

# Building a Transportation Information System Using Only GPS and Basic SMS Infrastructure

Ruth E. Anderson, Anthony Poon, Caitlin Lustig, Waylon Brunette, Gaetano Borriello, Beth E. Kolko

**Abstract**—This work consists of two main components: (a) a longitudinal ethnographic study in Kyrgyzstan that demonstrates the importance of transportation resources in the developing world and how to plan for an appropriate ICT solution, and (b) the results of a proof-of-concept system engineered to create a bottom-up, transportation information infrastructure using only GPS and SMS. Transportation is a very important shared resource; enabling efficient and effective use of such resources aids overall development goals.

The system, \*bus, involved the development of a hardware device (a \*box) containing a GSM modem and a GPS unit, that can be installed on a vehicle and used to track its location. The \*box communicates via SMS with a server connected to a basic GSM phone. The server runs route a prediction algorithm and users can send SMS messages to the server to find when a bus will arrive at their location.

The paper discusses the system and early testing, as well as the development implications for a range of urban and rural environments where transportation is scarce or inefficient, and where a central authority or institution is not in a position to provide robust information resources for users. We describe how the solution is also situated within technology usage patterns common to the developing world.

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

## I. INTRODUCTION

TRANSPORTATION of goods and people is key to economic and human development. This paper discusses the development of the \*bus system (pronounced “Starbus”), a transportation information system that uses GPS, GSM, and SMS technologies. The \*bus system was developed in response to transportation challenges experienced in

Manuscript received September 22, 2008. 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).

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

C. Lustig, Dep’t of Computer Science & Engineering, University of Washington, Seattle, WA 98195 USA. She is now with Intermecc, Inc. (c.lustig@gmail.com).

W. Brunette is with the Computer Science & Engineering, University of Washington, Seattle, WA 98195 USA (wrb@cs.washington.edu).

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

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

Kyrgyzstan, a developing region with poor infrastructure and limited resources; consequently, the findings in this paper are extensible to other selected developing regions with limited resources, fragile road infrastructure, resource-constrained central government authority, and ad-hoc transportation resources for both inter- and intra-city transport.

## II. TRANSPORTATION AND THE KYRGYZ CONTEXT

### A. Background on the Transportation Challenge

Transportation has long been recognized as having an impact on economic development. Factors such as levels of investment in transportation have been shown to have a positive impact on economic growth [1] [2]. In the developing world, the availability of safe and reliable transportation can have even broader implications. Not only does access to transport improve access to markets and mobility of workers, but it is also critical to the timely and affordable delivery of services such as health and education, can serve to empower vulnerable groups by increasing their independence, and is key for maintaining social networks [3]-[5].

Along with providing access to transportation, it is also important to provide potential users with information about its availability. In the developed world, users can often access information about most bus and train schedules easily via printed schedules or web pages maintained by centrally-funded transportation authorities. In some 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 [6]-[8]. However, when a community lacks the infrastructure to provide such information resources, potential users can find themselves unable to take advantage of whatever (limited) transportation resources are available.

### B. Methods

The background provided in the rest of this paper is based on several years of research in Central Asia by the Central Asia Information and Communication Technologies (CAICT) project at the University of Washington. Our research has included multiple periods of ethnographic research including a design ethnography, interview studies, and a longitudinal survey in four countries of Central Asia.

Ethnography has been conducted since 2000 and includes a team of US and Central Asia researchers. Field notes and photography are both used to document findings. Survey results are based on a nationwide survey of 1000 respondents

in each of four countries (Kazakhstan, Kyrgyzstan, Tajikistan, Uzbekistan) of age 15 and older, administered in 2006, 2007, and 2008. The survey sample was based on census information for age, gender, ethnicity, and geographic location as released by each country's government. The survey includes multiple urban and rural sampling locations, and it was administered in Russian and other regional/local languages. In addition to the general sampling scheme, a three-stage process was used to select respondents:

- probability proportional to sample size of Primary sampling units (PSUs);
- consecutive random sampling of households in determined PSU; and
- selection of a household respondent using a Kish Grid method.

