ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
Available online at www.ijarbest.com

International Journal of Advanced Research in Biology Ecology Science and Technology (IJARBEST)
Vol. I, Special Issue I, August 2015 in association with VEL TECH HIGH TECH DR. RANGARAJAN DR. SAKUNTHALA ENGINEERING COLLEGE, CHENNAI
National Conference on Recent Technologies for Sustainable Development 2015 [RECHZIG'15] - 28th August 2015

Cellular Based Bus Tracking System Using Participatory Sensing

[1] Mrs W.ANCY BREEN [2] H SAMSUNHAR , [3] P LAKSHMI PRIYA ,[4] M SINDHUJA
[1] ASSISTANT PROFESSOR [234] UG SCHOLAR
Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College

Abstract-

The bus arrival time is primary information to most city transport travellers. Waiting for a long time for the bus very often discourages the travellers and make them reluctant to take buses. In this paper, we present a bus arrival time prediction system based on bus passengers participatory sensing. Using our mobile phones, the bus passengers can get the context of surrounding environment effectively and we can utilize to estimate the bus traveling routes and can predict the bus arrival time at various bus stops.The proposed system solely relies on the collaborative effort of participating users and is independent from the bus operating companies, so it can be easily adopted to support universal bus service system without support from particular bus operating companies. Instead of referring GPS-enabled location information, includes cell tower signals, movement statuses, audio recording, etc., which bring less burden to the participatory party and encourage their participation. We develop a prototype system with different types of Android-based mobile phones and this proposed system achieves outstanding prediction accuracy compared with those bus operator initiated and GPS supported solutions. At the sametime, the proposed solution is more generally available and energy friendly.

Index Terms—Bus arrival time prediction, participatory sensing, mobile phones, cellular-based tracking

INTRODUCTION

PUBLIC transport, especially the bus transport, has been well developed in many parts of the world. The bus transport services reduce the private car usage and fuel consumption, and alleviate traffic congestion. As one of the very most comprehensive and affordable means of public transport, in 2011 the bus system serves over 3.3 million bus rides every day on average in area with around 5million residents [1]. When traveling with buses, the travelers usually want to know the accurate arrival time of the bus. Excessively long waiting time at bus stops may drive away the anxious travelers and make them reluctant to take buses. Now a days, most bus operating companies,

have been providing theirtimetables on the web freely available for the travelers. The bus timetables, however, only provide very limited information (e.g., operating hours, time intervals, etc.), which are typically not timely updated. Other than those official timetables, many public services (e.g., Google Maps) are provided for travelers. Although such services offer useful information, they are far from satisfactory to the bus travelers.For example, the schedule of a bus may be delayeddue to many unpredictable factors (e.g., traffic conditions, harsh weather situation, etc). The accurate arrival timeof next bus will allow travelers to take alternative transport choices instead, and thus mitigate their anxiety and improve their experience. Towards this aim, many commercial bus information providers offer the realtime bus arrival time to the public . Providing such services, however, usually requires the cooperation of the bus operating companies.(e.g., installing special location tracking devices on the buses), and incurs

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
Available online at www.ijarbest.com

International Journal of Advanced Research in Biology Ecology Science and Technology (IJARBEST)
Vol. I, Special Issue I, August 2015 in association with VEL TECH HIGH TECH DR. RANGARAJAN DR. SAKUNTHALA ENGINEERING COLLEGE, CHENNAI
National Conference on Recent Technologies for Sustainable Development 2015 [RECHZIG'15] - 28th August 2015

substantial cost. In this paper, we present a novel bus arrival time prediction system based on crowd-participatory sensing. We interviewed bus passengers on acquiring the bus arrival time. Most passengers indicate that they want to instantly track the arrival time of the next buses and they are willing to contribute their location information on buses to help to establish a system to estimate the arrival time at various bus stops for the community. This motivates us to design a crowd-participated service to bridge those who want to know bus arrival time (querying users) to those who are on the bus and able to share the instant bus route information (sharing users). To achieve such a goal, we let the bus passengers themselves cooperatively sense the bus route information using commodity mobile phones. In particular, the sharing passengers may anonymously upload their sensing data collected on buses to a processing server, which intelligently processes the data and distributes useful information to those querying users. Our bus arrival time prediction system comprises three major components:

(1) Sharing users: using commodity mobile phones as well as various build-in sensors sense and report the lightweight cellular signals and thesurrounding environment to a backend server;

(2) Querying users: querying the bus arrival time for a particular bus route with mobile phones;

(3) Backend server: collecting the instantly reported information from the sharing users, and intellectually processing such information so as to monitor the bus routes and predict the bus arrival time. No GPS or explicit location services are invoked to acquire physical location inputs. Such a crowd-participated approach for bus arrival time prediction possesses the following several advantages compared with conventional approaches. First, through directly bridging the sharing and querying users in the participatory framework, we build our system independent of the bus operating companies or other third-party service providers, allowing easy and inexpensive adoption of the proposed approach over other application instances. Second, based on the commodity mobile phones, our system obviates the need for special hardware or extra vehicle devices,which substantially reduces the deployment cost.Compared with conventional approaches GPS supported, our approach is less demanding and much more energy-friendly, encouraging a broader number of participating passengers. Third, through automatically detecting ambient environments and generating bus route related reports, our approach does not require the explicit human inputs from the participants, which facilitates the involvement of participatory parties.Implementing such a participatory sensing based system, however, entails substantial challenges. (1) Bus detection: since the sharing users may travel with diverse means of transport, we need to first let their mobile phones accurately detect whether or not the current user is on a bus and automatically collect useful data only on the bus. Without accurate bus detection, mobile phones may collect irrelevant information to the bus routes, leading to unnecessary energy consumption or even inaccuracy in prediction results. (2) Bus classification: we need to carefully classify the bus

route information from the mixed reports of participatory users. Without users' manual indication, such automatic

