清華大學
Tsinghua University

Vehicle-based Mobile Urban Sensing

The Technical Solution and Theoretical Problems

Zhang Lin, Associate Professor
Department of Electronic Engineering,
Tsinghua University, Beijing, China

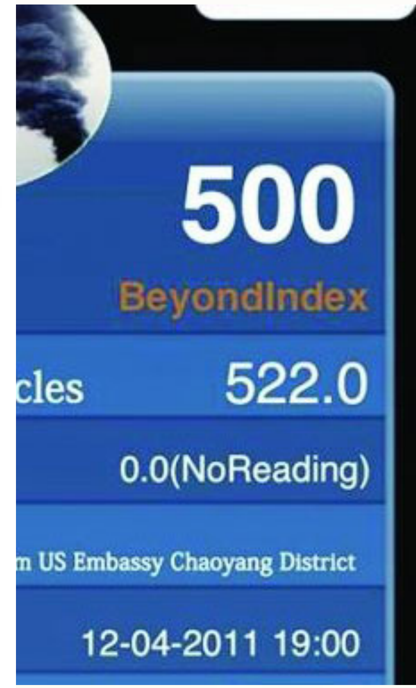
清華大學
Tsinghua University

Air Pollution in Big Cities

Tsinghua University

Difference of Opinion

西城区官园	95
西城区万寿西宫	120
朝阳区奥体中心	97
朝阳区农展馆	96
海淀区万柳	99
门头沟区龙泉镇	84
房山区良乡	160
通州区通州镇	158
顺义区仁和镇	100
昌平区定陵	74
昌平区昌平镇	71
大兴区黄村镇	168
大兴区榆垡	127
平谷区平谷镇	129
怀柔区怀柔镇	92



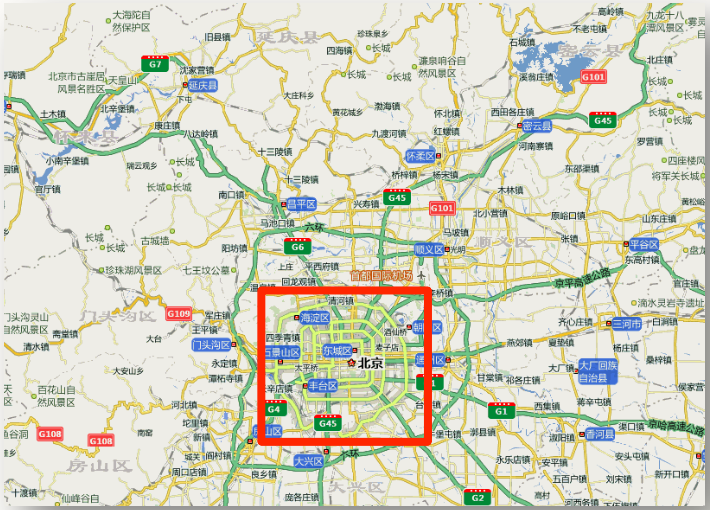
清華大學
Tsinghua University

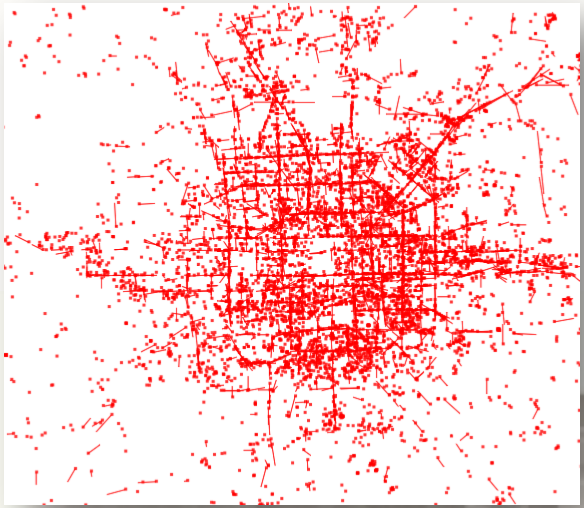
We don’t understand the city

- Challenges of urban sensing
  - Delivery of gigantic volume of data from heterogeneous sensors
  - Coverage of large areas with high spatial resolution

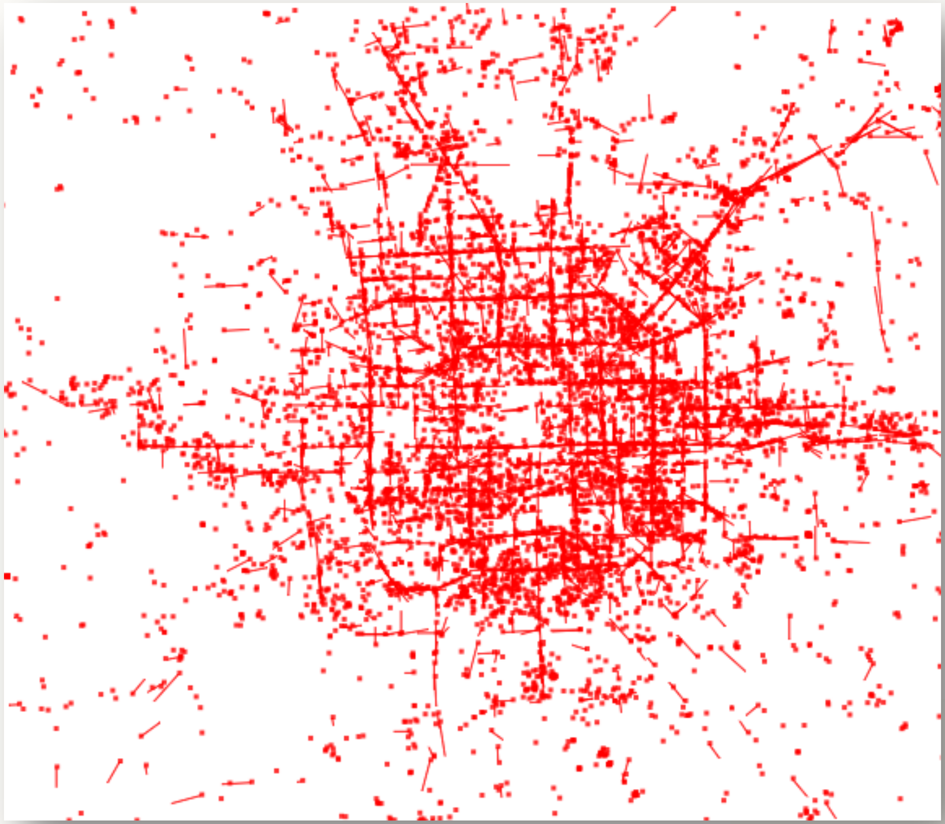
Design Objectives

- Large Coverage Area (Beijing: approx. 600 km^2)
- Scalability (number of sensors and data traffic)
- Versatility in applications
- Affordable deployment cost





The System Architecture and the Meats for Research



- System Architecture
  - Sensors (road side or on-board)
  - Carriers (taxi and buses)
  - Base station/Sinks (road side)
- Sensing Meets Mobility
  - Mobility improves the communication capacity
  - Mobility increases the system coverage

清華大學
Tsinghua University

System Deployment in Beijing

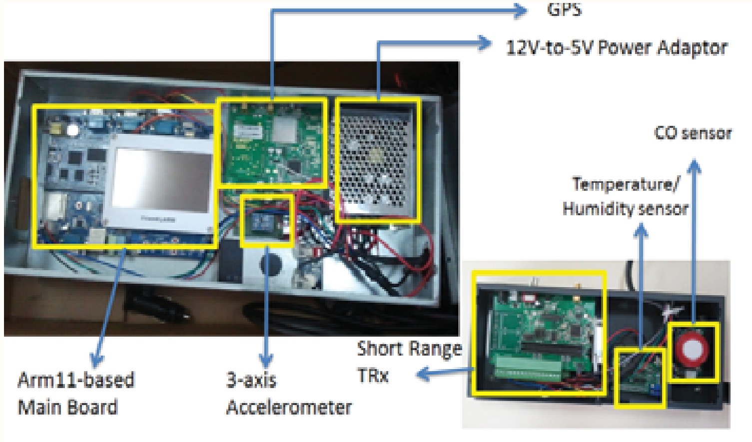


Fig. 1. Pervasive Urban Sensing Prototype (Left: sensing platform, size 15cm×40cm, Right: environmental sensing module, size 10cm×15cm)

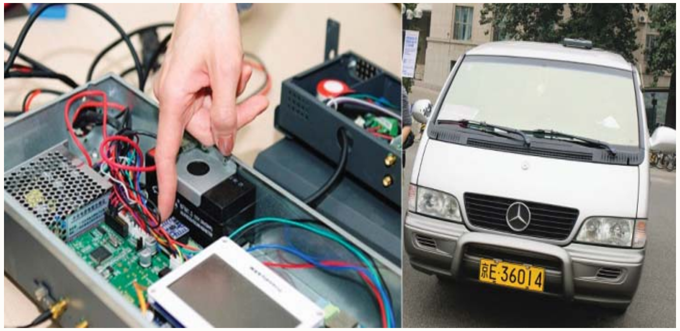
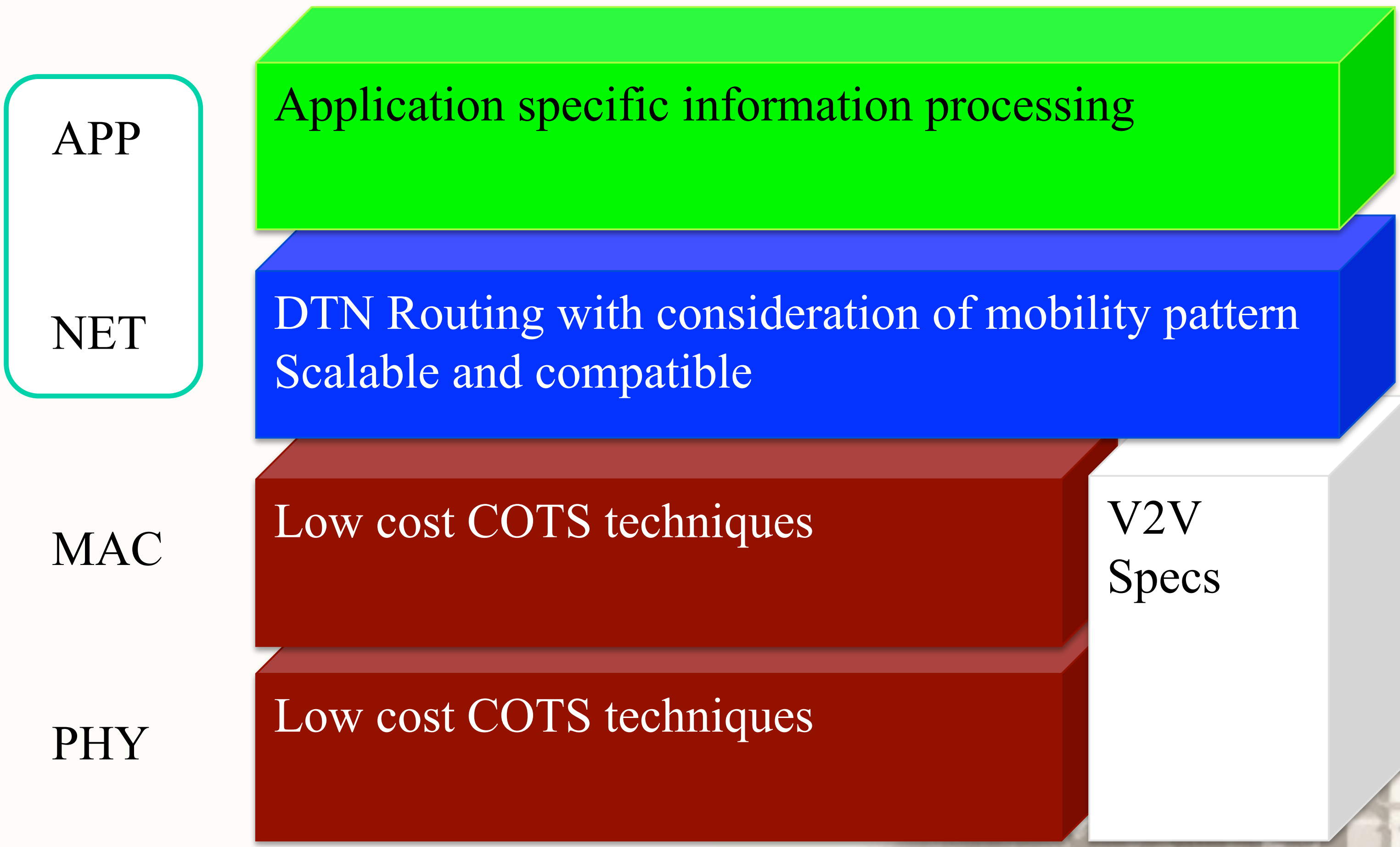


