Location Privacy-preserving in Crowdsensing System*

Professor Wanlei Zhou

Head of School of Software
University of Technology Sydney, Australia
wanlei.zhou@uts.edu.au
https://www.uts.edu.au/staff/wanlei.zhou

*Work is done by members of my group and through collaborations with colleagues in other universities.

Overview

- Background and related work
- Density-based location privacy preserving
- Blockchain-based location privacy preserving
- Economic model based location privacy preserving
- Conclusions

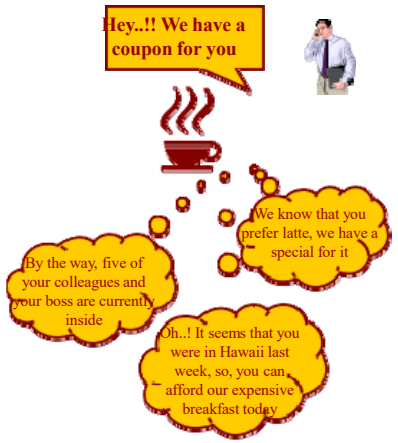
Background: Location based services in smart cities

Location based services (LBS)

1. Traffic reports and navigation
2. Place of interest (POI) finder
3. Advertisement
4. Mobile crowd sensing (MCS) applications

Privacy issues in LBS

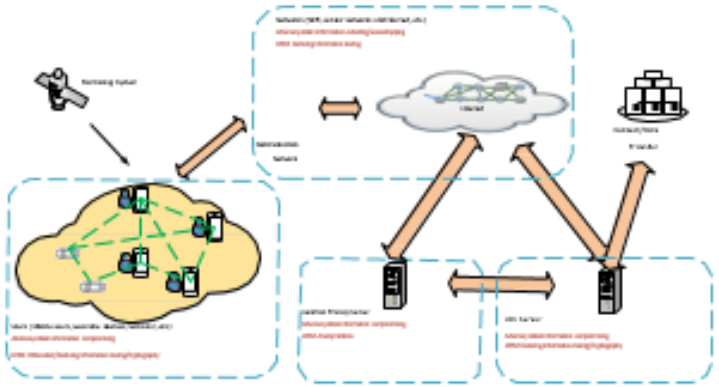
1. Location privacy – discrete physical locations.
2. Trajectory privacy – a path or trace in the geographical space.



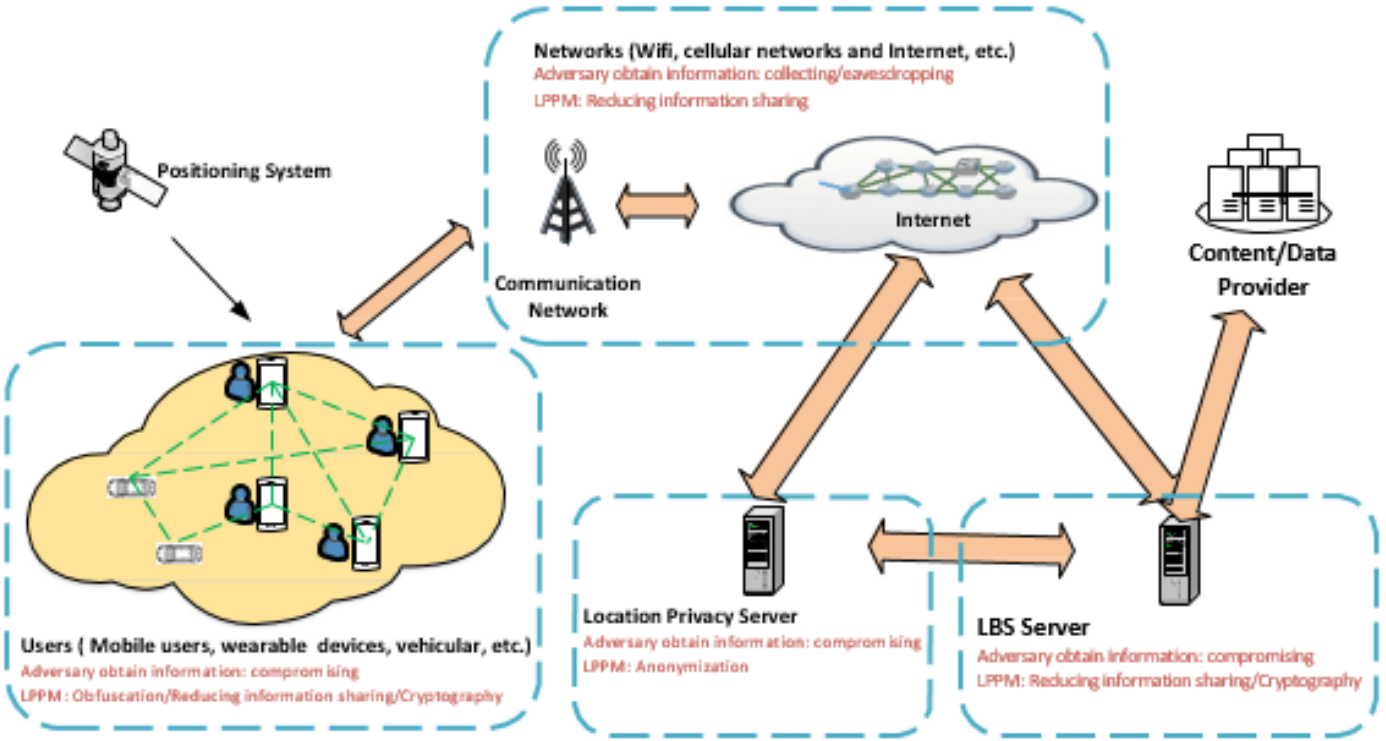
Background: LBS

Components of the LBS system

1. Positioning System
2. Users
3. Network
4. LBS Server
5. Content/Data Provider
6. Location Privacy Server



Background: LBS



Background: LBS

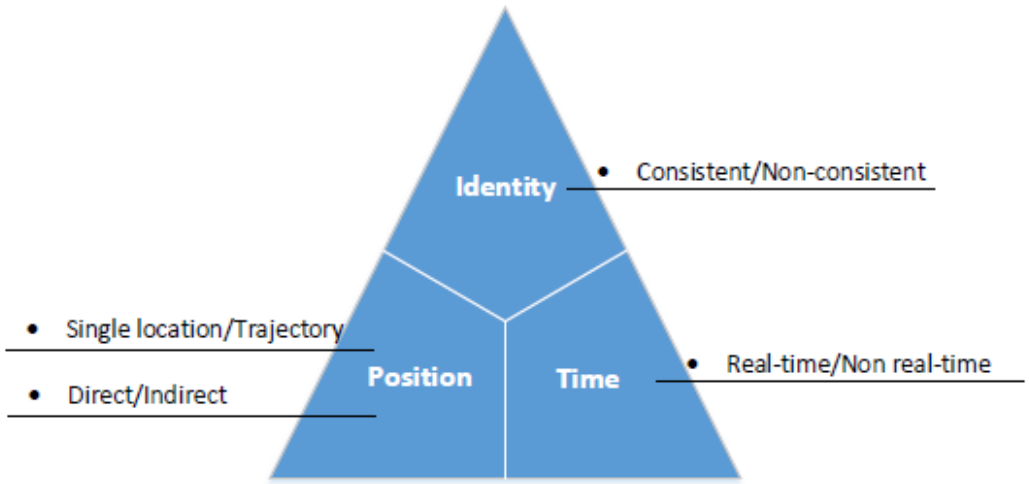
Five Aspects of Location Privacy Research

