ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

SPRING
SECURITY AND PRIVACY ENGINEERING LABORATORY

Information Security and Privacy (COM-402)

Part 2: Technical approaches to privacy

Carmela Troncoso

SPRING Lab

carmela.troncoso@epfl.ch

Some slides/ideas adapted from: George Danezis, Bart Preneel, Claudia Diaz, Seda Guerses, JP Hubaux, Bryan Ford

What is privacy in Privacy Enhancing Technologies

PETs

Not these PETs!!!

Two dimensions

1) **Privacy paradigms**: how privacy is defined

2) **Adversarial models**: who defines the problem and what are the goals

Gürses, Seda, and Claudia Diaz. "Two tales of privacy in online social networks." IEEE Security & Privacy 11.3 (2013): 29-37.
Diaz, Claudia, and Seda Gürses. "Understanding the landscape of privacy technologies." Information Security Summit (2012): 58-63.
Danezis, George, and Seda Gürses. "A critical review of 10 years of privacy technology." Surveillance cultures: a global surveillance society (2010): 1-16.

What is privacy: privacy paradigms

Privacy as
CONFIDENTIALITY

Privacy as
CONTROL

Privacy as
PRACTICE

FaKebook
Social Network

Privacy as

CONFIDENTIALITY

“The right to be let alone”
Warren & Brandeis (1890)

"the individual shall have full protection in person and in property."

PETs in this paradigm

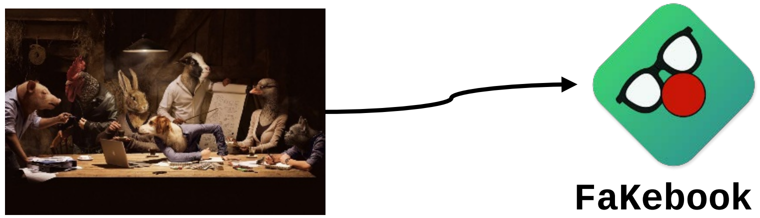
1) **Minimize data disclosure:** every bit counts
2) **Distribute trust:** avoid single points of failure
3) **Rely/require open source:** million eyes help security

In math we believe

strong proofs of security

Privacy as confidentiality

What would you do?



FaKebook
Social Network

Discuss with your neighbor(s) and propose one PET

> “The right to be let alone”
> Warren & Brandeis (1890)
>
> *"the individual shall have full protection in person and in property."*

Ts in this paradigm

1) **Minimize data disclosure:** every bit counts
2) **Distribute trust:** avoid single points of failure
3) **Rely/require open source:** million eyes help security

In math we believe
strong proofs of security

Privacy as

CONTROL

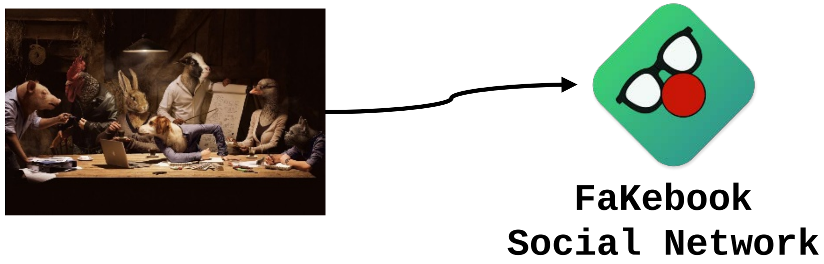
> "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others"
> Westin (1970)

PETs in this paradigm

1) **User participation:** let the user decide how data will be shared
2) **Transparency and Accountability:** let the user know how data is used, and if against his will point to who is responsible
3) **Organizational compliance:**
   General Data Protection Regulation (GDPR)
   Fair Information Practice Principles (FIPPs)

Privacy as control

What would you do?



Discuss with your neighbor(s) and propose one PET

> “the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others”
> Westin (1970)

Ts in this paradigm

1) **User participation:** let the user decide how data will be shared
2) **Transparency and Accountability:** let the user know how data is used, and if against his will point to who is responsible
3) **Organizational compliance:**
   General Data Protection Regulation (GDPR)
   Fair Information Practice Principles (FIPPs)

Privacy as

PRACTICE

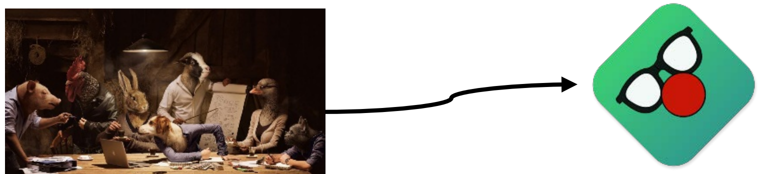
“the freedom from unreasonable constraints on the construction of one’s own identity”
Agre (1999)

PETs in this paradigm

1) **Improve user agency:** help them negotiate privacy
2) **Aid decision making and transparency of social impact:** help users understand the consequences of their actions
3) **Privacy as a collective practice**: help identify best practices for collectives

Privacy as practice

What would you do?



