17. Retrospective Security and Privacy

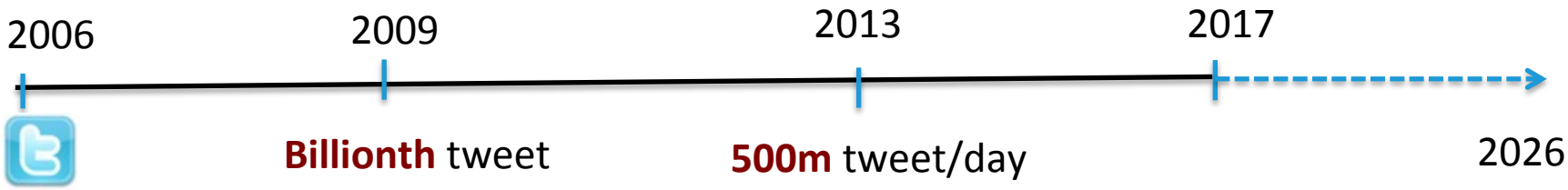
Blase Ur and Mainack Mondal
May 21st, 2018
CMSC 23210 / 33210

THE UNIVERSITY OF CHICAGO

Security, Usability, & Privacy Education & Research

Why should we care about retrospective security and privacy?

- Online social media sites (OSMs) are already around for a decade



- In sites like Twitter or Facebook
  - Users are content creators and managers
  - Might even need to change privacy preferences over time

OSM Users change privacy preferences over time

2009

Content posted in freshman year: shared with everybody on internet

2012

3 years later: Hiring manager and colleagues **should not** see this

They need to manage their data retrospectively: control **who can see old content**

What about cloud storage?

- At least one retrospective data management strategy is in the news for cloud storages

What about cloud storage?

- At least one retrospective data management strategy is in the news for cloud storages

https://www.usenix.org/sites/default/files/conference/protected-files/soups17_slides_ramokapane.pdf

Challenges of retrospective data management

- Why might users want to manage access settings of their data retrospectively?
  - Social media
  - Cloud storage

Challenges of retrospective data management

- Why might users want to manage access settings of their data retrospectively?
  - Social media
  - Cloud storage

- Life events, relationship change, relevance of content

Challenges of retrospective data management

- How would they do the management?

- How can we improve the usability?

Retrospective data management in social media

Why do users do retrospective data management in social media?

- Study by Ayalon et al., 2013
  - Between subjects experiment
  - 272 Amazon Mechanical Turk workers
  - Four condition groups – content age 0-1 years, 1-2 years, 2+ years, 0-2+ years (control group)
  - Total randomly selected 1,304 Facebook posts
  - Measured “willingness to share with friends” with likert scale

https://cups.cs.cmu.edu/soups/2013/proceedings/a4_Ayalon.pdf

Why would users do retrospective data management in social media?

- Analyzed the data with linear mixed model (LME)
  - Willingness to share decrease with age of the post
  - If a post had lesser relevancy to the social connections then users are less willing to share
  - Younger users are less willing to share old posts

- Similar study by Blase ("post anachronism study)
  - Willingness to share depends on life changes
  - With time for some posts users might want to share more and for others they want to share less

https://www.blaseur.com/papers/wpes13-temporality.pdf

Caveats of these studies?

Caveats of these studies?

- Have not measured what users actually did over long time for all their posts
- Considered random posts
  - Might missed the important content
- Did the study over relatively small period of time
  - At most two years

How about quantitative studies?

- Almuhmadi et al. did first study of deleted tweets in 2012
  - They listened to the tweets of 292,000 random users for one week
  - Total 67m tweets
  - Twitter also provided deletion notifications when the tweet was deleted
  - Total 2.4% of all tweets were deleted

https://www.cs.cmu.edu/~bliu1/Hazim_Almuhimedi_CSCW2013_Tweets.pdf

Difference between deleted and non-deleted tweets

- Almuhmadi et al. compared deleted and non-deleted tweets along multiple dimensions
  - Deleted tweets contain slightly more negative words than non-deleted ones
  - 17% of the tweets are deleted due to typos (how to detect?)

