

Experiences with a Transportation Information System that Uses Only GPS and SMS

Ruth E. Anderson, Waylon Brunette, Erica Johnson, Caitlin Lustig, Anthony Poon, Cynthia Putnam, Odina Salihbaeva, Beth E. Kolko, Gaetano Borriello

Abstract—We present our experiences with an SMS-based system for providing transit information based solely on existing cellular and GPS networks. The aim is to permit the development of information services that do not rely on a central authority or complex web hosting. We developed and applied our system to the network of privately-run marshrutka buses in Bishkek, Kyrgyzstan. However, our goal is to more broadly address issues of ad-hoc shared transportation systems in the developing world. A custom designed GPS-GSM unit is placed on a vehicle, and users can query our server over SMS with their own non-GPS-enabled cell phones. We report on the accuracy of our location naming approach and estimates of bus arrival times. In addition, we summarize interviews with bus drivers and bus riders relating their views of the system and outline directions for future work. Our system is a grassroots solution to the persistent lack of transport information in developing countries.

Index Terms— ICTD, transportation, information services, mobile phones, GPS, SMS, Kyrgyzstan, developing regions

I. INTRODUCTION

FINDING the right transportation at the right time is a universal problem. Transportation of goods and people is central to economic and human development and the availability of safe and reliable transportation can have broad implications. Not only does access to transport improve access to markets and mobility of workers, it is also critical to

Manuscript received July 30, 2010. This work was supported in part by U.S. National Science Foundation grants #0326101 and #0219350 and by the J. D. Noe Professorship at UW.

R. E. Anderson, Dep't of Computer Science & Engineering, University of Washington, Seattle, WA 98195 USA (rea@cs.washington.edu).

W. Brunette, Dep't of Computer Science & Engineering, University of Washington, Seattle, WA 98195 USA (wrub@cs.washington.edu).

E. Johnson, Global Studies, University of North Carolina, Chapel Hill, NC 27599 USA (ejj3@uw.edu).

C. Lustig, School of Information and Computer Sciences, University of California, Irvine, CA 92697 USA (celustig@ics.uci.edu).

A. Poon, Dep't of Computer Science & Engineering, University of Washington, Seattle, WA 98195 USA (anthop@cs.washington.edu).

C. Putnam, College of Computing and Digital Media, DePaul University, Chicago, IL 60604 USA (cputnam@cdm.depaul.edu).

O. Salihbaeva, Russian, East European and Central Asian Studies, Jackson School of International Studies, University of Washington, Seattle, WA 98195 USA (odinas@uw.edu).

B. E. Kolko, Dep't of Human Centered Design & Engineering, University of Washington, Seattle, WA 98195 USA (bkolko@u.washington.edu).

G. Borriello, Dep't of Computer Science & Engineering, University of Washington, Seattle, WA 98195 USA (gaetano@cs.washington.edu).

the timely and affordable delivery of services such as health care and education, can empower vulnerable groups by increasing their independence, and is key to maintaining social networks [1].

In addition to access, potential transportation users also need information about its availability. In the developed world, users can often access information about bus and train schedules easily via printed schedules or web pages maintained by centrally-funded transportation authorities. In many cases, users can view real time updates on the current location and expected arrival time of their bus or train via web, phone, or SMS/text message [2]-[5]. These solutions rely upon central infrastructure to provide the servers needed to collect data and answer queries. When a community lacks the infrastructure to provide such information resources, however, potential users can find transportation resources difficult to use, inefficient, and potentially unsafe. Consequently, we chose to investigate a design solution that would explore how current technology could be leveraged for small-scale, institutionally-independent transportation information systems.

This paper discusses our experiences developing and evaluating the *bus system (pronounced "star-bus"), a transportation information system that uses GPS and SMS technologies. This paper expands on the work reported previously in [6] by describing system modifications and results collected during an initial deployment in Bishkek, Kyrgyzstan in March 2009. While this project did not include a full-scale deployment, the design approach, technical troubleshooting, and initial usability tests provide a framework for how to create technological solutions that can be introduced on a grassroots level. The findings in this paper are extensible to other developing regions with limited resources, fragile road infrastructure, resource-constrained central governments, and ad-hoc transportation resources for both inter- and intra-city transport. We envision the technical components of this solution having broad potential relevance for shared transportation systems throughout the developing world, including car shares for long distance travel and hitchhiking.

II. CONTEXT AND FIELD WORK

The *bus system was developed in response to transportation challenges experienced in Kyrgyzstan, a

developing region with poor infrastructure and limited resources. Kyrgyzstan is a small, mountainous country in Central Asia with a population of about 5.5 million. Since independence from the Soviet Union in 1991, Kyrgyzstan has struggled economically with about 40% of the population living below the poverty line [7].

In this section, we describe background on the transportation problem and technology use in Kyrgyzstan generally. (More information on both can be found in [6].) In addition, we report on the results of field work done in Bishkek, Kyrgyzstan in March 2009, where we interviewed riders and drivers of the marshrutka system about their use of transportation and technology.

A. Transportation in Kyrgyzstan

In Kyrgyzstan, as in many developing countries, shared transportation often happens outside the realm of a national or metropolitan transit authority. As a result, there is no central body coordinating or allocating resources to ensure that the system is efficient and meets citizens' needs. Our work considered how, under such circumstances, one might aggregate and disseminate information about transportation options to make it more efficient and effective for both riders and drivers. Another way of stating the problem is how to build an information infrastructure for a shared public resource when a public government (or other centralized authority) is not positioned to do the job – either because of lack of resources, lack of will, or because the resource itself is decentralized and largely part of the informal sector.

