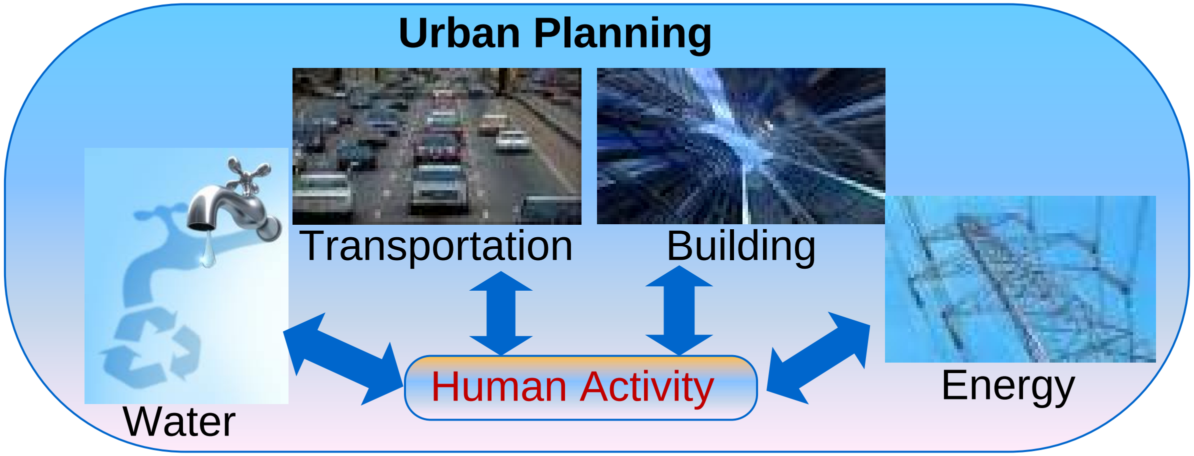
Transportation Activity Analysis Using Smartphones

Fang-Jing Wu
Intelligent Systems Centre
Nanyang Technological University
Singapore

Human Activity v.s. Urban Planning

- Human activity in urban space casts light on different aspects of urban planning.
  - Transportation activity
  - Energy/water usage
- Detailed knowledge of individuals' daily activities would help us understand the inter-relationships among activities in urban space better and thereby improve human wellbeing.

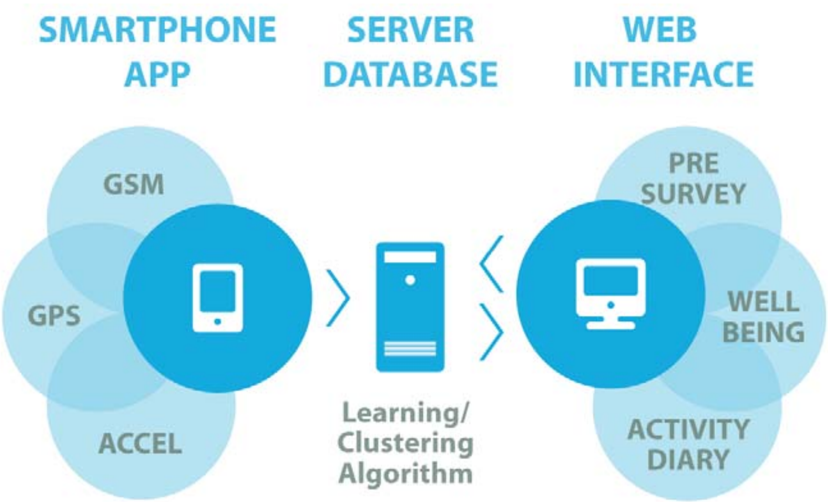


LTA's Travel Survey in Singapore

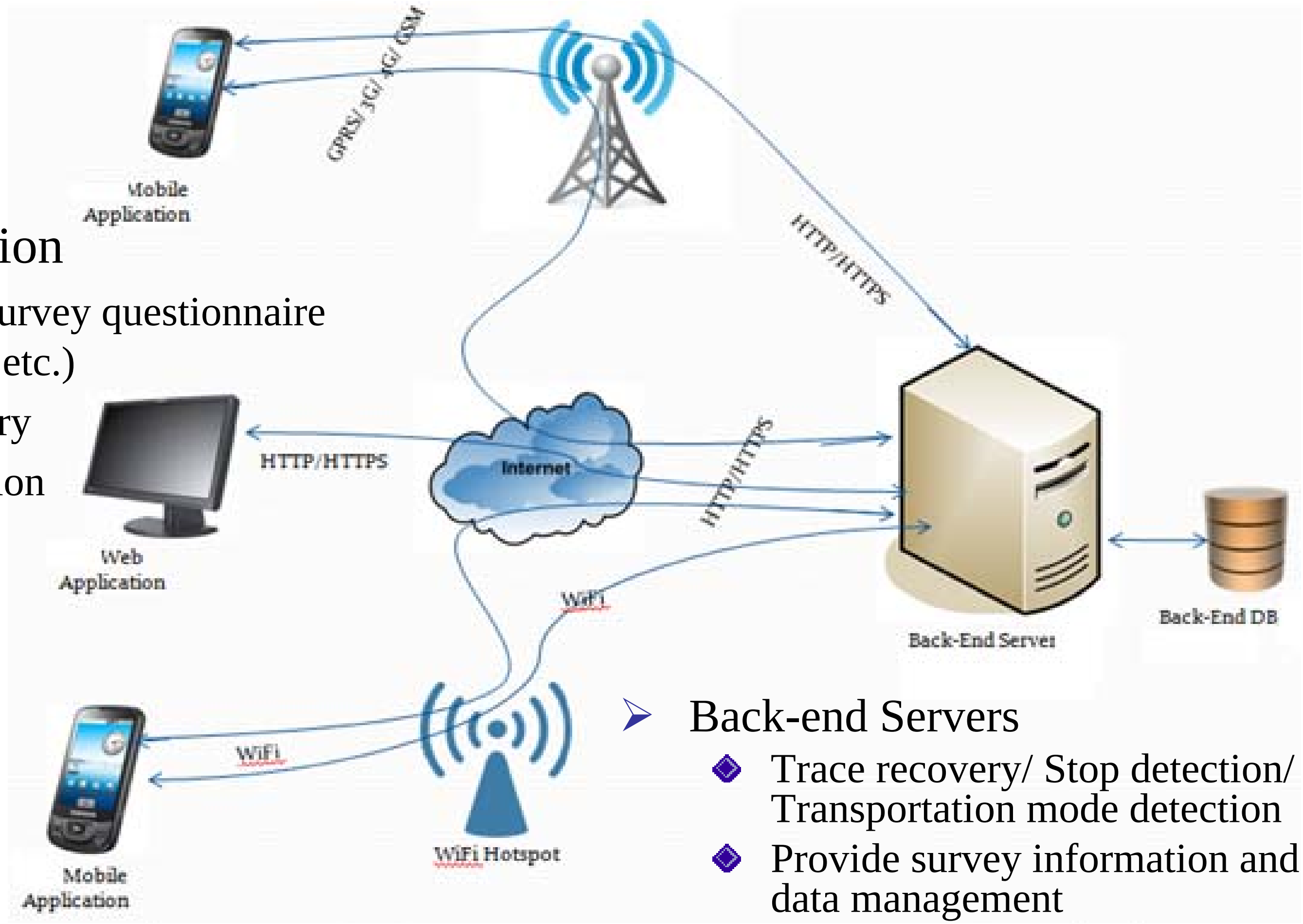
- Land Transport Authority (LTA) conducts surveys once every four years, collecting data on the transport activities of individual household.
  - Investigate when, where and how people travel in urban areas to provide information necessary for urban transportation planning
- Conventional data collection efforts usually involve surveys and questionnaires to be completed by participants.
- These complicated surveys are error-prone and time-consuming.

Future Mobility Survey System in SG

- The "Future Mobility Survey System", developed in collaboration with NTU, SMART, and MIT will be used to support the Land Transport Authority's (LTA) Travel Survey 2012
  - About 10,000 households will be surveyed for the Travel Survey 2012.
  - 1000 users will be involved in the smartphone survey.
- A non-intrusive approach (video)
  - Human mobility will be captured by our smart-phone mobility data logger automatically.
  - The backend analyzes the collected data to recover transportation behavior.
    - Trips / stops / activities/ modes of transportation
  - The web application prompts the participants to verify their mobility.



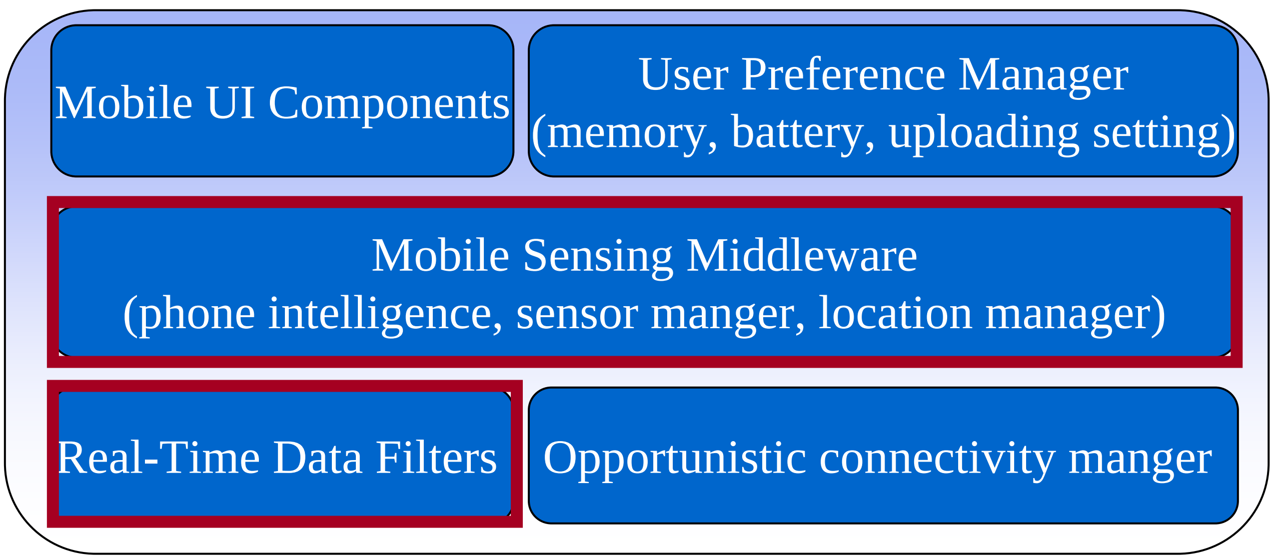
System Architecture



- Web Application
    - Household survey questionnaire (salary, age, etc.)
    - Activity Diary
    - User validation
- Back-end Servers
    - Trace recovery/ Stop detection/ Transportation mode detection
    - Provide survey information and data management
    - Handle both the Web and Mobile applications
- Mobile Application
    - Works as a background service
    - Collect participants' location / movement data
    - Upload data to back-end servers for processing

Mobile Application

- Mobile UI Components and User Preference Manager
  - Manage personal preferences to maximum memory used, battery limit for running mobile app., and how you want to upload data.
- Mobile Sensing Middleware
  - Design a phone intelligence process to reduce energy consumption for sensing
- Real-time Data Filters
  - Pick up high-quality and low-quantity sensing data to reduce energy consumption for uploading
- Opportunistic Connectivity Manger
  - Upload collected data if the connection is available



Reducing energy consumption for "sensing"

Reducing energy consumption for "uploading"

Observations on Energy Use

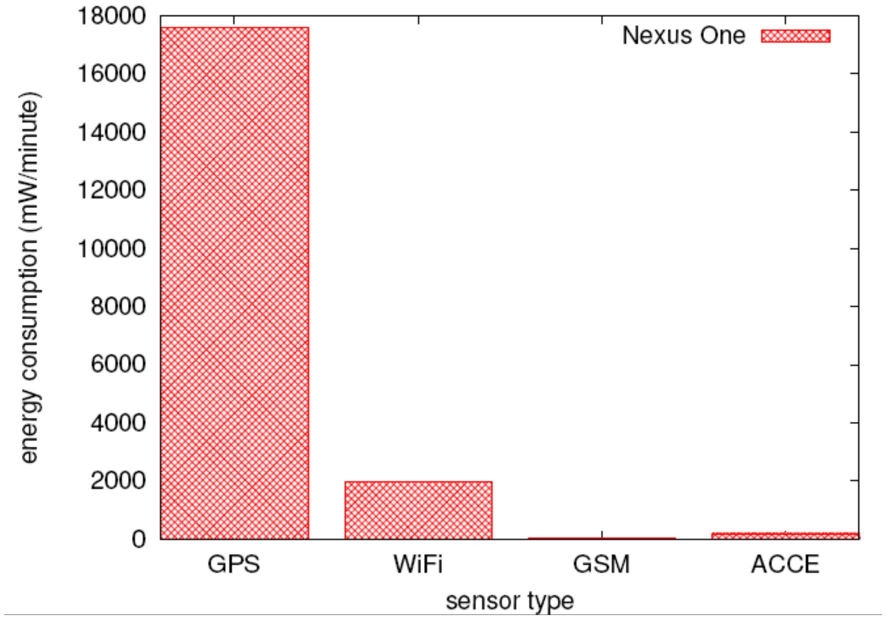
- Study energy use of smartphones to find out energy-hungry components for sensing and uploading tasks
  - For Sensing tasks:
    - GPS consumes much more energy than WiFi, GSM, and accelerometer sensors
  - For uploading tasks:
    - The 3G network takes long time to upload data and consumes much more energy than WiFi networks

Energy Consumption for Sensing (1/2)

- The measuring methodology
  - A Single phone for testing: Google Nexus One
  - Turn on each type of sensors individually to measure energy consumption
    - e.g., for GPS sensing, turn on GPS only (turn off other sensors)
  - Sensing duration: 2hr
  - Power Monitor: Power tutor
  - Measurement metric:
    - Energy/per minute (mW/minute)

Energy Consumption for Sensing (2/2)

- A comparison of energy consumption for different types of sensors



	Total energy consumption (mw)	Total size of collected sensing data (kB)	Avg energy consumption for sensing (mW/minue)
GPS	2110855	1060	17590.46
WiFi	235293	30.4	1960.775
GSM	3269	20.7	27.24167
ACCE	23089	312	192.4083

- For sensing, GPS consumes much more energy than other types of sensors.

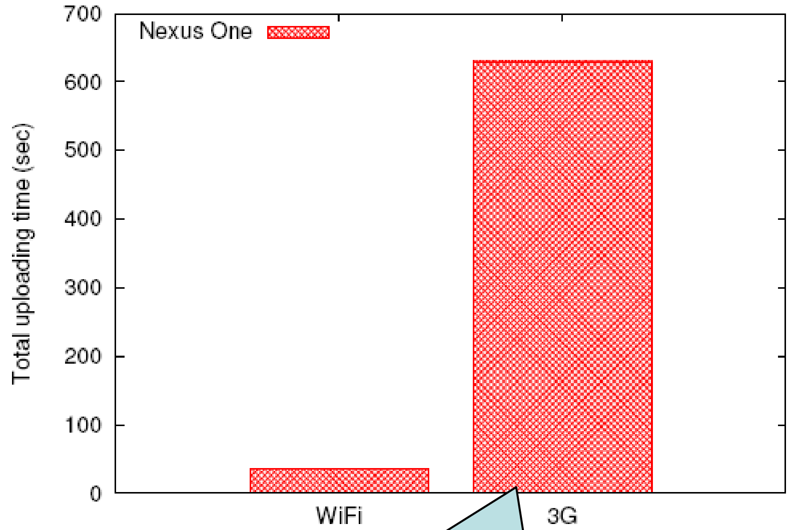
Energy consumption for Uploading(1/2)

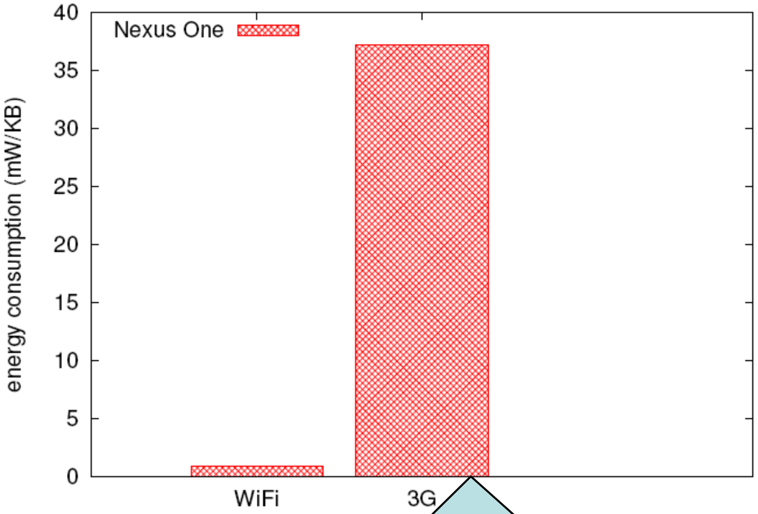
- The measuring methodology:
  - Uploaded data:
    - We prepare 200 JSON files based on the number of JSON files uploaded for a single user per day
    - Each of which has size 80KB,
    - Total uploaded size is 15.4 MB
    - Upload these files via WiFi or 3G
- Measurement metric:
  - Total uploading time (sec)
  - Energy/per unit data collected (mW/KB)

Energy consumption for Uploading(2/2)

- A comparison of energy consumption via different types of networks.

	Total uploading time (sec)	Avg. energy consumption for uploading (mW/KB)
WiFi	36	0.911234
3G	630	37.2422



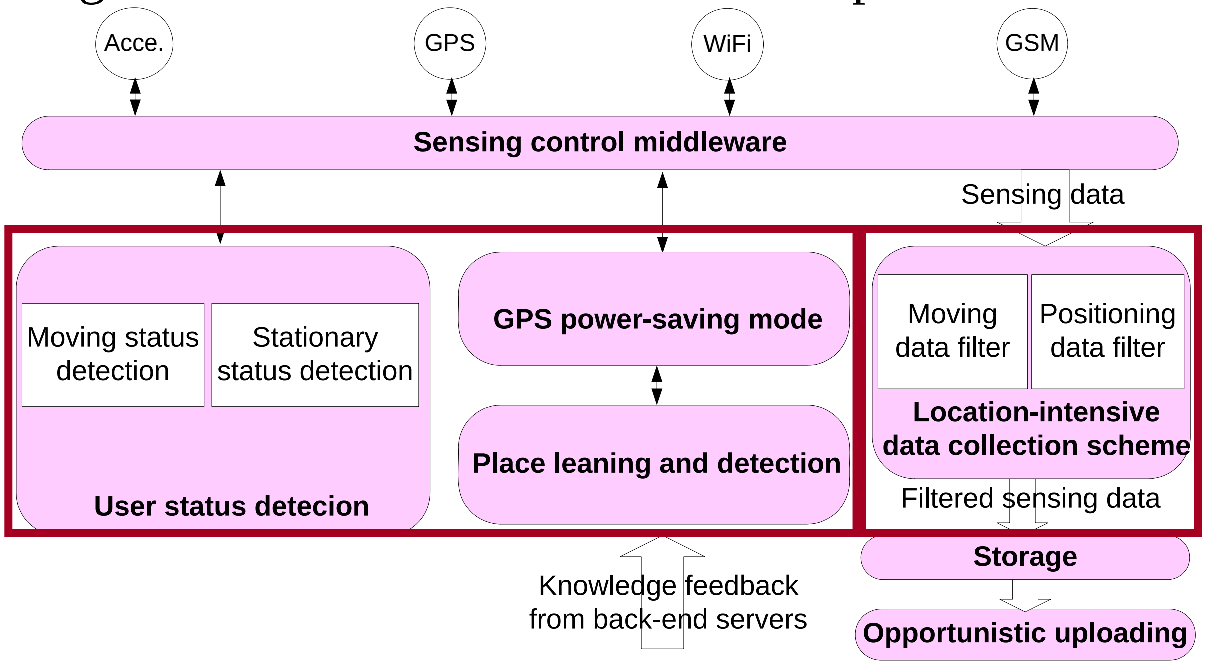


- For total uploading time, 3G is much slower than WiFi

- For energy consumption per unit data uploaded, 3G is much greater than WiFi.