FaKebook
Social Network

Discuss with your neighbor(s) and propose one PET

> “the freedom from unreasonable constraints on the construction of one’s own identity”
> Agre (1999)

Ts in this paradigm

1) **Improve user agency:** help them negotiate privacy
2) **Aid decision making and transparency of social impact:** help users understand the consequences of their actions
3) **Privacy as a collective practice**: help identify best practices for collectives

A different PETS classification

This classification is not better or worse that the paradigms, **just different**

Paradigms are great help to understand different conceptions of privacy

- they are somehow hard to connect to privacy in real scenarios
- do not make adversarial / threat model explicit

We will now see another classification according to:

- who defines the problem (**thus the adversary**)
- what are the privacy goals (**what should be protected and how**)

They allow to see PETs limitations and the challenges they pose

Gürses, Seda, and Claudia Diaz. "Two tales of privacy in online social networks." IEEE Security & Privacy 11.3 (2013): 29-37.
Diaz, Claudia, and Seda Gürses. "Understanding the landscape of privacy technologies." Information Security Summit (2012): 58-63.
Danezis, George, and Seda Gürses. "A critical review of 10 years of privacy technology." Surveillance cultures: a global surveillance society (2010): 1-16.

1 – PETs for “social” Privacy

CONCERNS - The privacy problem is defined by Users

Technology brings problems

“My parents discovered I'm gay”

“My boss knows I am looking for other job”

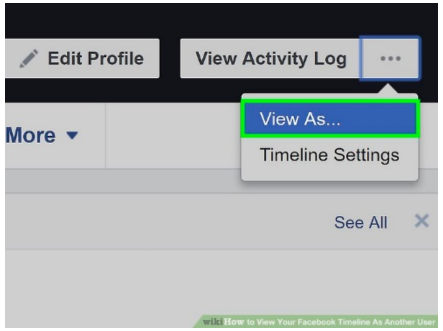
“My friends saw my naked pictures”

GOALS - Do not surprise the user

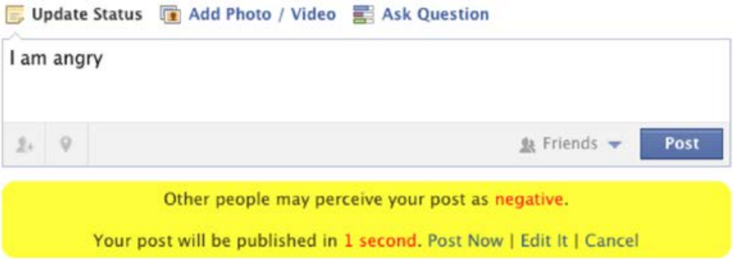
Two main approaches

Support decision making

Help identifying actions impact



Contextual feedback

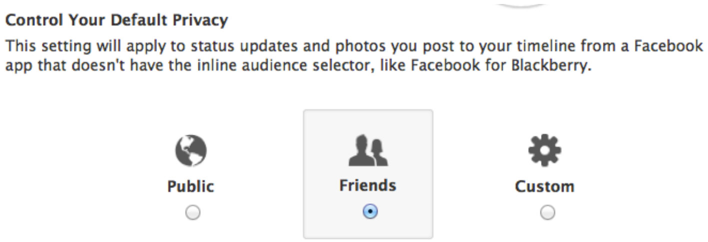


I am angry

Other people may perceive your post as negative.

Your post will be published in 1 second. Post Now | Edit It | Cancel

Privacy nudges



Control Your Default Privacy

This setting will apply to status updates and photos you post to your timeline from a Facebook app that doesn't have the inline audience selector, like Facebook for Blackberry.