Personal vehicle ownership rates are low in Kyrgyzstan, and the population depends heavily on shared and public transportation for both inter- and intra-city transportation. Cities have taxis and privately operated minibuses collectively known as the marshrutka system. The marshrutka system is a shared transportation system that runs on pre-arranged routes but without standard time schedules; the typical marshrutka (see Figure 1) seats about 15 and with standing passengers can fit 25 or more. Fares are 8 som (about US\$0.19) before 8pm and 10 som (about US\$0.24) after 8pm for central city routes, which is considered very affordable for the local population. For comparison, a taxi ride along the same route of a marshrutka ranges from 75-125 som (about US\$1.80 - \$3.00), an amount that represents 1-3% of the average Bishkek citizen's monthly income ($x = \$100-200$ USD, $N=118$). The marshrutka drivers are independent contractors of private companies as opposed to employees of a municipal or national transportation authority. Bus routes are part of local knowledge, and learning a new route is challenging even for a city resident. Numbered routes are marked by placards placed in the front windshield of the bus, with a list of major stops or landmarks indicating the rough routing. As buses whiz by, people try to glimpse the list of stops written on the placard to see if the bus is going in their direction, and quickly identifying an appropriate marshrutka can be difficult. There is a lack of formal bus stops, and people stand at areas known to be marshrutka stops and they can also hail a minibus from

other points on the street. There are no posted time schedules, so riders must wait indeterminate amounts of time for a given bus to pass by the stop.

The city government licenses independent companies to operate along specific bus routes. The private companies then contract drivers to provide service on a set of routes with a frequency they deem to be profitable. Because often only one company operates along each route, there generally are not alternative buses serving a particular location. This can be especially problematic at off-peak times when marshrutkas are less frequent. Women face extra challenges, as waiting for a long time in the evening can be dangerous.



Fig. 1. Marshrutka in Bishkek.

B. Mobile Phone Use in Kyrgyzstan

The rate of personal ICT usage in Kyrgyzstan is similar to many developing regions; that is, computer and Internet usage remains relatively low and has seen slow growth over the past 3 years (from 11% in 2006 to 15% in 2008), while mobile phone use is growing at a remarkable rate including in rural areas (from 21% in 2006 to 71% in 2008). People in Kyrgyzstan use SMS regularly (about one-third of mobile use is for SMS), and they are not using the Internet as their primary mode of ICT access [6]. Even when people do identify themselves as Internet users, their model of usage is very different from that of the developed world, with most users accessing the Internet once a week or less - this is likely due to cost factors as well as lack of locally relevant content. Meanwhile, mobile users use their phones much more frequently, with 89% reporting using them at least once a day. Survey data also shows that mobile users are becoming habituated to using their phones to seek information (21.4% in Kyrgyzstan ($N=716$) and 54% in Bishkek ($N=156$)) and send SMS (28.4% in Kyrgyzstan and 36.3% in Bishkek), while Internet use on phones is barely perceptible (0.9% in Kyrgyzstan and 2.4% in Bishkek). These usage patterns, in addition to our other research in the region [8], [9], guided our development process and drove our emphasis on SMS-based

information solutions as the current appropriate technology for our solution.

Elsewhere in the ICTD community, SMS-based solutions have also proven robust, flexible, and valuable in multiple contexts [10]-[12]. The success of platforms such as FrontlineSMS [13] and RapidSMS [14] demonstrate that information delivered via SMS can have a tangible, positive impact on individuals' lives.

C. Interviews with Bus Riders

In March 2009, we conducted 27 interviews with marshrutka riders in several areas of Bishkek. These interviews were designed to provide more specific feedback on marshrutka and SMS use in order to give us insight about the suitability of our proposed solution. We selected three sampling spots where we expected the socioeconomic status of riders would vary: a location in the city center, a middle-class residential location on the outskirts of town (12th Micro-region), and the popular Dordoi bazaar that draws affluent and poorer residents from both Bishkek and outlying suburban areas. We conducted interviews in each location at different times of the day to capture the attitudes of riders on commutes to work, from work, and during mid-day rides. In each location, we approached the passengers standing at the bus stop. We also interviewed passersby and vendors in the local shops who said they were habitual marshrutka riders. Interviews were conducted in either Russian or Kyrgyz and resulting notes were translated into English.

Fourteen women and thirteen men were interviewed on the street. Respondent ages ranged from 18-65 with an average of 33 years. Almost all (22 out of 27) had frequency rates of riding marshrutkas in the double digits each week. Rides on the minibuses ranged from 2-3 per month to 84 per week. The average weekly number of marshrutka trips was 20, with most people riding everyday, either 2 or 4 trips per day, depending on whether there was a direct route to their destination. Most respondents used marshrutkas to get to work or school; also cited were shopping trips (e.g. to bazaars on the outskirts of town) and to visit family. Six of our respondents had cars in the household, including four who had their own car. This demonstrates that the marshrutka system is not just used by those who have no other means of transport; it is more broadly a necessary piece of the transportation puzzle.

Two thirds (18 of 27) of our respondents specifically mentioned the expense of taxis. Even though some of our respondents did use taxis (especially at night, to address safety concerns), most of them said they found the fares very expensive. City buses, which are even cheaper than marshrutkas (6 som vs. 8 som) were reported to be very slow and inconvenient. These were used by respondents when they were low on money, or when they felt they had excess time and could be patient.

