Mobile and Sensor Systems

Lecture 7: Mobile Privacy

Prof Cecilia Mascolo

UNIVERSITY OF
CAMBRIDGE

Mobile and Sensor Privacy

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

UNIVERSITY OF
CAMBRIDGE

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

UNIVERSITY OF CAMBRIDGE

App Data Gathering

- More often than not apps collect sensing data beyond its technical needs (for advertising purposes)
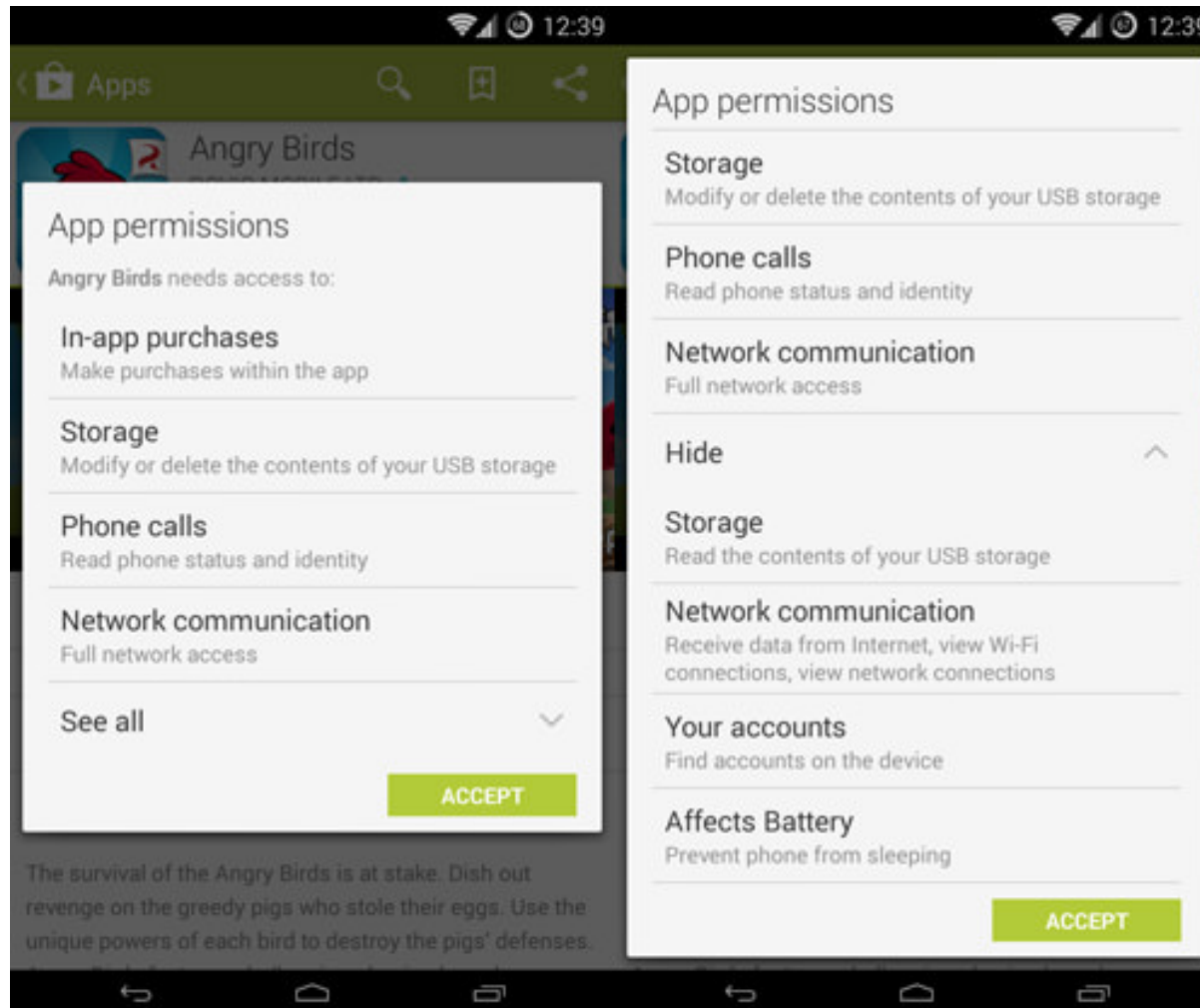
Where does the data go?