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
Available online at www.ijarbest.com

International Journal of Advanced Research in Biology Ecology Science and Technology (IJARBEST)
Vol. I, Special Issue I, August 2015 in association with VEL TECH HIGH TECH DR. RANGARAJAN DR. SAKUNTHALA ENGINEERING COLLEGE, CHENNAI
National Conference on Recent Technologies for Sustainable Development 2015 [RECHZIG'15] - 28th August 2015

classification is non-trivial. (3) Information assembling: One sharing user may not stay on one bus to collect adequate time period of information. Insufficient amount of uploaded information may result in inaccuracy in predicting the bus route. An effective information assembling strategy is required to solve the jigsaw puzzle of combining pieces of incomplete information from multiple users to picture the intact bus route status. In this paper, we develop practical solutions to cope with such challenges. In particular, we extract unique identifiable fingerprints of public transit buses and utilize the microphone on mobile phones to detect the audio indication signals of bus IC card reader. We further leverage the accelerometer of the phone to distinguish the travel pattern of buses to other transport means. Thus we trigger the data collection and transmission only when necessary .We let the mobile phone instantly sense and report the nearby cell tower IDs. We then propose an efficient and robust top-*k* cell tower set sequence matching method to classify the reported cell tower sequences and associate with different bus routes. We intellectually identify passengers on the same tower sequence concatenation approach to assemble their cell tower sequences so as to improve the sequence matching accuracy (Section **??**). Finally, based on accumulated information, we are then able to utilize both historical knowledge and the realtime traffic conditions to accurately predict the bus arrival time of various routes .

BACKGROUND AND MOTIVATION

The bus companies usually provide free bus timetables on the web. Such bus timetables, however, only provide very limited information (e.g., operating hours, time intervals, etc.), which are typically not timely updated according to instant traffic conditions. Although many commercial bus information providers offer the realtime bus arrival information,the service usually comes with substantial cost. With a fleet of thousands of buses, the installment of invehicle GPS systems incurs tens of millions of dollars . The network infrastructure to deliver the transit service raises the deployment cost even higher, which would eventually translate to increased expenditure of passengers. For those reasons, current research works explore new approaches independent of bus companies to acquire transit information. The common rationale of such approaches is to continuously and accurately track the absolute physical location of the buses, which typically uses GPS for localization. Although many GPS-enabled mobile phones are available on the market, a good number of mobile phones are still shipped without GPS modules Those typical limitations of the localization based schemes motivate alternative approaches without using GPS signal or other localization methods. Besides, GPS module consumes substantial amount of energy, significantly reducing the lifetime of power-constrained mobile phones due to the high power consumption, many mobile phone users usually turn off GPS modules to save battery power. The mobile phones in vehicles may perform poorly when they are placed without line-of-sight paths to GPS satellites [10] .To fill this gap, we propose to implement a crowd participated bus arrival time prediction system utilizing cellular signals. Independent of any bus companies, the system bridges

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
Available online at www.ijarbest.com

International Journal of Advanced Research in Biology Ecology Science and Technology (IJARBEST)
Vol. I, Special Issue I, August 2015 in association with VEL TECH HIGH TECH DR. RANGARAJAN DR. SAKUNTHALA ENGINEERING COLLEGE, CHENNAI
National Conference on Recent Technologies for Sustainable Development 2015 [RECHZIG'15] - 28th August 2015

the gap between the querying users who want to know the bus arrival time to the sharing users willing to offer them real time bus information

Fig.Absolute1. localization is unnecessary for arrival time prediction

Independent of any bus companies, the system bridges the gap between the querying users who want to know the bus arrival time to the sharing users willing to offer them realtime bus information. Unifying the participatory users, our design aims to realize the common welfare of the passengers. To encourage more participants, no explicit location services are invoked so as to save the requirement of special hardware support for localization. Compared with the high energy consumption of GPS modules, the marginal energy consumption of collecting cell tower signals is negligible on mobile phones. Our system therefore

utilizes the cell tower signals without reducing battery lifetime on sharing passengers' mobile phones. Our design obviate the need for accurate bus localization. In fact, since the public transit buses travel on certain bus routes (1D routes on 2D space), the knowledge of the current position on the route (1D knowledge) and the average velocity of the bus suffices to predict its arrival time at a bus stop. As shown in Fig. 1, for instance, say the bus is currently at bus stop 1, and a querying user wants to know its arrival time at bus stop 6. Accurate prediction of the arrival time requires the distance between bus stop 1 and 6 along the 1D bus route (but not on the 2D map) and the average velocity of the bus. In general, the physical positions of the bus We classify and track the bus statuses in such a logical space so as to predict the bus arrival time.

Querying user. As depicted in Fig. 2 (right bottom), a querying user queries the bus arrival time by sending the

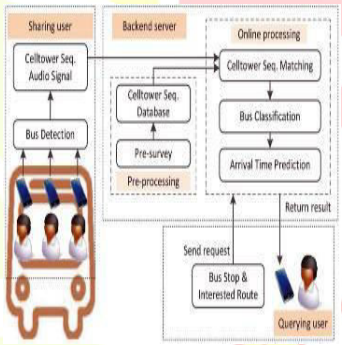


Fig2. System architecture

SYSTEM DESIGN

Though the idea is intuitive, the design of such a system in practice entails substantial challenges. In this section, we describe the major components of the system design. We illustrate the challenges in the design and implementation, and present several techniques to cope with them.

