A Study on the Unawareness of Shared Photos in Social Network Services

Benjamin Henne[*], Marcel Linke[*] and Matthew Smith[+]

[*] Distributed Computing & Security Group, Leibniz Universität Hannover, DE
henne@dcsec.uni-hannover.de

[+] Usable Security and Privacy Group, Friedrich-Wilhelm Universität Bonn, DE
smith@cs.uni-bonn.de

IEEE Security & Privacy Workshops: Web 2.0 Security & Privacy 2014 (W2SP'14), *May 18 2014*

The state of the photo privacy discussion

- Many privacy issues concerning photo privacy have been discussed at great length in the media…
  - Drunken pics, sexting, embarrasing locations
    - Accidentally published to more people than planed
    - Careless publishing “in the moment”
    - Malicious sharing by receiving party
  - Can be found and used by
    - News Corporations
    - Insurance Companies, etc.
    - Employers
    - Friends/partners

Social Media Threat

- Microsoft's Scott Charney offered a very good example during his Keynote speech at the RSA Conference 2012:

If a friend takes a picture of me during a volleyball game, shares this picture with other friends and one of them uploads the picture to the web, my insurance company can find and use that picture against me.

This is happening

- There have been reports that insurance companies are looking for just such information which could raise premiums or even deny claims.[1]
- The same is true for banks and credit rating companies.[2]

[1] http://abclocal.go.com/kabc/story?section=news/consumer&id=8422388

[2] http://www.betabeat.com/2011/12/13/as-banks-start-nosing-around-facebook-and-twitter-the-wrong-friends-might-just-sink-your-credit/

Privacy Threats & Metadata

1. Associate photo to person
   - Non-technical: person is recognizable on photo
   - Technical: image metadata contains link (name, unique identifier)
2. Photo contains objectionable content
   - Non-technical: image shows embarrassing actions or setting
   - Technical: image metadata contains objectionable entries like:
   - time, location, personal references

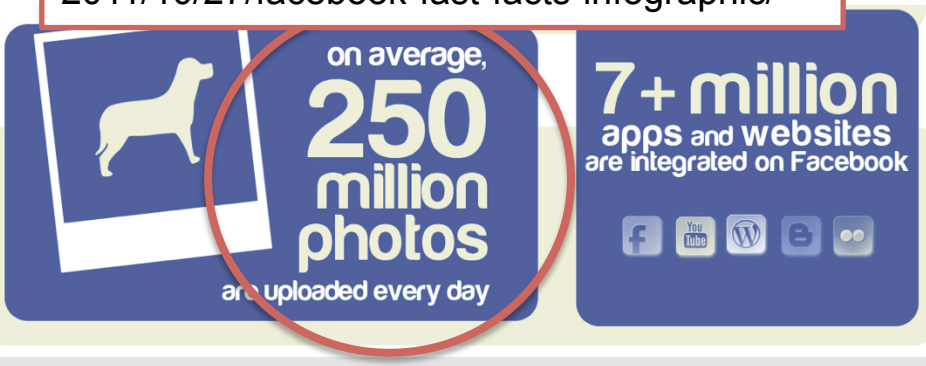
- Metadata increasingly is automatically added and users may not be aware of embedded metadata.

The problem of scale

"More than 250 billion photos have been uploaded to Facebook, and on average more than 350 million photos are uploaded every day

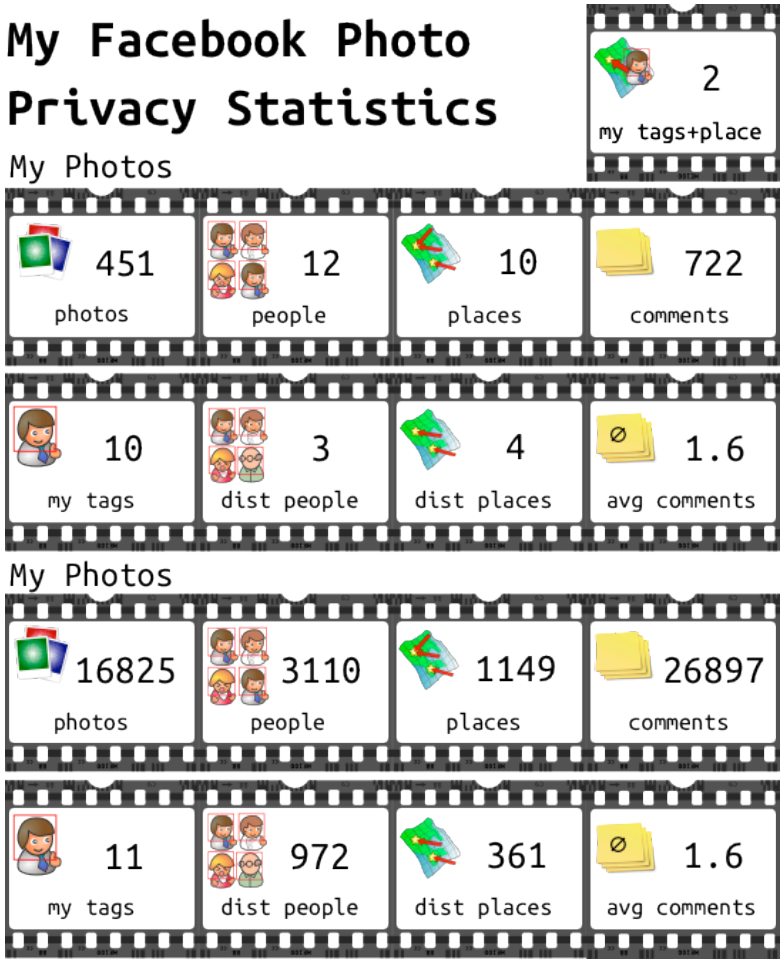
A Focus on Efficiency, whitepaper, Sept. 2013