1. Location information
2. Location attacks and adversaries
3. Location privacy preserving mechanisms (LPPMs)
4. Location privacy metrics
5. Location privacy application

Background: LBS

Location Information in LBS

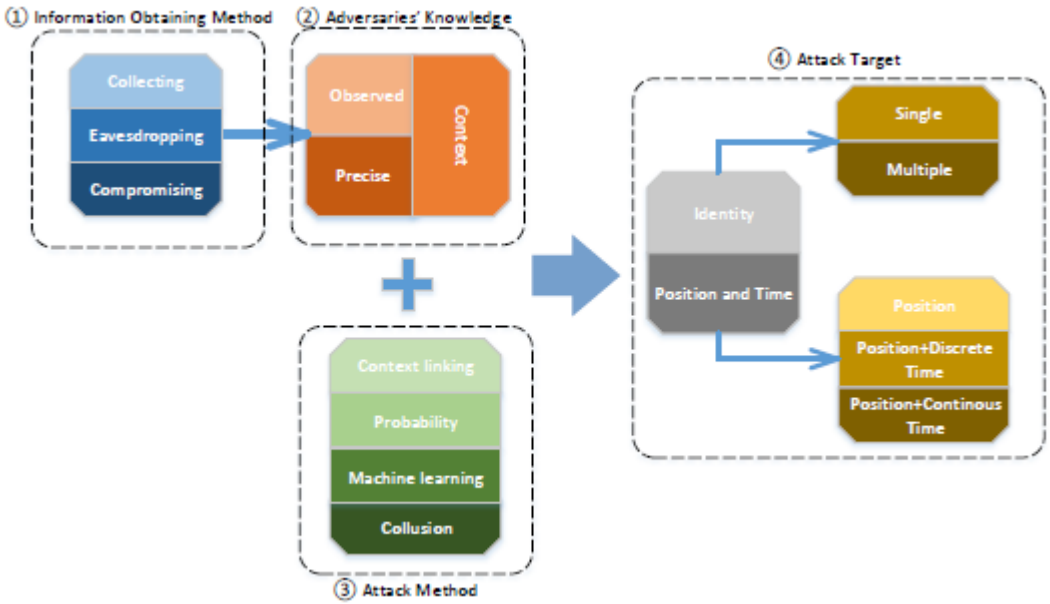
1. Identity: user's name or any feature that makes a person distinguishable.
2. Position: spatial information, described by a coordinate.
3. Time: time stamps associated with the location.



Background: LBS

Describe the Attacks and Adversaries

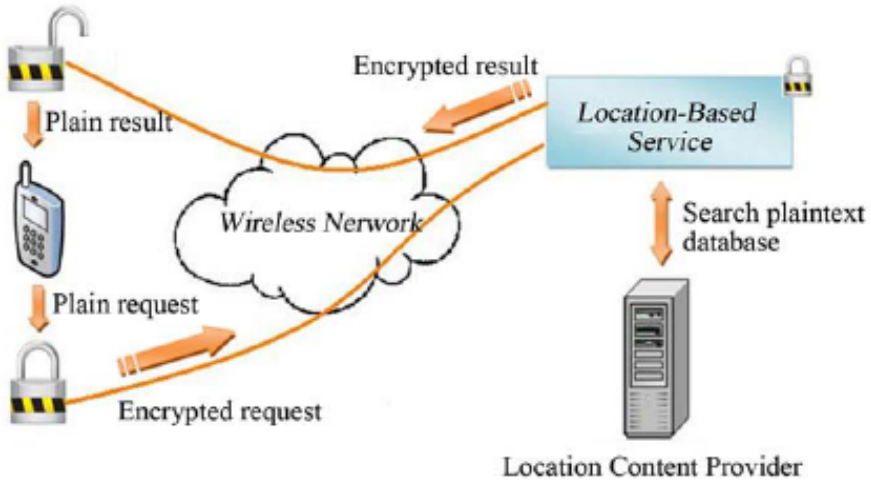
1. How they obtain the information.
2. How the attack is launched.
3. What the information they obtained (knowledge).
4. What is the target



