

User privacy in social media

Saptarshi Ghosh
and Mainack Mondal

CS 60017
Autumn 2019



Now we will talk about privacy

- Two broad dimensions
 - Preserving privacy from the background actors, e.g., advertisers or even the social media platform
 - Preserving privacy of data from other users, e.g., your ex

“What” of privacy?

Some slides borrowed from Blase Ur, UChicago

Fair Information practice principles (FIPP)

- **Notice/Awareness:** Consumers should be given notice of an entity's information practices before any personal information is collected from them

Fair Information practice principles (FIPP)

- **Notice/Awareness:** Consumers should be given notice of an entity's information practices before any personal information is collected from them
- **Choice/Consent:** Choice and consent in an on-line information-gathering sense means giving consumers options to control how their data is used

Fair Information practice principles (FIPP)

- **Notice/Awareness:** Consumers should be given notice of an entity's information practices before any personal information is collected from them
- **Choice/Consent:** Choice and consent in an on-line information-gathering sense means giving consumers options to control how their data is used
- **Access/Participation:** Not only a consumer's ability to view the data collected, but also to verify and contest its accuracy in inexpensive and timely manner

Fair Information practice principles (FIPP)

- **Notice/Awareness:** Consumers should be given notice of an entity's information practices before any personal information is collected from them
- **Choice/Consent:** Choice and consent in an on-line information-gathering sense means giving consumers options to control how their data is used
- **Access/Participation:** Not only a consumer's ability to view the data collected, but also to verify and contest its accuracy in inexpensive and timely manner
- **Integrity/Security:** Information collectors should ensure that the data they collect is accurate and secure

Fair Information practice principles (FIPP)

- **Notice/Awareness:** Consumers should be given notice of an entity's information practices before any personal information is collected from them
- **Choice/Consent:** Choice and consent in an on-line information-gathering sense means giving consumers options to control how their data is used
- **Access/Participation:** Not only a consumer's ability to view the data collected, but also to verify and contest its accuracy in inexpensive and timely manner
- **Integrity/Security:** Information collectors should ensure that the data they collect is accurate and secure
- **Enforcement/Redress:** In order to ensure that companies follow the Fair Information Practice Principles, there must be enforcement measures (self-regulation, sue by users, Government regulation)

Understanding privacy

We reviewed a number of definitions

Warren and Brandeis (1890)

Westin's definition (1967)

-
-
-

Solove's taxonomy of privacy (2008)

Nissenbaum's privacy as contextual integrity (2010)

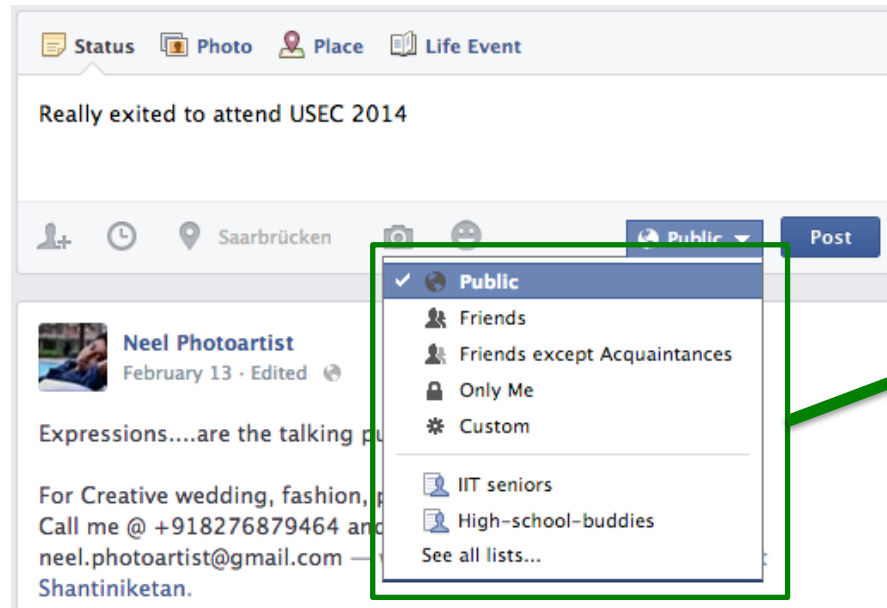


Identified different **aspects of privacy** from these definitions

A **subsequent** step is to **build mechanisms** to cover these aspects

“How” of privacy?