3.1 System Overview

Fig. 2 sketches the architecture of our system.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
Available online at www.ijarbest.com

International Journal of Advanced Research in Biology Ecology Science and Technology (IJARBEST)
Vol. I, Special Issue I, August 2015 in association with VEL TECH HIGH TECH DR. RANGARAJAN DR. SAKUNTHALA ENGINEERING COLLEGE, CHENNAI
National Conference on Recent Technologies for Sustainable Development 2015 [RECHZIG'15] - 28th August 2015

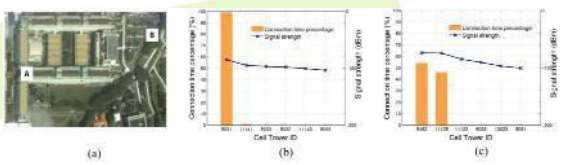
There are 3 major components.and the bus route on the 2D maps are not strictly necessary. In our system, instead of pursuing the accurate 2D physical locations, we logically map the bus routes to a space featured by sequences of nearby cellular towers. request to the backend server. The querying user indicates the interest bus route and bus stop to receive the predicted bus arrival time.

Sharing user. The sharing user on the other hand contributes the mobile phone sensing information to the system. After a sharing user gets on a bus, the data collection module starts to collect a sequence of nearby cell tower IDs. The collected data is transmitted to the server via cellular networks. Since the sharing user may travel with different means of transport, the mobile phone needs to first detect whether the current user is on a bus or not. As shown in Fig. 2 (left side), the mobile phone periodically samples the surrounding environment and extracts identifiable features of transit buses. Once the mobile phone confirms it is on the bus, it starts sampling the cell tower sequences and sends the sequences to the backend server. Ideally, the mobile phone of the sharing user automatically performs the data collection and transmission without the manual input from the sharing user.

Backend server. We shift most of the computation burden to the backend server where the uploaded information from sharing users is processed and the requests from querying users are addressed. Two stages are involved in this component. In order to bootstrap the system, we need to survey the corresponding bus routes in the offline pre-processing stage. We construct a basic database that associates particular bus routes to cell tower sequence signatures. Since we do not require the absolute physical location reference, we mainly war-drive the bus routes and record the sequences of observed cell tower IDs, which significantly reduces the initial construction overhead.The backend server processes the cell tower sequences from sharing users in the online processing stage. Receiving the uploaded information, the backend server first classifies the uploaded bus routes primarily with the reported cell tower sequence information. The bus arrival time on various bus stops is then derived based on the current bus route statuses.

3.2 Pre-Processing Cell Tower Data

The backend server needs to maintain a database that stores sequences of cell tower IDs that are experienced along different bus routes. Wardriving along one bus route, the mobile phone normally captures several cell tower signals at one time, and connects to the cell tower with the strongest signal strength. We find in our experiments that



and black in the figure. For example, the mobile phone initially connects to cell tower 5031 in

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
Available online at www.ijarbest.com

International Journal of Advanced Research in Biology Ecology Science and Technology (IJARBEST)
Vol. I, Special Issue I, August 2015 in association with VEL TECH HIGH TECH DR. RANGARAJAN DR. SAKUNTHALA ENGINEERING COLLEGE, CHENNAI
National Conference on Recent Technologies for Sustainable Development 2015 [RECHZIG'15] - 28th August 2015

Fig. 3. Cell tower connection time and received signal strength. (a) Cell tower coverage. (b) Connection at position A. (c) Connection at position B even if a passenger travels by the same place, the connected cell tower might be different from time to time due to varying cell tower signal strength. To improve the robustness of our system, instead of using the associated cell tower, we record a set of cell tower IDs that the mobile phone can detect. To validate such a point, we do an initial experimentWe measure the cell tower coverage at two positions A and B within the university campus, which are approximately 300 meters apart (Fig. 3(a) depicts the two positions on the map).

Fig. 3(b) and (c) report the cell tower that the mobile phone can detect, as well as their average signal strength and connection time at A and B, respectively. We find that position A and position B are both covered by 6 cell towers with divergent signal strength. In Fig. 3(b), we find that at position A the mobile phone is connected to the cell tower 5031 over 99% of the time, while its signal strength remains consistently the strongest during the 10-hour measurement. In Fig. 3(c), the mobile phone at position B observes two cell towers with comparable signal strength. We find that the mobile phone is more likely to connect to the cell tower with stronger signal strength, and also may connect to the cell tower with the second strongest signal strength. Therefore, in practice we choose the set of the top-*3* strongest cell towers as the signature for route segments.Fig. 4 illustrates the cell tower sequence collected on our campus bus traveling from our school to a rapid train station off the campus. The whole route of the bus is divided into several concatenated sub-route segments according to the change of the top-*3* cell tower set. They are marked alternately in red

Fig. 4. Cell tower sequence set along a bus route.

the first sub-route and the top-*3* cell tower set is {5031,5092, 11141}. Later the mobile phone is handed over to cell tower 5032 and the cell tower set becomes {5032, 5031, 5092} in the second sub-route. We subsequently record the top-*3* cell tower in each sub-route. Such a sequence of cell tower ID sets identifies a bus route in our database. By wardriving along different bus routes, we can easily construct a database of cell tower sequences associated to particular bus routes.

3.3 Bus Detection: Am I on a Bus?

During the on-line processing stage, we use the mobile phones of sharing passengers on the bus to record the cell tower sequences and transmit the data to the backend server. As aforementioned, the mobile phone should intelligently detect whether it is on a public transit bus or not and collect the data only when the mobile phone is on a bus. Some works study the problem of activity recognition and context awareness using

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
Available online at www.ijarbest.com

International Journal of Advanced Research in Biology Ecology Science and Technology (IJARBEST)
Vol. I, Special Issue I, August 2015 in association with VEL TECH HIGH TECH DR. RANGARAJAN DR. SAKUNTHALA ENGINEERING COLLEGE, CHENNAI
National Conference on Recent Technologies for Sustainable Development 2015 [RECHZIG'15] - 28th August 2015

various sensors. Such approaches, however, cannot be used to distinguish different transport modes public In this section, we explore multi-sensing resources to detect the bus environment and distinguish it from other transport modes. We seek a lightweight detection approach in terms of both energy consumption and computation complexity.

