

# Over-Exposed? Privacy Patterns and Considerations in Online and Mobile Photo Sharing

Onur Cobanoglu

09.28.2009

# Aim of the Study

- The study is focused on privacy decisions of Flickr/ZoneTag/users sharing photos taken by context-aware cameraphones.
- Trying to find answers to the following questions:
  - Which factors influence privacy decisions for online photos?
  - What is the effect of photo content on privacy decisions?
  - Do people change privacy decisions while being mobile afterwards and how often?
  - What is the user sensitivity against exposing the location of the photos?

# Why Do They Care?

- Such information may be used in the design of the sharing software such that the system:
  - May warn the user about her privacy decision when it finds the decision ‘potentially regrettable’.
  - May recommend privacy levels for the shared photos to uncertain users.
  - May warn the user about aggregations of disclosed information which the user may potentially find undesirable.

# Description of the Studied System

- **Flickr:** Online photo sharing service.
  - Provides five different privacy levels: Private, family-only, friends-only, friends-and-family, public.
  - In the study, these are abstracted to two levels: Public and non-public.
  - Privacy settings of the photos and the relationship level of the users can be changed at any time.

# Description of the Studied System

- **ZoneTag:** Cameraphone software for photo sharing via Flickr.
    - Photos can be uploaded with the previous or new privacy settings.
    - Tag assignment is encouraged via several features.
    - Uses cell-tower information to expose the capture location, converted human-readable location labels.  
**Any user who can see the photo can see the location.**
- LOCATION EXPOSURE PATTERNS IS PARTICULARLY IMPORTANT FOR CONSUMER MODELING!**

# Methodology

- **Methodology:** Quantitative statistical analysis over ZoneTag usage logs and qualitative analysis over interviews with users from selected groups.
- Basically, quantitative analysis aimed to discover user trends on privacy settings and qualitative analysis aimed to find underlying human factors affecting these settings and discover user trends which cannot be captured by the quantitative analysis.

# Quantitative Analysis

- Among the ZoneTag users, those having uploaded at least 40 photos are selected. This corresponds to:
  - 81 of the 350 users
  - 36,915 of the 44,000 photos
- Reasons for this decision:
  - The sample covers a large portion of the population due to the underlying power law distribution (see the numbers above).
  - Gives ability to study variation within a user's decision over time.
  - Recognizable behavior patterns cannot be established for lower photo per user ratios.

# Quantitative Analysis

- Data includes:
  - Time and location (at cell granularity) of the photo capture.
  - Privacy settings and tags for the photos set using ZoneTag **before the upload**.
  - Subsequent changes made to these settings via Flickr's web interface.
- Tried to find answers to the questions:
  - Does location predict privacy decisions?
  - Does content predict privacy decisions?
  - Are capture-time privacy decisions 'good'?
  - Are users concerned about exposing their location?

# Does Location Predict Privacy Decisions?

- Is divided into two questions:
  - Are there locations in which user behavior on privacy decisions significantly differ from the overall behavior? (H1)
  - Does the frequency of photographing a location significantly affects the privacy decision for that location? (H2)
- H2 was tested with the expectation that more frequently photographed locations are more private.

# Testing H1

- For each location-user pair, the quantity:

$$\frac{\text{\#of total public photos}}{\text{\#of total photos}}$$

Overall Behavior

$$\frac{\text{\#of public photos in the location}}{\text{\#of photos in the location}}$$