Background: LBS

[1] Cryptography-based approaches

1. High computational complexity
2. Rely on the trustworthy sever



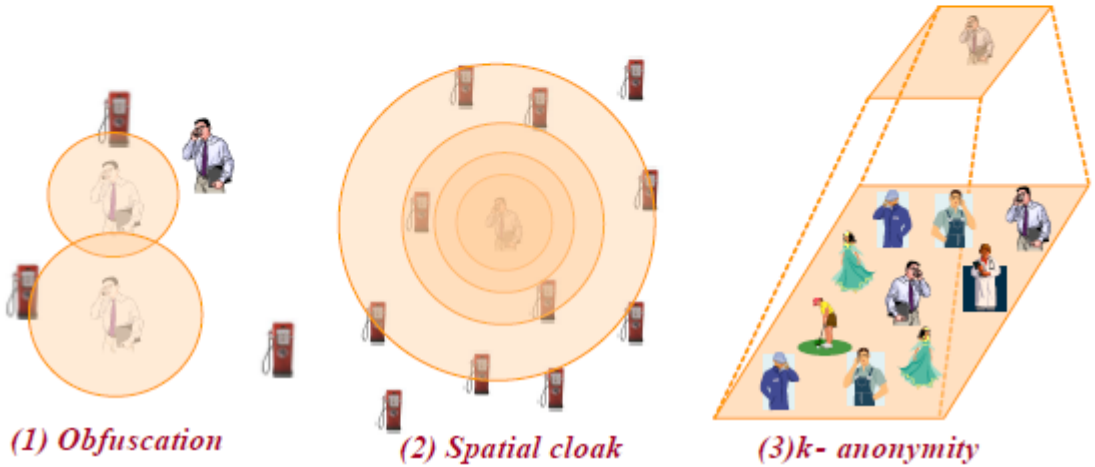
Background: LBS

[2] Anonymization Mechanisms

1. Mix-zone (spatial cloak)
2. *k*-anonymity

[3] Obfuscation Mechanisms

1. Location obfuscation
2. Dummy locations
3. Differential privacy based method



(1) Obfuscation *(2) Spatial cloak* *(3)k- anonymity*

Background: LBS

[4] Reducing Location Information Sharing

Caching: Pre-download before use

Game Theory

New Protocols

Background: Crowdsensing

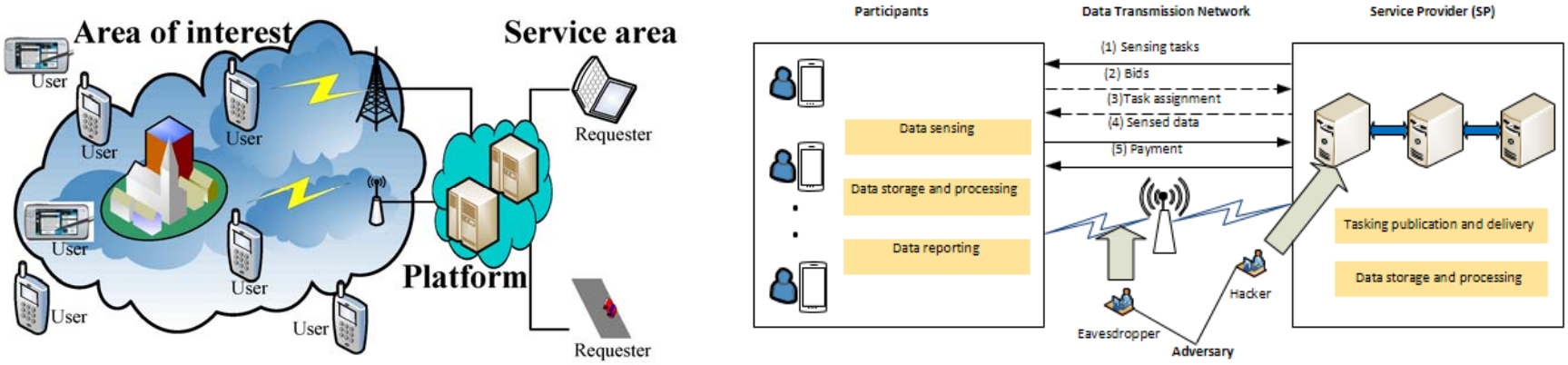


Figure: System architecture of mobile crowdsensing applications.

Mobile crowdsensing

Background: Crowdsensing

Two models of task assignment

- *Worker selected tasks* (WST): The platform (server) publishes the tasks and the workers autonomous select the ones they prefer.

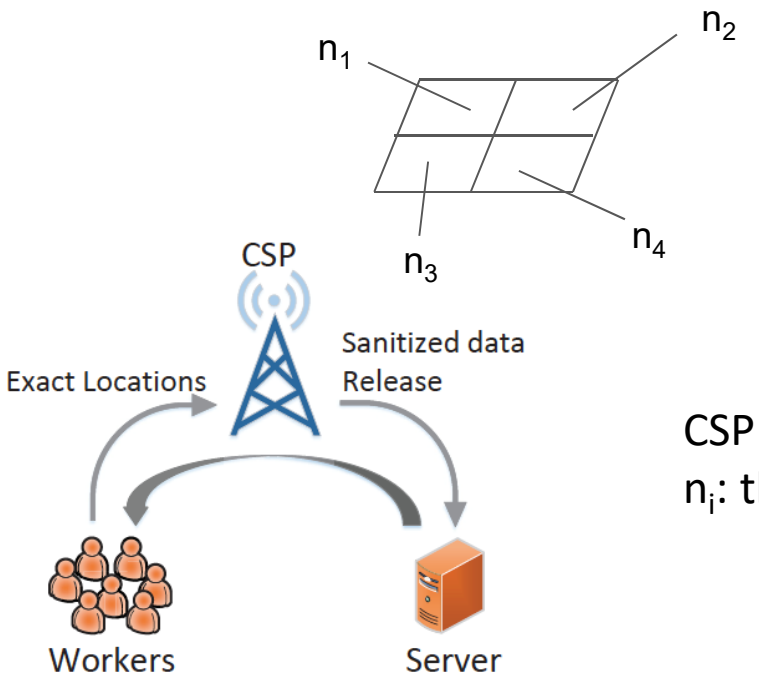
- *Server assigned tasks* (SAT): The worker first reports their location information to the platform (server), and the server assigns tasks according to the worker's location.

Background: Crowdsensing

Three ways of privacy disclosure:

- Workers submit their exact location to the server to be allocated tasks more efficiently. For example: Alice submits her location coordinates (23.23, 151.2), the server knows Alice is in hospital now.
- When a worker accepts an assigned task, the server knows that worker's future location. For example: Alice accepts the task A, then the server knows that Alice will go to task A's location.
- After completing the task, the server processes their payment, so it knows the task the worker completed. As such, completing a task reveals the worker's previous locations.

Density-based location privacy preserving



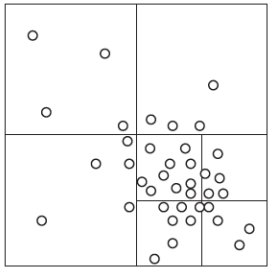
CSP: Trusted cell service provider.

n_i: the noisy number of workers in cell i.

Density-based location preservation

Problems:

- Assuming the distribution of the workers uniform is not practical.

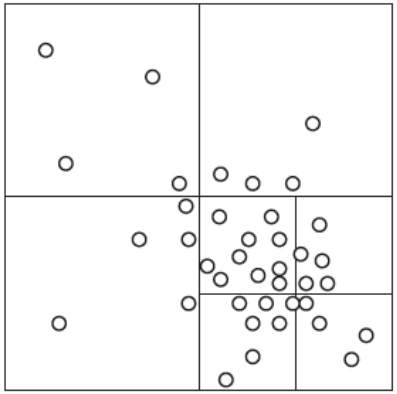


The partition is data-independent.
The midpoint is always chosen to partition the parent cell

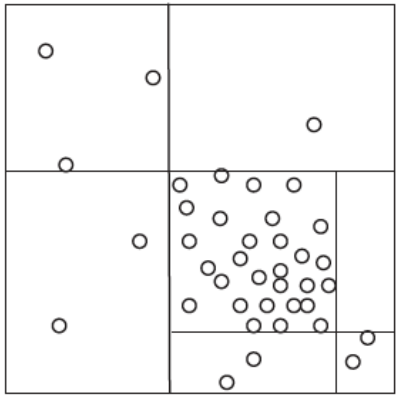
- Cannot prevent the location privacy disclosure in third way (payment process).

