Location Privacy

Cyrus Shahabi, Ph.D.
Professor of Computer Science, Electrical Engineering & Spatial Sciences
Viterbi School of Engineering
University of Southern California
Los Angeles, CA 900890781
shahabi@usc.edu

Privacy i… …d Service…


BBC NEWS
ONE-MINUTE WORLD NEWS
Page last updated at 02:52 GMT, Thursday, 14 August 2008 03:52 UK
E-mail this to a friend Printable version
Privacy worry over location data
By Maggie Shiels
Technology reporter, BBC News, Silicon Valley
Privacy advocates are warning of the dangers of rushing headlong into using location based services.
The Centre for Digital Democracy told the BBC that "while these services will be a powerful force in our lives they are a potential privacy nightmare."


February 12, 2009 7:27 AM PST
Obama's BlackBerry brings personal safety risks
by Chris Soghoian

When the mainstream media first announced Barack Obama's "victory" in keeping his BlackBerry, the focus was on the security of the device, and keeping the U.S. president's e-mail communications private from spies and hackers.

The news coverage and analysis by armchair security experts thus far has failed to focus on the real threat attacks against President Obama's location privacy, and the potential physical security risks that come with someone knowing the president's real-time physical location.

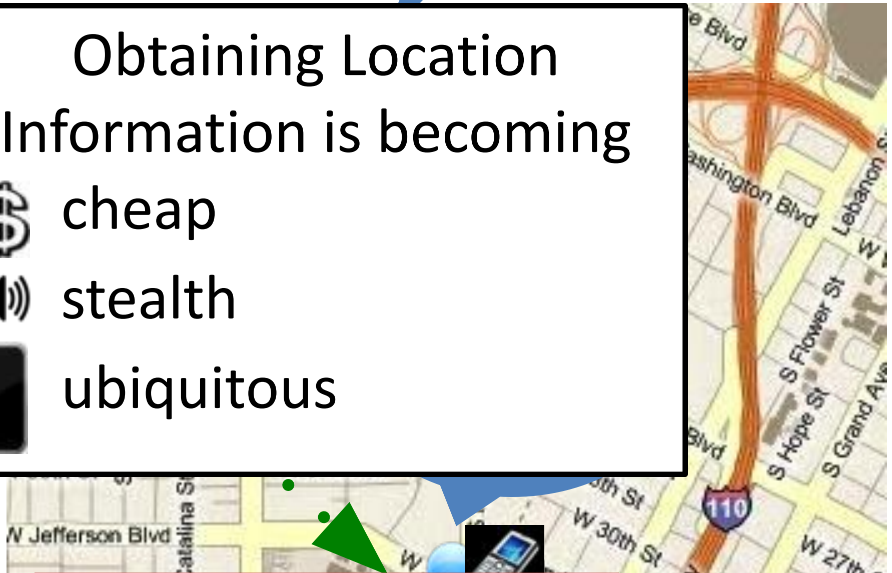
Serial numbers
Before we dive in, let's take a moment to note that each mobile phone has a unique serial number, known as an IMEI, or MEID. This unique number is transmitted in clear text every time the phone communicates with a nearby cell tower. Thus, while the contents of a phone call or the data session (for e-mail) are usually encrypted, anyone with the right equipment can home in on a particular IMEI and identify the location of the source of that signal.

The most common device used to locate a phone by its IMEI is a "Triggerfish", a piece of equipment that is routinely used by law enforcement and intelligence agencies. This kind of device tricks nearby cell phones into transmitting their serial numbers and other information by impersonating a cell tower.

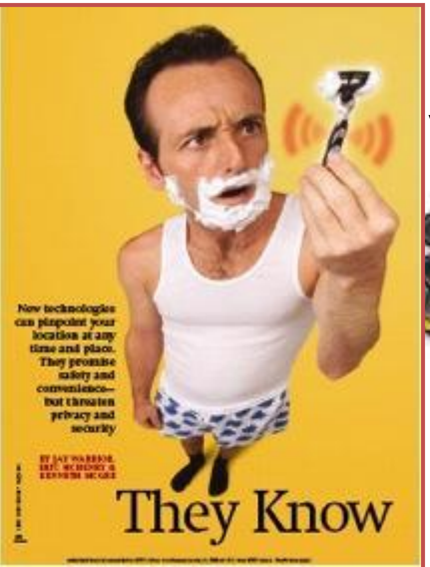
The devices, which are actually fairly low-tech, were used to hunt down famed hacker Kevin Mitnick back in the 1990s.

POI

Location Server (LS)



Obtaining Location Information is becoming
- cheap
- stealth
- ubiquitous




Bits
Business ▪ Innovation ▪ Technology ▪ Society
April 27, 2010, 4:27 PM
Senators Ask Facebook for Privacy Fixes
By MIGUEL HELFT
4:09 p.m. | Updated Added a report from the senators' meeting with Facebook executives.
Four senators are raising privacy concerns over new features that Facebook introduced last week and have asked the social networking company to roll back some of the changes.

The New York Times
Monday, May 3, 2010


Google Latitude Spurs Privacy Backlash
A consumer-advocacy group already sees five scenarios in which the Google Maps add-on could be abused.
By Thomas Claburn
InformationWeek
February 5, 2009 06:25 PM
Google's new Latitude location-sharing service "could be a gift to stalkers, prying employers, jealous partners, and obsessive friends," Privacy International warned Thursday.
Google introduced Latitude on Wednesday. It's a new Google Maps feature that lets users share location data with friends, using either a mobile phone or Google Gears-equipped computer.
Google knows well that it has a privacy problem, exemplified by its quixotic campaign last summer to avoid adding a link on its home


The Washington Post
Online Data Gets Personal: Cell Phone Records for Sale
By Jonathan Krim
Washington Post Staff Writer
Friday, July 8, 2005; D01
They're not just after your credit card or Social Security numbers.
Fueled by the ease of online commerce, snoops are on the trail of other personal information: records of phone calls, especially from cell phones.

Location Privacy Threats

FOX NEWS .com

U.S. & WORLD

Updated: 3-28-06 9:42pm ET

SEARCH GO

E-MAIL STORY PRINTER FRIENDLY FOXFAN CENTRAL

Man Accused of Stalking Ex-Girlfriend With GPS

Saturday, September 04, 2004

Associated Press

GLENDALE, Calif. — Police arrested a man they said tracked his ex-girlfriend's whereabouts by attaching a global positioning system (search) to her car.

Ara Gabrielyan, 32, was arrested Aug. 29 on one count of **stalking** (search) and three counts of making criminal threats. He was being held on $500,000 bail and was to be arraigned Wednesday.

"This is what I would consider stalking of the 21st century," police Lt. Jon Perkins said.

http://www.foxnews.com/story/0,2933,131487,00.html

USA TODAY Classifieds: cars.com | careerbuilder.com | eHarmony.com |

Home
News
Travel
Money
Sports
Life
Tech
Weather
Search
powered by YAHOO! GO

Tech Products
Products home
Edward C. Baig
Kim Komando
Ask Kim
Gaming
Gaming home
Arcade
Jinny Gudmundsen
Marc Saltzman
Science & Space
Science & Space
April Holladay
Dan Vergano

Tech

• E-MAIL THIS • PRINT THIS • SAVE THIS • MOST POPULAR • SUBSC

Posted 12/30/2002 7:57 PM

Authorities: GPS system used to stalk woman

KENOSHA, Wis. (AP) — A man was charged Monday with stalking his former live-in girlfriend with help from a high-tech homing device placed under the hood of her car.

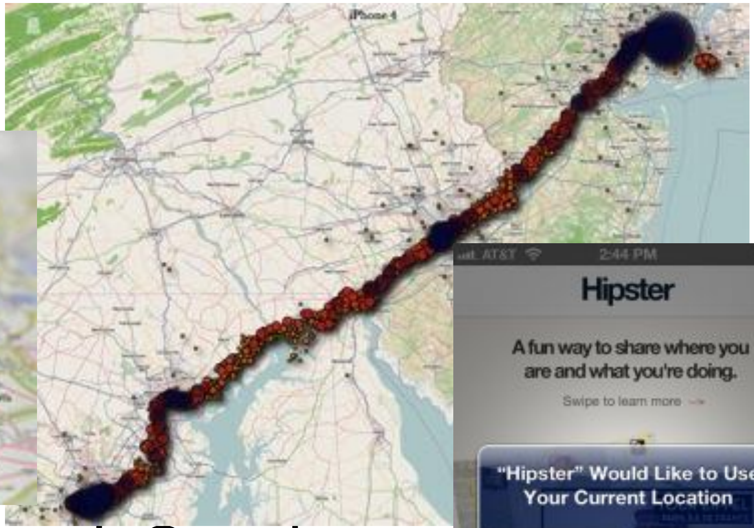
Paul Seidler, 42, was arrested during the weekend. On Monday, he was charged with stalking, burglary, second-degree reckless endangerment and disorderly conduct, and ordered held on $50,000 bail.

According to a criminal complaint, Connie Adams asked Seidler to move out of her apartment Oct. 25 after a three-year relationship. Prosecutors say he immediately began following her, including when she ran errands and went to work.

http://www.usatoday.com/tech/news/2002-12-30-gps-stalker_x.htm

Location Privacy in Industry

- ~ 26,000 persons are victims of GPS stalking annually, including by cellphone
  - [Jan 2009 report by the Department of Justice]



- ~ *50% top apps for Apple iPhones* and Google Android smartphones disclosed a user's location to third parties without his or her consent
  - [Dec 2010 investigation by the *Wall Street Journal]*

Location Privacy in Industry

- In April 2011, consumers learned that their smartphones were automatically sending out information about their smartphone's location

[The Location Privacy Protection Act of 2011]

f Reg router, courtesy of Google Maps

Google location tracking can invade privacy, hackers say

Unique IDs + router addresses = potential abuse

Got an iPhone or 3G iPad? Apple is recording your moves

A hidden file in iOS 4 is regularly recording the position of devices.

Mobile Apps for Kids:

Current Privacy Disclosures are Dis*app*ointing

Location Privacy Protection Act 2011

- **The Location Privacy Protection Act of 2011** requires any company that may obtain a customer's location information from his smartphone to
  1) Get that customer's express consent before collecting his location data
  2) Get that customer's express consent before sharing his or her location data with third parties

Apple, Google, And Others Agree To Mobile App Privacy Policy Guidelines

Location Priv

Location Infor

- [x] Include
- [x] Let me
- [x] When m
Faceboo
- [x] Let venu
custome

venue ?

ir check-in tweets or

one of their best

- All these
- One is p … an follow
from a f … tos, current
location

Briefing

Tech

NEW BEST FRIEND!

Where Everybody Knows Your Name. Apps tell strangers what they have in common

By Harry McCracken

SOCIAL NETWORKS FIRST persuaded millions of us to start cataloging our friends, family members and high school classmates. The networks got us to post photos, tweet our every thought and tend our virtual farms. Now the next wave wants to cross over into the real world and introduce us to nearby strangers with common interests—and perhaps a desire to make a new friend.

There are at least 11 new smart-phone apps pushing this concept, which techies call ambient social networking. Silicon Valley is rushing to fund these "people discovery" start-ups, and everybody at South by Southwest (SXSW) Interactive—the annual nerdfest in Austin that famously gave Twitter its big break in 2007—seemed to be tinkering with one of them: Highlight, an eight-week-old iPhone app, is designed to reveal real-life connections you didn't know you had, as well as alert you to the presence of friends you might otherwise miss. Co-founder Paul Davison calls it a "sixth sense."

Highlight, which has yet to make public how many people are using the app, works by rummaging through your Facebook account to see whom you know and what topics you like. Then it uses your iPhone's GPS to inform you when, say, a fellow conference attendee who's a former co-worker's buddy is in your immediate vicinity or when a good-looking patron who loves the same bands you do sits down at the other end of the bar.

It's a big shift for the tech industry. Unlike Foursquare—2009's SXSW darling, which now has 15 million members sharing their locations by "checking in" so they can earn discounts and other rewards—Highlight monitors your whereabouts continuously and automatically shares them with fellow members both in and outside your existing circle of friends. That introduces new privacy concerns and strikes some critics as enabling a form of high-tech stalking.

In its current form, Highlight is a rough draft of a powerful idea. Some problems are minor: Highlight has an odd habit of telling you who's nearby even when you're passing in a moving vehicle. It also drains your phone's battery as it constantly sends location data back to its servers, a problem the company says it is addressing.

But getting Highlight's algorithm to highlight people you actually want to meet is the biggest challenge of all. "We're just scratching the surface," says Davison. "If we both went to the same high school, it's more interesting if the school is 4,000 miles away than if it's two miles away." At SXSW, I wasn't moved to track down any of the individuals Highlight identified as people of interest. I did, however, keep striking up rewarding conversations with folks I encountered in hotel lobbies and at parties, no app required. Serendipity in its natural form is a wonderful thing—and manufacturing it won't be easy.

Come Here Often? Apps that help you meet people

Highlight Launched Jan. 24, this buzzed-about app uses GPS to let you know when a friend of one of your Facebook friends is close by

Glancee Gaining steam, with 10,000 members, the app pinpoints others who are "steps away" and share your interests

Foursquare One of the first location-sharing apps, it has 15 million users "checking in," a.k.a. broadcasting where they are, to earn perks

OkCupid Matchmaking gone mobile. This site's Locals feature sorts nearby singles by compatibility
—KEITH WAGSTAFF

18

TIME March 26, 2012

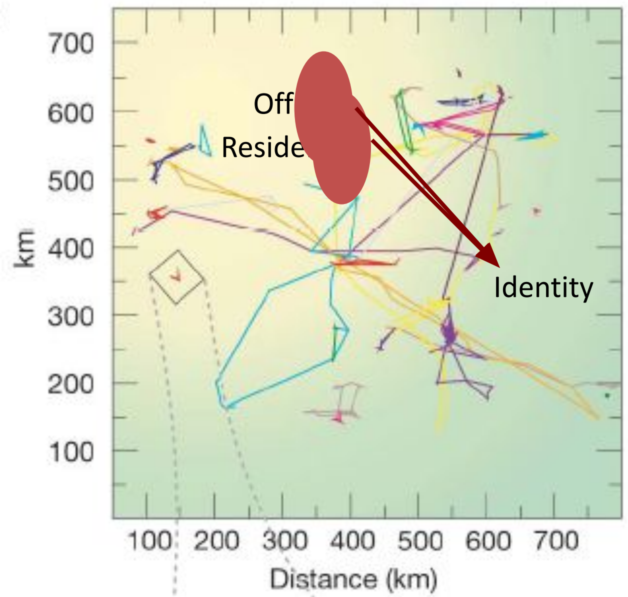
Isn’t Confidentiality Enough?

Sensitive information obtained by anonymous location data

- Baraba´si et al., Nature’08

Human Mobility ⟹ Spatial Probability Distribution

- Four spatiotemporal points are enough to uniquely re-identify 90% of individuals



- Anonymous queries leak information

Location Queries ⟹ Affiliations (political, religious, etc.)

