COMP90018
Location-based Services & Location Privacy

Anthony Quattrone

Location-based Services

Introduction to LBS

- **Location-based services (LBS)**
  - Services that integrate a mobile's device location with other information
  - Available at least since the 1970s (GPS: US military)
- **Mobile operators**
  - Voice, data (SMS, MMS, Video), location information
- **Push versus pull**
  - User receives information without an active request
  - User actively pulls information from the network

Applications I

- **Infotainment services**
  - Driving directions
  - Where is a point of interest (POI): ATM, hotel, restaurant, …
  - Where am I? Location on a map
  - Where are my friends?
- **Tracking services**
  - Fleet management, taxi monitoring and dispatching
  - Children, elderly people, sick persons
  - Goods and package tracking

Applications II

- **Information dissemination services**
  - Content delivery wrt. the user's context & profile
  - Advertisement & e-coupons
  - Hazard warnings
- **Emergency support systems**
  - Police, ambulance, fire brigades
  - Roadside assistance
- **Location-sensitive billing**
  - Call billing based on vicinity from base/home
  - Toll payment

AT&Ts Find Friend

- **User interface**
  - User can add friends to a list
  - Send a SMS with a tracking request
- **Privacy**
  - “Invisible” mode allows to hide from all users
  - Tracking requests always generated an alert
- **Services**
  - Driving directions or meeting point between friends
  - Search for all friends

Location-based Services

- **1st Generation**
  - Manual user input of location information
  - Driving directions, nearby POIs, weather information
- **2nd Generation**
  - Location information is acquired automatically within a couple of kilometers
  - Similar services as in 1st generation
- **3rd Generation**
  - High position accuracy & automatic initiation of services
  - Asset tracking, street-level routing and positioning, …

LBS and their required Accuracy

- **High accuracy**
  - Asset tracking
  - Directions
  - Emergency
- **Medium to high accuracy**
  - Advertising
  - Car navigation
  - POI
- **Low accuracy**
  - Fleet management
  - News
  - Traffic Information

Location Engine

- **(Reverse) Geocoding**
  - Translate street address to latitude & longitude and vice versa
  - Difficult if not complete information available
- **Routing & navigation**
  - Compute best route: A*, Dijkstra, …
  - Best: shortest, fastest, simplest, …
- **Proximity search**
  - Spatial DBs: POIs such as ATMs, hotels, gas stations, …

The Importance of Location Privacy

Location-based Services

- **Nearest neighbor queries**
  - Heart patient continuously monitors the closest hospitals
- **Monitoring for traffic applications**
  - Immediate warnings about oil spills or icy patches
- **Location-aware social networking**
  - Finding friends (Foursquare, Google Latitude, Gowalla, Loopt, …)
- **Location-based advertising**
  - Send 20% off coupons to all mobile devices close to a store (range query)

Privacy Concerns …

- **IEEE Spectrum, July 2003**
  - "They know where you are: new technologies can pinpoint your location at any time and place.
    They promise safety and convenience –
    but threaten privacy and security."
- **Lack in privacy-aware systems**
  - Might inhibit the growth of location-based services

Location Privacy

- **Status quo of current mobile systems**
  - Able to *continuously* monitor, communicate, and process information about a person's location
  - Have a high degree of spatial and temporal precision and accuracy
  - Might be linked with other data
- **Important research issue**
  - Techniques for protecting location privacy are required

The Importance of Location Privacy

- **Location-based spam**
  - Unsolicited advertising
- **Personal safety**
  - Stalking
  - Assault
- **Intrusive inferences**
  - Person's political views
  - Individual preferences
  - Health conditions

Sightings of Celebrities in Real-Time

http://gawker.com/stalker (not active anymore)

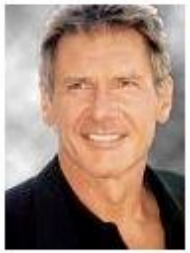
Harrison Ford
AVENUE OF THE AMERICAS & W 44TH ST
Jun 23rd, 2009 @ 4pm
Shooting a scene with Rachel McAdams for the new movie “Morning Glory/"

2

Common
W 14TH ST & 7TH AVE
Jun 23rd, 2009 @ 3pm
Jumping out of a cab on his cell in a building. Passer by exclaimed, "Oh shit!" and he acknowledged with a "hey".

Lance Bass
BROADWAY & CENTRAL PARK S
Jun 23rd, 2009 @ 12pm
I just sat across from lance bass on the 3 train going uptown, we both got off at columbus circle. he was with another attractive man being very affectionate. and smily.

4



Harrison Ford
5TH AVE & W 53RD ST
Jun 22nd, 2009 @ 5pm
Harrison Ford, filming right now. Large crowd.

Tracking of Individuals

- **Deutsche Telekom (telecommunication operator)**
  - Deutsche Telekom handed over six months of Malte Spitz's phone data
  - Tracked position, phone calls, SMS, Internet access
  - http://www.zeit.de/datenschutz/malte-spitz-data-retention

Do people care about location privacy?

Study about Rendezvousing

- **Study of Colbert, 2001**
  - Friend finding:
    - Asked participants to estimate the number of people with whom they would have long-standing agreements to obtain position information or permit their position to be obtained
  - Participants: 17 male and 17 female students
- **Results**
  - Many students had at least 10 individuals on their lists!
  - Colbert questioned their accuracy
  - Participants might have different or unstated expectations about the possible, future use of this service

User Needs for Location-Aware Mobile Services

- **Study of Kaasinen, 2001**
  - Participants: 55 people from different backgrounds
  - Group interviews
- **Results**
  - Worried about their privacy and evoked ‘‘big brother’’ association for services that locate people
  - But: not worried about their privacy using location-aware services
  - Most people did not realize that they can be located using location-aware services
  - Trust in regulations for mobile phone operators

How much is Location Privacy Worth?

- **Study of Danezis et al., 2005**
  - Collection of precise location information (via a mobile phone) for a (fictional) study running for one month
  - Competition auction: participants state their required monetary compensation
  - Participants: 74 undergraduate computer science students
- **Results**
  - Median price: £10; £20 pounds if data is used commercially
  - Two people lost interest if the data is commercially used
  - 9 people asked for more than £30

Gathering GPS Data from Drivers

- **Krumm's Study, 2007**
  - Convinced more than 250 people from Microsoft to provide recorded GPS data for two weeks
  - Incentive: a 1 in 100 chance of winning a $200 MP3 player
  - 97 of them were asked if the data could be used outside Microsoft and only 20% denied this

Privacy Concerns: Presence Services

Survey by Harris Interactive, 2006: 1028 adults, online

Do you believe any of these types of services are an invasion of privacy?

Presence service	Yes
The ability to look at your contact list and determine if they were available to talk, busy on a call or unavailable.	34%
The ability for friends and family to see this information about you.	59%
The ability to determine the location of persons on your contact list (snapshot of where they are now).	70%
The ability to determine what locations individuals on your contact list were over the last few hours (map of their whereabouts).	73%
The ability for your employer to see this information about you.	83%
None of these	10%

Source: http://www.harrisinteractive.com/news/allnewsbydate.asp?NewsID=1184

Microsoft in 2011 (Data Privacy Day)

Awareness of Location Based Services (LBS)

62% of people say they are aware of LBS.

Use of Location Based Services

51% report having ever used LBS.

31% use LBS less than once a month. One in 20 use these services every day, rising to one in 10 among those in the U.S.

Only 18% of respondents have used LBS to tell others where they were or connect with individuals.

Location Based Services: Value Perception

Of those who use LBS, 94% considered them valuable.

"Practical" services like GPS, weather alerts, traffic updates and finding restaurants are most commonly used.

People Are Concerned About Privacy and Want Control

52% of respondents expressed concern with sharing their location with other people or organizations.

49% would be more comfortable with LBS if they could easily and clearly manage who sees their location information.

A clear majority of respondents are concerned about sharing their location without consent (84%), having personal information or identity stolen (84%), and suffering loss of privacy (83%).

Responses Vary by Location
Awareness and use of location-based services was similar across Canada, Germany, the U.K. and U.S. However, twice as many respondents in Japan are familiar with LBS. Japanese respondents have the highest usage and fewest privacy concerns.

Japanese respondents (42%) were less concerned than the overall population (52%) with allowing people to view their location.

Research from Location-based Services Usage & Perceptions Survey conducted by Cross-Tab Marketing Services for Microsoft Corporation. Interviews were conducted online between Dec. 13-Dec. 22, 2010, among 1,500 respondents across the Canada, Germany, Japan, the U.K and U.S.

Techniques for Location Privacy

Stealth

- **Stealth**
  - Ability to be at a location without anyone knowing you are there
- **How**
  - Use of passive devices such as GPS
- **Disadvantages**
  - Active devices such as mobile phones cannot preserve stealth
  - Access of information overrides stealth

Anonymity-Based Approaches

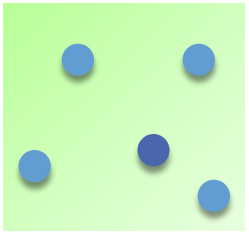
- **Idea**
  - Separate location information from an individual's identity
  - Special type is *pseudonymity*: an individual is anonymous but has a persistent identity (a pseudonym)
- **Disadvantages**
  - Vulnerability to data mining: a person's identity can be inferred from his or her location over time, in particular from homes or offices
  - A barrier to authentication and personalization
  - Efficacy in sparsely populated areas?

K-Anonymity

- **Sweeney (2000)**
  - Analysis of microdata
  - Anonymized health records: DOB, Gender, ZIP code
  - Uniquely identified medical records of the governor of Massachusetts => DOB, Gender, ZIP code are *quasi identifiers*
  - ~87% of population can be identified!
- **Databases: k-anonymity (Sweeney 2002)**
  - A table is k-anonymous if every record in the table is indistinguishable from at least k − 1 other records with respect to quasi identifiers

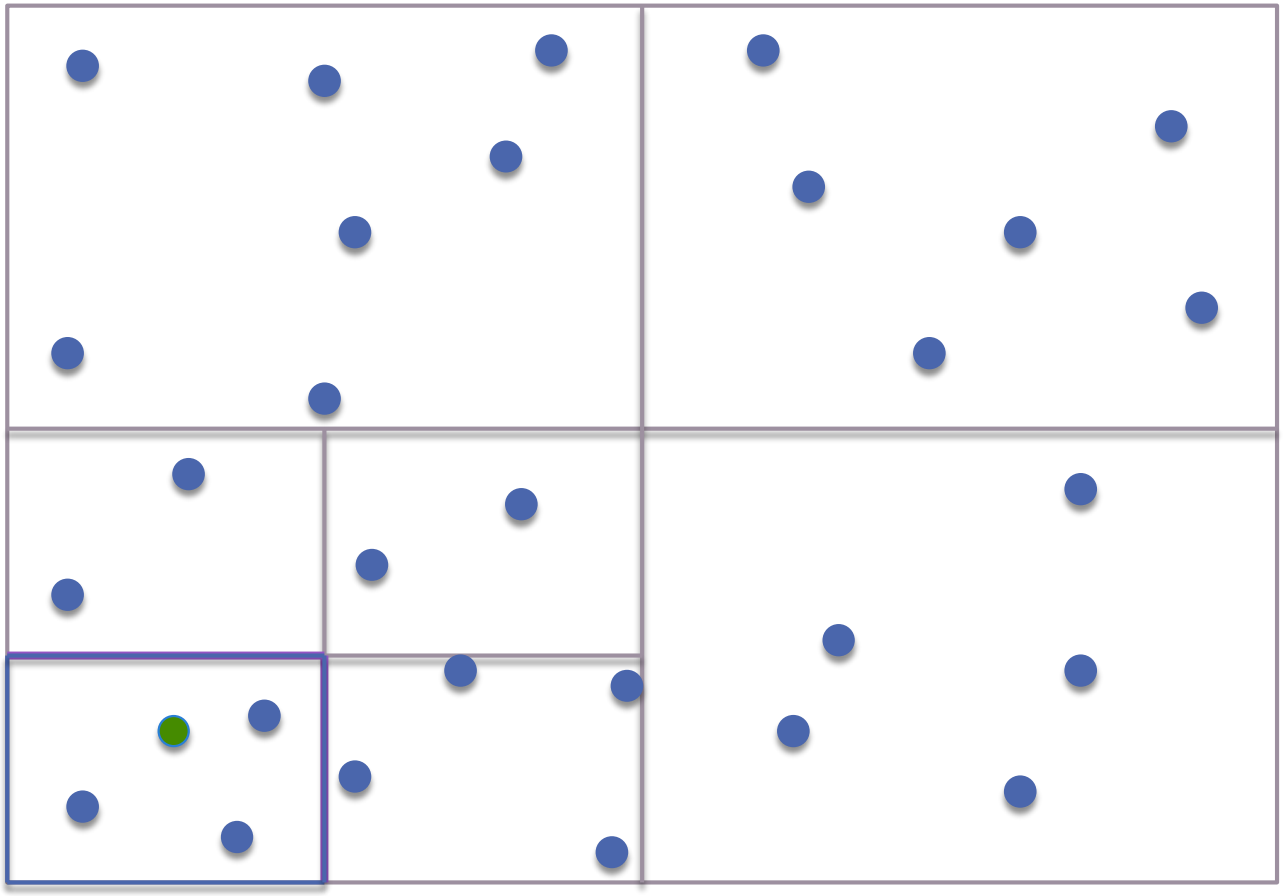
Anonymity: Cloaking

- **Spatial cloaking**
  - Gruteser & Grunwald use quadtrees
  - Adapt the spatial precision of location information about a person according to the number of other people in the same quadrant
- ***k*-anonymity**
  - Individuals are *k*-anonymous if their location information cannot be distinguished from *k*−1 other individuals



