memorial Day picnic was a success 😊 @greenfield, Pittsburgh, Pennsylvania
https://t.co/prMpLMydRD

NickyBelke: My favorites ❤️ @greenfield, Pittsburgh, Pennsylvania https://t.co/1DdT2pbSJ

cvciking82: Pup’s 76th birthday😊 #tornamore https://t.co/0lfZ71156
jssaylor: When you’re too tired to cook, there’s always the liquid dinner. #wine @greenfield https://t.co/QWXA0pqkrk

Monster_C_Lane: New morning view on fleek #dafaq #pittsburgh #niceyellowsuzuki @greenfield, Pittsburgh, Pennsylvania https://t.co/CnyX4dUxIH
EmojiMap of Pittsburgh
EmojiMap of Pittsburgh
Reflections on Urban Analytics
Few Privacy Issues on the Surface

• Publicly visible data with no expectations of privacy
  – No formal IRB issues

• Removed venues labeled as “home”
  – We only received one request to remove a venue (wasn’t labeled as a home)

• We only show data about geographic areas vs individuals

• So far, so good for privacy, but...
Reflections on Urban Analytics
But Lots of Deeper Questions

• Many discussions on how data like this *might* be used negatively in the future
  – Urban planners, Yahoo, Facebook, Google Maps, Startups, and more

• Lots of hard questions
  – Not so many great answers
  – Representative of many challenges we will face with lots of rich data
How Much Can Be Inferred?

- Tourists – far away point is airport
- The “music” cluster
- The GayBar cluster
How Much Can Be Inferred?

• Much more can probably be inferred using rich data like this
  – Demographics, socioeconomic, friends
  – Physical and mental health (depression)
  – How “risky” you are (bars, clinics, etc)

• Unclear how far inferencing can go
  – Very hard to detect (and thus to regulate)
  – Not much can stop advertisers, NSA, GCHQ, startups
  – Even if you disclose little, lots can be inferred based on what others similar to you (homophily)
How Much Can Be Inferred?

Built a logistic regression to predict sexuality based on what your friends on Facebook disclosed.
“[An analyst at Target] was able to identify about 25 products that... allowed him to assign each shopper a ‘pregnancy prediction’ score. [H]e could also estimate her due date to within a small window, so Target could send coupons timed to very specific stages of her pregnancy.” (NYTimes)
How Much Can Be Inferred?

**More Risk**

Chrome-skull accessories were in the top 1 percent of products signaling a risk of default among 85,000 types of purchases analyzed.

**Less Risk**

Premium wild birdseed was in the bottom 1 percent of products signaling a risk of default among 85,000 types of purchases analyzed.
A New Kind of Redlining?

• “denying, or charging more for, services such as banking, insurance, access to health care, ... supermarkets, or denying jobs ... against a particular group of people” (Wikipedia)
Huge Risk of New Kind of Redlining

Map of Philadelphia showing redlining of lower income neighborhoods. Households and businesses in the red zones could not get mortgages or business loans. (Wikipedia)
Johnson says his jaw dropped when he read one of the reasons American Express gave for lowering his credit limit:

"Other customers who have used their card at establishments where you recently shopped have a poor repayment history with American Express."
Recap So Far

• More rich data about people than ever before
• Can do lots of inferencing of people’s behaviors
  – Hard to detect this inferencing
  – Homophily very powerful
• Already leading to problematic behaviors by companies
  – Others: Insurance, employee hiring, banking?
• Hold thoughts for now, will get to reflections and recommendations at the end
Talk Overview

• Story 1: Livehoods

• Story 2: PrivacyGrade
  – How can we help people understand what data their apps are collecting?
  – Why are developers collecting so much data?

• Reflections on improving the privacy ecosystem
Story 2: PrivacyGrade
or, What Are Your Apps Really Doing?

