(K, T) - Route Privacy

KAN QI

KQI@USC.EDU

PRIVACY IN THE WORLD OF BIG DATA (CSCI 599)

04/07/2017

Why do we care about our routes?

Routes provide information about our daily lives

- Most people have a fixed set of routes for their daily lives
  - High chance of recurrence
  - Hacks: Addresses, Professions, Frequent Venues, or On-site Spy.

Our locations are sensed by the apps installed on our phones every second.

- Routes can be easily inferenced at this frequency.

What we can do about it?

Don’t allow sampling our locations

- Location information provides good utility – weather, navigation, and local recommendations.

Limit sampling frequency as policy

- What frequency is appropriate?
- How to get the data to understand an appropriate frequency?

(K, T)-route privacy and algorithm

Definition:

- The largest averaged frequency for sampling rate (F) to ensure at least K routes can be inferenced over a time series of location points $\{p_1, p_2, \ldots, p_T\}$

Assumptions:

Multiplicative Effect of routes between two points:

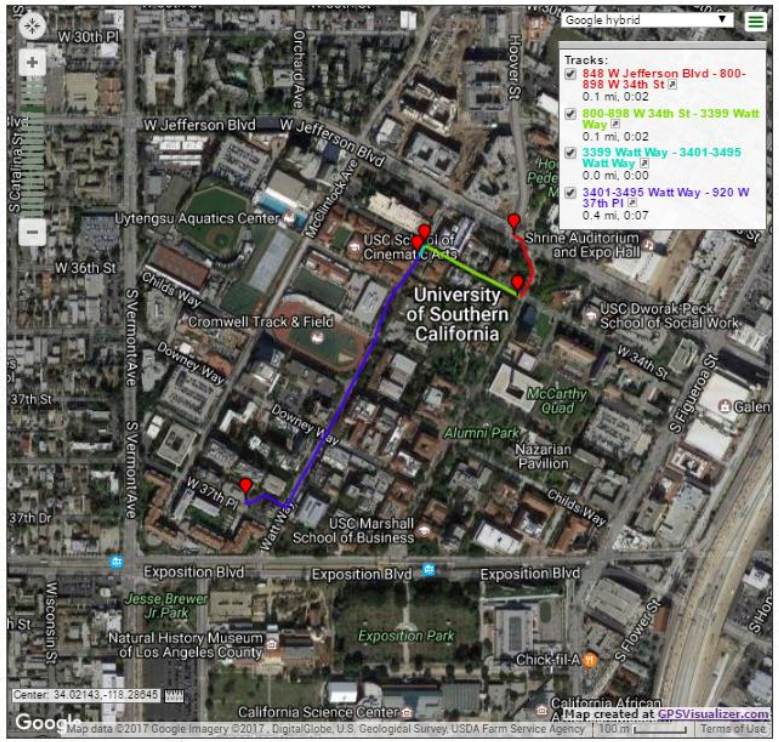
$$r(p_i, p_{i+2}) = r(p_i, p_{i+1}) * r(p_{i+1}, p_{i+2})$$

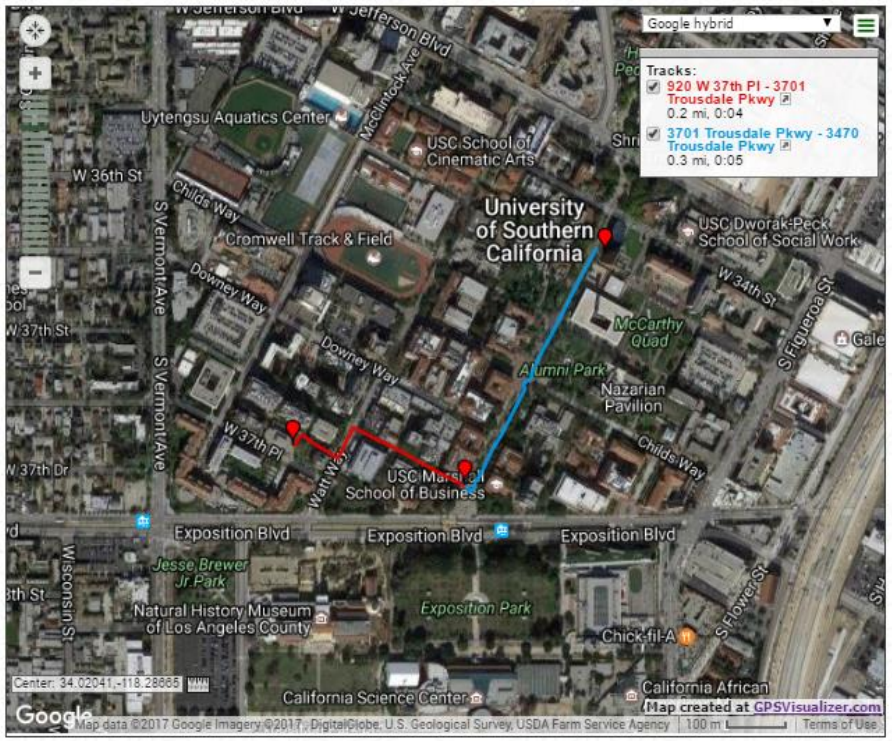
Simple Version of Algorithm is implemented