User Perception of Location Privacy

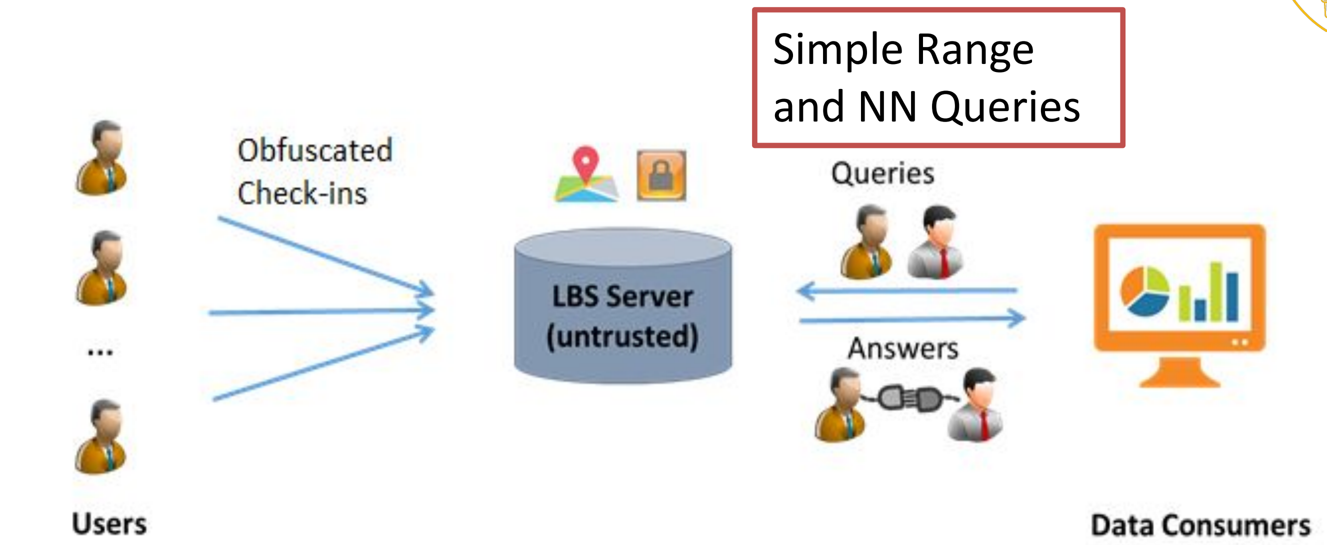
One World – Two Views

- Location-based services rely on the *implicit* assumption that users agree on revealing their private user locations
- Location-based services *trade* their services with privacy
  - If a user wants to keep her location privacy, she has to turn off her location-detection device and (temporarily) unsubscribe from the service
- *Pseudonymity* is not applicable as the user location can directly lead to its identity

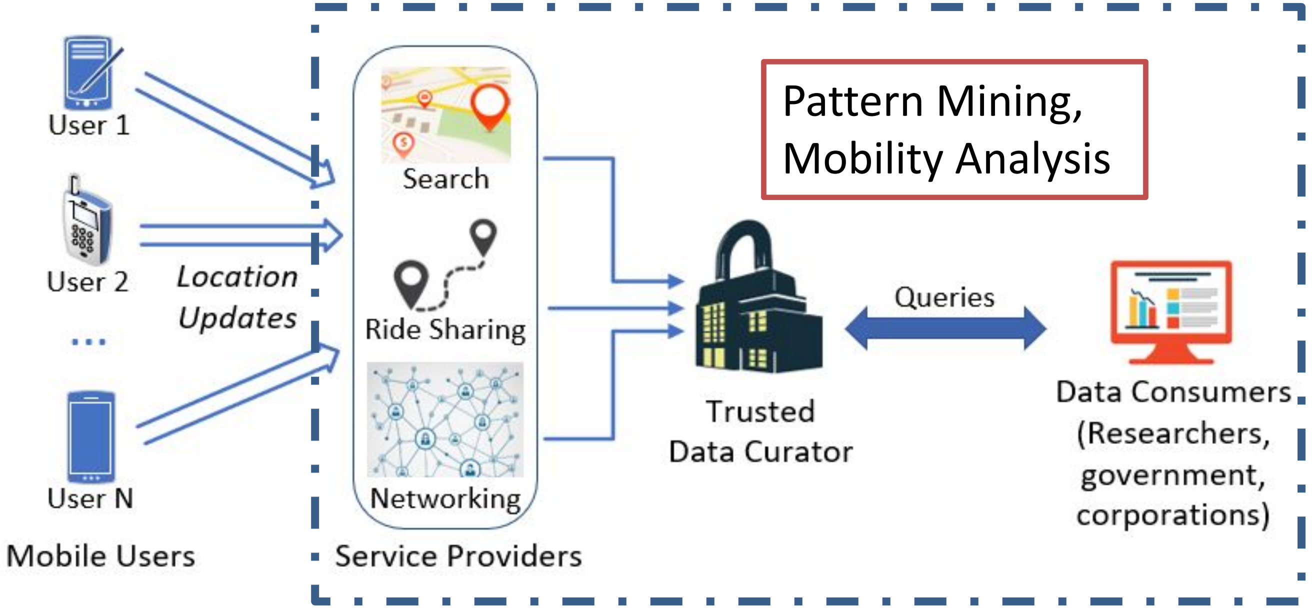
Several social studies report that users become more aware about their privacy and may end up not using any of the location-based services

System Models

Online Setting:



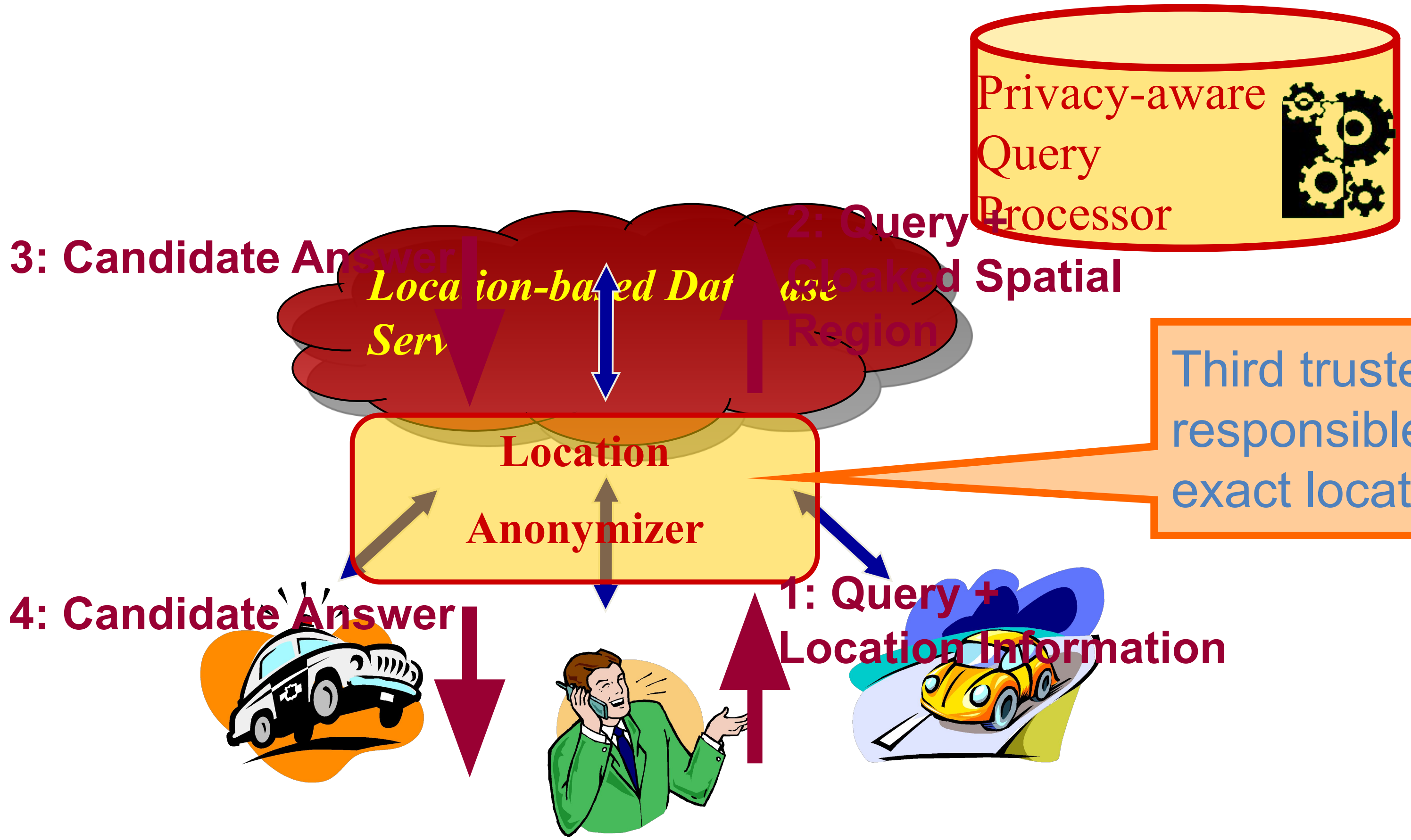
Offline Setting:



System Architectures for Online Location Privacy

- *Third trusted party architecture*
  - A centralized trusted entity is responsible for gathering information and providing the required privacy for each user
  - Analogous to output perturbation
- *Client-Server architecture*
  - Users communicate directly with the sever with noisy locations.
  - Analogous to input perturbation
- *Peer-to-Peer cooperative architecture*
  - Users collaborate with each other without the interleaving of a centralized entity to provide customized privacy for each single user

Third Trusted Party Architecture

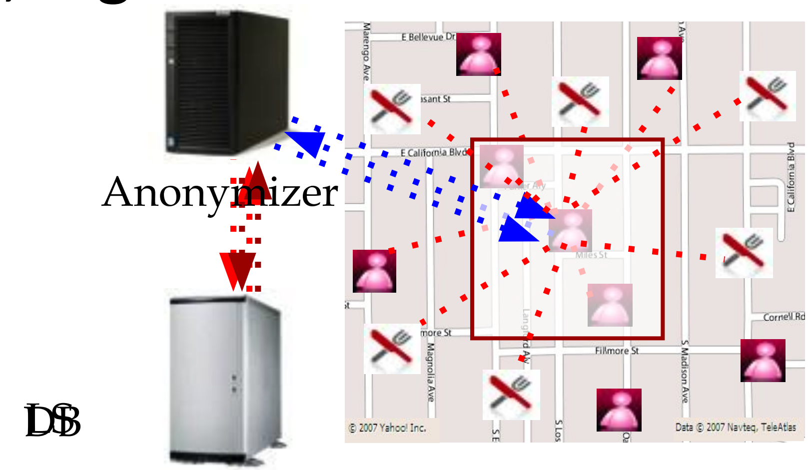


Third trusted party that is responsible on *blurring* the exact location information.

Location Privacy

Encryption-Based (reduce spatial characteristics to 1D, e.g. Hilbert curve)

- Private information retrieval *[Ghinita et al. SIGMOD 2008]*
- Space transformation *[Khoshgozaran & Shahabi SSTD 2007]*

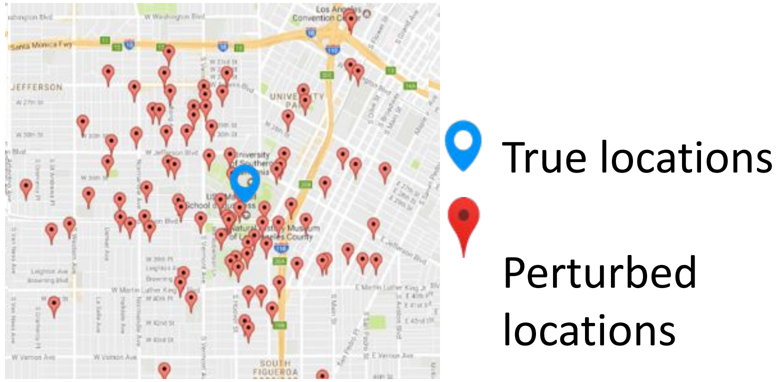


Anonymity based ("privacy in the crowd")

- Location Anonymizers *[Pfitzmann et al. 2010]*
- Spatial K-anonymity/Cloaking region *[Ghintita'10]*

Perturbation (e.g., differential privacy)

- Geo-indistinguishability *[Andrés et al CCS 2013]*
- δ-location set-based differential privacy *[Xiao & Xiong CCS*



Third Trusted Party Architecture

- A *trusted third party* receives the exact locations from clients, blurs the locations, and sends the blurred locations to the server

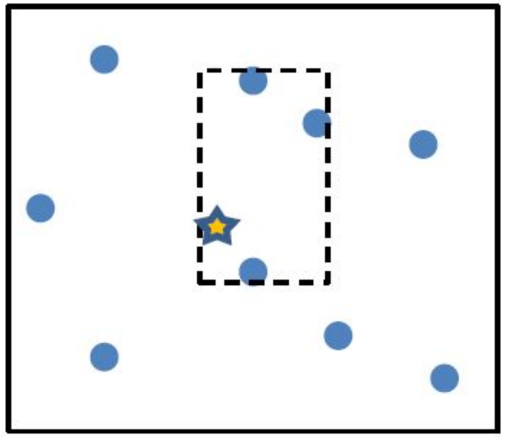
- Provide powerful privacy guarantees with high-quality services

- System bottleneck and sophisticated implementations

- *Examples:* Casper, CliqueCloak, and spatio-temporal cloaking

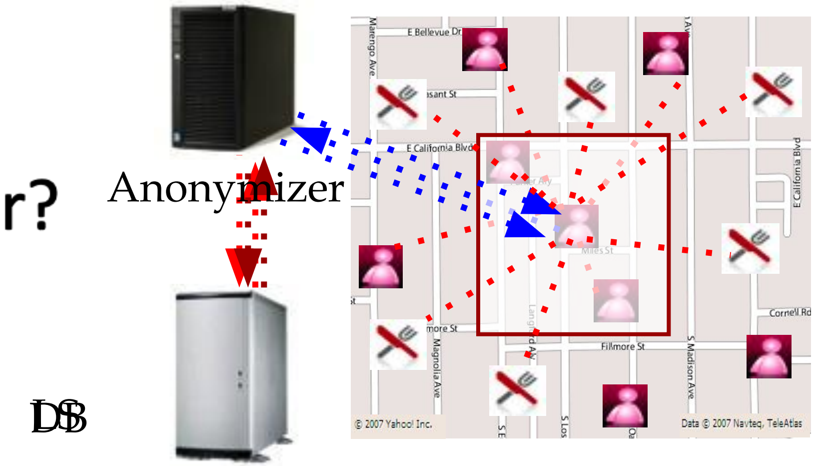
Location *k*-Anonymity

- Submitted cloaked region must contain at least *k* users
  - Called the *Anonymized Spatial Region* (ASR)
  - Collect and submit *k* queries together