http://www.socialmediadelivered.com/2011/10/27/facebook-fast-facts-infographic/



How much do these concern me?

Facebook App: Photo Privacy Statistics — a user's result



My 295 friends share at least 16825 photos with 3110 person tags and 1149 place tags. 26897 comments have been made to those photos.

I was tagged 11 times, 2 times on photos with a location tag. 972 other people were tagged as well. 361 different places were tagged.

18.6 % of my friends do not share photos or deny access to photos for apps others use.

What about you?
Try the Photo Privacy Statistics app!

Blind spots – “apps others use”

113 *initial* users – research group friends – mostly academics

79 users – recruited via *radio* broadcast

2561 users – recruited via *yellow press* online news article

- 30% of friends shared no photos (with our app)
  - those potentially activated privacy setting denying access for “apps others use”, since only few people share absolutely no photo on FB

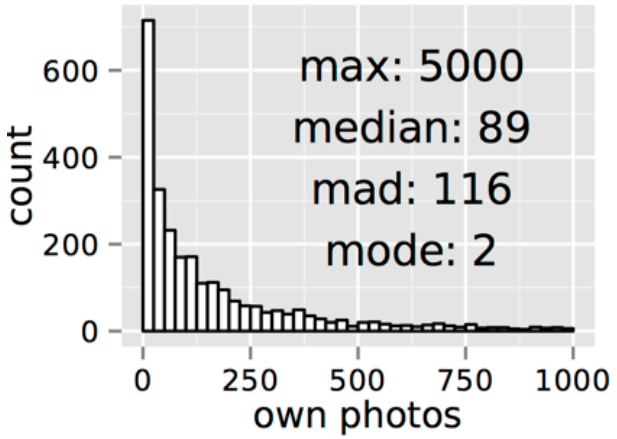
  35.1% of initial (academics) group

  32.7% of radio group

  26.2% of yellow press readers *(differed significantly from others)*

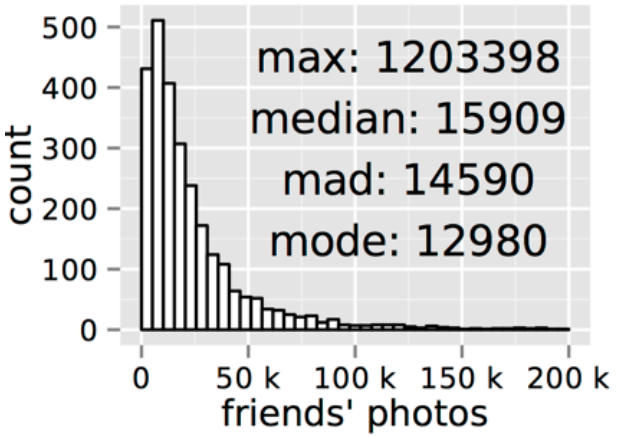
Facebook App: Photo Privacy Statistics — the dataset

- 2753 app users
  - 84.4% male, 15.1% female
  - age: 13–77 years, mode = 26
  - avg. 296 friends



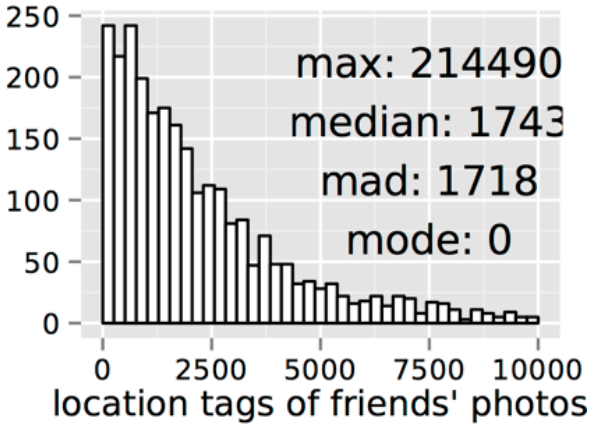
- 572K of 817K direct friends shared photos with app
  - 30% did not – potentially disallowed for “apps others use”

