Cameron Kerry

Visiting Scholar

MIT Media Lab

Human Dynamics
Big Data Privacy Workshop
Advancing the State of the Art in Technology and Practice

March 3, 2014 | Cambridge, Massachusetts

Big Data Privacy: Advancing the State of the Art in Technology and Practice
Organized by the MIT Big Data Initiative at CSAIL and the MIT Information Policy Project

The White House Office of Science and Technology Policy (OSTP) and MIT co-hosted a public workshop entitled “Big Data Privacy: Advancing the State of the Art in Technology and Practice” on March 3, 2014. The event was part of a series of workshops on big data and privacy organized by the MIT Big Data Initiative at CSAIL and the MIT Information Policy Project. The workshop was also the first in a series of events being held across the country in response to President Obama’s call for a review of privacy issues in the context of increased digital information and the computing power to process it.

The workshop convened key stakeholders and thought leaders from across academia, government, industry, and civil society for a thoughtful dialogue on the future role of technology in protecting and managing privacy.

Speakers included:
- MIT President Rafael Reif
- White House Counselor John Podesta (Keynote Speaker)
- Secretary of Commerce Penny Pritzker (Keynote Speaker)
- Cynthia Dwork, Microsoft Research
- Shafi Goldwasser, MIT CSAIL
- Michael Stonebraker, MIT CSAIL

The agenda page includes video clips of each speaker and selected slide presentations.

MIT would like to acknowledge the generous support of The Alfred P. Sloan Foundation in making this event possible.
MIT Information Policy Project

The Internet is now the central nervous system of our global economy and essential infrastructure for communication, commerce, and civic discourse. Yet at this transformative moment, many important public debates concerning information policy occur without adequate technical understanding and scholarship. The new MIT Information Policy Project seeks to fill this gap with technically-informed research and politically-engaged dialogue, aimed at guiding Internet policymakers around the world.

The Information Policy Project will build on MIT’s proven approach to engineering research and education: We view policy development, like technology development, as a design discipline that should be driven by methodical study. The Initiative will take a long-term, foundational approach to public policy, with a five- to ten-year horizon. We will research options from across the political spectrum, with attention to the impacts that policy and technical choices have on innovation, speech, economic development, democracy and human rights. In the process, the Initiative will train a new generation of technology policy leaders in government, civil society, academia and industry.

In its first three years, the Project will tackle research challenges such as:
The Human Dynamics Lab at the MIT Media Laboratories pioneered the idea of a society enabled by Big Data. The Lab has developed technologies such as reality mining, which uses mobile phone data to extract patterns that predict future human behavior, a ‘nervous system’ framework for dramatically more efficient transportation, health, energy, and financial systems, the New Deal on Data policies which are now enshrined in the US Consumer Privacy Bill of Rights, and a Trust Network communication architecture that ensures that this new data driven society is secure and fair.
CONSUMER DATA PRIVACY IN A NETWORKED WORLD:
A FRAMEWORK FOR PROTECTING PRIVACY AND PROMOTING INNOVATION IN THE GLOBAL DIGITAL ECONOMY

FEBRUARY 2012
February 23, 2012

Americans have always cherished our privacy. From the birth of our republic, we assured ourselves protection against unlawful intrusion into our homes and our personal papers. At the same time, we set up a postal system to enable citizens all over the new nation to engage in commerce and political discourse. Soon after, Congress made it a crime to invade the privacy of the mails. And later we extended privacy protections to new modes of communications such as the telephone, the computer, and eventually email.

Justice Brandeis taught us that privacy is the "right to be let alone," but we also know that privacy is about much more than just solitude or secrecy. Citizens who feel protected from misuse of their personal information feel free to engage in commerce, to participate in the political process, or to seek needed health care. This is why we have laws that protect financial privacy and health privacy, and that protect consumers against unfair and deceptive uses of their information. This is why the Supreme Court has protected anonymous political speech, the same right exercised by the pamphleteers of the early Republic and today's bloggers.

Never has privacy been more important than today. In the age of the Internet, the World Wide Web and smart phones. In just the last decade, the Internet has enabled a renewal of direct political engagement by citizens around the globe and an explosion of commerce and innovation creating jobs of the future. Much of this innovation is enabled by novel uses of personal information. So, it is incumbent on us to do what we have done throughout history: apply our timeless privacy values to the new technologies and circumstances of our times.

