Privacy-by-Design for the Security Practitioner

Richard Chow
richard.chow@intel.com


PbD
www.privacybydesign.ca

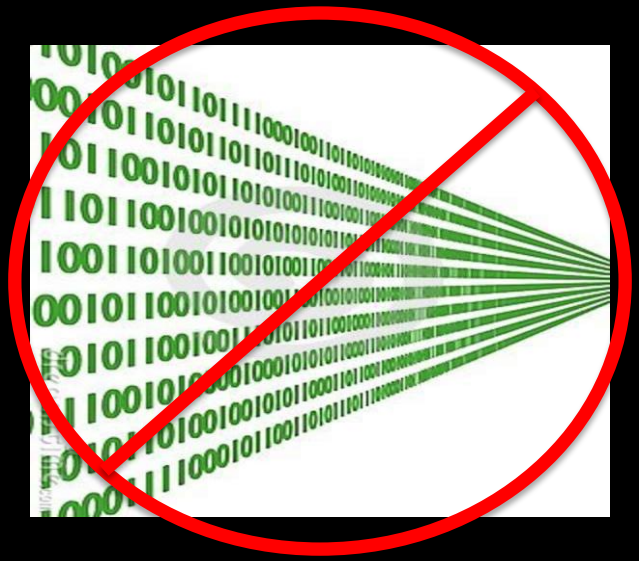
Agenda

- Security vs Privacy
- Personally Identifying Information (PII)
- Privacy-by-Design
  - Data minimization
  - Does the user understand?

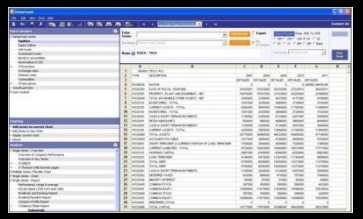
SECURITY VS PRIVACY

Security

Alice

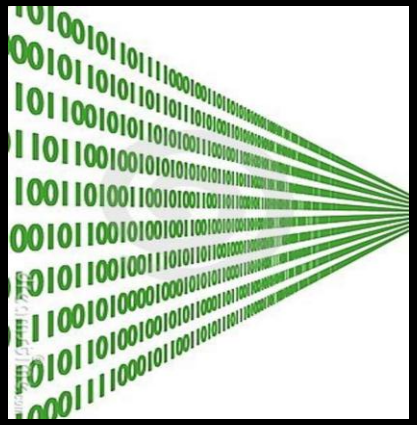


Privacy



Auxiliary Data

Alice



Inference

PERSONALLY IDENTIFYING INFORMATION (PII)

What is PII?

Asian-Pacific Economic Cooperation:

“any information about an identified or identifiable individual”

Really, what is PII?

- What does “identifiable” mean?
- Depends on all data collected
  - For example: browser user-agent, time-zone
- Depends on auxiliary data

- PII construct is based on policy and law
  - Not a technical construct!

Myth of PII

“Just as medieval alchemists were convinced a (mythical) philosopher's stone can transmute lead into gold, today's privacy practitioners believe that records containing sensitive individual data can be "de-identified" by removing or modifying PII.”

Narayanan and Shmatikov in “Myths and Fallacies of PII” in Communications of the ACM

What to do about PII polices?

- Risk of data depends on ease of identifiability and sensitivity
- Similar risk means similar methods of safeguarding and handling

PRIVACY-BY-DESIGN

Privacy-by-Design Principles

- Proactive not Reactive; Preventative not Remedial
- Privacy as the Default
- Privacy Embedded into Design
- Full Functionality – Positive-Sum, not Zero-Sum
- End-to-End Security – Lifecycle Protection
- Visibility and Transparency
- Respect for User Privacy

Privacy-by-Design History

1970's

Fair Information Practices: US Dept of HEW

1980's

Organization of Economic Cooperation and Development (OECD)

1990's

Privacy-by-Design

EU Data Protection Directive

2000's

Asian Pacific Economic Cooperation (APEC)

2010's

US FTC and EU Data Protection Regulation mention PbD

Guideline #1

Is the data secure?

Guideline #2

Have we minimized the data collected?

Guideline #3

Does the user understand?

PRIVACY-BY-DESIGN: DATA MINIMIZATION

The Problem with IDs

- Glues data together
- Silos good for privacy!