- 75.7M photos in sum
  - 99.2% shared by direct friends

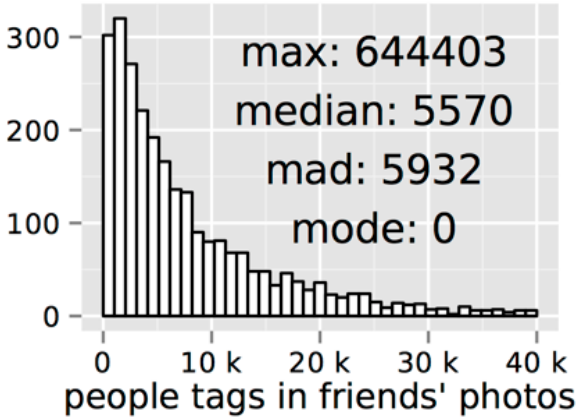


Facebook App: Photo Privacy Statistics — the dataset

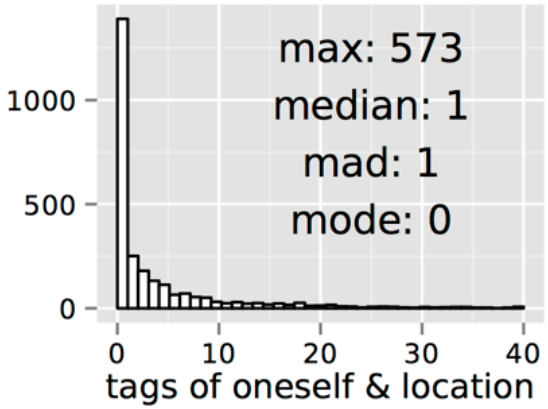
- 11.3% – 8.5M photos had a **location tag**
  - 610K different locations



- 22.4% – 17M photos contained **person tags**
  - 34M tags with profile links
  - 6M different people



