Truthful Mobile Crowdsensing for Strategic Users with Private Qualities

Xiaowen Gong, Ness Shroff

Department of Electrical and Computer Engineering
Auburn University

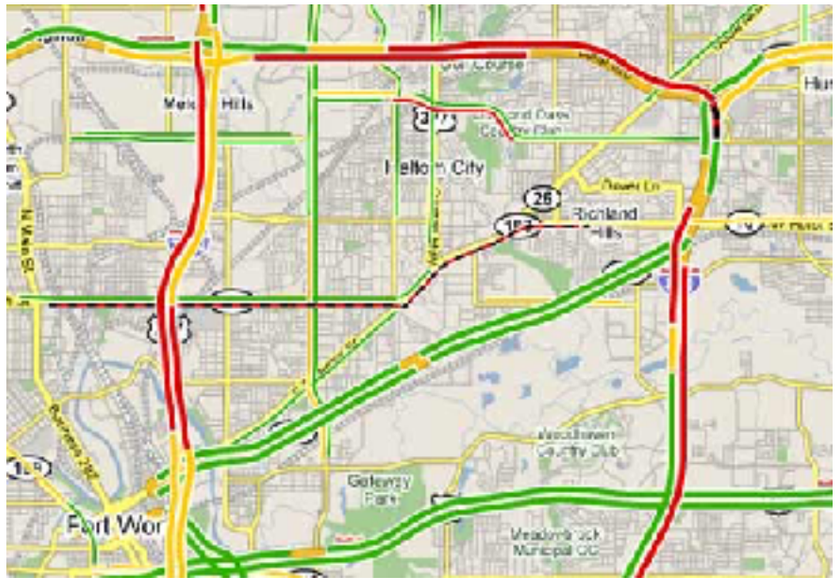
Department of Electrical and Computer Engineering &
Department of Computer Science and Engineering
The Ohio State University

Presenter: Morteza Hashemi

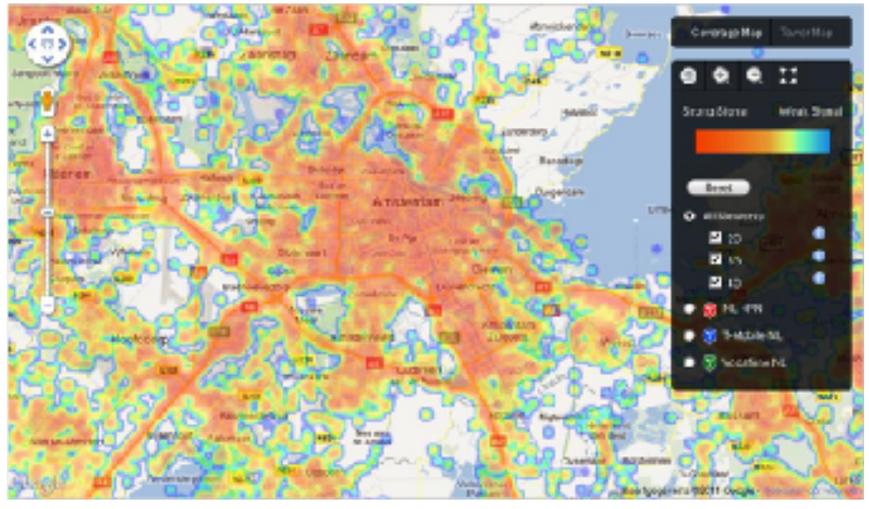
WiOpt 2017, Paris, France
May 17th, 2017

Mobile Crowdsensing

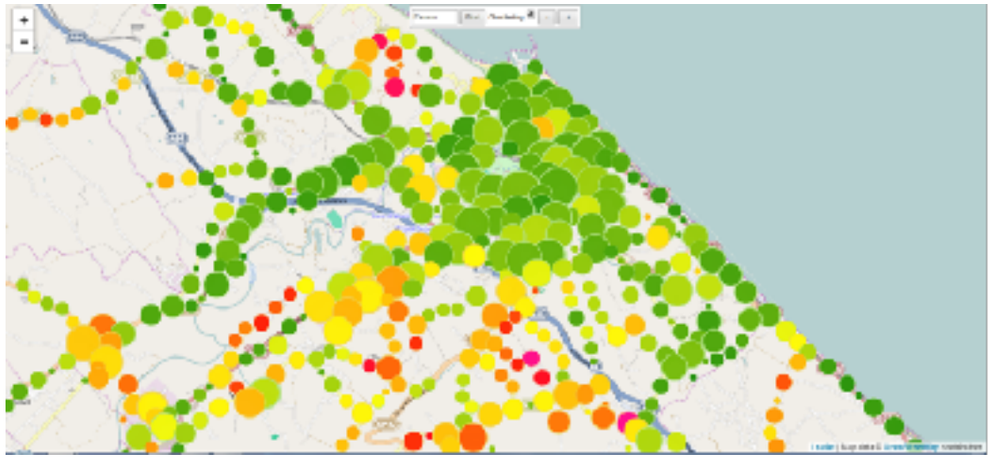
- Mobile crowdsensing: leverage "wisdom" of a crowd of mobile users by collecting their sensing data
  - ✓ Enabled by powerful mobile devices and pervasive wireless connectivity
  - ✓ Wide range of applications: environmental monitoring (e.g., weather condition, air pollution), traffic monitoring, spectrum sensing…
  - ✓ Important paradigm to exploit big data opportunities
  - ✓ Provide enormous potential via machine learning/data mining tools



Google Map
(traffic condition)



Open Signal
(cellular coverage)



Smart Road Sense
(road quality)