- Who are the users who deleted posts?

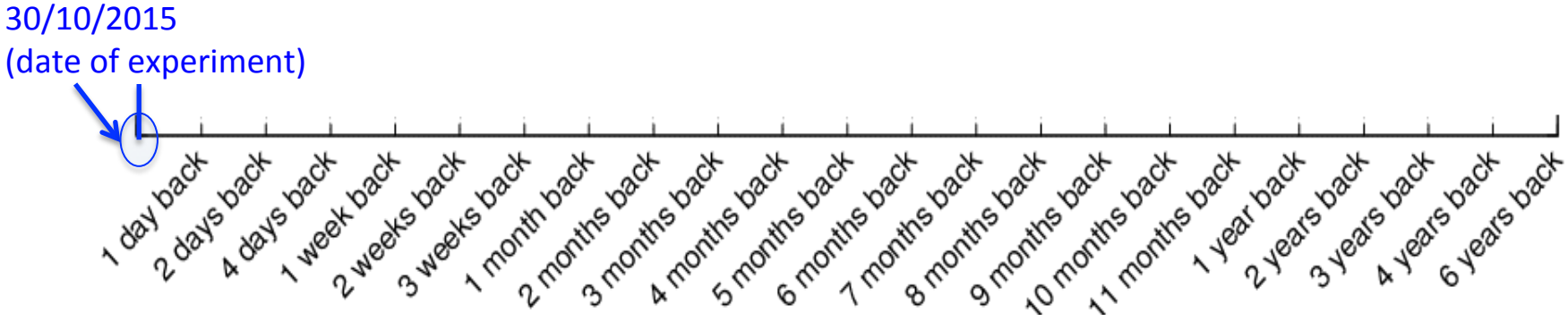
Difference between users deleting and not-deleting tweets

- Bhattacharya et al. did a similar study on 2016
  - 194k users are monitored over one month
  - Total 17m tweets out of which 7% were deleted
  - Measured big five personality traits (how?)
  - Deleters are less conscientious and more neurotic in their dataset

http://parantapa.net/mypapers/bhattacharya-icwsm16.pdf

How about a study for a longer period of time?

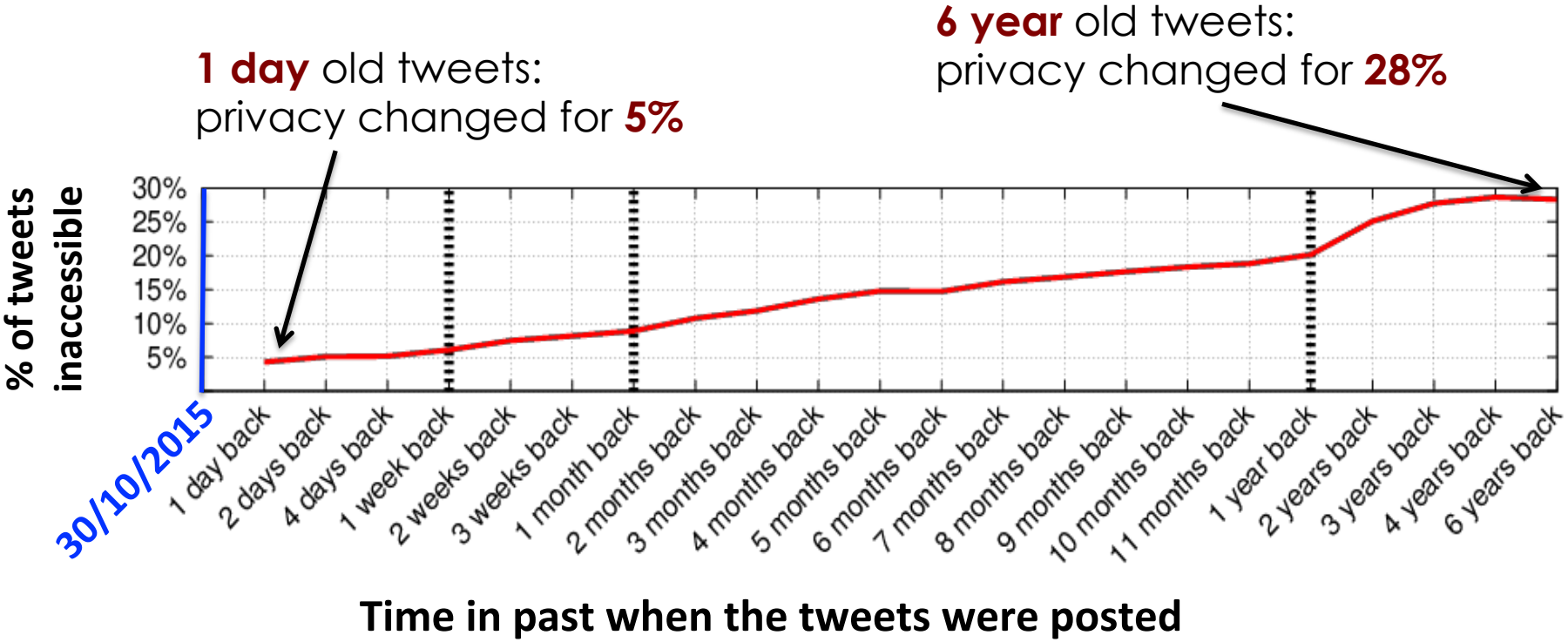
How to collect data?