- Tags of a user
  – 63.9% were tagged
  22.5% 1x
  14.2% 2x
  25.4% >10x



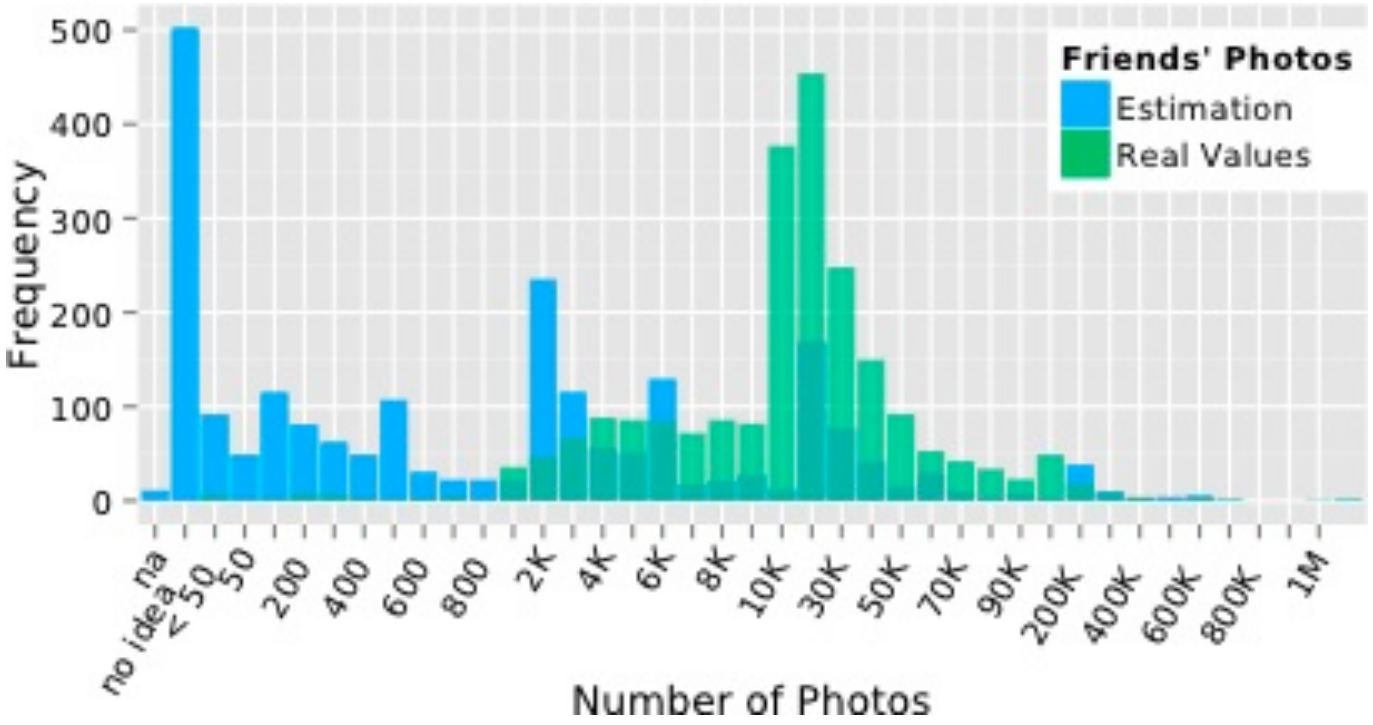
User Study: Preceding Questionnaire

- 2245 participants, demographics virtually identical to app users'

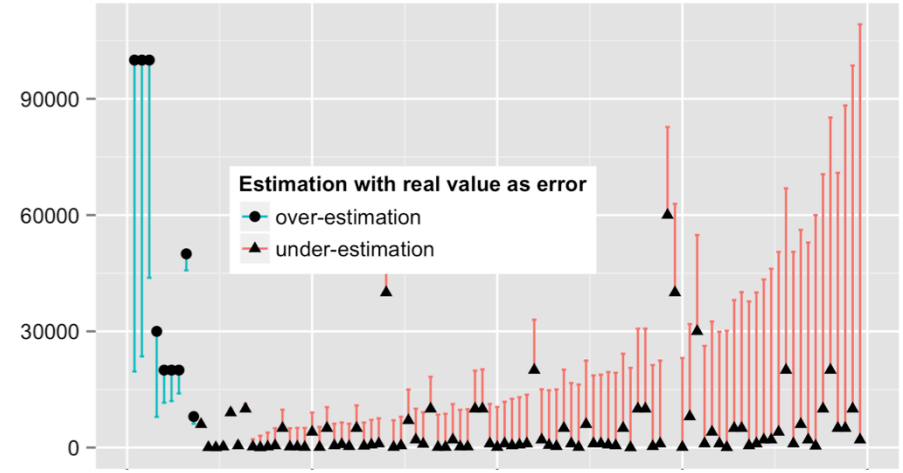
1. How many photos shared by **all** your friends can you **altogether** view?
   no answer, no idea, 50, 100, ..., 1000, 2000, ..., 10000, 20000, ..., 1M, >1M
2. How many photos that your friends share have a **location tag**?
   no answer, no idea, <10%, 10%, 20%, 33%, 50% >50%, 100%
3. How many photos that your friends share have a **person tag**?
   no answer, no idea, <10%, 10%, 20%, 33%, 50% >50%, 100%
4. How many **people** are **tagged in a photo with people tags** on the average?
   no answer, no idea, 1, 2, ..., 10, >10