Density-based location preservation

To solve the problem of uneven distribution problem.



Traditional quadtree



Density-based quadtree

Density-based location preservation

The density-based partition method

- Select several initial partition points in the location domain
- Calculate the differences in density between the subcells partitioned by each partition point.
- Choose the subcells with the biggest differences in density as partitions.
- Repeat the whole process for each subcell until the stop condition is met.

Three stop conditions

- No workers exist in the cell.
- The cell is too small to be further partitioned
- The distribution of workers in the cell is relatively uniform.
  The maximum density difference Δd is used to measure whether the worker location distribution is uniform.

$$\Delta d(cell, p) = \max_{c_i \in C}\{den(c_i)\} - \min_{c_i \in C}\{den(c_i)\}$$

Density-based location preservation

Differential privacy data release

Algorithm 2 Differential privacy data release

Require: Spatial decomposition SD
Ensure: Sanitized data SSD
1: **for** $c_i \in SD$ **do**
2: $n_i \leftarrow$ number of workers in c_i;
3: $N_i = n + Laplace(\frac{s}{\epsilon})$;
4: **end for**
5: **return** $SSD = \{r_1, r_2, ..., r_m\}$

Density-based location preservation

Task assignment

- Worker acceptance probability

$$p_w = f(d_{w,t})$$

$$f(d_{w,t}) = \begin{cases} \frac{d_{mtd} - d_{w,t}}{d_{mtd}}, & d_{w,t} \leq d_{mtd} \\ 0, & d_{w,t} > d_{mtd} \end{cases}$$

$$p_c = 1 - (1 - p_w)^n$$

- Geocast region selection

Standard:

1. The distance between the selected cell and the task should be as small as possible and within the maximum travel distance of the workers.
2. The acceptance probability of the geocast region should reach the expected task assignment success rate.
3. The number of notified workers should be as small as possible.

Density-based location preservation

Task assignment

- Geocast region selection

Algorithm 3 Geocast region selection

Require: SSD
Ensure: GR
1: Order $SSD = \{r_1, r_2, ..., r_n\}$, where $d_{r_1,t} \leq d_{r_2,t} \leq ... \leq d_{r_n,t}$
2: Choose r_1 as the initial geocast region GR.
3: **repeat**
4: Expanding GR by adding the closest cell in the remaining cells one by one;
5: **until** $p_{ar} < ES$ or $d_{r_i,t} > d_{mtd}$
6: Calculate the expectation of travel distance: $Ed = \sum_{i=1}^{m} p_{r_i} d_{r_i,t}$;
7: Calculate the number of workers in GR: $N = \sum_{i=1}^{m} n_{r_i}$;
8: $S \leftarrow$find c_i that $p_{r_i} > ES$, $n_{r_i} \leq N$ and $d_{r_i} \leq d_{mtd}$;
9: **if** $S \neq \emptyset$ **then**
10: **for** $r_i \in S$ **do**
11: find the cell r_i has the shortest distance to l_t;
12: **end for**
13: **if** $p_{r_i} d_{r_i,t} < Ed$ **then**
14: $GR = r_i$;
15: **end if**
16: **end if**
17: **return** GR

Density-based location preservation

Simulation settings

- Metrics
  1. Task assignment success rate (TASR)
  2. Average travel distance (ATD)
  3. Average notified workers (ANW)
- Comparison
  1. DP-GRB: the proposed method
  2. UP-GRB: the method with uniformed partitions and a balanced geocast region construction.
  3. UP-GRS: gives priority to the task assignment success rate.

Density-based location preservation

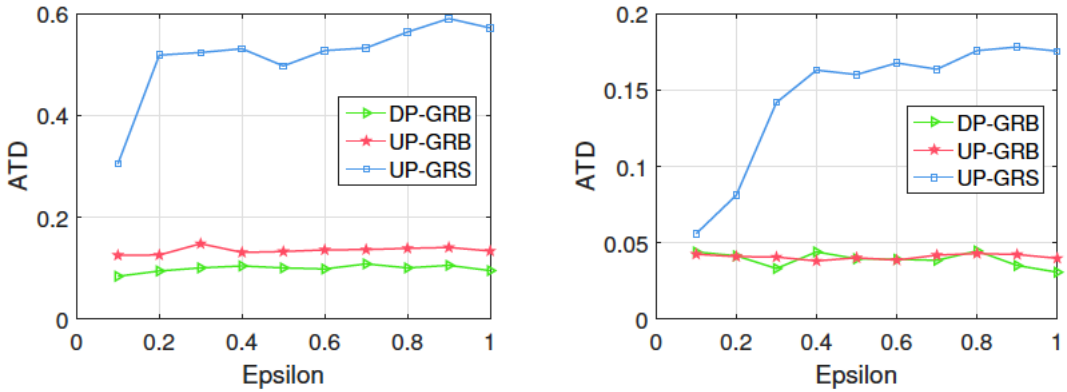
Simulation settings

- Parameters

Parameter	Description	Value	Default
ϵ	Privacy budget	0.1 – 1.0	1.0
MTD	Maximum travel distance	1km - 5km	5km
ESR	Expected success rate	0.3 - 0.9	0.8

