

ACM SIGSPATIAL PhD 2016

Task Assignment in Spatial Crowdsourcing: Challenges and Approaches

Hien To

Supervised by Prof. Cyrus Shahabi

Department of Computer Science
University of Southern California

Outline

- **Introduction**
- A taxonomy and research challenges
- Privacy-preserving spatial crowdsourcing
 - Risks of location leaks
 - Privacy threats during tasking and reporting
 - TaskRabbit: a case study
 - Privacy countermeasures: the state-of-the-art
- Conclusion and future works

Spatial Crowdsourcing (SC)

Crowdsourcing: outsourcing a set of tasks to a set of workers.  **mechanicalturk™**
Artificial Artificial Intelligence

Spatial crowdsourcing (SC): requires workers to *physically* travel to 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)

[1] <http://www.gartner.com/newsroom/id/2665715>

Spatial Crowdsourcing Apps (a.k.a Mobile Crowdsourcing)

There's an app that needs you: the rise of mobile crowdsourcing



The Importance of Mobile Crowdsourcing in 2015



→ The Wiley-AAG International Encyclopedia of Geography



Mobile Crowdsourcing:
The Next Chapter, 2016

THE
HUFFINGTON
POST

Microsoft Research

[Our research](#) [Connections](#) [Careers](#) [About us](#)

All Downloads Events Groups News People [Projects](#) Publications Videos

Microsoft Translator | Choose language

Search Microsoft Research

Spatial Crowdsourcing

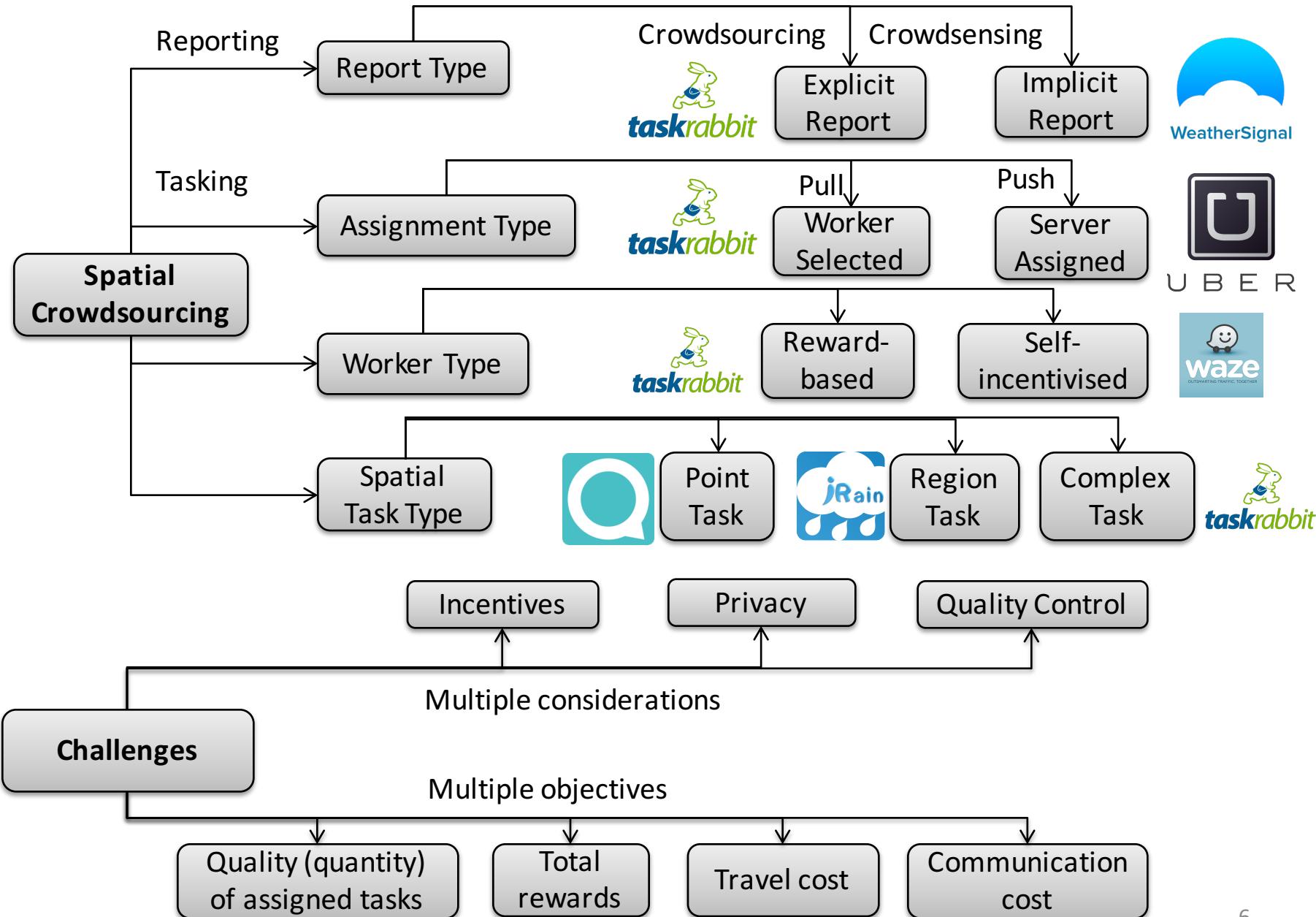


We are studying how we can get regular people to do simple tasks at specific locations. An example task is to take a picture of a sign at a certain location. We are interested in who to ask and how much to pay.

