

Tracking movements with mobile phone billing data: a case study with publicly-available data

Matthew Kwan matthew.kwan@student.rmit.edu.au

William Cartwright william.cartwright@rmit.edu.au

Colin Arrowsmith colin.arrowsmith@rmit.edu.au

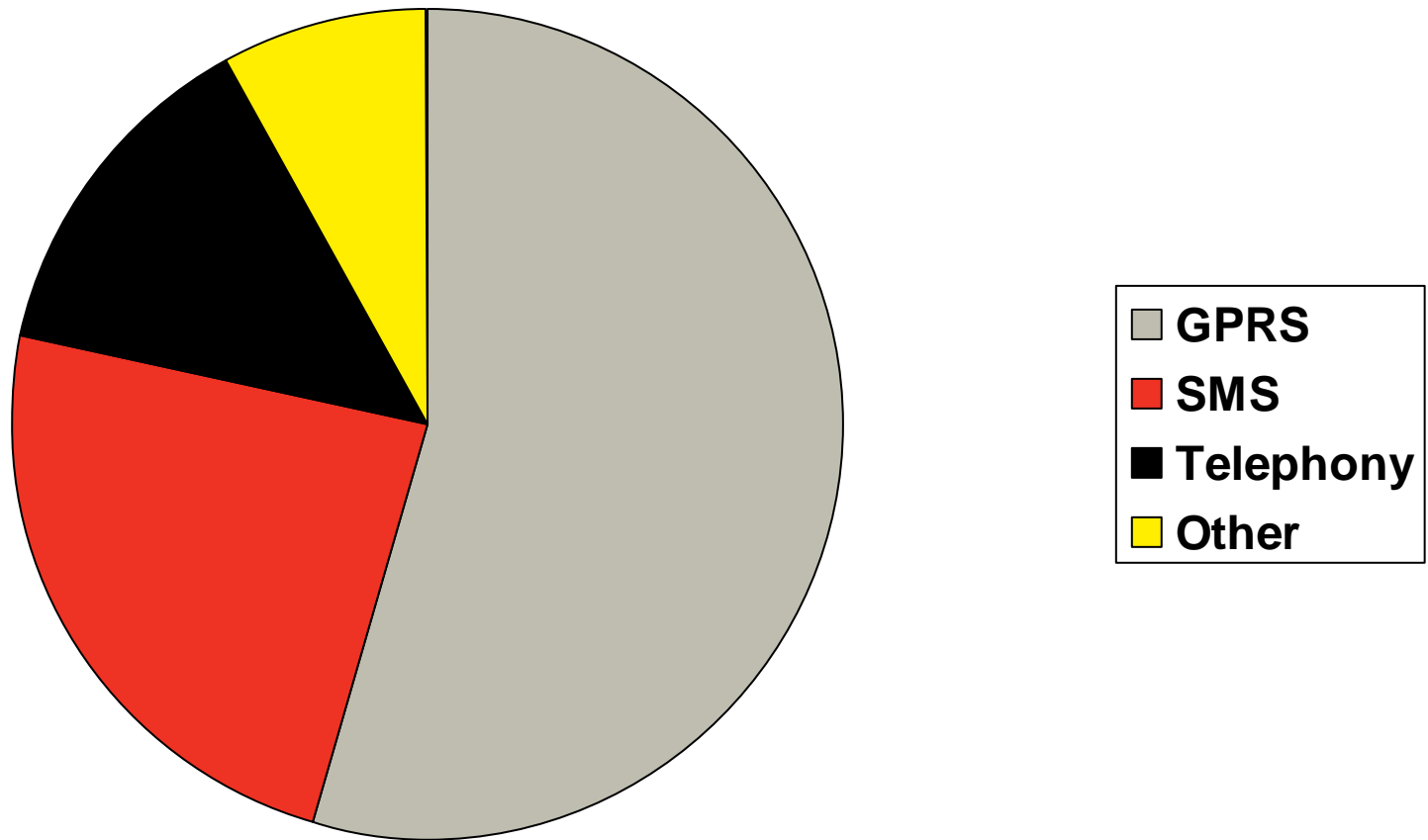
RMIT University, Melbourne, Australia

What mobile phone billing records look like

Beginn,Ende,Dienst,ein/ausgehend,Laenge,Breite,Richtung,Cell-Id_A,Cell-Id_B,Hintergrundinfo:
<http://www.zeit.de/vorratsdaten>