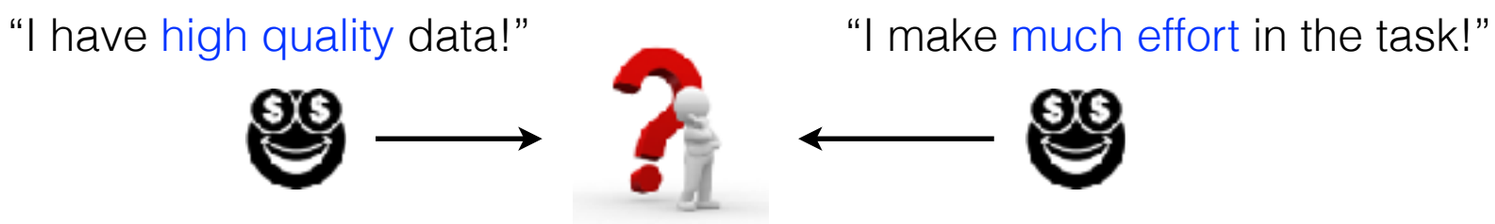
Quality-Aware Crowdsensing

- Data quality captures how accurate is a user's data compared to ground truth
  - ✓ Data accuracy is a key performance metric for crowdsensing
    - Data observed by individual users are inherently inaccurate due to noise, interference, error…
    - Exploit diversity gain by aggregating data from the crowd to improve data accuracy
  - ✓ E.g., to measure transmit signal strength of a transmitter, the received SNR determines data quality
  - ✓ Users have diverse data quality depending on their specific situations (e.g., location)

- Information of users' data quality is important for the crowdsensing requester
  - ✓ Assign more work to users with high quality data
  - ✓ Assign larger weights to high quality data in data aggregation
  - ✓ Know the accuracy of aggregated data

Challenges in Quality-Aware Crowdsensing

- A user's data quality can be its private information
  - ✓ A user can learn its data quality based on its relevant private information (e.g., location)
  - ✓ Cannot be known or verified by the requester
  - ✓ A user may have incentive to manipulate its data quality revealed to the requester
- A user's effort exerted in the crowdsensing task can be its hidden action
  - ✓ Effort captures how much work is devoted to the task (e.g. sensing time, number of sensing samples)
  - ✓ A user can control its effort which affects its data quality and its cost in the task
  - ✓ Cannot be known or verified by the requester
  - ✓ A user may have incentive to manipulate its effort made in the task



How to incentivize users to behave truthfully?

Data Quality, User Quality and Effort

interested variable

data → $D_i \triangleq X + W_i$ ← noise

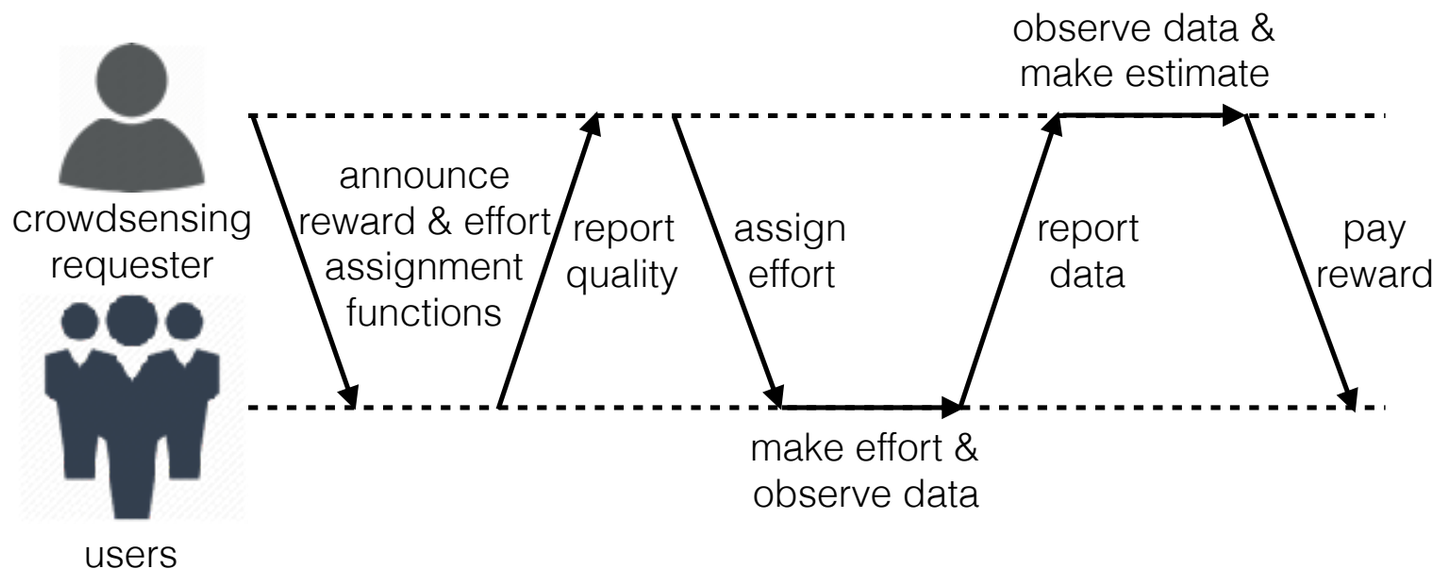
$$W_i \sim \mathcal{N}\left(0, \frac{q_i}{e_i}\right)$$