Outlines

- Introduction
- **A taxonomy and research challenges**
- Privacy-preserving spatial crowdsourcing
 - Risks of location leaks
 - Privacy threats during tasking and reporting
 - TaskRabbit: a case study
 - Privacy countermeasures: the state-of-the-art
- Conclusion and future works

Spatial Crowdsourcing Taxonomy



Research Challenges in Spatial Crowdsourcing

Task assignment (or worker selection)

Process of identifying which tasks should be assigned to which workers

- **Asghari et al. SIGSPATIAL 2016**
- Bessai and Charoy ISCRAM '16
- Hassan and Curry ESA'16
- Zhang et al. TVT '16
- Gao et al. WAIM '16
- Cheng et al. TKDE '16
- Tong et al. VLDB '16
- Liu et al. DASFAA '16
- Hu et al. ICDE '16
- Tong et al. ICDE'16
- Zhang et al. WCMC '16
- Liu et al. UbiComp '16
- Guo et al. THMS '16
- **To et al. PerCom '16**
- **To et al. TSAS '15**
- **Alfarrarjeh et al. MDM '15**
- Fonteles et al. MoMM '15
- Hassan and Curry. SIGSPATIAL '15
- Xiao et al. INFOCOM '15
- Xiong et al. PerCom '15
- Pournajaf et al. ICCS '14
- Hassan and Curry. UCI'14
- He et al. INFOCOM '14
- Fonteles et al. SIGSPATIAL '14
- Zhang et al. UbiComp '14
- **Dang et al. iiWAS '13**
- **Kazemi and Shahabi. SIGSPATIAL '12**

Research Challenges in Spatial Crowdsourcing

Privacy-preserving SC

Protect locations of workers and tasks

- **To et al. TMC '16**
- Boutsis and Kalogeraki. UbiComp'16
- Zhang et al. CN '16
- Zhang et al. ATIS '15
- Shen et al. GLOBECOM '15
- Gong et al. IoT'15
- Gong et al. TETC'15
- Hu et al. APWeb '15
- Pournajaf et al. MDM'14,
SIGSPATIAL'15
- **To et al. VLDB '14, ICDE '15**
- Boutsis and Kalogeraki PerCom '13
- Vu et al. INFOCOM '12
- **Kazemi and Shahabi SIGKDD '11**

Task scheduling

Workers' path planning to perform tasks

- Wang et al. 2016
- Fonteles et al. JLBS '16
- **Deng et al. GeoInformatica '16**
- Mrazovic et al. ICDMW '15
- Chen et al. IJCAI '15
- Chen et al. AAMAS '15
- Hadano et al. HCOMP '15
- **Deng et al. SIGSPATIAL '15**
- Chen et al. HCOMP '14
- **Deng and Shahabi. SIGSPATIAL '13**
- Sadilek et al. AAAI'2013

Research Challenges in Spatial Crowdsourcing

Trust and quality

Consider quality of report data or trustworthiness of workers

- Liu et al. Sensor '16
- Zhang et al. TETC'16
- Miao et al. DSS '16
- Fan et al. SOSE '15
- An et al. HPCC '15
- Kang et al. MASS '15
- Cheng et al. VLDB '15
- Zhao et al. MDM '15
- Wang et al. UbiComp '15
- Song et al. TVT '14
- Boutsis et al. ICDCS '14
- Feng et al. INFOCOM '14
- **Kazemi et al. SIGSPATIAL '13**

Incentive mechanism

Incentivize workers to perform tasks

- Zhang et al. TTV '16
- Kandappu et al. CSCW '16
- Kandappu et al. UbiComp '16
- Micholia et al. IJHCS '16
- **To et al. GeoRich '16**
- Li and Cao TMC '16
- Guo et al. IJMHCI '16
- Thebault-Spieker et al. CSCW '15
- Jin et al. MobiHoc '15
- Fan et al. ICTAI '15
- Shah-Mansouri et al. ICC '15
- Teodoro et al. CSCW '14
- Rula et al. HotMobile '14
- Musttag et al. CHI '13
- Heimerl et al. CHI '12
- Jainmes et al. PerCom '12
- Yang et al. MobiCom '12
- Lee and Hoh PMC '10
- Alt et al. NordiCHI '10

