Peer-Prediction in the Presence of Outcome Dependent Lying Incentives

https://goelnaman.github.io naman.goel@alumni.epfl.ch

Naman Goel

Aris Filos-Ratsikas

Boi Faltings

IJCAI-PRICAI
YOKOHAMA
2020

MOTIVATION

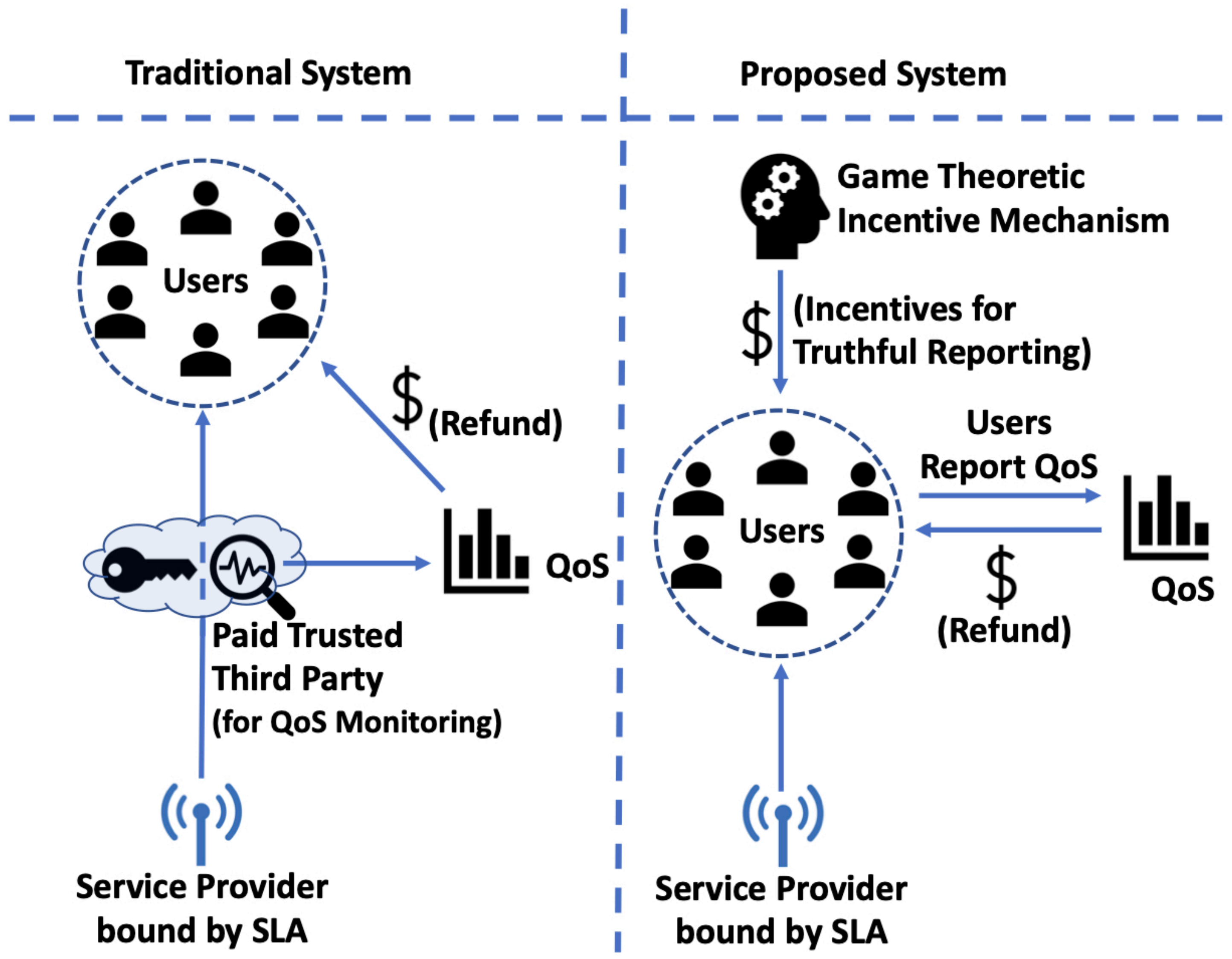
❑ In many real world scenarios, information collected from agents is used to make a decision or to determine some kind of outcome.

❑ Agents may have external incentives (as shown in the example below) to manipulate the outcome by misreporting the information.

