Mobile and Sensor Systems

Lecture 7: Mobile Privacy

Prof Cecilia Mascolo

Mobile Privacy

- In this lecture we will discuss some issues and solutions related to mobile systems and data privacy.

Mobile Phone Data

- Mobile phones generate user data which can reveal a lot about the user
- Where does this data go?

- Initially collected by a few "trusted" companies
- More and more smaller companies have built their business model around data

App Data Gathering

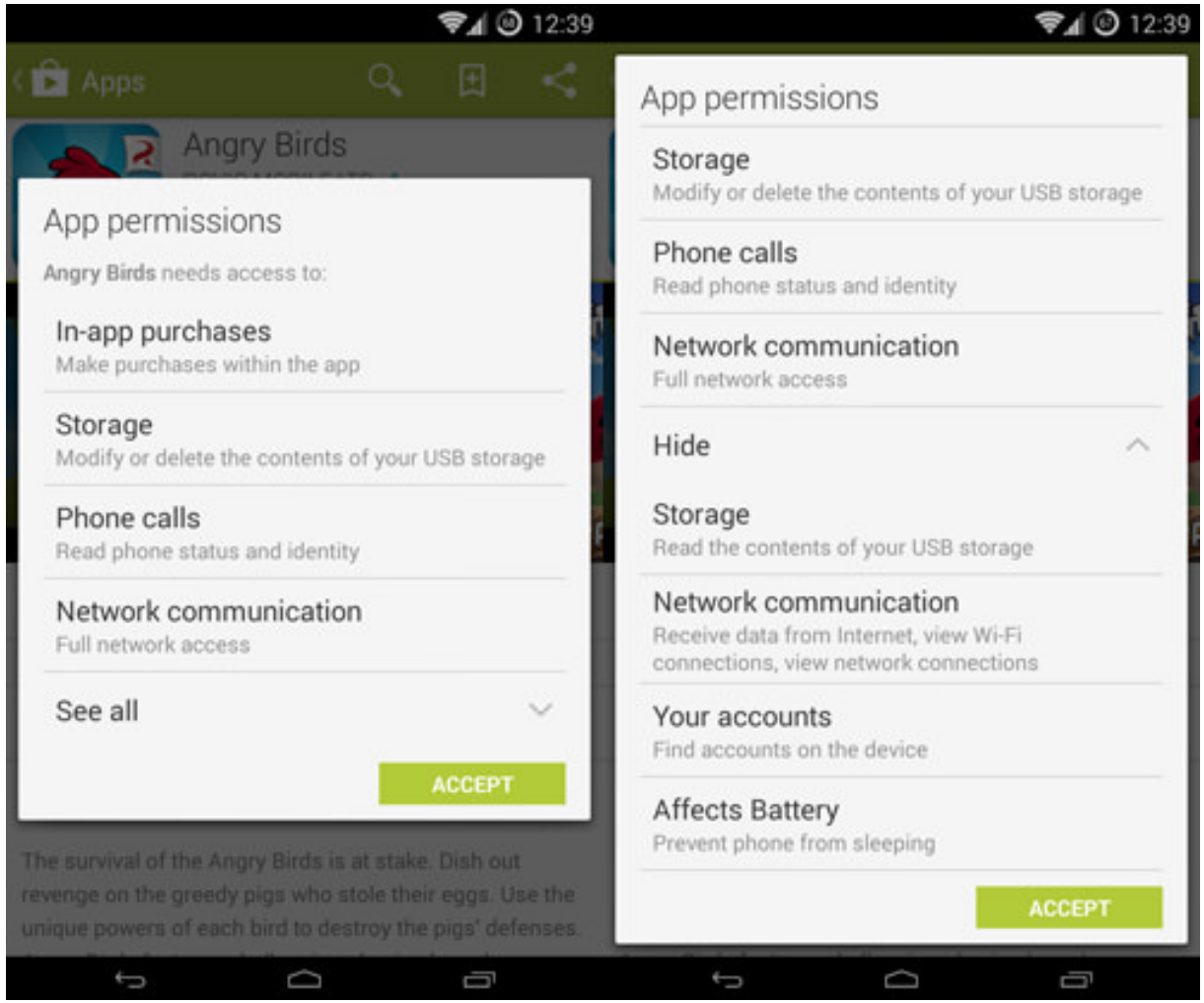
- More often than not apps collect sensing data beyond its technical needs (for advertising purposes)

Where does the data go?

- The app collects location. Does it then go to:
  - Location services
  - Advertisers
  - Developers

  - The OS does not offer support to know this…

Permissions: Android



UNIVERSITY OF CAMBRIDGE

Permissions: iOS

- Apple doing the vetting
- Now notifying users about location data usage

Privacy Data Breach Detection

- Monitor behavior of apps to determine when privacy sensitive information leaves the phone.

