Monitoring and Mining Sensor Data in Cloud Computing Environments

Wen-Chih Peng and Yu-Chee Tseng

Dept. of Computer Science

National Chiao Tung University
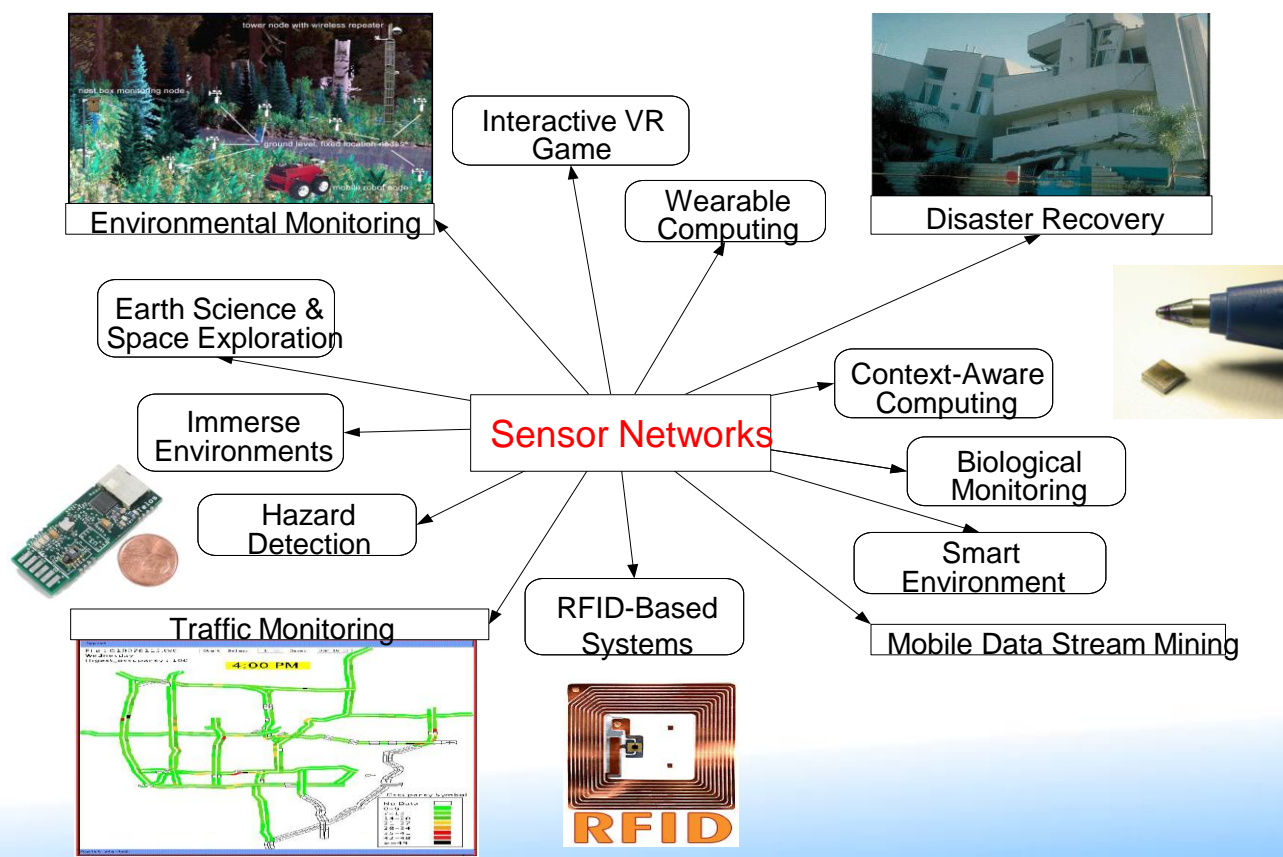
{wcpeng, yctseng}@cs.nctu.edu.tw

Outline

- Sensor applications
  - Tai-Chi platform
  - CarWeb platform
- Sensor applications meet "Cloud Computing"
  - Our implementation
  - Observations and issues
- Conclusions

Sensor Networks

- Several widely deployed HW/SW platforms
  - Low power radio, small processor and RAM/Flash
  - Collect a large amount of sensing data

Environmental Monitoring

Interactive VR Game

Wearable Computing

Disaster Recovery

Earth Science & Space Exploration

Context-Aware Computing

Immerse Environments

Sensor Networks

Biological Monitoring

Hazard Detection

Smart Environment

Traffic Monitoring

RFID-Based Systems

Mobile Data Stream Mining

Sensor Applications

- Sensing physical worlds via sensor networks

- Monitoring and mining sensor data from "servers"

Cloud Computing Project@NCTU

- ABC (Always Best Connected) cloud service and access platforms
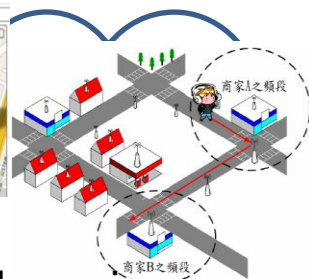
Cloud devices

Applications
- Location-based services
- Vehicular applications
- WSN applications

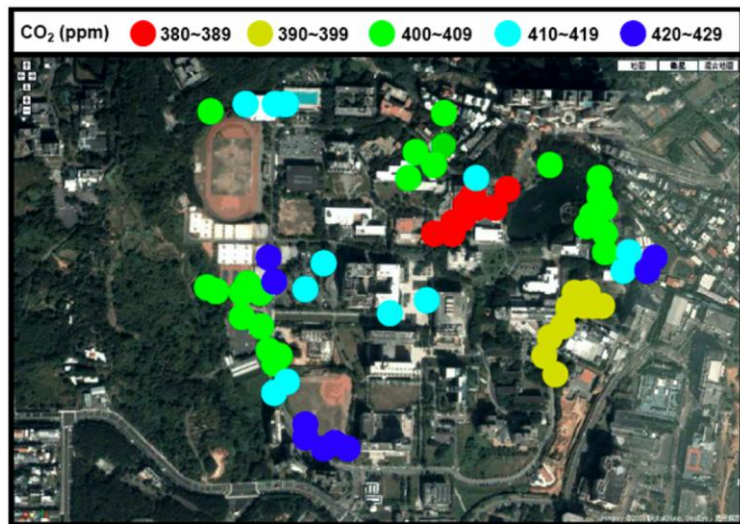**Flexible access**
- Software Defined Radio (SDR)

Cloud

Service platform

Windows Live | Office Live | Exchange Online | SharePoint Online | Microsoft Dynamics CRM Online

Azure Services Platform

Live Services | .NET Services | SQL Services | SharePoint Services | Microsoft Dynamics CRM Services
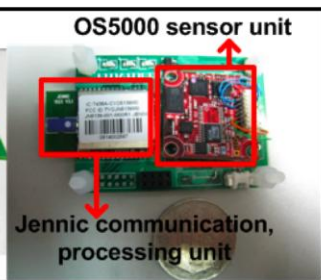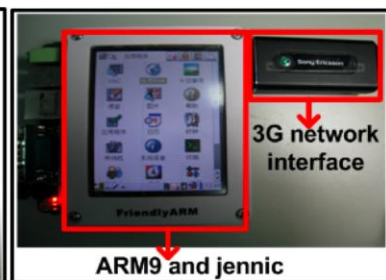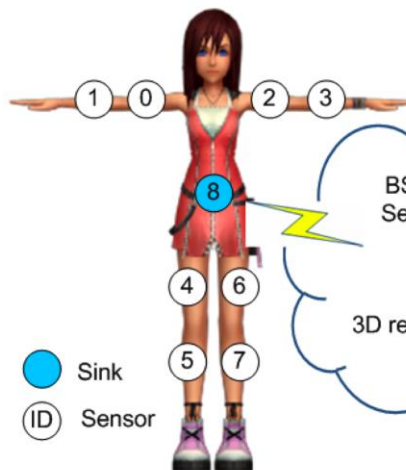
Windows Azure