8/31/09 7:57,8/31/09 8:09,GPRS,ausgehend,13.39611111,52.52944444,30,45830,XXXXXX XXXX
8/31/09 8:09,8/31/09 8:09,GPRS,ausgehend,13.38361111,52.53,240,59015,XXXXXXXXXX
8/31/09 8:09,8/31/09 8:15,GPRS,ausgehend,13.37472222,52.53027778,120,1845,XXXXXX XXXX
8/31/09 8:15,8/31/09 8:39,GPRS,ausgehend,13.37472222,52.53027778,120,1845,XXXXXX XXXX
8/31/09 8:20,,,,ausgehend,,,,,XXXXXXXXXX
8/31/09 8:20,,SMS,ausgehend,13.38361111,52.53,240,9215,XXXXXXXXXX
8/31/09 8:39,8/31/09 9:09,GPRS,ausgehend,13.37472222,52.53027778,120,1845,XXXXXX XXXX
8/31/09 9:09,8/31/09 9:39,GPRS,ausgehend,13.37472222,52.53027778,120,1845,XXXXXX XXXX
8/31/09 9:12,8/31/09 9:12,Telefonie,ausgehend,13.37472222,52.53027778,120,1845,XXXXXXXXXX

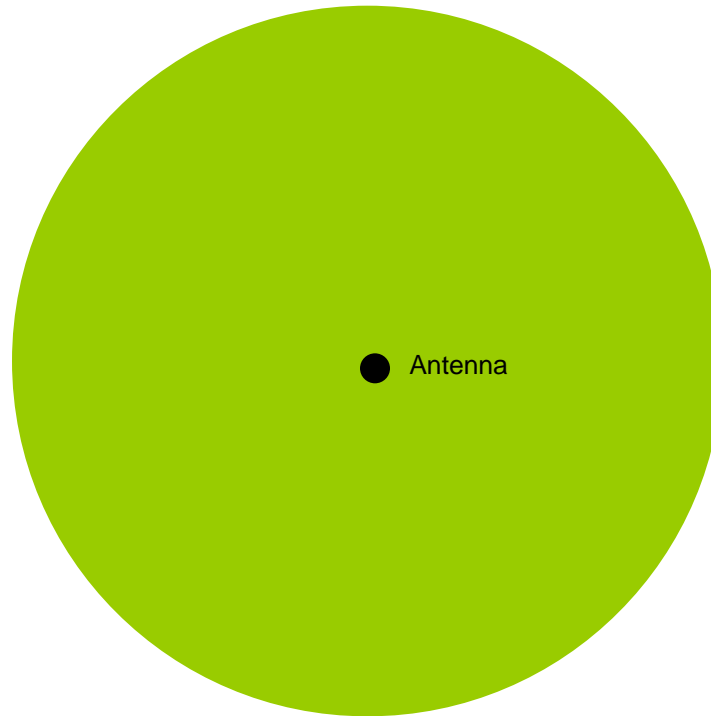
Breakdown of record types



Location data in billing records

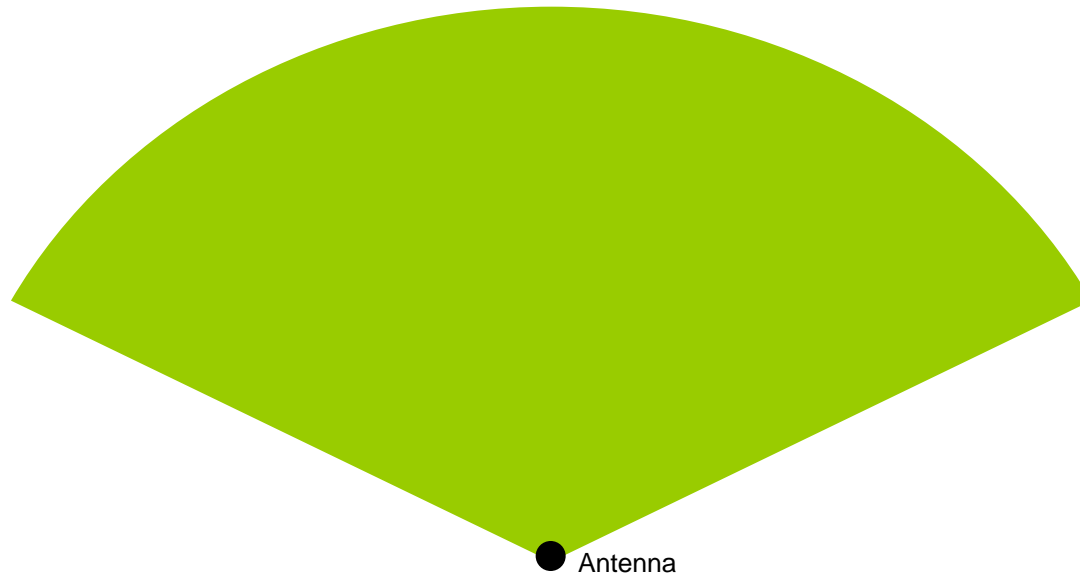
- 85% of records contain a latitude and longitude
- Equates to an average of 168 records per day with location data
- Records without location data probably occurred while roaming
- Zeit Online has an animation of the data

Estimating handset position in an omni-directional cell



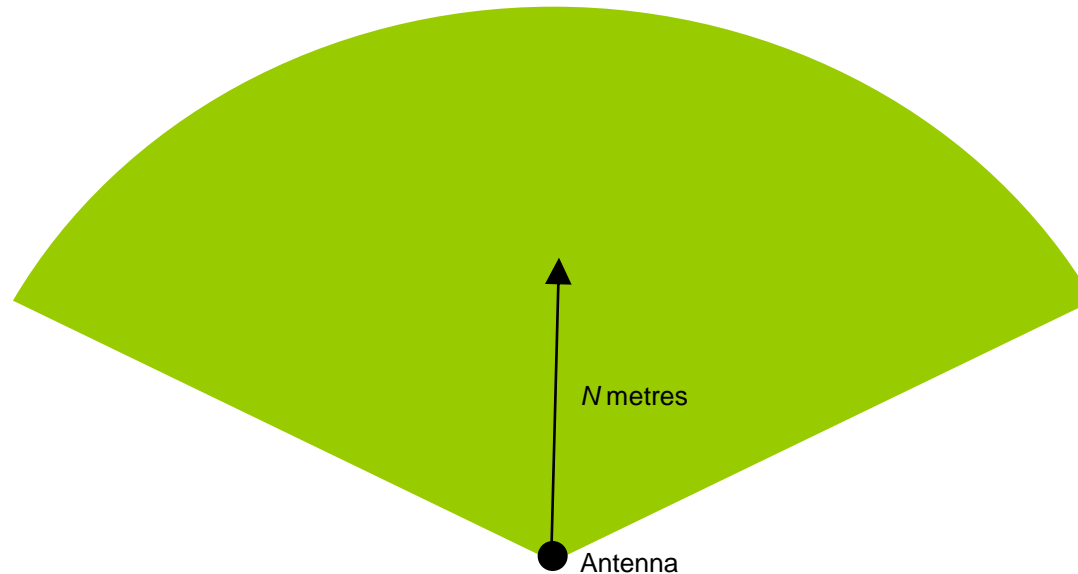
Antenna location is a good estimate of position

