You Can Hide, But Your Periodic Schedule Can't

Minghua Ma, Kai Zhao, Kaixin Sui, Lei Xu, Yong Li, Dan Pei

清華大學

Tsinghua University

Trajectory in a Large-scale Wi-Fi Network

Wi-Fi access point

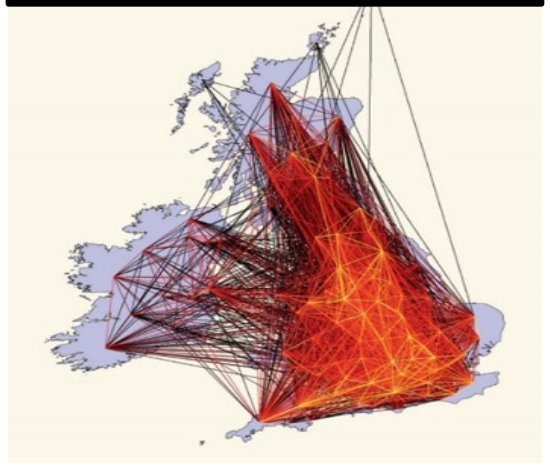
Tsinghua Campus

Trajectory in a Large-scale Wi-Fi Network

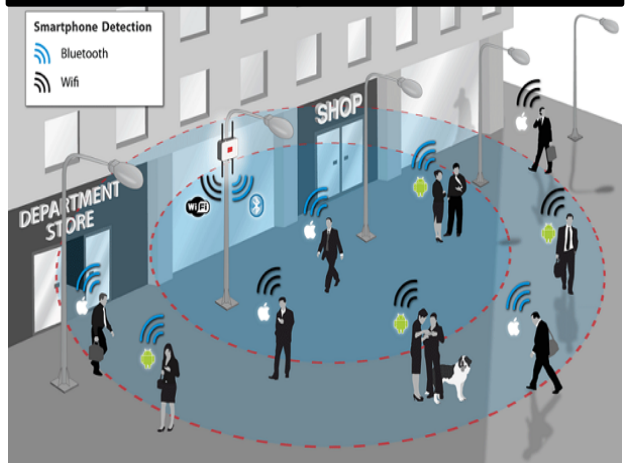
Trajectory in a Large-scale Wi-Fi Network