Density-based location preservation

Simulation results

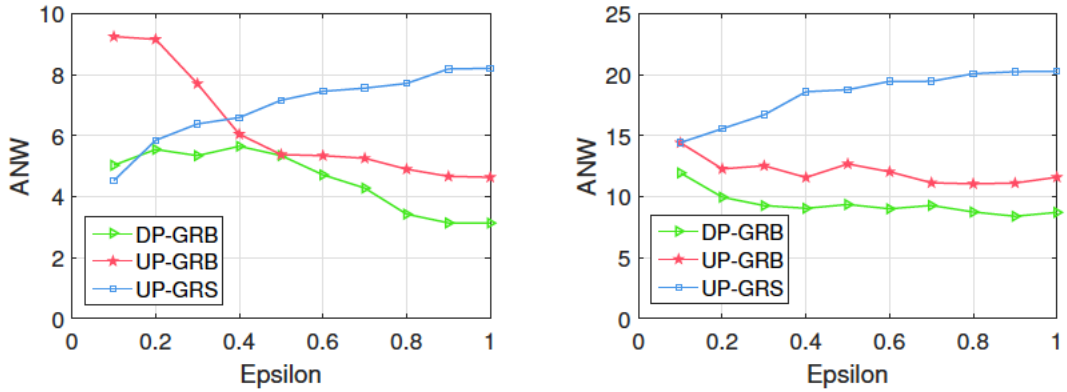


(e) Yelp-linear (f) SimpleGeo-linear

Figure. Performance by varying ϵ

Density-based location preservation

Simulation results

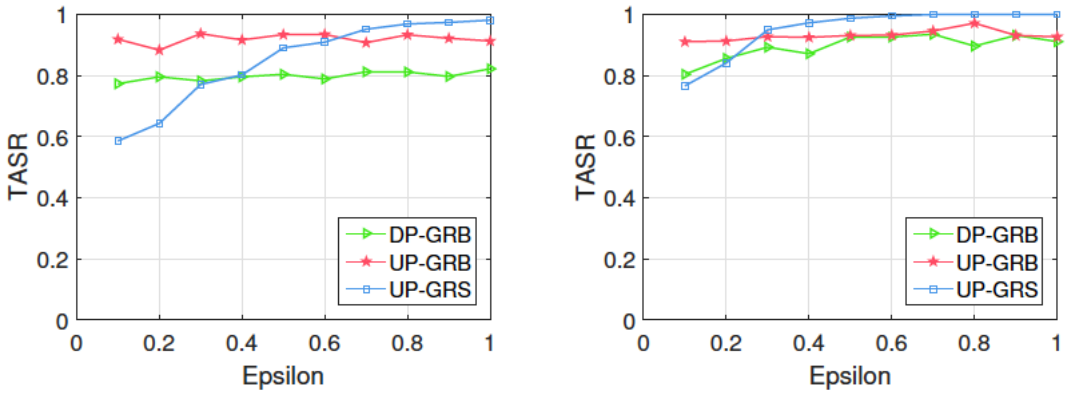


(a) Yelp-linear (b) SimpleGeo-linear

Figure. Performance by varying ϵ

Density-based location preservation

Simulation results



(i) Yelp-linear (j) SimpleGeo-linear

Figure. Performance by varying ϵ

Density-based location preservation

Simulation results

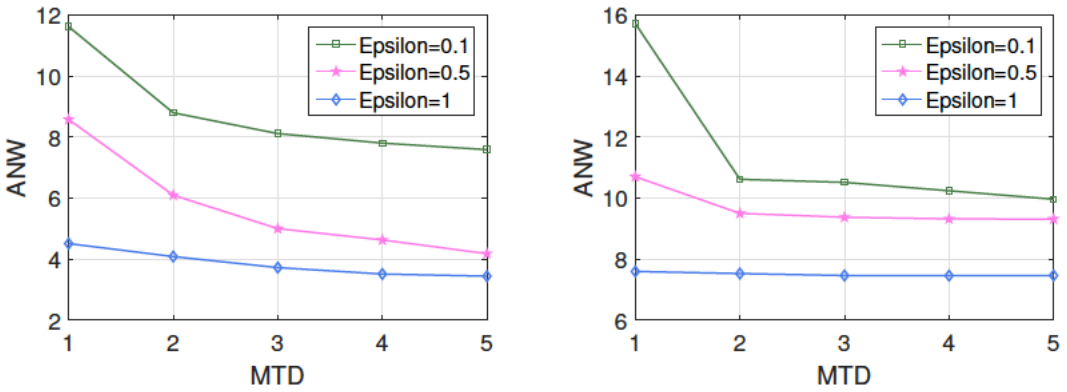
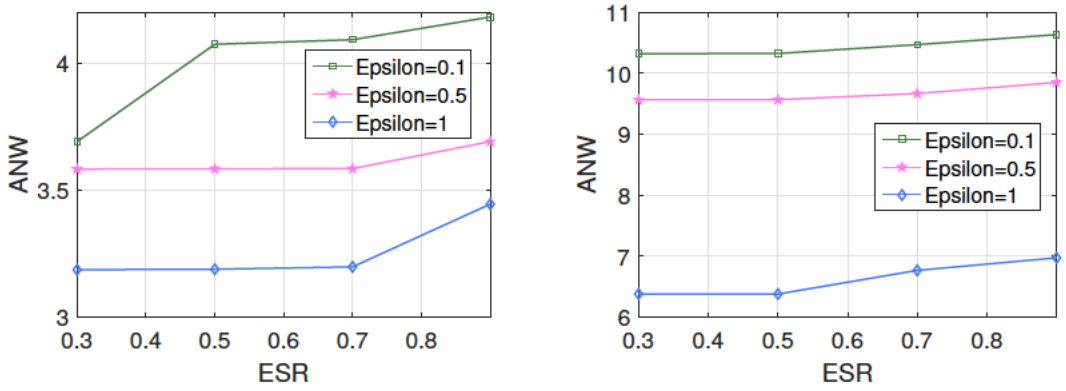


Figure. Performance by varying MTD

Density-based location preservation

Simulation results



(a) Yelp-linear (b) SimpleGeo-linear

Figure. Performance by varying ESR

Blockchain-based location privacy preserving

Possible solutions to tackle the privacy disclosed problem in the payment process

- Involve a trustworthy third party in the payment process.
  Challenges
  How to guarantee that the third party's payment process is precise and secure?
  How to instil the worker trust in the third party?

Use blockchains:

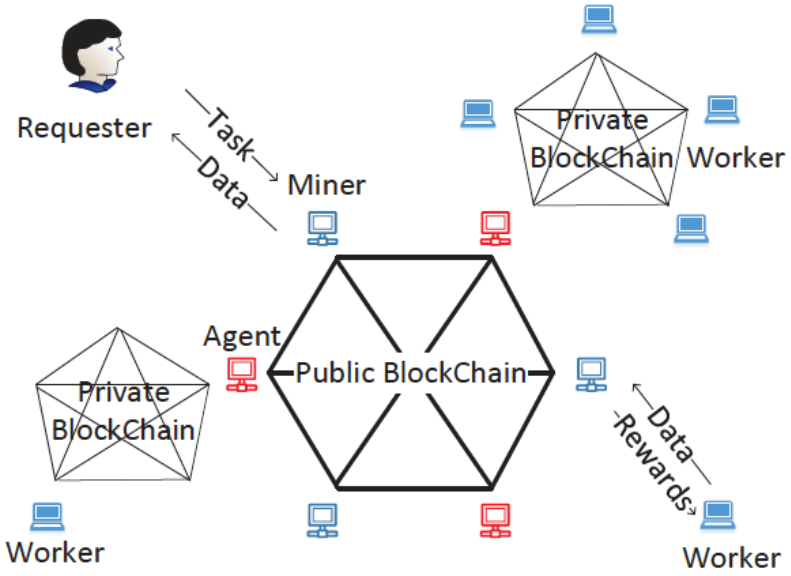
- Anonymous account information.
- Payment information is not associated with the worker's real identity.
- Character of immutability.

  Concern

  Character of transparency.

