Internet Privacy and Ethics in a Digital Society

Ben Zevenbergen
Oxford Internet Institute
University of Oxford

This Lecture

- Privacy - Theory, Law, Policy
- Practical Guidelines for Internet Research
- Ethics - Ethical Justification in Networked Systems Research & Emerging Themes
- Exercise on Ethical Justifications in Internet Projects

Part I: What is Privacy?

“Privacy is an illusion”

“We haven’t got the privacy definition right”

“Lawyers and politicians don’t get technology”

Part I: What is Privacy?

"Privacy is a difficult term to define" &
"[...] there is no single definition or analysis or meaning of the term"

- Every paper about privacy, ever.

Historically: right to solitude

Most prominent now: Informational privacy

Caspar Bowden 19 August 1961 – 9 July 2015

Information privacy theory

1. *"The right to be **let alone**"* (Warren and Brandeis, 1890),
2. *"The claim of persons to **determine for themselves** when, how, and to what extent information about them is communicated to others"* (Westin 1967),
3. *"[...] in the context of modern data processing, [...] respect the capacity of the **individual to determine in principle** the disclosure and use of his/her personal data. Limitations to this **informational self-determination** are allowed only in case of overriding public interest"* (German Constitutional Court, 1983),
4. *"Information privacy theory constantly **evolves** with the introduction and implementation of new information technologies"* (Solove, 2012).

Privacy violation moments

1. Information **collection** (surveillance and interrogation);
2. Information **processing** (identification, aggregation, storing, second uses, exclusion);
3. Information **dissemination** (disclosure, breach of confidentiality, exposure, increased accessibility, distortion, blackmail and appropriation);
4. Invasion (**intrusion** or interference into one's life).

Information Collection

Information Processing

'Like' curly fries on Facebook? Then you're clever

'Like' curly fries? Then there's a good chance you've got a high IQ, according to a Cambridge University project to discover what we unwittingly reveal about ourselves on Facebook.

Facebook users are unwittingly revealing intimate secrets - including their sexual orientation, drug use and political beliefs - using only public "like" updates, according to a study of online privacy.

How Depressives Surf the Web

JUNE 15, 2012

Gray Matter

By SRIRAM CHELLAPPAN and RAGHAVENDRA KOTIKALAPUDI

Email

IN what way do you spend your time online? Do you check your e-mail compulsively? Watch lots of videos? Switch frequently among multiple Internet applications — from games to file downloads to chat rooms?

We believe that your pattern of Internet use says something about you. Specifically, our research suggests it can offer clues to your mental well-being.

Information Dissemination

The UK's 10 most infamous data breaches

Software vulnerabilities, lost hard drives and CDs, malicious insiders, poor security - the UK's most important data breaches reveal just how many ways data can be put at risk.

By John E Dunn | Mar 17, 2015

ING Plan to Share Customer Payment Data Spurs Privacy Concerns

Intrusion

TARGET®

TECH 2/16/2012 @ 11:02AM | 2,780,171 views

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

Right to Privacy & Data Protection

ARTICLE 8

Right to respect for private and family life

1. Everyone has the right to respect for his private and family life, his home and his correspondence.

2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic wellbeing of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

Article 8

Protection of personal data

1. Everyone has the right to the protection of personal data concerning him or her.

2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.

3. Compliance with these rules shall be subject to control by an independent authority.

Fair Information Practice Principles

- **Data quality** – relevant, accurate, & up-to-date
- **Collection** - limited, lawful & fair; with consent or knowledge
- **Purpose specification** at time of collection
  - [Notice of purpose and rights at time of collection implied]
- **Uses & disclosures limited to purposes** specified or compatible
- **Security** through reasonable safeguards
- **Openness** regarding personal data practices
- **Access** – individual right of access
- **Correction** – individual right of correction
- **Accountable** – data controller with task of compliance
- **Data Minimisation** – minimal collection
- **Data retention limits** – relating to purpose
- **Sensitive information** – religion, political or sexual orientation, etc.

