# Designing Data Markets for Competitive Industries: Structure, Stability and Fairness

Tyler Westenbroek
with Roy Dong, Lillian Ratliff and
Shankar Sastry

Motivating Problems

* The value of a particular data source to a firm is difficult to measure
  * Data-driven firms use different methods to process data and are looking to estimate different things
  * A firm has no way to ascertain the 'quality' of data before purchasing it
* Data sources have incentive to misrepresent the 'quality' of the data they are providing
  * Buying data from an new source is a potentially risky transaction

$
Data
Buyer

FORCES
FOUNDATIONS OF RESILIENT
CYBER-PHYSICAL SYSTEMS

3/7/2017
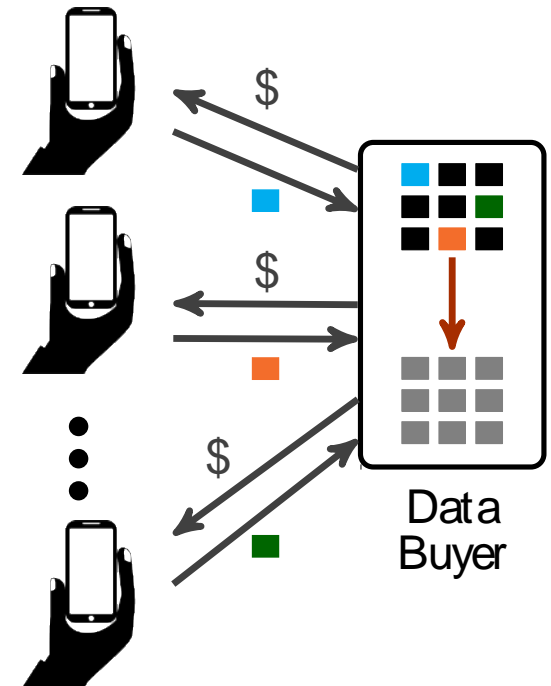
Previous Work in Literature

* Previous work primarily focuses on a single data buyer
* The buyer is interested in estimating some phenomena $f(x)$
* The buyer uses the readings from the sensors to construct the estimate $\hat{f}(x)$
    * Ex: Flow of traffic on a road
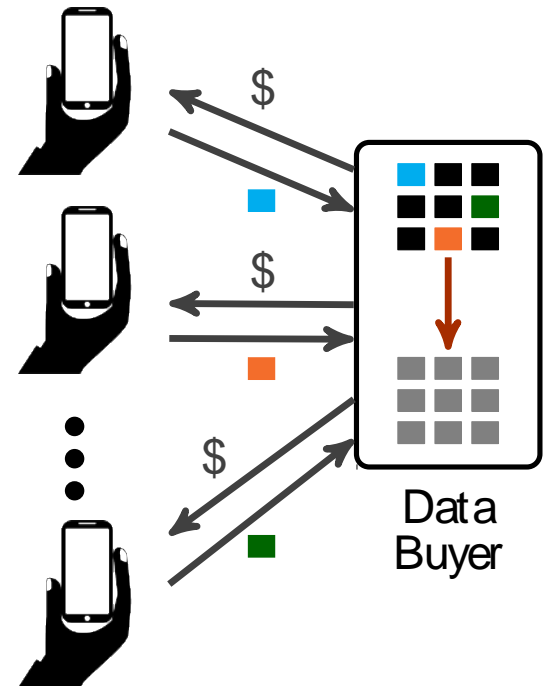
Strategic Sensing Framework

* Framework proposed in [1]:

  * Data source i exerts effort e_i to produce an estimate of $f(x)$

    $y_i(e_i) = f(x) + \epsilon_i$, where ϵ_i is distributed with zero mean and variance $\sigma_i(e_i)^2$

  * The buyer agrees to pay each data source i according to some pricing scheme, $p_i(x, \vec{y_i})$

  * Data source i chooses to exert effort

    $$e_i^* = \arg max_{e_i \in \mathcal{E}_i} \mathbb{E}[p_i(x, y(e_i))] - e_i$$

  * The buyer wants to pick the pricing scheme which minimizes

    $$\mathbb{E}_{x^*, \vec{y}(\vec{e}^*)} \left[\left(\hat{f}_{(\vec{x}, \vec{y}(\vec{e}))}(x^*) - f(x^*) \right)^2 + \eta \sum_{i \in W'} p_i \left(\left(x_j, y_j(e_j^*) \right)_{j \in W'} \right) \right]$$

[1] Cai, Daskalakis, Papadimitriou (2015)

FORCES
FOUNDATIONS OF RESILIENT
CYBER-PHYSICAL SYSTEMS

3/7/2017

Strategic Sensing Framework (Cont.)