State of the art: Access control model



Allow others
access to content

Privacy violation from Access Control point of view:

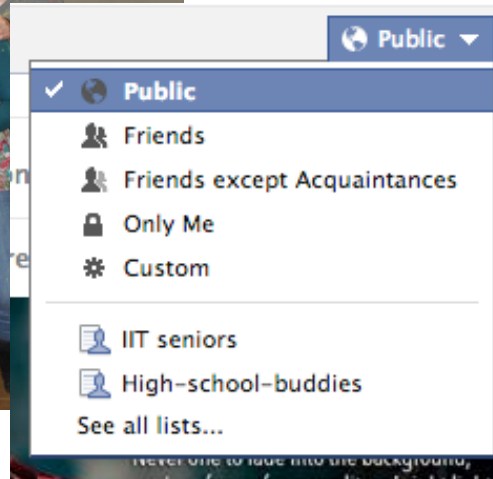
If someone accesses content who the user did not allow

Access control is inadequate to capture privacy

Exposure: A different concept to capture information privacy

Discussion: How to **manage privacy via exposure**

Privacy violations in the real world



Privacy violation in real world from user's point of view:

If someone accesses content who the user **did not intend**

ACLs are inadequate to capture many such privacy

Scenario 1: Facebook newsfeed

Facebook pushes your content as updates

Others **automatically get your content**
when they login to their Facebook page



After Newsfeed: **More** people actually saw the content

Users complained of **privacy violation** [Boyd et al. '08]