Contextual integrity (Nissenbaum)

- Violation of information privacy can occur when information **moves across contexts**.
- *Context-relative informational norms*, where the flow and use of specific information is considered to be inappropriate:
  - Actors (subject of information, capacity of recipient and power-balance with regards to the sender);
  - Attributes (data types of information);
  - **Transmission principles (constraints/rules under which information flows).**

Nissenbaum explains that these parameters should be imagined to be

“[…] juggling balls in the air, moving in sync: contexts, subjects, senders receivers, information types, and transmission principles.”

Designing Information Privacy

- **Information privacy can be designed** by applying an interdependent construct of technical tools, legal agreements and project governance to the information system (Bennett & Raab 2006).
- This requires a **sophisticated approach** whereby the combination of Transmission Principles used can vary greatly depending on the project, the type of data collected and their intended purpose (Altman et al. 2014).
- The method by which privacy is designed into an information system from the outset is called **Privacy-by-Design** (Cavoukian 2009).

Transmission principles

- Legal
  - Data protection law,
  - Informed consent and data sharing agreements.
- Technical
  - *“a system of ICT measures protecting informational privacy by eliminating or minimising personal data thereby preventing unnecessary or unwanted processing of personal data, without the loss of the functionality of the information system”* (van Blarkom et al. 2003)
- Organisational
  - Access limitations, privacy audits, etc.

Not one size fits all

- “Privacy is a socially created need, the substance of which depends on the specific context” (Moor 1984)
- “An individual’s relationship to society and cultural contexts is dynamic and changes over time, privacy is not a fixed condition” (Cohen 2013)
- ‘Privacy’ does not mean the same thing for different individuals or groups of people (Sheehan 2002; Bennett & Raab 2006)

Understanding Informational Norms

- Impact assessments..
- Best through inclusive, cross-disciplinary discussions
- What (surprising) privacy considerations have you dealt with in your work or life?

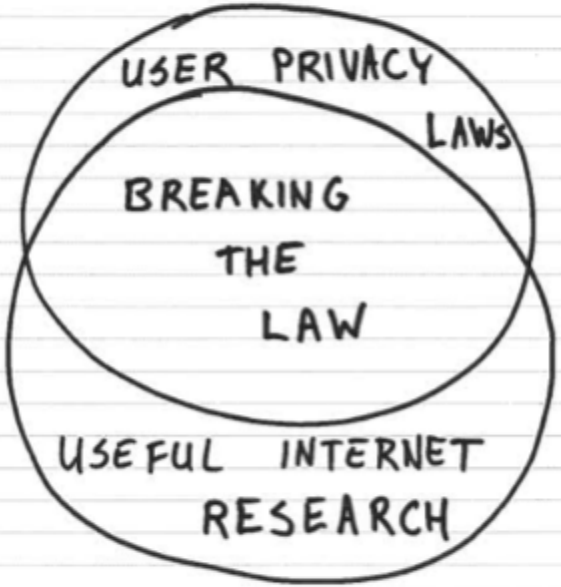
Part III: Mobile guidelines

Part III: Ethics in Networked Systems Research

Internet as socio-technical system

- Society & technology have become intertwined
- Internet is the technical backbone for modern society
- Measuring data flows = measuring behaviour (often)
- Ethical impact assessment and justification therefore important.

SAN DIEGO SUPERCOMPUTER CENTER
UNIVERSITY OF CALIFORNIA, SAN DIEGO
9500 Gilman Drive # 0505
La Jolla, CA 92093-0505
858-534-5000
www.sdsc.edu



SDSC

Human subject research?

- Collecting personally identifiable data
- Human subject research
  - Medical, social science, psychology, etc.
  - Relatively strict and broadly-understood ethical traditions
- Internet engineering
  - Consideration of these issues is relatively new,
  - Not obvious whether existing best practices from other fields can be successfully imported.
  - Wrong: “Privacy = security”

Laws can be confusing/outdated

- For example: Google/Spain,
  - Right to be Forgotten

Existing law and policy

- Even when relevant policies do exist,
- They are often ambiguous or inadequate
- As the were **designed for times with a less versatile technology** than computing.