Blockchain-based location preservation

The proposed framework



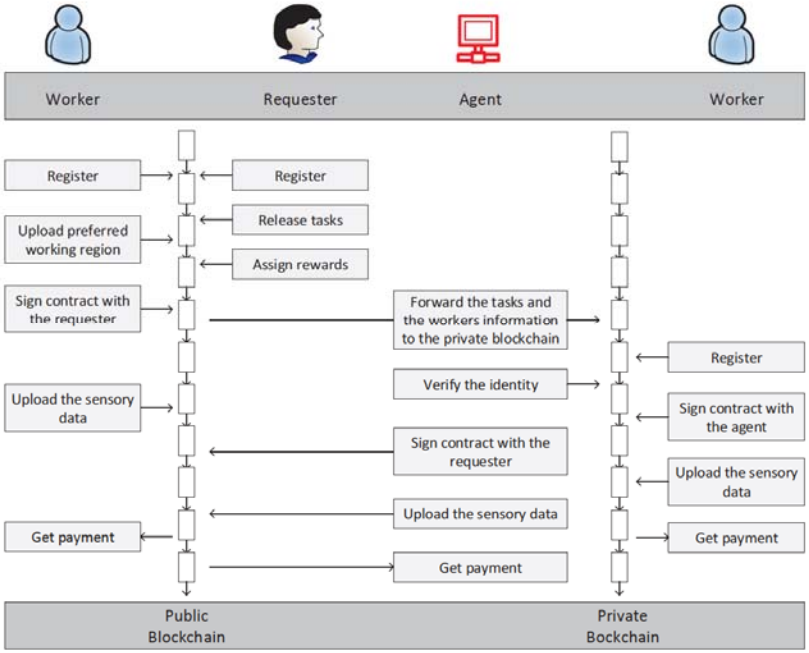
Blockchain-based location preservation

Adversary model

- **Assumptions.**
  All the participants on the blockchain network are untrusted except agents with private blockchans.
- **Attackers.**
  Participants in the public blockchain.
  Workers
  Agents

Blockchain-based location preservation

Overview of executive process



Blockchain-based location preservation

Overview of executive process

- Register (Public blockchain)
- Task assignment
- Smart contract creation in the public blockchain
- Transfer tasks to the private blockchain
- Register (Private blockchain)
- Smart contract creation in the private blockchain
- Upload sensory data
- Payment

Blockchain-based location preservation

- Privacy

  1. Cloaking region technology is used to upload worker's location information

  2. The workers are anonymized and the private blockchain is used to distribute the transaction records.

  3. Cryptocurrency is not authenticated with the worker's real identify.

- Security

  The consensus protocol of the blockchain ensures that the smart contracts are executed correctly. The combination of the smart contract the a deposit-based mechanism ensures fair trading.

Economic Model based Trajectory Privacy Preserving

Mobile Crowdsensing (MCS) Applications

1. Participants
2. Data Transmission Network
3. Service Provider (SP)

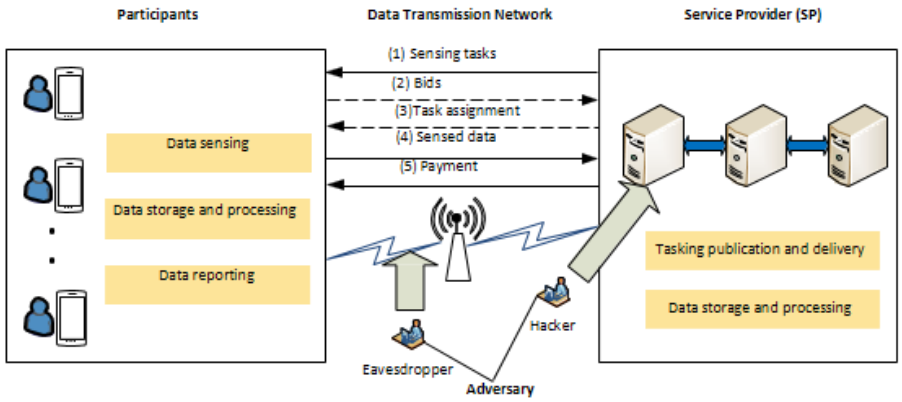


Figure: System architecture of mobile crowdsensing applications.

Proposed solution: enhancing the location privacy of mobile crowdsensing participants by eliminating the general bidding (step 2) and task assignment (step 3) processes

Problem formulation

Economic Modeling for MCS Application

An analogy between MCS applications and economic activities

- SP– consumer who buys useful data packets
- Participants – provider who can provide the needed outputs

Optimization Problems

$$\begin{cases} \min \sum_{l \in \mathscr{L}} m_l \\ \max \sum_{l \in \mathscr{L}} (m_{l,p} - \alpha \cdot TPL_{l,p}) \end{cases}$$

$$s.t.\ Q \geq Q_{th}$$

$$m_l = \sum_{p \in \mathscr{P}_l} m_{l,p}$$

Location privacy loss (LPL)

$$LPL_p = \sum_{r \in \mathscr{R}_p} [H_0(r) - H_T(r)] = \sum_{t \in \mathscr{T}, l \in \mathscr{L}} I(s^{p,d\langle l,t,TCl\rangle}, Y) \log_2 L$$

The Monopoly Model Based Scheme (MMBS)

Monopoly– all participants in the MCS task collude to maximize the overall profit.

- Participants' total output at a given location: q.
- Participant i's output at a given location: q_i.
- Constant marginal costs of Participants for each packet upload (caused by location privacy loss): $c = \alpha \cdot TPL_p$.
- Inverse demand function: $m(q) = A - Bq$.

Algorithm 1: Monopoly Model Based Scheme (MMBS).

Initialization: Define the targeted QoS Q_{th}, and the participants' cost c.
SP's Pricing Strategy:
for $l \in \mathcal{L}$ **do**
  Allocate the budget $m = \beta c q$, where $\beta > 1$ is the payment factor;
  Find the expected number of participants n at location l under the budget m;
  Calculate $q = n(1 - \sqrt[n]{1 - Q_{th}})$;
  Defining pricing strategy:

$$m(q) = A - Bq,$$

  where

$$A = \frac{2m}{q} - c, \; B = \frac{A - c}{2q}.$$

  Broadcast A, B and n to all participants.
end
Participant i's upload strategy:
for $l \in \mathcal{L}$ **do**
  Calculate $q_i = \frac{1}{n}\frac{A-c}{2B}$;
  Set the upload strategy which satisfies: $Pr\{s^{i,d(l,t,TC1)} = 1\} = q_i$.
end

Figure: Monopoly model based scheme.

Cournot's Oligopoly Model Based Scheme (COMBS)

Oligopoly– competitions among participants. All seek to maximize his own profit given other members' decisions.

- Participant i's output: $q_i = \mathbb{E}[s^{i,d\langle l,t,TCI\rangle}]$.
- Total output: $q = q_1 + q_2 + ... + q_n$.
- Opponents' output: $q_{-i} = q - q_i = \sum_{j \neq i} q_j$.
- Constant marginal cost of participant i: c_i.
- Inverse demand function: $m(q) = A - Bq$.

