VLDB 2017 Munich · Germany

Spatial Crowdsourcing: Challenges, Techniques, and Applications

Yongxin Tong
Beihang University

Lei Chen
Hong Kong University of Science and Technology

Cyrus Shahabi
University of Southern California

We get some pictures and materials from the Internet for this tutorial

Outline

- Overview of Spatial Crowdsourcing (20min)
  - Motivation
  - Workflow
  - Core Issues
  - Difference from Related Tutorials
- Fundamental Techniques (50min)
  - Task Assignment
  - Quality Control
  - Incentive Mechanism
  - Privacy Protection
- Spatial Crowdsourced Applications (15min)
  - Spatial Crowdsourcing Intrinsic Applications
  - Crowd-powered Spatial Applications
- Open Questions (5 min)

Crowdsourcing: Concept

- **Crowdsourcing**
  - **Organizing the crowd (Internet workers) to do micro-tasks in order to solve human-intrinsic problems**

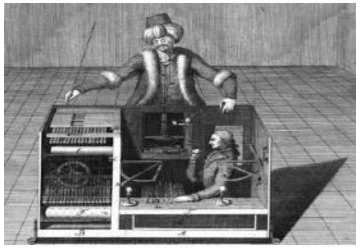
Crowdsourcing: Applications

- Wikipedia
  - Collaborative knowledge
- reCAPTCHA
  - Digitalizing newspapers

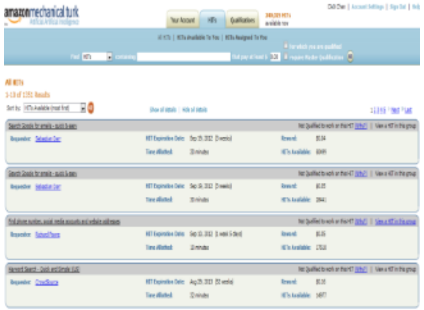
- Yahoo Answer
  - Question & Answer
- ImageNet
  - Image labelling
- Amazon Mechanical Turk (AMT)
  - General purpose crowdsourcing

Crowdsourcing: An Example of AMT

Requesters



AMT



MTurk workers
(Photo By Andrian Chen)

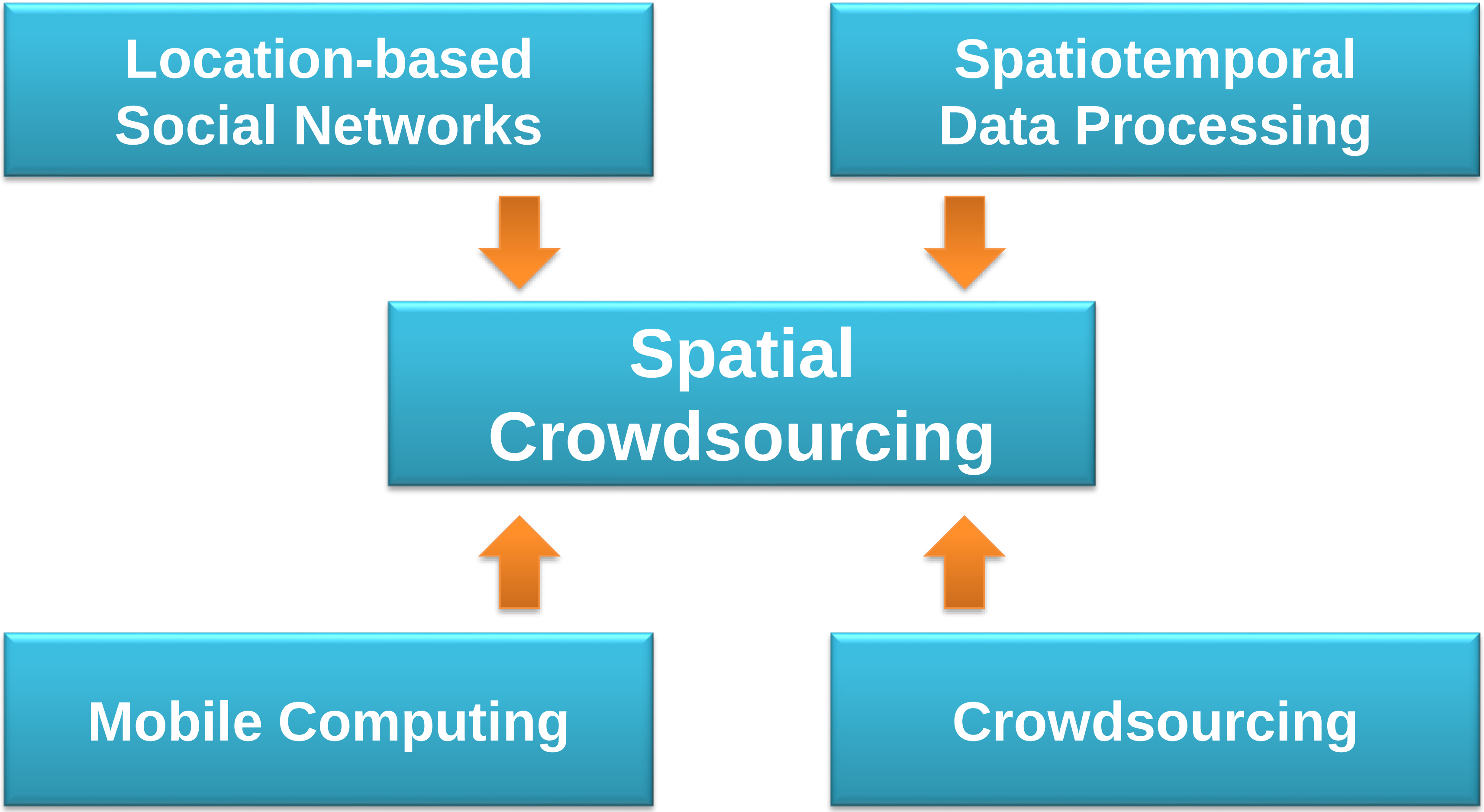
Crowdsourcing <u>in the Internet era</u> connects requesters and workers on the Internet

Spatial Crowdsourcing: Concept

- **Crowdsourcing**
  - **Organizing the crowd (Internet workers) to do micro-tasks in order to solve human-intrinsic problems**

- **Spatial Crowdsourcing**
  - **Organizing the crowd (Mobile Internet workers) to do spatial tasks by physically moving to other locations**

Spatial Crowdsourcing: Concept



a.k.a. mobile crowdsourcing, mobile crowdsensing, participatory sensing, location-based crowdsourcing

Spatial Crowdsourcing: Applications

- Open Street Map
  - Collaborative map
- Waze
  - Live traffic
- Facebook
  - Check-in
- Uber
  - Smart transportation
- Gigwalk
  - General purpose spatial crowdsourcing

Application Comparison

	(Internet) Crowdsourcing	Spatial Crowdsourcing
Information Collection	WIKIPEDIA The Free Encyclopedia	
Passive Participation		waze
Question & Answer	YAHOO! ANSWERS	facebook
Specialized Platforms	IMAGENET	UBER
General Platforms	amazon mechanical turk beta	Gigwalk

Spatial Crowdsourcing: Other Apps

Local Search-and-discovery Service

Mobile Market Research & Audits

Repair & Refresh Your Home

Intelligent Transportation

Food Delivery

Location-based Game

Outline

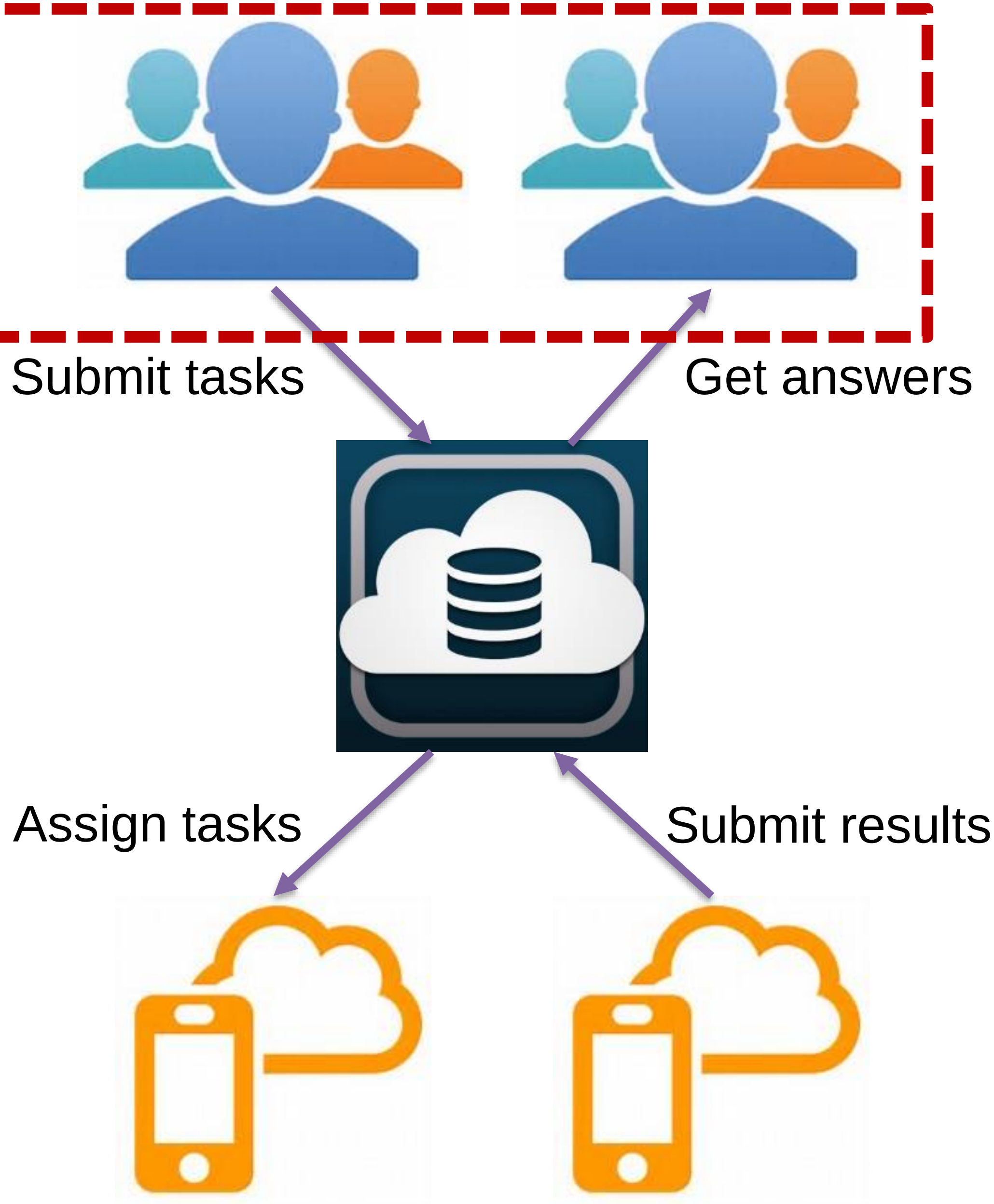
- Overview of Spatial Crowdsourcing (20min)
  - Motivation
  - Workflow
  - Core Issues
  - Difference from Related Tutorials
- Fundamental Techniques (50min)
  - Task Assignment
  - Quality Control
  - Incentive Mechanism
  - Privacy Protection
- Spatial Crowdsourced Applications (15min)
  - Spatial Crowdsourcing Intrinsic Applications
  - Crowd-powered Spatial Applications
- Open Questions (5min)

Spatial Crowdsourcing: Workflow

- Requesters
  - Submit tasks
- Platforms
  - Task management
- Workers
  - Perform tasks

Spatial Crowdsourcing: Workflow

- **Requesters**
  - **Submit tasks**
- **Platforms**
  - **Task management**
- **Workers**
  - **Perform tasks**



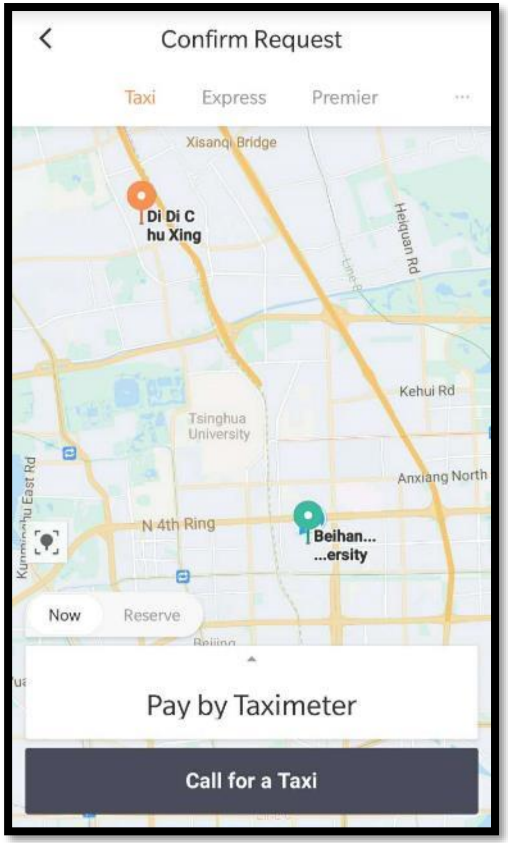
Spatial Crowdsourcing: Task

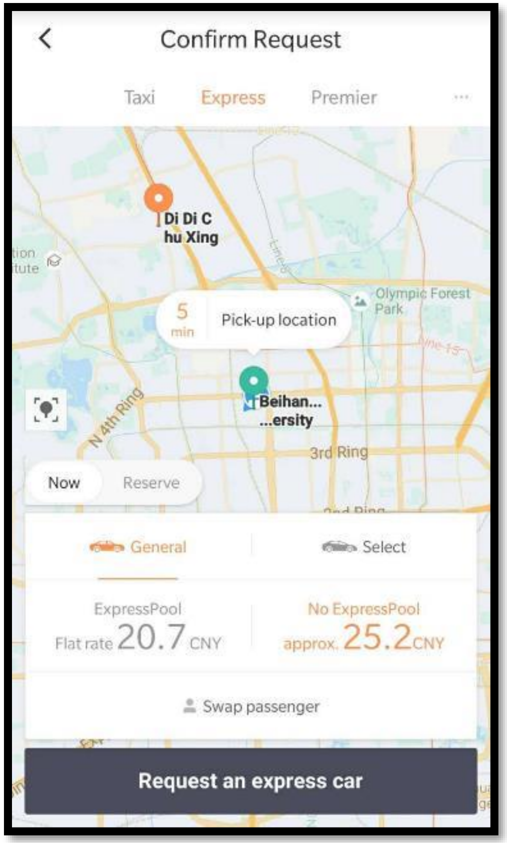
- General spatial tasks
  - Inventory identification
  - Placement checking
  - Data collection
  - Service evaluation

Gigwalk

Spatial Crowdsourcing: Task

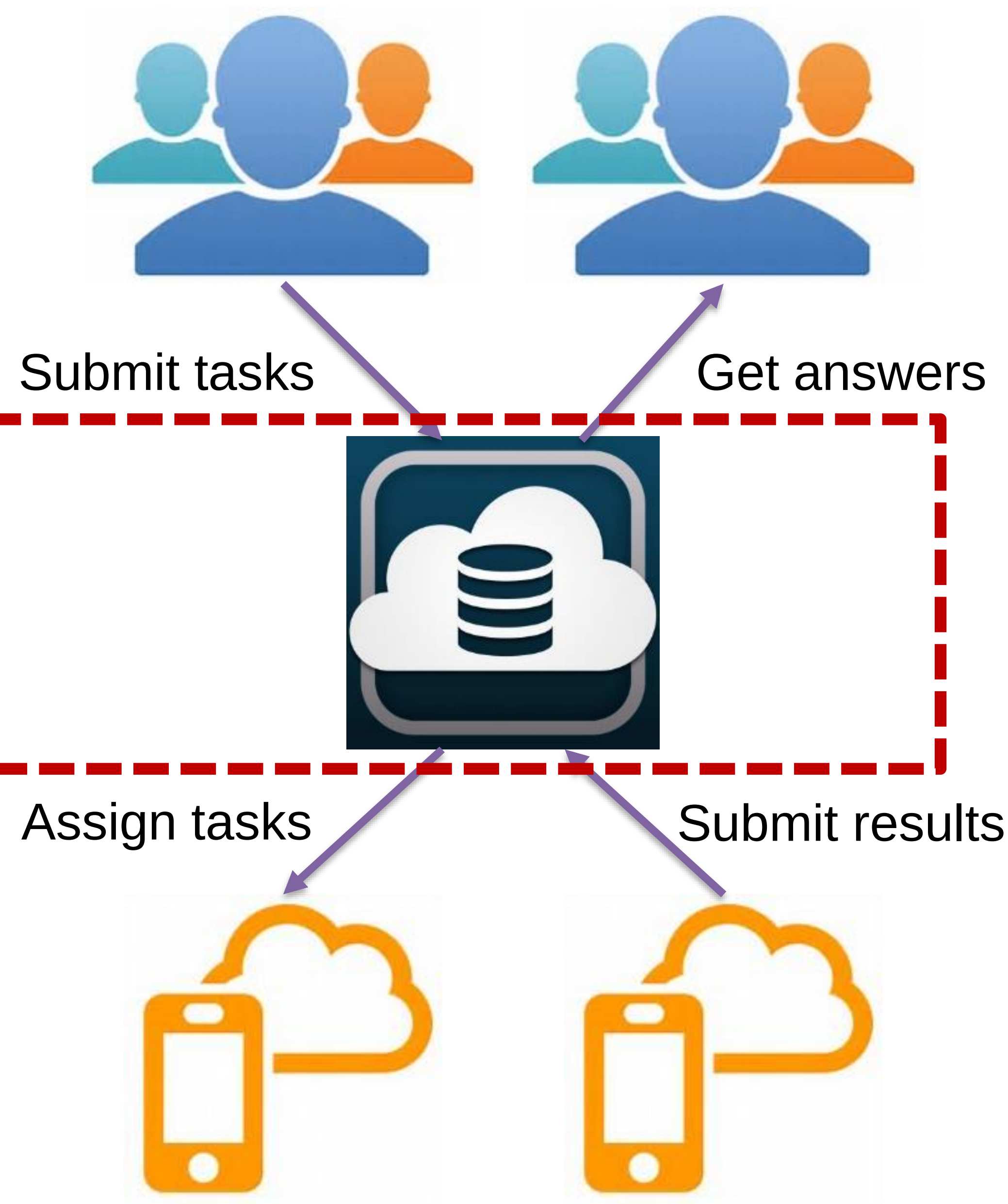
- Specific spatial tasks
  - Taxi calling service
  - Ridesharing service





Spatial Crowdsourcing: Workflow

- Requesters
  - Submit tasks
- Platforms
  - Task management
- Workers
  - Perform tasks



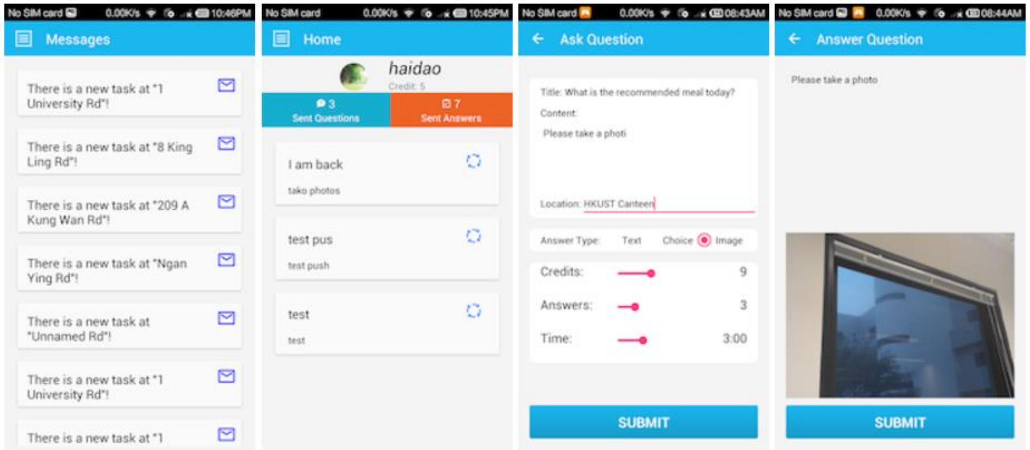
Spatial Crowdsourcing: Platform

- Management Modes
  - Worker Selected Tasks (WST)
    - Workers actively select tasks
  - Server Assigned Tasks (SAT)
    - Workers passively wait for the platform to assign tasks

Most studies focus on SAT mode because more optimization techniques can be designed by platforms

Other Platforms

- gMission
  - An open sourced spatial crowdsourcing platform
  - http://gmission.github.io

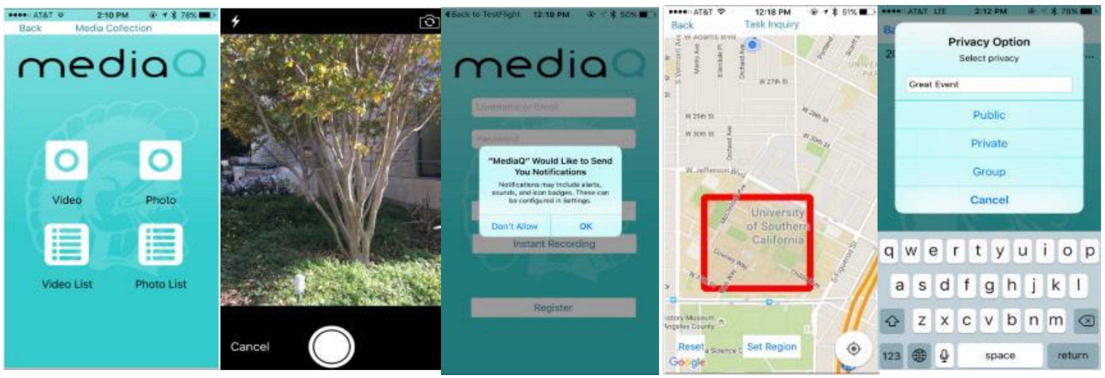


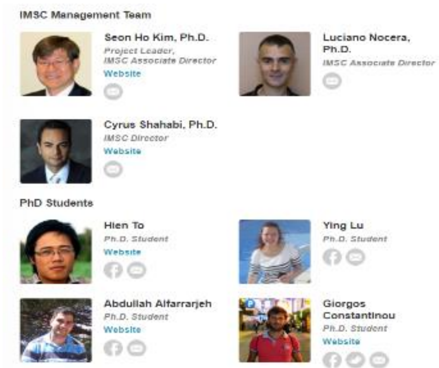
Available on the App Store

ANDROID APP ON Google play

Other Platforms

- **MediaQ**
  - **An online spatial-crowdsourcing-based media management system**
  - **http://mediaq.usc.edu/**



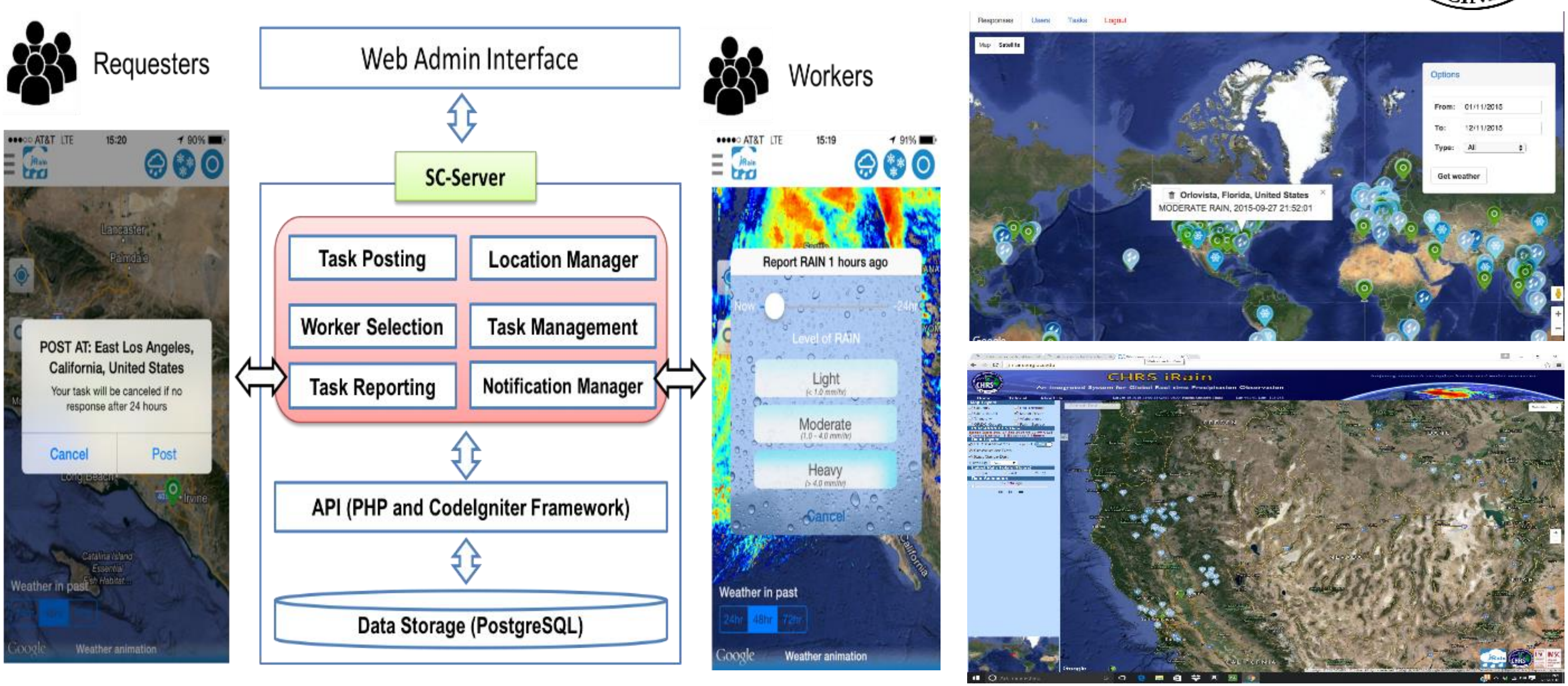


Available on the App Store

ANDROID APP ON Google play

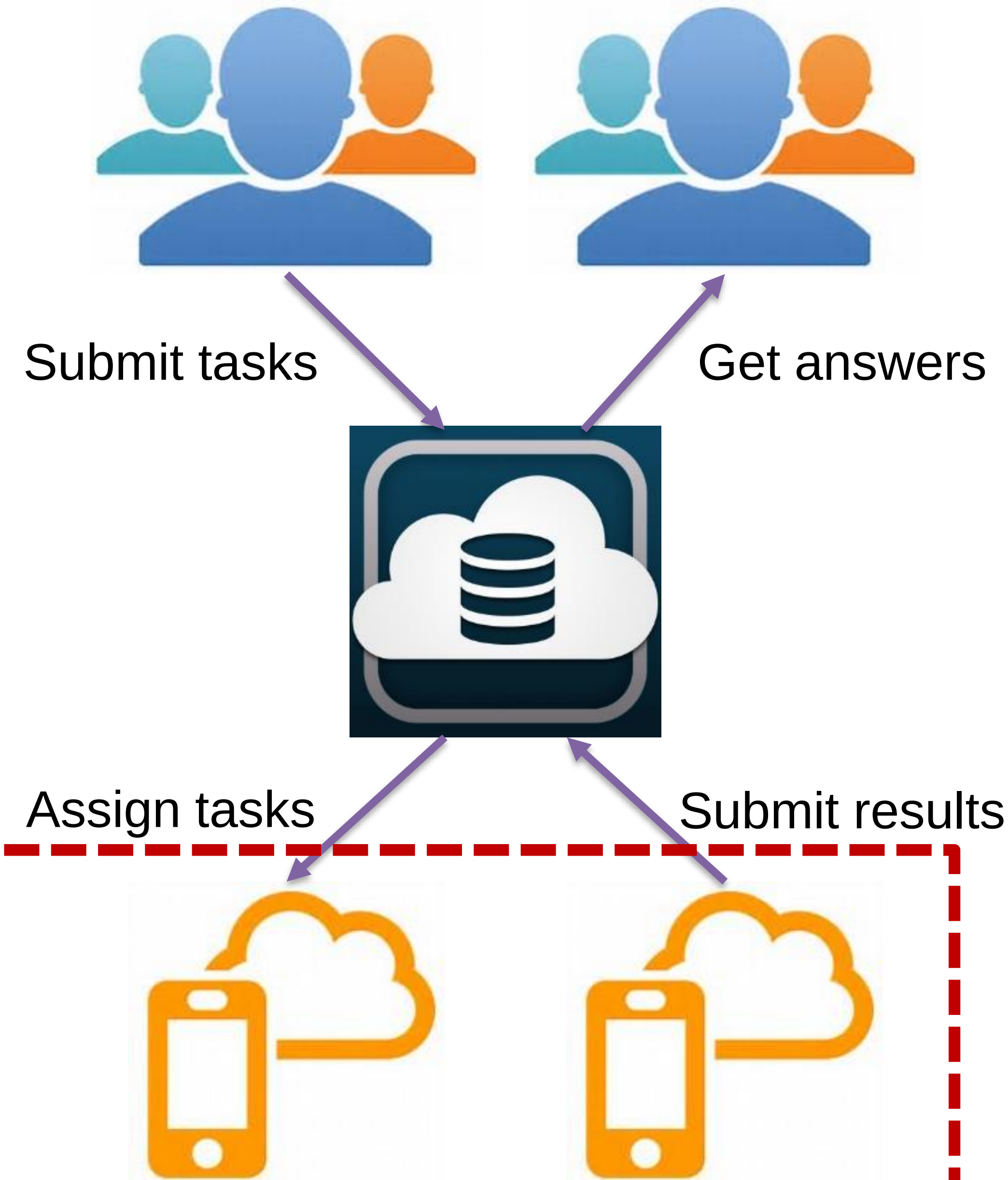
Other Platforms

- iRain
  - A Spatial Crowdsourcing System for Real-time Rainfall Observation
  - http://irain.eng.uci.edu/



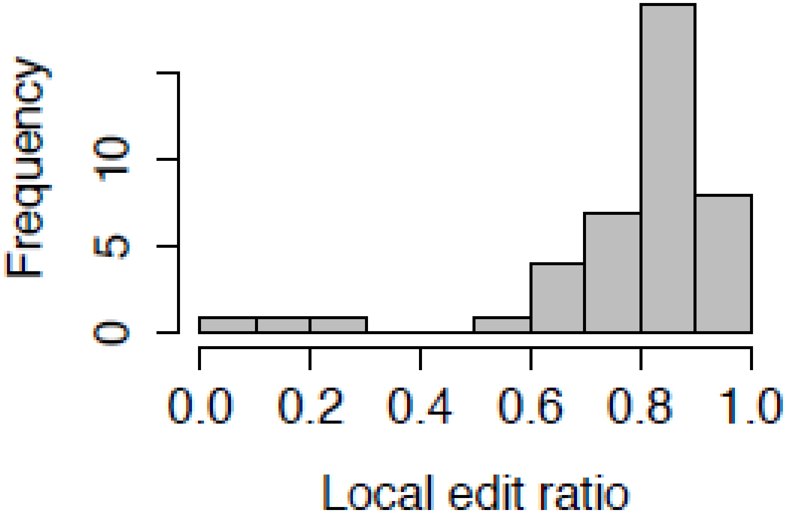
Spatial Crowdsourcing: Workflow

- Requesters
  - Submit tasks
- Platforms
  - Task management
- Workers
  - Perform tasks



Spatial Crowdsourcing: Worker

- **Influential factor**
  - **Distance**
  - **Socioeconomic status**

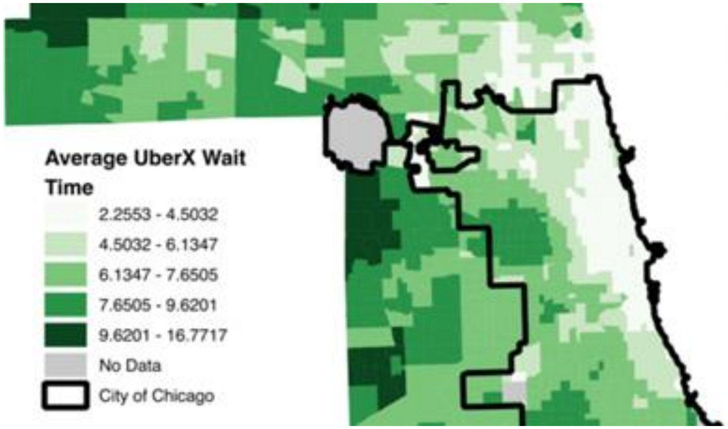


Workers tend to perform nearby tasks

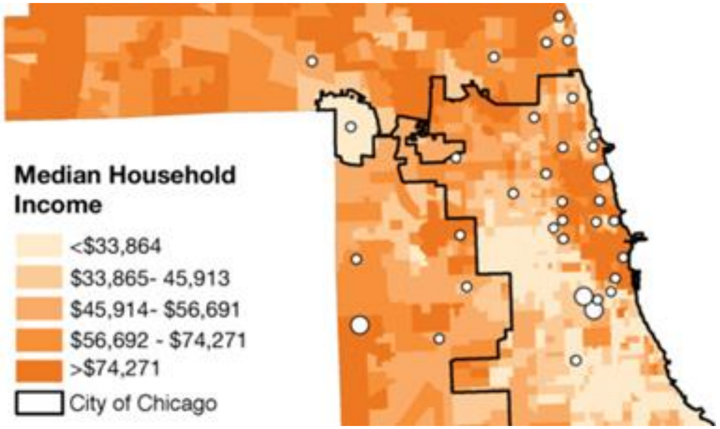
G. Quattrone, A. J. Mashhadi, L. Capra. Mind the map: the impact of culture and economic affluence on crowd-mapping behaviours. CSCW 2014.

Spatial Crowdsourcing: Worker

- Influential factor
  - Distance
  - Socioeconomic status



Average UberX Wait Time in Chicago



Average Median Household Income in Chicago

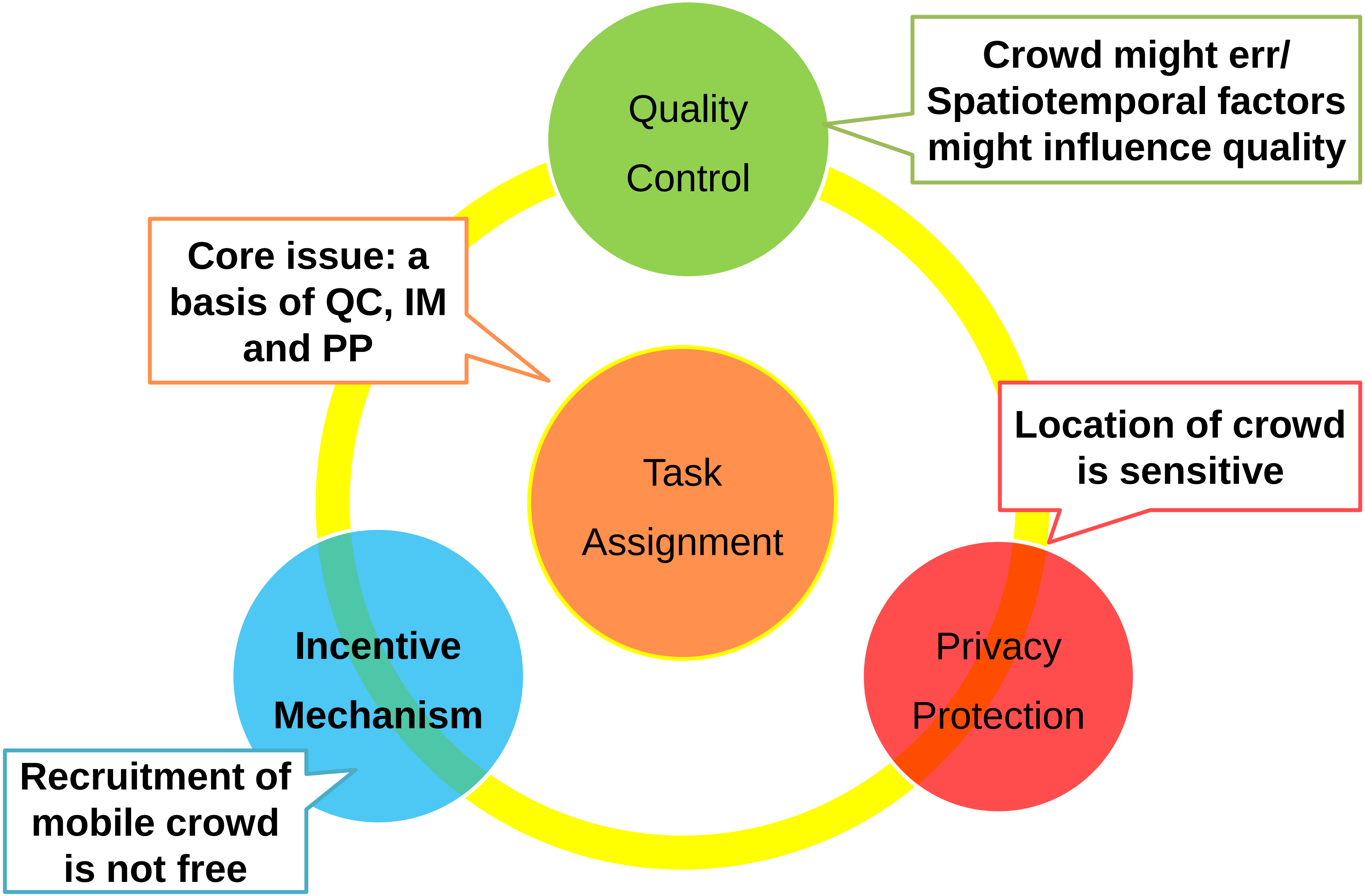
Workers tend to perform tasks in high income regions

J. Thebault-Spieker, L. G. Terveen, B. J. Hecht. Toward a Geographic Understanding of the Sharing Economy: Systemic Biases in UberX and TaskRabbit. TOCHI 2017.