Fig. 2. Deployment on Tour Buses

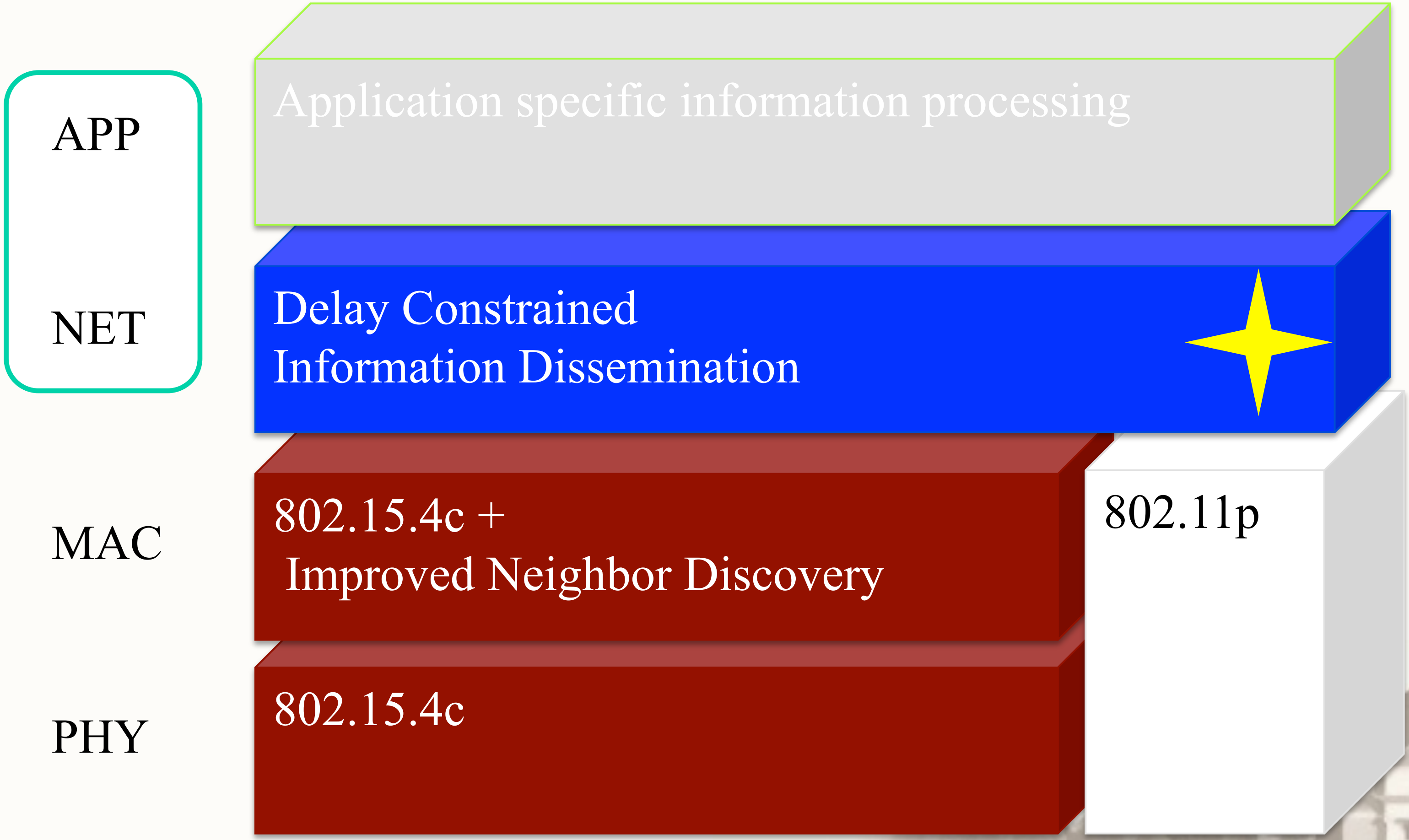
- Sensors deployed on tour buses in Beijing
  - CO, 3-axis Acceleration, GPS, Temperature, Humidity, Light
- Location data collected from 27,000 taxi cabs

清華大學 Tsinghua University

System Layers



The Network Layer



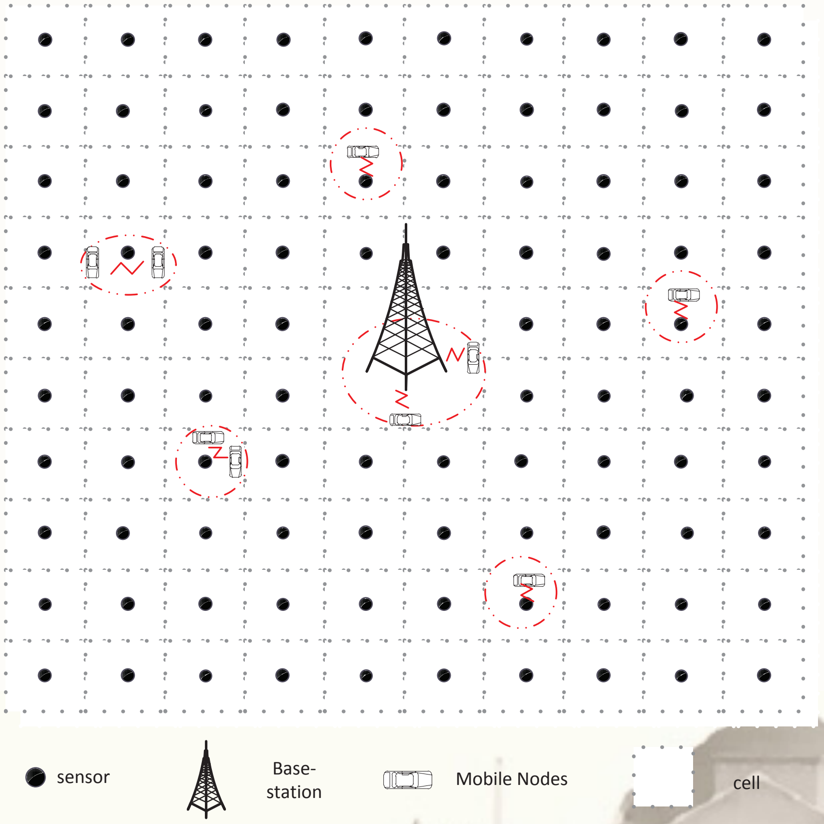
An infrastructure-free network

- Using Vehicular to Relay Information to Sink(s)
- Delay Tolerant Networks
  - Pros:
    - Vehicle has constant electric supply
    - Storage on-board is cheap
    - The mobility pattern are predictable
  - Cons & Challenges:
    - Frequently changed network topology
    - Limited inter-contact duration and local capacity
    - Density fluctuation over time and space

清華大學
Tsinghua University

Setup of Problem

- Network Elements
  - Sensors
  - Mobile Relays
  - Sink
- Channel Capacity is the Major Constraint



Existing Protocols

- Existing Routing algorithms
  - Routing based on **replication** (Epidemic, SW)
    - Increase delivery possibility within delay constrain
    - Require tremendous cost of network resource
  - Routing with network **resource constraint** (MaxProp, Rapid)
    - Optimize performance with resource constrain
    - Perform suboptimal as density fluctuates
  - Routing with **density adaptive** nature (DA-SW, ECAM)
    - Optimize performance as density changes
    - Not consider capacity limitation in VANET

Tsinghua University

DAWN: Utility-based Heuristics

- Single Packet Utility Function
  - A packet company is the totality of all the copies of a packet s
  - The delivery of the packet from source to the base station need to be within a time limit
  - $U_s(t)=P\{T_s(t)<T_{MAX}\}$

- System Utility Function
  - Average utility function values over all the packet companies

Tsinghua University

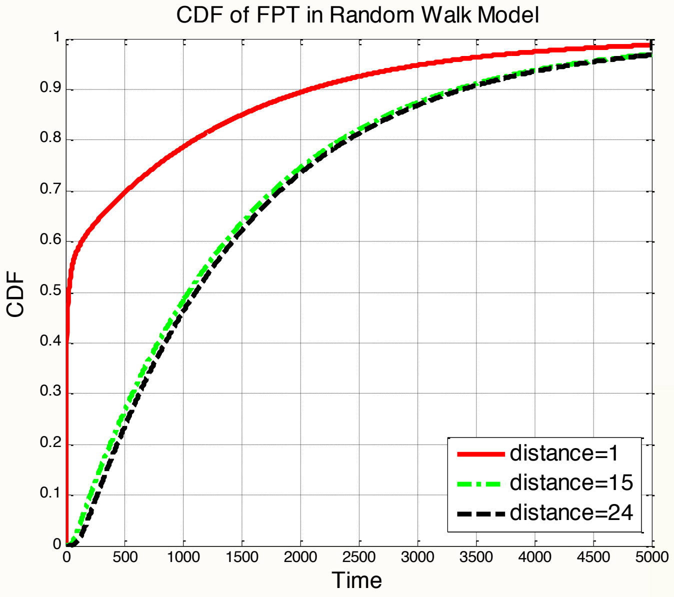
The Protocol Rationale

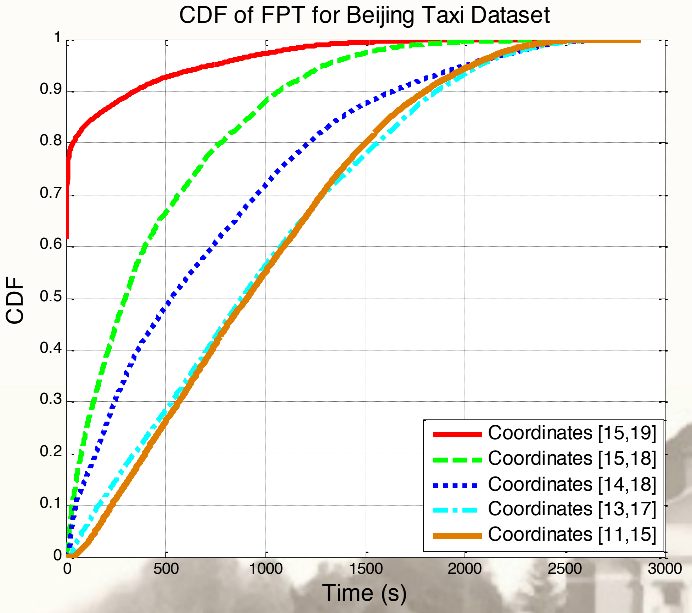
- Trade-off in the Duplication Behavior of Packets
  - Increase the probability of successful delivery / System Utility
  - Replications are subject to channel constraint
  - We need to find the best local replication strategy, but it depends on the future.
- Heuristics of the protocol
  - Packet utility gain from replication:

$$\Delta Up_s^{t,l} = (1 - Up_s^t)(1 - (1 - \Phi_l(T_{MAX} - t))^{\lambda_l})$$

First Arrival Time of Random Walk

- The interval between the time a mobile node start from cell *i* to the time that it first hits the base station
- No clean solution, even for the simple random walk on torus
- We use the empirical data from Beijing Taxi Database to estimate CDF of FPT in DAWN: $\Phi_i(t)$