Algorithm 2: Cournot's Oligopoly Model Based Scheme (COMBS).

Initialization: Define the targeted QoS Q_{th}, and the participants' cost c.
SP's Pricing Strategy:
for $l \in \mathcal{L}$ **do**
  Allocate the budget $m = \beta cq$, where $\beta > 1$ is the payment factor;
  Find the expected number of participants n at location l under the budget m;
  Calculate $q = n(1 - \sqrt[n]{1 - Q_{th}})$;
  Defining pricing strategy:

$$m(q) = A - Bq,$$

  where

$$A = \frac{(n+1)m}{q} - nc, \; B = \frac{n}{n+1}\frac{A-c}{q}.$$

  Broadcast A, B and n to all participants.
end
Participant i's upload strategy:
for $l \in \mathcal{L}$ **do**
  Calculate $q_i = \frac{1}{n+1}\frac{A-c}{B}$;
  Set the upload strategy which satisfies: $Pr\{s^{i,d\langle l,t,TCI\rangle} = 1\} = q_i$.
end

Simulation Setups

The Geolife Data Set

- Area: 20km× 30km area is divided into 5×5 small cells.
- Total trajectory: 312.

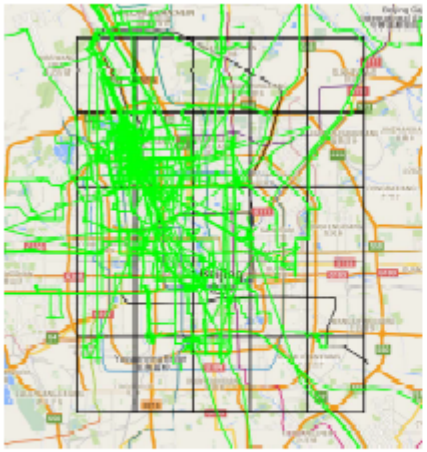


Figure: Map of the area for simulation (green lines indicate the trajectories).

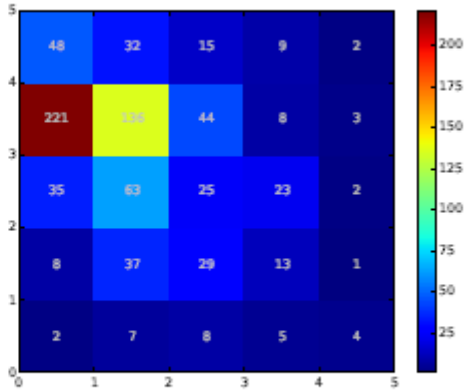


Figure: Number of available data packets in each cell.

Simulation Results

Figure: SP's budget with different QoS targets.

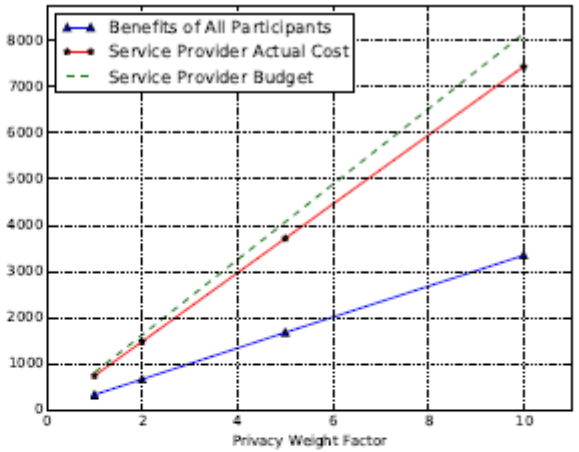


Figure: SP's budget with different privacy weight factor α.

1. Dramatic budget increase as the QoS grows from 0.99 to 0.999
2. linear relationship between the SP's budget and participants' privacy weight factor α

Simulation Results

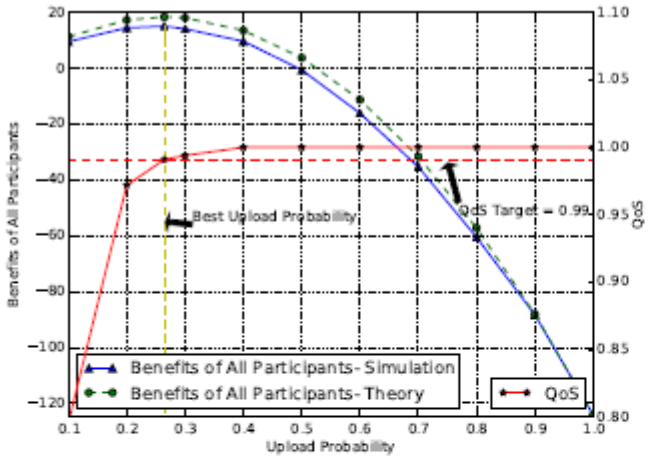


Figure: Participants' overall benefits with different upload probabilities in cell (4,2) using MMBS.

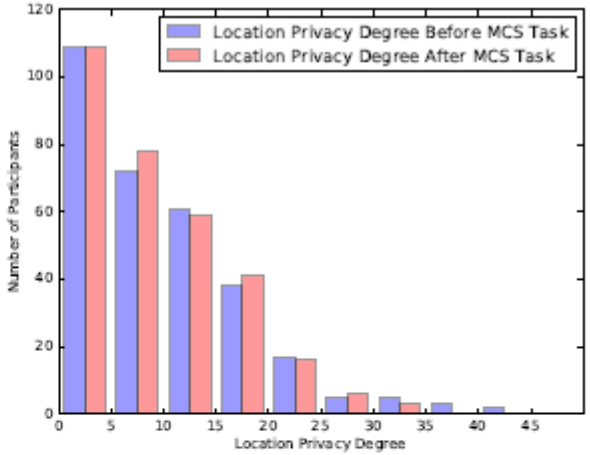


Figure: Histogram of participants' privacy degrees before and after the MCS task using MMBS.

1. The participants' overall benefit peaks at the best upload probability
2. The number of participants with high location privacy degrees (bigger than 20) declines after the task, which means that the uploaded data packets are prone to come from the long location information sets.

Simulation Results

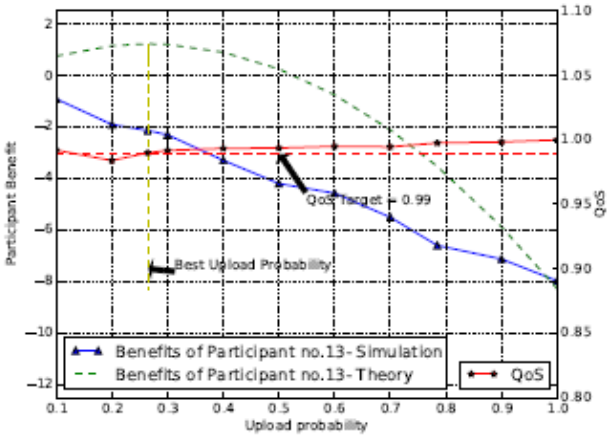


Figure: One participant's (no. 13) benefit with different upload probabilities in cell (4,2) using COMBS.

