

MINING DATA FOR TRAFFIC DETECTION SYSTEM USING GPS_ENABLE MOBILE PHONE IN MOBILE CLOUD INFRASTRUCTURE

Swe Swe Aung and Thinn Thu Naing

University of Computer Studies, Yangon, Myanmar

ABSTRACT

The increasing need for traffic detection system has become a vital area in both developing and developed countries. However, it is more important to get the accurate and valuable data to give the better result about traffic condition. For this reason, this paper proposes an approach of tracking traffic data as cheap as possible in terms of communication, computation and energy efficient ways by using mobile phone network. This system gives the information of which vehicles are running on which location and how much speed for the Traffic Detection System. The GPS sensor of mobile device will be mainly utilized to guess a user's transportation mode, then it integrates cloud environment to enhance the limitation of mobile device, such as storage, energy and computing power. This system includes three main components: Client Interface, Server process and Cloud Storage. Some tasks are carried out on the Client. Therefore, it greatly reduces the bottleneck situation on Server side in efficient way. Most of tasks are executed on the Server and history data are stored on the Cloud Storage. Moreover, the paper mainly uses the distance based clustering algorithm in grouping mobile devices on the same bus to get the accurate data.

KEYWORDS

GPS, Mobile Phone, Client, Server, Cloud

1. INTRODUCTION

Traffic prediction is becoming an important role both in the developed and in the developing countries. There are also plenty of techniques to be used in getting data for the traffic prediction system all over the world. Cameras, road sensors on roadside, HD traffic (traffic information service), GPS devices and Accelerometers in vehicles, and Traffic signals have been being used in almost traffic prediction system. However, these techniques are very expensive in term of initial installation cost as well as motivation and initiative of communities such as country infrastructure and policies concern and greatly need to emphasize on system maintenance. Because of the consequence of these ways, some developing countries are not able to utilize it and also not able to support these systems enough in their region. Fortunately, mobile phones with the riches of sensor, such as an accelerometer, digital compass, gyroscope, GPS, microphone and camera, develop immediately in the mobile market. Many clues, such as position, speed, direction and time, can be looked for by using an accelerometer and GPS enable mobile phone. Besides, these sensors are enabling new applications across a wide variety of domains, such as healthcare [3], social networks [4], safety environmental monitoring [5], and transportation [6], traffic detection system, and give arise to a new area of research called mobile phone sensing [2]. Tracking the traffic clues using raw GPS from mobile device, the current situation of traffic jam

International Journal on Cloud Computing: Services and Architecture (IJCCSA) ,Vol. 4, No. 3, June 2014
 can be predicted on the specific road. In these cases, mobile information need to be accurate, otherwise the performance of the traffic prediction system may be at zero level.

Motivation for transportation mode detection is transportation surveys and traffic clues for traffic detection system. In the way of getting traffic clue, there are multiple formats, such as via telephone and media such as Television and radio. These data collection strategies rely on manual labelling of data after the trip, and thus, inaccuracies are introduced. For example, a traveller may not recall the exact time that she/he boarded a transportation mode. Using GPS devices is more reliable for reporting accurate location, trip time, and trip duration. Hence, if the precise transportation modes of individual users are recognized, it is possible to provide more realistic traffic clues.

Mode of transportation data (walking or driving) provides the valuable information and services for traffic detection system. To analyse transportation mode, data filtering, mode stage determination and mode identification approaches are needed to be carried out. However, current mobile devices are not able to perform these CPU intensive processes for ages. Moreover, continuously growing sample sizes lead to an increasing demand for storage and for automated procedures with low computational cost. For this weakness, only the Cloud Computing can help to improve the reliability and efficiency of the Mining Traffic Data System. Mobile devices are not suitable for computation intensive applications. It still has the limitation of storage, power and CPU. To cover these weaknesses, this system takes the advantages of the Cloud Computing to improve the quality. This task greatly improves in the reduction of the power usage and amount of storage on the mobile phone as well as the leverage of Cloud storage.