- **Temporal cloaking**
  - Reduce the frequency of temporal information

Spatial Cloaking ($k_{min} = 4$)



Basic Model



Wireless Network

Location

Location and query

Location Server (Cloaking algorithms)

Response

Anonymous data:
location,
time of query

Location Based Service Provider

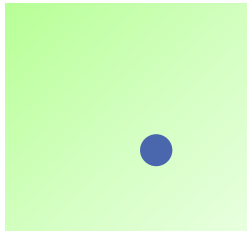
Location Privacy through Obfuscation

Workshop on Location-aware Computing

- *I would be happy if a third party who I sometimes do business with knew my location on a Saturday afternoon's shopping trip to an accuracy of (pick all that apply):*
  - None at all: 96%
  - Existence but no accuracy: 43%
  - **Country: 35%**
  - **City: 35%**
  - **Street: 17%**
  - **10m: 13%**
  - 1m: 4%
  - 1cm: 4%
- People are more prepared to reveal their location to some degree the less precise the location is

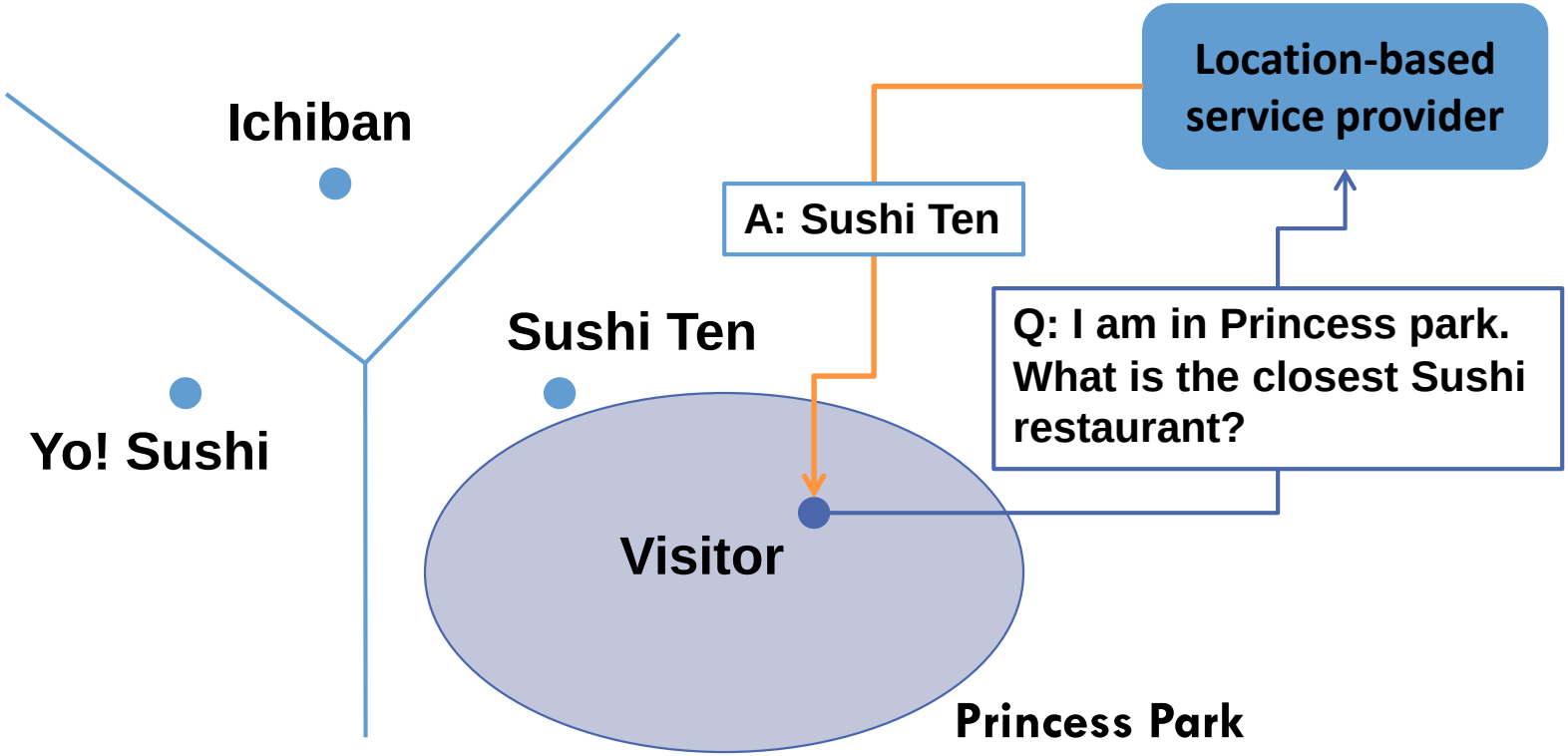
Obfuscation

- **Idea**
  - Mask an individual's precision
  - Deliberately degrade the quality of information about an individual’s location (imperfect information)
  - Identity can be revealed
- **Assumption**
  - Spatial imperfection ≈ privacy
  - The greater the imperfect knowledge about a user’s location, the greater the user’s privacy



Motivation for Obfuscation

- Finding the closest Sushi restaurant

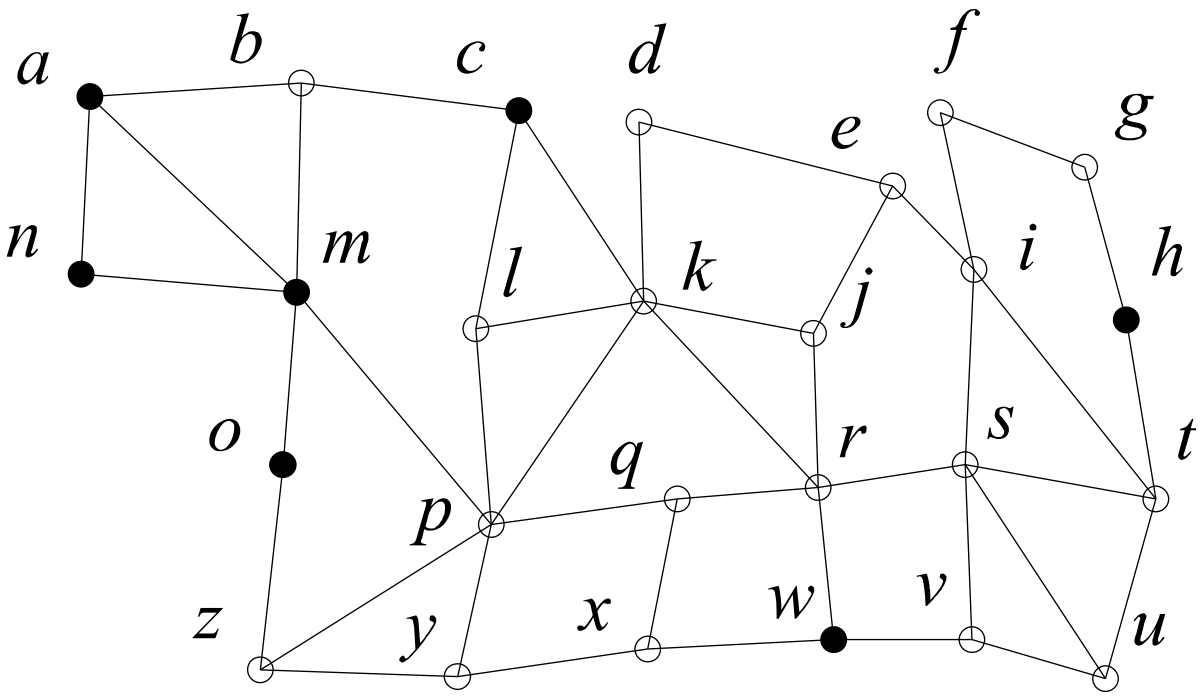


Strategies for Imperfection

- **Types of imperfection**
  - Accurate and precise: $l \in O$ and $|O| = 1$
  - Inaccurate and precise: $l \notin O$ and $|O| = 1$
  - Accurate and imprecise: $l \in O$ and $|O| > 1$
  - Inaccurate and imprecise: $l \notin O$ and $|O| > 1$
- **Consequences**
  - The larger O, the less information is revealed about the true location l
    - The greater the level of privacy
  - The greater the distance between O and the true location l
    - The greater the level of privacy

Example Obfuscation Set

- **Obfuscation set**
  - O = {*a*,*c*,*h*,*m*,*n*,*o*,*w*} does not need to be contiguous
  - An individual is located at one of the vertices in O



Obfuscation Algorithm

- **Input of the Algorithm**
  - Weighted graph $G = (V, E)$ representing the geographic environment
  - Obfuscation set $O \subseteq V$ of the client's location $l \in O$
  - Points of interest $Q \subseteq V$
- **Output of the Algorithm**
  - Pair $\langle q, C \rangle$, $C \in (0, 1]$ is the confidence that $q \in Q$ is the nearest target to the client's current location

Prerequisites

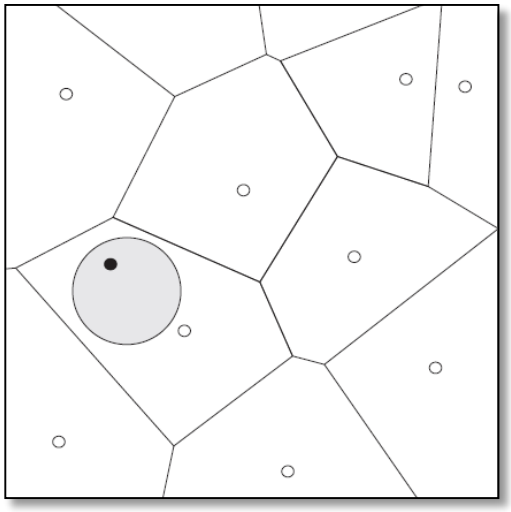
- **Definitions**
  - $d(v_1, v_2)$ is the shortest path distance of $v_1, v_2 \in V$ in G
  - Nearest point of interest:
    - NP : $O \rightarrow Q, o \mapsto q$
      such that $\forall q' \in Q\ [d(o, q) \leq d(o, q')]$
      (assumption: $d(o, q_1) = d(o, q_2) \Rightarrow q_1 = q_2$)
- **Equivalence relation**
  - $\delta \subseteq O \times O$: $o_1 \,\delta\, o_2 \quad \Leftrightarrow \quad \mathrm{NP}(o_1) = \mathrm{NP}(o_2)$