Before and **after** Newsfeed: **access control did not change!**

Scenario 2: Facebook timeline

Sort your content by upload time

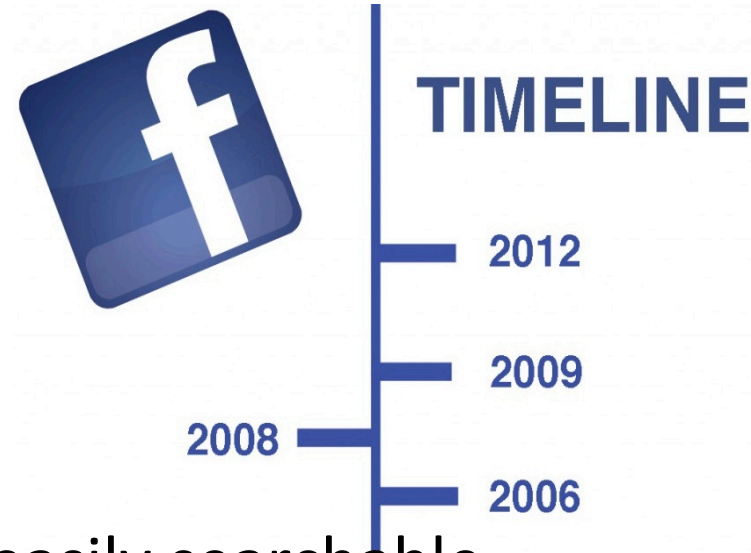
Others can **search by time**

After timeline: **Old** content became easily searchable

Users felt **privacy** was **violated**



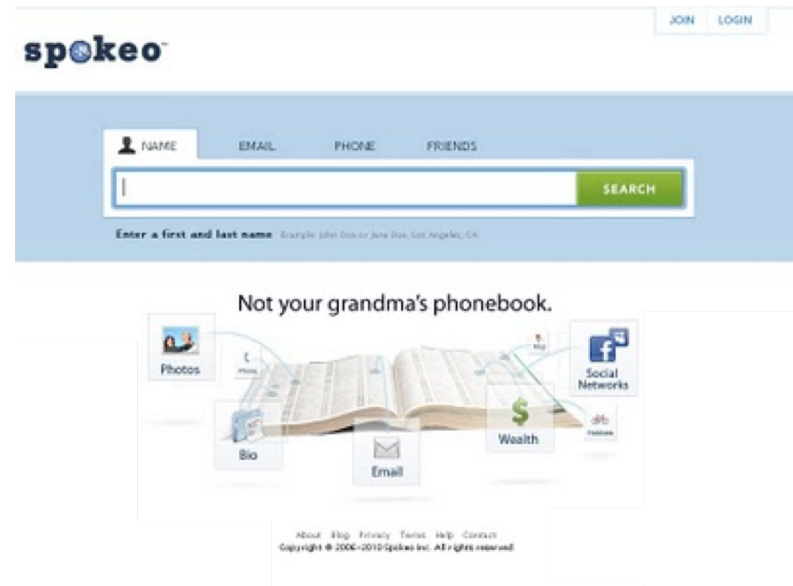
Before and **after** Timeline: **access control did not change!**



Scenario 3: Spokeo

Service aggregating public data from web

