User-Controllable Security & Privacy: Are the Expectations Realistic?

Norman Sadeh
Director, Mobile Commerce Lab.
School of Computer Science & CyLab
Carnegie Mellon University

Collaborators:
Faculty: Jason Hong, Lorrie Cranor, Lujo Bauer, Tuomas Sandholm
Post-Docs: Paul Hankes Drielsma, Eran Toch, Jinghai Rao
PhD Students: Patrick Kelley, Jialiu Lin, Janice Tsai, Michael Benisch and Ram Ravichandran
Staff: Jay Springfield, David Eggerschwiler, and Linda Francona

Outline
- User-Controllable Security & Privacy: The Expectations
- Location Sharing Applications: A Representative Domain
- What Are Users Really Capable of?
- How Can We Help Users?
  - Auditing Functionality
  - User-Controllable Policy Learning
  - Expressiveness
  - Default Policies
- Concluding Remarks

User-Controllable Security & Privacy
- Users are increasingly expected to set up security and privacy policies,
  - Home computer
  - Flatter, more agile organizations
  - Social networks
- Is this realistic?
  - Potential vulnerabilities

Privacy Policies on Social Networks

Copyright © Norman Sadeh, 2005-2009
Mobile Social Networking Apps As a Case Study

- **Desire to share data** with others
- Mitigated by **privacy concerns**
- **Location sharing** as a “hot” application
  - Tens of apps over the past several years
  - ...but adoption seems rather limited

Some Questions

- Can users be expected to effectively **specify their policies**?
  - Do people understand their own policies?
  - Can they articulate their policies?
  - **Tradeoffs** between user burden and accuracy
  - Do policies evolve?

- Can we develop technologies that **empower users** to more accurately & efficiently specify their policies?

High-Level Architecture

- **Location Sharing Server**
  - Combines GPS, GSM and WiFi
  - Available on **cell phones** and **laptops**
  - **PEA** = Policy Enforcing Agent

Time Line

- 2003-2005: **Early prototypes & Lab studies**
- 2006-2007: **“People Finder”** application
  - Laptops and some cell phones
  - Multiple pilots up – a couple of hundred users in total
- 2008: first Facebook application: **“Locyoution”**
  - Laptops
  - Piloted by a little over 100 users
- 2009: New Facebook application: **“Locaccino”**
  - Launched in mid February 2009: **www.locaccino.org**
  - Laptops and some cell phones
  - Could scale to 100,000s of users – if successful
Location Sharing Policies

Users Are Not Good At Defining Policies

Early Lab Study:
- 19 users
- 30 queries per user

<table>
<thead>
<tr>
<th></th>
<th>Mean (sec)</th>
<th>Standard Deviation (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule Creation</td>
<td>321.53</td>
<td>206.10</td>
</tr>
<tr>
<td>Rule Maintenance</td>
<td>101.15</td>
<td>110.02</td>
</tr>
<tr>
<td>Total</td>
<td>422.69</td>
<td>213.48</td>
</tr>
</tbody>
</table>

What is Going On?

- Is it because we have a **bad interface**?
- Do people who define **more rules** do better?
- Do people who spend **more time** defining & refining rules do better?

It’s Not Because of the Interface

Modifying rules was easy using the system's rule interface
Only Slight Correlation with # Rules

- Total of 30 requests
- Post-hoc accuracy

Could Auditing Help?

- Users do not always know their own policies
- Users do not fully understand how their rules will operate in practice
- Auditing (‘feedback’) functionality may help users better understand the behaviors their policies give rise to
Evaluating the Usefulness of Feedback: Before/After Surveys – Facebook Study

56 Facebook users divided into 2 groups: one w. (“F”) and one w/o (“NF”) access to a history of requests for their location.

Evaluating the Usefulness of Feedback: Do People Want it?

- 76.9% of people who had “feedback” indicated they wanted to keep it
- 83.3% of those who didn’t have said they would like to have it

Examining Users’ Privacy Rules at the end of the study

Policy Evolution – with feedback

Data for 12 most active users across 3 pilots of PeopleFinder Application.
How Expressive Should Policies Be?