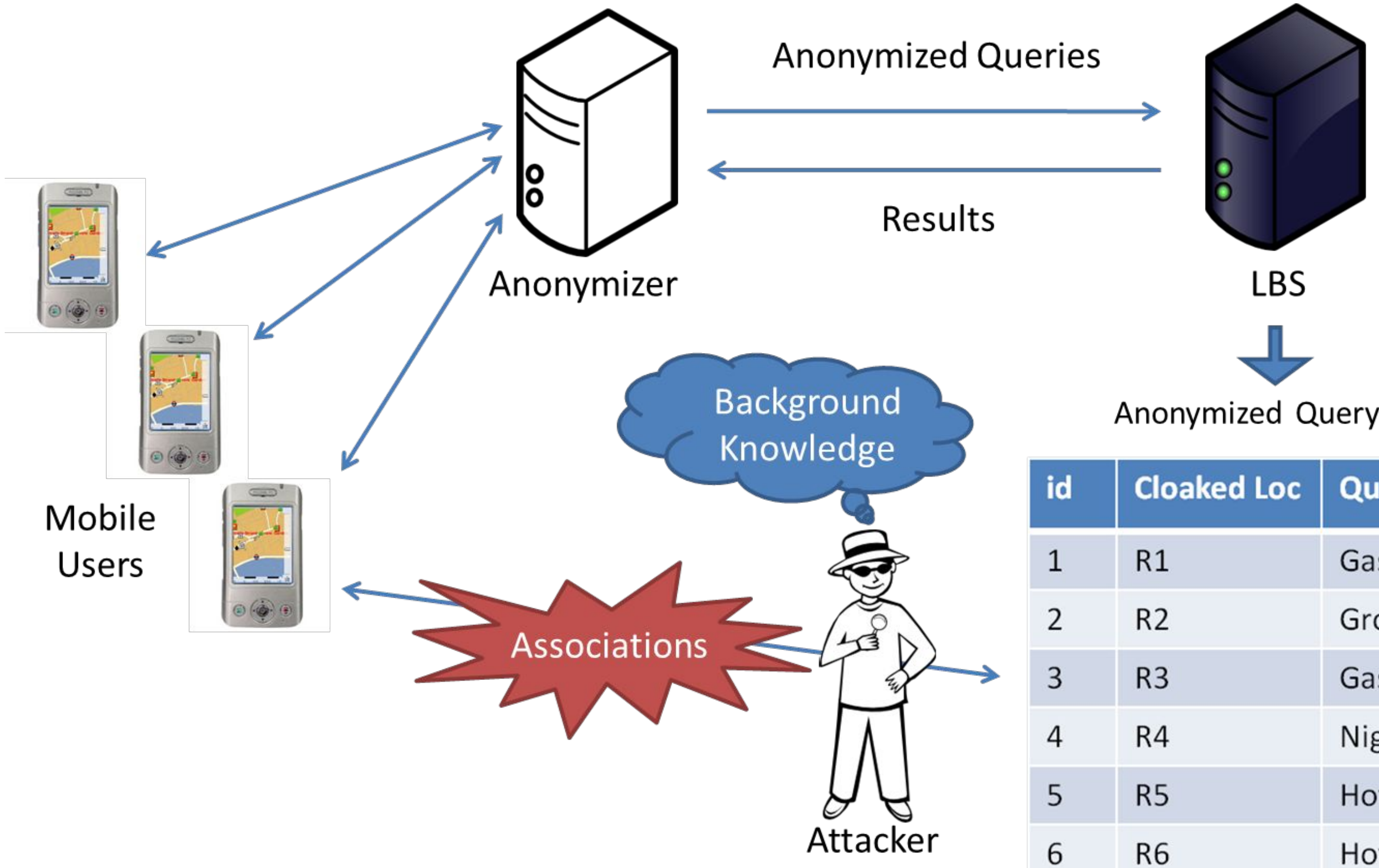
  - If not enough queries to group with
    - Drop the query (may not be acceptable)
    - Generate enough dummy (fake) queries (raises service cost)

- What if k other users are too close to each other?



What if in a sparse area? | Hybrid

LBS Anonymization: Threat Model



id	Cloaked Loc	Query
1	R1	Gas station
2	R2	Grocery store
3	R3	Gas station
4	R4	Night club
5	R5	Hotel
6	R6	Hotel

Service-Privacy Trade-off

- First extreme:
  - A user reports her exact location 100% service
- Second extreme:
  - A user does NOT report her location 0% service

Desired Trade-off: A user reports a perturbed version of her location *x*% service

The Privacy-aware Query Processor

Dealing with Cloaked Regions

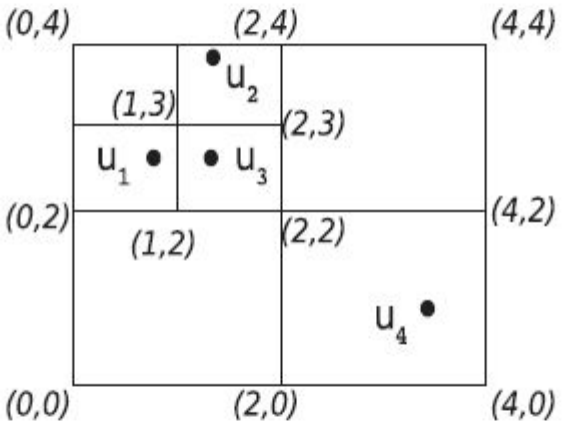
- A new privacy-aware query processor will be embedded inside the location-based database server to deal with spatial cloaked areas rather than exact location information

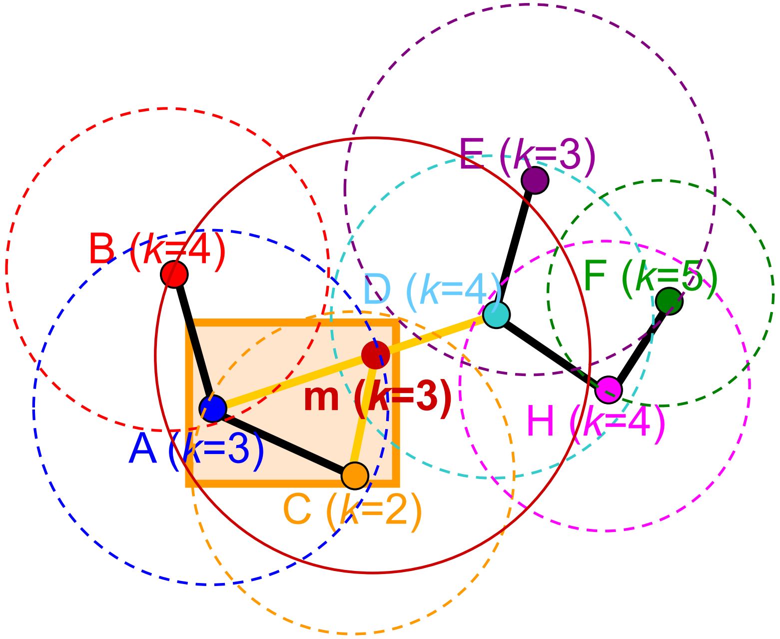
- Traditional Query:
  - ***What is my nearest gas station* given that *I am in this location***

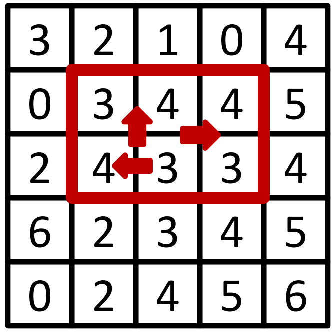
- New Query:
  - ***What is my nearest gas station* given that *I am somewhere in this region***

Location *k*-Anonymization

- Various algorithms
  - Nearest neighbor *k*-anonymization
  - Quad-tree spatial cloaking
  - CliqueCloak
  - Privacy Grid

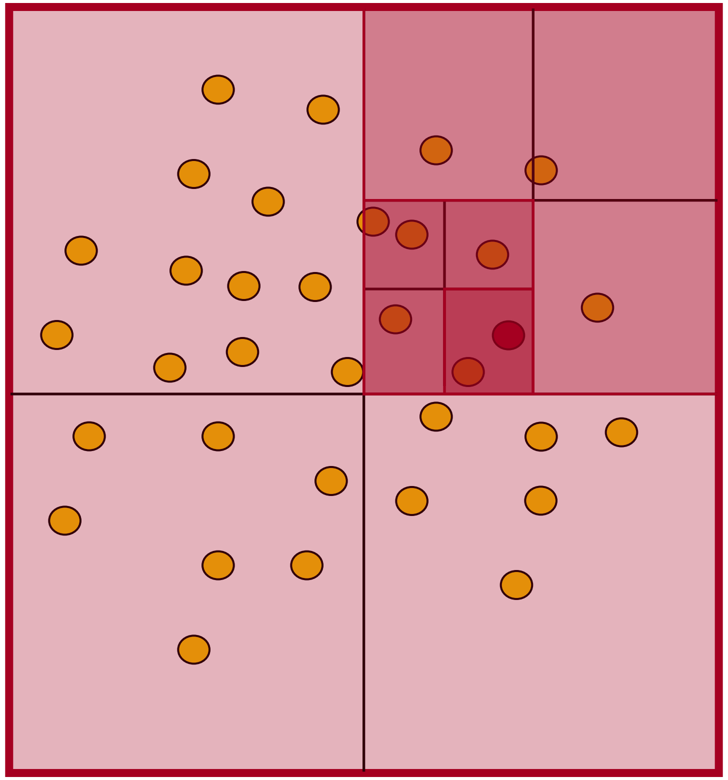






Third Trusted Party Architecture:
Quadtree Spatial Cloaking

- Achieve *k-anonymity*, i.e., a user is indistinguishable from other *k-1* users
- Recursively divide the space into quadrants until a quadrant has less than *k* users.
- The previous quadrant, which still meet the *k-anonymity* constraint, is returned

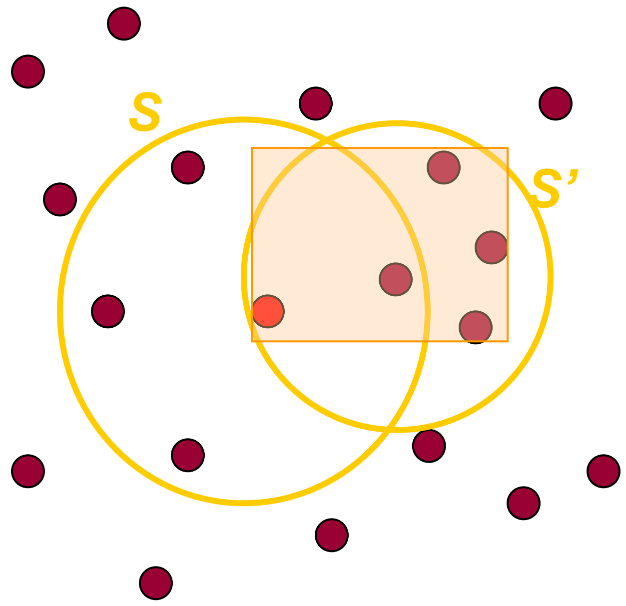


Achieve 5-anonmity for

Third Trusted Party Architecture:

Nearest-Neighbor k-Anonymizing

- STEP 1: Determine a set *S* containing *u* and *k* - 1 *u*'s nearest neighbors.
- Can we return the MBR of set S as anonymity region ?

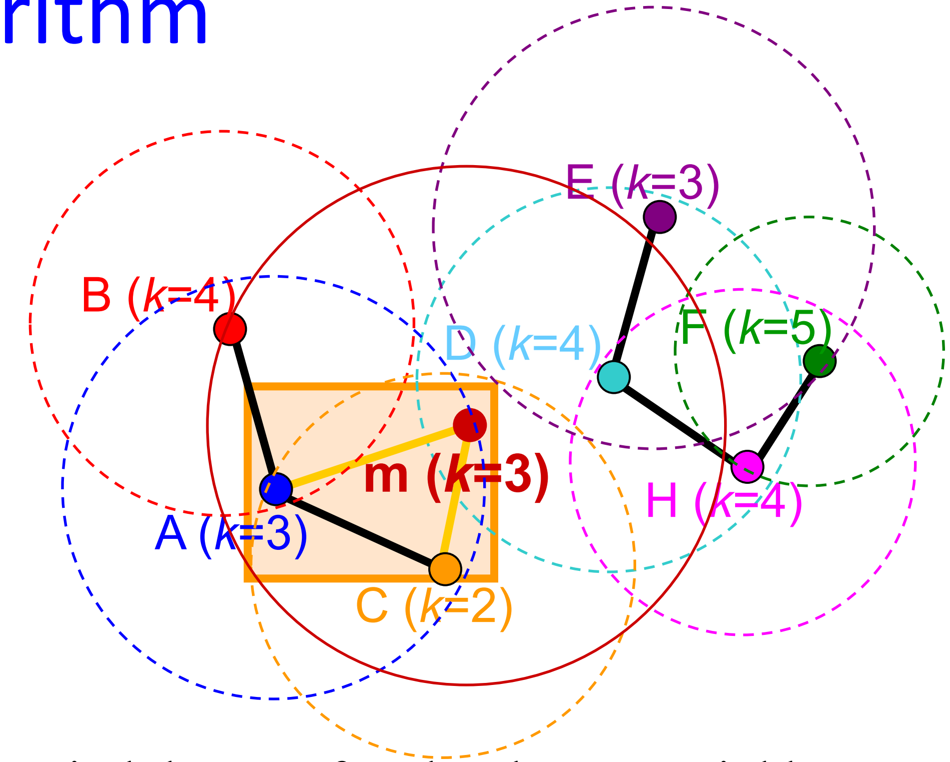


- STEP 2: Randomly select *v* from *S*.
- STEP 3: Determine a set *S'* containing *v* and *v*'s *k* - 1 nearest neighbors.
- STEP 4: A cloaked spatial region is an MBR of all users in *S'* and *u*.
- The main idea is that randomly selecting one of the k nearest neighbors achieves the k-anonymity

What if different users have different privacy requirements, service level needs

Third Trusted Party Architecture:
CliqueCloak Algorithm

- Each user requests:
  - A level of *k* anonymity
  - A maximum cloaked area
- Build an undirected constraint graph. Two nodes are neighbors, if their maximum areas contain each other.