Value of Trajectory Datasets

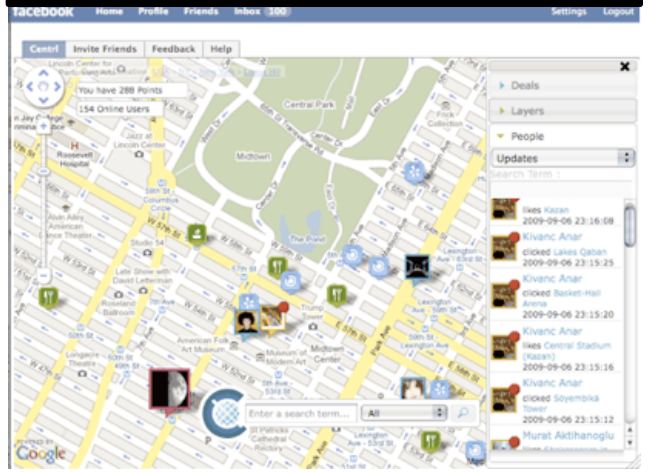
Mobility pattern mining



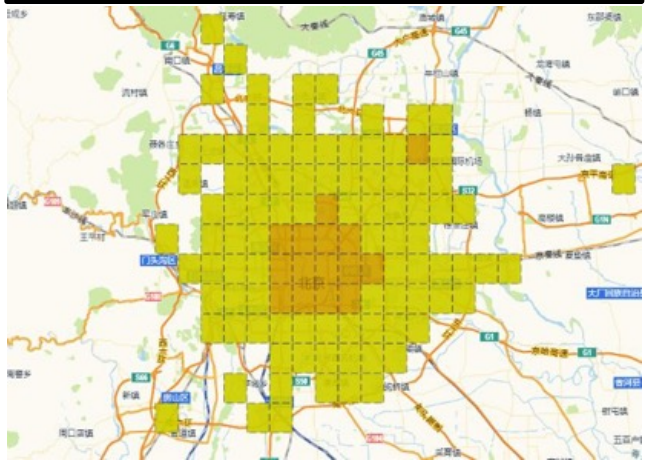
Proximity marketing



Location-based service



Transportation control



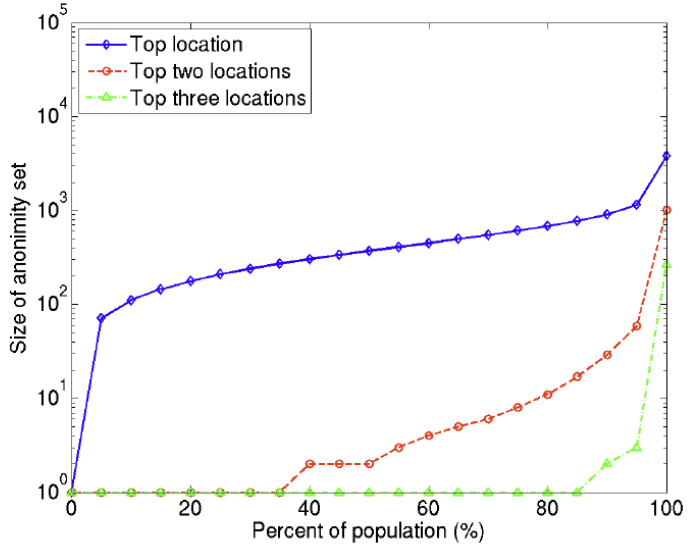