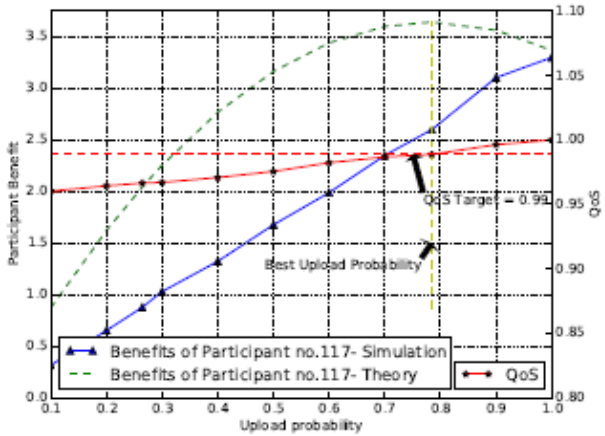


Figure: One participant's (no. 117) benefit with different upload probabilities in cell (3,4) using COMBS.

1. Theoretically, a certain participant's benefit should be maximized with the best Oligopoly upload probability.
2. The actual upload number for a certain participant is either 1 or 0, instead of the expectation value q_i.
3. When the theoretical best Oligopoly upload probability in a given cell is smaller than 0.5 (no. 13), the corresponding participant's benefit curve will be a descending one. The curve is ascending when the theoretical probability is bigger than 0.5 (no. 117).

Simulation Results

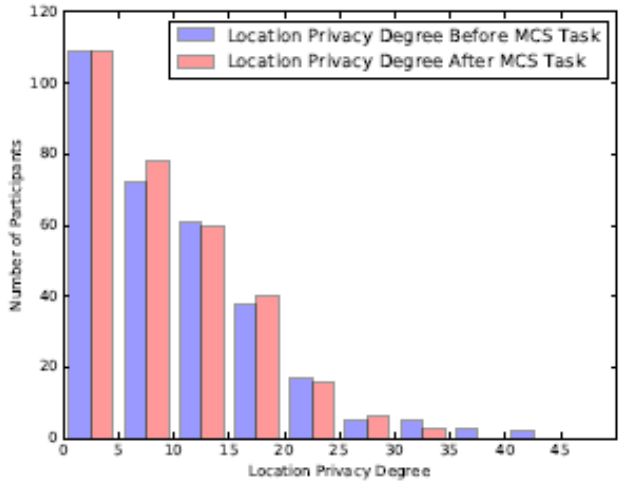


Figure: Histogram of participants' location privacy degrees before and after the MCS task using COMBS.

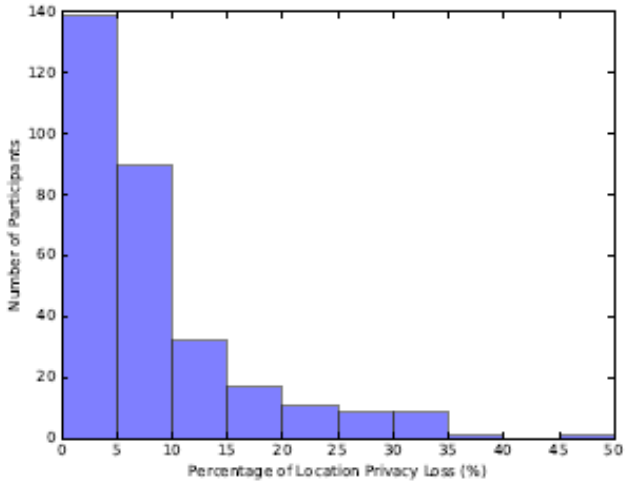


Figure: Histogram of participants' location privacy degree loss in percentage using COMBS.

1. From the perspective of location privacy degree loss, COMBS achieves similar privacy degree preservation with that of MMBS

Conclusions

- We investigate the location privacy problem from a different angle: reducing the information sharing. This new category of privacy preservation mechanism can be used with methods in other categories at the same time.
- We proposed a density-based location preservation mechanism for crowdsensing techniques that satisfies differential privacy, providing rigorous protection of worker locations. The partitioning method is based on worker density and considers non-uniform worker distribution.
- We propose a private Blockchain based method for task payment that effectively preserves individual privacy in the entire crowdsensing system.
- We propose a framework which enhances the location privacy of mobile crowdsensing participants by eliminating the general bidding and task assignment processes. And economic model is used to solve the optimization problem.

Selected publications

- Bo Liu, Wanlei Zhou, Tianqing Zhu, Yong Xiang, and Kun Wang***, Location Privacy in Mobile Applications***. Springer 2018, ISBN 978-981-13-1704-0.
- Youyang Qu, Shui Yu, Longxiang Gao, Wanlei Zhou, Sancheng Peng, A Hybrid Privacy Protection Scheme in Cyber-Physical Social Networks. ***IEEE Trans. Comput. Social Systems*** 5(3): 773-784 (2018)
- Mengmeng Yang, Tianqing Zhu, Yang Xiang, Wanlei Zhou, Density-Based Location Preservation for Mobile Crowdsensing With Differential Privacy. ***IEEE Access*** 6: 14779-14789 (2018)
- Bo Liu, Wanlei Zhou, Tianqing Zhu, Longxiang Gao, Yong Xiang, Location Privacy and Its Applications: A Systematic Study. ***IEEE Access*** 6: 17606-17624 (2018)
- Bo Liu, Wanlei Zhou, Tianqing Zhu, Haibo Zhou, Xiaodong Lin, Invisible Hand: A Privacy Preserving Mobile Crowd Sensing Framework Based on Economic Models. ***IEEE Trans. Vehicular Technology*** 66(5): 4410-4423 (2017)
- Tianqing Zhu, Gang Li, Wanlei Zhou, Philip S. Yu, Differentially Private Data Publishing and Analysis: A Survey. ***IEEE Trans. Knowl. Data Eng***. 29(8): 1619-1638 (2017)
- Bo Liu, Wanlei Zhou, Tianqing Zhu, Longxiang Gao, Tom H. Luan, Haibo Zhou, Silence is Golden: Enhancing Privacy of Location-Based Services by Content Broadcasting and Active Caching in Wireless Vehicular Networks. ***IEEE Trans. Vehicular Technology*** 65(12): 9942-9953 (2016)

Thank You!