- Data Flow Analysis (DFA): looks for routes between data sources and sinks
  - Any of these routes without user consent is classified as leak
- Capability Leak:
  - Explicit: malicious app manages to hijack permissions granted to other trusted apps
  - eg apps can have sharedUserID: apps by same developer have same ID so permissions for that ID are shared

[Haris et al]

Sources and Sinks

Sources
Location Data: GPS, last base station location, WLAN
Unique,Identifiers: IMEI, IMSI
Authentication,Data: Cashed password data
Contact and Calendar, Contacts, address and schedule
Call State, Start and end of incoming call, number of incoming call

Sinks
SMS, Communication: data can be transferred by SMS
File Output: Applications can write data to files that are globally readable
Network: Applications can access network by sockets or HTTP
Intents, objects: applications can send data objects to other apps
Content, Resolver Apps can use API to edit shared memory of device

Dynamic Analysis: TaintDroid

- **Challenges** …
  - Smartphones are resource constrained
  - Third-party applications are entrusted with several types of privacy sensitive information
  - Context-based privacy information is dynamic
  - Applications can share information
- TaintDroid is a modification to the Android OS which allows for dynamic tracking of sensitive data movements from an app to other apps and sinks

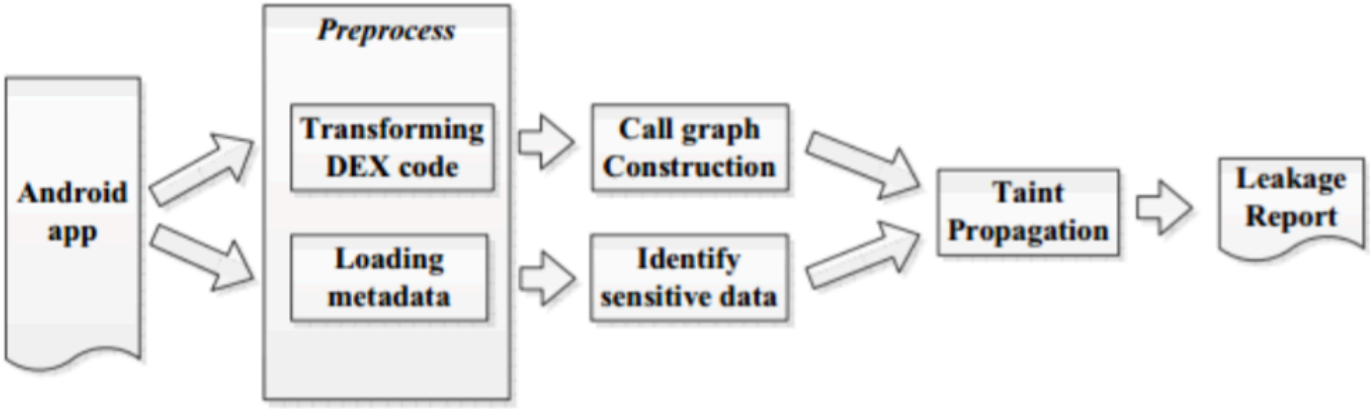
[Enck 2014]

TaintDroid

- TaintDroid automatically labels data from privacy-sensitive sources and transitively applies labels as sensitive data propagates through program variables, files, and interprocess messages.
- When tainted data are transmitted over the network, or otherwise leave the system, TaintDroid logs the data labels, the application responsible for transmitting the data, and the data destination.

Static Analysis

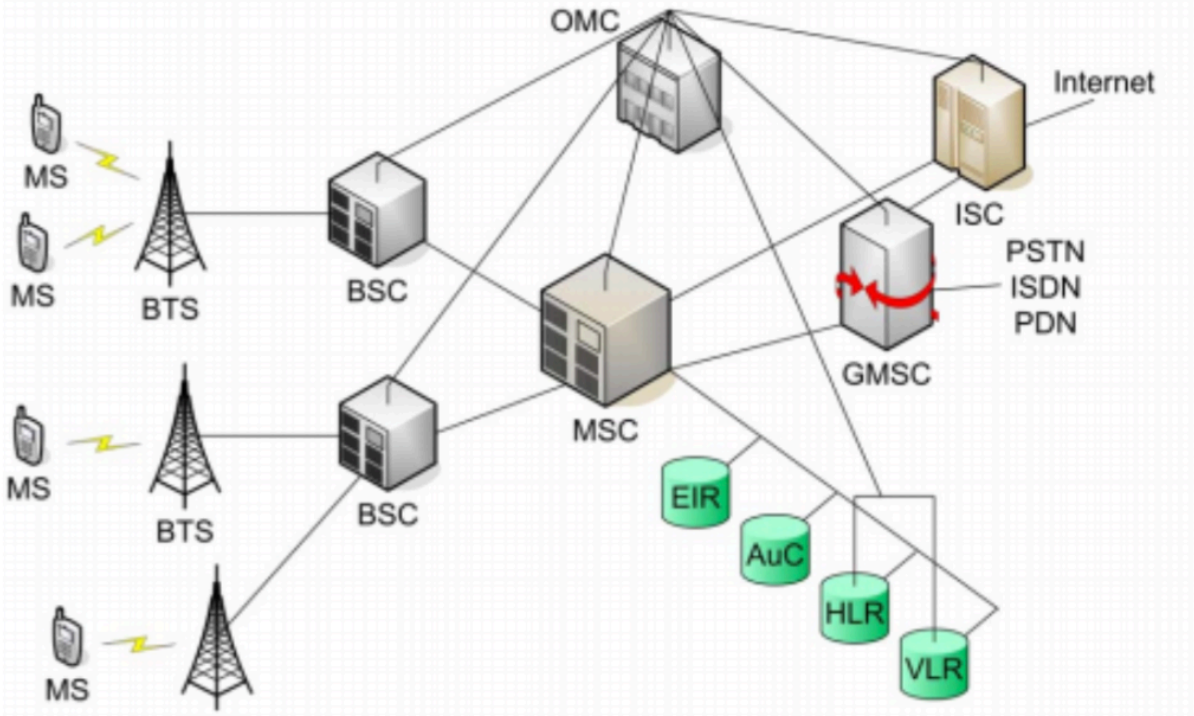
- Static Analysis covers all the paths from sources to sinks
  - Takes much longer than dynamic analysis
  - No time or efficiency overhead (done offline)
  - Example (LeakMiner):



UNIVERSITY OF CAMBRIDGE

Cellular Network Leaks

- Various parts of a cellular network are prone to leaks and attacks…



UNIVERSITY OF CAMBRIDGE

Cellular Network Leaks

- Mobile devices roam around and register with BTSs: they can be identified from these records and pre-existing location profiles. One paper identifies 80% of users.
- the network interface itself can allow listening attacks/leaks (GSM).

