Privacy engineering, privacy by design, privacy impact assessments, and privacy governance

Lorrie Faith Cranor

October 29, 2013

8-533 / 8-733 / 19-608 / 95-818:
Privacy Policy, Law, and Technology

Carnegie Mellon University
CyLab

institute for SOFTWARE RESEARCH

Engineering & Public Policy

Course schedule announcements

- http://cups.cs.cmu.edu/courses/pplt-fa13/
- No more homework except reading summaries
- Reading summaries due November 19
- No more reading assignments after November 19
- Work on your projects!

Engineering Privacy

- Sarah Spiekermann and Lorrie Faith Cranor. Eningeering Privacy. IEEE Transactions on Software Engineering. Vol. 35, No. 1, January/February, 2009, pp. 67-82. http://ssrn.com/abstract=1085333

Privacy spheres

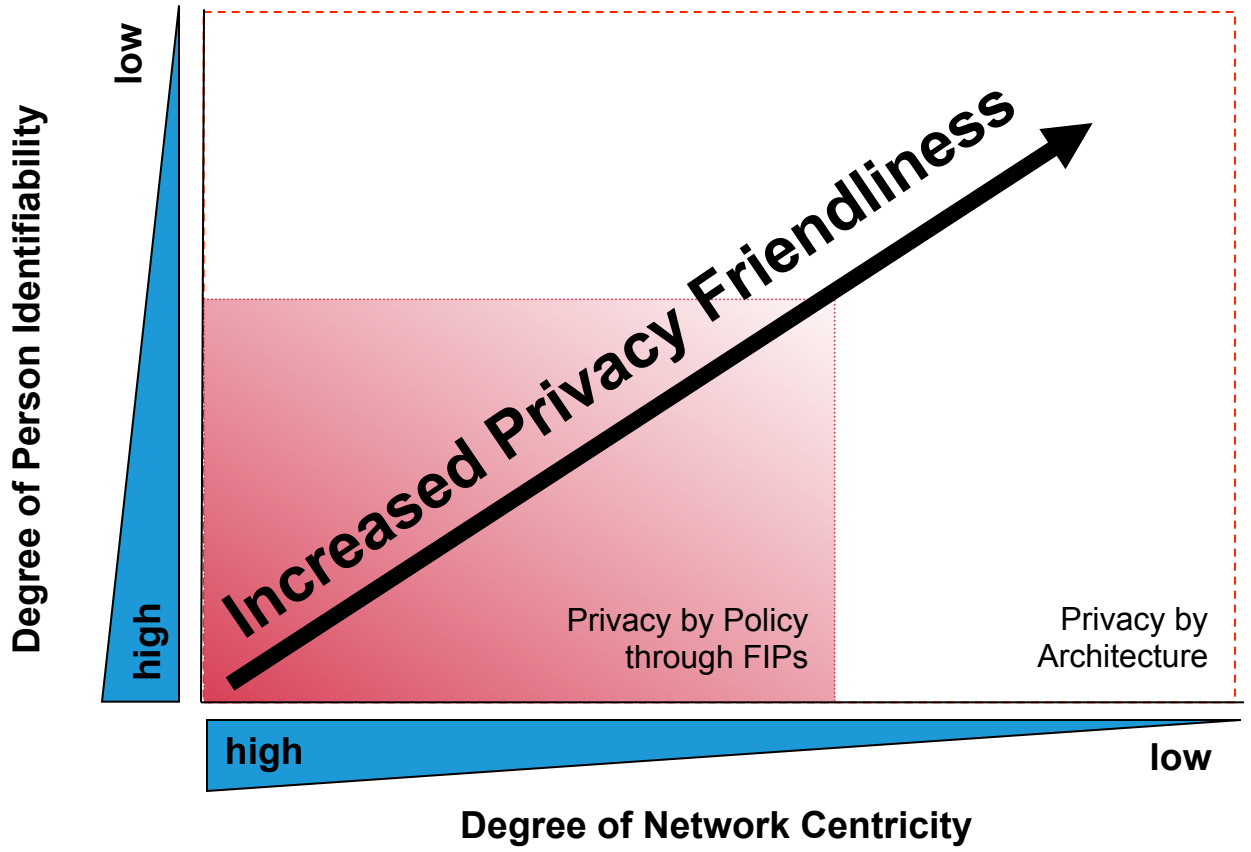
Privacy Spheres	Where Data is Stored	Engineer's Responsibility	Engineering Issues
User Sphere	Users' desktop personal computers, laptops, mobile phones, RFID chips	• Give users control over access to themselves (in terms of access to data and attention)	• What data is transferred from the client to a data recipient? • Is the user explicitly involved in the transfer? • Is the user aware of remote and/or local application storing data on his system? • Is data storage transient or persistent?
Joint Sphere	Web service provider's servers and databases	• Give users some control over access to themselves (in terms of access to data and attention) • Minimize users' future privacy risks	• Is the user fully aware of how his data is used and can he control this?
Recipient Sphere	Any data recipients: servers and databases of network providers, service providers or other parties with whom data recipient shares data	• Minimize users' future privacy risks	• What data is being shared by the data recipient with other parties? • Can the user expect or anticipate a transfer of his data by the recipient? • Is personal data adequately secured? • Is data storage transient or persistent? • Can the processing of personal data be foreseen by the user? • Are there secondary uses of data that may not be foreseen by the user? • Is there a way to minimize processing? (e.g. by delegating some pre-processing to User Sphere)

User privacy concerns

Sphere of Influence	User privacy concerns
User Sphere	Unauthorized collection Unauthorized execution Exposure Unwanted inflow of data
Joint Sphere	Exposure Reduced Judgment Improper access Unauthorized secondary use
Recipient sphere	Internal unauthorized use External unauthorized use Improper access Errors Reduced judgment Combining data