Pandora
Shares your location, gender, unique phone ID, phone# with advertisers

Pinterest

Facebook
Uploads your entire contact list to their server (including phone #s)
Many Smartphone Apps Have “Unusual” Permissions

- Location Data
- Microphone
- Unique device ID

- Location Data
- Network Access
- Unique device ID

- Location Data
- Unique device ID
PrivacyGrade.org

PrivacyGrade: Grading The Privacy Of Smartphone Apps

We're a team of researchers from Carnegie Mellon University. We have assigned privacy grades to Android apps based on some techniques we have developed to analyze to their privacy-related behaviors. Learn more here or browse our analyzed apps.

- Improve transparency
- Assign privacy grades to all 1M+ Android apps
Drag Racing

Developer: Creative Mobile
Category: Game Racing

Privacy Grade: Poor D

Related Apps
- Drag Racing
- Drag Racing

App Description

The following description comes from the Google Play Store description of the app:

- Drive 50+ officially licensed cars, from hot hatches to american muscle and 1000HP supercars
- Buy your dream car, install performance upgrades and show your skills in 1/4 or 1/2 mile races
- Challenge millions of players online: race 1 on 1, drive your opponent's

Read More

Privacy Analysis

App was last analyzed by Privacy Grade on:
08/19/2014

Why does this app have this grade?
Our method for grading apps uses a privacy model that we built. This model is based on crowdsourced surveys that we conducted to capture people's expectations and comfort levels with various app behaviors.

You can see more information about how we grade apps in our...
<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>App Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Can use user's phone id or phone number</td>
<td>It appears this app uses this data to identify users for market/customer analysis</td>
</tr>
<tr>
<td></td>
<td>Can use user's phone id or phone number</td>
<td>It appears this app uses this data to connect with independent/secondary mobile app stores</td>
</tr>
<tr>
<td></td>
<td>Can use user's phone id or phone number</td>
<td>It appears this app uses this data to identify users for delivering targeted advertisement</td>
</tr>
<tr>
<td></td>
<td>Full network access</td>
<td>Can access the Internet</td>
</tr>
<tr>
<td></td>
<td>Can access the Internet</td>
<td>It appears this app uses this data for internal use within the app's functionality</td>
</tr>
<tr>
<td></td>
<td>Can access the Internet</td>
<td>It appears this app uses this data for mobile analytics</td>
</tr>
<tr>
<td></td>
<td>Can access the Internet</td>
<td>It appears this app uses this data for a secondary app market</td>
</tr>
<tr>
<td></td>
<td>Can access the Internet</td>
<td>It appears this app uses this data for targeted advertising</td>
</tr>
<tr>
<td></td>
<td>Can access the Internet</td>
<td>It appears this app uses this data for utility purposes</td>
</tr>
<tr>
<td></td>
<td>Precise location (gps and network-based)</td>
<td>Can use user's precise location</td>
</tr>
<tr>
<td></td>
<td>Can use user's precise location</td>
<td>It appears this app uses this data for market/customer analysis</td>
</tr>
<tr>
<td></td>
<td>Can use user's precise location</td>
<td>It appears this app uses this data for advertising</td>
</tr>
</tbody>
</table>

See the sensitive permissions table below the app description to see the permissions that made users of the app uncomfortable.

If you have issues about the information on this page, click the button below to send us feedback.
<table>
<thead>
<tr>
<th>Permission</th>
<th>Functionality</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find accounts on the device</td>
<td>Can use user’s accounts information stored on the phone</td>
<td>It appears this app uses this data to connect with independent/secondary mobile app stores</td>
</tr>
<tr>
<td>Retrieve running apps</td>
<td>Can retrieve information on currently and recently running apps on the device</td>
<td>It appears this app uses this data for a secondary app market</td>
</tr>
<tr>
<td>Approximate location (network-based)</td>
<td>Can use user’s approximate location</td>
<td>Not analyzed yet</td>
</tr>
<tr>
<td>Modify or delete the contents of your USB storage</td>
<td>Can write to your phone’s USB storage</td>
<td>Not analyzed yet</td>
</tr>
</tbody>
</table>

