Synergy of social networks
defeats online privacy

Eleonora Petridou
Marek Kuczyński

System And Network Engineering
University of Amsterdam

February 2, 2011
Research question

“What are the privacy risks associated with social network user profiling?”

Sub questions

1. What is online privacy?
2. Current exposure and privacy policies
3. Business models and the future
4. Attack vectors and combining acquired information
5. Countermeasures
Overview

Concent discussed during this presentation

- Evolution Of Social Networks
- Characteristics Of Social Networks
- Business Models
- Proof Of Concept
- Risk Scenarios
- Countermeasures
- Conclusion

Please ask your questions at the end of the presentation!
Evolution Of Social Networks

Based on Web 2.0, founded between 2003 - 2008

<table>
<thead>
<tr>
<th>Incentives to use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication with friends all over the world</td>
</tr>
<tr>
<td>Quick and indirect overview of other people’s lifes</td>
</tr>
<tr>
<td>Social pressure</td>
</tr>
</tbody>
</table>
Monetizing Online Social Networks

Interested parties

- Third-parties
- Individual users
- Social networks themselves
Social Contact Oriented
Social Contact Oriented

Features

- 500+ million users
- Revenue $800 million (2010)

Privacy observations

- Personalised advertising - cookies
- Instant personalisation of other websites
- Anonymizing the data → Unknown algorithm → Who receives the data?
Synergy Of Social Networks Defeats Online Privacy, www.socialsynergy.nl

Social Contact Oriented

Business model

- Facebook Pages
  → Users connect with a brand or name
  → Shop online

- Social Ads
  → Targeted advertising
  → Attribute filtering
  → “Like” by friends

- Facebook credits and credit cards
Features

- 90 million users
- Profit $10.1 million in Q1 2010

Privacy observations

- Not selling personally identifiable information to third parties
  → How is data anonymized?
- Job history cannot be private
Business model

- Direct Ads
  - Specific audience
  - Search based on job, education, location etc.

- Premium accounts
  - For job seekers
  - For recruiters
Communication Oriented
Communication Oriented

Features

- 175 million users
- Applications extending Twitter’s functionality → Twitpic

Privacy observations

- Everything is public by default
- No disclosure to third parties → Exception only the legal cases
Communication Oriented

Business model

- Promoted Tweets
  → Top of results
Synergy Of Social Networks Defeats Online Privacy, www.socialsynergy.nl
Sports Oriented

Features

- Software package Endotracker
- Personal profile
  - Statistics
  - Image of the route in Google Maps

Privacy observations

- Public profile by default
- Data exposed at user’s own risk
- Data anonymization → Distribution
Sports Oriented

Business model

- Online shop with training equipment
- Premium accounts for businesses
- Pro version
Synergy Of Social Networks Defeats Online Privacy, www.socialsynergy.nl
Features

- GPS enabled devices required
- Checkins at locations
- Discounts
- Tips and suggestions

Privacy observations

- Personalised advertisements and emails
- Data not sold to third parties
Location Oriented

Business model

- Still in the process
- Some of the venues are already paying
Privacy Incidents And Enumeration Strategies

Recent privacy related incidents

- Lawsuits over Facebook/Myspace messages
- Middle-East countries blocking Facebook
- Subpoenas for Twitter information

Enumerate information based on...

- ... one individual → e.g. Job interviews, credit check?
- ... an interest → e.g. Soda producer, political parties?
- ... everything, see what comes out → e.g. Google, FBI?
Connecting The Dots...
Created Data Sources

Data gathered in POC

- Facebook → Tree of targetted user
- LinkedIn → Tree of targetted connection
- Twitter/Twitpic → Celebrity Twitpic GPS coordinates
- Foursquare → Who was at a venue at certain time?
- Visualizing it all

Easy accessible sources

- Twitter → Marketing relevant data
- Endomondo → Plot of user workouts
Plotting Twitpic EXIF Data
Crawling Facebook, LinkedIn
Visualizing Data With Vizster - Overview
<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Last Name</td>
<td></td>
</tr>
<tr>
<td>User ID</td>
<td></td>
</tr>
<tr>
<td>Nickname</td>
<td></td>
</tr>
<tr>
<td>Email</td>
<td>e8dc0f229455b01cea4aaef3b49e0a0e</td>
</tr>
<tr>
<td>Profile Pic</td>
<td></td>
</tr>
<tr>
<td>Mobile phone</td>
<td></td>
</tr>
<tr>
<td>Address</td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td>95a46ba40e643b4f636efb992ce9c6ae</td>
</tr>
<tr>
<td>Workplace</td>
<td></td>
</tr>
<tr>
<td>Groups</td>
<td></td>
</tr>
<tr>
<td>Last tweet</td>
<td>c3656271b6328fb26d9b10dbb894857f</td>
</tr>
<tr>
<td>GPS last tweetpic</td>
<td></td>
</tr>
<tr>
<td>Timestamp last tweetpic</td>
<td></td>
</tr>
<tr>
<td>Date of last tweetpic</td>
<td></td>
</tr>
<tr>
<td>GPS last Foursquare checkin</td>
<td></td>
</tr>
<tr>
<td>Timestamp last Foursquare checkin</td>
<td></td>
</tr>
<tr>
<td>Date of birth</td>
<td>83589779bbfe9fd85a5835b586e8c10b</td>
</tr>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>Sport</td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Status</td>
<td></td>
</tr>
</tbody>
</table>
Answering The Research Question

Research question

“What are the privacy risks associated with social network user profiling?”