3.3.1 Audio Detection

Nowadays, IC cards are commonly used for paying transit fees in many areas (e.g., EZ-Link cards in Area [2], Octopus cards in Hong Kong [3], Oyster cards in Area [4], etc). On a public bus in Area, several card readers are deployed for collecting the fees (as depicted in Fig. 5(a)). When a passenger taps the transit card on the reader, the reader will send a short beep audio response to indicate the successful payment. In our system, we choose to let the mobile phone detect the beep audio response of the card reader since such distinct beeps are not widely used in other means of transportation such as non-public buses and taxis.

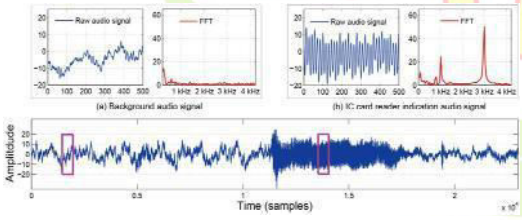


Fig. 6. Bus detection using audio indication signal

In order to exploit the unique beeps of IC card readers, in our initial experiment we record an audio clip on the bus at the audio sampling rate of 44.1kHz with Samsung Galaxy S2 i9100 phone.

Such a sampling rate is more than sufficient to capture the beep signals [22]. Fig. 6 (bottom) plots the raw audio signal in the time domain, where the IC card reader starts beeping approximately from 11000*th* sample and lasts to 18000*th* sample. We crop the section of the beep audio signal and depict the section in Fig. 6(b). After we convert the time domain signal to the frequency domain through 512pt Fast Fourier Transform (FFT) (Fig. 6(b)), we observe clear peaks at 1kHz and 3kHz frequency bands. For comparison we depict the audio clip as well where no beep signal is sent. Both time domain and the frequency domain signals are plotted in Fig. 6(a). We find no peaks at 1kHz and 3kHz frequency bands. With the knowledge of the frequency range of the dualtone beep signal sent by the IC card reader, in our system we can lower down the audio sampling rate of the mobile phone to 8kHz (8000 samples/s) which is sufficient to capture the beep signals with maximum frequency of 3kHz.We find that in practice 128pt FFT suffices to detect the IC card reader on the bus with tractable computation complexity on commodity mobile phones. We use the standard sliding window averaging technique with window size w = 32 samples to filter out the noises in both 1kHz and 3kHz frequency bands. We use an empirical threshold ε of three standard deviation (i.e., 99.7% confidence level of noise) to detect beep signals. If the received audio signal strengths in 1kHz and 3kHz frequency bands both exceed the threshold, the mobile phone confirms the detection of the bus. Fig. 7 depicts the beep signal detection process. When the IC card reader starts beeping, the signal strengths in both 1kHz

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
Available online at www.ijarbest.com

International Journal of Advanced Research in Biology Ecology Science and Technology (IJARBEST)
Vol. I, Special Issue I, August 2015 in association with VEL TECH HIGH TECH DR. RANGARAJAN DR. SAKUNTHALA ENGINEERING COLLEGE, CHENNAI
National Conference on Recent Technologies for Sustainable Development 2015 [RECHZIG'15] - 28th August 2015

and 3kHz frequency bands

jump significantly and therefore can be detected. The audio detection module is running all the time on mobile phones. We test the audio indication based bus detection method with various scenarios, and the experiments show encouraging results for bus detection (Section 4.2.1). As the dual-tone responsive signal is universally used in almost all public transit buses in Area, we can use it as an identifiable signature to distinguish the buses from other vehicles. Therefore, we use the dual-tone

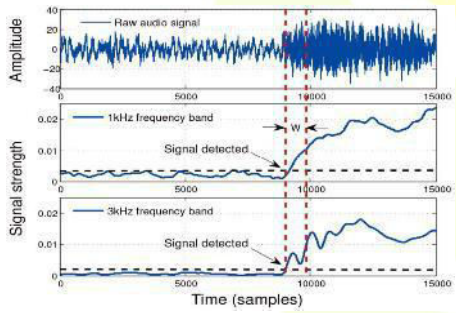


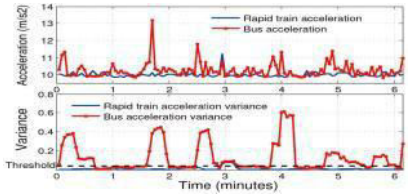
Fig. 7. Detecting audio beeps in the frequency domain.

as the acoustic trigger for the successive cell tower data collection and transmission of the mobile phones of sharing users. We can easily adopt similar techniques [19] to detect certain audio indications to identify the public transports as well in other areas (e.g., the bell ringing tunes in Hong Kong buses)

3.3.2 Accelerometer Detection: Bus vs. Rapid Train

For the audio detection technique, there may be false positives in our daily lives. Some similar beep signal may exist in other scenarios when users are tapping other types of cards like the cash card and employee's card. In some noisy

environments, the background sound or music may cause false positives. These kinds of false positives do not influence the system performance because the collected data can be filtered out at the backend server using bus classification algorithm which we will introduce later in Section 3.4. Besides such cases, the most possible false positives are from Rapid Train systems (MRT [4] in Area) because the IC card systems are also deployed in rapid train stations where the IC card readers in the entrances may send the same beep signal (Fig. 5(b)). Many other cities in the world have the similar situation as well. Solely relying on the audio detection the mobile phones may falsely trigger the cell tower ID collection when they go with the rapid trains. Since the train routes have substantial above-ground segments that overlap with bus routes, simply using cell tower signals does not effectively differentiate the two transit means. We expect to leverage the accelerometer sensor on the mobile phone to reduce such false detection. Intuitively, the rapid trains are moving at relatively stable speeds with few abrupt stops or sharp turns. On the contrary, the buses are typically moving with many sharp turns and frequent acceleration and deceleration. We collect the accelerometer data at a moderate sampling rate of 20Hz. The raw accelerometer readings are first made orientation independent by computing the *L*2-norm (or magnitude) of the raw data Fig. 8 (top) plots the accelerometer readings on a rapid train and a public transit bus which suggest that the accelerometer reading on the bus .We explore such acceleration features to distinguish the buses from the rapid trains. We measure the statistics of the accelerometer readings during 12.5 seconds (250 samples) to reduce the impact of noise, such as average and variance of the acceleration.



ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
Available online at www.ijarbest.com

International Journal of Advanced Research in Biology Ecology Science and Technology (IJARBEST)
Vol. I, Special Issue I, August 2015 in association with VEL TECH HIGH TECH DR. RANGARAJAN DR. SAKUNTHALA ENGINEERING COLLEGE, CHENNAI
National Conference on Recent Technologies for Sustainable Development 2015 [RECHZIG'15] - 28th August 2015

Fig. 8. Accelerometer readings on rapid train and busreadings on the rapid train and the public transit bus, respectively. According to the figure, the variance on the bus is significantly larger than that on the train. Therefore, we distinguish the buses from the trains using the variance of accelerometer readings by setting a proper threshold. We confirm the detection of buses if the measured acceleration variance is above the threshold, and the detection of rapid trains otherwise. We vary the threshold from 0.005 to 0.2 and calculate the detection accuracy. If the threshold is small, most buses will be correctly detected, while many trains will be misdetected as buses as well, which may lead to noisy inputs to the backend server and energy waste of mobile phones in collecting cell tower IDs. On the other hand, if threshold is too big, most rapid trains will be filtered out, while we will miss the detection of many actual buses, which may lose the opportunities in collecting useful cell tower information on the buses. We select an empirical threshold 0.03 to balance the falsenegative and false positive. In practice, we find that accelerometer based detection can distinguish the buses from the trains with an accuracy of approximately 90% (Section 4.2.2). The error rate of falsely detecting rapid trains as buses is even smaller. The detection error of falsely classifying public buses into rapid trains is mainly due to the abnormality of the bus routes (e.g., long straight routes) especially during non-peak hours. Such a detection error is tolerable in the bus classification
stage, where the backend server has information redundancy to handle the noisy reports.

3.4 Bus Classification

When a sharing user gets on the bus, the mobile phone samples a sequence of cell tower IDs and reports the information to the backend server. The backend server aggregates the inputs from massive mobile phones and classifies the inputs into different bus routes. The statuses of the bus routes are then updated accordingly.

3.4.1 Cell Tower Sequence Matching

We match the received cell tower sequences to those signature sequences store in the database. Fig. 9 shows an illustrative example where a sharing passenger gets on the bus at location A. The backend server will receive a cell tower sequence of _7, 8, 4, 5_ when the sharing user reaches location B. Say that the cell tower sequence of the bus route stored in the database is _1, 2, 4, 7, 8, 4, 5, 9, 6_, then the sequence _7, 8, 4, 5_ matches the particular bus route as a sub-segment as shown in Table 1.

In practical scenarios, the sequence matching problem becomes more complicated due to the varying cell tower signal strength. Recall that for each sub-route we record the top-*3* cell tower IDs instead of the connected cell tower ID in the pre-processing period. We let each mobile phone

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
Available online at www.ijarbest.com

International Journal of Advanced Research in Biology Ecology Science and Technology (IJARBEST)
Vol. I, Special Issue I, August 2015 in association with VEL TECH HIGH TECH DR. RANGARAJAN DR. SAKUNTHALA ENGINEERING COLLEGE, CHENNAI
National Conference on Recent Technologies for Sustainable Development 2015 [RECHZIG'15] - 28th August 2015

Fig. 9. Cell tower sequence matching

the top-*3* cell tower IDs instead of the connected cell tower ID in the pre-processing period. We let each mobile phone send back the sequence of cell towers that the mobile phonehas connected to. In the matching process on the server, we accordingly devise a top-*k* cell tower sequence matching scheme by modifying the Smith-Waterman algorithm. Smith-Waterman is a dynamic programming algorithm for performing local sequence alignment which has been widely used in bioscience (e.g., to determine similar regions between two nucleotide or protein sequences). We make concrete modifications on the original algorithm to support the top-*k* cell tower sequence matching. We weigh a matching of a cell tower ID with a top-*k* set according to the cell tower signal strength. Say that in a top-*k* set $S = \{c1, c2, \ldots, ck\}$ ordered by signal strength (i.e., $si \geq sj$, $1 \leq i \leq j \leq k$), where *ci* and *si* denote cell tower and its signal strength, respectively. We denote the uploaded cell tower sequence from a sharing user as *Seq*upload = _*u*1*u*2 . . . *um*_ where *m* is the sequence length. We also denote a cell tower set sequence in database as *Seq*database = _*S*1*S*2 . . . *Sn*_ where *n* is the set sequence length. If *ui* = *cw* ∈ *Sj*, *ui* and *Sj* are considered matching with each other, and mismatching otherwise. We assign a score *f (sw)* for a match, where *f (sw)* is a positive non-decreasing scoring function and *w* is the rank of signal strength. In practice, we use *f (sw)* = 0.5*w*−1 as the scoring function according to the signal strength order in the set. The penalty cost for mismatches is set to be an empiricalvalue of −0.5 which balances the robustness and accuracy in practice. We choose top-*3* cell tower IDs with strongest cell tower signal strength to form a set based on our initial observations (Section 3.2). The distinctive advantage of the proposed classification algorithm is its robustness to the variation of cell tower signal strength. Table 2 shows a cell tower set sequence matching instance. In the example,the uploaded cell tower sequence is *Seq*upload = _1, 8, 10, 15, 16_, and the cell tower ID set is shown in the first three rowscell tower signal strength sorted in decreasing order of the associated