How privacy rights are protected

- By policy
  - Protection through laws and organizational privacy policies
  - Must be enforced
  - Transparency facilitates choice and accountability
  - Technology facilitates compliance and reduces the need to rely solely on trust and external enforcement
  - Violations still possible due to bad actors, mistakes, government mandates
- By architecture
  - Protection through technology
  - Reduces the need to rely on trust and external enforcement
  - Violations only possible if technology fails or the availability of new data or technology defeats protections
  - Often viewed as too expensive or restrictive


Degree of Person Identifiability
low
high
Increased Privacy Friendliness
Privacy by Policy through FIPs
Privacy by Architecture
high
low
Degree of Network Centricity

Privacy stages	identifiability	Approach to privacy protection	Linkability of data to personal identifiers	System Characteristics
0	identified	privacy by policy (notice and choice)	linked	• unique identifiers across databases • contact information stored with profile information
1	pseudonymous		linkable with reasonable & automatable effort	• no unique identifies across databases • common attributes across databases • contact information stored separately from profile or transaction information
2		privacy by architecture	not linkable with reasonable effort	• no unique identifiers across databases • no common attributes across databases • random identifiers • contact information stored separately from profile or transaction information • collection of long term person characteristics on a low level of granularity • technically enforced deletion of profile details at regular intervals
3	anonymous		unlinkable	• no collection of contact information • no collection of long term person characteristics • *k*-anonymity with large value of *k*

Privacy by architecture techniques

- Best
  - No collection of contact information
  - No collection of long-term person characteristics
  - k-anonymity with large value of k
- Good
  - No unique identifiers across databases
  - No common attributes across databases
  - Random identifiers
  - Contact information stored separately from profile or transaction information
  - Collection of long-term personal characteristics w/ low granularity
  - Technically enforced deletion of profile details at regular intervals

De-identification and re-identification

- Simplistic de-identification: remove obvious identifiers
- Better de-identification: also k-anonymize and/or use statistical confidentiality techniques
- Re-identification can occur through linking entries within the same database or to entries in external databases

Examples

- When RFID tags are sewn into every garment, how might we use this to identify and track people?
- What if the tags are partially killed so only the product information is broadcast, not a unique ID?
- How can a cellular provider identify an anonymous pre-paid cell phone user?
- Other examples?

Privacy by policy techniques

- Notice
- Choice
- Security safeguards
- Access
- Accountability
  - Audits
  - Privacy policy management technology
    - Enforcement engine

User concerns	Notice should be given about…
Marketing Practices	
Combining Data	Notice about data combination practices • external data purchases? • linking practices?
Reduced Judgment	Notice about segmentation practices • type of judgments made? • personalization done? • what does personalization lead to for the customer? • sharing of segmentation information?
Future attention consumption	• contact plans (i.e. through newsletters, SMS)
IS Practices	
External unauthorized transfer	• is data shared outside the initial data recipient? • if yes, with whom is data shared?
External unauthorized processing	• is data processed externally for other purposes than initially specified? • if yes, for what purposes?
Internal unauthorized transfer	• is data transferred within a company conglomerate? • if yes with whom within the comglomerate?
Internal unauthorized processing	• is data processed internally for other purposes than initially specified? • if yes, for what purposes?
Unauthorized collection of data from client	• use of re-identifiers (i.e. cookies, stable IP address, phone number, EPC) • collection of information about device nature (i.e. browser, operating system, phone type) • collection of information from the device (i.e. music library, cache information)
Unauthorized execution of operations on client	• installation of software? • updates?
Exposure	• cached information (i.e browser caches, document histories) • collection of information from the device (i.e. music library, cache information)

Privacy Impact Assessment

A methodology for

- assessing the impacts on privacy of a project, policy, program, service, product, or other initiative which involves the processing of personal information and,
- in consultation with stakeholders, for taking remedial actions as necessary in order to avoid or minimize negative impacts

D. Wright and P. De Hert, eds. *Privacy Impact Assessment*. Springer 2012.

PIA is a process

- Should begin at early stages of a project
- Should continue to end of project and beyond

Why carry out a PIA?

- To manage risks
  - Negative media attention
  - Reputation damage
  - Legal violations
  - Fines, penalties
  - Privacy harms
  - Opportunity costs
- To derive benefits
  - Increase trust
  - Avoid future liability
  - Early warning system
  - Facilitate privacy by design early in design process
  - Enforce or encourage accountability

Who has to carry out PIAs?

- US administrative agencies, when developing or procuring IT systems that include PII
  - Required by E-Government Act of 2002
- Government agencies in many other countries
- Sometimes done by private sector
  - Case studies from Vodaphone, Nokia, and Siemens in PIA book

Data governance

- People, process, and technology for managing data within an organization
- Data-centric threat modeling and risk assessment
- Protect data throughout information lifecycle
  - Including data destruction at end of lifecycle
- Assign responsibility

Homework 5 discussion

- Pick a consumer software product or service that may collect information from or about its users and may transmit some or all of that information off the consumer's device or share information collected by a service with other parties.
- Use the Microsoft Privacy Guidelines to analyze this software. List all the applicable guidelines and try to determine whether/how the software complies with each one by using the software and reading its documentation. Make a table showing each guideline and how the software complies with or violates it (or explaining why you are unable to determine this). In the case of violations, what changes would you recommend to comply with these guidelines.
- Use the approaches described by Rubinstein and Good to expand your analysis to address issues not addressed by the Microsoft guidelines.

Carnegie Mellon University
CyLab

isr institute for SOFTWARE RESEARCH

Engineering & Public Policy