Service Level Agreement (SLA)

❖ web services, Amazon AWS

e.g., the *response time* of the service will be less than 2 seconds.



Traditional System
Proposed System
Users
$ (Refund)
QoS
Paid Trusted Third Party (for QoS Monitoring)
Service Provider bound by SLA
Game Theoretic Incentive Mechanism
$ (Incentives for Truthful Reporting)
Users Report QoS
Users
$ (Refund)
QoS
Service Provider bound by SLA

❑ The proposed system can be implemented as a smart contract as shown by Goel et al. in *Infochain: A decentralized, trustless and transparent oracle on blockchain (IJCAI 2020)*.

RESEARCH QUESTIONS

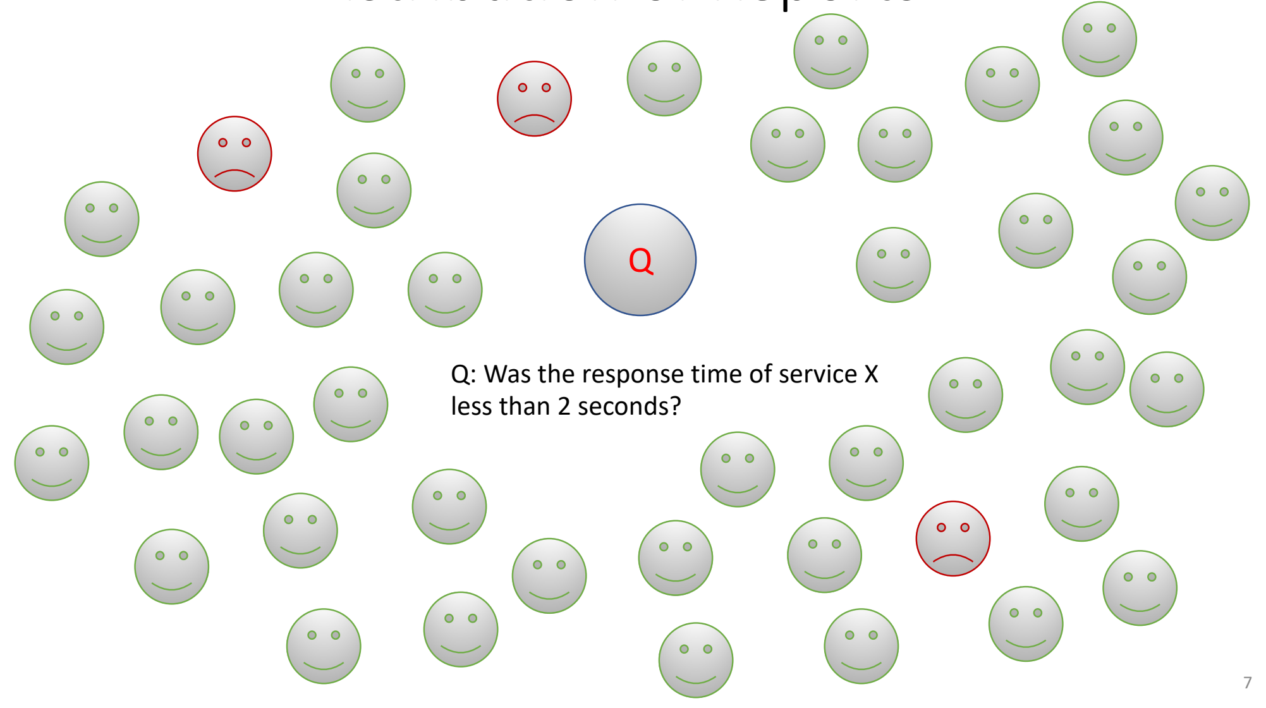
❑ Peer-prediction is a well known method to elicit effort and truthful information from rational agents.

❑ But what happens when the agents have outcome dependent lying incentives? Does this method still work?