Microsoft Sora Platform

WSN Applications
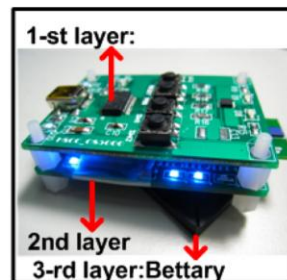
(a) Three-layer sensor node

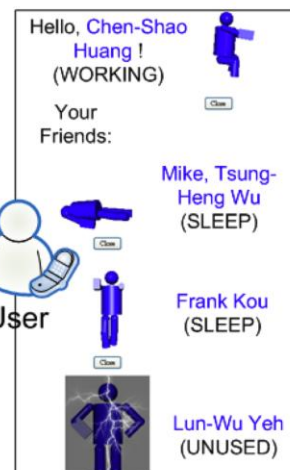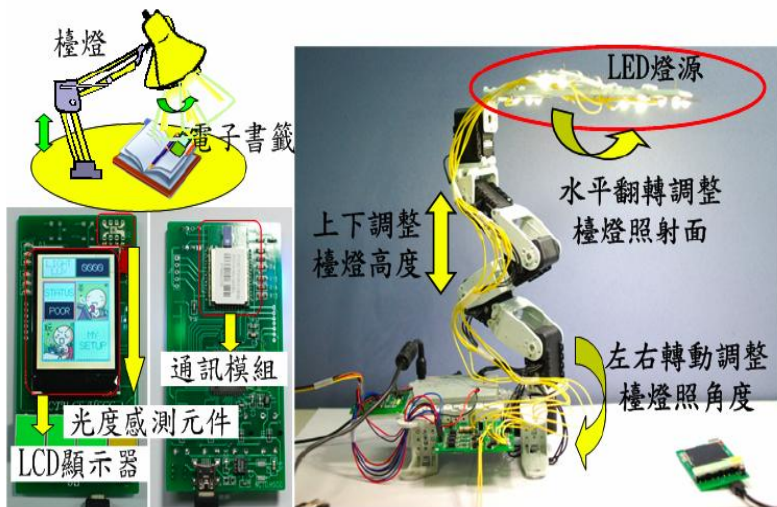(b) 2nd layer: communication, processing and sensing units

(c) sink node

1-st layer:

2nd layer

3-rd layer:Bettary

OS5000 sensor unit

Jennic communication, processing unit

3G network interface

ARM9 and jennic

CO_2 (ppm) ● 380~389 ● 390~399 ● 400~409 ● 410~419 ● 420~429

檯燈

電子書籤

上下調整 檯燈高度

水平翻轉調整 檯燈照射面

左右轉動調整 檯燈照角度

LED燈源

通訊模組

光度感測元件

LCD顯示器

電子書籤

Sink

Sensor

(d) sensor deployment

Cloud
BSN community Server/Database

3D rendering/recognition engine

User

Hello, Chen-Shao Huang ! (WORKING)

Your Friends:

Mike, Tsung-Heng Wu (SLEEP)

Frank Kou (SLEEP)

Lun-Wu Yeh (UNUSED)

(e) status of friends

CarWeb
暢行天下!

Share "*physical*" activities on "*virtual*" social network platform

(virtual)
Tai-Chi on-line game

(virtual-physical)
Tai-Chi through BSNs

(physical)
Tai-Chi group exercise

Tai-Chi master

student A

student B

Share *"virtual"* context on a *"virtual"* platform

Share "*physical*" activities via "*physical*" interaction

System Architecture
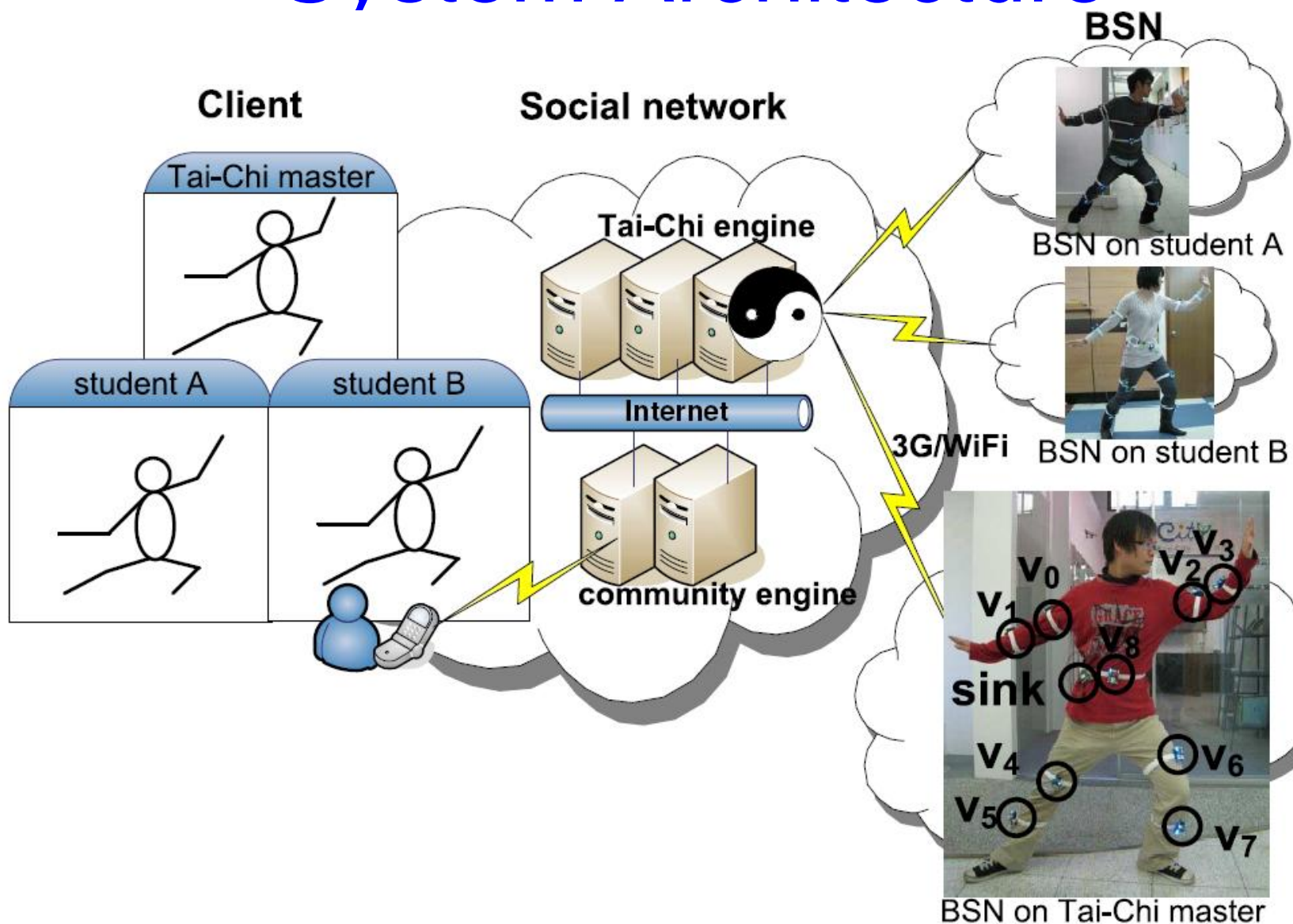
Three main components:

- BSN:
  - Each BSN has 9 sensor nodes and one sink node

- Social Network:
  - Tai-Chi engine
    - Motion database
    - Tai-Chi activities information
    - Motion-mapping module
  - Community engine
    - Tai-Chi Web service
      - Fetch and display motion data through 3D graphic libraries
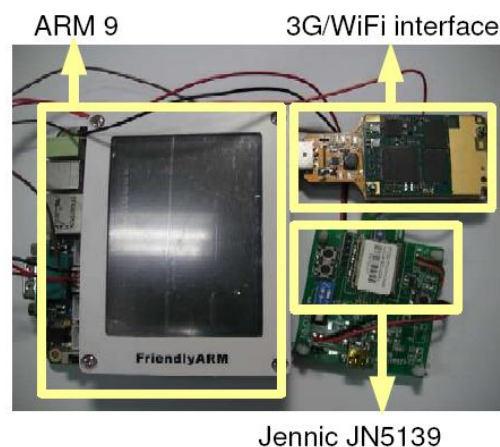
- Client:
  - Web service on Facebook
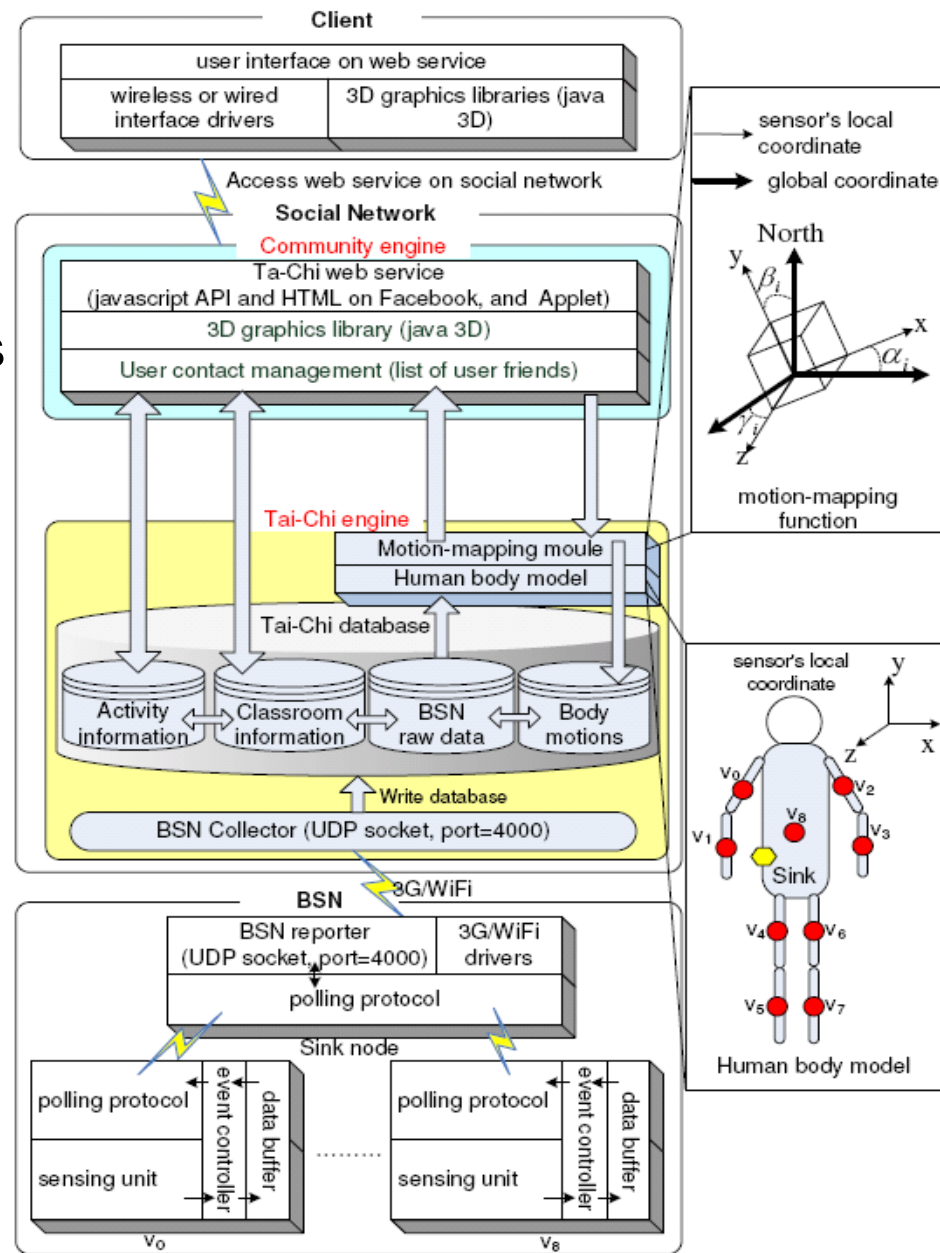
Hardware Design in BSN

- Sensor node:
  - Sensor (os5000):
    - 3-axes accelerometer and a digital compass
  - Micro-controller (Jennic JN5139)
- Sink node:
  - ARM9 microprocessor board
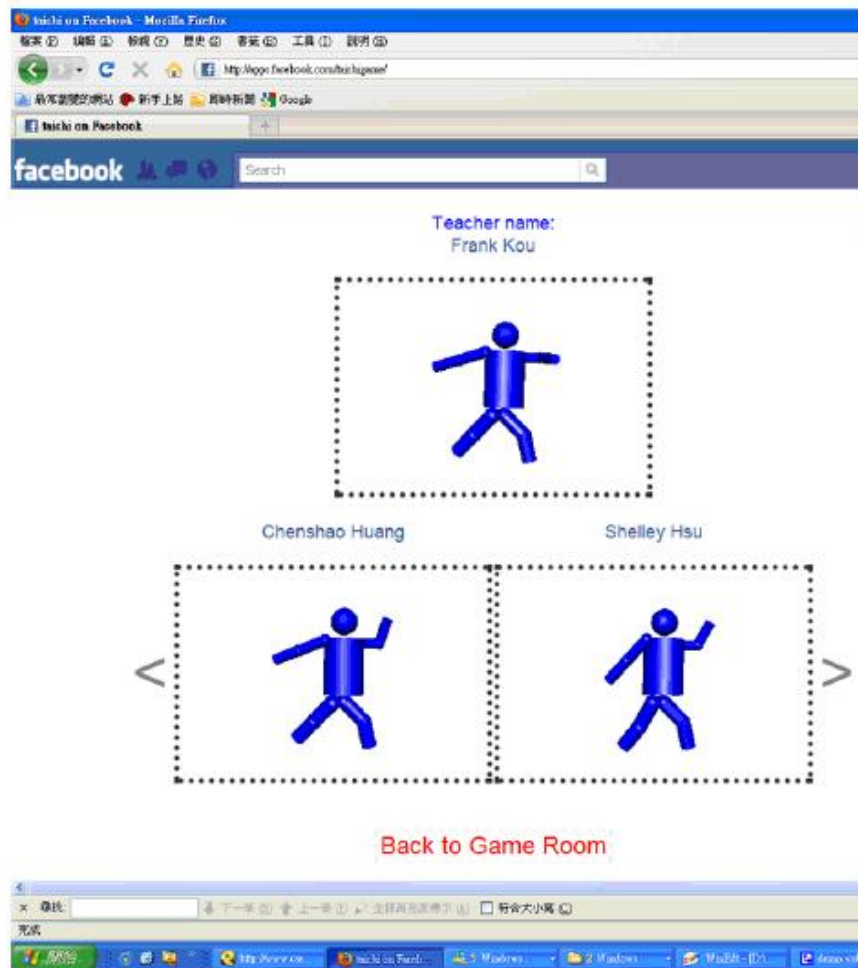   - Coordinate BSN and the 3G/WiFi network
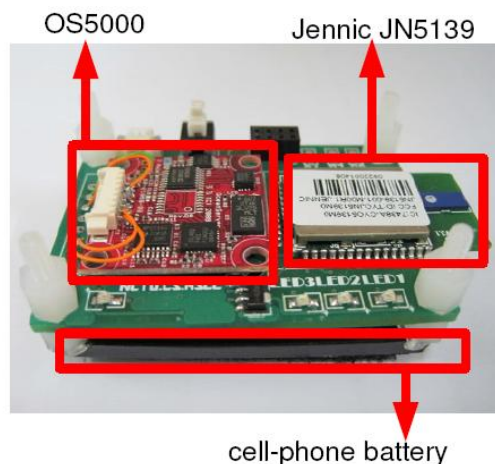
(a) Sensor node   (b) Sink node

Tai-Chi Demo

- Tai-Chi Web service on Facebook

(a) User interface on Facebook.   (b) Poses of 3 users.