user quality

effort

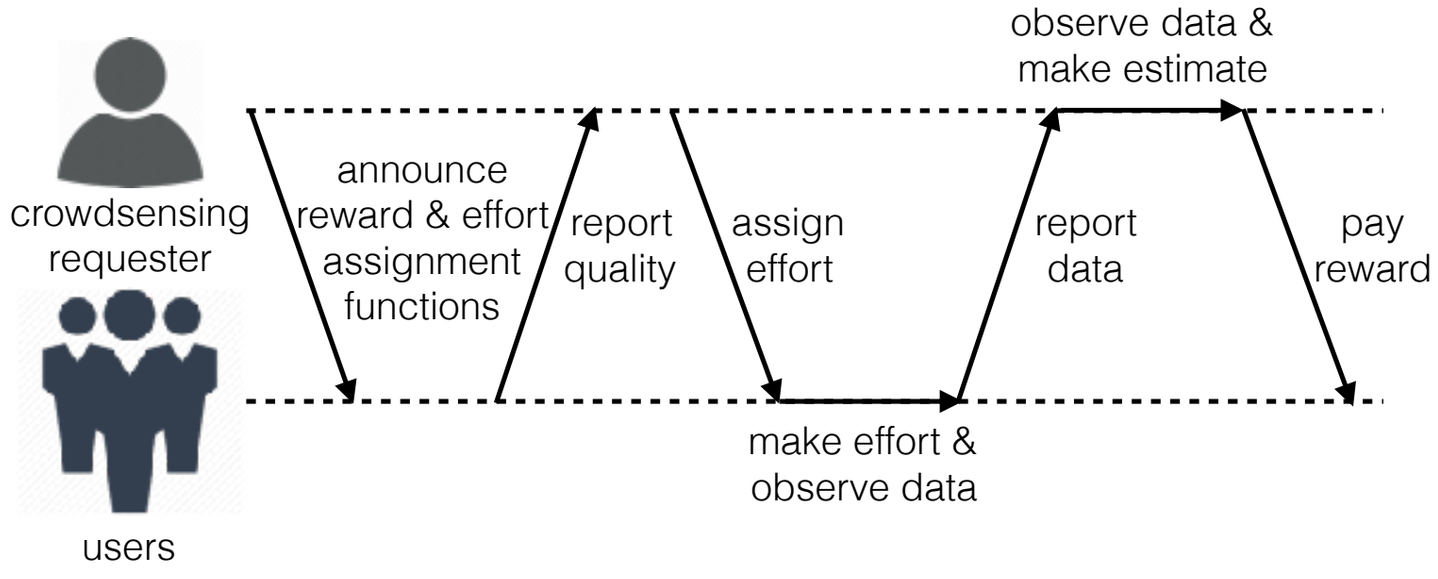
- Data quality is quantified by noise variance depending on user quality and effort
- User quality q_i quantifies how accurate is a user's data given its effort
  - ✓ Capture a user's intrinsic capability for the task
  - ✓ Private information of the user, unknown to the requester
- Effort e_i quantifies how much work is devoted by a user to the task
  - ✓ If e_i is the number of i.i.d. sensing samples taken and averaged, then the noise variance of the average is inversely proportional to e_i
  - ✓ Hidden action of the user, unknown to the requester

Quality Reporting, Effort Assignment



- Each user reports its quality q_i' to the requester
  - ✓ May have incentive to misreport $q_i' \neq q_i$
- The requester assigns a desired effort $e_i'(\boldsymbol{q}')$ to each user based on the reported qualities
  - ✓ The requester may have desired effort assignments (e.g., to maximize social welfare)
  - ✓ A user may have incentive to not make the desired effort $e_i \neq e_i'$
  - ✓ Effort assignment function is pre-defined and announced to users first
- Each user makes an effort to observe data and reports it to the requester
  - ✓ Assume that data are reported truthfully
    - Difficult for a user to misreport data to gain an advantage

Data Estimation, Reward Payment



- The requester obtains the MMSE estimate based on the reported qualities and data, and assigned efforts

$$x^* \triangleq \arg\min_{x'} E_{X|\boldsymbol{d}(\boldsymbol{q}',\boldsymbol{e}')}\left[(x'-X)^2\right]$$

- The requester's crowdsensing utility is the estimation loss which is the MSE

$$l(\boldsymbol{d},\boldsymbol{q}',\boldsymbol{e}') \triangleq E_{X|\boldsymbol{d}(\boldsymbol{q}',\boldsymbol{e}')}\left[(x^*-X)^2\right]$$

- The requester pays a reward $r_i(d_0, d_i, \boldsymbol{q}', e_i')$ to each user based on its own observed data d_0, the reported qualities and data, and assigned efforts
  - ✓ Reward function is pre-defined and announced to users first
    - Can only depend on information known to the requester
    - Ground truth of interested variable X is unknown

Mechanism Design Objective

- Each user's payoff $u_i(d_0, d_i, \boldsymbol{q}', e_i) \triangleq r_i(d_0, d_i, \boldsymbol{q}', e_i') - c_i e_i$
  - ✓ Cost coefficient c_i quantifies how much resource is consumed for each unit of effort
    - Assume that it is known and uniform (can be relaxed to diverse case)
  - Requester's payoff $u_0(x, \boldsymbol{d}, \boldsymbol{q}', \boldsymbol{e}') \triangleq -l(\boldsymbol{d}, \boldsymbol{q}', \boldsymbol{e}') - \sum_{i \in \mathcal{N}} r_i(d_0, d_i, \boldsymbol{q}', e_i')$
- Dominant incentive compatible (DIC): Given any qualities reported by other users, the optimal strategy of each user for maximizing its expected payoff is to truthfully report its quality and make the effort desired by the requester

$$E_{D_0, D_i(q_i, e_i)}\left[u_i(D_0, D_i, q_i, \boldsymbol{q}'_{-i}, e_i')\right] \geq E_{D_0, D_i(q_i, e_i)}\left[u_i(D_0, D_i, q_i', \boldsymbol{q}'_{-i}, e_i)\right], \forall (q_i', e_i), \forall \boldsymbol{q}'_{-i}$$

  - ✓ If users are not truthful, the requester's effort assignment are not desired, data estimation is not optimal, and its information of data accuracy is incorrect!!
  - ✓ Dominant is strong and desirable
- Individual rational (IR): Given that a user truthfully reports its quality and makes the effort desired by the requester, its expected payoff is nonnegative
  - ✓ Each user's reward can compensate its cost