Time in past when the tweets were posted (relative to the date of experiment**)**
ranges from tweets posted **1 day back** to **6 years back**

- All of these past tweets were public when they were posted
- If inaccessible on experiment date, privacy preferences changed over time

Do users employ retrospective mechanisms via deletion?



- Users change privacy for increasing amount of old data with time
- How do these users change privacy of this content?

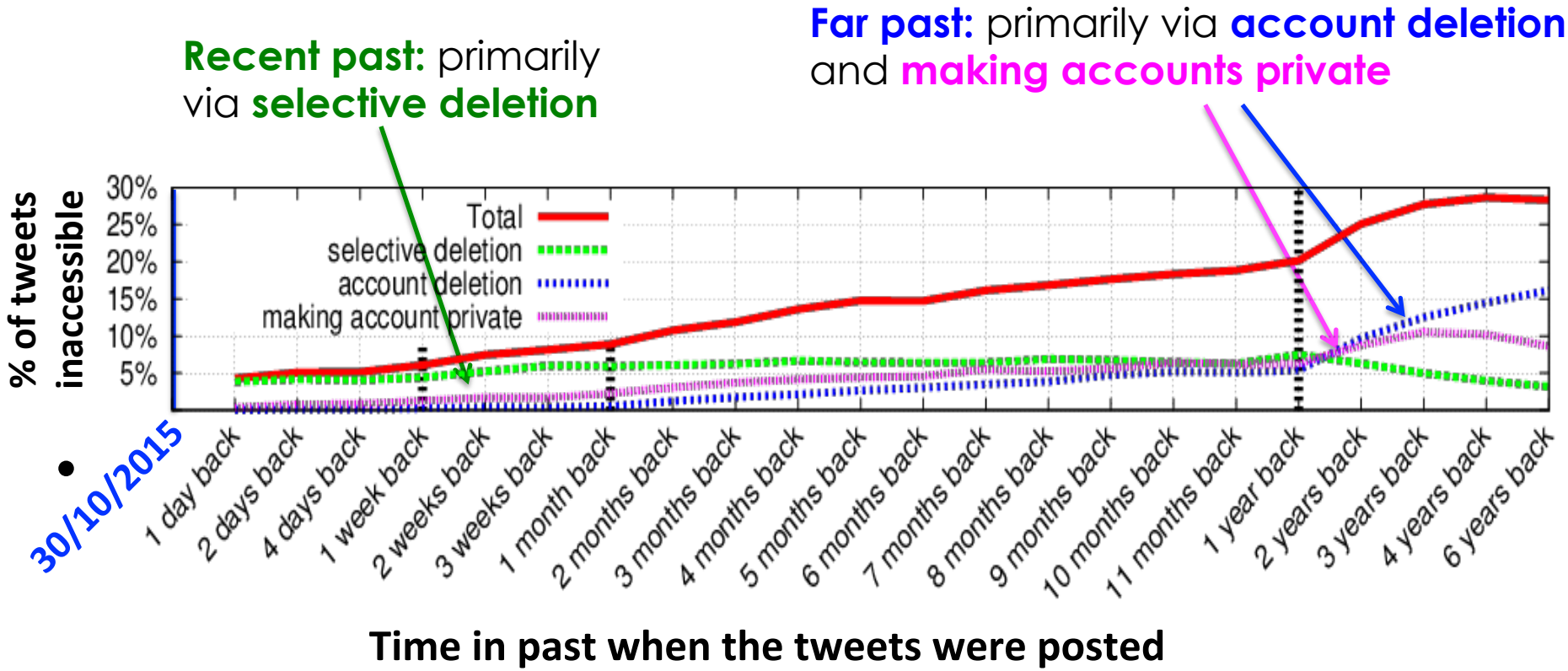
Retrospective mechanisms in Twitter

Three ways users change privacy of old content in Twitter

They are the longitudinal exposure control mechanisms

Mechanism	Description
Selective deletion	Selectively withdraw some old tweets to control exposure
Account deletion	Withdraw all old tweets to control exposure in bulk
Making account private	Withdraw all old tweets to control exposure in bulk

Relative usage of these mechanisms



Very different mechanisms to change privacy for content from far past compared to recent past

What fraction of users manage their old data?

- We randomly sample 100k active users from 2009
- Out of 8.9m random old tweets from these users 29.1% is inaccessible
- What fraction of users change privacy of their content?

User type	% of all users
Selectively deleted tweets	8.3%
Deleted their account	15.9%
Made their account private	10.4%
Users who take actions that changes privacy of their content	**34.6%**

Limitations of current deletion mechanims

- Two limitations
  - Limitation 1: Retaining residual activities
  - Limitation 2: Creating signal to identify possibly sensitive content

Limitation 1: Retained residual activities

These **conversations from** other users **remain public** even **after** a user **remove** her tweets/account

- These conversations are residual activities
  - Residual activities contain information about withdrawn old content
  - Anybody online can collect and analyze them by a username search

Sensitive user interests revealed by residual activities

- We checked user interests revealed for deleted/private accounts from 2009

Deleted/private accounts	Topics of interest from hashtags	Hashtags revealed by residual activities
Account 1	Politics, Sports, Technology	#iranelection, #prisoners, #strike, #frenchopen, #tech
Account 2	Sports, LGBTQ issues	#daviscup, #samesexsunday, #india, #lgbt, #followfriday
Account 3	Sports	#grandrapids, #nascar