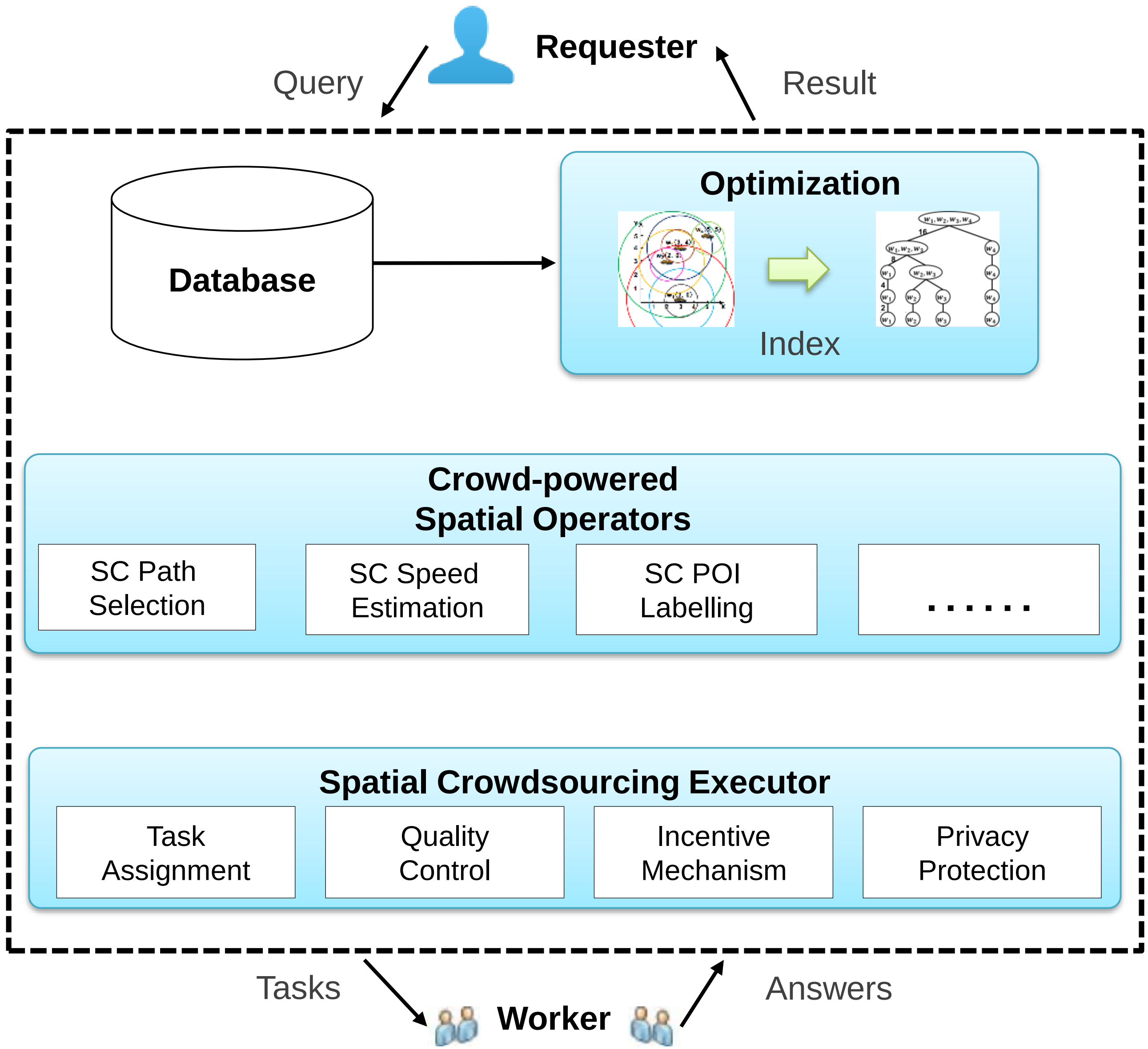
Outline

- Overview of Spatial Crowdsourcing (20min)
  - Motivation
  - Workflow
  - Core Issues
  - Difference from Related Tutorials
- Fundamental Techniques (50min)
  - Task Assignment
  - Quality Control
  - Incentive Mechanism
  - Privacy Protection
- Spatial Crowdsourced Applications (15min)
  - Spatial Crowdsourcing Intrinsic Applications
  - Crowd-powered Spatial Applications
- Open Questions (5min)

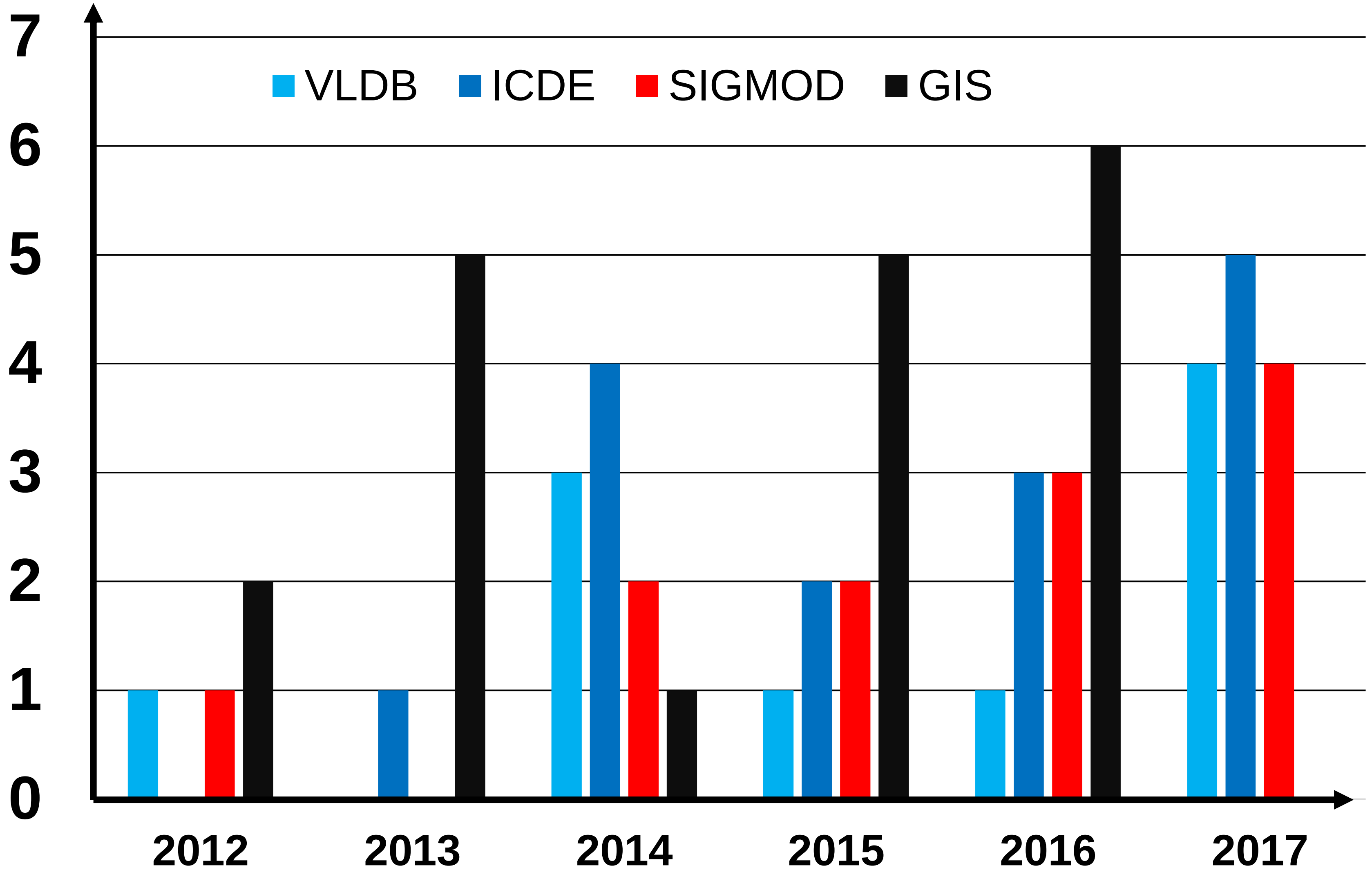
Core Issues in Spatial Crowdsourcing



A Spatial Crowdsourcing System



Existing Works



Outline

- Overview of Spatial Crowdsourcing (20min)
  - Motivation
  - Workflow
  - Core Issues
  - Difference from Related Tutorials
- Fundamental Techniques (50min)
  - Task Assignment
  - Quality Control
  - Incentive Mechanism
  - Privacy Protection
- Spatial Crowdsourced Applications (15min)
  - Spatial Crowdsourcing Intrinsic Applications
  - Crowd-powered Spatial Applications
- Open Questions (5min)

Difference from Related Tutorials

- VLDB 2012
  - Crowdsourced platforms and principles
- CIKM 2014
  - Traditional crowdsourced knowledge management
- ICDE 2015
  - Traditional crowdsourced queries, mining and applications
- VLDB 2016
  - Human factors involved in task assignment
- SIGMOD 2017
  - Traditional crowdsourced data management
- Our Tutorial
  - Spatial crowdsourced data management
  - Theories and applications in spatial crowdsourcing

Reference: Overview

1. L. Chen, C. Shahabi. Spatial crowdsourcing: challenges and opportunities. IEEE Data Engineering Bulletin, 39(4):14-25, 2016.
2. G. Li, J. Wang, Y. Zheng, J. Franklin. Crowdsourced data management: a survey. TKDE, 28(9): 2296-2319, 2016.
3. H. Garcia-Molina, M. Joglekar, A. Marcus,A. Parameswaran,V. Verroios. Challenges in Data Crowdsourcing. TKDE, 28(4):901-911, 2016.
4. A. Chittilappilly, L. Chen, S. Amer-Yahia. A Survey of General-Purpose Crowdsourcing Techniques. TKDE, 28(9):2246-2266, 2016.
5. G. Chatzimilioudis, A. Konstantinidis, C. Laoudias, D. Zeinalipour-Yazti: Crowdsourcing with Smartphones. IC,16(5): 36-44, 2012.
6. R. K.Ganti, F. Ye,H. Lei, Mobile crowdsensing: current state and future challenges. IEEE Commun. Mag. , 49(11), 32-39 ,2011.
7. J. Thebault-Spieker, L. G.Terveen, B. J.Hecht. Toward a geographic understanding of the sharing economy: systemic biases in uberX and taskrabbit. TOCHI, 24(3): 21:1-21:40, 2017.
8. Z. Chen, R. Fu, Z. Zhao, Z. Liu, L. Xia, L. Chen, P. Cheng, C. Cao, Y. Tong, C. Zhang. gMission: a general spatial crowdsourcing platform. PVLDB, 7(13): 1629-1632, 2014.
9. Kim, S. Ho, Y. Lu, G. Constantinou, C. Shahabi, G. Wang, R. Zimmermann. Mediaq: mobile multimedia management system. In MMSys, pages 224-235, 2014.
10. F. Alt, A. Sahami Shirazi, A. Schmidt, U. Kramer, Z. Nawaz. Location-based crowdsourcing: extending crowdsourcing to the real world. In NordiCHI, pages 13–22, lorian 2010.

Outline

- Overview of Spatial Crowdsourcing (20min)
  - Motivation
  - Workflow
  - Core Issues
  - Difference from Related Tutorials
- Fundamental Techniques (50min)
  - Task Assignment
  - Quality Control
  - Incentive Mechanism
  - Privacy Protection
- Spatial Crowdsourced Applications (15min)
  - Spatial Crowdsourcing Intrinsic Applications
  - Crowd-powered Spatial Applications
- Open Questions (5min)

Why Task Assignment?

- Almost all other core issues in spatial crowdsourcing depend on task assignment
- The core operation connects tasks, workers and the platform

DiDi

Gigwalk

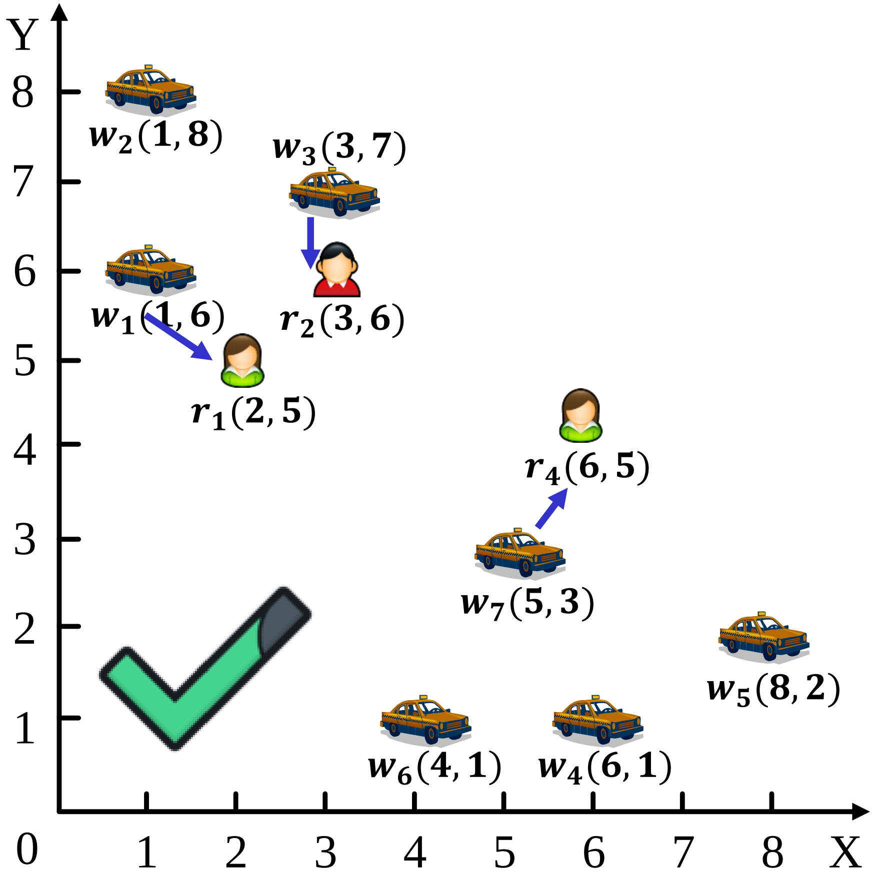
UBER

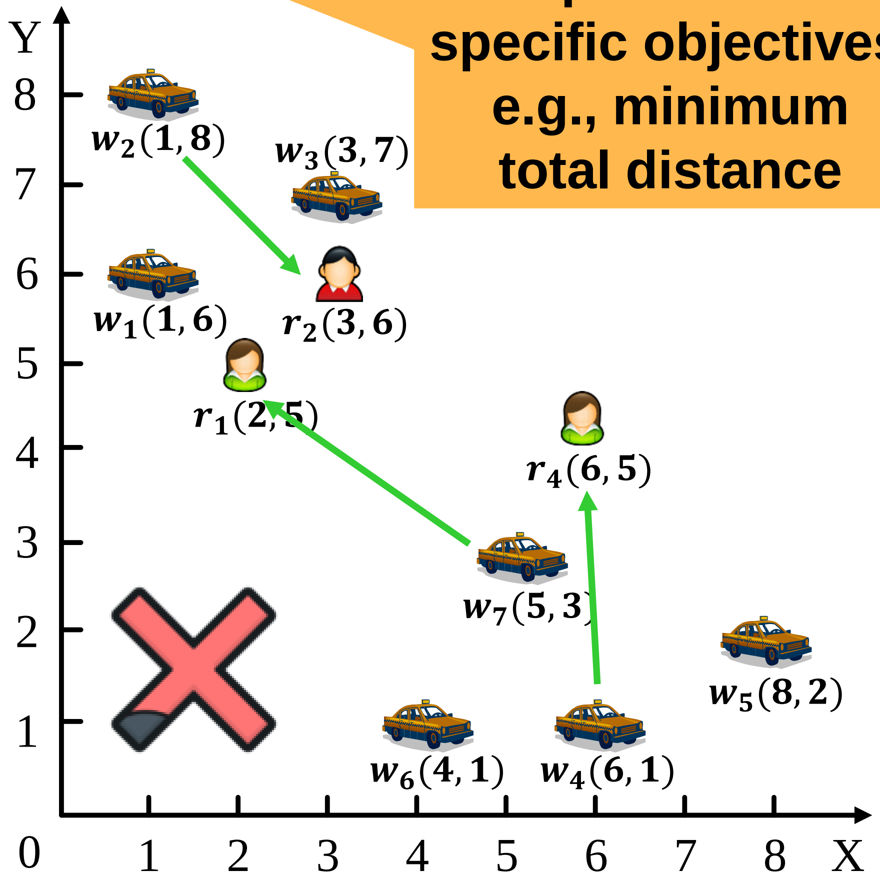
BOSTON taskrabbit
Life is busy. We can help.

Problem Definition

- Given **a set of workers and a set of tasks distributed in the physical space**, **the objective is to assign tasks to proper workers**

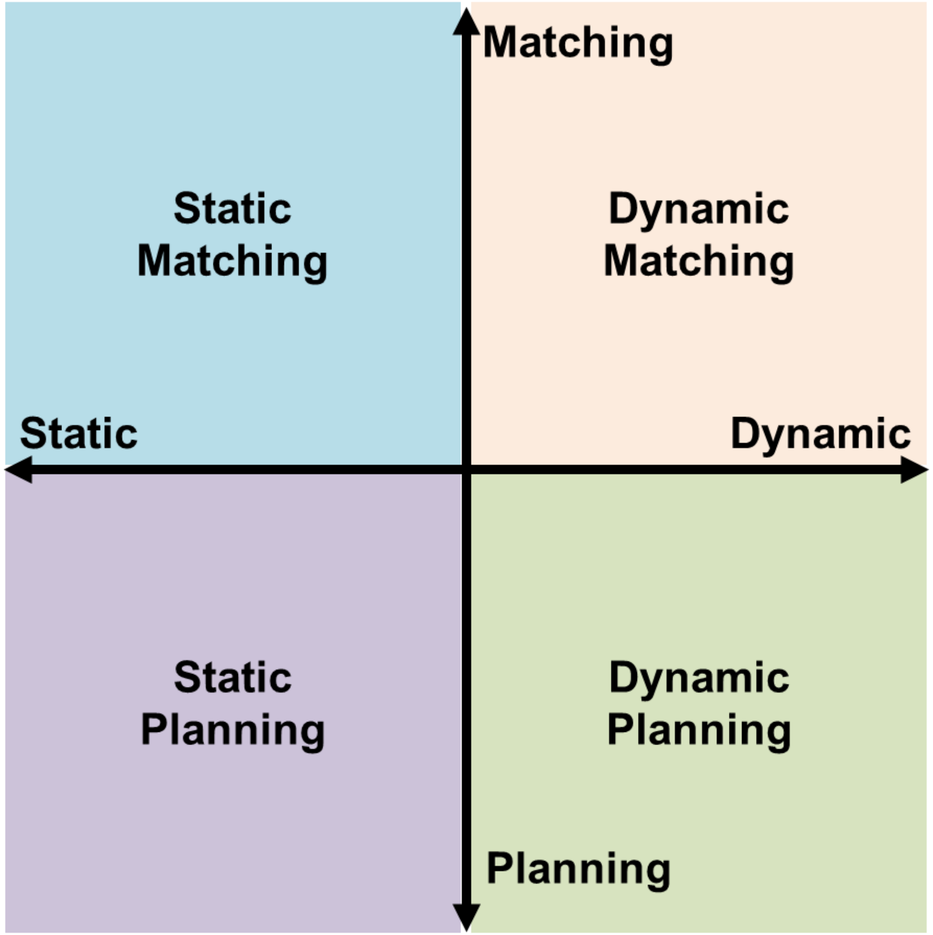
Dependent on specific objectives: e.g., minimum total distance





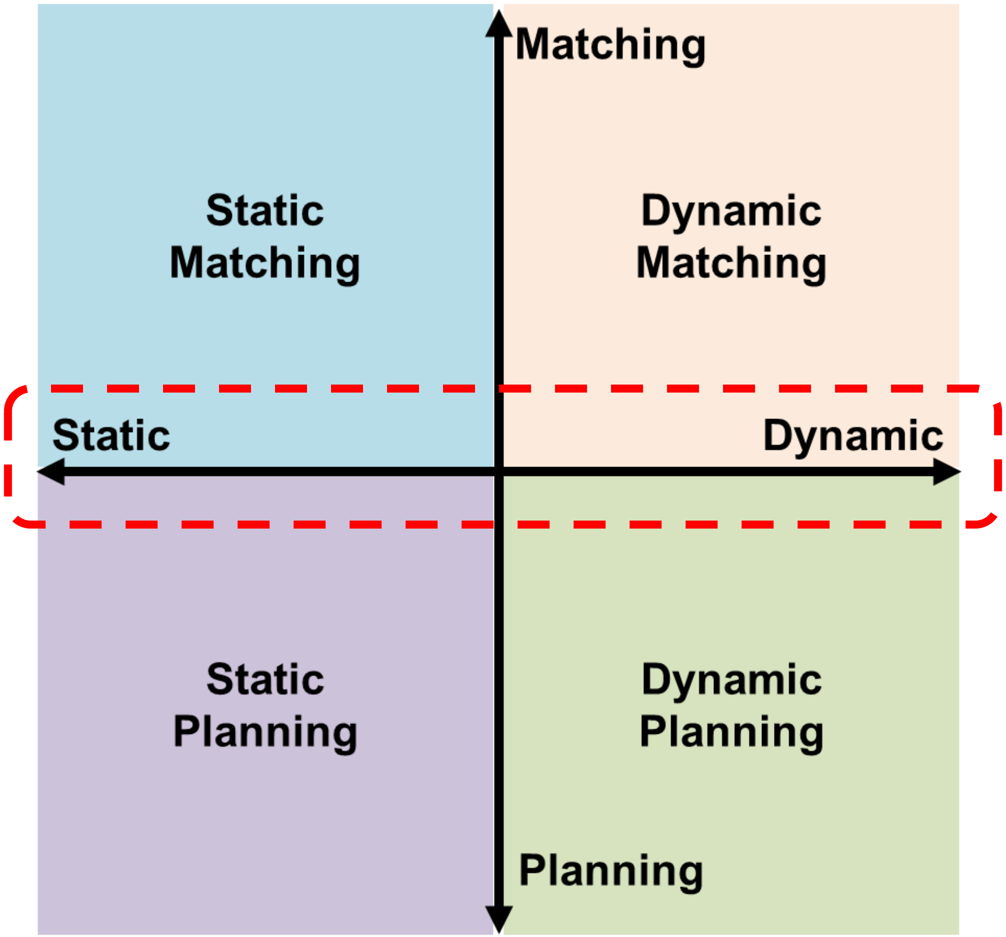
Categories of Existing Research

- Arrival Scenarios: Static vs. Dynamic
- Algorithmic Models: Matching vs. Planning



Categories of Existing Research

- Arrival Scenarios: Static vs. Dynamic
- Algorithmic Models: Matching vs. Planning



Static Scenario vs. Dynamic Scenario

- **Static Scenario**
  - **The platform is assumed to know all spatiotemporal information of tasks and workers at the beginning**

Refer to all arrival times and locations of tasks/workers

Platform

The platform knows all information when assigning tasks.

Arrival Time	8:00	8:01	8:02	8:07	8:08	8:09	8:09	8:15	8:18
	t_1	w_1	t_2	t_3	w_2	t_4	w_3	w_4	t_5

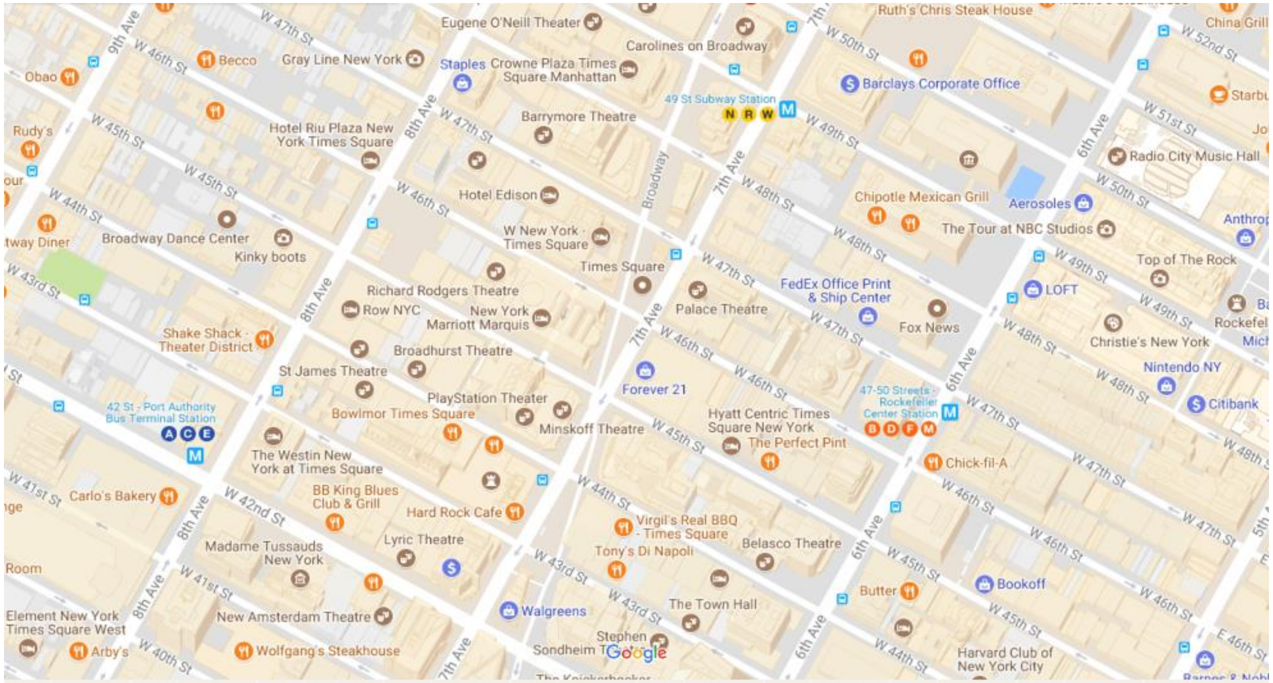
Static Scenario vs. Dynamic Scenario

- **Dynamic Scenario**
  - **Tasks/workers usually dynamically appear**
  - **They usually need to be assigned immediately based on real-time partial information**
    - **Fast Food Delivery Service**
    - **Real-Time Taxi-Calling Service**

Static Scenario vs. Dynamic Scenario

- **Dynamic Scenario: An Example**
  - **New tasks and workers dynamically arrive at the platform during task assignment**

Platform



Static Scenario vs. Dynamic Scenario

- **Dynamic Scenario: An Example**
  - **New tasks and workers dynamically arrive at the platform during task assignment**

Platform

Platform gradually knows information as time goes on.

Arrival Time	8:00
	t_1

Static Scenario vs. Dynamic Scenario

- **Dynamic Scenario: An Example**
  - **New tasks and workers dynamically arrive at the platform during task assignment**

Platform

Platform gradually knows information as time goes on.

Arrival Time	8:00	8:01
	t_1	w_1

Static Scenario vs. Dynamic Scenario

- **Dynamic Scenario: An Example**
  - **New tasks and workers dynamically arrive at the platform during task assignment**

Platform

Platform gradually knows information as time goes on.

Arrival Time	8:00	8:01	8:02
	t_1	w_1	t_2

Static Scenario vs. Dynamic Scenario

- **Dynamic Scenario: An Example**
  - **New tasks and workers dynamically arrive at the platform during task assignment**

Platform

Platform gradually knows information as time goes on.

Arrival Time	8:00	8:01	8:02	8:07
	t_1	w_1	t_2	t_3

Static Scenario vs. Dynamic Scenario

Dynamic Scenario: An Example

- New tasks and workers dynamically arrive at the platform during task assignment

Platform

Platform gradually knows information as time goes on.

Arrival Time	8:00	8:01	8:02	8:07	8:08
	t_1	w_1	t_2	t_3	w_2

Static Scenario vs. Dynamic Scenario

- **Dynamic Scenario: An Example**
  - **New tasks and workers dynamically arrive at the platform during task assignment**

Platform

Platform gradually knows information as time goes on.

Arrival Time	8:00	8:01	8:02	8:07	8:08	8:09
	t_1	w_1	t_2	t_3	w_2	t_4

Static Scenario vs. Dynamic Scenario

- **Dynamic Scenario: An Example**
  - **New tasks and workers dynamically arrive at the platform during task assignment**

Platform

Platform gradually knows information as time goes on.

Arrival Time	8:00	8:01	8:02	8:07	8:08	8:09	8:09
	t_1	w_1	t_2	t_3	w_2	t_4	w_3

Static Scenario vs. Dynamic Scenario

- **Dynamic Scenario: An Example**
  - **New tasks and workers dynamically arrive at the platform during task assignment**

Platform

Platform gradually knows information as time goes on.

Arrival Time	8:00	8:01	8:02	8:07	8:08	8:09	8:09	8:15
	t_1	w_1	t_2	t_3	w_2	t_4	w_3	w_4

Static Scenario vs. Dynamic Scenario

- **Dynamic Scenario: An Example**
  - **New tasks and workers dynamically arrive at the platform during task assignment**

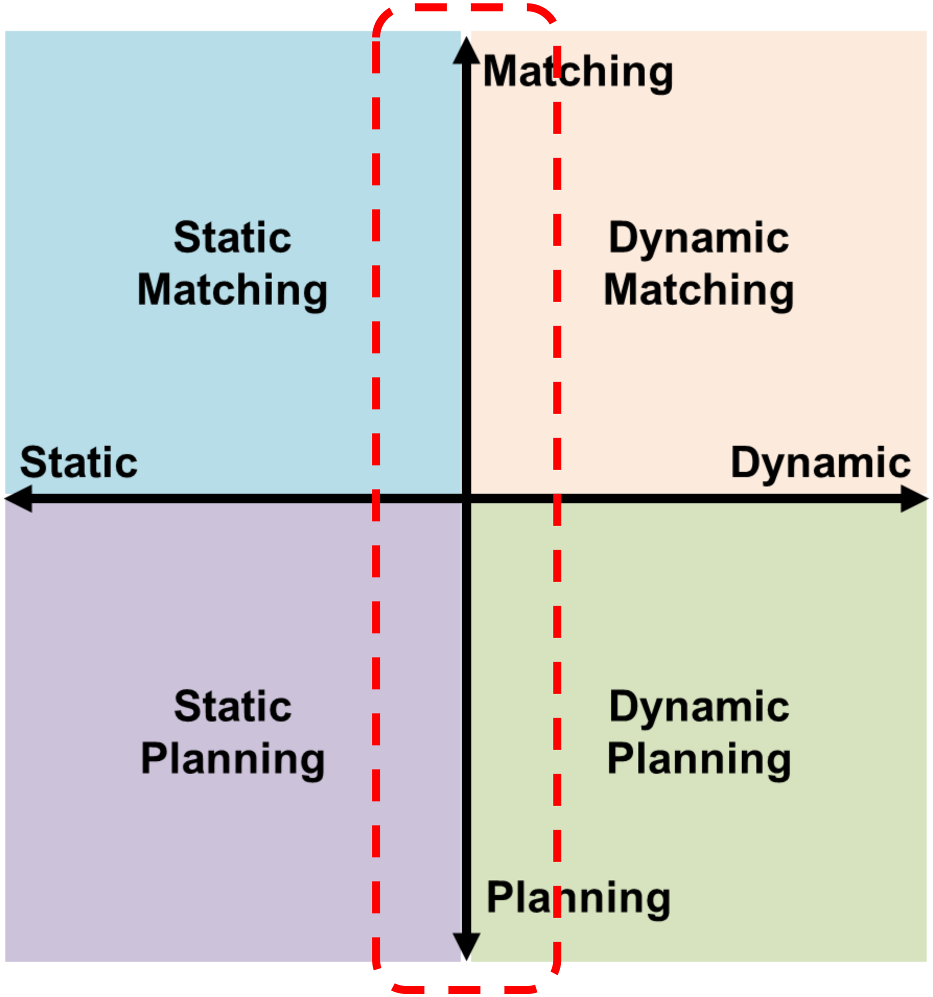
Platform

The platform does not know information of subsequent tasks/workers when it performs task assignment dynamically

Arrival Time	8:00	8:01	8:02	8:07	8:08	8:09	8:09	8:15	8:18
	t_1	w_1	t_2	t_3	w_2	t_4	w_3	w_4	t_5

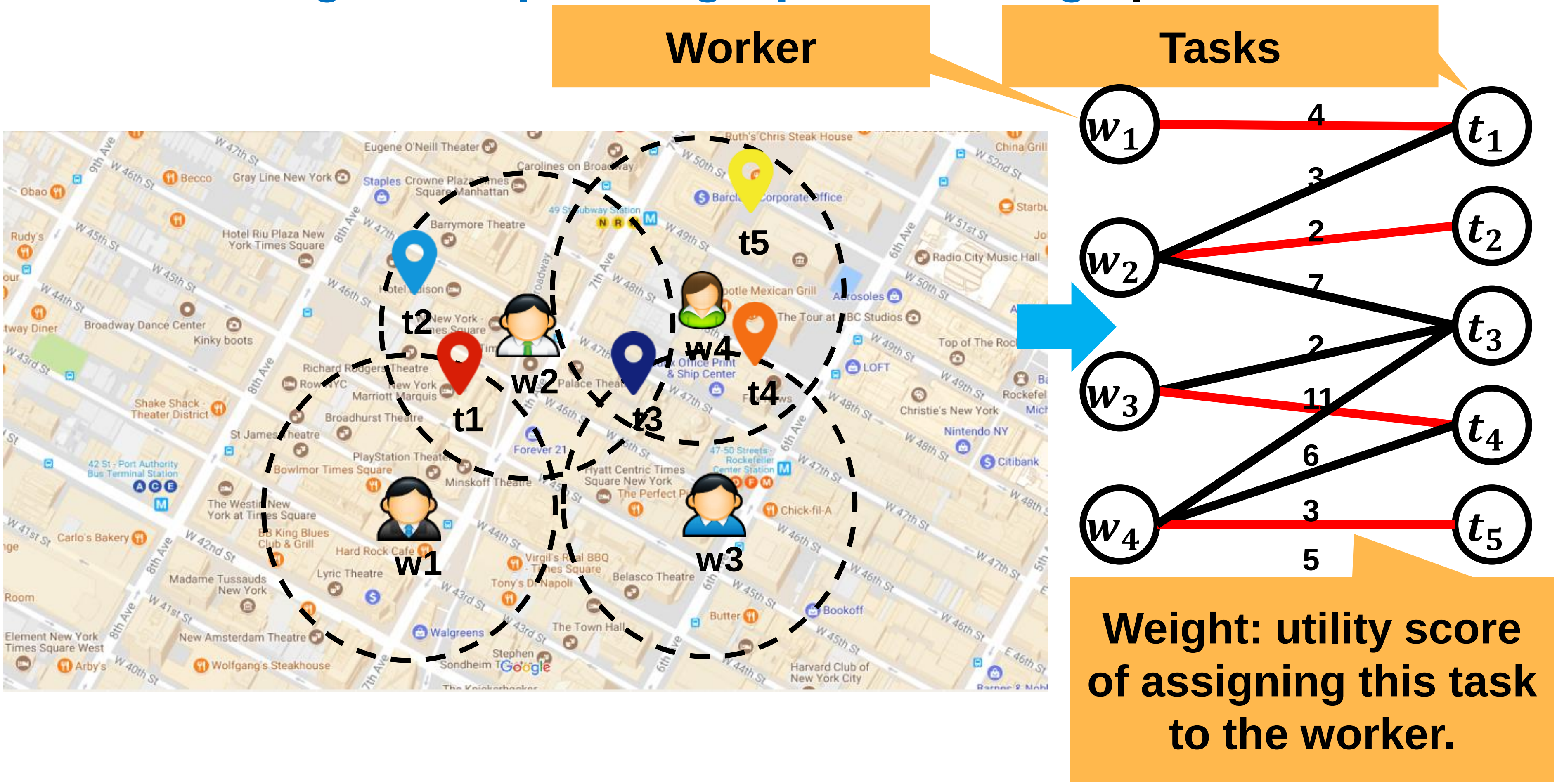
Categories of Existing Research

- Arrival Scenarios: Static vs. Dynamic
- Algorithmic Models: Matching vs. Planning



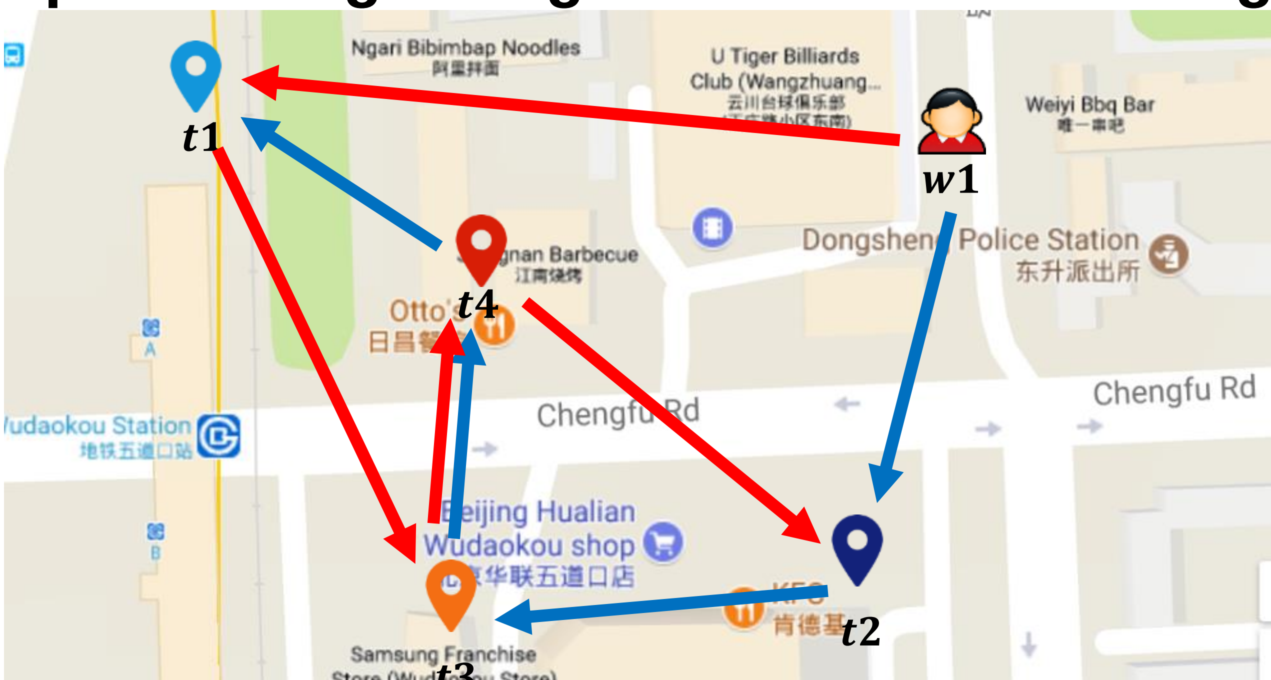
Matching vs. Planning

- **Matching Model**
  - **Formulate task assignment problem as classic "weighted bipartite graph matching" problem**