BSN-based Social Network Analysis

- Three phases:

Motion evaluation
•Extract motion feature
•Evaluate wave similarity

→

Classifications

→

Recommendations:
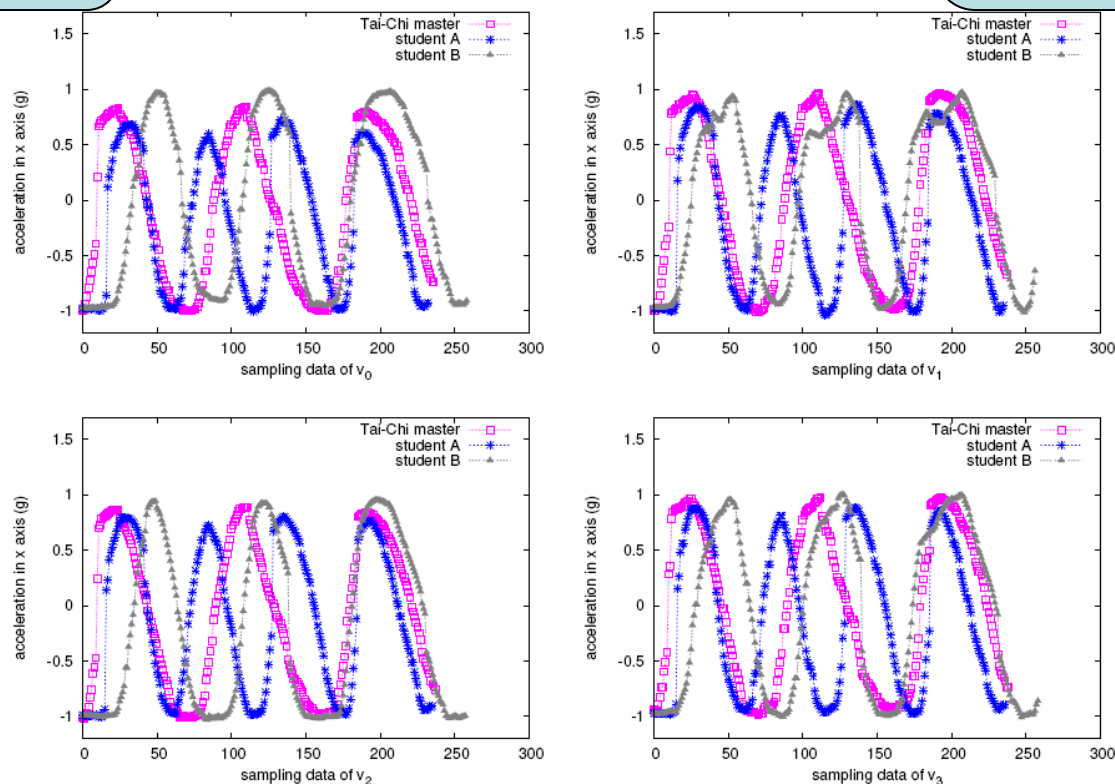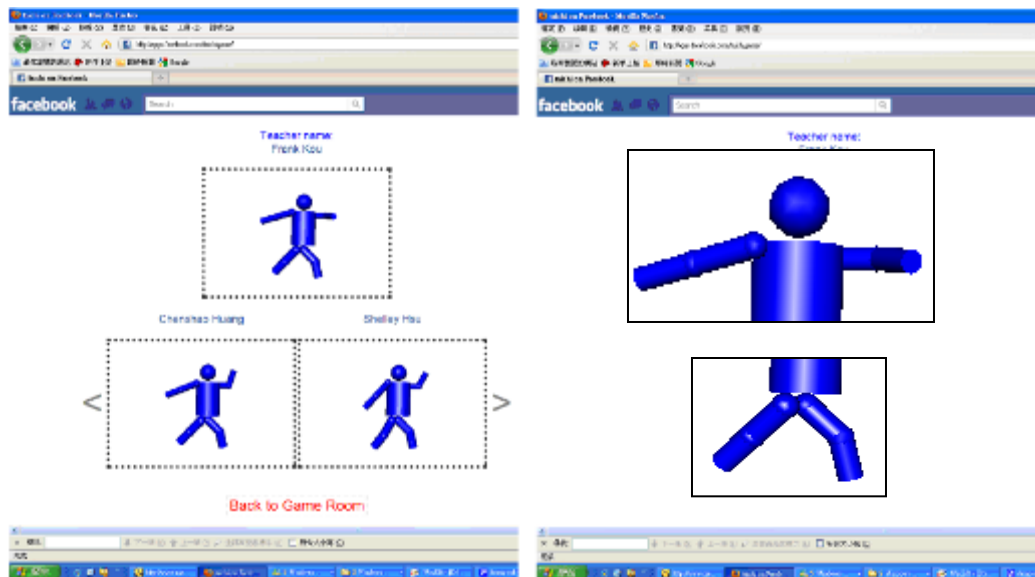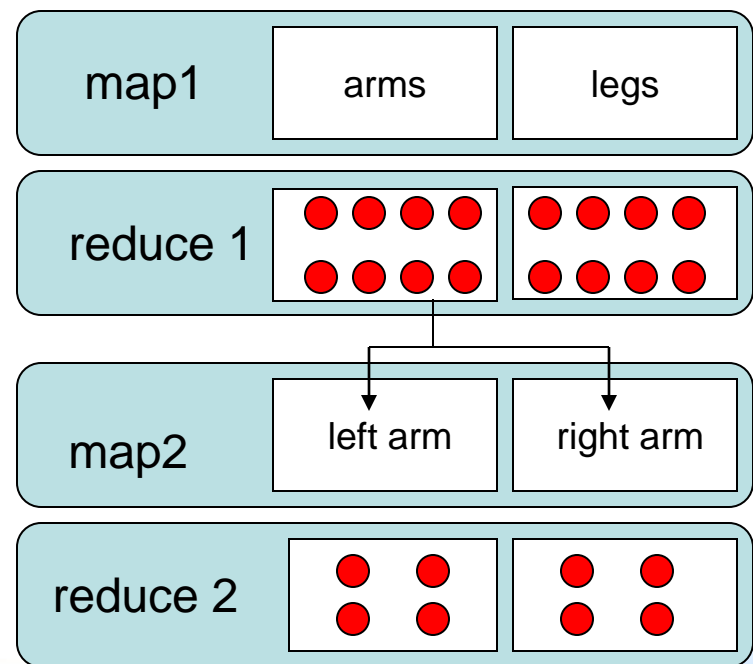•Social link establishment
•Tai-Chi Activity
•Upgrade skills

Fig. 7. The sensory data of v_0, v_1, v_2, and v_3.

Cloud Computing for Ta-Chi

- Support zoom in/out and different view for rendering user motion
  - Use local/partial sensor reading to render motions
- Similarity computation of sensor readings from large-scale of users
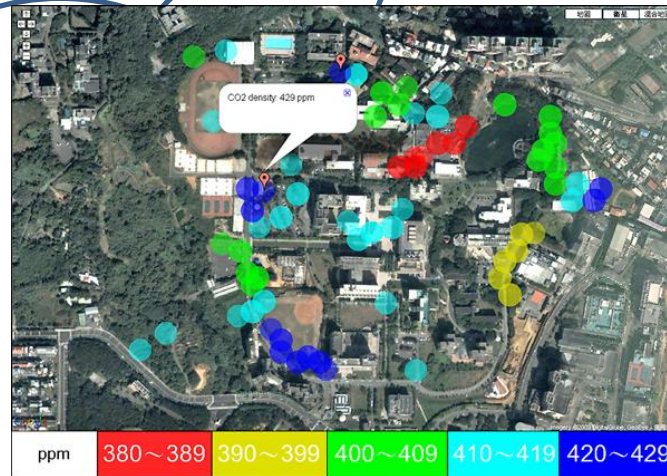

Diversity of motion rendering

Similarity matching computation

Environmental monitoring service on the clouds

Urban Air Quality Measure:CO2 MAP

- CO_2 density/temperature database and data mining
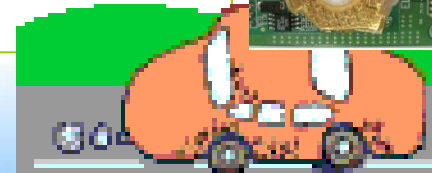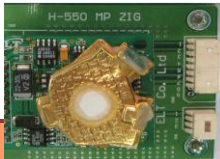- Urban air quality community

subscribe   publish

CO_2 sensor module

Environmental information can be published and subscribed:
CO2 monitoring, position locating, and reporting system

Vehicles are equipped with CO2 sensor, GPS module, and GSM/3G module

CarWeb
暢行天下!