The DAWN Heuristics

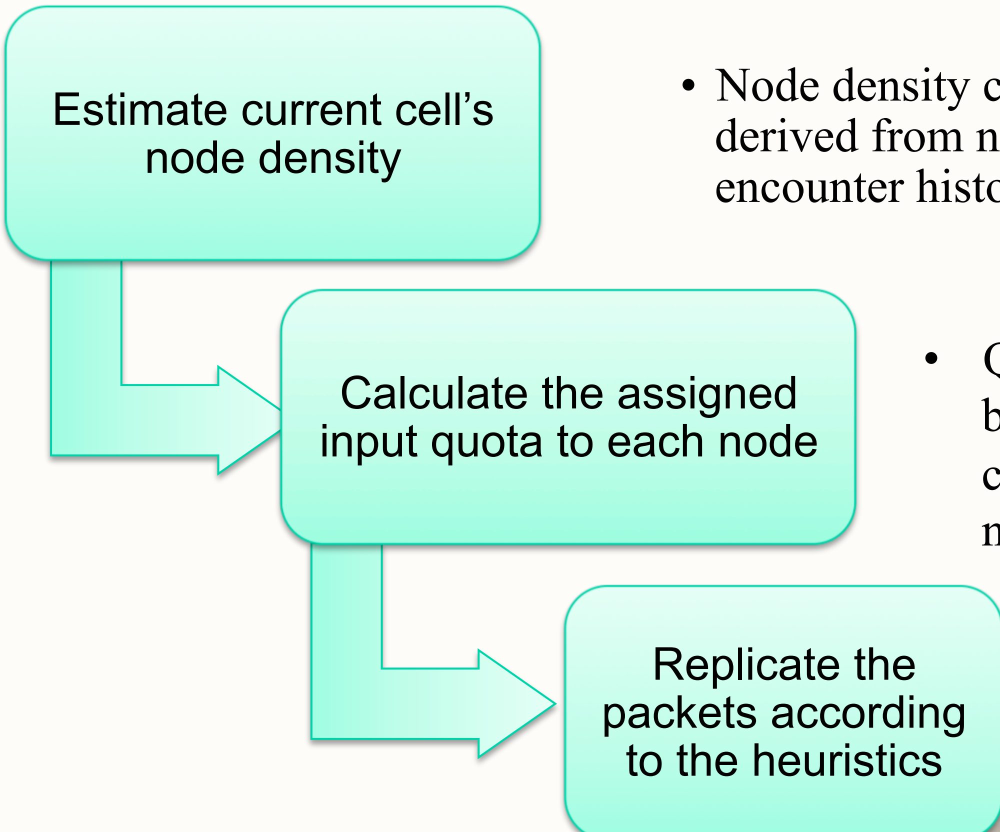
- The total number of duplication is decided by the optimal input K_{OPT} is decided by the cell capacity
- Packets with higher Utility Incremental Value will be duplicated on the broadcast channel with higher priority

$$(1 - U_{s,t})(1 - (1 - \Phi_i(T_{MAX} - t + t_0))^{\lambda_i})$$

  - λ_i : estimated by counting neighbor number
  - Φ_i : calculated based on the empirical data
  - $U_{s,t}$: estimated locally

DAWN: The protocol

❑ Density Adaptive Routing With Node Awareness

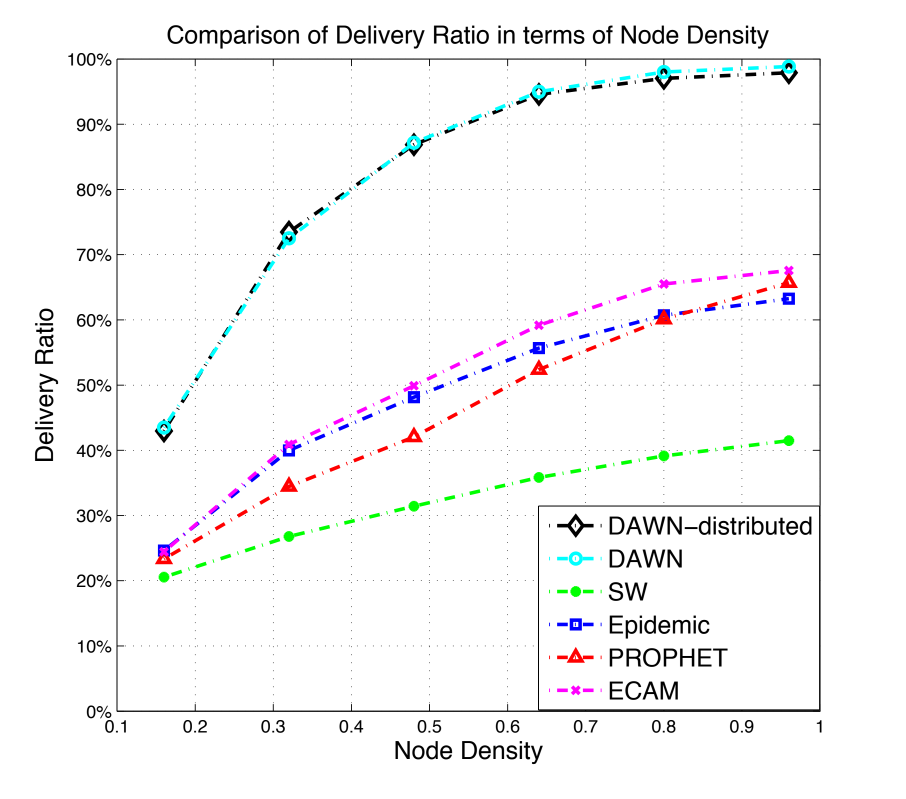


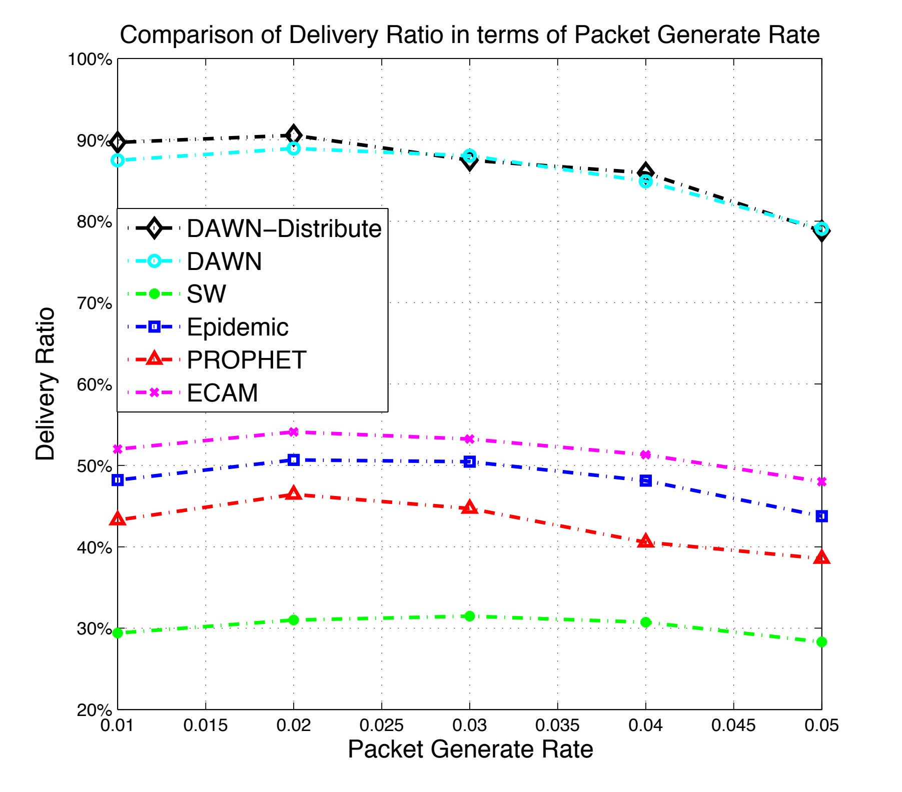
- Node density can be derived from node encounter history
- Quota for each node can be derived from the channel capacity and local node density
- Packets with higher Utility gain will be replicated with high priority

清華大學
Tsinghua University

Simulation on Manhattan Grids

- Simulation on Manhattan Grids
  - Delivery Ratio vs Node Density
  - Delivery Ratio vs Source Rate

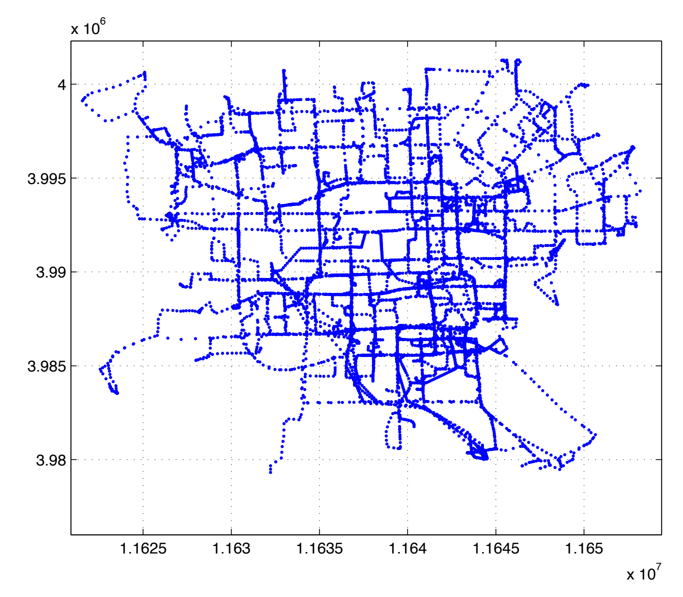


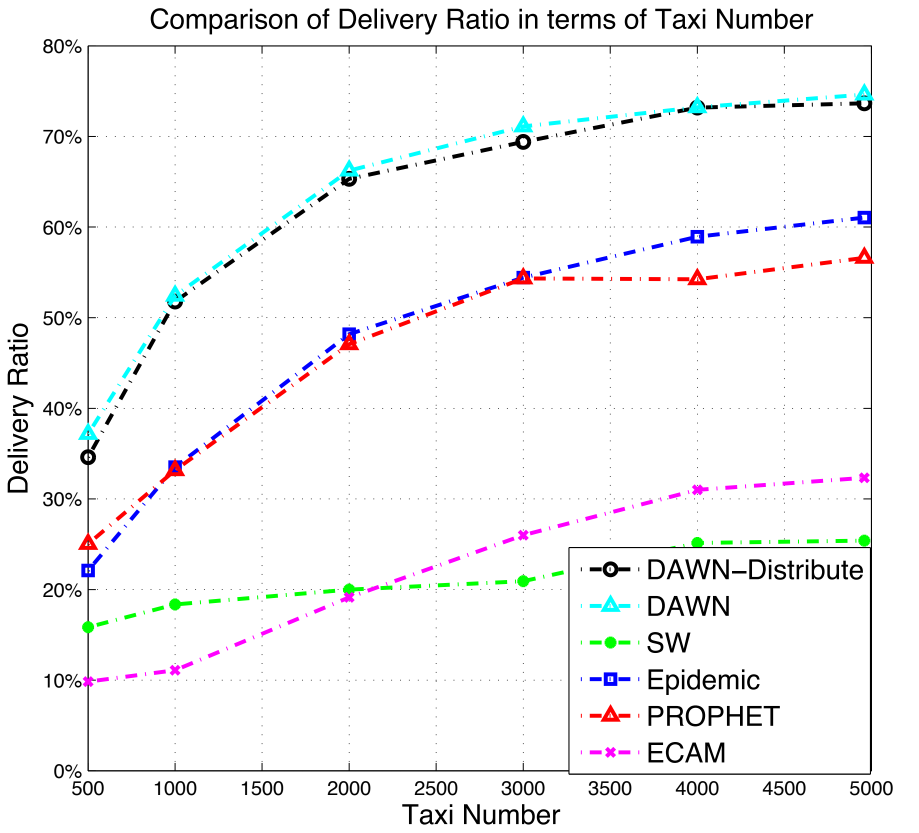


Beijing field data

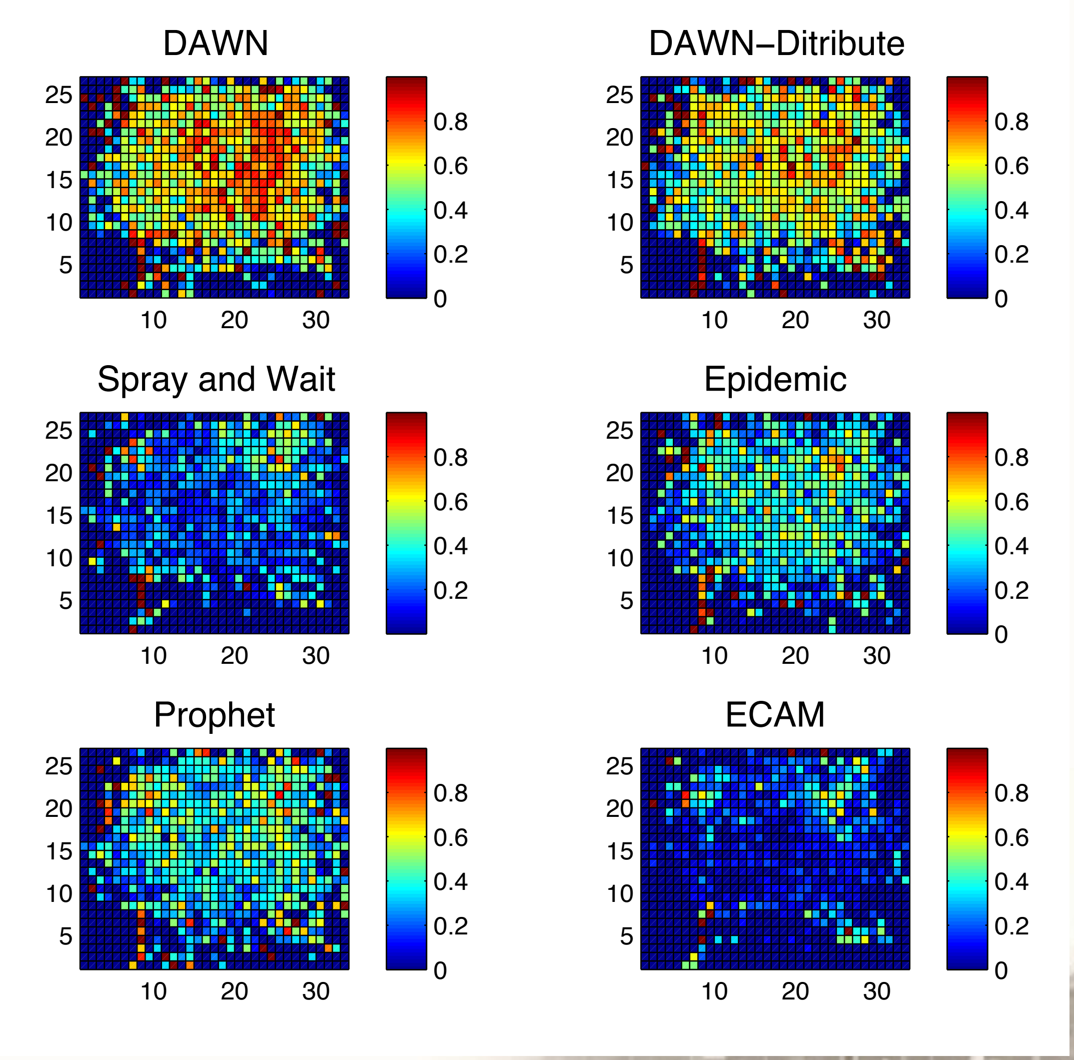
❑ Evaluate on the Beijing Taxi Dataset

- ☆ Within the 5th ring road (26.3x33.5km)
- ☆ Comm. range: 100m
- ☆ Max Delay: 250-minute
- ☆ May 1st ~ 30th, 2009