TABLE 1
Cell Tower Sequence Matching

Database seq.	124	7845	96
Uploaded seq.		7845	
Matched seq.		7845	

misclassification rate of such short sequence is high and thus the backend server postpones the classification and the updating process until the sequence excesses the empirical threshold (which will be elaborated later).One problem of the cell tower sequence matching is that some bus routes may overlap with each other. The mobile phones on the overlapped road segments are likely to observe similar cell tower sequences. Since many busestypically arrive at and depart from several major transit centers, such overlapping road segments among different bus routes are common.

We survey 50 bus routes in Area and measure their overlapped road segments using Google Maps. Fig. 10(a) plots the distribution of the lengths of overlapped road

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
Available online at www.ijarbest.com

International Journal of Advanced Research in Biology Ecology Science and Technology (IJARBEST)
Vol. I, Special Issue I, August 2015 in association with VEL TECH HIGH TECH DR. RANGARAJAN DR. SAKUNTHALA ENGINEERING COLLEGE, CHENNAI
National Conference on Recent Technologies for Sustainable Development 2015 [RECHZIG'15] - 28th August 2015

segments, which suggests that over 90% of the overlapped route segments are shorter than 1400 meters, and over 80% are less than 1000 meters. Considering that the coverage range of each cell tower in urban area is about 300-900 meters, we set the empirical threshold of cell tower sequence length to 7.

Fig. 10(b) plots the cell tower sequence matching accuracy in classifying the bus routes. We vary the length of uploaded cell tower sequence from 2 to 9. We find that the matching accuracy is low when the cell tower sequence length is small (e.g, <4) largely because of the problem of route overlap. We observe that when the cell tower sequence length reaches 6, the accuracy increases substantially to around 90%.

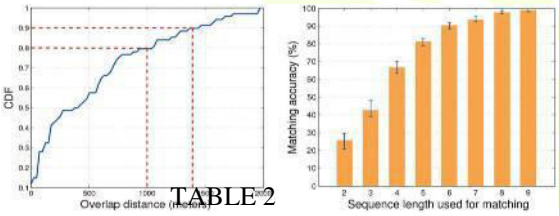


TABLE 2
Top-3 Set Sequence Matching

Database cell tower set seq.	19	**1**	4	7	**10**	13	**16**	22	∑
	20	2	5	**8**	11	14	17	23	
	21	3	6	9	12	**15**	18	24	
Uploaded seq.		1	–	8	10	15	16		
Score	0	+1	-0.5	+0.5	+1	+0.25	+1	0	3.25

After running the sequence matching algorithm across all bus route sequences in the database, the backend server selects the bus route with the highest score. If the highest matching score is smaller or the sequence length is shorter than our empirical thresholds, the backend server postpones the updates to avoid errors. Intuitively, the small highest matching score would be due to mistriggering of sharing phones uploading cell tower sequence not from interested bus routes (e.g., rapid trains, private cars, etc). Too short cell tower sequence may not be informative since the

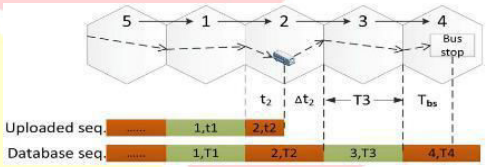


Fig. 10. Overlaped routes and matching accuracy with varying sequence length. (a) CDF of the overlapped route length. (b) Matching accuracy with varying sequence length

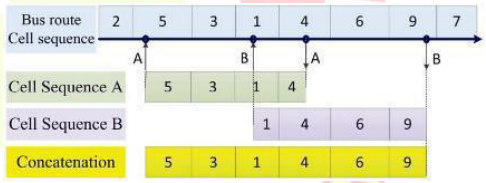


Fig. 11. Cell tower sequence concatenation

When the cell tower sequence length is larger than 8, the experimental results are reasonably accurate and robust.

3.4.2 Cell Tower Sequence Concatenation:

Solving Jigsaw Puzzles

In many practical scenarios, the length of the cell tower sequence obtained by a single sharing user, however, may be insufficient for accurate bus route classification. An intuitive idea is that we can concatenate several cell tower sequences of different sharing users on the same bus to form a longer cell tower

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
Available online at www.ijarbest.com

International Journal of Advanced Research in Biology Ecology Science and Technology (IJARBEST)
Vol. I, Special Issue I, August 2015 in association with VEL TECH HIGH TECH DR. RANGARAJAN DR. SAKUNTHALA ENGINEERING COLLEGE, CHENNAI
National Conference on Recent Technologies for Sustainable Development 2015 [RECHZIG'15] - 28th August 2015

sequence. In Fig. 11, both cell tower sequences of sharing user A and B are short, while by concatenating the two cell tower sequences the backend server may obtain an adequately long cell tower sequence which can be used for more accurate bus classification. A simple way of concatenating the cell tower sequences is to let the mobile phones of sharing passengers locally communicate with each other (e.g., over Bluetooth) This approach, however, mandates location exposure among sharing passengers and might raise privacy concerns. 1kHz frequency band around 0.8s, the signal strengths in 3kHz frequency band remain low).We thereby shift such a job to the backend server. Recall that the mobile phone needs to collect audio signals for bus detection (Section 3.3.1). Here, we reuse such information to detect whether the sharing passengers are on the same bus for cell tower sequence concatenation. Ateach bus stop, normally several passengers enter a bus and multiple beeps of the IC card readers can be detected. The time intervals between the consecutive beep signals fingerprint each bus in the time domain. Fig. 12 depicts an instance of the audio signals captured by three different mobile phones on the same bus. We depict the raw audio