Quality and Effort Elicitation Mechanism

Quality and Effort Elicitation (QEE) mechanism: A pair of any effort assignment function $e'_i(\boldsymbol{q}')$ satisfying

$$e'_i(q'_i, \boldsymbol{q}'_{-i}) \geq e'_i(q''_i, \boldsymbol{q}'_{-i}),\ \forall q'_i \leq q''_i,\ \forall \boldsymbol{q}'_{-i}$$

and a reward function based on that $e'_i(\boldsymbol{q}')$ given by

$$r_i(d_0, d_i, \boldsymbol{q}', e'_i) = c\int_{q'_i}^{\bar{q}} \frac{e'_i(q, \boldsymbol{q}'_{-i})}{q} dq + 2ce'_i(\boldsymbol{q}') \frac{ce'^2_i(\boldsymbol{q}')}{q'_i} \left[(d_0 - d_i)^2 - \frac{q_0}{e_0}\right].$$

- ✓ The requester has quality q_0 and makes effort e_0
- ✓ The condition on the effort assignment function: assigns more effort for higher quality
  - General, natural (e.g. satisfied for maximizing social welfare)

Properties of QEE Mechanisms

Lemma: Under the QEE mechanisms, given that any user reports any quality q_i' and its data d_i and makes any effort e_i, its expected payoff is given by

$$E_{D_0,D_i(q_i,e_i)}[u_i(D_0,D_i,\boldsymbol{q}',e_i)]=c\int_{q_i'}^{\bar{q}}\frac{e_i'(q,\boldsymbol{q}'_{-i})}{q}dq+2ce_i'(\boldsymbol{q}')\frac{ce_i'^2(\boldsymbol{q}')q_i}{e_iq_i'}-ce_i.$$

- ✓ A user's expected payoff can be expressed as a function of its true quality and actual effort

Lemma: Under the QEE mechanisms, given that any user reports any quality q_i' and its data d_i, its optimal actual effort is given by

$$e_i=\sqrt{\frac{q_i}{q_i'}}e_i'(\boldsymbol{q}').$$

- ✓ The optimal actual effort e_i is equal to the desired effort e_i' when the reported quality q_i' is equal to the true quality q_i
  - If quality elicitation is achieved, then effort elicitation is also achieved
  - A key step to overcome intricate coupling between quality elicitation and effort elicitation

Properties of QEE Mechanisms

Lemma: Under the QEE mechanisms, given that any user reports its data d_i and makes its optimal actual effort, its optimal reported quality is its true quality $q_i' = q_i$.

Theorem: The QEE mechanisms are dominant incentive compatible and individual rational.

- A user's expected payoff when it truthfully reports its quality and make the desired effort

$$c \int_{q_i}^{\bar{q}} \frac{e_i'(q, \boldsymbol{q}'_{-i})}{q} dq$$

  - ✓ Always nonnegative
    - The requester pays "information rent" due to uncertainty about users' qualities
  - ✓ Decreases when the upper bound $\bar{q}$ of quality decreases
    - Less "information rent" when having less uncertainty

Optimal Effort Assignment for QEE Mechanisms

- The requester's optimal (CO) effort assignment maximizes the requester's expected payoff among all the QEE mechanisms

$$\{\boldsymbol{e}^*(\boldsymbol{q}), \forall \boldsymbol{q}\} \triangleq \arg \max_{\{\boldsymbol{e}(\boldsymbol{q}), \forall \boldsymbol{q}\}} E_{X, \boldsymbol{D}(\boldsymbol{Q}, \boldsymbol{e})}[u_0(X, \boldsymbol{D}, \boldsymbol{Q}, \boldsymbol{e})]$$

Theorem: When $c \geq 1/(9\underline{q})$, the requester's optimal effort assignment is given by

$$e_i^*(\boldsymbol{q}) = \begin{cases} \max\left\{ q_i \left(\frac{1}{\sqrt{\alpha(q_i)}} - 1 \right), 0 \right\}, & i = \arg\min_j \alpha(q_j) \\ 0, & \text{otherwise} \end{cases}, \forall \boldsymbol{q},$$

where

$$\alpha(q_i) \triangleq c \left(\frac{F(q_i)}{f(q_i)} + q_i \right), \ \forall i$$

✓ Assume that users' qualities are i.i.d. with PDF $f(q)$ and CDF $F(q)$ known by the requester

✓ Assign effort only to the "best" user

- Appear to be "single-sensing" rather than "crowdsensing"
- Actually exploit diversity gain in users' qualities
- In contrast to statistical estimation that use data from multiple sources
- Simplify implementation: only need to collect data from the best user

Optimal Effort Assignment for QEE Mechanisms

Theorem: When $c \geq 1/(9\underline{q})$, the requester's optimal effort assignment is given by

$$e_i^*(\boldsymbol{q}) = \begin{cases} \max\left\{ q_i \left(\frac{1}{\sqrt{\alpha(q_i)}} - 1 \right), 0 \right\}, & i = \arg\min_j \alpha(q_j) \\ 0, & \text{otherwise} \end{cases}, \forall \boldsymbol{q},$$ where $$\alpha(q_i) \triangleq c\left(\frac{F(q_i)}{f(q_i)} + q_i \right), \ \forall i$$

- ✓ The best user has the smallest "virtual valuation" $\alpha(q_i)$ rather than the highest quality
  - Depend on both the quality and the quality distribution
  - In the same spirit as the virtual valuation introduced by Myerson
- ✓ Assign more effort to the best user when its quality improves, or the cost coefficient decreases, or the quality range $\Delta q \triangleq \bar{q} - \underline{q}$ (and thus the virtual evaluation) decreases
  - Assign no effort when its quality is too low or when the cost is too high