I am pleased to present this new Consumer Privacy Bill of Rights as a blueprint for privacy in the information age. These rights give consumers clear guidance on what they should expect from those who handle their personal information, and set expectations for companies that use personal data. I call on these companies to begin immediately working with privacy advocates, consumer protection enforcement agencies, and others to implement these principles in enforceable codes of conduct. My Administration will work to advance these principles and work with Congress to put them into law. With this Consumer Privacy Bill of Rights, we offer to the world a dynamic model of how to offer strong privacy protection and enable ongoing innovation in new information technologies.

One thing should be clear, even though we live in a world in which we share personal information more freely than in the past, we must reject the conclusion that privacy is an outdated value. It has been at the heart of our democracy from its inception, and we need it now more than ever.
The Consumer Privacy Bill of Rights

INDIVIDUAL CONTROL
TRANSPARENCY
RESPECT FOR CONTEXT
SECURITY
ACCESS AND ACCURACY
FOCUSED COLLECTION
ACCOUNTABILITY
Consumer Privacy Bill of Rights and Big Data: 
Response to White House Office of Science and Technology Policy 
Request for Information

April 4, 2014 - submitted to bigdata@ostp.gov

Daniel J. Weitzner, MIT Computer Science and Artificial Intelligence Lab 
Hal Abelson, MIT Department of Electrical Engineering and Computer Science 
Cynthia Dwork, Microsoft Research 
Cameron Kerry, MIT Media Lab 
Daniela Rus, MIT Computer Science and Artificial Intelligence Lab 
Sandy Pentland, MIT Media Lab 
Salil Vadhan, Harvard University

I. Introduction and Overview

In response to the White House Office of Science and Technology Policy Request for Information on Big Data Privacy we offer these comments based on presentations and
Consumer Privacy Bill of Rights

Individual Control
Transparency
Respect for Context
Security
Access and Accuracy
Focused Collection
Accountability

INDIVIDUAL CONTROL:
Consumers have a right to exercise control over what personal data companies collect from them and how they use it.
Consumer Privacy Bill of Rights

Individual Control

Transparency

Respect for Context

Security

Access and Accuracy

Focused Collection

Accountability

TRANSPARENCY:
Consumers have a right to easily understandable and accessible information about privacy and security practices.
Consumer Privacy Bill of Rights

Individual Control

Transparency

Respect for Context

Security

Access and Accuracy

Focused Collection

Accountability

RESPECT FOR CONTEXT:
Consumers have a right to expect that companies will collect, use, and disclose personal data in ways that are consistent with the context in which consumers provide the data.
Consumer Privacy Bill of Rights

Individual Control
Transparency
Respect for Context

Security
Access and Accuracy
Focused Collection
Accountability

SECURITY:
Consumers have a right to secure and responsible handling of personal data.
Consumer Privacy Bill of Rights

Individual Control
Transparency
Respect for Context
Security
Access and Accuracy
Focused Collection
Accountability

ACCESS AND ACCURACY:
Consumers have a right to access and correct personal data in usable formats, in a manner that is appropriate to the sensitivity of the data and the risk of adverse consequences to consumers if the data is inaccurate.
Consumer Privacy Bill of Rights

Individual Control
Transparency
Respect for Context
Security
Access and Accuracy
Focused Collection
Accountability

FOCUSED COLLECTION:
Consumers have a right to reasonable limits on the personal data that companies collect and retain.
Consumer Privacy Bill of Rights

Individual Control
Transparency
Respect for Context
Security
Access and Accuracy
Focused Collection
Accountability

ACCOUNTABILITY:
Consumers have a right to have personal data handled by companies with appropriate measures in place to assure they adhere to the Consumer Privacy Bill of Rights.
openPDS

Design Principles

- **Interoperability**: All of the data is your store in a well-defined format

- **Auditability**: User can verify what data is read and how it is used.

- **Distributed Computing**:  
  - No single point of attack  
  - Leverages data from multiple PDSs in a distributed and privacy-preserving fashion.  
  - Can reside on a machine the user legally owns
Current Transparency Mechanisms

Detailed Transparency model access to both **hardware** and **software**-based functionality as well as **read** and **write** access to information stored on the phone.

Semi-Transparency model access to both **hardware** and **software**-based functionality. Access to information stored on the phone.