Policy vacuums

- The **malleability** of computer allow them to be used in novel ways,
- Ways for which we frequently do not have formulated policies for controlling their use,
- Advancing computer technology produces policy vacuums in its wake.

Conceptual muddle

- Need an analysis which provides a **coherent conceptual framework** within which to formulate a policy for action.
- First understand how different disciplines understand their conceptual framework - identify opportunities for translation.

Task of ethics:

A basic job of computer/information ethics to:

- **identify** these policy needs,
- **clarify** related conceptual confusions,
- **formulate** appropriate new policies,
- and ethically **justify** them.

Computer ethics:

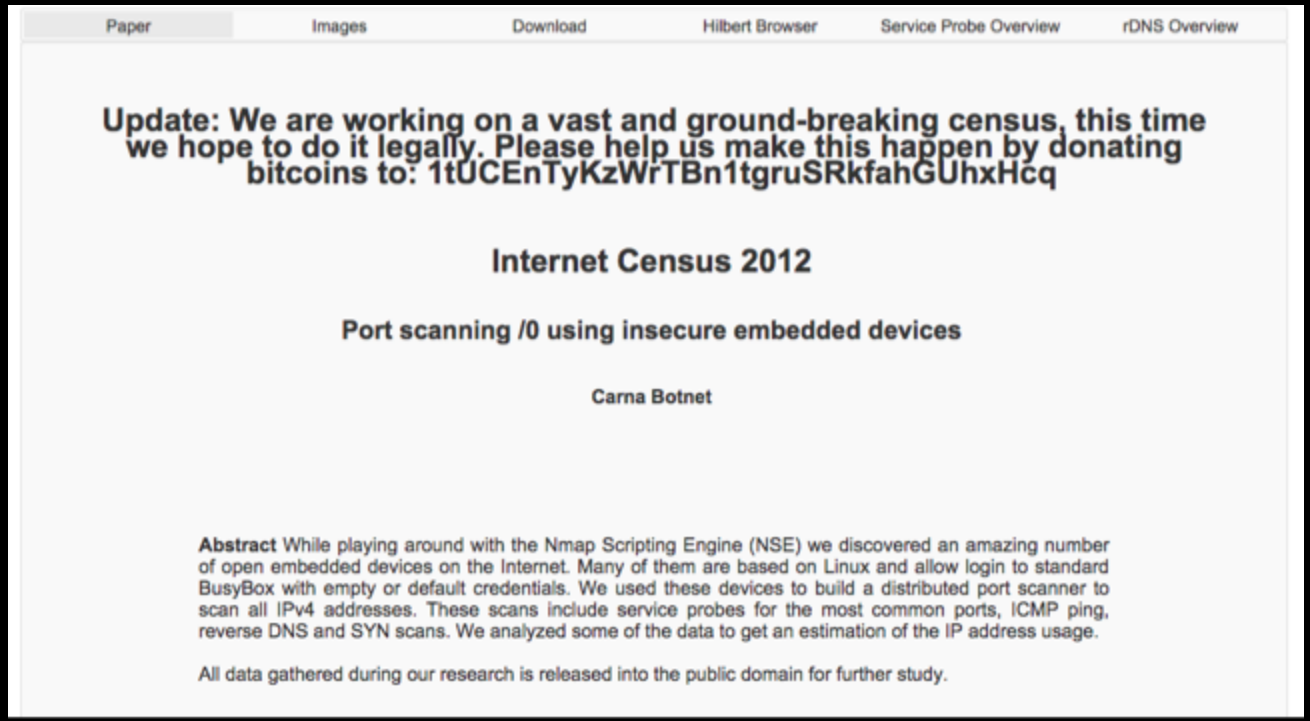
- The analysis of the nature and **social impact** of computer technology and,
- The corresponding **formulation** and **justification** of policies for the ethical use of such technology.

- (Moor 1985)

Ethics in Networked Systems Research

- Not to prescribe policies, technology design, nor to be considered as a definitive ethics guide.
- But to serve as a starting point for a discussion between engineers and ethical boards, lawyers, social scientists, and affected communities.