Matching vs. Planning

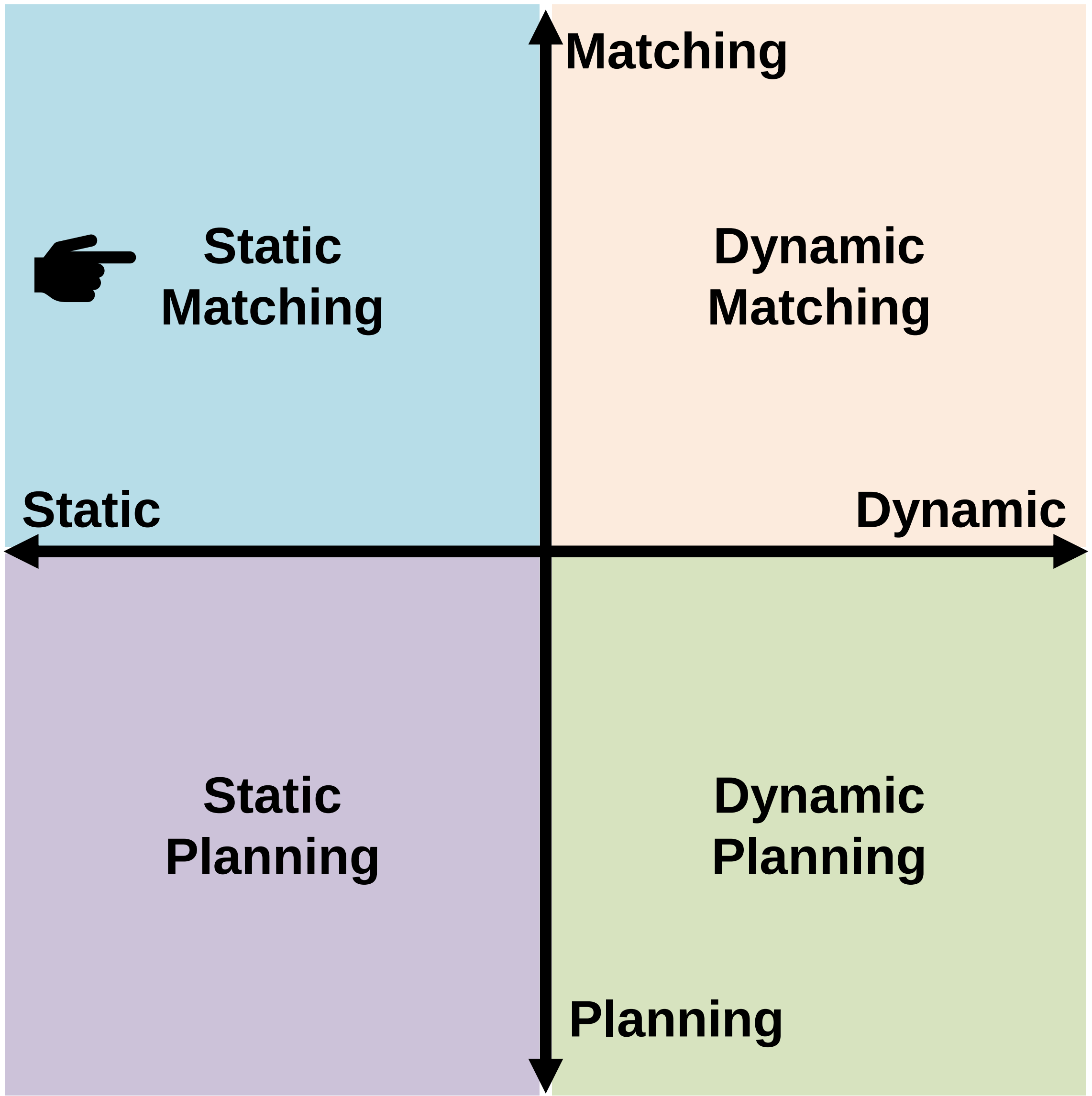
- **Planning Model**
  - **Plan a route for a worker or workers to complete a number of tasks**
  - **Each worker usually performs multiple tasks instead of performing a single task in the matching model**



Plan 1: $t2 \rightarrow t3 \rightarrow t4 \rightarrow t1$

Plan 2: $t1 \rightarrow t3 \rightarrow t4 \rightarrow t2$

Static Matching



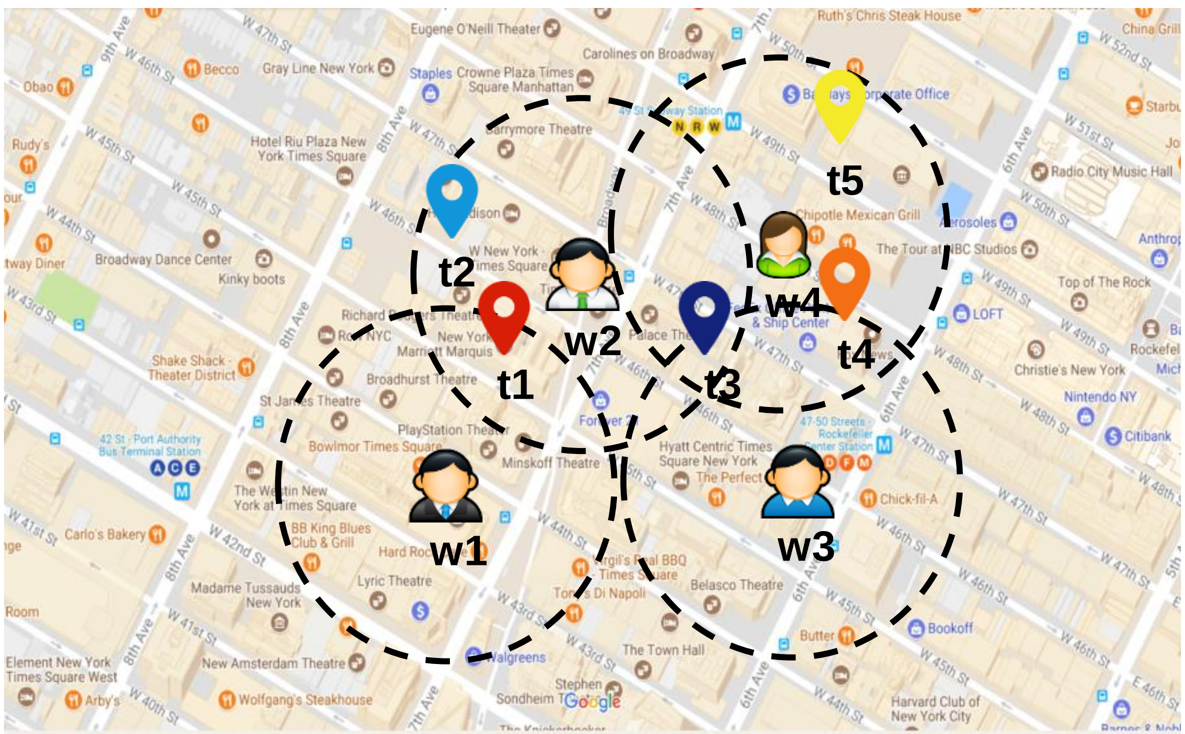
Static Matching

- Problem Definition

- Existing Research
  - Objective 1: MaxSum Matching
  - Objective 2: MinSum Matching
  - Objective 3: Stable Marriage Matching

Static Matching

- Problem Definition
  - Given all spatiotemporal information of a set of workers and tasks, the problem is to assign the tasks to the proper workers based on the specific objective function



Static Matching

- Problem Definition

- Existing Research
  - Objective 1: MaxSum Matching
  - Objective 2: MinSum Matching
  - Objective 3: Stable Marriage Matching

Spatial information acts as constraints

Spatial information is the optimization goal

Static Matching

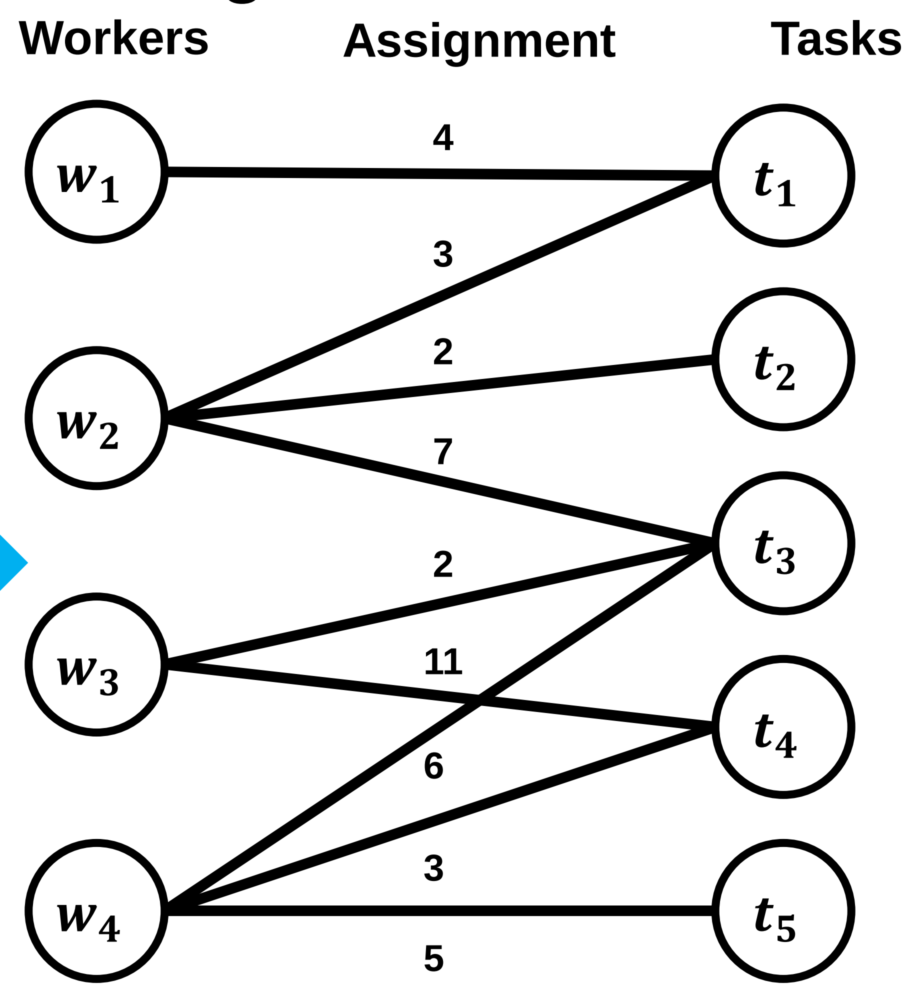
- Problem Definition

- Existing Research
  - Objective 1: MaxSum Matching
  - Objective 2: MinSum Matching
  - Objective 3: Stable Marriage Matching

MaxSum Matching

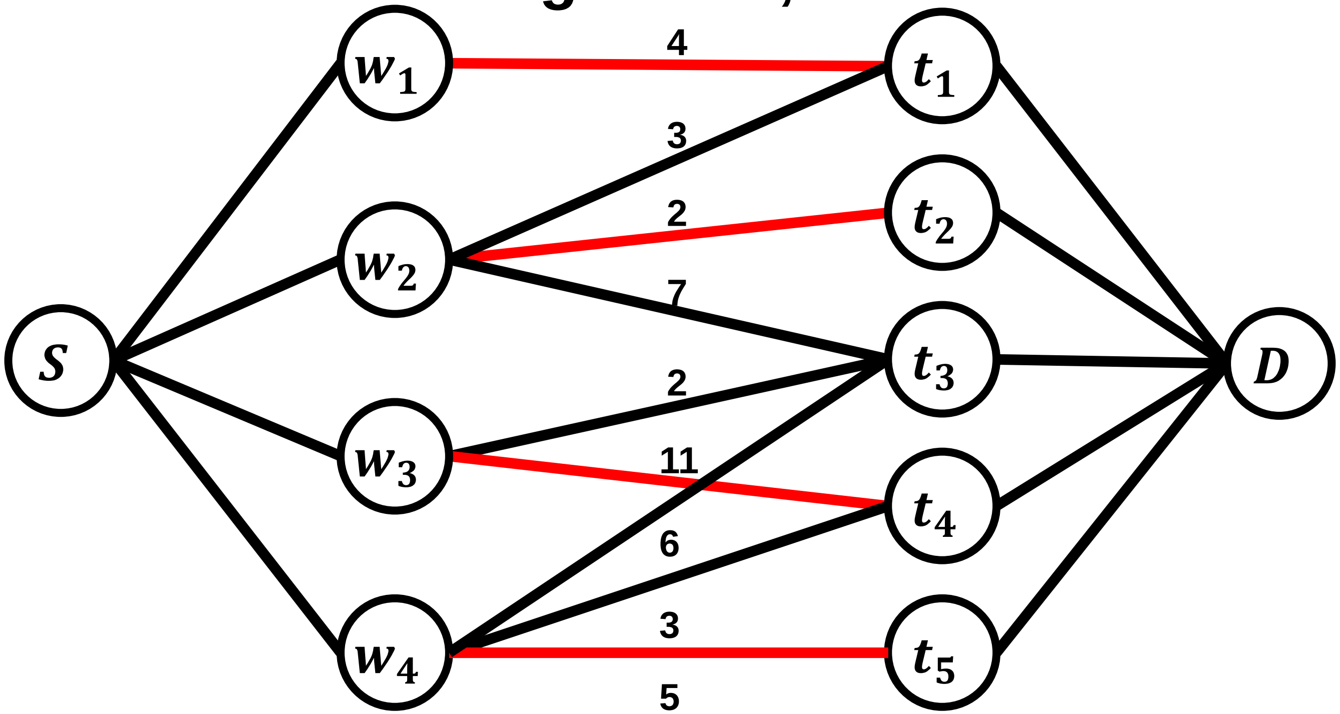
- **Objective 1: Maximizing the total utility/number of the matching**

The problem of static matching with Objective 1 equals to the Maximum Weighted Bipartite Matching (MWBM) problem.



MaxSum Matching

- The MWBM problem can be solved by various of classical Max-Flow algorithms
  - Ford-Fulkerson Algorithm, etc.



L. Kazemi, C. Shahabi. Geocrowd: enabling query answering with spatial crowdsourcing. GIS 2012.

H. To, C. Shahabi, L. Kazemi: A server-assigned spatial crowdsourcing framework. TSAS 2015

Static Matching

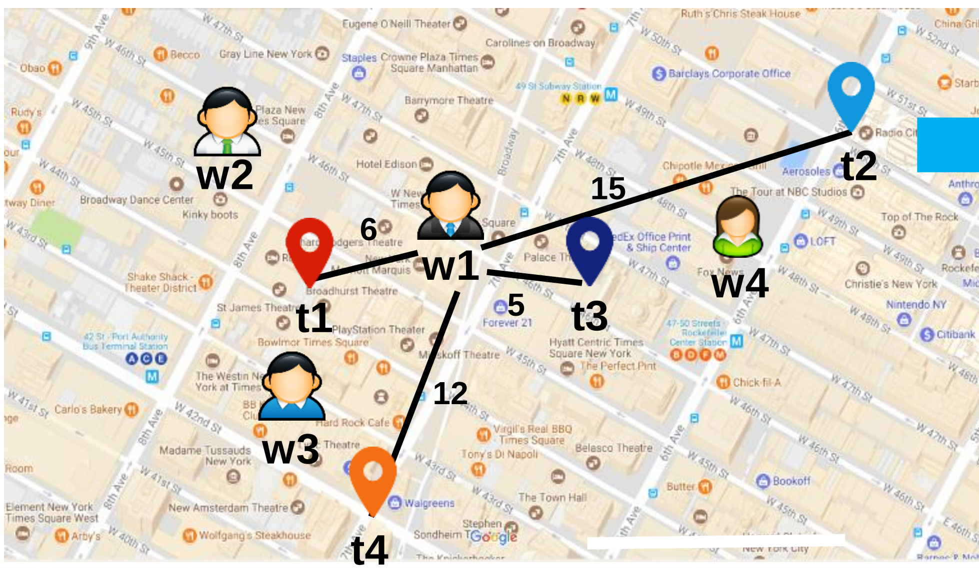
- **Problem Definition**

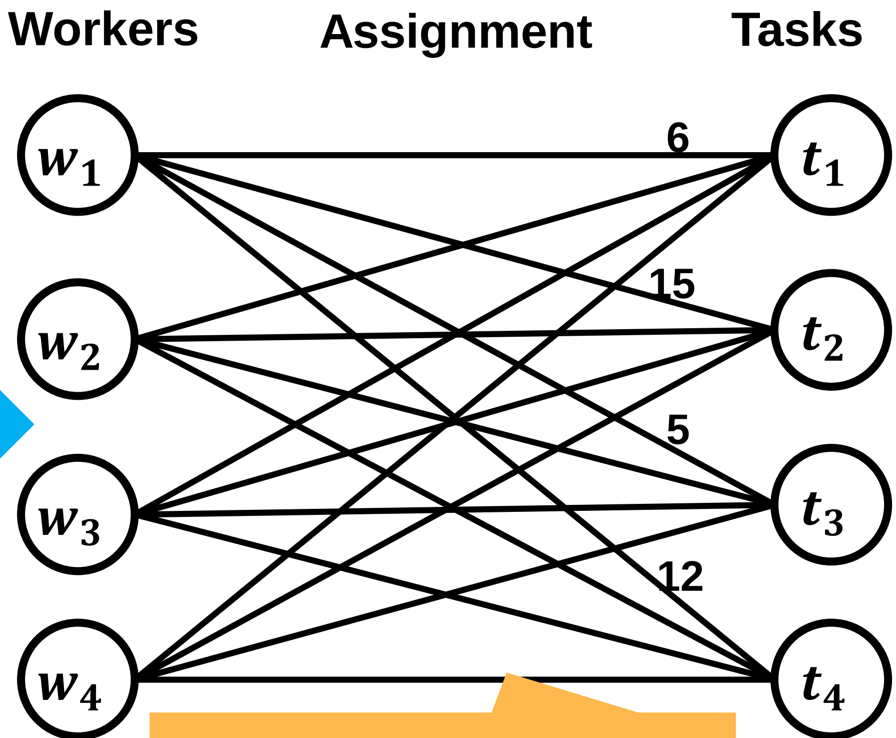
- **Existing Research**
  - **Objective 1: MaxSum Matching**
  - **Objective 2: MinSum Matching**
  - **Objective 3: Stable Marriage Matching**

MinSum Matching

- Objective 2: Minimizing the total distance cost of the maximum-cardinality matching (prefect matching)

The problem of static matching with Objective 2 equals to the Maximum Flow with Minimum Cost problem.



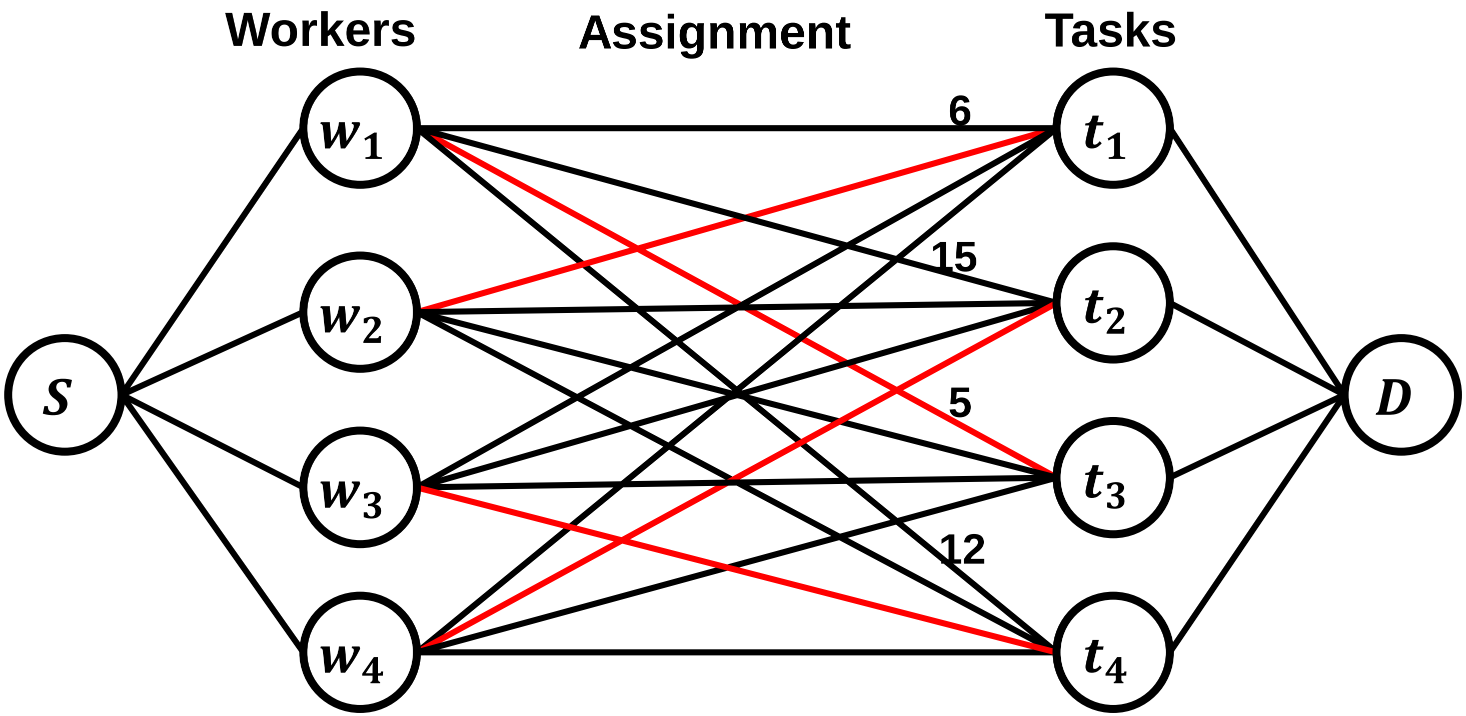


Weight: distance

L. U, M. Yiu, K. Mouratidis, N. Mamoulis. Capacity constrained assignment in spatial databases. SIGMOD 2008.

MinSum Matching

- The MimSum problem can be solved by a serial of classical assignment algorithms
  - Successive Shortest Path Algorithm (SSPA)
  - Leverage index and I/O optimization techniques



L. U, M. Yiu, K. Mouratidis, N. Mamoulis. Capacity constrained assignment in spatial databases. SIGMOD 2008.

Static Matching

- Problem Definition

- Existing Research
  - Objective 1: MaxSum Matching
  - Objective 2: MinSum Matching
  - Objective 3: Stable Marriage Matching

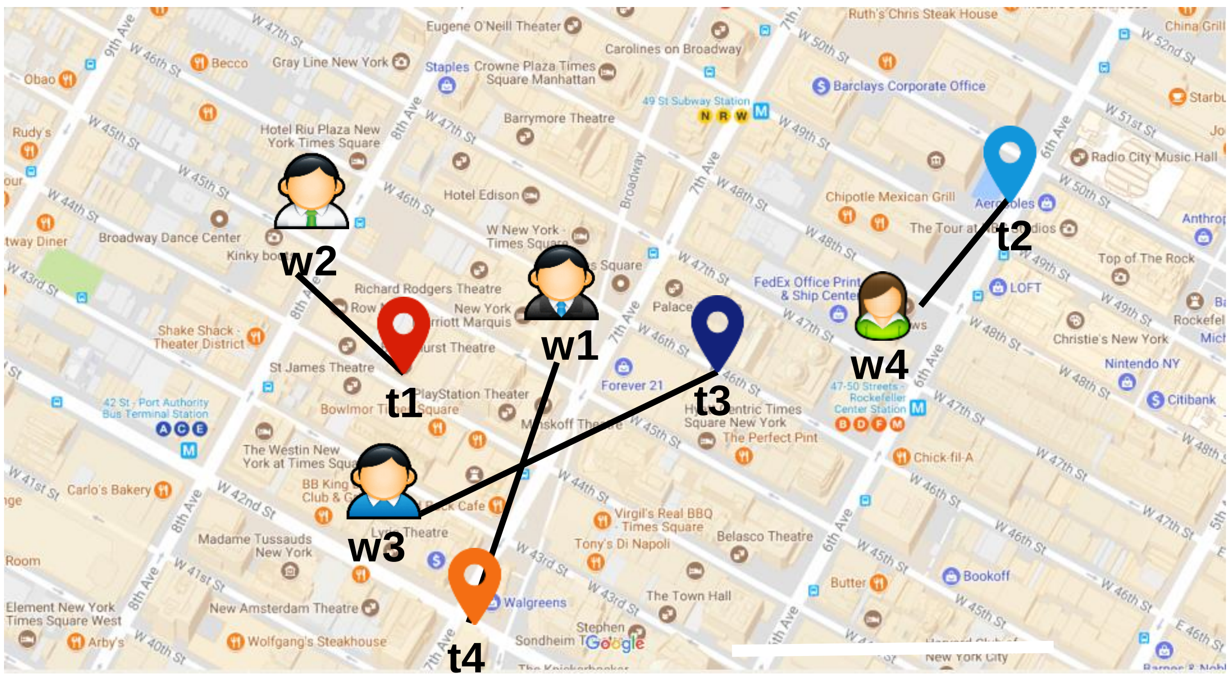
Stable Marriage Matching

- **Stable Marriage: Given the preference lists of n women and n men, the problem is to find a perfect matching with no unstable pairs**

- **Spatial Stable Marriage: Given the preference lists of n tasks and n workers based on the order of their distances, the problem is to find a perfect matching with no unstable pairs**

Stable Marriage Matching

- A worker-task pair (w, t) is an **unstable pair** if
  - |w, t| < the distance between t and t's partner in the given matching
  - |w, t| < the distance between w and w's partner in the given matching

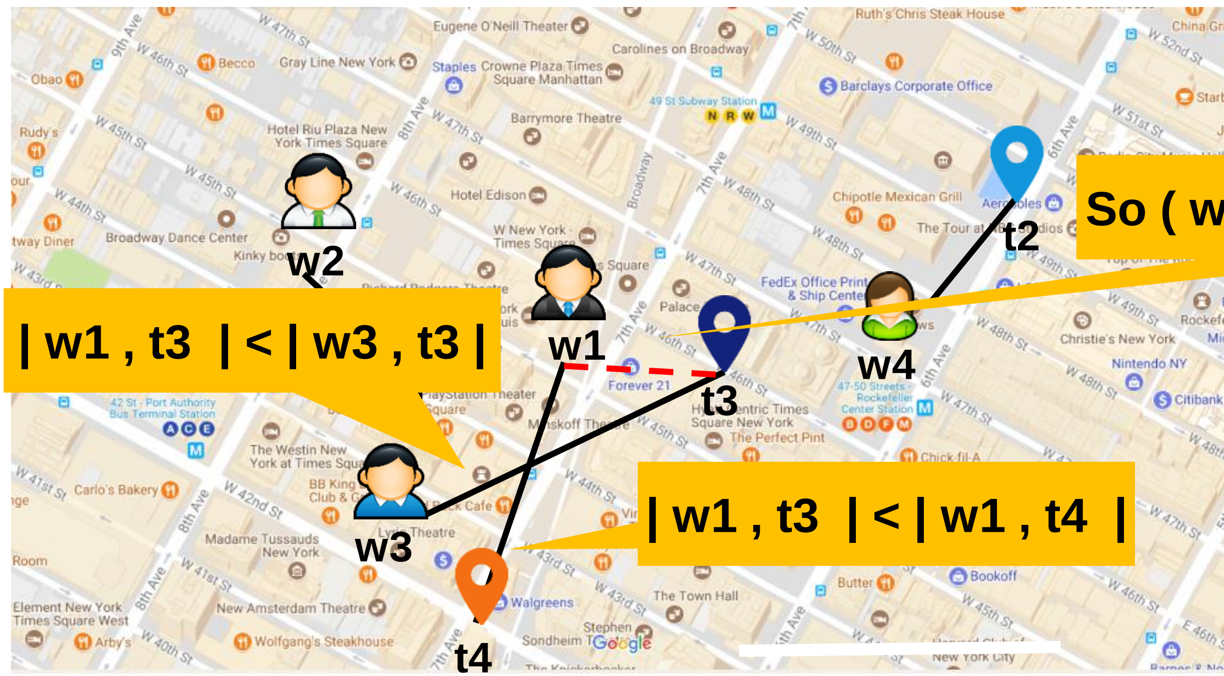


The given matching

R. Wong, Y. Tao, A. Fu, X. Xiao. On Efficient Spatial Matching. VLDB 2007.

Stable Marriage Matching

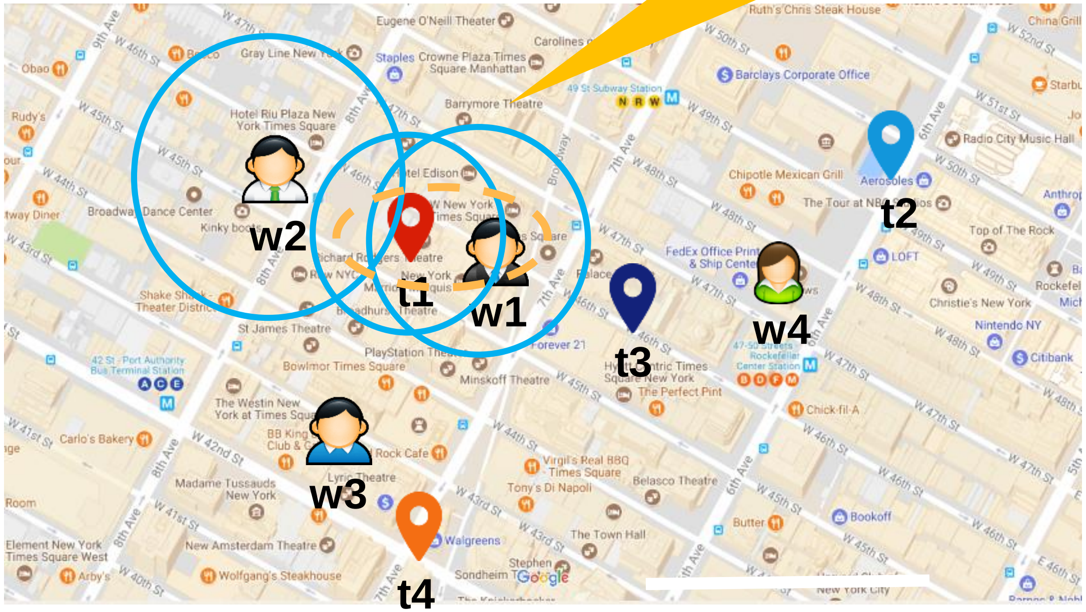
- A worker-task pair (w, t) is an **unstable pair** if
  - |w, t| < the distance between t and t's partner in the given matching
  - |w, t| < the distance between w and w's partner in the given matching

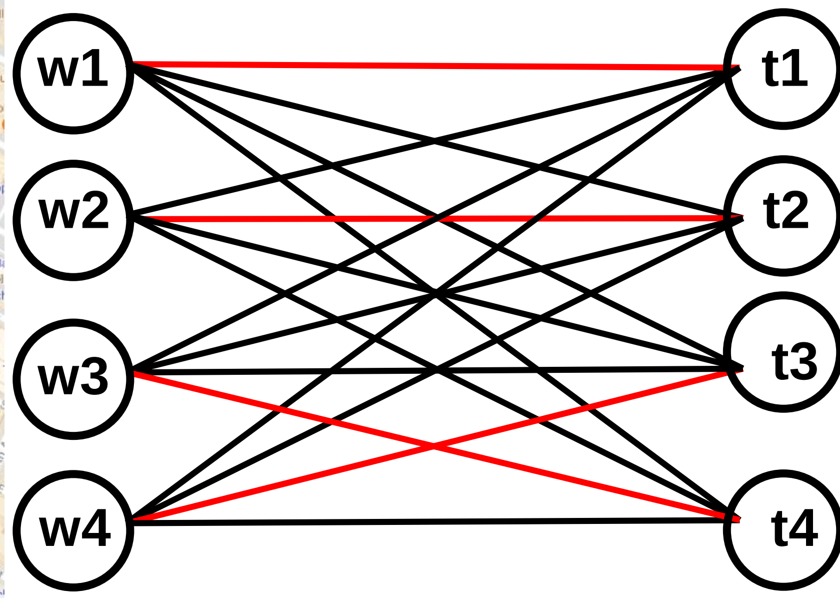


Stable Marriage Matching

- Classical Gale-Shapley algorithm is impractical for large-scale spatial data
- Design a Chain algorithm, which iteratively utilizes NN operation to enhance the efficiency

Iteratively find NN



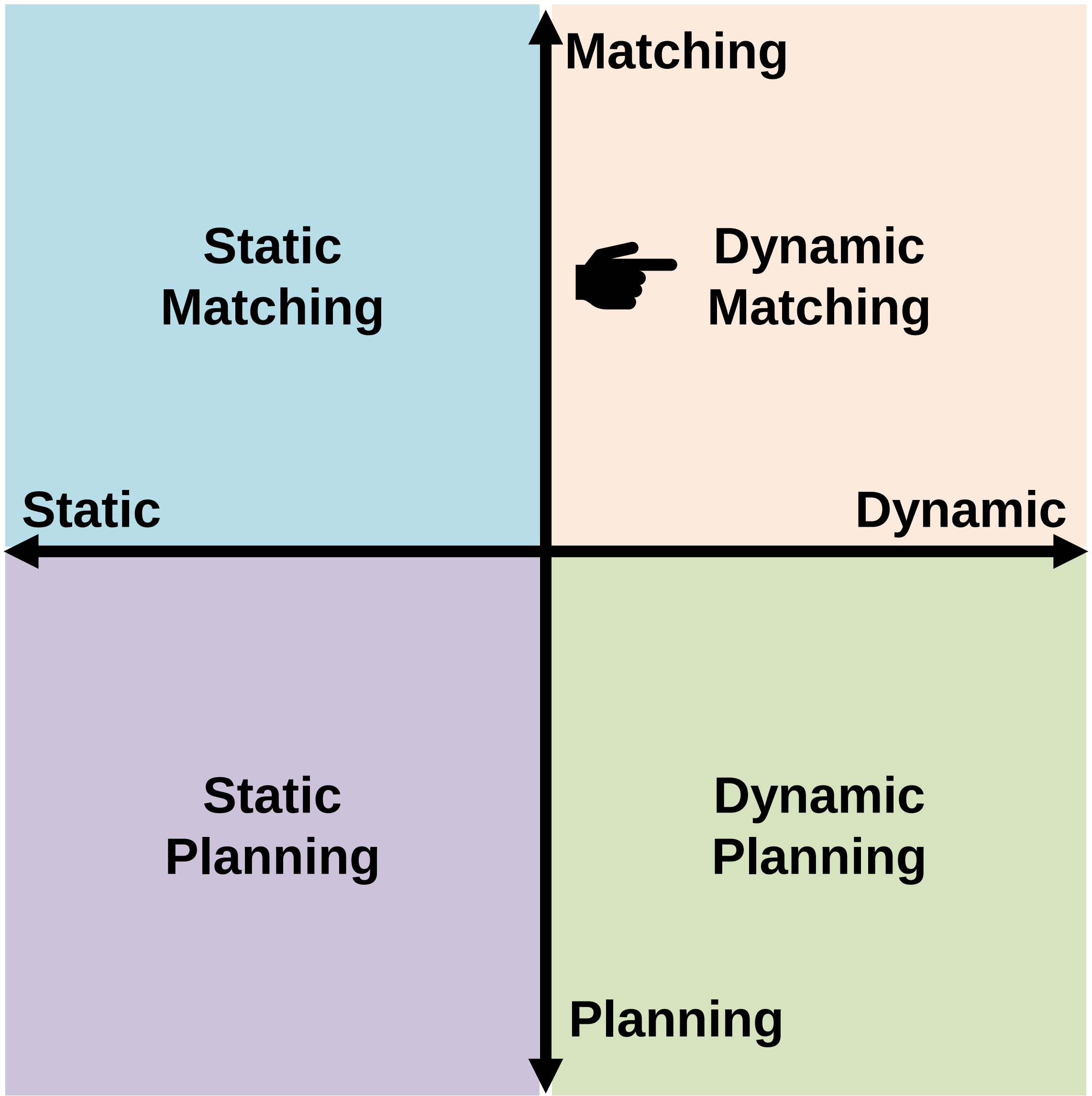


R. Wong, Y. Tao, A. Fu, X. Xiao. On Efficient Spatial Matching. VLDB 2007.

Reference: Static Matching

1. R. Burkard, M. Dell'Amico, S. Martello. Assignment problems. Society for Industrial and Applied Mathematics, 2009.
2. L. Kazemi, C. Shahabi. Geocrowd: enabling query answering with spatial crowdsourcing. In SIGSPATIAL/GIS, pages 189-198, 2012.
3. H. To, C. Shahabi, L. Kazemi. A server-assigned spatial crowdsourcing framework. TSAS, 1(1):2, 2015.
4. L. U, M. Yiu, K. Mouratidis, N. Mamoulis. Capacity constrained assignment in spatial databases. In SIGMOD, pages 15-28, 2008.
5. L. U, K. Mouratidis, M. Yiu, N. Mamoulis. Optimal matching between spatial datasets under capacity constraints. TODS 35(2): 9:1-9:44, 2010.
6. L. U, N. Mamoulis, K. Mouratidis: A fair assignment algorithm for multiple reference queries. PVLDB, 2(1): 1054-1065, 2009.
7. R. C. Wong, Y. Tao, A. Fu, X. Xiao. On efficient spatial matching. In VLDB, pages 579-590, 2007.
8. G. Dan, and R. W. Irving. The stable marriage problem: structure and algorithms. MIT press, 1989.
9. C. Long, R. Wong, P. S. Yu, M. Jiang. On optimal worst-case matching. In SIGMOD, pages 845-856, 2013.