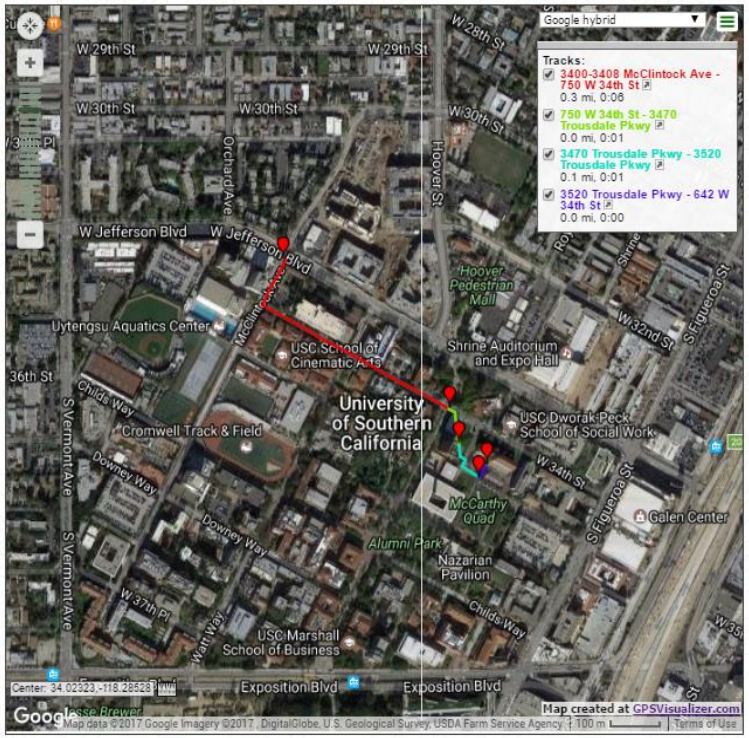
1. Given K and $\{p_1, p_2, \ldots, p_T\}$
2. Increase F from 1s to calculate R = $\prod_{(p_i, p_{i+1} \mid i, i+1 \in T)} r(p_i, p_{i+1})$, until R >= K

Experiment settings

1. Based on map of University Park Campus
2. 5 walking itineraries (speed 1m – 3m/s, about 10-15 minutes)
3. Apply ({1,2,…,15}, 15minutes)-route-privacy on the 5 itineraries by the algorithm.


Google hybrid
Tracks:
848 W Jefferson Blvd - 800-898 W 34th St
0.1 mi, 0:02
800-898 W 34th St - 3399 Watt Way
0.1 mi, 0:02
3399 Watt Way - 3401-3495 Watt Way
0.0 mi, 0:00
3401-3495 Watt Way - 920 W 37th Pl
0.4 mi, 0:07
University of Southern California
Uytengsu Aquatics Center
USC School of Cinematic Arts
Shrine Auditorium and Expo Hall
USC Dworak-Peck School of Social Work
Cromwell Track & Field
McCarthy Quad
Alumni Park
Nazarian Pavilion
USC Marshall School of Business
Natural History Museum of Los Angeles County
Exposition Park
Chick-fil-A
California Science Center
Center: 34.02143,-118.28645
Map created at GPSVisualizer.com
Map data ©2017 Google Imagery ©2017 , DigitalGlobe, U.S. Geological Survey, USDA Farm Service Agency
Terms of Use