Limitation 2: Creating signals to identify possibly sensitive content

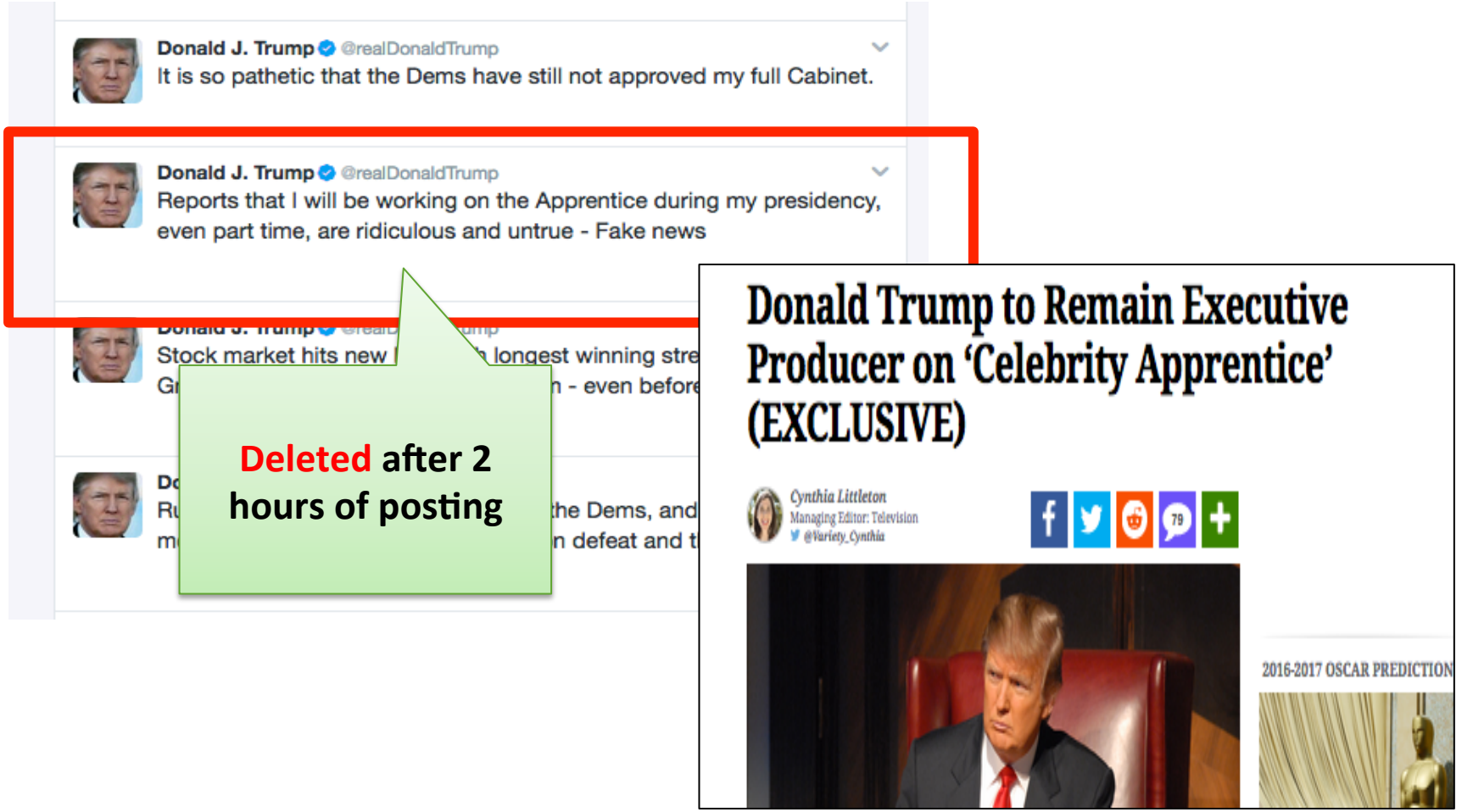
Donald J. Trump @realDonaldTrump
It is so pathetic that the Dems have still not approved my full Cabinet.

Donald J. Trump @realDonaldTrump
Reports that I will be working on the Apprentice during my presidency, even part time, are ridiculous and untrue - Fake news

Donald J. Trump @realDonaldTrump
Stock market hits new high with longest winning streak in decades. Great level of confidence and optimism - even before tax plan rollout!

Donald J. Trump @realDonaldTrump
Russia talk is FAKE NEWS put out by the Dems, and played up by the media, in order to mask the big election defeat and the illegal leaks!

Limitation 2: Creating signals to identify possibly sensitive content



Dealing with the limitations is difficult

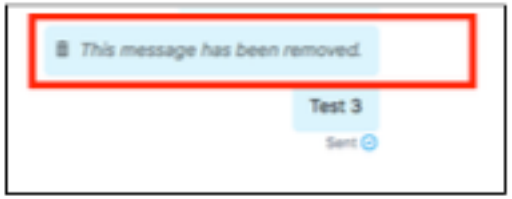
- Straw man:
  - Withdraw all the residual activities with original tweet/account
- Problem:
  - Residual activities are not “owned” by the original poster
  - Some OSMs solve it by age based withdrawal
  - Snapchat, Cyber dust
  - How does it solve the problem?
  - What additional problem can come up?