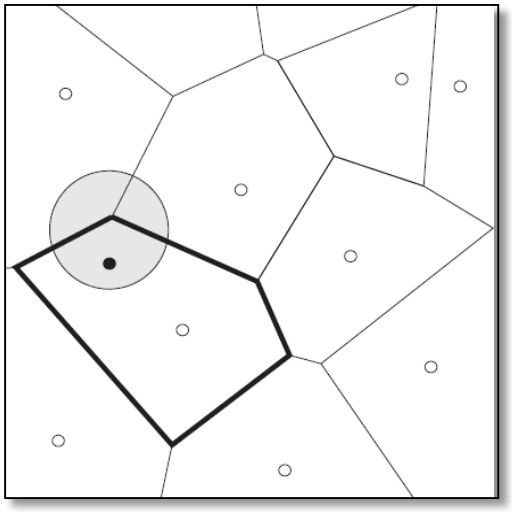
Negotiation Algorithm (NN Queries)

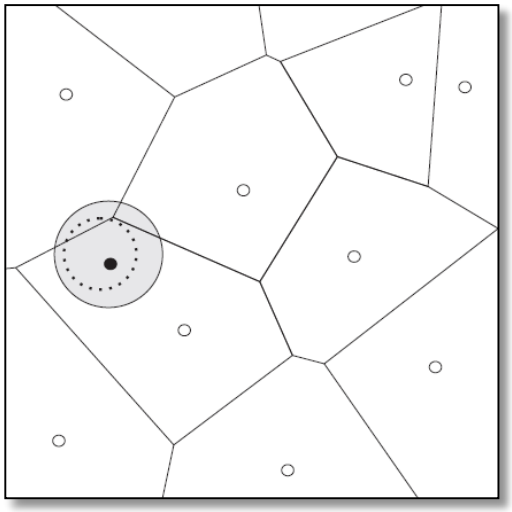
Data: $G = (V,E)$, O, Q
Result: $\langle q, C \rangle$

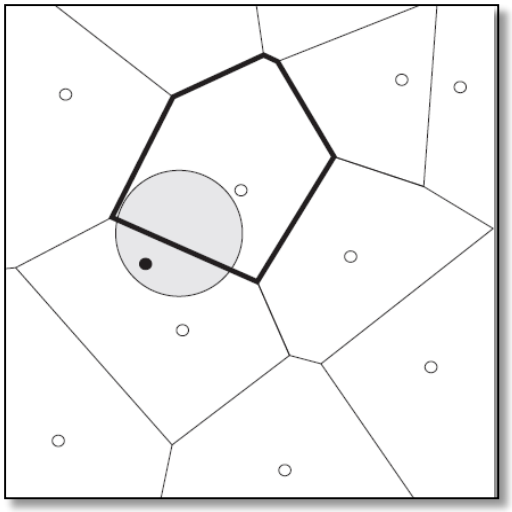
Construct the partition O/δ;
if $O \in O/\delta$ **then**
  Return $\langle q, 1.0 \rangle$ with $q = \text{NP}(o)$ for an arbitrary $o \in O$;
else
  if *Client identifies current location l as equivalence class* $[l] \in O/\delta$ **then**
    Return $\langle q, 1.0 \rangle$ where $q = \text{NP}(o)$ for an arbitrary $o \in [l]$;
  else
    if *Client identifies a new obfuscation* $O' \subset O$ **then**
      Reiterate algorithm with O' in place of O;
    else
      Return $\langle q, C \rangle$ with $C = |[o]|/|O| \wedge q = \text{NP}(o)$ for some $o \in O$ such that C is maximized;

Visualizing the Algorithm





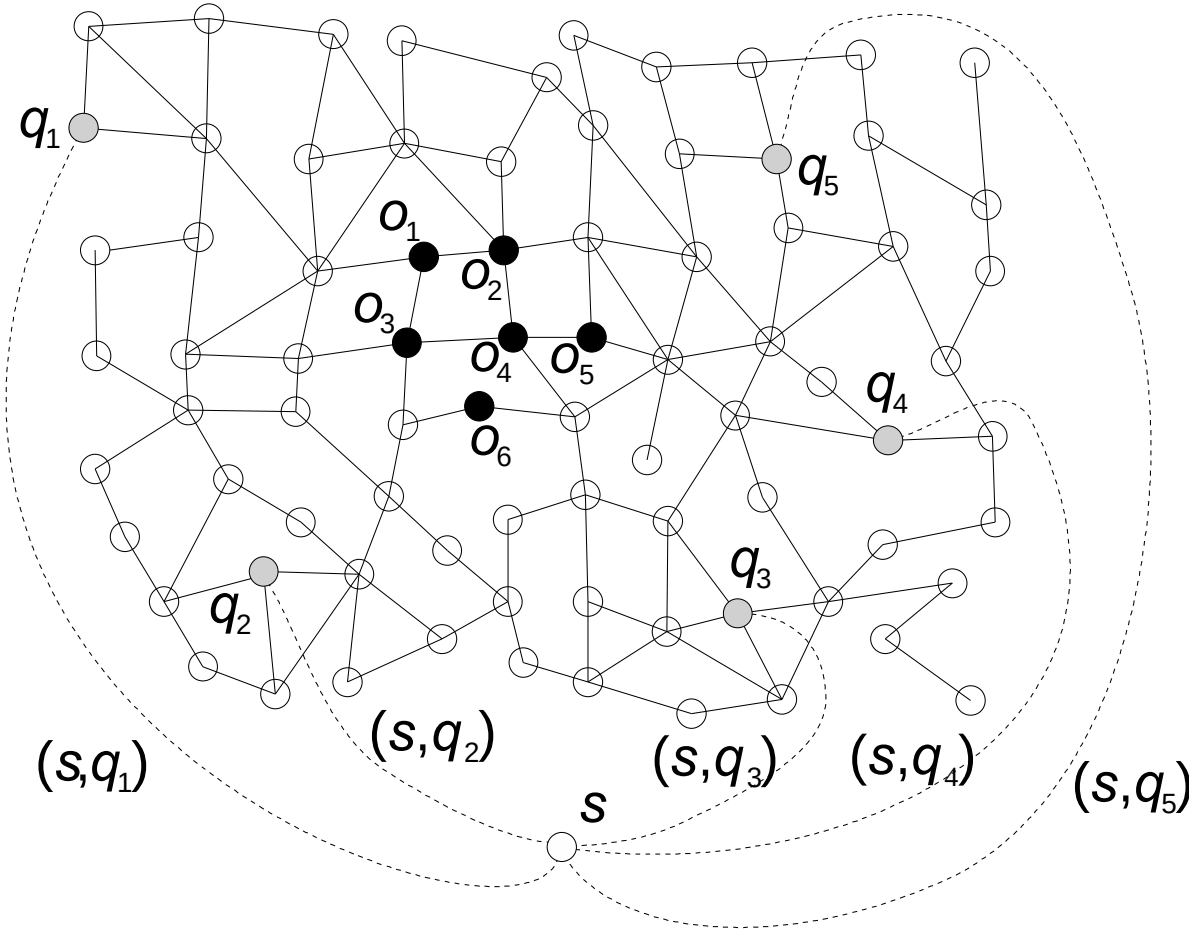




How to Compute δ

- **Naive Approach**
  - Compute the shortest path from every element of O to every element of Q
  - Complexity of a single source shortest path algorithm (Dijkstra): $O(n^2)$
  - In each step: $O(|O|n^2)$, i.e. if $|O|$ approaches n: $O(n^3)$
- **Our approach**
  - Introduce a dummy vertex such that every POI is the second vertex
  - Compute all the shortest paths on the new graph
  - Complexity per step: $O((n + 1)^2) = O(n^2)$

Computing δ Using Dijkstra



L-Diversity I

- **Homogeneity attack**
  - 31-year-old American who lives in the zip code 13053
- **Background attack**
  - 21 year old Japanese in 13068;
    Japanese are unlikely to have a heart disease

	Non-Sensitive			Sensitive
	Zip Code	Age	Nationality	Condition
1	13053	28	Russian	Heart Disease
2	13068	29	American	Heart Disease
3	13068	21	Japanese	Viral Infection
4	13053	23	American	Viral Infection
5	14853	50	Indian	Cancer
6	14853	55	Russian	Heart Disease
7	14850	47	American	Viral Infection
8	14850	49	American	Viral Infection
9	13053	31	American	Cancer
10	13053	37	Indian	Cancer
11	13068	36	Japanese	Cancer
12	13068	35	American	Cancer

	Non-Sensitive			Sensitive
	Zip Code	Age	Nationality	Condition
1	130**	< 30	*	Heart Disease
2	130**	< 30	*	Heart Disease
3	130**	< 30	*	Viral Infection
4	130**	< 30	*	Viral Infection
5	1485*	≥ 40	*	Cancer
6	1485*	≥ 40	*	Heart Disease
7	1485*	≥ 40	*	Viral Infection
8	1485*	≥ 40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

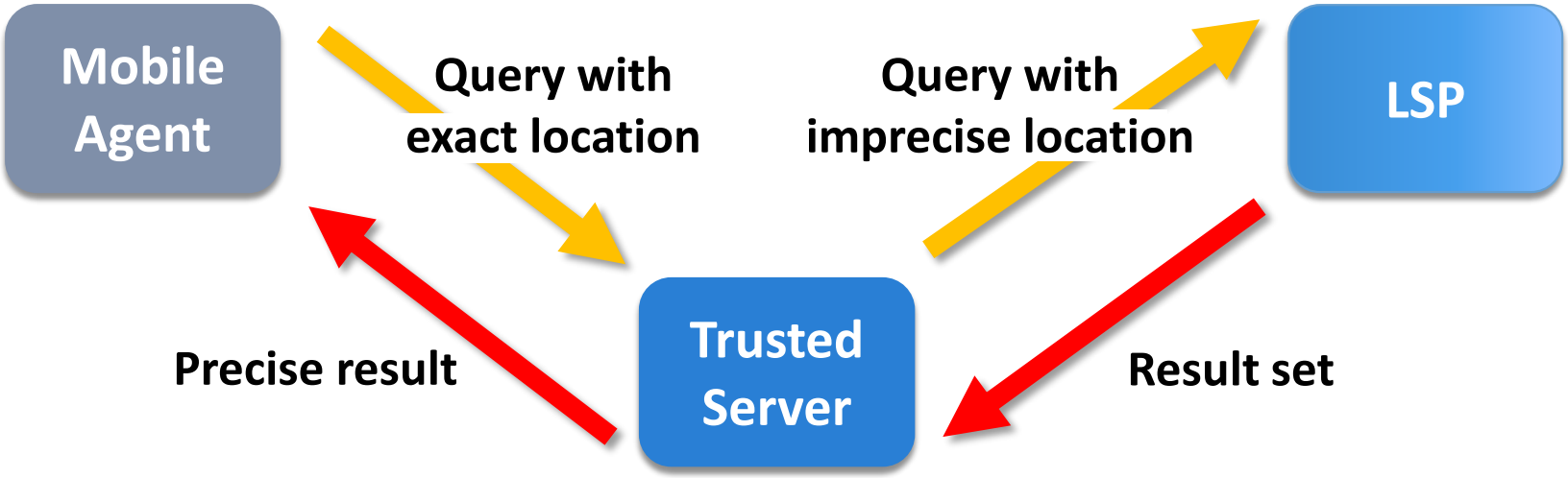
L-Diversity II

- **l-Diversity Principle**
  - A q*-block is l-diverse if contains at least l "well-represented" values for the sensitive attribute S.
  - A table is l-diverse if every q*-block is l-diverse.
  - An attacker needs l-1 damaging pieces of background knowledge to eliminate all l-1 possible sensitive values
- **LBS**
  - POI or content of queries

	Non-Sensitive			Sensitive
	Zip Code	Age	Nationality	Condition
1	1305*	≤ 40	*	Heart Disease
4	1305*	≤ 40	*	Viral Infection
9	1305*	≤ 40	*	Cancer
10	1305*	≤ 40	*	Cancer
5	1485*	> 40	*	Cancer
6	1485*	> 40	*	Heart Disease
7	1485*	> 40	*	Viral Infection
8	1485*	> 40	*	Viral Infection
2	1306*	≤ 40	*	Heart Disease
3	1306*	≤ 40	*	Viral Infection
11	1306*	≤ 40	*	Cancer
12	1306*	≤ 40	*	Cancer

Decentralized Approach to Location Privacy

Centralized Approaches



- **Limitations**
  - Communication overheads
  - Security threats
  - Single point of failure

Do Not Trust Anyone: Go Decentralized!