Data Minimization with IDs

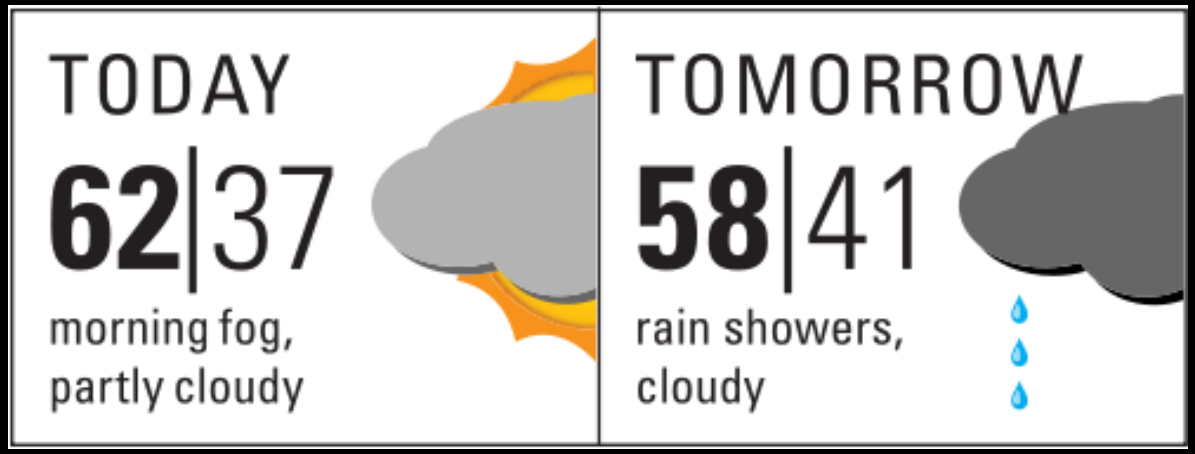
- Project: Collect data for trouble-shooting and diagnostics
- Need to correlate data from same device

Data Minimization with IDs

- Correlation only needs to be local in time
- Design: Periodically change identifier