CarWeb Platform

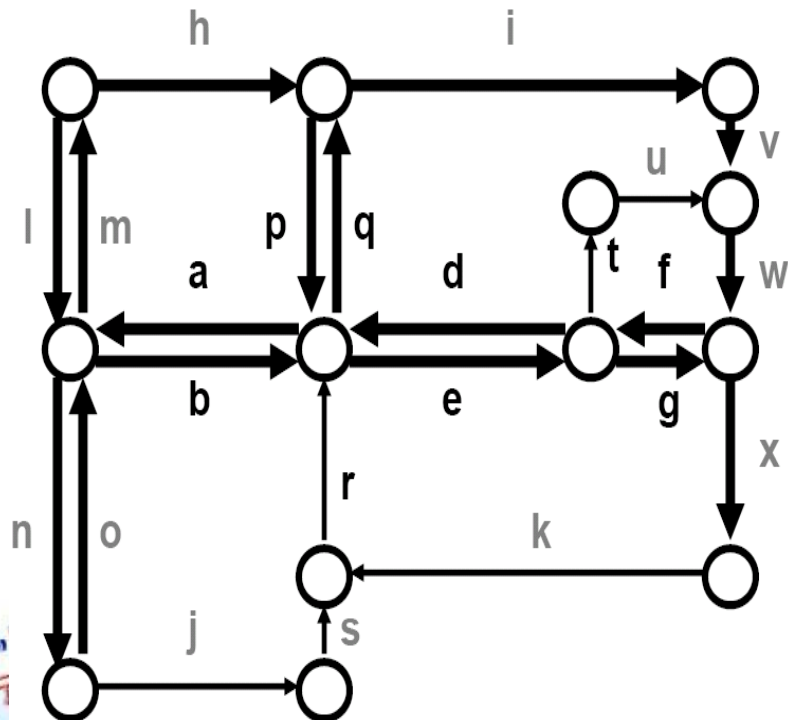# Probe and Sense via Vehicles

Assumption

- Traffic database
  - Road network
    - Road category: freeway, urban road, street and lane
  - Car information

CID	Location	Speed	Time
1	(120.99988, 24.794255)	52	08:12:09
2	(121.00237, 24.792445)	10	08:12:13
3	(121.01853, 24.798822)	53	08:12:16
4	(121.00853, 24.798001)	40	08:12:11
5	(120.99800, 24.796028)	80	08:12:07

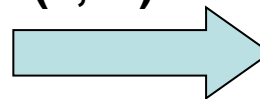# Problem

- Input
  - A traffic database
  - Query (road segment, time)
- Output
  - A speed of the query road segment

CID	T1	T2	T3	T4
1	(e, 50)	(e,50)	(e,52)	(e,52)
2	(e,0)	(e,0)	(e,0)	(e,0)
3	-	(b,50)	(b,54)	(b,53)
4	(e,40)	(e,48)	-	-
5	(a,80)	(a,79)	(a,77)	(e,80)

Query:

(e,T4)

EID	T4
e	50

Observation

- Less traffic information in real time

CID	T1	T2	T3	T4
1	(e, 50)	(e,50)	(e,52)	(e,52)
2	(e,0)	(e,0)	(e,0)	(e,0)
3	-	(b,50)	(b,54)	(b,53)
4	(e,40)	(e,48)	-	-
5	(a,80)	(a,79)	(a,77)	(e,80)

Road Segment e →

CID	T1	T2	T3	T4
1	50	50	52	52
2	0	0	0	0
3	-	-	-	-
4	40	48	-	-
5	-	-	-	80

Spatio-Temporal Weighted Approach

Query
(Time, Road Segment)

Data Selection
• Temporal Factor
• Spatial Factor

WMA-Like

Traffic Database

Historical Data

Status of the Query Road
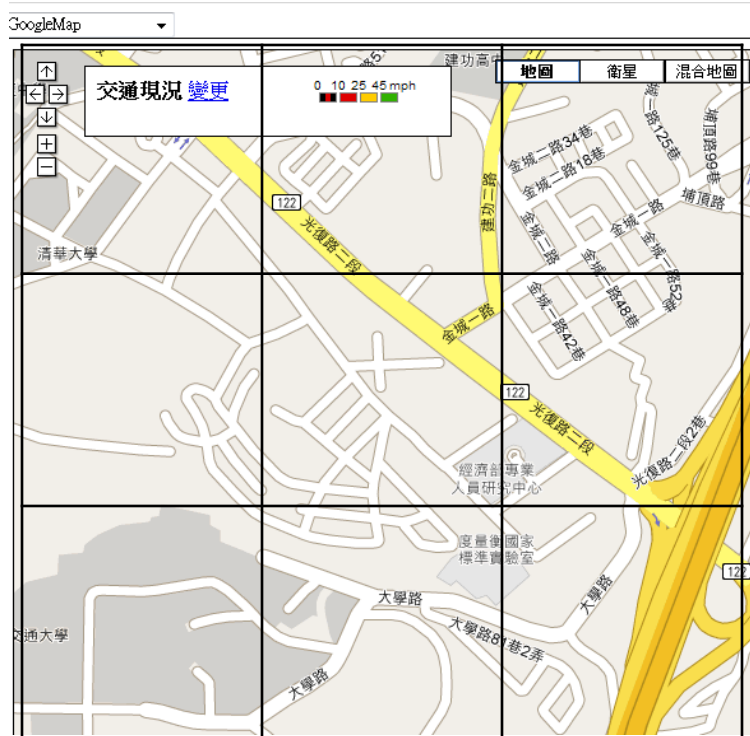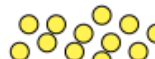
• Temporal Factor
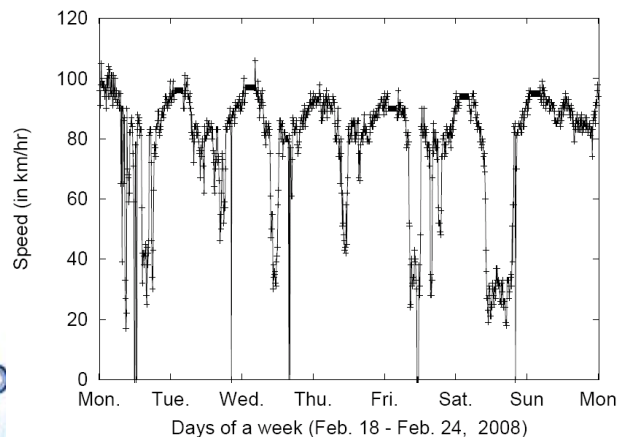


• Spatial Factor
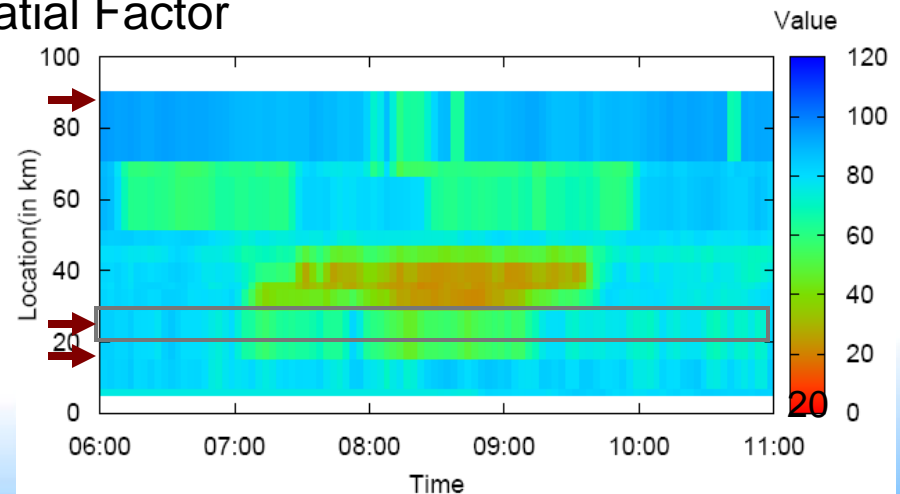
CarWeb Demo

- Users use their smart phones to report and get traffic status via our servers (video demo)