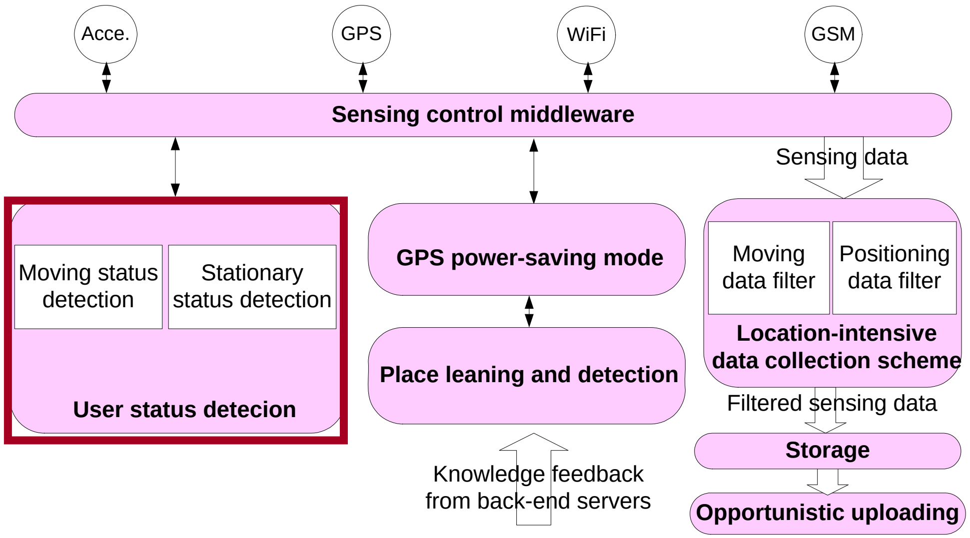
Mobile Client Design

- Reduce energy consumption for sensing
  - Avoid using GPS at users' long-stay places
    - Design a place learning and detection scheme to turn on GPS at an appropriate moment
  - Design a GPS power-saving mode to improve the energy usage of GPS sensor while user moving
- Reduce energy consumption for uploading
  - Design data filters to reduce amount of uploaded data.



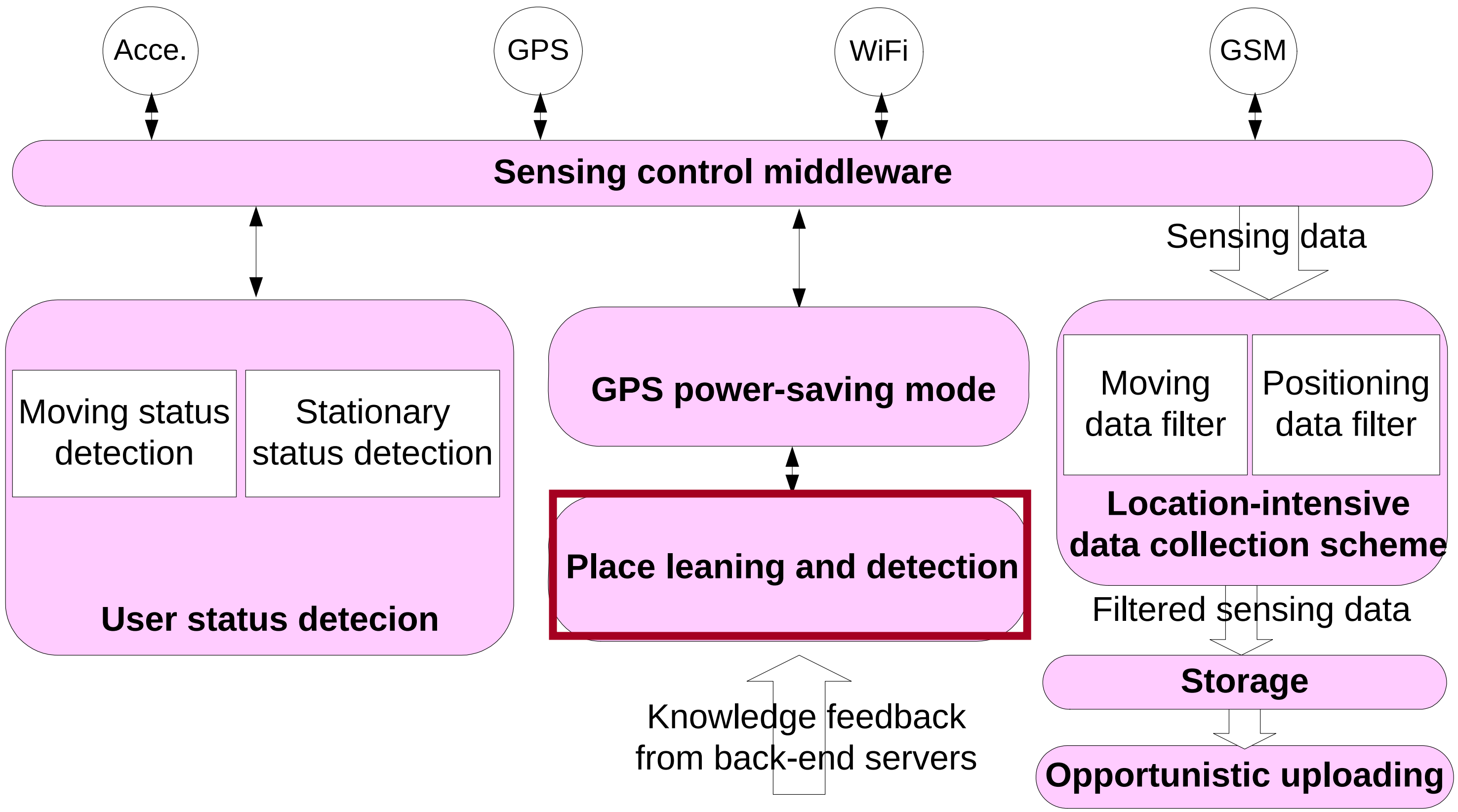
User Status Detection

- Moving status detection
  - A movement accelerometer sample :
    - The accelerometer reading has significant changes
  - Moving detection:
    - Consecutive movements are detected for a long period.
- Stationary status detection
  - A static accelerometer sample:
    - The accelerometer reading has no significant changes
  - Stationary detection
    - Consecutive static samples are detected for a long period.



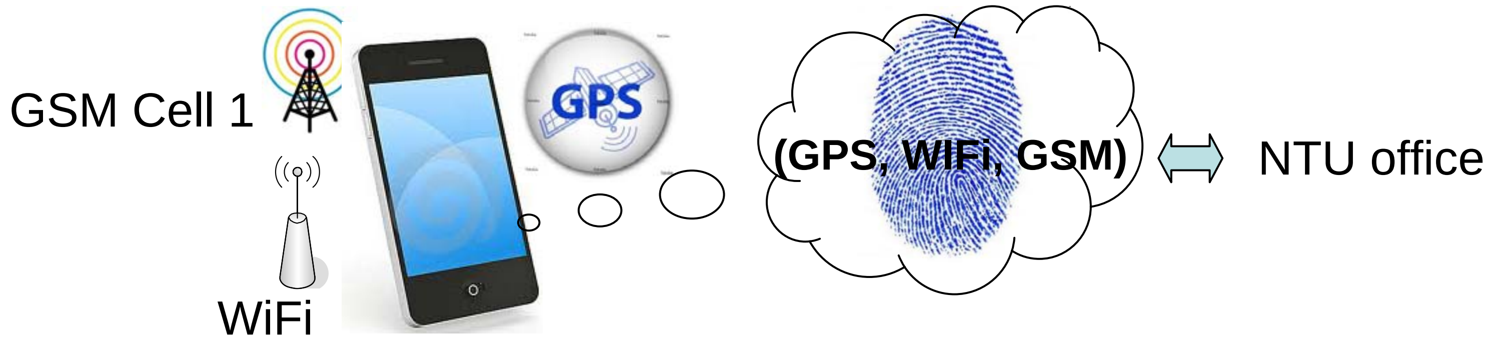
Components of Mobile Clients

- Place learning and detection



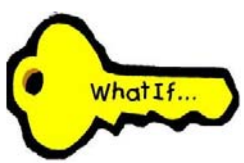
Place Learning and Detection

- Ambiance Signatures v.s. Places
  - Human intelligence:
    - People know a particular 'place' based on UNIQUE ambiance signatures at the place
    - ambiance signatures: buildings, sound, etc.
  - Phone intelligence:
    - How does the phone know 'places' ?
    - A place ←→ a UNIQUE network fingerprint



Place Information Facilitates Sensing

- Network signature is a key clue to detect whether a user is entering/leaving into/from a place so as to turn on/off GPS at an appropriate moment.



Places of Interest v.s. Privacy

- Places of Interest: long-stay places
  - Private places: Home, Offices
  - Public places: Parks, Bus stops, MRT stations

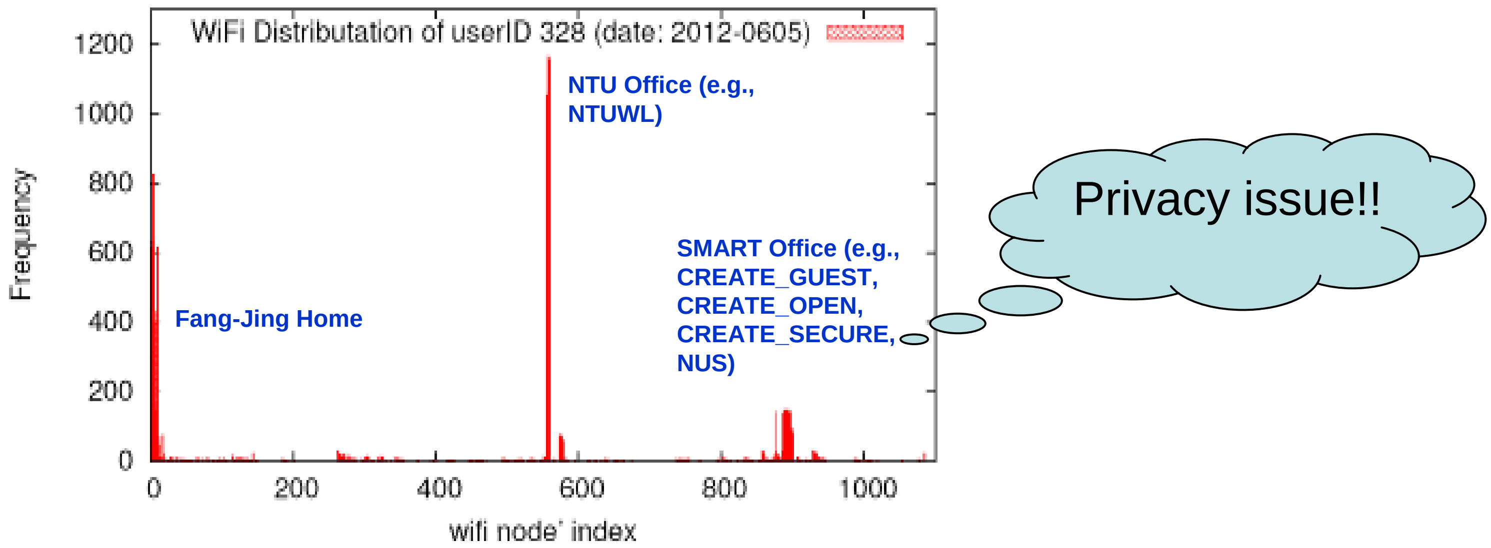
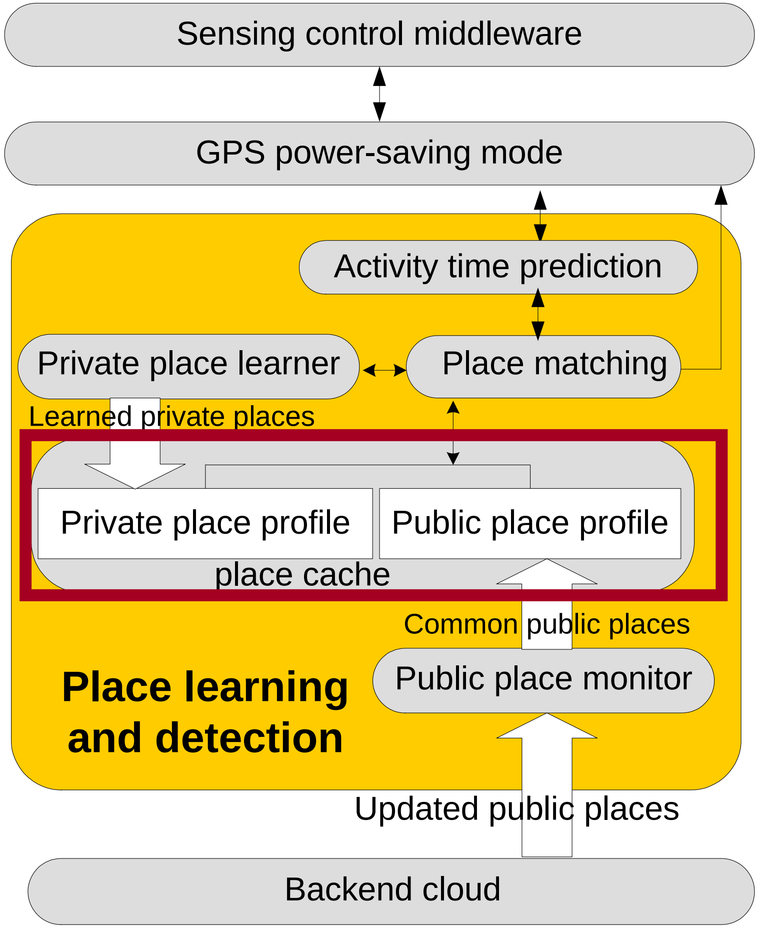


Figure: The frequency distribution of WiFi points for a single user during a single day.

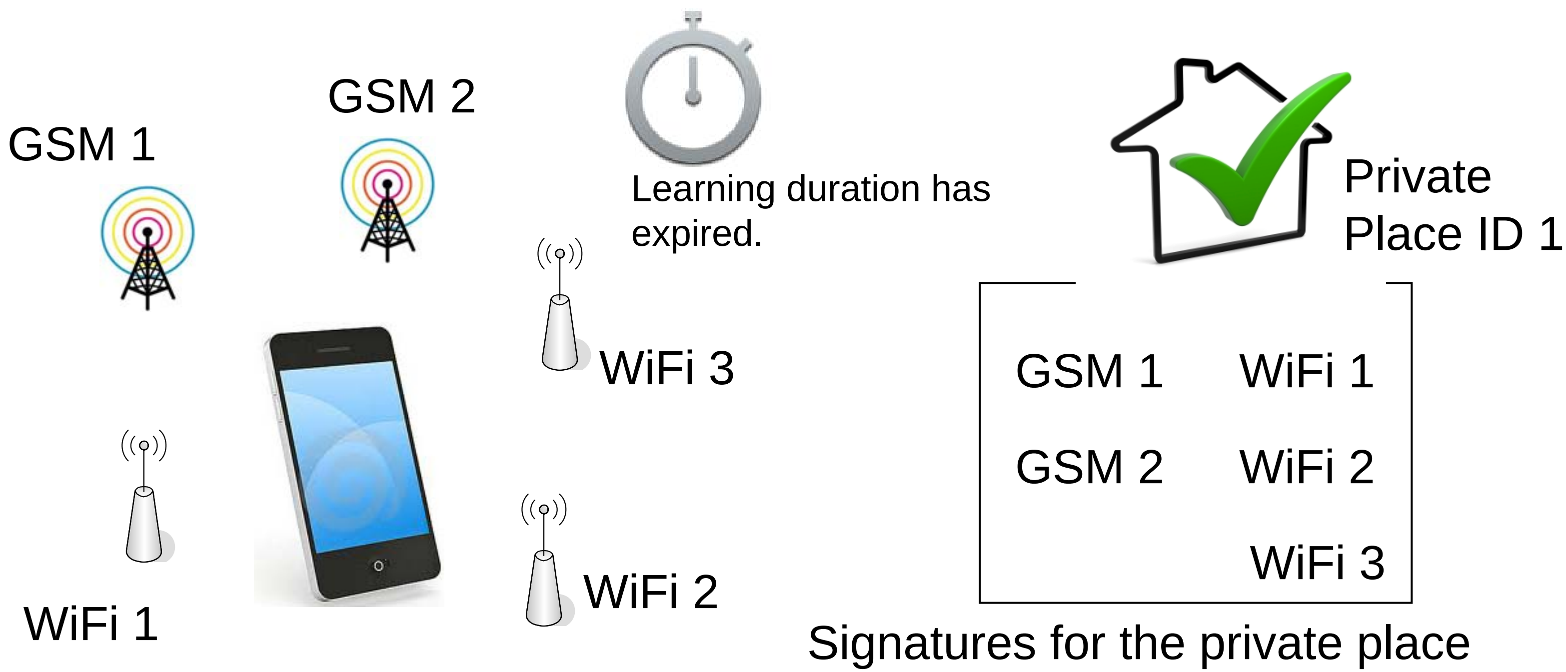
Private/Public Place Profiles

- Two kinds of place profiles:
  - Private places: (high-privacy)
    - Based on the ambient Wifi/GSM signals, the phone keeps learning each long-stay place for 1 hr in real-time way.
    - The private place profile is only kept by each individual for privacy consideration.
  - Public places: (low-privacy)
    - The backend will identify a public place profile based on GPS density in the database
    - The public place profile is shared between all of users (specifically, download the public profile from servers if there is a freshest one).



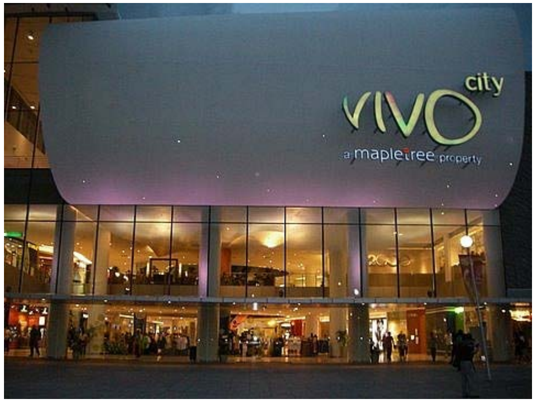
Private Place Learner

- An incremental learning mechanism to find a set of network signatures for a particular place
- During the learning process, there is no need to conduct localization at the place.



Public Place Identification

- Considering all of GPS data points in the whole database, we apply the density-based Spatial Clustering of Applications with Noise (DBSCAN) for identifying the top 100 public places (i.e., ‘hot spots’ for Singapore residents)
  - Each public place is a pair of latitude and longitude.
  - MRT stations, food courts, neighborhood of shopping malls.



Place Matching and Detection

- Private place detection
  - Similarity-based detection
    - We say "MATCHING" if the similarity $(S_G, S_w) > (\tau_G, \tau_w)$, where (τ_G, τ_w) is a pair of predefined thresholds.
    - $(S_G, S_w) > (\tau_G, \tau_w)$ if $S_G > \tau_G$ and $S_w > \tau_w$
- Public place detection
  - For each public place, a fixed activity range is defined by the circle centered at the public place with radius R_a
  - The phone will conclude the user at the public place if it gets a real-time GPS fix the activity range of the public place.