The survey is a general social survey that asks a series of questions about technology usage, as well as information-gathering habits and patterns of trust and confidence. Because Internet usage rates in the region tend to be low, the survey was not designed to target Internet users specifically, but, rather, gauge general trends among the overall population.

### C. Inter-city and Intra-city Transportation within Kyrgyzstan

Kyrgyzstan is a small, mountainous country in Central Asia. Its population is about 5.5 million and concentrated in the capital of Bishkek in the north and in the heavily agricultural Ferghana Valley in the south. The country has a youth bubble, with the median age of 24.2 years (for comparison, the US median age is 36.7). The population is defined as 98.7% literate; ethnically the population is largely Kyrgyz (64.9%), followed by Uzbek (13.8%) and Russian (12.5%).

Since independence from the Soviet Union in 1991, Kyrgyzstan has struggled economically. Unemployment is high and about 40% of the population lives below the poverty line [9].

Personal vehicle ownership rates are low, and the population depends heavily on shared and public transportation for inter-city and intra-city transportation. Cities have taxis and an independently operated bus system known as the marshrutka system. *Marshrutnoye taksi* is a Russian phrase for 'routed taxicab'; the marshrutka system is a shared taxi transportation system that runs on largely pre-plotted routes but without standard schedules; the standard bus is a mini-van that 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. The drivers are often independent contractors 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 the rough routing listed as major stops or landmarks. 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. Quickly identifying an appropriate marshrutka is a challenging step when a new route is being attempted. There is a lack of official bus stops, and people stand at areas known

to be places where the marshrutkas stop. There is no official schedule, so riders are not always sure how long they will have to wait. Sometimes buses are full and will not stop, and then riders have another unknown stretch of time to wait. Routes can shift slightly, and because stops are often demarcated by landmarks rather than actual routings, users sometimes have to guess what route a bus will take to get from listed landmark A to listed landmark B. Overall, the system can work well if one is traveling a familiar route that does not get too crowded, but if one travels at peak times, or off hours, or wants to get somewhere new, the system can be frustrating to navigate because there is no clear pathway to gather the information relevant to a rider.

Marshrutkas operate on inter-city as well as intra-city routes. In Kyrgyzstan the inter-city system is relatively transparent, if not especially efficient. In order to travel between towns, one goes to the bus depot where marshrutkas are lined up. In Bishkek, the bus depot is large and well signed, with destinations labeled on placards hanging over the bay where the minibuses are parked. In smaller cities the depots can be parking lots on the edge of town.

At the Bishkek depot, drivers stand in front of their buses with microphones, calling out their destination to the crowd. There is no departure time listed, and drivers will not leave until the bus is full. If your destination is a popular one, you may not have long to wait. But if you are headed to a smaller town on a not particularly well-traveled route, you could face a significant wait before the bus leaves. It is these kinds of inefficiencies that make the system ripe for technology-assisted solutions that could disseminate information to users and prospective users of the transportation infrastructure.



Figure 1 A bus depot in Bishkek with marshrutkas and car taxis.

In addition to the bus depots, there are also more informal ways of traveling between cities. In neighboring Uzbekistan, for example, if you want to travel and not take the buses (which can be overcrowded and perilous on the mountain roads), you can travel to an acknowledged site – in the case of the capital Tashkent, this is a subway stop furthest in the direction in which you want to travel. Or, in Tajikistan, a mountainous country with limited public transportation resources, you similarly go to an established place and look for a driver headed in your direction; at times these trips can

last for days. There, drivers with cars congregate, looking for passengers to fill up their cars; some drivers are professionals, some are there to help defray the costs of getting home (often after delivering agricultural produce or visiting relatives).

The characteristic these countries have in common, and that they share with many developing countries, is that shared transportation often happens outside the realm of a national or metropolitan transit authority. This means there is no central body coordinating or allocating resources to ensure efficiency or at least moderate attempts to meet the needs of citizens. Our research considered how, under such circumstances, one might aggregate and disseminate information about transportation options to make it more efficient or effective for both riders and drivers. Another way of stating the problem is how might one 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 resources, will, or because the resource itself is decentralized and largely part of the informal sector.