All but three of our respondents reported using SMS, although five of those who used SMS said they prefer to use voice and that they tended to mostly receive SMS messages. Ten respondents said they find SMS convenient, especially

when they are in meetings or need to leave a message. (Most mobile accounts in Kyrgyzstan do not support voice mail.) Eight respondents preferred SMS because it was cheaper than voice calls. Overall, these findings demonstrate that SMS is a viable mode of information dissemination, and that even those who find it inconvenient or confusing do still tend to use it for reading, if not writing. We believe that current patterns of SMS usage in Bishkek are conducive to adoption of an SMS-based transportation information system. To make writing SMS queries simple, habitual queries can be stored as template text messages, making it easy for users to send structured SMS queries.

D. Interviews with Bus Drivers

We also conducted 6 interviews with marshrutka drivers, representing 3 private marshrutka companies. The interviews were intended to inform our understanding of the marshrutka system (some of which is reported in Section II.A) and thus how our system might fit economically and logistically within that system. We also asked drivers about their mobile phone and SMS usage patterns in an attempt to determine if we might leverage that usage for future system enhancements (e.g., for reporting of people waiting at a stop, how full the next bus is). Interviews lasted approximately 15 minutes and were conducted at Dordoi bazaar and at the terminal point of the outlying residential area. Drivers were recruited by approaching them and explaining our purpose.

In general, all the drivers we interviewed use mobile phones regularly in their work and for personal use. While most of the drivers primarily use their mobile phones for voice calls, all but one also regularly used SMS for work or personal communication. The five drivers that use SMS all commented that it is cheaper than voice calls.

The marshrutka drivers reported that they communicate with each other by phone when they notice abnormalities along their routes (e.g., speed traps, roadblocks, construction) and to get updates about waiting passengers. Additionally, drivers commented that their phones are especially important for work when they use their vehicles for inter-city travel or private hire. In those cases, the mobile phone is an indispensable tool for communicating with clients.

E. Designing an Appropriate Solution

Based on these technology usage patterns, we designed a system to improve access to transit information for potential bus riders at a minimum cost to users. We chose to utilize existing technologies appropriate for Bishkek (SMS and GPS) for our solution. We postulate that a transit information system can leverage existing technology to automate information gathering, allowing for easier and more accurate information sharing between drivers and riders without the assistance of a central authority.

III. THE *BUS SYSTEM

Our solution, called *bus (star-bus), takes into consideration

technology usage patterns and technology infrastructure available in resource-constrained environments. An architectural overview of the solution is shown below in Figure 2. The *bus system requires each bus driver in the transportation network to have a *box (star-box) device installed on their bus. Additionally, at least one server (in our case a laptop in a hotel room), connected to a cell phone, needs to be running to receive location update messages from *boxes and to accept and respond to queries from bus riders.

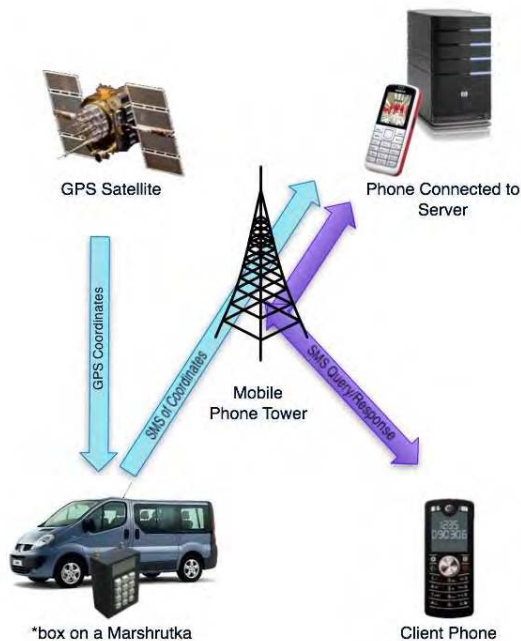


Fig. 2. Overview of the *bus system.

A. The *box

The primary purpose of a *box, is to communicate the marshrutka's location and other status. We built three prototype devices at a unit cost of about US\$200, although we believe mass production would bring the cost per box closer to US\$50. We chose to build our own prototype device to allow flexibility during our trial deployments. Another reason we chose to build a new device rather than use an existing GPS-enabled phone was to minimize theft risk, as the *box has a street value close to zero unlike a new high-end phone. The *box uses an embedded module that provides a GPS receiver, an SMS module into which a SIM can be inserted, and general-purpose I/O to control peripheral devices such as the keypad and LCD [6]. In the future, as GPS-enabled phones become less expensive, the functionality of a *box could potentially be replaced by an application running on a driver's own GPS-enabled cell phone.

The *box was designed to require minimal interaction from the bus driver. The driver prepares a *box for operation by entering a route number using the keypad, which is then displayed on an LCD. Once a driver is ready to begin his route, he simply flips the active switch to 'on' causing the

*box to start sending location update messages that include the route number and GPS coordinates to the *bus server. If the driver wants to take a lunch break or wants the system to quit tracking his vehicle's location for privacy reasons, he simply turns the active switch to 'off' causing the *box to stop sending update messages. Using a route number is a good fit for the marshrutka system in Bishkek where buses have set routes indicated by numbers displayed clearly and in big print in bus windows. We decided to require drivers to specify their route via the keypad interface rather than forcing the server to do the more difficult job of determining it automatically based on the path of the bus, although this is also a future possibility. One could imagine a simpler device including a hard-coded route number, although we chose the keypad and LCD interface for our prototype to allow for ease of testing on different routes.