### THIRD PARTY LIBRARIES USED BY THIS APP

<table>
<thead>
<tr>
<th>LIBRARY</th>
<th>USAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flurry</td>
<td>Mobile analytics libraries allow app developers to gather usage data about their app. It might gather data on how often an app user uses the app or in what ways is the app being used. Say for example usage data on a map app could be exemplified by a user using the map app twice a day in the morning and evening to find the quickest way to and from work with regards of the effects of traffic on certain roads.</td>
</tr>
<tr>
<td>Chartboost</td>
<td>Targeted ad libraries allows developers to monetize their app by allowing their app to serve advertisements to app users.</td>
</tr>
<tr>
<td>Millennialmedia</td>
<td>Targeted ad libraries allows developers to monetize their</td>
</tr>
<tr>
<td>LIBRARY</td>
<td>USAGE</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Flurry</td>
<td>Mobile analytics libraries allow app developers to gather usage data about their app. It might gather data on how often an app user uses the app or in what ways is the app being used. Say for example usage data on a map app could be exemplified by a user using the map app twice a day in the morning and evening to find the quickest way to and from work with regards of the effects of traffic on certain roads.</td>
</tr>
<tr>
<td>Chartboost</td>
<td>Targeted ad libraries allows developers to monetize their app by allowing their app to serve advertisements to app users.</td>
</tr>
<tr>
<td>Millennialmedia</td>
<td>Targeted ad libraries allows developers to monetize their app by allowing their app to serve advertisements to app users.</td>
</tr>
<tr>
<td>Sponsorpay</td>
<td>Targeted ad libraries allows developers to monetize their app by allowing their app to serve advertisements to app users.</td>
</tr>
<tr>
<td>Amazon</td>
<td>Utility libraries provide app developers tools to add useful functionality to app that may be more for aiding the developer rather than the end user. They can be authentication frameworks, or logging frameworks.</td>
</tr>
<tr>
<td>Getjar</td>
<td>Secondary market libraries allow users to connect to other app markets.</td>
</tr>
</tbody>
</table>
Expectations vs Reality
Privacy as Expectations

Use crowdsourcing to compare what people *expect* an app to do vs what an app *actually* does

App Behavior
(What an app actually does)

User Expectations
(What people think the app does)
How PrivacyGrade Works

• We crowdsourced people’s expectations of core set of 837 apps
  – Ex. “How comfortable are you with Drag Racing using your location for ads?”

• We generated purposes by examining what third-party libraries used by app

• Created a model to predict people’s likely privacy concerns and applied to 1M Android apps
How PrivacyGrade Works
How PrivacyGrade Works

<table>
<thead>
<tr>
<th>THIRD PARTY LIBRARY NAME</th>
<th>CATEGORY</th>
<th>NO. Of APPS USING THIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Admob</td>
<td>Targeted ads</td>
<td>121944</td>
</tr>
<tr>
<td>Facebook</td>
<td>Social networking service</td>
<td>107131</td>
</tr>
<tr>
<td>Slf4j</td>
<td>Utility</td>
<td>70227</td>
</tr>
<tr>
<td>Actionbarsherlock</td>
<td>Utility</td>
<td>67890</td>
</tr>
<tr>
<td>Millennialmedia</td>
<td>Targeted ads</td>
<td>60437</td>
</tr>
<tr>
<td>Inmobii</td>
<td>Targeted ads</td>
<td>57116</td>
</tr>
<tr>
<td>Twitter4j</td>
<td>Social networking service</td>
<td>54065</td>
</tr>
<tr>
<td>Flurry</td>
<td>Mobile analytics</td>
<td>47129</td>
</tr>
</tbody>
</table>

- Long tail distribution of libraries
- We focused on top 400 libraries
Overall Stats on PrivacyGrade
April 2015

• No sensitive permissions used means A+
• Other grades set at quartiles of grade range
# Changes in Grades Over Time

October 2014 to April 2015

<table>
<thead>
<tr>
<th>Previous Grade (Oct 2014)</th>
<th>Current Grade (Apr 2015)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A+</td>
</tr>
<tr>
<td>A+</td>
<td>99.20</td>
</tr>
<tr>
<td>A</td>
<td>0.03</td>
</tr>
<tr>
<td>B</td>
<td>0.03</td>
</tr>
<tr>
<td>C</td>
<td>0.01</td>
</tr>
<tr>
<td>D</td>
<td>0.00</td>
</tr>
</tbody>
</table>
### Changes in Grades Over Time