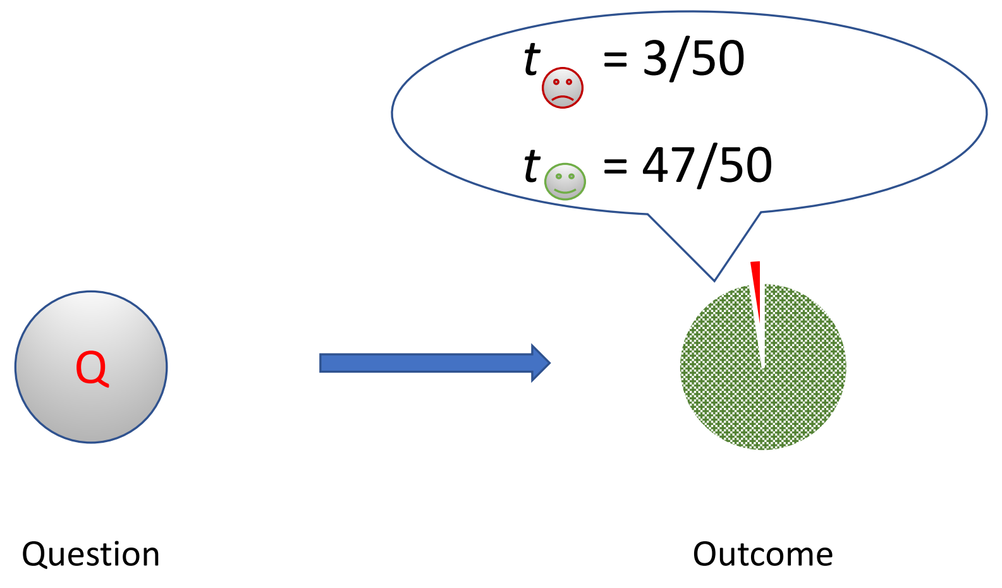
❑ How large do the incentives have to be, to counteract the lying incentives, and is the approach economically feasible?

MODEL

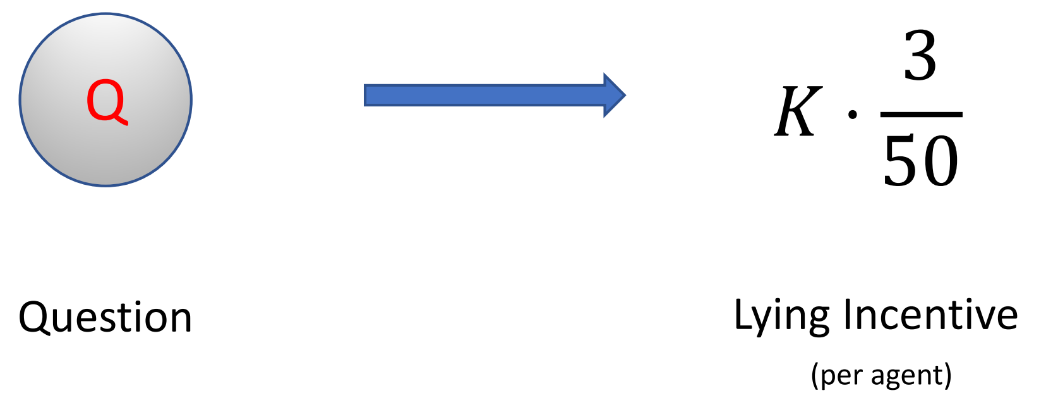
Distribution of Reports



Outcome Determination



Lying Incentives $= K \cdot t_{☹}$



Generous Refund Model:

Refund for everyone (irrespective of the report)

Conservative Refund Model:

Refund for only those who report ☹

> ➢ In both models, reporting ☹ is obviously the dominant strategy.
>
> ➢ We show that it is possible to get truthful information from agents in a profitable way, even in such challenging settings.

The Peer Truth Serum for Crowdsourcing

(Radanovic, Faltings and Jurca, 2016)

answer submitted by agent = y

answer submitted by another agent (peer) for the same question = y'

Payment Rule:

pay $\frac{1-p}{p}$ if $y = y'$ **charge 1 otherwise.**

where p is the relative frequency of y in the answers collected for statistically similar questions.

RESULTS

Making truth-telling an equilibrium

Theorem : Given δ and a scaling constant $\alpha > \frac{K}{n \cdot \delta}$, the truth-telling strategy profile is a strict equilibrium if $\beta \leq 0$, and is a $(\frac{\beta \cdot K}{n \cdot \delta})$-approximate equilibrium if $\beta > 0$.

where, δ is an approximation of δ^*, such that $\delta = \delta^* + \beta$

δ^* is the self-predictor value: a measure of correlation strength between the observations of agents.

Theorem : The expected relative saving in payments made in the truth-telling equilibrium is at least $Pr(☺) - \frac{1}{n \cdot \delta}$, where $Pr(☺)$ is the actual probability of a *random observation being* ☺.