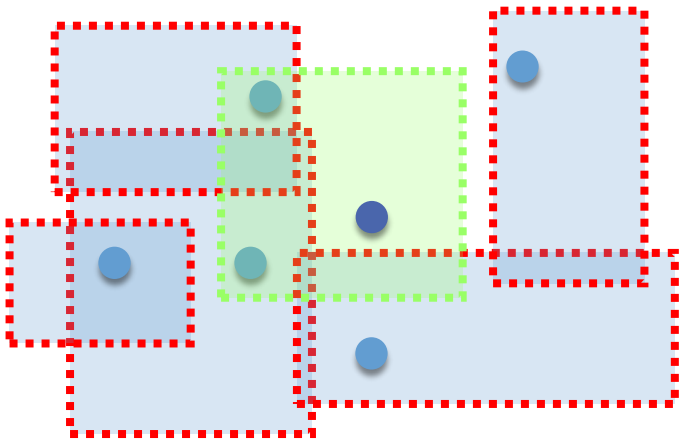
- **Idea**
  - Use WPANs (spontaneous local ad-hoc networks), e.g. Bluetooth or 802.11
  - Clique: form spontaneous local ad-hoc networks that are wirelessly connected
  - Do not disclose your precise position to anyone (obfuscation) including your neighbors
  - Be k-anonymous to your location-based service provider
- **Two roles**
  - Hide service request from mobile phone operator, i.e., separate an agent's request (query requestor) from the agent requesting this service (query initiator)

Decentralized Approach

- **Locally cloaked area (LCA)**
  - Represents imprecise location using obfuscation
  - Periodically broadcast to neighbors
- **Globally cloaked area (GCA)**
  - Obfuscation: query requestor computes its obfuscated area using the LCAs of its neighbors
- **Selection**
  - Anonymity: query initiator randomly selects a query requestor to forward its request to the LSP

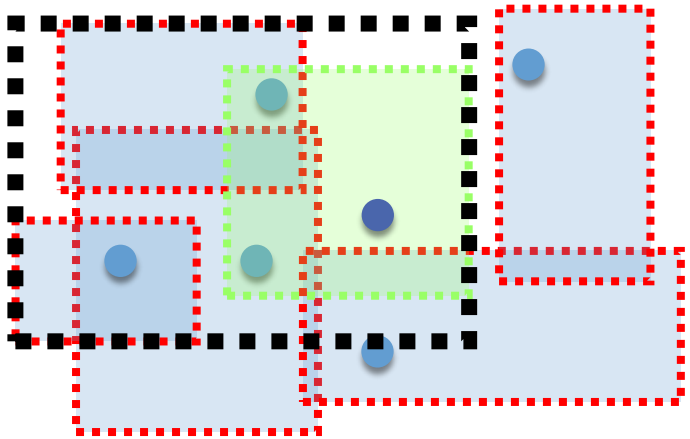
Locally Cloaked Area (LCA)

- **Agent specification**
  - Area
  - Parameter affecting ratio of length and width
  - Parameter for the agent's position relative to area's boundary
- **LCA**
  - Rectangle is described by $(x_{min}, x_{max}, y_{min}, y_{max})$



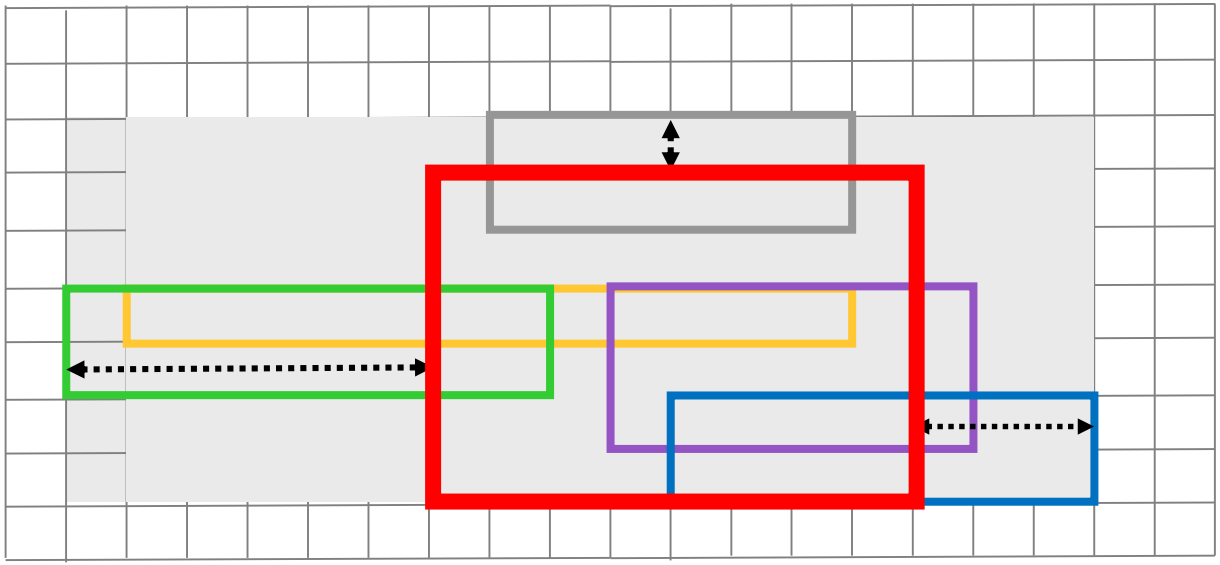
Globally Cloaked Area (GCA)

- **GCA computation**
  - An agent specifies anonymity level *k* and a required minimum area
  - Requests LCAs from its neighbors
- **Geometric problem**
  - Find the minimum bounding box of a *k*-subset (including the agent's own LCA) from *n* possible LCAs



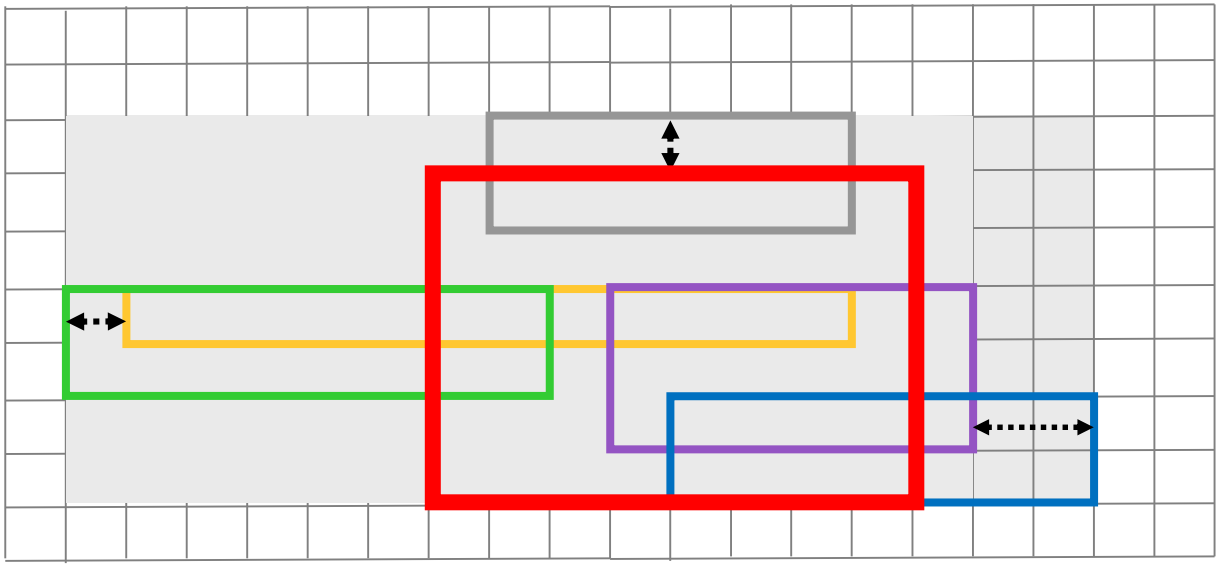
Approximating the GCA (1)

- Eliminate the rectangle whose edge has the greatest distance to the closest edge of the agent's rectangle



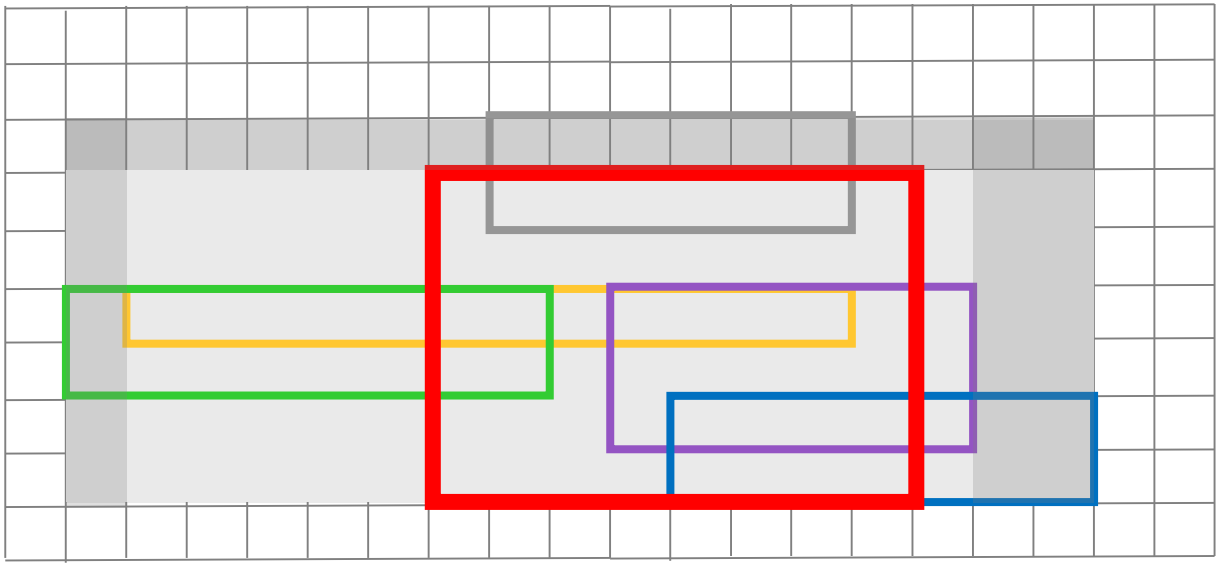
Approximating the GCA (2)

- Evaluate pairs of rectangles: those with the first and second maximum for x_{min}, x_{max}, y_{min}, y_{max}, respectively
- Compute the distance for each pair
- Eliminates the rectangle that maximizes the distance



Approximating the GCA (3)