Third-party Weather Service

Scenario: Web-site or app that incorporates a 3rd-party weather web service

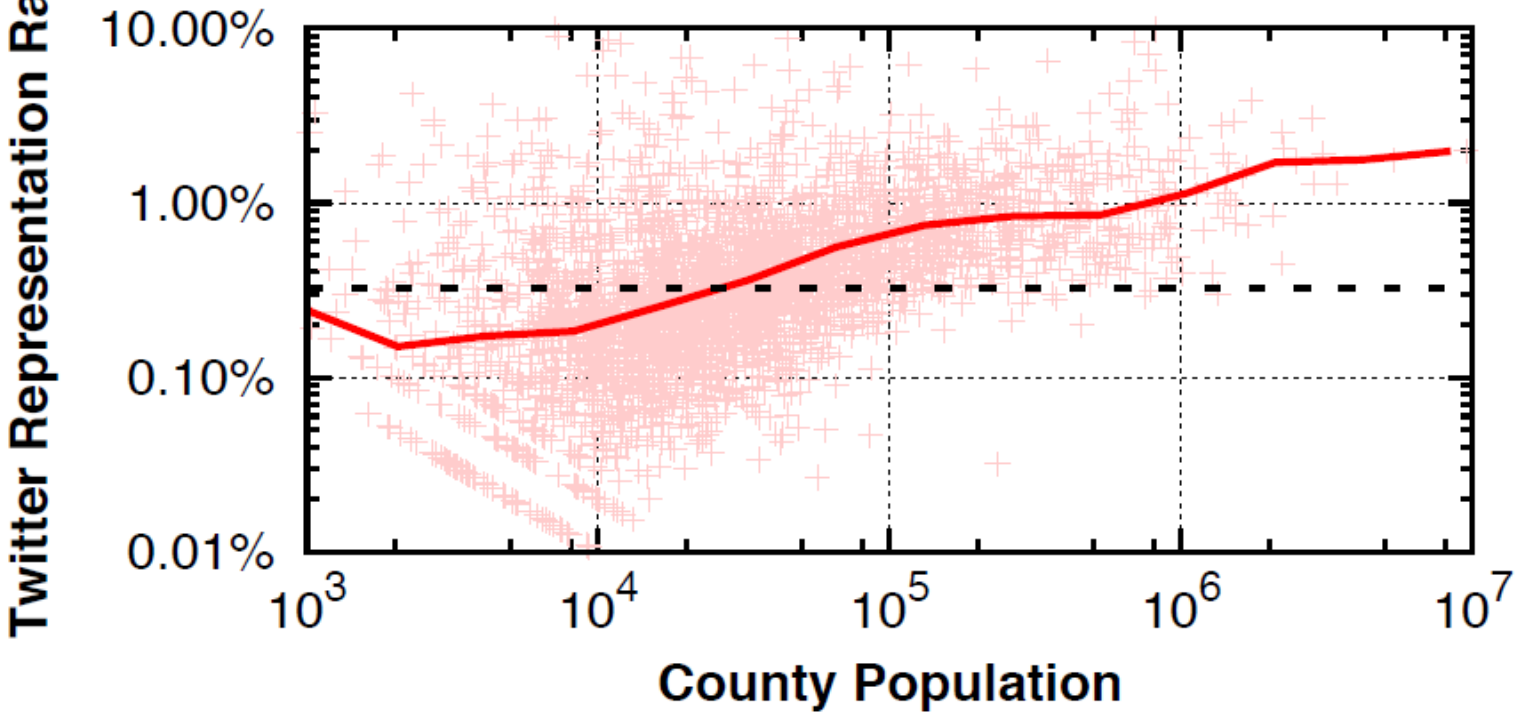


Third-party Weather Service

- Provides weather, given user location
- Does not learn user ID

Possible?

3.2 million Twitter users

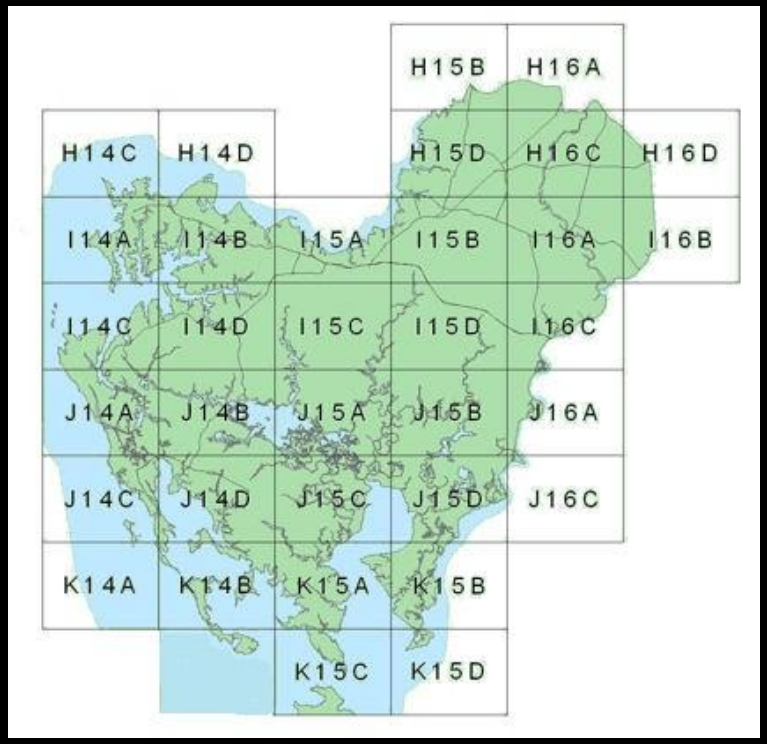


From: “Understanding the Demographics of Twitter Users” by Mislove et al.

Minimize data sent to web service

For example:

- coarsen latitude, longitude
- send aggregate data only



Anonymization

- Keep data around by de-personalizing?
- Example: Google and Yahoo de-personalize search data after X months

How to de-personalize?