Research Challenges in Spatial Crowdsourcing

Generic frameworks

Discuss components, architecture, programming framework of SC apps

- To et al. CROWDBENCH '16
- Fonteles et al. RCIS '16
- Peng et al. ASE '16
- Kucherbaev et al. SIGCHI '16
- Sakamoto et al. COMPSAC '16
- Fernando et al. MOBIQUITOUS '13
- Tamilin et al. UbiComp '12
- Ra et al. MobiSys '12
- Yan et al. SenSys '09

Related surveys

- Pournajaf et al. SIGMOD Rec. '16
- Guo et al. Comp Survey '15
- Zhao and Han 2016
- Christin JSS '15

Applications

- Konomi and Sasao Urb-IoT '16
- Jaiman et al. UbiComp/ISWC '16
- Fan and Tseng MOBIS '15
- Konomi and Sasao UbiComp/ISWC '15
- Harburg et al. CHI '15
- Chen et al. SenSys '15
- Kim CHI '15
- Aubry et al. CROWDSENSING '14
- Chen et al. VLDB '14
- **Kim et al. MMSys'14**
- Benouaret et al. IEEE IC '13
- Coric and Gruteser DCOSS '13
- Koukoumidis et al. MobiSys '11
- Goodchild and Glennon IJDE '10

Outlines

- Introduction
- A taxonomy and research challenges
- Privacy-preserving spatial crowdsourcing
 - **Risks of location leaks**
 - Privacy threats during tasking and reporting
 - TaskRabbit: a case study
 - Privacy countermeasures: the state-of-the-art
- Conclusion and future works

Risks of Location Leaks

[Stalkers Exploit Cellphone GPS](#)

“25,000 adults in the U.S. are victims of GPS stalking annually, including by cellphone.” by U.S. Justice Department, Aug 2010.



[Apple, Google Collect User Data](#)

“~50% top iPhones and Android smartphones disclosed a user’s location to third parties without user consent.” by Wall Street Journal, Dec 2010.

[The Hidden Privacy Threat of ... Flashlight Apps?](#)



[Security Flaw In Gay Dating App Grindr Reveals Precise Location Of 90% Of Users](#)

[Could using gay dating app Grindr get you arrested in Egypt?](#)



[Grindr's Location Security Update](#)

“Any user who connects to Grindr in these countries will have their distance hidden automatically by default, which include Russia, Egypt, Saudi Arabia, Nigeria, Liberia, etc.”



PRIVACY ROAD KILL 4/26/16 2:40 PM

If you use Waze, hackers can stalk you

Privacy Implication of Location Release

“Understanding individual human mobility patterns” [Barabási et al., Nature’08]

- 100,000 mobile phone users, 6 months
- **User mobility → spatial probability distribution**

“Unique in the Crowd:” [Montjoye et al., Nature’13]

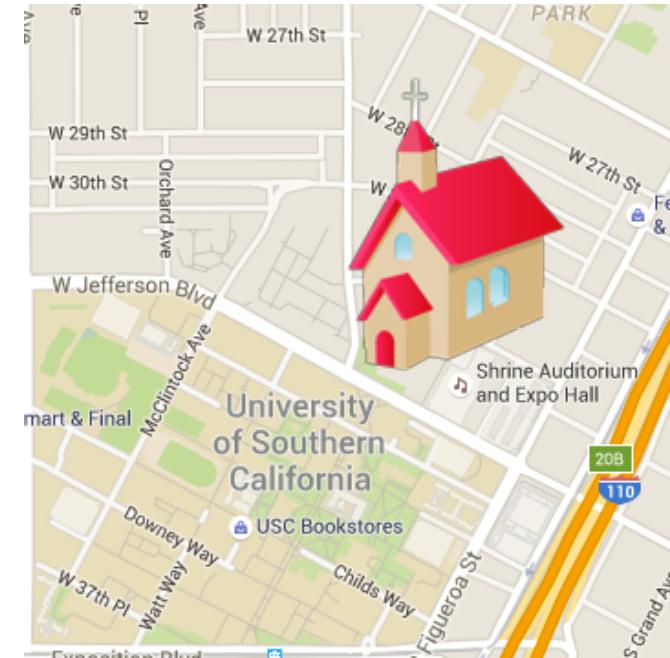
- 1.5 million individuals, 15 months
- **4 locations uniquely identify 95% individuals**

“*Inference Attacks* on location tracks” [John Krumm, PERVASIVE’07]

- GPS data from 172 people
- Find **home locations** with median error of **60 meters**

Privacy Implication of Location Release

- Location leaks sensitive information
 - Political and religious view
 - Individual's health status
 - Alternative lifestyles
 - Affiliations



- Attacks based on locations
 - Physical surveillance
 - Stalking
 - Identity theft