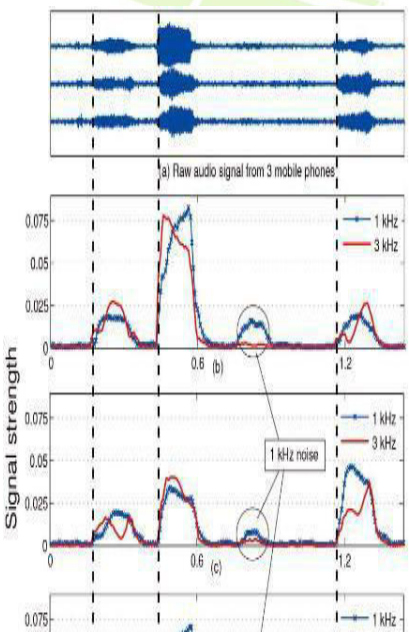


Fig. 12. Time intervals of audio indication signals.

signals in Fig. 12(a), and corresponding frequency domain signals in Fig. 12(b)–(d). Compared with the time domain signal, the frequency domain signal is robust against the background noise (e.g.,

We can see that in the frequency domain the signals are highly crosscorrelated and thus can be used to determine whether the phones are on the same bus. Specifically, the time intervals observed by three mobile phones are all approximately $dT1$ and $dT2$ in Fig. 12.We therefore use the time intervals between the detected beeps to determine whether multiple mobile phones are on the same bus. In our system, the mobile phones of sharing users keep sampling the audio signal and record the time intervals between the detected beeps. Such beep interval

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
Available online at www.ijarbest.com

International Journal of Advanced Research in Biology Ecology Science and Technology (IJARBEST)
Vol. I, Special Issue I, August 2015 in association with VEL TECH HIGH TECH DR. RANGARAJAN DR. SAKUNTHALA ENGINEERING COLLEGE, CHENNAI
National Conference on Recent Technologies for Sustainable Development 2015 [RECHZIG'15] - 28th August 2015

information is reported along with the cell tower sequences to the backend server. Receiving the uploaded sensing data from sharing passengers, the backend server detects and groups the sharing passengers on the same bus by comparing both cell tower sequences and the signals. The backend server concatenates the pieces of cell tower sequences from the same bus and forms a longercell tower sequence.

3.5 Arrival Time Prediction

After the cell tower sequence matching, the backend server classifies the uploaded information according to different bus routes. When receiving the request from querying users the backend server looks up the latest bus route status, and calculates the arrival time at the particular bus stop.

Fig. 13 illustrates the calculation of bus arrival time prediction. The server needs to estimate the time for the bus to travel from its current location to the queried bus stop. Suppose that the sharing user on the bus is in the coverageof cell tower 2, the backend server estimates its arrivaltime at the bus stop according to both historical data as well as the latest bus route status. The server first computes the dwelling time of the bus at the current cell (i.e., cell 2 in this example) denoted as $t2$. The server also computes the traveling time of the bus in the cell that the bus stop is located denoted as tbs. The historical dwellingtime of the bus at cell 3 is denoted as $T3$. The arrival time of the bus at the queried stop is then estimated as follows,

$$T = T2 - t2 + T3 + tbs.$$

Without loss of generality, we denote the dwelling time in cell i as Ti, $1 \leq i \leq n$, the bus's current cell number as k, and the queried bus stop's cell number

as q. The server can estimate the arrival time of the bus as follows,

$$T = q-1\_i=k\ Ti - tk + tq.$$

The server periodically updates the prediction time according to the latest route report from the sharing usersand responds to querying users. The querying users may indicate desired updating rates and the numbers of successive bus runs to receive the timely update.

4.5 System Overhead

Mobile phone. In order to maintain the sample resolution and remove the noise, we extract the audio signal with sliding widows with the window size of 32. We record the audio signal at the sampling rate of 8kHz, and use n = 128pt FFT to convert the time domain audio signals to frequency domain signals. The major computational complexity is attributed to performing FFT on mobile phones which is $O(n \log n)$. Current mobile phones can finish the computation task in realtime. For example, it takes approximately 1.25ms and 1.8ms on average to finish to 128pt FFT on Samsung Galaxy S2 i9100 and HTC Desire, respectively. We measure the power consumption of continuously sampling microphone accelerometer, GPS, and cellular

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
Available online at www.ijarbest.com

International Journal of Advanced Research in Biology Ecology Science and Technology (IJARBEST)
Vol. I, Special Issue I, August 2015 in association with VEL TECH HIGH TECH DR. RANGARAJAN DR. SAKUNTHALA ENGINEERING COLLEGE, CHENNAI
National Conference on Recent Technologies for Sustainable Development 2015 [RECHZIG'15] - 28th August 2015

signals. Table 5 illustrates the measured battery lifetime when the mobile phones continuously trigger different sensors.The experiments were performed with the screen set to minimum brightness. We report the average results over 10 independent measurements.The battery duration was quite similar for sampling accelerometer at 20Hz, sampling audio signal at 8kHz with 128pt FFT, and sampling no sensors.Sampling the cell tower signal consumes limited extra battery power as well. On the other hand the battery life time is substantially reduced when the GPS module in the phone is enabled.