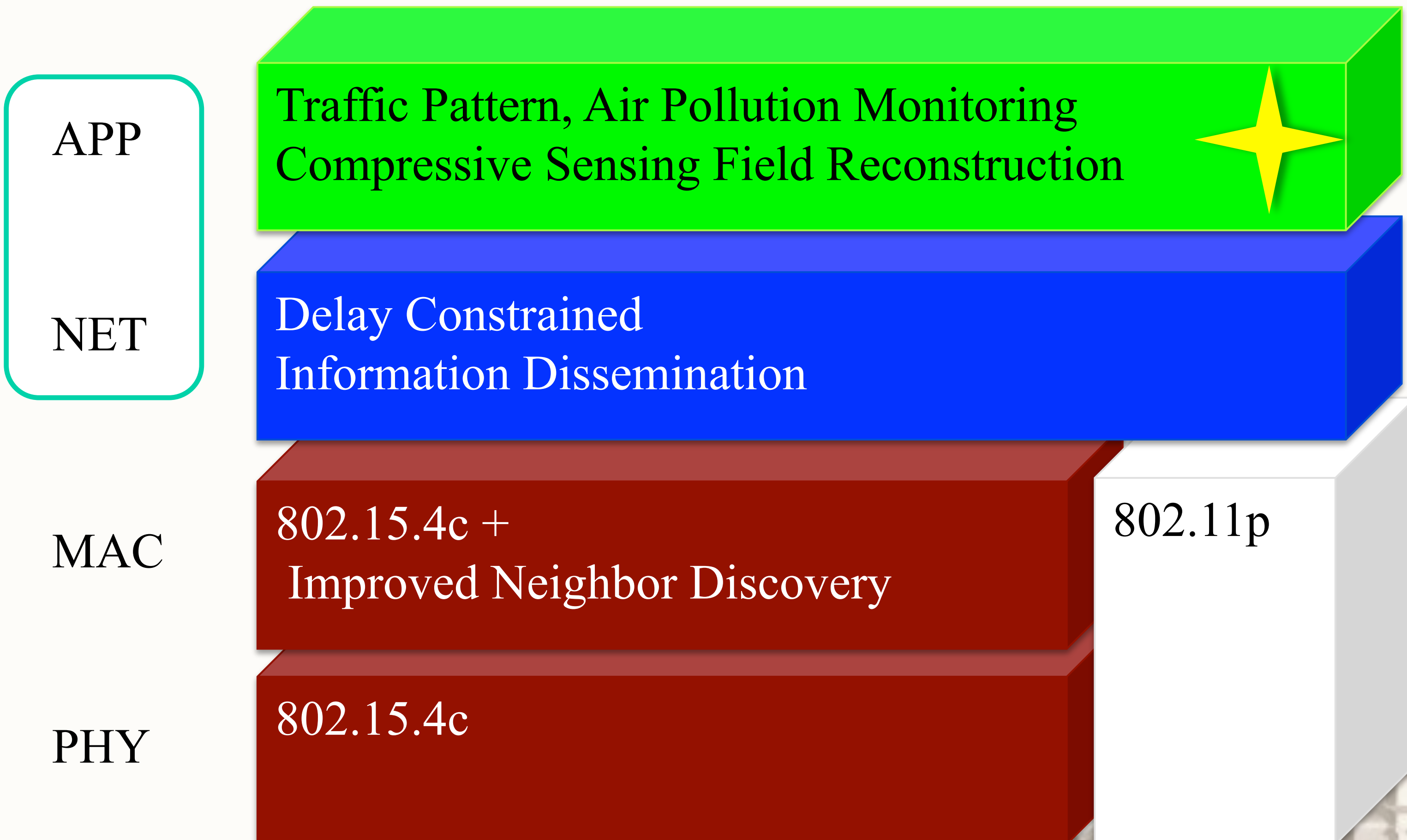


Delivery Ratio Geo-fairness



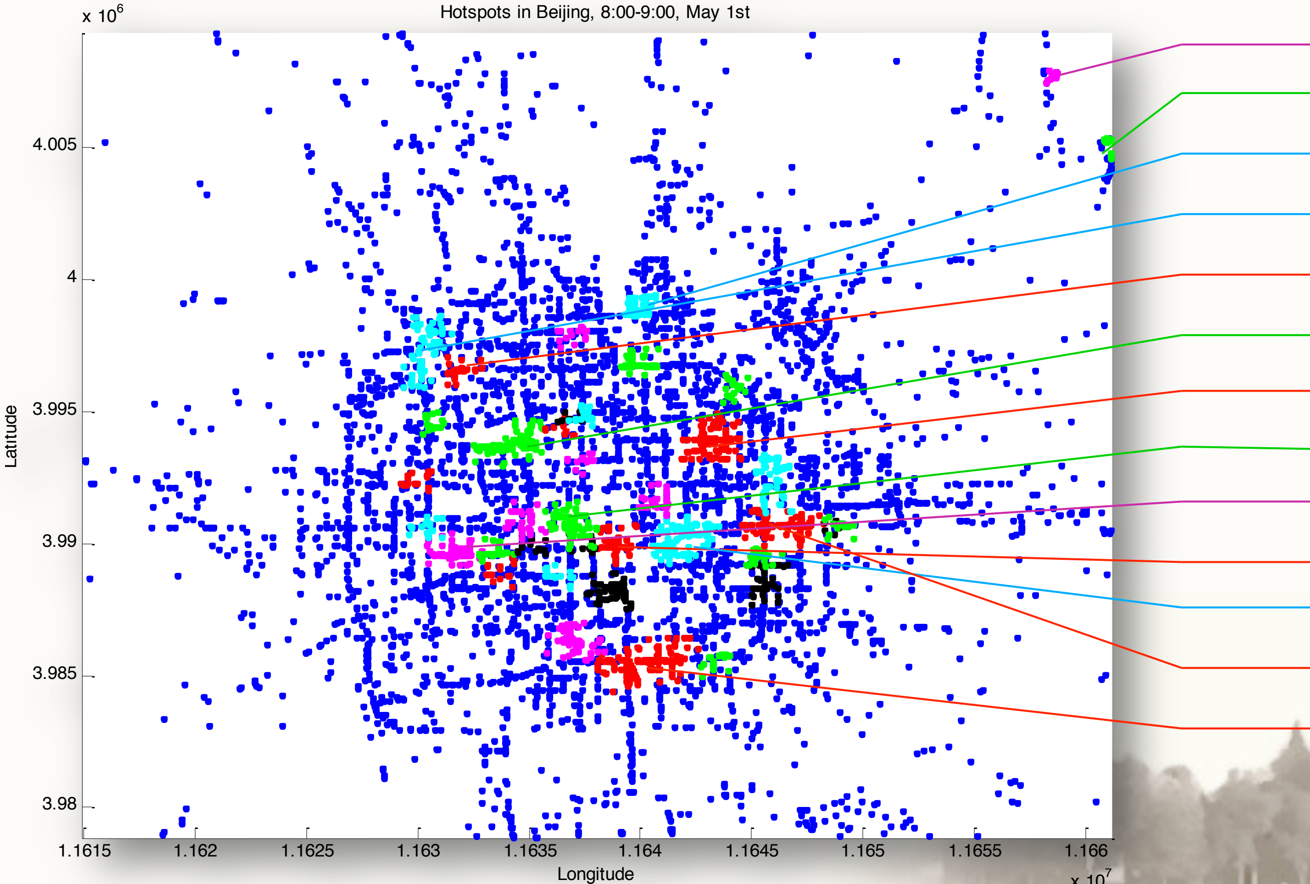
Tsinghua University

The *Sensing* Layer



Hot spots in Beijing

- Density-based Clustering (DBSCAN)
- 2009-05-01, 8:00-9:00, 3250 Rides
- 37 Hot zone identified



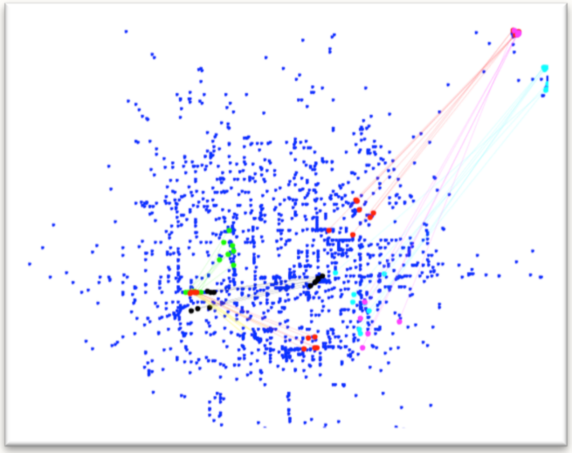
- T2航站楼
- T3航站楼
- 鸟巢
- 西苑-颐和园
- 中关村
- 西直门
- 东直门-三里屯
- 西单
- 北京西站
- 天安门-前门
- 王府井-北京站
- 国贸
- 刘家窑
- ……

清華大學 Tsinghua University

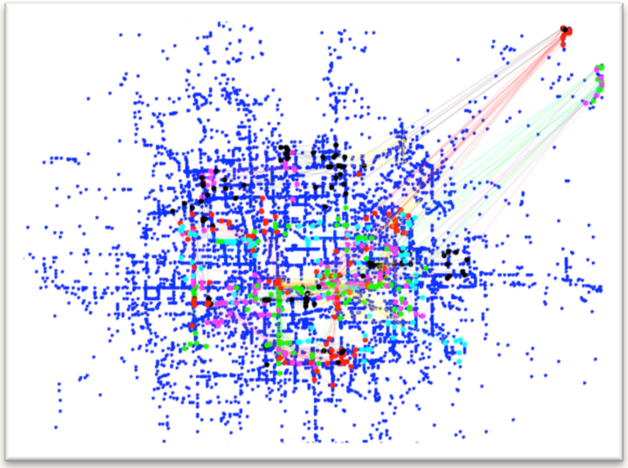
Origin-Destination Pair Cluster

Define the distance between two (O-D) pairs as

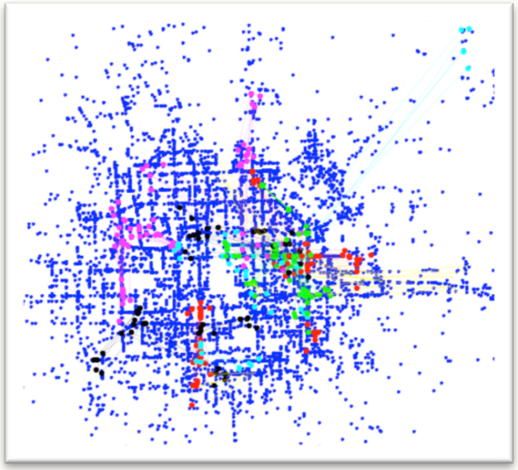
$$dist(O_1D_1, O_2D_2) = \frac{\left(||O_1 - O_2||_2^2 + ||D_1 - D_2||_2^2\right)^{\frac{1}{2}}}{\left(||O_1 - D_1||_2 + ||O_2 - D_2||_2\right)^{\alpha}}$$



5 a.m. ~ 6 a.m.