Outlines

- Introduction
- A taxonomy and research challenges
- Privacy-preserving spatial crowdsourcing
 - Risks of location leaks
 - **Privacy threats during tasking and reporting**
 - **TaskRabbit: a case study**
 - Privacy countermeasures: the state-of-the-art
- Conclusion and future works

TaskRabbit: A Case Study



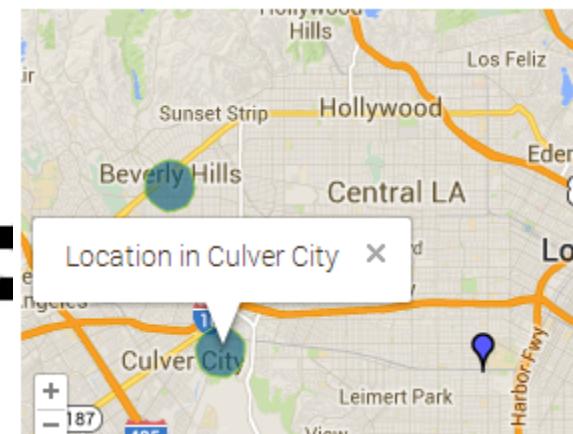
is an **online** and **mobile marketplace** that matches freelance labor with **local demand**, allowing consumers to find immediate help with everyday tasks, including **cleaning, moving, delivery** and **handyman work**

Leak of **task location** during tasking

A delivery task

Description:

Please pick up a box of mini-muffins from [REDACTED]
[REDACTED], promptly at 8am on Tues., 9/4, and drive them straight to me at
[REDACTED] ([REDACTED]).



Cloaked task locations

TaskRabbit: A Case Study

Leak of worker location tasking and reporting



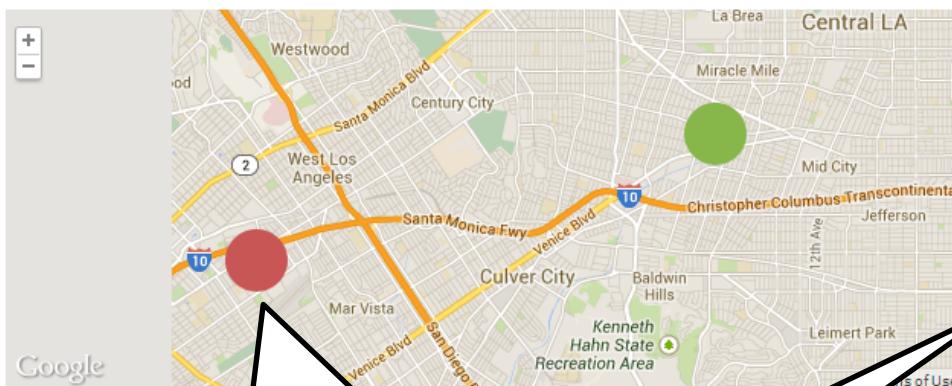
Super easy task! Deliver a bag to the doorstep of a sick friend

Wed. Dec 18, 3:59 PM for [REDACTED]

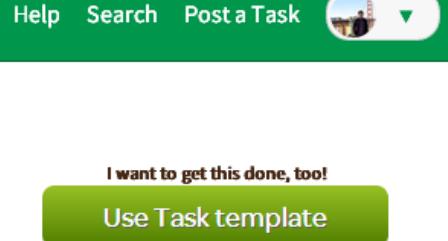
i Details

Description:

Hey guys! I need someone to do a super easy task in the spirit of the holidays! Pick up 2 bags (a holiday gift & groceries) from my house and leave them on the doorstep for a sick friend. I would realllllly appreciate it! It should take less than 30 minutes. Easy-peazy.



Cloaked task locations

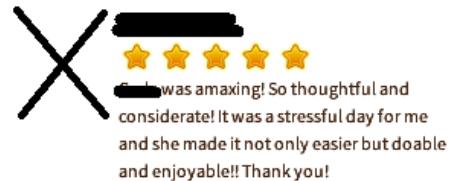


Task Price Estimate:

\$15 - \$21 Price range you can expect to pay for Tasks similar to this one.