Scenario 1: (Big) Data Collection by Unlawful Intrusion



Paper | Images | Download | Hilbert Browser | Service Probe Overview | rDNS Overview

Update: We are working on a vast and ground-breaking census, this time we hope to do it legally. Please help us make this happen by donating bitcoins to: 1tUCEnTyKzWrTBn1tgruSRkfahGUhxHcq

Internet Census 2012

Port scanning /0 using insecure embedded devices

Carna Botnet

Abstract While playing around with the Nmap Scripting Engine (NSE) we discovered an amazing number of open embedded devices on the Internet. Many of them are based on Linux and allow login to standard BusyBox with empty or default credentials. We used these devices to build a distributed port scanner to scan all IPv4 addresses. These scans include service probes for the most common ports, ICMP ping, reverse DNS and SYN scans. We analyzed some of the data to get an estimation of the IP address usage.

All data gathered during our research is released into the public domain for further study.

Tech Reasoning

- “Created a huge map of the Internet through the illegal use of half a million devices.“
- Best dataset to understand the Internet network.
- Design principle was “be nice and don’t break things”
- “All data gathered during our research is released into the public domain for further study.“

Ethics Reasoning

- Is this a precedent to set?
  - Standards stick for a very long time
  - Good bugs can be exploited
  - So this should not be encouraged
- Trade-off benefits and harms
  - Problem defining risk/harm
  - Subsequent problem identifying risk/harm
- But: What are the ethical costs of not having this information?

Outcome

- The dataset is widely hosted and used to influence policy debates.
- Investigators are now re-designing their methods so that the Means are ethically just, too (or at least legal).

Scenario 2:
Censorship Measurement

Technologist Reasoning

- Informed consent from the users
- Rely on URL lists of respectable organisations
  - (may include Falun Gong, pornography sites, etc.)
- “No one has yet been harmed”
  - Although data has been used against people
- “We will not know what governments think of these systems until someone ends up in jail”

Ethics Reasoning 1/2

- Internet as socio-technical system
  - Internet designed by a homogeneous group
  - Access to the Network is democratised
- Inherent knowledge & power imbalance/asymmetry
  - Relevant social norms often not understood by engineers
- Informed consent is meaningless if:
  - No intuitive understanding of personal data ecosystem,
  - No technical understanding of devices.