Risks of social profiling

- Theft, stalking (i.e. Twitpic + Endomondo)
- Identity theft (i.e. LinkedIn + Facebook)
- Insurance issues (i.e. Foursquare + Facebook)
- Police, corrupt governments (i.e. Facebook + Twitpic)
- Corporate spying, headhunters (i.e. Facebook + LinkedIn)
Countermeasures

Make the users aware…

- ... of what options are enabled by default
- ... who is interested in their data and why
- ... of data retention on social networks
- ... of GPS logging in (mobile) photo cameras

Restricting Access To Data

- Limit the exposure relations and bound accounts
- Make privacy settings part of the sign-up (opt-in!)
- Block accounts with excessive data volumes
Conclusion

What’s next?

- Maturity and exposure of social networks is growing
- No real leads on data selling, but advertising business is huge
- Risks are increasing as features and dependance on social networks grows
- Legislation will not save or protect the user, this issue is global
- User awareness seems to be the only viable option

How will we help?

- Newbie guide to "secure" social networking → http://www.socialsynergy.nl
Thanks for your attention

Questions?

http://www.socialsynergy.nl

eleonora.petridou@os3.nl&marek.kuczynski@os3.nl
<table>
<thead>
<tr>
<th>#</th>
<th>How to match the data items?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The Facebook &quot;Wall&quot; may contain an application that links directly to someone’s Endomondo profile.</td>
</tr>
<tr>
<td>2</td>
<td>The Facebook &quot;Wall&quot; may contain an application that links directly to someone’s Foursquare profile.</td>
</tr>
<tr>
<td>3</td>
<td>The name, nickname and location retrieved from Foursquare may lead to a positive match with Facebook. The account linking option of Foursquare offers a direct link to someone’s Facebook account.</td>
</tr>
<tr>
<td>4</td>
<td>The name, nickname and location retrieved from Foursquare may lead to a positive match with Twitter. The account linking option of Foursquare offers a direct link to someone’s Twitter account.</td>
</tr>
<tr>
<td>5</td>
<td>Postings on the &quot;wall&quot; of someone may be put there through Twitter. In this case, the Twitter account name of someone can be retrieved. Additionally, a match can be made based on full name and city.</td>
</tr>
<tr>
<td>6</td>
<td>A match between Twitter and Facebook can be made if the biography and name of a Twitter user match to an individual on Facebook.</td>
</tr>
<tr>
<td>7</td>
<td>The Twitpic ID of someone is identical to that of Twitter, so this option can always be checked reliably.</td>
</tr>
<tr>
<td>8</td>
<td>LinkedIn provides the name and geographical area of a person by default, but a Twitter account can be bound to someone’s LinkedIn profile as well.</td>
</tr>
<tr>
<td>9</td>
<td>A match between Twitter and LinkedIn can be made if the biography and name of a Twitter user match to an individual on LinkedIn.</td>
</tr>
<tr>
<td>10</td>
<td>The e-mail of a user is exposed by default on LinkedIn, it can be used on Facebook to retrieve a profile if it was registred with the same e-mail address.</td>
</tr>
<tr>
<td>11</td>
<td>The bi-directional connection between Facebook and LinkedIn can be created if the full name, date of birth and/or city match.</td>
</tr>
<tr>
<td></td>
<td>Facebook</td>
</tr>
<tr>
<td>------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>First name</td>
<td>Yes*</td>
</tr>
<tr>
<td>Last name</td>
<td>Yes*</td>
</tr>
<tr>
<td>Nick name</td>
<td>Yes(URL)</td>
</tr>
<tr>
<td>Profile pic</td>
<td>Yes</td>
</tr>
<tr>
<td>E-Mail</td>
<td>No*</td>
</tr>
<tr>
<td>Age</td>
<td>No</td>
</tr>
<tr>
<td>Date of birth</td>
<td>No*</td>
</tr>
<tr>
<td>Height</td>
<td>No</td>
</tr>
<tr>
<td>Gender</td>
<td>Yes*</td>
</tr>
<tr>
<td>Relationship status</td>
<td>Yes</td>
</tr>
<tr>
<td>City</td>
<td>Yes</td>
</tr>
<tr>
<td>Country</td>
<td>Yes</td>
</tr>
<tr>
<td>Address</td>
<td>No</td>
</tr>
<tr>
<td>Mobile Phone</td>
<td>No</td>
</tr>
<tr>
<td>Workplace</td>
<td>Yes</td>
</tr>
<tr>
<td>Groups</td>
<td>Yes</td>
</tr>
<tr>
<td>Connections</td>
<td>All</td>
</tr>
<tr>
<td>Family Relations</td>
<td>Yes</td>
</tr>
<tr>
<td>Tweets</td>
<td>No</td>
</tr>
<tr>
<td>GPS locations</td>
<td>No</td>
</tr>
<tr>
<td>Height</td>
<td>No</td>
</tr>
<tr>
<td>Sport</td>
<td>No</td>
</tr>
<tr>
<td>User content</td>
<td>Yes (wall)</td>
</tr>
<tr>
<td>Apps bound to other social networks</td>
<td>No</td>
</tr>
<tr>
<td>Search engine</td>
<td>Yes</td>
</tr>
</tbody>
</table>
START

Start.sh
Type: BASH control file
Input: target ID
Output: Result XML file

fb_lib.py / li_lib.py
Type: Python lib
Input: $option, $ID, $friends
Output: HTML page

fb_scrapy.py / li_scrapy.py
Type: Python lib
Input: HTML from libs
Output: XML tags

$targetID.xml
Type: XML content file
Input: supplied by start.sh
Output: the XML itself

END