Many GPS trace sharing social networks have been implemented by many researchers. These social networks enable friends to upload and share their GPS traces. Knowledge of transportation mode, added to these GPS traces, will enable the users to reflect on their past motion more meaningfully. It also allows users to obtain additional information from their friends' travel experience. Additionally, awareness of transportation mode of a user may help to determine the user's carbon footprint, or track the amount of calories burnt. Another application of transportation mode detection is crowd sourced real-time traffic information in which traffic speeds are aggregated from probes such as mobile phones carried by travellers. Transportation mode detection enables the aggregation system to filter out the speed data submitted by non-motorized travellers or travellers on trains. [10]

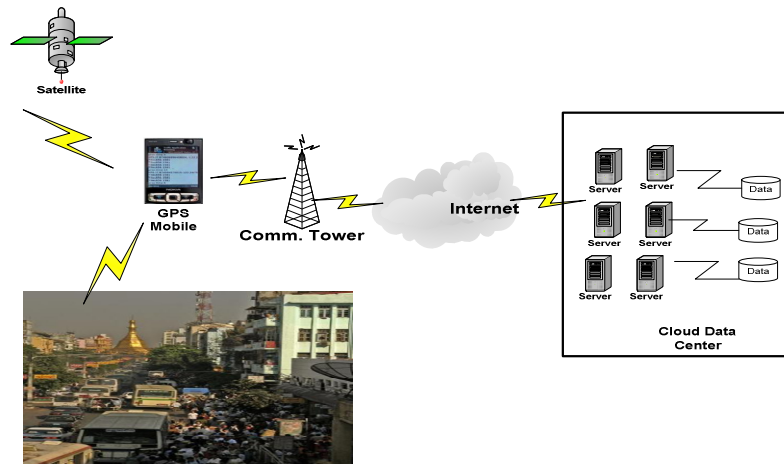


Figure 1. General Architecture of Mining Data for Traffic Detection System

2. DESIGN OF SYSTEM ARCHITECTURE

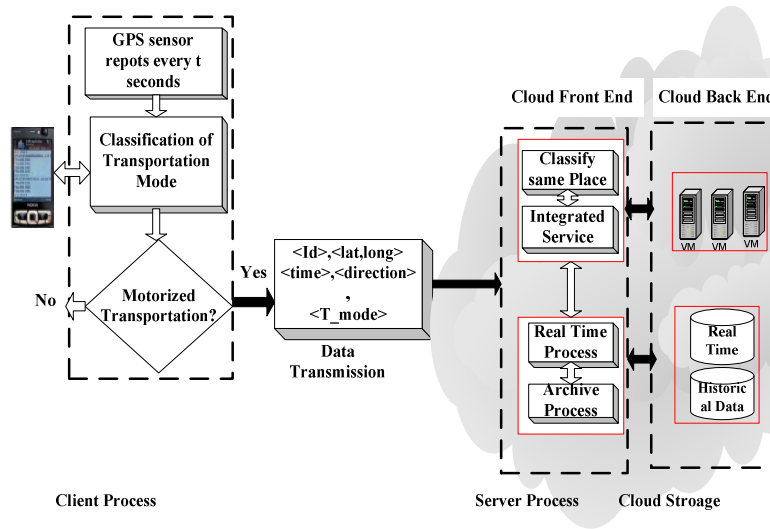


Figure 2. Design of System Architecture

This system focuses on analysing whether the user is travelling with motorized transportation mode or non-motorized transportation mode by only using GPS and an accelerometer data. Moreover, there are two main components deployed in this system such as: Server Process and Client Process.

For estimating the situation of traffic, it does not need to analyse the raw GPS data to get the detail information whether bike or car or motorbike. While bus, taxi, biking and motorbike are defined as the group of motorized transportation mode, only walking man is recognized as non-motorized mode in this system. This system does not take into account every GPS as traffic probe. If the average speeds to the given time window is less than the predefined threshold, it automatically rejects this probe. Only the assurance data, e.g. motorize transportation mode, is sent to server. This work is only done on Client, mobile device, to be accurate and be far away from the incorrect data submission due to the Internet connection. This way greatly reduces enormous amount of ineffective jobs, such as communication cost, sending data repeatedly and Server bottleneck problems, for the system.

On the Server side process, it detects whether the submitted data comes from the same vehicle or not as an initial stage. If so, all data from the same bus is taken as only one traffic probe. Later, these probes are going to be used as a historical data to improve the quality of the system. Furthermore, this system is divided into two parts, Client side and Cloud side. Some tasks are done on the Client and the rests which are computation intensive. The data storage parts are carried out on the Cloud storage. In these cases, all history data are stored on the Cloud. On the other word, this idea can be named as Mobile Cloud Computing.

