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 Eggerschwiller, and Linda Francona

Outline

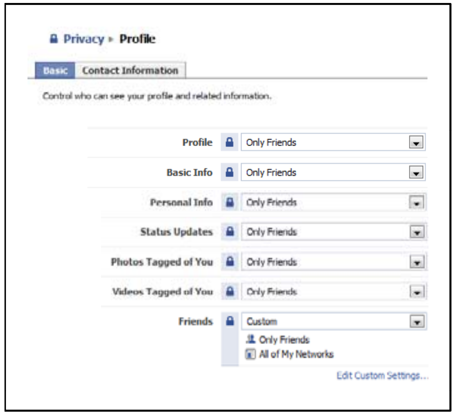
- User-Controllable Security & Privacy: The Expectations
- Location Sharing Applications: A Representative Domain
- What Are Users Really Capable of?
- How Can We Help Users?
  - Auditing Functionality
  - User-Controllable Policy Learning
  - Expressiveness
  - Default Policies
- Concluding Remarks

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

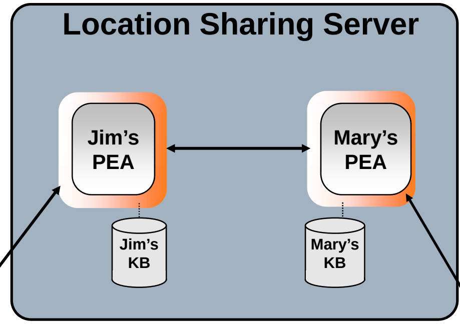
facebook

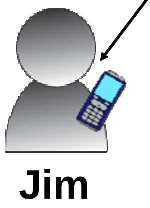


Mobile Social Networking Apps As a Case Study

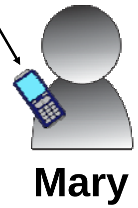
- **Desire to share data** with others
- Mitigated by **privacy concerns**
- **Location sharing** as a "hot" application
  - Tens of apps over the past several years
  - …but adoption seems rather limited

High-Level Architecture





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



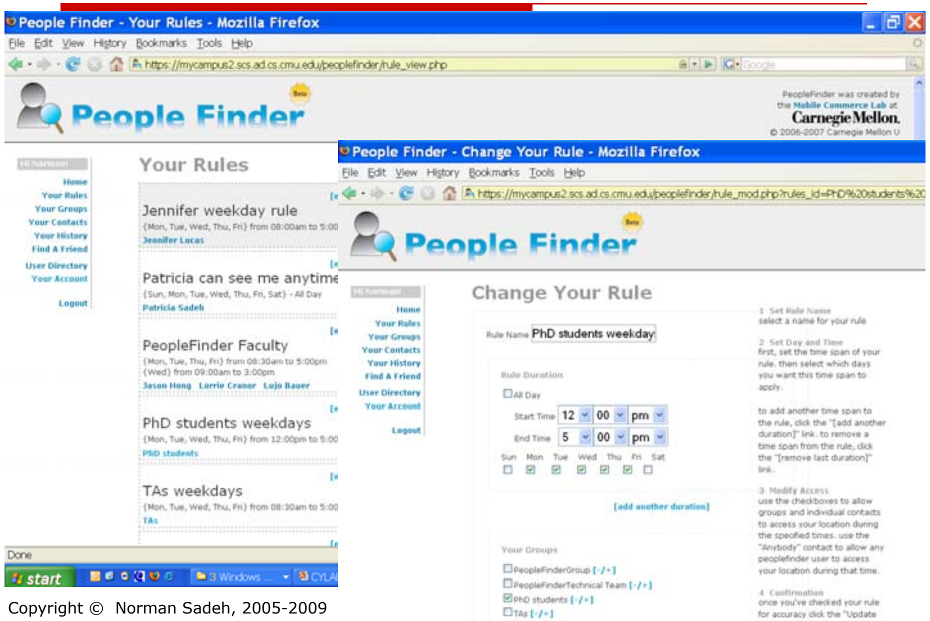
Some Questions

- 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?