To further minimize driver interaction and to improve maintainability of the system, an interface for the *box was created that allows for remote control of key parameters via SMS messages sent to the box. Available parameters include: the rate at which the *box sends location updates, the server phone number or email address that location updates are sent to, and overriding of any setting that a driver has entered using the keypad. Additionally, the remote interface allows a user to request a *box location update message to make sure everything is working properly.

The test deployment revealed the costs of transmitting location updates via SMS to be a major system constraint. A potentially useful solution is the use of USSD messages instead of SMS [15]. USSD are network maintenance messages that carriers use and require a port to be opened on the cellular network. This does require the cooperation of the local carrier; however, increased SMS traffic from users may be an attractive reason to provide free messages to drivers for location updates. Advertising provides another attractive revenue source because when the precise location of users is known highly targeted advertising messages can be sent (e.g., 10% discount at the store at the corner).

Another approach would be to redesign the *bus system into a more distributed design to lower transmission costs. For example, a poll model could be used to lower costs by having the *bus server request an update from the *box only when somebody wants a prediction thereby eliminating unused location update messages. Yet another solution could be to have the *box report if the marshrutka is deviating arrival time from a normal predictive model, allowing the *bus to only message the server when it starts and ends a route or if the marshrutka is running early or late so the model on the server can adjust its arrival estimations. Smartphone platforms as well as the embedded modules we used easily provide the computational resources required for these approaches.

B. The *bus Server

The primary purpose of the *bus server is to accept and process SMS messages from *boxes and queries from bus riders. Since the system is SMS-based, the server does not

need to be connected to the Internet, facilitating its deployment in areas with low Internet connectivity. System requirements are intentionally minimal: a laptop or desktop computer running a Java virtual machine and a MySQL database, connected to a phone capable of sending and receiving SMS messages. For our deployments we used a Dell Inspiron 1420 Intel Core 2 Duo @ 2GHz, connected to a Sony Ericsson Walkman Phone (W580i). Lower system requirements allow the server to run using older and less expensive hardware (at a cost as low as \$200-300). Our prototype server was implemented using MySMS [16], an SMS application framework built on top of SMSLib [17]. Though we did our initial testing in English, the server has been designed to accept queries and send messages in alternate character sets – in this case, in Cyrillic to support both the Russian and Kyrgyz languages.

Currently, the server is responsible for handling the following types of requests, which arrive via SMS messages.

1) **Location Update Messages** sent from the *boxes to the server have the following format:

```
!LOC,<GPS-obtained-time>,<GPS-obtained-
location>,<bus-id>,<route-#>
```

Location update messages are periodically sent (typically every 30 seconds) to the server by the *boxes. These messages update the server on the GPS coordinates of a bus traveling a certain route at the given time. This information is recorded in a database for use in responding to user queries. The bus-id field is the unique identifier assigned to each individual bus equipped with a *box. We would expect that bus drivers would display this bus id in addition to their route number in a way that it is accessible to riders. In Kyrgyzstan, one example of a bus-id is the vehicle's license plate number, which is posted in large letters on the back of the marshrutka as seen in Figure 3. Riders need the bus-id in order to assign a name to a location as described later.



Fig. 3. Marshrutkas in Bishkek displaying bus-id.

2) **Queries from Users about Bus Arrival Time** are continuously received and replied to via SMS using the

information gathered about the current location of buses in the system. Users are not required to have GPS-enabled phones and only need to be on a network that supports sending SMS messages. Potential bus riders simply text to a *bus server telephone number, and the server responds by sending an SMS back to the user's phone number. Although many types of queries could be supported, the most basic example is to ask the server to predict the arrival time of the next bus of a given route number to a given location:

```
<route-#> TO <location>
```

The server immediately replies to the user with an SMS predicting the arrival time of the next bus to that location:

```
Next <route-#> arrives to <location>
at about <estimated-arrival-time>.
```

Location names are either provided by the system or geocoded by users.

A user may also wish to be notified *automatically* if a bus of the given route is near. In these cases, a user may send a query asking to be updated when the bus is close to arriving:

```
<route-#> TO <location> UPDATE
```

The server replies to the user with at least one prediction. The first response arrives immediately, while up to two others may arrive when the estimated arrival time is ten and five minutes away.

3) **User Geo-coding Requests** are used to tag buses when they are at a particular location with a user-personalized name for that location. This design allows users to leverage the GPS unit on the bus rather than having to have a GPS-enabled phone themselves. These named locations can then be used to query about the arrival time of buses to that location. This facility is critical for a system such as the marshrutka system, where there are few official stops. When a user sees a bus at a location of interest, the user sends a message containing 1) the bus-id, and 2) a personal location name to the server as follows:

```
STORE <bus-id> AS <my-location-name>
```

The server finds the current location of the bus and stores it with the user-specified location name and the user's phone number. This association between location and personal name and phone number is maintained by the server and can then be used in arrival time queries coming from that same phone number. We expect that there will be some locations frequented by many users that will have well-known names and can be associated with GPS coordinates in the server's database and can be shared among all users. Our design allows the system to work in areas where little or no geo-coded street information exists – GPS coordinates do not need to be associated with locations on maps in order for users to tag locations or receive arrival predictions. User tags generated by the crowd mapping approach could potentially also be used to create maps in regions where few maps exist. In Kyrgyzstan where the street names changed after independence from the Soviet Union and a combination of old and new names are used, the *bus system could be used to map streets in a way that is comprehensible to all residents regardless of which naming system they use. Such a usage pattern would be an

additional contribution of the system, resulting in data similar to OpenStreetMap [18] or other grassroots mapping initiatives.