○ Public ● Friends ○ Custom

Easy defaults

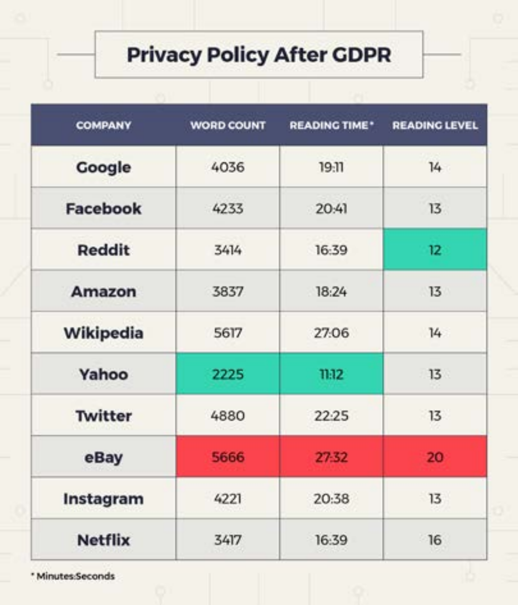
1 – PETs for “social” Privacy

LIMITATIONS

Only protects from other users: **trusted service provider**!

Limited by user’s capability to understand policies

Common Industry approach
Make users comfortable

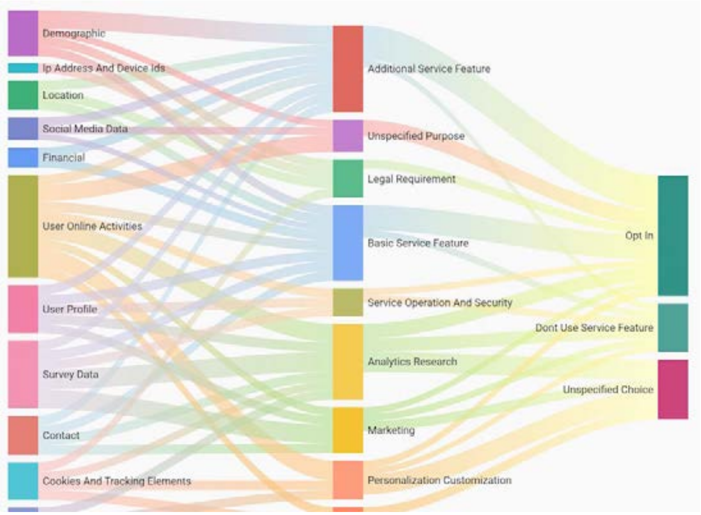


Privacy Policy After GDPR

COMPANY	WORD COUNT	READING TIME*	READING LEVEL
Google	4036	19:11	14
Facebook	4233	20:41	13
Reddit	3414	16:39	12
Amazon	3837	18:24	13
Wikipedia	5617	27:06	14
Yahoo	2225	11:12	13
Twitter	4880	22:25	13
eBay	5666	27:32	20
Instagram	4221	20:38	13
Netflix	3417	16:39	16

* Minutes:Seconds

https://www.varonis.com/blog/gdpr-privacy-policy/



https://pribot.org/polisis

1 – PETs for “social” Privacy

LIMITATIONS

Only protects from other users: **trusted service provider**!

Limited by user’s capability to understand policies

Common Industry approach

Make users comfortable

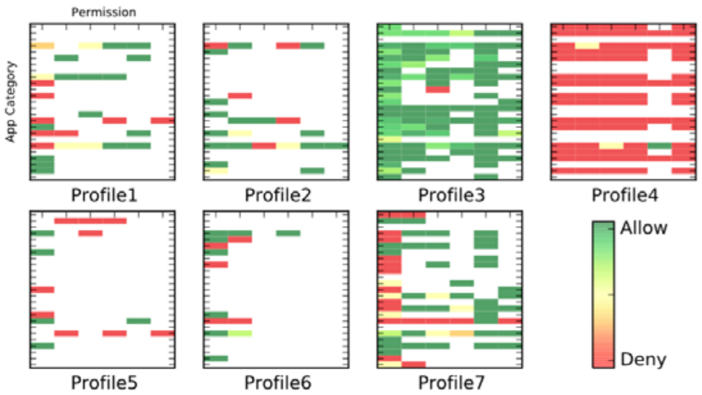


Figure 2: Privacy profiles learned from collected app privacy settings. Profile 1 is more protective on Location and Productivity apps than other profiles. Profile 2 denies phone call log permission more. Profile 3 is generally permissive. Profile 4 denies most permission requests. Profile 5 generally denies contacts, message, phone call log and calendar access, with only location and camera allowed for some apps. Profile 6 denies location and contact access of Social apps and Finance apps. Profile 7 is stricter regarding Social apps and location access in general.