Expressiveness and Efficiency

- Security & privacy **mechanism**: $f(\theta, a)$ decides on an outcome based on a user’s stated preferences $\theta$ and the context $a$ of a request
- **Rational user assumption**: users define policies that take full advantage of available expressiveness $h^*(t) = \arg \max_{\bar{a}} \int \bar{a} P(\bar{a}) u(t, \bar{a}, f(\theta, \bar{a}))$
- **Efficiency**: How well do we capture the ground truth preferences of a user population given an expected distribution of requests $E[\mathcal{E}(f)] = \int_t P(t) \int_{\bar{a}} P(\bar{a}) u(t, \bar{a}, f(h^*(t), \bar{a}))$

**Expected efficiency of best policies**

- Data from 30 users over 1 week – cell phones – GPS & WiFi
- Assumes that an erroneous disclosure is 5x worse than an erroneous non-disclosure & fully “rational” user
Capturing Location-Sensitive Preferences

Observations
- Can be applied to the design of any security or privacy mechanism
- ...but real users are not fully rational
  - User burden
    - Cognitive
    - Time

Could Machine Learning Help?

Early Experiment with Case-Based Reasoning
More Recent Pilots – 12 most active target users

3 Pilots – total of over 60 participants

User-Defined Rules: 79% vs. ML: 91%

Note: Includes benefits of auditing

User-Controllable Policy Learning (patent pending)

- Learning traditionally configured as a “black box” technology
- Users are unlikely to understand the policies they end up with
  - Major source of vulnerability
- Can we develop technology that incrementally suggests policy changes to users?
  - Tradeoff between rapid convergence and maintaining policies that users can relate to

User-Controlled Policy Learning (patent pending)

Legend:
- Access granted
- Suggested Rule Change
- Audited Request
- Audit says Deny Access
- Audit says Grant Access
Exploring Neighboring Policies: Users Are More Likely to Understand Incremental Changes

Rate neighboring policies based on:
- **Accuracy**
- **Complexity**
- **Distance from current policy**

Expressiveness & User Bruden

Average number of rules a user would have to define to achieve optimal efficiency

<table>
<thead>
<tr>
<th></th>
<th>Friends</th>
<th>Family</th>
<th>University community</th>
<th>Anyone</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black list</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>1</td>
</tr>
<tr>
<td>Time</td>
<td>1.97</td>
<td>2.03</td>
<td>1.50</td>
<td>0.70</td>
<td>6.20</td>
</tr>
<tr>
<td>Location</td>
<td>6.90</td>
<td>6.23</td>
<td>3.30</td>
<td>1.37</td>
<td>17.80</td>
</tr>
<tr>
<td>Time/Location</td>
<td>7.97</td>
<td>7.97</td>
<td>5.23</td>
<td>2.73</td>
<td>23.90</td>
</tr>
</tbody>
</table>
Identifying Default Policies (Ongoing Work)

- Location sharing with members of the campus community – 30 different users

Green: Share  
Red: Don’t

Clustering Canonical Policies

- Canonical locations, days of the week and times of the day: Morning, home, work, weekday

Social Networking View of Location Sharing
Adding More Functionality

Overall Vision

- **New Technology**
- **Policy Creation**
  - Jane: My colleagues can see my location on weekdays between 8am and 5pm
- **Policy Visualization**
  - Jane and Eric are late for our meeting. Show me where they are!
  - Bob: Why couldn't Bob see where I was?
- **Policy Enforcement**
  - Jane is in Oakland but I can't access Eric's location
  - Bob: Bob is a colleague. So far only your friends can see where you are
- **Policy Auditing & Refinement**
  - Eric: What if my colleagues could see my location too?
    - In the past you denied access to your colleague Steve
    - OK, make it just my superiors

Are the Expectations Realistic?

- Users are not very good at specifying policies
  - **Vulnerability**
- Tradeoffs between expressiveness and user burden
  - **Quantifying the benefits of additional expressiveness**
    - can help
- Auditing functionality
  - Understanding the set of behaviors entailed by a given policy
  - Asking questions
    - Why/Why not? What if?
  - User-Controllable Learning
    - Moving away from machine learning as a black box
    - In security & privacy, users have to remain in control
Location Sharing: Lessons Learned

- Users have complex privacy preferences
  - Simple “black list” approaches only capture a small fraction of scenarios
  - Application becomes less useful: users err on the safe side -> little sharing
- Time and location are important attributes
  - Other attributes still to be quantified
- Auditing functionality increases user comfort and contributes to more, albeit selective sharing
- User-controllable learning seems to make a difference
- Default policies are not easy to find but can help

Selected References


Selected Press Coverage

- BusinessWeek blog, March 2009
- Numerati blog, March 2009
- The Piper, March 2009
A Video

http://www.screentoaster.com/watch/stUkxdQERIR1dSVleWIJZUIR_U/specific_find_demo