- For a new user *m*, add *m* to the graph. Find the set of nodes that are neighbors to *m* in the graph and has level of anonymity <= *k*
- The cloaked region is the MBR that includes the user and neighboring nodes. All users within an MBR use that MBR as their cloaked region

Third Trusted Party Architecture:

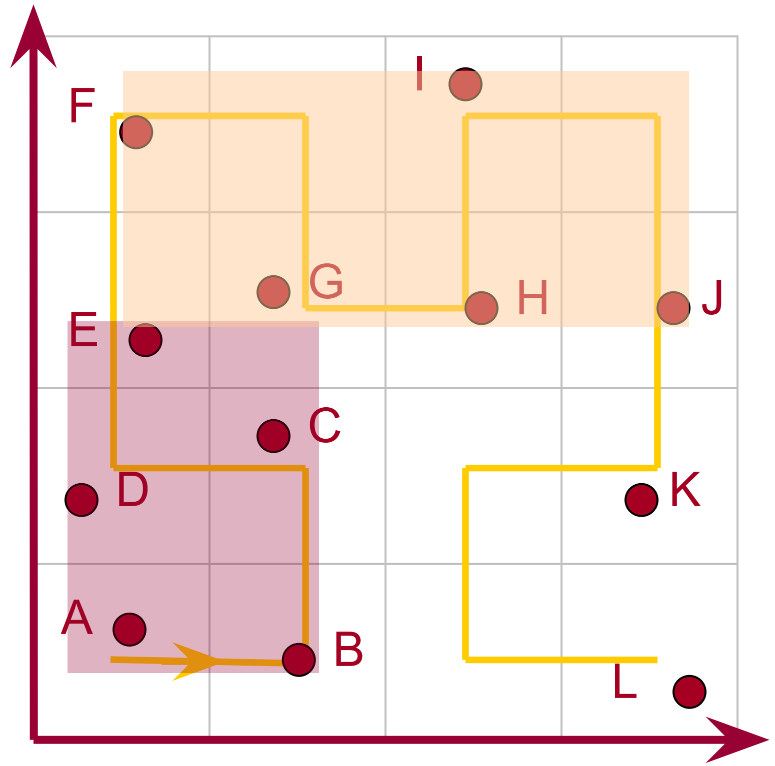
Hilbert k-Anonymizing

- All user locations are sorted based on their Hilbert order

- To anonymize a user, we compute *start* and *end* values as:
  - $start = rank_u - (rank_u \bmod k_u)$
  - $end = start + k_u - 1$

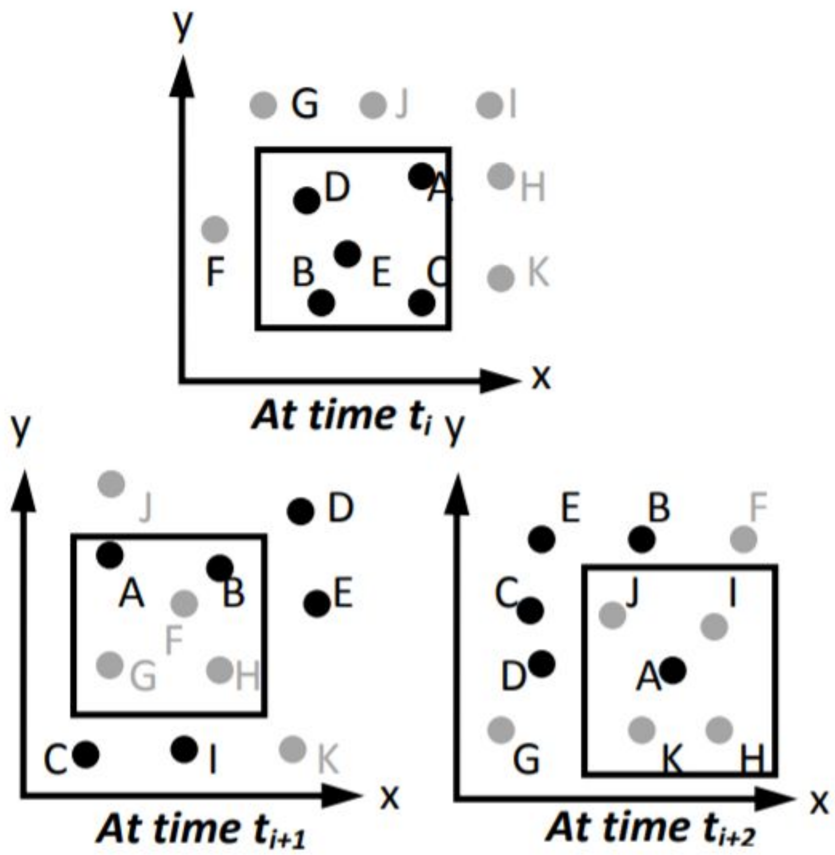
- A cloaked spatial region is an MBR of all users within the range (from *start* to *end*).

- The main idea is that it is always the case that k_u users would have the same *[start,end]* interval

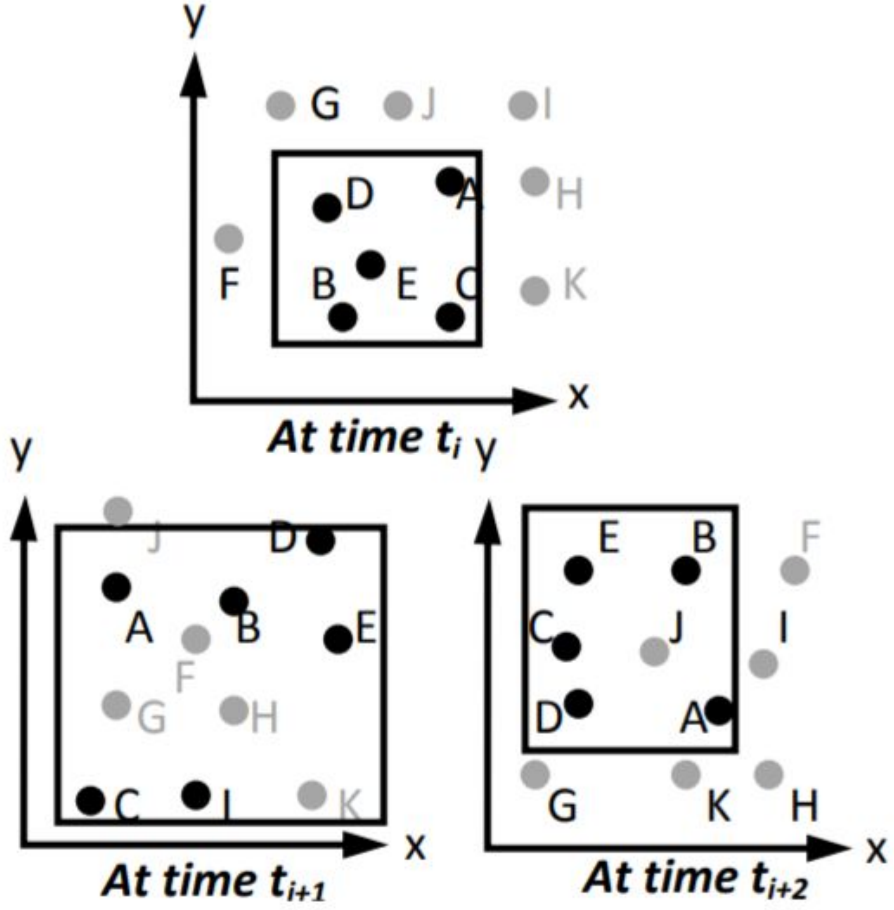


	A	B	C	D	E	F	G	H	I	J	K	L
k_u	6	5	4	5	4	5	6	5	7	4	5	4

Anonymizing Trajectories ?



- Correlation Attack
  - User A submits query at time i for k = 5
  - At time i + 1, his anonymity reduces to ½
  - At time I + 2, his identity is revealed.



Possible Solution.
But need a lot of noise.

System Architectures for Online Location Privacy

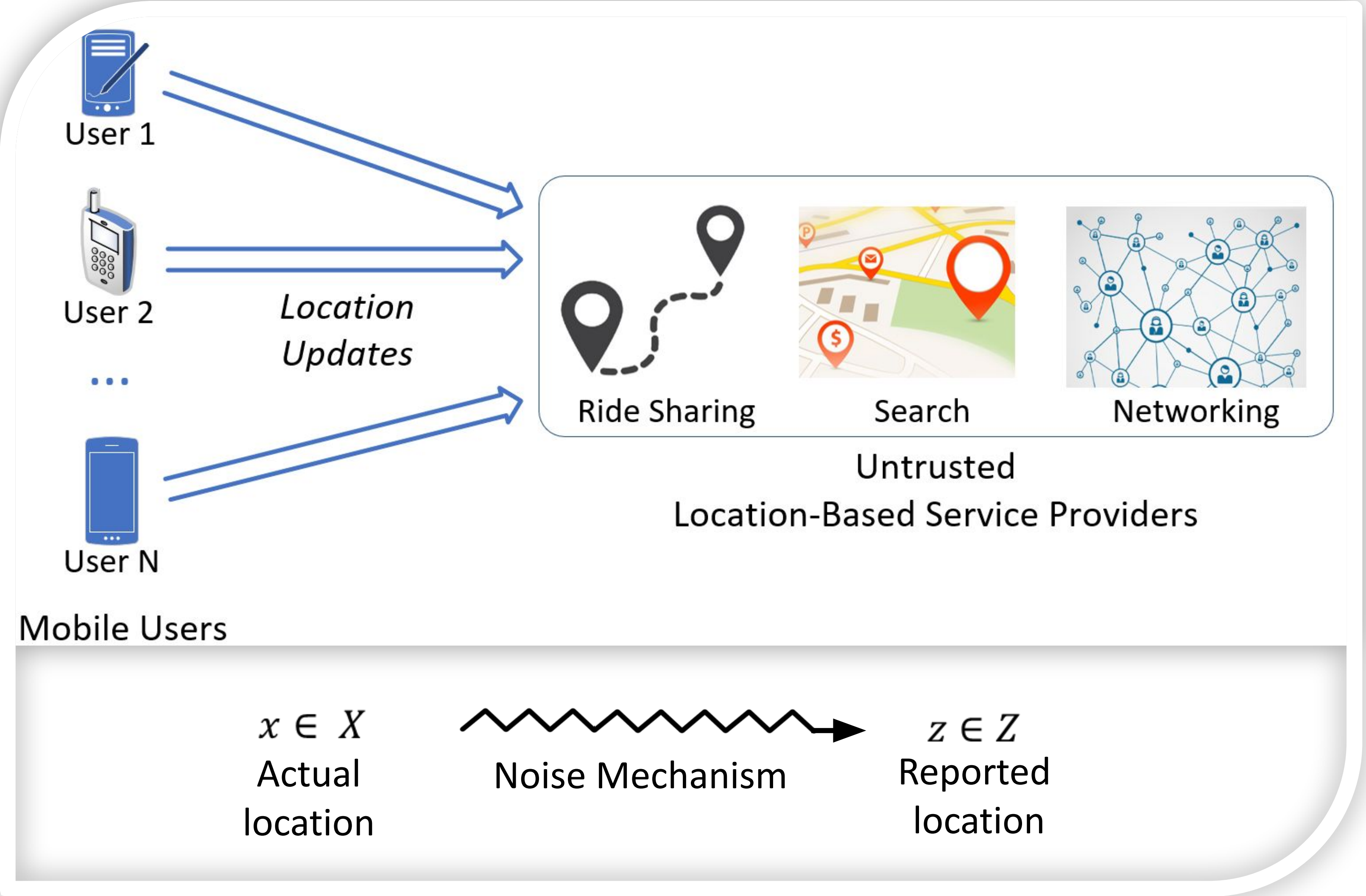
- *Third trusted party architecture*
  - A centralized trusted entity is responsible for gathering information and providing the required privacy for each user
  - Analogous to output perturbation
- *Client-Server architecture*
  - Users communicate directly with the sever with noisy locations.
  - Analogous to input perturbation
- *Peer-to-Peer cooperative architecture*
  - Users collaborate with each other without the interleaving of a centralized entity to provide customized privacy for each single user

Client-Server Architecture

- Users randomly perturb their inputs.
- No need for a trusted centralized party.
- More obfuscation means

Better Privacy ⬄ Utility Loss

e.g. requesting Uber.



Client-Server Architecture

- Clients try to *cheat* the server using either fake locations or fake space

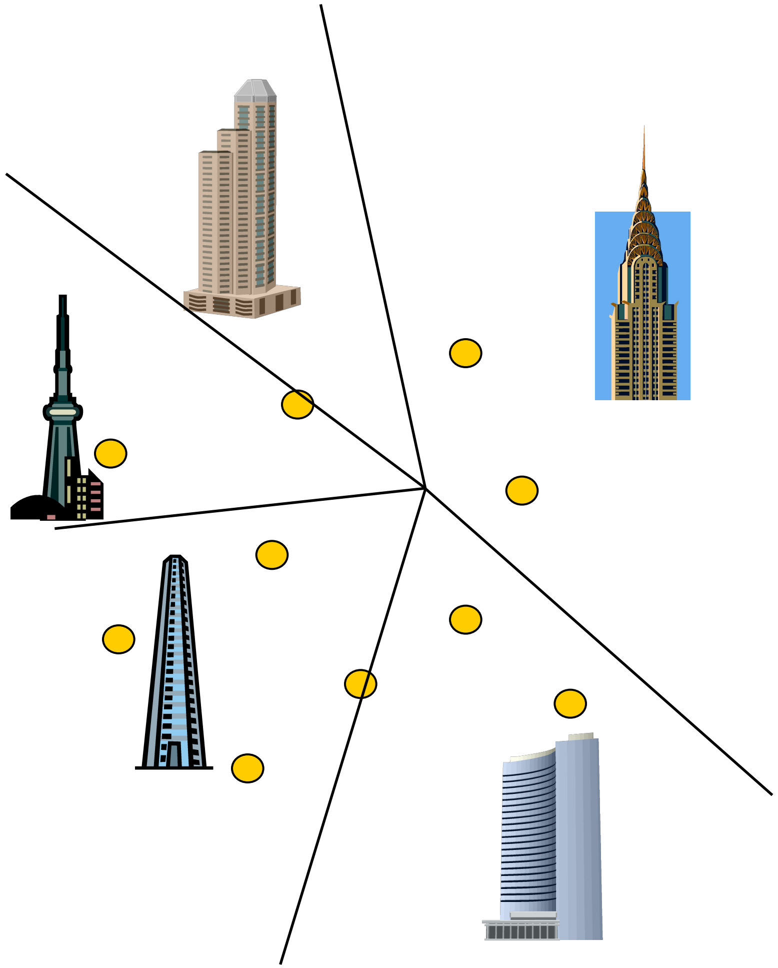
- Simple to implement, easy to integrate with existing technologies

- Lower quality of service

- *Examples*: Landmark objects, false dummies, and space transformation

Client-Server Architecture:
Landmark objects

- Instead of reporting the exact location, report the location of a closest landmark
- The query answer will be based on the landmark
- Voronoi diagrams can be used to identify the closest landmark



Moving to a better privacy definition

- Early efforts
  - Location Generalization.
  - Location Cloaking, k-anonymity models.

Lack of a formal privacy guarantee

- Geo-Indistinguishability [Andres et. al., CCS 2013]
  - A powerful model that mimics traditional Differential Privacy.
  - Broadens the scope, over distance metric.
  - prevents an adversary from inferring with high probability the user's whereabouts.