Others get all of this data by searching Spokeo



After aggregation: Inferring non public data become easier

Users complained of **privacy violation**



Before and **after** aggregation: **access control did not change!**

User reaction suggests each of the cases violated privacy

However access control was not violated in any of the cases

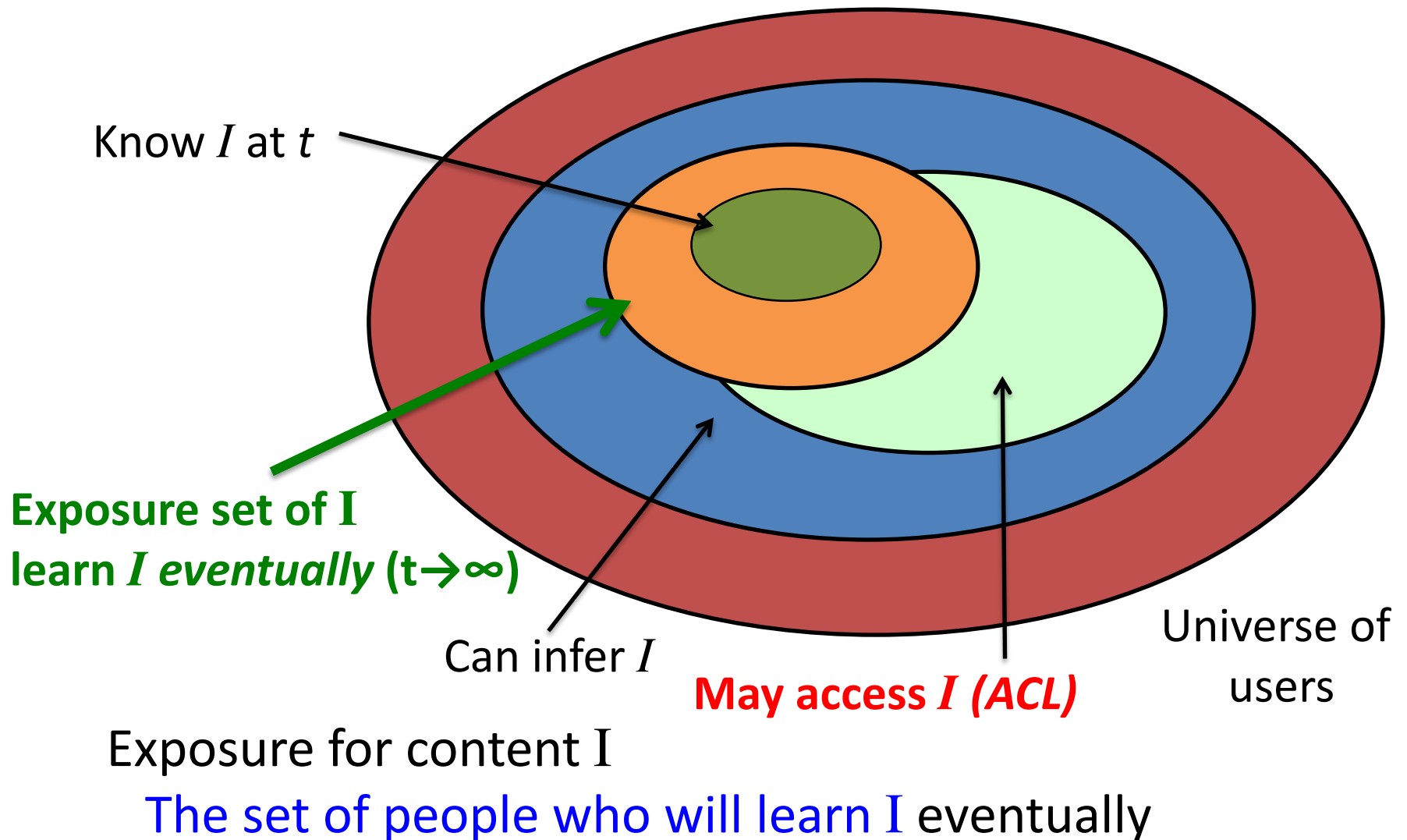
Take away 1: Access control is inadequate to capture user intention