Performance

Table 5: Performance: all four identification periods

	Dec 2004	**Jan 2005**
one hour	40.91%	48.98%
one day	65.15%	65.31%
one week	74.24%	75.51%
one month	77.27%	87.76%

Yoni De Mulder, George Danezis, Lejla Batina, and Bart Preneel. 2008. Identification via location-profiling in GSM networks. In *Proceedings of the 7th ACM workshop on Privacy in the electronic society* (WPES '08). ACM, New York, NY, USA, 23-32.

WiFi based Leaks

- WLAN fingerprints can be used to infer social relation between the users.
- Mobile devices broadcast their Wi-Fi information that contain device ID or MAC address, it is possible for adversaries to track locations of devices.
- Users' names can also be detected by analyzing applications, websites and ad content in traffic data through WiFi hot-spots

WiFi based Leaks (2)

- By having WiFi connection info, applications can get the MAC address of devices, which is a unique ID. Since this ID is permanent, it allows third parties to track users.
- Applications can also learn about the last scanned list of WiFi hot spots. This includes MAC address, name, signal strength, operating channels and so on. This information can later be used to geo-locate user positions.
- It is also possible for applications to determine configured network lists on devices. Moreover, by comparing these lists, social relationship between individuals can also be inferred, such as professional, family, interest groups and the like.

Mobile Sensing Leaks

- Sensors are powerful (eg mood automatic recognition)
- Accelerometer is different on all devices (so it can be used as a "device signature").
- Or it can be used to identify different users of same device
- Touch sensor usage behaviour, keying behaviour can give away user identity

Mobile Sensing De-anonymization

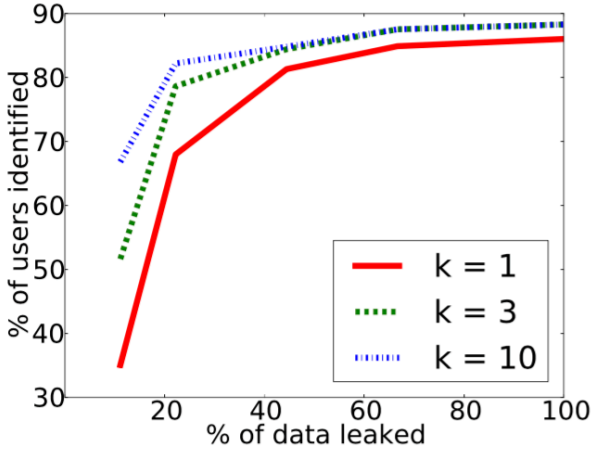
- Even in large anonymized datasets there is a risk that a user can be identified.
  - Example of the Netflix challenge
  - Mobile sensing datasets exhibit data "sparsity"
- With some user information it is possible to single out the user in a dataset that just contains activity profiles of users.

De-anonimyzation

- Netflix type datasets contains narrow range of behaviour (preferences of movies)
- Sensing data contain a wider range:
- Correlation of activities can be a "key"
  - Eg you go to the gym at a certain time after a train ride
  - This increases data sparsity
  - Broad range of auxiliary information which can be used by an adversary

Example

- Auxiliary information of the adversary is a collection of activities by a user.
- Aim: Identify which of the anonymized activity streams belongs to the target user.
- Auxiliary information may be collected by observing the user or from available public sources.



UNIVERSITY OF CAMBRIDGE

Lane et al 2012]

Location Data Privacy

by Tom Spring

October 20, 2016 , 9:48 am

UNIVERSITY OF CAMBRIDGE

Location Data Leaks

- Location aware apps collect data which is (sometimes) provided to third parties
  - Advertisement?
- Location data can be used to infer more personal information (or user identity from anonymous datasets)
- Profile of users can be built from solely location tracks

