Meanings of "Privacy" in Privacy Enhancing Technologies

Claudia Diaz
KU Leuven – COSIC

Digital Identity Management (DIM)
November 8, 2013
Overview

• Review of (three) different families of privacy technologies focusing on:
  – the concept of “privacy” they embed
  – their goals
  – their challenges and limitations
  – incentives/obstacles for deployment

• Content based on ongoing work with Seda Gürses (NYU) on Privacy Research Paradigms in CS
“Social privacy”: Privacy concerns

• Technology mediation of social interactions leads to problems in the immediate social context of the user
  – Examples:
    • “My parents discovered I’m gay”
    • “My boss found out that I hate him”
    • “My friends saw my naked pictures OMG!”

• Self-presentation and identity construction towards friends, family, colleagues
  – Particularly relevant in social media applications
  – Tension between privacy and publicity

• Decision making: cognitive overload, bounded rationality, immediate gratification, hyperbolic discounting, behavioral biases

• **Who** defines the privacy problem:
  – Users
“Social privacy”: Goals

• Meet privacy expectations: system behaves as expected by the user:
  – “don’t surprise the user!”

• Make privacy controls (e.g., settings) visible and easy to use

• Assist users in privacy-relevant decision making:
  – users can predict the outcomes of their actions, such that they do not regret their actions after the fact

• Help users develop appropriate privacy practices
  – e.g., etiquette: use “Bcc:” instead of “Cc:” when sending email to a large number of people
“Social privacy”: Examples

• Appropriate defaults
  – “only friends”

• Usable privacy settings, tools for audience segregation
  – automated grouping of friends

• Contextual feedback mechanisms
  – “how others see my profile”

• *Privacy nudges*
  – timer nudge, audience visualization nudge, content analysis nudge
Social privacy technologies: challenges and limitations

• Focus on volitional actions (e.g., user-generated content)
  – Concerns relate to harms that are direct consequences of user actions
  – Typically leaving out implicit data, more abstract privacy risks

• Focus on the front-end
  – Making abstraction of: how the back-end is implemented, what information is disclosed to the service provider, how it can be (stealthily) used by the provider

• Research methodology: user studies
  – Mostly conducted in Europe and North America
  – Focus on the “average consumer”
  – Limited by users’ understanding and perception of the system

• Focus on “privacy expectations”
  – Slippery slope if expectations erode
    • Example: prisoners in the Panopticon have no expectation of privacy, thus, the system design perfectly meets their privacy expectations

• Paradox of control (affects all types of privacy technologies)

• Incentives for deployment: strong
  – Aligned with industry’s interests: make users comfortable with sharing information in their systems
“Institutional privacy”: Privacy concerns

• Data collection without user awareness or *informed consent*

• **Use** of data for illegitimate purposes

• Data security:
  – Information becoming public (or widely available to third parties)
  – Safety, protection from crime: identity theft, stalking, etc.

• Data correctness, integrity, deletion

• **Who** defines the privacy problem:
  – Legislation, organizations (through policies)
“Institutional privacy”: Goals

• Ensure compliance with data protection principles:
  – informed consent, purpose limitation, data minimization, data security obligations, subject access rights

• Data security:
  – prevent (or mitigate the consequences of) data breaches

• Auditability and accountability
“Institutional privacy”: Examples

• appropriate defaults and privacy controls
  – again, but here towards organizations instead of peers
• tools to make privacy policies easier to understand and negotiate
  – P3P, DNT
• tools help organizations define and enforce access control policies
  – purpose-based access control
• auditing systems
• database privacy technologies
  – data anonymization and differential privacy techniques
Institutional privacy technologies: challenges and limitations

• Assumes the collection and processing of personal information by organizations is good and necessary
• The organization is (semi-)trusted to be honest, competent, and act in the best interest of the user
  – Reliance on the legal system to punish lack of compliance
  – No (technical) protection guarantees towards organizations that want to violate user privacy by stealthily abusing the data that they hold
• Focus on limiting (mis)use of personal data, rather than collection
  – In spite of data minimization principles in data protection, it is easy to justify mass collection and/or obtain consent for it: does not preempt the creation of large databases
  – Auditing and legal compliance mechanisms may result in more data being recorded
• Who has the power to define and enforce the policies on data use?
  – Do whatever we wanted to do with the data while being compliant
• Focus on “personal data”
  – Does not address inferences from anonymized or aggregated data (discrimination concerns)
• Limits on transparency posed by IP (proprietary software, algorithms, databases)

• Incentives for deployment: strong
  – Legal compliance is a very strong driver
Anti-surveillance technologies (PETs): Privacy concerns

• Data disclosure by default through the use of the ICT infrastructure

• Threat model:
  – surveillance by (possibly colluding) service providers and governments
  – not unreasonable given recent revelations.