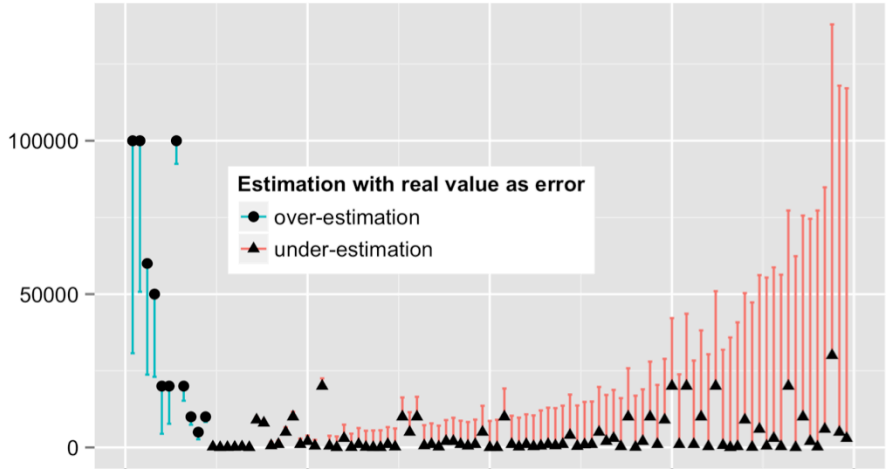
Estimation of Friends‘ Photos

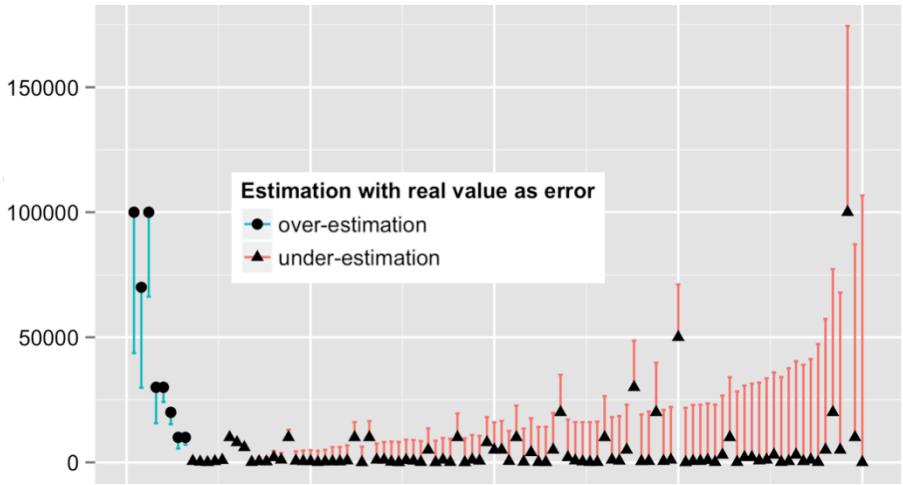
- Users‘ estimations
  - median/modus = 1,000
  - $Q_{25\%}$ = 400
  - $Q_{75\%}$ = 8,000
- Real values
  - median/modus = 15,909
  - $Q_{25\%}$ = 7,722
  - $Q_{75\%}$ = 30,687

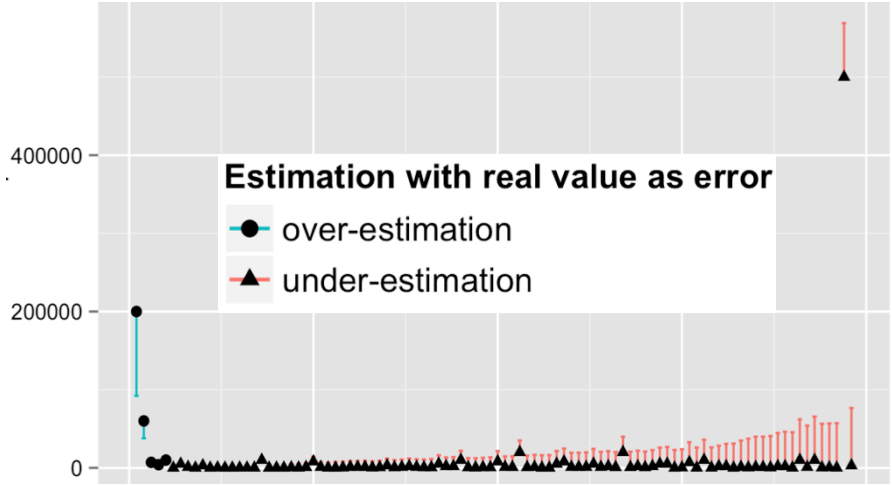


Friends‘ Photos: Estimations vs. Real values – absolute
4x 100 users random subsample









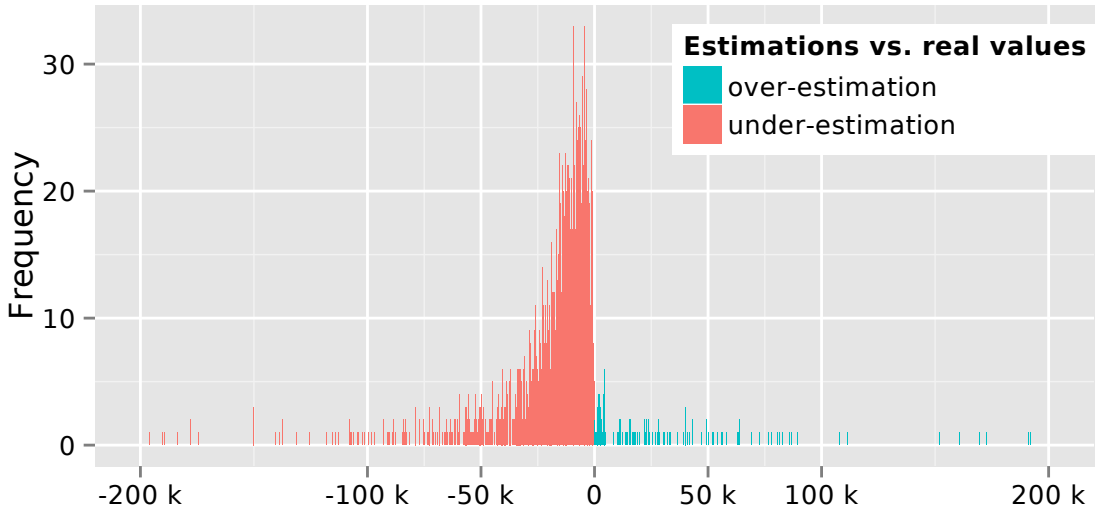
Misestimation of Friends' Photos

- estimation $e=anwsers_n$ **correct, iff** $anwsers_{n-1} <$ real value $\leq anwsers_{n+1}$,

  answers = 0, 25, 50, 100, 200, 300, …, 900, 1000, 2000, …, 9000, 10000, …

  - 8.2% of estimations were correct

- *Misestimation = estimation - real value*
  - 8.6% over-estimated
  - **91.4% under-estimated**