Privacy Issue of Trajectory datasets

Mobile Trajectories are **highly unique**

→ **Potential re-identification**

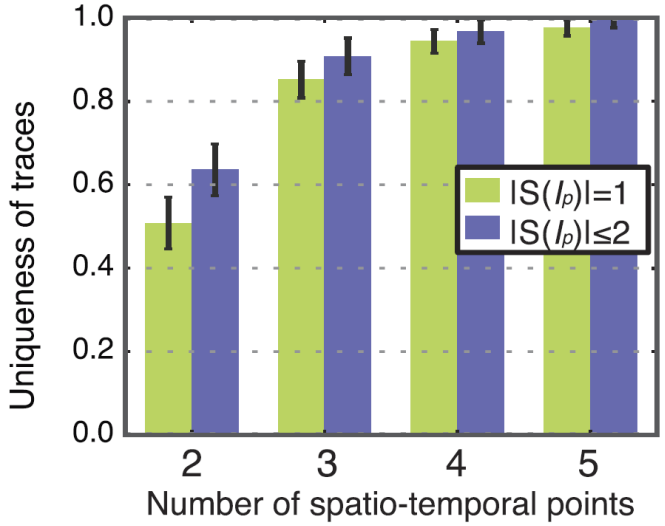
→ **Privacy risk**

MobiCom 2011



Three top locations pinpoint 50% users

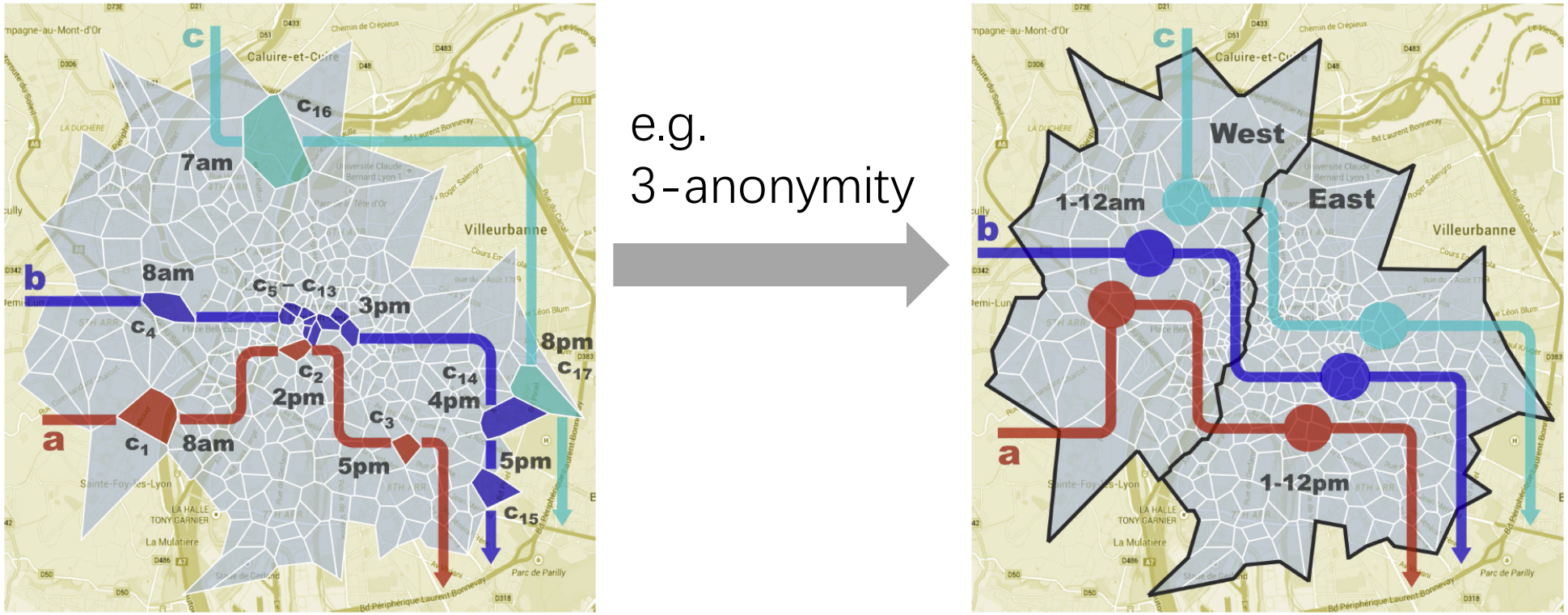