Location Specific Behavior

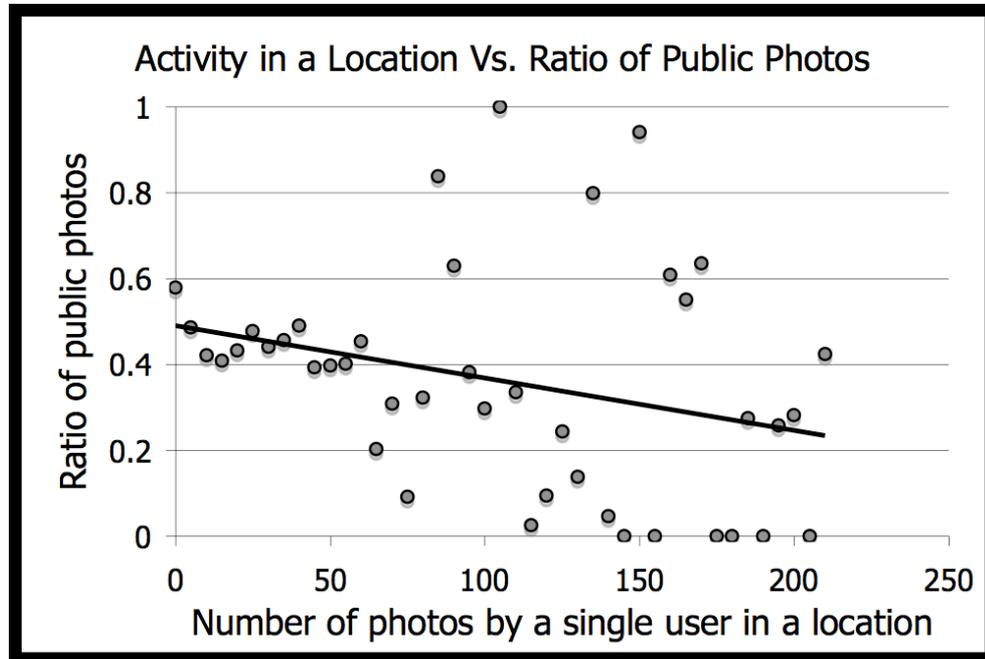
is calculated. If the absolute value is:

- Less than 0.1, the location is *typical*
- Between 0.1 and 0.25, *public* or *private* (depending on the sign)
- Greater than 0.25, *very public* or *very private* (depending on the sign)

# Results

- 81 users in total:
  - 5 had only private photos.
  - 14 had only public photos.
  - 30 took fewer than half of their photos in typical locations.
  - 19 took at least half of their photos in very public or very private locations.
- **Conclusion:** H1 is true for a significant portion of users

# Testing H2

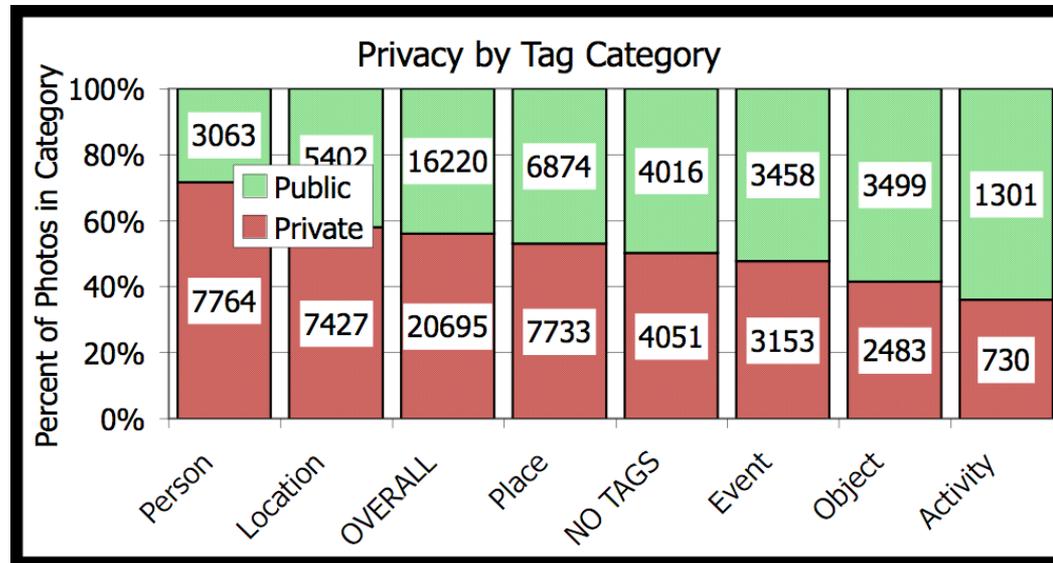


- Negative correlation ( $r(118)=-0.213, p < 0.05$ )
- **Conclusion:** H2 seems to be true.

# Does Content Predict Privacy Decisions?

- Privacy decisions over the content **estimated** from user-assigned tags (without any image examination) are examined.
- Content is abstracted into 6 categories: *Person, location, place, object, event* and *activity*.
- Classification tried on top-fifth most frequently tags (making a total of 1538). Out of this, 1295 tags were classified into one of the six categories.