Access control is inadequate to capture privacy


Exposure: A different concept to capture information privacy

Discussion: How to **manage privacy via exposure**

Exposure : Definition



How accurately do users estimate exposure?

Facebook researchers did a study with 589 users 
[Bernstein et al. 2013]

Perceived exposure grossly underestimates actual exposure



There may be a feeling of privacy violation when actual exposure is different from perceived exposure

Exposure in more detail

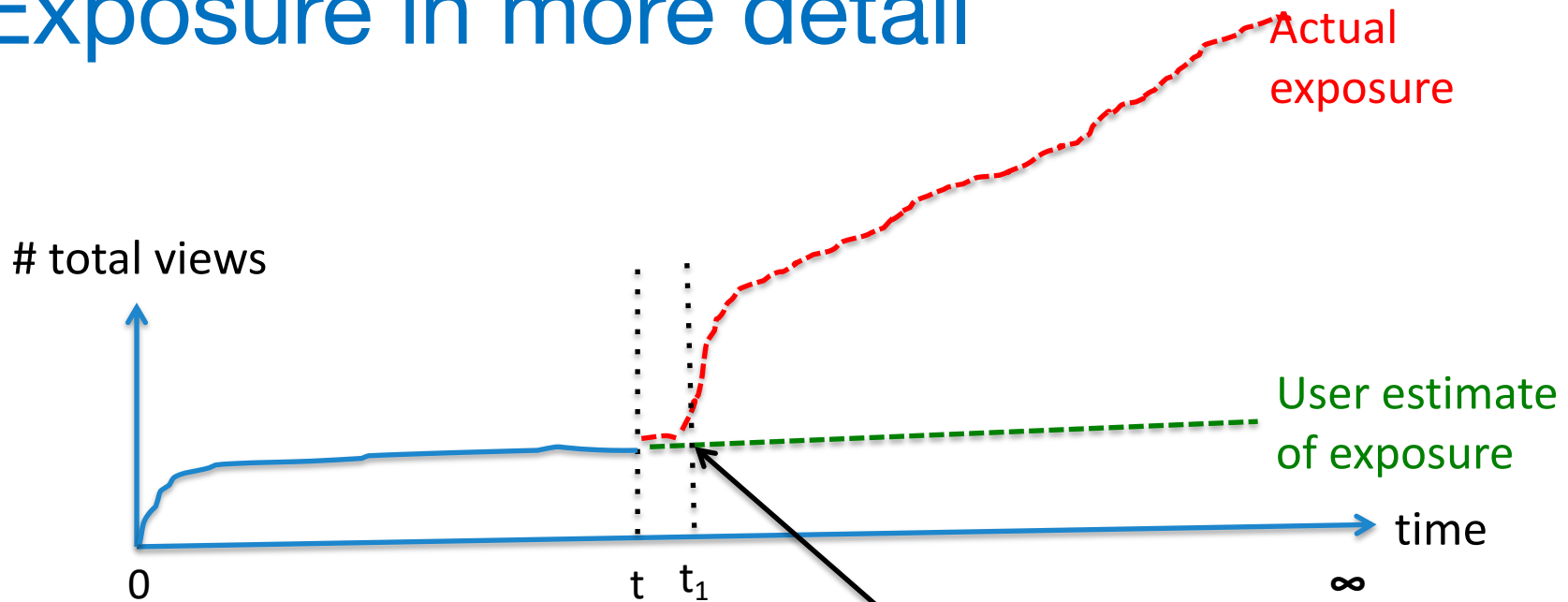


Photo uploaded and shared with public

 **reddit**
Posted in reddit

This is when users possibly start feeling their privacy is violated

Revisiting scenario 1: Facebook newsfeed

Exposure before newsfeed
Friends who visit profile



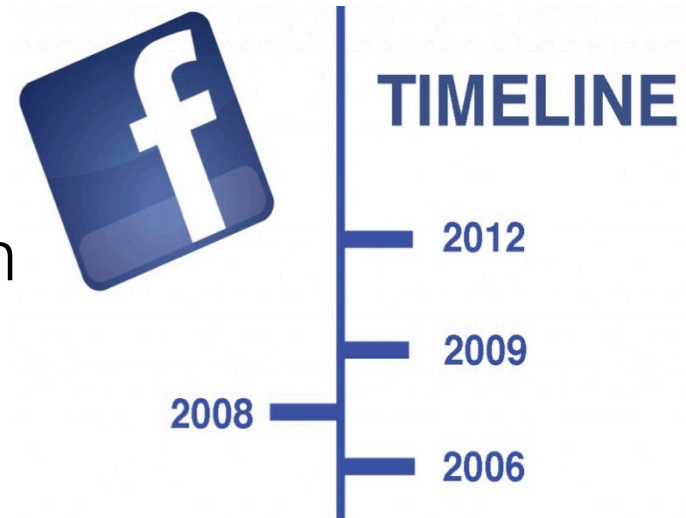
Exposure after newsfeed
All the friends who are logged into Facebook

Exposure of uploaded information **after newsfeed** > **Exposure** of uploaded information **before newsfeed**

Revisiting scenario 2: Facebook timeline

Exposure of old content **before** timeline
Users who will **scroll down**
thousands of content

Exposure of old content **after** timeline
All users who **search** by time



Exposure of old
information **after**
timeline

>

Exposure of old
information **before**
timeline

Revisiting scenario 3: Spokeo

Exposure before aggregation

Users who collect content
themselves from multiple sources



Exposure after aggregation

Any user who searches in Spokeo

Exposure of inferred
information **after**
aggregation

>

Exposure of inferred
information **before**
aggregation

Take away 2: Exposure based privacy model can capture violations which are not captured by access control

Recap

Access control is inadequate to capture privacy

Exposure: A different concept to capture information privacy

Discussion: How to **manage privacy via exposure**

Discussion: Managing privacy via exposure

Challenge 1:

How to estimate exposure for a content?

Challenge 2:

How to make users aware of the estimated exposure?

Challenge 3:

How to allow users more control over exposure?

Challenge 1: Estimating exposure

Situations where predicting exposure is very hard

Cross site prediction, exposure of inferred information

Situations where predicting exposure is possible

Predicting exposure of content in a site

Lots of research in content popularity growth

[Borghol et al] [Figueiredo et al.]

[Hong et al.] [Zaman et al]

[Bernstein et al.]



Challenge 1: Who can best estimate exposure

OSN operators are in the **best position to predict** exposure accurately with the data they collect

- They log who is accessing what content

- They collect historical data for content access

OSN operators can also **control** exposure

- They decide which content to show other users



Challenge 2: How to make users aware of the exposure?

Prediction can be shown to users at different granularity

- List of predicted people for a content

- Number of predicted people for a content

- Showing the prediction for a certain time period

- Showing the prediction with error bounds

- Showing how a specific dissemination mechanism changes the prediction

 - e.g., 200 more people are likely to see your content due to newsfeed

Challenge 3: How to allow users more control over exposure?

Different “knobs” can be provided to the user

- Change access control to a more restrictive setting

- Disabling particular dissemination mechanisms, e.g. search

- Enabling tripwires

 - Take content offline if more than 50 people view

 - Take content offline after two months

Take away 3: There are lots of open challenges and substantial research opportunities in how to design and deploy exposure based systems