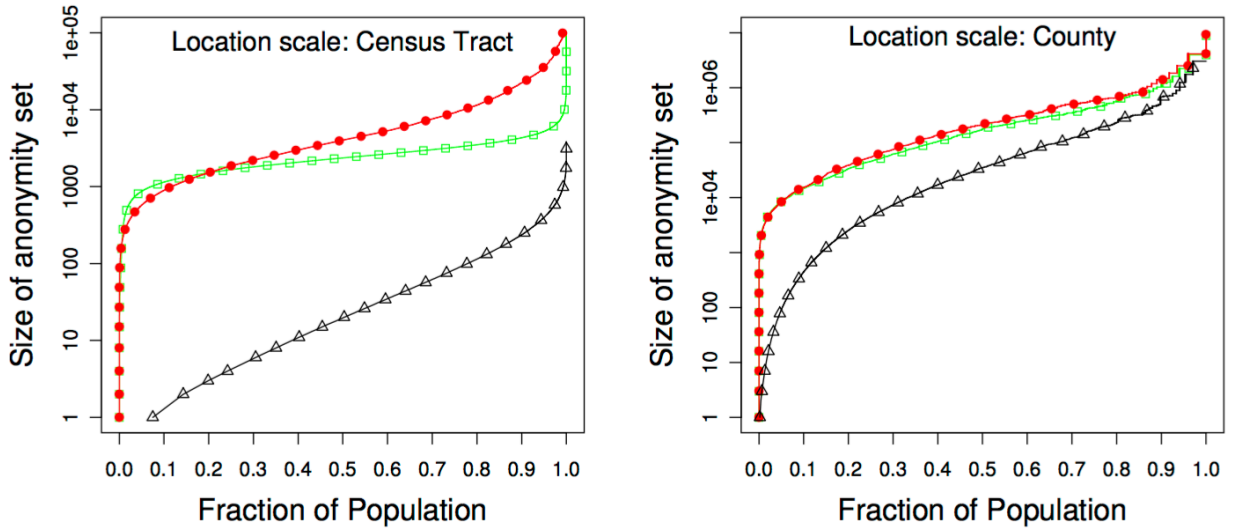
Examples

- GPS of 172 users find their home location with median error of 60m.
  - Features (last destination of day, long stay, time..)
- Similar features have been used to detect "significant places" for a user (work, gym etc)
- Speed of travelling/transport modality
- Data can be cross correlated with social network data for more deanonymization

[Golle et al 2014]

And more

- If the approximate locations of a user's home and workplace can be deduced from a location trace, then the median size of the individual's anonymity set in the U.S. working population is 1, 21 and 34,980, for locations known at the granularity of a census block, census tract and county respectively.



UNIVERSITY OF CAMBRIDGE

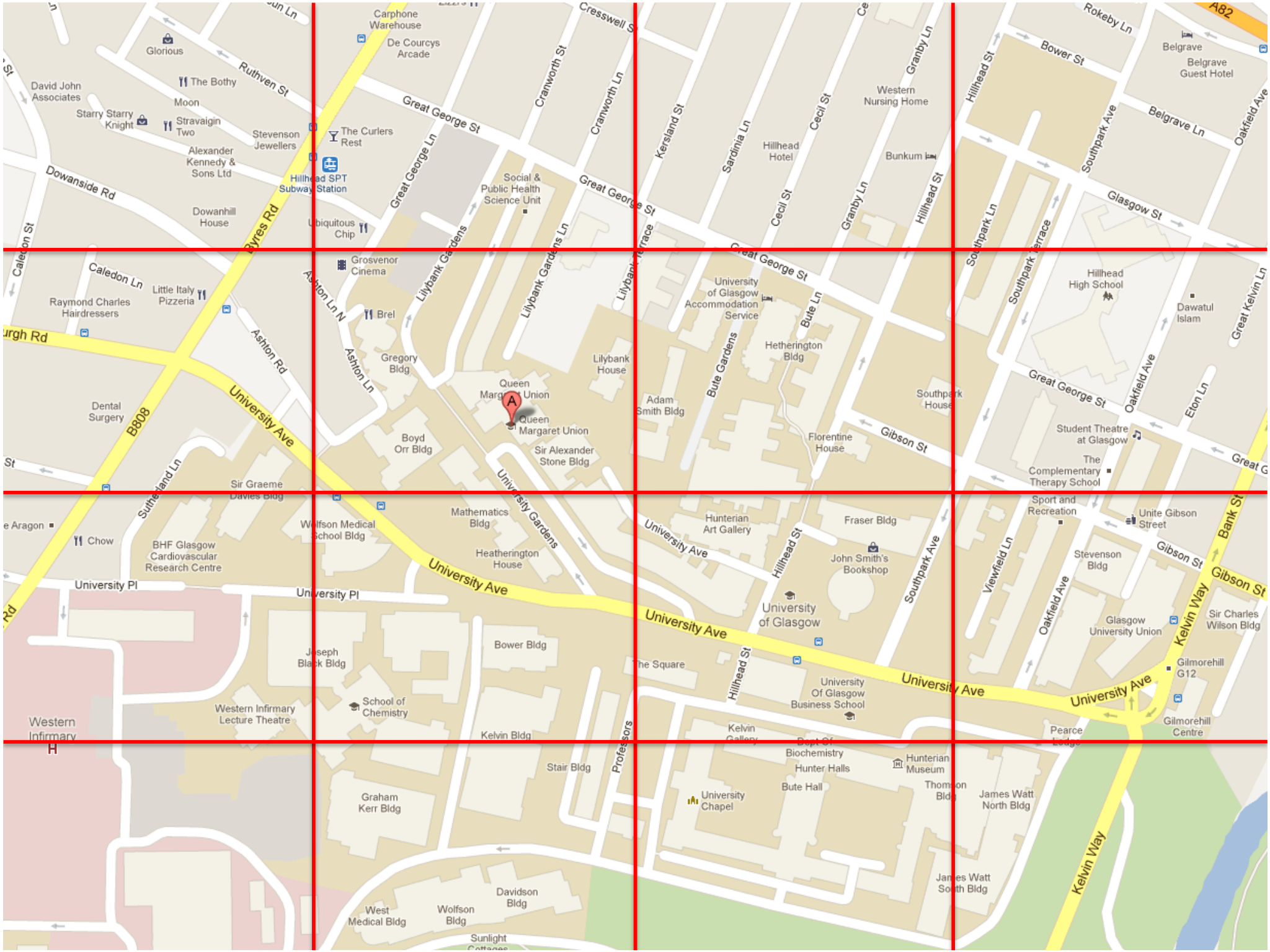
Fig. 1. Size of anonymity set under disclosure of work location (red circles), home location (green squares) or both (black triangles). Location granularity is either census tract (left graph) or county (right graph). Note the different scales on the Y-axes.