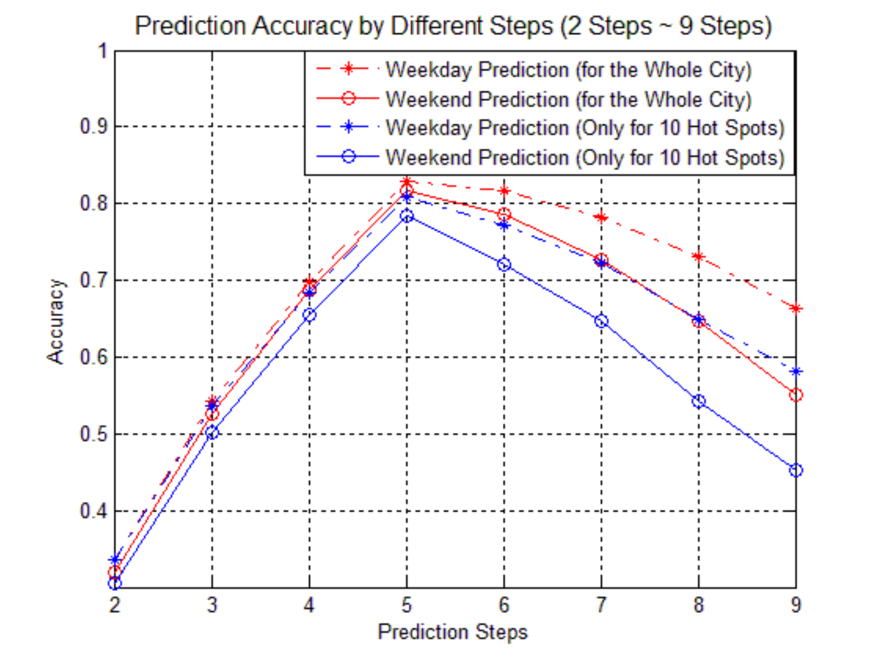
12 p.m. ~ 1 p.m.

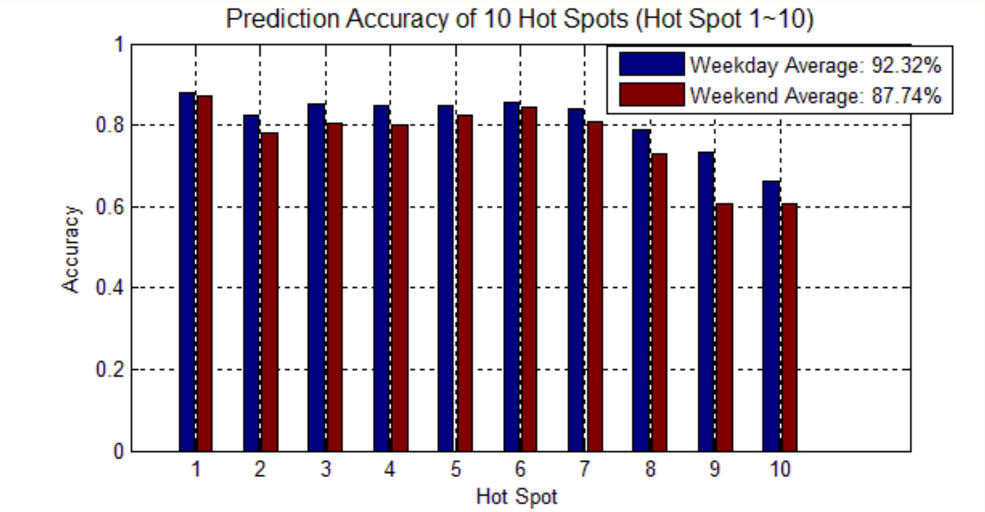


11 p.m. ~ 12 a.m.

Taxi Destination Estimation

- Can we forecast the drop-off location from the GPS data
- Learn a Markov chain from the dataset
- Good for mobile ads business





Prediction Accuracy of 10 Hot Spots
Weekday Average Accuracy: 92.32%
Weekend Average Accuracy: 87.74%

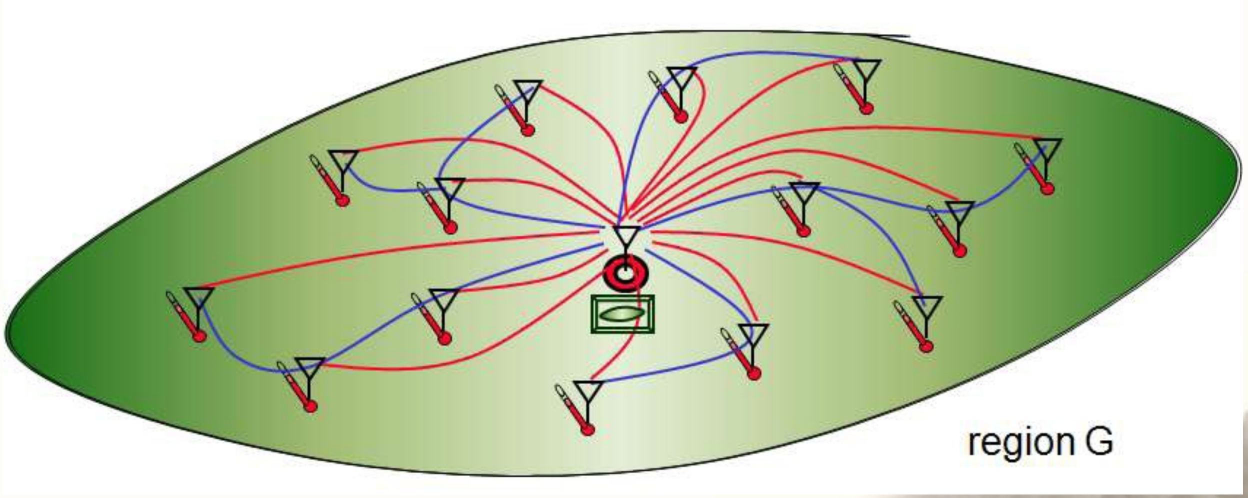
Traffic Pattern Mining

- Sample density near an intersection within 15 minutes
- Colors indicate the number of sampled locations

- Traffic Pattern Mined from GPS Data
- The spatial resolution is much higher than the loop-based detection
- The granularity can be further improved to lane-level
- Using lifted space method to detect cars' turning behavior at the intersection

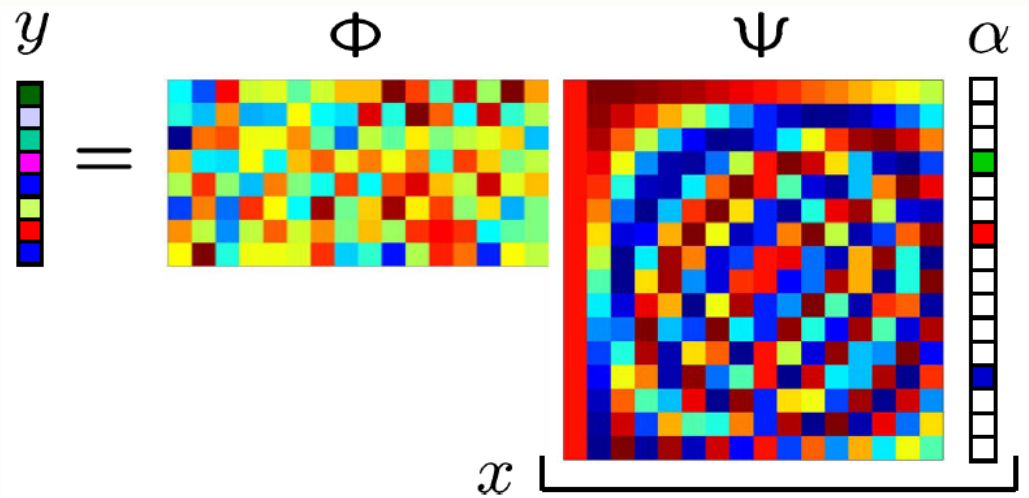
Most Applications: Field Gathering System

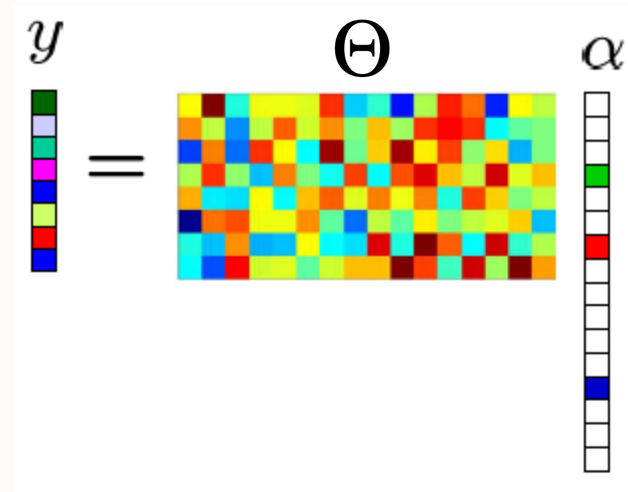
- A physical field on G: $X(u,v,t)$
- Distributed sensor nodes embedded in G to sample $X(u,v,t)$
- Sensor nodes encode the data and transmit back to BS
- Reconstruct $\hat{X}(u,v,t)$
- **Objectives**: Minimal cost, minimal distortion



Exploiting the Sparsity in the Signal

$$y = \Phi x = \Phi \Psi \alpha = \Theta \alpha$$





Φ : measurement matrix $M \times N$

Ψ : orthonormal basis $N \times N$

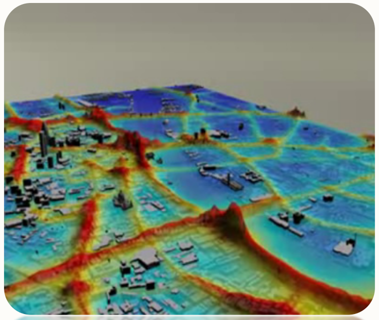
Θ : Compressed Sensing reconstruction matrix $M \times N$

From R. G. Baraniuk, Compressive sensing, IEEE Signal processing magazine July 2007

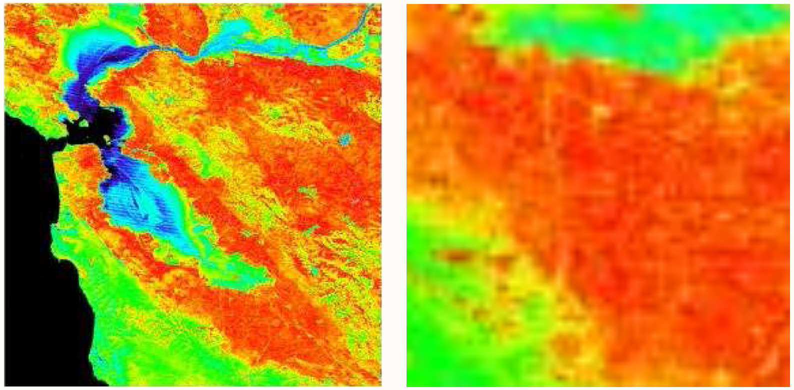
Information is sparse in physical fields