7 TaskRabbits submitted offers on this Task

TaskRabbit for this Task



Task Status:

Posted	Dec 18, 1:06 PM	
Assigned	Dec 18, 1:25 PM	
Completed	Dec 18, 3:47 PM	
Closed	Dec 21, 2:49 PM	

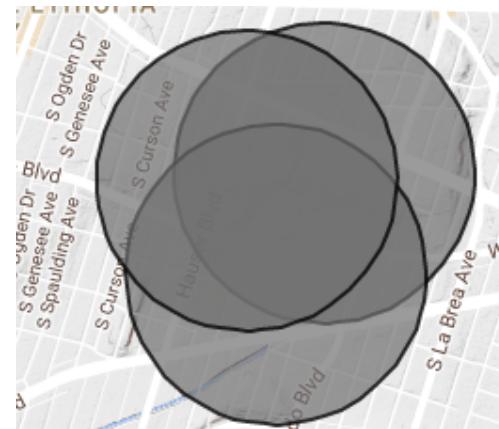
Task status

TaskRabbit: A Case Study

Learn a **requester's home** by tracking his requests

- Hypothesis: cloaked task regions cover actual task location
- Validate the hypothesis: a known location of another task is actually in the overlapped area

Task description	Corresponding JavaScript
Quick post-party dishwashing clean up needed	<pre>"radius": "0.5", "geo_center" : {"lat": "33.885951", "lng": "-118.094682"}</pre>
Take down light Christmas decorations	<pre>"radius": "0.5", "geo_center" : {"lat": "33.885951", "lng": "-118.094682"}</pre>
Put up 20 yard sale signs in Mid-Wilshire area	<pre>"radius": "0.5", "geo_center" : {"lat": "33.885951", "lng": "-118.094682"}</pre>



Triangulation attack

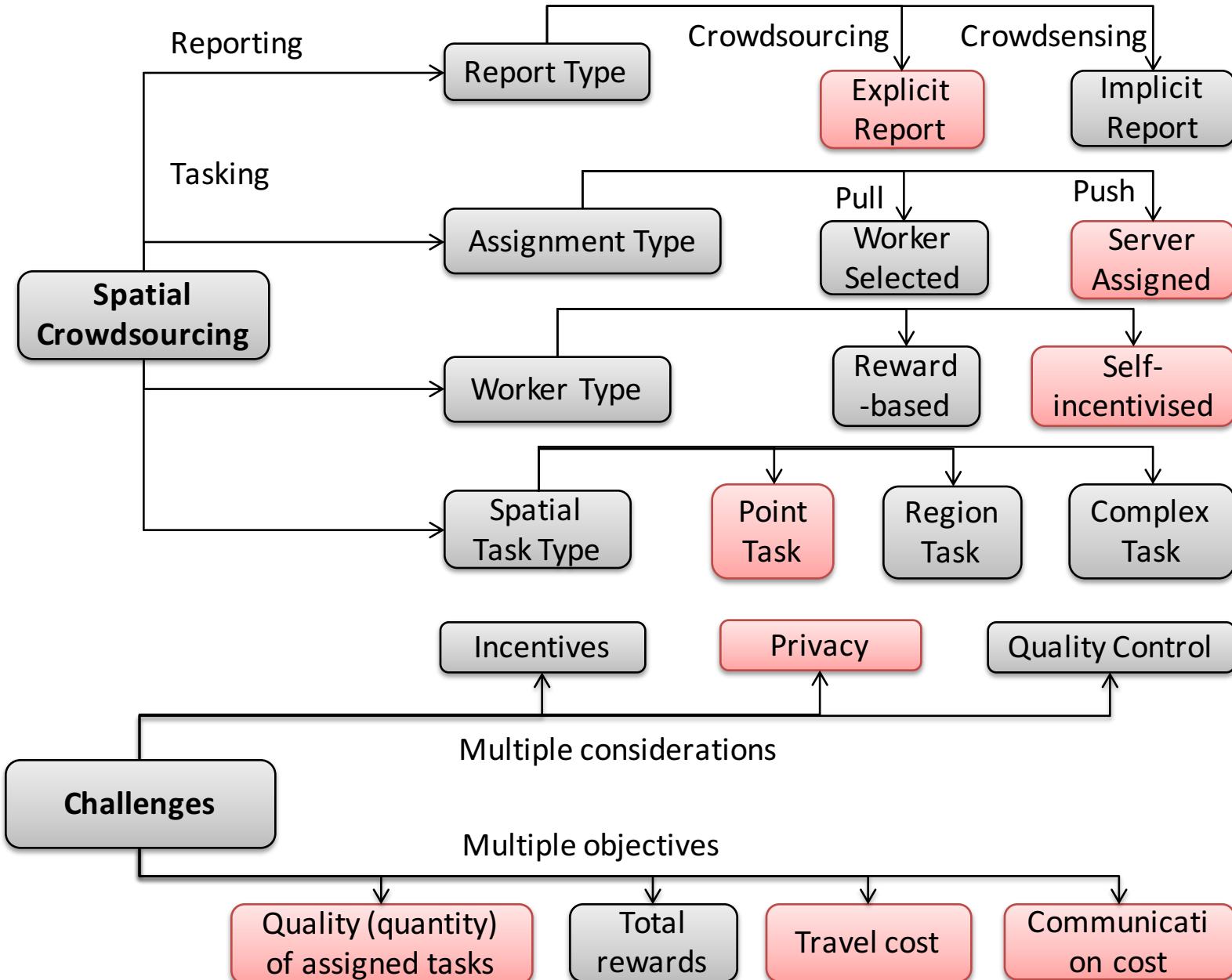
Outlines

- Introduction
- A taxonomy and research challenges
- Privacy-preserving spatial crowdsourcing
 - Risks of location leaks
 - Privacy threats during tasking and reporting
 - TaskRabbit: a case study
 - **Privacy countermeasures: the state-of-the-art**
- Conclusion and future works

