Social Computing Systems

Walter S. Lasecki

EECS 498/598, Winter 2018
(http://tiny.cc/socsClass)
Today

- Privacy and security risks/threats
- Privacy-related design considerations in social computing
- Privacy in human computation and crowdsourcing
Why do we care about privacy?
“Privacy is the ability of an individual or group to seclude themselves, or information about themselves, and thereby express themselves selectively.”

— Wikipedia
Definitionz

“Privacy is the ability of an individual or group to seclude themselves, or information about themselves, and thereby express themselves selectively.”

— Wikipedia

… that doesn’t sound very social. So why do we care?
Why do we want things to be private?

- **Unintentional social disclosure**
  - Some things are just embarrassing (private vs public disclosure)

- **Direct threat to our (or our ‘liked-ones’) safety or assets**
  - The ability to steal or harm people via information or access

- **Overall:** Because there are risks to having all information public in an imperfect world.
What do we want to be private?

- Embarrassing or personal content (e.g., photos / audio)
  - Things we just don’t WANT people to see

- Account or status information (e.g., location logs / shopping history)
  - Information that people could use to track or attack us

- Access information (e.g., work servers / PCs / smart-home controls)
  - Controls people could use to gain over parts of our lives

- ...?
Privacy vs. Security

- **Privacy**: your data is not being seen / shared with others
- **Security**: your data is not being manipulated / stolen
Sources of Privacy Leaks
Surveillance / Monitoring
[Web] Surveillance / Monitoring
Government Agencies

U.S. Spy Rocket Has Octopus-Themed 'Nothing Is Beyond Our Reach' Logo. Seriously.

The Office of National Intelligence was very excited about sending a rocket into space Thursday with a bunch of new satellites and live tweeted its launch. This would usually be a cute display of social media, along the lines of NASA getting the world excited about its Mars Curiosity Rover, except these are spy satellites that will likely be used to gather communications flotsam and who knows what else from people around the world. Dragnet surveillance is a touchy...
“Non-government” Organizations

● Groups with a cause (e.g., Anonymous)
  ○ Stealing info/access, denial of service attacks, etc.
  ○ Further a mission or belief. Mostly. (You can debate the details of their vigilantism)

● Groups / collectives with... fewer clear causes (e.g., 4chan)
  ○ Stealing info/access, denial of service attacks, etc.
  ○ Now, for bullying, retaliation, or ‘teh LOLs’ (← clearly a technical term)

● Individuals
  ○ Lone wolf hackers, identity thieves, con artists, etc.
  ○ Motivation and methods vary widely
  ○ “Surveillance” → “stalking”
Types of Data Risk

(some examples. not a complete list.)
TYPE: Personal Content

● Examples:
  ○ Erin Andrews case: stalker recorded nude video and posted online
    ■ Particularly high profile case + $75million civil suit
  ○ Countless examples of cyberbullying in online settings (often w/ K-12 children)
    ■ Often involve sharing embarrassing content or information (real or ‘plausible’ fake)
    ■ Many have lead to suicides (or attempts) due to emotional toll
  ○ Publicly visible grades
  ○ Private/semi-private comments made public
    ■ E.g., Mitt Romney’s 47% comment (those who feel ‘entitled’)
    ■ E.g., that kinda-mean thing you said about one friend to another via private message

● Overall: most personal content leaks affect social capital / relationship status
TYPE: Personal Information

- Examples:
  - Account info (e.g., bank, etc.)
  - Usage / access permissions (e.g., servers, etc.)
  - Social media access (e.g., ‘hacked’ FB or Twitter accounts)
  - ... many more

- Overall: Information that can be used to do direct harm / fraud
TYPE: Personal Status

For example…
STATUS: Location Risks

The location of our Bat Cave is meant to be secret, so STOP CHECKING IN!
Finding You Using Your Friends

Sadilek et al.
STATUS: Records

- Medical records
- Purchase records
- Interaction records
- Court records
- Search logs
- ...

Ex. attack:

CC statements saying you paid for a flight in Dec., and bought an inflatable reindeer
What else?

[ There are many examples. Let’s name some. ]
Privacy / Threats in Crowdsourcing
Types of Threats: Extraction
Types of Threats: Extraction

Exploitation
Types of Threats: Extraction

Exposure
Types of Threats: Extraction

Reconstruction
Types of Threats: Extraction

Reconstruction
Types of Threats: Manipulation

Individuals

worker aggregation

"length"

"length"

"sun"

"length"

"lenht"
Types of Threats: Manipulation

Uncoordinated Groups

- "weight"
- "sun"
- "potato"
- "length"

worker aggregation
Types of Threats: Manipulation

Coordinated Groups

worker aggregation

“length”

“sun”

“sun”

“sun”
Types of Threats: Manipulation of Iterative Workflows
Sources of Threats: Individuals

Extract unfiltered info
View sensitive info
Steal partial info

Add noise to responses
Disrupt iterative workflows
Sources of Threats: Groups

Large-scale individual attacks
Reconstruct info

Disrupt systems’ responses
Produce an undesired answer
Raising an Army

Could malicious requesters hire workers for group attacks?
Raising an Army

Could malicious requesters hire workers for group attacks?

