Incentives for Data Gathering in Community Sensing


Adish Singla


Andreas Krause

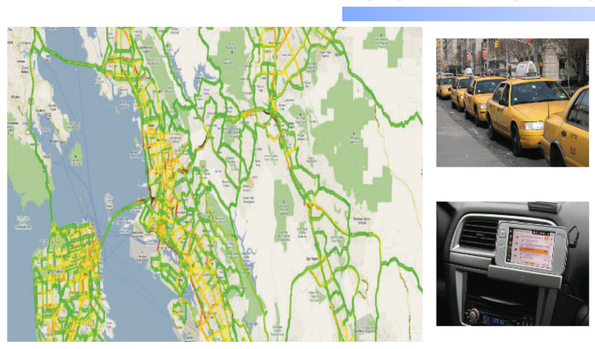
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Contribution: **Privacy aware**, **Truthful** and **Adaptive** mechanism **SEQTGREEDY** for recruiting participants in **Community sensing** application.

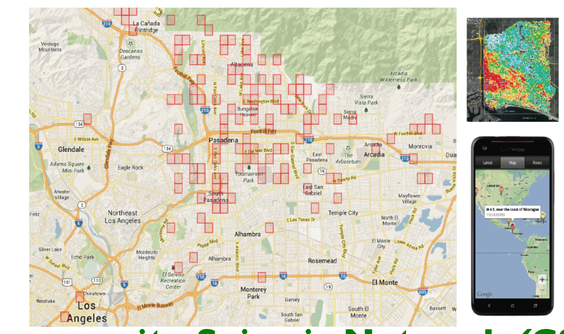
Community Sensing

Estimate Spatial phenomenon

- Community owned devices
- Low-cost sensors
- Dense sensing network



Mobile Millenium: Traffic monitoring
Berkeley, California



Community Seismic Network (CSN)
Earthquake monitoring Pasadena, California

Open Sense: Air Quality monitoring
Lausanne/Zurich, Switzerland

Privacy and Incentives

Privacy Concerns

- Sharing of private attributes, e.g. location
- Inference of location from GPS traces
- General anxieties


Krumm'07

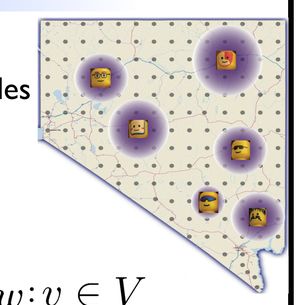
KEY QUESTION : How to **valuate** and **negotiate** access to **private information**?

Use **monetary incentives** to **compensate** for **information shared**

Sensing Phenomena

Environment

- Set of discrete locations V, e.g. zip codes



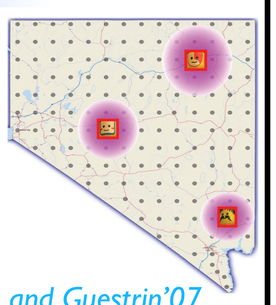
User's sensing location

- Location of the user w is given by $y_w : v \in V$

Sensing Utility

Utility acquired

- Sensing locations $A \subseteq V$: $f(A)$



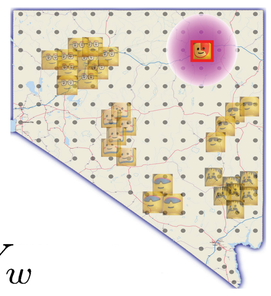
Submodular Set functions

- Notion of diminishing return
- Captures many complex utilities: Krause and Guestrin'07
  - *E.g.* reduction of predictive uncertainty in a probabilistic model
- Near-opt polynomial-time solutions: Nemhauser'78, Feige'98

Privacy through Obfuscation

Obfuscation

- Users share obfuscated location
- Reduce the risk of identifiability
  - Sweeney'02, Dwork'06



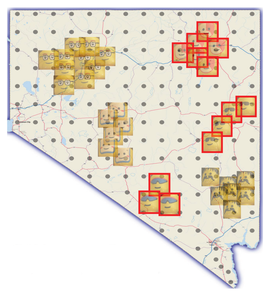
User's privacy profile

- Model user's location as random variable Y_w
- Exact user location y_w revealed after recruitment and payment
- y_w sampled from user's shared distribution $P(Y_w = v)$

Adaptive Selection of Participants

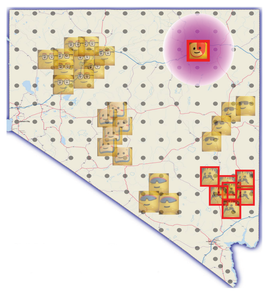
Non-Adaptive

- Select without observations



Adaptive

- Sequential selection policies



Strategic Users and Truthfulness

Bidding Model

- True cost: c_w , declared bid: b_w

Strategic Users

- Aim to maximize profit by bids
- Mechanism can do **arbitrarily bad**

Truthful Mechanisms

- Dominant strategy for users to declare $b_w = c_w$

Protocol: Mechanism and Users

Mechanism

I

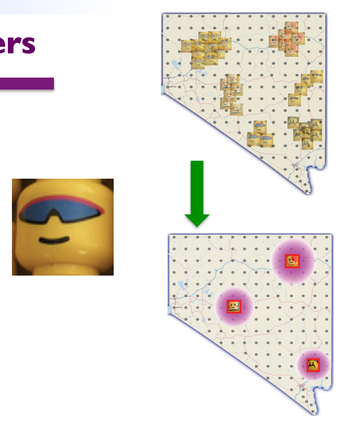
Privacy profiles and bids from all users