**Most Grades Remained the Same**

<table>
<thead>
<tr>
<th>Previous Grade (Oct 2014)</th>
<th>Current Grade (Apr 2015)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A+</td>
</tr>
<tr>
<td>A+</td>
<td>99.20</td>
</tr>
<tr>
<td>A</td>
<td>0.03</td>
</tr>
<tr>
<td>B</td>
<td>0.03</td>
</tr>
<tr>
<td>C</td>
<td>0.01</td>
</tr>
<tr>
<td>D</td>
<td>0.00</td>
</tr>
</tbody>
</table>
## Changes in Grades Over Time

A Fair Number of Apps Improved

<table>
<thead>
<tr>
<th>Previous Grade (Oct 2014)</th>
<th>A+</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A+</td>
<td>99.20</td>
<td>0.79</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>A</td>
<td>0.03</td>
<td>99.80</td>
<td>0.11</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>B</td>
<td>0.03</td>
<td>10.22</td>
<td>83.52</td>
<td>5.99</td>
<td>0.24</td>
</tr>
<tr>
<td>C</td>
<td>0.01</td>
<td>2.15</td>
<td>1.62</td>
<td>95.93</td>
<td>0.29</td>
</tr>
<tr>
<td>D</td>
<td>0.00</td>
<td>0.31</td>
<td>0.22</td>
<td>0.91</td>
<td>98.56</td>
</tr>
</tbody>
</table>
### Changes in Grades Over Time

Lots of Apps Deleted

<table>
<thead>
<tr>
<th>Latest Grade</th>
<th>Number of Apps Removed</th>
<th>Percentage of Apps with Same Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>A+</td>
<td>24463</td>
<td>30.6%</td>
</tr>
<tr>
<td>A</td>
<td>359490</td>
<td>37.7%</td>
</tr>
<tr>
<td>B</td>
<td>17424</td>
<td>39.1%</td>
</tr>
<tr>
<td>C</td>
<td>26809</td>
<td>53.1%</td>
</tr>
<tr>
<td>D</td>
<td>24089</td>
<td>53.4%</td>
</tr>
</tbody>
</table>

- Not sure why deleted yet
  - Some apps were re-uploaded
Impact of this Research

• Popular Press
  – NYTimes, CNN, BBC, CBS, more
  – Consumer Reports interest

• Government
  – Earlier work helped lead to FTC fines

• Google
  – Google replicating PrivacyGrade internally

• Developers

In light of the results from a recent study by PrivacyGrade, Halfbrick is going to still greater lengths to inform players about what data the game is collecting during the in-game experience and why.
What Do Developers Know about Privacy?

• Interviews with 13 app developers
• Surveys with 228 app developers

• What tools? Knowledge? Incentives?
• Points of leverage?

Summary of Findings
Third-party Libraries Problematic

• Use ads and analytics to monetize
Summary of Findings
Third-party Libraries Problematic

• Use ads and analytics to monetize
• Hard to understand library behaviors
  – A few didn’t know they were using libraries (inconsistent answers)
  – Some didn’t know they collected data
  – “If either Facebook or Flurry had a privacy policy that was short and concise and condensed into real English rather than legalese, we definitely would have read it.”
Summary of Findings
Developers Don’t Know What to Do

• Low awareness of existing privacy guidelines
  – Often just ask others around them

• Low perceived value of privacy policies
  – Mostly protection from lawsuits
  – “I haven’t even read [our privacy policy]. I mean, it’s just legal stuff that’s required, so I just put in there.”
Developers Need a Lot of Help!

• Good set of best practices for security
  – SSL, hashing of passwords, randomization
  – Common attacks: SQL, XSS, CSRF

• What are best practices for privacy??

• Developers also have many problems
  – App functionality, bandwidth, power, making money... privacy pretty far down the list
Talk Overview

• Story 1: Livehoods
• Story 2: PrivacyGrade
• Reflections on improving the privacy ecosystem
Reflections on Privacy
Consider Entire Ecosystem for Privacy

• I’m increasingly skeptical of end-user privacy
  – Places too much burden on people
  – Lets developers get away with too much

• Instead, leverage rest of ecosystem
  – Developers
  – Third-party developers
  – App stores
  – OS and H/W manufacturers
  – Regulators
  – Third party advocates
  – Journalists
Reflections on Privacy
Developers Really Need Help!