Mobility Prediction

- Given user patterns are highly regular, a user can not only be identified in a dataset but his previous patterns can be used to predict her future movements
- In addition her friends patterns if known can considerably aid this prediction

Different Types of Prediction

- Location can be:
  - Logical location (workplace, home)
  - Geographic location
    - Discrete areas (e.g., square in a grid)
    - GPS locations
- Not only *where* but also *when*
  - Spatio-temporal prediction is hard

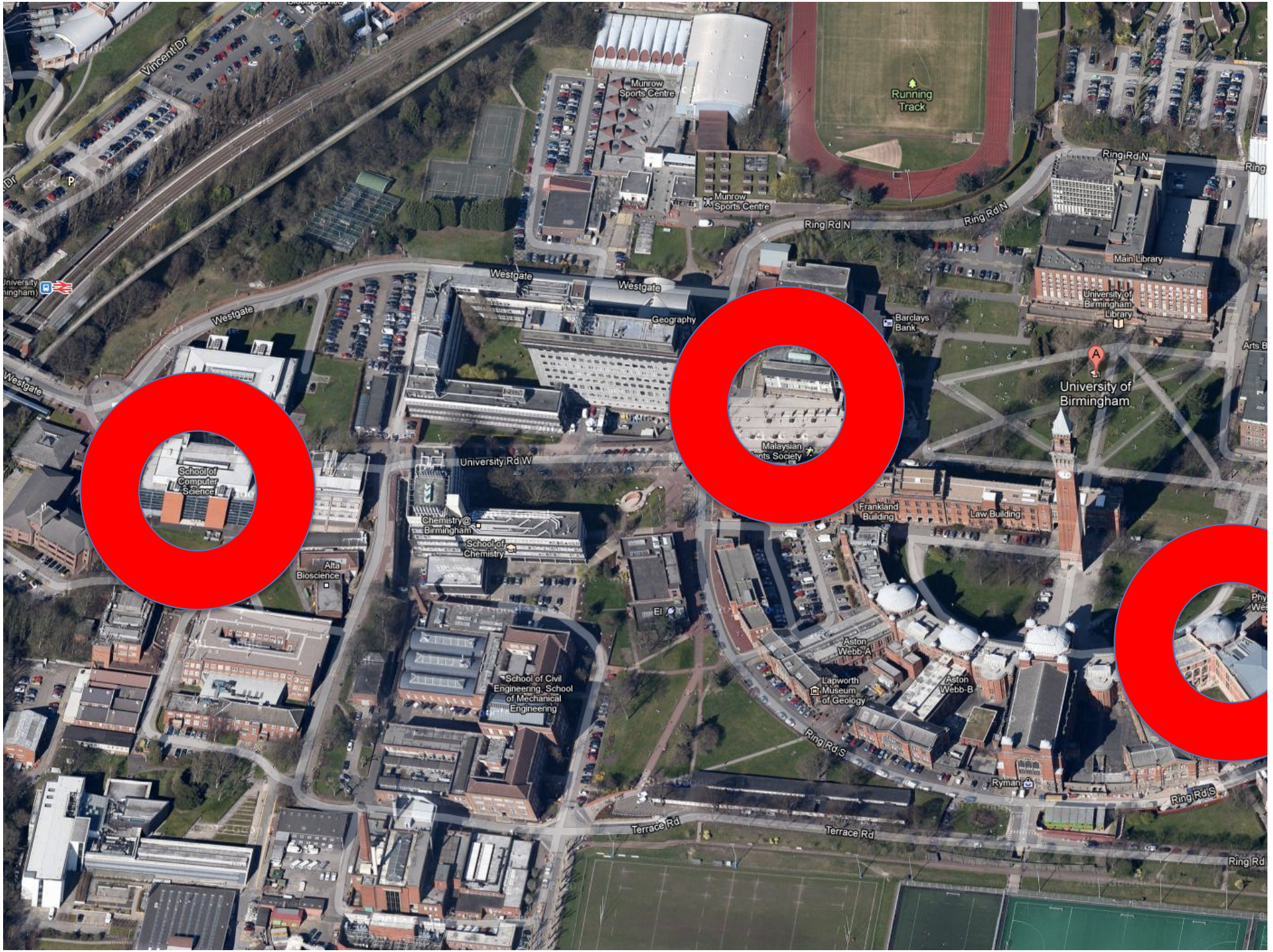


Carphone Warehouse
De Courcys Arcade
Glorious
The Bothy
Ruthven St
David John Associates
Moon
Starry Starry Knight
Stravaigin Two
Stevenson Jewellers
The Curlers Rest
Alexander Kennedy & Sons Ltd
Hillhead SPT Subway Station
Dowanside Rd
Dowanhill House
Ubiquitous Chip
Great George St
Great George Ln
Cranworth St
Cranworth Ln
Cresswell St
Social & Public Health Science Unit
Kersland St
Sardinia Ln
Hillhead Hotel
Cecil St
Granby Ln
Western Nursing Home
Bunkum
Hillhead St
Rokeby Ln
Bower St
Belgrave
Belgrave Guest Hotel
Belgrave Ln
Southpark Ave
Oakfield Ave
Glasgow St
Byres Rd
Caledon St
Caledon Ln
Little Italy Pizzeria
Raymond Charles Hairdressers
Grosvenor Cinema
Ashton Ln N
Brel
Lilybank Gardens
Lilybank Gardens Ln
Lilybank Terrace
University of Glasgow Accommodation Service
Bute Ln
Southpark Ln
Southpark Terrace
Hillhead High School
Dawatul Islam
Great Kelvin Ln
Ashton Rd
Ashton Ln
Gregory Bldg
Lilybank House
Hetherington Bldg
Bute Gardens
Queen Margaret Union
Adam Smith Bldg
Southpark House
Eton Ln
Dental Surgery
B808
University Ave
Boyd Orr Bldg
Sir Alexander Stone Bldg
Florentine House
Gibson St
Student Theatre at Glasgow
The Complementary Therapy School
Sutherland Ln