Automated configuration

A. Is good for any user

B. Only works for average users

C. Only works for average apps

D. Only works for outliers

E. Has problems for everyone

Liu et al., Follow My Recommendations: A Personalized Privacy Assistant for Mobile App Permission, USENIX, 2016.

1 – PETs for “social” Privacy

LIMITATIONS

Only protects from other users: **trusted service provider**!
Limited by user’s capability to understand policies
Based on user expectations – What if the expectations are null?

Common Industry approach
Make users comfortable

YOU SEEM MUCH NICER THAN YOUR FACEBOOK PAGE.
SLANE

2 – PETs for “institutional” Privacy

CONCERNS - The privacy problem is defined by Legislation

Data **should not** be collected without user consent or processed for illegitimate uses

Data **should** be secured: correct, integrity, deletion

Personal data: any information that relates to an identified or identifiable living individual.

≠

Personal Identifiable Information (PII)
NIST Special Publication 800-122

any information about an individual maintained by an agency, including
(1) any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and
(2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information.

(1) is directly sensitive data
(2) combines data from same service or from different services

2 – PETs for “institutional” Privacy

GDPR

The General Data Protection Regulation

CONCERNS - The privacy problem is defined by **Legislation**

Data **should not** be collected without user consent or processed for illegitimate uses

Data **should** be secured: correct, integrity, deletion

GOALS – Compliance with data protection principles

informed consent

ty of data

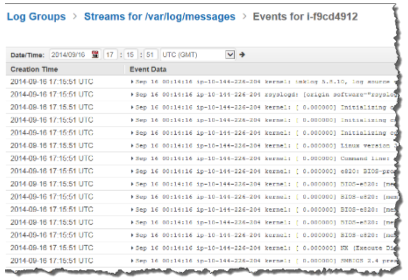
untability

Personal data

any information that relates to an identified or identifiable living individual.

Wouldn't it be nice if... you could take a dataset full of personal data, and transform it into one with no personal data – while keeping all the value of the data? **THIS IS HARD!!!!**

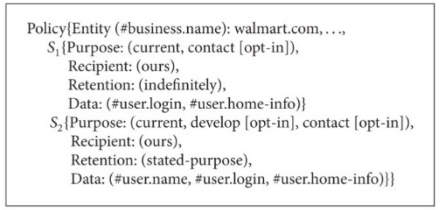
Access control


Log Groups > Streams for /var/log/messages > Events for i-f9cd4912
Date/Time: 2014/09/16 17 : 15 : 51 UTC (GMT)

Creation Time	Event Data
2014-09-16 17:15:51 UTC	Sep 16 00:14:16 ip-10-144-226-204 kernel: imklog 5.8.10, log source
2014-09-16 17:15:51 UTC	Sep 16 00:14:16 ip-10-144-226-204 rsyslogd: [origin software="rsyslog
2014-09-16 17:15:51 UTC	Sep 16 00:14:16 ip-10-144-226-204 kernel: [0.000000] Initializing c
2014-09-16 17:15:51 UTC	Sep 16 00:14:16 ip-10-144-226-204 kernel: [0.000000] Initializing c
2014-09-16 17:15:51 UTC	Sep 16 00:14:16 ip-10-144-226-204 kernel: [0.000000] Initializing c
2014-09-16 17:15:51 UTC	Sep 16 00:14:16 ip-10-144-226-204 kernel: [0.000000] Linux version
2014-09-16 17:15:51 UTC	Sep 16 00:14:16 ip-10-144-226-204 kernel: [0.000000] Command line:
2014-09-16 17:15:51 UTC	Sep 16 00:14:16 ip-10-144-226-204 kernel: [0.000000] e820: BIOS-pro
2014-09-16 17:15:51 UTC	Sep 16 00:14:16 ip-10-144-226-204 kernel: [0.000000] BIOS-e820: [me
2014-09-16 17:15:51 UTC	Sep 16 00:14:16 ip-10-144-226-204 kernel: [0.000000] BIOS-e820: [mem
2014-09-16 17:15:51 UTC	Sep 16 00:14:16 ip-10-144-226-204 kernel: [0.000000] BIOS-e820: [me
2014-09-16 17:15:51 UTC	Sep 16 00:14:16 ip-10-144-226-204 kernel: [0.000000] BIOS-e820: [me
2014-09-16 17:15:51 UTC	Sep 16 00:14:16 ip-10-144-226-204 kernel: [0.000000] BIOS-e820: [me
2014-09-16 17:15:51 UTC	Sep 16 00:14:16 ip-10-144-226-204 kernel: [0.000000] NX (Execute Di
2014-09-16 17:15:51 UTC	Sep 16 00:14:16 ip-10-144-226-204 kernel: [0.000000] SMBIOS 2.4 pre