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

    - The OS does not offer support to know this…

UNIVERSITY OF CAMBRIDGE

Permissions: Android

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 could at some point have sharedUserID: apps by same developer have same ID so permissions for that ID were shared

UNIVERSITY OF CAMBRIDGE

[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

Static Analysis

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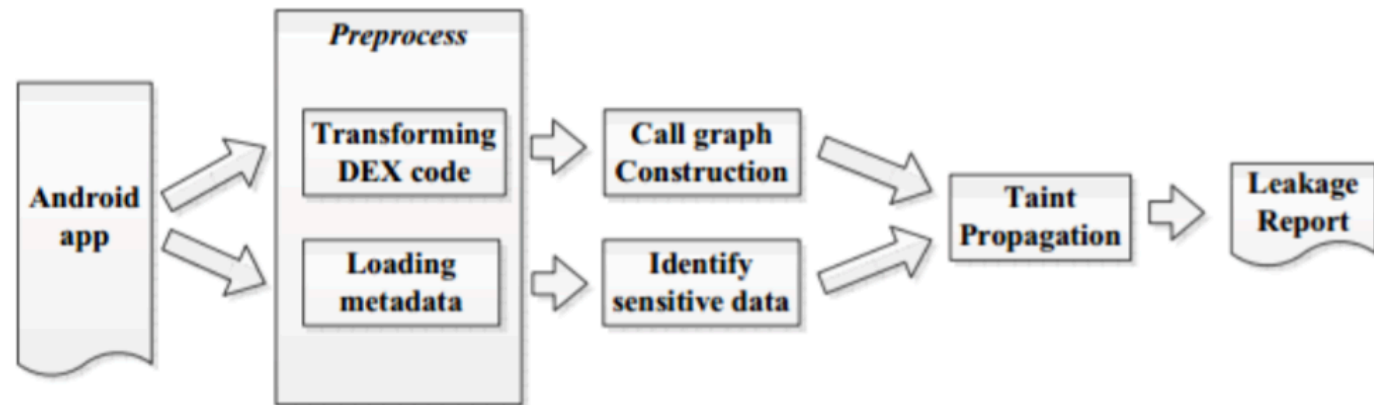
Dynamic Analysis: TaintDroid

- **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

UNIVERSITY OF CAMBRIDGE

[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.

Cellular Network Leaks

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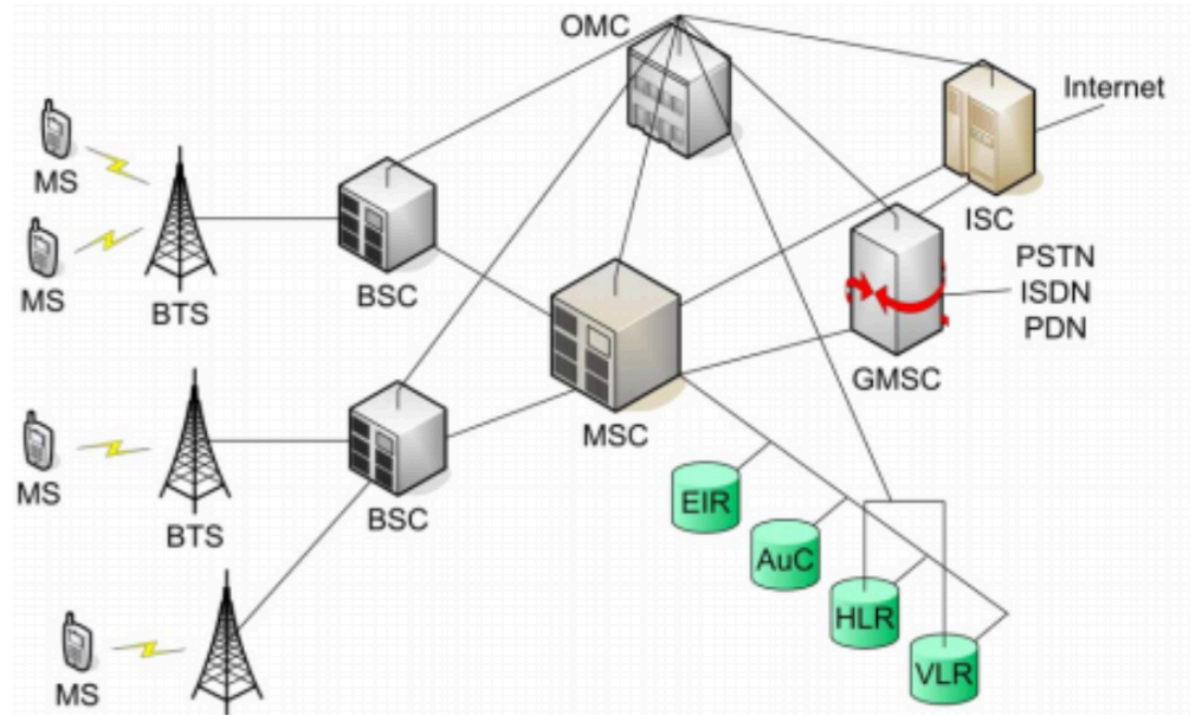
Cellular Network Leaks

- 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 could in the past allow listening attacks/leaks (GSM).

UNIVERSITY OF
CAMBRIDGE

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.