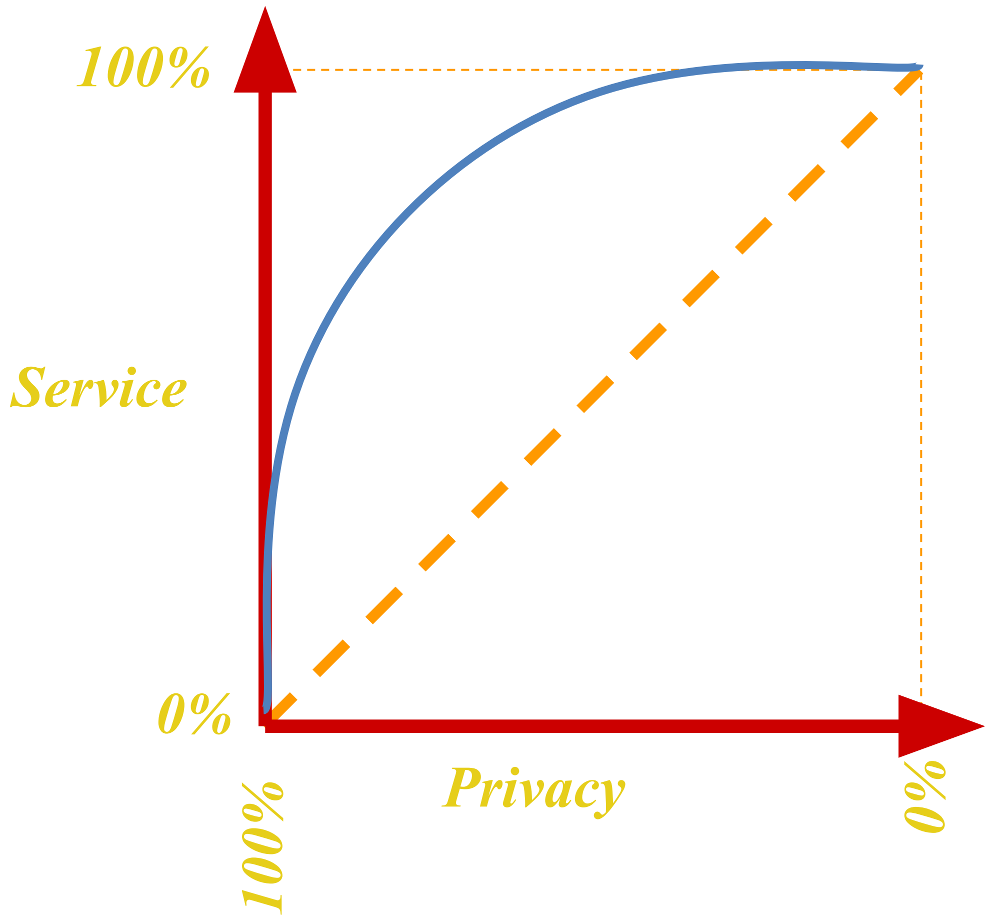
Protecting geo-coordinate with DP

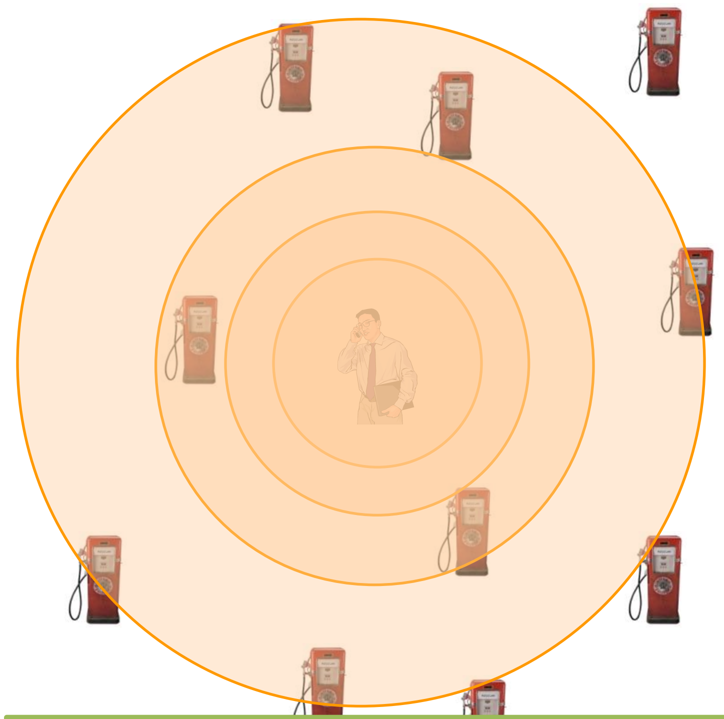
- What is the sensitivity of the following queries:
  - "Count of users who are taller than 6 feet?"
  - "Count of users present in this classroom?"

- Given a database of each users geo-coordinate:
  - "What is the location of a user ?"
  - Sensitivity is over the entire globe. Too high to be useful.

Need to relax privacy constraint.

Service-Privacy Trade-off
What sort of utility do we want ?





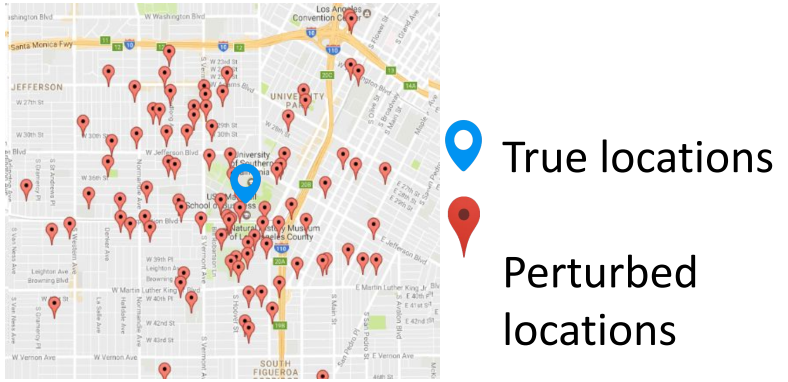
Example:: *What is my nearest gas station*

ε-Geo-Indistinguishability (GeoInd)

Let X, Z be the set of all possible user locations.
A randomized mechanism $K(X)(Z)$
satisfies ε-GeoInd iff for all $x, x', z \subseteq Z$:

$$\frac{K(x)(z)}{K(x')(z)} \leq e^{\varepsilon\, d(x,x')}$$

where ε is the privacy parameter.



$\epsilon = \log(2)$, r=1 km

A GeoInd mechanism should produces similar results when applied to locations that are geographically close.

The uncertainty of the adversary increases as he tries to narrow down your location.

E.g. LA ok, USC not ok.

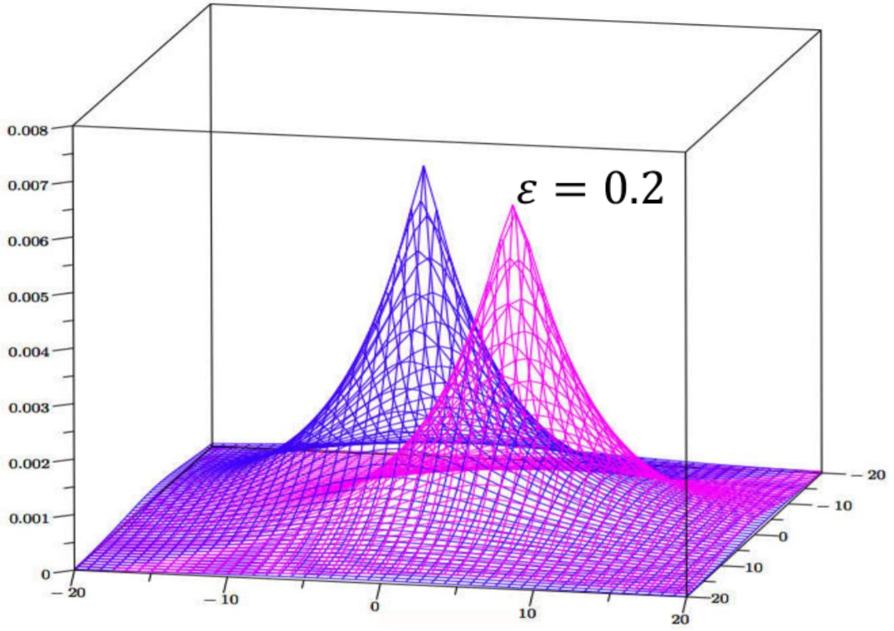
Planar Laplace Mechanism (PL)

The bi-variate pdf of PL noise mechanism is:

$$D(z)(x) = \boxed{\frac{\varepsilon^2}{2\pi}} e^{-\varepsilon \boxed{d(z,x)}}$$

Normalization factor.

Distance to reported location.



Method to obtain GeoInd:

I. Sample a 2D displacement vector $\boldsymbol{v}$ from the pdf.

II. Report $z = x + \boldsymbol{v}$

How to sample ?

Laplace vs. Normal

The Laplace and normal (Gaussian) distributions are both symmetric and centered around a mean, but they have distinct characteristics:

<u>Shape and Tails:</u>

- The normal distribution has a bell-shaped curve, with thinner tails that decrease rapidly. This distribution is defined by its mean and standard deviation and has "lighter" tails, meaning that extreme values are less likely.
- The Laplace distribution has a sharper peak at the mean and heavier tails than the normal distribution. This distribution's tails decrease more slowly, making extreme values more probable compared to the normal distribution.

2. **Mathematical Definition:**

   - For a **normal distribution** with mean μ and standard deviation σ, the probability density function (PDF) is:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

   - For a **Laplace distribution** with mean μ and scale parameter b, the PDF is:

$$f(x) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$$

Here, b controls the spread, similar to σ in the normal distribution, but with a heavier tail decay rate.

Planar Laplace Mechanism (PL) contd.

- Not equivalent to generating the two coordinates independently from a standard (one dimensional) Laplace distribution.
- Correct way to sample:
  - Convert to polar coordinates $D_\epsilon(r,\theta) = \dfrac{\epsilon^2}{2\pi} r\, e^{-\epsilon r}$
  - Determine Angular and Radial Marginals: $D_{\epsilon,R}(r) = \int_0^{2\pi} D_\epsilon(r,\theta)\, d\theta \;=\; \epsilon^2 r\, e^{-\epsilon r}$

    $D_{\epsilon,\Theta}(\theta) = \int_0^{\infty} D_\epsilon(r,\theta)\, dr \;=\; \frac{1}{2\pi}$
  - Draw a point (r, θ), by drawing separately r and θ from
  - $D(r)$ and $D(\theta)$ respectively

Planar Laplace Mechanism (PL) contd.

- The closer (geographically) two points are, the less distinguishable we would like them to be.

- The planar Laplace mechanism offers no optimality guarantees for the quality loss of the reported location

Efficient, BUT poor Utility in practice.

Can you achieve better utility by using some knowledge of user check-in behavior ?

What about protecting trajectories

- Let a trajectory $x = [x_1, \ldots, x_n]$
- We can simply apply a noise mechanism independently to each secret x_i .

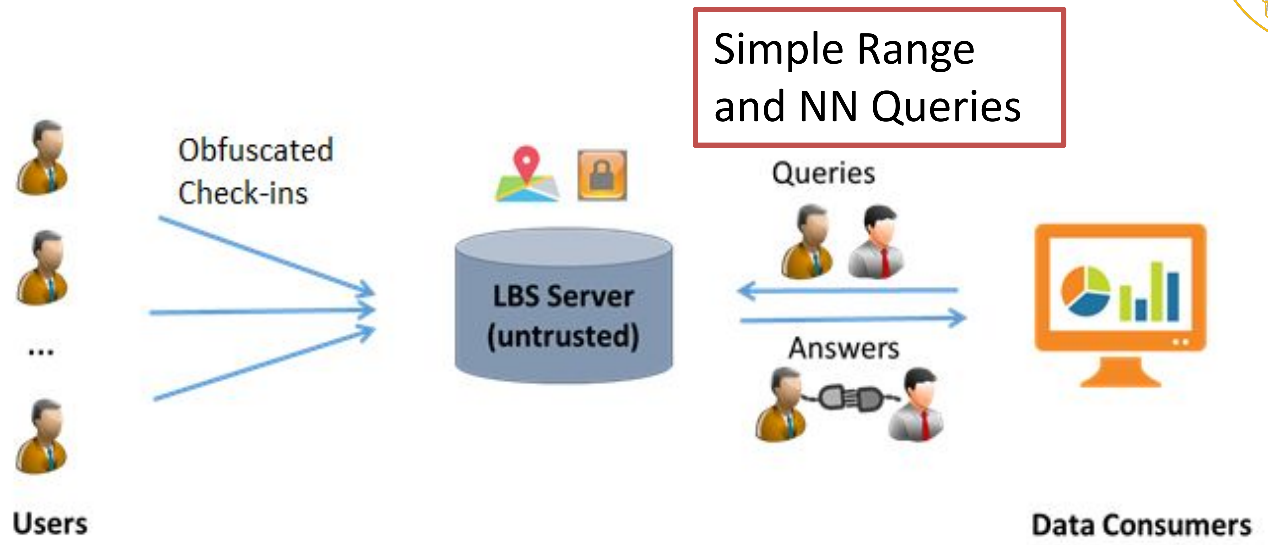
$$
\begin{array}{l}
\textbf{mechanism} \ \ \mathrm{IM}(\mathbf{x}) \\
\quad \mathbf{z} := [\,] \\
\quad \textbf{for} \ \ i := 1 \ \ \textbf{to} \ \ |\mathbf{x}| \\
\quad\quad z := N(\epsilon_N)(\mathbf{x}[i]) \\
\quad\quad \mathbf{z} := z :: \mathbf{z} \\
\quad \textbf{return} \ \ \mathbf{z}
\end{array}
$$

- Total privacy leakage ?
- $n\epsilon$ according to Composition Theorem of DP

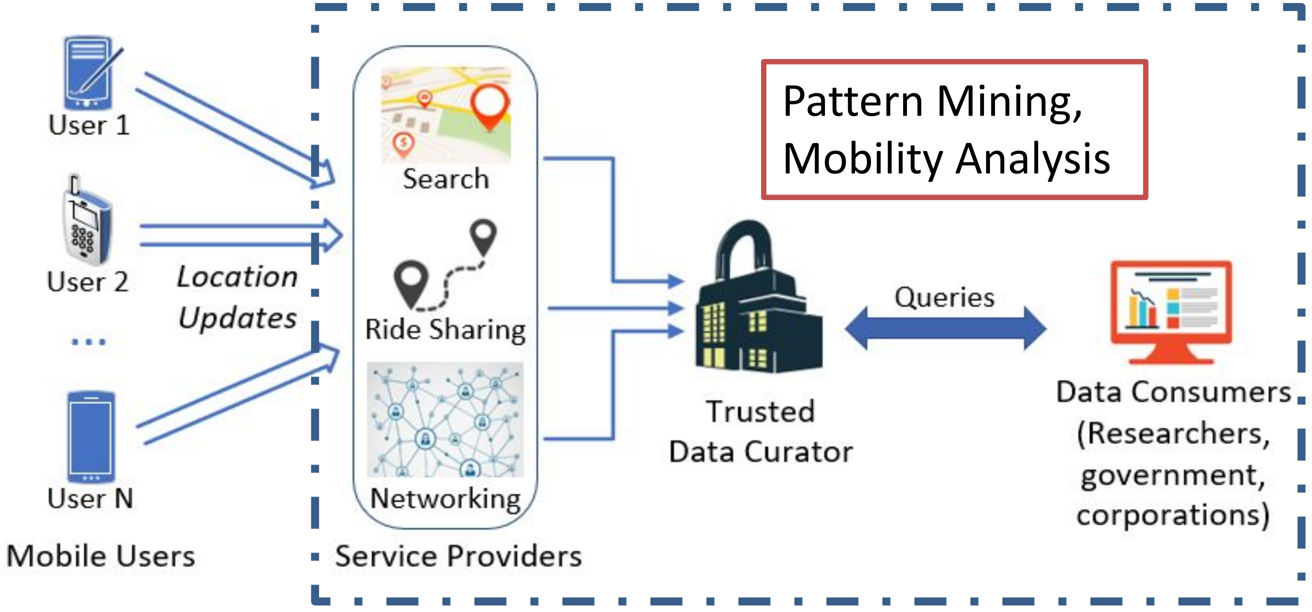
$\boxed{\text{Can we do better ?}}$

System Models

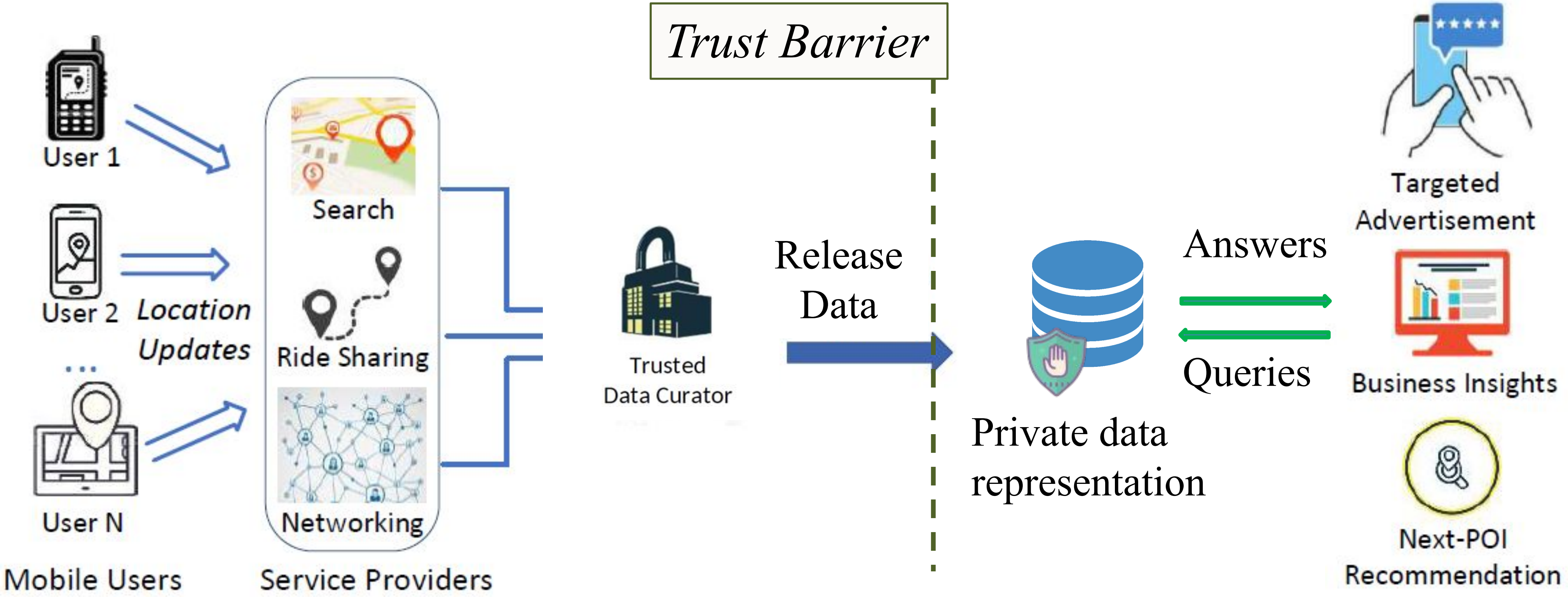
Online Setting:



Offline Setting:



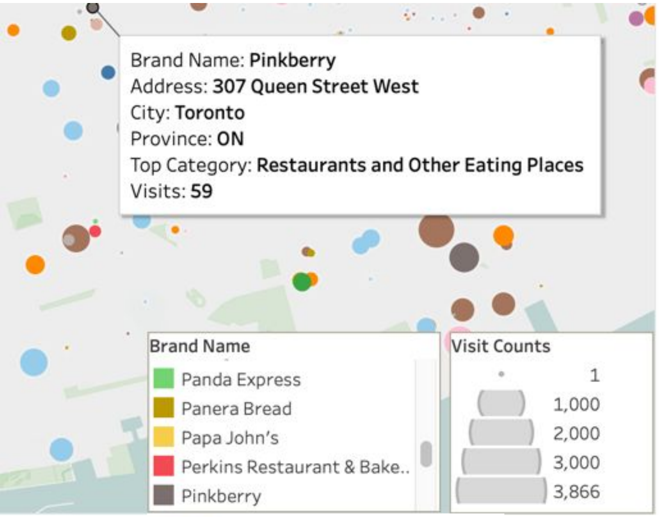
Privacy-Preserving Services Offline Setting (Publishing)



Published data representations must preserve user's privacy.

Privacy-Preserving Release of Aggregate Location Data

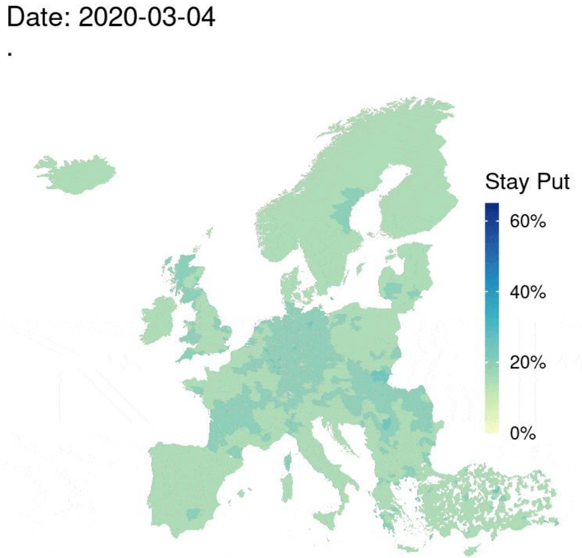
POI Visits Pattern



SAFEGRAPH

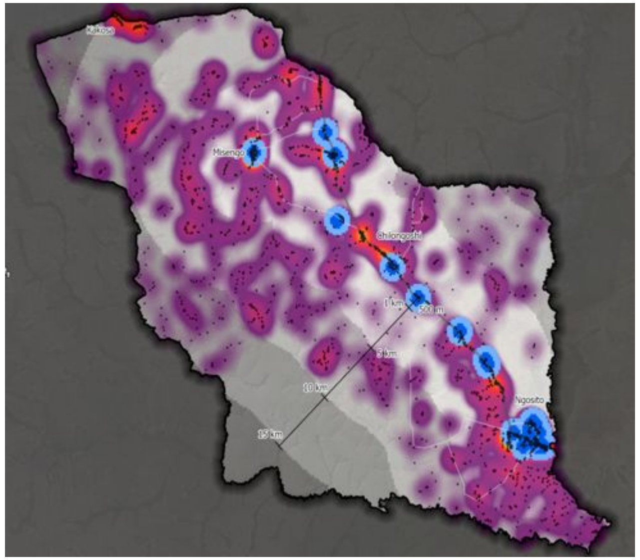
Google Mobility Reports

Europe's COVID-19 response efforts



facebook data for good

World Vision Project
for Clean Water Access



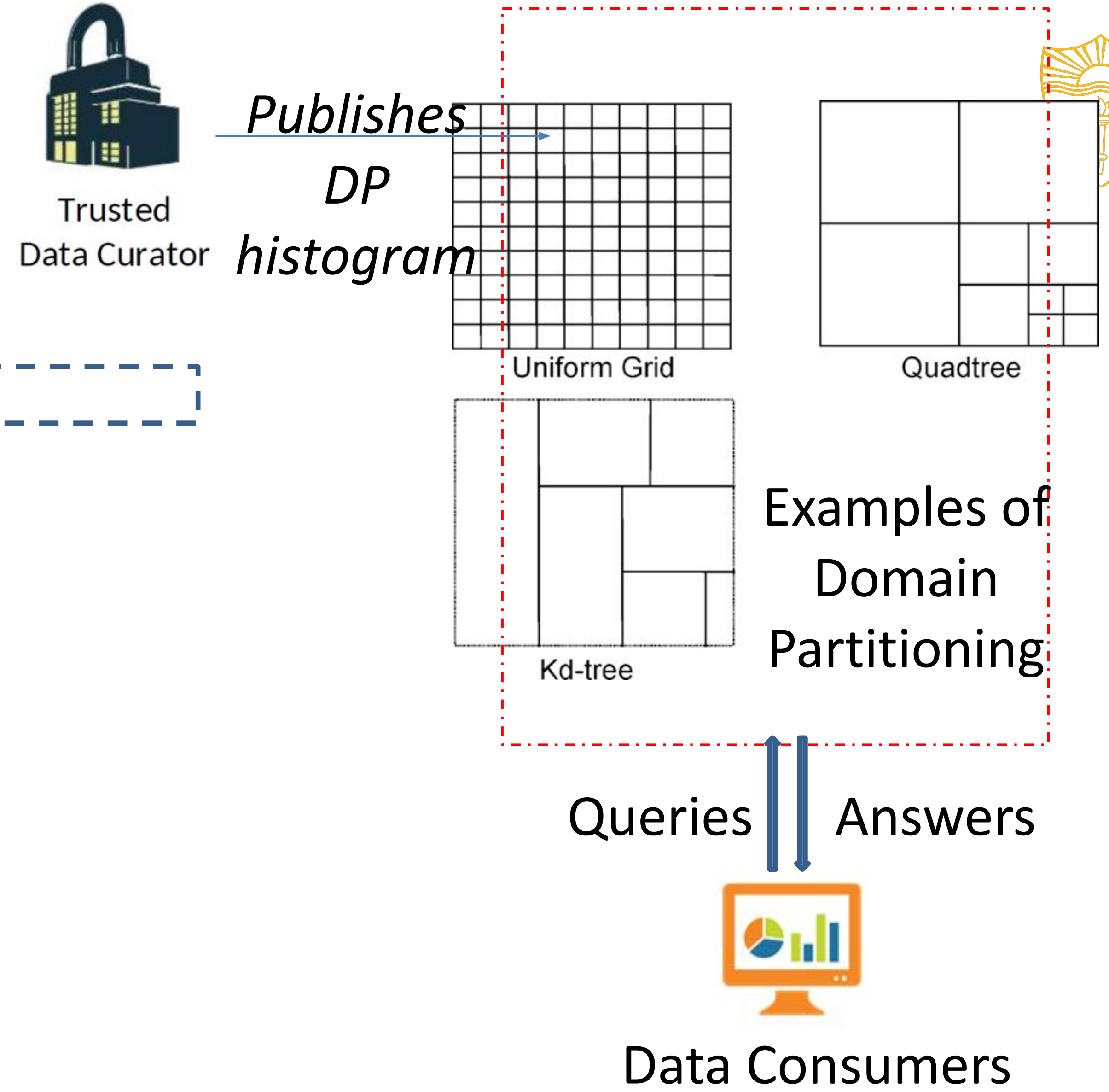
World Vision

United States
Decennial Census

United States Census Bureau

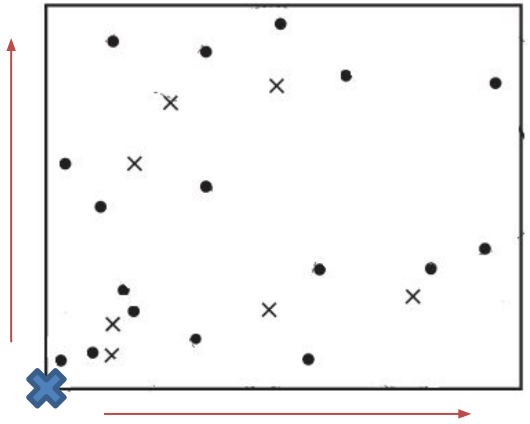
Apportionment
Redistricting
Funding allocation

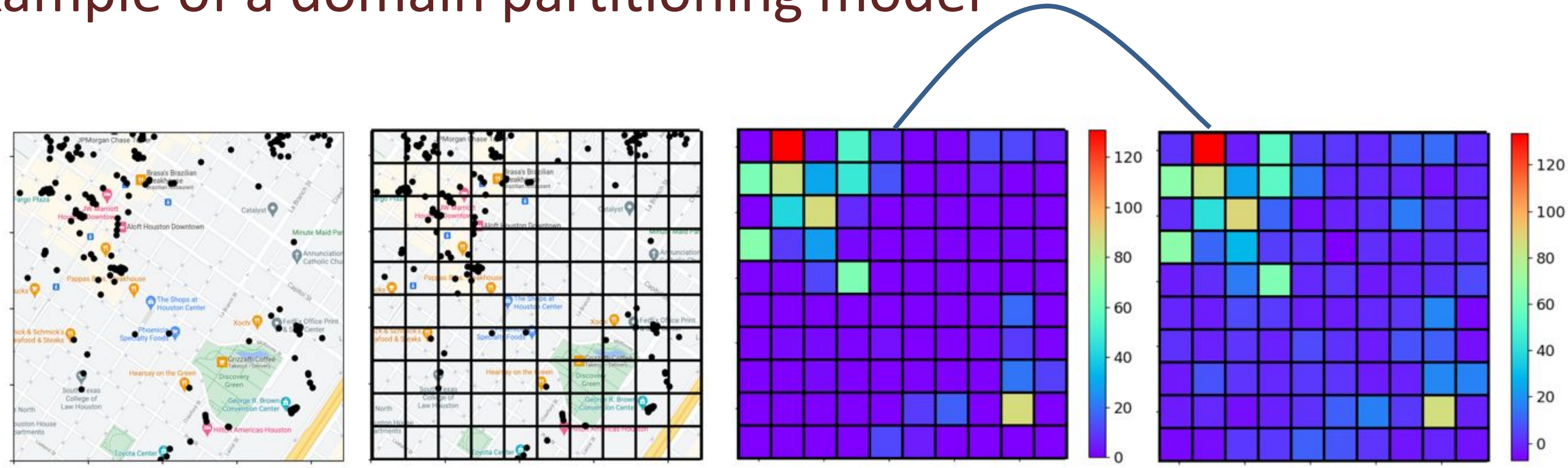
Problem and Related Work
Privately Answering RCQs



Sina Shaham, Gabriel Ghinita, Ritesh Ahuja, John Krumm, Cyrus Shahabi:
HTF: Homogeneous Tree Framework for Differentially-Private Release of
Location Data. SIGSPATIAL/GIS 2021: 184-194

DP location data release
An example of a domain partitioning model





True database

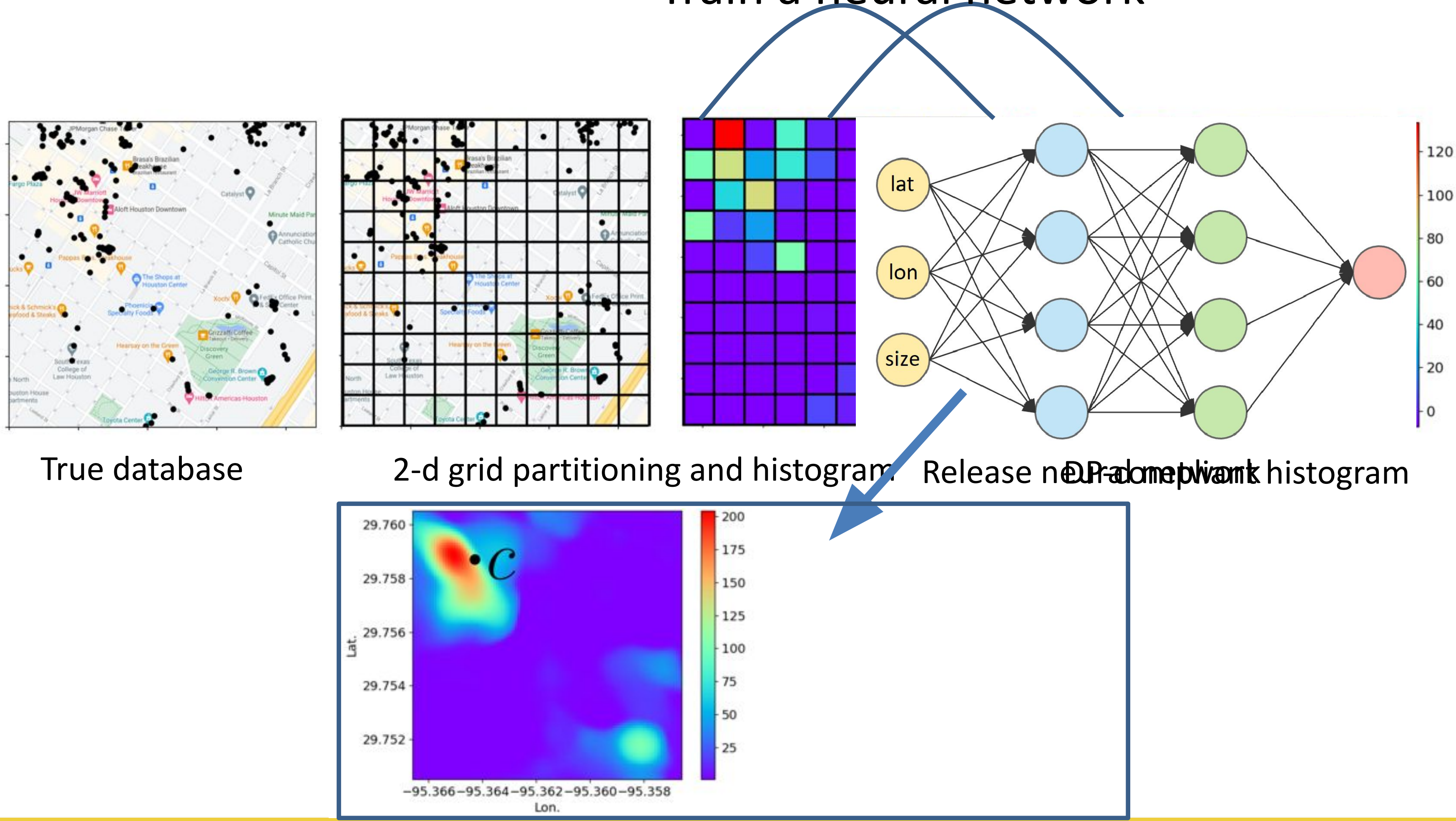
2-d grid partitioning and histogram

DP-compliant release

Once sanitized, *post-processing property* of DP ensures any further computation cannot cause privacy leakages.