Mobile Cloud Computing is widely accepted as a concept that can significantly improve the user experience when accessing mobile service. By removing the limitation of mobile devices with respect to storage and computing capabilities and providing a new level of security by a centralized maintenance of security-critical software for e.g. mobile payment applications, it is expected that it will fine brad acceptance on the business as well as consumer side. Research

International Journal on Cloud Computing: Services and Architecture (IJCCSA) ,Vol. 4, No. 3, June 2014 indicates that Mobile Cloud Computing will additionally help to make visions of context-aware services become reality [7].

3. RELATED WORK

D. Patterson, et.al [1] and L. Liao, et.al [11] used an unsupervised learning technique to detect the transportation mode of a traveller. The transportation modes that were detected in this system include buses, cars and walk. The work in the system was able to predict the traveler's goals, such as trip destination and trip purpose. In addition to GPS data, the system used historical information about the user. Historical information includes, past user trips and information about where the users parked their cars.

N. D. Lane, et.al [2] also determined the user's current mode of transportation by applying a Boosted Naïve Bayes classifier to data collected from the sensors of the mobile device for the analysis of traffic patterns and CO2 emissions. For generating this Boosted Naïve Bayes classifier, it needed to develop an application to collect training data firstly. Importantly, the propose approach was different than that in [2]. The system did not give bother to get the training data previously. It kept the true data from the result of mobile device as a historical data and afterward these history data are integrated with the live result to improve the accuracy.

Liu [8] has a different focus; it considered extracting the semantic location from outdoor positioning systems. Likewise, Hightower [9] learnt and recognized the places a mobile user visited by observing the Wi-Fi and GSM radio fingerprints. This work did not consider Wi-Fi or GSM information.

4. SYSTEM MODEL

The system is composed of two parts, Client and Server side.

4.1. Client Side

4.2. Algorithm for filtering raw GPS data

To illustrate the flow of filtering raw GPS data, this system needs to identify the following notations for system parameters.

Table 1. GPS sensor log descriptions.

Notation	Description
P	GPS point which includes latitude, longitude, speed and time.
pi	set of GPS point(lat,lon,v,t) , $pi \in P(i=1,2, \dots ,n)$
lat	Latitude
lon	Longitude
t	the timestamp of the sensor report in seconds
v	the current ground speed of the device
h	the direction of travel
acc	the accuracy level of the latitude and longitude coordinates
tm	transportation mode(non walking)
nm	non transportation mode(walking)
GPS log	a sequence of GPS points
R	<id,lat,lon,v,t,s,tm>
D	threshold for calculating distance is 0.6
V _t	threshold for calculating average speed is 2 m/s.

Assume it computes the average speed every 15 seconds.

Table 2. Client Process Algorithm.

Client Process Algorithm	
Input : GPS data= <id,lat,lon,v,t> Output : R=<id,lat,lon,v,t,s,tm>	
Begin	
Step 1.	Receipt $p=\{p1,p2, \dots , pn\}$ within a time window
Step 2.	Average Speed (P.V)= $(\sum_{i=1}^n piv) / n$ If(P.V > V _t)
Step 3.	Submit R to backend Server
	Else
Step 4.	Clculate distance (D) within five minutes If(D>threshold)
Step 5.	Submit R to backend Server
	Else
Step 6.	Go to step 1
End	

This algorithm is for computing the filtering process on the mobile device (Client side). Responsively it has to receipt the raw GPS data from the device and then filters out mode of transportation (driving or bus) by calculating these raw GPS data. In other word, it classifies that the mobile device carrier is taking bus or walking. However, there are some problems in analyzing the Transportation mode (walking or not). In real world, the running buses have to stop

International Journal on Cloud Computing: Services and Architecture (IJCCSA), Vol. 4, No. 3, June 2014
 at the any bus stop and at the traffic light. Consequently, it sometimes makes the speed of the buses be under 1.0 constantly. Even though the bus has traveled for much distance, the velocity is almost the same with the walking man's speed. To prevent this real time error, it filters out again to get more accurate result although the speed is lower than the threshold. If the final result is Transportation mode (driving or bus), it will send location, speed of device and time to the backend server. Otherwise, the system will begin its process.