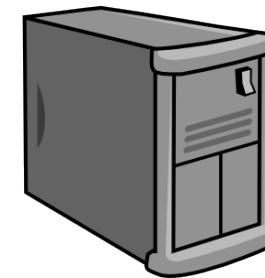
Current State-of-the-art

Category	Paper	Description
Tasking & Reporting	Shin PMC '11	Protecting workers' identity and <i>location (IP address)</i> during <i>task download</i> (use Tor) and <i>report submission</i> (use anonymity service)
Reporting	Boutsis PerCom '13 • Zhang CN '16	Workers <i>exchange</i> spatial task reports before uploading to SC-server so that the server cannot tell which workers upload which reports
Tasking	Kazemi SIGKDD Ex'11 • Vu INFOCOM '12	SC-server distributes spatial tasks to workers such that each worker is assigned a subset of tasks that are closer to himself than to any other worker
	To VLDB '14 • Gong TETC'15 • Zhang ATIS '15 • To TMC '16	SC-server assigns a spatial task to a set of workers <ul style="list-style-type: none"> use <i>differential privacy</i> to hide workers' location with low travel cost and low communication overhead without compromising assignment success rate
	Pournajaf MDM'14 • Hu APWeb '15	SC-server <i>matches</i> a set of spatial tasks to a set of workers given their <i>cloaked regions</i> <ul style="list-style-type: none"> maximize task assignment while satisfying travel budget constraint of each worker
	Shen GLOBECOM '15	SC-server assigns a spatial task to the nearest workers on the <i>encrypted domain</i>

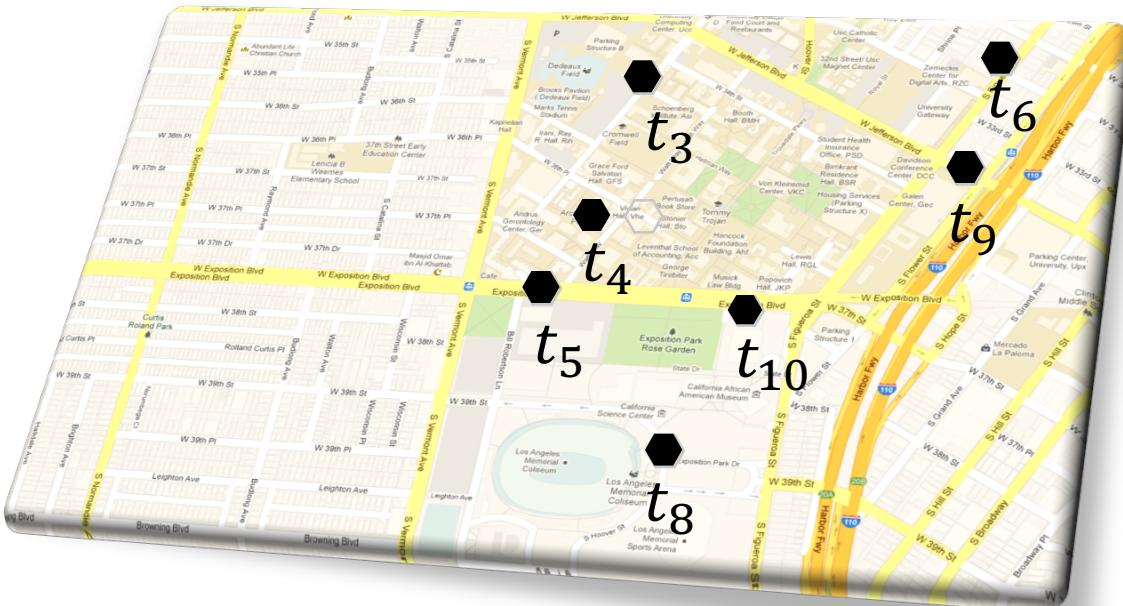
Spatial Crowdsourcing Taxonomy



SC-Server Performs Task Assignment [*]