Optimal Effort Assignment for QEE Mechanisms

- Social welfare is the crowdsensing utility minus the total cost of all users

$$v(\boldsymbol{e}(\boldsymbol{q})) \triangleq -E_{\boldsymbol{D}(\boldsymbol{q},\boldsymbol{e})}[l(\boldsymbol{D},\boldsymbol{q},\boldsymbol{e})] - \sum_{i\in\mathcal{N}} c_i e_i$$

- The socially optimal (SO) effort assignment maximizes the social welfare among all the QEE mechanisms

$$\boldsymbol{e}^{so}(\boldsymbol{q}) \triangleq \max_{\boldsymbol{e}(\boldsymbol{q})} v(\boldsymbol{e}(\boldsymbol{q}))$$

Theorem: The socially optimal effort assignment is given by

$$e_i^{so}(\boldsymbol{q}) = \begin{cases} \max\left\{q_i\left(\frac{1}{\sqrt{cq_i}}-1\right),0\right\}, & i=\arg\min_j q_j \\ 0, & \text{otherwise} \end{cases}.$$

✓ Assign effort only to the "best" user with the highest quality

- Assign more effort than the requester's optimal effort assignment
- The requester's optimal effort assignment is not socially optimal due to the requester's uncertainty about users' qualities

Optimal Effort Assignment for QEE Mechanisms

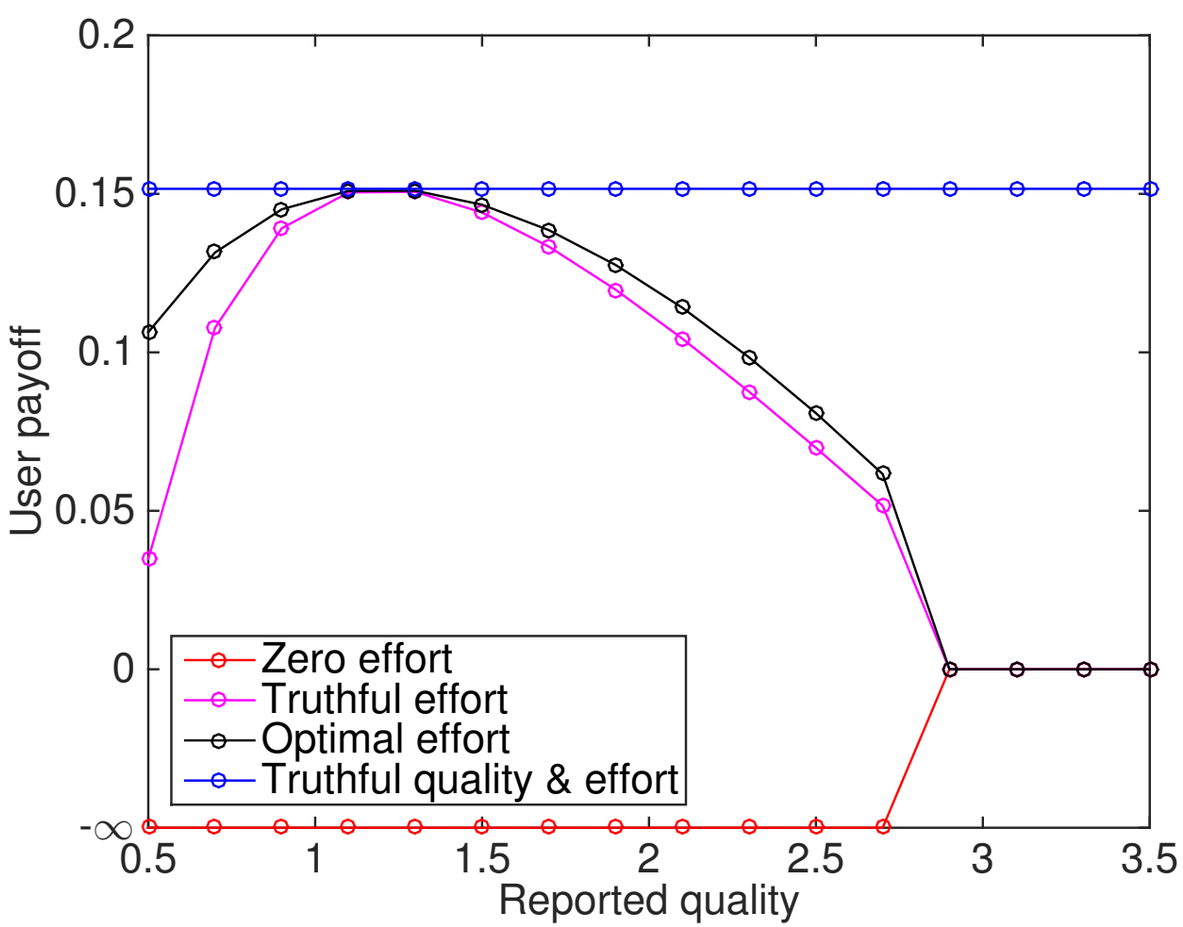
Proposition: The requester's optimal payoff $E_{\boldsymbol{Q}}[u_0(\boldsymbol{e}^*(\boldsymbol{Q}))]$, the socially optimal social welfare $E_{\boldsymbol{Q}}[v(e_1^{so}(\boldsymbol{Q}))]$, and the social welfare $E_{\boldsymbol{Q}}[v(e_1^*(\boldsymbol{Q}))]$ attained by the requester's optimal effort assignment increase when the number of users N increase, or the cost coefficient c decreases.