*D. Personal Technology Use in the Region*

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, while mobile phone use is growing at a remarkable rate. As Figure 2 shows, the mobile usage rate in Kyrgyzstan has grown from 21% to 71% from 2006 to 2008. During that same time, Internet usage has grown almost imperceptibly. Even in neighboring Kazakhstan, with a rapidly developing economy rich with oil and gas money, Internet use is holding flat, with only 19% of the population reported as Internet users in 2007 and 2008. And even though computer use climbed from 29% to 43% in Kazakhstan over the same time frame, Internet usage showed no increase.

The rapid increase in mobile usage is not unique to Kyrgyzstan, and adoption of mobiles in and of itself is not enough to establish that a mobile solution is the best avenue for a transportation solution. However, in addition to mobile usage, what we also see in Kyrgyzstan is that people use their phones for personal calls more than business calls, they are using SMS in measurable amounts (about one-third of users use SMS; see Figure 3 ), and they are **not** using the Internet as their primary mode of ICT access.

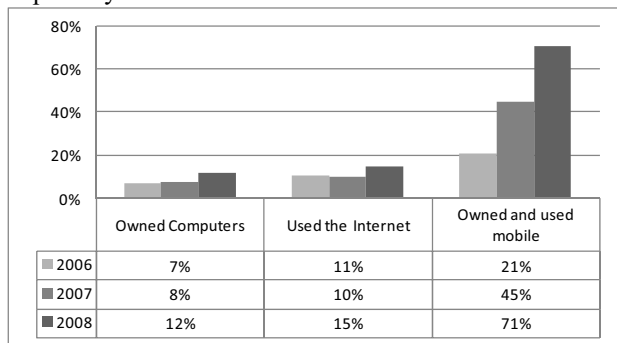


Figure 2 Personal Technology Use in Kyrgyzstan.

As Figure 3 illustrates, approximately one third of mobile phone users report using SMS; interestingly, this number held rather steady between 2007 and 2008, unlike overall mobile adoption numbers. This rate of SMS usage is larger than those using their phones to find information, and the percentage of people using their phones to connect to the Internet decreased, but is so low as to be irrelevant (4% in 2007, 1% in 2008). There are information numbers one can call in Kyrgyzstan to get information about local businesses, so it is possible for mobile phone users to seek information via voice.

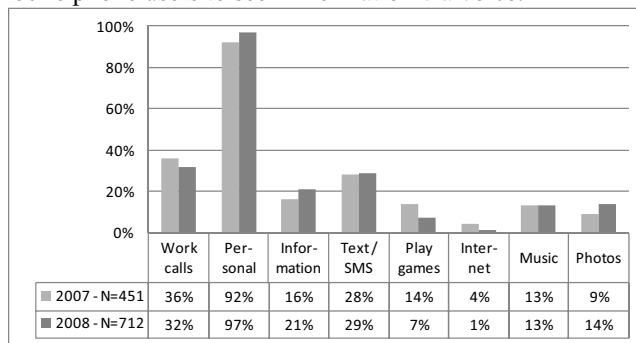


Figure 3 Activities people engage in with mobile phones

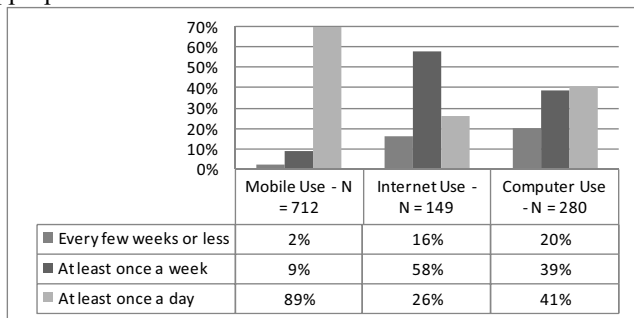
The issue of Internet use is especially key to the particular solution this team has created. Survey findings in the region illustrate that not only is mobile usage nearly five times that of Internet usage (see Figure 2), but even when people do identify themselves as Internet users, their model of usage is very different from that of the developed world. This point is crucial, and it feeds directly into the \*bus system developed as part of this study.

