Privacy and Social Networks

Jessica Staddon
Google
10/11/11
UNC Charlotte Cyber Security Symposium
Overview

- Privacy goals
- Inputs
- Input detail: privacy survey
- Google+ and privacy
Privacy Goals

- Transparency
- Choice

UX

Attacks
Social Challenges

- Transparency
- Choice

UX

Attacks
Social Challenges
Social Challenges
Overview

• Privacy goals
• Inputs
• Input detail: privacy survey
• Google+ and privacy
Inputs (1/5)

• Threat analysis
  – Security attacks $\rightarrow$ privacy compromise
  – Statistical inference
  – Fake accounts
Inputs (2/5)

- Direct User Feedback
  - In-product feedback options
  - Help Forums/Discussion Boards
Inputs (3/5)

• Interviews
  • Field studies
  • Usability studies (cognitive walkthroughs)
Inputs (4/5)

- Online studies
  - Task-oriented evaluations
  - Surveys

<table>
<thead>
<tr>
<th>Question</th>
<th>Options</th>
</tr>
</thead>
</table>
| How often do you refer to the record of a purchase you made several weeks ago? | - Several times a day  
- About once a day  
- A few times a week  
- A few times a month  
- A few times a year  
- Almost never                                                                 |
| How often do you refer to the record of a purchase you made several weeks ago? | - Several times a day  
- About once a day  
- A few times a week  
- A few times a month  
- A few times a year  
- Almost never                                                                 |
| How frequently do you forward or otherwise share (e.g., by printing and giving the printed copy) an online purchase with your close friends or close family members? | - Several times a day  
- About once a day  
- A few times a week  
- A few times a month  
- A few times a year  
- Almost never                                                                 |
Inputs (5/5)

- Indirect inputs
  - Usage statistics (logs analysis)
  - Web queries
Overview

- Privacy goals
- Inputs
- Input detail: privacy surveys
  - (paper: http://research.google.com/pubs/pub37128.html)
- Google+ examples
More detail: privacy surveys

Example privacy survey statistics:

• More than 70% of respondents are concerned about tracking online (Turow, June 2003)
  
  Question: I am nervous about websites having information about me
  
  Question: I sometimes worry that members of my family give information they shouldn’t about our family to web sites

• 93% of respondents are concerned about company and government access to health records (UPI Zogby, 2007)
  
  Question: If you have health records in electronic systems, do you want to decide which companies and government agencies can see and use your sensitive data?

• 60% of respondents have considered quitting Facebook over privacy concerns (Sophos, May 2010)
What’s going on here?

- Education
  - Respondents learn about privacy risks through the survey
- Difficulty evaluating trade-offs
  - What do I get in return for my data?
- Survey bias
This talk

- Demonstrating bias: 3 surveys of content privacy showing impact of wording on responses
- Avoiding bias: Approach to indirectly measuring content privacy
  - Applied to the first survey
Surveys

• 3 studies of the “importance of different online information sources”
  – Study 1: No mention of privacy/data sensitivity
  – Study 2: Privacy mentioned in the instructions only
  – Study 3: Privacy mentioned in the instructions *and* 2 questions
Survey Instructions

– Study 1 (200 respondents)
  • Instructions: “We are studying the importance of different online information sources in daily life. Please answer a few questions about your use of the given information source or sources.”

– Study 2 (200 respondents)
  • Instructions: “We are studying the importance of different online information sources, many of which are privacy-sensitive and common targets of phishers and others who commit online fraud. Please answer a few questions about your use of the given information source or sources.”

– Study 3 (200 respondents)
  • Instructions: “We are studying the importance of different online information sources, many of which are privacy-sensitive and common targets of phishers and others who commit online fraud. Please answer a few questions about your use of the given information source or sources keeping in mind the potential privacy risks of sharing or otherwise revealing sensitive information.”
Example Question

- Study 1
  - How many of your [source] would you be willing to show close friends or family members?

- Study 2
  - Same as survey 1

- Study 3
  - Keeping in mind that [source] may contain sensitive information, how many of your [source] would you be willing to show close friends or family members?
Some results

Instructions: We are studying the importance of different online information sources in daily life. Please answer a few questions about your use of the given information source or sources.

4. How many of your online purchase records would you be willing to show your close friends and close family members?

Study 1, (no privacy warnings)
Some results

Instructions: We are studying the importance of different online information sources, many of which are privacy-sensitive and common targets of phishers and others who commit online fraud. Please answer a few questions about your use of the given information source or sources keeping in mind the potential risks of sharing or otherwise revealing sensitive information.

4. Keeping in mind that purchase records may contain sensitive information, how many of your online purchase records would you be willing to show your close friends and close family members?
Another example

**Instructions:** We are studying the importance of different online information sources, many of which are privacy-sensitive and common targets of phishers and others who commit online fraud. Please answer a few questions about your use of the given information source or sources keeping in mind the potential risks of sharing or otherwise revealing sensitive information.