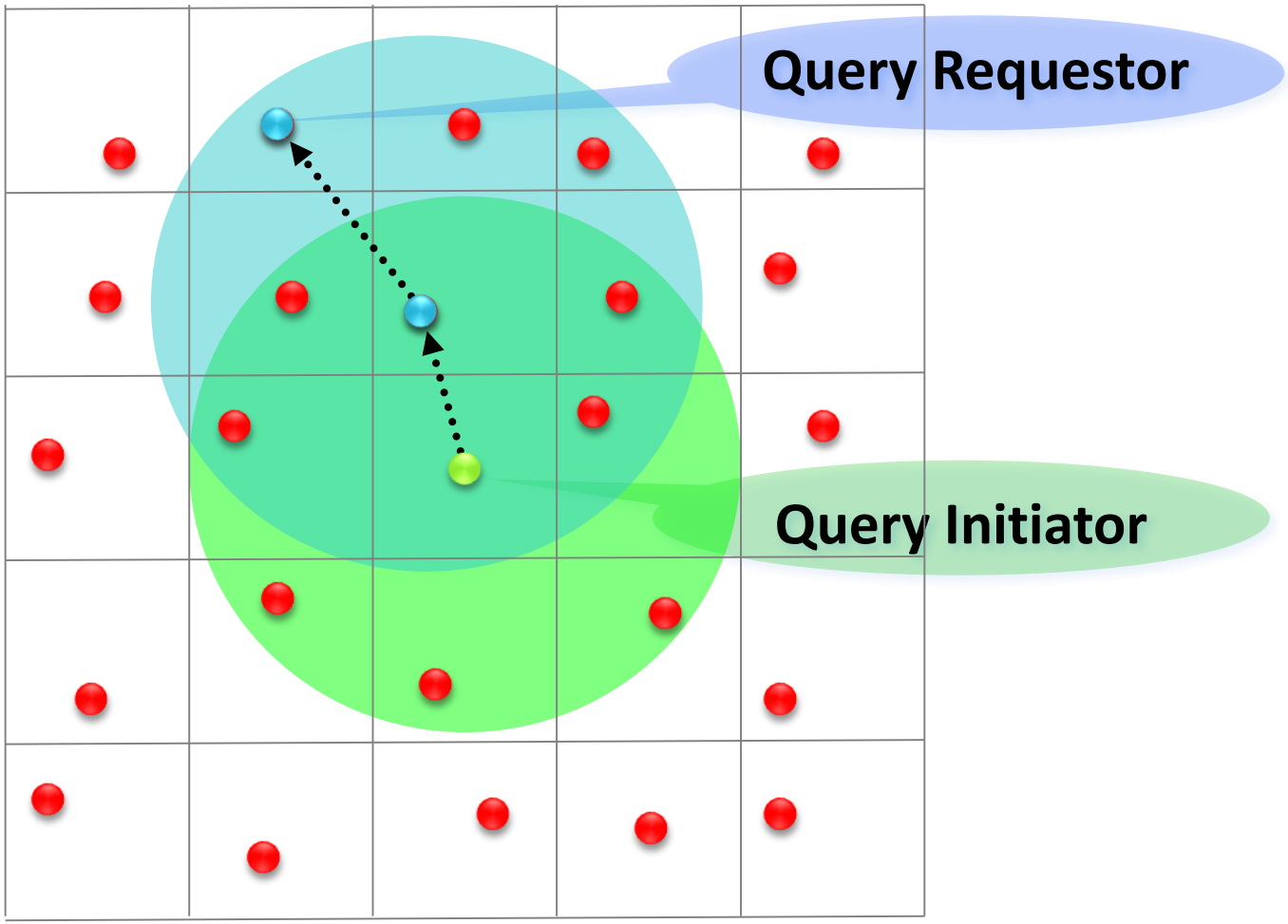
- Similar to the second criterion
- Eliminate the maximum area to be discarded from the current GCA



Random Selection

- **Problem**
  - How to select a query requestor if the hop distance is greater than 1?
- **Anonymous selection algorithm**
  - We propose a near-uniform random selection process for a query initiator to select a query requestor
  - Even from agents that are not in communication range
  - Without disclosing the IDs of the involved agents

Random Selection



Frequency Computation

- **Broadcasting message (query initiator)**
  - Unique message ID
  - Count initialized to 1
  - Hop count initialized as the maximum hop count

	0		0	0		0
0	0			0	0	
		1		1		0
	0	1	1		0	0
	0	1		1	0	
0	0	0		0	0	
0			0	0	0	

	0		0	0		0
0	1			1	1	
		3		2	0	0
	3	4	6		2	0
	2	3		2	1	
0	1	1		1	1	
0			0	0	0	

	1		1	2		1
1	4			4	4	
		17		12		3
	15	21	20		7	3
	14	20		13	7	
3	7	7		5	5	
1			2	2	2	

Naïve Versus Near-Uniform Selection

	0.83		0.67	1.22		0.56
0.83	2.61			2.44	2.44	
		9.09		6.36		1.57
	7.23	**10.12**	9.84		3.63	1.48
	6.32	9.22		6.16	3.27	
1.13	2.74	2.74		2.25	2.25	
0.40			0.87	0.86	0.86	

	3.47		2.53	2.35		2.18
3.47	5.22			3.31	3.31	
		4.93		2.91		**2.11**
	5.54	4.79	4.23		3.39	2.60
	5.25	4.55		2.85	2.79	
3.49	4.01	4.01		3.08	3.08	
2.67			3.51	2.18	2.18	

Secure Communication

Query initiator sends a message including its public key and the service request (encrypted using the public key of the LSP)

Query requestor sends the encrypted message to the LSP

LSP decrypts the message with its private key

LSP encrypts the requested information using the public key of the query initiator

LSP broadcasts the encrypted message in the query initiator's GCA

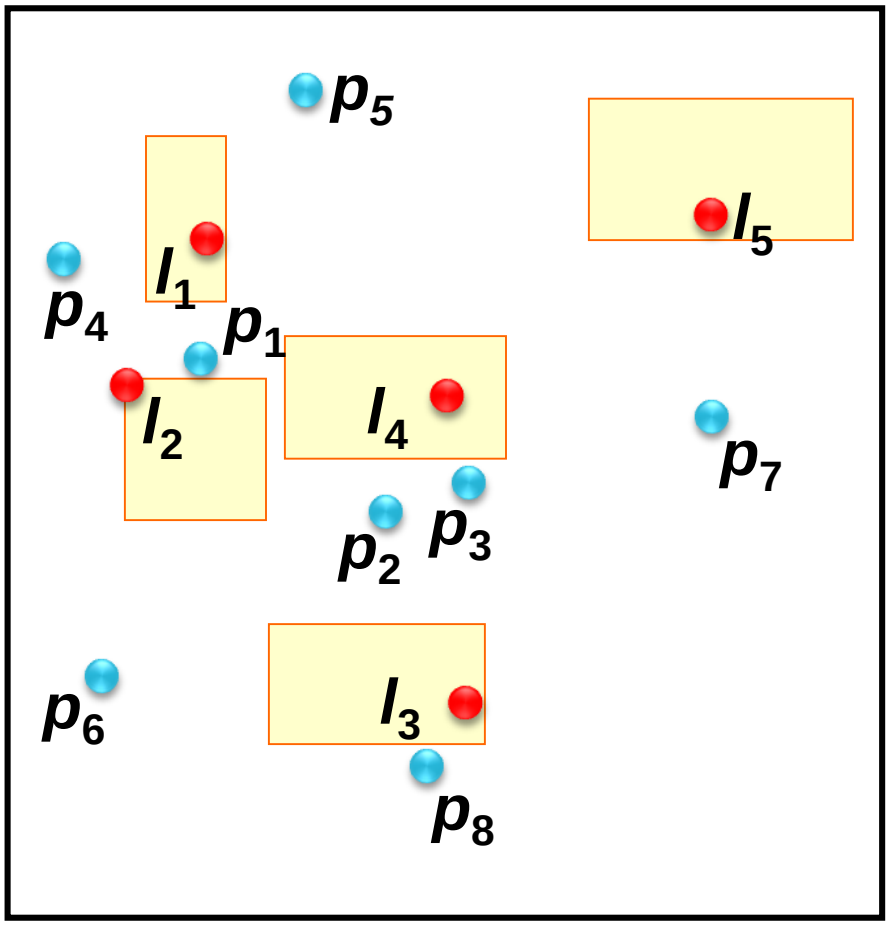
Privacy Preserving Group NN Queries

Privacy Preserving *k*-GNN Queries

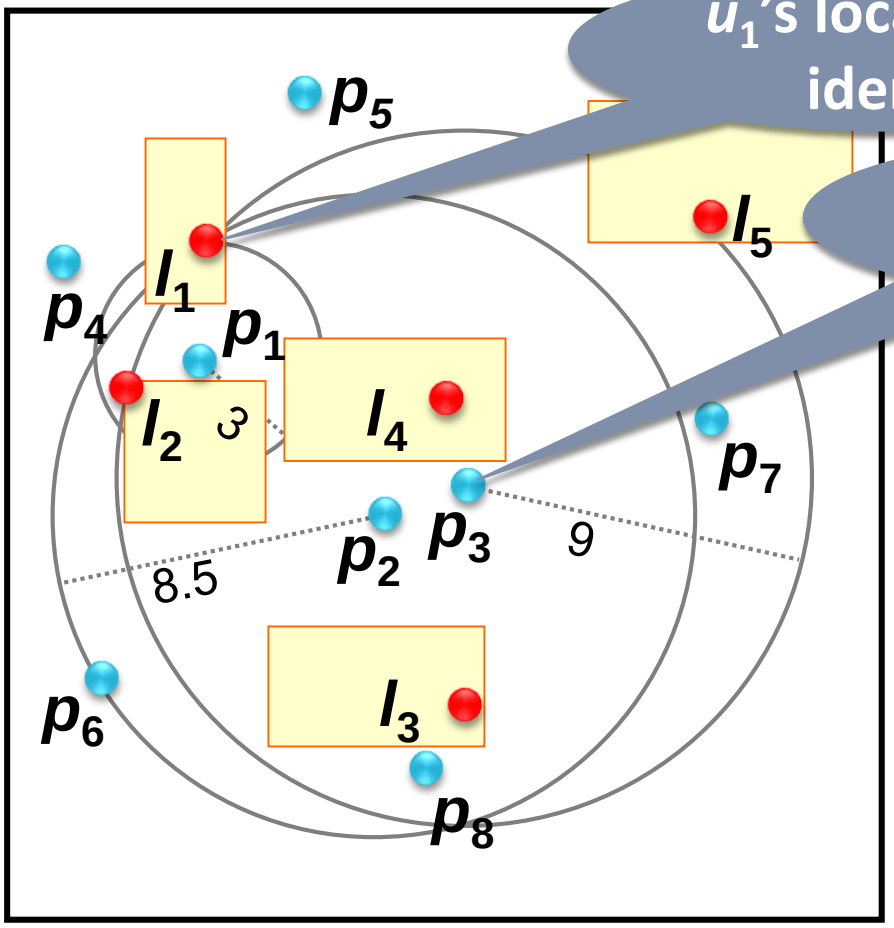
- **K-group nearest neighbor queries**
  - Returns the location of a meeting place that minimizes the aggregate distance (SUM or MAX) for a group of users
  - Example: find a restaurant with the smallest travel distance (SUM) or time (MAX) for friends located at different positions
- **Challenge**
  - Do not reveal exact locations to the location service provider and not even to your friends!

Private Group Nearest Neighbor Queries

- How do we find the actual group nearest neighbor without revealing the locations of users to others?
- How do we efficiently evaluate group nearest neighbor queries with respect to a set of rectangles?



Distance Intersection Attack I



	p_3	p_4	p_5	p_6	p_7	p_8
	9	3.5	4.5	11.5	13.5	14
	8.5	3.5	8.5	7.5	14.5	12
			15.5	9	9	2
			8.5	11	6.5	9
u_5	8.5	15.5	10	18.5	5	15.5

Distances from users to data objects

	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8
u_1	3	8.5	9	3.5	4.5	11.5	13.5	14
u_2	5	16	17.5	7	13	19	28	26
u_3	16	21	23	22	28.5	28	37	28
u_4	22	24.5	25	32	37	39	43.5	37
u_5	34.5	35	33.5	47.5	47	57.5	48.5	52.5

Updated distances after adding each user's actual distance to the data objects

Distance Intersection Attack II

- **Distance intersection attack**
  - If a user's distances from three or more data objects are revealed, the user's location is the intersection point of all circles

	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8
u_1	3	8.5	9	3.5	4.5	11.5	13.5	14
u_2	2	7.5	8.5	3.5	8.5	7.5	14.5	12
u_3	11	5	5.5	15	15.5	9	9	2
u_4	6	3.5	2	10	8.5	11	6.5	9
u_5	12.5	10.5	8.5	15.5	10	18.5	5	15.5

Distances from users to data objects

	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8
u_1	3	8.5	9	3.5	4.5	11.5	13.5	14
u_2	5	16	17.5	7	13	19	28	26
u_3	16	21	23	22	28.5	28	37	28
u_4	22	24.5	25	32	37	39	43.5	37
u_5	34.5	35	33.5	47.5	47	57.5	48.5	52.5

Updated distances after adding each user's actual distance to the data objects

Solution: Use Aggregate Information

- **LSP**
  - Set of candidate data objects
  - For each candidate data object p_h:

    $d_{max}(p_h) = \sum_{i=1...n} MaxDist(R_i, p_h)$

R_i = User u_i's rectangle (imprecise location)

l_i = User u_i's precise location

- **User**
  - u_i computes $d'_{max}(p_h)$ for p_h and updates $d_{max}(p_h)$

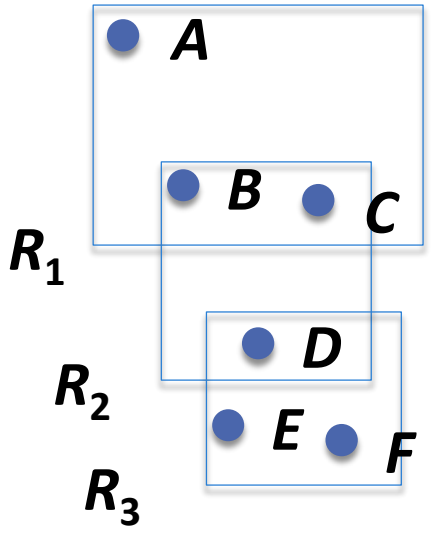
    $d_{max}(p_h) \leftarrow d_{max}(p_h) - MaxDist(R_i, p_h) + Dist(l_i, p_h)$
- **After all updates**
  - $d_{max}(p_h)$ represents the total distance of p_h to the group:

    $d_{max}(p_h) = \sum_{i=1...n} Dist(l_i, p_h)$

Attacks to Location Privacy

Query Sampling Attack

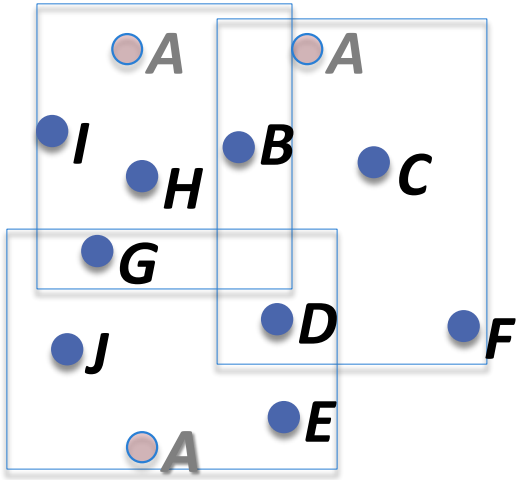
- **Assumptions**
  - Positions (but not identities) are known
  - User choice: smallest region for best QoS
- **Attack**
  - Outliers are easily identified



Users	*k*-Level	Region
A	3	R_1
B, C	3	R_2
D, E, F	3	R_3

Query Tracking Attack

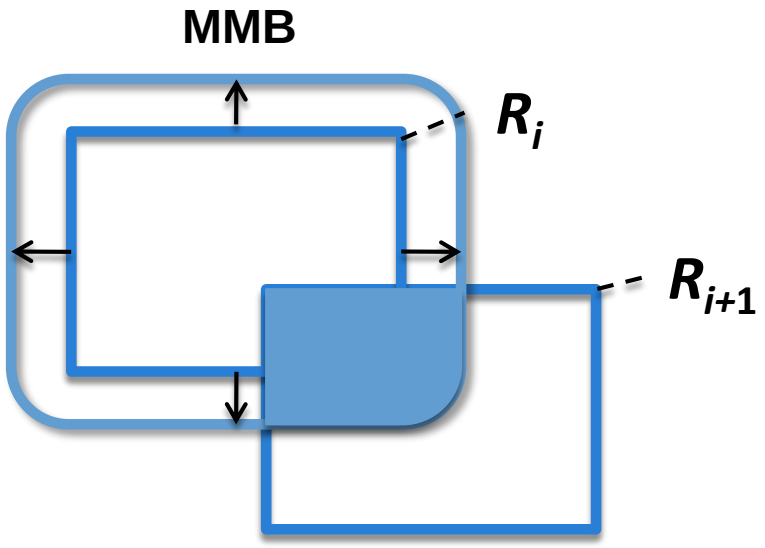
- **Assumptions**
  - Continuous query
  - Positions (but not identities) are known
  - Persistent pseudonym during updates
- **Attack**
  - Intersection of candidate lists



Time	Candidate List
t_0	*A, B, C, D, F*
t_1	*A, B, G, H, I*
t_2	*A, D, E, G, J*

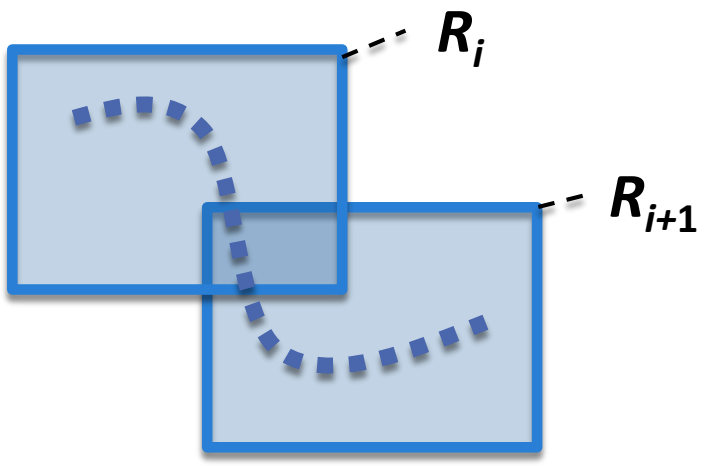
Maximum Movement Boundary Attack

- **Assumptions**
  - Continuous query
  - Maximum possible speed
  - Same pseudonym during two consecutive updates
- **Attack**
  - Maximum speed defines possible locations
  - Intersect MMB (maximum movement boundary) with updated query rectangle



Query Trajectory Attack

- **Assumptions**
  - Continuous query
  - Continuous updates necessary to ensure result is correct
- **Attack**
  - Intersect rectangles of two consecutive updates to refine the user location



Context Inference (Outdoor)

- **Patterson et al.**
  - GPS traces to infer in real-time a person's mode of transportation: foot, bus, car
  - Prediction of their future route based on their history
  - Enables privacy attacks on likely future locations
- **Krumm**
  - Inference attacks on 172 GPS traces
  - Correct computation of home address: 12%
  - Correctly revealed identity via reverse lookup: 5%

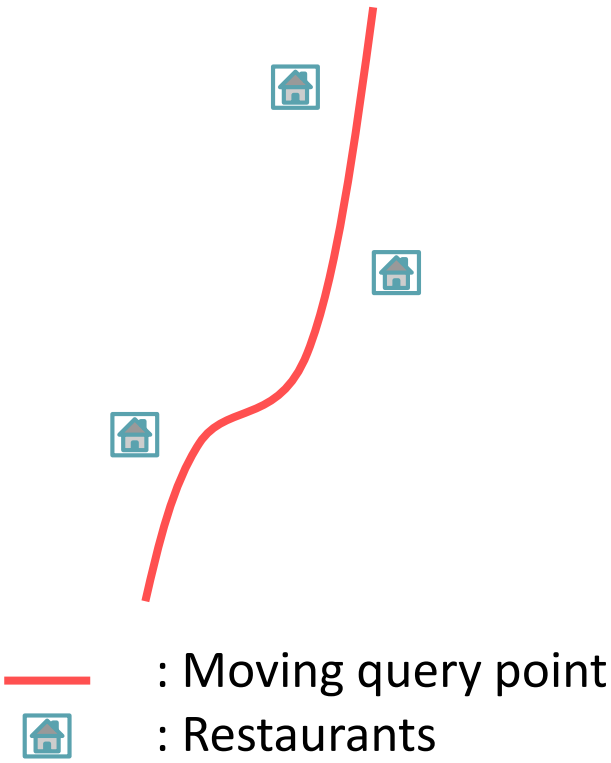
Context Inference (Indoor)

- **PEPYS (Newman et al.)**
  - Based on an indoor location sensing in offices
  - Filters to find significant events like gatherings of multiple people
- **Matsuo et al.**
  - Based on indoor location sensing
  - Inference of properties about people based on their visited locations:
    - Age
    - Work role, work frequency, …
    - Coffee/tea drinker, smoker, …

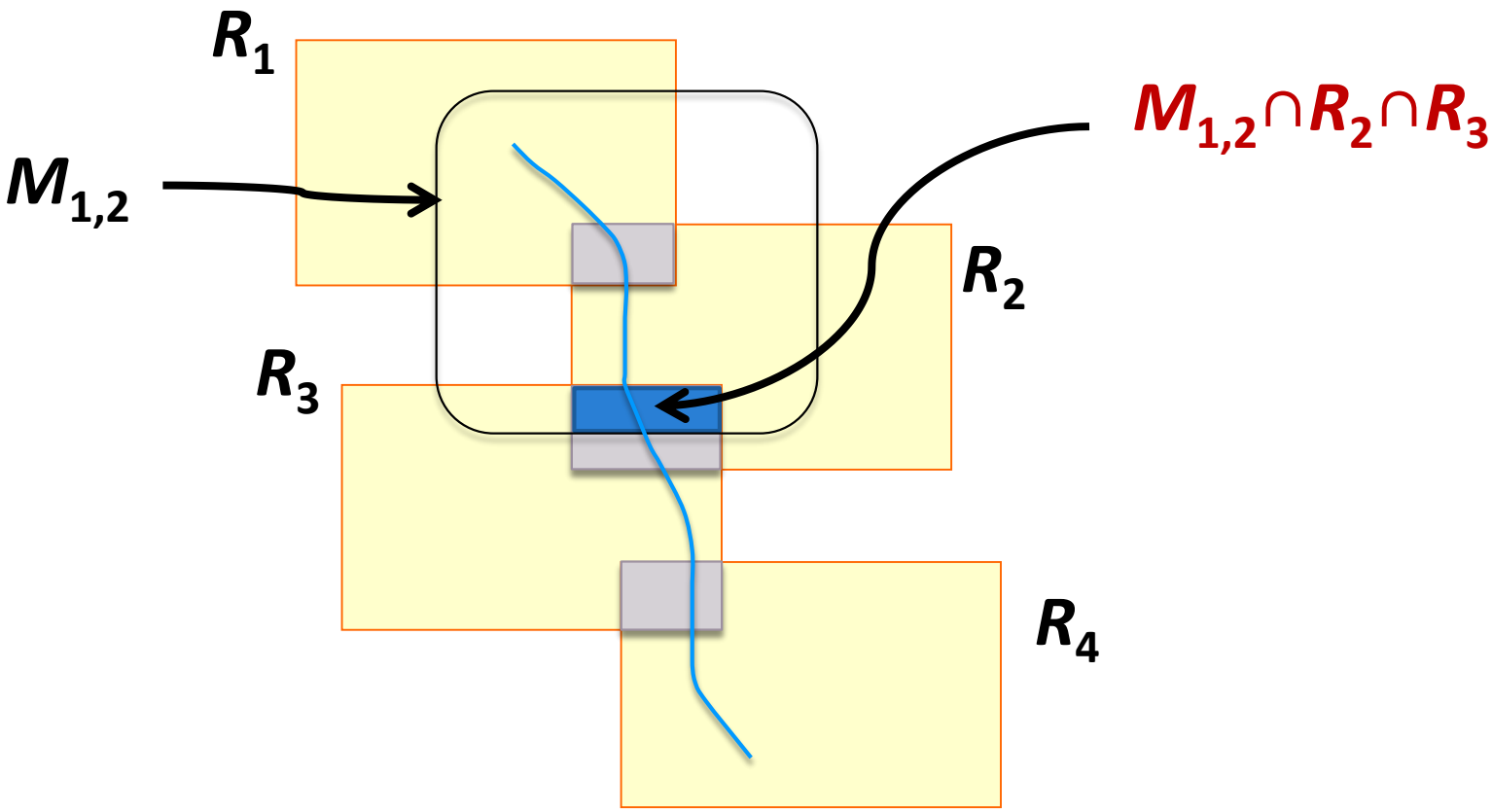
Privacy Preserving Moving NN Queries

Moving kNN Queries

- **Moving *k* Nearest Neighbor Queries**
  - Continuously returns the locations of *k* nearest data objects with regard to a moving query point
  - Scenario: the user reveals her identity for personalized service
- **Privacy Risks**
  - Continuous update of locations reveals a user's trajectory