In a city like Seattle, for example, extensive information about public transportation is available on the Internet, and people access this information either through a computer or a mobile phone. Information about bus routes, timetables, fares, re-routings, etc. is available on a comprehensive and user-friendly site. However, this model of information delivery is predicated on the reality that most users of the system have regular access to Internet resources. The percentage of Internet users in the US is estimated at 85%, and over half of those users are connected on a regular basis [10]. In Kyrgyzstan, however, even among Internet users, the majority report using the Internet once a week or less (see Figure 4). Meanwhile, mobile users use their phones on a much more frequent basis, with 89% reporting using them at least once a day.

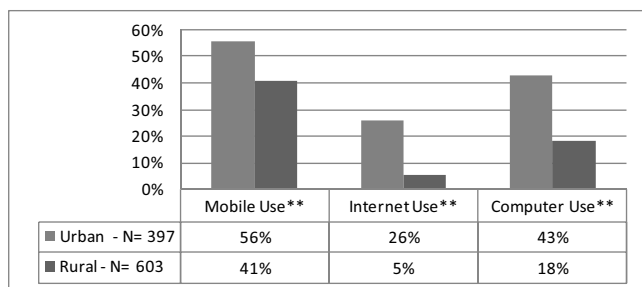
What this means is that developing and delivering a robust information delivery system that relies on people having Internet access in order to take advantage of the information resource is a fundamental mismatch between a technology and user habits and abilities. Recognizing what actually constitutes an “Internet user” then, in many parts of the developing world, helps to better envision an appropriate technological solution: delivery of information via mobile phone-based SMS.

In addition, it is worth noting that mobile phone usage rates are significant among rural as well as urban populations. As Figure 5 below illustrates, mobile usage is not only an urban phenomenon. In many ICT and development situations, there

is a gap between rural and urban users. In this project, we have taken care to include the usage behavior of both urban and rural populations in order to develop a solution that will be appropriate for the broadest swath of users.



**Figure 4 Frequency of technology use in Kyrgyzstan, 2008**



**Figure 5 Urban and rural use of technology in Kyrgyzstan, 2008**

\*\* =  $p < .001$  based on Chi-square test

### E. Design Ethnography Findings Regarding Transportation Challenges

During the summer of 2006, the CAICT team conducted design ethnography work in Kyrgyzstan to better understand people's habits of technology use within their personal communities, as well as the pain points of everyday life, which might best indicate potential solutions that leverage those preexisting technology usage patterns.

Four groups participated in the design ethnography: two multi-generational family groups, and two groups of young people (18-25). Interviews were conducted in an urban area, and a smaller, regional town. At each site one family and one group of young people were interviewed. Interviews covered issues related to technology use including Internet and mobile phones, challenges of everyday life, and social networks. In particular, interviews and questionnaires focused on patterns of reciprocity of goods and information exchange within the groups, as well as whether any technologies were used to facilitate such exchanges.

Findings from the design ethnography indicated mobile phones were extremely important to participants, and that both voice and SMS played a daily role in people's lives. Of the 12 participants in the study, only one did not own a mobile phone.

Transportation emerged as one of the significant problem spaces for our users. Although for most people, a general sense of lacking information resources outside of family and friend networks was prevalent, the transportation issue was of particular interest because it generated animated discussion among our users. This was the case because the way

transportation worked was in stark contrast to the public information resources available in many developed country communities, and because transportation was clearly a fundamentally shared resource in these settings.

As we continued to analyze the findings from the design ethnography, we found the following: (a) pricing issues for marshrutkas were not a concern, however the failure of the marshrutka system often meant people had to take taxis which were seen as prohibitively expensive, (b) safety and security were issues, especially for women, and this related to uncertain routings, riding at night, uncertain waiting times, and buses that would be full and unable to take new passengers, (c) learning new routes was seen as daunting, (d) there was some confusion over how much information about routes was available either through maps or online, and (e) inter-city travel, especially, often created a level of schedule uncertainty.

Among these issues, the ones that stood out as particularly addressable were making the shared ride system more efficient so that riders would not be forced to default to the more expensive taxi system, and increasing information awareness to give users the mechanism by which they could increase their personal safety but still make use of the shared ride system. Although the \*bus system currently focuses on intra-city travel, the development impact is perhaps even more significant for inter-city use where long-distance transportation resources are even more scarce.