Friends‘ Photos: Magnitude of Factor of Misestimation

- **Factor** of Misestimation ranged from -38,989 to 258

Magnitude of Factor

0.5% no answer

22.4% no idea

6.3% correct

5.6% overestimated magnitude 1

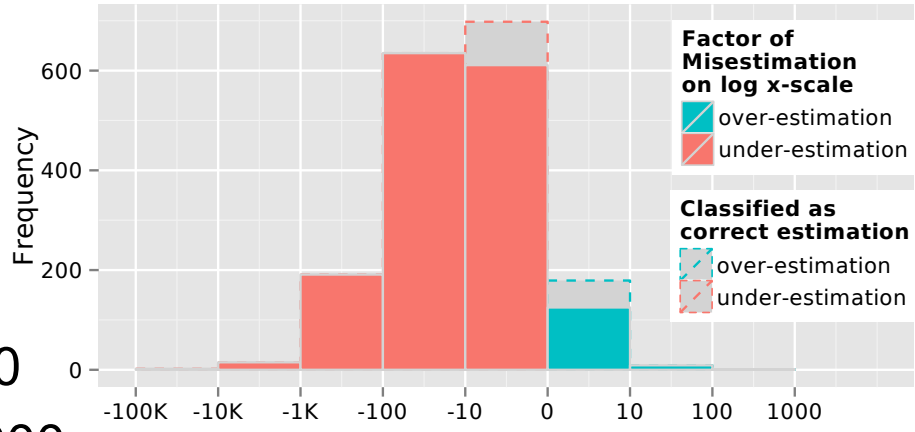
0.5% overestimated magnitude 10+

27.2% underestimated magnitude 1

28.3% underestimated magnitude 10

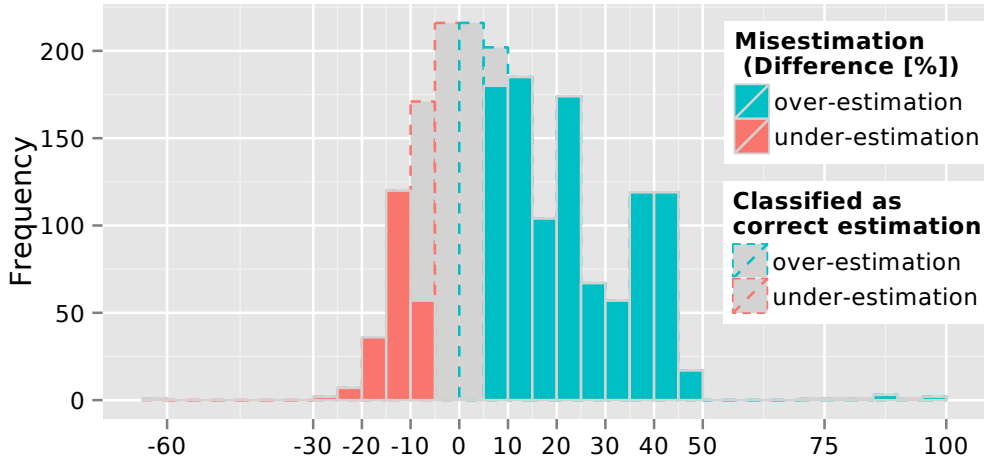
8.5% underestimated magnitude 100

0.7% underestimated magnitude 1,000



Estimation of Friends' Photos with Location Tag

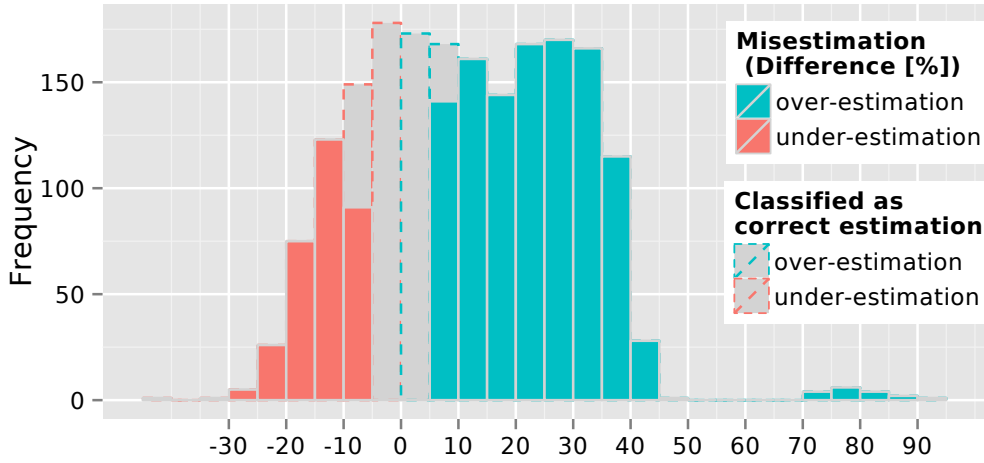
- Estimation: median = 20%, $Q_{25\%}$ = 10%, $Q_{75\%}$ = 33%
  Real values: median = 10.8%, $Q_{25\%}$ = 8.4%, $Q_{75\%}$ = 13.9%
- Correct, iff Real value closer to Estimation than to its neighbors or if interval matches
- All answers
  - 2.2% no answer
  - 16.6% no idea
  - 25.3% correct
  - 45.9% overestimation
  - 10.0% underestimation



- Participants seem to be more aware of location tags than of photo count
- Participants tend to overestimate the number of location tags

Estimation of Friends' Photos with Person Tags

- Estimation: median = 33%, $Q_{25\%}$ = 10%, $Q_{75\%}$ = 50%
  Real values: median = 17.8%, $Q_{25\%}$ = 13.5%, $Q_{75\%}$ = 22.7%
- Correct, iff Real value closer to Estimation than to its neighbors or if interval matches
- All answers
  - 2.8% no answer