Dynamic Matching



Dynamic Matching

- Batch-based Matching
  - Problem Definition
  - Existing Research
- Online Matching
  - Problem Definition
  - Existing Research
    - Objective 1: Online MaxSum Matching
    - Objective 2: Online MinSum Matching
- Summary

Batch-based Matching

- A set of workers and tasks dynamically appear in the physical space, this problem needs to perform a series of static matching for new arrival workers/tasks per time slot based on different objectives

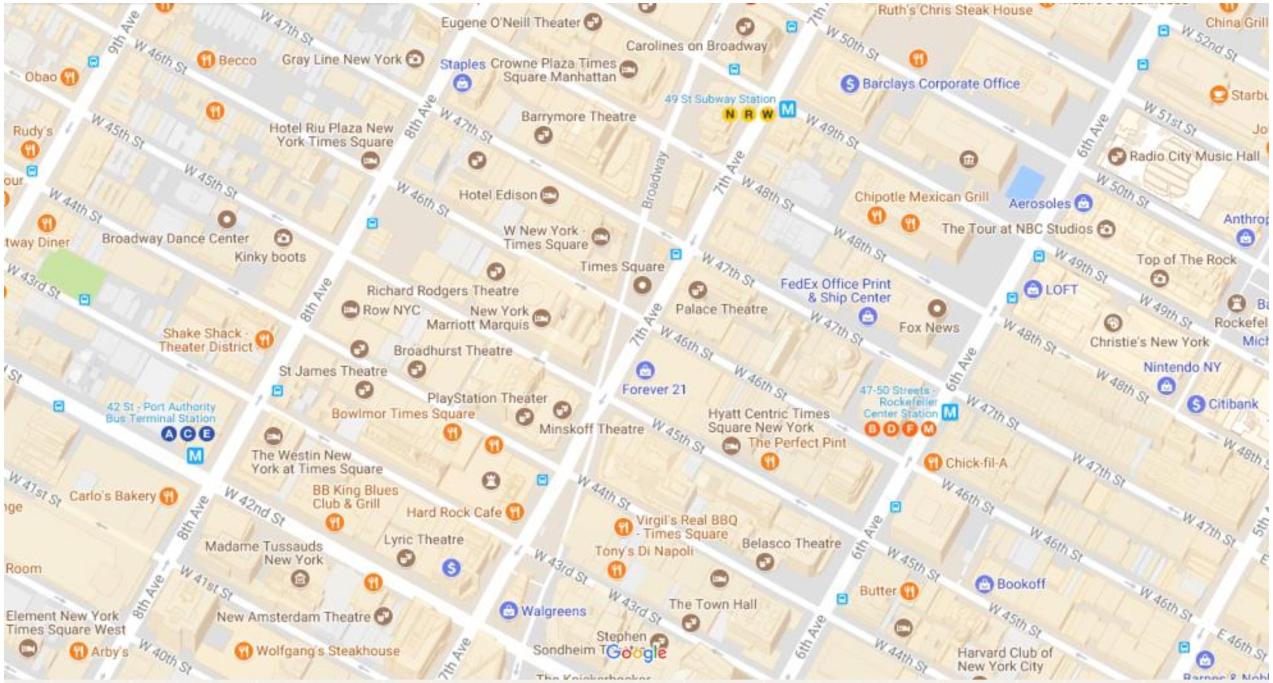
Dynamic Matching

- **Batch-based Matching**
  - **Problem Definition**
  - **Existing Research**
- **Online Matching**
  - **Problem Definition**
  - **Existing Research**
    - **Objective 1: Online MaxSum Matching**
    - **Objective 2: Online MinSum Matching**
- **Summary**

Batch-based Matching

- **Objective: Maximizing the total utility/number of matching in all batches (batch size=4min)**

Platform

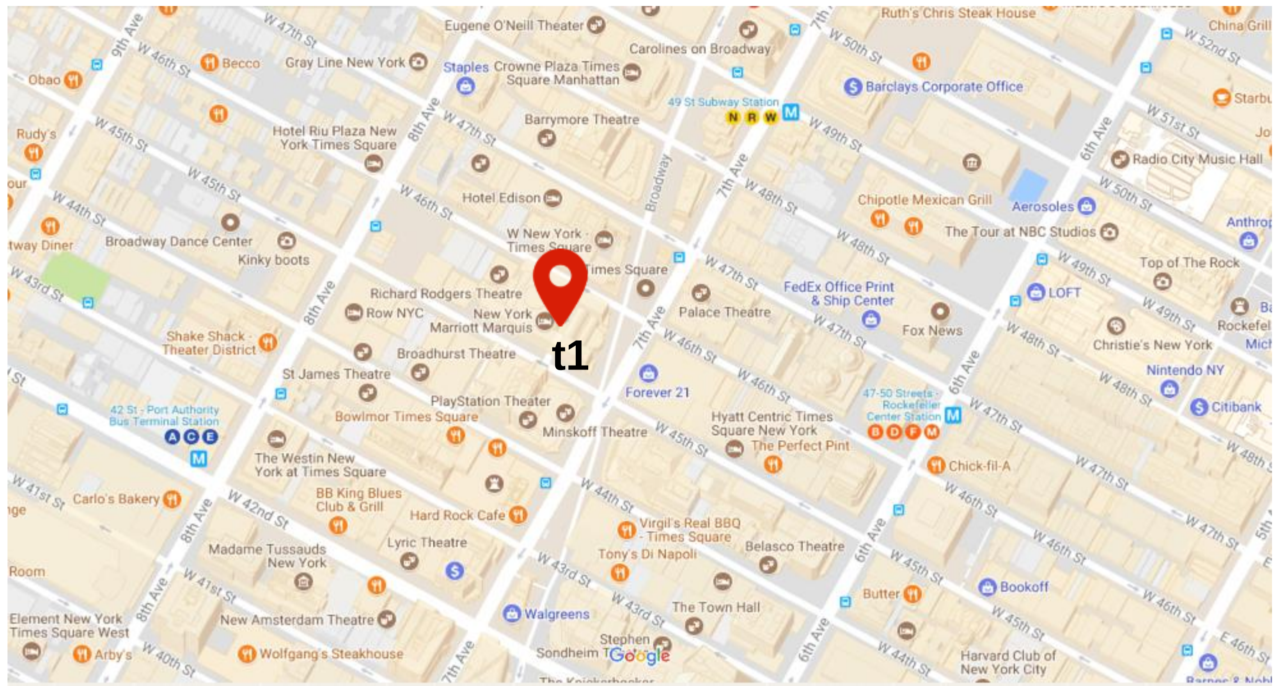


L. Kazemi, C. Shahabi. Geocrowd: enabling query answering with spatial crowdsourcing. GIS 2012.

Batch-based Matching

- **Perform a MaxSum static matching for new arrival tasks/workers per time slot (e.g., 4min)**

Platform



Platform gradually knows information (past and present) as time goes on

Arrival Time	8:00
	t_1

L. Kazemi, C. Shahabi. Geocrowd: enabling query answering with spatial crowdsourcing. GIS 2012.

Batch-based Matching

- Perform a MaxSum static matching for new arrival tasks/workers per time slot (e.g., 4min)

Platform

Platform gradually knows information (past and present) as time goes on

Arrival Time	8:00	8:01
	t_1	w_1

L. Kazemi, C. Shahabi. Geocrowd: enabling query answering with spatial crowdsourcing. GIS 2012.

Batch-based Matching

- Perform a MaxSum **static** matching for new arrival tasks/workers **per time slot** (e.g., 4min)

Platform

Platform gradually knows information (**past and present**) as time goes on

Arrival Time	8:00	8:01	8:02
	t_1	w_1	t_2

L. Kazemi, C. Shahabi. Geocrowd: enabling query answering with spatial crowdsourcing. GIS 2012.

Batch-based Matching

- Perform a MaxSum static matching for new arrival tasks/workers per time slot (e.g., 4min)

Platform

Platform gradually knows information (past and present) as time goes on

Perform an assignment at 8:04

Arrival Time	8:00	8:01	8:02
	t_1	w_1	t_2

L. Kazemi, C. Shahabi. Geocrowd: enabling query answering with spatial crowdsourcing. GIS 2012.

Existing Research

- Perform a local static assignment in each time slot
  - Exact methods: Ford-Fulkerson algorithm
  - Approximation methods: Greedy algorithm
- Aggregate the total number of the assignments

The total number of assignment is 3

Match $\{w_1, t_1\}$ at 8:04

Match $\{w_2, t_3\}$ at 8:08

Match $\{w_3, t_4\}$ at 8:12

Arrival Time	8:00	8:01	8:02	8:07	8:08	8:09	8:09	8:11
	t_1	w_1	t_2	t_3	w_2	t_4	w_3	w_4

L. Kazemi, C. Shahabi. Geocrowd: enabling query answering with spatial crowdsourcing. GIS 2012.

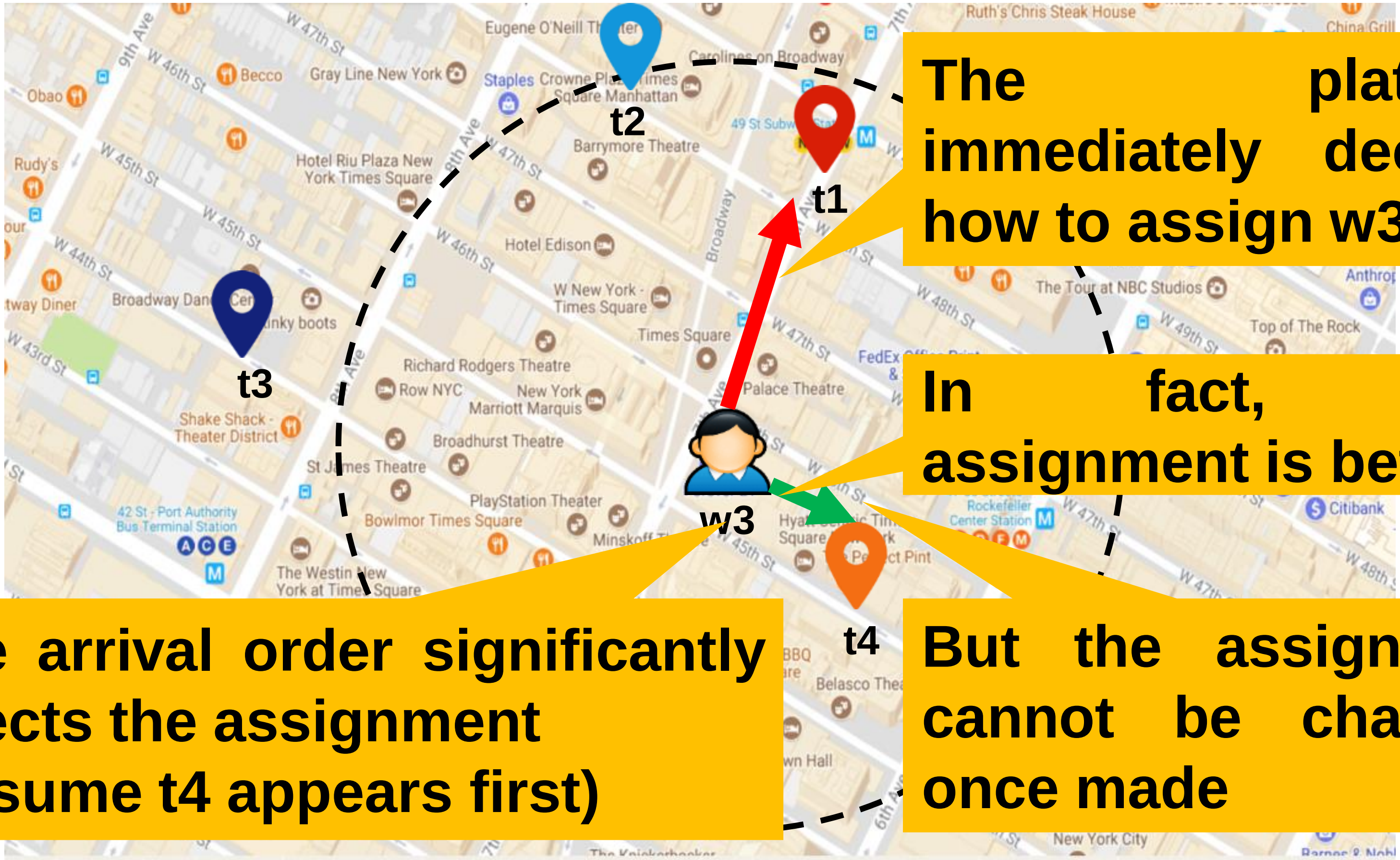
Dynamic Matching

- Batch-based Matching
  - Problem Definition
  - Existing Research
- Online Matching
  - Problem Definition
  - Existing Research
    - Objective 1: Online MaxSum Matching
    - Objective 2: Online MinSum Matching
- Summary

Online Matching

- A set of tasks/workers dynamically appear one by one in the physical space, the online matching based on a specific objective must be performed immediately and irrevocably, when a new task or worker arrives

Online Matching



The platform immediately decides how to assign w3

In fact, this assignment is better

The arrival order significantly affects the assignment (assume t4 appears first)

But the assignment cannot be changed once made

Online Matching

Evaluation of Online Matching Algorithms

Competitive Ratio (CR)

$$CR = \frac{Value\ of\ an\ online\ algorithm}{Value\ of\ the\ corresponding\ offline\ algorithm}$$

Input Models

- **Adversarial Model (Worst-Case Analysis)**

$$CR_A = min_{\forall G(T,W,U)\, and\, \forall v \in V} \frac{MaxFunc(M)}{MaxFunc(OPT)}$$

$$CR_A = max_{\forall G(T,W,U)\, and\, \forall v \in V} \frac{MinFunc(M)}{MinFunc(OPT)}$$

The worst input

The worst arrival order

- **Random Order Model (Average-Case Analysis)**

$$CR_{RO} = min_{\forall G(T,W,U)} \frac{\mathbb{E}[MaxFunc(M)]}{MaxSum(OPT)}$$

$$CR_{RO} = max_{\forall G(T,W,U)} \frac{\mathbb{E}[MinFunc(M)]}{MinSum(OPT)}$$

The expectation of the total utility of all possible arrival orders

The worst input

Dynamic Matching

- Batch-based Matching
  - Problem Definition
  - Existing Research
- Online Matching
  - Problem Definition
  - Existing Research
    - Objective 1: Online MaxSum Matching
    - Objective 2: Online MinSum Matching
- Summary

Online MaxSum Matching

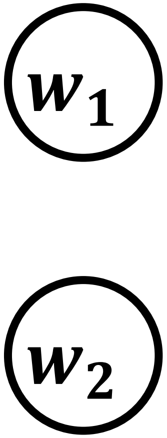
- Objective 1: Assign tasks to workers in the online scenario to maximize the total utility/number of the matching

Online MaxSum Matching

- **Baseline**
  - **Greedy**
    - **When a task/worker appears, the algorithm assigns it to a worker/task with the current maximum utility**

Online MaxSum Matching

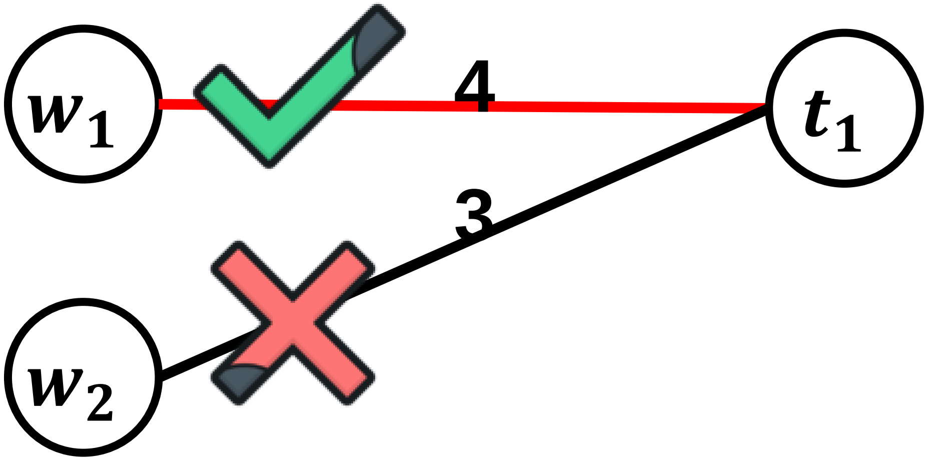
- Baseline
  - Greedy
    - When a task/worker appears, the algorithm assigns it to a worker/task with the current maximum utility



Y. Tong, J. She, B. Ding, L. Wang, L. Chen. Online mobile micro-task allocation in spatial crowdsourcing. ICDE 2016.

Online MaxSum Matching

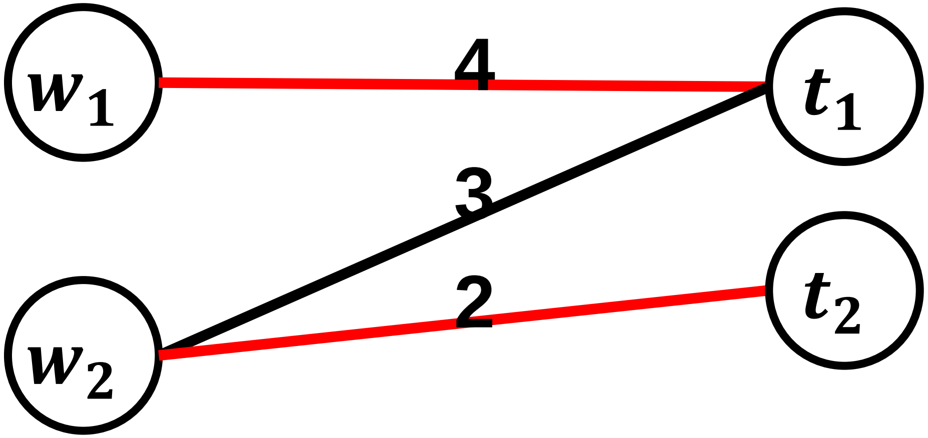
- Baseline
  - Greedy
    - When a task/worker appears, the algorithm assigns it to a worker/task with the current maximum utility



Y. Tong, J. She, B. Ding, L. Wang, L. Chen. Online mobile micro-task allocation in spatial crowdsourcing. ICDE 2016.

Online MaxSum Matching

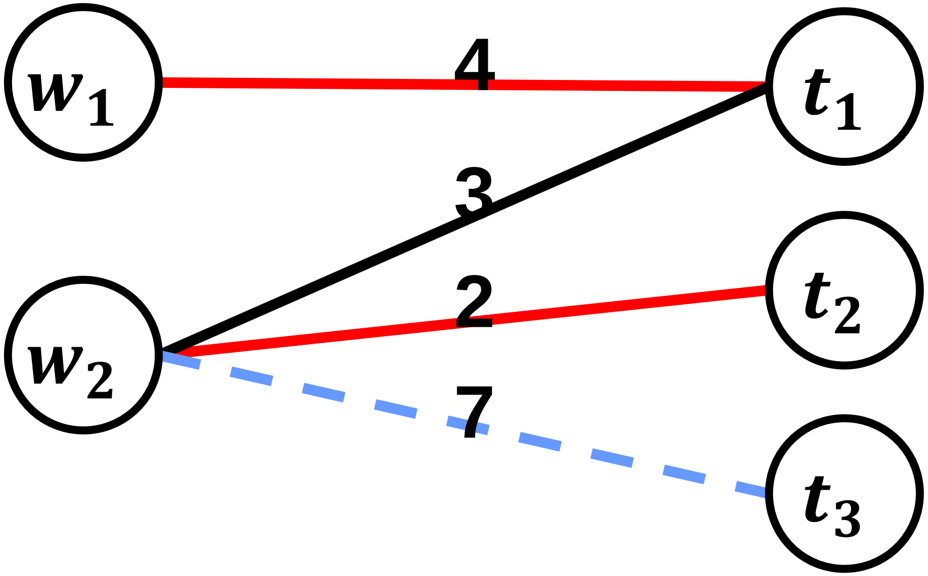
- Baseline
  - Greedy
    - When a task/worker appears, the algorithm assigns it to a worker/task with the current maximum utility



Y. Tong, J. She, B. Ding, L. Wang, L. Chen. Online mobile micro-task allocation in spatial crowdsourcing. ICDE 2016.

Online MaxSum Matching

- Baseline
  - Greedy
    - When a task/worker appears, the algorithm assigns it to a worker/task with the current maximum utility

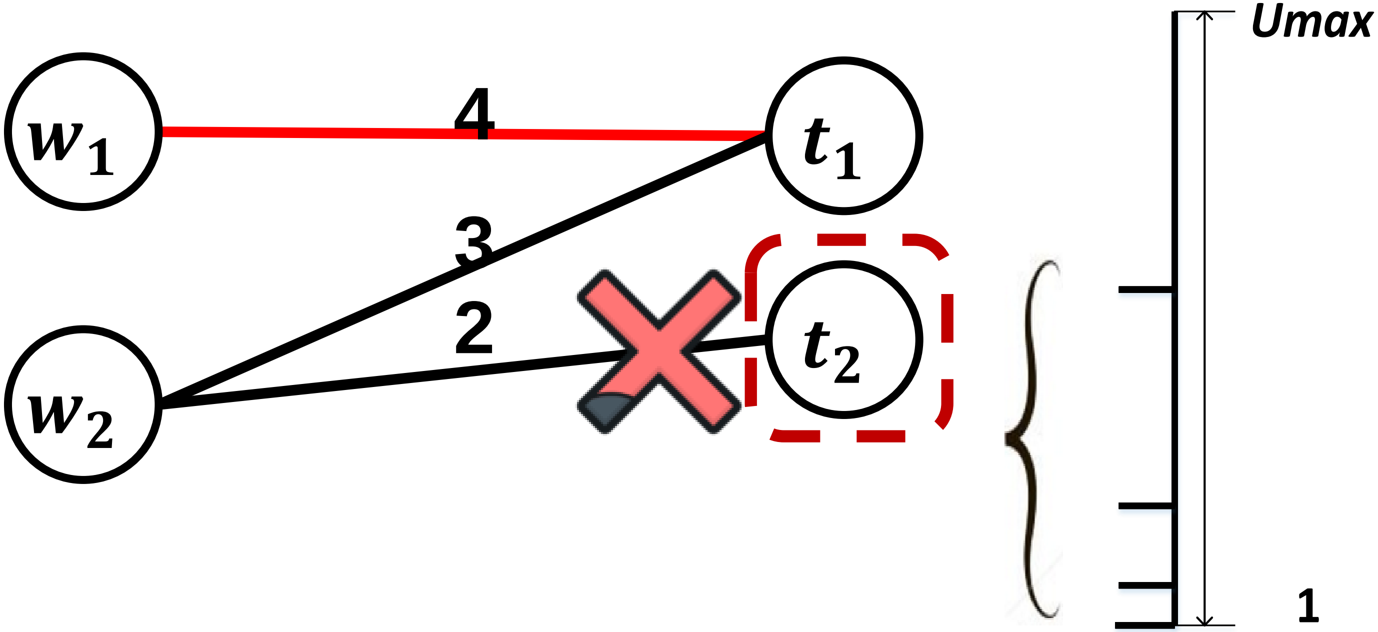


Y. Tong, J. She, B. Ding, L. Wang, L. Chen. Online mobile micro-task allocation in spatial crowdsourcing. ICDE 2016.

Online MaxSum Matching

- Improvements
  - Extended Greedy
    - Choose an integer k from 0 to $\lceil \ln(U_{max}+1) \rceil$ randomly.
    - Filter the edges with weights lower than the threshold e^k.
    - Use a greedy strategy on the remaining edges.

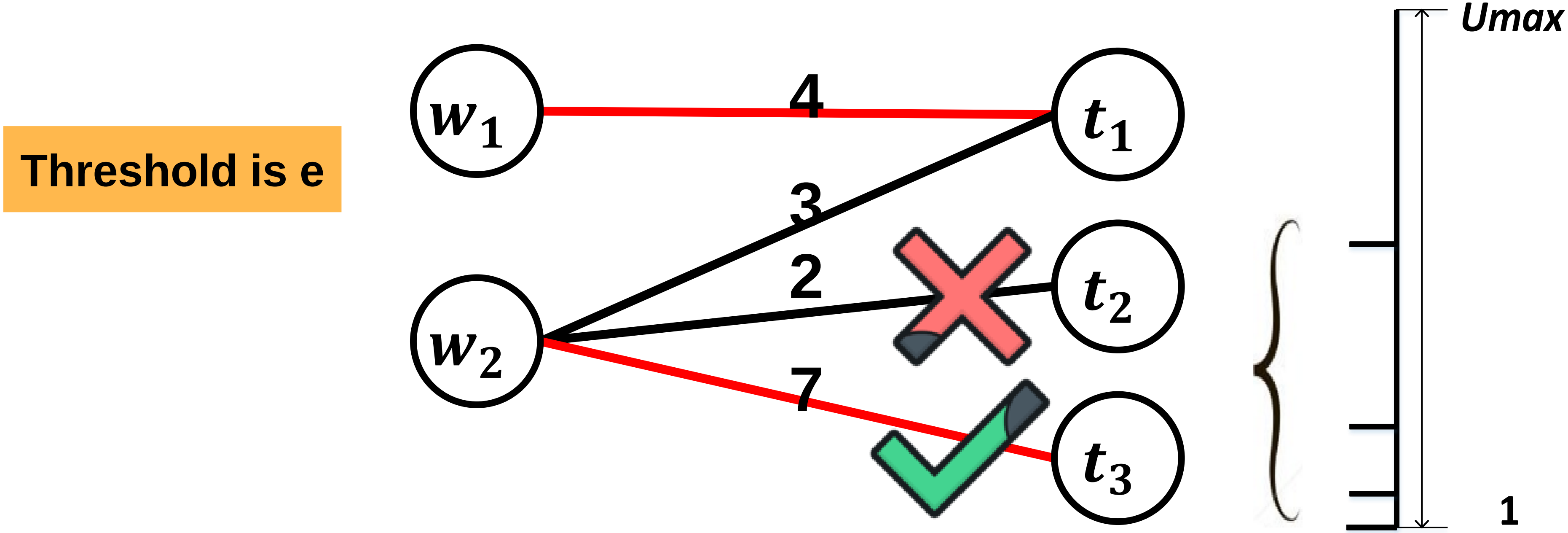
Threshold is e



Y. Tong, J. She, B. Ding, L. Wang, L. Chen. Online mobile micro-task allocation in spatial crowdsourcing. ICDE 2016.

Online MaxSum Matching

- **Improvements**
  - **Extended Greedy**
    - **Choose an integer k from 0 to $\lceil \ln(U_{max}+1) \rceil$ randomly.**
    - **Filter the edges with weights lower than the threshold e^k.**
    - **Use a greedy strategy on the remaining edges.**



Y. Tong, J. She, B. Ding, L. Wang, L. Chen. Online mobile micro-task allocation in spatial crowdsourcing. ICDE 2016.

Dynamic Matching

- Batch-based Matching
  - Problem Definition
  - Existing Research
- Online Matching
  - Problem Definition
  - Existing Research
    - Objective 1: Online MaxSum Matching
    - Objective 2: Online MinSum Matching
- Summary

Online MinSum Matching

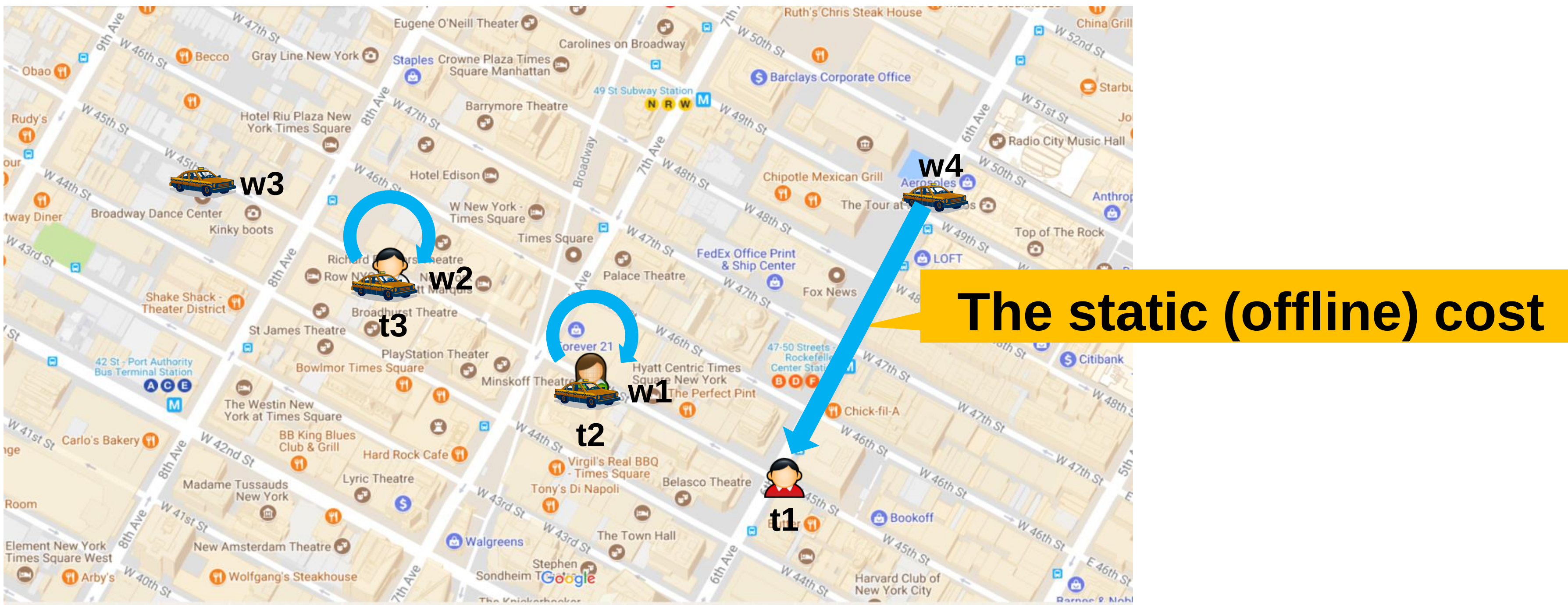
- **Objective 2: Assign tasks to workers in the online scenario to minimize the total utility**
  - **A task/worker is assigned immediately on arrival based on partial information**

Arrival Time	9:30	9:32	9:35
Order	t_1	t_2	t_3

Y. Tong, J. She, B. Ding, L. Chen, T. Wo, K. Xu. Online minimum matching in real-time spatial data: experiments and analysis. VLDB 2016.

Review: Static MinSum Matching

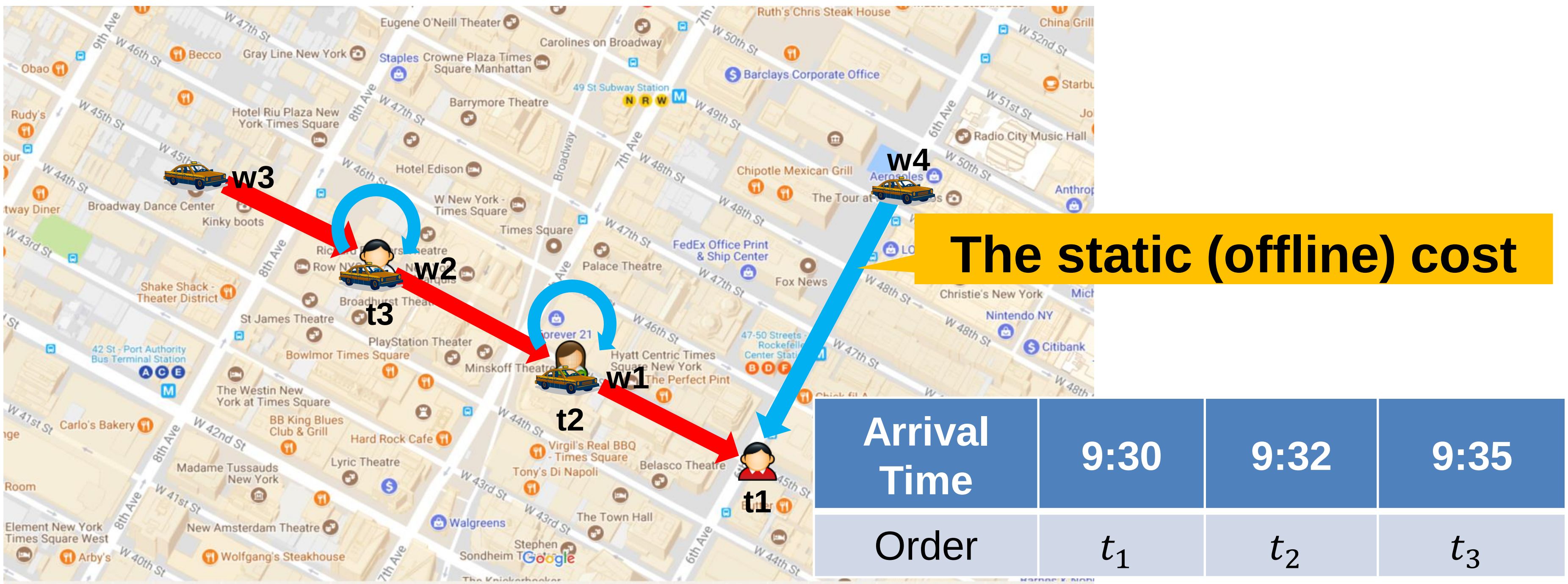
- Objective: Minimize the total distance cost of maximum-cardinality matching
  - Assume that all spatiotemporal information is known in advance (offline OPT)



Y. Tong, J. She, B. Ding, L. Chen, T. Wo, K. Xu. Online minimum matching in real-time spatial data: experiments and analysis. VLDB 2016.

Review: Static MinSum Matching

- Objective: Minimize the total distance cost of maximum-cardinality matching
  - Assume that all spatiotemporal information is known in advance (offline OPT)



Arrival Time	9:30	9:32	9:35
Order	t_1	t_2	t_3

Y. Tong, J. She, B. Ding, L. Chen, T. Wo, K. Xu. Online minimum matching in real-time spatial data: experiments and analysis. VLDB 2016.

Online MinSum Matching

Four representative algorithms

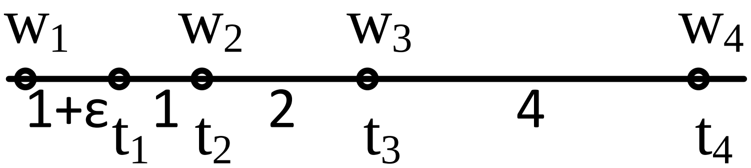
Algorithms	Time Complexity per Each Arrival Vertex	Randomization	Data Structure	Competitive Ratio (Worst-Case Analysis)
Greedy [SODA'1991]	O(k)	Deterministic	No	$O(2^k)$
Permutation [SODA'1991]	$O(k^3)$	Deterministic	No	O(2k-1)
HST-Greedy [SODA 2006]	O(k)	Randomized	HST	$O(\log^3 k)$
HST-Reassignment [ESA 2007]	$O(k^2)$	Randomized	HST	$O(\log^2 k)$

Hierarchically Separated Tree

Is the greedy algorithm really the worst?

Online MinSum Matching

- **Greedy Revisited**

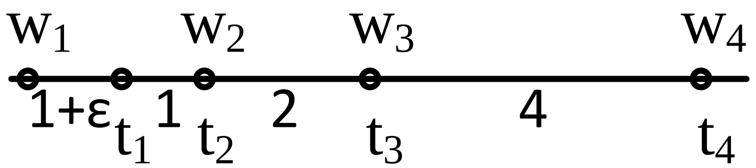


(a)Locations

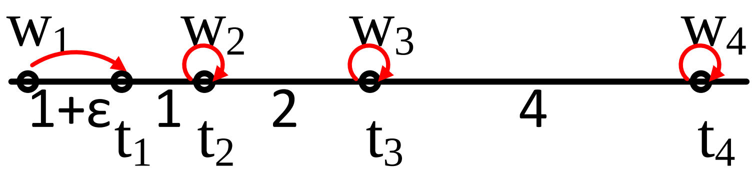
Y. Tong, J. She, B. Ding, L. Chen, T. Wo, K. Xu. Online minimum matching in real-time spatial data: experiments and analysis. VLDB 2016.

Online MinSum Matching

- **Greedy Revisited**



(a)Locations

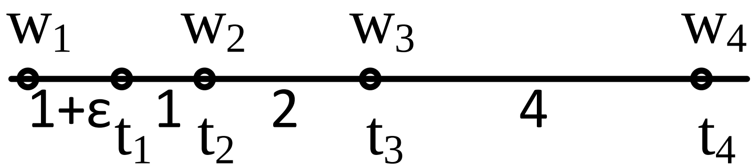


(b)Matching of Offline OPT

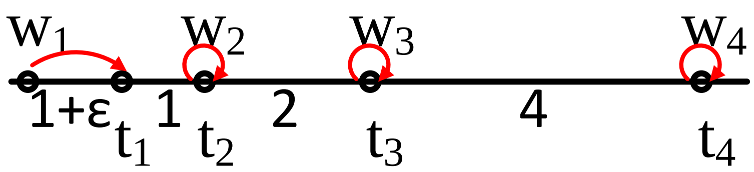
Y. Tong, J. She, B. Ding, L. Chen, T. Wo, K. Xu. Online minimum matching in real-time spatial data: experiments and analysis. VLDB 2016.

Online MinSum Matching

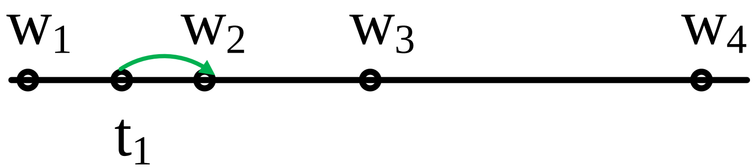
- **Greedy Revisited**



(a)Locations



(b)Matching of Offline OPT



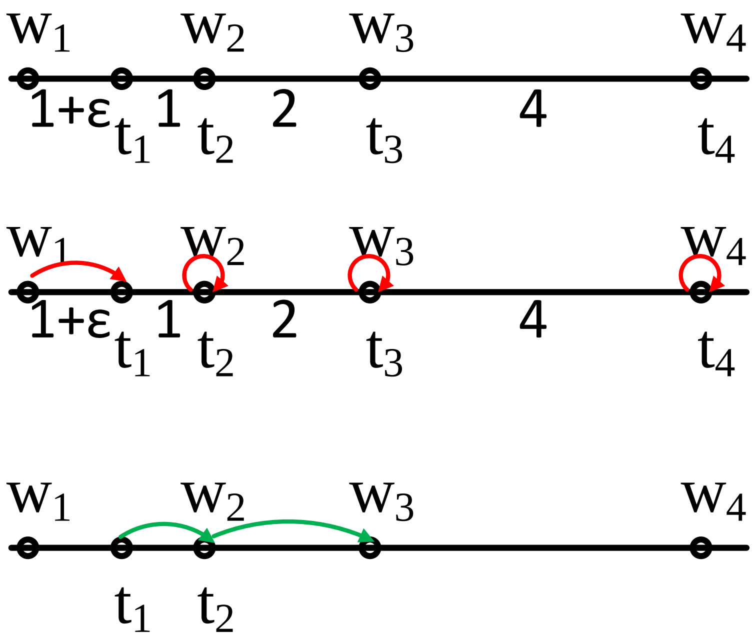
(c)Matching of Worst-Case Greedy

When t_1 appears, w_2 will be assigned to t_1 by Greedy

Y. Tong, J. She, B. Ding, L. Chen, T. Wo, K. Xu. Online minimum matching in real-time spatial data: experiments and analysis. VLDB 2016.

Online MinSum Matching

Greedy Revisited



(a)Locations

(b)Matching of Offline OPT

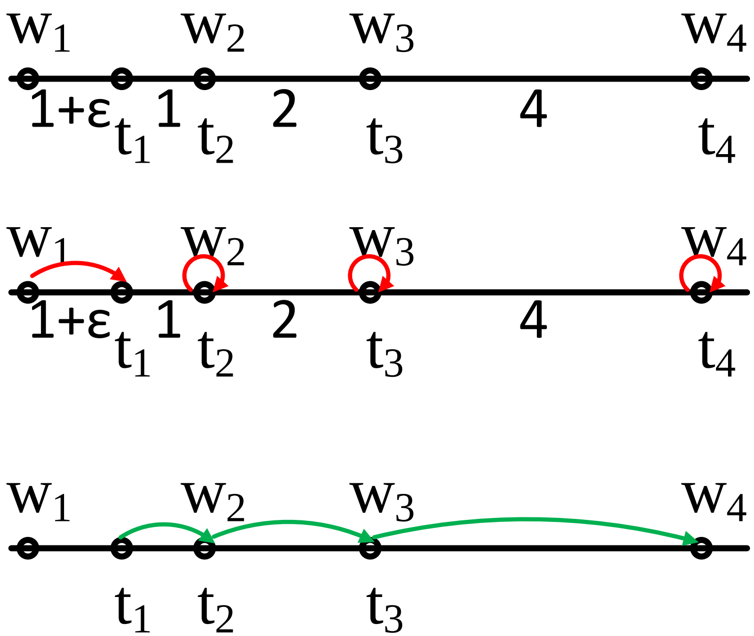
(c)Matching of Worst-Case Greedy

When t_2 appears, w_3 will be assigned to t_2 by Greedy

Y. Tong, J. She, B. Ding, L. Chen, T. Wo, K. Xu. Online minimum matching in real-time spatial data: experiments and analysis. VLDB 2016.

Online MinSum Matching

- **Greedy Revisited**



(a)Locations

(b)Matching of Offline OPT

(c)Matching of Worst-Case Greedy

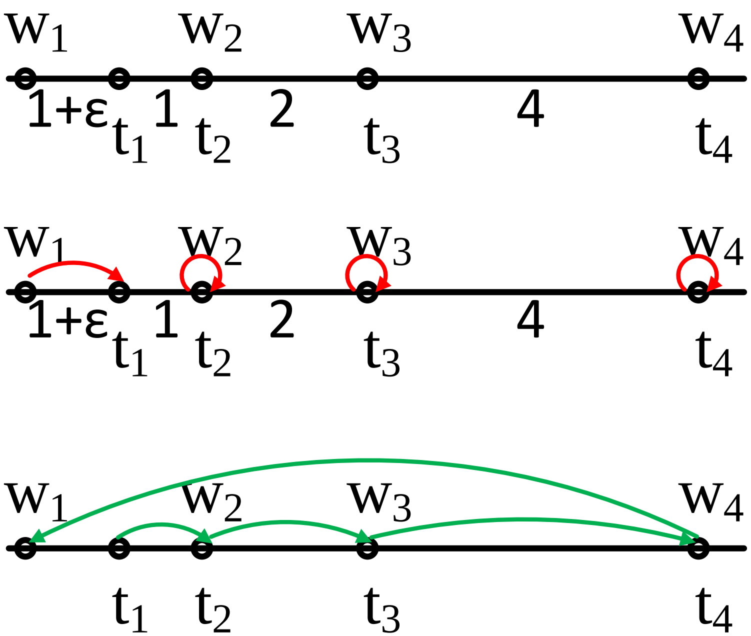
When t_3 appears, w_4 will be assigned to t_3 by Greedy

Y. Tong, J. She, B. Ding, L. Chen, T. Wo, K. Xu. Online minimum matching in real-time spatial data: experiments and analysis. VLDB 2016.

