Information Revelation and Privacy in Online Social Networks

By Ralph Gross and Alessandro Acquisti

George Boulos
gwf5@pitt.edu
Motivation

• Participation in social networking sites has dramatically increased in recent years.

• Study patterns of information revelation in online social networks and their privacy implications.

• Analyze the online behavior of more than 4,000 Carnegie Mellon University students who have joined Facebook.

• highlight potential attacks on various aspects of their privacy.
Online Social Networks

- First application was developed at 1960, yet really evaluated after the advent of the Internet
- A user has a “profile”
- Network of users
- Personally identifiable data
- Category based representations of a person’s broad interests
- Picture of the user
Facebook

• More than 300 million active users
• More than 8 billion minutes are spent on Facebook each day
• More than 2 billion photos, 14 million videos uploaded to the site each month
• More than one million developers

Patterns of Personal Information revelation

• The information available depends on the purpose of the network
• The use of real names
• Identifiable personal photos
• Semi-public information such as current and previous schools
• Private information such as drinking and drug habits and sexual preferences and orientation
• *Information is highly visible*
Social Network and Privacy

• We can tell Information about ourselves only to a small circle of close friends, and not to strangers

• we are willing to reveal personal information to anonymous strangers, but not to those who know us better
Offline Social Network VS. Online Social Network

- Offline ties classification are extremely diverse in terms of how close and intimate a subject perceives a relation to be.
- Online ties are categorized as a binary value (friend or not a friend).
- In offline networks, the number of weak ties one can form and maintain may be able to increase substantially.
Offline Social Network VS. Online Social Network (Cont.)

- Offline social network may include 1000 to 1700 interactions
- Online social network may include hundred of thousands of friends of friends

What does this mean?!

online social networks are both vaster and have more weaker ties, on average, than offline social networks

User have less control in an online social network
Privacy Implications

• Privacy implications associated with online social networking depend on the level of \textit{identifiability} of the information provided, its \textit{possible recipients}, and its \textit{possible uses}

• Face Identification

• Demographic data
Privacy Implications (Cont.)

Who can see the information stored on social sites?
Any one, it is relatively easy to gain access to it. by joining the network, hacking the site, or impersonating a user by stealing his password.

What it can be used for?
Practically anything, from stalking to identity theft.
Facebook Data

• Collecting Facebook profiles of 4500 CMU student
• Assessing how the data provided can be used to reidentify its users
Age Distribution
Provided Information

<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage Facebook Profiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real Name</td>
<td>89%</td>
</tr>
<tr>
<td>Partial Name</td>
<td>3%</td>
</tr>
<tr>
<td>Fake Name</td>
<td>8%</td>
</tr>
</tbody>
</table>
Identifiable photos

<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage Facebook Profiles</th>
<th>Percentage Friendster Profiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifiable</td>
<td>61%</td>
<td>55%</td>
</tr>
<tr>
<td>Semi-Identifiable</td>
<td>19%</td>
<td>15%</td>
</tr>
<tr>
<td>Group Image</td>
<td>8%</td>
<td>6%</td>
</tr>
<tr>
<td>Joke Image</td>
<td>12%</td>
<td>23%</td>
</tr>
</tbody>
</table>

(a) Network of CMU friends

(b) Network of Non-CMU friends
Facebook visibility settings

• Everyone on the Facebook appears in searches of everyone else. This is the “recommended” settings.

• Searchability: 1.2% change default search settings.

• Visibility: 0.03% change the default visibility settings.
Privacy Implications

• “Personal data is generously provided and limiting privacy preferences are sparingly used”

• “Due to the variety and richness of personal information disclosed in Facebook profiles, their visibility, their public linkages to the members’ real identities, and the scope of the network, users may put themselves at risk”
Privacy Threats

• Stalking
• Re-identification
  – Demographic
  – Face
  – Social Security Number
• Building Digital Dossier
Privacy Threats (Cont.)

<table>
<thead>
<tr>
<th>Risk</th>
<th># CMU Facebook Profiles</th>
<th>% CMU Facebook Profiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real-World Stalking</td>
<td>280 (Female) 580 (Male)</td>
<td>15.7 (Female) 21.2 (Male)</td>
</tr>
<tr>
<td>Online Stalking</td>
<td>3528</td>
<td>77.7</td>
</tr>
<tr>
<td>Demographics Re-Identification</td>
<td>1676</td>
<td>44.3</td>
</tr>
<tr>
<td>Face Re-Identification</td>
<td>2515 (estimated)</td>
<td>55.4</td>
</tr>
</tbody>
</table>
How To protect?

• Fake email address
• Accept only known friends
• Rethinking Facebook search options

*Putting ourselves in Facebook’s shoe; how to eliminate privacy threats?*
Strengths & Weaknesses

• Strengths:
  – The paper shows how real privacy threats are available in online social networking
  – Experimental data is provided on 4500 CMU student

• Weaknesses:
  – The paper does not propose any solution
Q&A

Thank you