Research into transportation information systems in the developed world turned up municipal ride share systems such as 511 Rideshare in California's Bay Area which allows riders and drivers to find one another; robust information resources available via the Internet include bus location information, and mobile-phone based systems (e.g., in Scandinavia) that provide up-to-the-minute schedule information so people can time their arrival at their bus, tram, or train stop.

### III. DESIGNING AN APPROPRIATE TECHNOLOGICAL SOLUTION

We designed a system with the goal of improving access to transit information for potential bus riders at a minimum of cost to users and without requiring the assistance of a central authority. For our solution, we chose three existing technologies appropriate for Bishkek (and for much of the developing world as well) and augmented them with two new components we developed. We describe these engineered components later in Section V. The existing technologies we used were SMS, GPS, and GSM, all described below.

#### A. Short Message Service (SMS)

SMS, also known as text messaging, is a protocol to exchange short messages between mobile phones. Our system uses SMS as its primary communication mechanism. We chose SMS for several reasons. As discussed in Section II.D, mobile phone ownership and text messaging is common in Bishkek. Text messages are perceived as relatively cheap, costing anywhere from US\$0.01 to \$0.06 to send a message, and for many plans it is free to receive messages. Cell coverage in Bishkek is also excellent, and the country overall

has good coverage. Good urban and rural cell coverage is becoming the case in much of the developing world.

Additionally, in the ICTD community, SMS-based solutions have proven robust, flexible, and valuable to multiple communities [11]-[13]. Work such as Frontline SMS and Warana Unwired demonstrate that information delivered via SMS can have a tangible, positive impact on individuals' lives [14] [15].

Although the mobile web is a compelling solution space for some, our research in Kyrgyzstan and elsewhere points to the value of solving for cheaper devices and more lightweight technology solutions that dovetail with technology patterns already in place. Text messages, as opposed to richer mechanisms for distributing information, such as web pages, can be supported by cheap and ubiquitous phones; additionally, because of the way the mobile phone integrates into how people already get information in their daily lives, it is a better solution space than Internet-based resources [16].

### B. Global Positioning System (GPS)

GPS satellites transmit microwave signals to GPS receivers that use data from the signal to determine the location of the receiver. Our system relies on information provided by GPS satellites to determine the current location of buses participating in the system. GPS was chosen over other methods of location determination, such as GSM triangulation [17], for its high level of accuracy – within 8-10 meters, and its appropriateness for outdoor usage, particularly in rural areas or urban areas with few tall buildings and mostly wide streets. Although GPS units continue to decrease in price, they currently are only included as part of high end cell phones. Thus, our system attempts to minimize the number of GPS units required by allowing bus riders to leverage GPS units in buses to “tag” locations they are interested in querying about in the future (see V.B.3)). In this way, we are leveraging an expensive piece of technology that is rare in the community (GPS enabled phones are not a significant part of the market in most parts of the world) in order to make its capabilities serve the needs of a larger group. In the \*bus system, multiple individuals can make use of the capabilities of GPS technology despite individuals not owning their own device.

### C. Global System for Mobile (GSM)

GSM communications is the most popular mobile phone network standard. Most mobile phones, including those commonly used in Kyrgyzstan, are on GSM networks. In 2005, the GSM family of technologies had more than 1.5 billion subscribers, more than the total number of Internet users world-wide [18].

## IV. PROPOSED SOLUTION

The overall architecture of the \*bus project, shown below in Figure 6, incorporates the three existing technologies mentioned above with two new components. These engineered components are: a GPS vehicle tracking box (*\*box* pronounced “Starbox”) and a back-end server that aggregates the location of the \*boxes, runs a route prediction algorithm,

and provides the querying service to riders. These are both described in more detail later in Section V. The goal of the \*bus system is to track buses' GPS coordinates and to provide users an SMS messaging system that will tell them when a bus should be arriving at their location.