Logging

Anonymization????



```
Policy{Entity (#business.name): walmart.com, . . .,
    S1{Purpose: (current, contact [opt-in]),
        Recipient: (ours),
        Retention: (indefinitely),
        Data: (#user.login, #user.home-info)}
    S2{Purpose: (current, develop [opt-in], contact [opt-in]),
        Recipient: (ours),
        Retention: (stated-purpose),
        Data: (#user.name, #user.login, #user.home-info)}}
```

Automated Policy Negociation

2 – PETs for "institutional" Privacy

LIMITATIONS

Assumes:

- collection and processing by organizations is necessary
- organizations are (semi)-trusted and honest
  - Reliance on punishment
  - No technical protection of the data

Focuses on limiting misuse, **not** collection

- Easy to circumvent minimization to collect in bulk
- Auditing may require more data!
- The danger of *informed consent*: if compliant is ok!

Limited

- Scope (personal data != all data)
- Transparency (proprietary software and algorithms)

Common Industry approach
Make users comfortable
+ Legal compliance!!

But does not prevent

3 – PETs for "anti-surveillance" Privacy

CONCERNS - The privacy problem is defined by **Security Experts**

Data is disclosed **by default** through the ICT infrastructure: **the adversary is anybody**

Concerned about: censorship, surveillance, freedom of speech,…

GOALS – Minimize

Default disclosure of personal information to anyone - both explicit and implicit!

Minimize the need to trust others

End-to-End encryption: Signal, PGP, OTR

Anonymous comms: Tor, mixnets

Obfuscation:
- dummy actions
- hiding
- generalization
- differential privacy

Advanced crypto:
- Private information retrieval
- Anonymous authentication
- Multiparty computation
- Blind signatures
- Cryptographic commitments

COM-402
and CS-523

3 – PETs for “anti-surveillance” Privacy

Limitations

Privacy-preserving designs are narrow – difficult to create “general purpose privacy”

Difficult to evolve – do not deal well with the Agile paradigm

Also difficult to combine

3 – PETs for “anti-surveillance” Privacy

LIMITATIONS

Privacy-preserving designs are narrow – difficult to create “general purpose privacy”

Difficult to evolve – do not deal well with the Agile paradigm

Also difficult to combine

Usability problems both for developers and users

how the @$%&#$Ŷ& do I program this?

performance hit

unintuitive technologies

3 – PETs for “anti-surveillance” Privacy

LIMITATIONS

Privacy-preserving designs are narrow – difficult to create “general purpose privacy”
Difficult to evolve – do not deal well with the Agile paradigm
Also difficult to combine

Usability problems both for developers and users
how the @$%&#$Ŷ& do I program this?
performance hit
unintuitive technologies

Lack of incentives:
- Industry: loses the data!
- Governments: national security, fraud detection, …

Takeaways

One can think about privacy technologies depending on:

- The privacy conception: confidentiality, control, practice

- The adversary model: other users, semi-trusted service provider, everyone

Each type of PETs present different challenges

Let’s exercise your privacy brain

You are developing a new app to rate beer bars in Lausanne. Rightfully, you decide that your target audience are **students**, your customers are the **bar owners**, and you will use a **Cloud provider** to host the application data.

Compare the following configurations in terms of privacy from the point of view of the students. Identify possible adversaries and what can they learn.

CONFIG A: The application gathers the recommendations from the students and then: lets other students see each other recommendations, and lets the bars see the student recommendations so that they can offer discounts to students that give good ratings.

CONFIG B: The application gathers the recommendations from the students and then: lets other students and the restaurant owners see the average rating for a restaurant**.**

Let’s exercise your privacy brain

From the second classification, what kind of privacy technologies would you use if:

- You want to protect the students social network (who is friends with whom) from the students they do not know

- You do not want the bar owners to learn which bar each student has visited, only the aggregates

- You do not want the cloud provider to learn what students connect

- You want the students to be able to evaluate how much other application users know about them

For each case, what privacy paradigm have you followed?