Other retrospective data management mechanisms

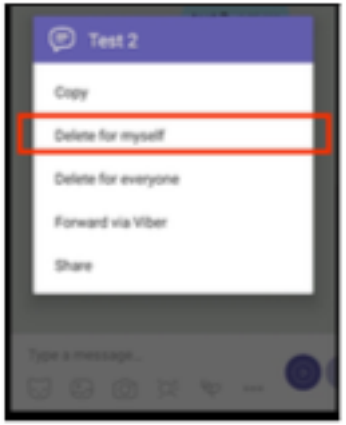
- So far we talked mostly about deletions
  - What other mechanisms can be there?

Other retrospective data management mechanisms

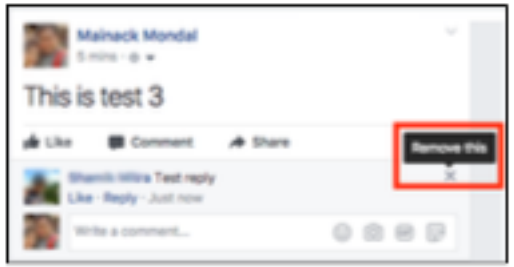
(a) Deleting individual post (no trace) in Facebook.



(b) Deleting individual post (with marker) in Skype.



(c) Deleting only owner's copy of content in Viber.

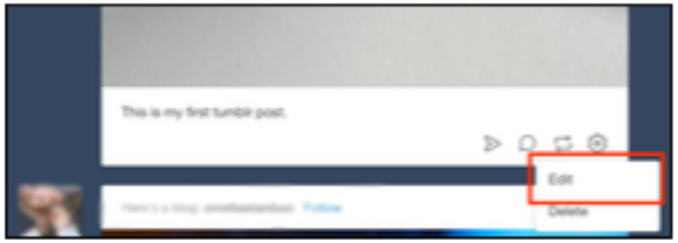


(d) Deleting content from the indirect exposure set in Facebook.

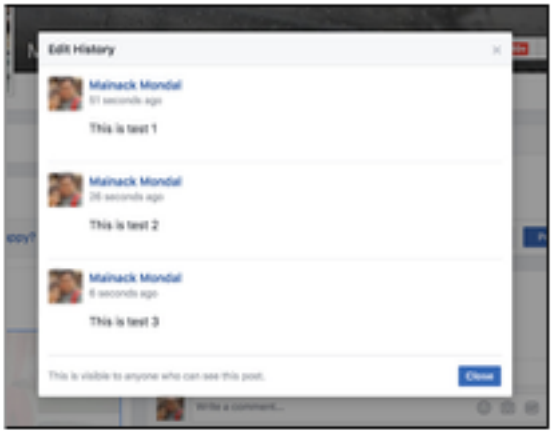
(e) Age-based deletion in Snapchat.

(f) Inactivity-based deletion in 4chan.

Other retrospective data management mechanisms

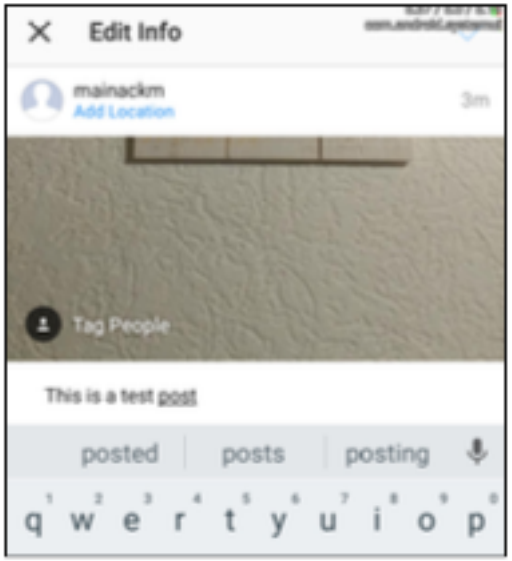


(a) **Edit post (no trace) in Tumblr.**



(b) **Edited post (with edit history) in Facebook.**

(c) **Edited post (with the small pencil icon as marker) in Skype.**



(d) **Editing only captions/tags in Instagram.**

(e) **Archive in Instagram.**

(f) **Anonymize in Reddit.**

Improving retrospective data management mechanisms

- What are the remaining challenges?