UNIVERSITY OF CAMBRIDGE

Location Data Privacy

Mobile Applications Leak Device, Location Data

MOBILE APPLICATIONS LEAK DEVICE, LOCATION DATA

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

UNIVERSITY OF CAMBRIDGE

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

UNIVERSITY OF
CAMBRIDGE

[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.
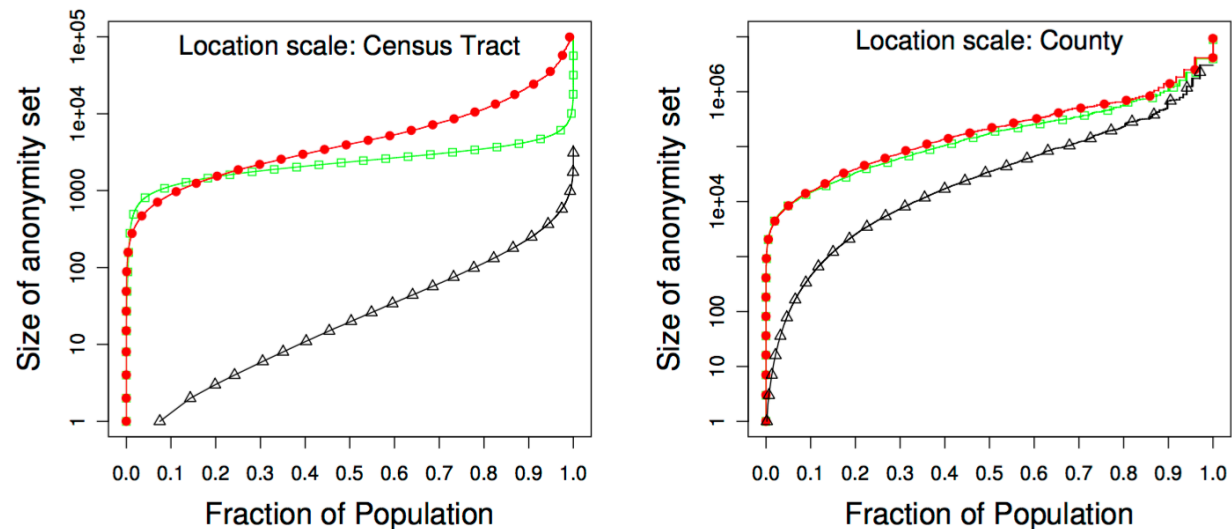

Fig. 1. Size of anonymity set under disclosure of work location (red circles), home location (green squares) or both (black triangles). Location granularity is either census tract (left graph) or county (right graph). Note the different scales on the Y-axes.

UNIVERSITY OF CAMBRIDGE

WiFi based Leaks

- WLAN fingerprints can be used to infer social relation between the users. Applications could 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 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.

UNIVERSITY OF
CAMBRIDGE

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

UNIVERSITY OF CAMBRIDGE

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.
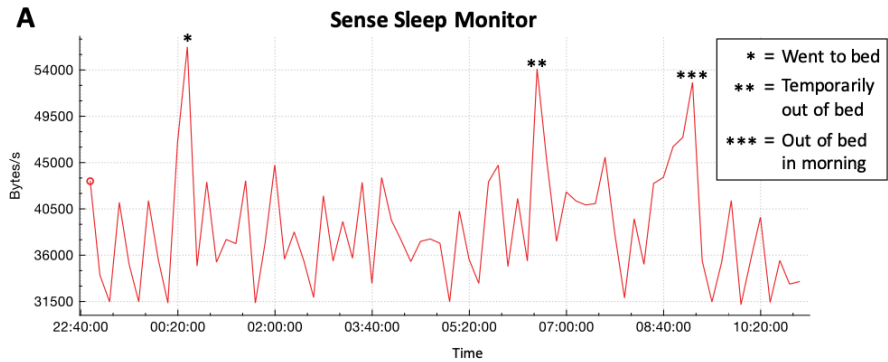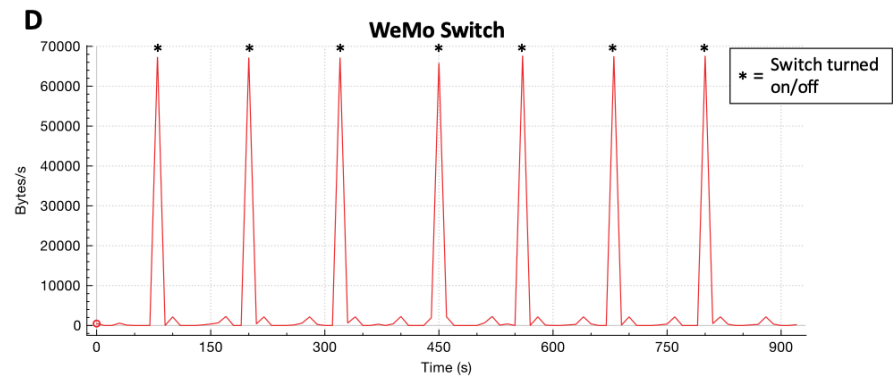