  - 13.9% no idea
  - 23.3% correct
  - 49.5% overestimation
  - 14.4% underestimation



- Similar to location tags; participants tend to overestimate
- Estimations seemed to be more informed than for location tags: visible trend that estimations corresponded to real values in this case

Summary: Estimations and Unawareness

- Many participants were not aware of the amount of shared photos
  - They were not aware of the mass of photos that might raise concerns
    - and we only considered photos of direct friends
  - They mostly underestimated, which is the worse option
  - Even allowing a misestimation factor of magnitude 1, only 39% of all participants did not make a substantial false estimation

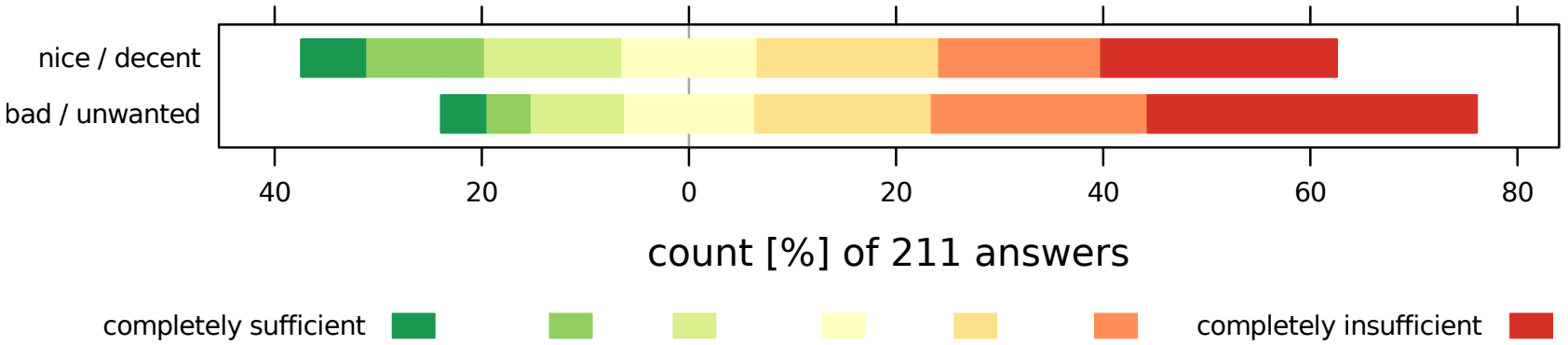
- Estimations of tags were more often correct
  - ➢ 25.3% for locations, 19.4% for person tags
  - Potential reason: Person tag notification – doesn't explain location tags
  - Participants mostly overestimated use of metadata
    - This could be dangerous because they expect to be notified of photos

User Study: Post-Questionnaire

- 269 participants
- demographics nearly identical to app users'
- invited via result notification email and at personal results page
- Time between result notification and participation
  median = 5 hours, $Q_{75\%}$ = 14 ours, $Q_{95\%}$ = 2 days