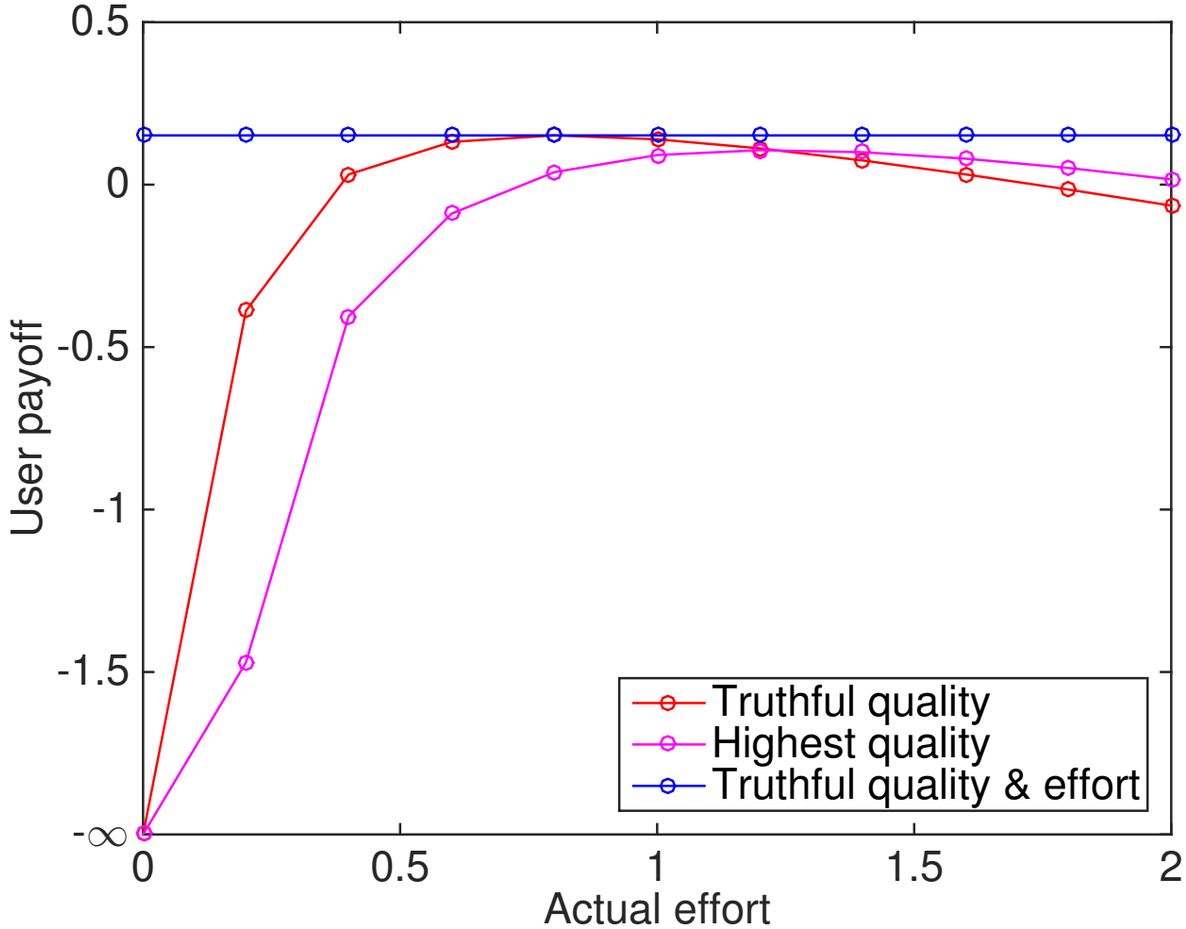
- ✓ The requester's payoff and social welfare both benefit from a greater diversity gain
  - When there are more users, the quality of the best user probably improves

Proposition: The gap between the social welfare attained by the requester's optimal effort assignment and the social optimal effort assignment $E_{\boldsymbol{Q}}[v(e_1^{so}(\boldsymbol{Q}))] - E_{\boldsymbol{Q}}[v(e_1^*(\boldsymbol{Q}))]$ decreases when the number of users N increase, and converges to 0 as N goes to infinity.

- ✓ The performance gap decreases to 0 asymptotically as the number of users increases