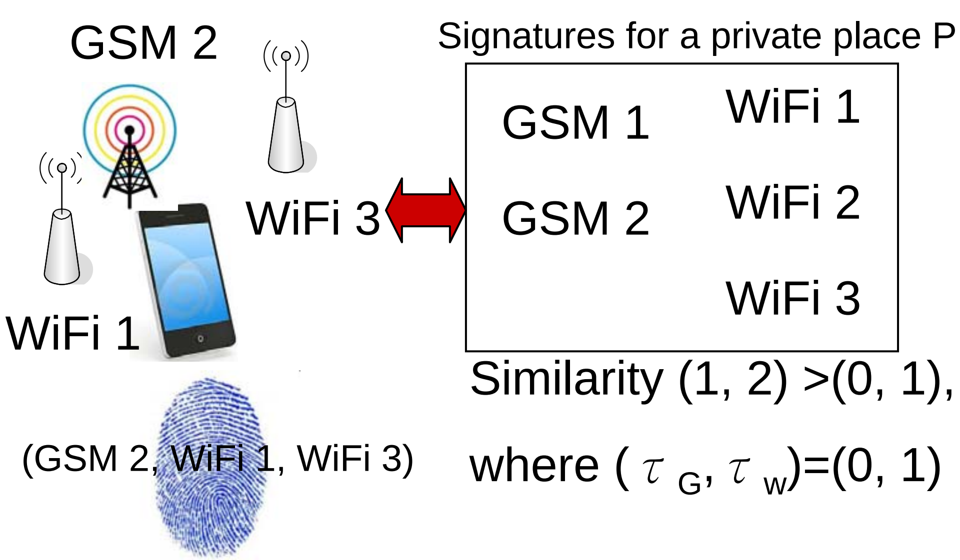


Fig (a): Private place detection

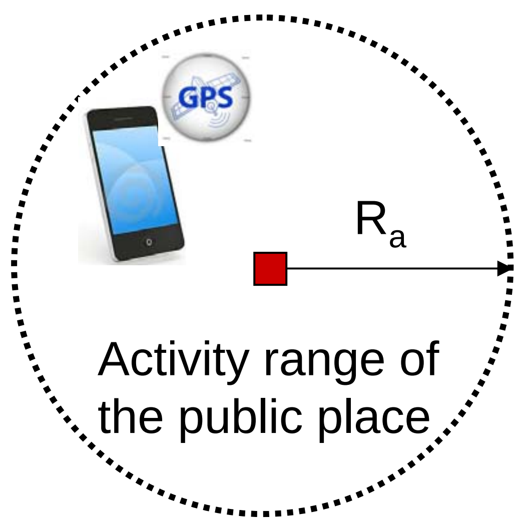
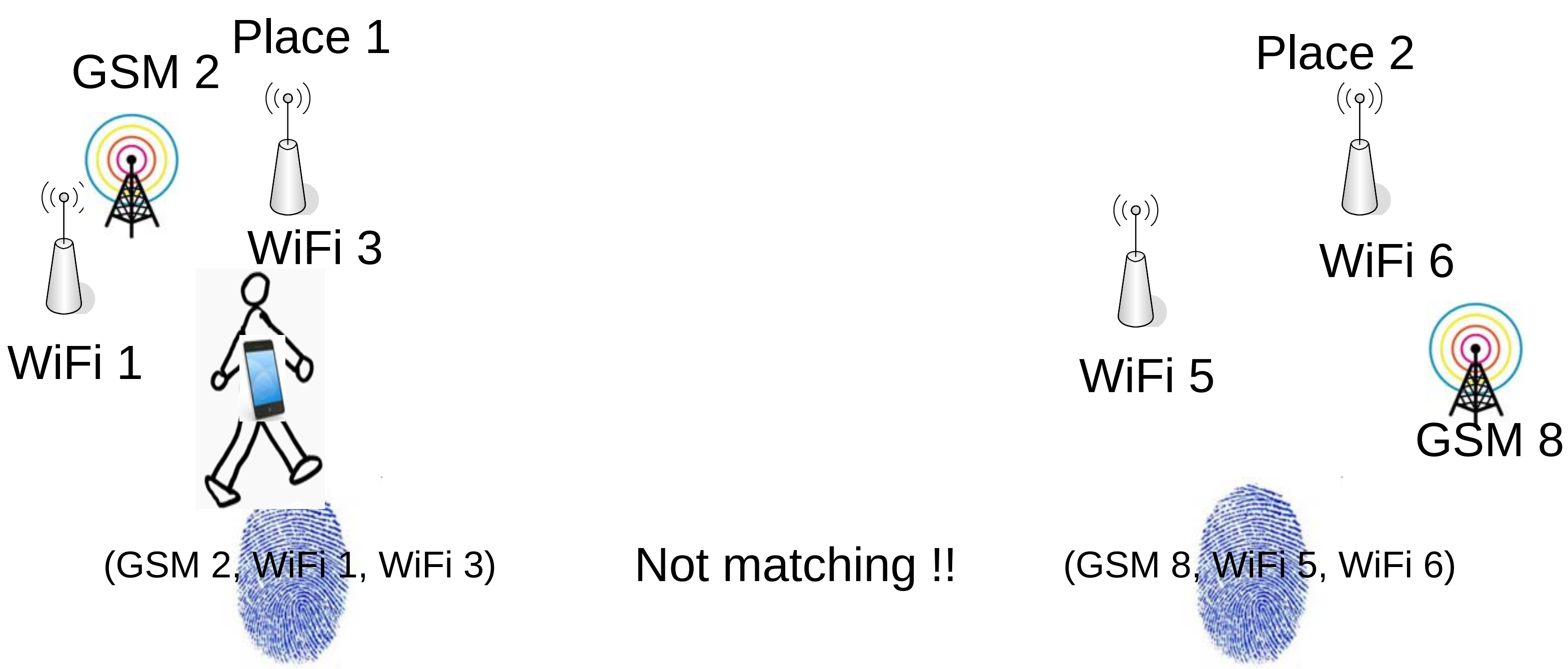


Fig (b): Public place detection

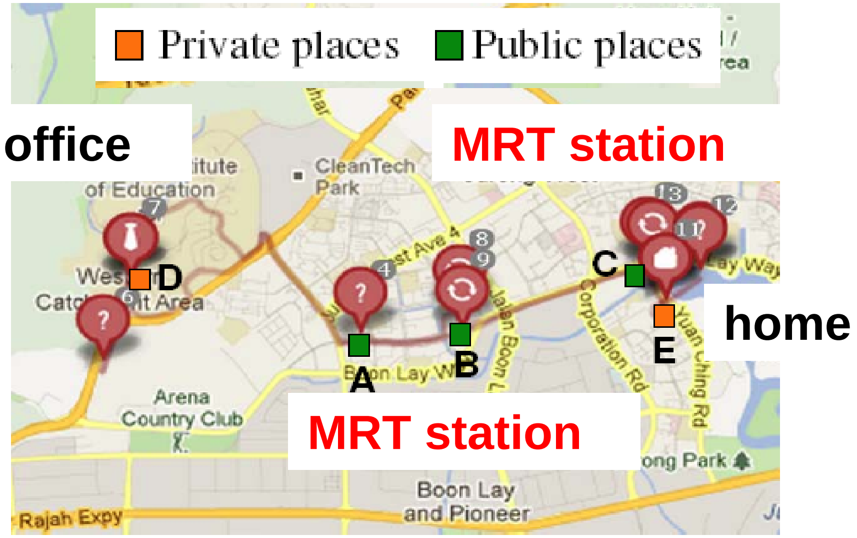
Moving Intention

- Entering a new place/Leaving from the current place is determined based on the changes of pattern in ambience signatures
  - The similarity between any two consecutive ambiance signatures is not grater than the predefined thresholds.



Place-aware GPS use

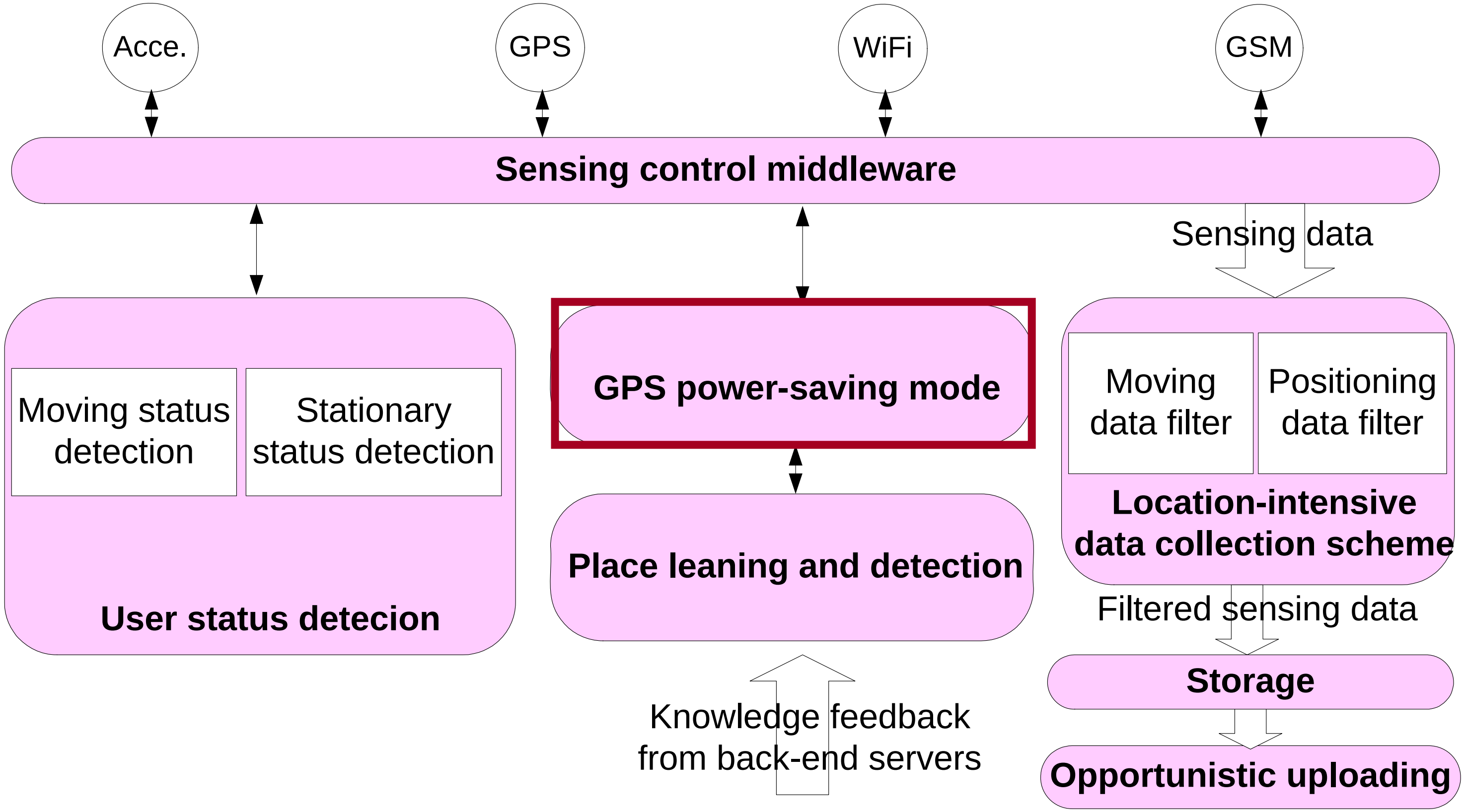
- Energy-saving principles for GPS use:
  - Delay to turn on GPS in a public place
    - the phone will predict how long the user takes to move out of the activity range of a public based on the current speed
  - Do not turn on GPS in a private place until detect the user leaving



A trace with 3 public places and 2 private places

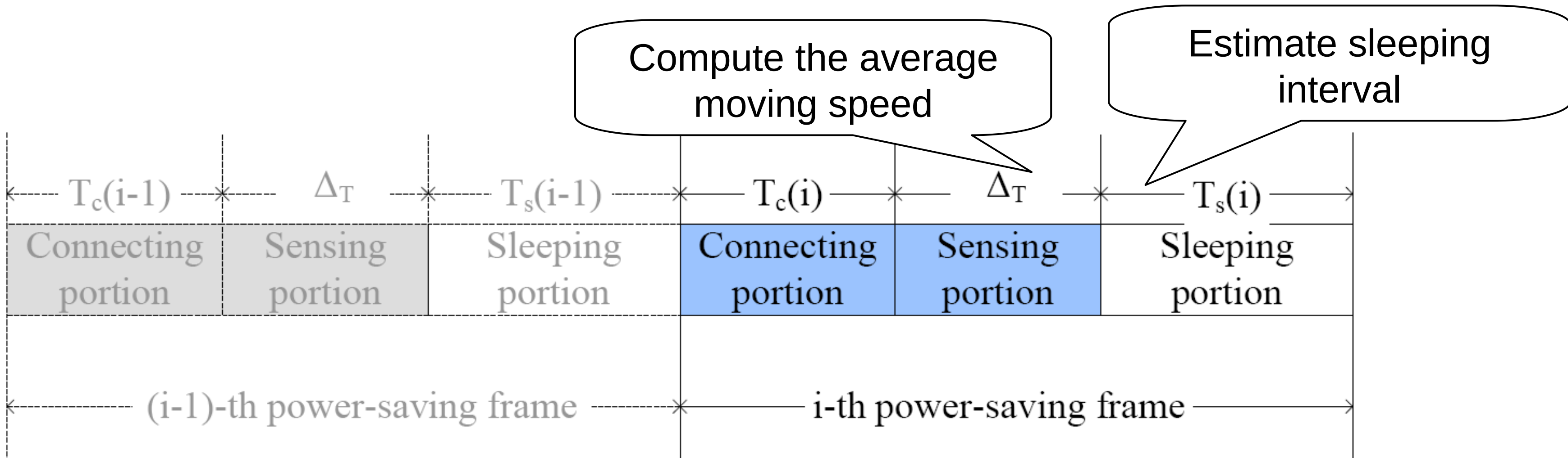
Components of Mobile Clients

- GPS power-saving mode
  - The GPS operates on the power-saving mode, once GPS is turned on



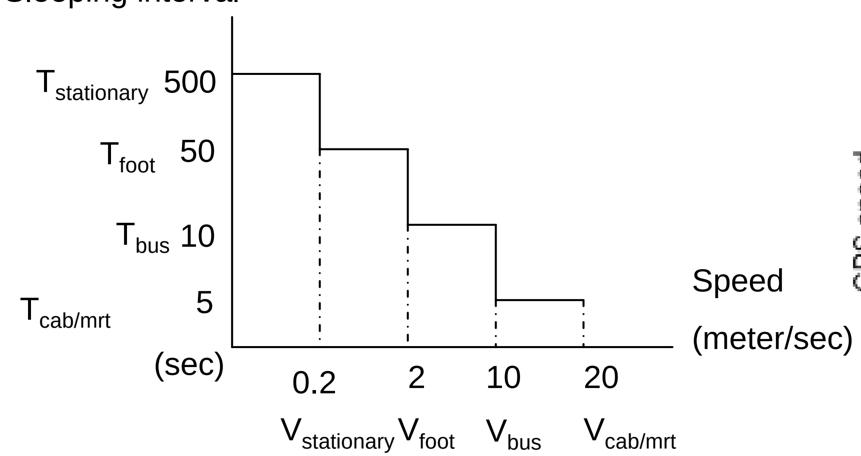
GPS Power-Saving Mode

- GPS power-saving frame:
  - Connecting portion
    - Time To First Fix (TTFF)
  - Sensing portion:
    - A fixed duration ΔT
    - Collect GPS fixes with speed information.
  - Sleeping portion
    - The sleeping interval is estimated based on the average moving speed during the sensing portion

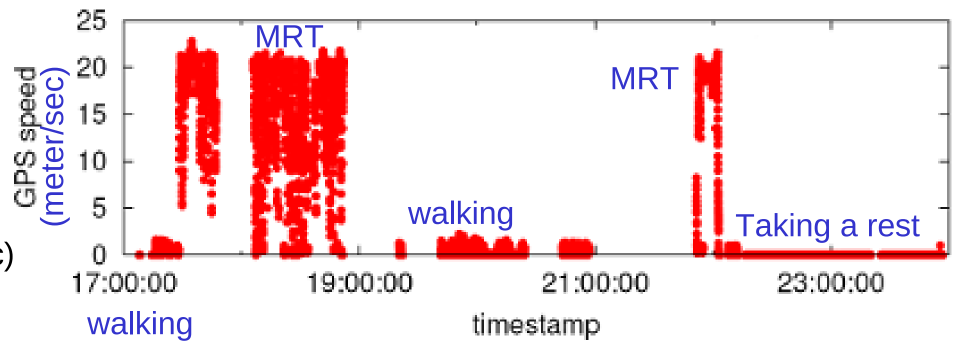


Estimate Sleeping Interval

- Consider MRT's speed to be the baseline because it is more stable and faster than other transportation modes
- The sleeping interval is determined by a discrete mapping function



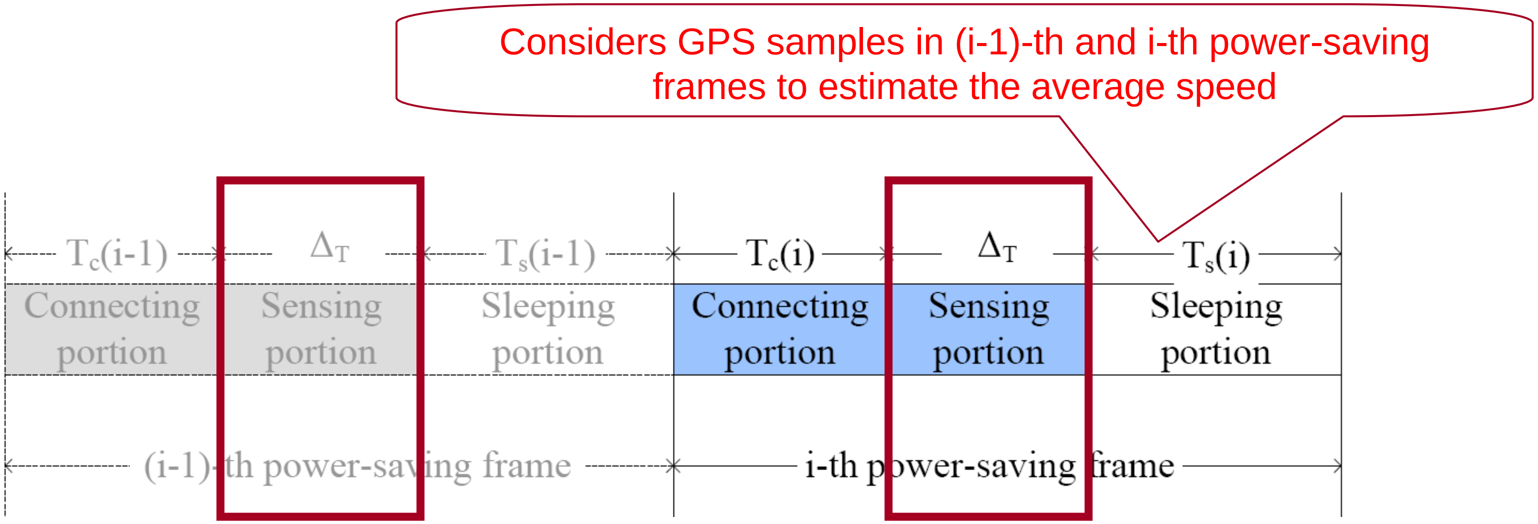
(a) A discrete mapping function from speeds to sleeping intervals



(b) Moving speed distribution for a user's trace.

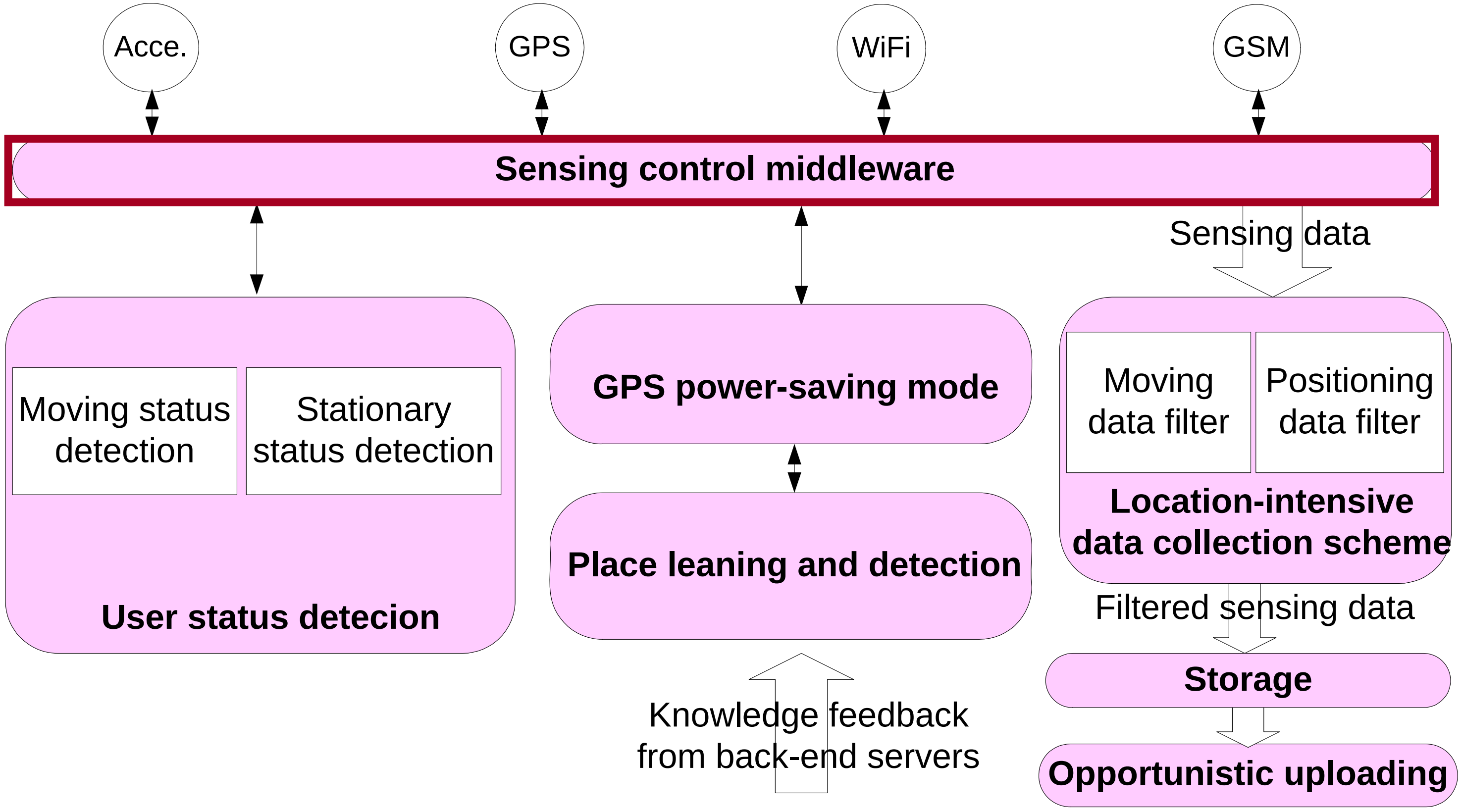
Speed Underestimated Problem

- Problem: the sleeping portion may be too long because of a zero speed when the sensing portion is at a special moment
  - waiting for traffic light
  - waiting for passengers alighting
- Current approach:
  - Consider multiple power-saving fames to estimate the average speed e.g., consider 2 power-saving frames.