Sample

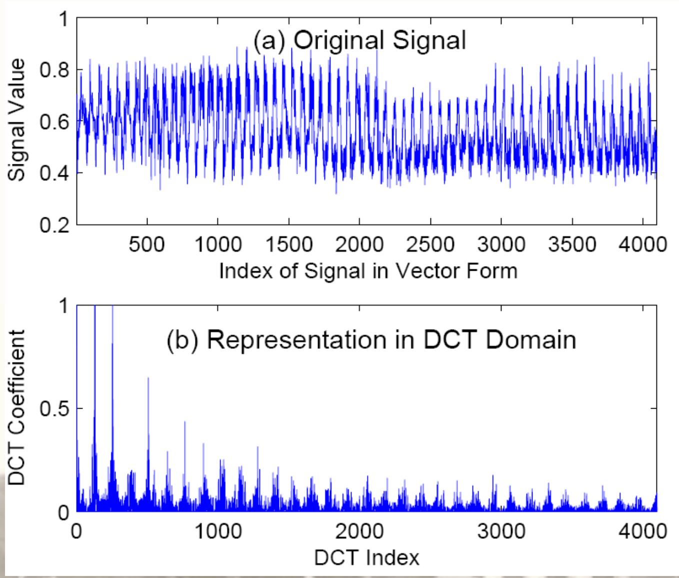
$$y = f(x_1, x_2, \ldots, x_n)$$



Reconstruction



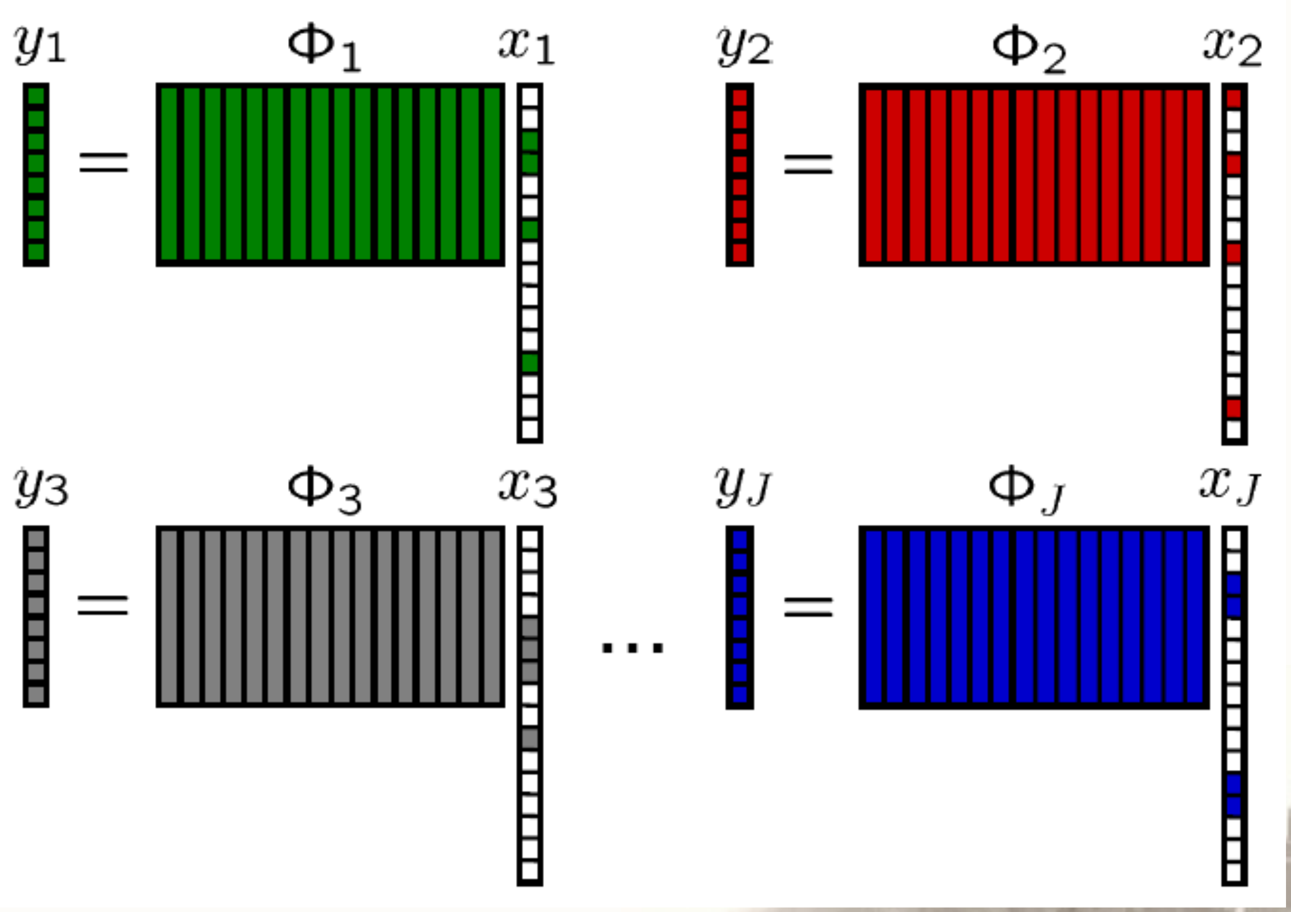
Temperature Distribution



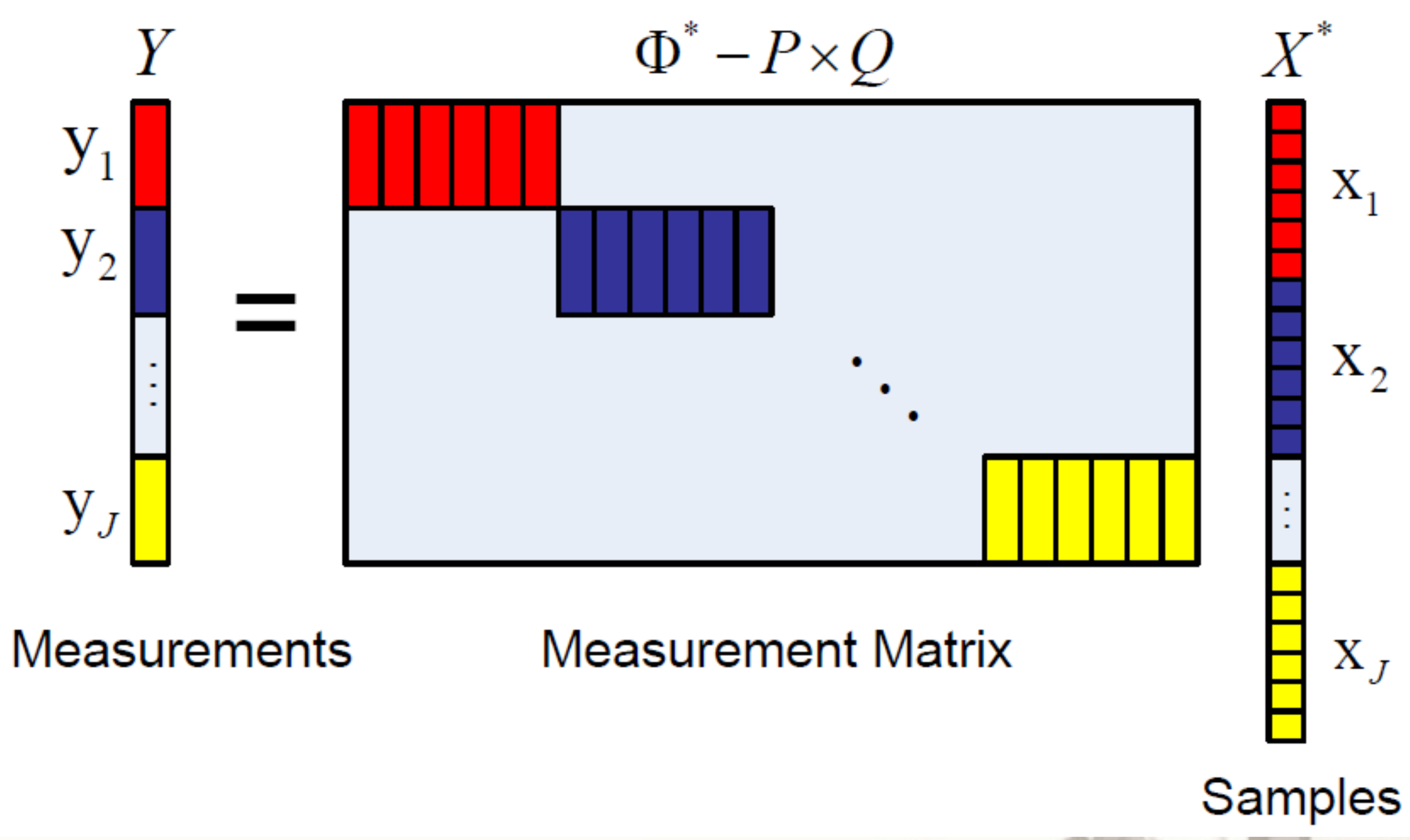
清華大学
Tsinghua University

Distributed CS

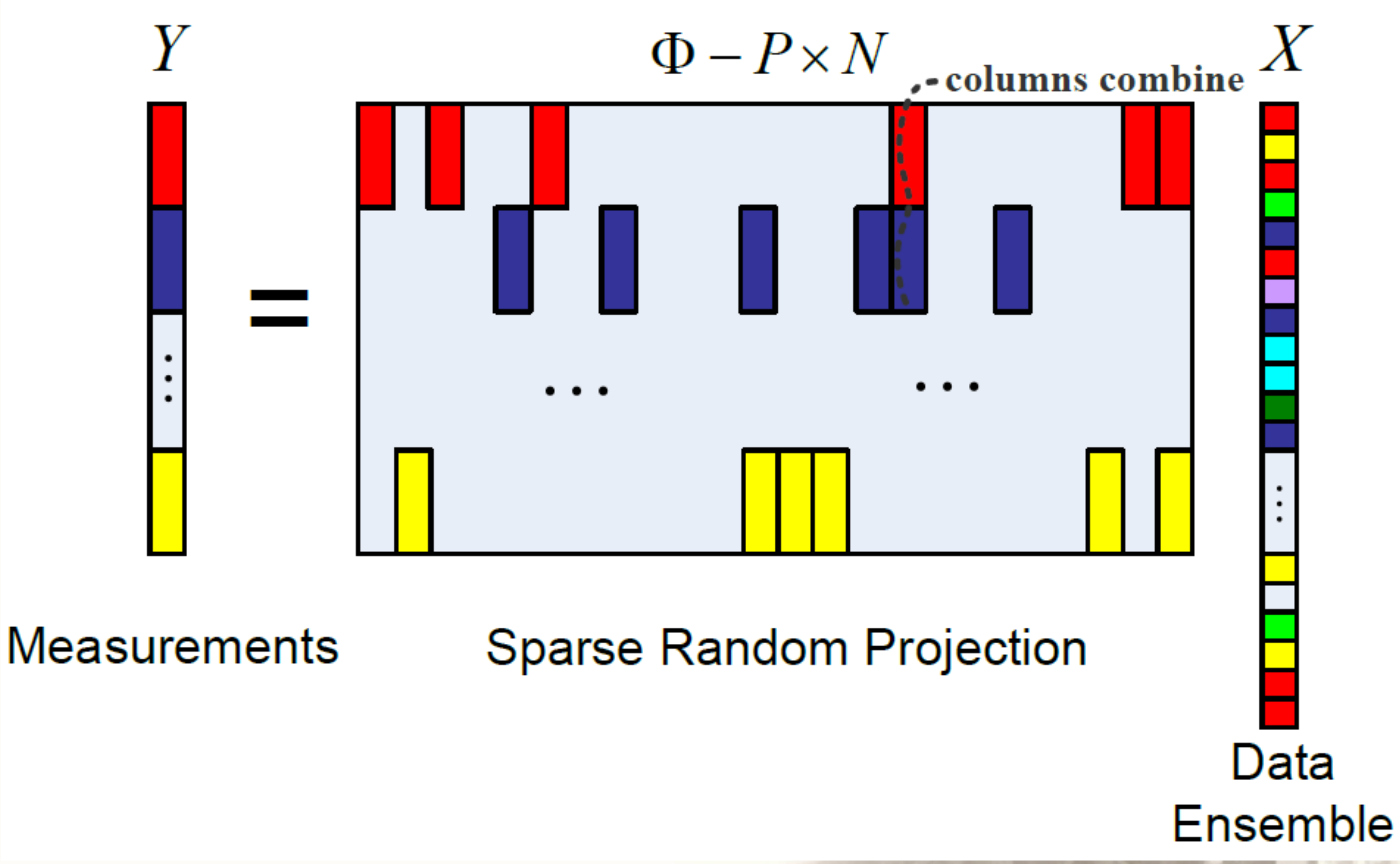
- Decompose sensing tasks to individual sensors