The \*box is a low-cost vehicle tracking solution that utilizes GPS satellites to locate \*box-equipped vehicles; the \*box hardware includes a GPS device and a GSM modem integrated into a single package that just requires a local SIM card. The GSM modem allows the \*box to send SMS messages with the GPS data to the server, taking advantage of the widespread GSM cellular phone networks.

On the back end, the central server continually collects the GPS location data from all \*boxes and stores it in a database. The server consists of a laptop computer connected to a mobile phone capable of sending SMS messages and that serves a gateway to the SMS services. The server does not need to be connected to the Internet, and the service does not require cooperation with mobile providers. Potential bus riders can use their cell phones to send SMS queries to the \*box server and receive transit information in response. Initially, the system will just support riders' queries for transit arrival times, but it has been designed to be easily expanded to handle more complex activities such as route planning or registration for notification of bus location updates.

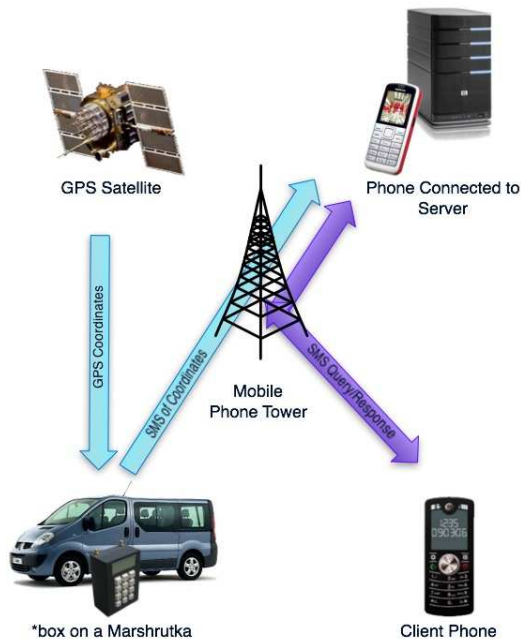
## V. ENGINEERED COMPONENTS

Each bus driver participating in the \*bus transportation network will need to have a \*box device installed or placed on their bus with them at all times to allow riders to receive updates on their expected arrival times. At least one server, connected to a cell phone, also needs to be running at all times to receive location update messages from \*boxes and to accept and respond to queries from users. That server can be located anywhere in the country with cell coverage.

### A. \*box

#### 1) \*box Driver Interface

The \*box was designed to require minimal interaction with the bus driver. In order to use the \*box, the driver only needs to turn it on (using a button located on the side), and then enter a route number using a keypad as shown below in Figure 7. The selected route number is displayed on a small LCD and stored locally on the \*box to be appended to each location update message sent to the server. A version of the display and button labels is under development in Russian and can easily be adapted for other languages.



**Figure 6 Overview of the \*bus System**

The notion of a route number is a good fit for the marshrutka system in Bishkek where buses have set routes (although no set timings along those routes), indicated by numbers displayed in bus windows. Changing routes is done easily by typing in the new route number. 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 attempting to determine 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. The \*box was designed to open easily to allow for upgrades of the internal components and changing of batteries, or inspection of device internals to allay fears of riders or drivers. Drivers can turn the \*box off when not in service and safeguard the privacy of off-duty movements. Since we would expect deployment models to build in driver accountability, we do not consider theft to be a major concern.

We built three prototype devices at a cost of about US\$200 each, although we believe mass production would bring the cost per box closer to US\$50. In the future, as GPS-enabled cell phones become less expensive, the functionality of a \*box could potentially be replaced by an application running on a driver's own cell phone.

Currently, the holder of the \*box, in this case the driver or private company that owns the marshrutka, would absorb the cost of the text messages sent out from the system. Our group has considered cost, and the barrier it presents to adoption. Increased efficiencies may or may not be significant enough for an individual driver to adopt the system, but there are other entities that might sponsor such a system for users, including a newly formed transit authority, mobile companies which are counting on increased SMS traffic from users, or the employing marshrutka company deciding to adopt the system.