Estimating handset position in a directional cell



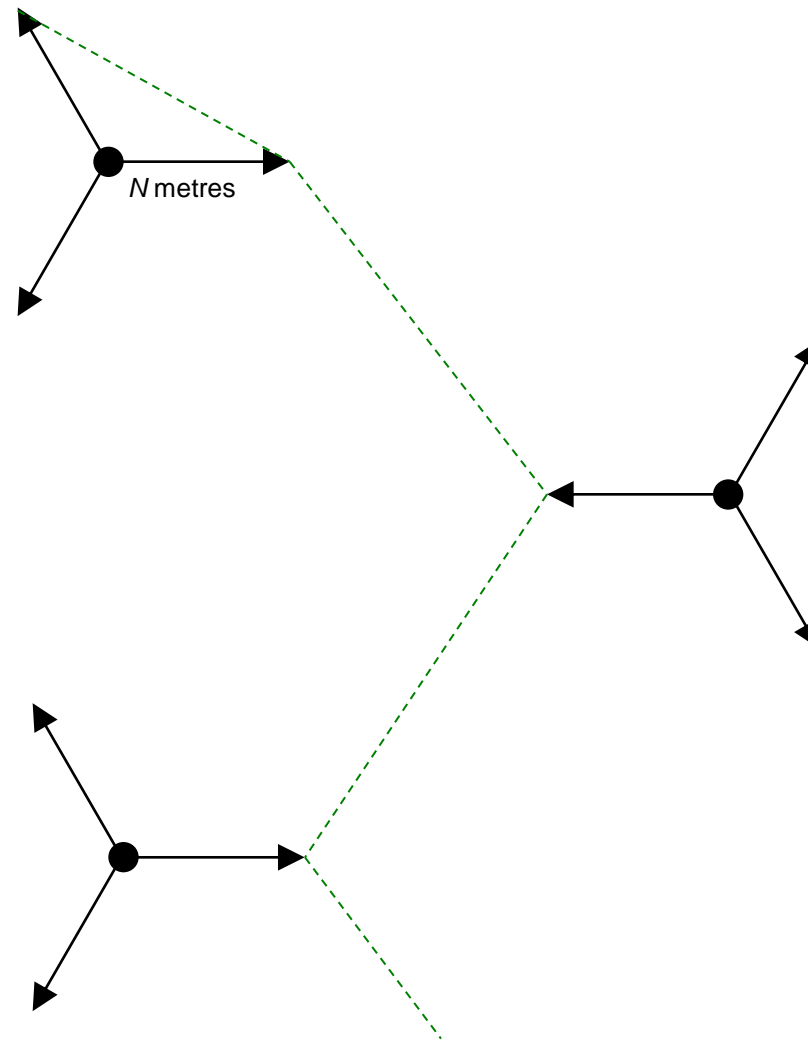
- Antenna location is a bad estimate of position
- 96.3% of cells in the data set are directional

Estimating handset position in a directional cell



- Option 1: Estimate the geometry of each cell and calculate its centroid
- Option 2: Crowd-source GPS readings within the cell and take the average
- New: Assume centroid lies along centre-line, N metres in front of antenna

Choose distance in front of antenna that minimizes total distance travelled



Best estimate of directional cell centroids

- On the Deutsche Telekom network, total distance travelled is minimized when $N = 320$ metres
- Best estimate of cell centroid is 320 metres in front of antenna
- In Australia the distance is around 700 metres

We need more public datasets of mobile billing records

- Private datasets are bad for science
- Results cannot be reproduced or verified
- Impossible to improve on past work
- Restricts the number of researchers in the field

What to do about it

- Lobby mobile phone companies to provide anonymized data to researchers who sign non-disclosure agreements
- Sue your carrier to get your own billing data – supported by EU law
- Make your data publicly available (anonymously perhaps)
- Build a public database of individual records

The end

- Questions?
- Comments?