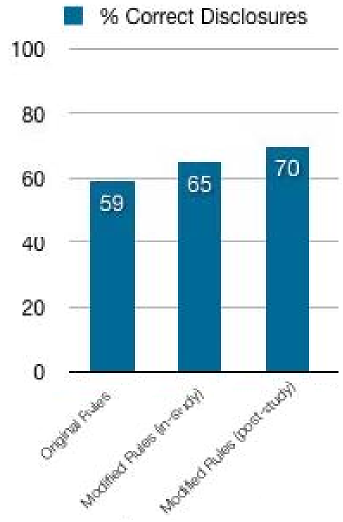
Time Line

- 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

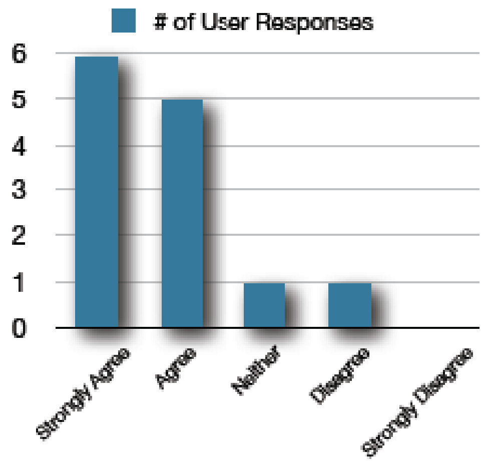
	Mean (sec)	Standard Deviation (sec)
Rule Creation	321.53	206.10
Rule Maintenance	101.15	110.02
Total	422.69	213.48

What is Going On?

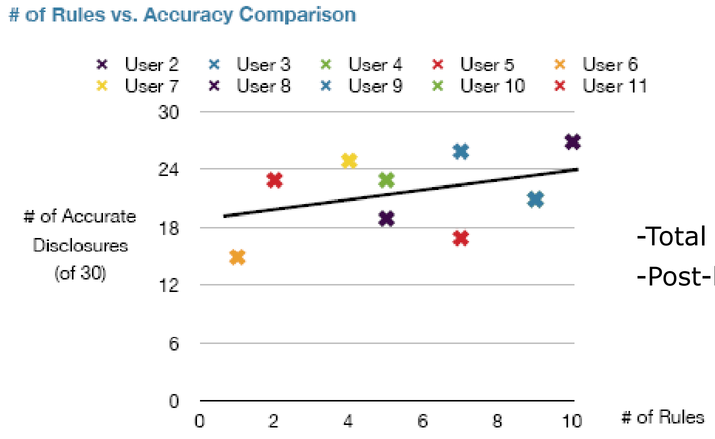
- Is it because we have a **bad interface**?
- Do people who define **more rules** do better?
- Do people who spend **more time** defining & refining rules do better?

It's Not Because of the Interface

Modifying rules was easy using the system's rule interface



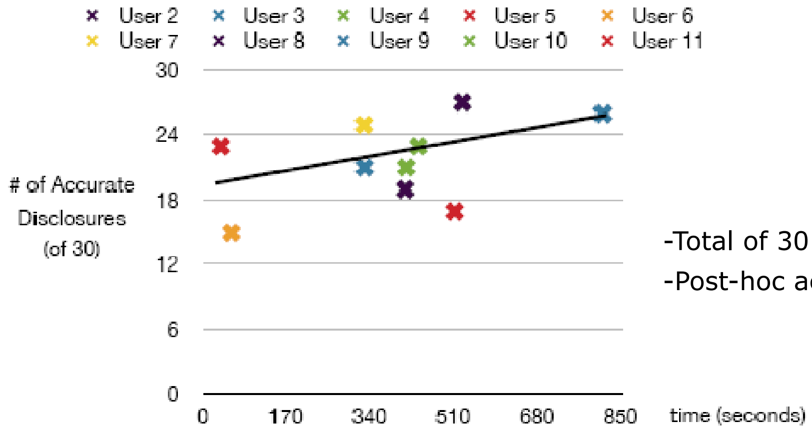
Only Slight Correlation with # Rules



-Total of 30 requests
-Post-hoc accuracy

Regression Line	R-squared
y = 0.4789x + 18.876	R2 = 0.1397

Only Slight Correlation with Time Spent



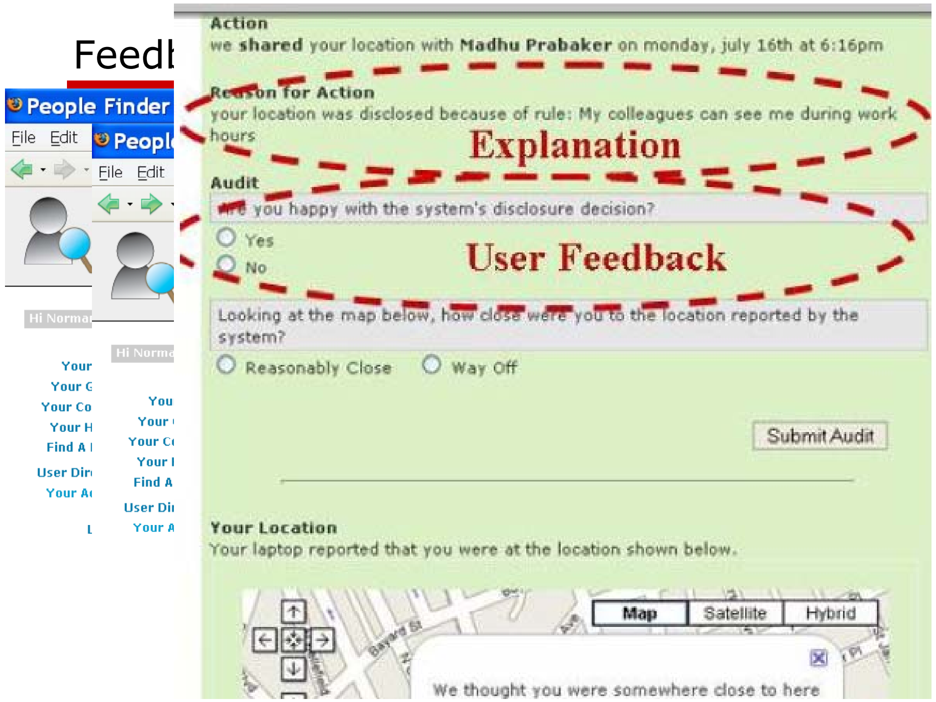
-Total of 30 requests
-Post-hoc accuracy