• Censorship by service providers and governments
  – protection of the public sphere

• Relationship to other democratic values:
  – Protection of dissent, free speech, freedom of association, freedom from government intrusion, protection of the democratic system itself (danger of moving towards a totalitarian system through mass surveillance)

• **Who** defines the privacy problem:
  – Security experts
Anti-surveillance technologies (PETs): Goals

• Prevent/minimize default disclosure of personal information to service providers and other third parties:
  – Only information explicitly disclosed is made available to intended recipients (confidentiality)
  – This includes user-generated content and implicit data
  – Minimize the need to trust others with appropriately handling data
    • Distribute trust by avoiding single points of failure

• Circumvent censorship
  – Availability properties
  – Circumvention might need to be undetectable (hard!)
Anti-surveillance technologies (PETs): Examples

• end-to-end encryption
  – PGP, OTR
• systems for anonymous communications
  – Tor
• advanced crypto protocols:
  – private information retrieval, oblivious transfer
  – anonymous authentication
  – privacy-preserving smart metering
• obfuscation approaches:
  – TMN: degrade data quality with noise
• Technologies that expose surveillance (transparency)
  – FPDetective
Anti-surveillance technologies (PETs): challenges and limitations

• Focus on (preventing) data disclosure
  – No protection for information after disclosure
• Making secure design and implementations is hard
  – Active research
  – Importance of public algorithms and open source: “it takes a village to keep systems secure”
  – Security of end-devices: big issue
• Research methodology:
  – Narrow privacy definitions
  – Driven by threat (adversarial) models
  – Explicit (sometimes implicit) assumptions that need to hold to guarantee privacy properties
    (mathematical, behavioral, available building blocks, trust assumptions)
• Making security usable is hard
  – Target: global user base, or users with stronger privacy concerns (e.g., activists, journalists)?
• Incentives for deployment: weak at best
  – Companies don’t want this: less data is bad for business
  – Governments neither: national security, law enforcement, social control, detection of fraud
PETs implemented by the Service Provider

• Example: advanced crypto protocols:
  – identity management systems, privacy-preserving smart metering, road tolling, etc.

• Requires
  – Designing the system with the PET integrated in it
  – Significant investment
  – Expertise in implementing and integrating the PET
  – Availability of software for review (trust in the implementation)
  – Interest/incentives from the SP
Unilateral PETs

- Example: encryption plug-ins (e.g., for gmail, facebook), OTR for instant messaging, obfuscation tools (e.g., TMN), anonymizing proxies
- SPs do not need to invest or modify their services, PET only at client-side
  - Often implemented as (research) open-source projects: expertise and review required!
- Requires:
  - That the SP “tolerates” the use of the PET
    - In the terms and conditions
    - In practice: e.g., that it does not take action to make the PET unusable (plausible deniability)
  - ... or that the PET is made undetectable: possible? desirable?
Collaborative PETs

- Example: anonymous communications networks (e.g., Tor hidden services), distributed social networks, community-based systems
- The service itself is implemented in a P2P fashion, often as a (research) open-source project
- Commercial SPs still involved: communications infrastructure
  - Possibility to make PET unusable by blocking its communications
  - Governments also sometimes interested in blocking these PETs
- Requires
  - An engaged community of users, security expertise, and software review
  - Tradeoffs performance/cost/security, particularly to protect against traffic analysis
  - Protection from being outlawed
Identity Management Systems

• What is identity management?
  – Many possible meanings
  – Complex systems meant to support a broad variety of applications

• Elements belonging to the various paradigms:
  – Anti-surveillance: minimize disclosure of data, narrow privacy definitions, hard guarantees
  – Institutional privacy: managing interactions with organizations
    • Who defines / verifies the policies that determine which attributes should be disclosed in a given context?
  – Social privacy: managing identity towards other peers

• Increased identity assurances may facilitate surveillance (in spite of data minimization)
Conclusions & Refs

- Diverse landscape of privacy technologies, in terms of goals, limitations, and assumptions (trust, dependencies on technology, law, social norms or third parties)
  - hard to approach for outsiders (and even for insiders!)
- Importance of understanding embedded concepts of privacy and who gets to define those concepts and fill them with meaning!
- How to integrate the different technological approaches?
- Incentives!! Particularly, how to incentivize and support the deployment of anti-surveillance technologies?

- Recent articles: