Privacy friendly location based services

Sotiris Michaelides
Giorgos Baroutas
Tobias Eidelpes
Christina Kreza
Mark Juvan

Radboud University

Agenda

- **Introduction** — What? How? Why? The Privacy Project
- **Threat Models** — Understanding our attackers
- **Privacy Metrics** — Quantifying Anonymity
- **Implementations** — Privacy friendly LBS implementations
- **PrETP** — Privacy-preserving Electronic Toll Pricing system
- **Attacks** — Location privacy attacks

01

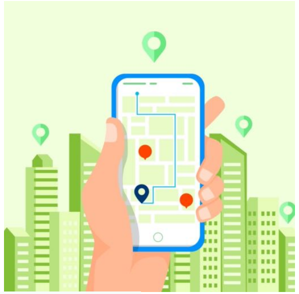
Introduction

Location Based Services

Specific services provided to users based on their location.

Then

Now



Location Based Services : HOW?

	GPS	Cellular networks	Wi-Fi positioning	Bluetooth low energy
Environment	Outdoor	Outdoor	Indoor & outdoor	Indoor mainly
Accuracy	10 - 100 m	0.75 miles	5 - 15 m	1 - 2 m
Battery consumption	High	Low	Medium	Low

Location Based Services : WHO & WHY?

WHO?

Big tech companies
Telecom companies
Surveillance operations
Location data companies

WHY?

Convenience
Safety
VS
$ PROFIT $

How law enforcement is using technology to track down people who attacked the US Capitol building

January 20, 2021 1.32pm GMT

Many of the people who broke into the U.S. Capitol building on Jan. 6 carried cellphones, which can be tracked, and posted photos of their activities on social media. Photo by Saul Loeb/AFP via Getty In

Email
Twitter 17
Facebook 500
LinkedIn
Print

After rioters flooded the U.S. Capitol building on Jan. 6, there was an immediate call for those who overran officers on the scene and swarmed the House and Senate floors, as well as congressional members' personal offices, to be identified, arrested and prosecuted. The coordinated law enforcement response to this incident is massive.

As researchers who study criminal justice, we see that law enforcement agencies are accessing large amounts of information via technological sources to investigate the attack on the U.S. Capitol building. High-definition security cameras, facial recognition technology, location services acquired from cellphones and third-party apps, and accessing archival evidence on social media are all used to identify perpetrators of crimes and tie them to specific places and

Gabon Fights Poaching with Elephant GPS Tracking collars

Technology is helping Gabon combat poachers with the latest GPS tracker collars. The country's National Parks Agency (APN) has started to fit elephants with the collars. In January of 2018, the team has already fitted 10 elephants in the Mwanga National Park and 8 elephants in the Ivindo National park. The initial project will fit 20 elephants with the tracking collars.

TRACKOMETER
Published on March 14, 2018

Bengaluru: BBMP to launch app to track waste collectors

Nithya Mandyam / TNN / Updated: Feb 19, 2022, 11:20 IST

Picture used for representational purpose only

BENGALURU: The solid waste management wing of Bruhat Bengaluru Mahanagara Palike is planning to launch a mobile application that tracks autorickshaws and other vehicles involved in garbage collection.

As of now, the civic body has already installed two monitoring systems in the vehicles — GPS and RFID. Now there will be another user interface for the existing GPS tracker. There are 5,500 auto- tippers and 600 compactors in the city.

Stalkers Use GPS to Track Victims

MILWAUKEE — Connie Adams found it impossible to escape her ex-boyfriend. He would follow her as she drove to work or ran errands. He would inexplicably pull up next to her at stoplights and once tried to run her off the highway, authorities said. When he showed up at a bar she was visiting for [...]

MILWAUKEE -- CONNIE Adams found it impossible to escape her ex-boyfriend.

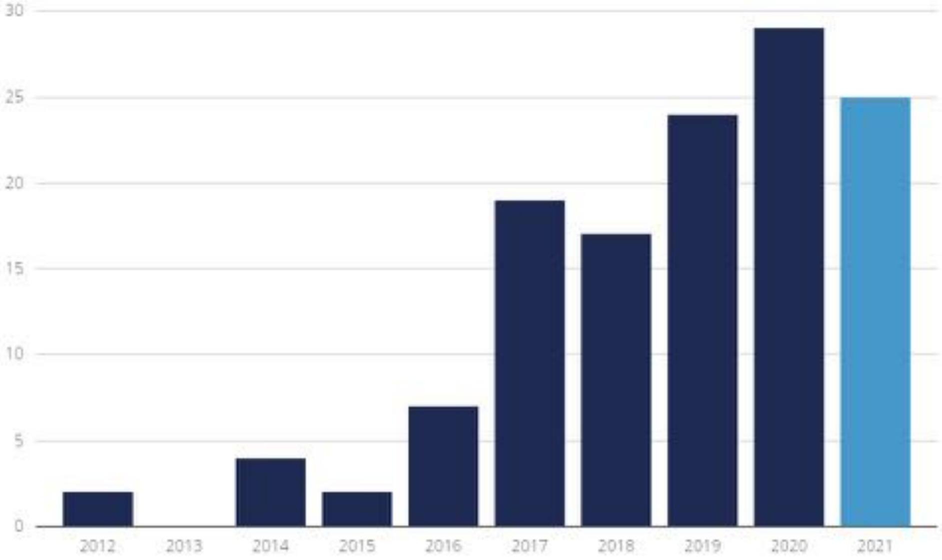
He would follow her as she drove to work or ran errands. He would inexplicably pull up next to her at stoplights and once tried to run her off the highway, authorities said.

When he showed up at a bar she was visiting for the first time, on a date, Adams began to suspect Paul Seidler wasn't operating on instinct alone.

GPS Trackers

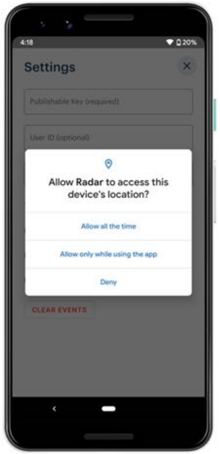
'More and more stalkers are using trackers'

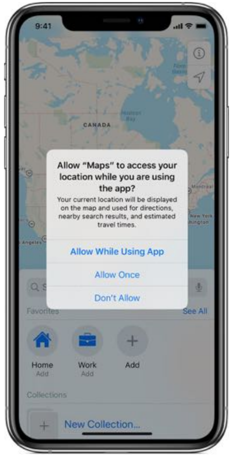
27 january 2022 10:30
Updated: 27 January 2022 12:40



Have you ever felt that you are tracked?

Allow OR Deny





- LBS are based on the implicit assumption that users consent to the disclosure of their private user locations.
- LBS trade their services with your location.
- Can the location of a user directly lead to its identity?

The New York Times : The Privacy Project

19 DEC 2019

- **Largest logging data**

  50B location pings
  12M mobile phones
  2016-2017

- **Anonymous Source**

  Concerned about the misuse of this information

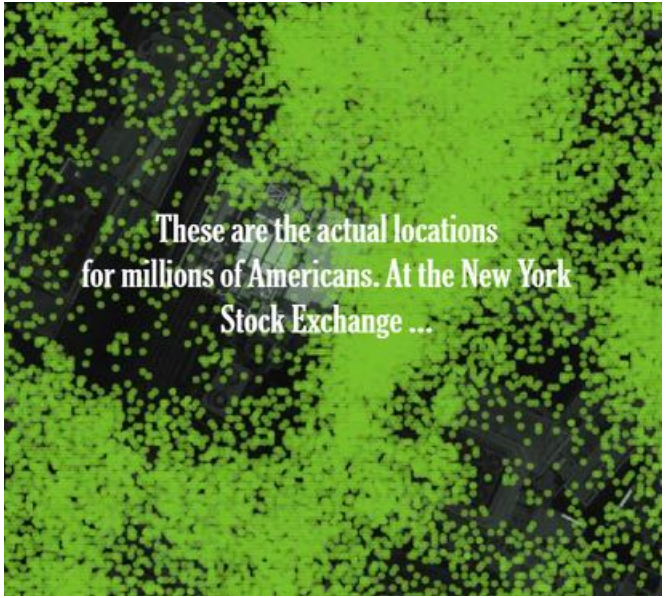
- **Location Data company**

  Software on mobile phone apps

- **Journalists tracked**
  - military officials
  - law enforcement officers
  - high-powered lawyers

- **Data is anonymous**

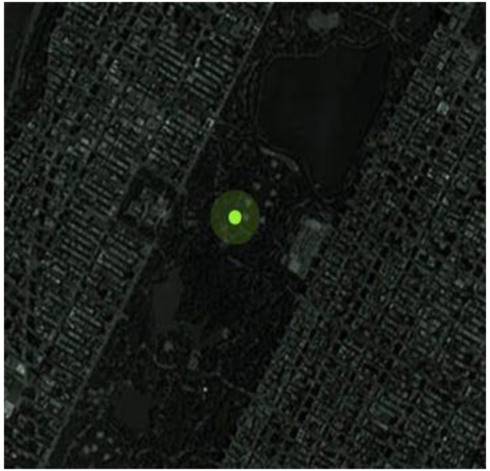
- **BUT HOW??**



The New York Times : The Privacy Project

BUT HOW ?

- Isolate one phone
- Gather all pings
- Connect the dots



WHAT ABOUT the connection with the physical owner of the phone??

The New York Times : The Privacy Project

ANONYMOUS DATA ?

"D.N.A. is probably the only thing that's harder to anonymize than precise geolocation information."

-Paul Ohm, law professor and privacy researcher at the Georgetown University Law Center

The New York Times : The Privacy Project
Laws & Regulations

- NO strict regulations / federal laws
  - “If a private company is legally collecting location data, they’re free to spread it or share it however they want,” Calli Schroeder, a lawyer for the privacy and data protection company VeraSafe.
  - Responsibility of company policies and individuals.

- Some steps on regulations
  - EU's General Data Regulation Protection (GDPR)
  - California Consumer Protection Act (CCPA)

Benefits to society

Transportation studies — The City Council of Portland approved a deal to study traffic and transit

Humanitarian purposes — Unicef study epidemics, natural disasters, and demographics using aggregated mobile location data provided by Cuebiq

Healthcare — Real Time LS provider Quuppa explores the potential impact of Adopting location tracking technologies in clinical settings

Market growth

Allied Market Research

Location-Based Services Market To Reach $318.64 Billion In 2030: Allied Market Research

Positive demand for location based services from several sectors such as agriculture, defense, transportation, energy, and transportation for navigation and traffic management propels the global location-based services market. The COVID-19 pandemic moderately impacted the growth of the location-based services market. Lockdown restrictions have increased the wide adoption of LBS software, enabling projects to continue in a virtual and digital world.

November 15, 2021 02:46 ET | Source: Allied Market Research

SURVEY : User Opinion on Location Privacy

Discount coupons on products from shops that you pass by regularly — **Service A**

Updates on the congestion of the roads — **Service B**

Tracking of your movement as an employee by your company — **Service C**

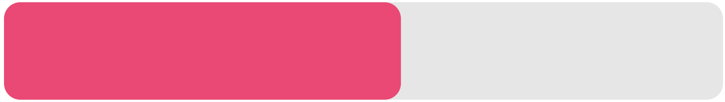
Alerts when predefined members of your family or friends are nearby — **Service D**

What people really want?

Useful location based services

WITHOUT

Revealing their private location information



Service - Privacy
Trade off

02

Threat models

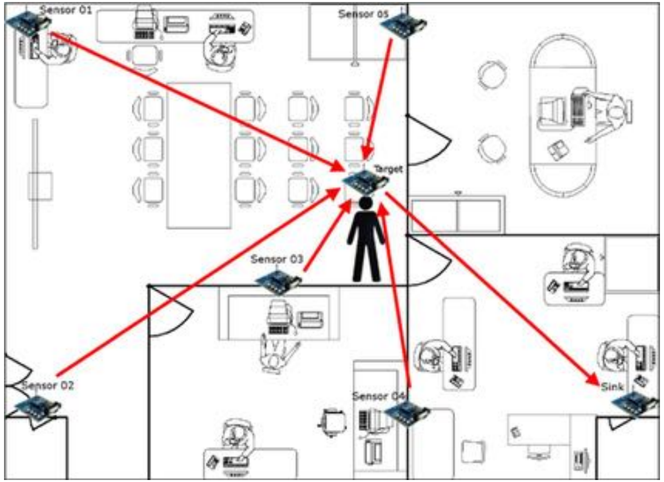
Categories of threats

- When LBSs/localization systems are being used, they collect location/user's tracking data.
- If these data are not properly processed and protected, our privacy is exposed if an attacker gain access to them
  - **Location privacy threats**
- Some localization systems depend on the exchange of data between nodes. The attacker can intercept
  - **Communication privacy threats**

Communication Privacy Threats

Some Localisation Systems are just as a network of devices/sensors that are used to track people/objects when satellite positioning is not available/lack of precision

- Altering Sensor Data attacks, including the positioning algorithms
- Spoofing Attacks, aka modification attacks, by injecting fake data to the nodes
- **Sinkhole Attack**, similar to BGP leak attacks where a node advertises fake routing paths



Location Privacy Threats

➔ Threats that occur because the attacker is able to get information about someone's location.

➔ **Profiling Threat** : Attacker is able to create a profile about a user without identifying him.

➔ **Tracking Threat** : Attacker receives continuous updates about user's location in real time.

➔ **Identification Threat** : Attacker is able to match location data to a certain person.

➔ ...others

Who collects your data?

- Biggest location based services providers:

  - **Foursquare** : provides location-based to some of the biggest retail in the world including Apple, Microsoft and Samsung, Uber, Coca Cola, Uber etc.. Collecting data: City guide application

  - **Here Technologies** : Here offers map/navigation services to multiple OSs, and collects data through those services. It is behind many other third party LBSs. Clients of Here: Amazon, Facebook, Yahoo, Oracle and etc...

  - **The usual suspects :** Google , Apple , Facebook. . .

  - . . . . And many other Companies as this market is booming.

Who wants to profile you?

Profiling threat: Who Would Like to create a profile for you but he is not interested to identify you?

- Advertising Companies
- Service providers (Netflix,Spotify ...)
- Companies that need data to train models
- **The main motive** for profiling is usually **money** and the Attackers are very powerful.

Who wants to track/identify you?

Tracking/Identification Threat: Who would like to identify you or track you?

- Authorities (Police , NSA etc)
- Secret services
- Foreign Governments
- Maybe even your Ex ?
- An assassin?
- **The power, the skills and motive** of the attacker in this case may vary. (From an individual stalker who just like to stalk people to governments with 'unlimited' resources who want to spy on their citizens)

03

Privacy Metrics

Privacy Metrics

- ➔ Query privacy
  - ◆ User identification
- ➔ Location privacy
  - ◆ Accurately locating users

Query Privacy vs. Location Privacy

- ➔ Area of *k* users
  - ◆ Identification impossible
- ➔ Restricted area
  - ◆ Location privacy compromised

Query Privacy vs. Location Privacy

- ➔ Single user
  - ◆ Identification possible
- ➔ Large area
  - ◆ Location privacy protected

k-Anonymity

"a release provides k-anonymity protection if the information for each person contained in the release cannot be distinguished from at least k-1 individuals whose information also appears in the release."

– Kang G. Shin et al.: Privacy Protection for Users of Location-Based Services

k-Anonymity

- ➔ Concept from database research
- ➔ Cloaking region with *k* users
- ➔ Indistinguishability from *k-1* users

k-Anonymity

- ➔ Multiple queries allow inference
- ➔ Render data sufficiently anonymous
- ➔ Retain usefulness

k-Anonymity Example

	Race	Birth	Gender	ZIP	Problem
t1	Black	1965	m	0214*	short breath
t2	Black	1965	m	0214*	chest pain
t3	Black	1965	f	0213*	hypertension
t4	Black	1965	f	0213*	hypertension
t5	Black	1964	f	0213*	obesity
t6	Black	1964	f	0213*	chest pain
t7	White	1964	m	0213*	chest pain
t8	White	1964	m	0213*	obesity
t9	White	1964	m	0213*	short breath
t10	White	1967	m	0213*	chest pain
t11	White	1967	m	0213*	chest pain

k-Anonymity for LBSs

➔ Trusted central location server

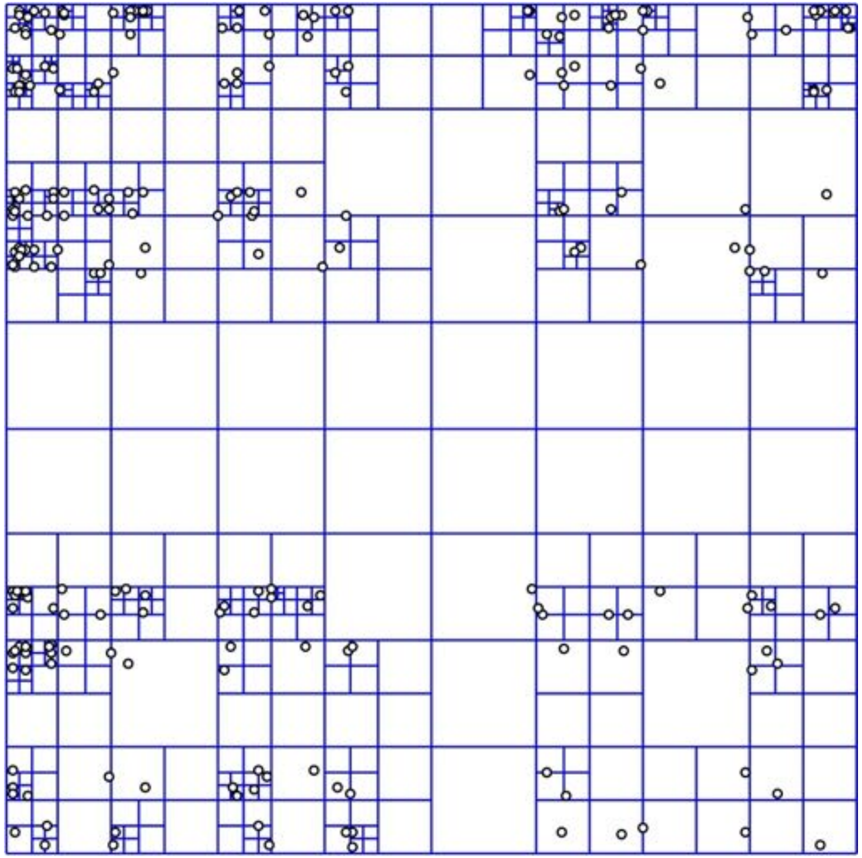
➔ Spatial cloaking

➔ Temporal cloaking

Spatial Cloaking

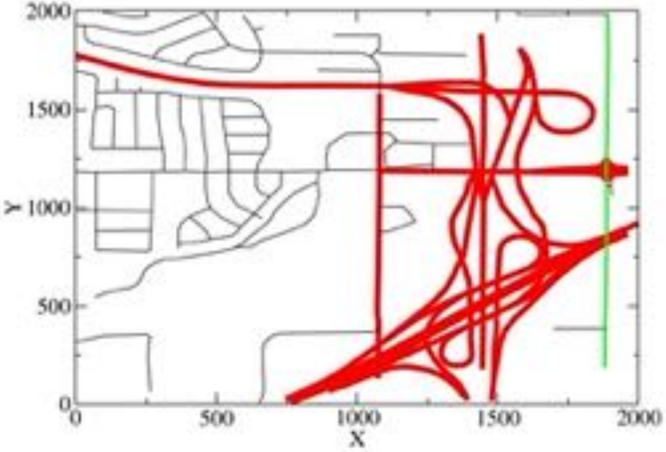
- ➔ Location information represented as tuple
  - ◆ $([x_1, x_2], [y_1, y_2], [t_1, t_2])$
- ➔ Increase anonymity by decreasing spatial accuracy
- ➔ Divide area until threshold *k*□¡□

Quadtree Algorithm



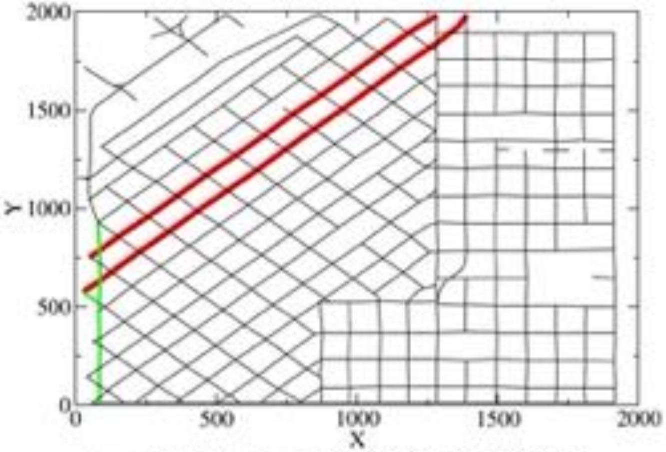
Temporal Cloaking

- ➔ Increase anonymity
- ➔ Delays increase location accuracy
- ➔ Result after threshold *k*□¡□

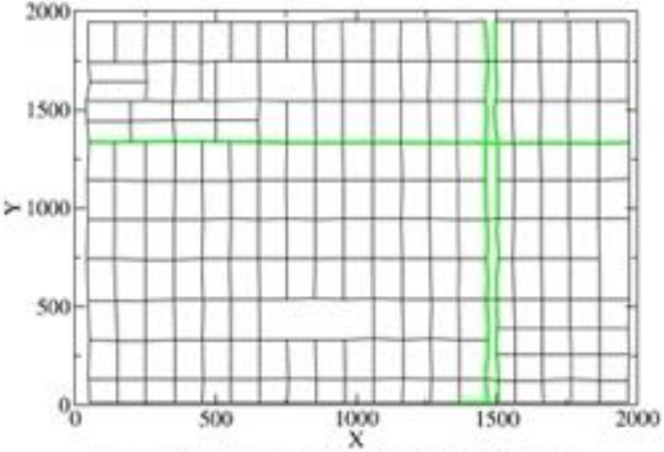


Area 1: Expressways (X:500000, Y:4407000)

Area 2: Expressways (X:500000, Y:4402000)



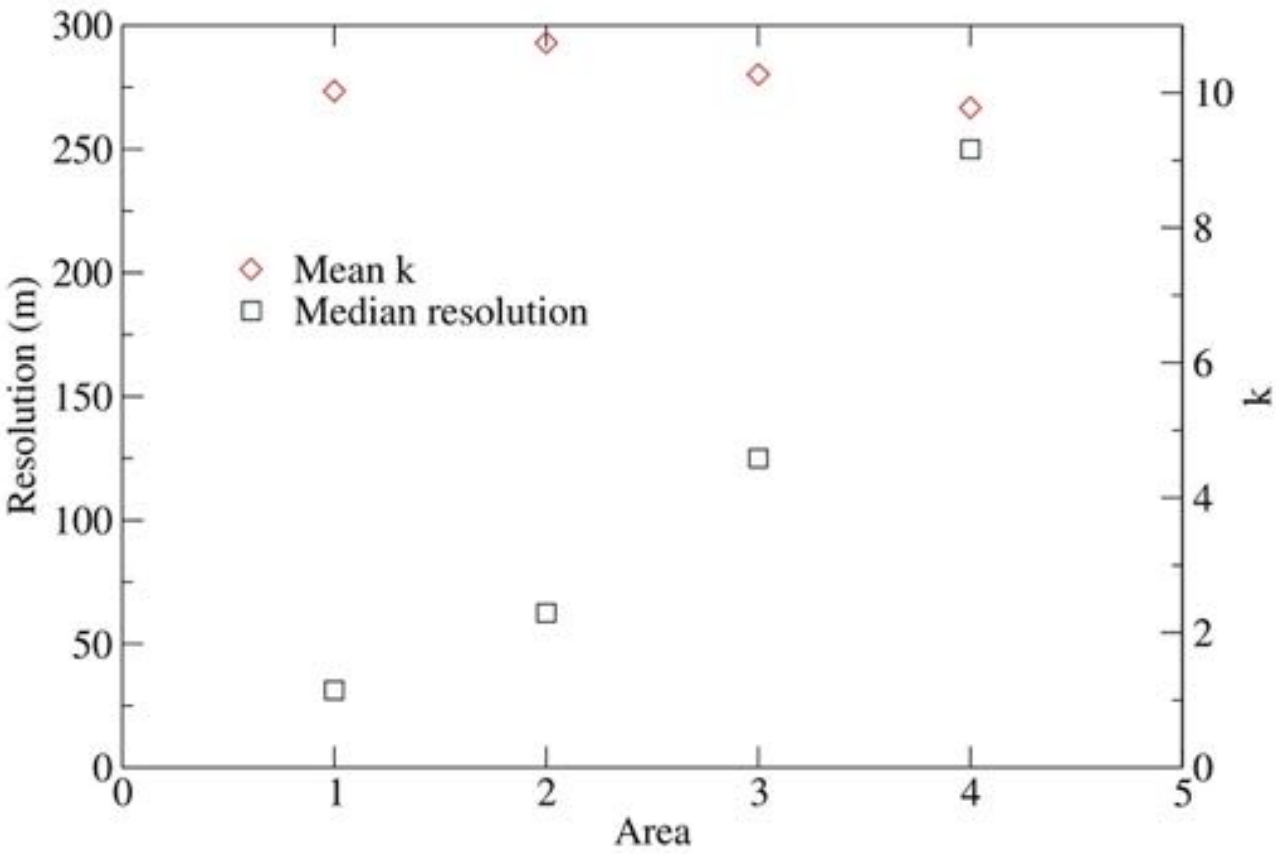
Area 3: Collectors (X:501000,Y:4400000)



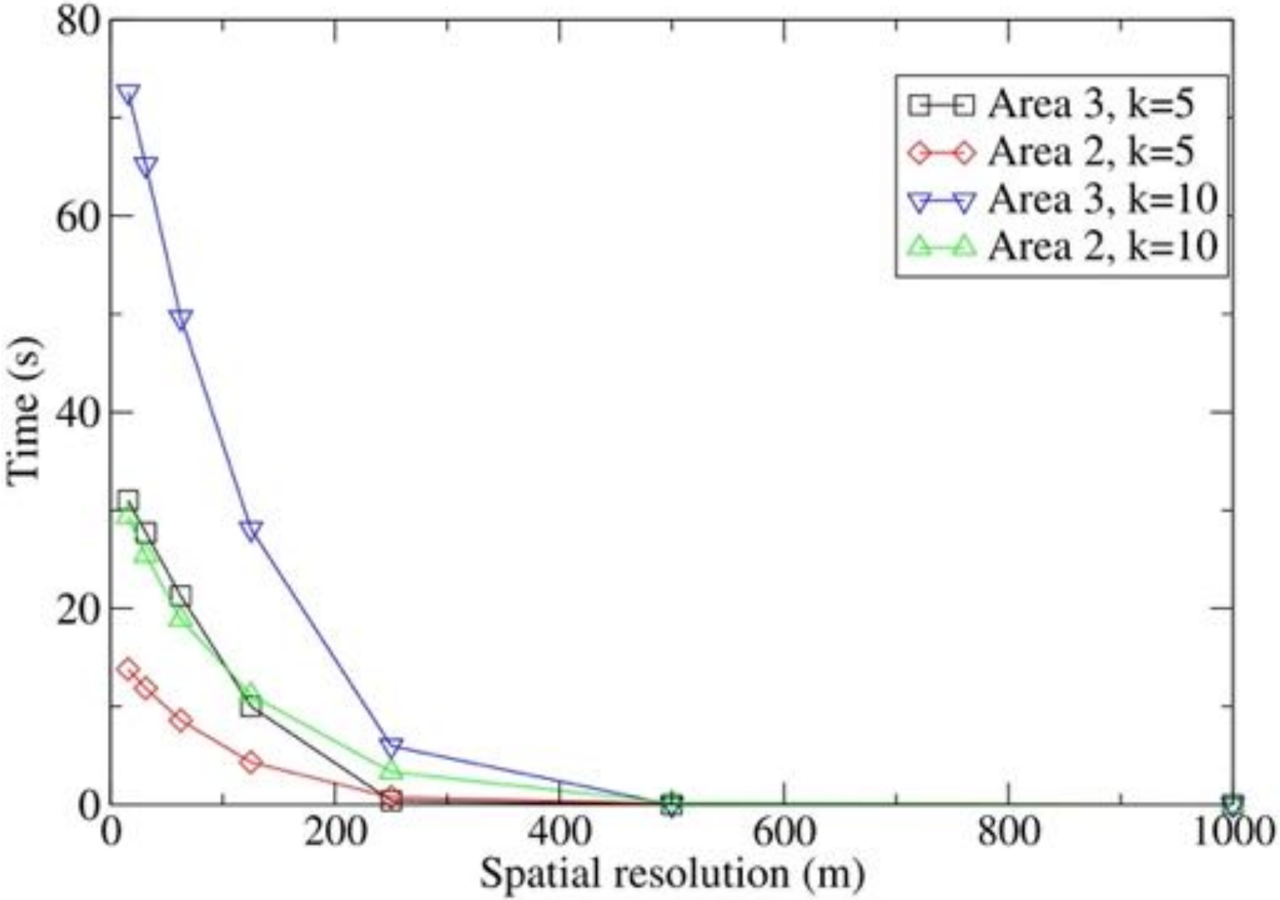
Area 4: Collectors (X:506000,Y:4400000)

USGS Class	Road Type	Traffic Volume
1	Expressway	70000
2	Arterial	22000
3	Collector	6000

Traffic count statistics per road type in Denver, Colorado



Mean resolution and *k*-anonymity per area



Temporal resolution vs. spatial resolution for different *k*

04

Algorithms and implementations

Overview

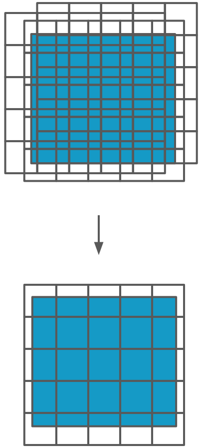
- General k-anonymity implementation:
  - Uniform grid caching
- k-anonymity + trajectory protection:
  - DKM
  - DPP
- Dummy generation:
  - PPCS
- Hiding in the Mobile Crowd: Location Privacy through Collaboration