Regression Line	R-squared
y = 0.0078x + 18.698	R2 = 0.2031

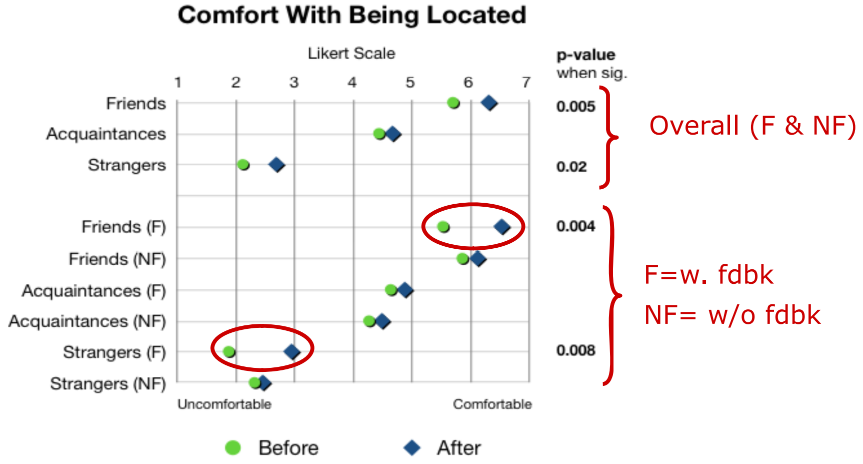
Could Auditing Help?

- 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

Feedb



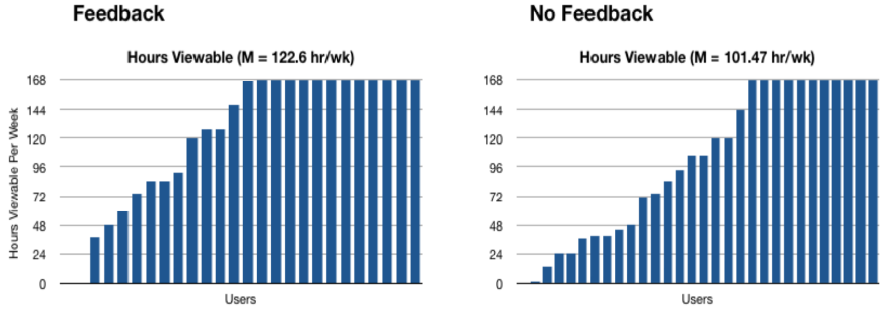
Evaluating the Usefulness of Feedback: Before/After Surveys – Facebook Study



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: Looking at People's Privacy Rules – Facebook Study

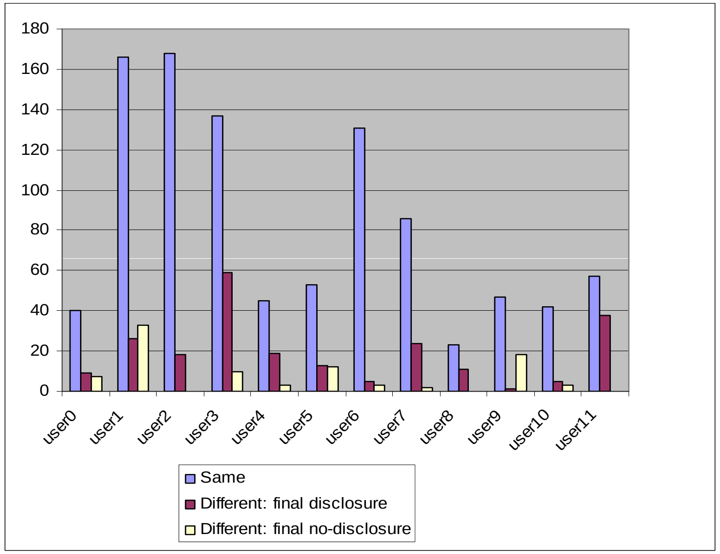
Examining Users' Privacy Rules **at the end** of the study