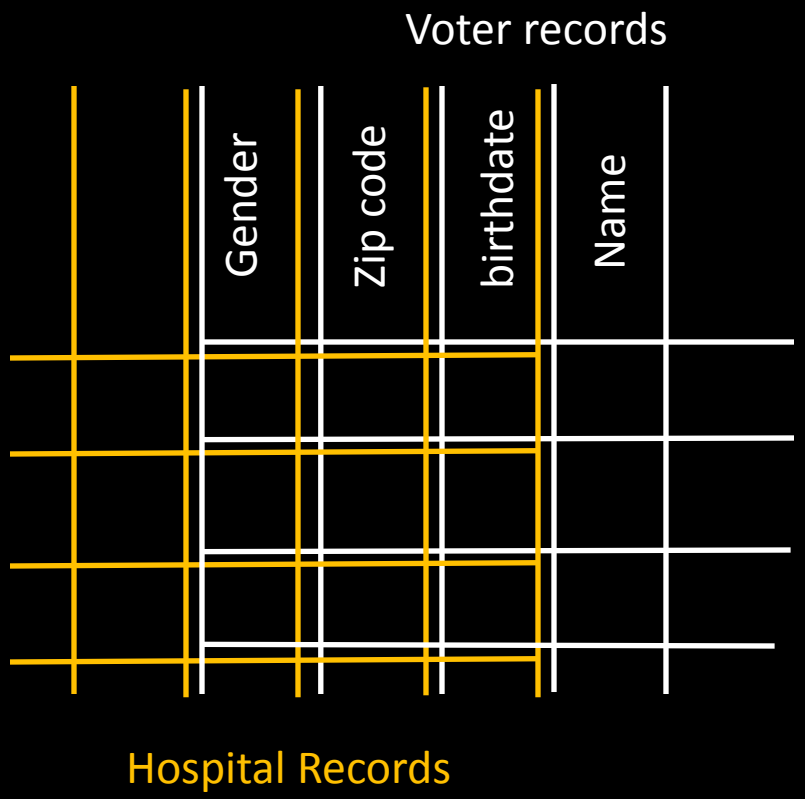
Not trivial…

- Location data
- Search data

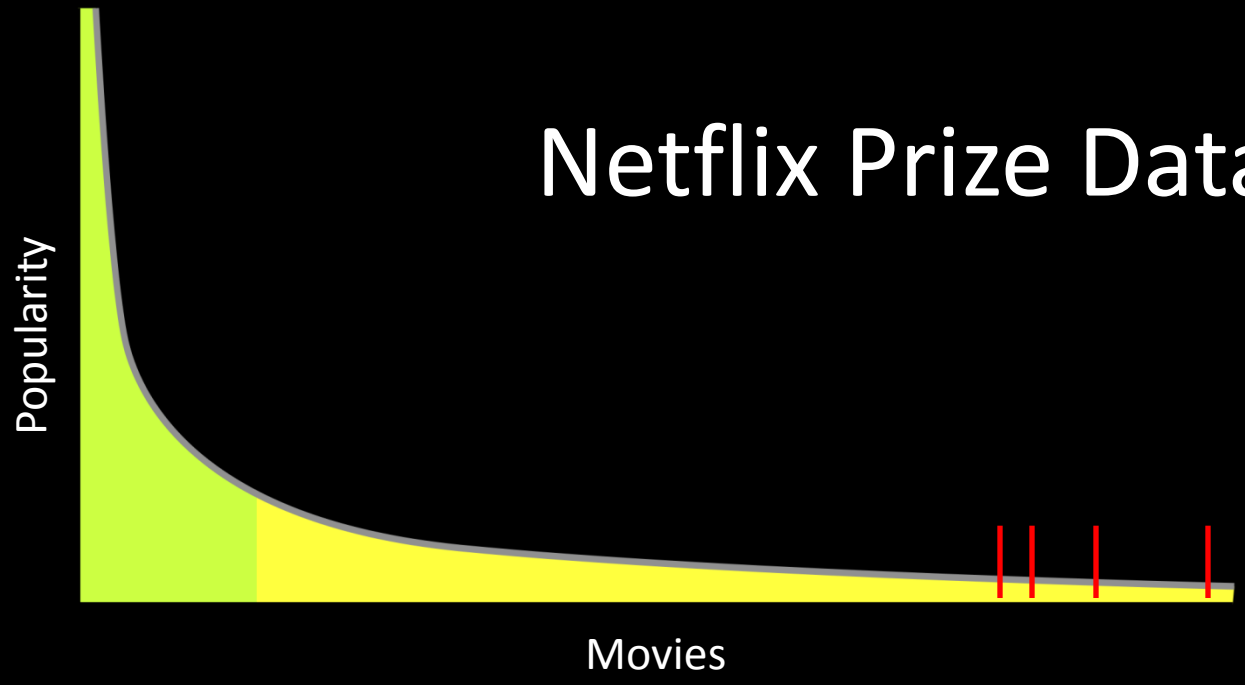
From The New York Times: A Face Is Exposed for AOL Searcher No. 4417749