5. If you were to leave a hard copy of your online newspapers on a restaurant table how likely are you to return to retrieve them?
Results by content type (1/2)

Average Willingness to Share Data: 4. "(Keeping in mind that [source] may contain sensitive information), how many of your [source items] would you be willing to share with your close friends and close family members?"

<table>
<thead>
<tr>
<th>Statistically significant difference?</th>
<th>Study 1, 3</th>
<th>Study 1, 2, 3</th>
<th>Study 1, 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>email</td>
<td>yes</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>photos</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>purchases</td>
<td>yes</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>web history</td>
<td>yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>news</td>
<td>yes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results by content type (2/2)

**Average Interest in Retrieving Data Left in a Public Place:** “(Keeping in mind that [source] may contain sensitive information), if you were to leave a hard copy of your [source items] on a restaurant table how likely are you to return to retrieve it?”

<table>
<thead>
<tr>
<th>Statistically significant difference?</th>
<th>Study 1, 3</th>
<th>Study 2, 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>docs</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>purchases</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>web history</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>news</td>
<td>yes</td>
<td></td>
</tr>
</tbody>
</table>
Linear mixed model restricted to control category, news (Q5 as dependent variable)

Reported retrieval of news (control) by survey version

Mean retrieval score

Study 1 Study 2 Study 3
Survey Version
An indirect approach

Conjecture: Data that is interesting to you and others but you don’t want to share → private
An indirect approach

1. How frequently do you check [content type]?
2. How often do you refer to a [content type] that is several weeks old?
3. Let’s say a server went down and you lost access to your [content type] for 2 weeks. How would this affect you?
4. Imagine you have lost access to all the following information sources... Please rank the information sources in the order the team should work on recovering them.
An indirect approach

5. If you were to leave a hard copy of one of your [content type] on a restaurant table, how likely are you to return to retrieve them?

- Important to you
- Important to some others
- Infrequently shared
An indirect approach

3. How frequently do you forward or otherwise share [content type] with your close friends or close family members?
4. How many of your [content type] would you be willing to show to your close friends and close family members?
How can we test the indirect approach?

We are arguing that there is no ground truth to compare against!
One evaluation approach

Assumptions

• Privacy ratings are sensitive to survey wording for certain data types
• Privacy rankings are less sensitive to survey wording, especially for strongly sensitive data types
Evaluation: Privacy “score”

Compare rankings based on score and direct privacy survey

*Privacy Score = F(your interest, others interest, willingness to share)*
Comparison: Direct Privacy Survey

Instructions: Suppose your computer had a virus that gave it access to ALL of your information on this computer and the Internet. Specifically, it would have access to: email, calendar, photos, documents, contacts, Buzz/Twitter, online purchases, web history

1. How *private* do you consider this information?
   
   Answers: extremely private, very private, moderately private, slightly private, not at all private, I don’t have this info on my computer/online or don’t use this service (0-5 point scale)
Example privacy score

Score:

\[
\text{Score} = \text{Q1: Frequency of viewing} \quad \text{Q2: Frequency of referring} \\
\quad + \text{Ave(importance to you)} \quad \text{Q3: Frequency of sharing} \\
\quad + \text{Ave(importance)} \quad \text{Q6: Disruptiveness of lost access} \\
\quad - \text{Ave (sharing)} \quad \text{Q4: Willingness to share} \\
\quad \text{Q5: Likelihood of retrie+} \\
\]

Q6: Disruptiveness of lost access
Indirect privacy score vs. direct survey rankings
Comparing results

<table>
<thead>
<tr>
<th>Direct Study</th>
<th>1. email</th>
<th>2. docs</th>
<th>3. Web history</th>
<th>4. purchases</th>
<th>5. photos</th>
<th>6. calendar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 3</td>
<td>1. email</td>
<td>2. docs</td>
<td>3. purchases</td>
<td>4. Web history</td>
<td>5. photos</td>
<td>6. calendar</td>
</tr>
<tr>
<td>Study 2</td>
<td>1. email</td>
<td>2. docs</td>
<td>3. purchases</td>
<td>4. photos</td>
<td>5. Web history</td>
<td>6. calendar</td>
</tr>
<tr>
<td>Study 1</td>
<td>1. email</td>
<td>2. docs</td>
<td>3. purchases</td>
<td>4. photos</td>
<td>5. calendar</td>
<td>6. Web history</td>
</tr>
</tbody>
</table>
Overview

• Privacy goals
• Inputs
• Input detail: privacy survey
• Google+ and privacy
Example: Profiles
Example: Profiles

Alma Whitten

Proud to be a Googer today. It’s great that web-based apps (“The Cloud”) not only provide great functionality but also use energy more efficiently than the alternatives. And, in the case of Google, they’re all carbon neutral.