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

Policy Evolution – with feedback

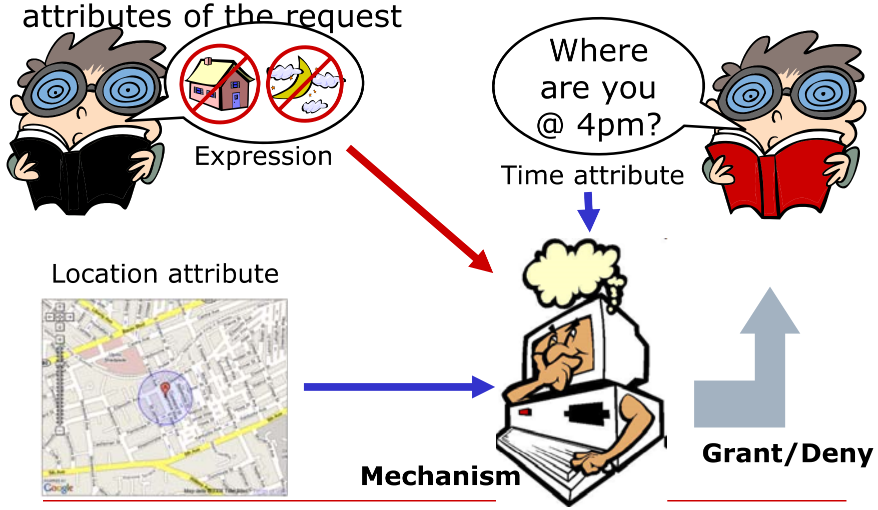


Data for 12 most active users across 3 pilots of PeopleFinder Application

How Expressive Should Policies Be?

What is a Privacy (Security) Mechanism?

- A function that chooses whether to deny a request for private info based on the expression of an agent & the attributes of the request

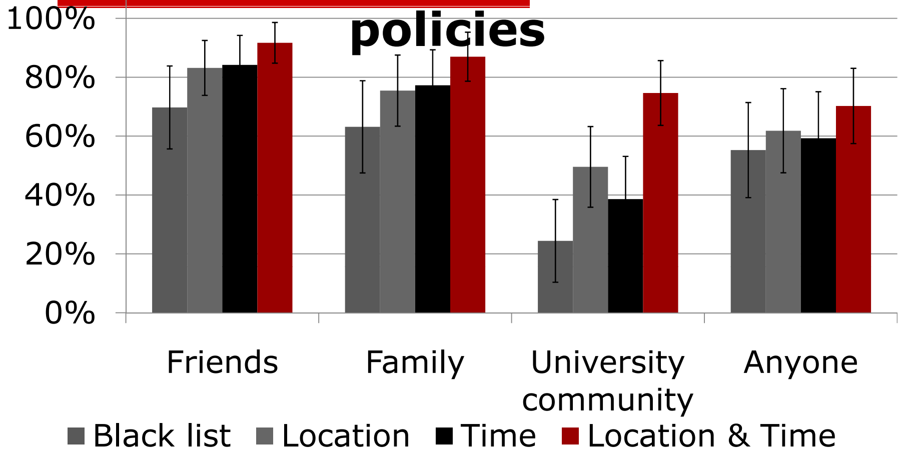


Expressiveness and Efficiency

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

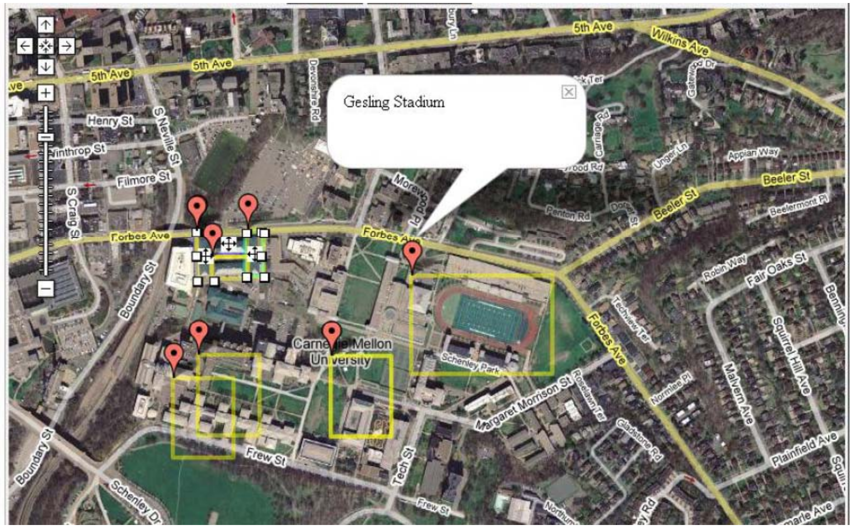
$$E[\mathcal{E}(f)] = \int_t P(t) \int_{\vec{a}} P(\vec{a})\ u(t, \vec{a}, f(h^*(t), \vec{a}))$$