Components of Mobile Clients

- Sensing control middleware



Sensing Control Middleware

- Control sampling rates of sensors
  - WiFi, GSM, and accelerometer sensors sample data in fixed rates
  - Coordinate components of status detection, place learning and detection, and GPS power-saving mode to control the GPS sensor

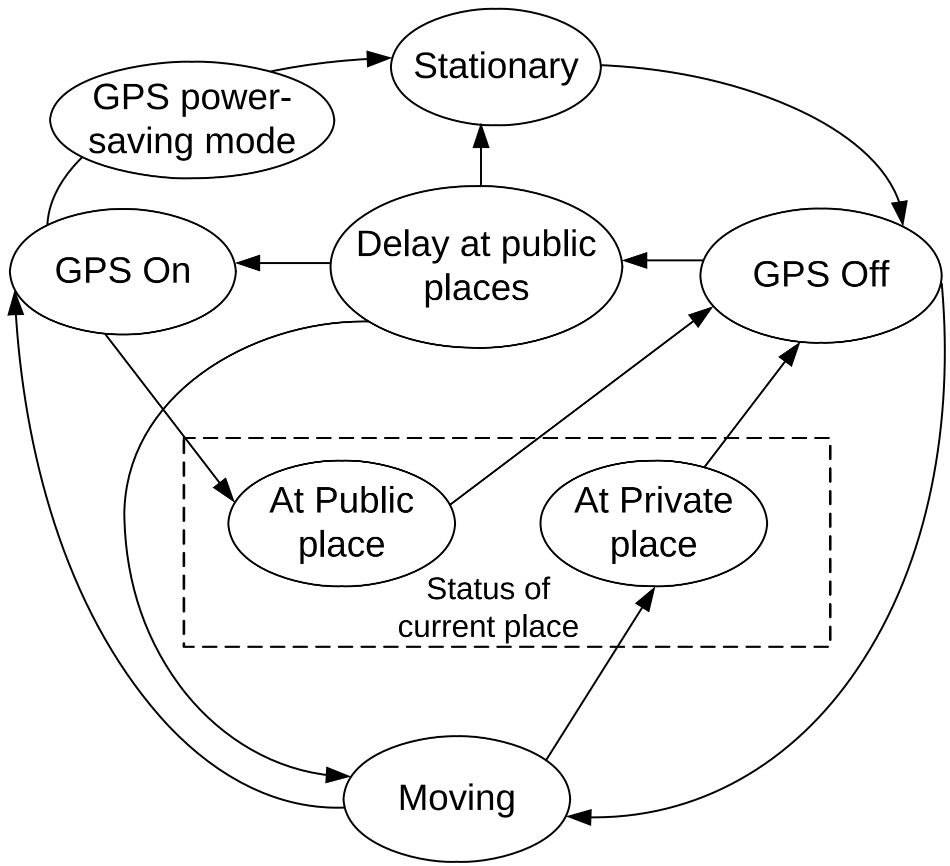
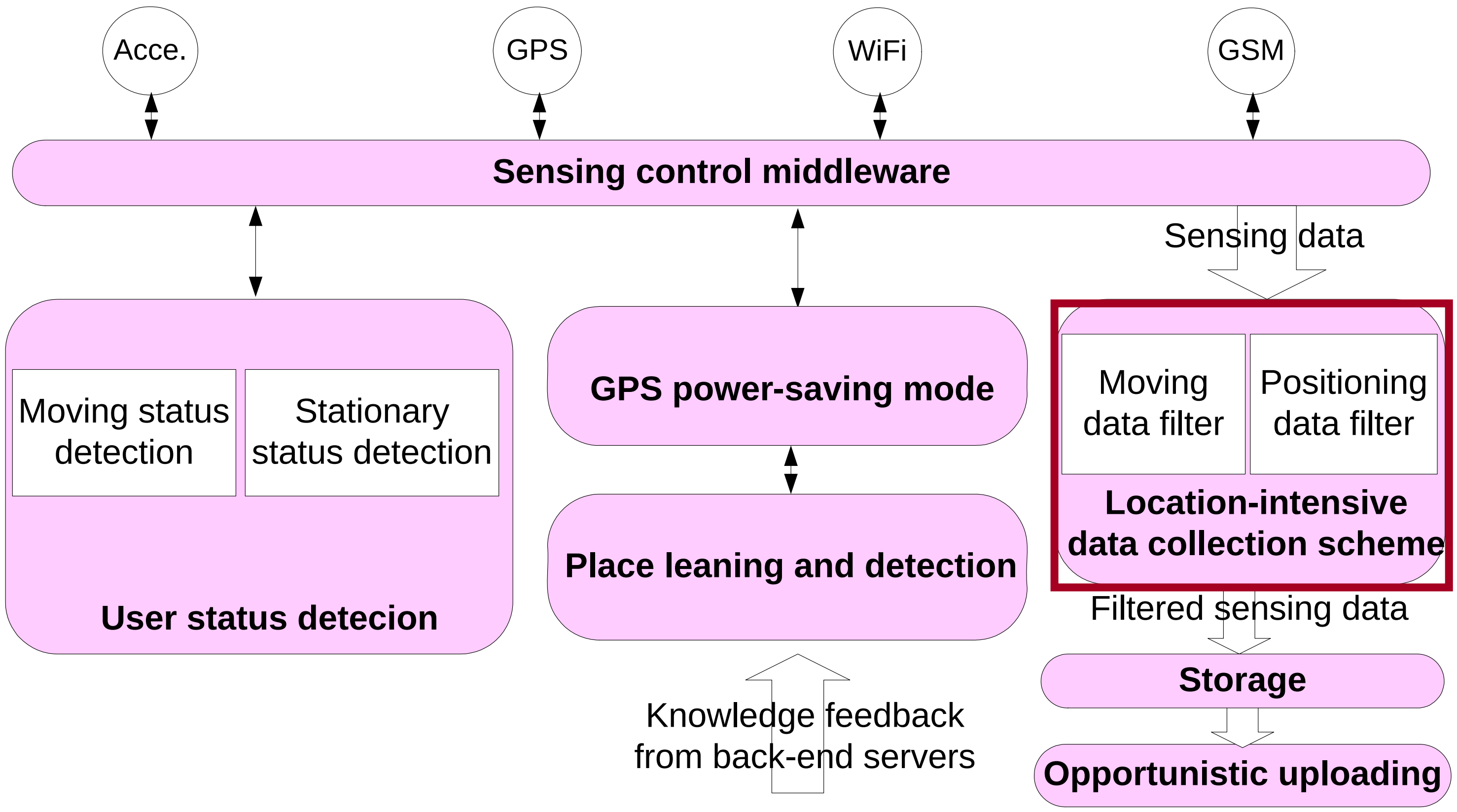


Fig: State transition of the GPS sensor.

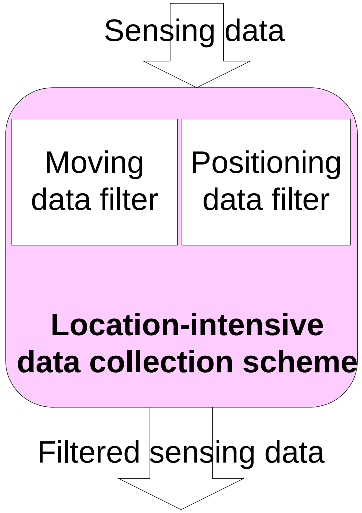
Components of Mobile Clients

- Location-intensive data collection scheme



Location-Intensive Data Collection

- Goal: Reduce the total amount of uploaded data for energy-conservation purposes.
- Approaches:
  - Upload "high-quality" and "low-quantity" data
    - The smaller size of data contains much more accurate location information and significant changes in movements
  - Data storage strategies:
    - Positioning data filter
    - Movement data filter



Positioning Data Filter

- Filter out the lower-accurate positioning data whilst multiple types of positioning information coexist.
  - Example:
    - Pick up GPS during [t1, t2]
    - Pick up WiFi data during [t1, t3]

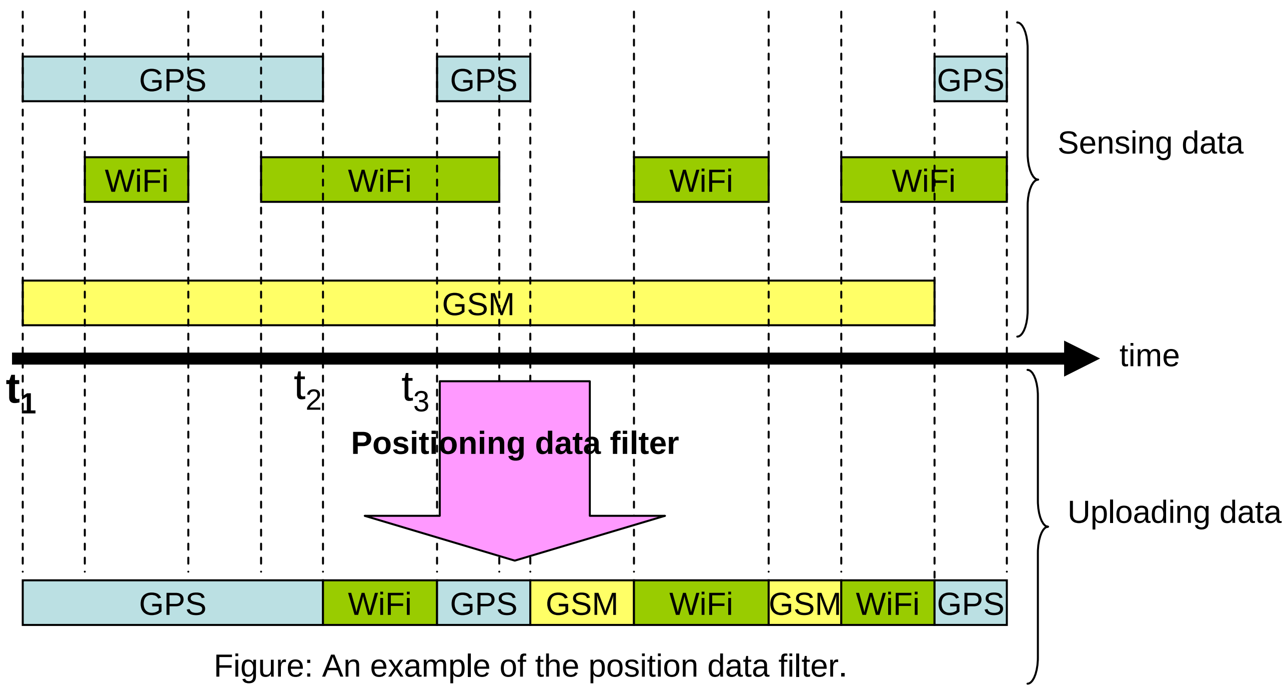
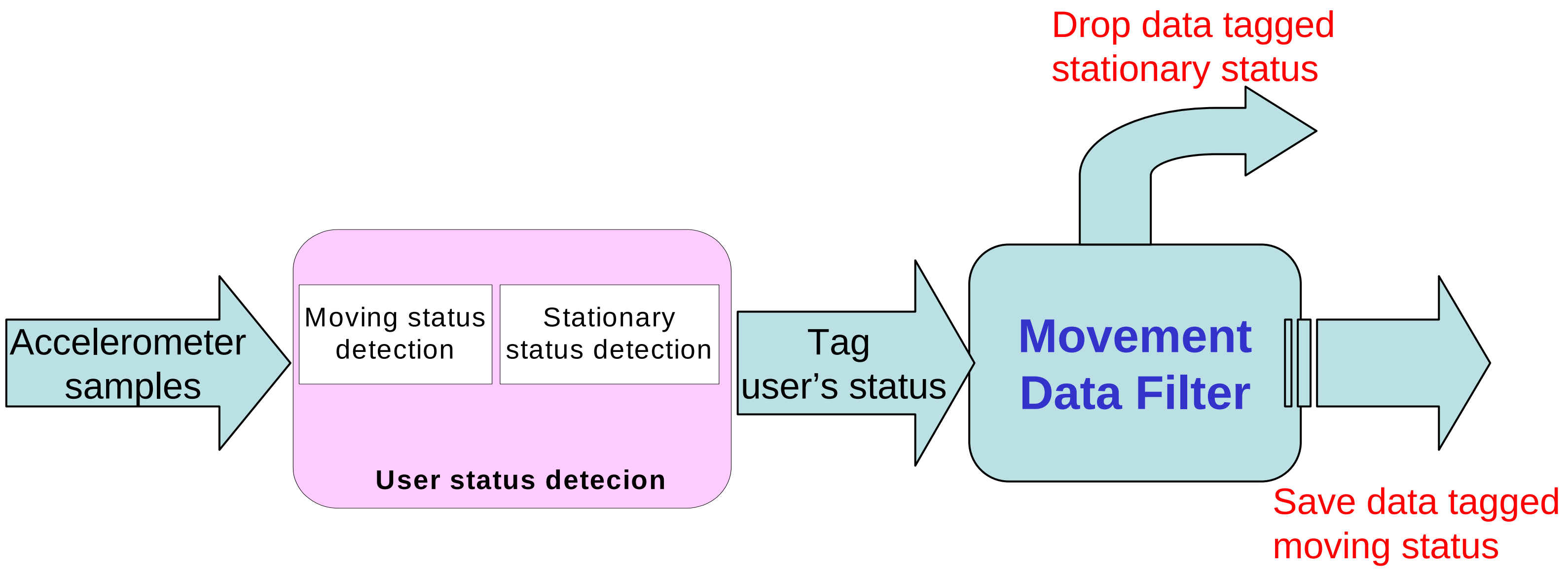


Figure: An example of the position data filter.

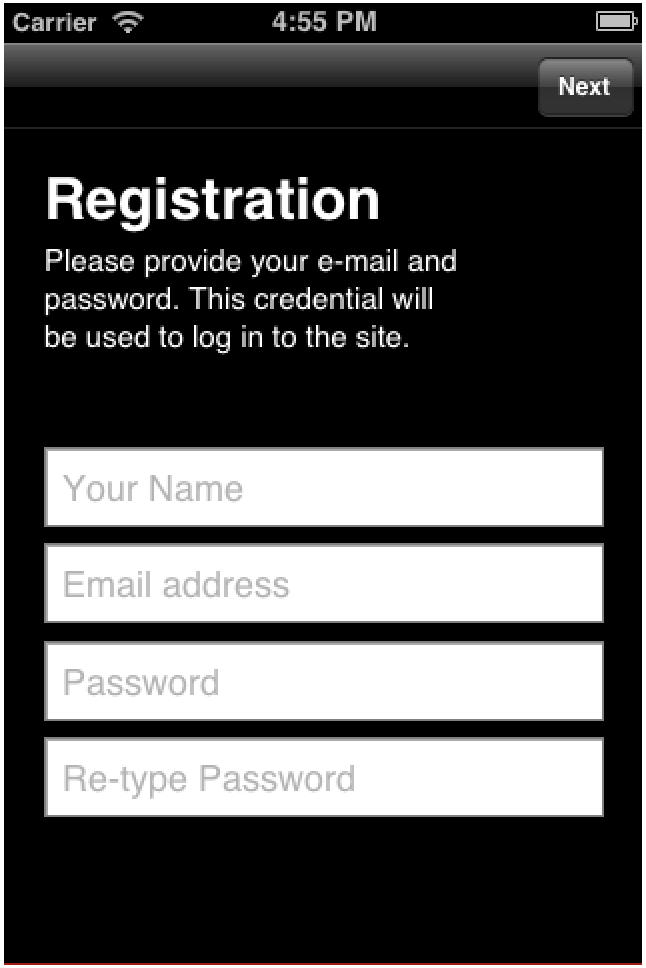
Movement Data Filter

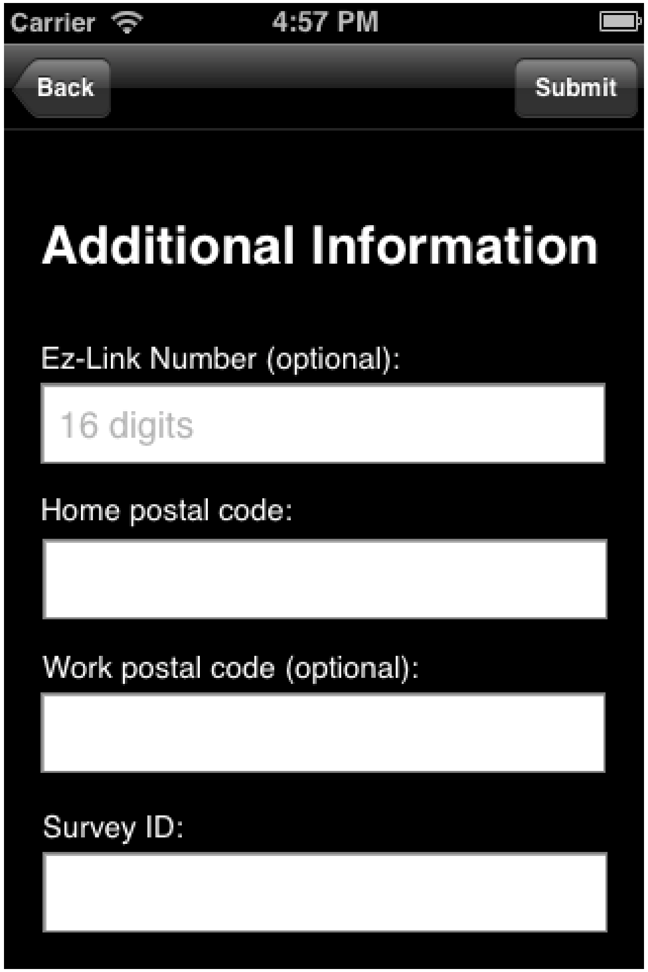
- Cooperate with the user status detection component
- Pick up those accelerometer samples which indicates the user moving



Mobile Application UI Design 1

- User interface for account registration

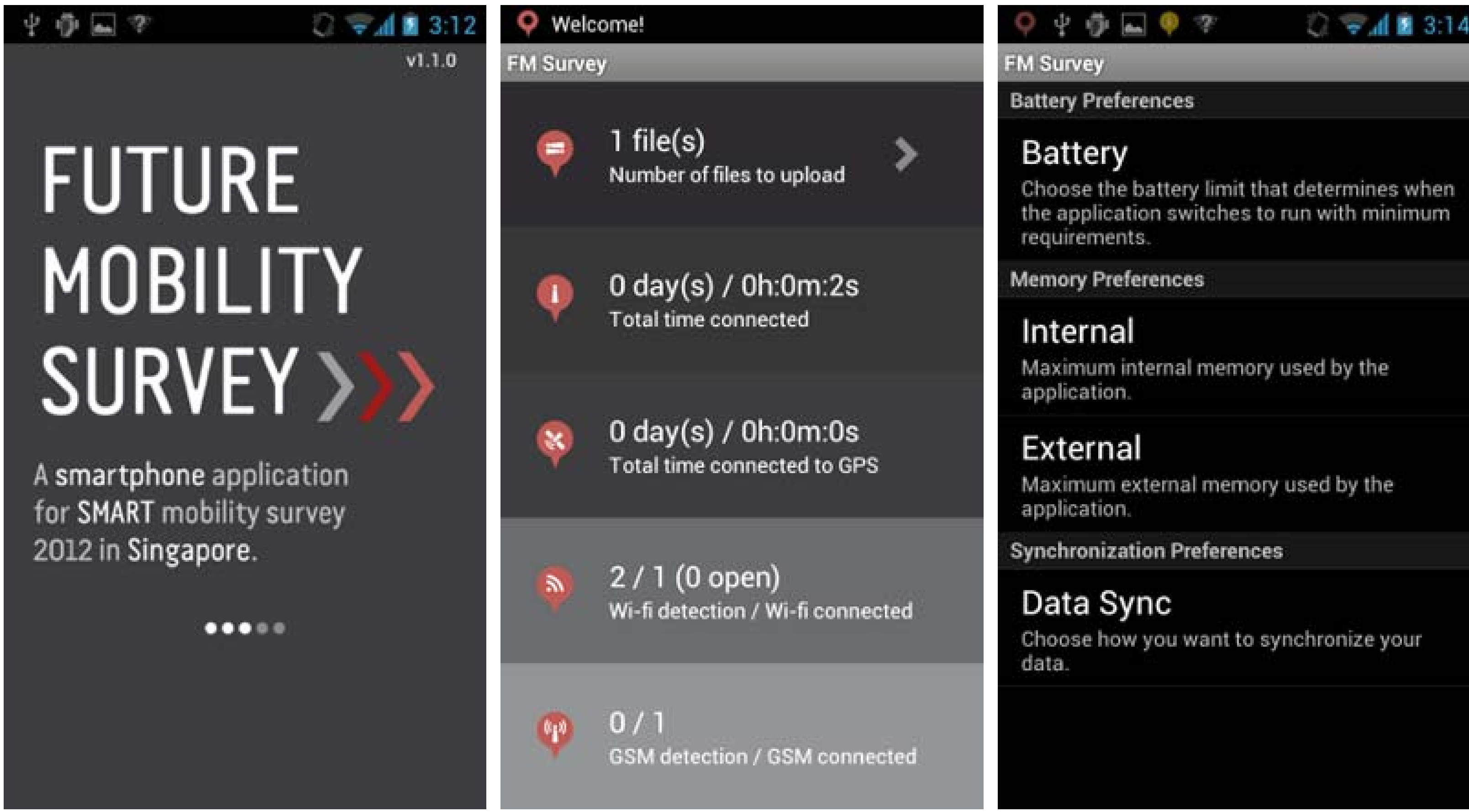




A simple registration process is available on our mobile application.