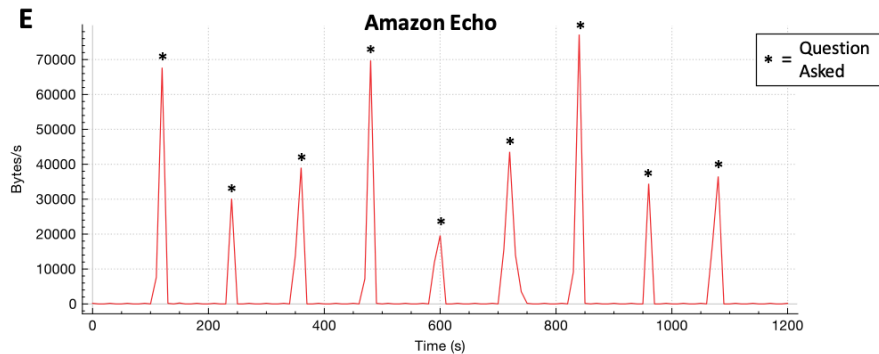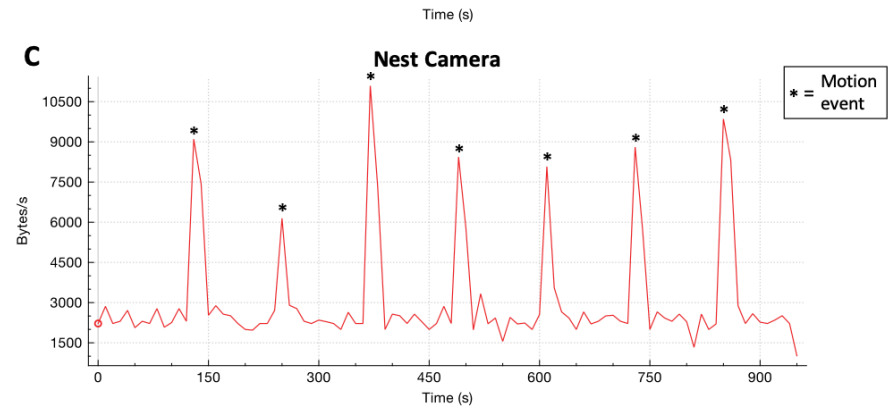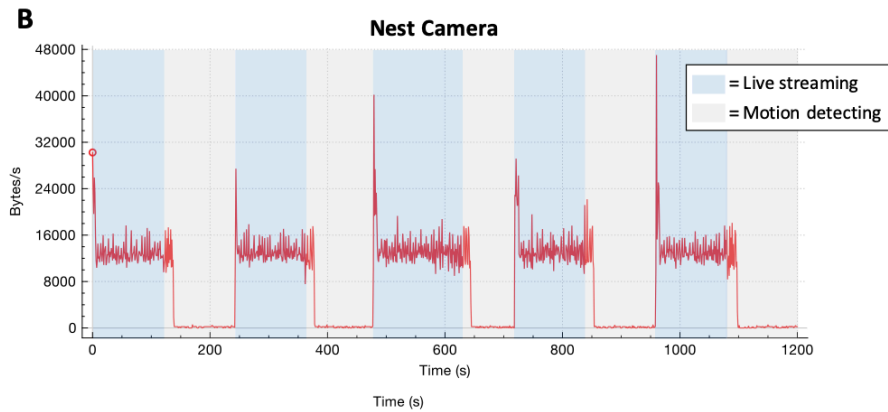UNIVERSITY OF CAMBRIDGE

[Lane et al 2012]
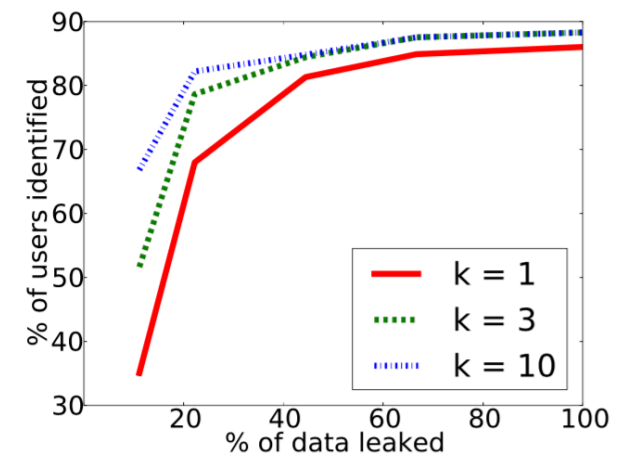
IoT Devices…

- IoT devices in our homes sense our behaviour constantly.

Home Devices' Patterns

"encryption alone does not provide adequate privacy protection"

IoT network addresses can be revealing...