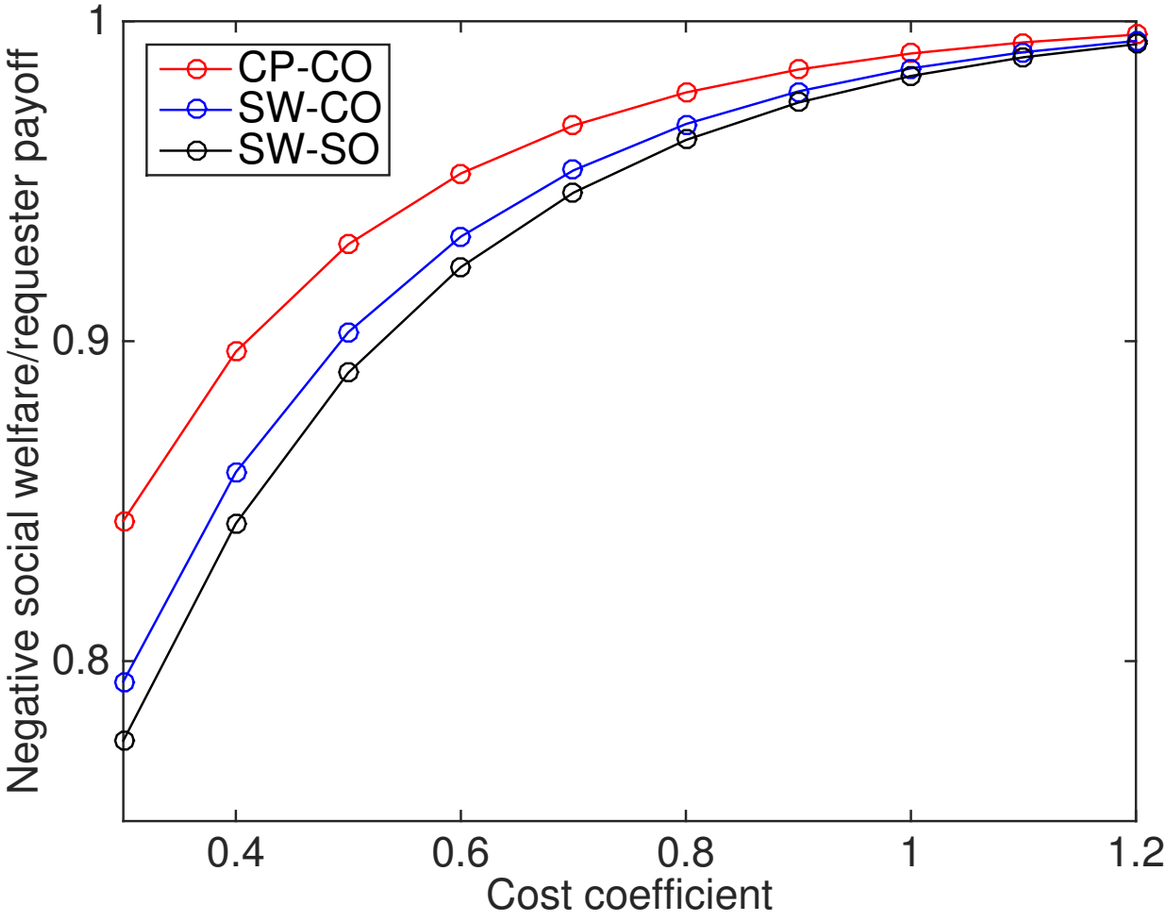
Impact of Untruthful Behavior

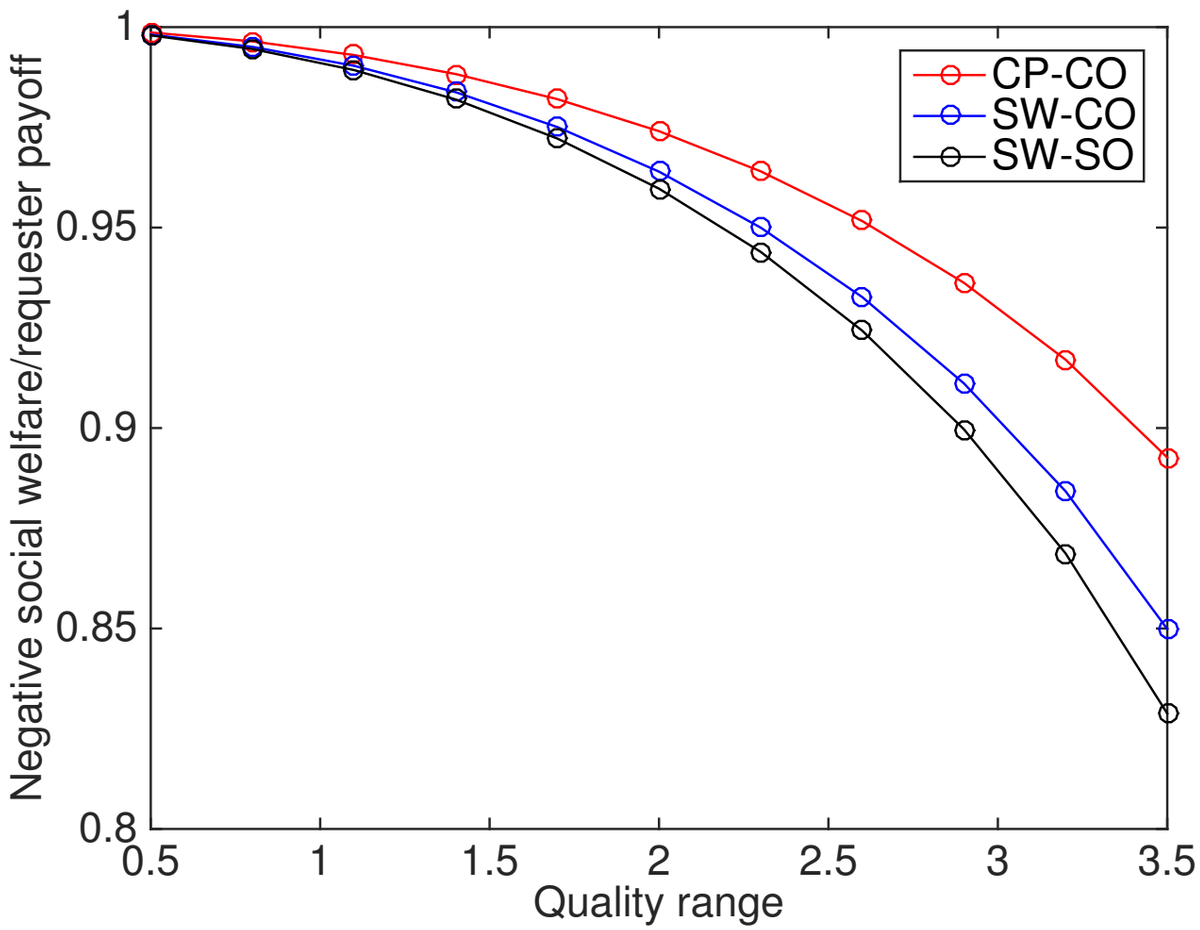




- ✓ Each user's payoff is maximized by truthfully reporting its quality and making the desired effort

