A Generic Framework for *Spatial Crowdsourcing*

QInF 2015 Presentation

**Students:**

Hien To  
Email: [hto@usc.edu](mailto:hto@usc.edu)  
Website: [http://www-scf.usc.edu/~hto](http://www-scf.usc.edu/~hto)

Giorgos Constantinou  
Email: [gconstan@usc.edu](mailto:gconstan@usc.edu)  
Website: [http://www-scf.usc.edu/~gconstan](http://www-scf.usc.edu/~gconstan)

**Advisor:**

Prof. Cyrus Shahabi  
Email: [shahabi@usc.edu](mailto:shahabi@usc.edu)  
Website: [http://infolab.usc.edu/Shahabi](http://infolab.usc.edu/Shahabi)
Crowdsourcing: outsourcing a set of tasks to a set of workers.  

Spatial crowdsourcing (SC): requires workers to **physically** travel at the task's location in order to execute the task.

**Ubiquity of mobile users**
- 6.5 billion mobile subscriptions, 93.5% of the world population \[1\]

**Technology advances on mobiles**
- Smartphone's sensors, e.g., video cameras

**Network bandwidth improvements**
- From 2.5G (up to 384Kbps) to 3G (up to 14.7Mbps) and recently **4G** (up to 100 Mbps)

**SC Applications**
- Uber
- Trapster.com
- FAVOR
- taskrabbit
- OnStar
- WeatherSignal
- WeGoLook
- Gigwalk
- OpenSignal
- TomTom
- FieldAgent

---

\[1\] http://www.gartner.com/newsroom/id/2665715
MediaQ helped PBS cover the Presidential Inauguration on Jan. 20, 2013
Maximum task assignment is reducible to max-flow problem

One time snapshot

MaxT$_1$ = 2

MaxT$_2$ = 3

MaxT$_3$ = 4

[Kazemi et.al ACM GIS’12]
Challenges of Task-assignment

• **Dynamism**
  - Tasks/workers arrive **without** our knowledge

• **Location Privacy**
  - Adversary can infer workers’ **sensitive details**

• **Trust**
  - Workers **cannot** always be trusted, i.e., malicious/spam users
**Dynamic Task Assignment**

- **Local task assignment** at every time snapshot

- **Learn worker distribution**
  - Prefer tasks in **worker-sparse areas**
  - Prefer **nearby** workers

  [To et.al TSAS’15]
  [Kazemi et.al ACM GIS’12]

- **Learn activity patterns**
  - Defer tasks arriving in **uphill** periods

  Submitted to [VLDB’15]

  From [Musthag et.al CHI’13]
Worker Location Privacy

**Objectives**
- Assign tasks to workers **without** knowing workers’ locations
- Ensure tasks will be performed, with **high probability**

**Solutions**
- Assign every task to all workers
  - **Does not scale!**
- Assign tasks to **sufficient nearby** workers

**Differential Privacy**
- Preserve privacy of individual workers

<table>
<thead>
<tr>
<th>Noisy worker count per grid cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>~0</td>
</tr>
<tr>
<td>~4</td>
</tr>
<tr>
<td>~4</td>
</tr>
<tr>
<td>~0</td>
</tr>
</tbody>
</table>

[To et.al VLDB’14]
[To et.al ICDE’15]

**The first** privacy study in spatial crowdsourcing!
Trustfulness of Workers

• Non-spatial metrics
  ▪ Rating
  ▪ # of transactions
  ▪ Fast response time
  ▪ Quality of work

• Spatial metrics
  ▪ Distance traveled
  ▪ Spatial coverage

• Use reputation-based trust
  ▪ To maximize the **quality** of the result in task assignment
  ▪ To direct requesters to “trusted” areas in task posting
Proposed Framework & Execution Plan

- Develop a **generic spatial crowdsourcing architecture** to ease the development of SC applications.

**Applications**
- MediaQ: Crowdsourcing Multimedia
- iRain: Crowdsourcing Weather Information
- Others: Uber, Taskrabbit, Waze, Google Shopping Cart

**Application Services**
- A5 B5 Geocrowd APIs
- A4 gAnalytics
- B4 gVisualizer

**Geocrowd Engine**
- A3 Generic Task
- B3 Assignments
- B1
- Account Management
- Task Management
- Query Processing
- Notification Management

**Resources**
- A1
- Databases
- Repository

**Phases**

<table>
<thead>
<tr>
<th>Phases</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
</tr>
<tr>
<td>A1B1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2B2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A4B4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A5B5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A6B6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Evaluation

• Use **real** data workload and generate **synthetic data**
  ▪ e.g. Gowalla, Yelp, Foursquare

• Build **real systems**, e.g., MediaQ, iRain
  ▪ Organize **events** at USC to collect SC data
  ▪ Deploy on **cloud** and monitor workload and system behavior
  ▪ Get feedback via **academic collaboration**

• Publish framework as an **open source project**
  ▪ e.g. github
Team

- **Prof. Cyrus Shahabi (advisor)**
  - Pioneered spatial crowdsourcing
  - Director of InfoLab; an active research group in spatial crowdsourcing

- **Hien To (4th year PhD student)**
  - Research on scalability/privacy of task assignment
  - Published papers in prestigious database conferences, e.g., ICDE, VLDB, CIKM
  - Completed two summer interns at Teradata in 2012 and 2014

- **George Constantinou (1st year PhD student)**
  - His research focuses on trustfulness in spatial crowdsourcing
  - Completed a summer intern at Amazon in 2014
  - Fulbright fellow

- **Partnership/Research Accomplishments**
  - SC for collecting media content: MediaQ (since 2012)
  - SC for collecting weather information: iRain (since 2014)
  - Experience in developing mobile platforms

- The Wiley-AAG International Encyclopedia of Geography