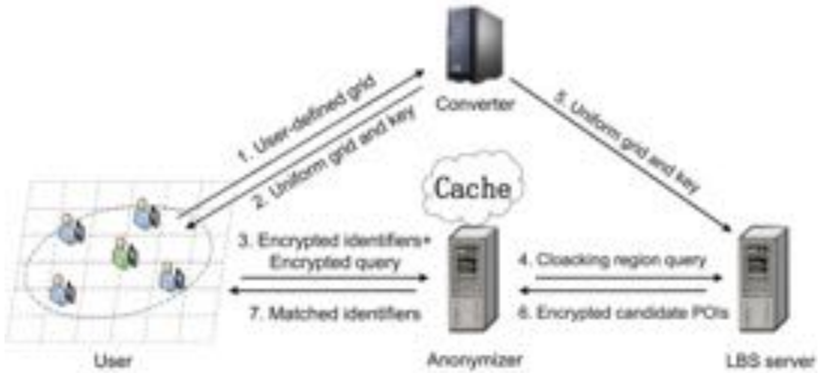
Uniform grid caching

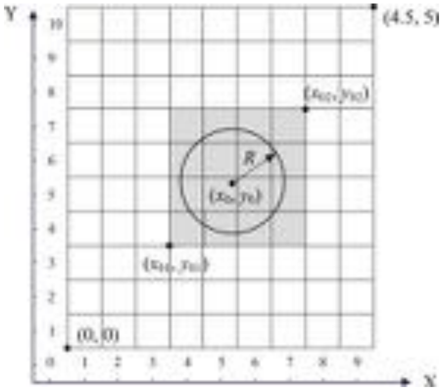
➔ Convert user defined to uniform grid

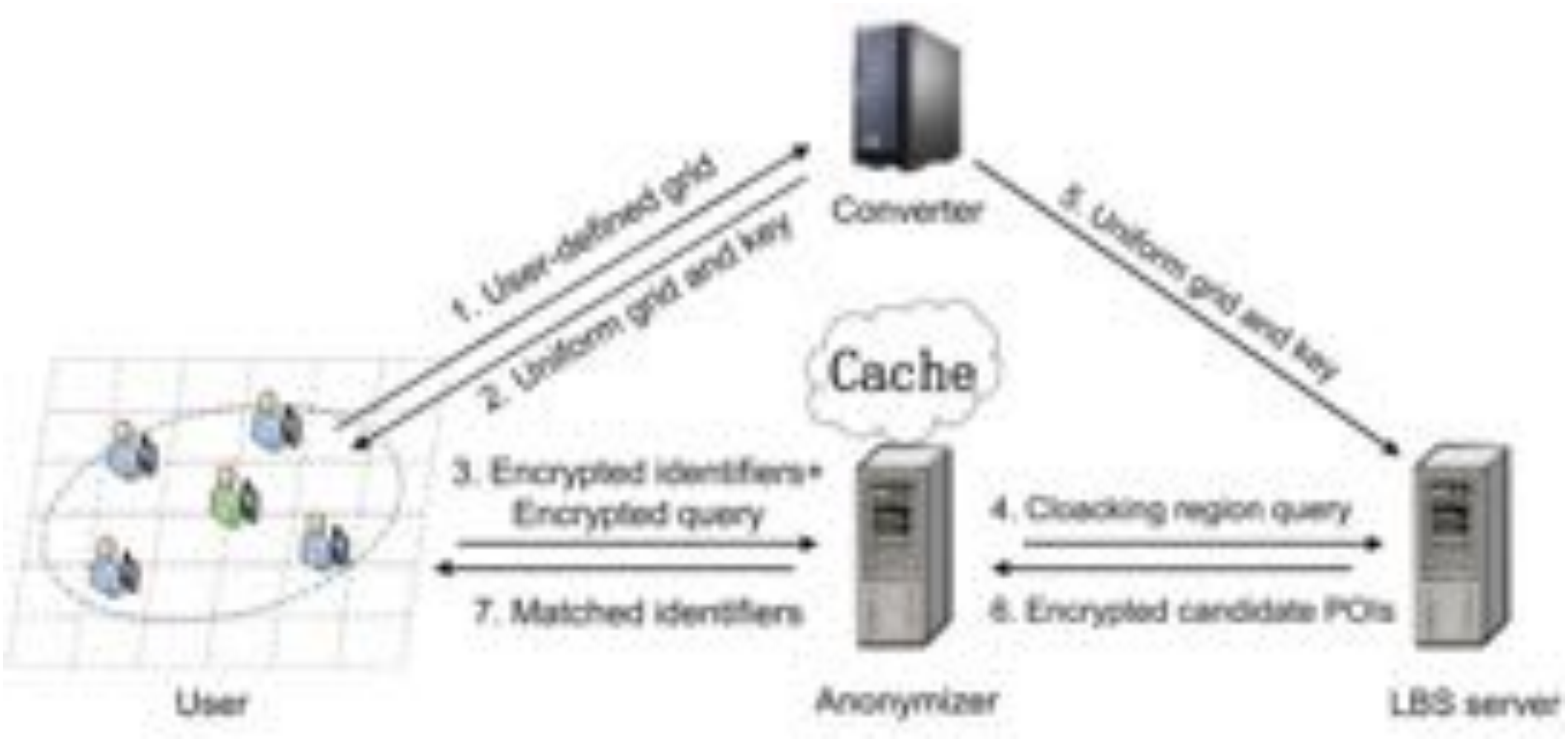
➔ OPSE (order preserving symmetric encryption)

➔ Cache of queries on semi-TTP (i.e. honest-but-curious): anonymizer

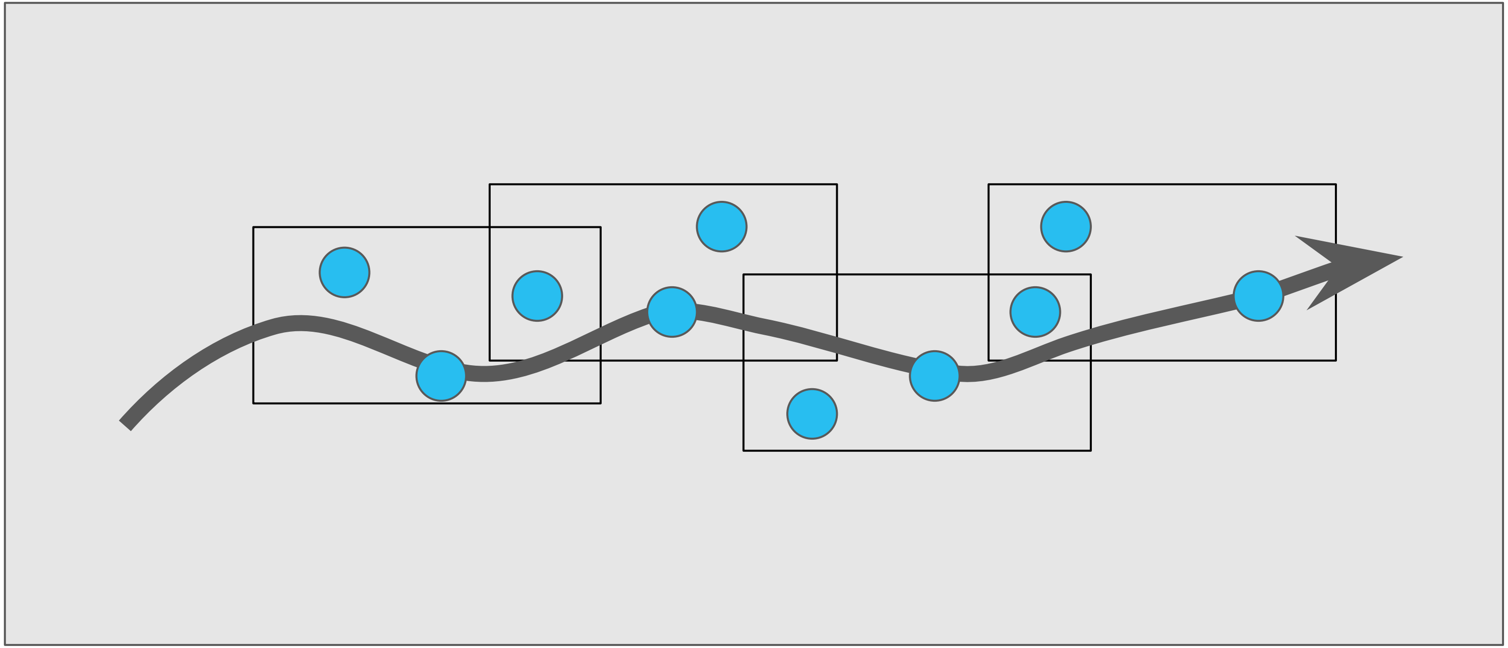






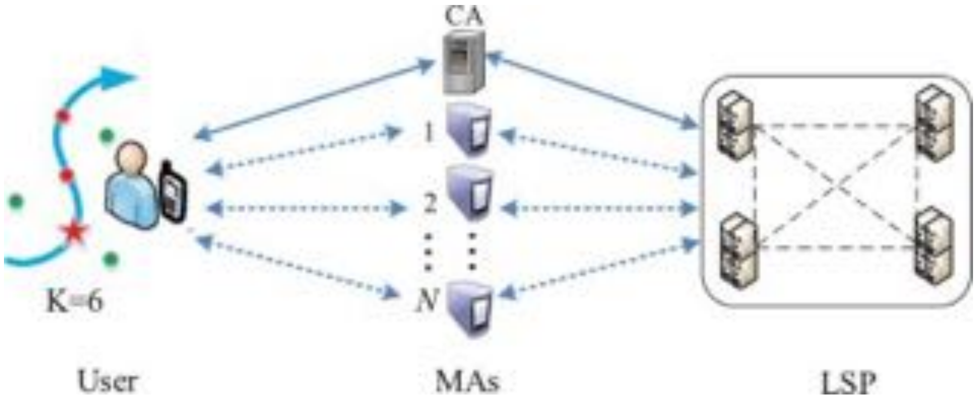

Converter
1. User-defined grid
2. Uniform grid and key
5. Uniform grid and key
Cache
3. Encrypted identifiers+
Encrypted query
7. Matched identifiers
4. Cloacking region query
6. Encrypted candidate POIs
User
Anonymizer
LBS server