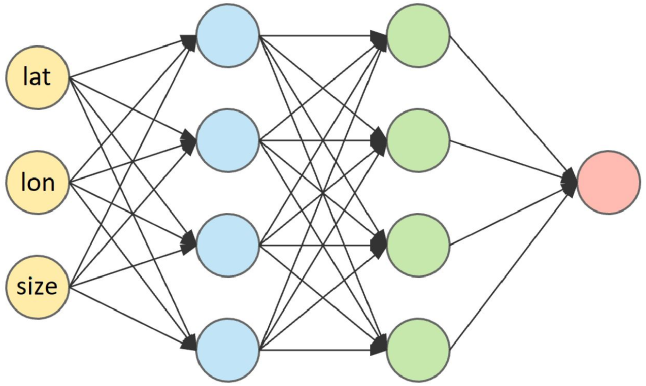
Spatial Neural Histograms (SNH)

Train a neural network



Sepanta Zeighami, Ritesh Ahuja, Gabriel Ghinita, Cyrus Shahabi:
A Neural Database for Differentially Private Spatial Range Queries.
In VLDB 2022

Neural Network Training



Histogram cells

A Neural Database (NeuroSketch):

- Model weights store data
- Answer query with forward pass

Also has applications in non-private setting

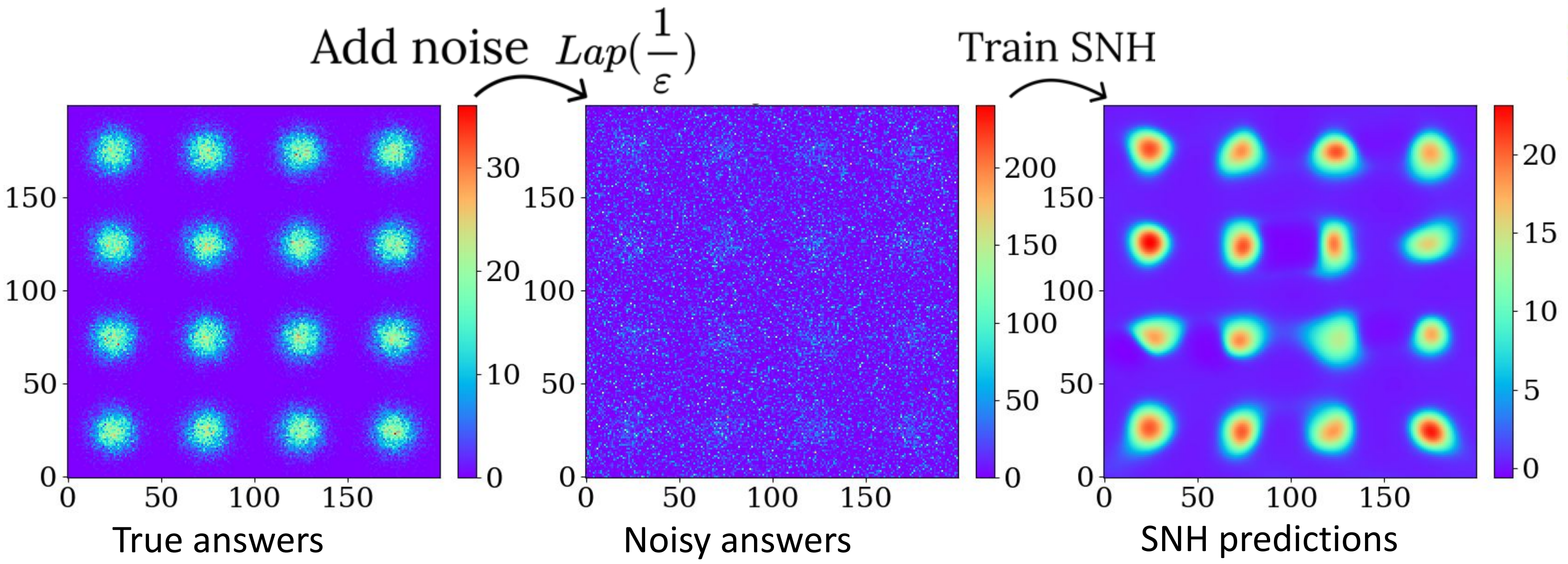
Train with SGD

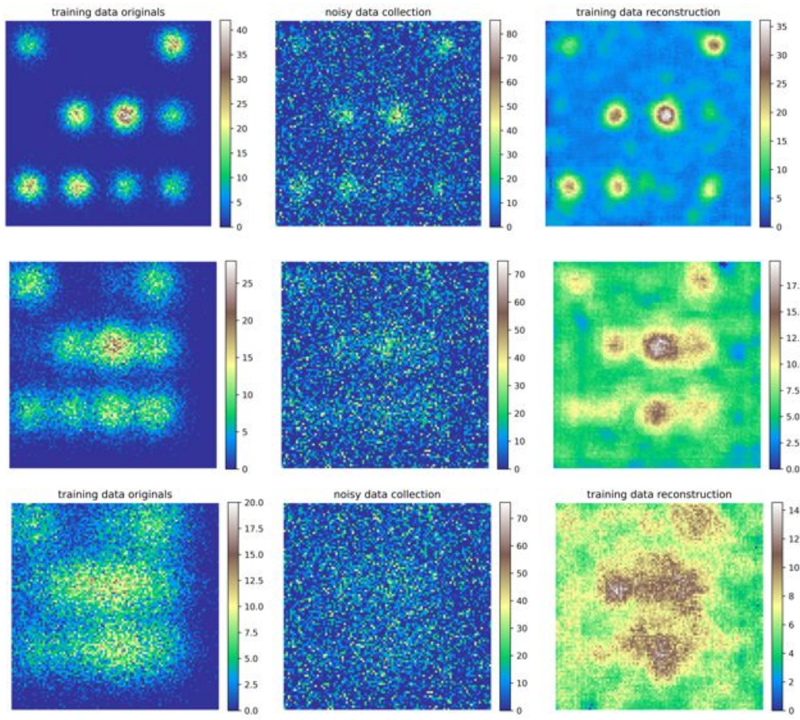
Disclaimer: Actual loss function and training set are more complicated, see paper.

But Why Does It Work?

Neural network fits to the patterns not noise

- Random noise difficult to fit
  - Highly non-smooth
- Neural network learns a smoother underlying function





Experimental Evaluation
Datasets

Veraset (VS)
- Covers 10% of U.S. mobile devices 2019
- 2.5B check-ins from 1.2M devices per day

Gowalla (GW)
- 6.4M records from 200k users
- From Feb 2009 – Oct 2010

San Francisco-CABS
- GPS coordinates of approximately 250 taxis collected over 30 days in San Francisco

SPD-VS
- Veraset dataset with StayPoint Detection algorithm to retrieve POI visits of users.

Low Pop. density	Medium Pop. density	High Pop. density
Fargo [46.877, -96.789]	Phoenix [33.448 -112.073]	**Miami** [25.801, -80.256]
Kansas City [39.09, -94.59]	Los Angeles [34.02, -118.29]	Chicago [41.880, -87.70]
Salt Lake [40.73, -111.926]	Houston [29.747, -95.365]	SF [37.764, -122.43]
Tulsa [36.153, -95.992]	$\boxed{\textbf{Milwaukee}\ [43.038, -87.910]}$	Boston [42.360 -71.058]

Default city

Wide range of location datasets, with application scenarios ranging for location networks, POI visitations, taxis, etc.

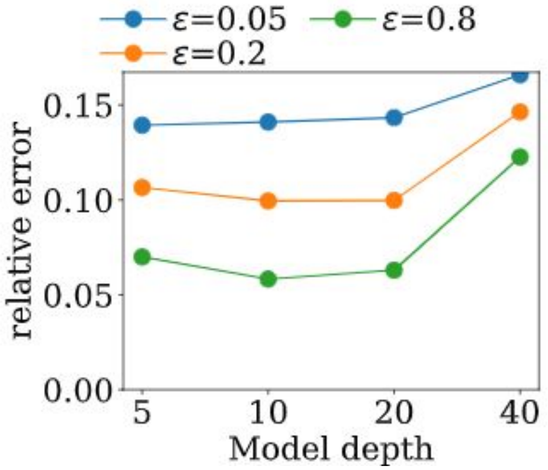
Experimental Evaluation
Parameters

Query Specification

- 5000 RCQs centered at uniformly random positions, size = [25 m to 200 m].
- Metric: relative error, with smoothing factor ψ = 0.1% of

Workload Queries

- 2000 RCQ more sampled from same distribution.

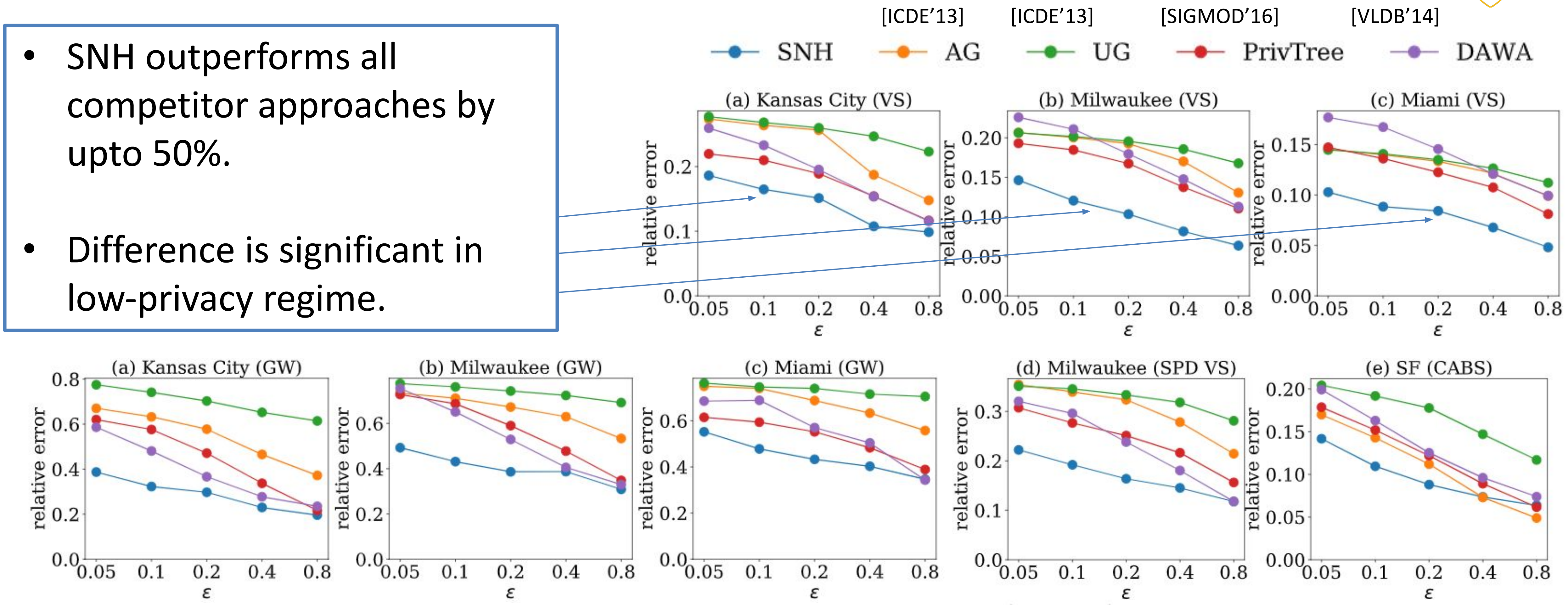


SNH model specification

- Fully connected neural networks is set to 20 layers of 80 units each

- SNH outperforms all competitor approaches by upto 50%.

- Difference is significant in low-privacy regime.



Conclusion

Aggregate Queries (AQ) are the most widely used query primitive and DP-compliant answering is crucial.

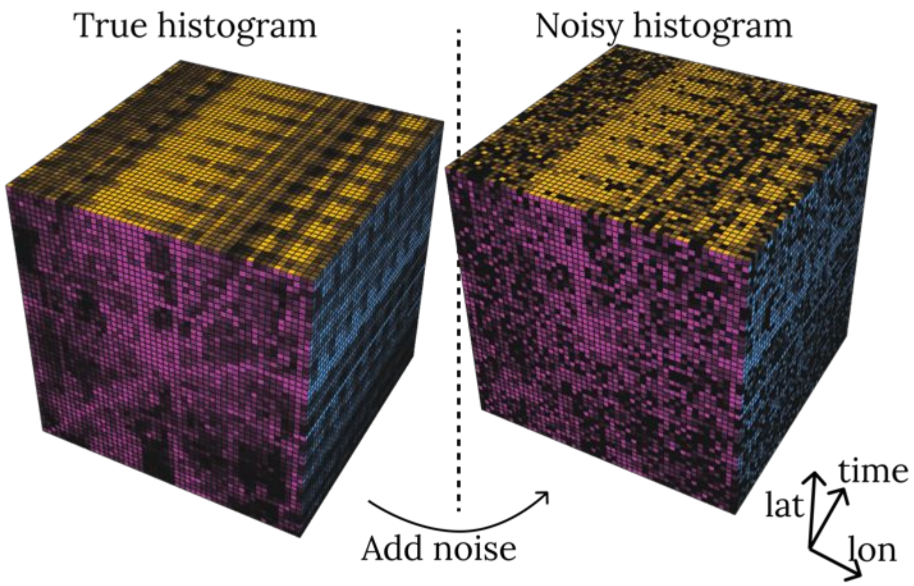
First work that leverages neural networks to answer AQs. It addresses shortcomings of existing methods by learning patterns from location and time datasets and de-noising local information.