Nature Scientific Reports 2013



Five random points pinpoint 95% users

Counter-measure: K-anonymity

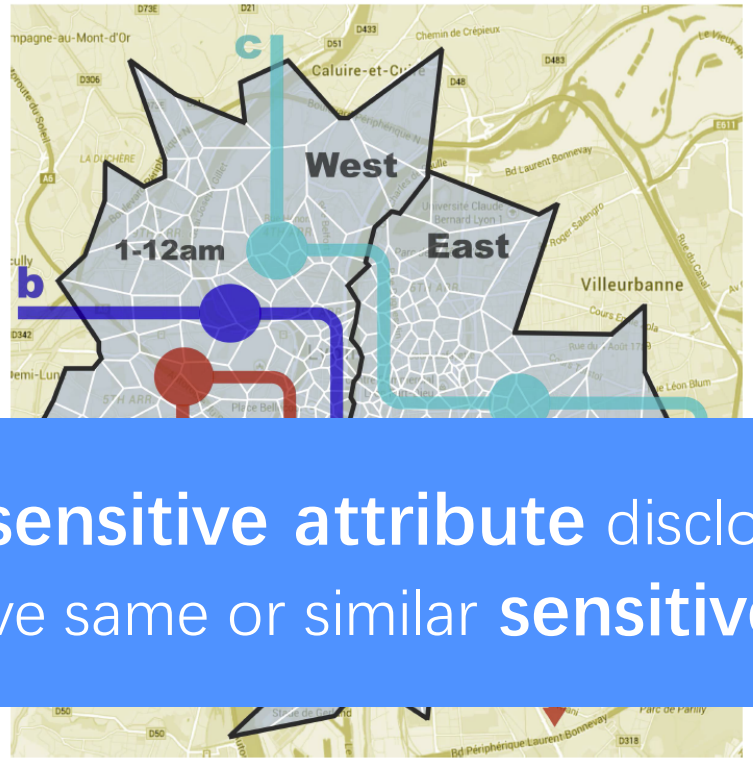
- **Key idea: each user should be indistinguishable from at least k − 1 others (hidden in an anonymity set no smaller than k)**