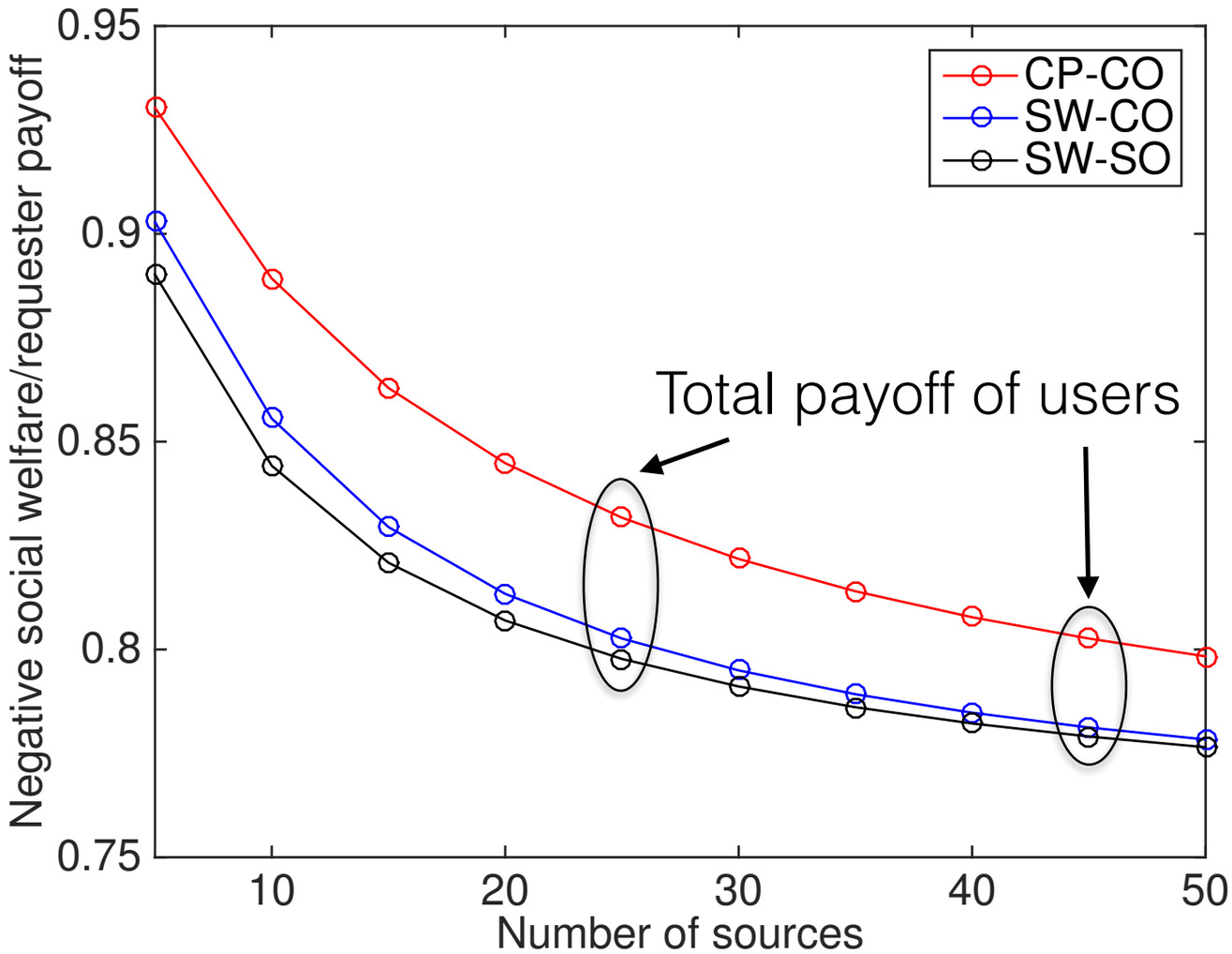
Impact on Requester's Payoff





- ✓ The gap between the requester's optimal (CO) and socially optimal (SO) effort assignments decreases as the quality range decreases
  - When less uncertain about users' qualities, assign more effort which is closer to the SO effort assignment

Impact on Requester/Users' Payoffs



- ✓ Total users' payoff can decrease when the number of users increases
  - Due to "competition" among users

Related Work

- Truthful crowdsensing with private cost
  - ✓ Many mechanisms to incentivize users to truthfully reveal their costs in crowdsensing
  - ✓ Existing designs for cost elicitation (e.g., VCG auction, the characterization of truthful mechanisms) cannot work for quality elicitation
- Mechanism design for hidden actions
  - ✓ [Luo et al, NetEcon'15] proposed incentive mechanisms for joint elicitation of effort, data, and cost in statistical estimation
  - ✓ Joint elicitation of quality and effort involves intricate coupling
- Quality-based crowdsourcing
  - ✓ Incentive mechanisms for users with diverse qualities and private costs in crowdsensing [Jin et al, Mobihoc'15,'16]
  - ✓ Learning quality of users' data in crowdsourcing [Karger et al, Allerton'11]
  - ✓ Quality elicitation has not been studied

Conclusion

- Summary
  - ✓ Proposed a quality-aware crowdsensing system that assigns efforts to users based on the reported user qualities and performs data estimation based on the reported data qualities
  - ✓ Designed Quality and Effort Elicitation (QEE) mechanisms that incentivize users to truthfully report their qualities and make the desired efforts
  - ✓ Characterized the optimal effort assignments for the requester's payoff and social welfare under the QEE mechanisms; Analyzed the impacts of quality range, user number, and cost coefficient on the performance
- Highlights
  - ✓ The first study of quality elicitation for crowdsensing
  - ✓ Untruthful quality reporting and effort exertion lead to undesired effort assignment, non-optimal data estimation, and incorrect information of data accuracy!!
  - ✓ Overcome intricate coupling between quality elicitation and effort elicitation in truthful mechanism design
- Future work
  - ✓ Many interesting problems in quality-based crowdsourcing…
  - ✓ Joint elicitation of quality, effort, and data
  - ✓ Learning quality/cost when it is unknown to the user/requester

Thank you!

Questions?
Please send questions to
xgong@auburn.edu