Online MinSum Matching

- **Greedy Revisited**

Competitive Ratio CR_{RO}=3.195



(a)Locations

(b)Matching of Offline OPT

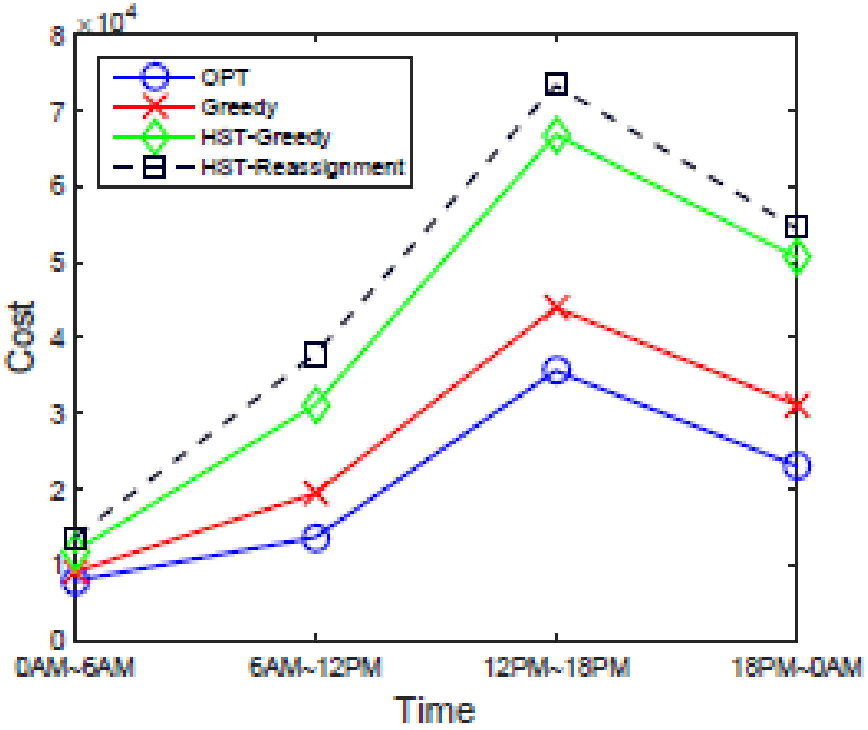
(c)Matching of Worst-Case Greedy

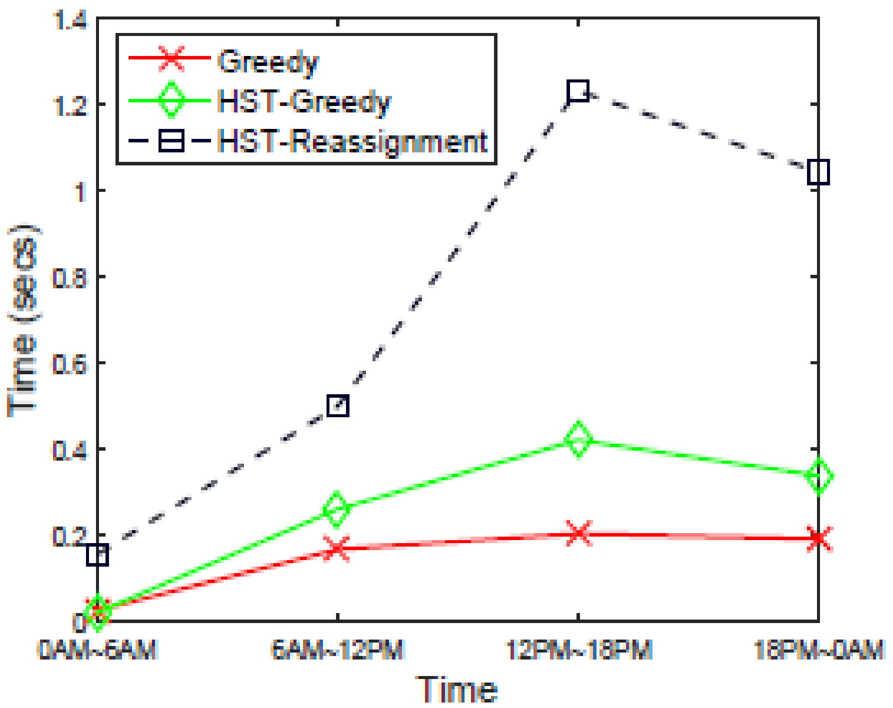
When t_4 appears, w_1 has to be assigned to t_4 by Greedy

The probability that the worst-case happens is only 1 over the factorial of k, which is the number of workers (taxis)

Y. Tong, J. She, B. Ding, L. Chen, T. Wo, K. Xu. Online minimum matching in real-time spatial data: experiments and analysis. VLDB 2016.

Online MinSum Matching

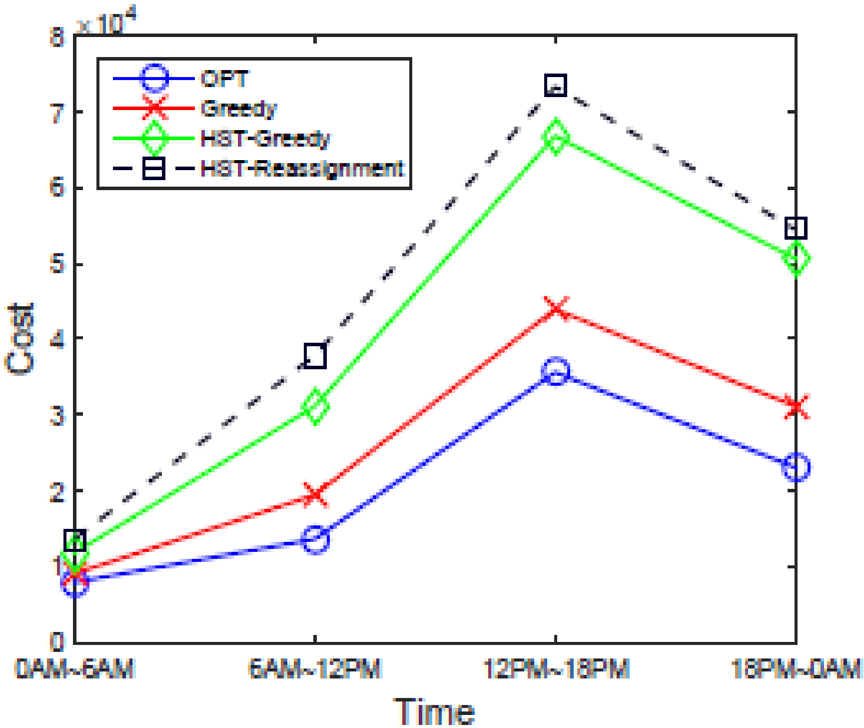


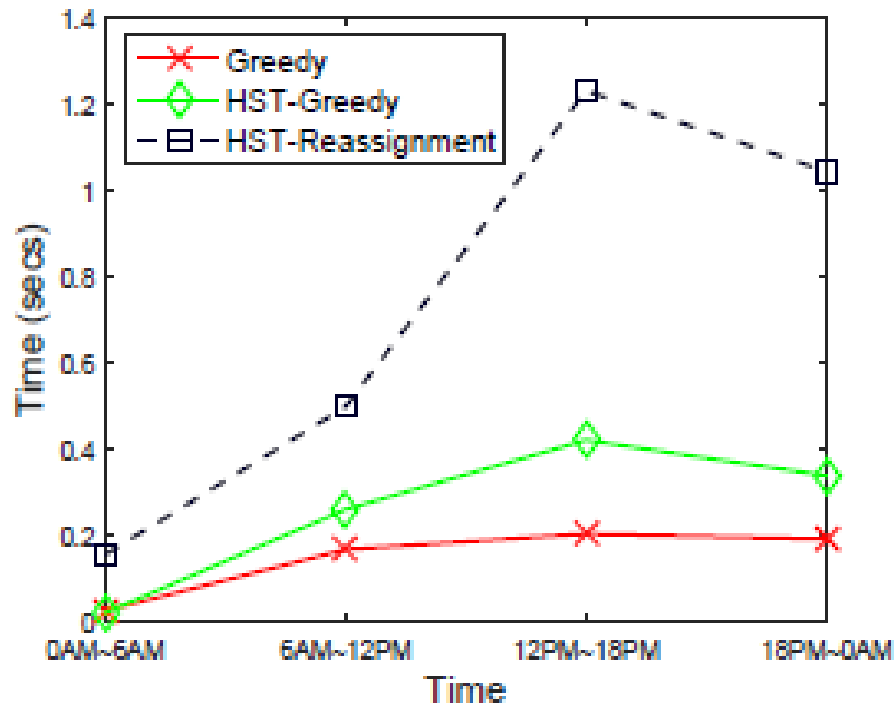


Their extensive experiments on real datasets and synthetic datasets shows that the greedy algorithm is not bad and always outperforms other state-of-the-art online algorithms!

Y. Tong, J. She, B. Ding, L. Chen, T. Wo, K. Xu. Online minimum matching in real-time spatial data: experiments and analysis. VLDB 2016.

Online MinSum Matching





Open question: the average-case competitive ratio under the random order model of Greedy for the online MinSum matching problem should be constant or O(k)?

Y. Tong, J. She, B. Ding, L. Chen, T. Wo, K. Xu. Online minimum matching in real-time spatial data: experiments and analysis. VLDB 2016.

Dynamic Matching

- Batch-based Matching
  - Problem Definition
  - Existing Research
- Online Matching
  - Problem Definition
  - Existing Research
    - Objective 1: Online MaxSum Matching
    - Objective 2: Online MinSum Matching
- Summary

Summary of Matching Models

	Static/Dynamic Scenario	Effect of Spatial Factors	Optimization Target
U et al. SIGMOD08	Static	Target	MinSum Distance
Wong et al. VLDB07	Static	Target	Spatial Stable Marriage
Long et al. SIGMOD13	Static	Target	MinMax Distance
She et al. ICDE15	Static	Constraint	Conflict-based MaxSum Utility
Alfarrarjeh et al. MDM15	Static	Target	MinSum Distance in Distributed SC
Kazemi et al. GIS12	Batch-based	Constraint	MaxSum Number
To et al. TSAS15	Batch-based	Constraint	MaxSum Utility
Liu et al. ICDE16	Batch-based	Constraint	MaxSum Utility
Cheng et al. ICDE17	Batch-based	Constraint	Prediction-based MaxSum Utility

Summary of Matching Models

	Static/Dynamic Scenario	Effect of Spatial Factors	Optimization Target
Tong et al. ICDE16	Online	Constraint	MaxSum Utility
Tong et al. VLDB16	Online	Target	MinSum Distance
She et al. TKDE16	Online	Constraint	Conflict-based MaxSum Utility
Song et al. ICDE 17	Online	Constraint	Trichromatic MaxSum Utility
Tong et al. VLDB 17	Online	Constraint	Prediction-based MaxSum Number
Hassan et al. UIC14	Online	Constraint	MaxSum Utility
To el al. Percom 16	Static and Online	Constraint	MaxSum Utility

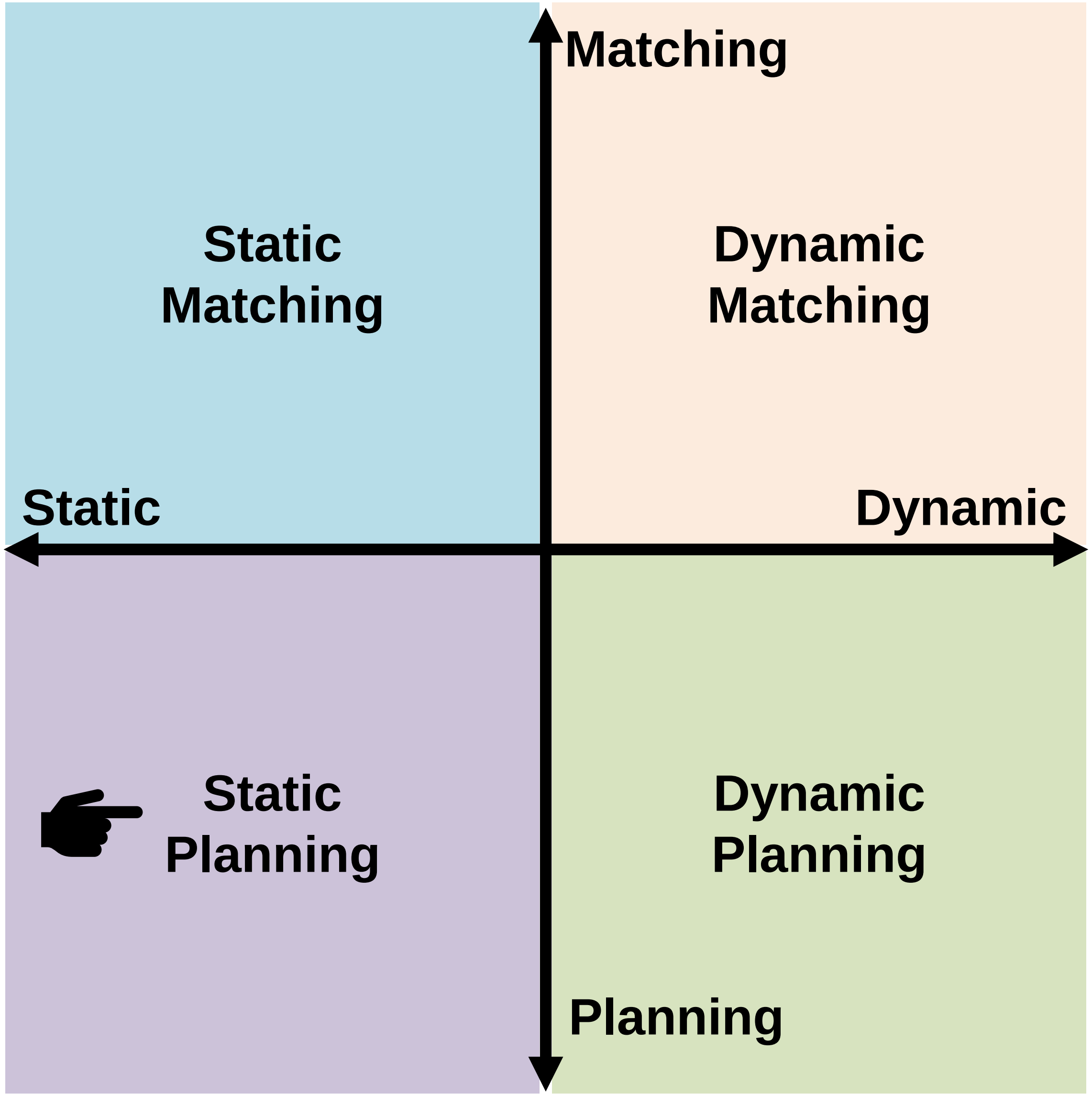
Reference: Dynamic Matching

1. L. Kazemi, C. Shahabi. Geocrowd: enabling query answering with spatial crowdsourcing. In GIS, pages 189-198, 2012.
2. H. To, C. Shahabi, L. Kazemi. A server-assigned spatial crowdsourcing framework. TSAS, 1(1):2, 2015.
3. Y. Tong, J. She, B. Ding, L. Wang, L. Chen. Online mobile micro-task allocation in spatial crowdsourcing. In ICDE, pages 49-60, 2016.
4. Y. Tong, J. She, B. Ding, L. Chen, T. Wo, K. Xu. Online minimum matching in real-time spatial data: Experiments and analysis. PVLDB 9(12): 1053-1064, 2016.
5. Y. Tong, L. Wang, Z. Zhou, B. Ding, L. Chen, J. Ye, K. Xu. Flexible online task assignment in real-time spatial data. PVLDB, 10(11): 1334-1345, 2017.
6. P. Cheng, X. Lian, L. Chen, C. Shahabi. Prediction-based task assignment in spatial crowdsourcing. In ICDE, pages 997-1008, 2017.
7. J. She, Y. Tong, L. Chen, C. Cao. Conflict-aware event-participant arrangement. In ICDE, pages 735-746, 2015.
8. J. She, Y. Tong, L. Chen, C. Cao. Conflict-aware event-participant arrangement and its variant for online setting. TKDE, 28(9): 2281-2295, 2016.
9. L. Zheng, L. Chen. Mutual benefit aware task assignment in a bipartite labor market. In ICDE, pages 73-84, 2016.

Reference: Dynamic Matching

10. T. Song, Y. Tong, L. Wang, J. She, B. Yao, L. Chen, K. Xu. Trichromatic Online Matching in Real-Time Spatial Crowdsourcing. In ICDE, pages 1009-1020, 2017.
11. H. To. Task assignment in spatial crowdsourcing: challenges and approaches. In SIGSPATIAL PhD Symposium, pages 1:1-4, 2016.
12. H. To, L. Fan, L. Tran, C. Shahabi. Real-time task assignment in hyperlocal spatial crowdsourcing under budget constraints. In PerCom, pages 1-8, 2016.
13. L. Tran, H. To, L. Fan, C. Shahabi. A real-time framework for task assignment in hyperlocal spatial crowdsourcing. TIST, 2017
14. A. Alfarrarjeh, T. Emrich, C. Shahabi. Scalable Spatial Crowdsourcing: A Study of Distributed Algorithms. In MDM, pages 134-144, 2015.
15. U. ul Hassan, E. Curry. A Multi-armed bandit approach to online spatial task assignment. In UIC, pages 212-219, 2014.
16. U. ul Hassan, E. Curry. Efficient task assignment for spatial crowdsourcing: A combinatorial fractional optimization approach with semi-bandit learning. ESA, 58: 36-56, 2016.

Static Planning



Static Planning

- **Problem Definition**

- **Background**
  - **Orienteering Problem (OP)**

- **Existing Research**
  - **One-Worker-To-Many-Task Planning**
  - **Many-Worker-To-Many-Task Planning**

Static Planning

- Problem Definition
  - Given a set of spatial tasks with some constraints (e.g., deadline), and the spatiotemporal information of tasks is known, the problem is to find a proper route of tasks for workers

Static Planning

- **Problem Definition**

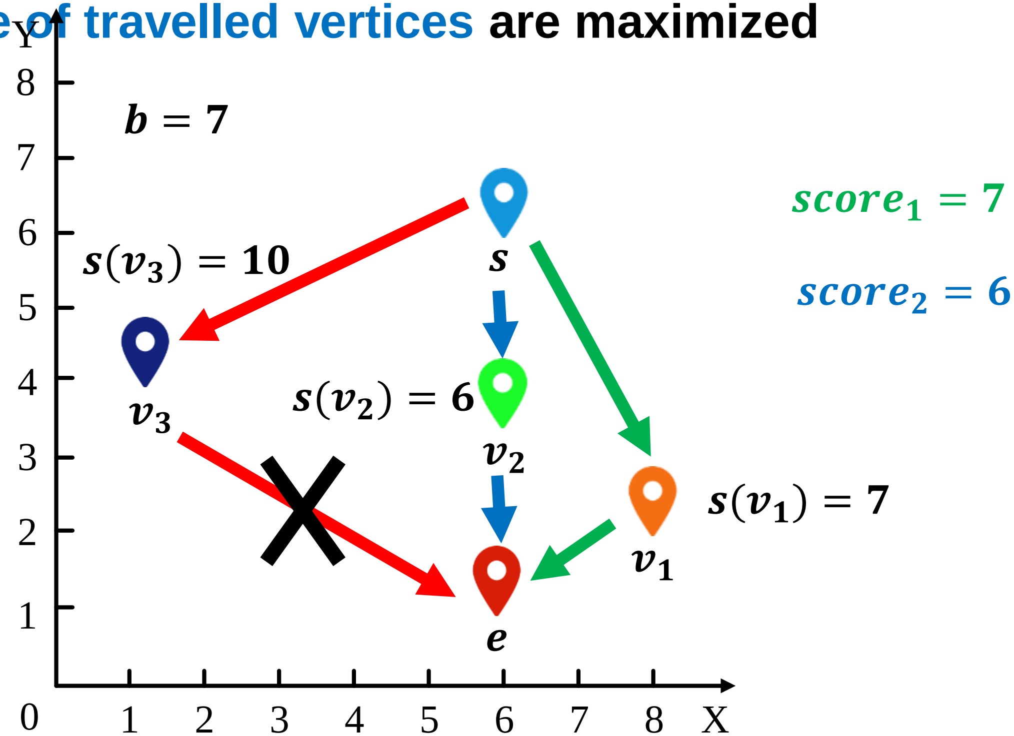
- **Background**
  - **Orienteering Problem (OP)**

- **Existing Research**
  - **One-Worker-To-Many-Tasks Planning**
  - **Many-Worker-To-Many-Tasks Planning**

Static Planning

- **Orienteering Problem (OP)**
  - **Given a set of vertices, each associated with a score, a travel budget b, a start vertex s, an end vertex e, and the distance matrix among them, the problem is find a route from s to e such that:**
    - **The total travel distance is no more than the travel budget b**
    - **The total score of travelled vertices are maximized**

A variant of the Traveling Salesman Problem (TSP)



Static Planning

- Problem Definition
- Background
  - Orienteering Problem (OP)
- Existing Research
  - One-Worker-To-Many-Task Planning
  - Many-Worker-To-Many-Task Planning

One-Worker-To-Many-Tasks Planning

- Objective: Find a route for one worker to maximize the number of the tasks satisfying travel/time budgets
- Difference from Orienteering Problem
  - The utility score of each task is zero or one
  - The end vertex of the path is not given

One-Worker-To-Many-Tasks Planning

- **Heuristic Methods**
  - **Nearest Neighbor first**
  - **Limitation first (e.g., time expiration)**

D. Deng, C. Shahabi, U. Demiryurek. Maximizing the number of worker's self-selected tasks in spatial crowdsourcing. GIS 2013.

One-Worker-To-Many-Tasks Planning

- Heuristic Methods
  - Nearest Neighbor first
  - Limitation first (e.g., time expiration)

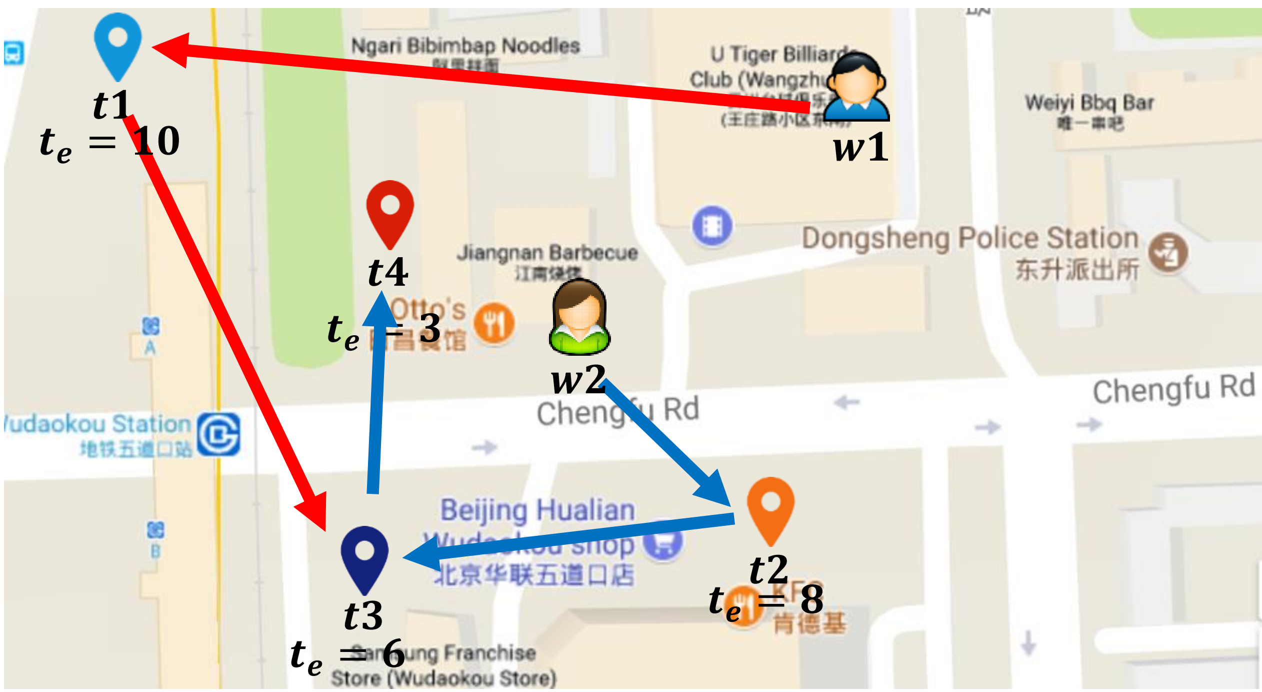
D. Deng, C. Shahabi, U. Demiryurek. Maximizing the number of worker's self-selected tasks in spatial crowdsourcing. GIS 2013.

Static Planning

- Problem Definition
- Background
  - Orienteering Problem (OP)
- Existing Research
  - One-Worker-To-Many-Task Planning
  - Many-Worker-To-Many-Task Planning

Many-Worker-To-Many-Tasks Planning

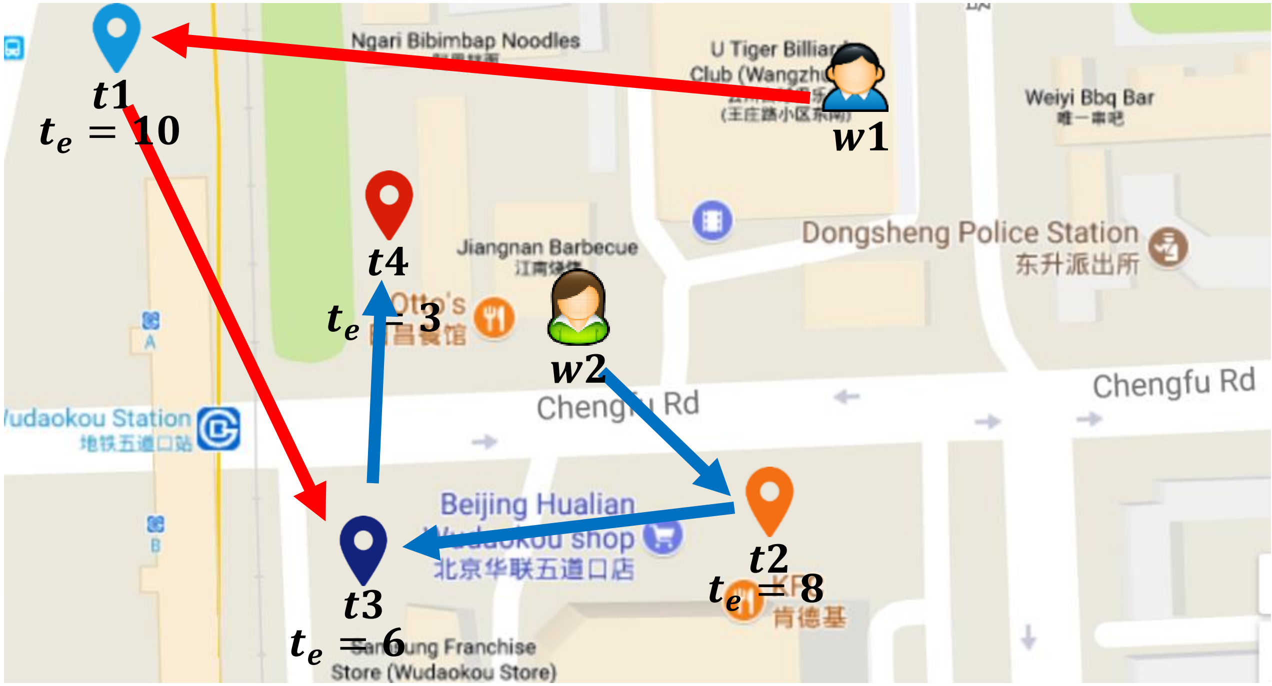
- **Objective: Find routes for all workers to maximize the total utility/number satisfying travel/time budgets**



J. She, Y. Tong, L. Chen. Utility-aware event-participant planning. SIGMOD 2015
D. Deng, C. Shahabi, L. Zhu. Task matching and scheduling for multiple workers in spatial crowdsourcing. GIS 2015.

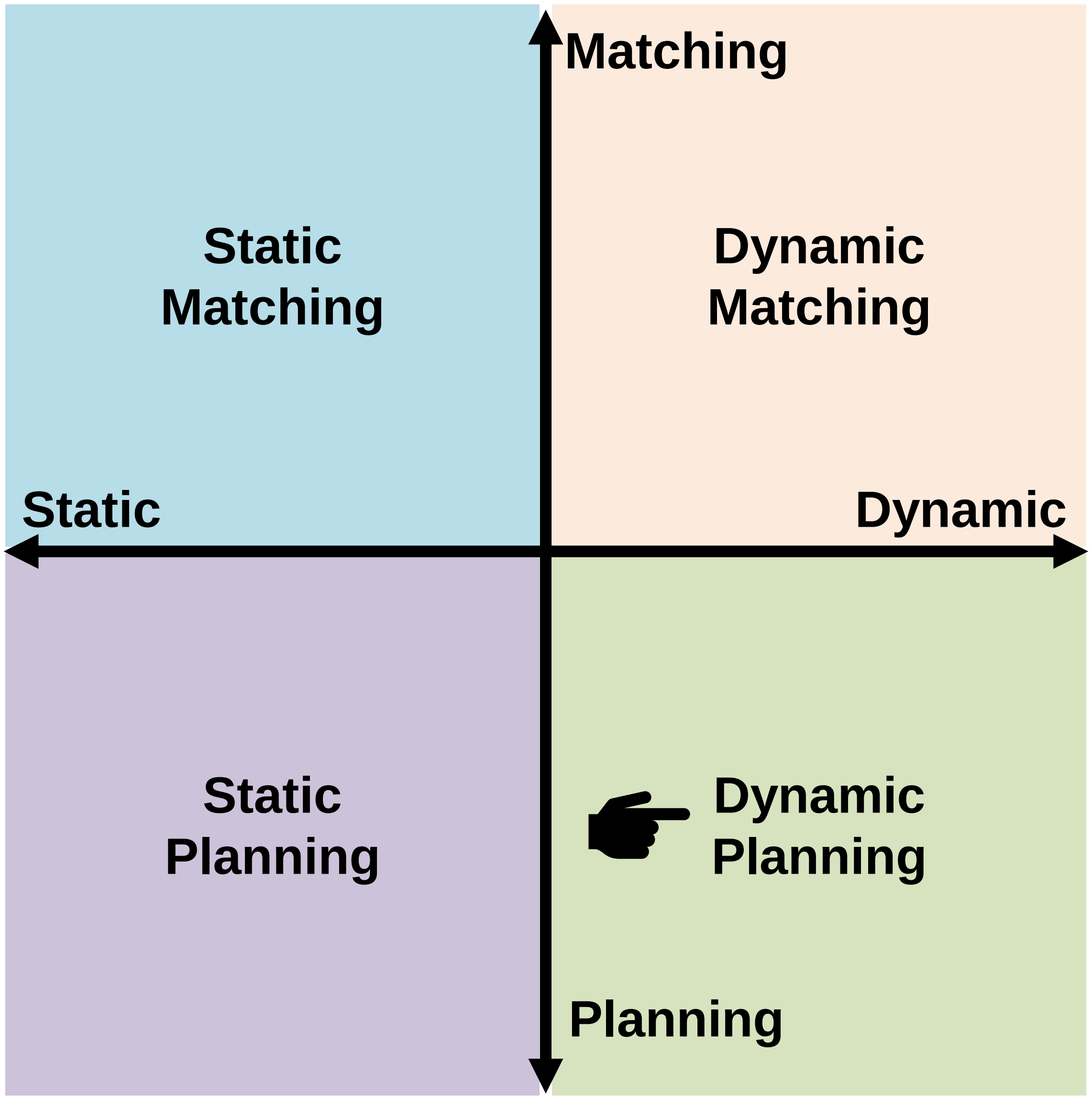
Many-Worker-To-Many-Tasks Planning

- Greedy-based Algorithm
  - Greedily choose the pair $< w, t >$ with the maximum utility/travel cost ratio for each worker



J. She, Y. Tong, L. Chen. Utility-aware event-participant planning. SIGMOD 2015
D. Deng, C. Shahabi, L. Zhu. Task matching and scheduling for multiple workers in spatial crowdsourcing. GIS 2015.

Dynamic Planning



Dynamic Planning

- **Problem Definition (One-worker-to-many-tasks)**
  - **Given one worker and a set of spatial tasks, which are dynamically released and have some constraints (e.g., deadline), the problem is find a route for the worker such that the number of the finished tasks is maximized**
  - **The route can only be updated when a task is released**

Dynamic Planning

- Task-layer Greedy
  - Greedily choose the next nearest task after finishing the current task

There is no constant competitive ratio

Y. Li, M. Yiu, W. Xu. Oriented online route recommendation for spatial crowdsourcing task workers. SSTD 2015.

Reference: Planning

1. C. H. Papadimitriou. The Euclidean travelling salesman problem is NP-complete. TCS, 4(3):237-244, 1977.
2. P. Vansteenwegen, W. Souffriau, D. V. Oudheusden. The orienteering problem: A survey. EJOR, 209(1): 1-10, 2011.
3. D. Deng, C. Shahabi, U. Demiryurek. Maximizing the number of worker's self-selected tasks in spatial crowdsourcing. In GIS, pages 314-323, 2013.
4. S. He, D. Shin, J. Zhang, J. Chen. Toward optimal allocation of location dependent tasks in crowdsensing. In INFOCOM, pages 745-753, 2014.
5. J. She, Y Tong, L. Chen. Utility-aware event-participant planning. In SIGMOD, pages 1629-1643, 2015
6. D. Deng, C. Shahabi, L. Zhu. Task matching and scheduling for multiple workers in spatial crowdsourcing. In GIS, pages 21:1-10, 2015.
7. D. Deng, C. Shahabi, U. Demiryurek, L. Zhu. Task selection in spatial crowdsourcing from worker's perspective. GeoInformatica, 20(3): 529-568, 2016.
8. Y. Li, ML. Yiu, W. Xu. Oriented online route recommendation for spatial crowdsourcing task workers. In SSTD, pages 137-156, 2015.
9. M. Asghari, D. Deng, C. Shahabi, U. Demiryurek, Y. Li. Price-aware real-time ride-sharing at scale: an auction-based approach. In GIS, pages 3:1-3:10, 2016.