5. SERVER SIDE

5.1. Server Side Process

Every data submitted form mobile device at every t time will be classified to define accurate traffic data. Firstly, these data are analyzed whether the corresponding devices are in the same place or not according the location (latitude and longitude). In this case, this system needs to group the numerous GPS_location into similar counterpart. For this case, one of the data mining functionalities, cluster analysis is going to be used in this research. In detail one of the clustering algorithm, distance based clustering is specially utilized. In this case, this clustering algorithm operates on Dissimilarity matrix (or object-by-object structure).

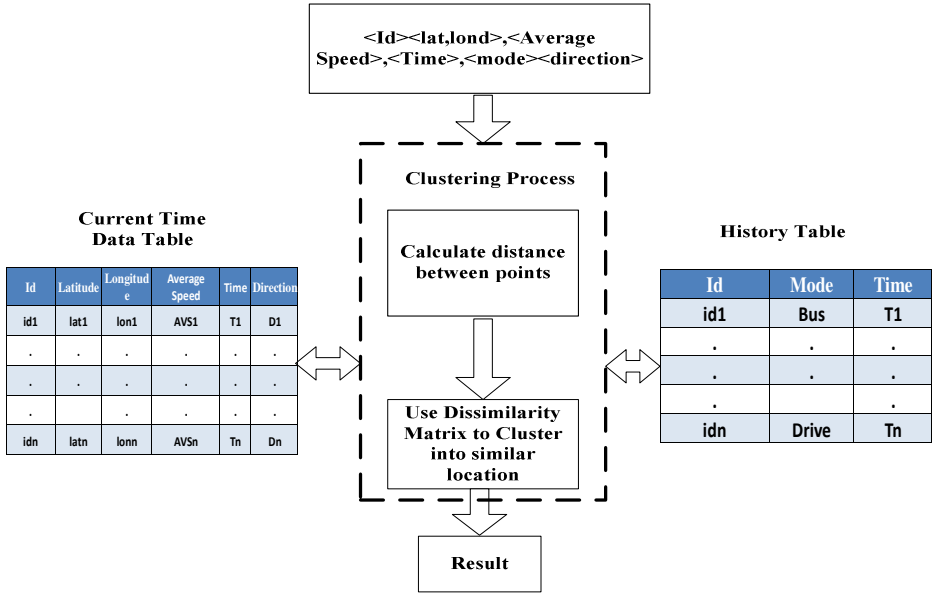


Figure 3. Design of System Architecture on Server Side

The data flow diagram shows the process of calculating the distance between the current input data and the previous data in the current time table by using Haversine formula and the clustering the GPS points into similar location with Dissimilarity Matrix. This system has to access to the current data table to know whether it already exist or not and history data for storing the result for future prediction. Finally, it produces the result, real time data result for traffic prediction system.

5.2. Algorithm for clustering GPS data on Server site

Table 3. Server Side Clustering Algorithm.

Server Side Clustering Algorithm	
Input : <id,lat,lon,v,t,s,tm> Output : real time data result	
Begin	
Step 1.	Receipt input <id,lat,lon,v,t,s,tm>
Step 2.	Clustering Process <ol style="list-style-type: none"> 1. Search already exist in the current time table 2. If(exist) go to Step 3 3. Else Calculate distance between the input GPS point and the previous data in the current table by using Haversine formula 4. Use dissimilarity Matrix to cluster into similar location.
Step 3.	Store as a new one or update the old data
Step 4.	Output real time data result
End	

The Server side algorithm intends to cluster the PGS points into similar counterpart to reduce non usable data. Firstly, it receipts the data passed from the Client side. Then, it explores the current time table whether already exit or not. If it already exists, it skips distance calculation and clustering process. Secondly, it store as a new one or update the old data in the history data. Finally, the data result from this system is ready to use for traffic prediction system.

6. SERVER SIDE CALCULATION

There are data collection and computing distance between points by using Haversine formula. After that, these data are clustered into similar counterpart with Dissimilarity matrix (or object-by-object structure).

Table 4. Data Matrix

Point	Latitude	Longitude
X ₁	lat ₁	lon ₁
X ₂	lat ₂	lon ₂
X ₃	lat ₃	lon ₃
X ₄	lat ₄	lon ₄
X ₅	lat ₅	lon ₅
X ₆	lat ₆	lon ₆

International Journal on Cloud Computing: Services and Architecture (IJCCSA) ,Vol. 4, No. 3, June 2014
 The table 4 denotes the location of devices submitted by the Client side. These data are input for the distance matrix.