Sir Graeme Davies Bldg
University Gardens
Mathematics Bldg
Hunterian Art Gallery
Fraser Bldg
Sport and Recreation
Unite Gibson Street
Bank St
Chow
BHF Glasgow Cardiovascular Research Centre
Wolfson Medical School Bldg
Heatherington House
John Smith's Bookshop
Viewfield Ln
Stevenson Bldg
University Pl
University of Glasgow
Glasgow University Union
Kelvin Way
Sir Charles Wilson Bldg
Joseph Black Bldg
Bower Bldg
The Square
Gilmorehill G12
University Of Glasgow Business School
Western Infirmary Lecture Theatre
School of Chemistry
Western Infirmary
Kelvin Bldg
Kelvin Gallery
Pearce Lodge
Gilmorehill Centre
Professors
Dept Of Biochemistry
Hunter Halls
Hunterian Museum
Stair Bldg
Bute Hall
Thomson Bldg
James Watt North Bldg
Graham Kerr Bldg
University Chapel
James Watt South Bldg
Davidson Bldg
West Medical Bldg
Wolfson Bldg
Sunlight Cottages
A82

Transition Matrix

$$T = \begin{pmatrix} 0.1 & 0.3 & 0.2 & 0.4 \\ 0.4 & 0.2 & 0.15 & 0.25 \\ 0.1 & 0.3 & 0.4 & 0.2 \\ 0.1 & 0.2 & 0.2 & 0.5 \end{pmatrix}$$

UNIVERSITY OF CAMBRIDGE


Vincent Dr
Munrow Sports Centre
Running Track
Ring Rd N
Main Library
University of Birmingham Library
Barclays Bank
Westgate
Geography
University of Birmingham
Malaysian Society
School of Computer Science
University Rd W
Chemistry @ Birmingham
School of Chemistry
Alta Bioscience
Frankland Building
Law Building
Aston Webb-A
Aston Webb-B
Lapworth Museum of Geology
School of Civil Engineering, School of Mechanical Engineering
Ryman
Ring Rd S
Terrace Rd

Various Prediction Techniques

- Techniques forecast
  - the next location of a user
  - his/her arrival and residence time, i.e., the interval of time spent at that location.

UNIVERSITY OF CAMBRIDGE

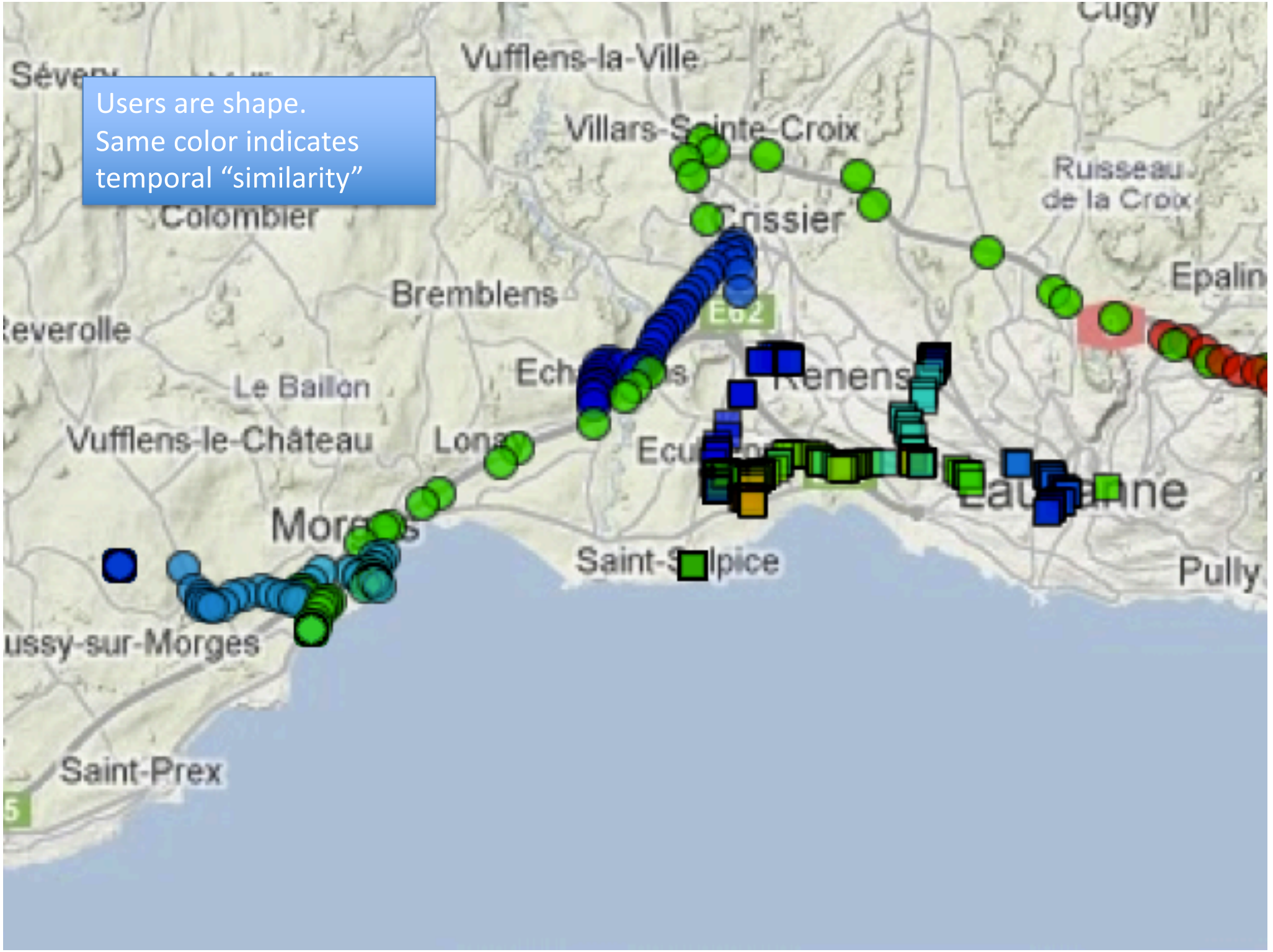
[Scellato et al 13]