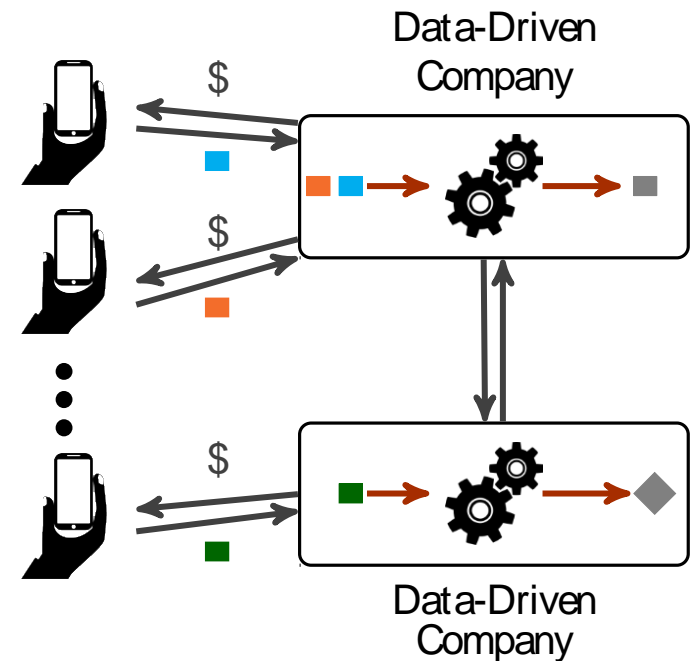* In particular, [1] utilizes pricing schemes of the form

$$p_i((\vec{x}, \vec{y})) = c_i - d_i \left(y_i - \hat{f}_{(\vec{x}, \vec{y})-i}(x_i) \right)^2$$

* Demonstrates a method for choosing c_i and d_i that allows the buyer to
  * Precisely set the level of effort each of the sensors exerts and
  * Exactly compensate each sensor such that

$$\mathbb{E}_{x^*, \vec{y}(\vec{e}^*)} \left[p_i((\vec{x}, \vec{y})) \right] = e_i$$

FORCES
FOUNDATIONS OF RESILIENT
CYBER-PHYSICAL SYSTEMS

[1] Cai, Daskalakis, Papadimitriou (2015)

3/7/2017

Adding Competition Between Firms

* Multiple firms buy data from the same pool of data sources

* Firms want to maximize the quality of data they receive and minimize the quality of data their competitors receive

* The value of data ultimately arises from the competition between the firms

Data-Driven Company

Data-Driven Company

FORCES
FOUNDATIONS OF RESILIENT
CYBER-PHYSICAL SYSTEMS

3/7/2017

Two Firm Example

* Firm j tries to minimize:

$$\mathbb{E}\left[\left(\hat{f}_{(\vec{x},\overrightarrow{y_j}(\vec{e}))}(x^*) - f(x^*)\right)^2 - \delta_j\left(\hat{f}_{(\vec{x},y_{-j}(\vec{e}))}(x^*) - f(x^*)\right)^2 + \eta\sum_{i\in W'} p_i^j\left((x_j, y_j(e_j^*))_{j\in W'}\right)\right]$$

* Where the sensors collect data as before and report reading y_j to firm j

* Each firm j commits to paying each data source i the payment

$$p_i^j((\vec{x}, \vec{y})) = c_i^j - d_i^j\left(y_i^j - \hat{f}_{(\overrightarrow{x,y_j})-i}(x_i)\right)^2$$

* In total, each data source receives payment

$$p_i((\vec{x}, \vec{y})) = p_i^1((\vec{x}, \vec{y})) + p_i^2((\vec{x}, \vec{y}))$$

