Unique in the Crowd: The privacy bounds of human mobility

Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel,
Scientific Reports Vol. 3, 2013

Presented by Lindsay Shaw

Summary

- The authors examined data containing human location traces
- Huge amounts of movement data is generated all the time
- No two people are likely to have the same movement patterns

Summary

- The authors show a method for calculating the uniqueness of movement data
- 4 random points in a trace can identify 95% of traces

Summary

- Still highly unique at low temporal resolution
- Still highly unique at low spatial resolution
- Fairly unique even with few data points

Temporal Resolution - The precision of the time aspect of a point in a trace.

Spatial Resolution - The precision of the location aspect of a point in a trace.

The Good

- The authors provide a mathematical analysis of the three factors affecting uniqueness
  - Spatial resolution
  - Temporal resolution
  - Number of data points

- So long as resolution remains high in one dimension, uniqueness remains high

The Good

- This provides an interesting problem when it comes to anonymising data
- In order to become truly anonymous, the data must be significantly stripped down
- How much “anonymous” data is truly anonymous?

The Bad

- How much of real world data looks like that data used in this article?
- The data the authors had access to had a moderately low resolution - 1 hour and 1 cell
- Do we encounter the case of low resolution data in the real world?

The Bad

- Most timestamping gives a specific date and time: 1:15:23.115 pm on the 23rd of July
- Not: Between 12pm and 3pm on the 23rd
- A lot of location data comes from GPS, not cell towers.
- 5 - 50 meter range, vs 5 - 50 kilometers

The Question

Given the large quantities of mobility data on most of the public, and the uniqueness of the data implied by this article, what damage could be done by a malicious individual or group with access to such data?