- Given a range query, how to efficiently estimate traffic status of all road segments within the range specified
  - Computation intensive

Cloud Computing for CarWeb

- Support zoom in/out for traffic status estimation
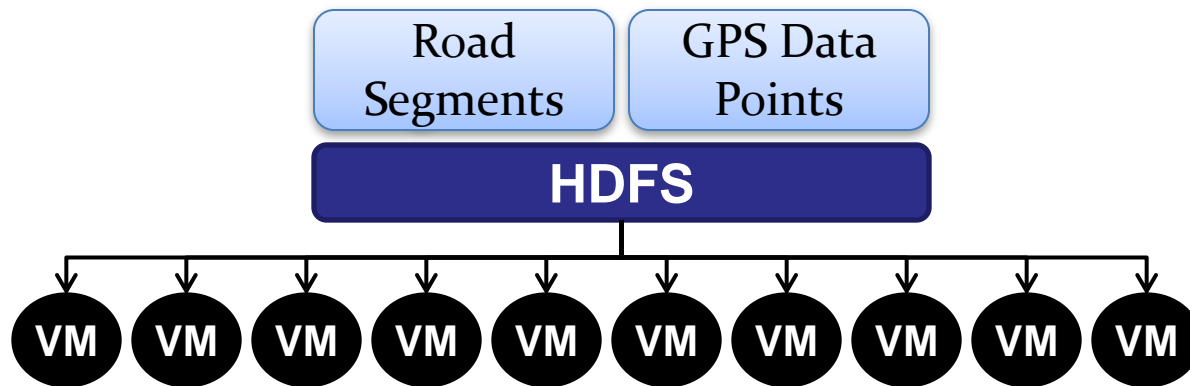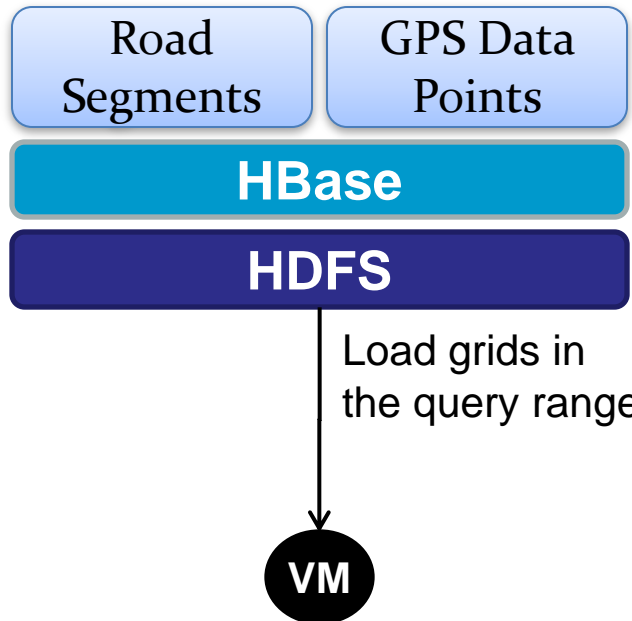- Scalable for calculating traffic status
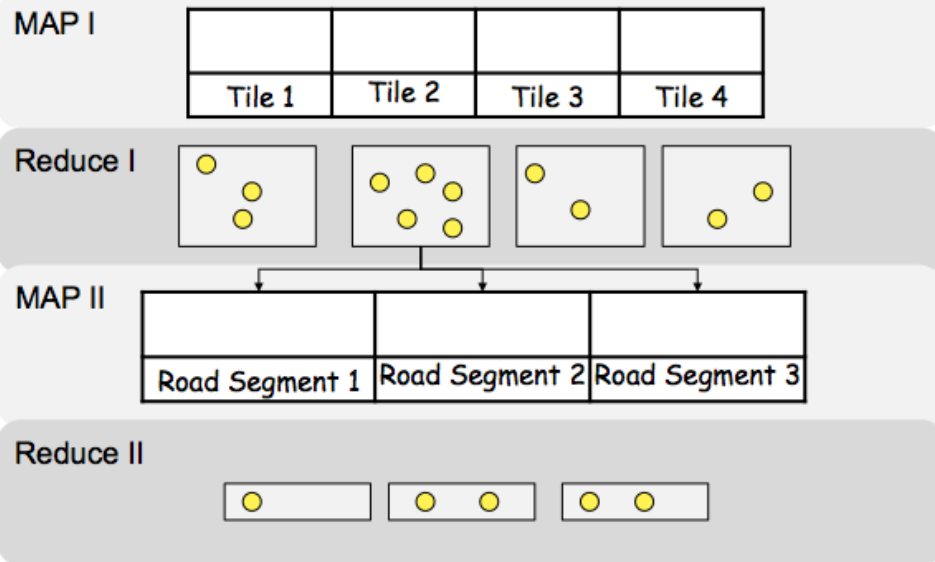
Our Implementation

- Execution time
  - HDFS + MapReduce(10 VMs): 12 min. 11 sec.