Improving retrospective data management mechanisms

- What are the remaining challenges?
  - Building systems to help users
  - Concern with shared nature of social content
  - Usability evaluations
  - …

Retrospective data management in cloud storage

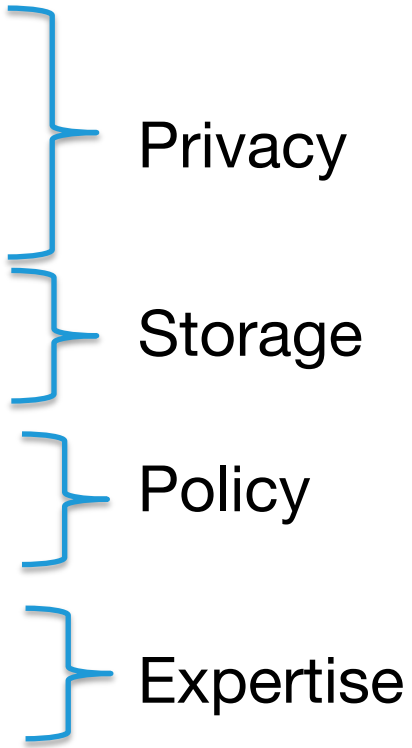
User perceptions of deletion in cloud storage

- Ramonkapane et al.: how do users retrospectively manage their data in cloud storage?
  - Qualitative approach
  - Semi-structured interview of 26 participants
  - Analysis by iterative coding
  - Goal: understanding perception about deletion in cloud storage

https://www.usenix.org/system/files/conference/soups2017/soups2017-ramokapane.pdf

User perceptions of deletion in cloud storage

- Motivations to delete from cloud storage
  - Lack of trust on provider
  - Avoid future conflicts
  - To forget
  - Tidying up cloud storage
  - Organizational policy compliance
  - Perceived value
  - Self-efficacy



User perceptions of deletion in cloud storage

- Reasons for deletion failures in cloud storage
  - Insufficient information in ToS, deletion messages,
  - Wrong mental models, e.g., in a shared folder *only* the user's copy will be deleted (like whatsapp)
  - User interface issues (specially in mobile)

How do users cope with deletion?

- Changing cloud providers
- Not even writing sensitive data to the cloud
- Deleting from one device because they can use the device better
- Seeking help from friends
- Deleting a different file to free up space
- Ad hoc strategies like buying more cloud storage
- Head in the sand – doing nothing

How about cloud storage deletion in real-world ?

- Ramonkapane et al. dealt with user perception
  - They have not checked what users might actually do
  - Blase and others did a survey study
  - Goal: To understand whether users manage access to old content
  - 100 Amazon mechanical users with 10 files (by stratified sampling)
  - They asked for each piece of content whether the user want to keep the content or delete or encrypt the content
  - They asked other questions regarding the content itself

How about cloud storage deletion in real-world ?

- They used a mixed effect logistic regression for data analysis
  - 48% of participants wanted to delete/encrypt half of more of all the files shown
  - Encrypt decisions were mainly driven by privacy
  - Decisions to delete a file was primarily driven by lack of future use
  - Similar file types should have similar management decision

https://www.blaseur.com/papers/chi18retrospectivecloud.pdf

Improving cloud storage data management

- Users want to manage a huge amount of data
  - They end up not managing much
  - They need better tools to browse through their online archive
  - Is it possible to come up with some automated system to help?
  - One relevant example: Cloudsweeper