# Does Content Predict Privacy Decisions? - Results



- Ratio of public photos significantly differs between *person* and all others (except *location*) ( $p < 0.05$ )
- Overall and *person* ratios of public photos are moderately significantly different ( $p < 0.1$ )

# Are Capture Time Privacy Decisions Good?

- Only 7% of the upload time privacy settings are changed afterwards:
  - Public → Non-public (2.4%)
  - Non-public → Public (4.6%)
- Regretted public decisions may originate from the fact that ZoneTag defaults to previous privacy setting:
  - In 95% of the photos changed from public to non-public, previous setting was also public.

# Are Users Concerned About Exposing Their Location?

- Short answer: No
- Long answer:
  - 2% of photos have location info suppressed.
  - Out of 81 users:
    - 18 ever suppressed location tags
    - 3 suppressed on more than 10% of their photos
- **Important Note:** The default behavior is inclusion.

# Quantitative Analysis - Limitations

- Possible self-censorships cannot be captured from the data
- Non-public privacy settings are not distinguished.

# Qualitative Analysis

- Conducted in the format of interviews with selected 15 users:
  - Mostly non-Flickr and ZoneTag users previously (to avoid the self selection bias?)
  - 11 had no technical background.
  - 3 in 20-25, 8 in 30-35, 3 in 35-40 and 1 in 50+ age groups.
  - 8 male, 7 female

# Qualitative Analysis

- 9 parents with young children
- 3 co-workers from the same department/floor (other people in the department and the company were Flickr users).
- 2 males and 1 female, all young, all friends.
- Aim of this deliberate variety is to cover as various scenarios of sharing, privacy and security as possible.
- Conducted interviews after users used ZoneTag for 2 weeks or more.
- Used photo elicitation in interviews.

# Qualitative Analysis - Results

		Theme			
		<i>Security</i>	<i>Identity</i>	<i>Social Disclosure</i>	<i>Convenience</i>
Object	Self	Exposing self to security hazards	Managing own on-line identity	Exposing socially sensitive information to contacts	Difficulty of sharing and viewing
	other	Exposing other to security hazards	Influencing other's on-line identity	Exposing other's socially-sensitive information	Other's ability to share and view

<b>Object</b>	<b>Theme</b>			
	<i>Security</i>	<i>Identity</i>	<i>Social Disclosure</i>	<i>Convenience</i>
<i>self</i>	3 [2]	13 [8]	8 [3]	6 [2]
<i>other</i>	8 [8]	14 [9]	12 [8]	7 [3]

# Qualitative Analysis – Important Concerns

- Privacy of others (compared to self).
- Photos of family (especially children).
- Exposing location of residence.
- Photos' giving a bad impression or exposing user interests being tried to be kept private.
- Disclosure of user activity and whereabouts to known people (sometimes leading to self-censorship).

# Qualitative Analysis – Important Findings

- Privacy of others are taken more seriously than the privacy of self.
- Security of children is an overwhelming concern for parents (only 1 non-parent considers security at all).
- Identity is a very important consideration for virtually everyone.

# Qualitative Analysis – Some Other Concerns

- Users are sometimes uncertain about the content of, audience for and norms regarding particular disclosures.
- Users sometimes find the privacy adjustment too complex and selects the same setting for all photos.
- Unsatisfactory decisions are much more frequent than regretted decisions.

# Qualitative Analysis – Location Privacy

- Users have little or no concern about exposure of aggregate location data.
- Different users are comfortable with different levels of location granularity.
- Participants talked different about location privacy before and after experiments:
  - Discomfort about zip code exposure dropped from 50% to almost none.

# Strengths and Weaknesses

- Strengths:
  - Human factors beneath the privacy decisions for online mobile photo sharing revealed.
  - Both qualitative and quantitative studies are conducted.
  - New suggestions for mobile photo sharing software development are put forth.
- Weaknesses:
  - Abstracting privacy decisions to binary (i.e. public vs. non-public) is too simple. Decisions at friends vs. family vs. others level not apparent.
  - Statistical method used in H1 experiment is kind of *ad hoc*.
  - ?