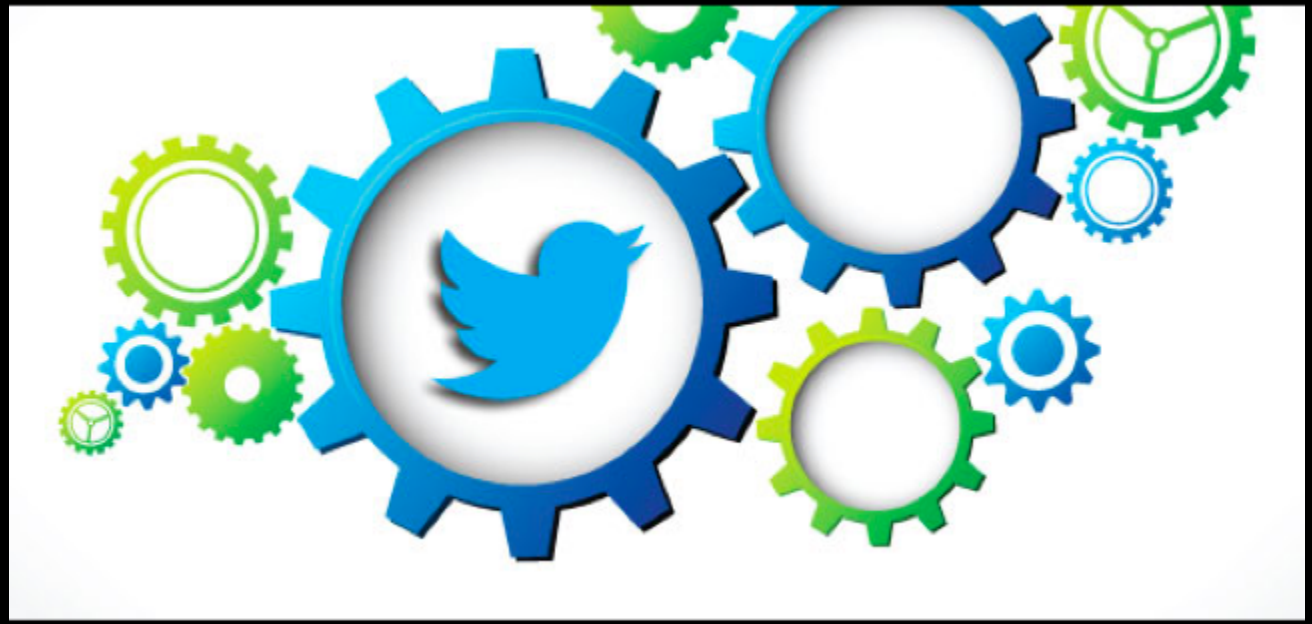
Ethics Reasoning 2/2

- Informed consent is also meaningless if engineers don’t understand:
  - Some websites may be particularly sensitive
  - Rule of law & Fair trial provisions may not exist in target country, or are secondary,
- Paying participants could lead to prosecution for (participating in) espionage
  - Therefore: speak to local lawyers or fellow engineers before deploying.

Outcome

- Some academic papers have been rejected on ethical grounds
  - However, now accepted because the data is particularly good,
  - Possibly with a note of caution attached.
- Projects have been scrutinised
  - But finding ways to operate

Scenario 3:
Publicly/Freely/Openly available data



Technologist Reasoning

- Data is available,
- Users publish data openly,
- Devices broadcast signals unprotected,
- Users should be aware of this,
- Therefore, it’s free to use.. Right?!

Ethics Reasoning

- Identifiers make it personal data – protected.
- New “public” space
  - Physical plane (observational research),
  - Digital plane (Wifi, Bluetooth, Tweets - technically mediated research).
- (Legal) Balancing test required.
- Unethical research - right reasons

Outcome

- Many projects collect tweets,
  - Three-tiered process agreed, opt-in or opt-out depending on sensitivity.
- Wifi/Bluetooth probes
  - Many commercial applications,
  - But can you expect users to know this amount of technical detail when they switch on the various antenna of their phone?

Emerging themes

- Internet as socio-technical system
  - Internet designed by a largely homogenous group,
  - Access has been democratised,
  - Relevant social norms of political context of data subjects may not be fully understood by researchers.

Emerging themes

- How to weigh facts and values for an ethical analysis
  - Defining and identifying potential harm
    - What is bad? How bad is it? In what way?
    - Complex, dynamic, technical environment
    - Fluid socio-political contexts
  - Clarification on the concept of privacy and other relevant values and rights,

Emerging themes

- Power imbalance/asymmetry,
  - Need for an increased awareness among engineers,
  - e.g. understanding consequences of behaviour in physical plane and the digital plane,
  - e.g. collecting data from non-participant's via mobile phones connections,
  - e.g. effects of creating direct links to participants (through remuneration).

Emerging themes

- Status of easily accessible data
  - Clarifying data protection law principles,
  - Applying the balancing test for 'legitimate interest'

- Meaningful informed consent
  - Make relevant to users,
  - What to include?
  - Which format?

Emerging themes

- Whether condoning potentially harmful research methods is desirable
  - Outlining some reasoning steps to take.
- Responsibility for all stakeholders in a research projects (inc. ethics boards)
  - To understand the consequences of a project, whereby the limitations should be set through conversation, rather than top-down decision making

Part IV: Exercise

- Create groups of 4
- Design an extremely useful project for Estonia
  - E-government
  - Internet research
  - Disregarding everything I just said
  - Based on much Estonian personal data collection and processing
- 20 minutes!

Exercise part 2

- Youngest two persons
  - Switch to nearby group
  - Become ethics board
- Explain project in detail to ethics board
  - Ethics board then scrutinises the design
  - Find compromises - re-design system together to make it ethically justifiable
- tinyurl.com/LaulasmaaEthics
- 20 minutes & Be ready to present!