FORCES
FOUNDATIONS OF RESILIENT
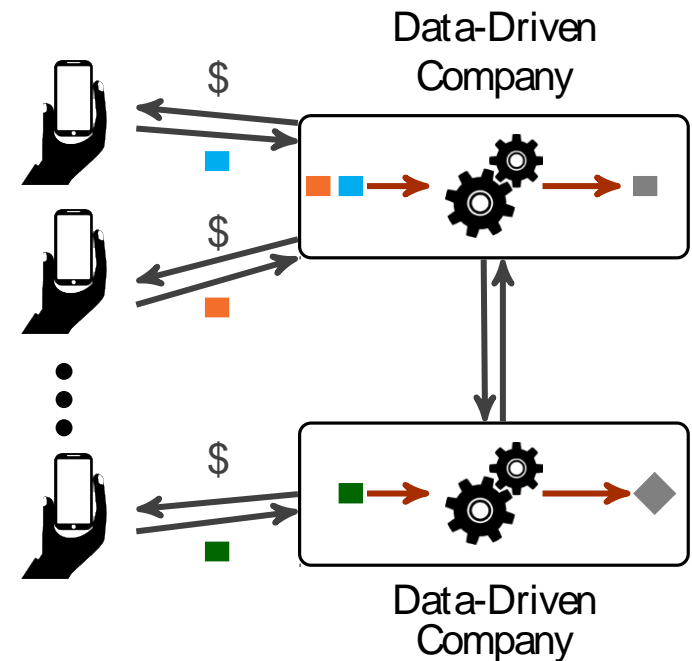CYBER-PHYSICAL SYSTEMS

3/7/2017

Two Firm Example (Cont.)

* Data source i exerts effort

$$e_i^* = \arg max_{e_i \in \mathcal{E}_i} \mathbb{E}\left[p_i^1(x, y_1(e_i)) + p_i^2(x, y_2(e_i))\right] - e_i$$

* For a given level of effort e_i, it is optimal for data source i to report the same reading to both firms

    * When a firm incentivizes a data source to exert more effort the firm's competitor benefits

* We look at two cases:

    * Symmetric Firms

    * Asymmetric Firms

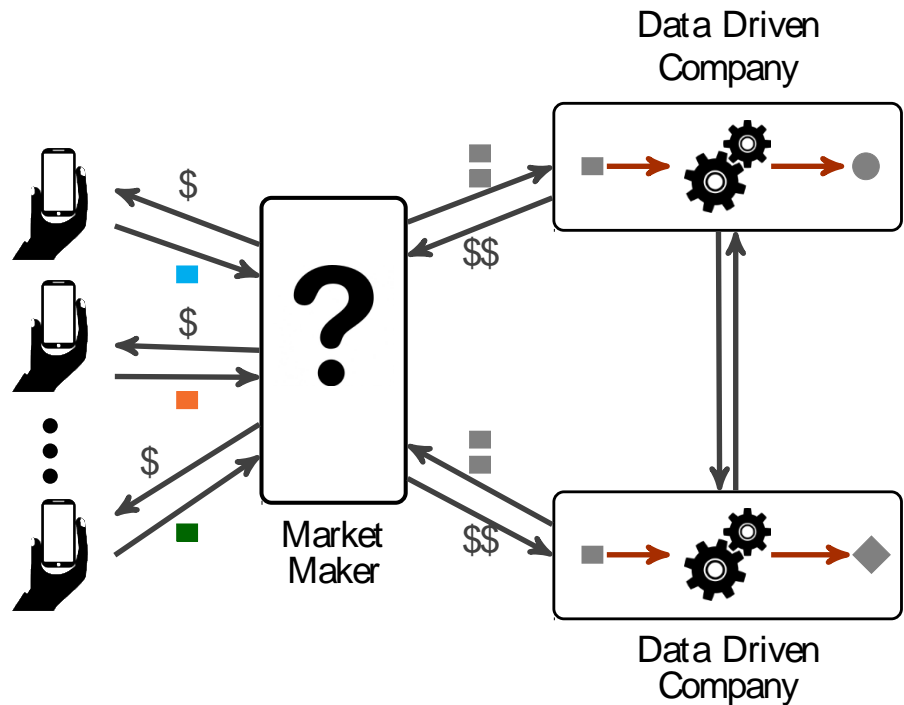FORCES
FOUNDATIONS OF RESILIENT
CYBER-PHYSICAL SYSTEMS

New Challenges Introduced by Competition

* If both firms use pricing schemes as in [1], the market unravels

  * Potential for free riding

  * Cannot guarantee individual rationality is maintained for all agents

  * Firms have incentive to make data sources sign exclusivity contracts

Data-Driven Company

Data-Driven Company

[1] Cai, Daskalakis, Papadimitriou (2015)

FORCES
FOUNDATIONS OF RESILIENT
CYBER-PHYSICAL SYSTEMS

3/7/2017

Future Work

* Devise richer (possibly nonlinear) pricing contracts between data sources and data buyers
  * Decouple the incentives to give each firm high quality estimates
* Add a third party to mediate these contracts

FORCES
FOUNDATIONS OF RESILIENT
CYBER-PHYSICAL SYSTEMS

Thanks!