Distributed CS

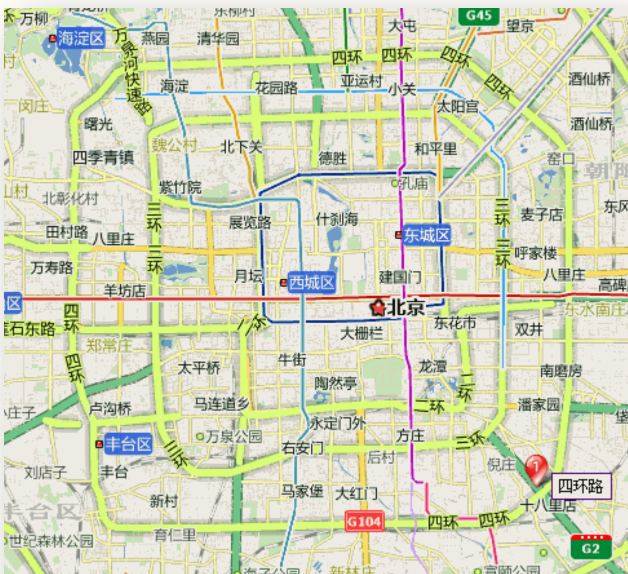


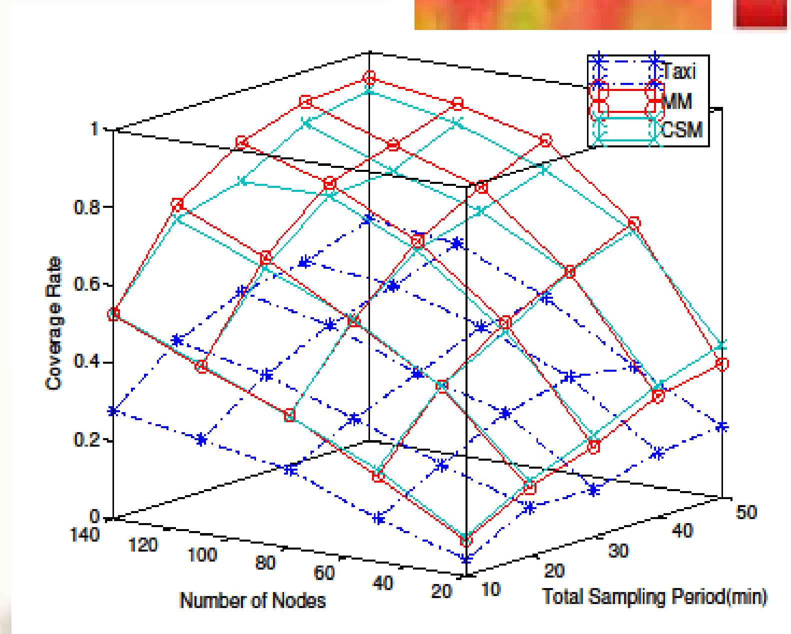
Distributed CS



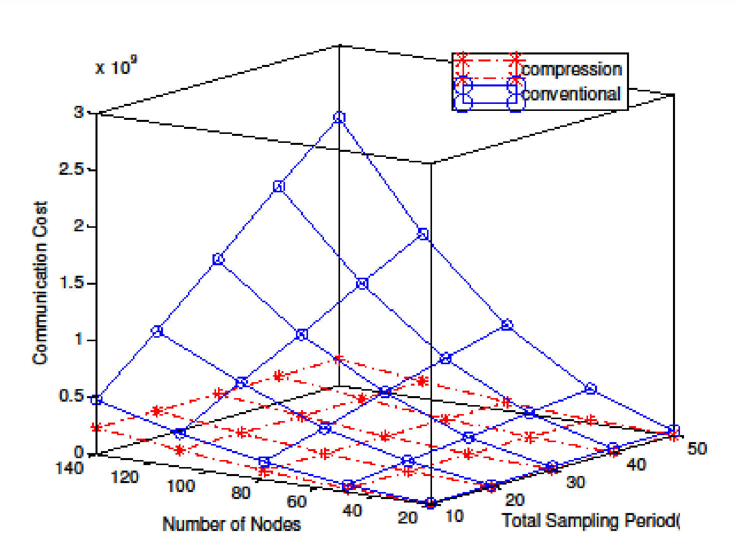
Performance Evaluation

- ❑ Temperature Field inside the 4th Ring Road (16x16 km)
- ❑ Mobility Model:
  - Manhattan Mobility Model
  - City Section Mobility Model
  - Real Taxi trajectory data

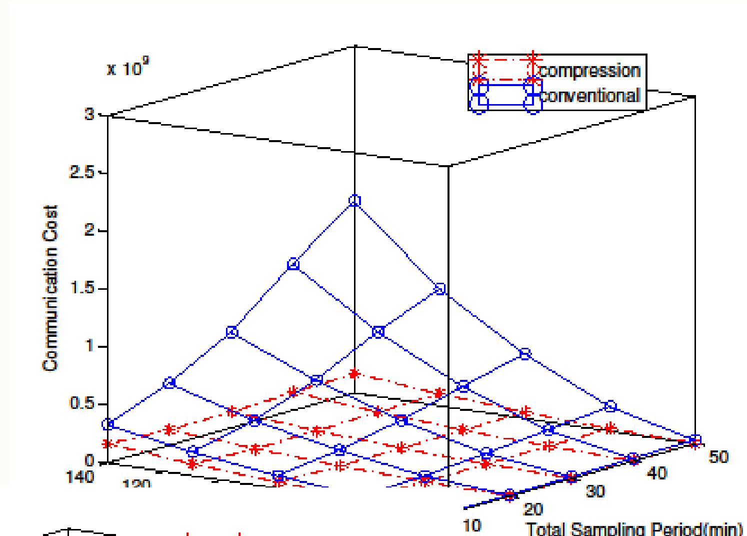




Performance of the distributed CS



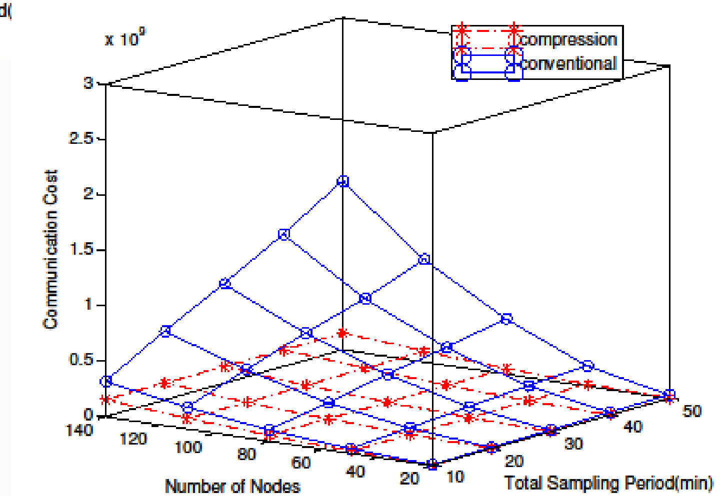
Random Walk



Random Walk with Section

Communication Cost:

$$\sum (bit)\times(distance)^2$$



Taxi Trajectories

清華大學
Tsinghua University

Traffic and Pollution Reconstruction

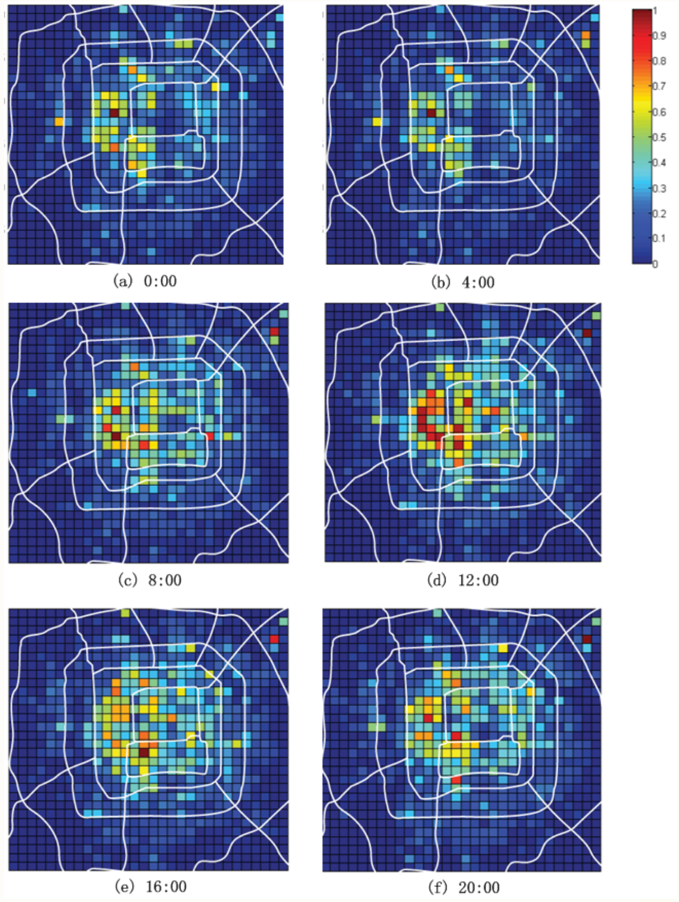


Fig. 4. Traffic Density in Beijing in 24 hours

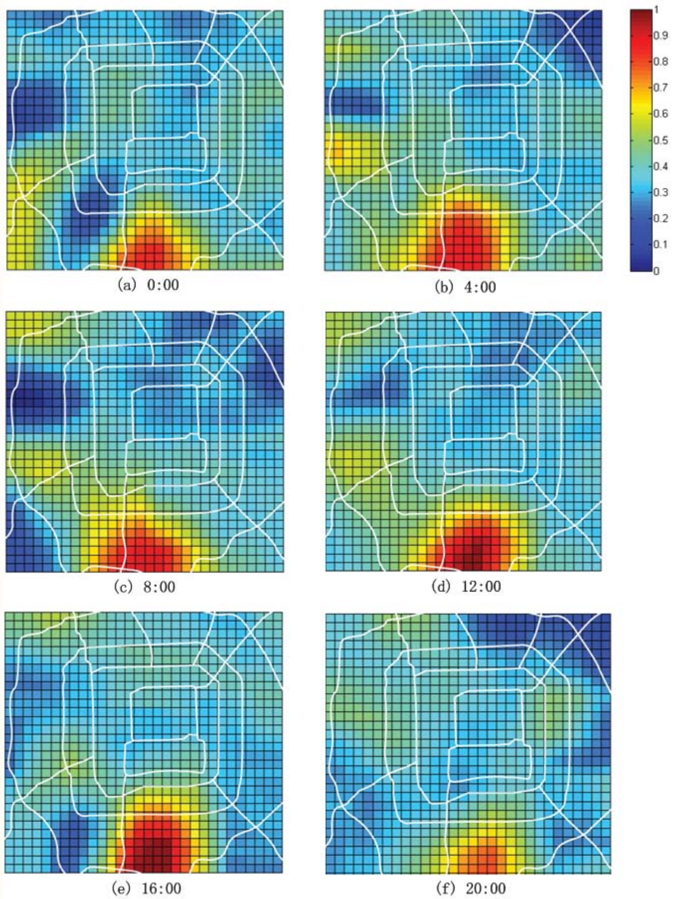
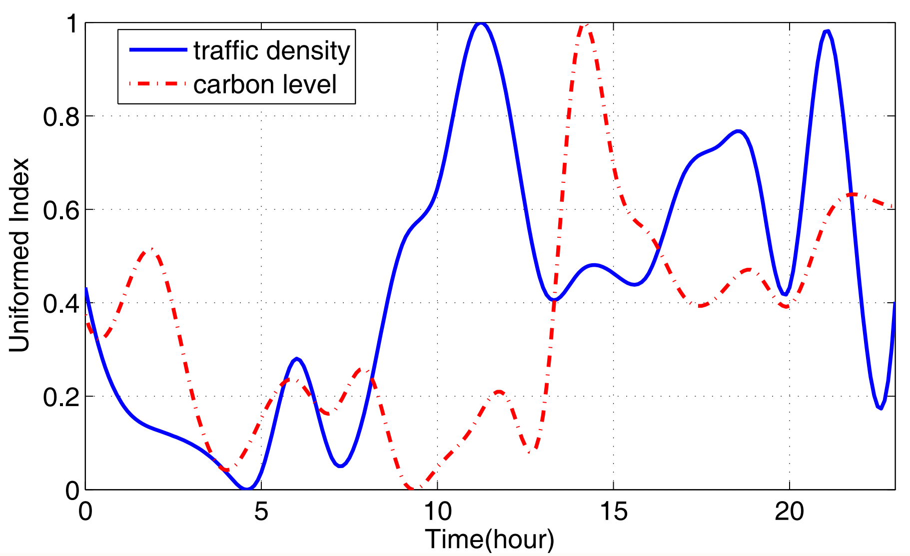
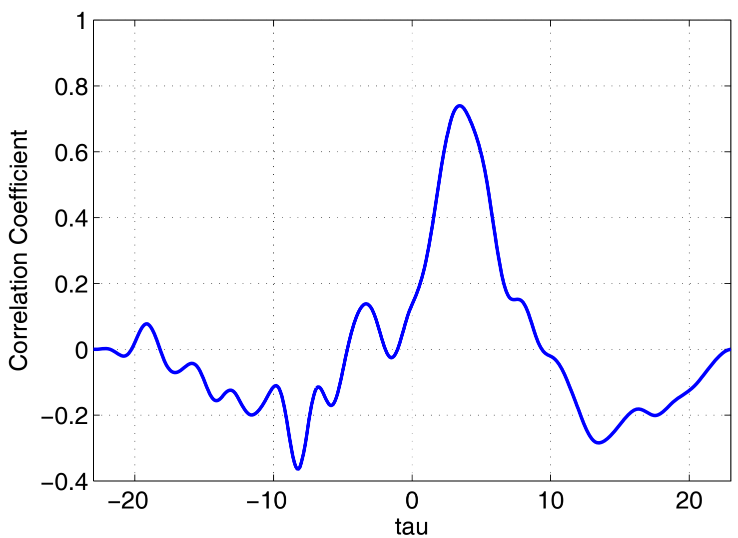