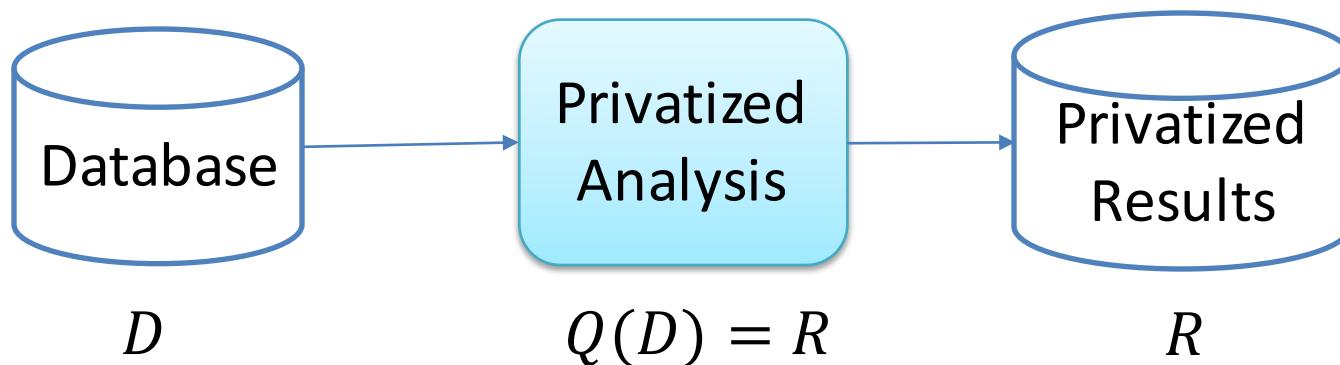


SC-server
(Crowdsourcing
Company)



- SC-server chooses best workers for task based on task-worker proximity
- **SC-server knows workers' locations**

Why Differential Privacy ^[*]?



Privacy guarantees

- ✓ Protection for **individuals**
- ✓ Protection **against background knowledge**
- ✓ Strong **theoretical guarantees** on privacy of released data

Differential privacy supports various aggregation queries, e.g., count, min, max, mean, etc.

Adoption of Differential Privacy

Apple and Google have adapted DP to discover usage patterns from a large number of users.



“**RAPPOR** enables learning statistics about the behavior of users’ software while guaranteeing client privacy. The guarantees of **differential privacy**, which are widely accepted as being **the strongest form of privacy**, RAPPOR introduces a practical method to achieve those guarantees.” *Úlfar Erlingsson*, Google’s Tech Lead Manager, Security Research

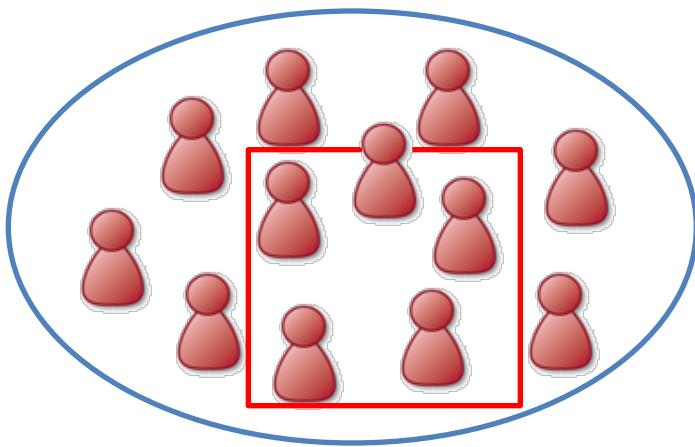


“iOS 10 adds Siri intelligence into **QuickType** and **Photos**, ..., and opens up **Siri**, **Maps**, **Phone** and **Messages** to developers — while **increasing security and privacy** with powerful technologies like **Differential Privacy**.” *Craig Federighi*, Apple’s senior vice president of Software Engineering, June 13, 2016.

Differentially Private Publication of **Location Entropy**
Thursday, November 3. 8:00 - 9:00 AM (Session 11)

Differential Privacy in Use

In spatial crowdsourcing, Q = “How many people in a particular area?”



Output of Q : $5 + \text{symmetric random Laplace noise}$

Problem Statement

Using Differential Privacy in Task Assignment phase of Spatial Crowdsourcing to hide the location of workers with low cost and low overhead without compromising performance.

- Low cost: travel cost
- Low overhead: communication overhead required to send task requests to workers
- Performance: # of performed tasks

Privacy-Preserving Task Assignment

Perform task assignment in DP-compliant manner

- Use worker dataset Private Spatial Decomposition (PSD)

Challenges:

1. DP introduces “fake” workers - may lead to **inefficient assignment**
 - Assign each task to all workers is **not scalable**
 - Ensure task completion, but keep **overhead low**, e.g., minimize the number of workers that need to be notified
2. Communication model significantly complicated
 - Establishing direct communication channels violates DP!
 - Instead, disseminate task requests within **geocast region**^[*]