IV. DEPLOYMENT

To test the system, we created prototypes of *boxes and the server software. The system was first deployed in Seattle [6], and in March 2009, the system was deployed in Bishkek, Kyrgyzstan. During the deployments the authors took *boxes on multiple bus routes and multiple runs of the same route. Usability testing of the system with potential riders was also done in a lab setting. In this section, we describe data collected from the Bishkek deployments including lessons learned about the local infrastructure and use of the *box hardware, quantitative data on the performance of the system, and insight gained from usability tests. These lessons provide a valuable perspective on potential difficulties of design approaches that must span multiple and varying development contexts – in this case both Seattle and Bishkek.

A. Infrastructure Differences and Hardware Lessons

Characteristics of local cellular infrastructure are not widely publicized by providers and details of SIM card interfaces also vary among providers. Here we mention several issues we discovered, describe measurements of performance and reliability on the GSM networks that we used in our deployment, and offer design suggestions that may be instructional for other researchers deploying devices in unfamiliar or remote geographic areas.

SIM Card and Mobile Provider Network Differences

The experience in Kyrgyzstan revealed several differences with various SIM cards. For example, SIM cards from different providers appeared to have different timing requirements, and an error with explanation “SIM busy” would occur when the *box issued commands too quickly. A SIM card’s PIN requirement also varied based on the provider with some SIMs requiring no PIN while others required the PIN to be entered every time a mobile device was turned on. Pay-per-use accounts are much more popular in Kyrgyzstan than the subscription models common in the United States, and providers in Kyrgyzstan required the balance associated with a SIM card to be above a certain amount to successfully send a message. (In our experience, this was 30 som.)

Another notable difference of Kyrgyzstan mobile providers was the lack of communication between some networks. Not all networks had interoperability agreements, and thus some mobile phones were unable to contact mobile phones on different networks. To combat this, the *box server would need to be modified to interface with multiple mobile phones - one per disconnected network, to enable communication with all mobile devices. A dedicated phone for each mobile network could potentially lower operational costs, as often SMS messages are cheaper to send within the same network, as was the case in Kyrgyzstan.

We also discovered that network transmit power requirements were higher in Bishkek compared to the United

States. This was problematic because of design optimizations made to maximize battery life based on our Seattle testing, design optimizations commonly adopted when designing for resource constrained environments. Although in some cases the *box could be plugged into the vehicle being tracked, we wanted to be able to deploy the system in environments where this may not be possible (e.g. vehicle outlet is already in use or does not exist).

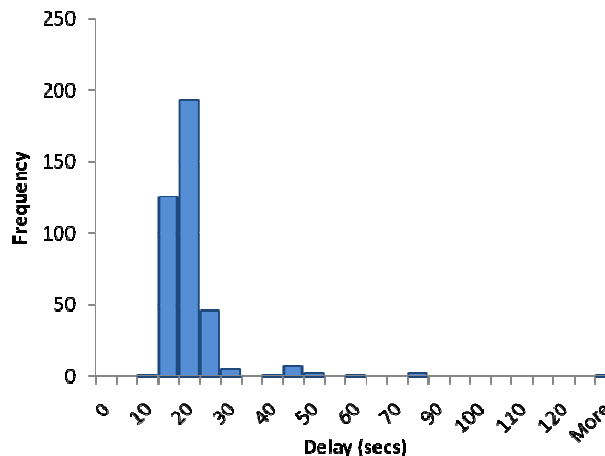


Fig. 4. Histogram of message delay within the Mobi network in Bishkek. The average delay is 18.6 seconds.

SMS Message Latency and Drop Rates

The time it takes for a user’s request to reach the server and for the prediction to arrive to the user affects the usefulness of the *bus system. If responses to a user’s query arrive later than the predicted arrival time, those responses are not helpful to the user. Furthermore, if messages are frequently delayed or dropped between *boxes and the server, the accuracy of the prediction may be adversely affected.

In Seattle, we measured the average SMS message latency between a *box and the server on the T-Mobile network to be approximately 10 seconds (standard deviation= \sim 7 seconds, $N=647$). In Bishkek, we repeated this process for the Mobi GSM network. Figure 4 shows the distribution of delay when both the sender and receiver are on the Mobi network. We found an average delay of approximately 19 seconds (standard deviation= \sim 11 seconds, $N=394$). The mean delay in Bishkek is longer than that in Seattle, and the larger standard deviation reflects the fact that text messages in Bishkek are more likely to experience a long delay of greater than 30 seconds.

We also measured dropped message rates within the Mobi network and from the Mobi to O! network in Bishkek. For the intra-network communication, we lost 5 messages from a total of 698 sent for a drop rate of less than one percent. In our T-Mobile measurements in Seattle, we lost 4 messages out of 716 sent, for a similarly small drop rate. We did not lose any of the 76 messages sent between Mobi to O! networks.