"Our findings indicate that millions of IoT devices are detectable and identifiable within hours, both at a major ISP as well as an IXP, using passive, sparsely sampled network flow headers. Our methodology is able to detect devices from more than 77% of the studied IoT manufacturers, including popular devices such as smart speakers "
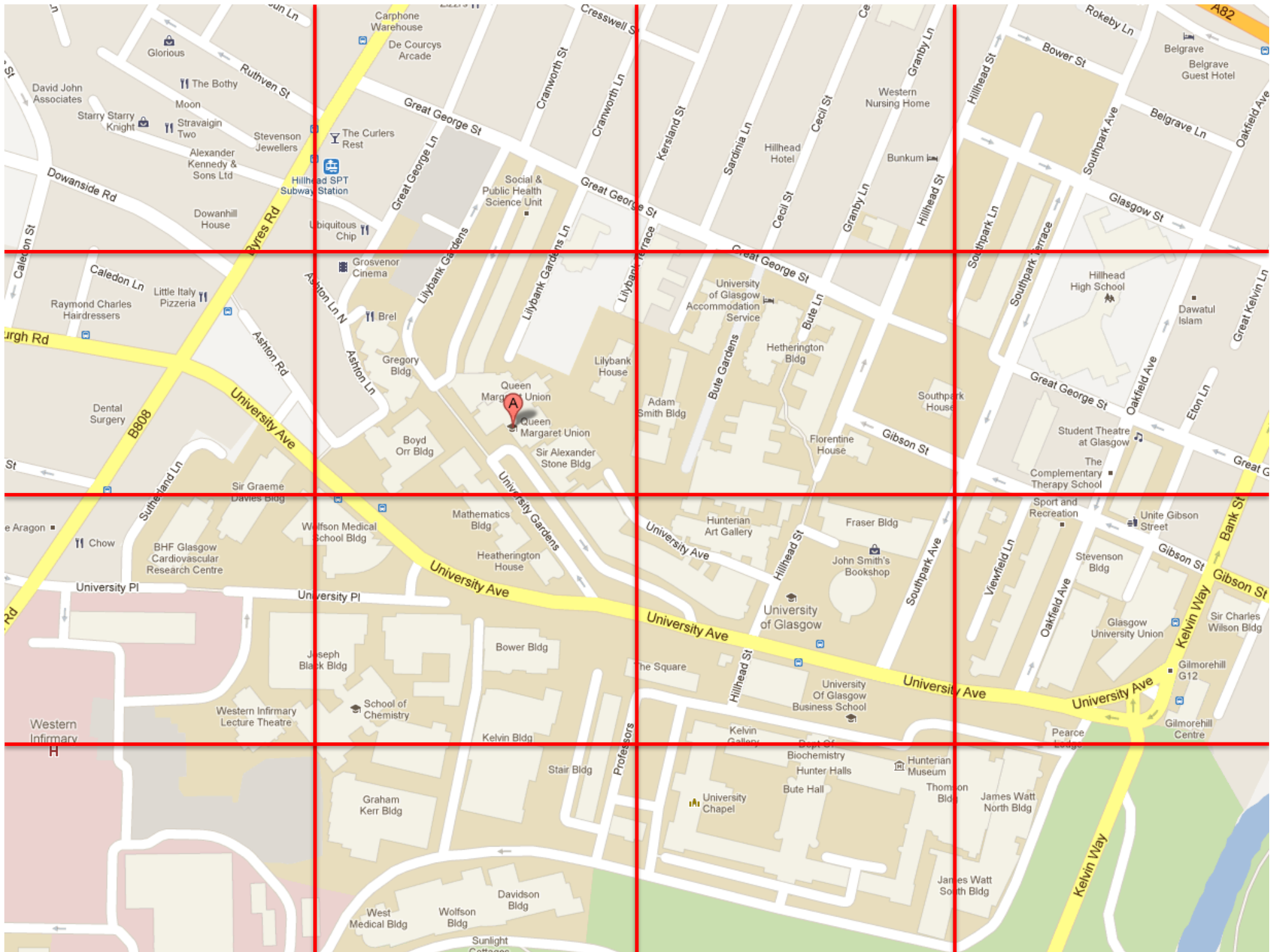