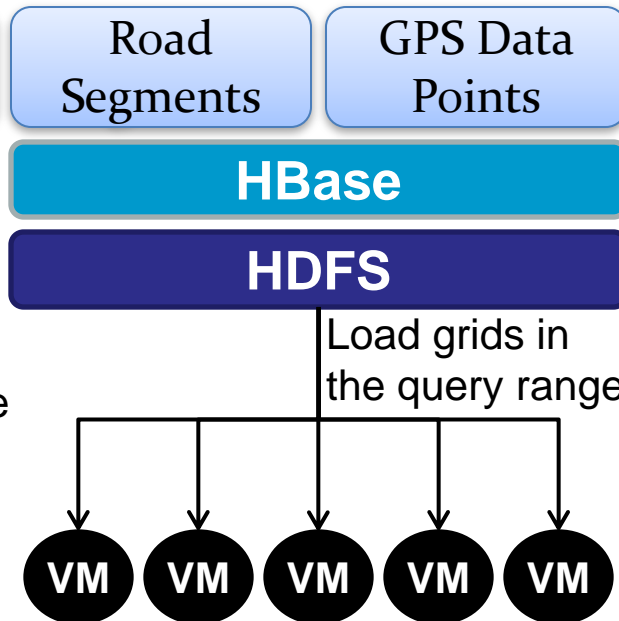- Performance improvement
  - Construct map index for road segments/GPS data points by grids
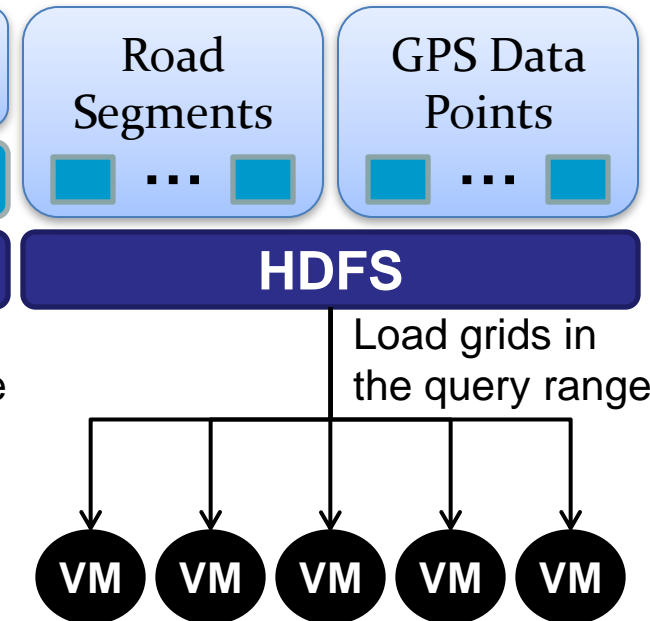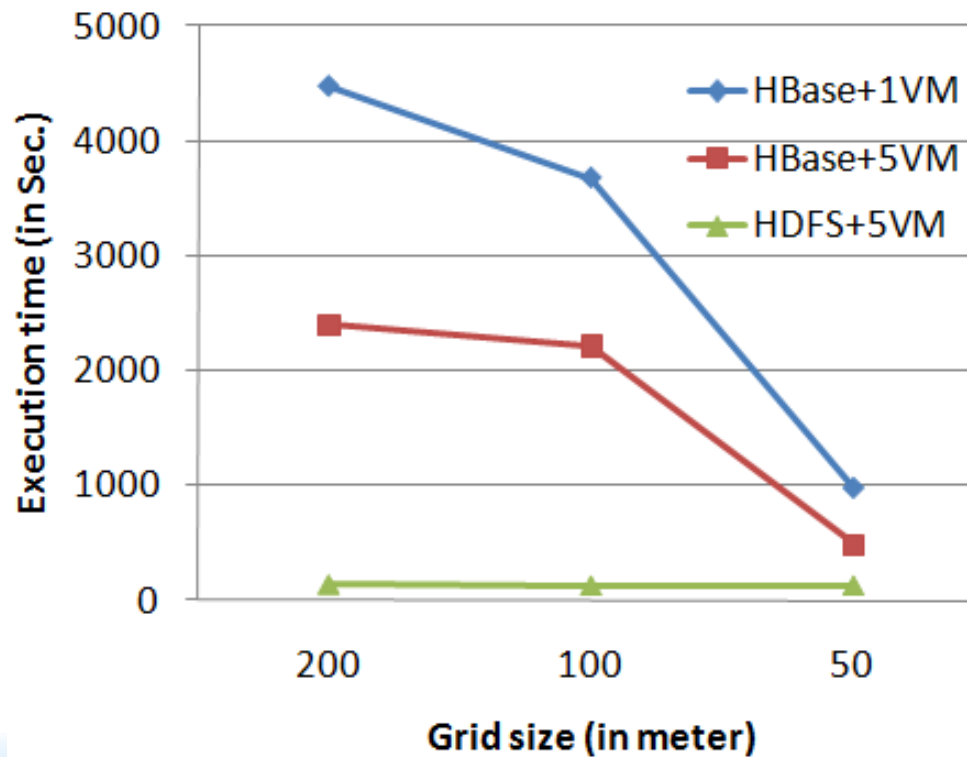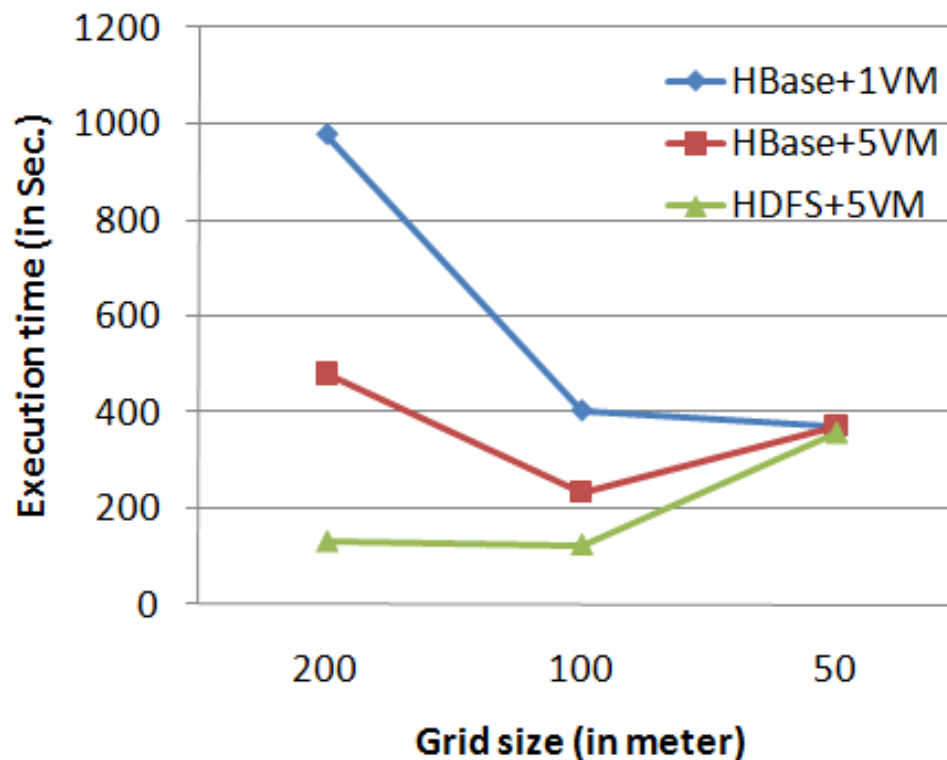 - Increase the number of VMs

Approaches

Performance Evaluation

- Effect of grid sizes for GPS data points
  - Spatial range: 3km x 3km
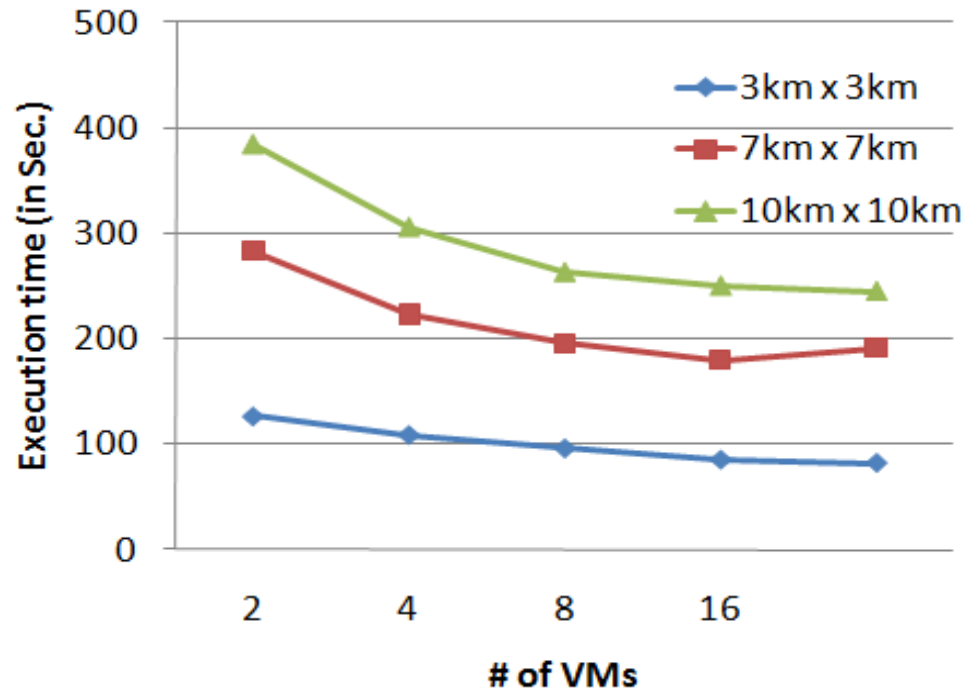 - Grid size for road segments: 2km