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

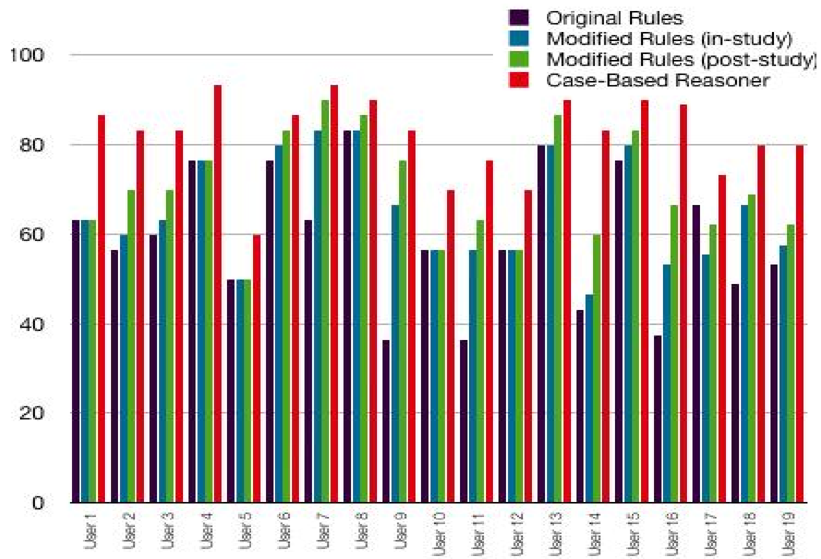


Observations

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

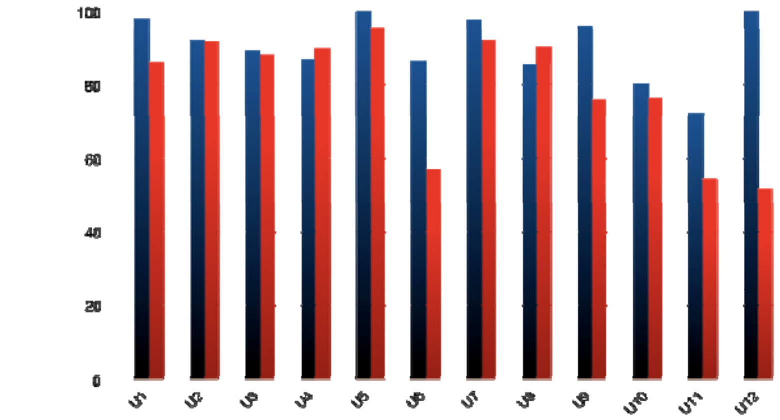
Could Machine Learning Help?