Summary

		Matching Model	Planning Model
Computation Complexity	Static Scenario	Weighted Bipartite Matching [Polynomial Time Solvable Problem]	Traveling Salesman Problem (TSP) or Orienteering Problem (OP) [NP-hard Problem]
	Dynamic Scenario	Constant-competitive-ratio Online Algorithms	Non-constant-competitive-ratio Online Algorithms
Workload per Worker		One Task	Multiple Tasks
Representative Applications		Taxi Calling Service Bike Sharing	Food Delivery Service Ridesharing Service

Summary

		Matching Model	Planning Model
Computation Complexity	Static Scenario	Weighted Bipartite Matching [Polynomial Time Solvable Problem]	Traveling Salesman Problem (TSP) or Orienteering Problem (OP) [NP-hard Problem]
	Dynamic Scenario	Constant-competitive-ratio Online Algorithms	Non-constant-competitive-ratio Online Algorithms
Workload per Worker		One Task	Multiple Tasks
Representative Applications		Taxi Calling Service Bike Sharing	Food Delivery Service Ridesharing Service

Summary

		Matching Model	Planning Model
Computation Complexity	Static Scenario	Weighted Bipartite Matching [Polynomial Time Solvable Problem]	Traveling Salesman Problem (TSP) or Orienteering Problem (OP) [NP-hard Problem]
	Dynamic Scenario	Constant-competitive-ratio Online Algorithms	Non-constant-competitive-ratio Online Algorithms
Workload per Worker		One Task	Multiple Tasks
Representative Applications		Taxi Calling Service Bike Sharing	Food Delivery Service Ridesharing Service

Summary

		Matching Model	Planning Model
Computation Complexity	Static Scenario	Weighted Bipartite Matching [Polynomial Time Solvable Problem]	Traveling Salesman Problem (TSP) or Orienteering Problem (OP) [NP-hard Problem]
	Dynamic Scenario	Constant-competitive-ratio Online Algorithms	Non-constant-competitive-ratio Online Algorithms
Workload per Worker		One Task	Multiple Tasks
Representative Applications		Taxi Calling Service Bike Sharing	Food Delivery Service Ridesharing Service

Summary

		Matching Model	Planning Model
Computation Complexity	Static Scenario	Weighted Bipartite Matching [Polynomial Time Solvable Problem]	Traveling Salesman Problem (TSP) or Orienteering Problem (OP) [NP-hard Problem]
	Dynamic Scenario	Constant-competitive-ratio Online Algorithms	Non-constant-competitive-ratio Online Algorithms
Workload per Worker		One Task	Multiple Tasks
Representative Applications		Taxi Calling Service Bike Sharing	Food Delivery Service Ridesharing Service

Outline

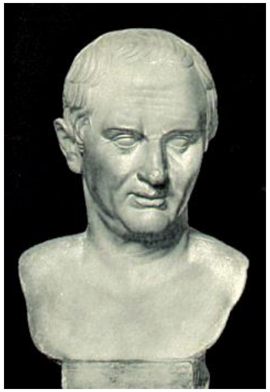
- Overview of Spatial Crowdsourcing (20min)
  - Motivation
  - Workflow
  - Core Issues
  - Difference from Related Tutorials
- Fundamental Techniques (40min)
  - Task Assignment
  - Quality Control
  - Incentive Mechanism
  - Privacy Protection
- Spatial Crowdsourced Applications (20min)
  - Spatial Crowdsourcing Intrinsic Applications
  - Crowd-powered Spatial Applications
- Open Questions (10min)

Quality Control

- Why Quality Control
- Existing Research
  - Spatiotemporal-constraint-based Quality Control
  - Spatiotemporal-target-based Quality Control
    - Minimum Latency
    - Maximum Spatiotemporal Diversity
- Summary

Why Quality Control

- Traditional crowdsourcing
  - Erroneous
    - "To err is human" by Marcus Tullius Cicero
  - Existing research
    - Estimate the quality of workers
    - Infer the truth of tasks



- Spatial crowdsourcing
  - Spatiotemporal factors as constraints
    - Workers still err / truth of tasks still need to be inferred
    - Spatiotemporal factors are only used as constraints
  - Spatiotemporal factors as goals
    - Quality of tasks is assessed by spatiotemporal factors

Quality Control

- Why Quality Control
- Existing Research
  - Spatiotemporal-constraint-based Quality Control
  - Spatiotemporal-target-based Quality Control
    - Minimum Latency
    - Maximum Spatiotemporal Diversity
- Summary

Spatiotemporal-constraint-based QC

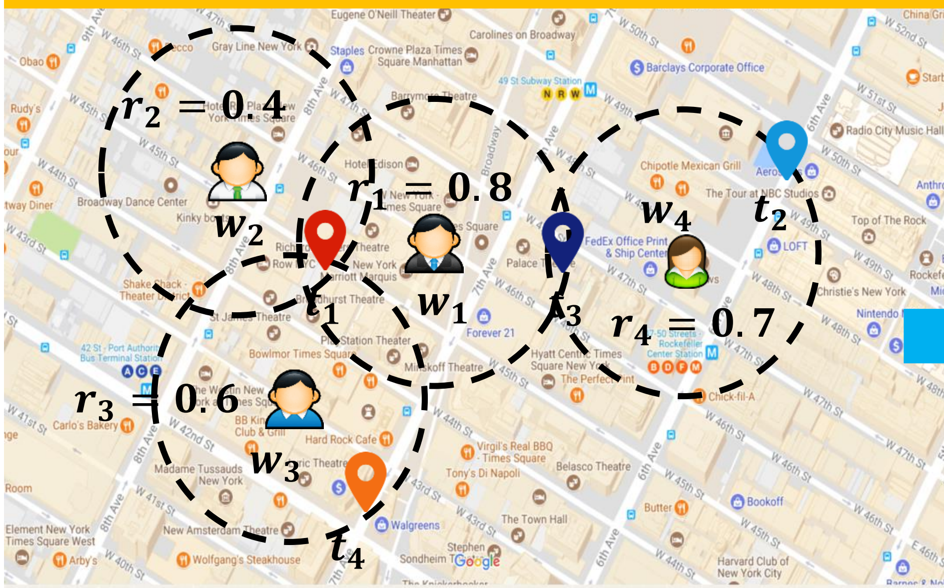
- Given a set of tasks and a set of workers with their **reputation**, the problem is to maximize the total number of assigned tasks such that
  - Quality requirement of each task is satisfied
  - **Spatiotemporal constraint:** tasks should locate in the service range of the assigned workers

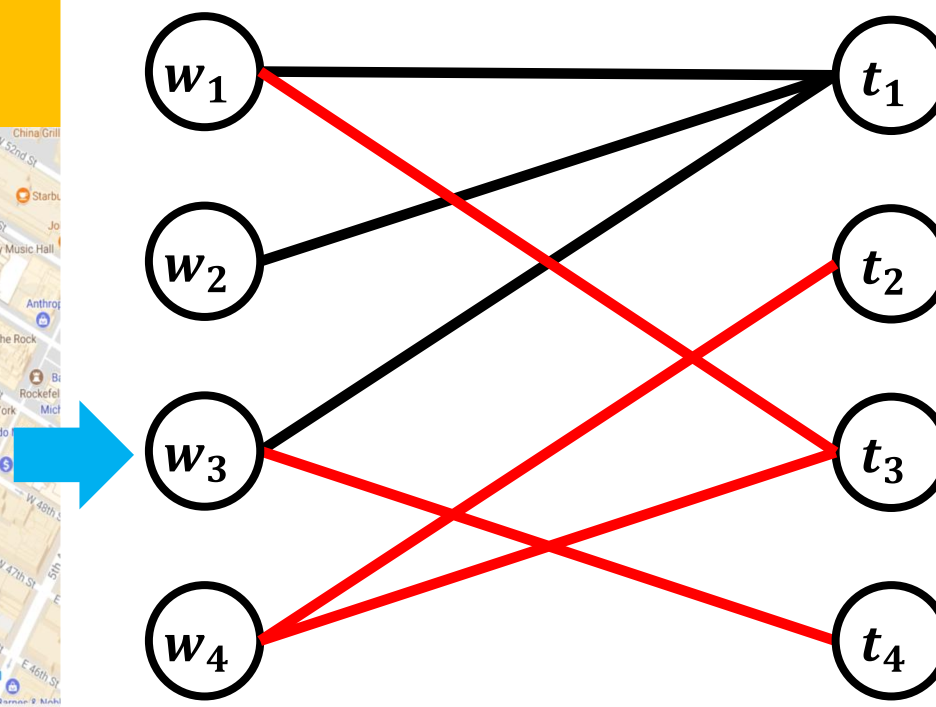
Spatial-constraint-based QC

- Quality requirement depends on quality control methods e.g. Majority Voting

Aggregate Reputation Score (ARS):

$$ARS(Q) = \sum_{k=\frac{|Q|}{2}+1}^{|Q|} \sum_{A \subset F_k} \prod_{w_j \in A} r_j \prod_{w_j \notin A} (1 - r_j)$$





L. Kazemi, C. Shahabi, L. Chen. GeoTruCrowd: trustworthy query answering with spatial crowdsourcing. GIS 2013.

Quality Control

- Why Quality Control
- Existing Research
  - Spatiotemporal-constraint-based Quality Control
  - Spatiotemporal-target-based Quality Control
    - Minimum Latency
    - Maximum Spatiotemporal Diversity
- Summary

Quality Control

- Why Quality Control
- Existing Research
  - Spatiotemporal-constraint-based Quality Control
  - Spatiotemporal-target-based Quality Control
    - Minimum Latency
    - Maximum Spatiotemporal Diversity
- Summary

Minimum Latency

- **Latency is a more important concern in spatial crowdsourcing than in traditional crowdsourcing**

(Internet) Crowdsourcing

Latency: only care the longest completion time of the tasks

Spatial Crowdsourcing

Latency: also care the total completion time of all the tasks

Minimum Latency

- **Latency is a more important concern in spatial crowdsourcing than in traditional crowdsourcing**

- **Solutions**
  - **Modelled and solved by online matching or online planning**

Online Matching

Online Planning

Quality Control

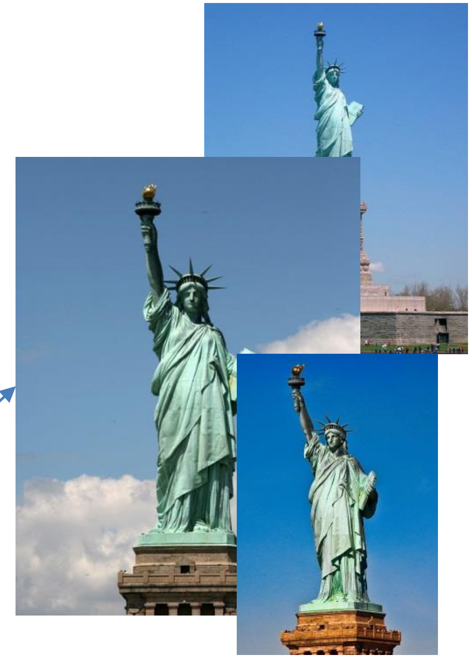
- Why Quality Control
- Existing Research
  - Spatiotemporal-constraint-based Quality Control
  - Spatiotemporal-target-based Quality Control
    - Minimum Latency
    - Maximum Spatiotemporal Diversity
- Summary

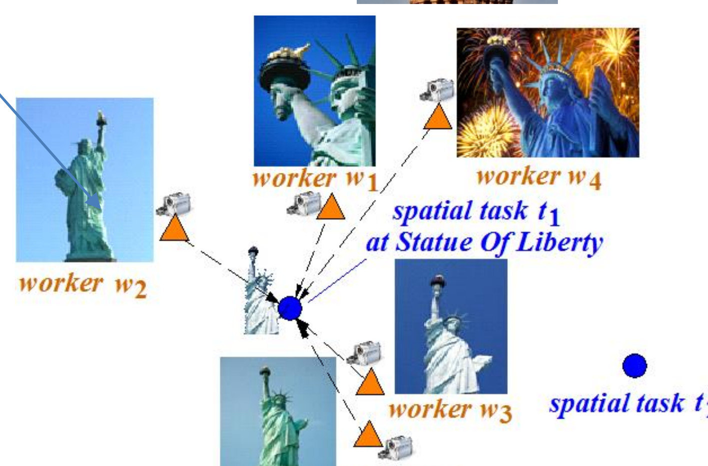
Maximum Spatiotemporal Diversity

◻ Spatial Diversity

Take photos of a landmark

Post Task

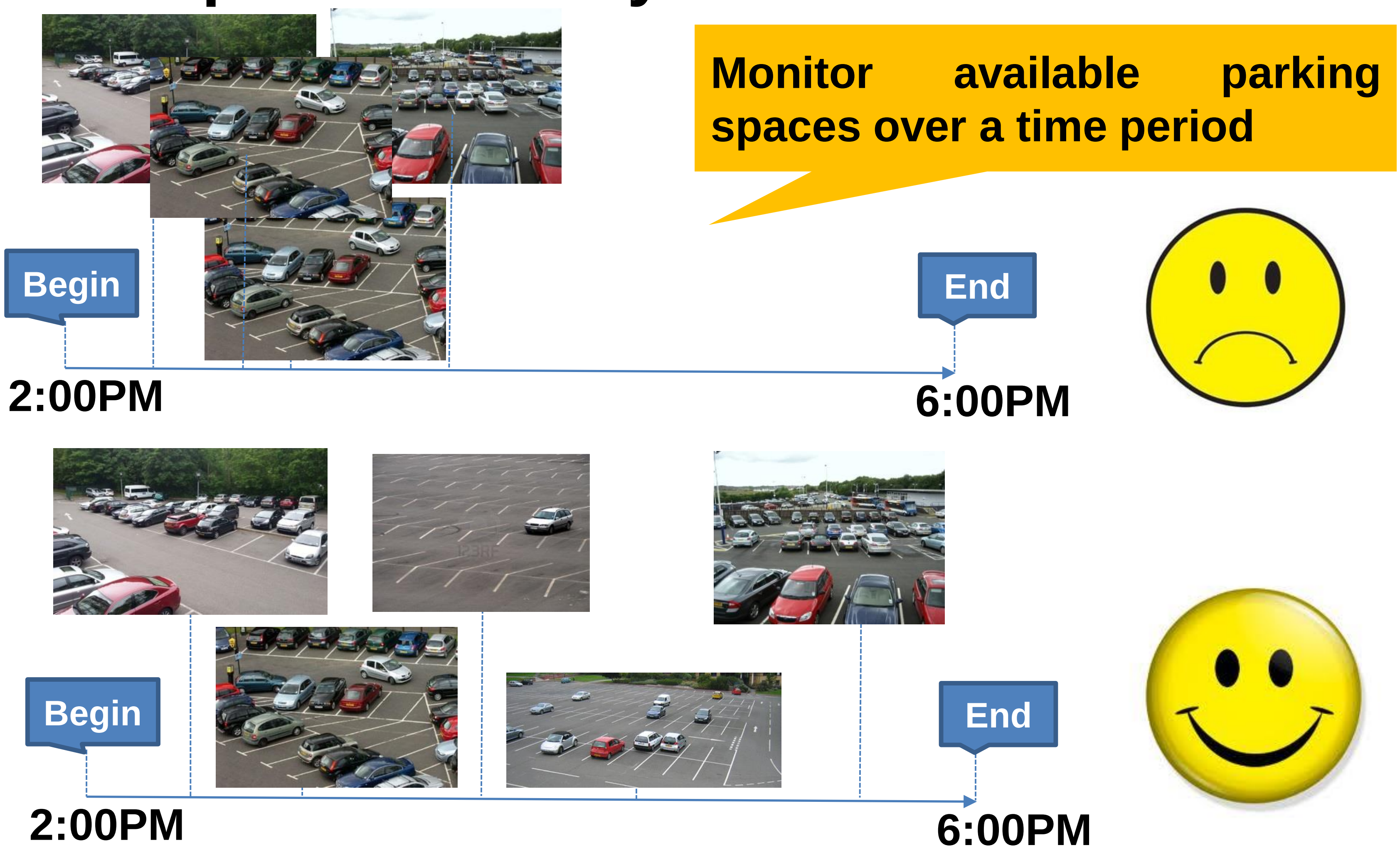




P. Cheng, X. Lian, Z. Chen, R. Fu, L. Chen, J. Han, J. Zhao. Reliable Diversity-Based Spatial Crowdsourcing by Moving Workers. VLDB 2015.

Maximum Spatiotemporal Diversity

- Temporal Diversity

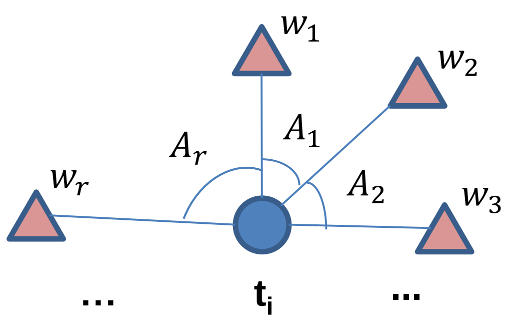


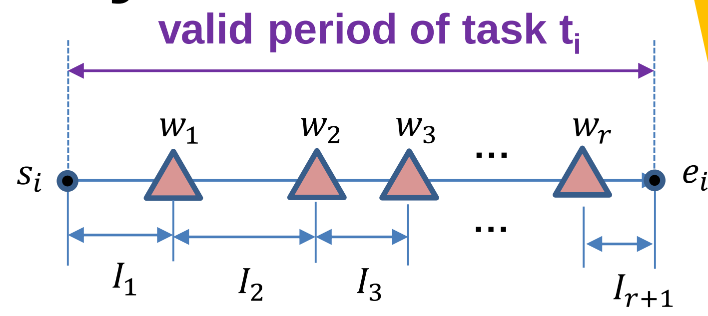
P. Cheng, X. Lian, Z. Chen, R. Fu, L. Chen, J. Han, J. Zhao. Reliable Diversity-Based Spatial Crowdsourcing by Moving Workers. VLDB 2015.

Maximum Spatiotemporal Diversity

Spatiotemporal Diversity

Entropy of temporal diversity





Spatial Diversity is given as

$$SD(t_i) = -\sum_{j=1}^{r} \frac{A_j}{2\pi} \cdot log\left(\frac{A_j}{2\pi}\right)$$

Entropy of spatial diversity

Temporal Diversity is given as

$$TD(t_i) = -\sum_{j=1}^{r+1} \frac{I_j}{e_i - s_i} \cdot log\left(\frac{I_j}{e_i - s_i}\right)$$

Balance Spatial Diversity and Temporal Diversity:

$$STD(t_i, W_i) = \beta \cdot SD(t_i) + (1 - \beta) \cdot TD(t_i)$$

The goal is to maximize the spatiotemporal diversity of a given task

P. Cheng, X. Lian, Z. Chen, R. Fu, L. Chen, J. Han, J. Zhao. Reliable Diversity-Based Spatial Crowdsourcing by Moving Workers. VLDB 2015.

Quality Control

- Why Quality Control

- Existing Research
  - Spatiotemporal-constraint-based Quality Control
  - Spatiotemporal-target-based Quality Control
    - Minimum Latency
    - Maximum Spatiotemporal Diversity

- Summary

Summary

- “To err is human”

- Spatiotemporal-constraint-based Quality Control
  - Tasks have quality requirement for truth of answers
  - Workers have spatiotemporal constraints

- Spatiotemporal-target-based Quality Control
  - Minimizing Latency
  - Maximizing Spatiotemporal Diversity

Reference: Quality Control

1. L. Kazemi, C. Shahabi, L. Chen. GeoTruCrowd: trustworthy query answering with spatial crowdsourcing. In GIS, pages 304-313, 2013.
2. B. Hecht, M. Stephens. A tale of cities: Urban biases in volunteered geographic information. In ICWSM, pages 197-205, 2014.
3. R. Teodoro, P. Öztürk, M. Naaman, W. Mason, J. Lindqvist. The motivations and experiences of the on-demand mobile workforce. In CSCW, pages 236-247, 2014.
4. P. Cheng, X. Lian, Z. Chen, R. Fu, L. Chen, J. Han, J. Zhao. Reliable diversity-based spatial crowdsourcing by moving workers. PVLDB, 8(10): 1022-1033, 2015.
5. J. Thebault-Spieker, L. Terveen, B. Hecht. Avoiding the south side and the suburbs: The geography of mobile crowdsourcing markets. In CSCW, pages 265-275, 2015.
6. T. Kandappu, N. Jaiman, R. Tandriansyah, A. Misra, S. Cheng, C. Chen, H. Lau, D. Chander, K. Dasgupta. TASKer: Behavioral insights via campus-based experimental mobile crowdsourcing. In UbiComp, pages 392-402, 2016.
7. C. Cai, S. Iqbal, J. Teevan. Chain reactions: The impact of order on microtask chains. In CHI, pages 3143-3154, 2016.
8. W. Ouyang, M. Srivastava, A. Toniolo, T. Norman. Truth discovery in crowdsourced detection of spatial events. TKDE, 28(4): 1047-1060, 2016.
9. P. Cheng, X. Lian, L. Chen, J. Han, J. Zhao. Task assignment on multi-skill oriented spatial crowdsourcing. TKDE, 28(8): 2201-2215, 2016.
10. D. Gao, Y. Tong, J. She, T. Song, L. Chen, K. Xu. Top-k team recommendation in spatial crowdsourcing. In WAIM, pages 191-204, 2016.

Reference: Quality Control

11. L. Zheng, L. Chen. Maximizing acceptance in rejection-aware spatial crowdsourcing. TKDE, 29(9): 1943-1956, 2017.
12. V. Sheng, F. Provost, P. Ipeirotis. Get another label? Improving data quality and data mining using multiple, noisy labelers. In KDD, pages 614-622, 2008.
13. P. Ipeirotis, F. Provost, J. Wang. Quality management on amazon mechanical turk. In KDD, pages 64-67, 2010.
14. X. Liu, M. Lu, B. Ooi , Y. Shen, S. Wu , M. Zhang. Cdas: A crowdsourcing data analytics system. PVLDB, 5(11): 1495-1506, 2012.
15. C. Cao, J. She, Y. Tong, L. Chen. Whom to ask? Jury selection for decision making tasks on micro-blog services. PVLDB, 5(11): 1495-1506, 2012.
16. V. Raykar, S. Yu. Eliminating spammers and ranking annotators for crowdsourced labeling tasks. JMLR, 13:491-518, 2012.
17. C. Cao, Y. Tong, L. Chen, H. Jagadish. Wisemarket: A new paradigm for managing wisdom of online social users. In KDD, pages 455-463, 2013.
18. N. Dalvi, A. Dasgupta, R. Kumar, V. Rastogi. Aggregating crowdsourced binary ratings. In WWW, pages 285-294, 2013.
19. M. Yin, Y. Chen. Predicting crowd work quality under monetary interventions. In AAAI, pages 259-268, 2017.

Outline

- Overview of Spatial Crowdsourcing (20min)
  - Motivation
  - Workflow
  - Core Issues
  - Difference from Related Tutorials
- Fundamental Techniques (40min)
  - Task Assignment
  - Quality Control
  - Incentive Mechanism
  - Privacy Protection
- Spatial Crowdsourced Applications (20min)
  - Spatial Crowdsourcing Intrinsic Applications
  - Crowd-powered Spatial Applications
- Open Questions (10min)

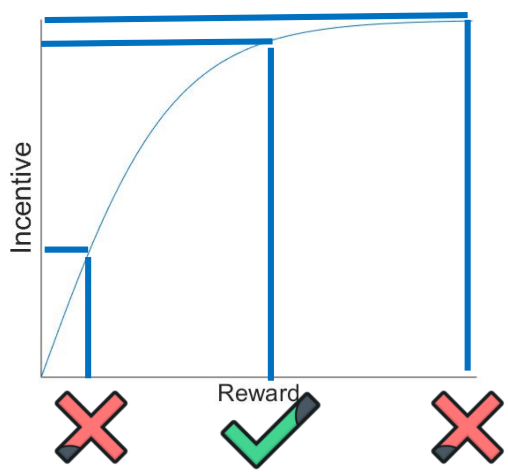
Incentive Mechanism

- Motivation
- Existing Research
  - Monetary Reward
    - Game Theory based incentive mechanism
    - Auction based incentive mechanism
  - Gamification
  - Volunteer
- Summary

Incentive Mechanism

- Stimulate workers to complete tasks efficiently using proper reward
- A tradeoff between workers and tasks
  - Excessively low reward discourages the workers to finish the task
  - Excessively high reward hurts the benefit of the platform

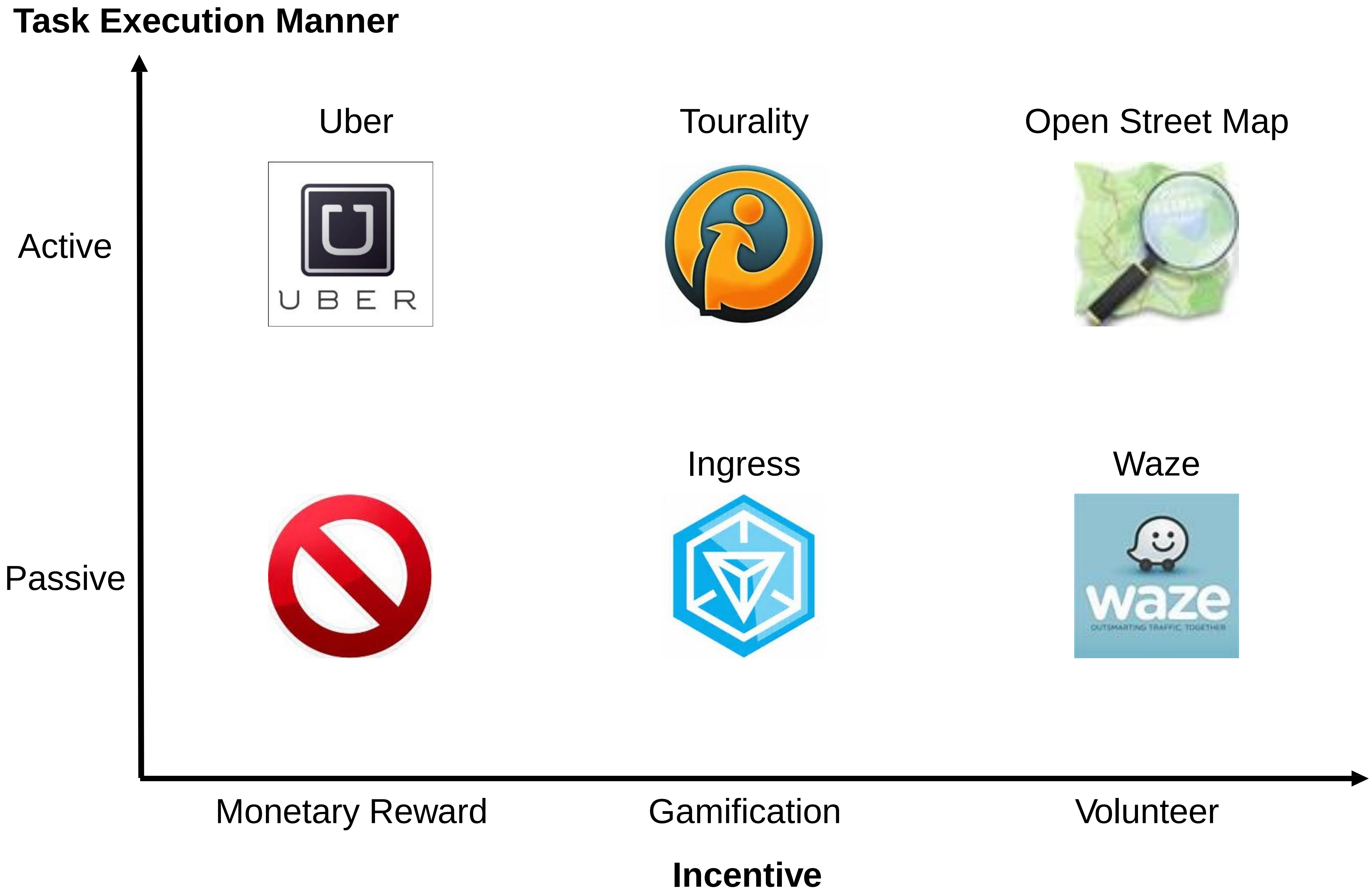
Platform



It hurts my benefit because of the high reward

Too little reward that I cannot work efficiently

Incentive Mechanism: Design Space



Incentive Mechanism

- Motivation
- Existing Research
  - Monetary Reward
    - Game Theory based incentive mechanism
    - Auction based incentive mechanism
  - Gamification
  - Volunteer
- Summary

Incentive Mechanism

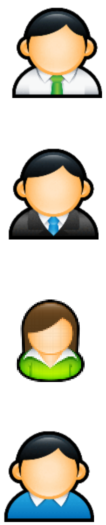
- Motivation
- Existing Research
  - Monetary Reward
    - Game Theory based incentive mechanism
    - Auction based incentive mechanism
  - Gamification
  - Volunteer
- Summary

Game Theory Based Mechanism

- ## Platform-Centric Model

We need to compete for the total budget

I have a spatial task that needs to be completed collaboratively by multiple workers



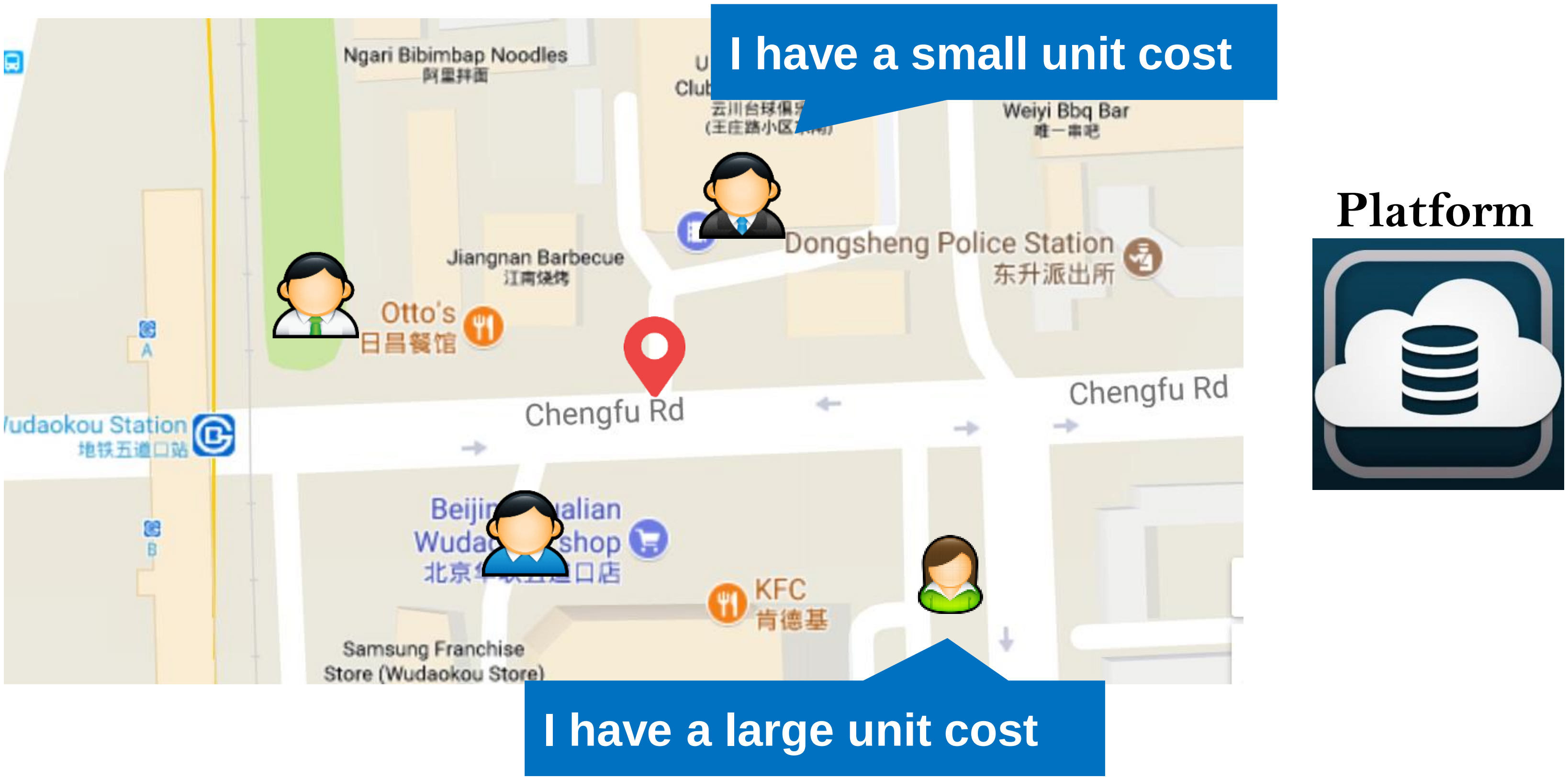
Platform

I also have a total budget

The total budget is allocated to the workers according to their contributions

Game Theory Based Mechanism

◻ Platform-Centric Model



Game Theory Based Mechanism

- **Stackelberg Game**

Leader ⇨ Followers

- **Stackelberg Equilibrium**

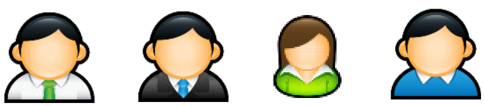
I have the knowledge of the followers' behavior, and want to maximize my utility

I know the leader's strategy, and want to maximize my utility

D. Yang, G. Xue, X. Fang, J. Tang. Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing. MobiCom 2012.

Game Theory Based Mechanism

Step 1: Estimate an optimal total budget and publish it



Step 2: Submit their actual plans of performing the tasks

Step 3: Allocate the budget to the workers according to their actual plans

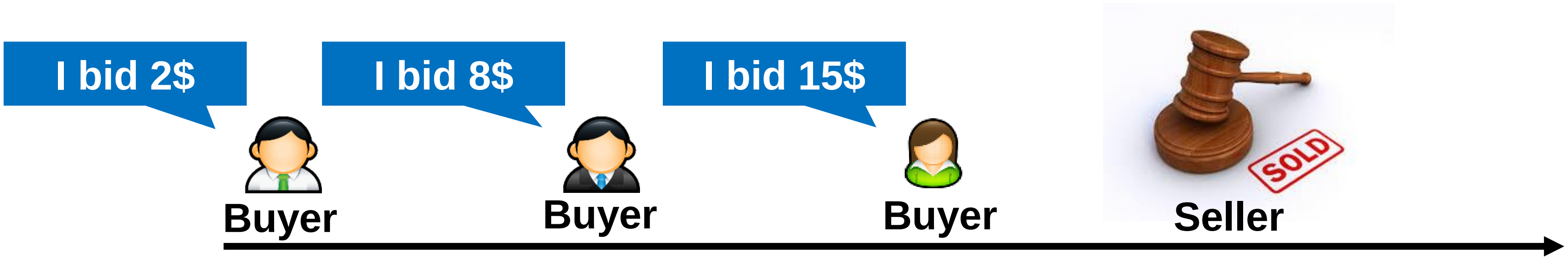
D. Yang, G. Xue, X. Fang, J. Tang. Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing. MobiCom 2012.

Incentive Mechanism

- Motivation
- Existing Research
  - Monetary Reward
    - Game Theory based incentive mechanism
    - Auction based incentive mechanism
  - Gamification
  - Volunteer
- Summary

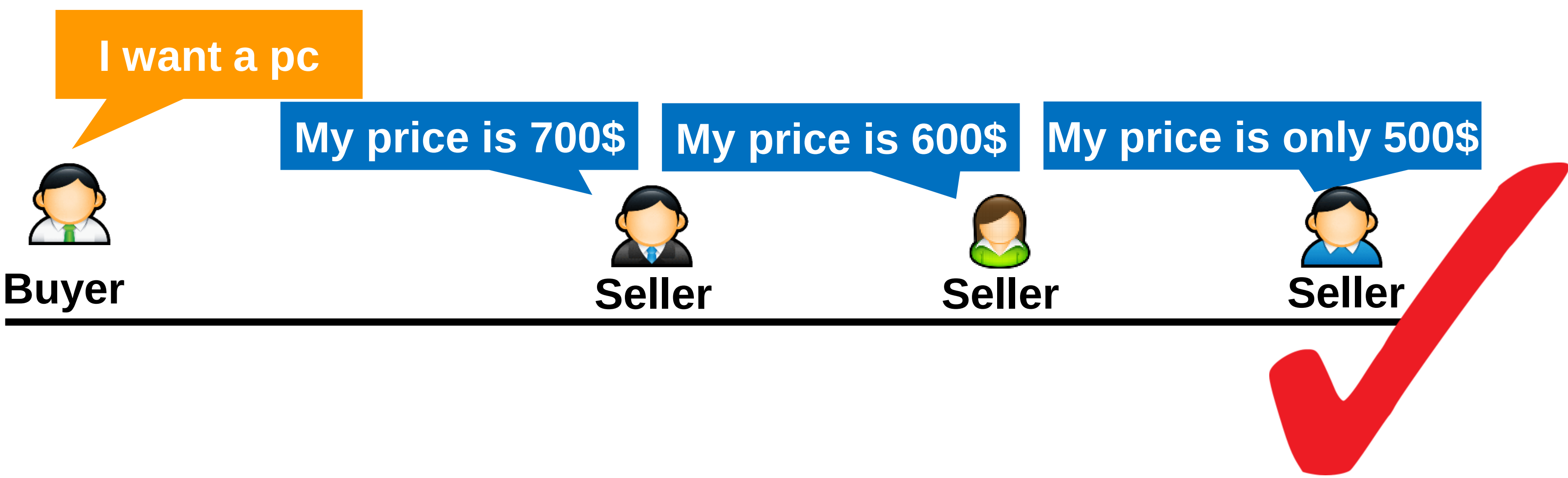
Background of Auction

- **Auction**
  - **Buyers compete to obtain goods or services by offering increasingly higher prices**



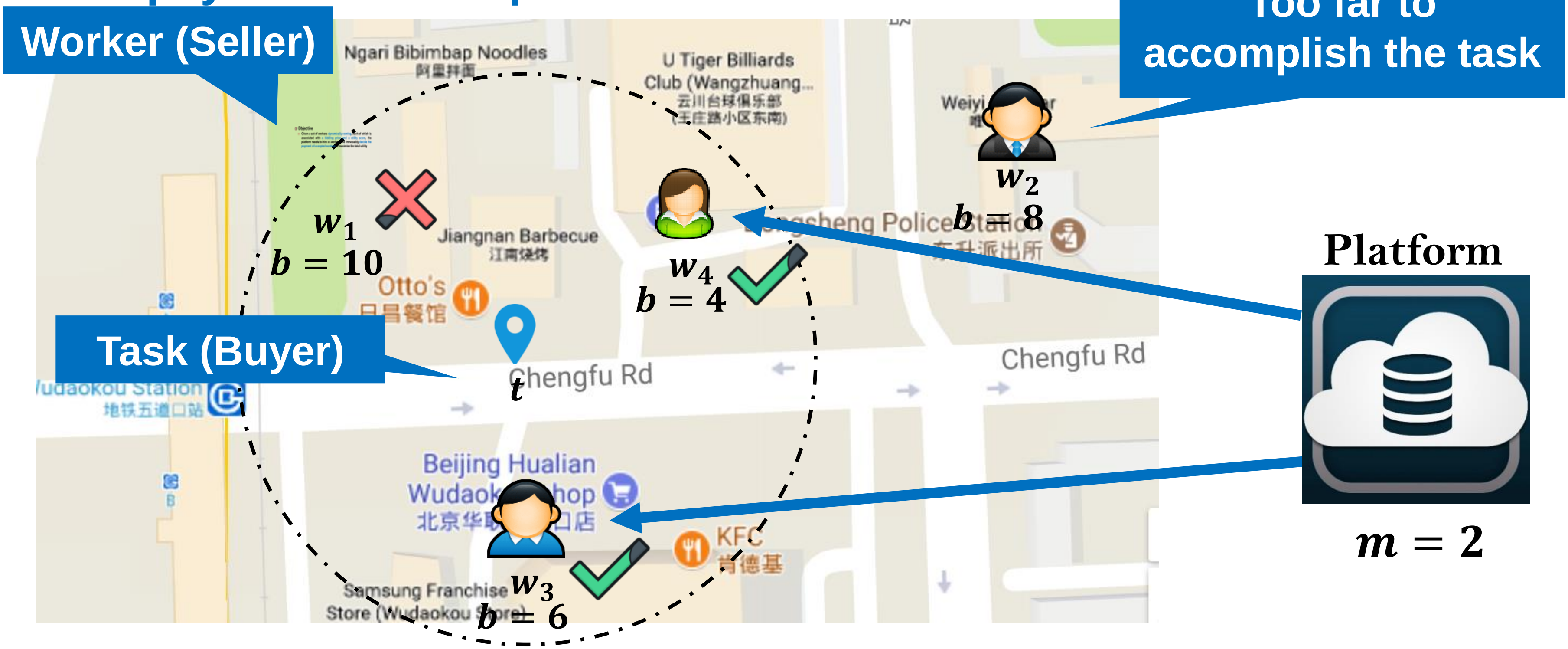
Background of Auction

- **Auction**
  - **Buyers compete to obtain goods or services by offering increasingly higher prices**
- **Reverse auction**
  - **The sellers compete to obtain a business from the buyer and prices will typically decrease as the sellers underbid each other**



Auction Based Mechanism

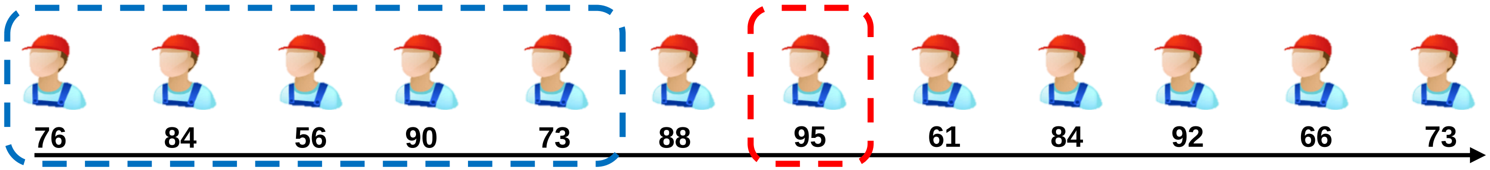
- **Objective**
  - **Given a set of workers dynamically coming, each of which is associated with a bidding price and a utility score, the platform needs to hire m workers and irrevocably decide the payment of accepted workers to maximize th**



X. Zhang, Z. Yang, Z. Zhou, H. Cai, L. Chen, X. Li. Free market of crowdsourcing: incentive mechanism design for mobile sensing. TPDS 2014.