General Feelings about Photo Awareness

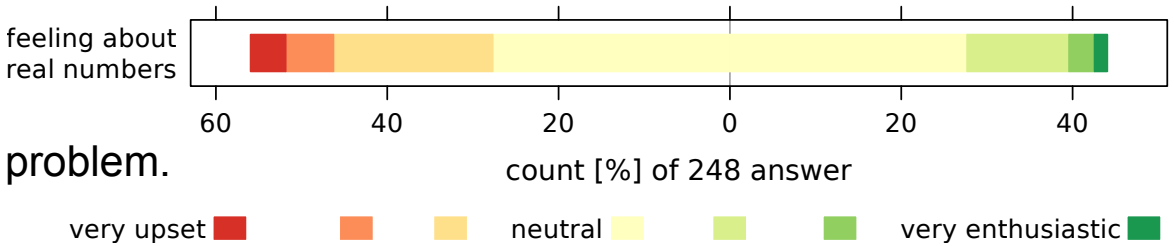
How well participants feel informed about photos on the Web



- Nice photos: 6% completely sufficient, 56% worse than neutral
- Bad photos: 4% completely sufficient, 70% worse than neutral

Interestingly, most stated not to be upset about app results.

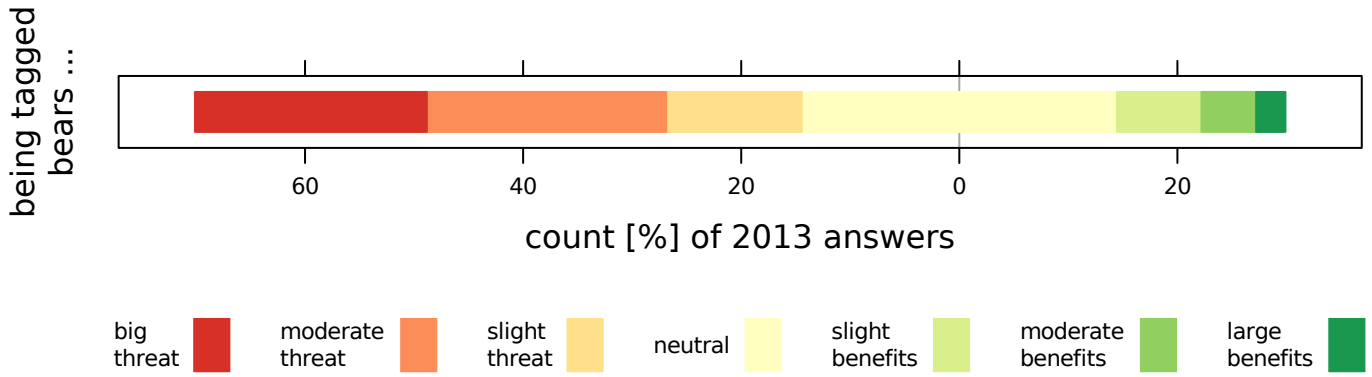
It seems, they have no concern here right now? Unaware of the problem.



Person Tags – Benefit or Threat to Privacy?

Prior work[1]: Being notified after having been tagged was the most often referred way (75% of participants) of getting to know of photos of oneself.

[1] = Henne and Smith: Awareness about Photos on the Web and How Privacy-Privacy-Tradeoffs Could Help, USec'13@FC'13



- Just 15.4% of 2,013 participants perceived person tags with notification as beneficial for their privacy.
  - 28.9% were neutral
  - 55,6% called it a threat

Reflecting on the Overall Dataset and Application

- The average app user's stats
  - 16K direct friends' photos, with a median of 5.5K person tags, 1.7K locations and 21.9K comments
  - ➢ Amount is already higher than what any use could manually review
- Privacy Invasion by Apps
  - Our app had just 2753 users. It had access to 75 million photos and had access to photos with person tags of 6.3 million people.
  - ➢ Such real world numbers are valuable and more are needed as basis for effective for privacy education
- Less than ⅓ used the privacy options "apps others use" to hide their data from apps. Does the lion's share not regards this as necessary? Or do they not know the option? This is worth working on.

Conclusion

- Participants' inability to estimate provides evidence for lack of awareness about the dimension of shared data and the potential threat to privacy.
  - Nr. of photos underestimated
  - Nr of tags overestimated

- Our empirical evidence highlights the need for new privacy-enhancing technology to cope with the huge amount of media shared by friends
- Current privacy settings do not deal with this topic particularly well
- Apps like the one presented in this work can provide people with valuable insights
  - as basis for re-thinking their habits on the Social web
  - As basis for privacy education

App removal

113 *initial* users – research group friends – mostly academics

79 users – gained via *radio* broadcast

2561 users – mostly caught via *yellow press* online news article

- Two weeks after app usage we tested for app removal
  - removal was suggested in results notifications and results page
  - 89.1% of users had **not** removed app/permissions

  92.1% of initial (academics) group removed it

  16.5% of radio group removed it

  7.1% of yellow press readers removed it *(all pairs differed significantly)*