Tabular Data

Latanya Sweeney identified hospital visits of MA governor



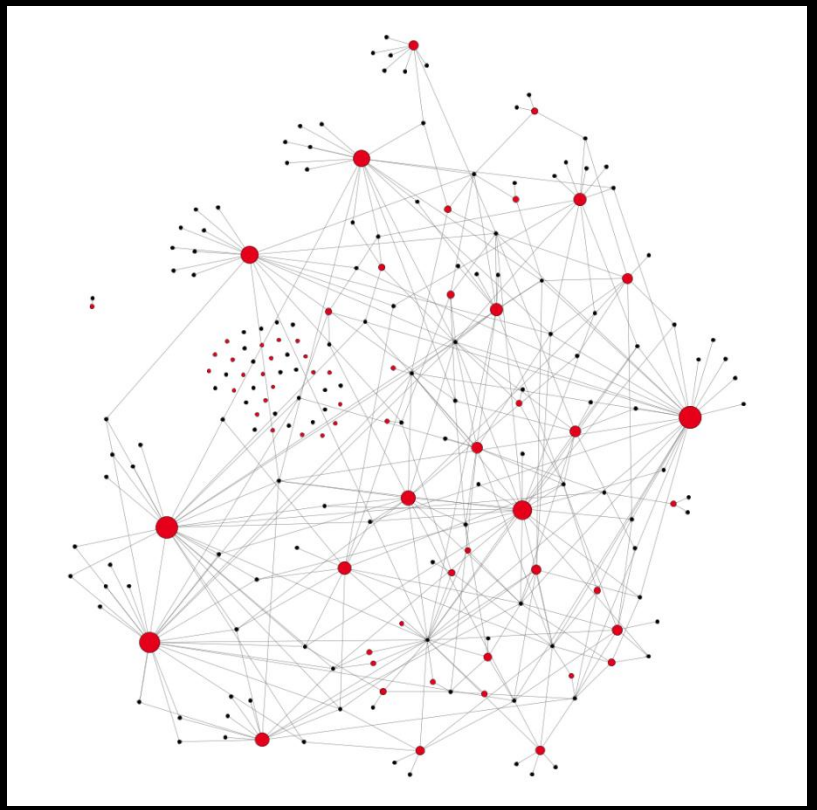
Long tail data



Netflix Prize Dataset

Graphical data

Social Network



Challenge

Data Minimization

Data Mining

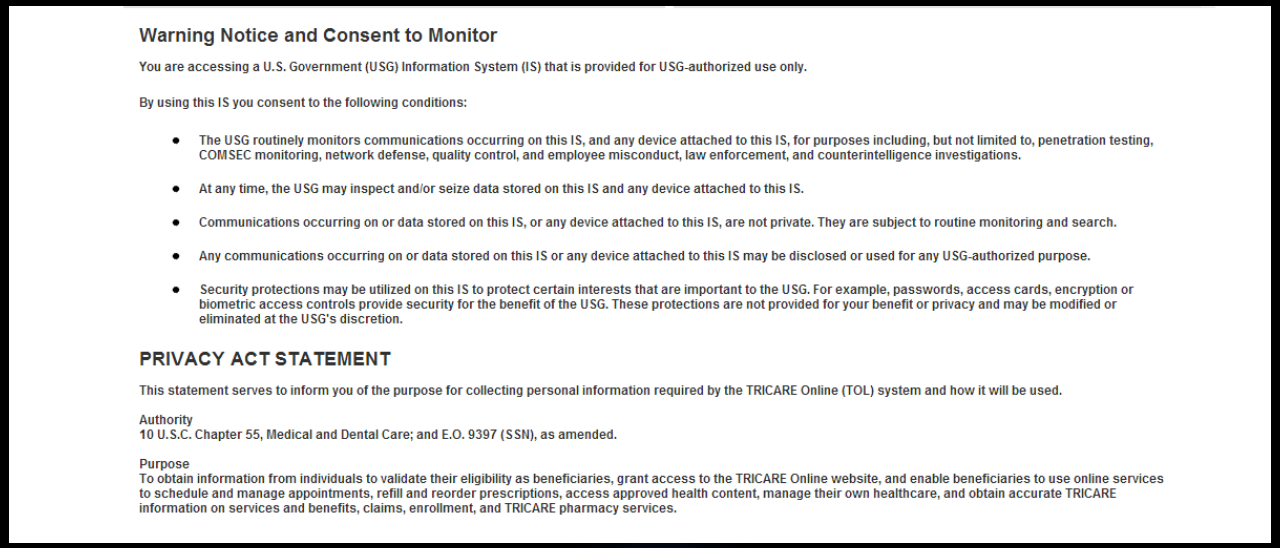
Data Mining

PRIVACY-BY-DESIGN:
DOES THE USER UNDERSTAND?

Traditional Notice and Consent

User consents to data collection after understanding:

- Which data is collected
- Why it is collected



Warning Notice and Consent to Monitor

You are accessing a U.S. Government (USG) Information System (IS) that is provided for USG-authorized use only.

By using this IS you consent to the following conditions:

- The USG routinely monitors communications occurring on this IS, and any device attached to this IS, for purposes including, but not limited to, penetration testing, COMSEC monitoring, network defense, quality control, and employee misconduct, law enforcement, and counterintelligence investigations.
- At any time, the USG may inspect and/or seize data stored on this IS and any device attached to this IS.
- Communications occurring on or data stored on this IS, or any device attached to this IS, are not private. They are subject to routine monitoring and search.
- Any communications occurring on or data stored on this IS or any device attached to this IS may be disclosed or used for any USG-authorized purpose.