*How many workers will be willing to help?*
Experiment: Credit Card Extraction

**Attack**

Instructions

1. Read all directions in this list FIRST!
2. Go to the link below
3. Give the answer 'sun' for the image.
4. When complete, please enter 'Done' in the box below and submit the HIT.
5. Submit this HIT once done.

**Target**

Type what you see written in the image

Enter your answer here:

Submit HIT
## Experiment: Credit Card Extraction

<table>
<thead>
<tr>
<th>Condition</th>
<th>Return rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>73.8%</td>
</tr>
</tbody>
</table>

## Experiment: Credit Card Extraction

<table>
<thead>
<tr>
<th>Condition</th>
<th>Return rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>73.8%</td>
</tr>
<tr>
<td>Innocent</td>
<td>62.1%</td>
</tr>
</tbody>
</table>

Experiment: Credit Card Extraction

<table>
<thead>
<tr>
<th>Condition</th>
<th>Return rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>73.8%</td>
</tr>
<tr>
<td>Innocent</td>
<td>62.1%</td>
</tr>
<tr>
<td>Malicious</td>
<td>32.8%</td>
</tr>
</tbody>
</table>

Experiment: OCR Manipulation

Instructions

1. Read all directions in this list FIRST!
2. Go to the link below
3. Give the answer 'sun' for the image.
4. When complete, please enter 'Done' in the box below and submit the HIT.
5. Submit this HIT once done.

## Results: OCR Manipulation

<table>
<thead>
<tr>
<th>Condition</th>
<th>Return rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>73.8%</td>
</tr>
</tbody>
</table>

## Results: OCR Manipulation

<table>
<thead>
<tr>
<th>Condition</th>
<th>Return rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>73.8%</td>
</tr>
<tr>
<td>Innocent</td>
<td>75.0%</td>
</tr>
</tbody>
</table>

Type “sun”

## Results: OCR Manipulation

<table>
<thead>
<tr>
<th>Condition</th>
<th>Return rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>73.8%</td>
</tr>
<tr>
<td>Innocent</td>
<td>75.0%</td>
</tr>
<tr>
<td>Malicious</td>
<td>27.9%</td>
</tr>
</tbody>
</table>

## Results: OCR Manipulation

<table>
<thead>
<tr>
<th>Response</th>
<th>Baseline</th>
<th>Innocent</th>
<th>Malicious</th>
</tr>
</thead>
<tbody>
<tr>
<td>“sun”</td>
<td>12%</td>
<td>75%</td>
<td>28%</td>
</tr>
<tr>
<td>“gun”</td>
<td>36%</td>
<td>16%</td>
<td>-</td>
</tr>
<tr>
<td>“fun”</td>
<td>26%</td>
<td>7%</td>
<td>-</td>
</tr>
<tr>
<td>“lun”</td>
<td>14%</td>
<td>1%</td>
<td>-</td>
</tr>
<tr>
<td>“jun”</td>
<td>9%</td>
<td>1%</td>
<td>-</td>
</tr>
<tr>
<td>other</td>
<td>3%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>“length”</td>
<td>-</td>
<td>-</td>
<td>72%</td>
</tr>
</tbody>
</table>
Convincing Workers to Attack

Some workers are already willing (~30%)

How do we convince the rest?

Money? Story? What else?
The Effects of Price

Idea: To increase participation, let’s pay more

In theory…
   All workers are better paid by helping us

In practice…
Experiment: Credit Card Extraction

**Attack**

**Instructions**

1. Read all directions in this list FIRST!
2. Goto the link below
3. Give the answer "sun" for the image.
4. When complete, please enter "Done" in the box below and submit the HIT.
5. Submit this HIT once done.

**Click to see the task in new tab**

Enter your answer here:

Submit HIT

**Target**

Type what you see written in the image

4551 7080 928 0304

SUBMIT
The Effect of Price

Response Rate (%)

Payment Amount ($)
The Effect of Price

Response Rate (%)

Payment Amount ($)
The Effect of Price

Response Rate (%)

Payment Amount ($)

- Attack-50
- Target-50
The Effect of Price

Response Rate (%) vs. Attack Task Payment Amount ($)

- Red line: Target-5
- Blue line: Target-50
Why do Some Workers Still Help?

Are workers just jerks?
Why do Some Workers Still Help?

Are workers just jerks? **Not really.**

Many even warned others, but…
Why do Some Workers Still Help?

Are workers just jerks? **Not really.**

Many even warned others, but…

“*You know what, I had the same issue today [... I] didn’t pay it much mind when it happened.* “
Can We Trust the Crowd?

Some worker won’t complete suspicious work

A few workers will complete tasks regardless

Very few will report issues

Might be susceptible to:
- monetary incentives
- false reasons
Pre-Filtering
In-class Attack Activity

< attack me (hypothetically, not for real!) >
Today

- Privacy and security risks/threats
- Privacy-related design considerations in social computing
- Privacy in human computation and crowdsourcing
Next Class

- Lecture Topic: NONE!

... but there is a BIG QUIZ coming in a week!