Backend server. Since our implementation is in a particular area of Area, we do not have the experience when the system scales to the entire city. We make mathematical analysis to forecast the computation capacity needed when the system scales. The computation overhead of backend server is mainly bounded by the bus classification algorithm,i.e., the uploaded cell tower sequence length l, the cell tower set sequence length k, and the number of cell tower set sequences in the database N. The computation

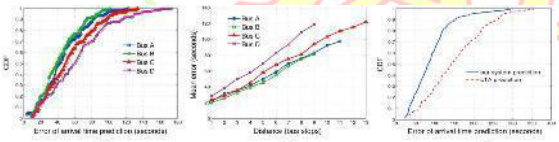


Fig. 14. Arrival time prediction performance. (a) Bus arrival time prediction error. (b) Bus arrival time prediction. (c) Our system v.s. LTA

complexity of sequence matching using dynamic programming is $O(lk)$, and as we need to compare with N candidate sequences in database the overall computation complexity is $O(lkN)$. Since in practice both m and n are usually small (e.g., $\max\{l,k\}$ is around 40 according to our experiments),the computation complexity increases almost linearly to the number of candidate cell tower sequences in the database.

6 RELATED WORK

Phone-based transit tracking. Our work is mostly related to recent works on the transit tracking systems.EasyTracker presents an automatic system for lowcost, real-time transit tracking, mapping and arrival time prediction using GPS traces collected by in-vehicle smartphones.It present a grassroots solution for transit tracking utilizing accelerometer and GPS modules on participating mobile phones. EEMSS [27] presents a sensor management framework which uses minimum number of sensors on mobile devices to monitor user states. VTrack [26] estimates road travel time based on a sequence of WiFi-based positioning samples using an HMM-based algorithm for map matching. CTrack [25] presents trajectory mapping using cell tower fingerprints and utilizes various sensors on mobile phones to improve the mapping accuracy. Our work differs from them in that it predicts the bus arrival time based on cell tower sequence information shared by participatory users. To encourage moreparticipants, no explicit location services (e.g., GPS-basedlocalization) are invoked so as to reduce the overhead of using such special hardware for localization.

Cell tower sequence matching. StarTrack [9] provides a comprehensive set of APIs for mobile application development. Applying new data structures, enhances StarTrack in efficiency,robustness, scalability, and ease

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
Available online at www.ijarbest.com

International Journal of Advanced Research in Biology Ecology Science and Technology (IJARBEST)
Vol. I, Special Issue I, August 2015 in association with VEL TECH HIGH TECH DR. RANGARAJAN DR. SAKUNTHALA ENGINEERING COLLEGE, CHENNAI
National Conference on Recent Technologies for Sustainable Development 2015 [RECHZIG'15] - 28th August 2015

of use. CAPS determines a highly mobile user's position using a cell-ID sequences matching technique which reduces GPS usages and saves energy on mobile phones.Unlike those proposals, our work does not aim to position the mobile users though similar in spirit to these existing works in utilizing the cell tower sequences.

Participatory sensing. Many recent works develop participatory platforms for people-centric mobile computing applications [8]. MoVi studies the problem of social activity coverage where participants collaboratively sense ambience and capture social moments through mobile phones. Escort obtains cues from social encounters and leverages an audio beacon infrastructure to guide a user to a desired person. WILL designs an indoor logical localization technique leveraging user mobility and WiFi infrastructure while avoiding site survey. Although targeted at totally different applications and problems, the common rationale behind these works and our design is that the absolute physical locations of users though sometimes sufficient are not always necessary to accomplish particular tasks.

CONCLUSION AND FUTURE WORK

In this paper, we present a crowd-participated bus arrival time prediction system. Primarily relying on inexpensive and widely available cellular signals, the proposed system provides cost-efficient solutions to the problem. We comprehensively evaluate the system through an Android prototype system.Our system can accurately predict the bus arrival time. Being independent of any support from transit agencies and location services, the proposed scheme provides a flexible framework for participatory contribution of the community. For a particular city, the only requirement of our system implementation is that there exist a backend server and an IC card based bus system.Future work includes how to encourage more participants to bootstrap the system because the number of sharing passengers affects the prediction accuracy in our system. This common issue of crowd-sourced solutions is largely influenced by the penetration rate and popularity of the services. One may actively promote the service to reach a critical penetration rate so as to ensure that at least one sharing user is on the bus willing to report the bus status .At the initial stage, we may encourage some specific passengers(like the bus drivers) to install the mobile phone clients.

REFERENCES

1. *Bus Transport in Area* [Online]. Available: http://en.wikipedia.org/wiki/Bus_transport_in_ Area *EZ-Link* [Online]. Available: http://www.ezlink.com.sg
2. *Octupus* [Online]. Available: http://www.octopus.com.hk/home/e
3. *Oyster* [Online]. Available: *h*ttps://oyster.tfl.gov.uk/tr
4. *PublicTransport@SG* [Online]. Available: http://www.publictransport.sg/
5. *Buses in Area* [Online]. Available: http://en.wikipedia.org/wiki/Area_bus
6. *Transport for Area* [Online]. Available: http://www.tfl.gov.uk/
7. T. Abdelzaher *et al.*, "Mobiscopes for human

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
Available online at www.ijarbest.com

International Journal of Advanced Research in Biology Ecology Science and Technology (IJARBEST)
Vol. I, Special Issue I, August 2015 in association with VEL TECH HIGH TECH DR. RANGARAJAN DR. SAKUNTHALA ENGINEERING COLLEGE, CHENNAI
National Conference on Recent Technologies for Sustainable Development 2015 [RECHZIG'15] - 28th August 2015

spaces," *IEEePervasive Comput.*, vol. 6, no. 2, pp. 20–29, Apr. 2007.

8. G. Ananthanarayanan, M. Haridasan, I.
9. Mohomed, D. Terry, andC. A. Thekkath, "Startrack: A framework forenabling track-based applications," in *Proc. ACM MobiSys*, 2009, pp.207–220.
10. [10] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Proc. IEEE INFOCOM*, 2000, pp. 775–784.