**Figure 7 Prototype of the \*box. Note the golden pigtail connectors on the top of the box to allow easy attachment of GSM and GPS antennas.**

## 2) \*box Internals

In addition to being able to accept route numbers via a keypad and display the provided route number to the driver, internally the \*box must also be able to determine its location (via GPS satellites) and send a time-stamped SMS to the server. For our prototype device, we chose the Telit GM862-GPS module. We chose to implement the \*bus system with this very compact module because it provides a low cost solution with the additional convenience of combining GPS and GSM capabilities on one chip (see Figure 8). Because no additional work is necessary to interface the GPS and GSM components, developing on this chip is relatively simple. The Telit module communicates with the keypad via general purpose IO to obtain the driver's currently selected route number and stores the number in memory. An ATmega16 micro-controller is used to coordinate the communication between the Telit chip, keypad and the LCD. The module continuously obtains its GPS location information from GPS satellites without any action by the driver and reports this information along with its currently selected route number to the server via an SMS message.

## B. \*bus Server

The primary purpose of the server is to accept SMS messages coming from \*boxes and bus riders, and to process them accordingly. 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 receiving SMS messages. Lower system requirements allow the server to run using older and less expensive hardware (such a system may cost around US\$200 to \$300). To minimize SMS latency, it is recommended that the phone connected to the server be on the same cell network as the \*boxes, although this is not required. The server does not need to be connected to the Internet, facilitating its deployment in areas with low Internet connectivity. Our prototype server was implemented using MySMS [19], an application framework built on top of SMSLib [20]. MySMS provides a transport layer, allows SQL queries, notifications, remote code execution and rapid application development to be executed over SMS.

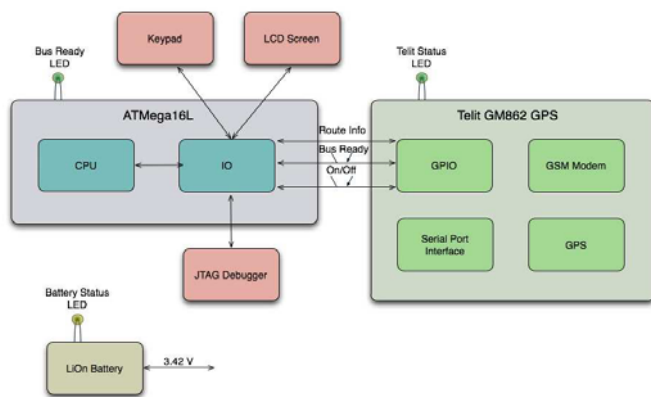


Figure 8 \*box Internals.

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

### 1) Location Update Messages from \*boxes

Messages sent from the \*boxes to the server have the following format:

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

The `bus-id` field is the unique identifier assigned to each individual bus that is 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. Riders will need the bus-id in order to associate a name to a location (described later in 3)).

The server records this information in a database for use in responding to user queries.

### 2) Arrival Time Queries from Users

Information gathered about the current location of buses in the system could be used to answer a variety of queries. The server continuously receives and replies to SMS queries from users. Potential bus riders simply text to a special \*bus server telephone number, and the server responds by sending an SMS back to the user's phone number. The most basic example is to ask the server to predict the arrival time of the next bus travelling on a given route number at a given location:

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

The server replies to the user with a SMS predicting:

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

These examples demonstrate the types of queries we support in English, but the exact format and language can be changed. The system has been designed so that it can accept queries and push out information in alternate character sets – in this case, in Cyrillic to support both the Russian and Kyrgyz languages. While some issues related to a multi-lingual society pose significant challenges (for example, when a user sends a query in a mixture of two languages), the system can be modified to support linguistic variation.

### 3) Geo-coding Requests

We expect that there will be some locations frequented by

many users that will have well-known names. These can potentially be associated with GPS coordinates in the server's database by the server operator. However, the \*bus system was designed to work in scenarios where little or no geo-coded street information exists. The solution to this problem is that we allow the clients to geo-code a location, on the fly, by using the GPS equipment already installed on the \*buses.