At Google, we’ve worked hard to minimize the environmental impact of our services. In fact, to provide you with Google products for a month, our servers use less energy per user than leaving a light a...
Example: Analytics
Example: Sharing

Stream

Looking forward to this

2011 Fall Cyber Security Symposium - Event Summary | Online Registration by Cvent

Remove description

Friends John Staddon + Add more people

Share

Also email 20 people not yet using Google+
Example: Sharing

[Image of a screenshot from a social media platform, showing privacy settings and a prompt for sharing information.]

- [Text: What happens if you block Alma Whitten]
  - The user is asked to confirm their intention to block the user.
  - The blocked user will be removed from your feed.
  - The user will no longer see your public posts.
  - You can always revive the blocked user.

[Options: Cancel, Report and block this service, Revive blocked user]
Example: Sharing
Thank you!

Feedback, questions?
Back-up slides
Does this approach generalize?

A survey about login questions

E.g. “Login by entering your birthdate”

<table>
<thead>
<tr>
<th>How willing are you to answer this question?</th>
<th>How likely are you to know the answer?</th>
<th>How easy is it to answer this question?</th>
<th>Any Comments?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Willing</td>
<td>Likely</td>
<td>Easy</td>
<td>Privacy</td>
</tr>
<tr>
<td>Not Willing</td>
<td>Likely</td>
<td>Easy</td>
<td>Privacy</td>
</tr>
<tr>
<td>Not Willing</td>
<td>Likely</td>
<td>Easy</td>
<td>Privacy</td>
</tr>
</tbody>
</table>
Questions from Study 1

We are studying the importance of different online information sources in daily life. Please answer a few questions about your use of the given information source or sources.

1. How frequently do you read/view [source]?
2. How often do you refer to [source] from several weeks ago?
3. How frequently do you share [source]?
4. How many of your [source] would you be willing to show close friends or family members?
5. If you were to leave a hard copy of your [source] on a restaurant table how likely are you to return to retrieve them?
6. Let’s say you lost access to your [source] for 2 weeks. How would this affect you? It would be...
7. Imagine you’ve lost access to all of the following information sources: email, online calendar, online photos, online documents, web history, online bank/credit card statements and online newspapers. That is, you can no longer access old emails, online calendar entries, online photos, online documents, web history, online bank/credit card statements or receive/create new instances of any of these. There is a team available to recover these materials for you, and they need to know how to focus their attention. Please rank the information sources in the order in which the team should work on recovering them (with #1 being the source the team focuses on first).
Questions from Study 2

We are studying the importance of different online information sources, many of which are privacy-sensitive and common targets of phishers and others who commit online fraud. Please answer a few questions about your use of the given information source or sources keeping in mind the potential privacy risks of sharing or otherwise revealing sensitive information.

1. How frequently do you read/view [source]?
2. How often do you refer to [source] from several weeks ago?
3. How frequently do you share [source]?
4. How many of your [source] would you be willing to show close friends or family members?
5. If you were to leave a hard copy of your [source] on a restaurant table how likely are you to return to retrieve them?
6. Let’s say you lost access to your [source] for 2 weeks. How would this affect you? It would be...
7. Imagine you’ve lost access to all of the following information sources: email, online calendar, online photos, online documents, web history, online bank/credit card statements and online newspapers. That is, you can no longer access old emails, online calendar entries, online photos, online documents, web history, online bank/credit card statements or receive/create new instances of any of these. There is a team available to recover these materials for you, and they need to know how to focus their attention. Please rank the information sources in the order in which the team should work on recovering them (with #1 being the source the team focuses on first).
Questions from Study 3

We are studying the importance of different online information sources, many of which are privacy-sensitive and common targets of phishers and others who commit online fraud. Please answer a few questions about your use of the given information source or sources keeping in mind the potential privacy risks of sharing or otherwise revealing sensitive information.

1. How frequently do you read/view [source]?
2. How often do you refer to [source] from several weeks ago?
3. How frequently do you share [source]?
4. Keeping in mind that [source] may contain sensitive information, how many of your [source] would you be willing to show close friends or family members?
5. Keeping in mind that [source] may contain sensitive information, if you were to leave a hard copy of your [source] on a restaurant table how likely are you to return to retrieve them?
6. Let’s say you lost access to your [source] for 2 weeks. How would this affect you? It would be...
7. Imagine you’ve lost access to all of the following information sources: email, online calendar, online photos, online documents, web history, online bank/credit card statements and online newspapers. That is, you can no longer access old emails, online calendar entries, online photos, online documents, web history, online bank/credit card statements or receive/create new instances of any of these. There is a team available to recover these materials for you, and they need to know how to focus their attention. Please rank the information sources in the order in which the team should work on recovering them (with #1 being the source the team focuses on first).

A This phrase “Keeping in mind...” did not appear with the “online news” information source.