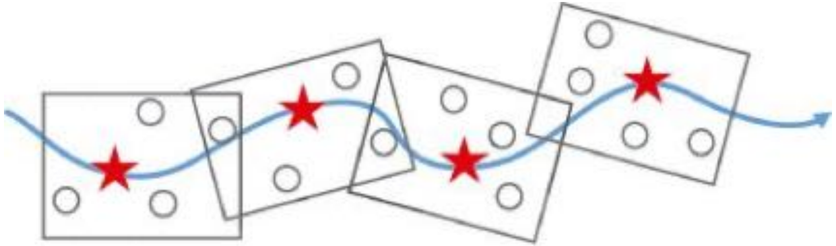
DKM scheme



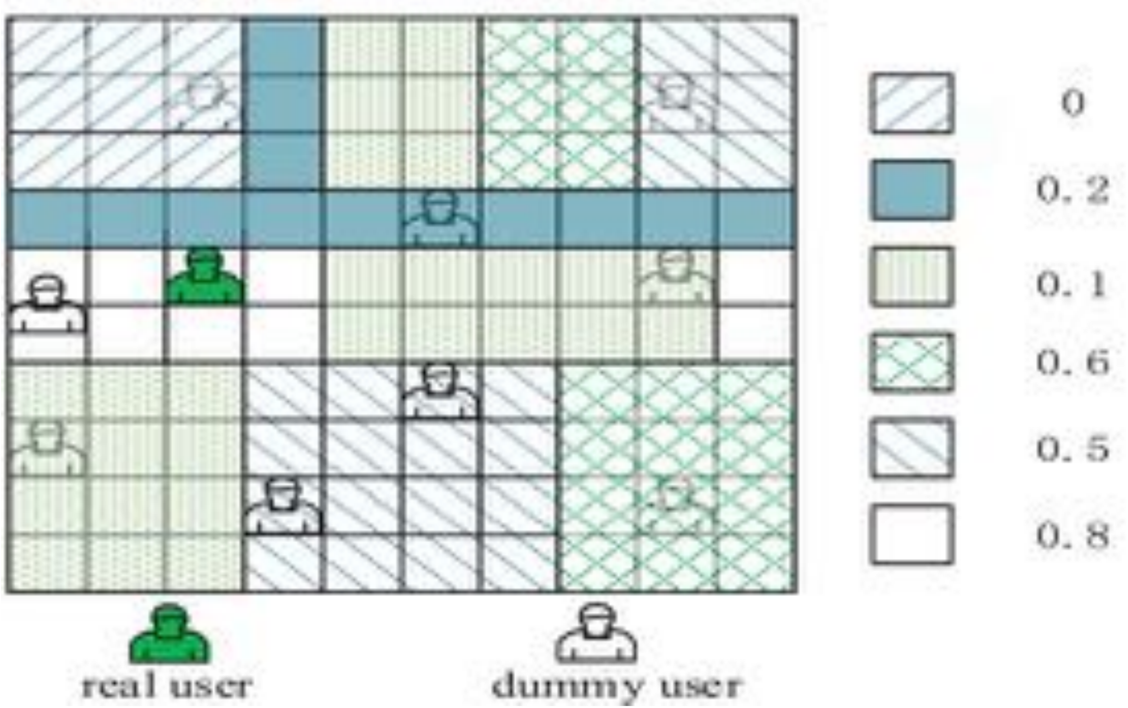
DKM (Dual-K mechanism) and DPP (Dual Privacy Preserving) scheme

- ➔ k-anonymity is a problem with continuous LBSs
- ➔ Protect user trajectory and query privacy





PPCS



PPCS (Preserve the location Privacy of mobile device users in location-based Cyber Services)

- ➔ Select l location types, for each generate (k-1) candidates
- ➔ Select k-1 candidates until l/2 location types are in selection
- ➔ Take dummies with the probability most similar
- ➔ If the actual location is not in selection, we extend the ROI radius and select new locations

Hiding in the Mobile Crowd

- ➔ Users in same area mostly require query results from the same area
- ➔ If each user stores a buffer of information, it is possible for other users to get it without contacting LBS
- ➔ MobiCrowd evaluation model

05

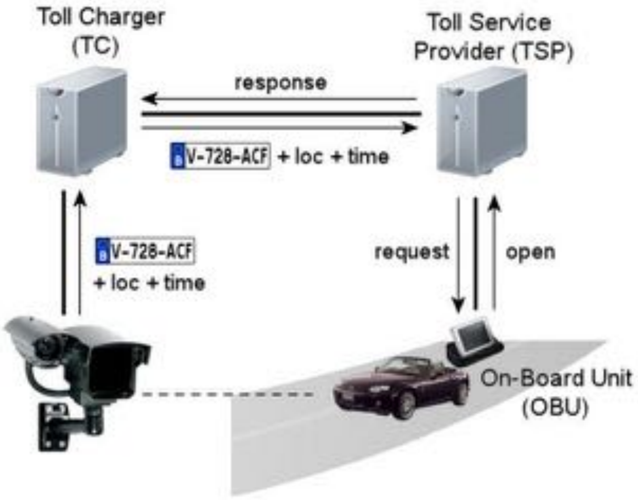
PrETP: Privacy-Preserving Electronic Toll Pricing

Electronic Toll Pricing System

- ➔ Transponders or on-board units used
- ➔ On-board unit receives radio signal from toll reader device and sends back an identifying number
- ➔ In other implementations, on-board units compute the corresponding fee at the end of each tax period and send it to the service provider alongside the actual location data
- ➔ A way to apply congestion pricing

PrETP System Model

- ➔ On-Board Unit (**OBU**): electronic device installed in vehicles subscribed to an ETP service
- ➔ Toll Service Provider (**TSP**): offers the ETP service, provides vehicles with OBUs and monitors their performance and integrity
- ➔ Toll Charger (**TC**): the organization that levies tolls for the use of roads and defines the correct use of the system



Basic idea behind PrETP

- ➔ Compute the fee locally
- ➔ Send along with the final fee, commitments to the locations and prices used in the fee computation
- ➔ Commitments ensure that drivers cannot claim they were at any other position
- ➔ Commitments ensure that drivers cannot claim that different prices were used during the calculation of the fee

Security Goals

False final fees

Drivers should not be able to report an arbitrary fee, but only the result from the correct calculations

Incorrect road prices

Drivers should not be able to assign arbitrary prices to the roads which they are driving

False GPS location data

Drivers should not be able to spoof the GPS signal and simulate a cheaper route than the actual roads

Inactive OBUs

Drivers should not be able to shut down their OBUs at will to simulate they drove less

Technical Requirements

Signatures

SigKeygen, **SigSign**(*sk*, *x*) and **SignVerify**(*pk*, *x*, s_x)

The signature scheme must be correct and unforgeable

Commitments

ComSetup, **Commit**(*params*, *x*) and **Open**(*params*, c_x, *x*, $open_x$)

The commitment scheme has a hiding and a binding property

Additively homomorphic property:
if $c = c_{x1} * c_{x2}$, then Open(*params*, *c*, *x1* + *x2*, $open_{x1} + open_{x2}$)

Proofs

The prover proves to the verifier **knowledge** of some **secret** values that fulfill some statement without disclosing the secret values to the verifier

Optimistic Payment Protocol

During each tax period:

- ➔ OBU slices the trajectories of the driver in segments - tuple *(loc, time)*

- ➔ TSP establishes $f : (loc, time) \rightarrow Y$ that maps every tuple to a price $p \in Y$

- ➔ OBU computes a payment tuple that consists of :
  - ◆ randomized hash h on the data
  - ◆ homomorphic commitment c_p
  - ◆ proof π that the committed price belongs to Y

Optimistic Payment Protocol

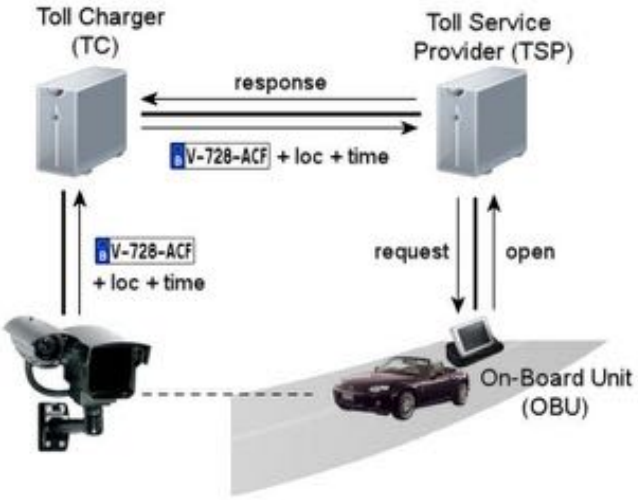
At the end of the tax period:

- ➔ Two-party protocol between OBU and TSP
- ➔ OBU adds all the segments to obtain a total fee *fee*
- ➔ OBU adds all the openings to obtain an opening $open_{fee}$
- ➔ Payment message *m* consists of $(tag, fee, open_{fee})$ and all the payment tuples (h, c_p, π)
- ➔ OBU signs *m* and sends both m and s_m to TSP
- ➔ TSP add the commitments c_p to obtain c'_{fee} and checks that $(fee, open_{fee})$ is a valid opening for c'_{fee}

Optimistic Payment Protocol

Given a proof φ from TC:

- ➔ TSP relays φ to OBU
- ➔ OBU verifies that the request is signed by the TC
- ➔ OBU searches for a tuple for which $\mu(\varphi, (loc, time))$ outputs *accept*
- ➔ OBU sends the number of the tuple together with the preimage *(loc, time)* of *h* and the opening $(p, open_p)$
- ➔ TSP checks if:
  - ◆ $(p, open_p)$ is a valid opening for c_p
  - ◆ *(loc, time)* is the preimage of *h*
  - ◆ $\mu(\varphi, (loc, time))$ outputs accept



Security Goals

False final fees

The homomorphic property ensures that the final fee is the sum of all the committed subfees

Incorrect road prices

TSP can check whether the correct price for a segment was used once the commitments are opened

False GPS location data

Risk of not having committed to a segment containing the *(loc, time)* in the challenge φ

Inactive OBUs

Risk of not having committed to a segment containing the *(loc, time)* in the challenge φ

Optimistic Payment Protocol

OBU
Pay() algorithm

// **Main loop**
For all $1 \leq k \leq N$ *tuples do:*
$p_k = f(loc_k, time_k)$
// **Hash computation**
$h_k = H((loc_k, time_k))$
// **Commitment computation**
$open_{p_k} \leftarrow \{0,1\}^{l_n}$
$c_{p_k} = g_0^{p_k} g_1^{open_{p_k}} \pmod{n}$
// **Proof computation**
$open_w, w \leftarrow \{0,1\}^{l_n}$
$\hat{A} = A g_0^{w} \pmod{n}$
$c_w = g_0^{w} g_1^{open_w} \pmod{n}$
$r_\alpha \leftarrow \{0,1\}^{l_\alpha}$
$t_{c_{p_k}} = g_0^{r_{p_k}} g_1^{r_{open_{p_k}}}$
$t_Z = \hat{A}^{r_e} R^{r_{p_k}} S^{r_v} (g_0^{-1})^{r_{w \cdot e}}$
$t_{c_w} = g_0^{r_w} g_1^{r_{open_w}}$
$t = c_w^{r_e} (g_0^{-1})^{r_{w \cdot e}} (g_1^{-1})^{r_{open_w \cdot e}}$
$ch = H(\beta \| t_{c_{p_k}} \| t_Z \| t_{c_w} \| t)$
$s_\alpha = r_\alpha - ch \cdot \alpha$
$\pi_k = (\hat{A}, c_w, ch, s_\alpha)$
End for
// **Fee reporting**
$fee = \sum_{k=1}^{N} p_k$
$open_{fee} = \sum_{k=1}^{N} open_{p_k}$
$m = [tag, fee, open_{fee}, (h_k, c_{p_k}, \pi_k)_{k=1}^{N}]$
$s_m = \text{OBUsign}(sk_{OBU}, m)$

$\xrightarrow{(m, s_m)}$

TSP
VerifyPayment() algorithm

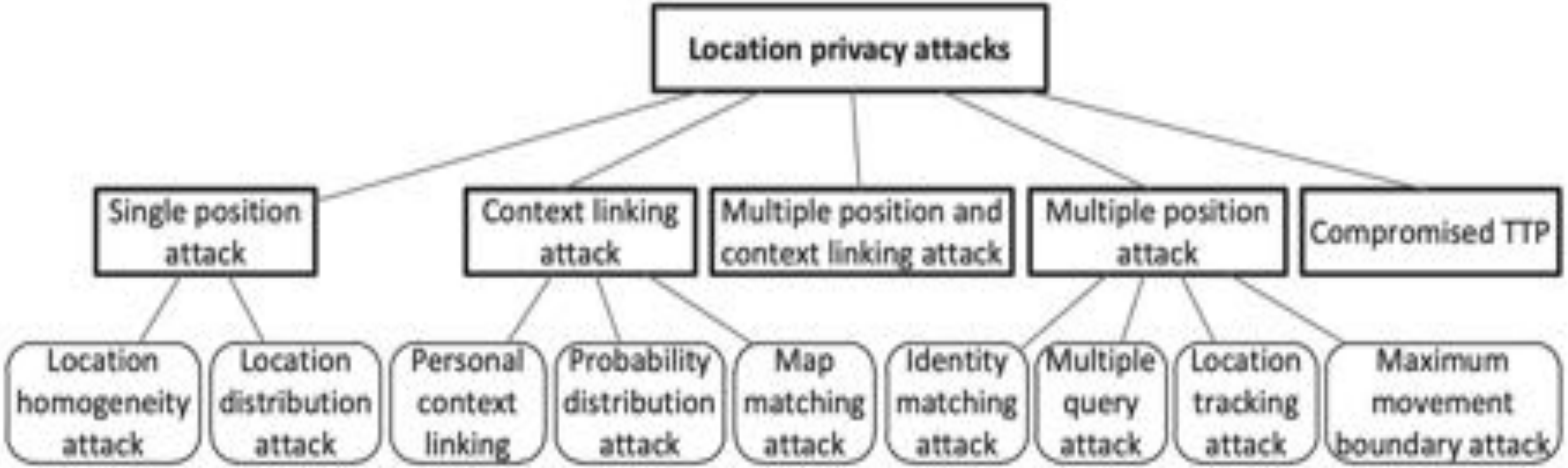
$\text{OBUverify}(pk_{OBU}, m, s_m)$
// **Main loop**
For all $1 \leq k \leq N$ *tuples do:*
$t'_{c_{p_k}} = c_{p_k}^{ch} g_0^{s_{p_k}} g_1^{s_{open_{p_k}}}$
$t'_Z = Z^{ch} \hat{A}^{s_e} R^{s_{p_k}} S^{s_v} (1/g_0)^{s_{w \cdot e}}$
$t'_{c_w} = c_w^{ch} g_0^{s_w} g_1^{s_{open_w}}$
$t' = C_w^{s_e} (1/g_0)^{s_{w \cdot e}} (1/g_1)^{s_{open_w \cdot e}}$
$ch' = H(\beta \| t'_{c_{p_k}} \| t'_Z \| t'_{c_w} \| t') ? = ch$
$s_e \in \{0,1\}^{l_e + l_c + l_z}$
$s_{p_k} \in \{0,1\}^{l_p + l_c + l_z}$
End for
// **Commitment validation**
$c'_{fee} = \prod_{k=1}^{N} c_{p_k}$
$c_{fee} = g_0^{fee} g_1^{open_{fee}} \pmod{n}$
$c_{fee} ? = c'_{fee}$

$\alpha \in \{p_k, open_{p_k}, e, v, w, open_w, w \cdot e, open_{w \cdot e}\}$
$\beta = (n \| g_0 \| g_1 \| \hat{A} \| R \| S \| g_0^{-1} \| g_1^{-1} \| c_{p_k} \| Z \| c_w \| 1)$

06

Attacks

Attacking K-Anonymity



Attacker's knowledge about location information

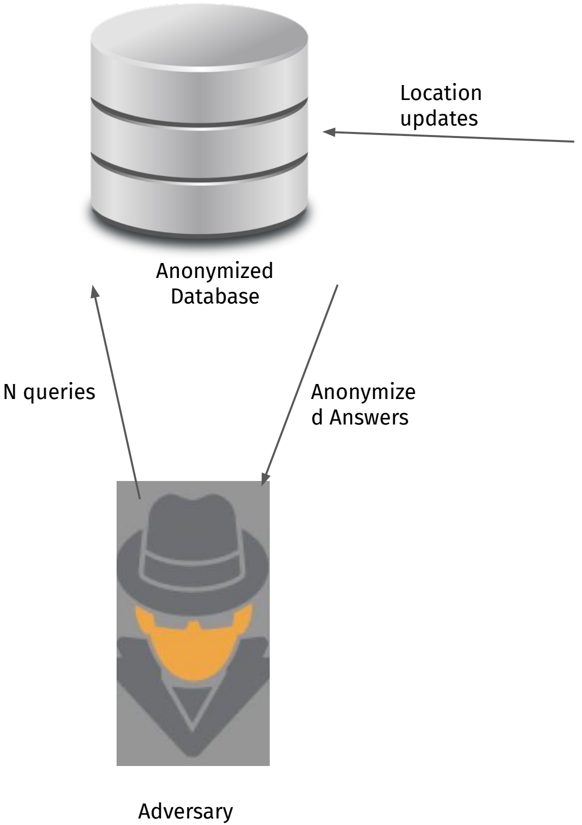
2 Types of information Knowledge

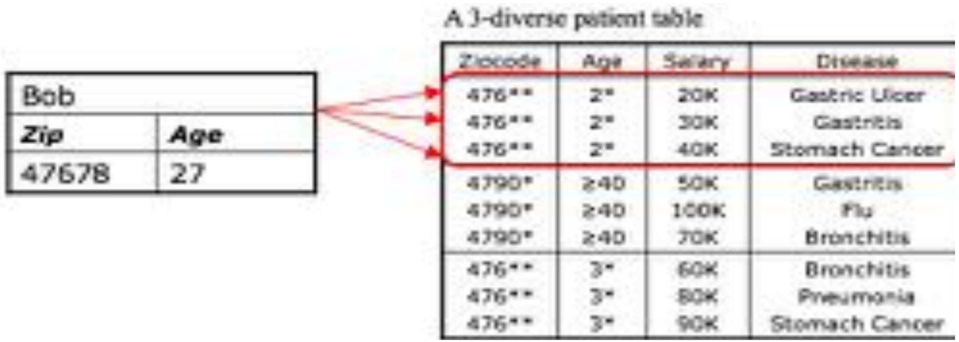
- **Spatiotemporal information** : Attacker has access only to location data of a user (either a single snapshot of users' position, or multiple snapshots over a period of time or even the complete trajectory of the user).