• Long tail of developers
  – Most developers just don’t know what to do
  – Also very easy to make mistakes

• Developers need better tools
  – PrivacyGrade: Upload your app to our site, we analyze and offer suggestions to fix
  – Understand what a given library does
  – Make it easy to get only minimal data
    • Don’t need your location, just if at “work” or “home”
    • Don’t need microphone, just level of sound
Reflections on Privacy
Developers Really Need Help!

• Developers need canonical online resources
  – Establish Virtual Centers of Excellence
  – Good programming examples
  – Good design patterns and UI examples
  – Good best practices and processes
  – If a developer has a privacy question, make it much easier for them solve it
Reflections on Privacy
Other Parts of the Ecosystem

• Third-party developers
  – If most data collected by third parties, push to restrict (either app store, OS, regulators)

• Hardware manufacturers
  – H/W manufacturers care a lot about privacy
  – Try to get them to adopt privacy software

• Third party advocates
  – Better tools to help them analyze
Reflections on Privacy
Other Parts of the Ecosystem

• Regulators
  – Often lawyers without strong technical support
  – Ex. Forthcoming work on identifying apps for kids

• Journalists
  – Shame can be very powerful
  – Angry Birds stopped collecting location data after multiple popular press about their app

• Earlier, hard to detect undesired uses
  – These are possible points of leverage
How can we create a connected world we would all want to live in?
Thanks!

More info at cmuchimps.org or email jasonh@cs.cmu.edu

• Shah Amini
• Justin Cranshaw
• Afsaneh Doryab
• Jialiu Lin
• Song Luan

Special thanks to:
• DARPA
• Google
• CMU Cylab
• Army Research Office
• NSF
• Alfred P. Sloan
• NQ Mobile
Bezerkeley, CA
Few people read privacy policies

- We want to install the app
- Reading policies not part of main task
- Complexity (the pain!!!)
- Clear cost (time) for unclear benefit
Story 2: Google Glass

• Why there has been so much negative backlash about Google Glass?
• Are there lessons we can learn here about privacy and adoption of tech?
Origins of Ubicomp

Mark Weiser (1952-1999)
The Founder of Ubiquitous Computing

Mark Weiser was the chief technology officer at Xerox’s Palo Alto Research Center (PARC). He is often referred to as the father of ubiquitous computing. He coined the term in 1988 to describe a future in which invisible computers, embedded in everyday objects, replace PCs. Other research interests included garbage collection, operating systems, and user interface design. He received his SIA and Ph.D. in computer and communications science at the University of Michigan, Ann Arbor. After completing his Ph.D., he joined the computer science department at the University of Maryland, College Park, where he taught for 32 years. He wrote or cowrote over 75 technical publications on such subjects as the psychology of programming, program sharing, operating systems, programming environments, garbage collection, and technological ethics. He was a member of the ACM, IEEE Computer Society, and American Association for the Advancement of Science. Weiser passed away in 1999. Visit www.parc.xerox.com/about/; members/ewebserver or contact communications@parc.xerox.com for more information about him.

The Computer for the 21st Century

Specialized elements of hardware and software, connected by wires, radio waves and infrared, will be so ubiquitous that no one will notice their presence.

The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life, and they are indistinguishable from it.

Consider writing, perhaps the true information technology. The ability to express spoken language symbolically for long-term storage and retrieval of information from the limits of individual memory. Today, this technology is ubiquitous in我们必须 know and use.

Not only does books, magazines, newspapers, television, radio, and the Internet contain information, but so do signs, billboards, shop signs and store signs, street signs and the signs we are now.

The constant presence of these pieces of technology and the active attention it demands, is critical to be remembered at every level.

It is difficult to imagine modern life in a world where books have become part of our daily lives.

Most of the information we need and access is not through the printed page, but through the symbols we create and the technology that allows us to interact with them.