Mobile Application UI Design 2

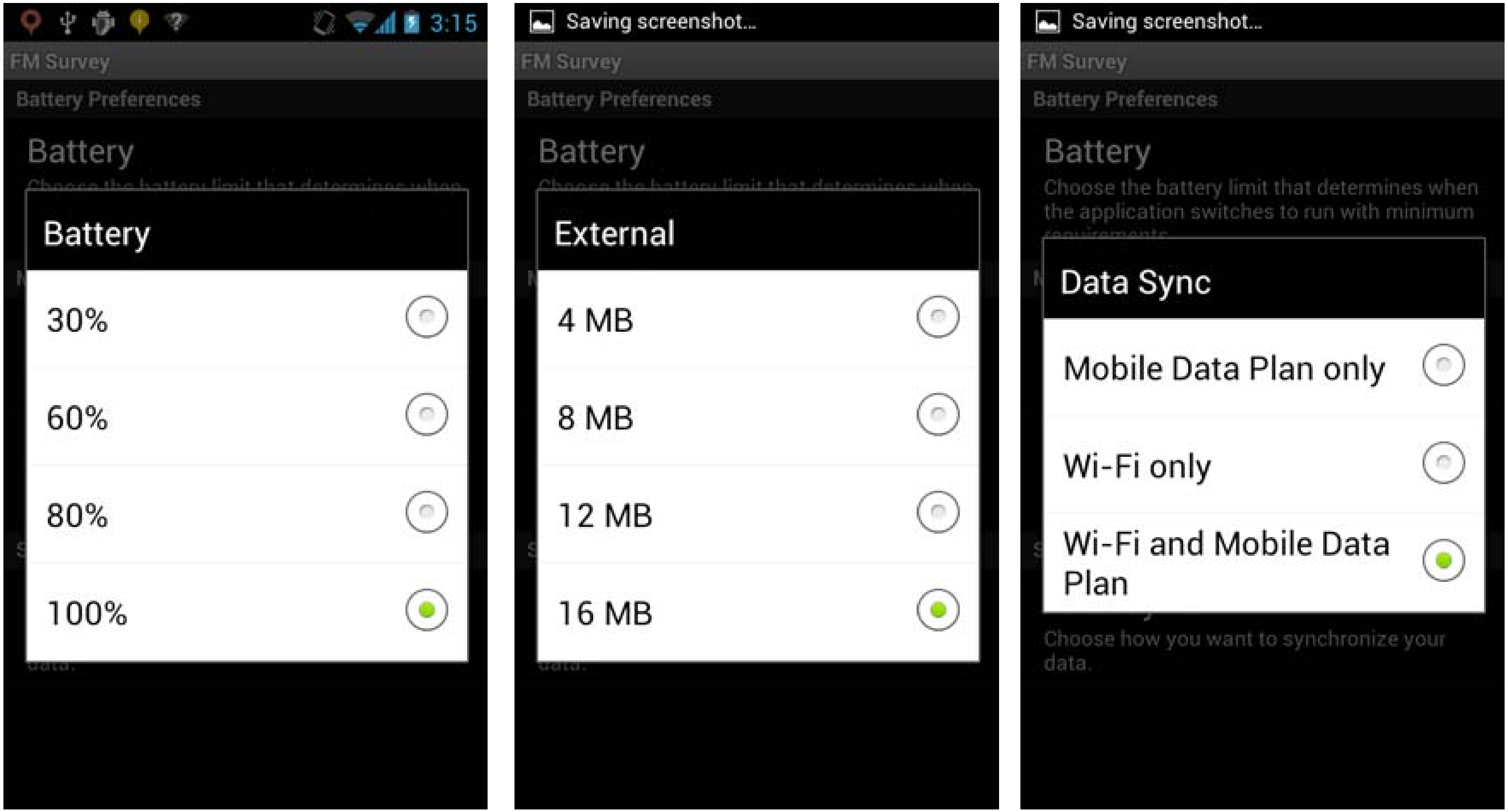
- Splash screen (Left) and main interface (Center and right)



(a) Splash screen (b) detailed information of data collected, (c) user preference.

Mobile Application UI Design 3

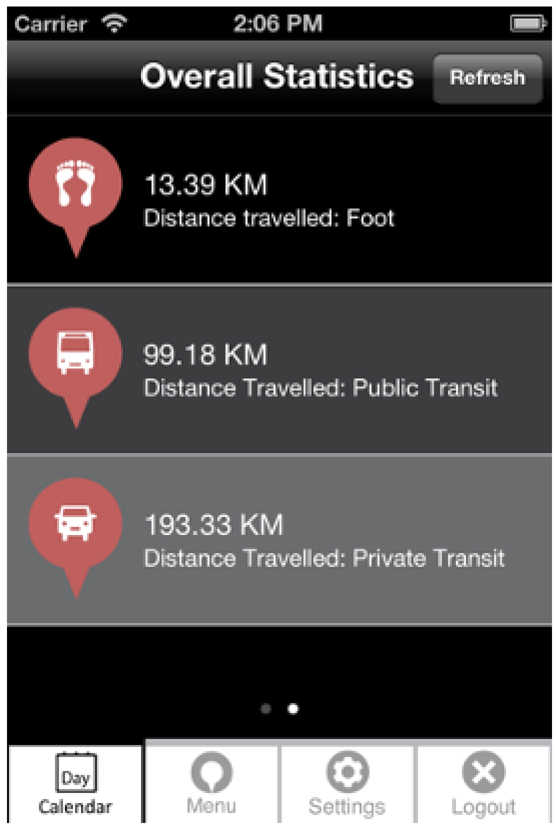
- Mobile UI Components and User Preference Manager

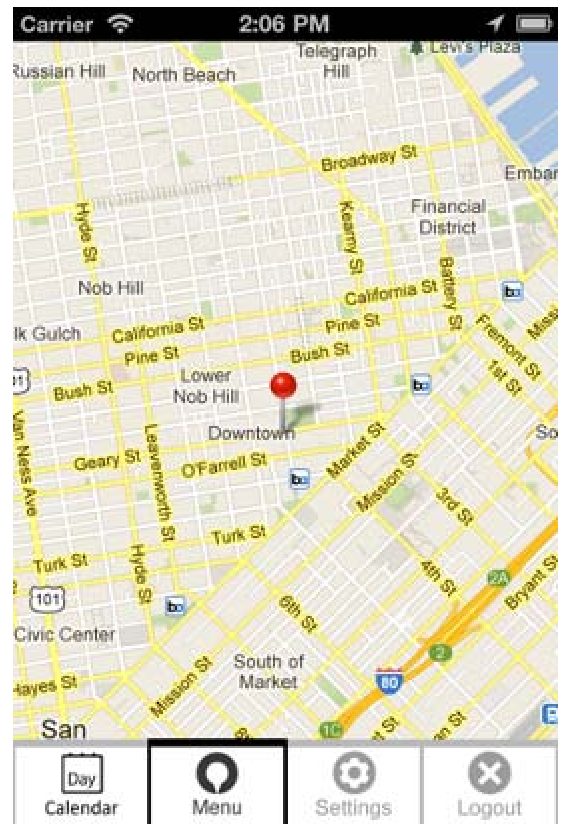


User Preference to (a) battery, (b) memory, (c) uploading.

Mobile Application UI Design 4

- User’s validated date on calendar (Left)
- Overall statistics display (Center)
- User’s current location (Right)



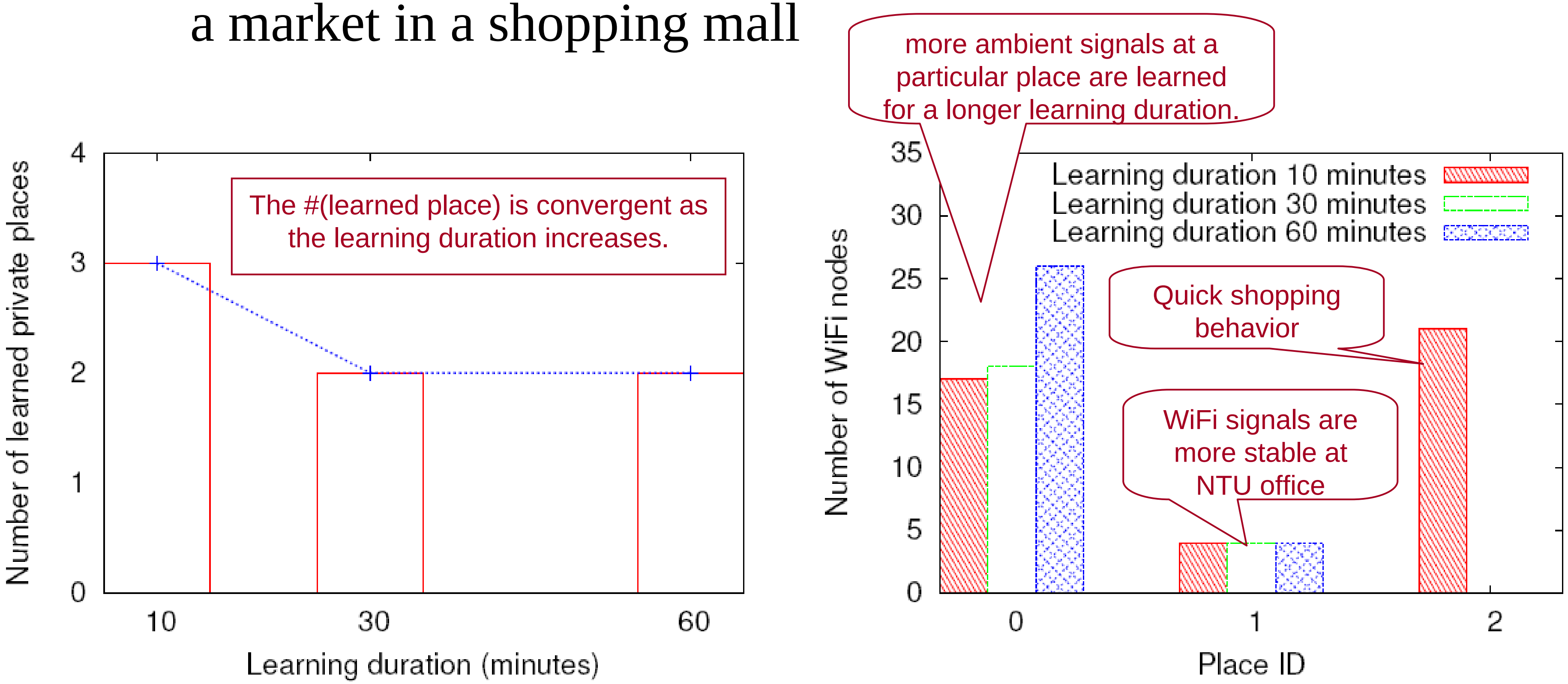


Experiments and Results: Mobile App.

- We compare each component of our mobile app. against a naïve data collection scheme
  - For the naïve data collection scheme, only user status detection is considered to control GPS
    - Turn on GPS if moving status
    - Turn off GPS if stationary status
  - 3 test scenarios
    - Place-aware scheme v.s. naïve scheme
    - GPS power-saving mode v.s. naïve scheme
    - Location-intensive data filters v.s. naïve scheme

Exp. Results: Place-aware Scheme

- Change the learning duration of place learning
  - Place ID 0: home, Place ID 1: NTU office, and Place ID 2: a market in a shopping mall

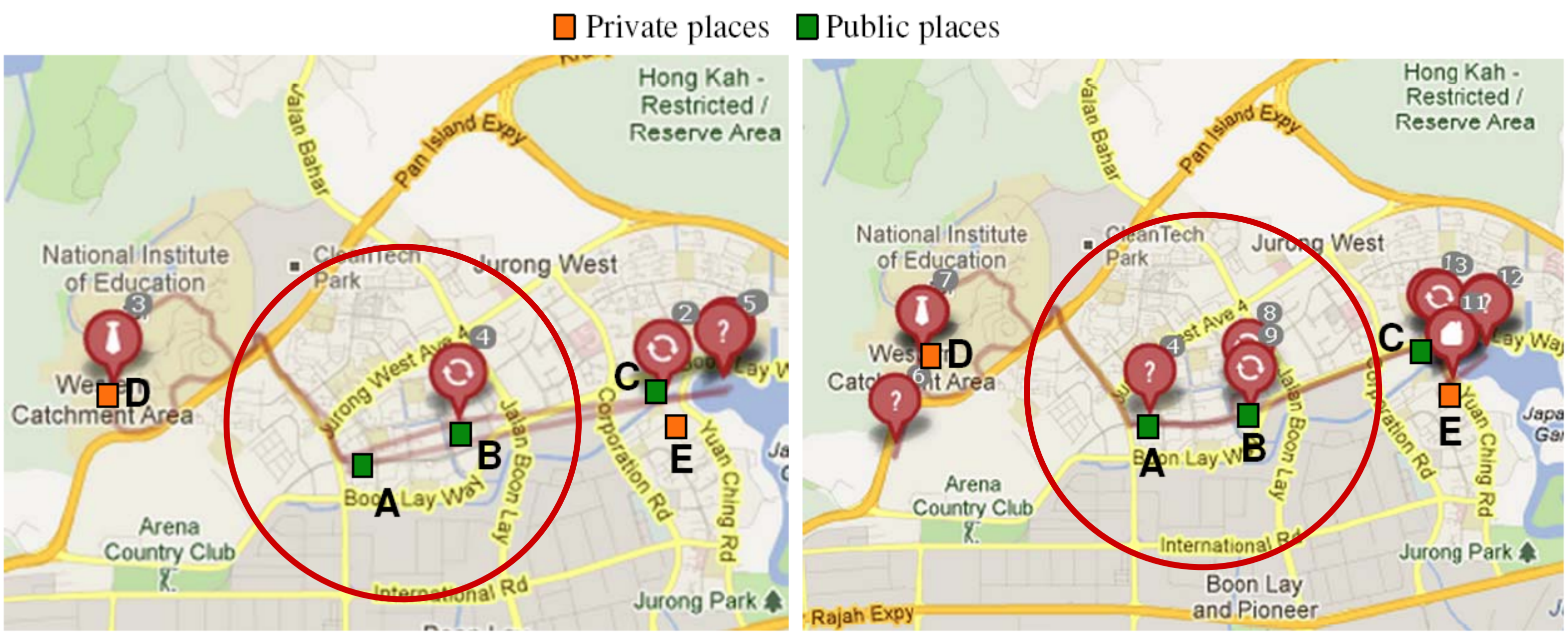


(a) Number of private places learned by different learning duration, (b) Number of WiFi nodes at private places.

Exp. Results: Place-aware Scheme

- A comparison of traces collected.

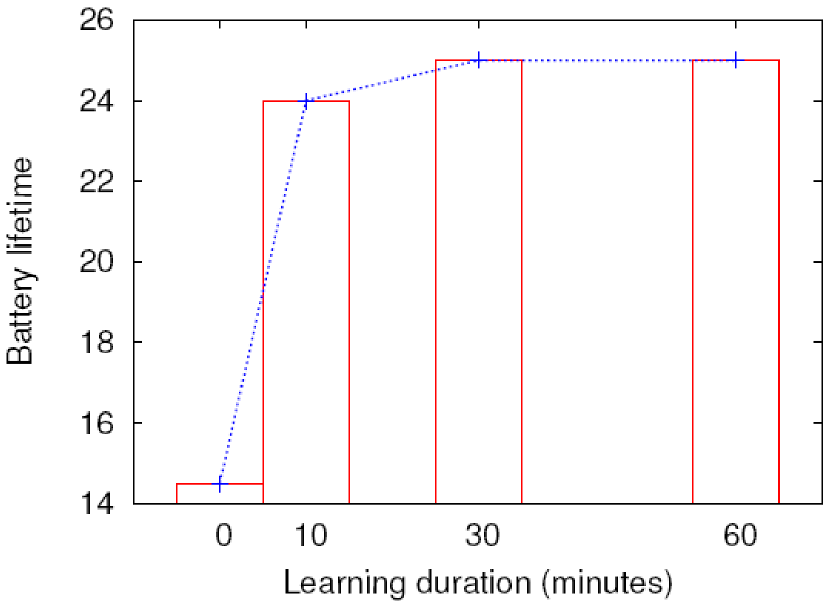
Rough traces at public places are collected because of delay time of GPS



(a) with place learning and detection (b) without place learning and detection

Exp. Results: Place-aware Scheme

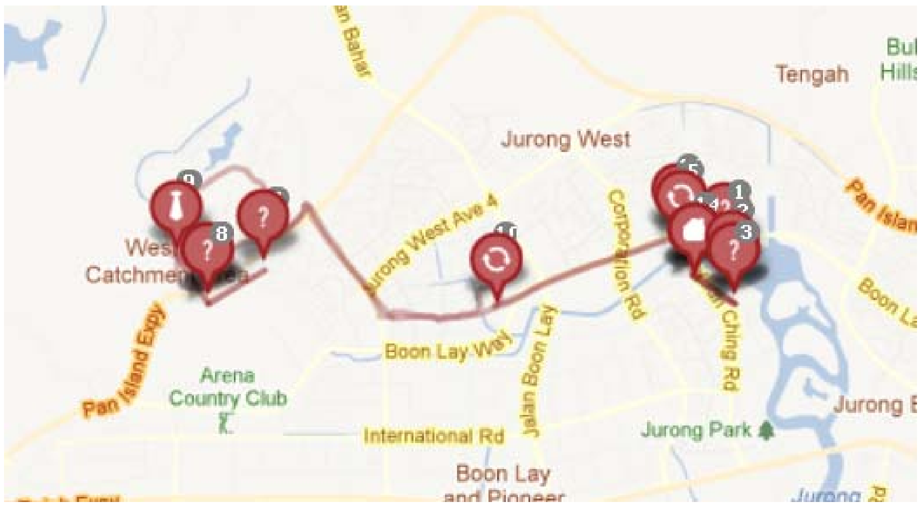
- A bound of battery lifetime which is resulted from the limited number of long-stay places for a user.