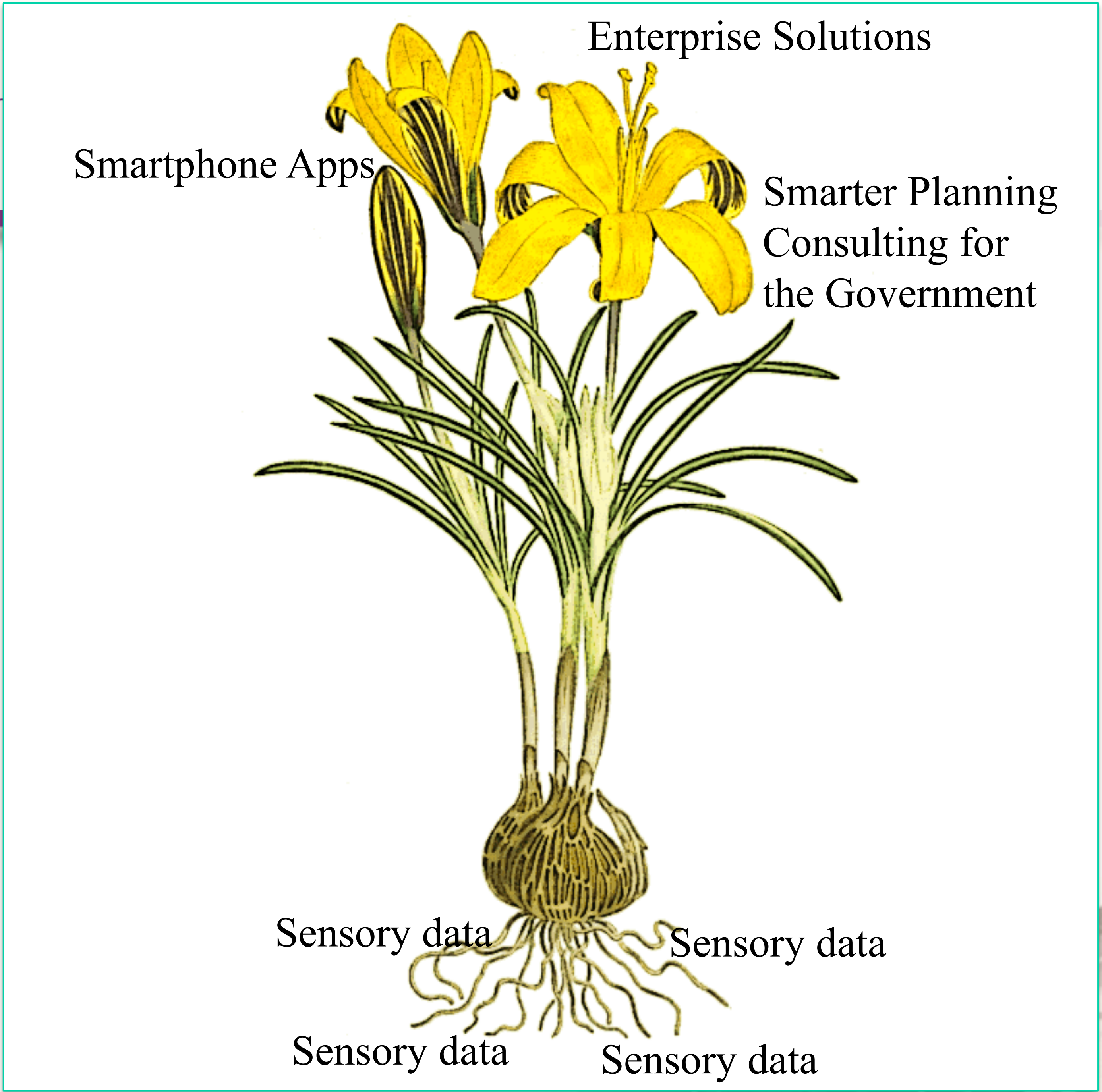
Fig. 5. Carbon monoxide dynamics in Beijing in 24 hours

Correlation between traffic and pollution



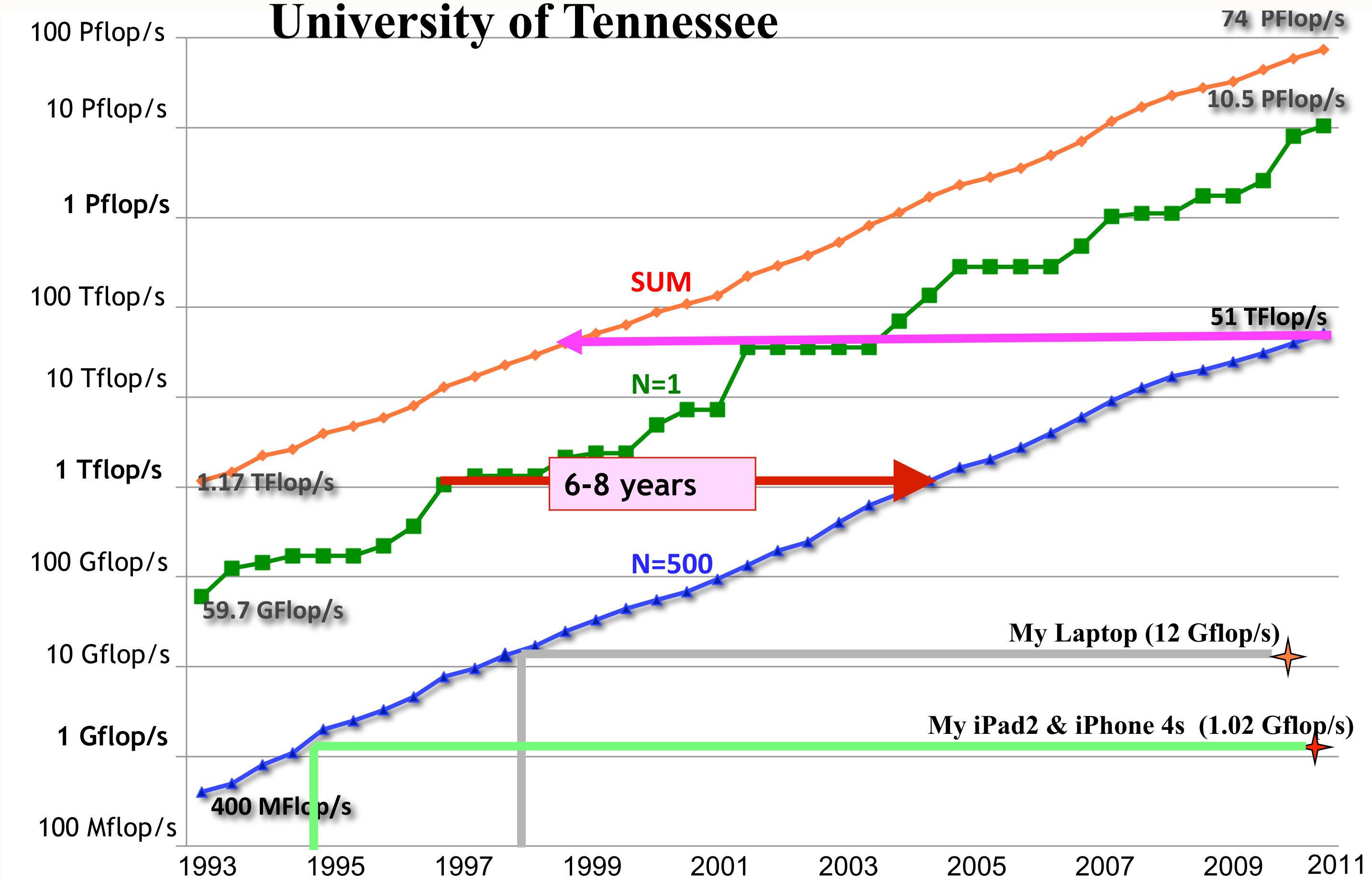


ALL THE EXAMPLES REQUIRE LARGE AMOUNT OF SENSORY DATA AND STRONG COMPUTATIONAL POWER

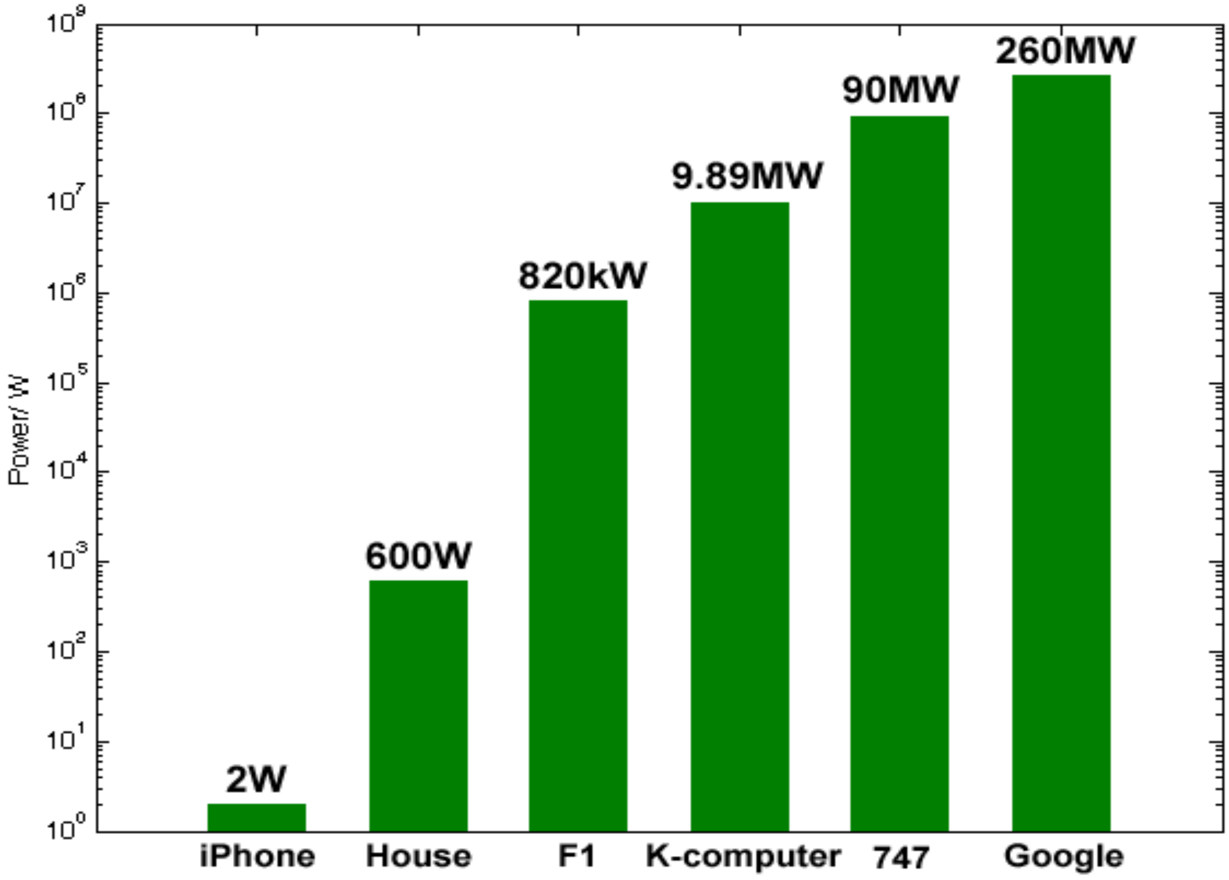

Enterprise Solutions
Smartphone Apps
Smarter Planning
Consulting for
the Government
Sensory data
Sensory data
Sensory data
Sensory data

COMPUTING IS GETTING FASTER

Courtesy Prof. Jack Dongarra
University of Tennessee



…AND CHEAPER


Power/ W
10^9
10^8
10^7
10^6
10^5
10^4
10^3
10^2
10^1
10^0
2W
600W
820kW
9.89MW
90MW
260MW
iPhone
House
F1
K-computer
747
Google

IDEAS IN A NUTSHELL

- Understanding urban dynamics is helpful to increase city life quality.
- The moving vehicle can be an ideal carrier for sensing tasks.
- Computation plays an important role in the information mining.

清華大學
Tsinghua University

ACKOWLEDGEMENTS

Thanks

For information and datasets :
http://sensor.ee.tsinghua.edu.cn