Performance Evaluation

- Effect of grid sizes for road segments
  - Spatial range: 3km x 3km
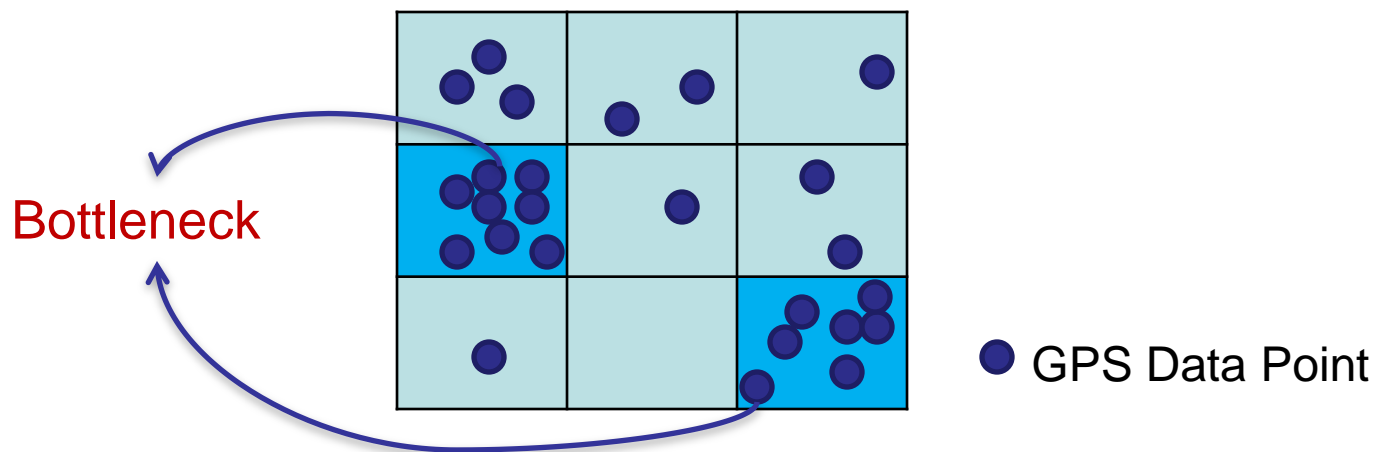 - Grid size for GPS data points: 50m

Performance Evaluation

- Effect of # of VMs with spatial ranges varied
  - Approach is based on HDFS
  - Grid size for GPS data points: 50m
  - Grid size for road segments: 2km

Possible Issues

- GPS data points do not uniformly distributed

Bottleneck

● GPS Data Point

- In MapReduce, design a method to split a data file into several smaller data files while considering data distributions (load balance)

Possible Issues

- Index structures for data stored in cloud computing platforms
  - The time for loading data for MapReduce programming model is much costly
- For real-time sensor data, a batch mode should be re-designed

Conclusions

- A new paradigm for cloud computing (Sensor Clouds): using sensors to collect information o physical things and putting all sensor data into cloud computing platforms

- How to efficiently handle sensor data for monitoring and mining in clouds
    - Efficient storage and retrievals
    - Partition schemes for MapReduce
    - Deal with real-time sensor data

Temporal Feature

- Neighboring time slots
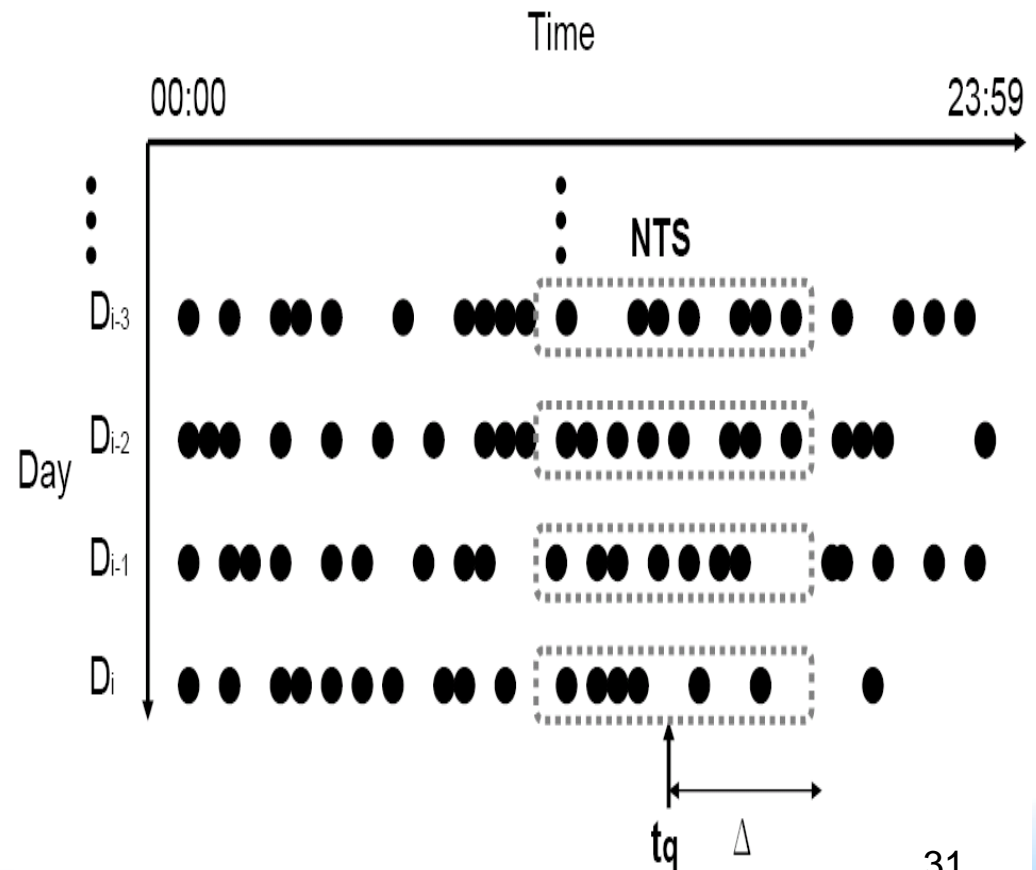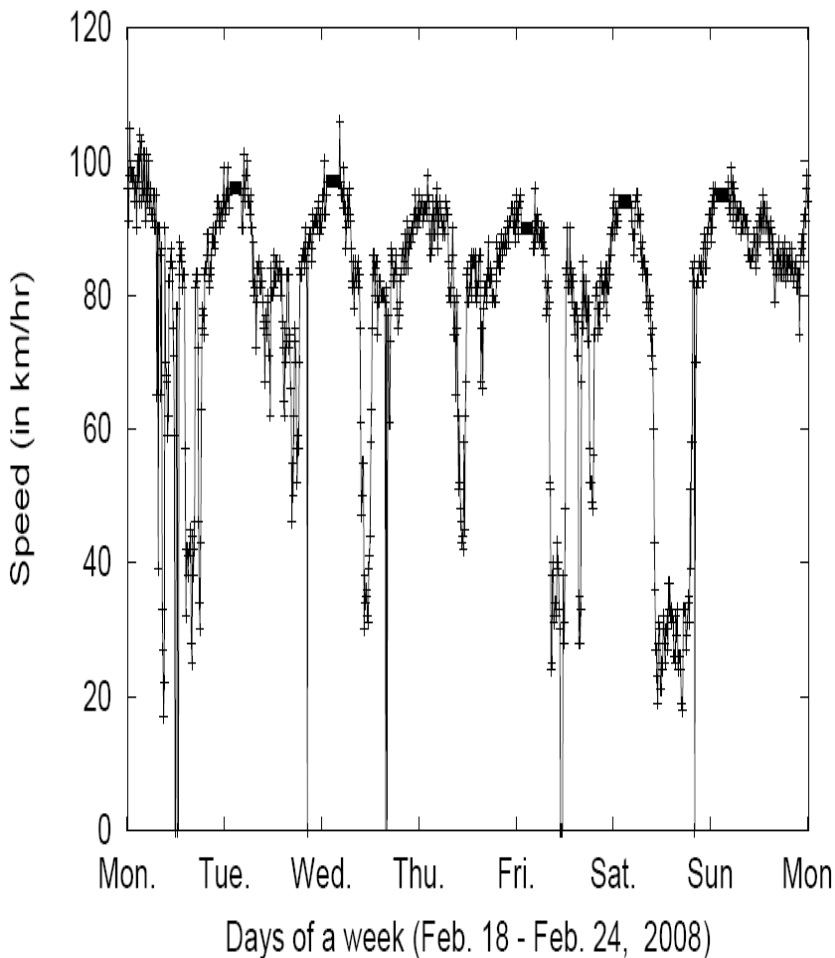
Spatial Feature

- Nearby road segments
- Similar road types