Semi-Transparency model access to **personal data** after installation.
Google Play Store

635,264 apps
• 483,277 free (76 %)
• 151,987 paid (24 %)

52%

10% privacy policy

52% (331,529 apps)
• 86% (286,376) free (59%)
• 14% (45,153) paid (29%)
Improved Permission Interface

- Identify **safe** apps (apps that collect no personal information);
- Focus attention on permissions that grant access to personal information;
- Identify possible changes in the permission set when new versions are released;
- Make it easier to identify when an app transitions from being safe to having the potential to disclose data. Users can decide not to update the app and hence keep the “previous" version on their phone, which will still function for a time.
User Study Outcome

Users, who are otherwise unable to do so, are able to understand access to personal permissions at glance and have the ability to make more informed choices, both when choosing and updating apps.
Relevant Advanced at MIT
Privacy Fingerprints – type of content accessed by an app in different contexts

Angry Birds

idle (87%)

- sensor rotation vector
- sensor accelerometer
- ip address (not moving)
- neighboring cell location (not moving)
- cell location (not moving)
- wifi (not moving)
- msisdn
- sim serial
- device id
- gps (not moving)
- voicemail
- browsing
- phone number
- email
- contact
- photo
- audio
- sms

active (13%)

- sensor rotation vector
- sensor accelerometer
- ip address (not moving)
- neighboring cell location (not moving)
- cell location (not moving)
- wifi (moving)
- msisdn
- sim serial
- device id
- gps (not moving)
- voicemail
- browsing
- phone number
- email
- contact
- photo
- audio
- sms

Increasing intrusiveness

proportion of data accesses
Dashboard showing App’s Privacy Fingerprints in App store

PlacesForYou

<table>
<thead>
<tr>
<th>Sensor/Device</th>
<th>Idle (61%)</th>
<th>Active (39%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>sensor rotation</td>
<td>0.5%</td>
<td>1.0%</td>
</tr>
<tr>
<td>sensor acc</td>
<td>4.0%</td>
<td>2.2%</td>
</tr>
<tr>
<td>ip address(not moving)</td>
<td>0.3%</td>
<td>0.2%</td>
</tr>
<tr>
<td>neighboring cell(not moving)</td>
<td>3.8%</td>
<td>2.5%</td>
</tr>
<tr>
<td>ip address(moving)</td>
<td>91%</td>
<td>94%</td>
</tr>
<tr>
<td>neighboring cell(moving)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cell location(not moving)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cell location(moving)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wifi(not moving)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wifi(moving)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>msisdn</td>
<td></td>
<td></td>
</tr>
<tr>
<td>imsi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sim serial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>device id</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gps(not moving)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gps(moving)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>voicemail</td>
<td></td>
<td></td>
</tr>
<tr>
<td>browsing</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Intrusiveness Score: 18.7

Compare with Similar Apps

<table>
<thead>
<tr>
<th>Intrusiveness Ranking</th>
<th>Name</th>
<th>Intrusiveness Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PlacesForYou</td>
<td>18.7</td>
</tr>
<tr>
<td>3</td>
<td>FriendHangout</td>
<td>17.2</td>
</tr>
<tr>
<td>20</td>
<td>HotSpots</td>
<td>11.0</td>
</tr>
</tbody>
</table>
Understand People’s Privacy Preferences

1. Collect sensor data + user semantic annotations (location, situation)
2. Generate personalized surveys to probe user privacy preferences
3. Model a user’s contextual preferences

Databases

Questions for time April 18, 2012, 1:39 p.m.

At April 18, 2012, 1:39 p.m., you labeled your location as Restaurant/Miscellaneous Mess Hall Dining and your situation as Having lunch. If, at that time, your smartphone had prompted you with questions:

Would you have disclosed your location to an application developed by Medallia for research purposes?
- Yes
- No

Would you have disclosed your location to an application developed by First Bank? If you receive a coupon worth 2 dollars?
- Yes
- No

Would you have disclosed your device scans (digital devices around you) to an application developed by Walmart if you receive a coupon worth 2 dollars?
- Yes
- No

Understanding People's Preferences for Disclosing Contextual Information to Smartphone Apps, HCII 2013
(best paper)
Results (from one user)

Location
0 Home
1 School:Miscellaneous:stata 5th floor lunch place
2 Work:Office:marisol office
3 Work:Office:My Office
4 School:Miscellaneous:stata 1st floor entrance
5 Fun Stuff:Miscellaneous:peter house
6 Other:Religious Center:alewife meeting hall
7 Errands:Store:cmart

Situation
0 sleeping
1 praying
2 Chilling alone
3 Chatting informally
4 Working alone
5 Hanging out with friends
6 eating a meal
7 Watching TV
8 In meeting
9 Having lunch
10 In car
11 Having dinner
12 In lecture
13 exercise step climbing
14 In discussion
15 On bus
16 read story
17 waiting for bus
18 Walking
19 Reading
20 On subway
21 wash dishes
22 singing
23 laundry
24 run and leave
25 playing with kids
Thank you

Cameron Kerry
Visiting Scholar
MIT Media Lab
Human Dynamics
ckerry@media.mit.edu