Need for Feedback on State of Hardware

Because of the previously mentioned issues with SIM cards and other sources of fallibility, there was an increased emphasis on the need for feedback to the staff on the status of

the hardware. As our hardware was implemented, it was often difficult to detect *box operational issues such as low battery power, inadequate currency on the prepaid mobile account, inability to obtain a GPS lock, and no GSM network reception. Additional visual information about the *box's operational state would enable staff to better recognize problems and facilitate timely corrective action. Currently, the GSM chip used in the *box has an external LED that provides feedback by blinking at a different rate based on the chip's state; however, this single visual output did not provide enough detailed information for effective monitoring or efficient troubleshooting. Our deployment team was often unclear if the chip was blinking quickly because the chip was constantly resetting due to low power, unable to contact a GSM network, or had run out of money on the pay-per-use accounts. The system would have benefited from clear and separate visual indicators of error conditions that prevented the *box from broadcasting location updates as expected.

B. Measurements of System Accuracy

In this section, we report on the accuracy of our system. We describe mapping routes as well as the accuracy of route prediction and geo-coding in Bishkek, Kyrgyzstan.

*Mapping Marshrutka Routes with *boxes*

*boxes were hand-carried on several runs of two marshrutka

routes in Bishkek. Route 114 (5.5km of travel distance) (see Figure 5) is an inner-city route that runs frequently but also travels slower than other routes because of more frequent stops. By contrast, Route 122 (7.5km of travel distance), mapped in Figure 6, travels from the center of town to the 12th micro-region, a residential area on the outskirts of the city. The 122 runs less frequently but also generally travels more quickly and makes fewer stops. The two routes demonstrate properties common to many routes in Bishkek. We rode each route at least five times during various times of the day to gain a feel for the consistency of different runs on each route.

One section of the Route 122 varied from run to run, as each driver sought to find a shortcut around a small congested area. Other factors, such as time of day, number of passengers, and planned or unexpected roadblocks, also affected the predictability of the run. Moreover, because they are private contractors, drivers have some liberty to make personal stops along the route, as happened during one of our runs on Route 114. Drivers are, however, aware of the duration of their runs, and some drivers time their runs using a digital timer displayed on the dashboard. Our observations and interviews suggest that drivers try to finish their routes as quickly as possible in order to make as many runs (and as much money) in a day as possible.



Fig. 5. Route 114 (5.5 km total) in Bishkek.



Fig. 6. Route 122 (7.55 km total) in Bishkek.

Run Timing Similarities & Prediction Accuracy

We intentionally chose an uncomplicated prediction algorithm for ease of implementation. It simply treats a single run specified by the server operator to be a timing model on which to base predictions of other runs. Because of the simplicity of our algorithm and its dependence on the timing similarities between runs of the same route, measures of accuracy may better be considered to reflect the timing regularity of runs and are a lower-bound on the accuracy possible with more sophisticated algorithms (e.g. [19]). Many variables can be incorporated into such an algorithm (time of day, day of week, real time information on weather and congestion) to improve its accuracy.

To measure this accuracy, we attempted to ascertain the error of a prediction compared to actual arrival time. For each route, we chose one of the collected runs to create a model. Once a model run was selected, the remaining runs were treated as a bus in motion. (For these accuracy measurements, we repeated the following process multiple times, allowing each of our collected runs to serve once as the model.) We chose a few locations along the route spaced at around two to three minute intervals. For each location, we were able to ascertain the actual arrival time of a run to the given location. To isolate a possible separate source of error, the locations created for this test were considered perfectly coded; these locations referred to a well-defined point along our route and were accurately given GPS coordinates through Google Earth.

We provided the prediction algorithm with the model, location, and simulated system time that was a known number of minutes prior to the actual arrival time to the location. We also assumed a conservative SMS message delay of 20 seconds. In order to create a prediction, the algorithm uses linear interpolation to find the closest points along the model path to the destination and the last known position of the bus. The time difference between these two points in the model is the estimated travel time of the bus to the location. This is used to construct an estimated arrival time and we were able to ascertain the error between the estimated and the actual arrival times in seconds. We repeated this method, for all locations with various known times to arrival on any unique combination of runs, for a total of 697 measurements of error.

As seen in Figure 7, the amount of error in the prediction increases with the amount of time the bus is actually away from the location (as expected). For example, when the current run is around five minutes away, the average error in prediction returned by our current algorithm would be about a minute and fifteen seconds. When the bus is fifteen minutes away, this mean error increases to about two minutes and fifteen seconds. Given the simplicity of our algorithm, this error is relatively small and demonstrates that prediction based on data provided by the *bus system is feasible.

Another way to view the error in our prediction is as a proportion of the actual time to arrival. The dotted line in Figure 7 provides a graph of this proportion as a percentage of the actual time the bus is away from the user’s location. As previously mentioned, bus drivers in Bishkek do not make an attempt at ensuring timing consistency between runs of a route. Some runs may be slower or faster than others. The error in our algorithm’s predictions provides a measure of the differences in timing between runs of the same route. As seen in Figure 7, the proportional error decreases as the actual time to arrival increases.

