UNSW
THE UNIVERSITY OF NEW SOUTH WALES

NICTA

Exposing and Mitigating Privacy Loss in Crowdsourced Survey Platforms

Thivya Kandappu, Vijay Sivaraman, Arik Friedman, Roksana Boreli

Crowdsourcing Platforms

- Amazon Mechanical Turk (AMT)
  - >500K users
  - Widely used in psychology and social studies
  - Trusted curator
  - User's unique id, ip-address, city and country are revealed to the surveyor

- Google Consumer Surveys
  - Trusted curator
  - One question at a time

User de-anonymization is easy!

- We launched a series of survey tasks in AMT
- Survey 1: astrology services
  - Star sign, date/month of birth, beliefs in astrology, …
- Survey 2: online match making services
  - Gender, age, marital status, usage of match-making, …
- Survey 3: mobile phone coverage
  - Zip code, phone signal strength and quality, …
- 100 respondents for each survey, 3 hours, $30
- 72 users took all 3 surveys: got their DoB, gender, Zip
  - Can de-anonymize these users with high probability (~76%)

Private Information is Easy to Extract

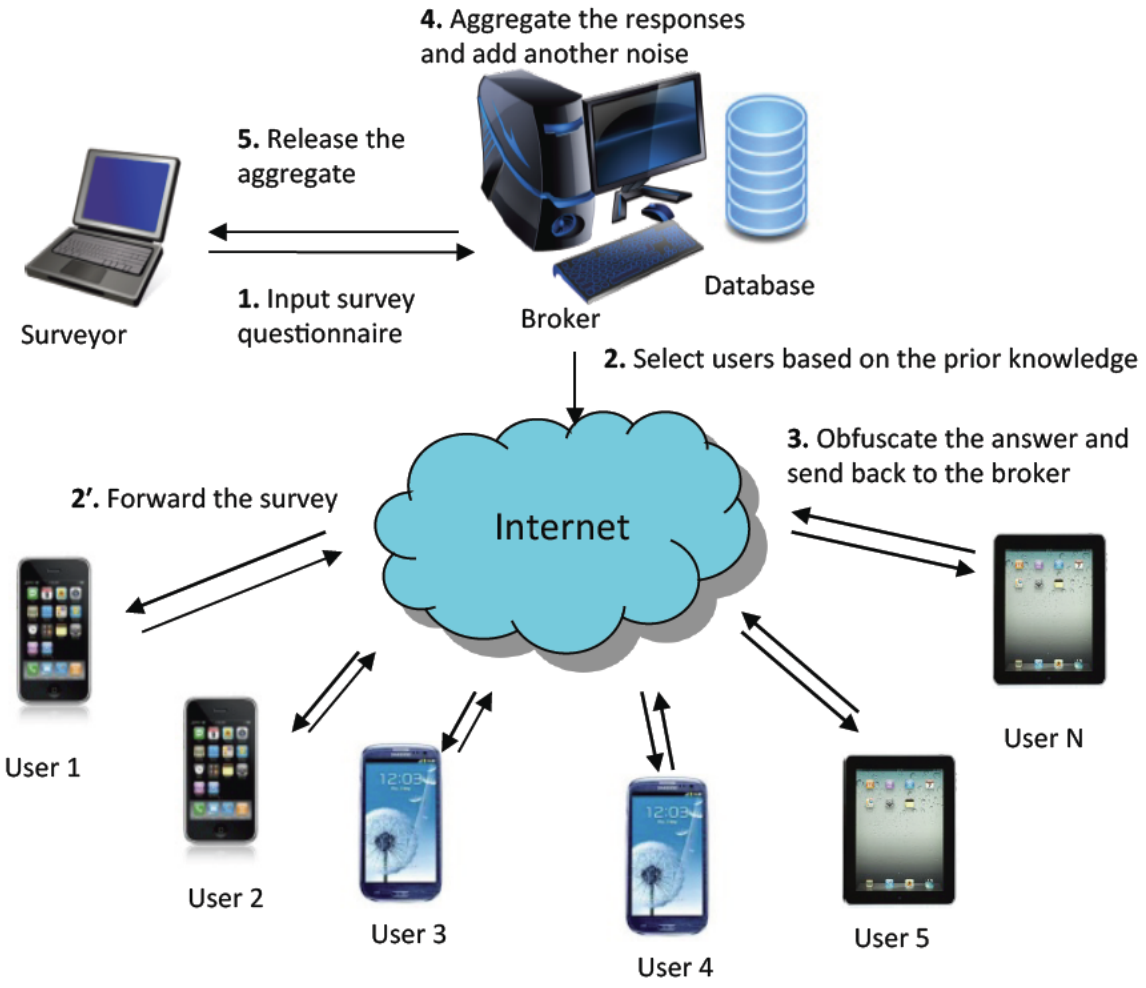
- Survey 4: smoking habits
  - Smoking intensity, coughing frequency, income,…
- 18 of the 72 de-anonymized participants took this survey
- Got highly personal information for these individuals
  - Respiratory health, income, …
- Easy to obtain personal information on these platforms!
- Survey 5: user perception of privacy in such platforms (would you do this if you know you can be de-anonymized?)
  - 73 out of 100 users said they would not have participated