Counter-measure: K-anonymity

- **Key idea: each user should be indistinguishable from at least k − 1 others (hidden in an anonymity set no smaller than k)**

e.g.
3-anonymity



However, k-anonymity cannot prevent **sensitive attribute** disclosure
Because users in the same anonymity set may have same or similar **sensitive attributes**

Example

The adversary

knows his **three spatiotemporal samples** e.g. from social networks

and wants to know his **periodic schedule** e.g. curriculum schedule

Example

The adversary

finds **four** people with the same **three spatiotemporal samples**

Trajectory Dataset

Example

The adversary

cannot identify the target **(thanks to 4-anonymity!)**

Example

The adversary

cannot identify the target (thanks to 4-anonymity!)

But, what if these people all attend class at **Monday 8:00-9:30 in Building 6 ?**

Trajectory Dataset

Problem

The adversary

low diversity of the sensitive attributes

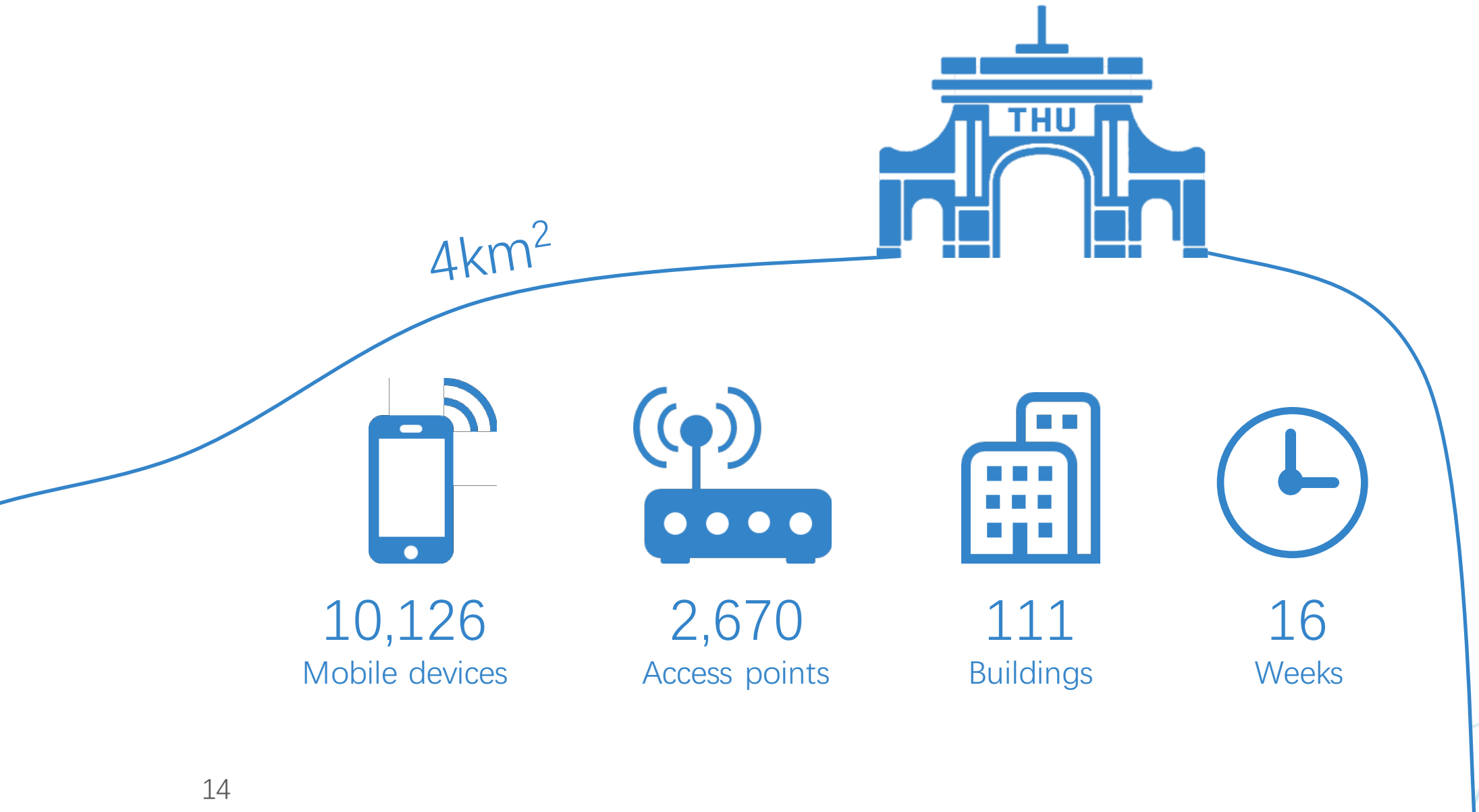
But, what if these people all attend class at **Monday 8:00-9:30 in 6 Jiao?**