UNIVERSITY OF CAMBRIDGE
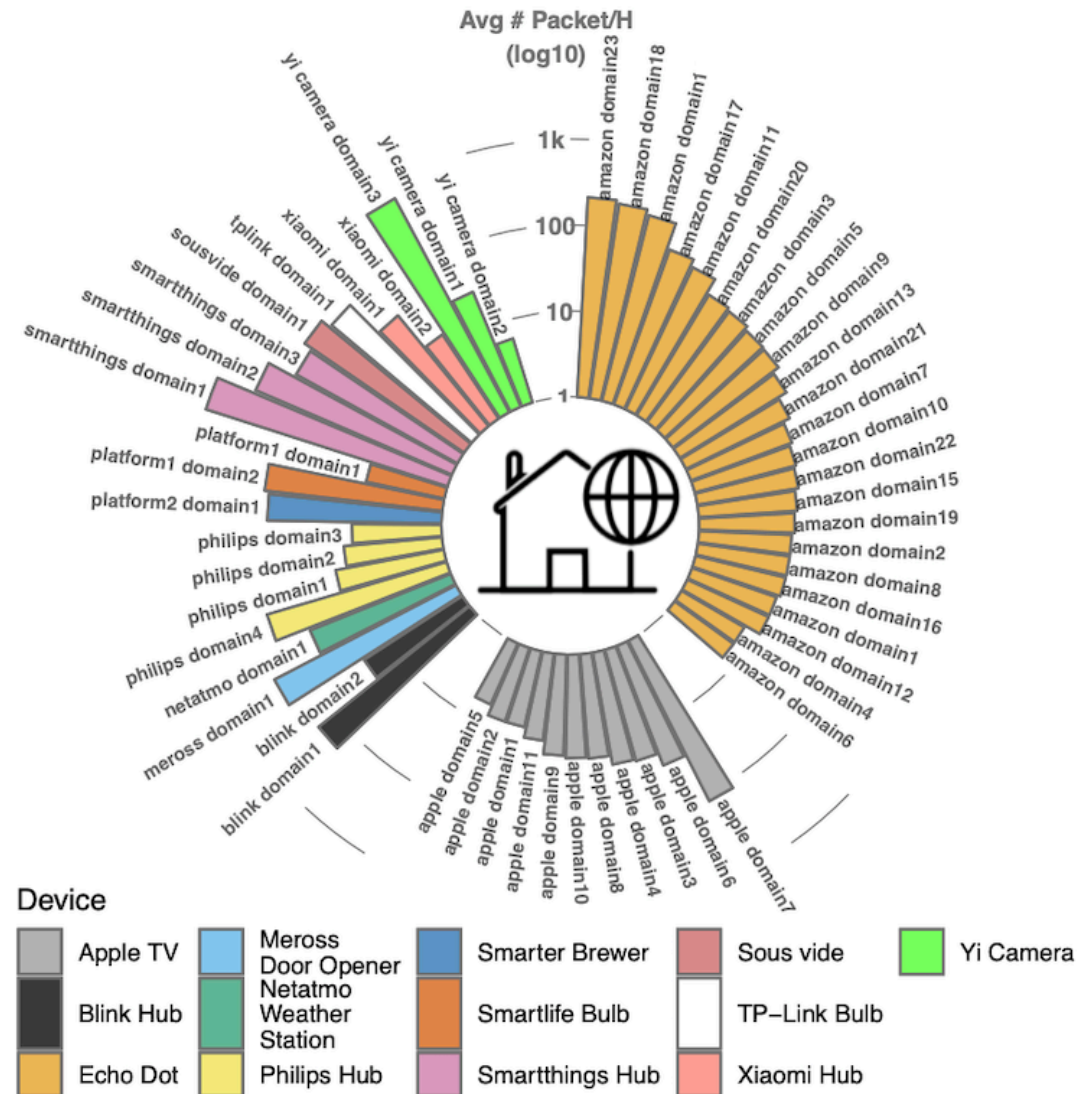
Solutions?

- Recognizing the problem is a good start…

- Traffic patterns obfuscation should help

- Technical solution to this is not hard
  - Flighting the business model: governance?

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

UNIVERSITY OF
CAMBRIDGE

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

Rokeby Ln

A82

Glorious
Carphone Warehouse
De Courcys Arcade
Cresswell S
Bower St
Belgrave
Belgrave Guest Hotel

The Bothy
Moon
Ruthven St
Great George St
Cranworth St
Cranworth Ln
Western Nursing Home
Hillhead St
Belgrave Ln
Oakfield Ave

David John Associates
Starry Starry Knight
Stravaigin Two
Alexander Kennedy & Sons Ltd
Stevenson Jewellers
The Curlers Rest
Great George Ln
Kersland St
Sardinia Ln
Cecil St
Hillhead Hotel
Granby Ln
Bunkum
Hillhead St
Southpark Ave
Southpark Terrace

Dowanside Rd
Byres Rd
Hillhead SPT Subway Station
Great George St
Cecil St
Great George St
Glasgow St

Dowanhill House
Ubiquitous Chip
Social & Public Health Science Unit
Great George St
Southpark Ln

Caledon St
Caledon Ln
Little Italy Pizzeria
Grosvenor Cinema
Lilybank Gardens
Lilybank Gardens Ln
Lilybank Terrace
University of Glasgow Accommodation Service
Bute Ln
Hillhead High School
Dawatul Islam
Great Kelvin Ln

Raymond Charles Hairdressers
Ashton Ln N
Brel
Gregory Bldg
Lilybank House
Hetherington Bldg
Great George St