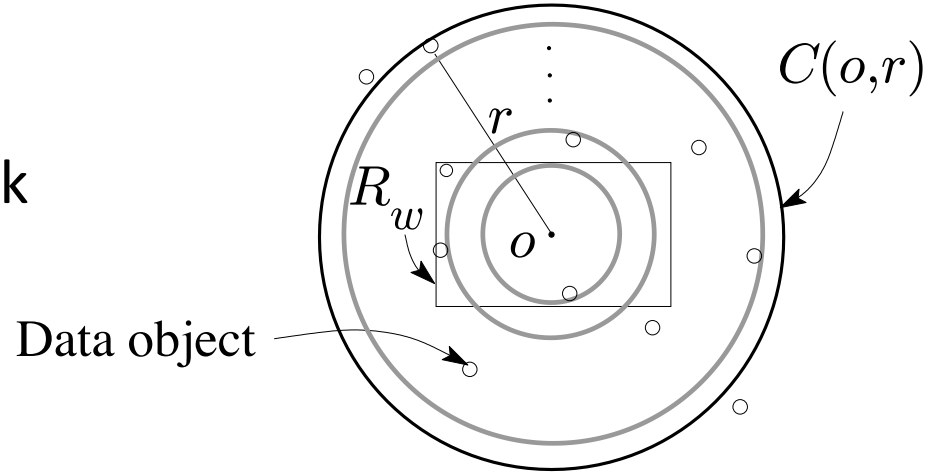


Combined Attack

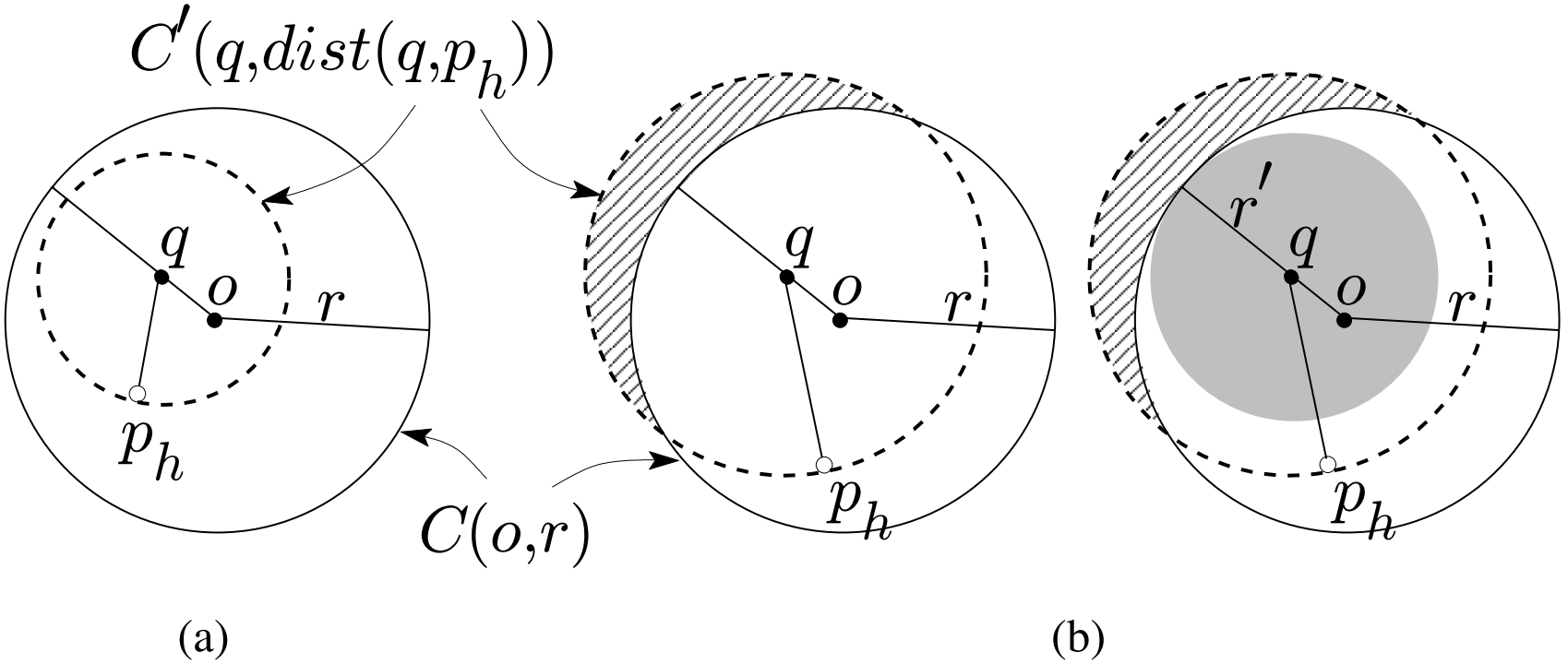


Problem Setup

- **Confidence level**
  - Provides a user with an option to trade the accuracy of the query answers for trajectory privacy
  - Guarantees that the distance of the data object to the user's location is within a bound of the actual nearest data object's distance
- **Idea**
  - Pay more: ask for a larger k
  - Travel more: confidence level cl (travel at 1/cl)



Confidence Level



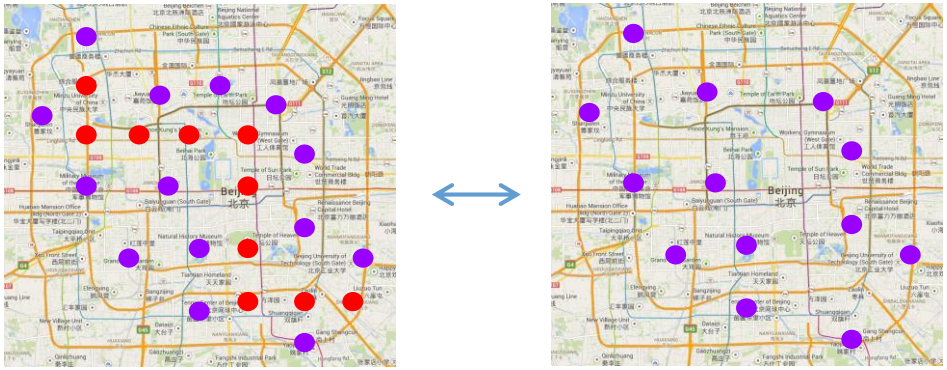
Computing Consecutive Rectangles

- **Key Idea**
  - Specify higher values for confidence level or the number of nearest data objects (i.e., *k*) or both
  - Not reveal the required ones
- **An algorithm to compute the user's consecutive rectangles**
  - Resists the overlapping rectangle attack
  - Prevents the identification of the user's trajectory
  - Efficient evaluation of a kNN query with respect to a rectangle and a specified confidence level

Private Data Exchange

Tell Me What You Want And I Will Tell Others Where You Have Been

- **Data sanitization**
  - Protection of privacy through removal



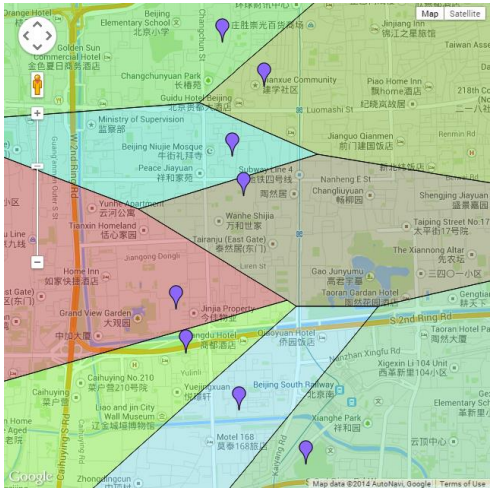
- Sampled GPS Points
- POI

- **Attacker Model**

Edge Centrality:	The more central an edge, the more frequently it occurs in a set of paths
Edge Frequency:	The adversary may have access to the trips users have take in the past
Maximum Velocity:	The adversary may assume there is a maximum velocity a user can travel

Experiments using POI Datasets

- **Dataset**
  - Use of Beijing routes from Microsoft's Geolife dataset for performance evaluation
  - Road network data was from OpenStreetMap

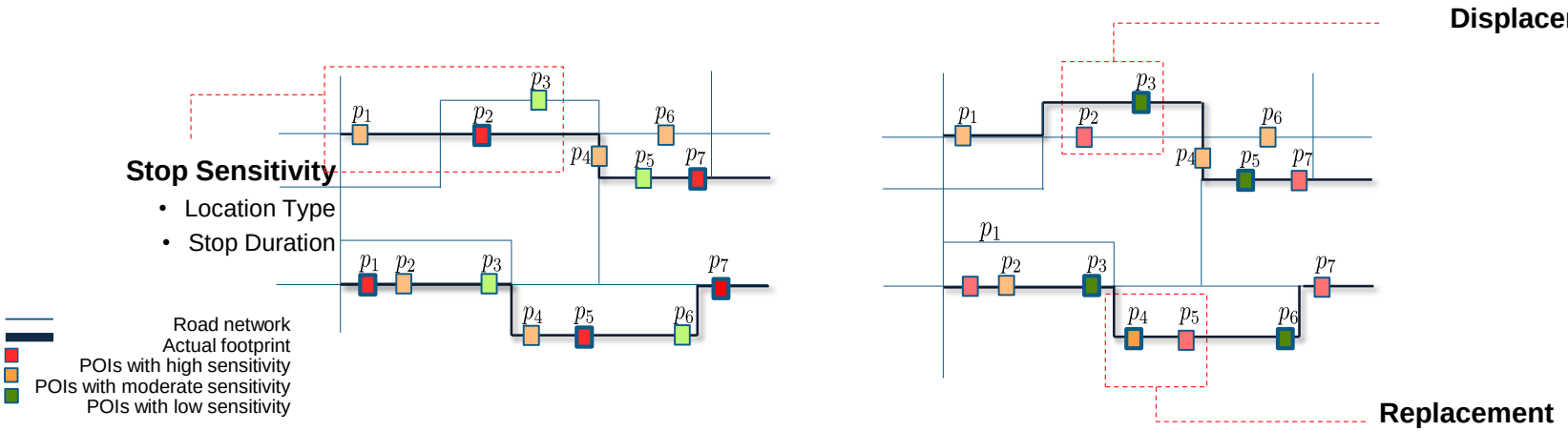


- **Evaluation**
  - Proximity circles were varied 50, 100, 250, 500 meters

POI	r = 50m (%)	r = 100m (%)	r = 250m (%)	r = 500m (%)
400	27.73	39.10	51.83	64.74
800	35.10	47.97	61.31	73.76
1,600	39.00	53.90	69.63	80.84

Sensitive Trajectory Datasets

- **Spatial and temporal exchange**
  - Aim: preserve privacy prior to exchanging trajectory datasets
  - Focuses stops as a trip can better be learnt from stops
  - Less data distortion due to local (partial) changes
  - Resilient to inference attacks
  - Exchanges sensitive stops with less sensitive POIs
  - Balances trajectory privacy against data utility



Measuring Privacy and Utility

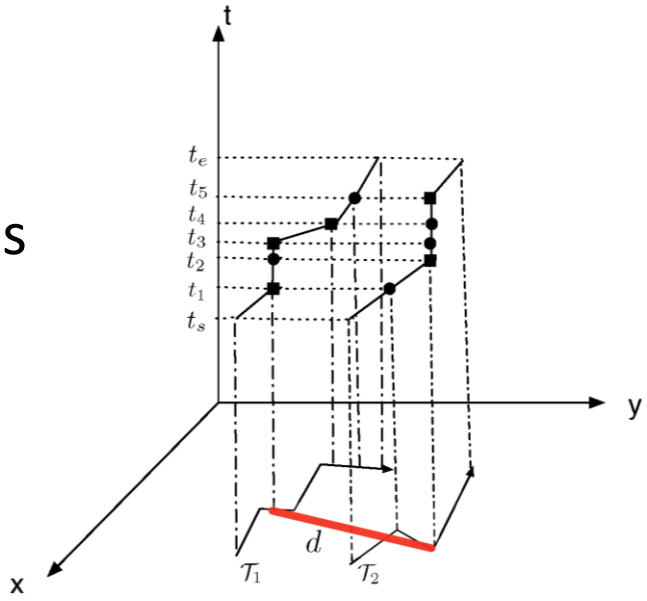
- **The sensitivity of a stop**
  - Place sensitivity rank r_p, the duration of a stop d_s, and the total duration of a trip d_t

$$r_s = r_p^{\frac{d_t - d_s}{d_t}}$$

- **Privacy and utility**
  - max_{sd} is the maximum sensitivity deviation for the k least sensitive POIs

$$Privacy\ \ Gain = \frac{\sum_{i=1}^{k}(r_{s_i} - r_{s_i}^*)}{max_{sd}} \quad \in [0,1]$$

$$Utility = 1 - \frac{Distortion(\mathcal{T}, \mathcal{T}^*)}{Distortion_{max}} \quad \in [0,1]$$



Negative Information

Monitoring Moving Objects

- **Applications**
  - Discovering movement patterns in shopping malls
  - Monitoring traffic
- **Aggregate query processing**
  - Summarized information from a number of locations for more than one moving object
- **Examples**
  - Number of distinct entries to a shopping mall
  - Traffic between two suburbs
  - Traffic on a ring road

What is Negative Information?