Early Experiment with Case-Based Reasoning



More Recent Pilots – 12 most active target users

3 Pilots – total of over 60 participants



Machine Learning
User-Defined Rules

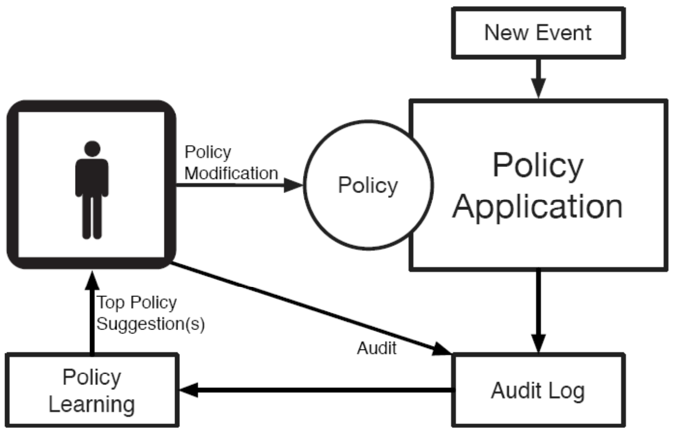
User-Defined Rules: 79% vs. ML: 91%

Note: Includes benefits of auditing

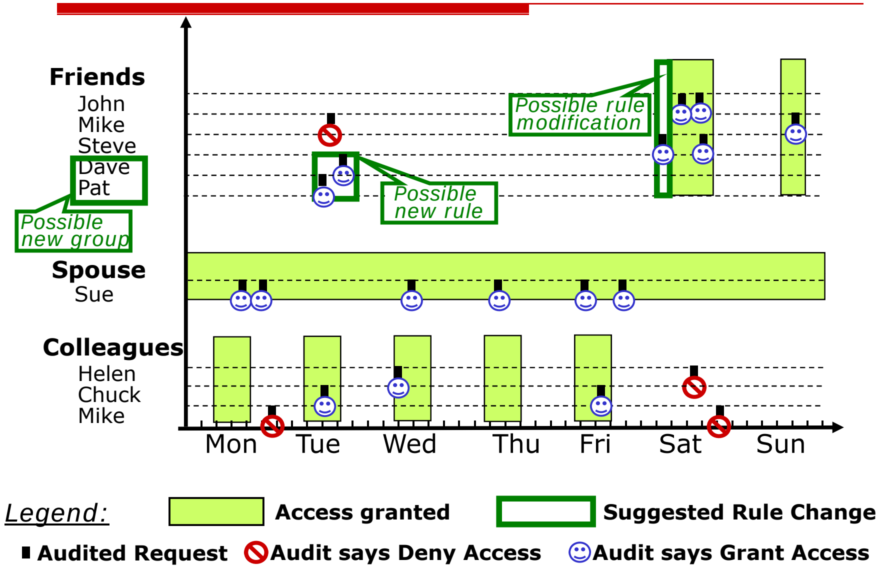
User-Controllable Policy Learning (patent pending)

- 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)



Suggesting Rule Modifications based on User Feedback (patent pending)



Exploring Neighboring Policies: Users Are More Likely to Understand Incremental Changes

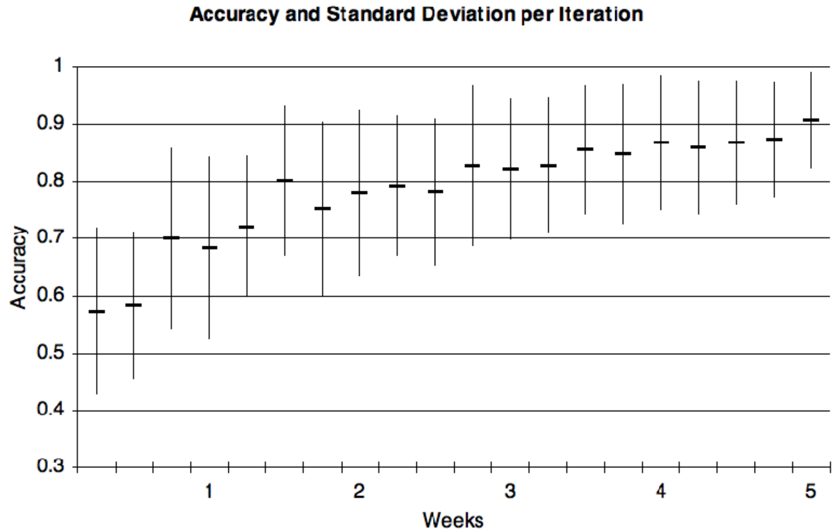
$$
\begin{array}{ll}
\text{Transform} :: \text{Restriction} \longrightarrow \mathbb{P}(\text{Restriction}) & \text{Restriction transformation function} \\
\text{GenActions} :: \text{Action} \longrightarrow \mathbb{P}(\text{Action}) & \text{Action transformation function} \\
\text{GenRules} :: \text{Rule} \longrightarrow \mathbb{P}(\text{Rule}) & \text{Rule generation function, where} \\
\quad \text{GenRules}((R, A)) = \bigcup_{r \in R} \bigcup_{r' \in \text{Transform}(r)} (r', A) \cup & \\
\qquad \bigcup_{a \in A} \bigcup_{a' \in \text{GenActions}(a)} (R, a') \cup & \\
\qquad \bigcup_{r \in R} R \setminus \{r\} \cup \bigcup_{r \in \text{Restriction}} R \cup \{r\} & \\
\qquad \bigcup_{a \in A} A \setminus \{a\} \cup \bigcup_{a \in \text{Action}} A \cup \{a\} & \\
\text{Neighbor} :: \text{Policy} \longrightarrow \mathbb{P}(\text{Policy}) & \text{Neighbor generation function, where} \\
\quad \text{Neighbor}(P) = \bigcup_{p \in P} \text{GenRules}(p) & \\
\qquad \bigcup_{r \in P} P \setminus \{r\} \cup & \\
\qquad P \cup \{(\emptyset, \emptyset)\} &
\end{array}
$$

Rate neighboring policies based on:

- **Accuracy**
- **Complexity**
- **Distance from current policy**

Emphasis on keeping changes understandable

With Suggestions for Policy Refinement



Expressiveness & User Bruden

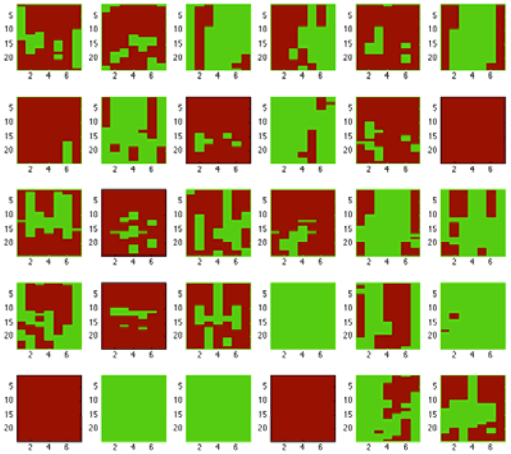
Average number of rules a user would have to define to achieve optimal efficiency

	Friends	Family	University community	Anyone	Total
Black list	N/A	N/A	N/A	N/A	1
Time	1.97	2.03	1.50	0.70	6.20
Location	6.90	6.23	3.30	1.37	17.80
Time/Location	7.97	7.97	5.23	2.73	23.90

Could Default Policies Help?