We present a large-scale measurement study to answer:

- What is the schedule-leakage risk?
- How to evaluate the schedule-leakage risk?
- How serious is the schedule-leakage risk in the campus?

Trajectory Dataset

Wi-Fi Based Trajectory Dataset from Tsinghua University



10,126
Mobile devices

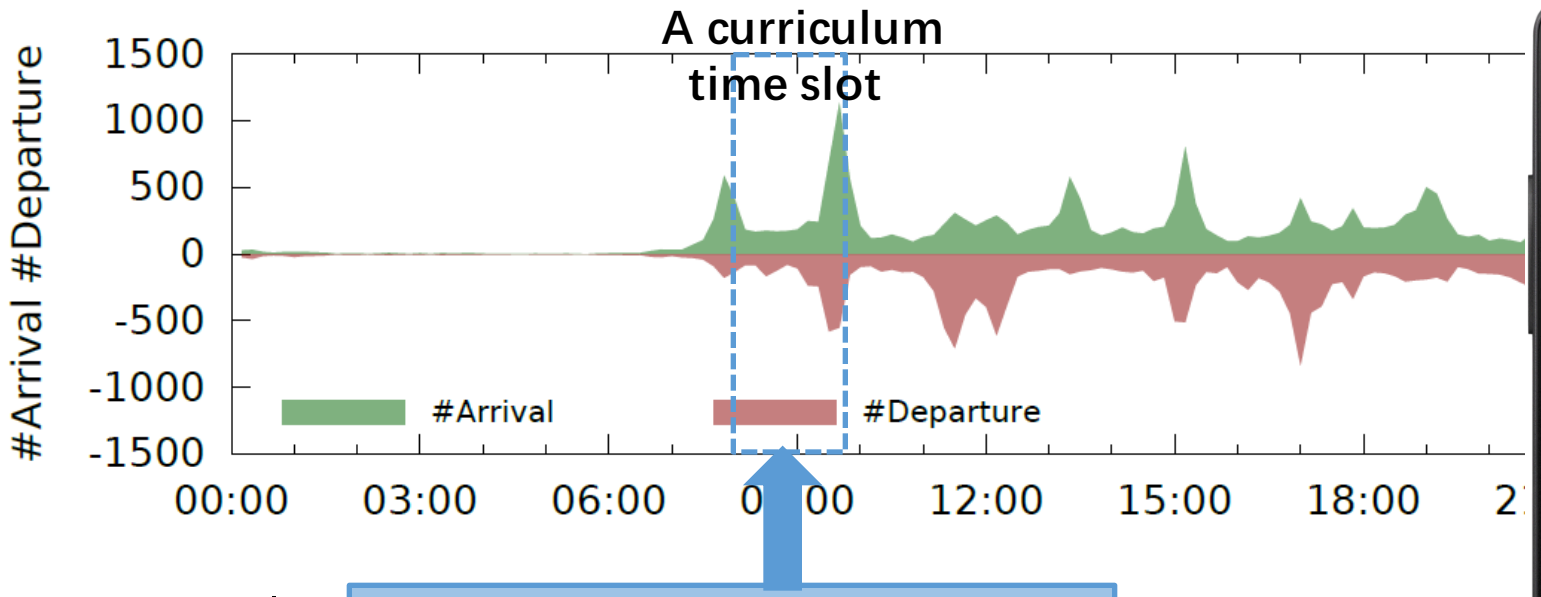
2,670
Access points

111
Buildings

16
Weeks

Observation #1

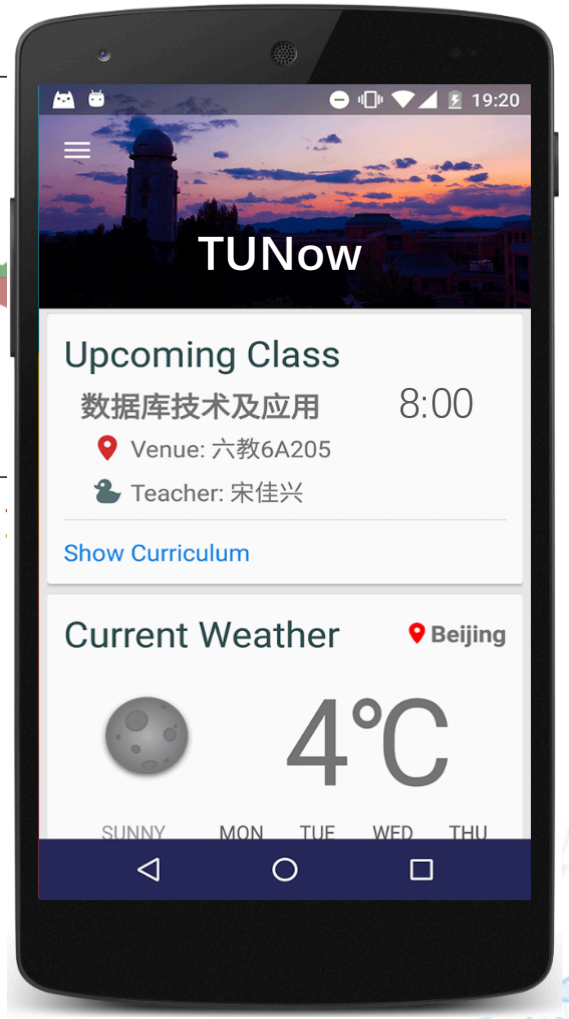
Periodic schedule is easy to extracted from trajectory datasets.



Example:

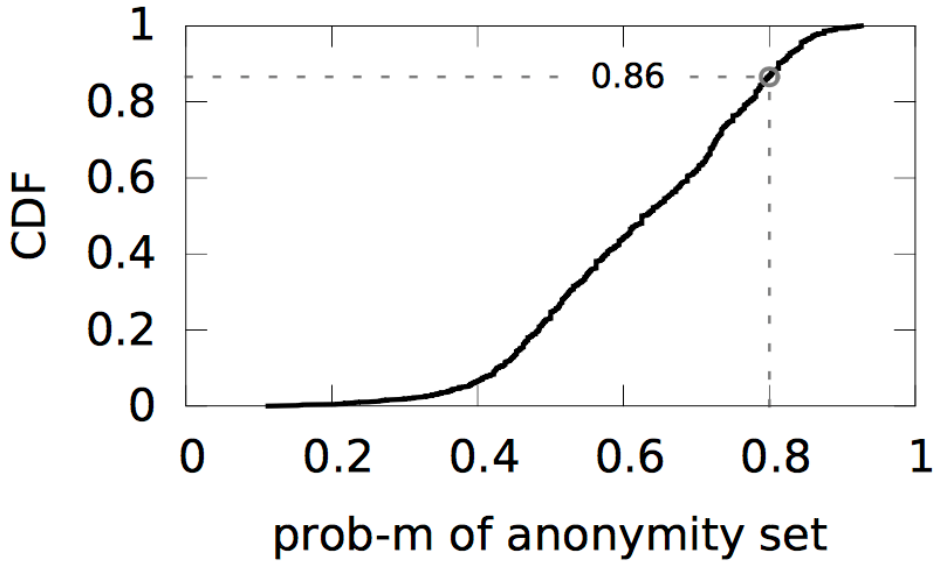
```
ID1, 20170515_0830, 6 jiao
ID1, 20170522_0820, 6 jiao
ID1, 20170529_0825, 6 jiao
```

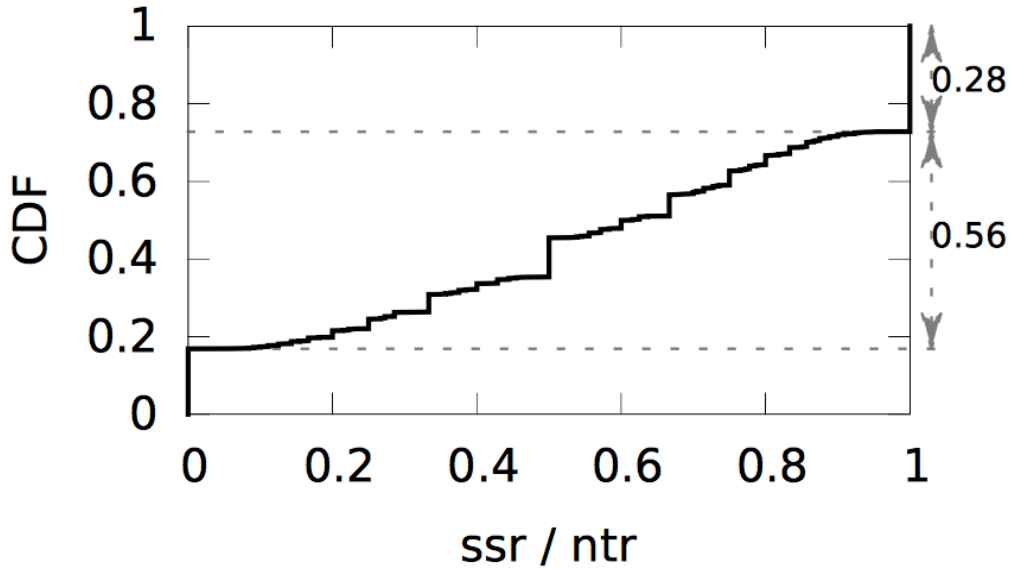
Periodic schedule extracted compares to the ground truth from TUNow app



Observation #2

The risk of schedule-leakage is high, even for K>3





prob-m: Max Attack Probability

ssr/ntr: Sensitive Schedule Rate / Non-empty Time slots Rate

Two metrics that measure the diversity

See more results in the paper

Conclusion

- A large-scale trajectory measurement shows that
  - A very high risk of the schedule-leakage
  - Trade off between data utilities and privacy
  - Calls for schedule diversity-oriented solutions (future work)

Gracias~

mmh16@mails.tsinghua.edu.cn

清華大學

Tsinghua University