Ashton Rd
Ashton Ln
Queen Margaret Union
Bute Gardens
Southpark House
Oakfield Ave
Eton Ln

urgh Rd
University Ave
Dental Surgery
A
Queen Margaret Union
Adam Smith Bldg
Florentine House
Student Theatre at Glasgow
Great G

Sutherland Ln
Boyd Orr Bldg
Sir Alexander Stone Bldg
Gibson St
The Complementary Therapy School

B808
Sir Graeme Davies Bldg
Mathematics Bldg
University Gardens
Hunterian Art Gallery
Fraser Bldg
Sport and Recreation
Unite Gibson Street

e Aragon
Chow
BHF Glasgow Cardiovascular Research Centre
Wolfson Medical School Bldg
Heatherington House
University Ave
Hillhead St
John Smith's Bookshop
Southpark Ave
Viewfield Ln
Stevenson Bldg
Gibson St
Bank St

University Pl
University Pl
University Ave
University of Glasgow
Oakfield Ave
Gilmorehill G12
Gibson St

Rd
Joseph Black Bldg
Bower Bldg
The Square
Hillhead St
University Of Glasgow Business School
University Ave
Glasgow University Union
Sir Charles Wilson Bldg

Western Infirmary Lecture Theatre
School of Chemistry
Kelvin Bldg
Kelvin Gallery
Pearce Lodge
Kelvin Way
Gilmorehill Centre

Western Infirmary
H
Graham Kerr Bldg
Stair Bldg
Professors
Dept Of Biochemistry
Hunter Halls
Bute Hall
Hunterian Museum
Thomson Bldg
James Watt North Bldg

West Medical Bldg
Wolfson Bldg
Davidson Bldg
University Chapel
Sunlight Cottages
James Watt South Bldg
Kelvin Way

Transition Matrix

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

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.

[Scellato et al 13]