Auction Based Mechanism

- **State-of-the-art**
  - **Based on the secretary problem**



- **Secretary Problem**
  - **n applicants, which can be ranked in strictly total order, apply for a single position. The applicants are interviewed sequentially and immediately after an interview, the interviewed applicant is either accepted or rejected irrevocably. The objective is to have the highest probability of selecting the best applicant of the whole group**
- **Solution of Secretary Problem**
  - **Reject the first n/e applicants and hire the first applicant thereafter who has a higher score than all preceding applicants.**
  - **With probability $1/e$ we can select the best applicant.**

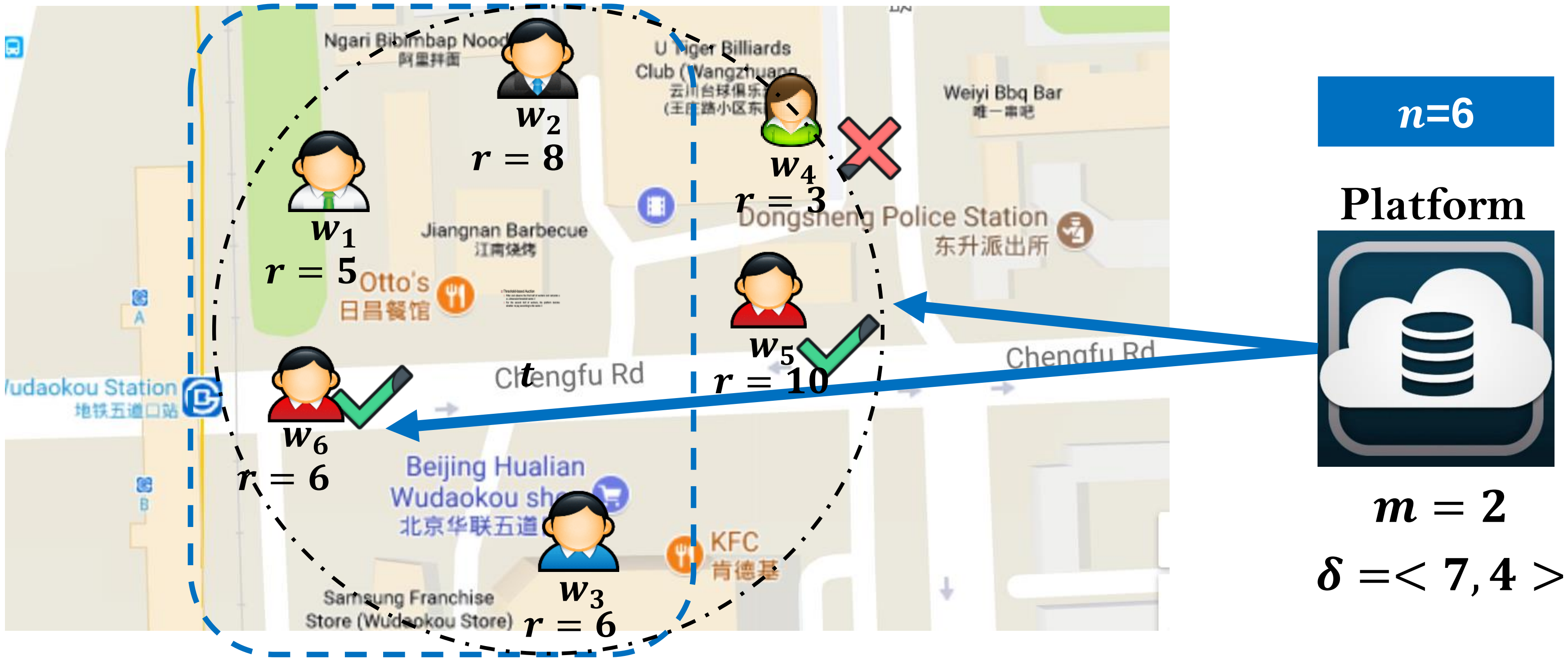
Auction Based Mechanism

- **Threshold-based Auction**
  - **Filter and observe the first half of workers and calculate a m –dimension threshold vector δ**
  - **For the second half of workers, the framework decides whether the ratio of utility over bidding price by the new arrival worker is larger than the corresponding threshold in the vector$\boldsymbol{\delta}$**
  - **If so, the platform hires the worker, otherwise rejects the worker.**

X. Zhang, Z. Yang, Z. Zhou, H. Cai, L. Chen, X. Li. Free market of crowdsourcing: incentive mechanism design for mobile sensing. TPDS 2014.

Auction Based Mechanism

- **Threshold-based Auction**
  - **Filter and observe the first half of workers and calculate a m–dimension threshold vector δ**
  - **For the second half of workers, the platform decides whether to pay according to the vector δ**



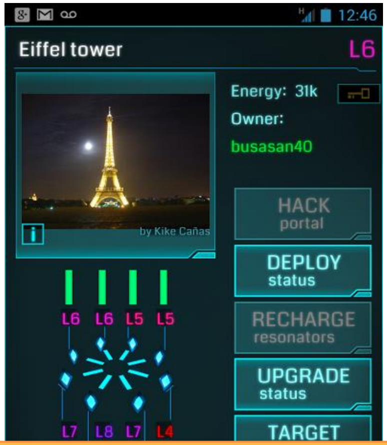
X. Zhang, Z. Yang, Z. Zhou, H. Cai, L. Chen, X. Li. Free market of crowdsourcing: incentive mechanism design for mobile sensing. TPDS 2014.

Incentive Mechanism

- Motivation
- Existing Research
  - Monetary Reward
    - Game Theory based incentive mechanism
    - Auction based incentive mechanism
  - Gamification
  - Volunteer
- Summary

Incentive Mechanism: Gamification

- **Ingress (Niantic, Google)**
  - **A location-based mobile game**
  - **The gameplay consists of capturing "portals" at places of cultural significance such as public art**
  - **Portals are set at specific positions to collect data during the capturing process using phones' camera**



Data from Ingress was used to populate the locations for Pokéstops and gyms in Pokémon Go, released in July 2016

Incentive Mechanism: Gamification

- Pokémon Go (Niantic, Nintendo)
  - A location-based, augmented-reality mobile game
  - The game utilizes the player's mobile device to locate and capture virtual creatures called Pokémon
  - Pokémon can be set at specific positions to collect data during the capturing process using phones' camera

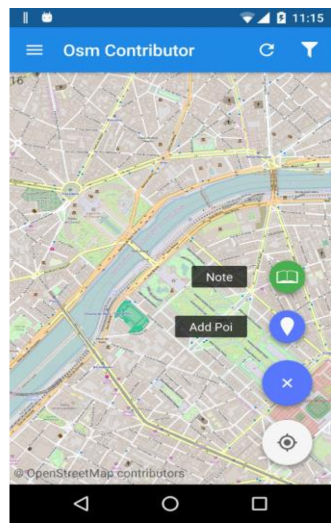
A. Colley, J. Thebault-Spieker, A. Yilun Lin, et al. The Geography of Pokémon GO: Beneficial and Problematic Effects on Places and Movement. CHI 2017.

Incentive Mechanism

- Motivation
- Existing Research
  - Monetary Reward
    - Game Theory based incentive mechanism
    - Auction based incentive mechanism
  - Gamification
  - Volunteer
- Summary

Incentive Mechanism: Volunteer

- **Open Street Map**
  - **A collaborative project to create a free editable map**
  - **A prominent example of volunteered geographic information**
  - **Map data is collected from scratch by volunteers using tools such as a mobile phone, handheld GPS unit, a notebook, digital camera, or a voice recorder**



M. Haklay, P. Weber. Openstreetmap: User-generated street maps. IEEE Pervasive Computing 2008.

Incentive Mechanism

- Motivation
- Existing Research
  - Monetary Reward
    - Game Theory based incentive mechanism
    - Auction based incentive mechanism
  - Gamification
  - Volunteer
- Summary

Summery

- Workers always need to be motivated
- Tradeoff between workers and tasks
- State-of-the-art techniques
  - Monetary-Reward-based incentives
  - Gamification-based incentives
  - Volunteer-based incentives

Reference: Incentive Mechanism

1. M. Musthag, D. Ganesan. Labor dynamics in a mobile micro-task market. In CHI, pages 641-650, 2013.
2. G. Hui, C. Liu, W. Wang, J. Zhao, Z. Song, X. Su, J. Crowcroft, K. Leung. A survey of incentive mechanisms for participatory sensing. IEEE Communications Surveys and Tutorials 17(2): 918-943, 2015.
3. X. Zhang, Z. Yang, W. Sun, Y. Liu, S. Tang, K. Xing, X. Mao. Incentives for Mobile Crowd Sensing: A Survey. IEEE Communications Surveys and Tutorials, 18(1): 54-67, 2016.
4. B. Morschheuser, J. Hamari, J. Koivisto, A. Maedche. Gamified crowdsourcing: Conceptualization, literature review, and future agenda. IJHCS, 106: 26-43, 2017.
5. D. Yang, G. Xue, X. Fang, J. Tang. Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing. In MobiCom, pages 173-184, 2012.
6. I. Koutsopoulos. Optimal incentive-driven design of participatory sensing systems. In INFOCOM, pages 1402-1410, 2013.
7. A. Singla, A. Krause. Incentives for privacy tradeoff in community sensing. In HCOMP, pages 1-9, 2013.
8. D. Zhao, X. Li, H. Ma. How to crowdsource tasks truthfully without sacrificing utility: Online incentive mechanisms with budget constraint. In INFOCOM, pages 1213-1221, 2014.
9. X. Zhang, Z. Yang, Z. Zhou, H. Cai, L. Chen, X. Li. Free market of crowdsourcing: incentive mechanism design for mobile sensing. TPDS, 25(12): 3190-3200, 2014.

Reference: Incentive Mechanism

10. W. A. Mason, . J. Watts. Financial incentives and the performance of crowds. SIGKDD Explorations, 11(2): 100-108, 2009.
11. J. Horton, L. Chilton. The labor economics of paid crowdsourcing. In EC, pages 209-218, 2010.
12. Y. Singer, M. Mittal. Pricing mechanisms for crowdsourcing markets. In WWW, pages 1157-1166, 2013.
13. A Singla, A Krause. Truthful incentives in crowdsourcing tasks using regret minimization mechanisms. In WWW, pages 1167-1178, 2013.
14. Gao Y, Parameswaran AG. Finish them!: pricing algorithms for human computation. PVLDB, 7(14):1965-1976, 2014.
15. M. Yin, Y. Chen. Bonus or not? learn to reward in crowdsourcing. In IJCAI, pages 201-208, 2015.
16. D. Difallah, M. Catasta, G. Demartini, P. Ipeirotis, P. Cudré-Mauroux. The dynamics of micro-task crowdsourcing: The case of Amazon MTurk. In WWW, pages 238-247, 2015.

Outline

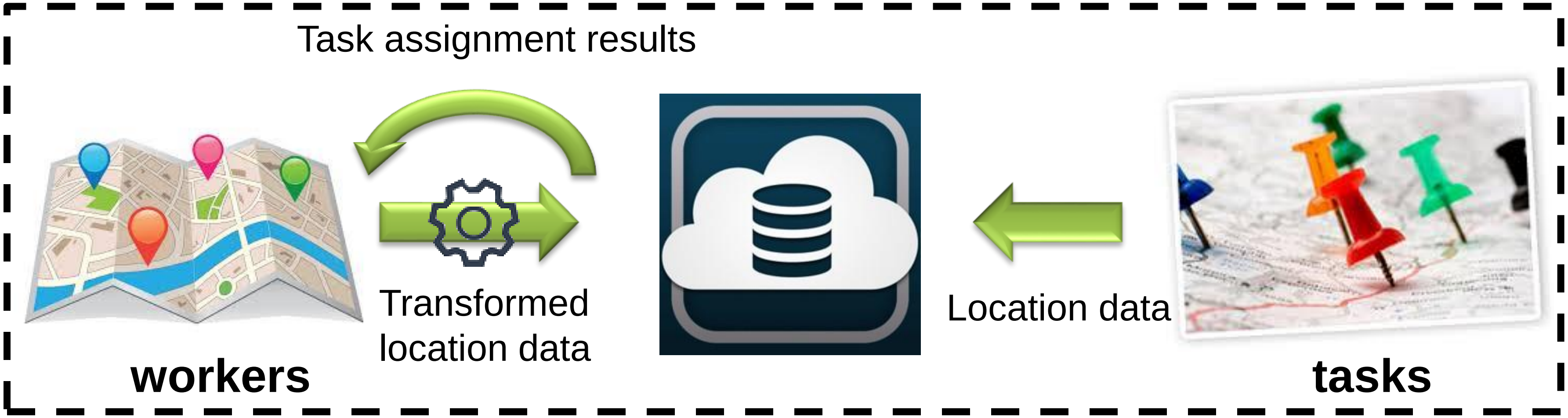
- Overview of Spatial Crowdsourcing (20min)
  - Motivation
  - Workflow
  - Core Issues
  - Difference from Related Tutorials
- Fundamental Techniques (40min)
  - Task Assignment
  - Quality Control
  - Incentive Mechanism
  - Privacy Protection
- Spatial Crowdsourced Applications (20min)
  - Spatial Crowdsourcing Intrinsic Applications
  - Crowd-powered Spatial Applications
- Open Questions (10min)

Privacy Protection

- Motivation & A Unified Framework
- Existing Research
  - Cloaked Locations-based Protection
  - Differential Privacy-based Protection
  - Encrypted Data-based Protection
- Summary

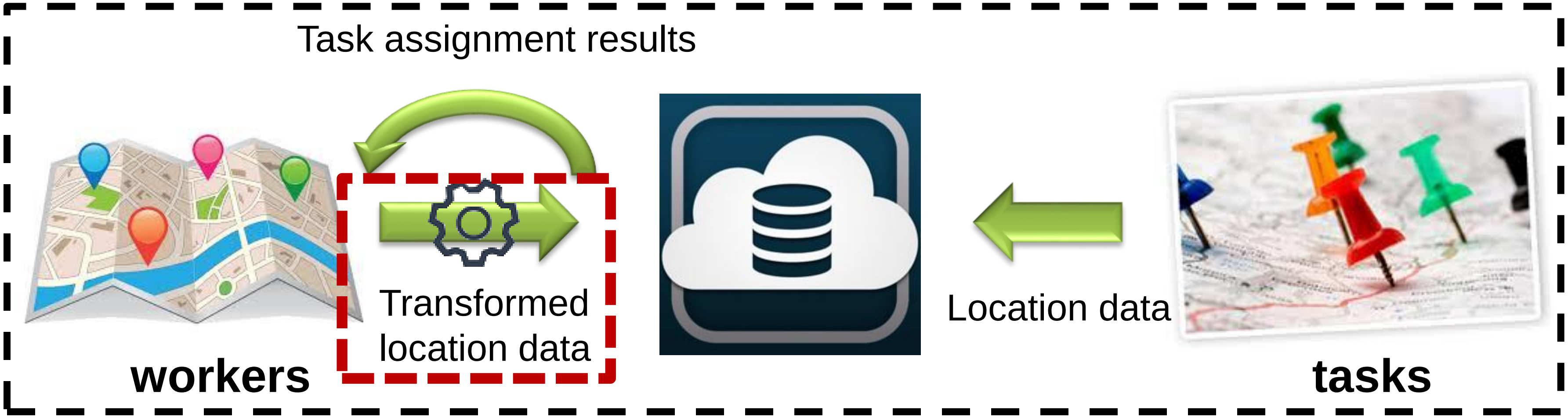
Motivation and Framework

- Motivation
  - Protect the privacy of workers' locations
- A unified framework
  - The locations of the workers are transformed by some techniques
  - The platform performs task assignment based on the transformed locations of the tasks
  - The workers confirm/refine the task assignment results based on their true locations



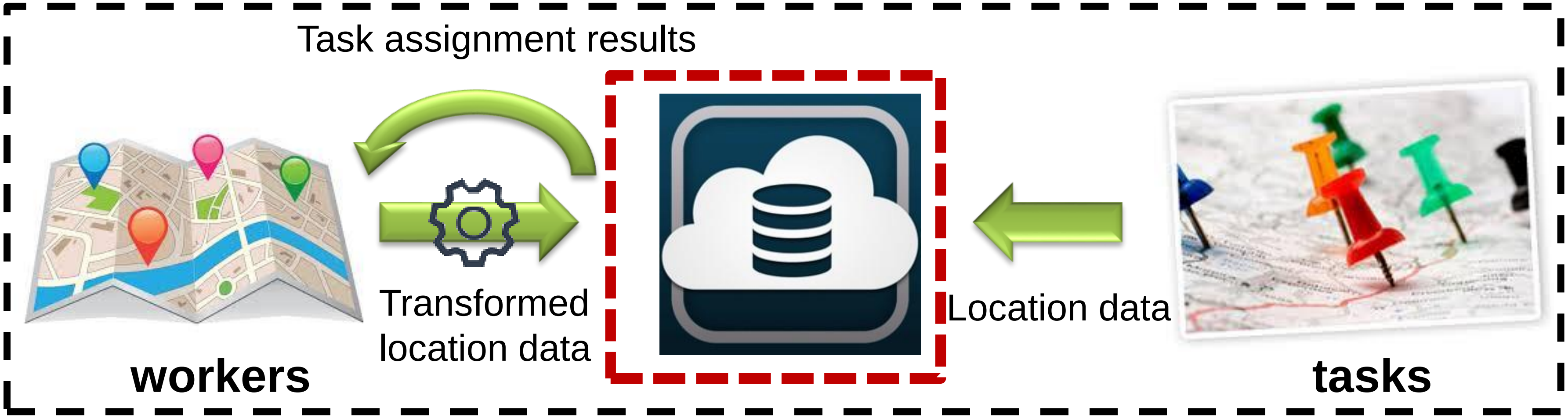
Motivation and Framework

- Motivation
  - Protect the privacy of workers' locations
- A unified framework
  - The locations of the workers are transformed by some techniques
  - The platform performs task assignment based on the transformed locations of the tasks
  - The workers confirm/refine the task assignment results based on their true locations



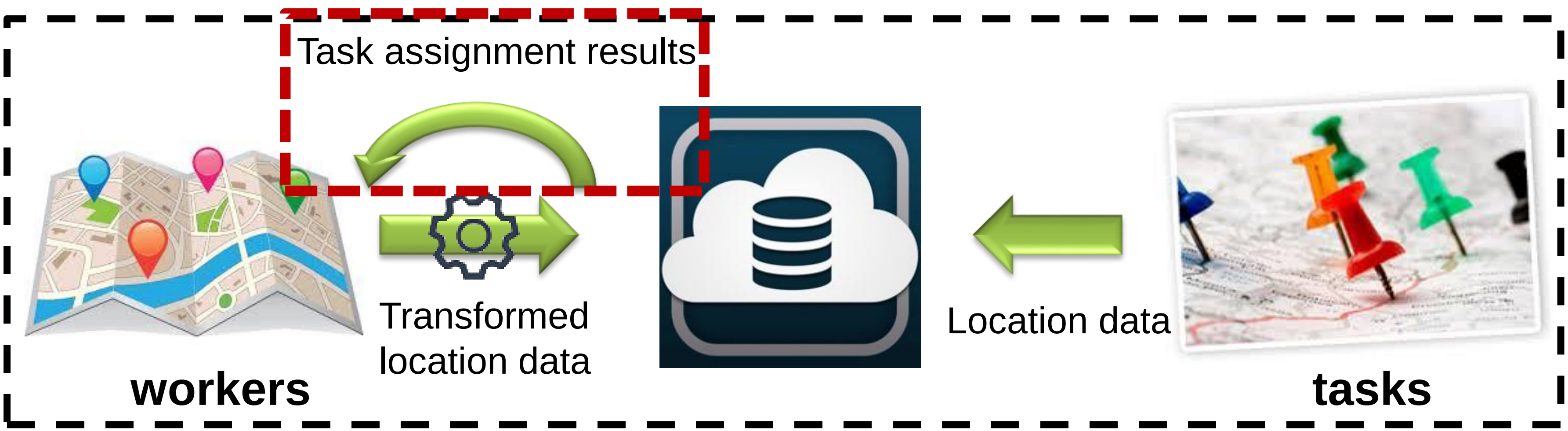
Motivation and Framework

- Motivation
  - Protect the privacy of workers' locations
- A unified framework
  - The locations of the workers are transformed by some techniques
  - The platform performs task assignment based on the transformed locations of the workers
  - The workers confirm/refine the task assignment results based on their true locations



Motivation and Framework

- **Motivation**
  - **Protect the privacy of workers' locations**
- **A unified framework**
  - **The locations of the workers are transformed by some techniques**
  - **The platform performs task assignment based on the transformed locations of the tasks**
  - **The workers confirm/refine the task assignment results based on their true locations**

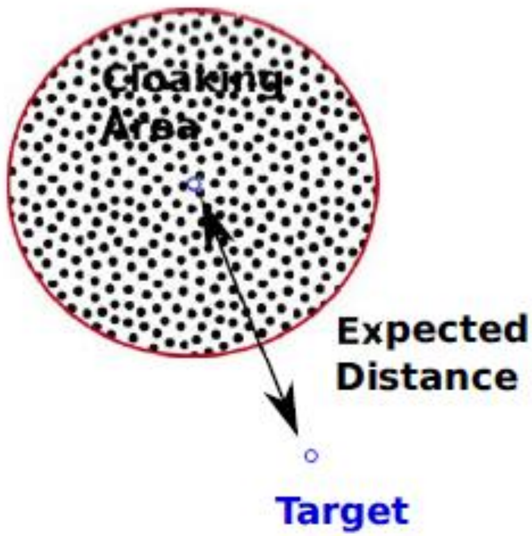


Privacy Protection

- Motivation & A Unified Framework
- Existing Research
  - Cloaked Locations-based Protection
  - Differential Privacy-based Protection
  - Encrypted Data-based Protection
- Summary

Cloaked Area Based Protection

- **The location of the worker is transformed as a cloaked area**
  - **A cloaked area is a pair $<a, f>$, where a is a spatial range and f is the probability density function of the worker at each point in a**



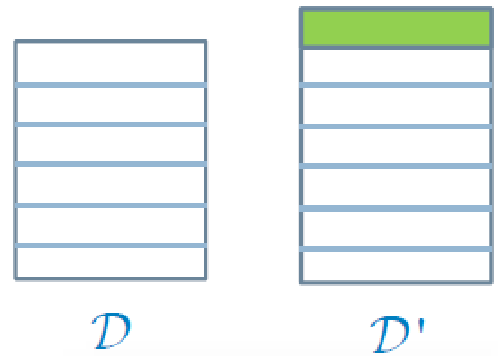
L. Pournajaf, L. Xiong, V. S. Sunderam, S. Goryczka. Spatial Task Assignment for Crowd Sensing with Cloaked Locations. MDM 2014.
L. Pournajaf, L. Xiong, V. S. Sunderam, X. Xu. STAC: spatial task assignment for crowd sensing with cloaked participant locations. GIS 2015.

Privacy Protection

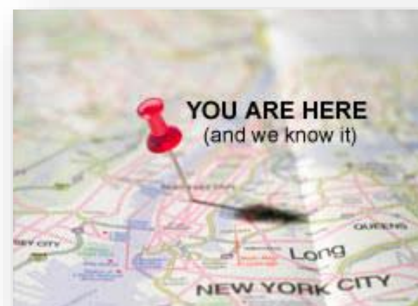
- Motivation & A Unified Framework
- Existing Research
  - Cloaked Locations-based Protection
  - Differential Privacy-based Protection
  - Encrypted Data-based Protection
- Summary

Differential Privacy based Protection

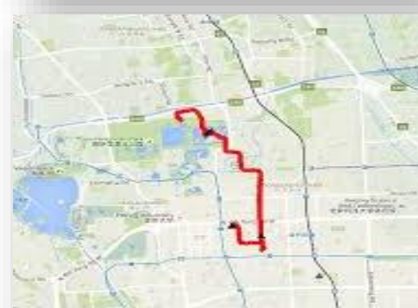
- **Differential Privacy (DP)**
  - **Basic idea: limit the impact of any particular record on the output**
  - **Definition**
    - $\Pr[\quad(\quad)\in\quad]\le e^{\varepsilon}\Pr[\quad(\quad{}^{\prime})\in\quad]$
  - $\varepsilon > 0$ **controls the level of privacy**



- **DP in spatial data**
  - **Protect location information of points**

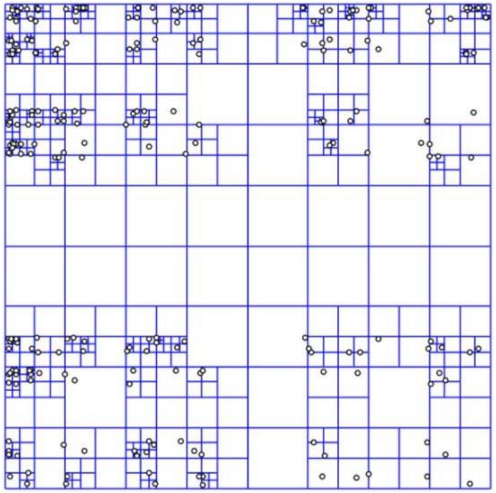


  - **Protect location information of trajectories**

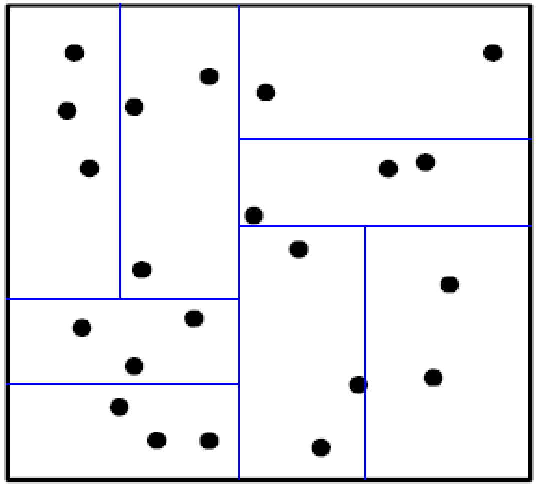


DP: Point Location Protection

- **Point** location protection
- Core Technique
  - Private spatial decompositions (PSD)
  - Decompose a geometric space, data points partitioned among the leaves



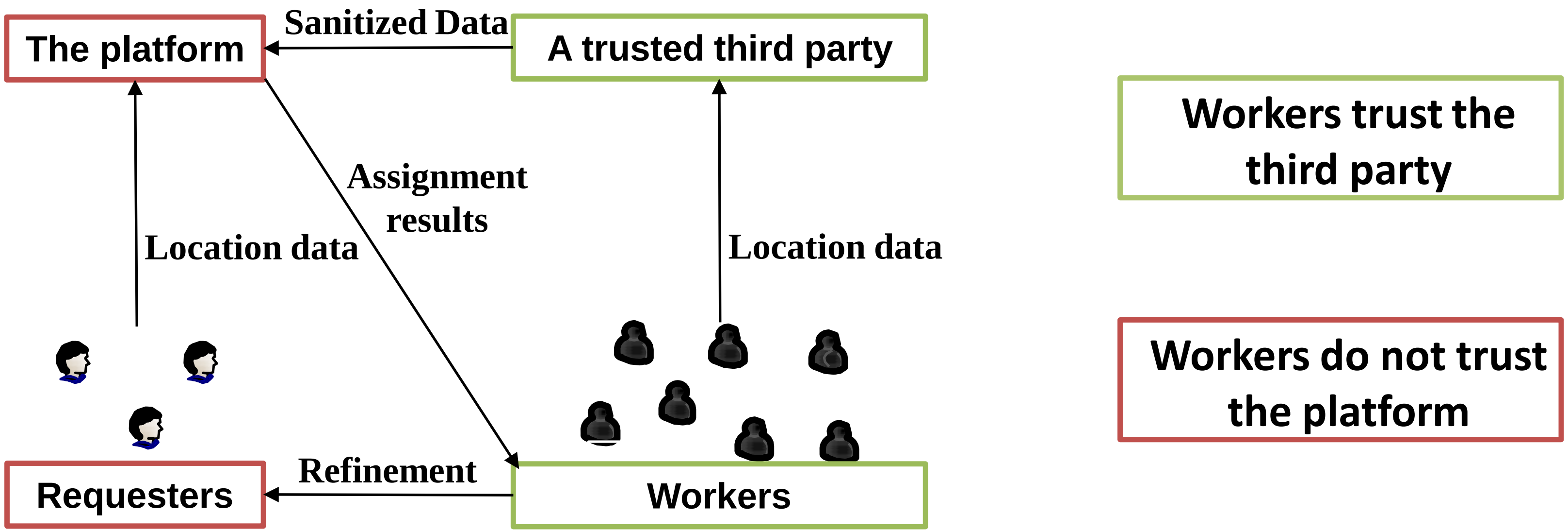
quadtree



kd-tree

G. Cormode, C. M. Procopiuc, D. Srivastava, E. Shen, T. Yu. Differentially Private Spatial Decompositions. ICDE 2012.

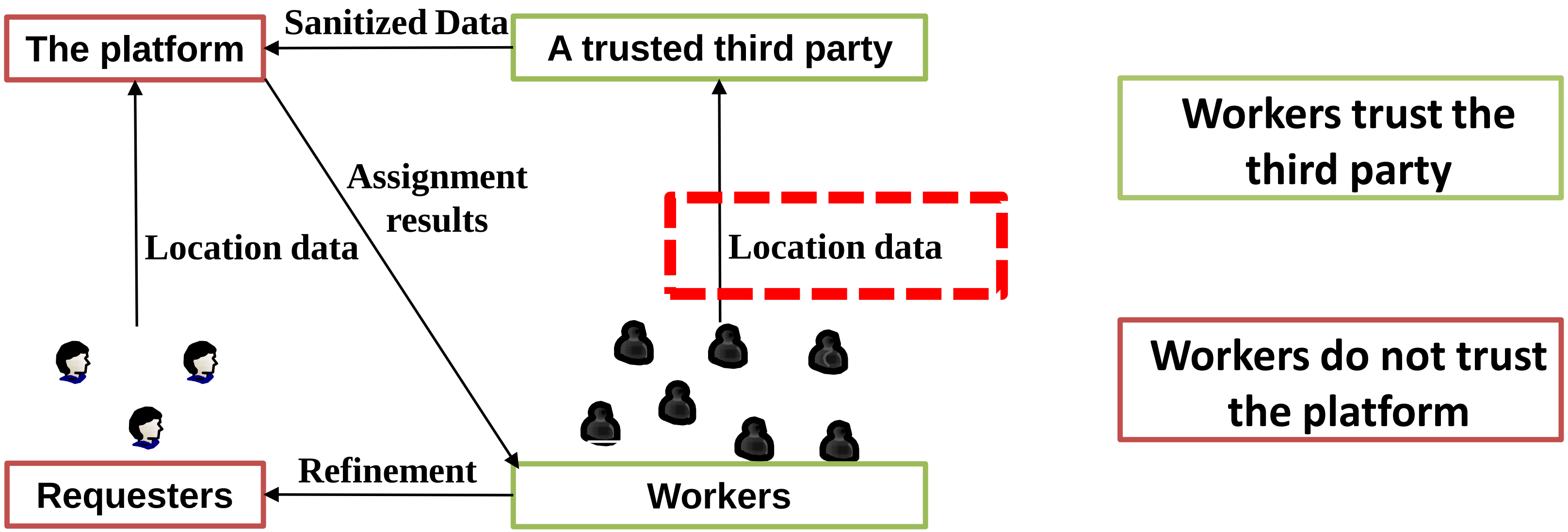
DP: Point Location Protection



- Workers send their locations to a trusted third party
- The third party sanitizes the location of workers according to differential privacy techniques and release it
- The platform performs task assignment according to PSD
- Workers refine the assignment using their exact locations

H. To, G. Ghinita, C. Shahabi. A Framework for Protecting Worker Location Privacy in Spatial Crowdsourcing. VLDB 2014.

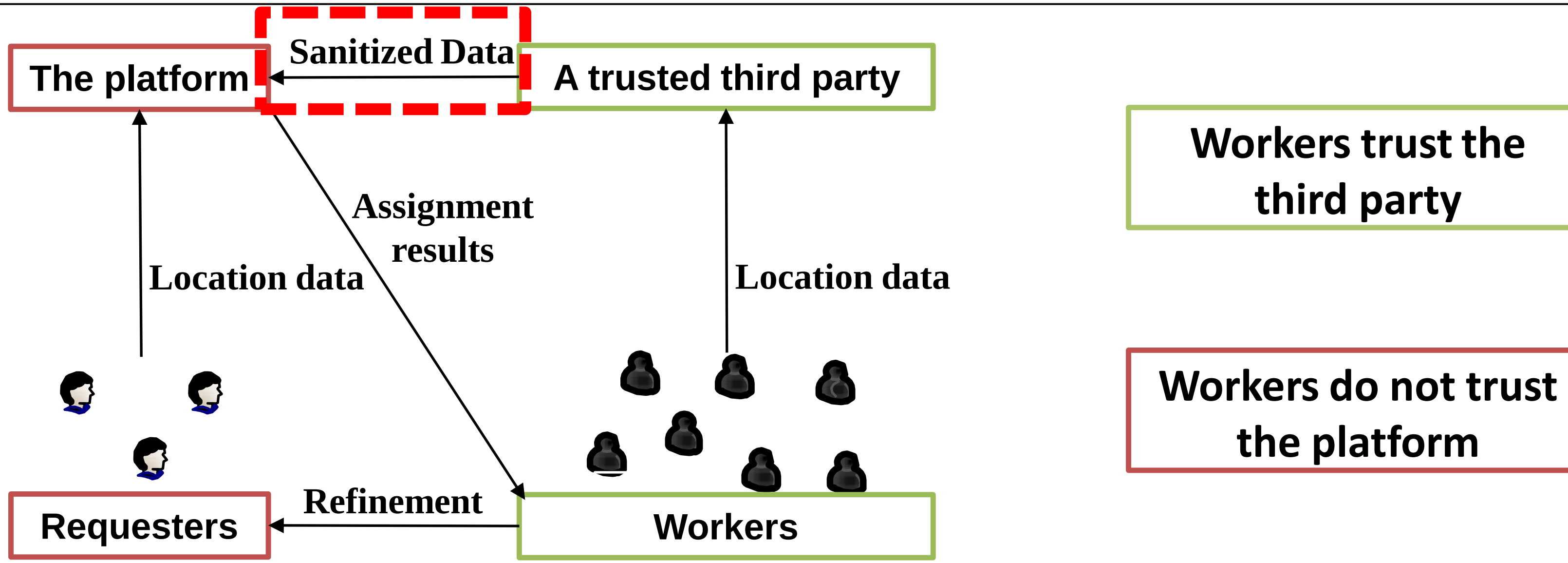
DP: Point Location Protection



- Workers send their locations to a trusted third party
- The third party sanitizes the location of workers according to differential privacy techniques and release it
- The platform performs task assignment according to PSD
- Workers refine the assignment using their exact locations

H. To, G. Ghinita, C. Shahabi. A Framework for Protecting Worker Location Privacy in Spatial Crowdsourcing. VLDB 2014.

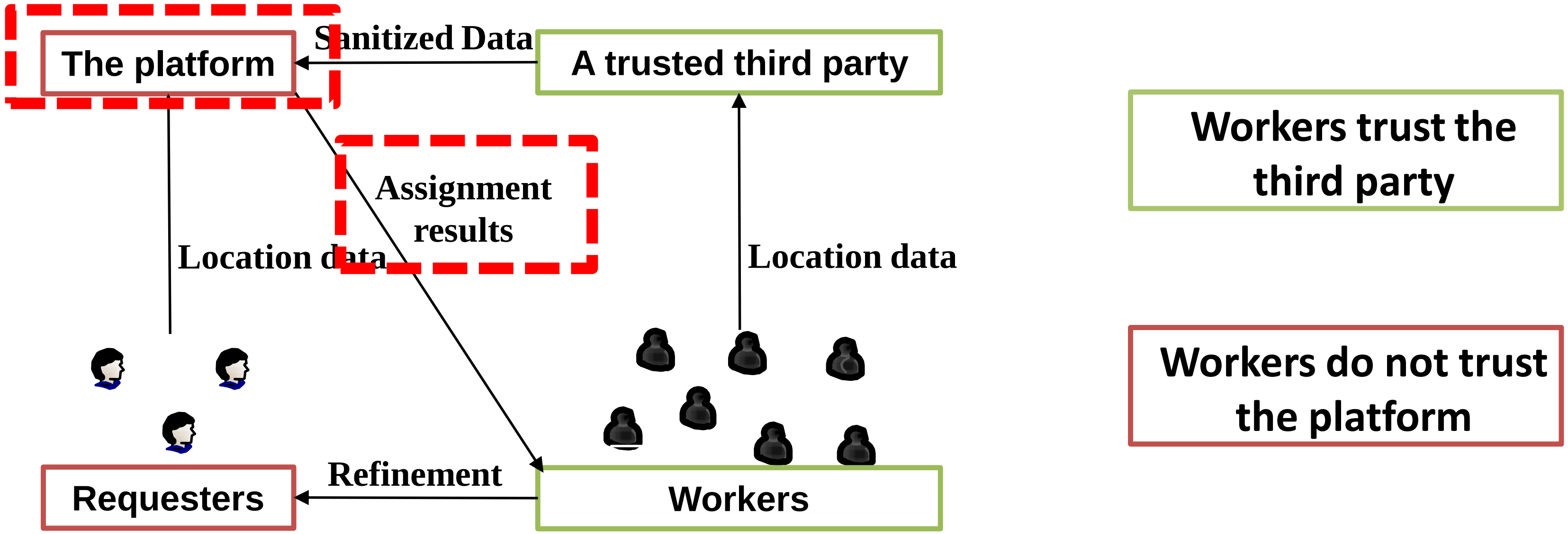
DP: Point Location Protection



- Workers send their locations to a trusted third party
- The third party sanitizes the location of workers according to differential privacy techniques and release it
- The platform performs task assignment according to PSD
- Workers refine the assignment using their exact locations

H. To, G. Ghinita, C. Shahabi. A Framework for Protecting Worker Location Privacy in Spatial Crowdsourcing. VLDB 2014.

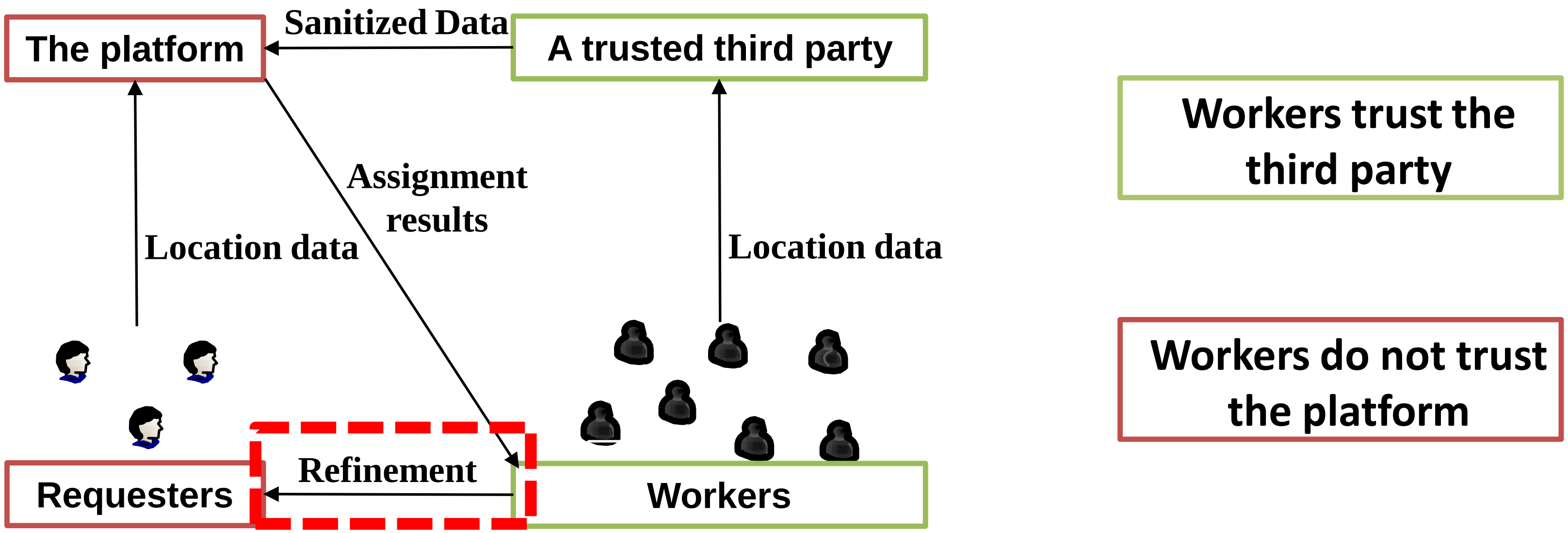
DP: Point Location Protection



- Workers send their locations to a trusted third party
- The third party sanitizes the location of workers according to differential privacy techniques and release it
- The platform performs task assignment according to PSD
- Workers refine the assignment using their exact locations

H. To, G. Ghinita, C. Shahabi. A Framework for Protecting Worker Location Privacy in Spatial Crowdsourcing. VLDB 2014.

DP: Point Location Protection

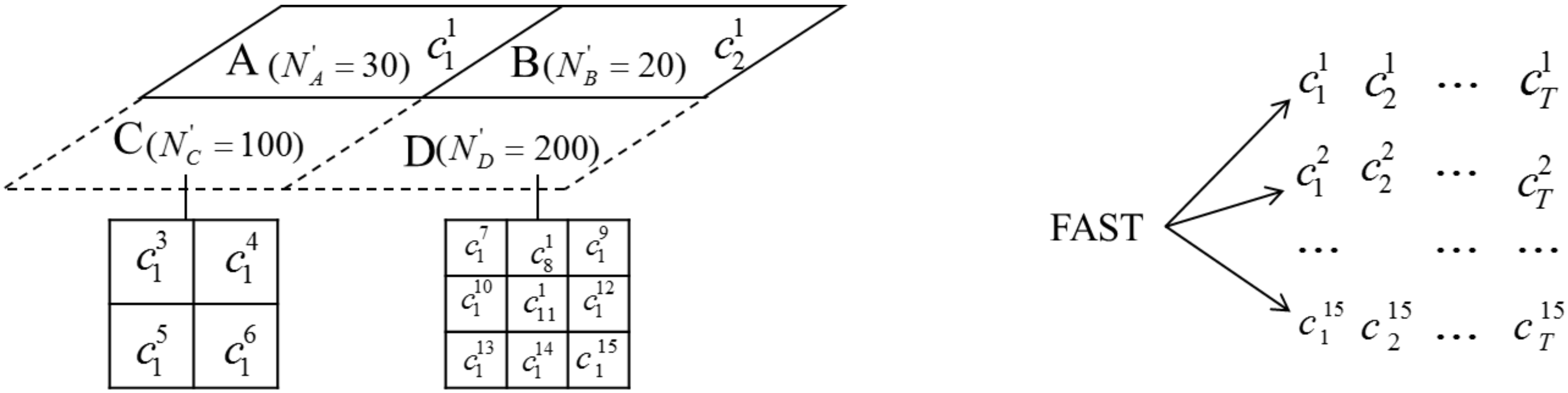


- Workers send their locations to a trusted third party
- The third party sanitizes the location of workers according to differential privacy techniques and release it
- The platform performs task assignment according to PSD
- Workers refine the assignment using their exact locations

H. To, G. Ghinita, C. Shahabi. A Framework for Protecting Worker Location Privacy in Spatial Crowdsourcing. VLDB 2014.

DP: Trajectory Location Protection

- **Trajectory location protection**
  - **Structure of PSD is computed at the first time instance and is reused for all remaining instances (use a fixed privacy budget)**
  - **The remaining budget is used for each grid cell across multiple time instances using FAST – a time-series perturbation technique**



H. To, G. Ghinita, L. Fan, C. Shahabi. Differentially Private Location Protection for Worker Datasets in Spatial Crowdsourcing. TMC, 16(4): 934-949,2017.