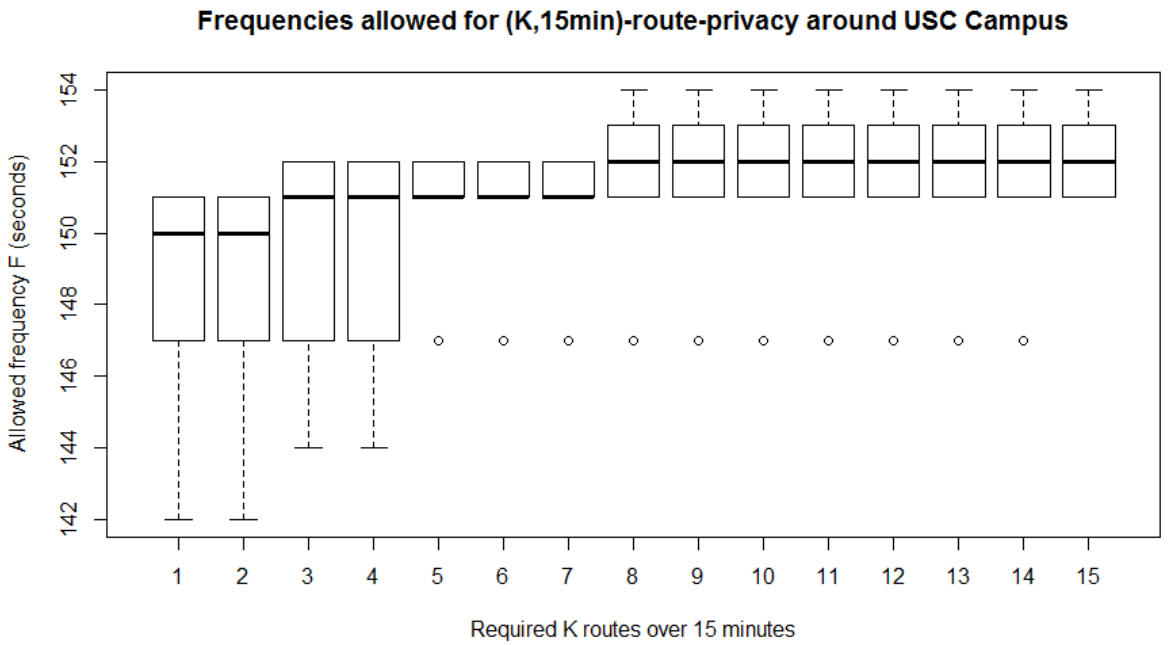

Google hybrid
Tracks:
920 W 37th Pl - 3701 Trousdale Pkwy
0.2 mi, 0:04
3701 Trousdale Pkwy - 3470 Trousdale Pkwy
0.3 mi, 0:05
University of Southern California
Uytengsu Aquatics Center
USC School of Cinematic Arts
USC Dworak-Peck School of Social Work
Cromwell Track & Field
McCarthy Quad
Alumni Park
Nazarian Pavilion
USC Marshall School of Business
Natural History Museum of Los Angeles County
Exposition Park
Chick-fil-A
California Science Center
Center: 34.02041,-118.28665
Map created at GPSVisualizer.com
Map data ©2017 Google Imagery ©2017 , DigitalGlobe, U.S. Geological Survey, USDA Farm Service Agency
Terms of Use


Google hybrid
Tracks:
3400-3408 McClintock Ave - 750 W 34th St
0.3 mi, 0:06
750 W 34th St - 3470 Trousdale Pkwy
0.0 mi, 0:01
3470 Trousdale Pkwy - 3520 Trousdale Pkwy
0.1 mi, 0:01
3520 Trousdale Pkwy - 642 W 34th St
0.0 mi, 0:00
University of Southern California
Uytengsu Aquatics Center
USC School of Cinematic Arts
Shrine Auditorium and Expo Hall
USC Dworak-Peck School of Social Work
Cromwell Track & Field
McCarthy Quad
Alumni Park
Nazarian Pavilion
USC Marshall School of Business
Exposition Park
Galen Center
Center: 34.02323,-118.28528
Map created at GPSVisualizer.com
Map data ©2017 Google Imagery ©2017 , DigitalGlobe, U.S. Geological Survey, USDA Farm Service Agency
Terms of Use


Google hybrid
Tracks:
1068-1098 W 36th Pl - 851 Downey Way
0.2 mi, 0:05
851 Downey Way - 899 W 36th St
0.1 mi, 0:01
899 W 36th St - 3600 Trousdale Pkwy
0.1 mi, 0:02
3600 Trousdale Pkwy - 3520 Trousdale Pkwy
University of Southern California
Shrine Auditorium and Expo Hall
USC Marshall School of Business
Natural History Museum of Los Angeles County
Exposition Park
California Science Center
Exposition Blvd
Center: 34.02119,-118.28769
Map created at GPSVisualizer.com
Map data ©2017 Google Imagery ©2017 , DigitalGlobe, U.S. Geological Survey, USDA Farm Service Agency
Terms of Use


Google hybrid
Tracks:
920 W 37th Pl - Bloom Walk
0.2 mi, 0:04
Bloom Walk - 688-698 Hellman Way
0.2 mi, 0:04
688-698 Hellman Way - 642 W 34th St
0.1 mi, 0:03
642 W 34th St - 622-642 USC McCarthy Way
University of Southern California
Shrine Auditorium and Expo Hall
USC Marshall School of Business
Natural History Museum of Los Angeles County
Exposition Park
California Science Center
LA Memorial Coliseum
Exposition Blvd
Center: 34.02000,-118.28519
Map created at GPSVisualizer.com
Map data ©2017 Google Imagery ©2017 , DigitalGlobe, U.S. Geological Survey, USDA Farm Service Agency
Terms of Use

Results on the 5 sample routes



An empirical finding:

An appropriate frequency of sampling pedestrian's location Information on USC campus is 150.5733s (one/2.5minutes).

Limitations

1. Only modeled for pedestrian data.
   - Implications in route identification algorithm for driving and bicycling.
2. Route data depends on the GIS systems.
   - More strict than actual scenarios, even for walking mode. False negatives often occur.
3. Depends on the maps.
   - Work as a sampling technique to gather data to set up a privacy policy for a local area.
4. The “Average” frequency condition in definition doesn’t hold in the simple algorithm.
   - Involves false negatives.

Applications

1. Setting up location sampling policy for local areas.
   - Work as sampling technique to understand averaged frequency should be enforced to ensure K route privacy over time T when developing a policy.
   - T could be hyper parameter from further study - the average walking time for a local area.
2. Providing an option for users
   - understand whether their routes can be inferred.
3. Providing data for app developers
   - understand the implications on privacy of the frequency of sampling they set in the app.

Thanks & Questions