UNIVERSITY OF
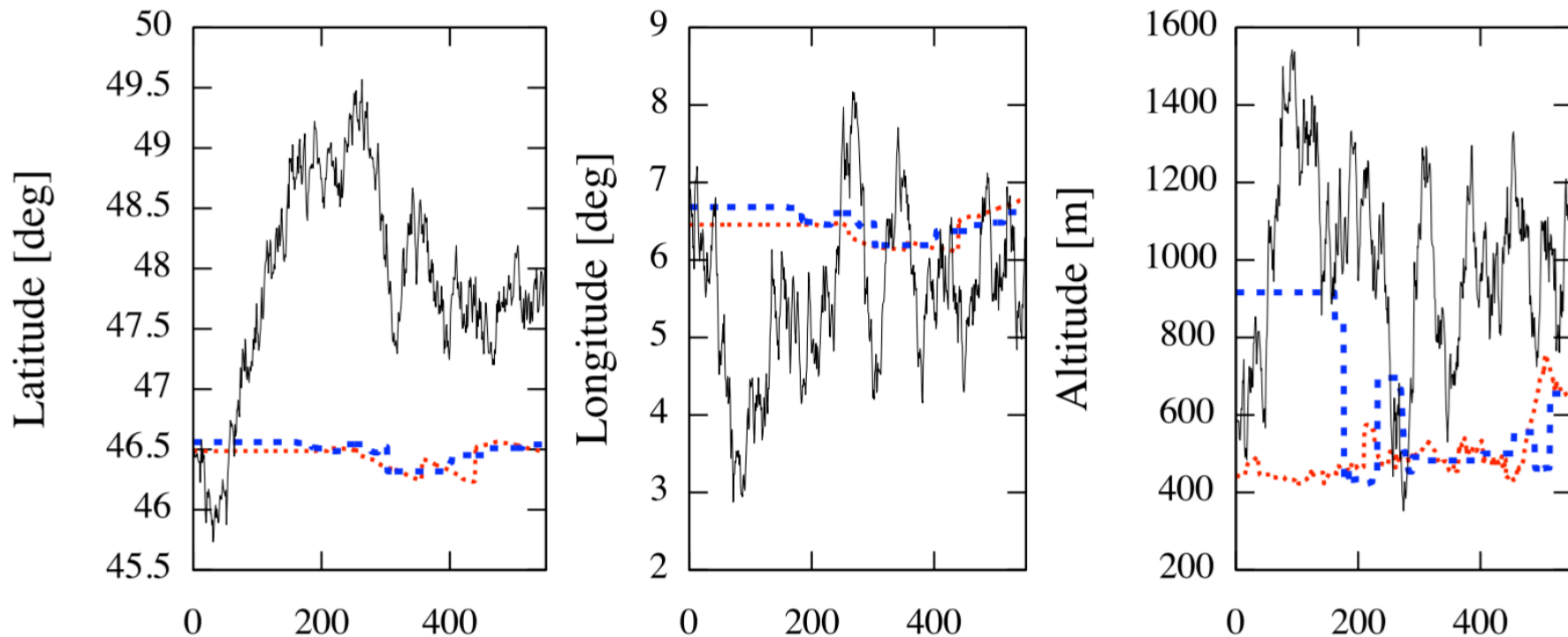CAMBRIDGE

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"
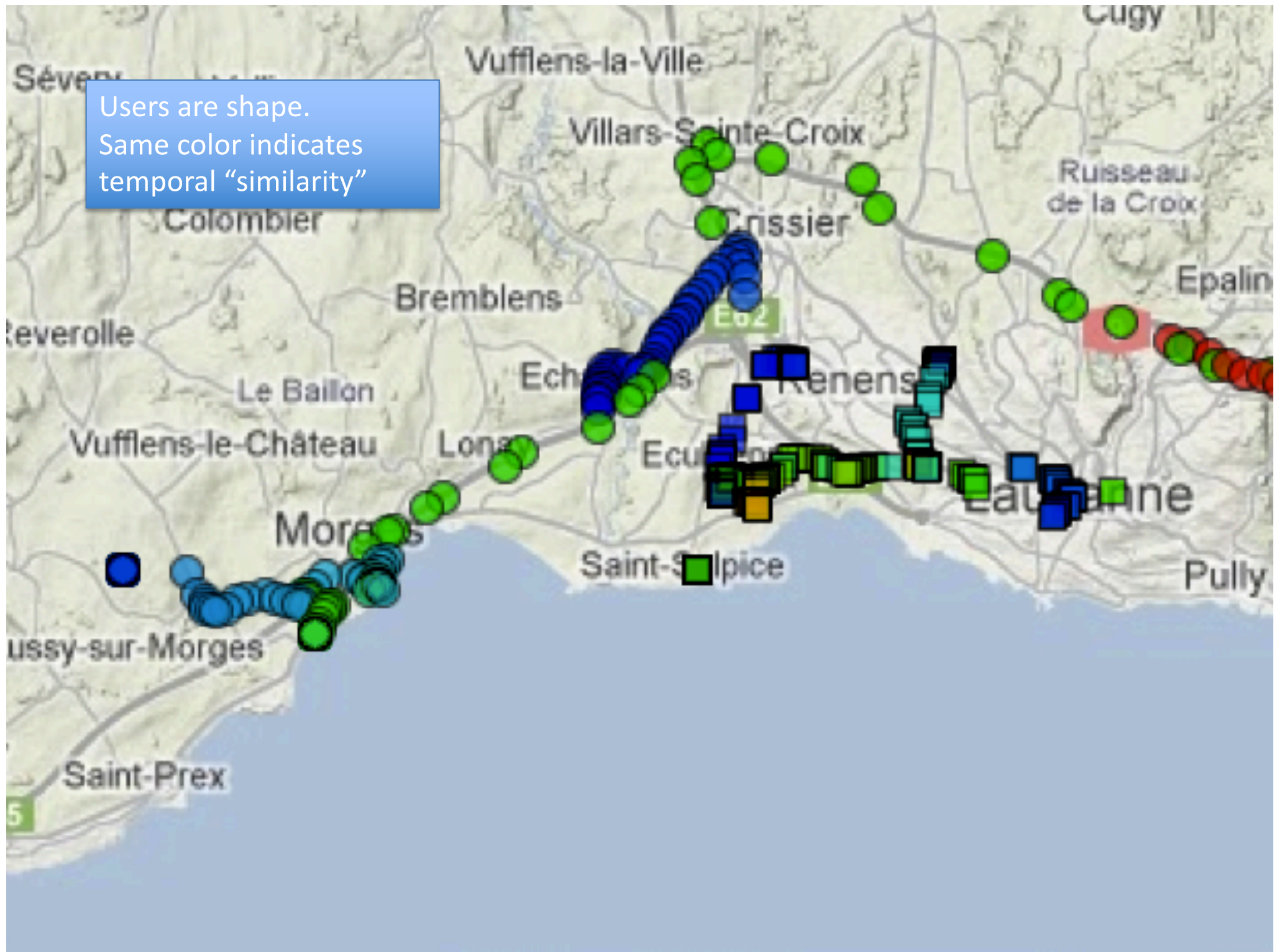
Example

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

Video…

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 (1)

- 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.

- Said Jawad Saidi, Anna Maria Mandalari, Roman Kolcun, Hamed Haddadi, Daniel J. Dubois, David, Choffnes, Georgios Smaragdakis, Anja Feldmann. A Haystack Full of Needles: Scalable Detection of IoT Devices in the Wild. nternet Measurement Conference (IMC) 2020

UNIVERSITY OF CAMBRIDGE

References (2)

- Noah Apthorpe, Dillon Reisman, Nick Feamster. A Smart Home is No Castle: Privacy Vulnerabilities of Encrypted IoT Traffic. https://arxiv.org/abs/1705.06805

- 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/

UNIVERSITY OF CAMBRIDGE