Table 5. Matrix for Calculating Distance.

	X1	X2	X3	X4	X5	X6
X1	0	d(X1,X2)	d(X1,X3)	d(X1,X4)	d(X1,X5)	d(X1,X6)
X2	d(X2,X1)	0	d(X2,X3)	d(X2,X4)	d(X2,X5)	d(X2,X6)
X3	d(X3,X1)	d(X3,X2)	0	d(X3,X4)	d(X3,X5)	d(X3,X6)
X4	d(X4,X1)	d(X4,X5)	d(X4,X3)	0	d(X4,X5)	d(X4,X6)
X5	d(X5,X1)	d(X5,X2)	d(X5,X3)	d(X5,X4)	0	d(X5,X5)
X6	d(X6,X1)	d(X6,X2)	d(X6,X3)	d(X6,X4)	d(X6,X5)	0

Where $d(x_i, x_j)$ is the measured difference between points i and j . $d(x_i, x_j)$ is a nonnegative number that is close to 0 when two points i and j are near each other.

The distance between two points is calculated by using Haversine formula.

Haversine formula: $a = \sin^2(\Delta\phi/2) + \cos(\phi_1) \cdot \cos(\phi_2) \cdot \sin^2(\Delta\lambda/2)$
 $c = 2 \cdot \text{atan2}(\sqrt{a}, \sqrt{1-a})$
 $d = R \cdot c$

where ϕ is latitude, λ is longitude, R is earth’s radius (mean radius = 6,371km)
 note that angles need to be in radians to pass to trig functions!

After they have been clustered into the similar group, their average speeds are executed again to predict that they are on the same vehicle. After that, these outcomes are integrated with the history data. The history data means the mobile device frequently takes the bus or drives a car. The latest outcome results are stored altogether with the previous history data on the Cloud storage with the corresponding mobile id.

6.1. Data Collection

The table 6 represents one GPS point (Latitude and Longitude) collected from the three devices, Sony, Note 8 and Galaxy Age, on the same bus at the same time along the Pyay Road in Yangon, Myanmar.

Table 6. Real Data Form

Point(X)	Latitude	Longitude
Sony	16.85573266	96.13369122
Note 8	16.85600044	96.13369122
Galaxy Age	16.85602302	96.13387954

Let Sony be X_1 , Note 8 be X_2 , Galaxy Age be X_3

Table 7. Distance Matrix based on Real Time Data

	X ₁	X ₂	X ₃
X ₁	d(X ₁ ,X ₁)	d(X ₁ ,X ₂)	d(X ₁ ,X ₃)
X ₂	d(X ₂ ,X ₁)	d(X ₂ ,X ₂)	d(X ₂ ,X ₃)
X ₃	d(X ₃ ,X ₁)	d(X ₃ ,X ₂)	d(X ₃ ,X ₃)

The table 7 shows the distance matrix based on real time data. The matrix computes the distance among devices. For example, d(X₁, X₂) is a function that calculates the distance between X₁ device and X₂ device and so on. In this calculation, function d(X_n, X_n) accomplishes its task by using Haversine formula.

Table 8. Distance Matrix based on Real Time Data

	X1	X2	X3
X1	0 km	0.03589 km	0.03234 km
X2	0.03589 km	0 km	0.02020 km
X3	0.03234 km	0.02020 km	0 km

The table 8 gives the result of the table 5. It computed the distance between X1 and X2, X1 and X3, and X2 and X3. It shows these devices be on the same bus because the distance between these devices is lower than 0.01 km.

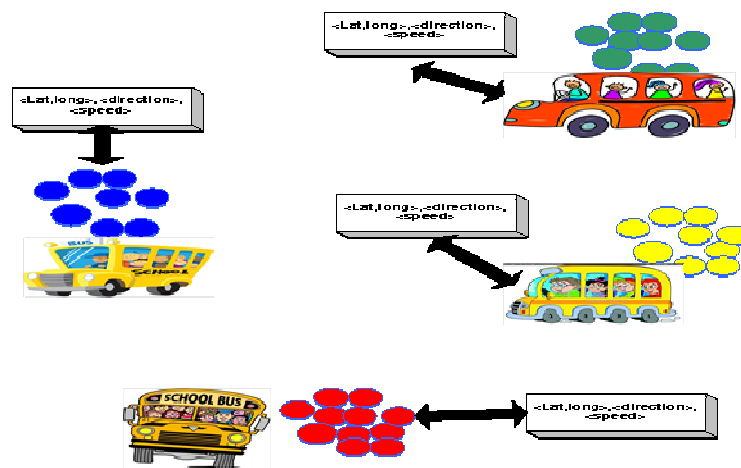


Figure 4. the simple example of clustering latitude and longitude into similar group

In the traffic detection system, all populated data should not be taken into account as traffic data. It needs to filter these data into the actual one to calculate the percentage of traffic condition. For example, four mobile devices on the same bus send their GPS data to the backend Server.