Identifying Default Policies (Ongoing Work)

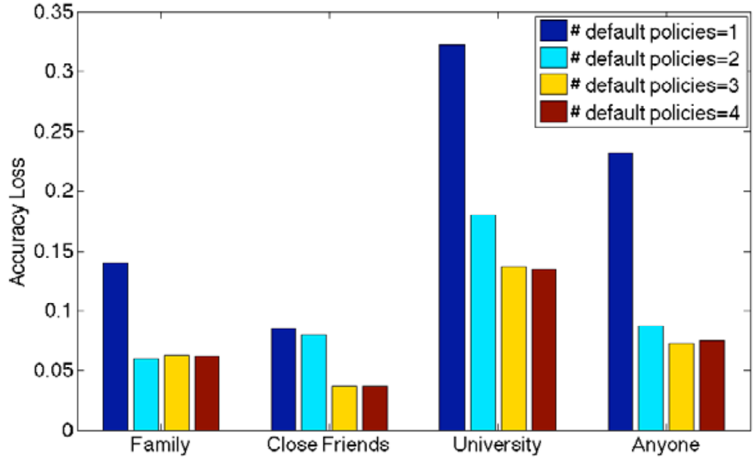
- Location sharing with members of the campus community – 30 different users



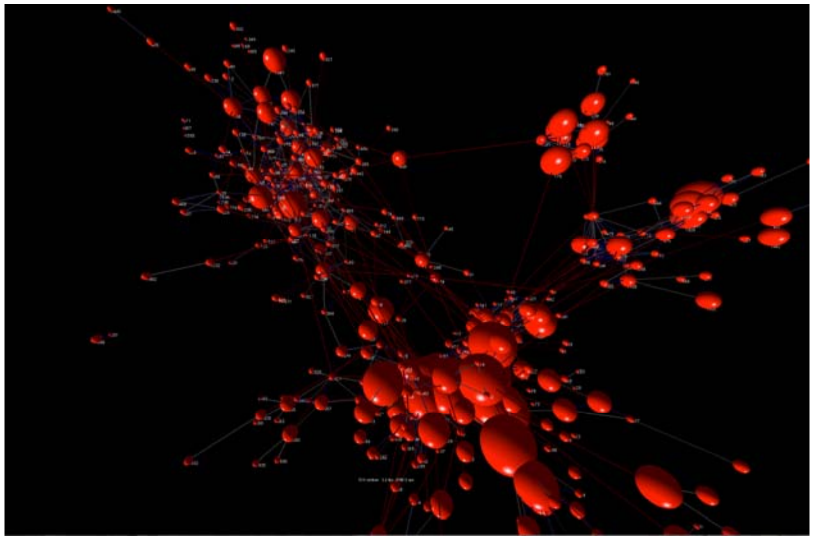
Green: Share
Red: Don't

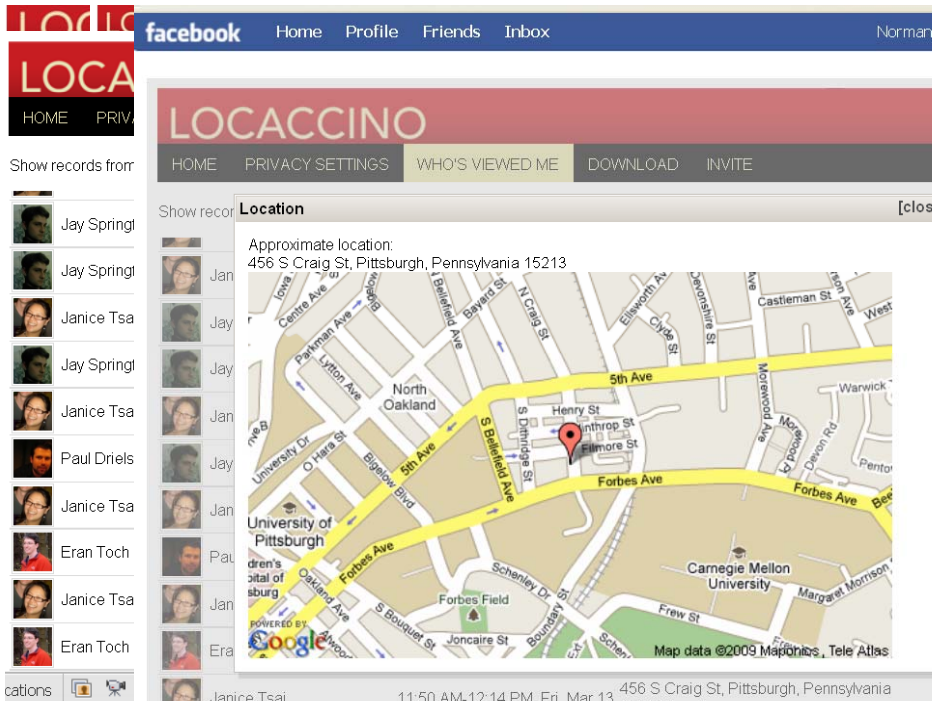
Clustering Canonical Policies

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



Social Networking View of Location Sharing





Nokia N95 Client

Adding More Functionality

Overall Vision

New Technology

Policy Creation

Jane

My colleagues can see my location on weekdays between 8am and 5pm

Policy Visualization

Policy Enforcement

Bob

Jane and Eric are late for our meeting. Show me where they are!

Jane is in Oakland but I can't access Eric's location

Bob's Phone

Policy Enforcing Engines

Policy Auditing & Refinement

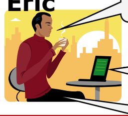
Eric

Why couldn't Bob see where I was?

Bob is a colleague. So far only your friends can see where you are

Explanation

Eric



What if my colleagues could see my location too?

Dialog

In the past you denied access to your colleague Steve

OK, make it just my superiors

Learning from the past

Time

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

Q&A

Support for this project has been received from the US National Science Foundation, the Army Research Office, CyLab,Microsoft, Nokia, FranceTelecom, Nortel, FCT through the CMU/Portugal Information and Communication Technologies Institute, and Skyhook Wireless.

Selected References

- **Norman Sadeh, Jason Hong, Lorrie Cranor, Ian Fette, Patrick Kelley, Madhu Prabaker, and Jinghai Rao. Understanding and Capturing People's Privacy Policies in a Mobile Social Networking Application *Journal of Personal and Ubiquitous Computing 2008.***
- **Janice Tsai, Patrick Kelley, Paul Hankes Drielsma, Lorrie Cranor, Jason Hong, and Norman Sadeh. Who's Viewed You? The Impact of Feedback in a Mobile-location System. Proceedings of *CHI '09.***