[*] Navas and Imielinski. Geocast: geographic addressing and routing, MOBICOM'97

System Architecture

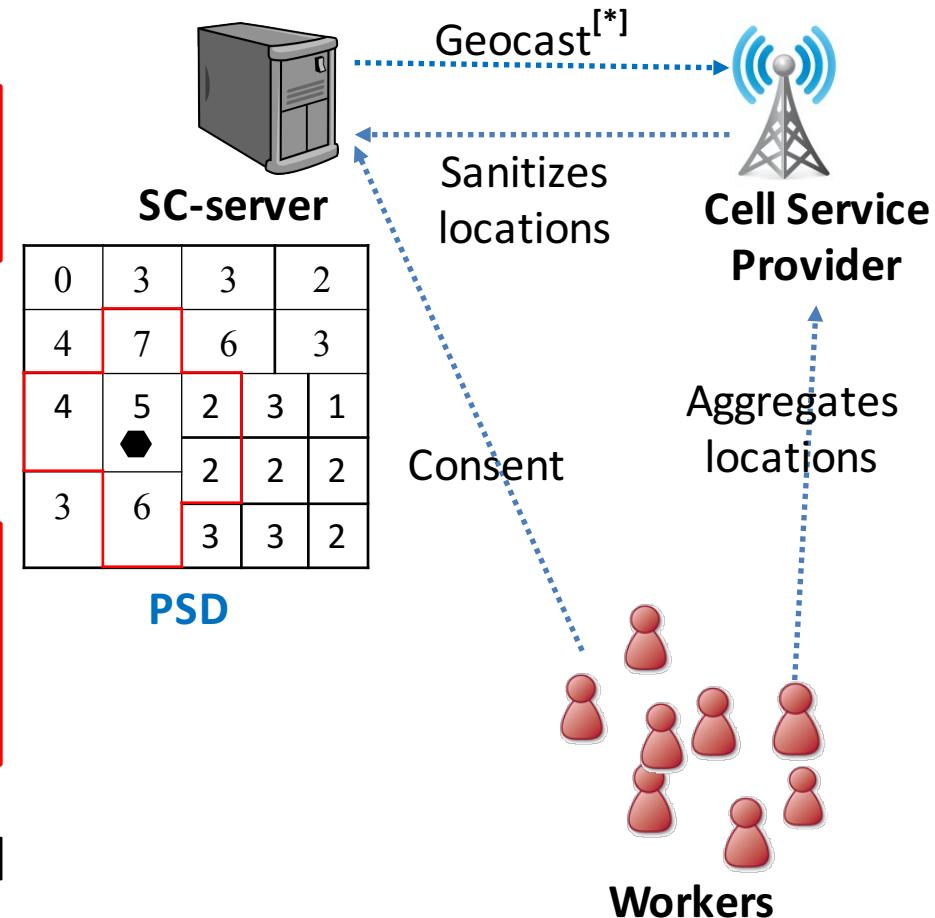
1. CSP aggregate locations

2. CSP releases **PSD**, according to DP

3. SC-server receives task request

4. SC-server queries **PSD** to find **geocast region (GR)** that encloses *sufficient* workers

5. SC-server geocast the task to all workers within *GR*



[*] Navas and Imielinski. Geocast: geographic addressing and routing, MOBICOM'97

Performance Metrics and Results

Assignment Success Rate (ASR)

- Ratio of tasks accepted by a worker

Similar performance compared to non-private case

Worker Travel Distance (WTD)

- Captures travel cost or assignment quality

20% increase compared to non-private case

Average Notified Workers (ANW)

- Captures overhead (communication required to geocast task requests or computation of assignment algorithm)

150% increase compared to non-private case

Publishing Results

[To et al. VLDB 2014]

- Differentially-private framework for task assignment

[To et al. ICDE 2015]^[1]

- Toolbox for turning spatial crowdsourcing parameters

[To et al. TMC 2016]^[2]

- Protecting worker locations across multiple timestamps

[1] To et al. PrivGeoCrowd: A Toolbox for Studying Private Spatial Crowdsourcing, ICDE 2015

[2] To et al. Differentially Private Location Protection for Worker Datasets in Spatial Crowdsourcing, TMC 2016

Outlines

- Introduction
- A taxonomy and research challenges
- Privacy-preserving spatial crowdsourcing
 - Risks of location leaks
 - Privacy threats during tasking and reporting
 - TaskRabbit: a case study
 - Privacy countermeasures: the state-of-the-art
- **Conclusion and future works**

Conclusions and future works

- Taxonomy of spatial crowdsourcing
- Most studies focus on **task assignment** problem
 - Multiple objectives: number (quality) of assigned tasks, total rewards, travel cost, communication overhead, etc.
 - Multiple considerations: privacy, trust, quality, incentive
- Privacy issue
 - Privacy risks occur to **workers** and **requesters** during tasking and reporting phases
 - Various techniques have been used: pseudonymity, cloaking, perturbation, encryption, exchange-based
- Task assignment with **rigorous privacy protection**
 - The cost of privacy is practical
 - But none solves privacy risks for both workers and tasks

Q/A

Hien To

Email: hto@usc.edu