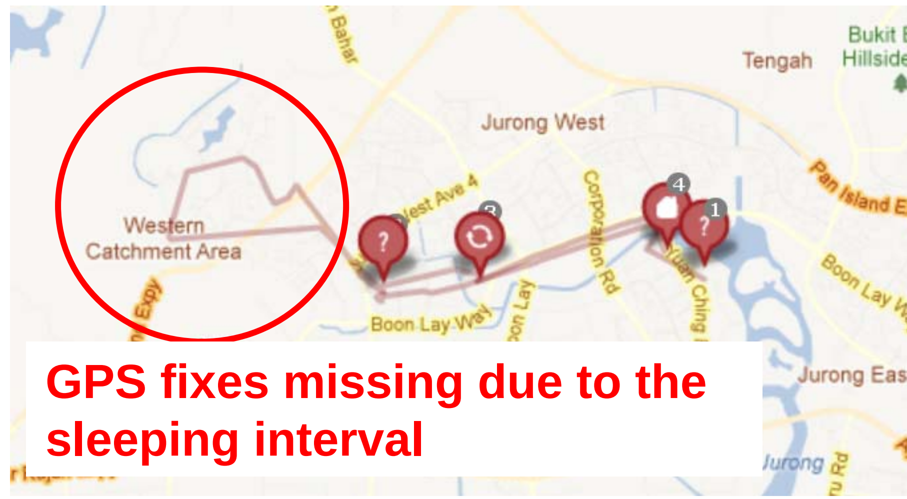
Battery lifetime for different learning durations.

Exp. Results: GPS Power-saving Mode

- A comparison of traces collected with and without GPS power-saving mode



(a) Trace collected by the naïve scheme



(b) Trace collected by GPS power-saving mode

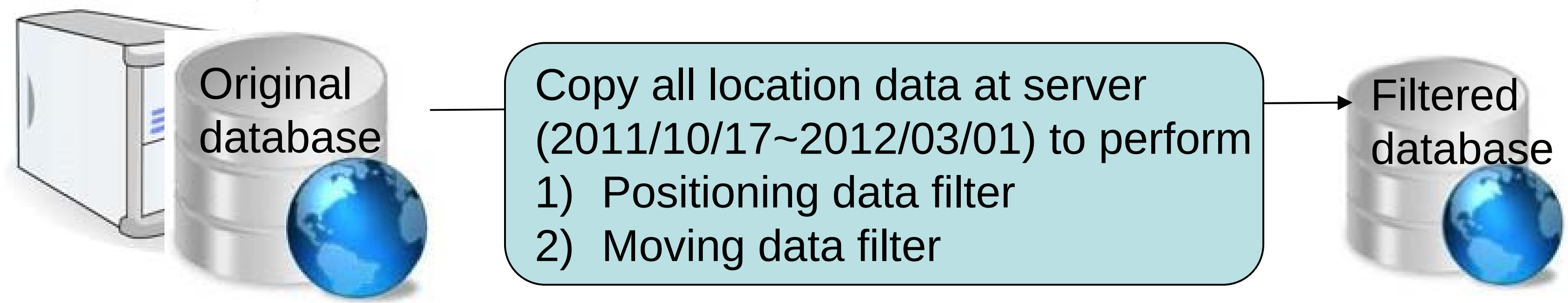
Exp. Results: Overall Battery Lifecycle

- GPS power-saving mode
  - The collected traces may miss something important, the battery lifecycle is good though
- Place-aware scheme
  - A good enough battery lifecycle and traces
- Tradeoff between trace accuracy and battery lifecycle
  - Place-aware scheme avoids charging during 1-day trips and has a good enough trace accuracy

Phone types	Battery lifetime (hr)			
	Without app	Naïve scheme	GPS power-saving mode	Place-aware scheme
Samsung Galaxy S II	≈48h	≈ 12	≈ 40	≈ 25

Exp. Results: Data Filters

- Improvement of Data Quantity



Data size	Before (G)	After (G)	Reduction of the amount of data %
GPS	1.48	1.48	0
WiFi	0.265	0.239	9.811321
GSM	0.153	0.0964	36.99346
ACCE	5.97	0.121	97.9732
Total data size	7.89	1.95	75.28517

Table: Reduction of data size

Exp. Results: Data Filters

➢ Improvement of Data Quantity

Copy all location data at server (2011/10/17~2012/03/01) to perform
1) Positioning data filter
2) Moving data filter

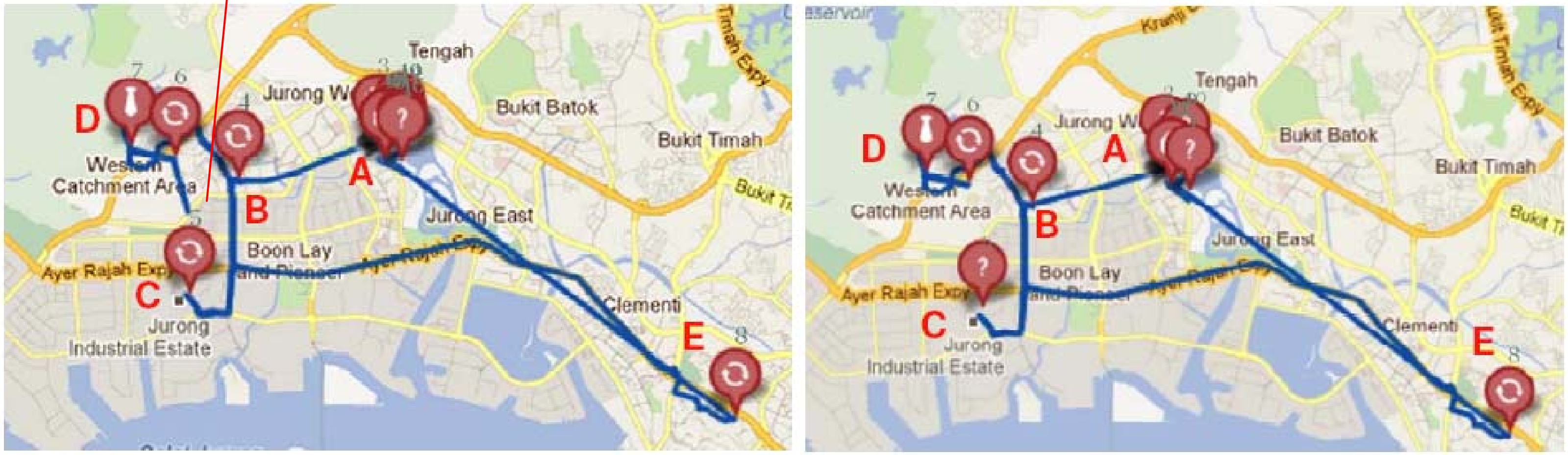
Number of samples	Before	After	Reduction of the amount of data %
#(GPS samples)	10496856	10496856	0
#(WiFi samples)	957058	811409	15.21841
#(GSM samples)	1533659	932922	39.17018
#(ACCE samples)	94427189	1585569	98.32086
Total number of samples	107414762	13826756	87.12769479

Table: Reduction of number of samples

Exp. Results: Data Filters

- Case Studies for Data Quality

Jumping traces are resulted from the low-accurate GSM positioning technologies.



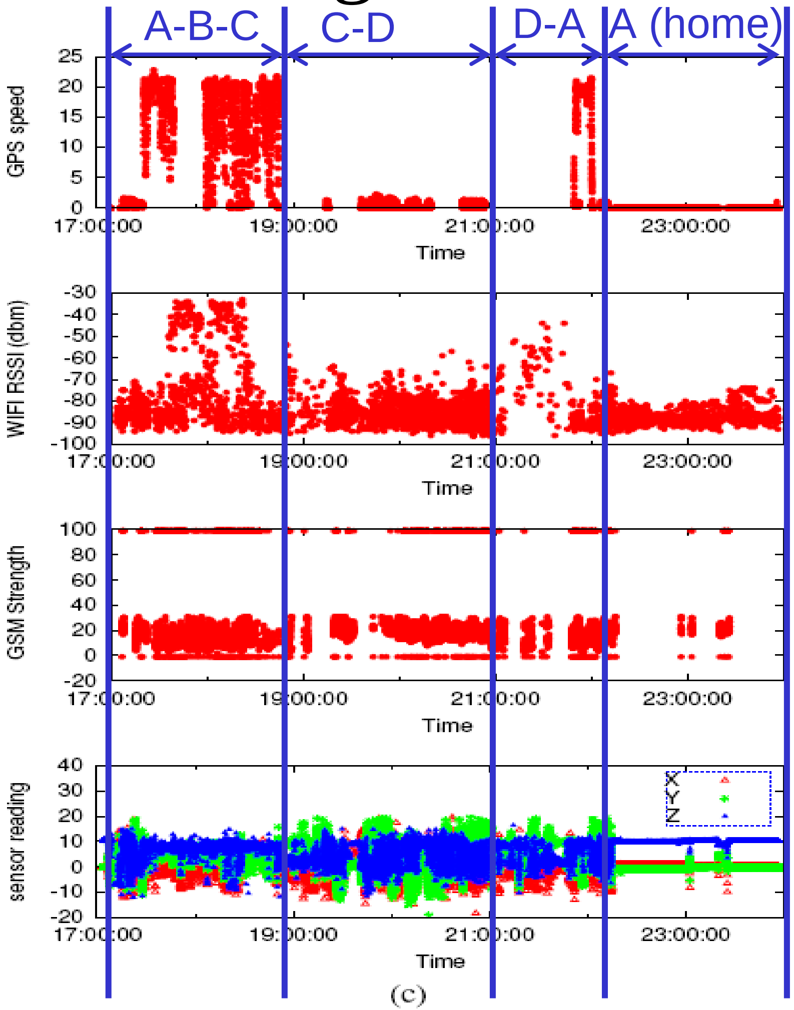
(a) Traces collected by the naive data collection scheme. (b) Traces collected by our data collection scheme.

Exp. Results: Data Filters

- A case study of a long trace
  - A-B-C: taking MRT (through 37 MRT stations)
  - C-D: walking
  - D-A: taking MRT (through 11MRT stations)

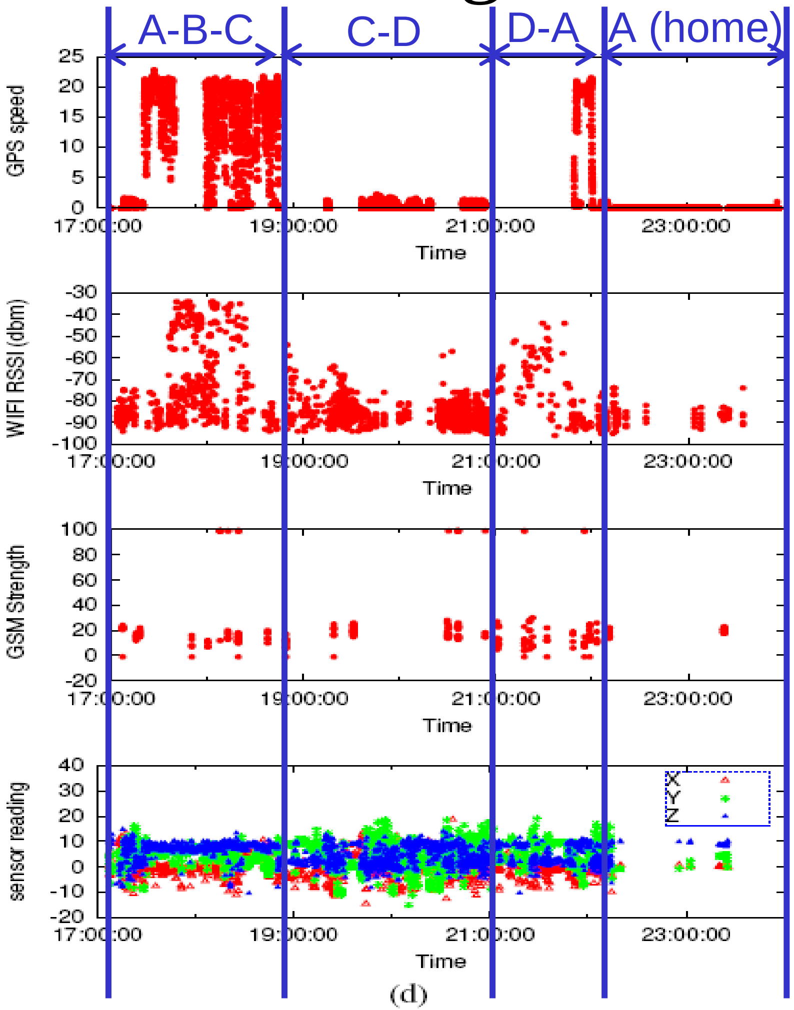
Exp. Results: Data Filters

- Sensing data distribution for the long trace



(c)

Without data filters

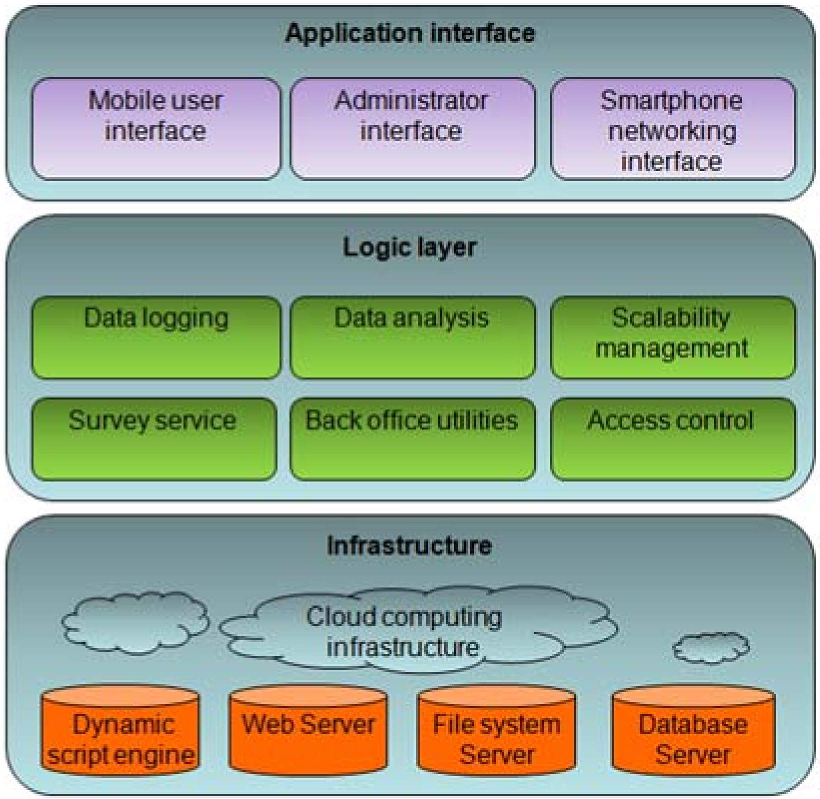


(d)

With data filters

Initial backend Implementation

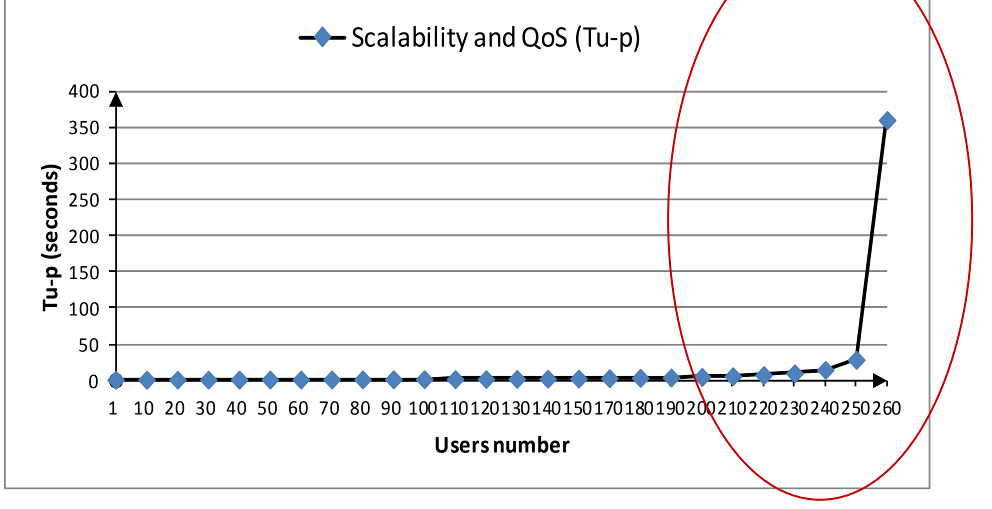
- A Standalone Backend Server
  - Hardware configure
    - CPU: Intel Xeon E5520/2.27GHz/quad cores/8 threads.
    - Memory: 8GB.
  - Software configure
    - Ubuntu Linux operating system.
    - MySql database.
    - Apache HTTP server.
    - Ruby+Rails web application framework.



QoS Analysis for a Standalone Server

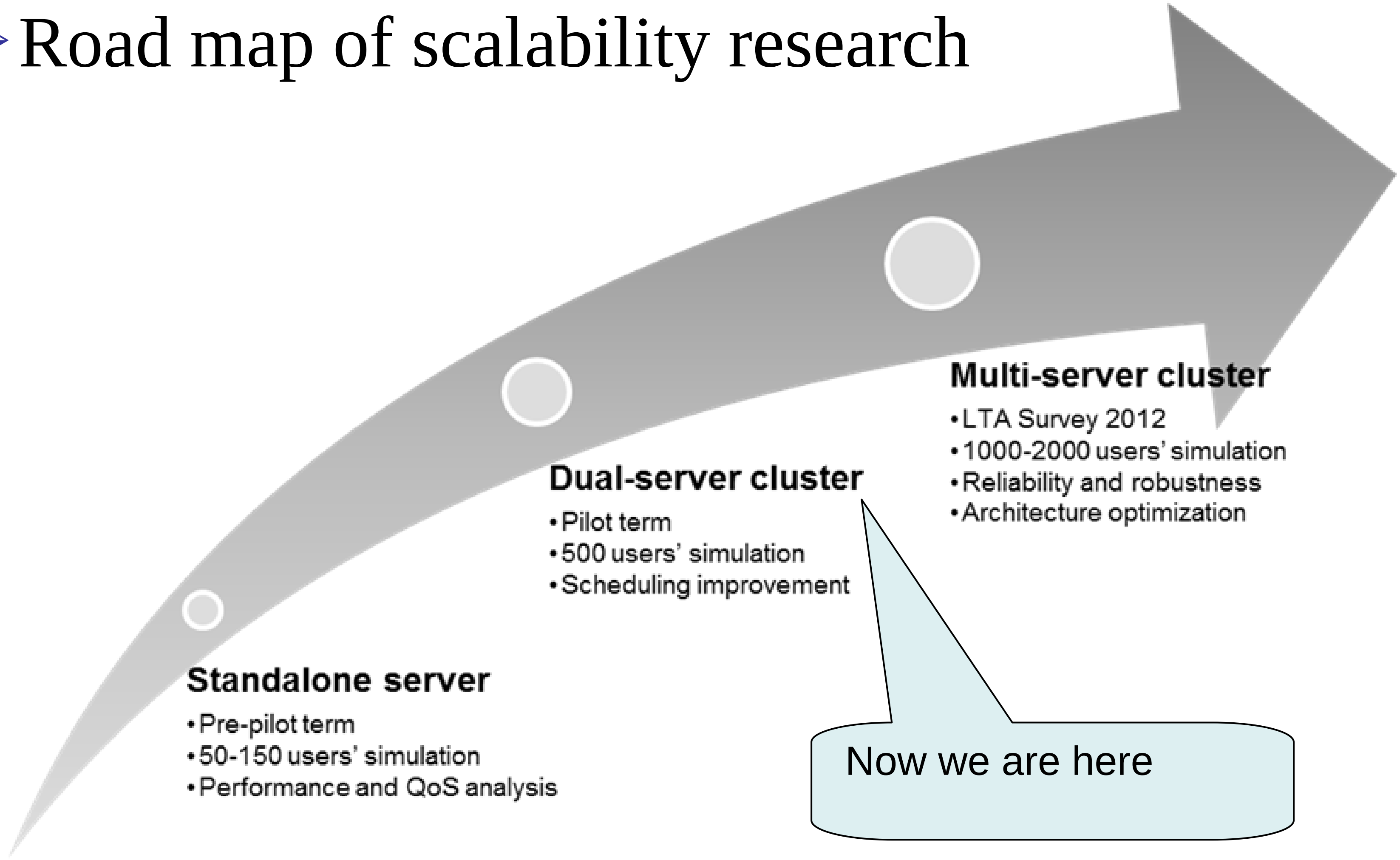
- QoS definition
  - Tu-p: Time from data uploading completion to processing completion.
- A robot's typical job arrival rate evolution model
  - $\lambda \approx 11.5$ (i.e., a mobile user uploads 11.5 JSON file every hour)
- If the QoS requirement is *Tu-p* <= 30 seconds, the number of mobile users that the back-end servers can support is less than 250.