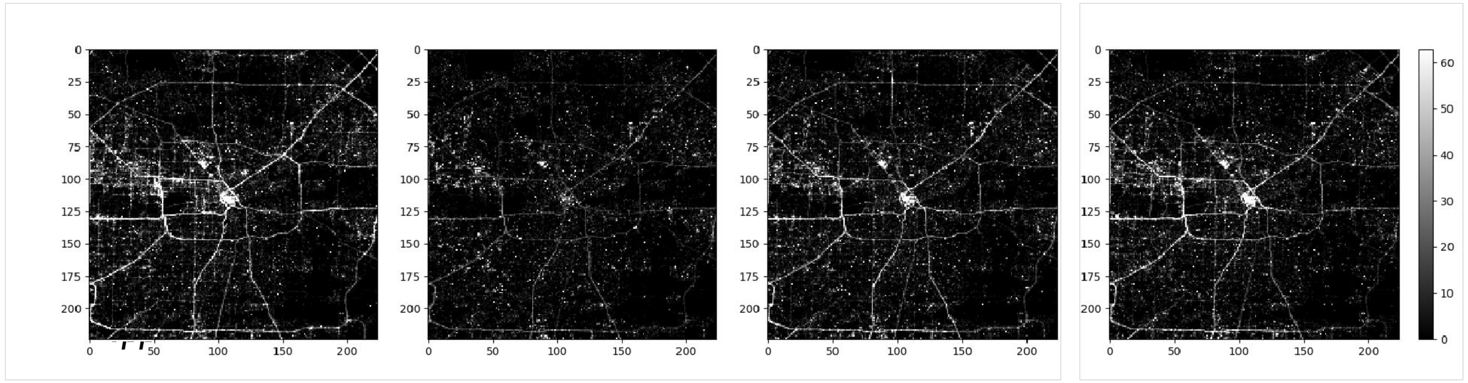
Paramselect, a private parameter tuning model specifies how to avoid using privacy budget to learn system parameters.

Spatio-Temporal Data Release

- Release a differentially private 3-dimensional histogram

User_id	Latitude	Longitude	Timestamp
John	37.7920	-122.3927	10/11 20:32
Kyle	37.7930	-122.3827	10/11 20:33
John	37.7936	-122.3224	10/11 21:45
…	…		…
John	37.7143	-122.3687	10/11 23:50



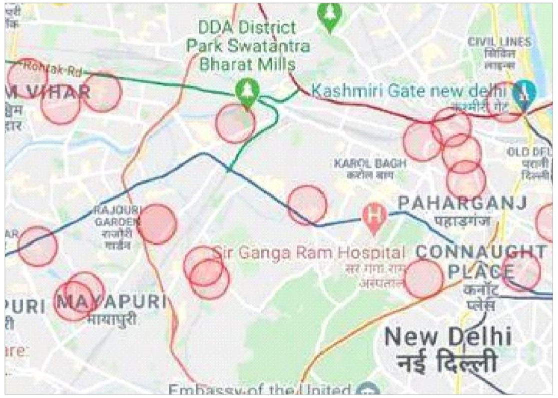


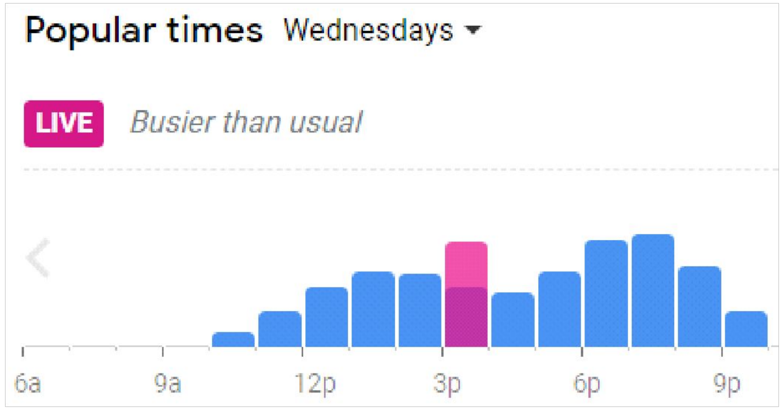
Sepanta Zeighami, Ritesh Ahuja, Gabriel Ghinita, Cyrus Shahabi:
A Neural Approach to Spatio-Temporal Data Release with User-Level Differential Privacy, In SIGMOD 2023

Variational Autoencoder-Based Density Release (VDR)

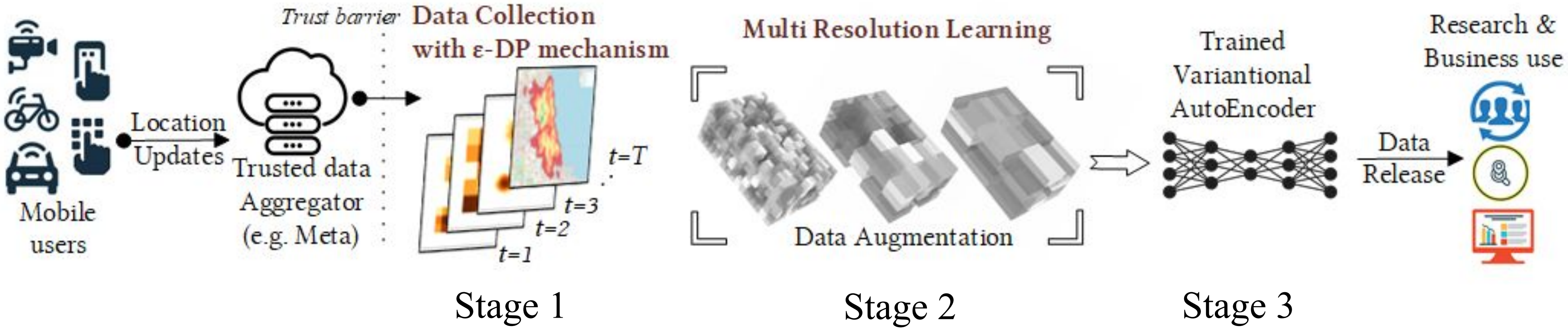
- Allows arbitrary query types, e.g., Range Count Queries at time instances and more:

Hotspot discovery

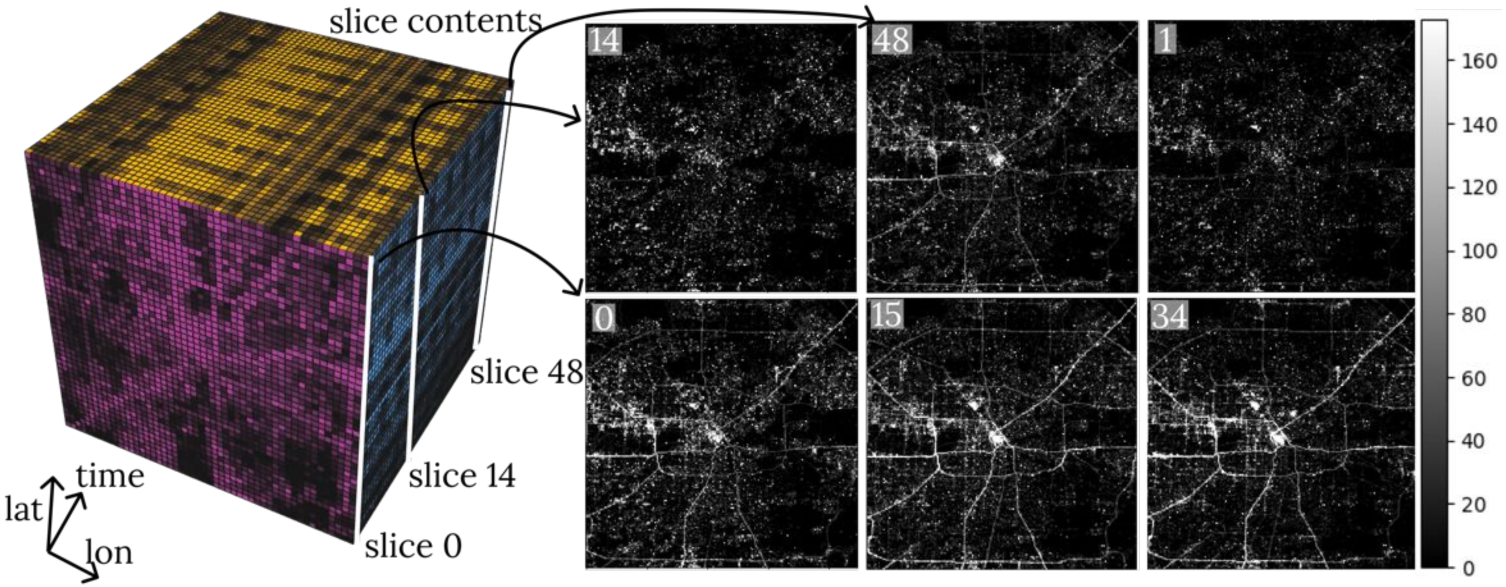


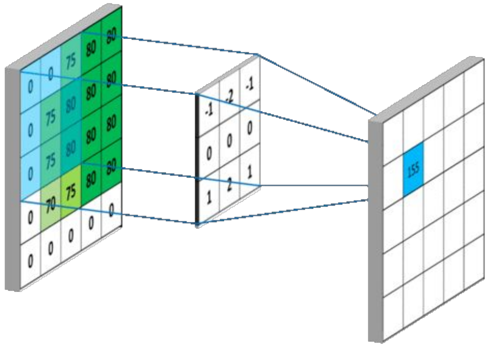


POI visits forecasting



Stage 3: Learned Denoising
Spatial Patterns as Visual patterns





Utilize CNNs to learn spatial patterns.

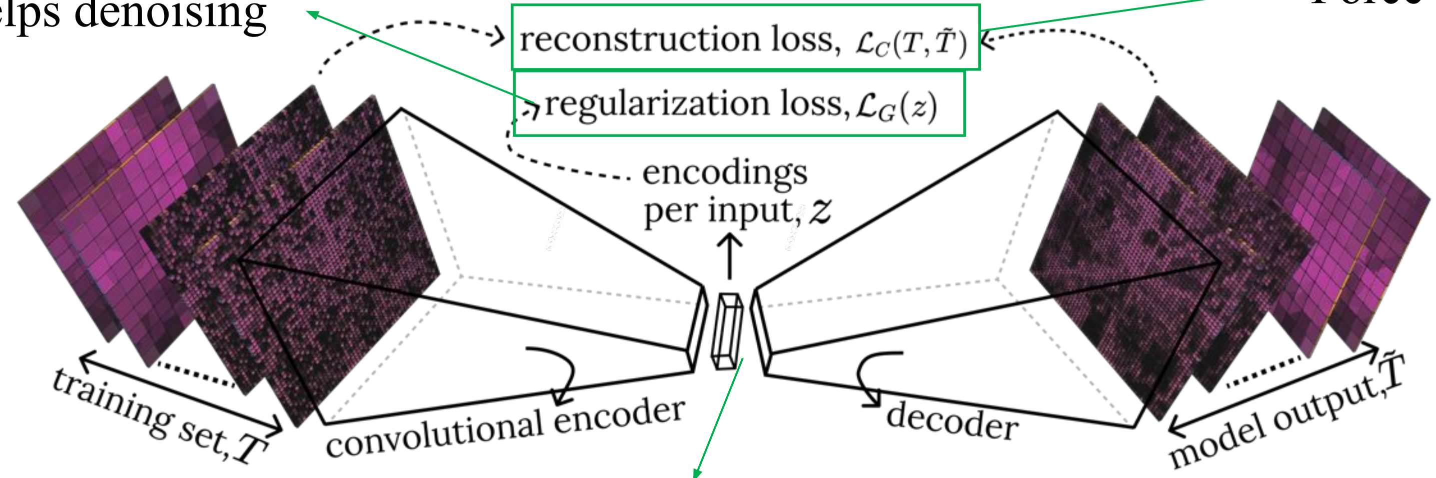
Spatio-temporal location data can be viewed as a series of images.
We utilize lessons from image feature extraction literature.

Stage 3 : Learned Denoising

Final histogram released publicly

Further helps denoising

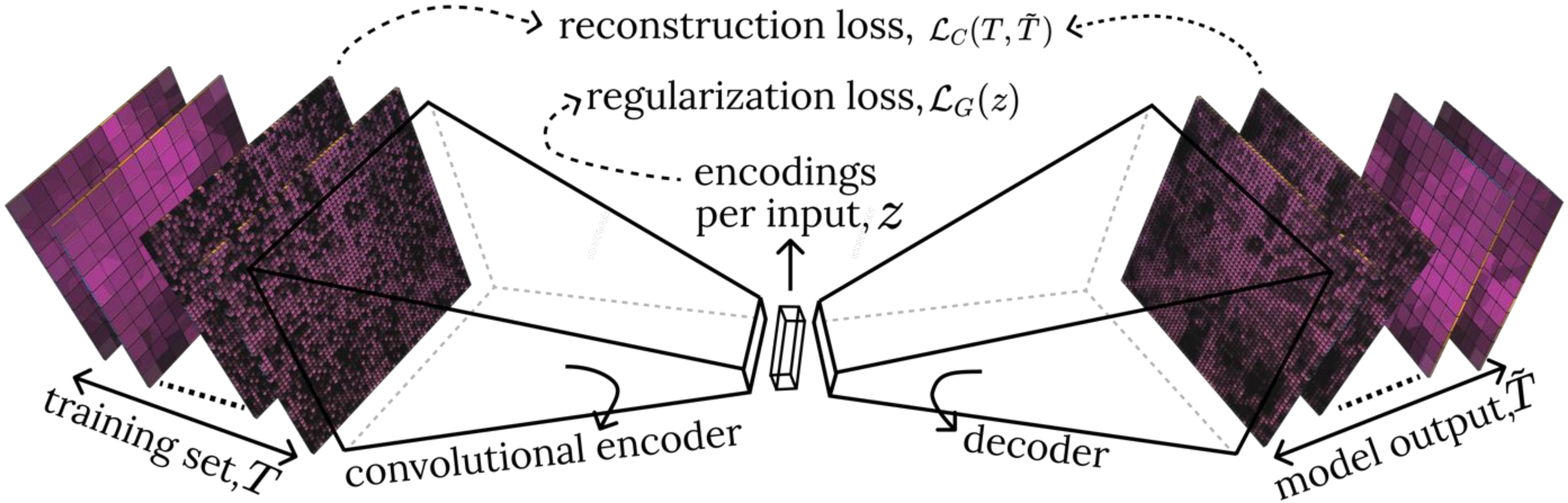
Force accurate reconstruction



Need to learn repeatable patterns to maximize accuracy

VDR Features

- Does not introduce bias from complex domain partitioning
- Exploit spatial patterns to reduce variance (i.e., denoise) by learning a VAE
- Explicitly account for user-level privacy (compared with event-level privacy)



Q&A

Thanks!

References

- [Bordenabe et. al., CCS 2014] Bordenabe et. al. “Optimal Geo-Indistinguishable Mechanisms for Location Privacy”. CCS 2014
- [Andres et. al., CCS 2013] Andres et. al. “Geo-indistinguishability: differential privacy for location-based systems” CCS 2013
- Sepanta Zeighami, Ritesh Ahuja, Gabriel Ghinita, Cyrus Shahabi: A Neural Database for Differentially Private Spatial Range Queries. In VLDB 2022
- Sepanta Zeighami, Ritesh Ahuja, Gabriel Ghinita, Cyrus Shahabi: A Neural Approach to Spatio-Temporal Data Release with User-Level Differential Privacy, In SIGMOD 2023

- Some slides borrowed from ICDM08 tutorial by Mohamed F. Mokbel