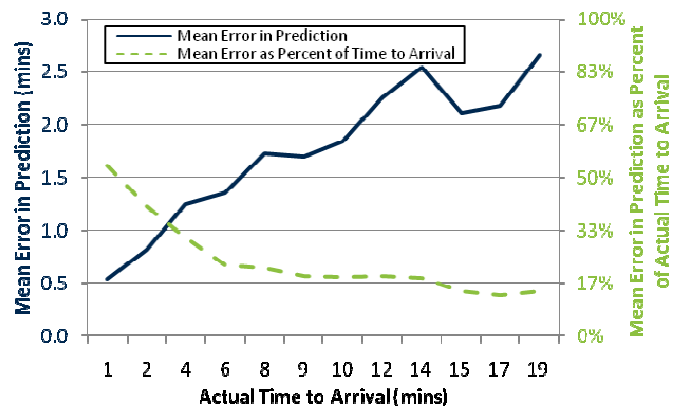


Fig. 7. Mean error in predictions by actual time to arrival.

Geo-coding Locations with *boxes

We tried naming several locations in Bishkek using *boxes on moving marshrutkas in order to separately test the accuracy of the geo-coding algorithm ("tagging"). Locations were geo-coded as would have been done by users, by sending a "store

location” query to the server at the time when a *bus was passing. These locations were compared with the GPS coordinates obtained by marking the intended location on Google Earth. Two metrics were calculated to gauge the accuracy of the geo-coded locations: physical distance and the average time it would take the bus to travel along its route between the coded and intended locations.

For convenience, geographic distance was measured “as the crow flies.” The time difference was measured by finding the points along a run of the route closest to the coded and intended locations. The time it took for that run of the route to travel between these points was then measured, using linear interpolation when points appeared between those given in the stream of location updates. This process was repeated for each of the runs for which we had collected data to obtain an average time difference between the coded and intended locations.

Table 1 presents data measured to gauge the accuracy of geo-coding. Errors on Route 122 tended to be larger than errors produced by locations coded on Route 114, as Route 122 buses had less frequent stops and often traveled faster. However, neither route produced coded locations with errors large enough to negatively affect predictions of bus arrival. While the average geographic error is over 200 meters, the time difference between the coded and intended locations is approximately thirty seconds. As predictions of bus arrival to a location are given to a minute accuracy, most users are unlikely to notice any inaccuracies in prediction due to errors in the coded locations. Thus, our geo-coding algorithm is accurate for the purposes of predicting arrival of buses, but more work can be done to improve its geographic accuracy.

TABLE I
ACCURACY MEASURES OF GEO-CODED LOCATIONS
(N=6 PER LOCATION)

Location Name	Route	Distance (meters)	Avg. Time Difference (secs)	Std. Dev Time Diff (secs)
East AV	114	134	28	7
Dordoi Mini	114	136	44	17
Toys	114	21	3	2
Beer	122	229	38	12
Hospital	122	448	51	11
Windmill	122	393	53	13
Ferris Wheel	122	178	20	4
Average	---	220	34	20

C. Usability Tests

Evaluation of user reactions to the system in use on actual buses will not be feasible without broader deployment – requiring more *boxes than we currently have constructed. However, in our visit to Bishkek we attempted to evaluate our proposed system with potential riders. We conducted usability tests to obtain user feedback on the functionality of the system

in a controlled setting. We created task-based scenarios that asked users to conduct several queries of the system, including route prediction and location tagging. We had 5 participants: two men and three women, students and professionals, ranging in age from 21 to 25.

The test materials focused on four tasks. The first task asked users to find out when the next bus following their route would arrive at their location. The second task asked the user to geo-code their current location and then find out when the next bus would arrive at that location. The third task asked the user to find out the names of publicly available locations. The final task asked users to request multiple updates on the status of a bus. While users performed these tasks, a ‘dummy bus’ was running on the server to mimic real-time route information.

While full results of the usability tests are not reported here, initial findings provide important feedback in three areas. First, the system and its functionality were relatively easy for potential users to grasp. Time on task was extremely short across all participants, although one study participant experimented with commands he thought “should” be in the system which slowed his completion of tasks. Overall, though, all participants easily grasped the purpose of geo-coding locations and predicting bus arrival times.

Second, the largest problem participants had conceptually with the system was related to command syntax. The query “<route-#> to <location>” generated a significant amount of confusion across 3 of the 5 participants; during the course of the tests, each of these participants spoke of the query as if it would tell them when the bus was going to arrive at their *destination*, rather than simply the next bus arriving at their current location. This problem can easily be solved by changing the language of the query so that it reads “next bus at” rather than “next bus to.” In the future, we may also be able to support multiple formats for the same query.

The final preliminary finding from the usability tests served to reinforce our initial goal in creating the system. That is, each participant was enthusiastic and asked whether the system was live and/or when they would be able to use it (in fact, some sent SMS messages to the system after the usability session ended). Although a general design ethnography provided the initial impetus for the transportation solution, the usability tests confirmed the significance of the problem space related to information about transportation resources in the community.

V. FUTURE WORK

Future work on *bus will focus on developing a location-based business model in preparation for a large-scale deployment in Bishkek, Kyrgyzstan. In addition, we would like to generalize *bus to fit other transportation needs and regions. In particular, we are interested to assess the potential of this type of grassroots, small-scale implementable solution as a contrast to large, municipally sponsored transportation

information systems that are increasingly common throughout the world. We can also envision our system being generalized for use in the transportation of goods—helping transporters connect with people who need to move goods—and a sort of Craigslist that includes reputation management and bidding for resources. Reputation management could also be used for feedback on bus drivers (something requested by one of our usability test participants). This would be particularly useful in the inter-city scenario where the bus drivers are not employees of a central transit authority and riders of vulnerable populations may not feel safe riding with strangers. It would also be useful for informal car-sharing that happens for long-distance travel in much of the developing world. An additional way to connect riders to drivers is a service that would alert drivers when riders are waiting for them such as [20]. One could envision extending our system to facilitate communication of other information we learned was useful to drivers (such as the location of speed traps, roadblocks, construction, etc.) using SMS. Sharing of location names would allow users access to much more data and a larger sense of investment in the system as well as opening up advertising opportunities [21], [22]. These changes would serve to create a more robust and marketable system for a larger scale deployment. Finally, we envision experimenting with using GPS-enabled mobile phones as *boxes and harnessing the increased processing power of smart phones for use as the *bus server.