International Journal on Cloud Computing: Services and Architecture (IJCCSA) ,Vol. 4, No. 3, June 2014
 However, the Server takes one traffic clue (one vehicle) instead of four vehicles because it automatically knows that they are on the same bus. The figure 4 represents the characteristic of filtering GPS data into valuable data.

Table 9. Final Real Time Data Result

Id	Latitude	Longitude	Speed	Time	Vehicle
1	16.84818682	96.13054588	5.702219	12/18/13 7:59 AM	V1
2	16.84810627	96.1305606	5.808751	12/18/2013 7:59	V2
3	16.84805336	96.13063597	6.255025	12/18/2013 7:59	V3
4	16.84802346	96.13067158	6.678388	12/18/2013 7:59	V4
5	16.84804854	96.130737	6.580866	12/18/2013 7:59	V4
6	16.84808068	96.13082407	7.1498036	12/18/2013 7:59	V5
7	16.84808552	96.13090513	7.278797	12/18/2013 7:59	V6
8	16.84811254	96.13098263	7.532394	12/18/2013 7:59	V7
9	16.84811614	96.13106725	7.9681973	12/18/2013 7:59	V8
10	16.84812296	96.13114856	8.02197	12/18/2013 7:59	V9
11	16.84815176	96.13121787	8.276309	12/18/2013 7:59	V10
12	16.84816573	96.13130099	8.288028	12/18/2013 7:59	V11
13	16.84818595	96.13136981	8.507573	12/18/2013 7:59	V12

The table 9 illustrates the final result based on real time data. It also shows the vehicle's speed, location and time. The traffic detection system can compute the condition of traffic on the specified route by using these data.

The figure 5 displays the comparison between the speeds of two devices on the same bus. It also shows that the speed is not different completely if the devices are on the same bus.

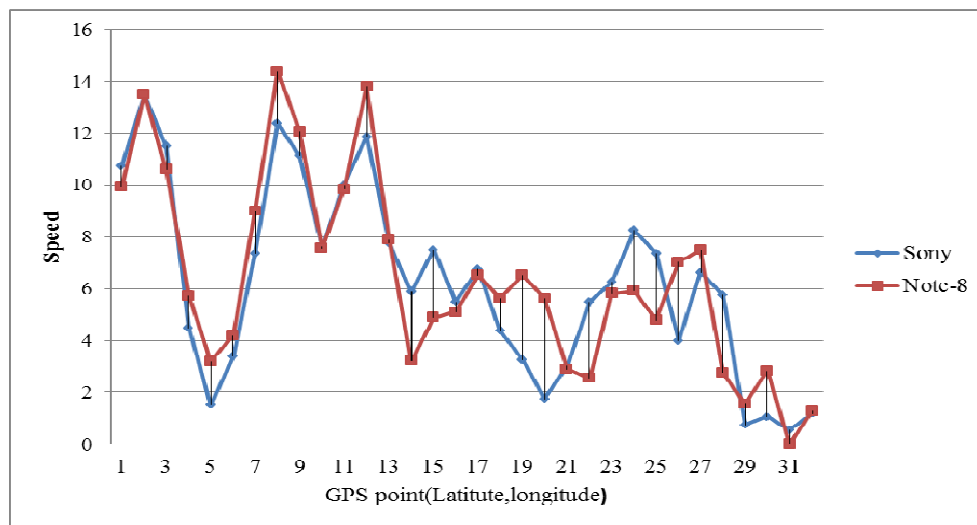


Figure 5. The detection of difference in speed of two mobile phones carrying on the same bus by using GPS

7. CONCLUSIONS

In this paper, we emphasize to provide the accurate traffic probe based on GPS data. This research intends to infer transportation modes from GPS logs based on unsupervised approach. Using our approach, it primarily provides a reduction in cost of approaching the way of getting the true traffic data. Moreover, some part of processes carried out on Client or backend Server also save in cost. Furthermore, it is impossible to assume all population to be traffic clues. For this reason, it computes data from all mobile devices to know whether on the same bus or not by using distance based clustering algorithm. From the standpoint of transportation agencies, the system comes at almost no installation and maintenance cost. Mobile phone sensing systems will ultimately provide both micro- and macroscopic views of cities, communities, and individuals, and help improve how society functions as a whole.