By Mark Weiser

Reprinted with permission. Copyright 1991 by Scientiﬁc American, Inc. All right reserved.
Popular Press Reaction

Orwellian Dream Come True: A Badge That Pinpoints You
By LEONARD SLOANE
Published: September 12, 1992

Is Big Brother right under our noses, or should we be concerned about the future of personal privacy? The first badge-sized personal computer, developed by Dr. John Jarvi of the Electronic Communications Laboratory in Menlo Park, Calif., promises to revolutionize the way we think about personal identity. The device, which can be worn on a belt or tucked behind a shirt collar, contains a mini-version of a standard personal computer. It is used to store information about the wearer and to provide a backup in the event of a computer failure.

Businessweek Archives
Big Brother, Pinned To Your Chest
August 16, 1992

Top of the News

BIG BROTHER
Alarms do not sound near the Laboratory

You're Not Paranoid: They Really Are Watching You

Surveillance in the workplace is getting digitized - and getting worse.
Why Such Negative Press?

• PARC knew privacy a big problem
  – But didn’t know what to do
  – So didn’t build any privacy protections

• Unclear value proposition
  – Focused on technical aspects
  – What benefits to end-users?

• Google Glass is replaying the past
Why Such Negative Press?

• Grudin’s law
  – Why does groupware fail?
  – “When those who benefit are not those who do the work, then the technology is likely to fail, or at least be subverted”

• Privacy corollary
  – When those who bear the privacy risks do not benefit in proportion to the perceived risks, the tech is likely to fail
But Expectations Can Change

- Initial perceptions of mobile phone users
  - Rude, annoying
  - Casual chat, driving
- Six weeks later...
  - Had same behaviors
- People with more exposure to mobile phones better
But Expectations Can Change

• Famous 1890 article defining privacy as “the right to be let alone” was about photography
But Expectations Can Change

- People objected to having phones in their homes because it “permitted intrusion... by solicitors, purveyors of inferior music, eavesdropping operators, and even wire-transmitted germs”
But Expectations Can Change

One resort felt the trend so heavily that it posted a notice: “PEOPLE ARE FORBIDDEN TO USE THEIR KODAKS ON THE BEACH.” Other locations were no safer. For a time, Kodak cameras were banned from the Washington Monument.

The “Hartford Courant” sounded the alarm as well, declaring the “the sedate citizen can’t indulge in any hilariousness without the risk of being caught in the act and having his photograph passed around among his Sunday School children.”
And Framing **Really** Matters

- Ubicomp -> Invisible Computing
  - Talked less about the tech
  - More about how it could help people
  - More positive press
Privacy Hump

- A lot of our concerns about tech fall under umbrella term “privacy”
  - Value, fears, expectations, what others around us think

Many legitimate concerns
Many alarmist rants
“Right” way to deploy?
Value proposition?
Rules on proper use?

Things have settled down
Few fears materialized
People feel in control
People get tangible value
But How to Get Over Hump?

- Still a big gap in knowledge on best ways of mitigating privacy issues
- Prime example: Facebook news feed
“We’d put an ad for a lawn mower next to diapers. We’d put a coupon for wineglasses next to infant clothes. That way, it looked like all the products were chosen by chance.”
• Privacy Placebos: Things that make people feel better about privacy, but doesn’t offer much.
• Other examples: Privacy policies, access logs
• How ethical is it to use these approaches?
Understanding Individuals

Min et al, Toss ‘n’ Turn: Smartphone as Sleep and Sleep Quality Detector, CHI 2014.
Who Gains From this Data?

• Today, most data only flows one way
  – Mainly to advertisers
  – Also banks, insurance, credit cards
• Little value goes back to people or society
Who Gains From this Data?

• Can we design systems that share the value across more people?
  – People co-create data and gain value
  – Participatory design philosophy

• Can we also make people feel more invested in the cities they live in?
Stronger Emphasis on Getting Our Research Results Out There

• We are great at science, terrible at transfer

• Five major factors for successful innovations:
  – Relative Advantage
  – Trialability
  – Complexity
  – Compatibility
  – Observability