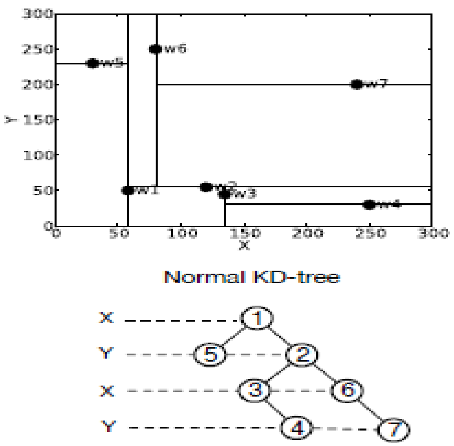
L. Fan, L. Xiong. An Adaptive Approach to Real-Time Aggregate Monitoring With Differential Privacy. TKDE, 26(9): 2094-2106, 2014.

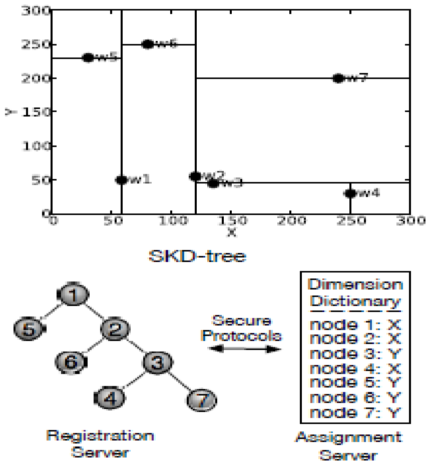
Privacy Protection

- Motivation & A Unified Framework
- Existing Research
  - Cloaked Locations-based Protection
  - Differential Privacy-based Protection
  - Encrypted Data-based Protection
- Summary

Encryption Based Protection

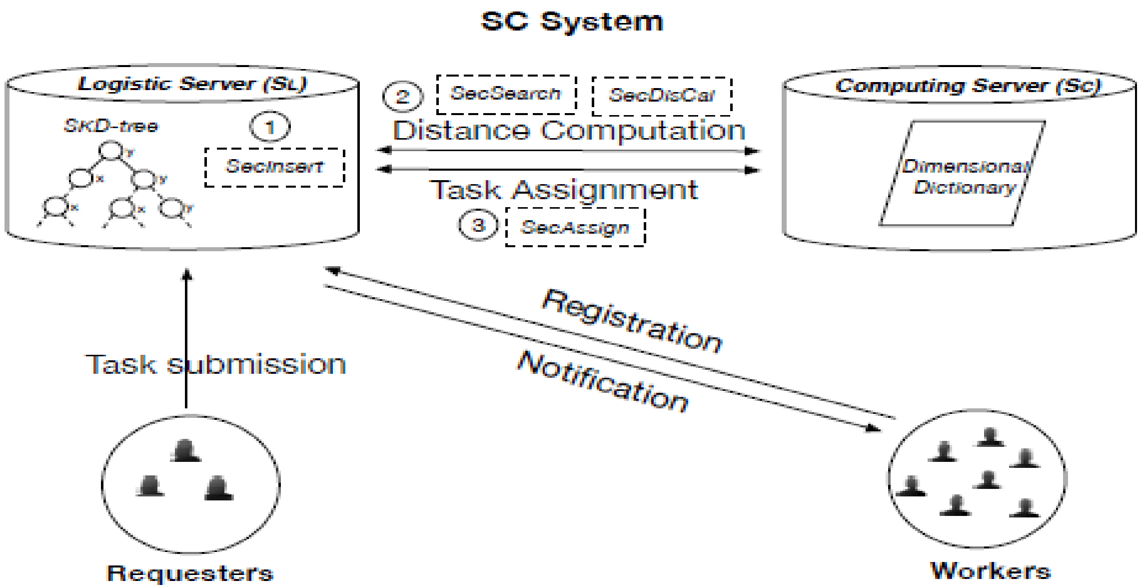
- **Homomorphic Encryption**
  - **The exact distances between tasks and workers can be computed based on their encrypted locations**
- **Secure Indexing**
  - **A secure indexing technique which combines homomorphic encryption with KD-tree**





B. Liu, L. Chen, X. Zhu, Y. Zhang, C. Zhang, W. Qiu. Protecting Location Privacy in Spatial Crowdsourcing using Encrypted Data. EDBT 2017.

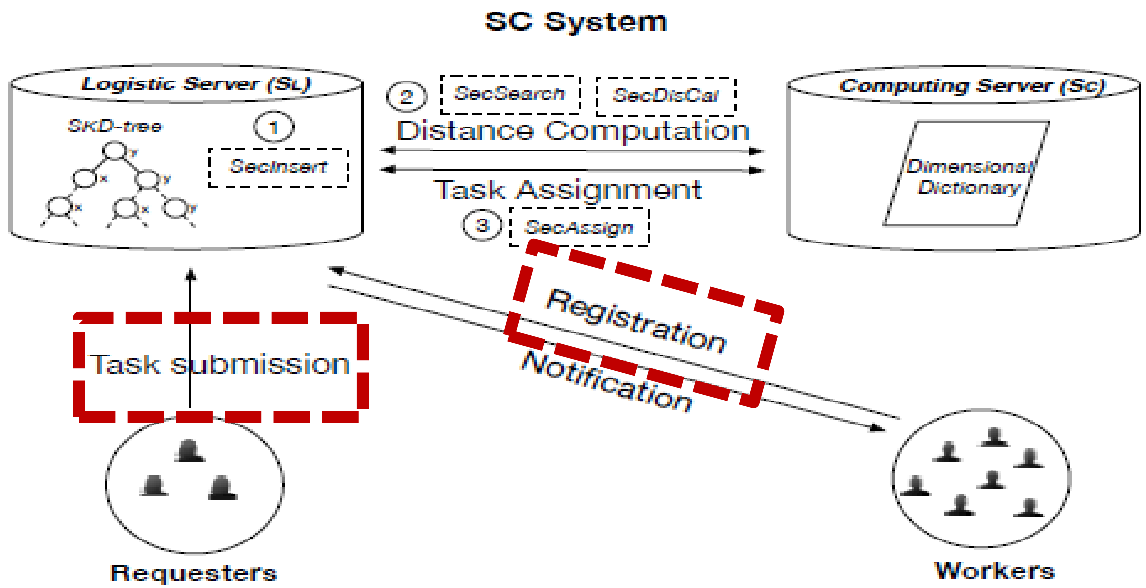
Encryption Based Protection



- The locations of tasks and workers are encrypted by homomorphic encryption
- The platform performs task assignment based on the distances computed by the encrypted data
- The workers receive the encrypted location of the task assigned by the platform, and decrypts it to get the location

B. Liu, L. Chen, X. Zhu, Y. Zhang, C. Zhang, W. Qiu. Protecting Location Privacy in Spatial Crowdsourcing using Encrypted Data. EDBT 2017.

Encryption Based Protection

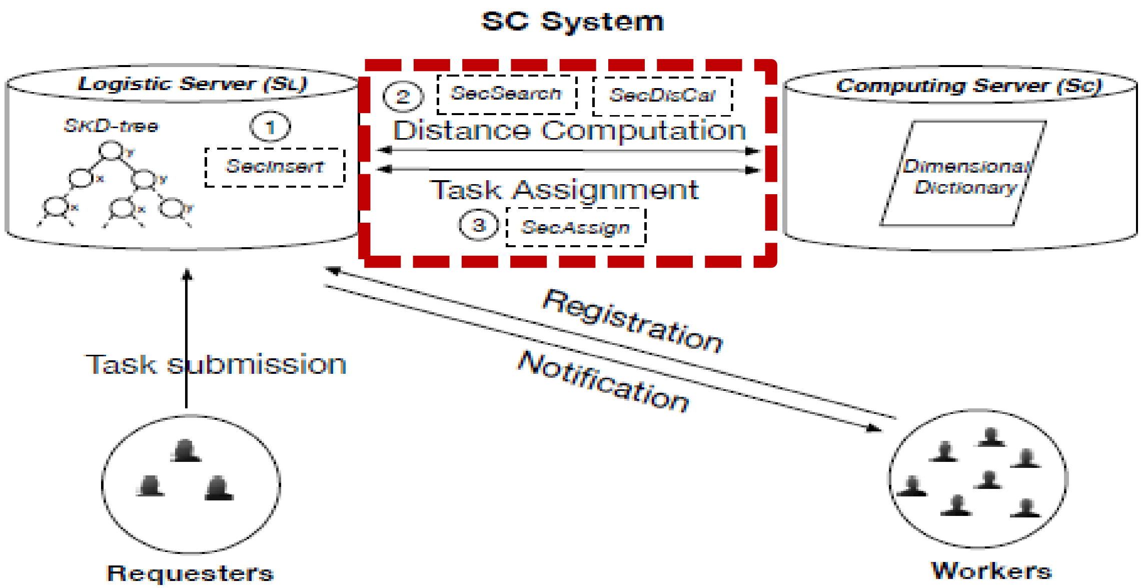


- The locations of tasks and workers are encrypted by homomorphic encryption
- The platfo … based on the distances c…
- The workers receive the encrypted location of the task assigned by the platform, and decrypts it to get the location

Locations of both the workers and the tasks are protected

B. Liu, L. Chen, X. Zhu, Y. Zhang, C. Zhang, W. Qiu. Protecting Location Privacy in Spatial Crowdsourcing using Encrypted Data. EDBT 2017.

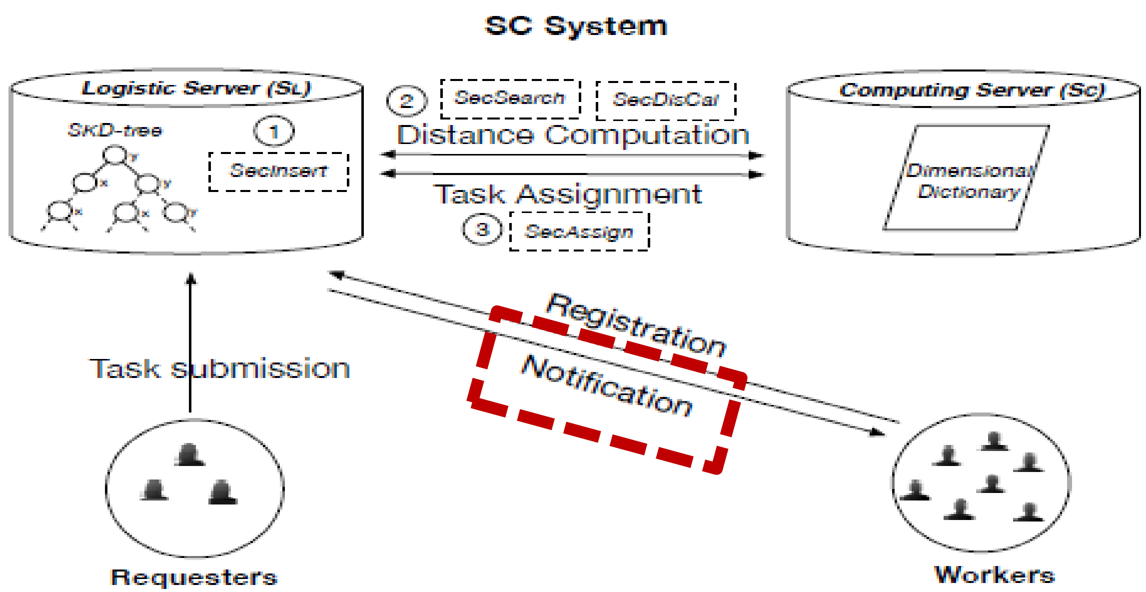
Encryption Based Protection



- The locations of tasks and workers are encrypted by homomorphic encryption
- The platform performs task assignment based on the distances computed by the encrypted data
- The workers receive the encrypted location of the task assigned by the platform, and decrypts it to get the location

B. Liu, L. Chen, X. Zhu, Y. Zhang, C. Zhang, W. Qiu. Protecting Location Privacy in Spatial Crowdsourcing using Encrypted Data. EDBT 2017.

Encryption Based Protection



- The locations of tasks and workers are encrypted by homomorphic encryption
- The platform performs task assignment based on the distances computed by the encrypted data
- The workers receive the encrypted location of the task assigned by the platform, and decrypts it to get the location

B. Liu, L. Chen, X. Zhu, Y. Zhang, C. Zhang, W. Qiu. Protecting Location Privacy in Spatial Crowdsourcing using Encrypted Data. EDBT 2017.

Privacy Protection

- Motivation & A Unified Framework
- Existing Research
  - Cloaked Locations-based Protection
  - Differential Privacy-based Protection
  - Encrypted Data-based Protection
- Summary

Summary

- **Challenge**
  - **Location Security vs. Distance Accuracy**
- **Framework**
  - **Transform the locations of the workers**
  - **The platform performs task assignment based on the transformed locations of the workers**
  - **The workers confirm/refine the task assignment results based on their true locations**
- **State-of-the-art techniques**
  - **Cloaked Locations-based protection**
  - **Differential Privacy-based protection**
  - **Encrypted Data-based protection**

Reference: Privacy Protection

1. L. Pournajaf, L. Xiong, V. S. Sunderam, S. Goryczka. Spatial task assignment for crowd sensing with cloaked locations. In MDM, pages 73-82, 2014.
2. L. Pournajaf, L. Xiong, V. S. Sunderam, X. Xu. STAC: spatial task assignment for crowd sensing with cloaked participant locations. In GIS, pages 90:1-90:4, 2015.
3. G. Cormode, C. M. Procopiuc, D. Srivastava, E. Shen, T. Yu. Differentially private spatial decompositions. In ICDE, pages 20-31, 2012.
4. H. To, G. Ghinita, C. Shahabi. A framework for protecting worker location privacy in spatial crowdsourcing. PVLDB, 7(10):919-930, 2014.
5. H. To, G. Ghinita, L. Fan, C. Shahabi. Differentially private location protection for worker datasets in spatial crowdsourcing. TMC, 16(4):934-949, 2017.
6. L. Fan, L. Xiong. An Adaptive Approach to Real-Time Aggregate Monitoring With Differential Privacy. TKDE, 26(9): 2094-2106, 2014.
7. L. Pournajaf, D. A. Garcia-Ulloa, L. Xiong, V. S. Sunderam. Participant privacy in mobile crowd sensing task management: A survey of methods and challenges. SIGMOD Record, 44(4):23-34, 2015.
8. H. To, G. Ghinita, C. Shahabi. PrivGeoCrowd: A toolbox for studying private spatial crowdsourcing. In ICDE, pages 1404-1407, 2015.
9. Y. Xiao, L. Xiong. Protecting locations with differential privacy under temporal correlations. In CCS, pages 1298-1309, 2015.

Reference: Privacy Protection

10. X. Jin, R. Zhang, Y. Chen, T. Li, Y. Zhang. DPSense: differentially private crowdsourced spectrum sensing. In CCS, pages 296-307, 2016.
11. B. Liu, L. Chen, X. Zhu, Y. Zhang, C. Zhang, W. Qiu. Protecting location privacy in spatial crowdsourcing using encrypted data. In EDBT, pages 478-481, 2017.
12. L. Wang, D. Zhang, D. Yang, B. Y. Lim, X. Ma. Differential location privacy for sparse mobile crowdsensing. In ICDM, pages 1257-1262, 2016.
13. X. Jin, Y. Zhang. Privacy-preserving crowdsourced spectrum sensing. In INFOCOM, pages 1-9, 2016.
14. L. Wang, D. Yang, X. Han, T. Wang, D. Zhang, X. Ma. Location privacy-preserving task allocation for mobile crowdsensing with differential geo-obfuscation. In WWW, pages 627-636, 2017.
15. H. To, C. Shahabi. Location Privacy in Spatial Crowdsourcing. arXiv preprint arXiv:1704.06860, 2017

Outline

- **Overview of Spatial Crowdsourcing (20min)**
  - **Motivation**
  - **Workflow**
  - **Core Issues**
  - **Difference from Related Tutorials**
- **Fundamental Techniques (50min)**
  - **Task Assignment**
  - **Quality Control**
  - **Incentive Mechanism**
  - **Privacy Protection**
- **Spatial Crowdsourced Applications (15min)**
  - **Spatial Crowdsourcing Intrinsic Applications**
  - **Crowd-powered Spatial Applications**
- **Open Questions (5min)**

Outline

- Overview of Spatial Crowdsourcing (20min)
  - Motivation
  - Workflow
  - Core Issues
  - Difference from Related Tutorials
- Fundamental Techniques (50min)
  - Task Assignment
  - Quality Control
  - Incentive Mechanism
  - Privacy Protection
- Spatial Crowdsourced Applications (15min)
  - Spatial Crowdsourcing Intrinsic Applications
  - Crowd-powered Spatial Applications
- Open Questions (5min)

Intrinsic Applications

- What are the Intrinsic Applications
  - Naturally modelled by spatial crowdsourcing
  - Share the same core issues with spatial crowdsourcing
- Two kinds of representative applications
  - Real-time taxi calling service
  - Food delivery service

Intrinsic Applications

- Real-time taxi calling service

Passenger is the task requester

The task is to pick up the passenger at location A and send her/him to location B

Intrinsic Applications

- **Real-time taxi calling service**

Passengers and available taxis dynamically appear on the platform

A typical dynamic matching problem

UBER

Platform

Uber is the platform, and its core concern is how to assign taxis to pick up different passengers efficiently

Intrinsic Applications

- **Real-time taxi calling service**

Task Assignment

UBER

Platform

Incentive Mechanism

Quality Control

Privacy Protection

Intrinsic Applications

- Food delivery service

The task is to pick up the food and send the food to the user

User is the task requester

Intrinsic Applications

- Food delivery service

Orders and available deliverers dynamically appear on the platform

A typical **dynamic planning** problem

Platform

Grubhub is the platform, and its core concern is how to design effective plans for the deliverers

Intrinsic Applications

- Food delivery service

Task Assignment

Orders and available deliverers dynamically appear on the platform

Platform

Incentive Mechanism

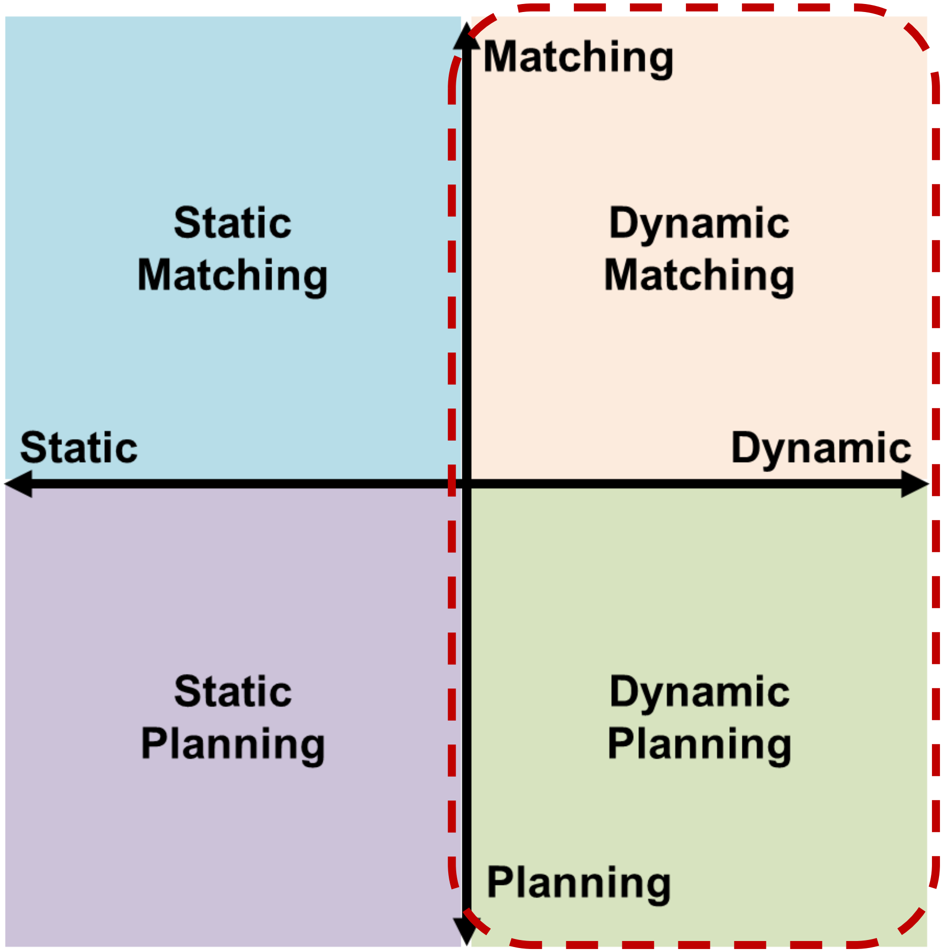
Quality Control

Privacy Protection

Intrinsic Applications

- **Ridesharing Service**

Intrinsic Applications



Outline

- Overview of Spatial Crowdsourcing (20min)
  - Motivation
  - Workflow
  - Core Issues
  - Difference from Related Tutorials
- Fundamental Techniques (50min)
  - Task Assignment
  - Quality Control
  - Incentive Mechanism
  - Privacy Protection
- Spatial Crowdsourced Applications (15min)
  - Spatial Crowdsourcing Intrinsic Applications
  - Crowd-powered Spatial Applications
- Open Questions (5min)

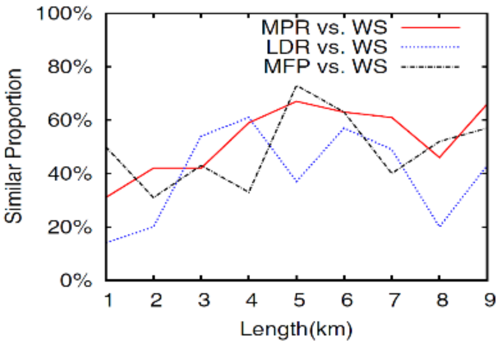
Outline

- Crowd-powered Spatial Applications
  - Crowdsourced Path Selection
  - Crowdsourced Speed Estimation
  - Crowdsourced POI Labelling

Path Selection

- Motivation
  - Route recommendation service is indispensable in daily life
- Solutions without crowdsourcing
  - Using popular routes mined from historical trajectories as recommended routes
    - Most Popular Route
    - Local Driver Route
    - Most Frequent Path

- Drawbacks
  - Sparsity: insufficient historical trajectories for inference
  - Diversity: different algorithms, different routes



Path Selection

- **Solutions using spatial crowdsourcing**
  - **Leverage crowds' knowledge to improve the recommendation quality**
- **Main Idea**
  - **Consolidate candidate routes from different sources (e.g., map service providers, popular routes)**
  - **Request experienced drivers to select amongst them**

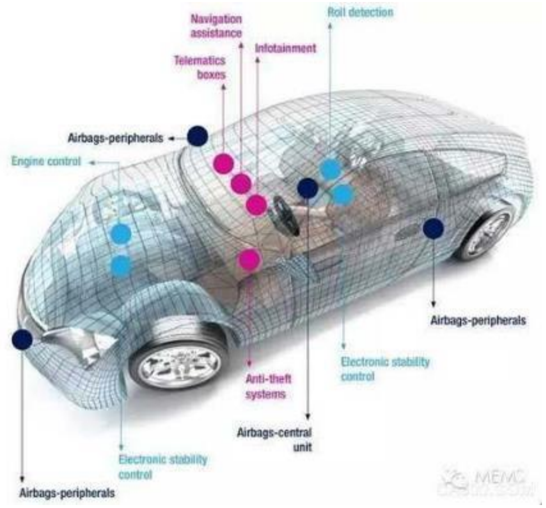
H. Su, K. Zheng, J. Huang, H. Jeung, L. Chen, X. Zhou. CrowdPlanner: A crowd-based route recommendation system. ICDE 2014.

Outline

- **Crowd-powered Spatial Applications**
  - **Crowdsourced Path Selection**
  - **Crowdsourced Speed Estimation**
  - **Crowdsourced POI Labelling**

Traffic Speed Estimation

- Motivation
  - Real-time traffic speed represents the congestion of road and is one of the most important aspects for traffic monitoring
- Speed Estimation without Crowdsourcing
  - Use sensor or trajectory data for speed estimation
  - It is hard to infer the speed of remote roads



Traffic Speed Estimation

- **Solutions using Spatial Crowdsourcing**
  - **Assign crowd workers to physically check traffic speed of some ambiguous roads or remote roads**
- **Key Idea**
  - **A two-layer framework for speed estimation**
    - **Assign k roads (called seed road) to crowd workers to generate their real speeds**
    - **Infer speeds of other roads according to seed roads and historical speed information**

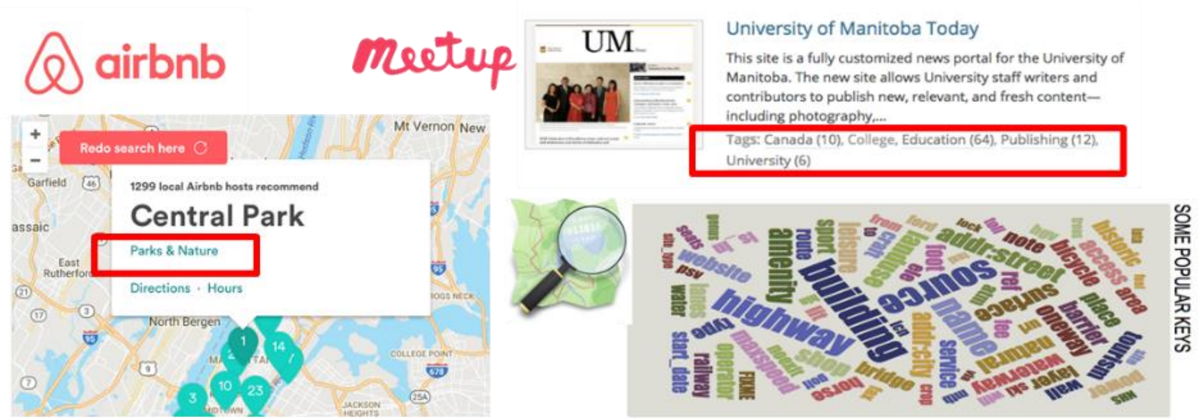
H. Hu, G. Li, Z. Bao, Y. Cui, J. Feng. Crowdsourcing-Based Real-Time Urban Traffic Speed Estimation: From Trends to Speeds. ICDE 2016.

Outline

- **Crowd-powered Spatial Applications**
  - **Crowdsourced Path Selection**
  - **Crowdsourced Speed Estimation**
  - **Crowdsourced POI Labelling**

POI Labelling

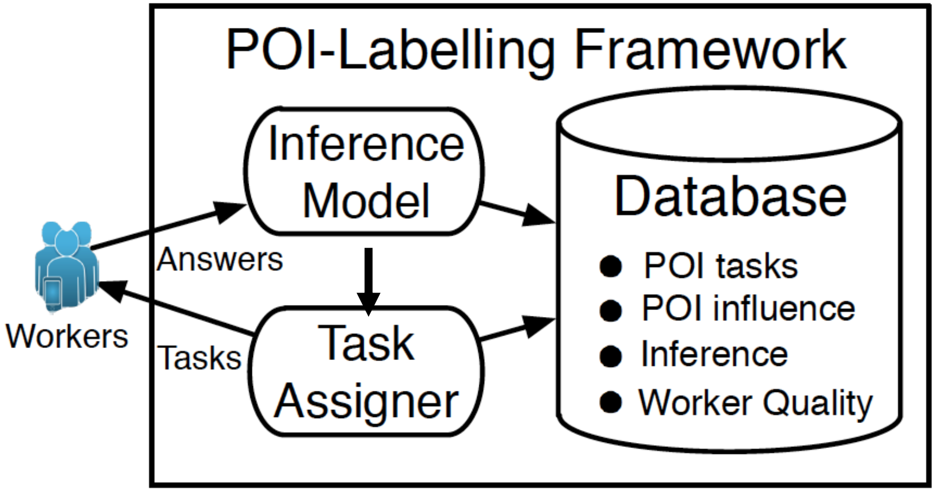
- **Motivation**
  - **POI (Point of Interest) is one of the most useful and fundamental spatial data**



- **Incorrect labels exist**
  - **Low quality POI labels from volunteers**
  - **Limited accuracies of AI algorithms**

POI Labelling

- **Solutions using Spatial Crowdsourcing**
  - **An iterative framework**
    - **Assign tasks to workers**
    - **Collect the results and infer the truth of POIs**
    - **Repeat until the budget is exhaustive**



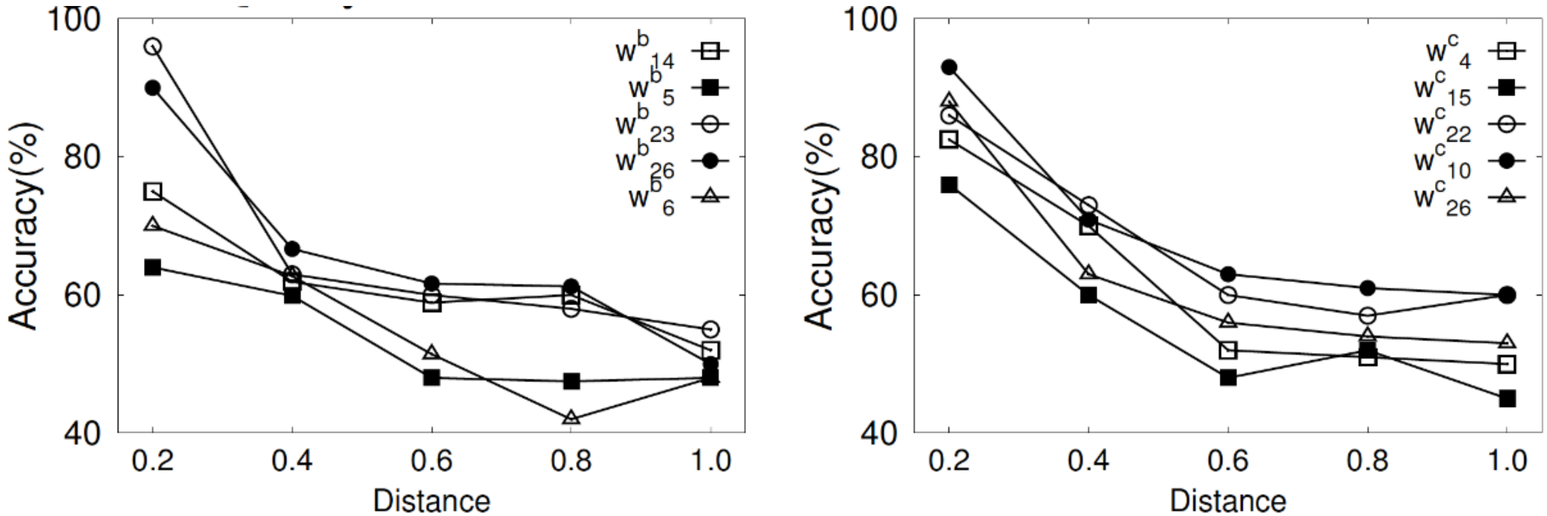
Framework

An Example

H. Hu, Y. Zheng, Z. Bao, G. Li, J. Feng, R. Cheng. Crowdsourced POI Labelling: Location-Aware Result Inference and Task Assignment. ICDE 2016.

POI Labelling

- **Experimental Findings**
  - **The distance between POIs and workers has a huge impact on POI labelling**



(a) Beijing (b) China

The longer distance, the lower accuracy

H. Hu, Y. Zheng, Z. Bao, G. Li, J. Feng, R. Cheng. Crowdsourced POI Labelling: Location-Aware Result Inference and Task Assignment. ICDE 2016.

Summary

	Spatial crowdsourced intrinsic applications			Crowd-powered spatial applications		
	Taxi Calling	Ride Sharing	Food Delivery	SC Path Selection	SC Speed Estimation	SC POI Labelling
Task Assignment	✓	✓	✓	✓	✓	✓
Quality Control	✓	✓	✓	✓	✓	✓
Incentive Mechanism	✓	✓	✓	✓	✗	✓
Privacy Protection	✓	✓	✓	✗	✗	✗

Spatial crowdsourced intrinsic applications: Pure human effort
Crowd-powered spatial applications: Hybrid human-machine effort

Reference: Applications

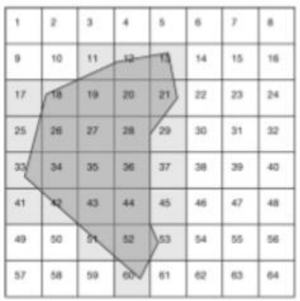
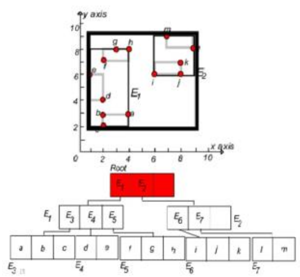
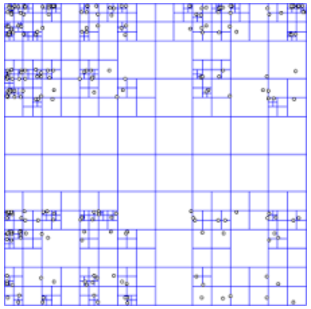
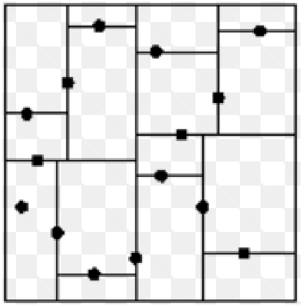
1. H. Su, K. Zheng, J. Huang, H. Jeung, L. Chen, X. Zhou. CrowdPlanner: A crowd-based route recommendation system. In ICDE, pages 1144-1155, 2014.
2. C. Zhang, Y. Tong, L. Chen. Where To: Crowd-Aided Path Selection. PVLDB, 7(14):2005-2016, 2014.
3. A. Artikis, M. Weidlich, F. Schnitzler, I. Boutsis, T. Liebig, N. Piatkowski, C. Bockermann, K. Morik ,V. Kalogeraki, J. Marecek, A. Gal, S. Mannor, D. Gunopulos, D. Kinane. Heterogeneous stream processing and crowdsourcing for urban traffic management. In EDBT, pages 712-723, 2014
4. H. Hu, G. Li, Z. Bao, Y. Cui, J. Feng. Crowdsourcing-based real-time urban traffic speed estimation: From trends to speeds. In ICDE, pages 883-894, 2016.
5. H. Hu, Y. Zheng, Z. Bao, G. Li, J. Feng, R. Cheng. Crowdsourced POI labelling: location-aware result inference and task assignment. In ICDE ,pages 61-72, 2014.
6. W. Ouyang, M. Srivastava, A. Toniolo, T. Norman. Truth discovery in crowdsourced detection of spatial events. TKDE, 28(4): 1047-1060, 2016.
7. D. A. Garcia-Ulloa, L. Xiong, V. S. Sunderam. Truth discovery for spatiotemporal events from crowdsourced data. PVLDB 10(11): 1562-1573, 2017.
8. Y. Huang, F. Bastani, R. Jin, X. Wang: Large scale real-time ridesharing with service guarantee on road networks. PVLDB 7(14): 2017-2028, 2014.
9. S. Ma, Y. Zheng, O. Wolfson. Real-time city-scale taxi ridesharing. TKDE, 27(7): 1782-1795, 2015
10. M. Asghari, D. Deng, C. Shahabi, U. Demiryurek, Y. Li. Price-aware real-time ride-sharing at scale: an auction-based approach. In GIS, pages 3:1-3:10, 2016.

Outline

- Overview of Spatial Crowdsourcing (20min)
  - Motivation
  - Workflow
  - Core Issues
  - Difference from Related Tutorials
- Fundamental Techniques (50min)
  - Task Assignment
  - Quality Control
  - Incentive Mechanism
  - Privacy Protection
- Spatial Crowdsourced Applications (15min)
  - Spatial Crowdsourcing Intrinsic Applications
  - Crowd-powered Spatial Applications
- Open Questions (5min)

Question 1: Index

- **Abundant indexes in spatial data management**
  - **Grid, R-tree, Quadtree, kd-tree, etc**

- **How to leverage existing indexes or design new indexes to optimize the efficiency of fundamental issues in spatial crowdsourcing?**

Question 2: Dynamic Scenarios

- **Although some recent research has focused on dynamic task assignment, there are several big gaps between theory and practice**
  - **E.g., good performance in practice versus bad theoretical results of the simple greedy algorithm for the online minimum matching problem**

Question 2: Dynamic Scenarios

- The research of other core issues of spatial crowdsourcing in dynamic scenario is limited
  - How to design adaptive incentive mechanisms which can handle the dynamic supply and demand between tasks and workers?
  - How to protect the location privacy of the dynamic arrival tasks and workers

Question 3: Benchmark

- Abundant benchmarks for many classical spatial data management
  - E.g., DIMACS for shortest path

- Although there are a few synthetic data generators for spatial crowdsourcing, public real datasets are still inadequate
  - Platforms owning abundant real data are usually commercial, and would not like to open their data
  - Open sourced platforms do not have enough money recruitment enough workers


dank u
ju faleminderit
Tack
Asante
谢谢
Tak
mulţumesc
kiitos
Salamat!
Gracias
Terima kasih
Aliquam
Merci
Dankie
Obrigado
ありがとう
köszönöm
grazie
Aliquam
Go raibh maith agat
děkuji
Thank you
gam