- **Negative representation of data**
  - Any category except the true category
  - Proposed by Esponda (2004)

- **Example**

Given a set of 4 colors

Positive information: actual color of an object

Negative information: any color but the true color

Negative Information for Privacy

- **Privacy**
  - Number of categories for negative information is much larger than for positive information
- **Scenario: wireless sensor networks**
  - Each sensor picks a negative category from *t* categories
  - Reported counts for all categories are $R_1, R_2, \ldots, R_t$
  - If *n* is the total number of reports, the actual count of category *i* is estimated as

$$A_i = n - (t - 1) R_i$$

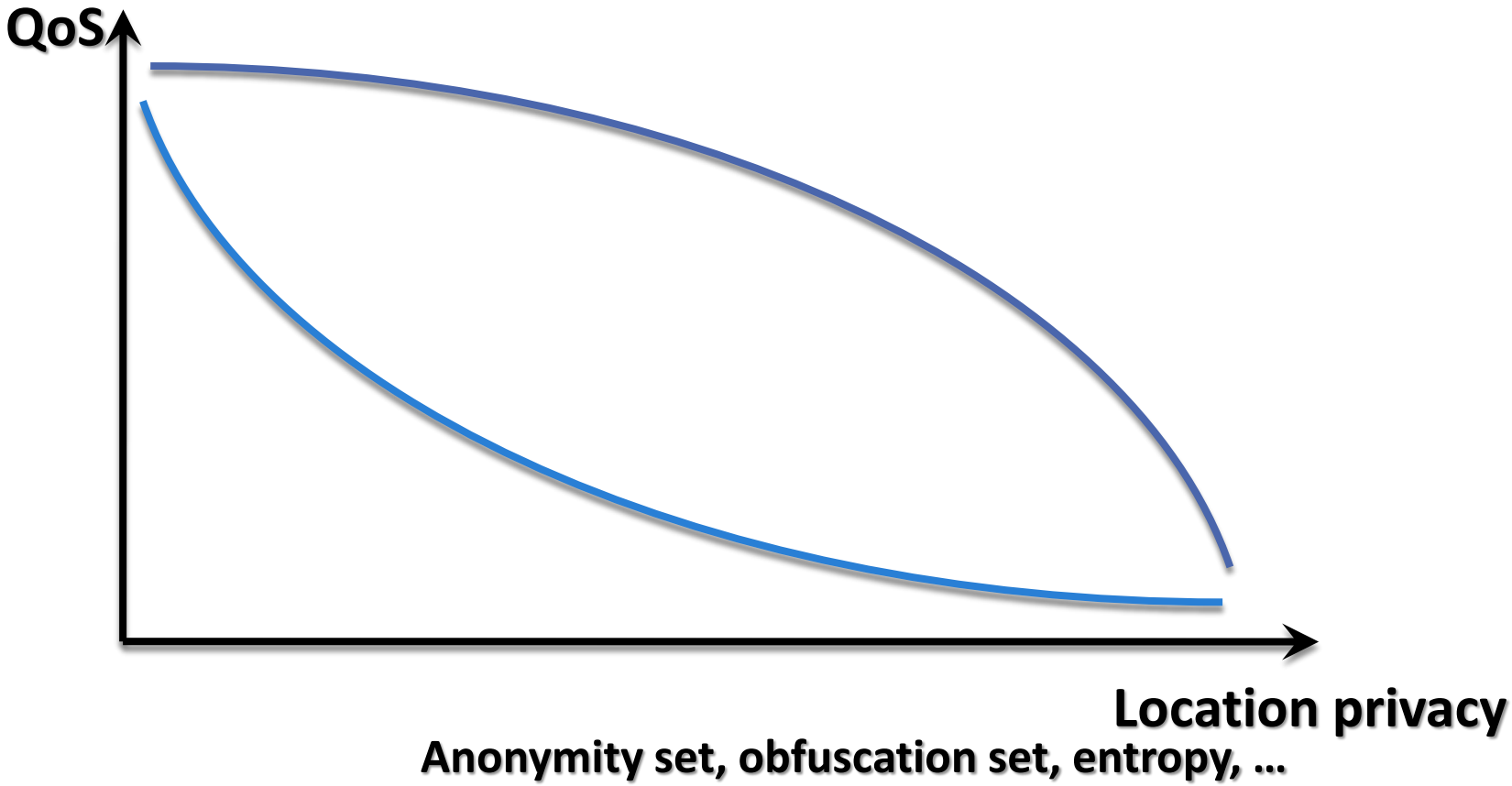
Estimating Average Velocity

- **Speed readings**
  - Divided into categories: 0 – 10 km/h, 10 – 20 km/h, …
  - Each sensor *randomly* picks a category that is NOT cover the true value
  - System counts the reported category
  - Based on the counts, system estimates the *true* count of each category
- **Traffic jam**
  - No counts in the category 0 – 10 km/h!

Metrics for Location Privacy

What is a Good Metric?

No agreement in the community on the x-axis!



Anonymity Sets

- **Definition (by Chaum)**
  - *The set of all possible subjects who might cause an action.* (following Chaum for anonymous communications)
  - Location privacy: action is requesting a LBS
- **Properties**
  - The larger the size of the set, the greater the anonymity
  - Example for location privacy: *k*-anonymity
  - Assumes all subjects are equally important!
  - Only true if pseudonyms are unlinkable

Obfuscation Sets

- **Definition (by Duckham and Kulik)**
  - The locations from which a user's position is indistinguishable/indiscernable
  - The user location is element of this set
- **Properties**
  - Location privacy is measured by the cardinality of the set
  - The larger the size of the set, the greater the location privacy

Distance Measures

- **Definition (by Hoh and Gruteser)**
  - Location privacy is the distance between a user's location and the estimated location
- **Properties**
  - Works well is user's report a false or dummy locations

Entropy

- **Shannon: Information theory (1948)**
  - Measure for the average information content
  - The less likely a character (element), the higher its information content
- **Application to location privacy (Beresford and Stajano)**
  - Entropy is maximal if all users (elements) are of equal interest to an attacker
  - Accounts for additional knowledge by conditional probabilities
  - Information content: knowledge about previous travel patterns such as the likelihood of a U-turn to link pseudonyms

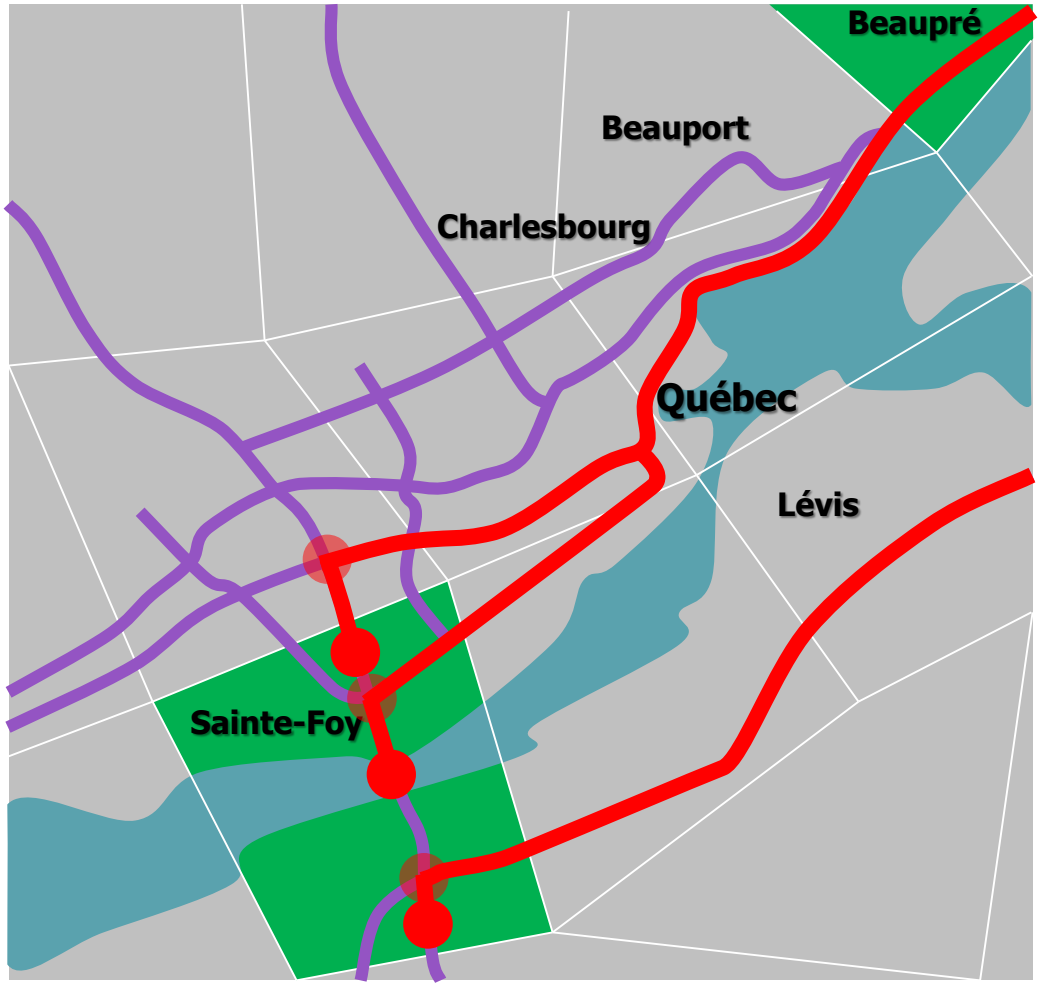
Navigation Under Imprecision

Navigation under Imprecision

- **Assumption**
  - An agent's location cannot be precisely determined
- **Contingency Strategy**
  - Search for an instruction sequence that improves an agent's chances of reaching its destination
  - Evaluate different paths from a given position to a destination location and select the one that fits as many routes as possible
  - Goal: Find a maximal set of instruction-equivalent paths

Contingency Strategy

- **Instruction sequence**
  - Turn right at the next intersection
- **Evaluation**
  - Only 1 of 3 instruction-equivalent paths leads to the destination



Contingency Strategy

- **Instruction sequence**
  - Pass the next two intersections
  - Turn right at the following intersection
- **Evaluation**
  - Provides in general not the shortest path
  - All instruction-equivalent paths lead to the goal

Contingency Strategy

- **Algorithm**
  - Based on (reversed) Dijkstra's algorithm
  - Find instruction sequences for shortest paths from all possible current locations to the destination
  - Select instruction sequence that most likely reaches the goal
  - Follow instruction sequence; ignore instructions that cannot be executed
  - If destination is reached then end, else repeat

Summary & Future

Summary

- **Obfuscation**
  - Enables an individual's identity to be revealed, facilitating authentication and personalization
  - An individual can balance the desired level of privacy against the desired quality of service
- **Decentralized privacy preserving approaches**
  - Nearest neighbor queries
  - Group nearest neighbor queries
- **Continuous location-based services**
  - Counters the combined Query Trajectory and Maximum Movement Boundary Attack

Private Location Authentication

- **Idea**
  - Provide verified locations to determine if users are where they claim they are
- **Applications**
  - PAYD (Pay As You Drive) insurance policy
  - Prevention of CC fraud
  - Restricting information access (mobile phone, laptop)
- **System**
  - Explicitly request verified locations in advance?
  - Continuous tracking of individuals with high spatial and temporal accuracy?
  - What should we do?

Information Privacy

- **Information privacy for ubiquitous systems**
  - Location privacy is a subset of information privacy
- **Examples**
  - Activity information from a wireless snooping attack
    - Cooking, showering, toileting, and sleeping by eavesdropping on the wireless transmissions of sensors
    - Works even if all of transmissions are encrypted
  - “Do not sweat your privacy”
    - Humidity sensors to infer the number of people in a room

Research Questions

- An accepted quantification of location privacy
- Types of inference attacks for continuous queries
- Approaches to protect location privacy for continuous queries
- A definitive study on the importance of location privacy
- Formal approach to trajectory privacy
- Decentralized frameworks