- Security protections may be utilized on this IS to protect certain interests that are important to the USG. For example, passwords, access cards, encryption or biometric access controls provide security for the benefit of the USG. These protections are not provided for your benefit or privacy and may be modified or eliminated at the USG's discretion.

PRIVACY ACT STATEMENT

This statement serves to inform you of the purpose for collecting personal information required by the TRICARE Online (TOL) system and how it will be used.

Authority
10 U.S.C. Chapter 55, Medical and Dental Care; and E.O. 9397 (SSN), as amended.

Purpose
To obtain information from individuals to validate their eligibility as beneficiaries, grant access to the TRICARE Online website, and enable beneficiaries to use online services to schedule and manage appointments, refill and reorder prescriptions, access approved health content, manage their own healthcare, and obtain accurate TRICARE information on services and benefits, claims, enrollment, and TRICARE pharmacy services.

Easier than it sounds…

- Privacy Notices
- EULA

Recommendation:

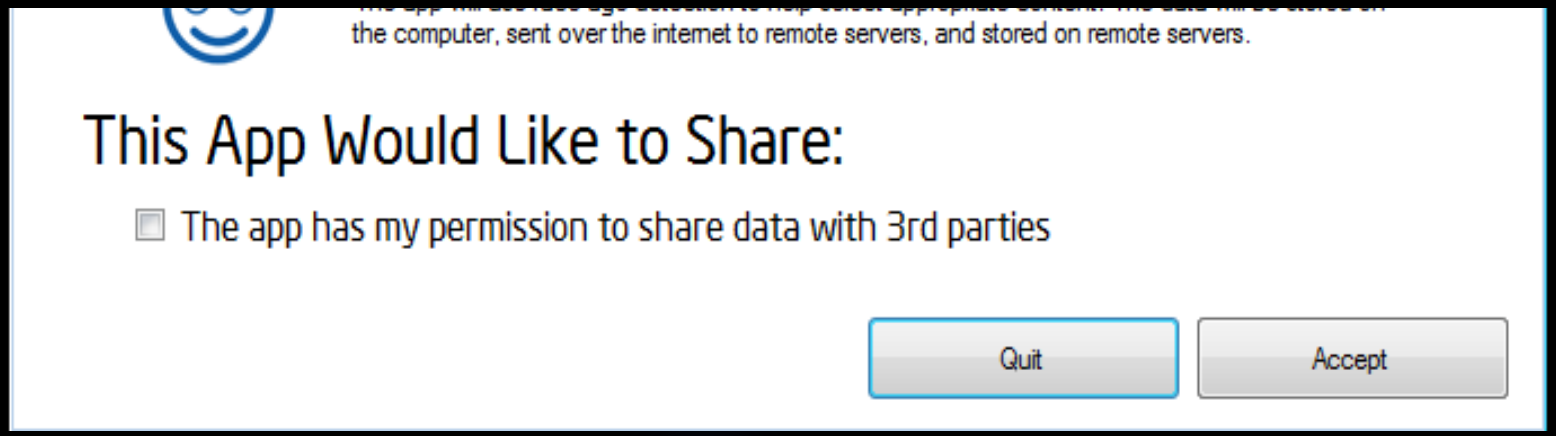
Do as much as possible in the area of data minimization; rely on user understanding as little as possible