Prediction through correlated movement

- Analysis of the correlation of mobility patterns of the users
- Movement correlation is measured through **mutual information**
- The resulting network can be considered as a network of "movements"

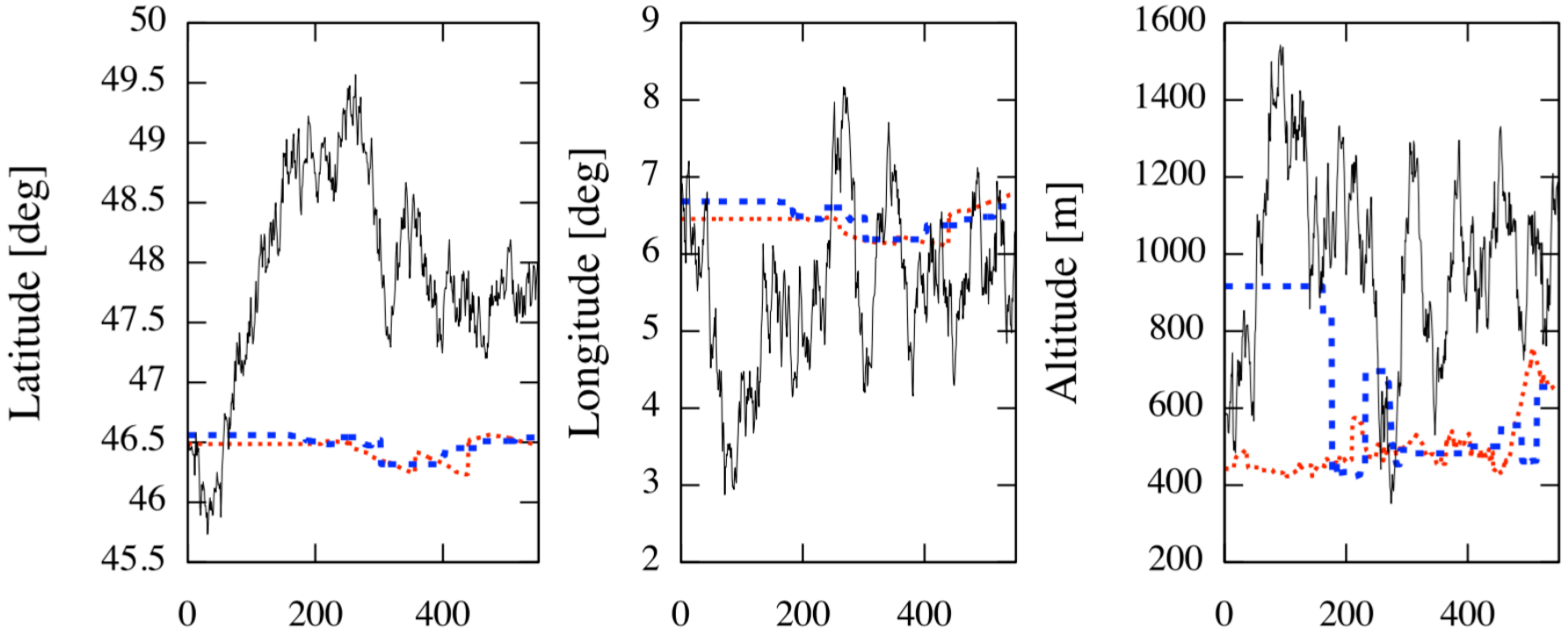
[De Domenico et al 2011]

UNIVERSITY OF CAMBRIDGE


Users are shape.
Same color indicates
temporal “similarity”
Vufflens-la-Ville
Villars-Sainte-Croix
Colombier
Crissier
Bremblens
Le Baillon
Vufflens-le-Château
Renens
Saint-Prex
Pully

