Using data traces to improve transport systems
Introduction and initial considerations

Joe Molloy

IVT
ETH
Zürich

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

June 2017

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Why use data traces for transport modeling?

Limitations of Travel diaries and methods
- Data collection is time consuming and expensive
- Small sample sizes
- Short term, infrequent – Mikrozensus every 5 years
- Challenges in recall (trips, routes, …)
- Evaluating real world infrastructure changes is difficult

Data traces – an alternative?

Data Trace Technologies: GPS v. GSM

Trace data are generated anyway, they just needs to be captured

	GPS	**GSM - CDR**
Nature	Active	Passive
Spatial Resolution	High	Low
Temporal Resolution	High	Low, very variable
Sample size	Small	High
Doesn't work:	In urban canyons, tunnels, buildings	Out of network range
Social demographic data	**Sometimes**, when linked to travel diary	**No, by law (some exceptions)**

The Swisscom Dataset

- Much more detailed than standard CDR (yellow vs red points)
- Includes roaming users
- Triangulation possible (*)
- Signal timing and delay propagation information also available
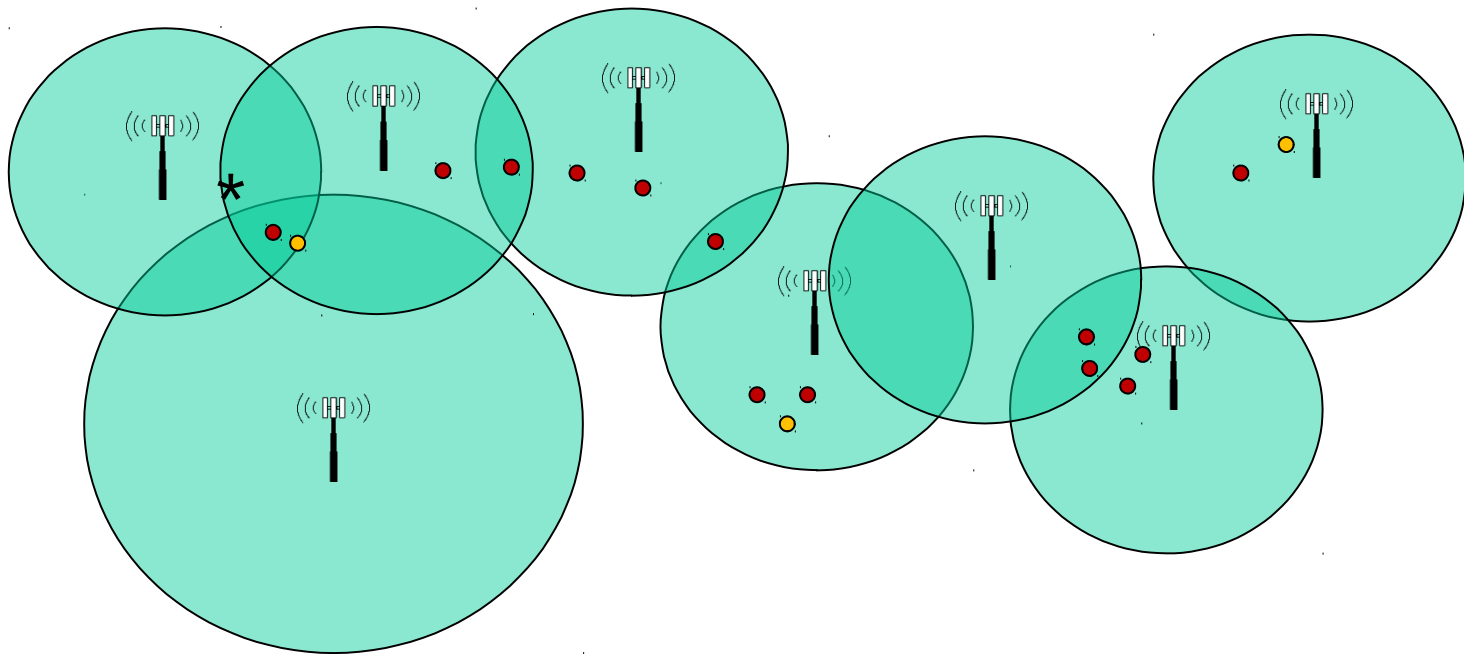
Sample Data Point

```
{ NetworkEvent : {
    "id": { "string": "228017219732835" },
    "startTime": 1455524318434,
    "endTime": null,
    "cells": {
        { "cellid": 23902, "lac": 606 }
        {"cellid": 23868, "lac": 606 }
        {"cellid": 40465, "lac": 606 }
    },
    "eventId": 128,
    "source": { "GPEH" },
    "auxiliary": {
        "map": {
            "propDelay": "7.0",
            "frameoffset": "112.0",
            "chipoffset": "33366.0"
        }
    }
}
```

[ID, time]

[Cell towers]

[Distance from Antenna]

Pros & Cons of High Resolution CDR

✓ Negates some of the disadvantages of CDR datasets
   ✓ Resolution, triangulation
✓ Can track same travellers over multiple days

- Cleaning instead of collection
- Very large datasets
  - 1 Week
  - 10,000 persons
  - ~14,000 samples per person/day

✗ Biases remain - i.e. Market share
✗ No social demographic information
✗ Operational Obstacles – Privacy, governance considerations

Our Objectives

Overarching questions:
- Can the added resolution can help understand travel behavior?
- How limiting is the lack of personal information?

Paradigm shift
- Finding questions for data to answer, not data to answer questions
- Geographic information observatories – Miller H.J. (in press)

Avenues of investigation:
1. Before/after assessment of a significant network change
2. Calibrate MATSim using mobile data
3. End goal of investigating road pricing schemes

Stage 1: Route Choice - Transjurane

Using Swisscom data to analyze the impact of the finished Transjurane

- Is traveler ● on the:
  a) Autobahn
  b) Old route
  c) Train

- Route/Mode Choice

- Traces from before & after opening of final section

- No test data to validate any algorithms

http://www.a16.ch

Discussion