VI. CONCLUSION

We presented an SMS-based solution for providing transit information based solely on existing cellular and GPS networks. The aim is to permit the development of grassroots information services that do not rely on a central authority or complex web hosting to deploy new applications. We relate our experience in developing our system for the marshrutka system of privately-run buses in Bishkek, Kyrgyzstan, but with an eye to looking more broadly at informal transportation networks that span developing regions. The findings in this paper are extensible to regions with limited resources, resource-constrained central governments, and ad-hoc transportation resources. We demonstrated that the system has an accuracy that supports our usage models; that riders and drivers recognize how our system could make a shared transportation system more transparent and meet important information needs; and that usability tests indicate the current interfaces are adequate. We are now turning our attention to business models and even simpler server models for lower maintenance deployment.

ACKNOWLEDGMENT

Thanks to our local research team, especially Mirgul Karimova, Indira Torogeldieva, Aidai Seidakmatova, and Medina Aitieva. Thanks also to our participants.

REFERENCES

- [1] "Transport's role in achieving the millennium development goals," Produced by the DFID Transport Resource Centre, August 2002, Available: http://siteresources.worldbank.org/INTWDR2004/Resources/22572_Tra nsportsRole.pdf
- [2] Metro Tracker, Last accessed: 27 July 2010. Available: <http://trackerloc.kingcounty.gov/>
- [3] MyBus, Last accessed: 27 July 2010. Available: <http://mybus.org>
- [4] NextBus, Last accessed: 27 July 2010. Available: <http://www.nextbus.com>
- [5] One Bus Away, Last accessed: 27 July 2010. Available: <http://onebusaway.org/>
- [6] R.E. Anderson, A. Poon, C. Lustig, W. Brunette, G. Borriello, and B.E. Kolko, "Building a transportation information system using only GPS and basic SMS infrastructure," in *Information and Communication Technologies for Development*, 2009.
- [7] CIA - The World FactBook, Last accessed: 27 July 2010. Available: <https://www.cia.gov/library/publications/the-world-factbook/>
- [8] B. Kolko, E. Rose, and E. Johnson, "Communication as information-seeking: the case for mobile social software for developing regions," Proc. of the *16th International Conference on World Wide Web*, 2007. ACM (2007), pp. 863-872.
- [9] C. Putnam, E. Johnson, E. Rose, and B. Kolko, "Adapting User-Centered Design Methods to Design for Diverse Populations," In *Special HCI4D Issue of Information Technologies and International Development (ITID)*, 5(4), 2009, pp. 51-73.
- [10] J. Donner, "Research approaches to mobile use in the developing world: a review of the literature," *The Information Society* 24(3), 2008.
- [11] MSR India SMS Toolkit, Last accessed: 30 July 2010. Available: <http://smstoolkit.codeplex.com/>
- [12] R. Veeraghavan, N. Yasodhar, and K. Toyama, "Warana unwired: replacing PCs with mobile phones in a Rural Sugarcane Cooperative," in *Information and Communication Technologies for Development*, 2007.
- [13] FrontlineSMS. Company Website, Last accessed: 30 July 2010. Available: <http://www.frontlinesms.com/>
- [14] RapidSMS, Last accessed: 30 July 2010. Available: <http://rapidsms.org/>
- [15] Unstructured Supplementary Services Data, Last accessed: 27 July 2010. Available: <http://www.telecomspace.com/messaging-ussd.html>
- [16] MySMS, Last accessed: 27 July 2010. Available: <http://mysms.sourceforge.net>
- [17] SMSLib, Last accessed: 27 July 2010. Available: <http://smslib.org>
- [18] OpenStreetMap, Last accessed: 29 July 2010. Available: <http://www.openstreetmap.org>
- [19] A. Karbassi and M. Barth, "Vehicle route prediction and time of arrival estimation techniques for improved transportation system management," Proc. of *Intelligent Vehicles Symposium*, 9, 11 (2003), pp. 511-516.
- [20] A. Kumar, N. Rajput, S. Agarwal, D. Chakraborty, and A. A. Nanavati, "Organizing the unorganized – employing IT to empower the underprivileged," Proc. of the *17th International Conference on World Wide Web*, 2008, ACM (2008), pp. 935-944.
- [21] F. Espinoza, P. Persson, A. Sandin, H. Nyström, E. Cacciatore and M. Bylund, "GeoNotes: social and navigational aspects of location-based information systems," Proc. of *UbiComp 2001*, Springer Berlin/Heidelberg (2001), pp. 2-17.
- [22] J. Stewart, S. Bauman, M. Escobar, J. Hilden, K. Bihani and M.W. Newman, "Accessible contextual information for urban orientation," Proc. of *UbiComp 2008*, ACM (2008), pp. 332-335.