User Studies on the Internet
Steven M. Bellovin
https://www.cs.columbia.edu/~smb

Goals
- Study how users control their privacy on Facebook
- Recruit subjects via the Internet
- In other words, using the Internet to study behavior on the Internet
- (Experiments done by PhD student Maritza Johnson)

Facebook
- Vastly popular — 800 million users
- Every item you post has its own privacy setting: shared with friends, friends of friends, “network”, or the world
- Custom settings are also possible
- Questions to study:
  - Do users care about privacy?
  - Is the interface usable?
  - If not, can we improve it?
Do Facebook Users Care About Privacy?

- Yes, they say so
  - Confirmed by later Pew Foundation Survey
- More important, people do take certain privacy-protecting actions
- But – they rarely change permission settings

Is There Really a Problem?

- In a study of CU undergraduates, every single subject confirmed at least one error in their permission settings
- Not surprisingly, people’s attitudes about privacy vary with subject matter
- But our subjects rarely fixed the problems...
- Hypothesis: permission-setting mechanisms are too complex

Another Experiment

- Show subjects who can see what
- If they’re unhappy about it, ask if they want to fix the problem, and tell them exactly how to do so
- Result: improved behavior; more people fixed the problematic permissions
- Still nowhere near perfect – but since people take other privacy-preserving measures (e.g., “untagging” photographs), we conclude that the permission-setting interface is fundamentally broken
Finding Experimental Subjects

- Advertising: Craig’s List and via Google Ad Words
- Amazon’s “Mechanical Turk”
  - People paid to do all sorts of things, including – in our case – going through our study
  - “Mechanical Turk” name comes from fake chess-playing automaton
- Research Match
  - Consortium of universities; used to match volunteers with IRB-approved studies

Conducting Research over the Net: Issues

- Demographic mix
- Verifying demographic information, especially age
- Terms of service for Facebook and Amazon
- Limitations of the experimental environment

Demographic Mix

- Does our pool of subjects match Facebook’s user population, in age, gender, educational level, etc.?
- Research Match did better – but we had some problems with non-uniform dropouts
Verifying Demographic Information
- Hard to verify subject-supplied data: “On the Internet, no one knows you’re a dog!”
- Must have age for consent purposes!
  - Is Facebook’s assertion good enough?
  - It’s probably reasonably accurate but by no means perfect
  - Research Match doesn’t verify self-reported demographic data, either
  - (It also sends far too much for our purposes – we weren’t doing a study where medical conditions are relevant)

Terms of Service
- Various services limit what you’re allowed to do
  - Amazon: You may not ask Turkers to install software on their machines
  - Makes sense from a computer security perspective, but limits the types of experiments we can do
  - Facebook: users may not share passwords

Environment Limitations
- Facebook apps could not encrypt data – gave us no safe way to transmit data to Columbia for more detailed analysis or reanalysis
- Facebook apps cannot directly access object permissions – forced clumsy workarounds, and (in one study) forced us to limit our population to Columbia students
- Apps cannot modify permissions – no way to experiment with better user interface
  - A front end? Maybe – but that would require subjects to enter their Facebook passwords into our code, which is risky for both us and them