- **Patrick Kelley, Paul Hankes Drielsma, Norman Sadeh, Lorrie Cranor. User Controllable Learning of Security and Privacy Policies. Proceedings of *AISec 2008.***
- **Ramprasad Ravichandran, Michael Benisch, Patrick Gage Kelley, and Norman M. Sadeh. "Capturing Social Networking Privacy Preferences: Can Default Policies Help Alleviate Tradeoffs between Expressiveness and User Burden?". PETS '09.**
- **Michael Benisch, Patrick Gage Kelley, Norman Sadeh, Tuomas Sandholm, Lorrie Faith Cranor, Paul Hankes Drielsma, Janice Tsai. The Impact of Expressiveness on the Effectiveness of Privacy Mechanisms for Location Sharing. CMU-ISR Tech Report 08-141**
- **Jason Cornwell, Ian Fette, Gary Hsieh, Madhu Prabaker, Jinghai Rao, Karen Tang, Kami Vaniea, Lujo Bauer, Lorrie Cranor, Jason Hong, Bruce McLaren, Mike Reiter, and Norman Sadeh. User-Controllable Security and Privacy for Pervasive Computing. Proceedings of IEEE *HotMobile* 2007.**
- **Norman Sadeh, Fabien Gandon and Oh Buyng Kwon. Ambient Intelligence: The MyCampus Experience Chapter 3 in "Ambient Intelligence and Pervasive Computing", Eds. T. Vasilakos and W. Pedrycz, ArTech House, 2006. (Also available as Tech Report CMU-ISRI-05-123, Sch. of Computer Science, Carnegie Mellon Univ)**

Selected Press Coverage

- J. Young, "Now You Can Track Colleagues and Students on Your Laptop", Chronicle of Higher Education, Feb 2009 - http://chronicle.com/free/v55/i25/25a01501.htm
- R. Power, "Q&A with Norman Sadeh", CyLab Chronicles, http://www.cylab.cmu.edu/research/chronicles/sadeh.html
- BusinessWeek blog, March 2009
- Numerati blog, March 2009
- The Piper, March 2009
- "Locaccino Enables the Watched to Watch the Watcher", CyBlog, March 2009 http://www.cyblog.cylab.cmu.edu/2009/03/cylab-research-update-locaccino-enables.html

A Video

http://www.screentoaster.com/watch/stUkxdQERIR11dSVleWlJZUlRU/specifind_demo