- **Context information** : where the attacker gains, apart from spatiotemporal information about user, extra information about user's address, place of interest, phonebook etc.

Why is this important? Based on the knowledge an attacker has, he can perform different kind of attacks, gain more knowledge and perform more attacks

Attack Modeling

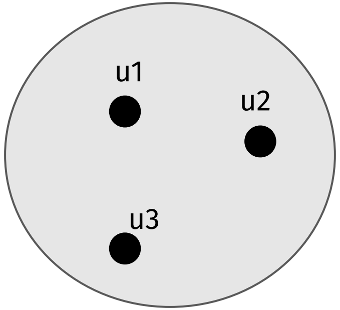




Bob	
Zip	*Age*
47678	27

A 3-diverse patient table

Zipcode	Age	Salary	Disease
476**	2*	20K	Gastric Ulcer
476**	2*	30K	Gastritis
476**	2*	40K	Stomach Cancer
4790*	≥40	50K	Gastritis
4790*	≥40	100K	Flu
4790*	≥40	70K	Bronchitis
476**	3*	60K	Bronchitis
476**	3*	80K	Pneumonia
476**	3*	90K	Stomach Cancer



Cloaking area

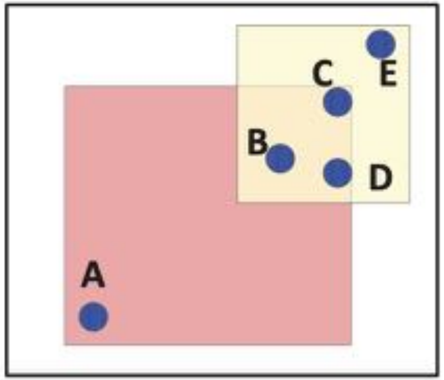
Single position Attacks

Location Homogeneity attack
In this attack, the attacker learns that the sensitives' use location, be learning that the each one of (k-1) other-users are located in an almost identical aera.

If the users are distributed over a large area then the attack is not possible.

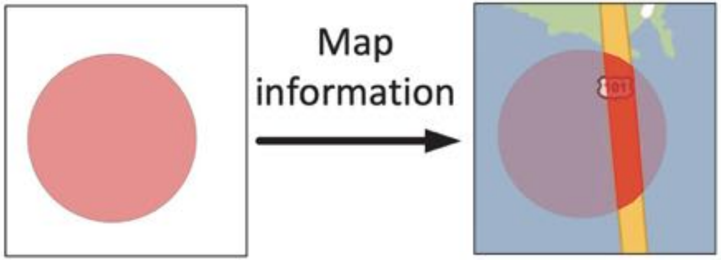
Location distribution attack

Attacker learns some information about the user's location by observing the behaviour of the algorithm. This happening when people are not equally distributed in the cloaking area.



Map Matching attacks

Attacker reduces the obfuscation area by using a map to remove all of the irrelevant areas.



Context Linking Attacks

Personal context linking attack

- Attacker has personal context information that can help him locate a user. For example an attacker can reduce the obfuscation area to locations of churches/clinics etc. (within the obfuscation area, aka area ok k-anonymity) if he knows that the user visits that place a specific time.

Probability distribution attack

- An attacker uses context information knowledge to calculate the possibility of the user being to different known -to the attacker- places. The places were extracted from the attacker who analysed the context/environmental information and statistics .

Multiple Position Attack

Shrink region attack

The attacker monitors location updates/queries and the corresponding members of the k-anonymity set. If the members of the set change, an attacker can infer which user sent the initial update or query.

Region intersection attack

The attacker uses several location updates/queries from a user to calculate their intersection. From the intersections if the attack succeed, the attacker can infer where the user is located.

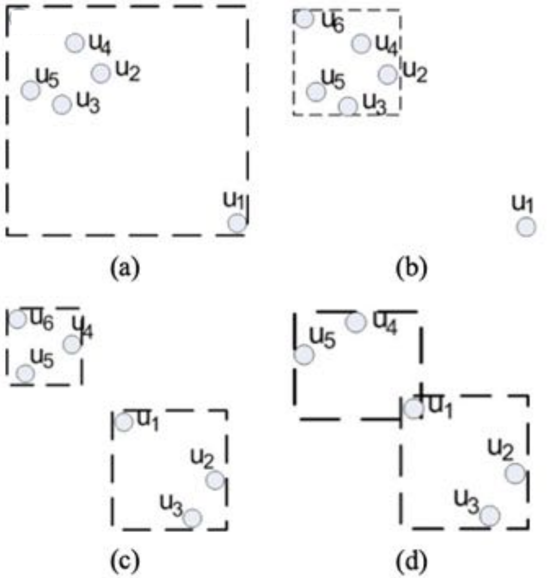
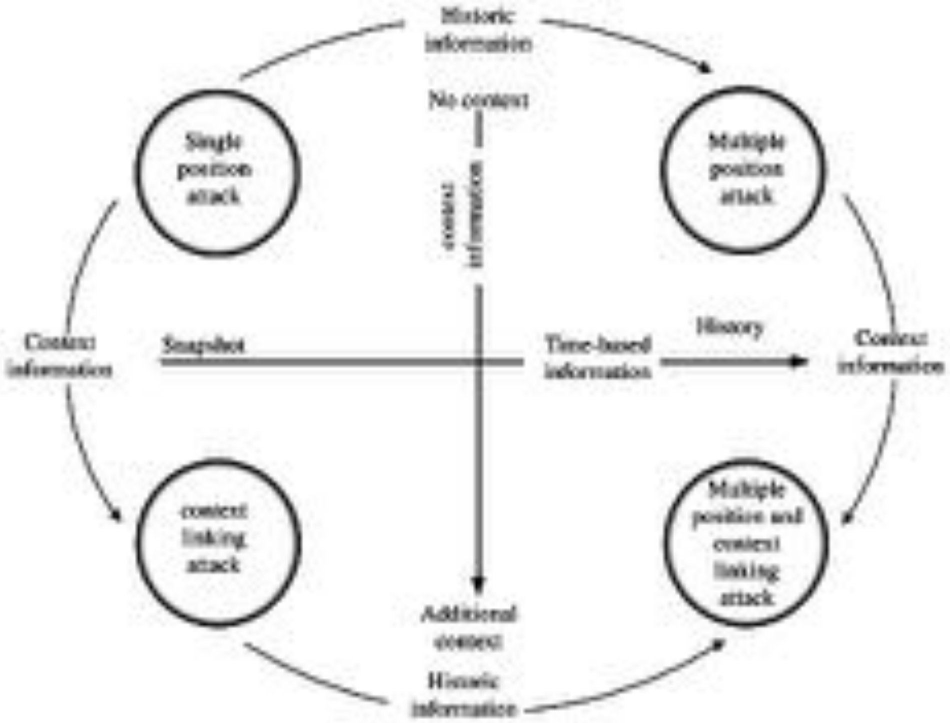


Figure 2. (a) (b) Shrink region attack (c) (d) Region intersection attack

Other Attacks And Transition Between Attacks

- Combined multiple position and context linking: Combination of different mentioned attacks

- Compromised trusted third-parties: Attacker gains access to data stored in (Usually) LBSs servers.



Conclusion

- ➔ Service vs privacy tradeoff
- ➔ Misalignment of incentives