Number of users: n	*QoS: Tu-p*
n -> 261	Tu-p -> ∞
$251 \leqslant n \leqslant 260$	32 seconds $\leqslant$ Tu-p $\leqslant$ 6 minutes
$201 \leqslant n \leqslant 250$	5 seconds $\leqslant$ Tu-p $\leqslant$ 29 seconds
$101 \leqslant n \leqslant 200$	2 seconds $\leqslant$ Tu-p $\leqslant$ 5 seconds
$1 \leqslant n \leqslant 100$	1.2 seconds $\leqslant$ Tu-p $\leqslant$ 2 seconds



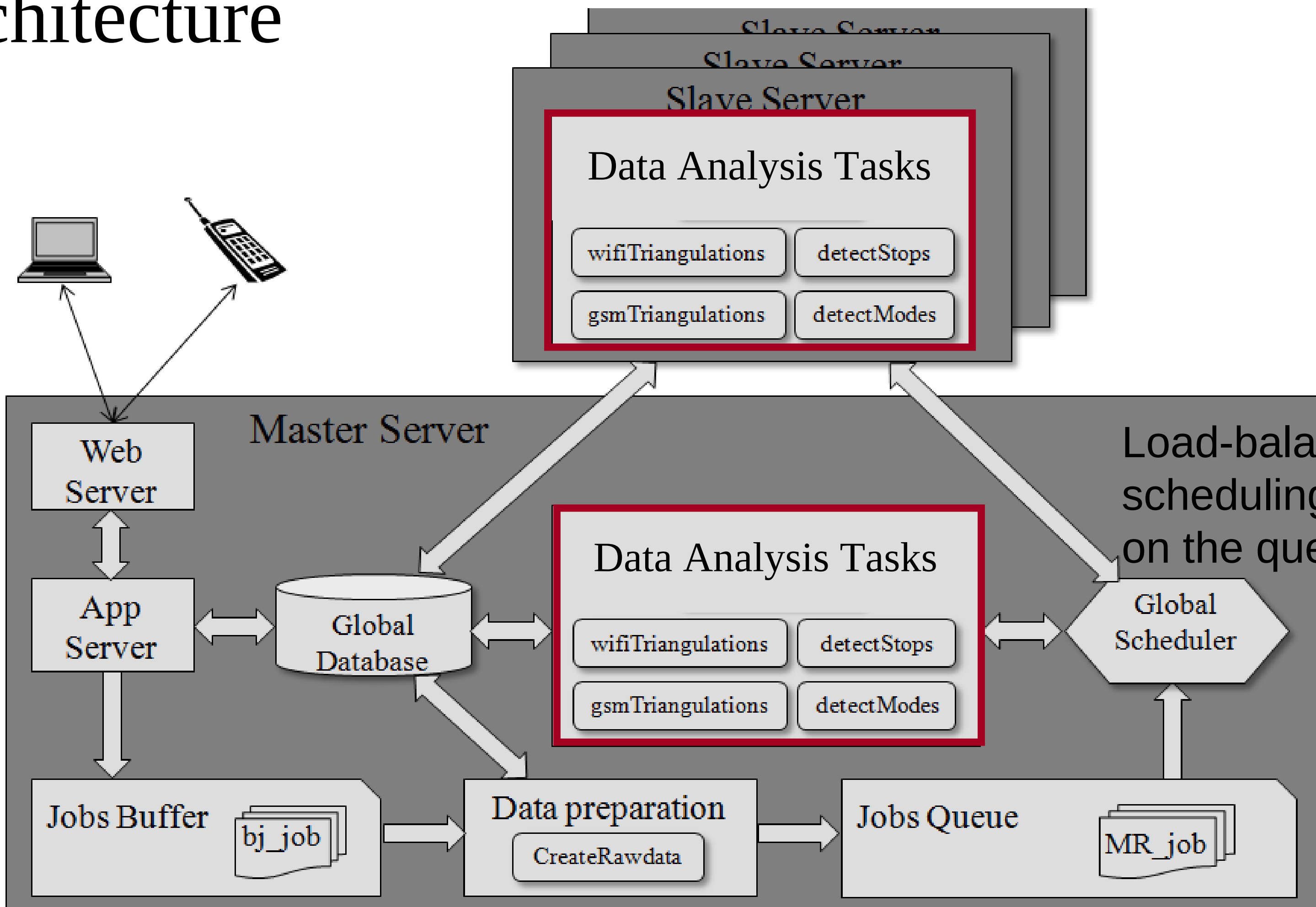
Resolution of our Backend Servers

- Road map of scalability research



Backend: Dual-Server Architecture

- MapReduce based dual-server cluster architecture



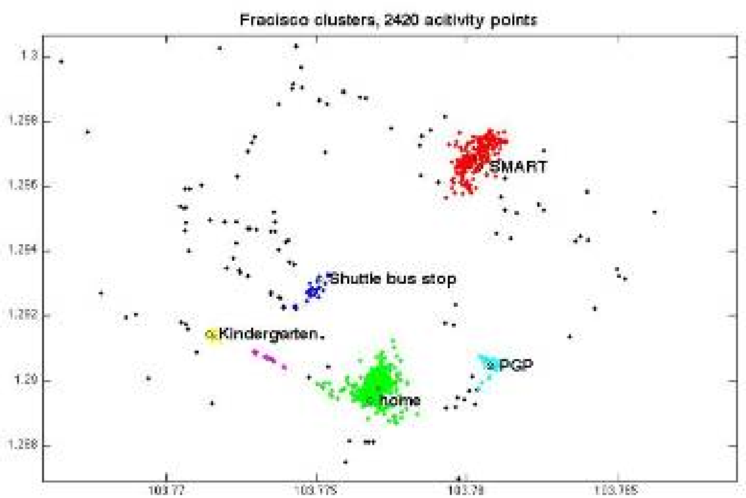
Load-balancing scheduling based on the queue size

Dual-server Cluster Implementation

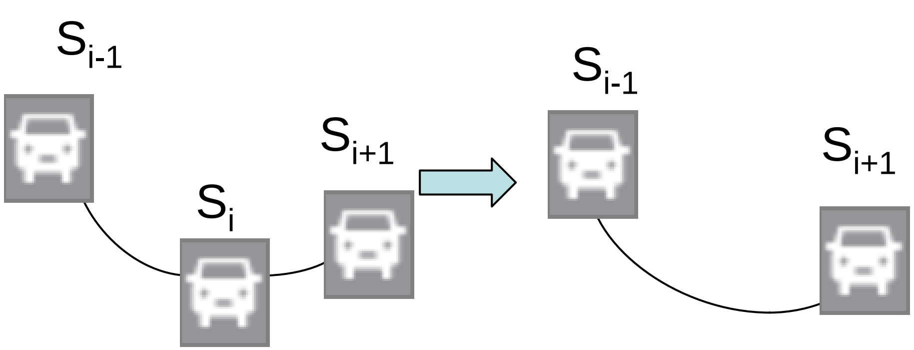
- Hardware config
  - Master server
    - CPU: Intel Xeon E5520/2.27GHz/quad cores/8 threads
    - Memory: 8GB
  - Slave server
    - CPU: Intel Xeon E5520/2.4GHz/quad cores/8 threads
    - Memory: 8GB
- Software config
  - Ubuntu Linux operating system
  - MySql database
  - Apache HTTP server
  - Ruby+Rails web application framework
  - MapReduce
    - Master server (JobTracker + Name node + TaskTracker + Data node)
    - Slave server (TaskTracker + Data node)

Data Analysis Tasks

- Trace generation:
  - Lower-accurate GSM/WiFi data points will fill the interval of no GPS fixes.
- Stop detection
  - Initial stop detection: clustering positions within a fixed time window into a single one stop if the maximum distance between any two of these positions is smaller than a threshold.
  - Stop merging: Any two consecutive stops will be merged if the sets of visible GSM ID at the two stops are the same
  - Stop calibration: A Stop S_i will be removed if the transpiration mode from S_{i-1} to S_i is the same as the transportation mode from S_i to S_{i+1}.



Stop merging: clustering positions based on spatial-and-temporal density



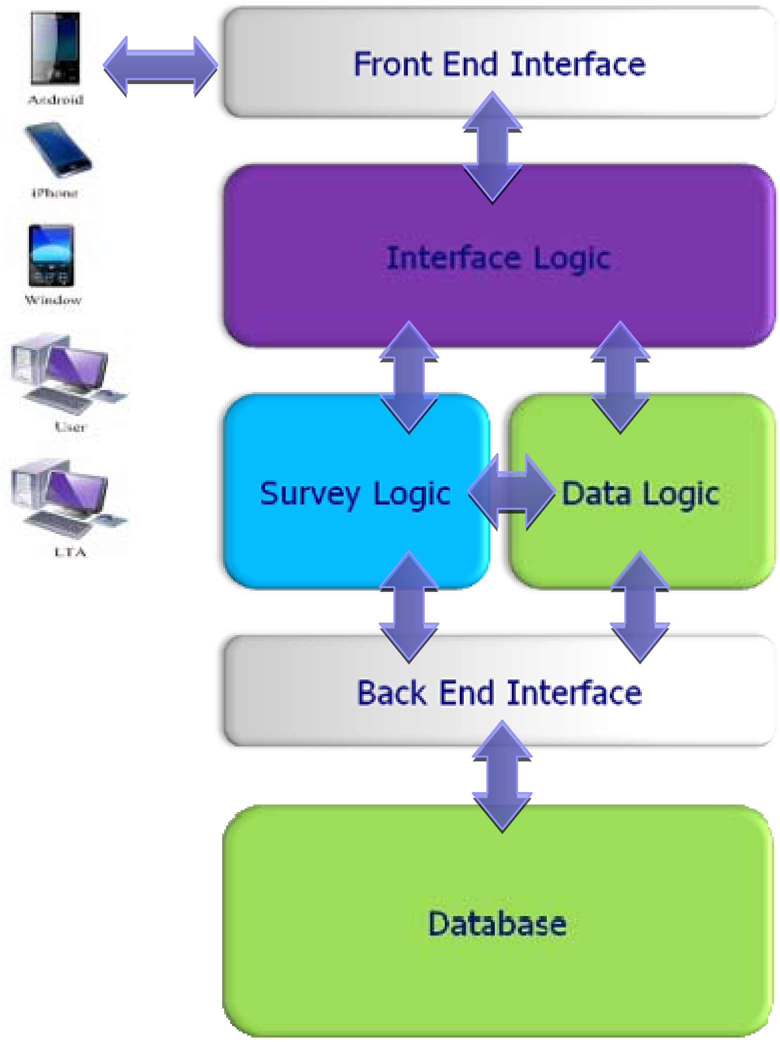
Stop Calibration based on transpiration modes

Data Analysis Tasks

- Transportation mode detection
  - Best Decision Tree (DT) model is based on GPS and accelerometer data with features
    - Max speed,
    - between stop avg speed,
    - Acce. force variance,
    - distance to closest bus and MRT stop.
  - Our internal members tag their transportation modes manually to collect the set of training data.

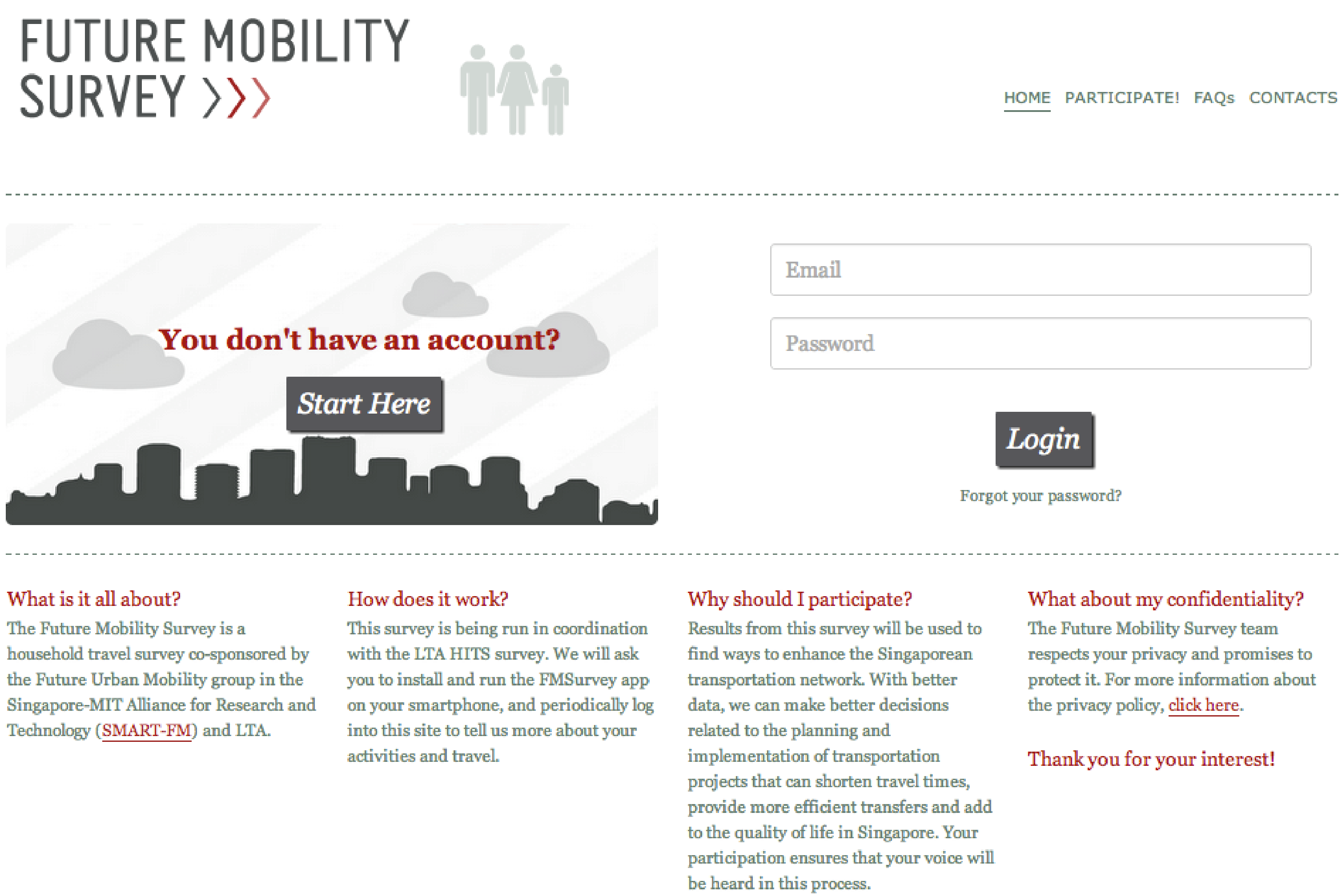
Web Application

- Mobile application uploads collected sensor data to web application.
- Users access web application to complete the survey questionnaire.
- Update/validate the travel information (trips, stops and activity).



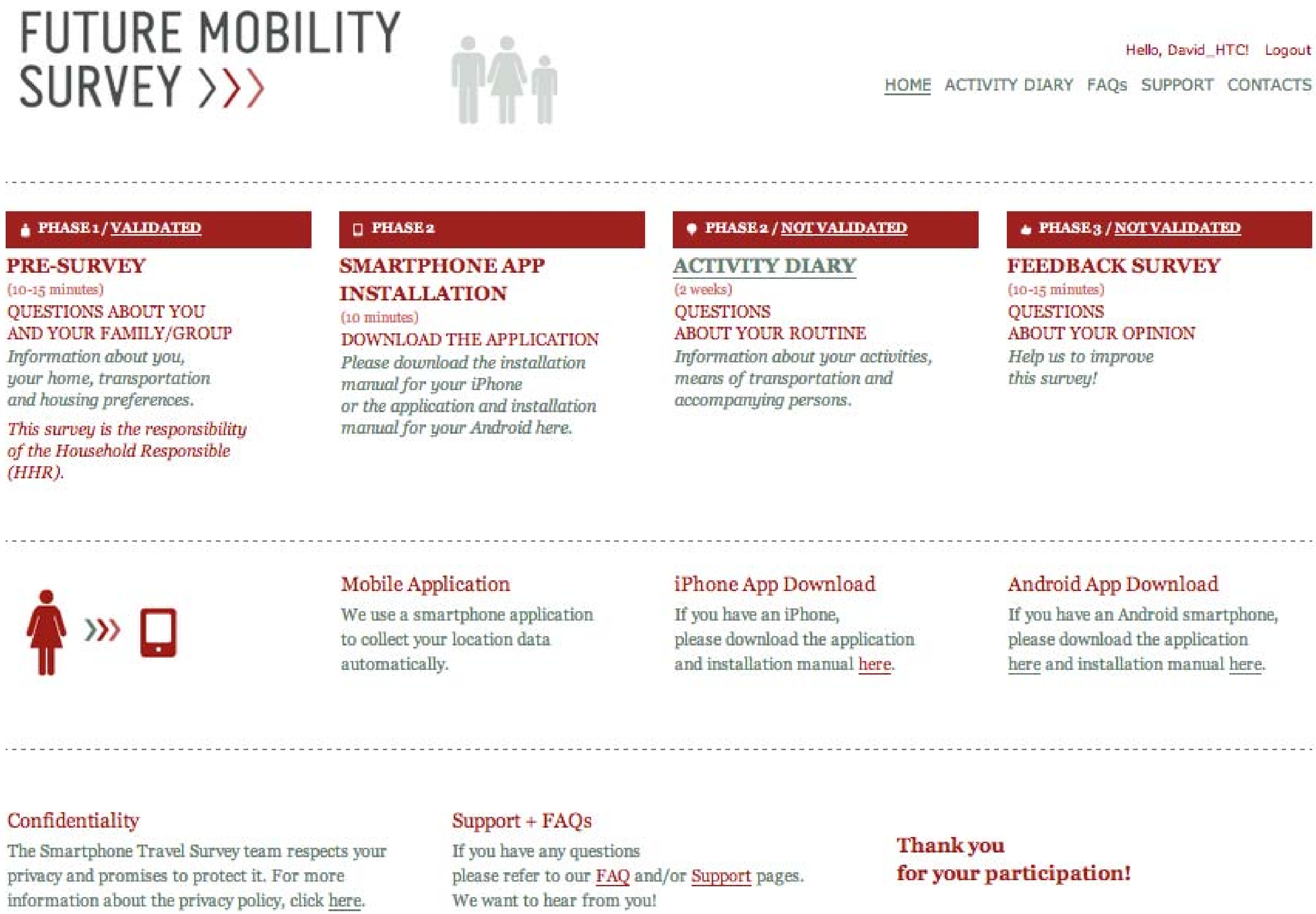
Web Application UI 1

- Survey website front-page design



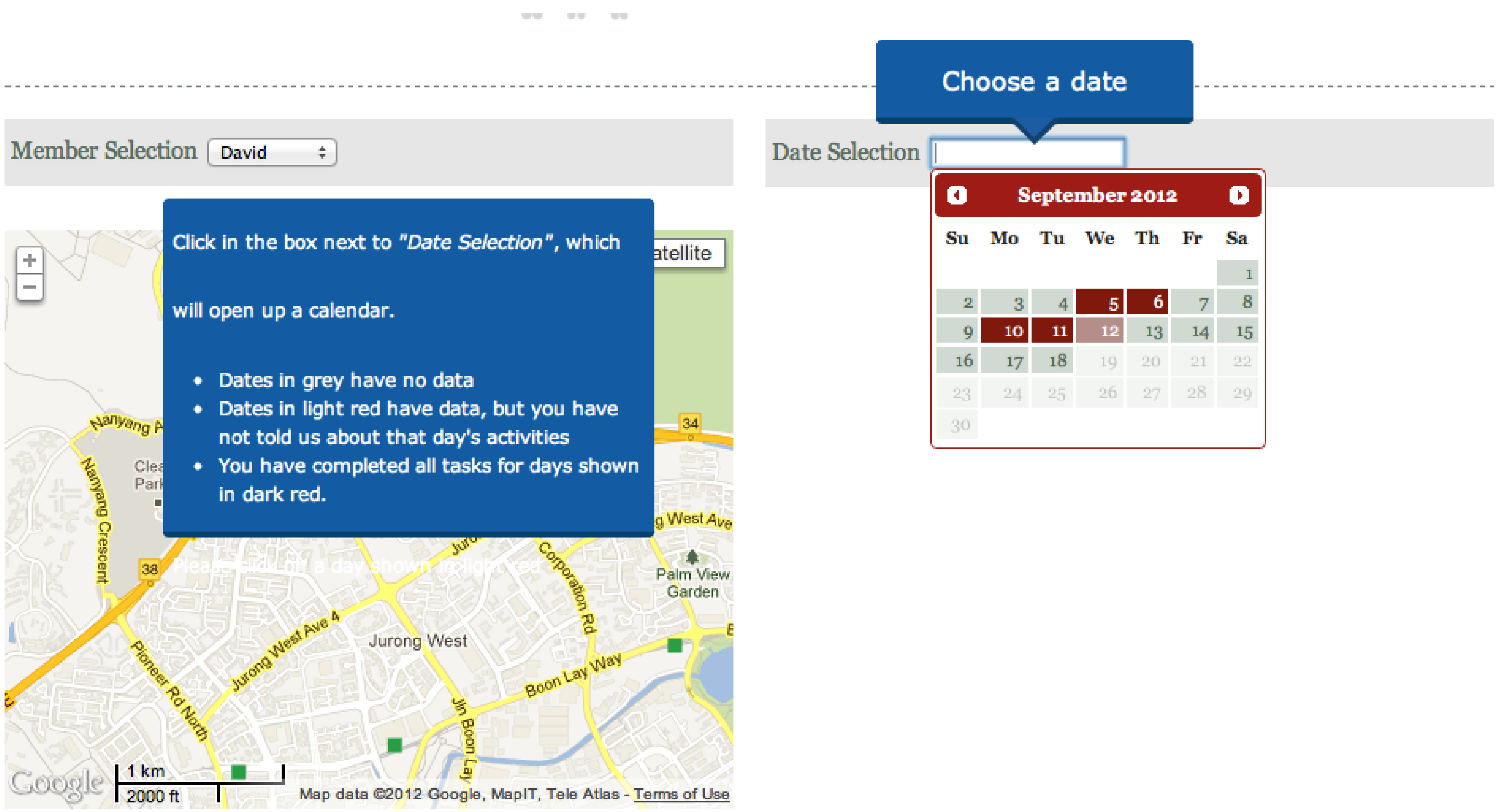
Web Application UI 2

- User page design (After logged in)