ACKNOWLEDGEMENTS

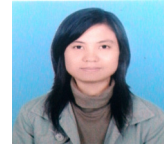
The authors would like to thank everyone, just everyone!

REFERENCES

- [1] D.Patterson, L. Liao, D. Fox, and H. Kautz, Inferring High-Level Behavior from Low-Level Sensors, ACM UBICOMP 2003.
- [2] Nicholas D. Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury, and Andrew T. Campbell. A survey of Mobile Phone Sensing, September 2010.
- [3] S.Consolvo et al., "Activity Sensing in the Wild: A Field Trial of Ubifit Garden," Proc. 26th Annual ACM SIGCHI Conf. Human Factors Comp. Sys., 2008, pp. 1797–1806.
- [4] E.Miluzzo et al., "Sensing meets Mobile Social Networks: The Design, Implementation, and Evaluation of the CenceMe Application," Proc. 6th ACM SenSys, 2008, pp. 337–50.
- [5] M. Mun et al., "Peir, the Personal Environmental Impact Report, as a Platform for Participatory Sensing Systems Research," Proc. 7th ACM MobiSys, 2009, pp. 55–68.
- [6] A. Thiagarajan et al., "VTrack: Accurate, Energy-Aware Traffic Delay Estimation Using Mobile Phones," Proc. 7th ACM SenSys, Berkeley, CA, Nov. 2009.
- [7] Andreas Klein, Christian Mannweiler, Joerg Schneider and Hans D. Schotten Chair for Wireless Communications and Navigation, "Access Schemes for Mobile Cloud Computing", 2010.
- [8] J. Liu, O. Wolfson, H. Yin. Extracting Semantic Location from Outdoor Positioning Systems. Int. Workshop on Managing Context Information and Semantics in Mobile Environments (MCISME), 2006.
- [9] J. Hightower, S. Consolvo, A. LaMarca, I. Smith, J. Hughes. Learning and Recognizing the Places We Go. ACM Conference on Ubiquitous Computing, 2005.
- [10] Leon Stenneth, Ouri Wolfson, Philip S. Yu, Bo Xu, "Transportation Mode Detection using Mobile Phones and GIS Information", November 2011.
- [11] L. Liao, D. Fox, H. Kautz. Learning and Inferring Transportation Routines. AAAI 2004.
- [12] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W. Ma, Understanding mobility based on GPS data. In Ubiquitous Computing, ACM New York, 2008, pp. 312-321.
- [13] S. Reddy, M. Mun, J. Burke, D. Estrin, M Hansen, and M. Srivastava. Using Mobile Phones to Determine Transportation Modes. ACM Transactions on Sensor Networks, Vol.6, No. 2, Article 13, 2010.
- [14] Parkka, J., Ermes, M., Korpipaa P., Mantjarvi J., Peltola, J., Activity classification using realistic data from wearable sensors, IEEE Transactions on Information Technology in Biomedicine 10, 1 (2006), 119-128.

Authors

Swe Swe Aung. She received the Bachelor of Computer Science degree in 2004 from Computer University, Loikaw, Myanmar and Master of Computer Science degree in 2009 from the Computer University, Taunggyi, Myanmar. She is currently working toward the Ph.D degree at University of Computer studies, Yangon, Myanmar. Her research interests in the mobile computing , cloud computing and transportation technologies.



Thinn Thu Naing. She obtained her Ph.D degree in Computer Science from University of Computer Studies, Yangon in 2004, and Bachelor of Computer Science degree and Master of Computer Science degree in 1994 and 1997 respectively from University of Computer Studies, Yangon. Currently, she is a Professor in Computer Science at University of Computer Studies, Yangon, Myanmar. Her specialization includes transportation technologies, cluster computing, grid computing, cloud computing, distributed computing and mobile computing.