Allocate next participant
Makes a payment p_w to the participant

t

Participant reveals the actual location
Sends the sensing data

T (budget *B* exhausts)



Main Research problem addressed

Mechanisms for recruiting participants in community sensing:

Privacy Aware ***Truthful Payments*** ***Adaptive Selection***

With following desirable properties:

- Budget feasibility
- Polynomial time computation

Our Mechanism: SEQTGREEDY

Greedy Selection Rule

Expected marginal gain — Conditioned on past observations

$$w^* \leftarrow \arg\max_{w \in \mathcal{W}'} \frac{\Delta_g(w|\mathbf{y}_\mathcal{S})}{b_w}$$

Bid

Early Stopping

Proportional share criteria

$$b^*_w \leq \frac{\mathcal{B}}{\alpha} \cdot \frac{\Delta_{w^*}}{\left(\left(\sum_{s \in \mathcal{S}} \Delta_s\right) + \Delta_{w^*}\right)}$$

Reduced budget

Truthful Payments

- Threshold Payments : Myerson'81
- Expected maximum raise in bid a user can do before being removed
- Payments depend on observations

SEQTGREEDY: Analysis

Main Results

- Achieves a utility at least $\frac{1}{3}\left(1 - \frac{1}{e}\right)$ i.e. ~**21%** compared to that of **SEQOPT** (non-polynomial with unrealistic access to true costs)
  - Under the assumption, users are independent
  - And, utility gain from each user is small
- Generalizes results of Singer'10, Chen'11 for adaptive submobularity.

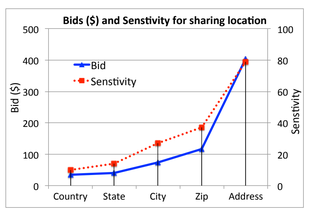
Case study of Air Quality Monitoring

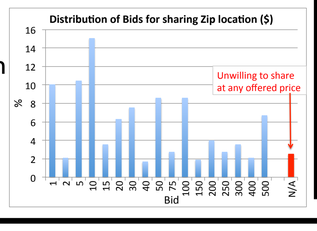
Environment

- Zip data for Nevada State (220 zip codes)
- Population statistics to simulate users' locations

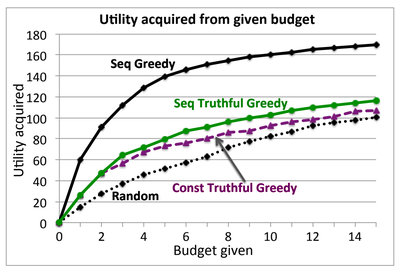
Data from Mechanical Turk

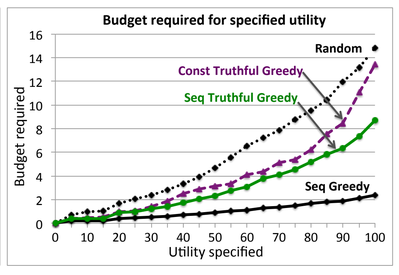
- Survey about participation in application
- **75%** users responded positively for participation
- Realistic cost distributions
- Mobility data to simulate sensing radius for user





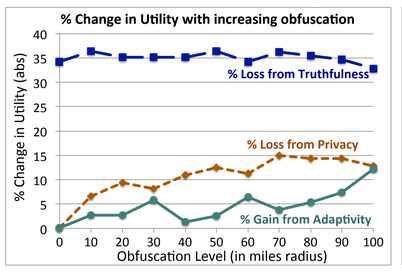
Results: Varying Budget and Utility

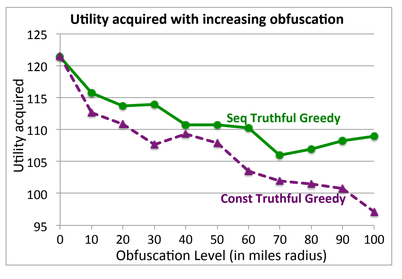




- **R**ANDOM : Naïve Baseline with unrealistic access to true costs
- **SEQG**REEDY : Near-optimal with unrealistic access to true costs
- **CONSTTG**REEDY : Non-adaptive truthful (Singer'10, Chen'11)

Results: Varying Obfuscation





- **% Gain from Adaptivity**
- **% Loss from Truthfulness**
- **% Loss from Privacy**