Example: Perceptual Computing SDK

- Bundle of algorithms for using cameras and microphones
- 3rd-party developers write apps on top of SDK

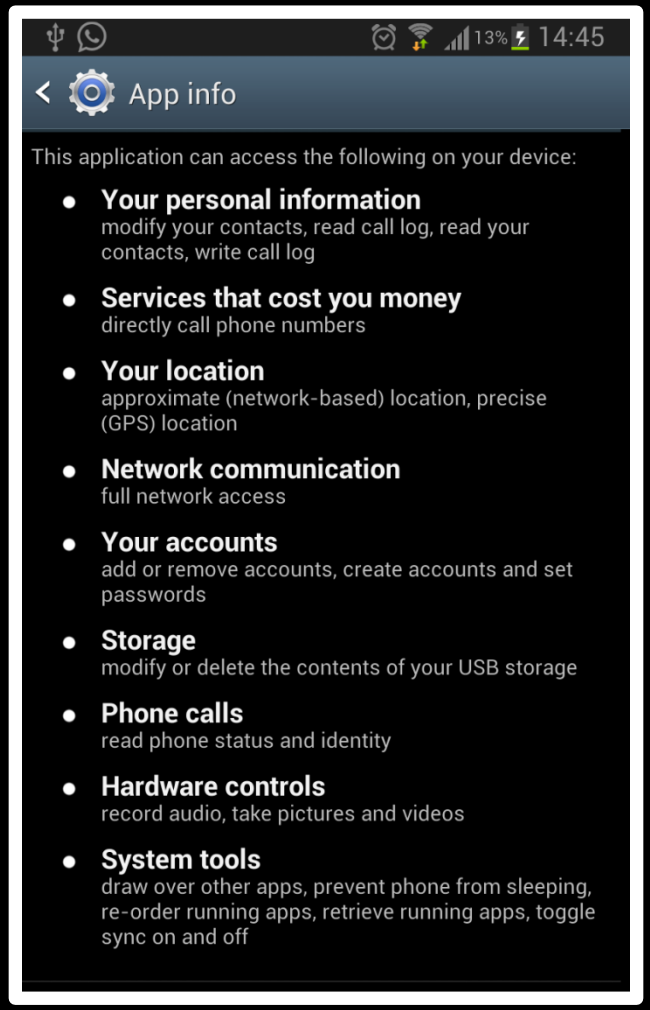
Private-by-Default

- Maintain privacy if the user doesn't do anything or not paying attention
- Similar to *fail-safe*



Effectiveness of Notice?

- Similar to Android install-time permissions
  - User does not want distractions
  - Like Privacy Notices and EULAs
- Contextual approaches?



Summary

- Privacy-by-Design now standard for privacy engineering
- For security practitioner, two less familiar areas
  - Data Minimization: Emphasizes machine learning
  - User Understanding: Emphasizes HCI


Thank you!
Richard Chow
richard.chow@intel.com
black hat
ASIA 2014

Please complete the Speaker Feedback Surveys

Richard Chow

richard.chow@intel.com

black hat ASIA 2014