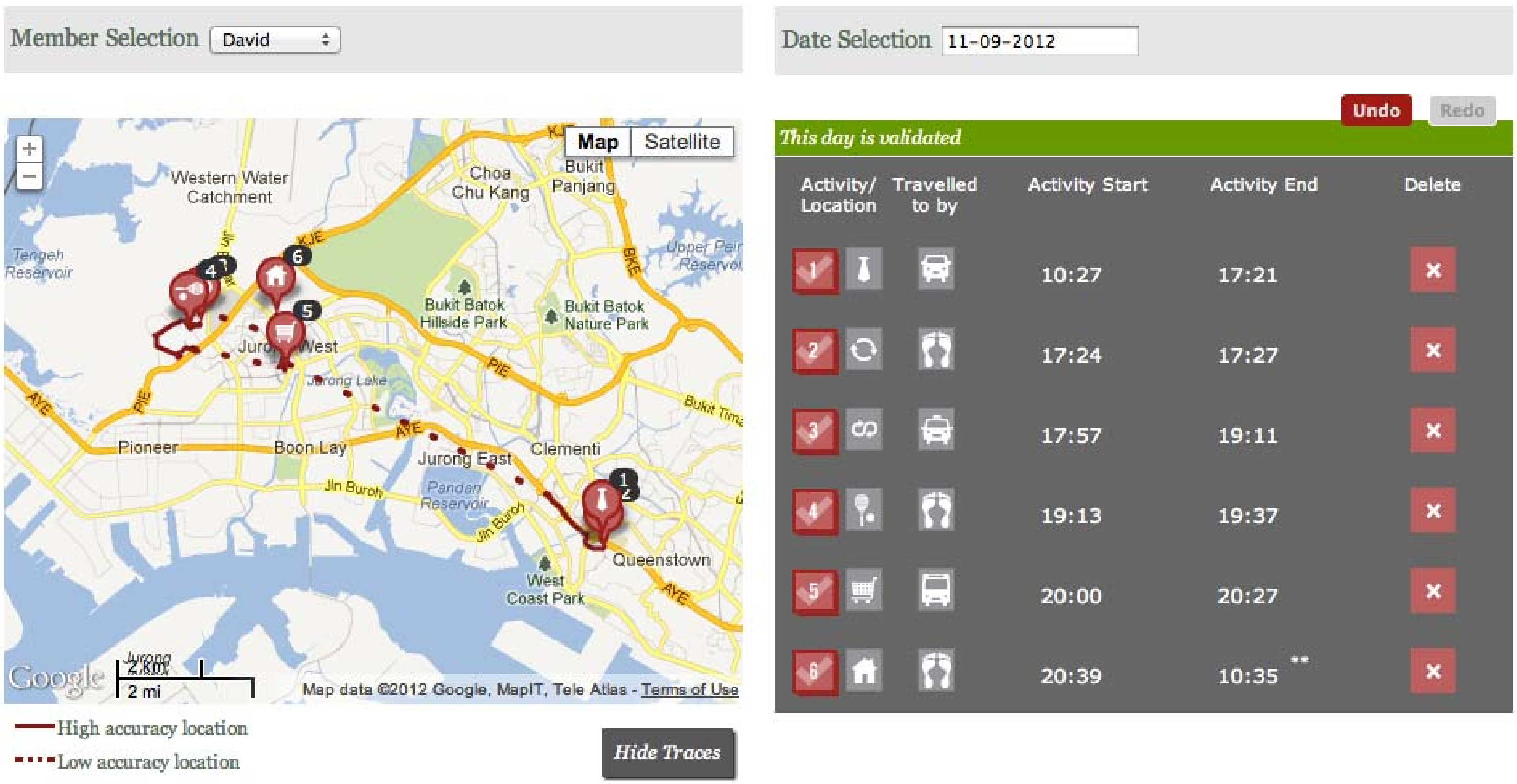
Web Application UI 3

- Activity diary date selection



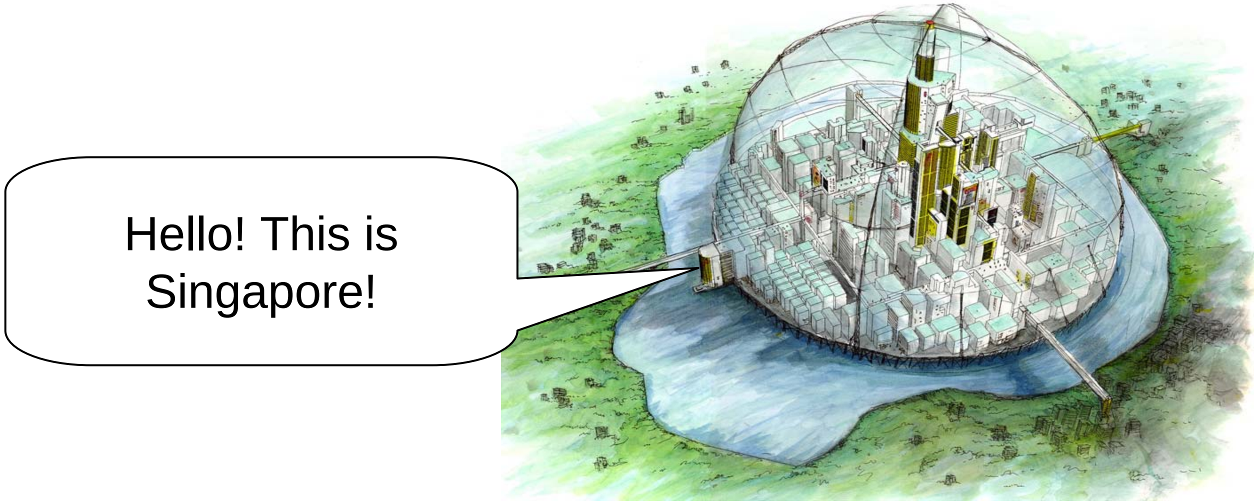
Web Application UI 4

- Trip validation (demo video)

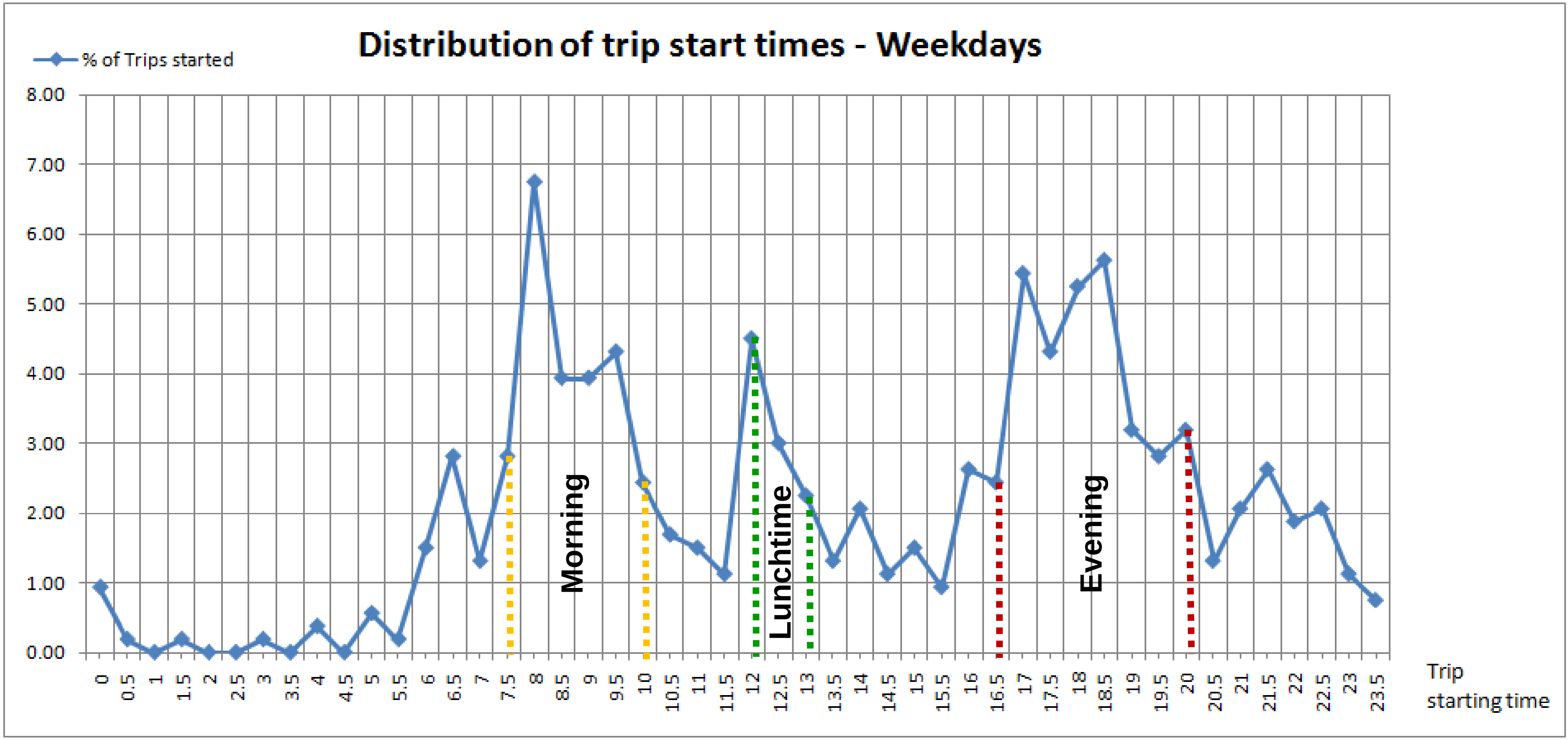


Live Singapore!

- Knowledge behind sensing data.
  - What is going on in the "City within a Shopping Mall“?
  - City Mobility: moving pattern of Singaporean.
    - When is the peak hour?
    - How long are Singaporean’s trips at weekdays?

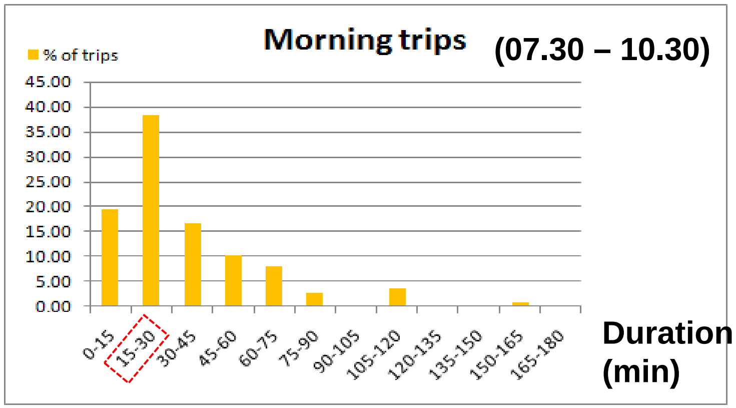


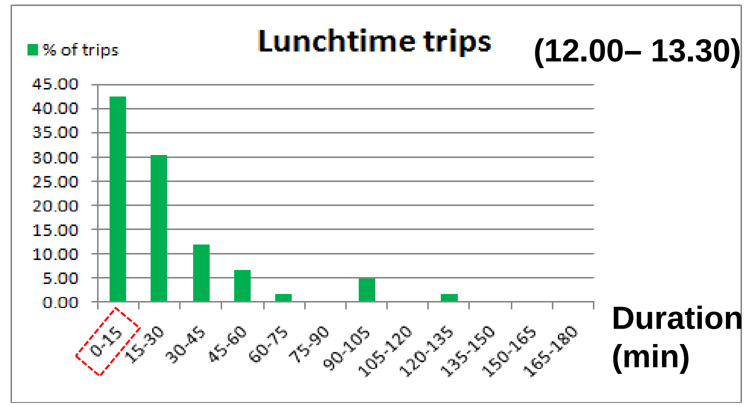
Distribution of Trips

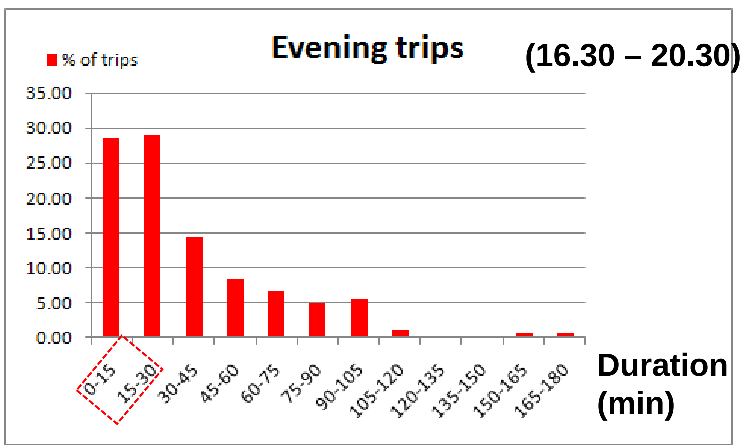


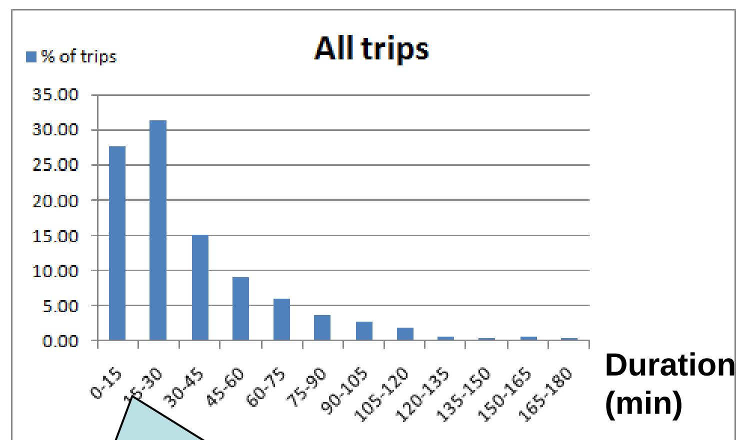
Identified peak times: Morning (07.30 - 10.30), Lunchtime (12.00 - 13.30), Evening (16.30 - 20.30)

Distribution of Trip Durations



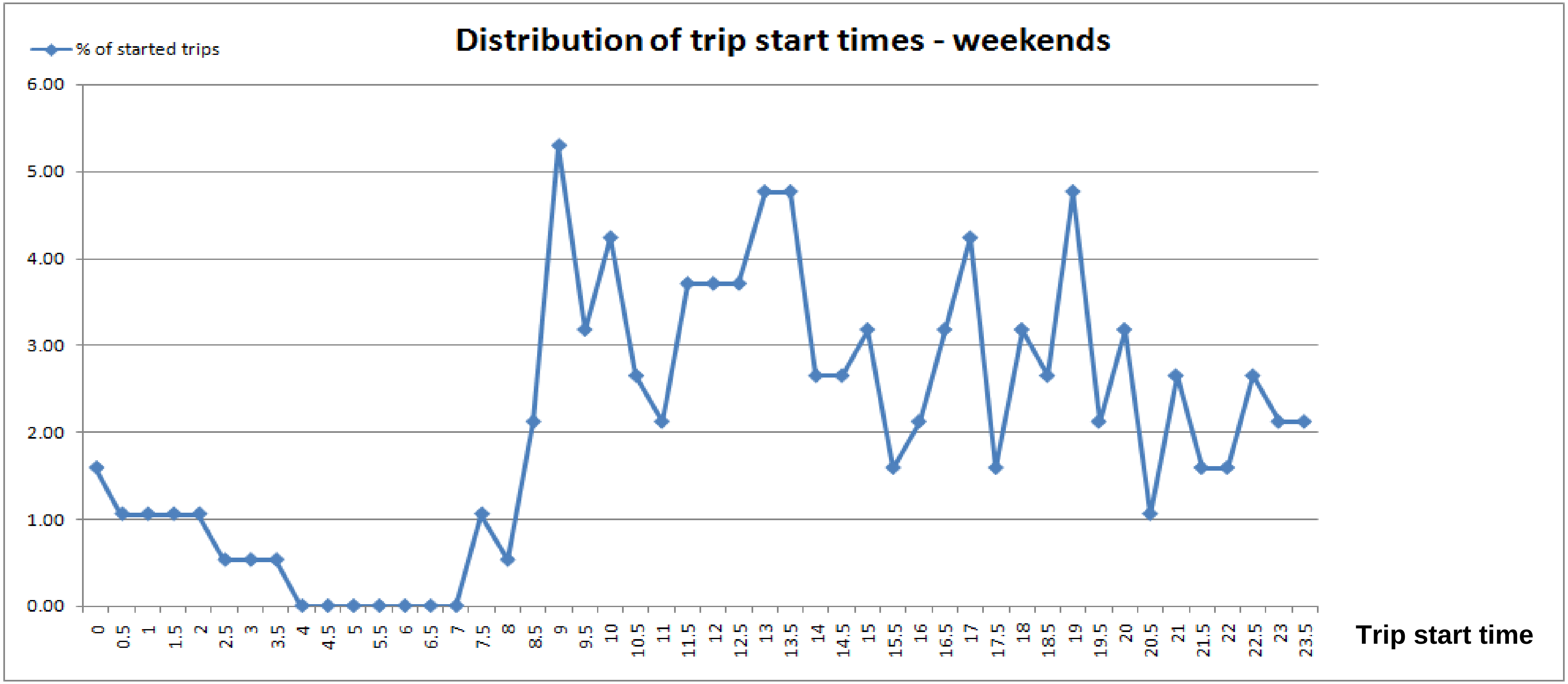






50% trips in SG are less than 30 minutes.

Distribution of Trips for Weekends



We don't see much strong patterns

Sensor-aided Urban Planning

- Green Singapore:
  - Turn off the street lights partially while Singapore goes to snore..
- Improvement on public transport systems
  - Dynamically add more MRT trains or Buses during the peak period.
- Transportation guide for personal trips
  - Taking Cabs/MRT during the peak period.

From Research towards Release

- Our app has been available on App store and Google play
- Testing targets at the current most popular phones in SG and US

Singapore – September '12 (M1, Starhub, Singtel)	(US) – August '12 (AT&T, Verizon, T-Mobile, Sprint
1. Samsung Galaxy Ace 2 2. Samsung Galaxy S 3 3. Samsung Galaxy Note 4. Apple iPhone 4S 5. Sony Xperia go 6. LG Optimas 4X HD 7. Samsung Galaxy S 2 8. HTC Desire V	1. Apple iPhone 4S 2. Samsung Galaxy S 3 3. Motorola RAZR MAXX 4. HTC One S 5. HTC One X 6. HTC EVO 4G LTE

- New/Upcoming Phones
  - Apple iPhone 5 – Popular in SG and US.
  - Samsung Galaxy S3 – Popular in SG and US.
  - Samsung Ace 2 – Popular in SG.
  - Sony Xperia T – Sony is becoming popular.
  - Motarola RAZR MAXX – Popular in US.

Conclusion and Future Work

- Our system has been deployed in Singapore to support Land Transport Authority's (LTA) travel survey in 2012.
  - Several algorithms have been developed to support the individual users' mobility analysis
- Enlarge the training set and enhance the algorithm accuracy, such as rule-based transportation mode classification
- Stop detection can be improved with users' historical information and other context information, such as bus route and social events;
- Research on cloud computing and P2P infrastructure to enhance the system scalability and flexibility.