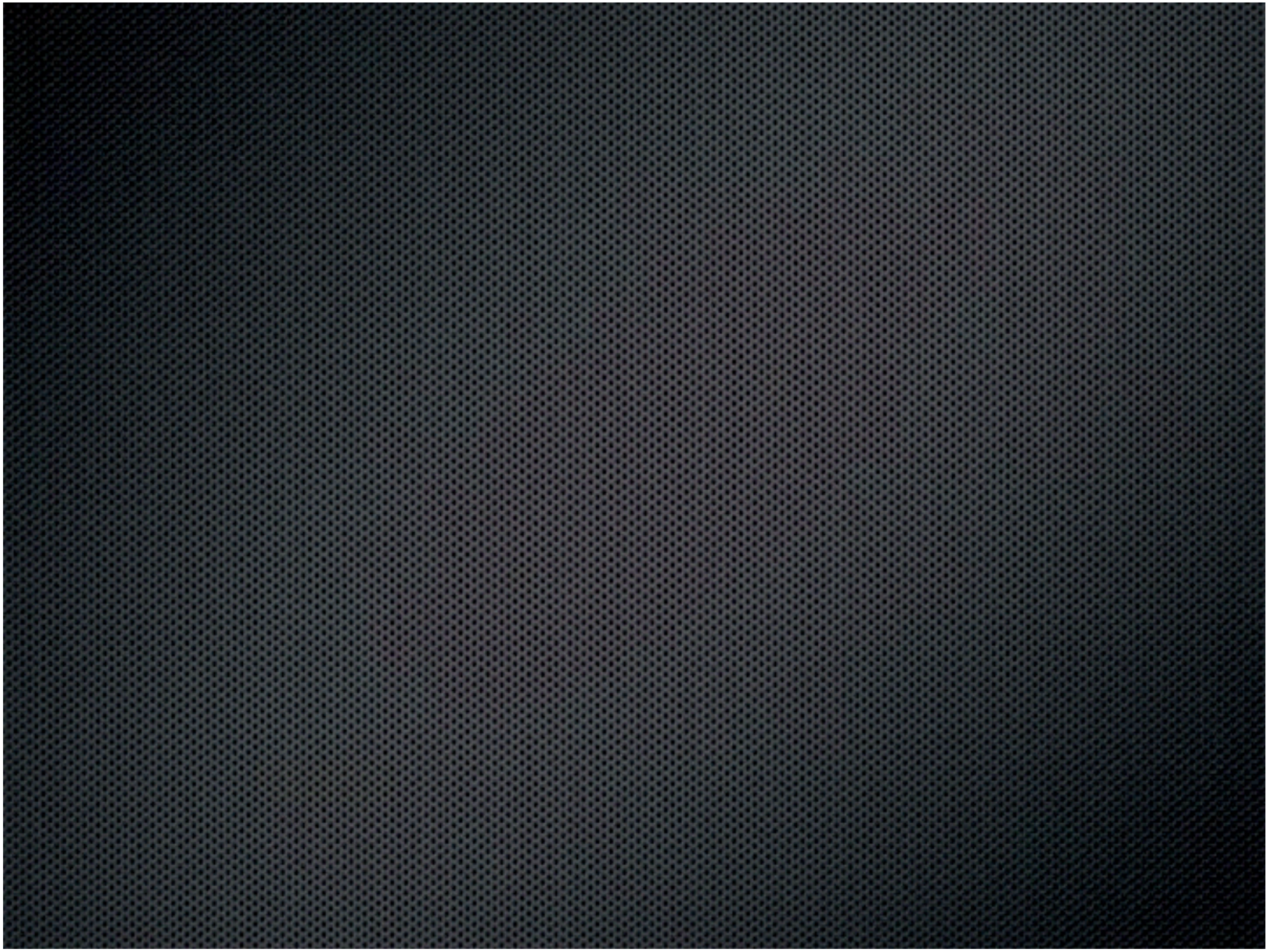
Jouxtens-Mézery
Bussigny-près-Lausanne
Bremblens
9
5
E62
1
E25
Renens
Lonay
Denges
Le Bief
Préverenges
Saint-Sulpice

Example



Blue dotted line is prediction with multivariate non linear predictor
Red line real data
Black line ARMA linear modelling

UNIVERSITY OF CAMBRIDGE



Summary

- We have described various privacy issues related to mobile devices and shown how they could be addressed.
- Challenges we discussed were
  - Systems related
  - Sensing related

References

- Muhammad Haris, Hamed Haddadi, Pan Hui, "Privacy Leakage in Mobile Computing: Tools, Methods, and Characteristics" https://arxiv.org/pdf/1410.4978v1.pdf
- W. Enck, P. Gilbert, B.-G. Chun, L. P. Cox, J. Jung, P. McDaniel, and A. N. Sheth, "Taintdroid: an information flow tracking system for real-time privacy monitoring on smartphones," *Communications of the ACM*, vol. 57, no. 3, pp. 99–106, 2014
- Golle, Philippe, and Kurt Partridge. "On the anonymity of home/work location pairs." International Conference on Pervasive Computing. Springer Berlin Heidelberg, 2009.
- N. D. Lane, J. Xie, T. Moscibroda, and F. Zhao, "On the feasibility of user de-anonymization from shared mobile sensor data," in *Proceedings of the Third International Workshop on Sensing Applications on Mobile Phones*. ACM, 2012.
- Salvatore Scellato, Mirco Musolesi, Cecilia Mascolo, Vito Latora and Andrew T. Campbell. NextPlace: A Spatio-temporal Prediction Framework for Pervasive Systems. Proceedings of the 9th International Conference on Pervasive Computing (Pervasive'11). San Francisco, California, USA. June 2011
- Manlio De Domenico, Antonio Lima and Mirco Musolesi, "Interdependence and Predictability of Human Mobility and Social Interactions", In Pervasive and Mobile Computing. Volume 9. Issue 6. December 2013. Elsevier.
- Video: https://www.cs.bham.ac.uk/research/projects/nsl/mobility-prediction/