Available Solutions

- Anonymize the user
  - Can still deduce from device-id, IP address
- Trust the surveyor (curator)
  - E.g. trust Google surveys not to sell your data!
  - Or trust lawyers to offer you legal protection
- Obfuscate your answer (hide in the crowd!)
  - Add noise to individual responses
  - Surveyor cannot get accurate individual information but can get accurate “on average” information about the population

System Architecture

- Semi-trusted broker
- Users choose their own comfort privacy level
- Broker tries to maximize the utility of the estimation



Loki: Privacy Preserving Mobile App

- User: https://itunes.apple.com/tr/app/loki/id767077965?mt=8
- Surveyor: http://loki.eng.unsw.edu.au/
- Evaluate the system with 130 volunteer students

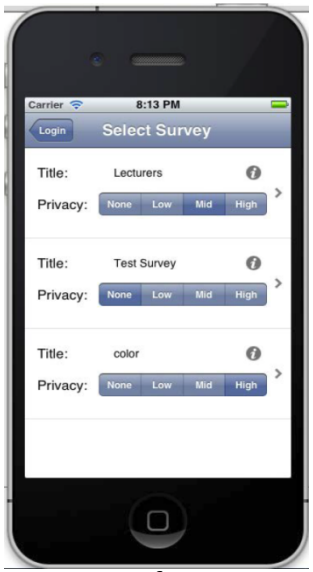


Fig 1. List of surveys and privacy levels available to the users

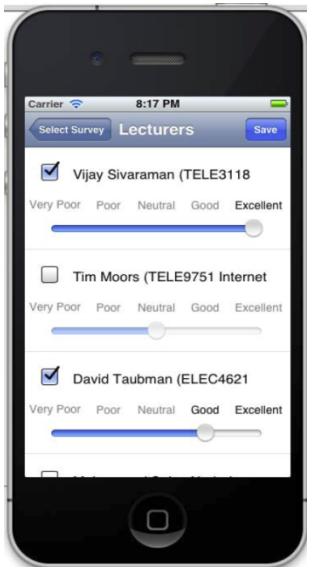


Fig 2. Questions and rating entered by the user

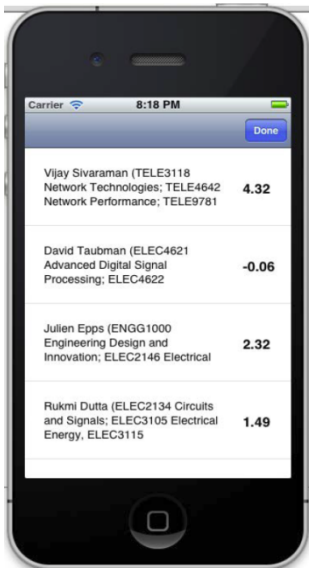


Fig 3. Uploaded user responses after noise addition

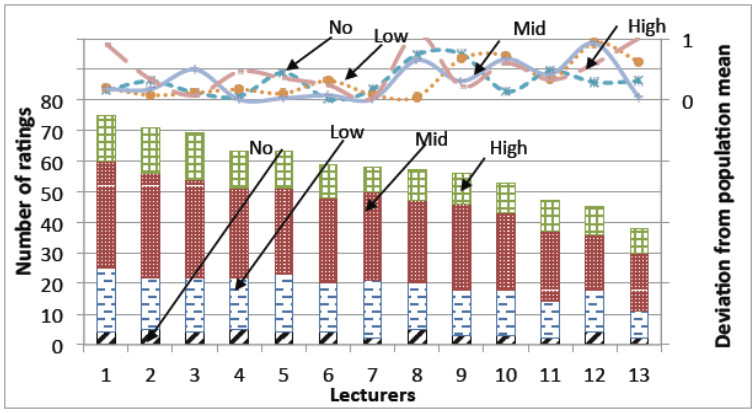


Fig 4. Deviation in mean across the bins for various lecturers

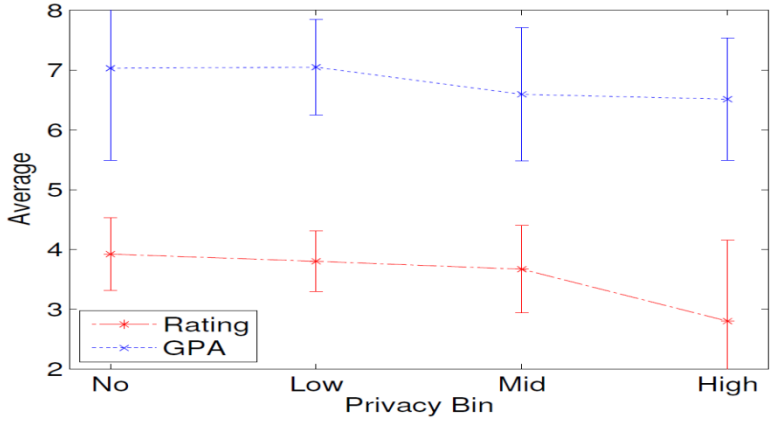


Fig 5. Correlation of privacy choice with average ratings and student GPA

Conclusion

- Demonstrated that privacy is easily compromised in current online crowdsourcing survey platforms
- Developed a new platform that dispenses with trusted entities, and allows users to obfuscate their answers at source
  - Ratings based and multiple-choice questions
  - Aggregated answers with confidence levels