➢ Relative saving is always positive if $n > \frac{1}{Pr(☺) \cdot \delta}$

➢ Approaches the optimal relative saving of $Pr(☺)$ as $n \to \infty$.

Eliminating denial strategy equilibrium

denial strategy = always reporting ☹ regardless of the true observation.

Theorem : Given that for any $f > 0$,

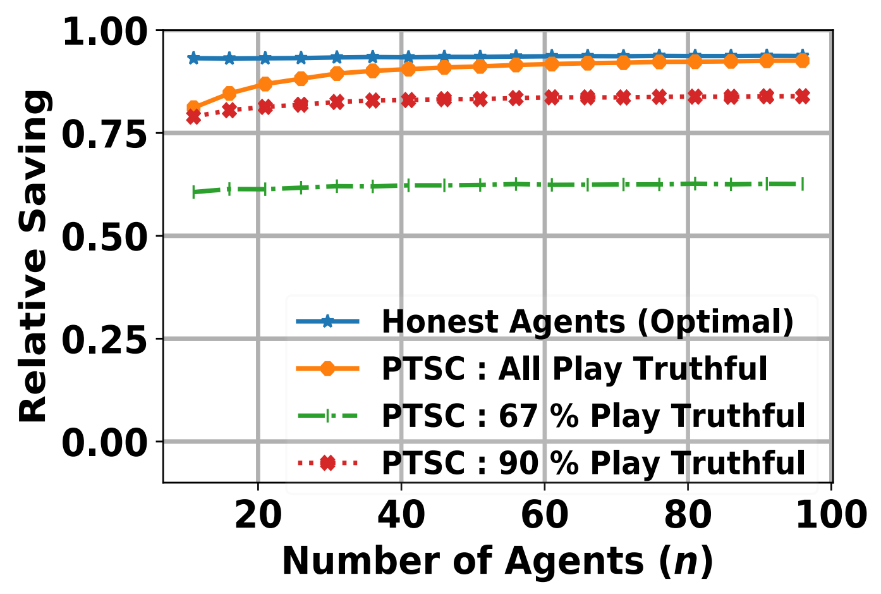
a) an f-fraction of agents are honest,

b) the remaining $1 - f$ adopt the denial strategy, and

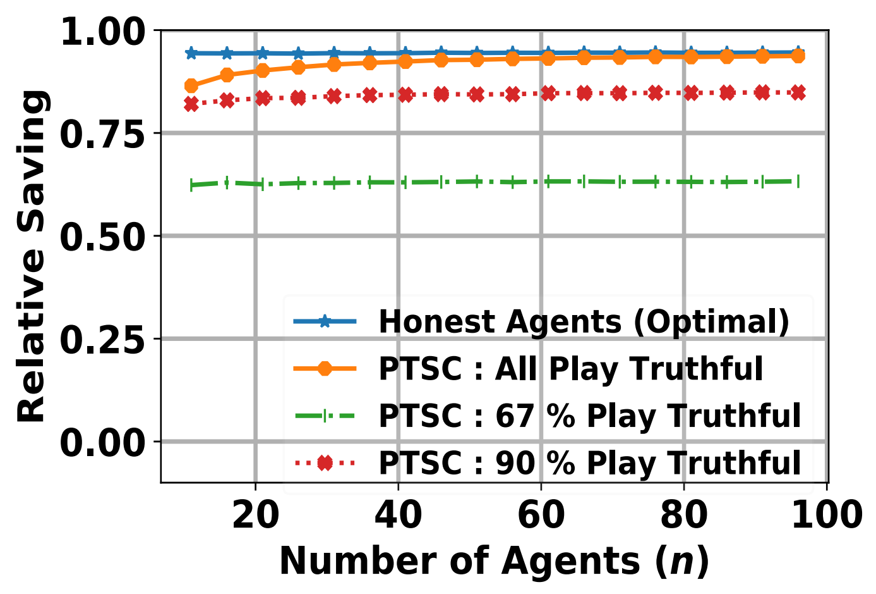
c) it holds that $\alpha > \frac{K}{n \cdot \delta_c}$

then the truth-telling strategy is strictly best response if $\beta_c \leq 0$, and is $(\frac{\beta_c \cdot K}{n \cdot \delta_c})$-approximate best response if $\beta > 0$.

Numerical Experiments



Response Time Data



Throughput Data