When a user wants to be able to refer to their current location in future \*bus queries, they can associate a name of their choosing with the GPS coordinates of their location; they do this by becoming a secondary user of the GPS on the bus. The user can instruct the server to associate an arbitrary name to their current location by waiting until a bus arrives at that location and immediately sending the following SMS message to the server:

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

`bus-id` is the \*bus id of the bus that has arrived at the user's location. In this case, then, a user's location that she knows only by street name is now given GPS coordinates thanks to the GPS receiver on the bus. The server takes the bus's GPS coordinates at the time that the message was sent and associates it with the location name specified by the user. Such geo-coded locations can be retrieved in the future based on the location name and the user's phone number. Associating location names with phone numbers prevents naming conflicts between users, although other schemes are also possible whereby location names can be shared with other users, thus organically growing a set of tags for places of common interest. User tags could also be used to create maps in regions where few maps exist. This is especially important 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.

## VI. PROTOTYPE DEPLOYMENT RESULTS

We have implemented prototype versions of the \*box hardware and the server software. In this section, we describe some preliminary data collected on the technical performance of the system.

We deployed a \*box on several buses in the University District area of Seattle. The \*box sent SMS messages to the server containing its GPS coordinates at an interval of one message every 30 seconds. Based on this data the server composed a model of each of the routes. We were interested in examining the accuracy of our basic route prediction algorithm as well as the amount of time required for users to receive a response to their queries.

Although we have not yet field tested the system in Kyrgyzstan, we have been careful to base the engineered solution on technology that will work in resource constrained environments. Specifically, fieldwork over the years has demonstrated to the research team that cell coverage and SMS latency in Bishkek is consistent with that experienced in Seattle. Additionally, the relatively open architectural space of Bishkek means that GPS readings will be easier to obtain.

Traffic patterns in the two cities are, of course, different, but the route prediction algorithm deployed (see next section) can be modified based on local traffic conditions.

### A. Route Prediction Accuracy

#### 1) Route Model

Much research has been done on route prediction algorithms for public transit and vehicles in general (e.g. [21]). For our prototype, we implemented a simple algorithm based on the concept of building a model by merging together multiple runs of the route. While one could imagine an implementation that requires the server to be pre-seeded with detailed sets of GPS coordinates for each route, our method provides the advantage of low setup costs for the \*bus server, as no special action is required to enter routes into its database. Furthermore, prediction models can be made more accurate by merging additional runs of the route, and more models for varying traffic patterns can easily be created. More sophisticated algorithms can be used in future versions, such as ones which automatically account for varying traffic flow and are tolerant of changing routes.

We use our route prediction model to provide responses to user queries about when the next bus travelling on a certain route is due to arrive at a particular location. To make this prediction, the server takes recent coordinates from current runs of the route and the location entered by the user and attempts to find these points in the model. The time between the corresponding locations in the model provides an estimate of the amount of time until the bus arrives at the user's location.

The accuracy of these estimated arrival times depends on two things: first, the fidelity of the model to the path of the route and secondly, the timing similarity between the route model and the buses whose status is being queried. The model fidelity is largely determined by the number of points in the model. A point consists of a set of GPS coordinates and the GPS satellite-provided time of day that the data point was measured. Accuracy of data points should be quite good (GPS accuracy is within 8-10 meters). Thus, data points collected for earlier runs on a given route should provide an accurate picture of where those buses were at the time provided.

Number of data points collected per route is determined by the frequency that \*boxes are configured to send location update messages to the server. More frequent location update messages incur more cost and put more load on the server, so it is desirable to keep this frequency low, yet high enough to give accurate arrival estimates.

Figure 9 below shows a plot of \*box location data showing the level of accuracy provided by different update message frequencies. As the figure shows, in general, sending update messages more frequently is more likely to create a model that follows the actual path of the route. Particularly around curves and turns, less frequent updates can make it seem as though the bus has cut a corner.



(a) Plot of the acquired GPS data for the whole route (approximately 3 miles) (Google Earth)



(b) Detailed view of the plot of acquired GPS data (Google Earth)

**Figure 9 Accuracy and resolution of GPS data collected at different message send intervals: red – 10 seconds yellow – 20 seconds, blue – 30 seconds, purple -40 seconds, green – 60 seconds.**

One option is to first produce one run with finer granularity, for example, a location update every 10 seconds. This can also be done manually to specify the exact path of the route. The method we employ takes points from multiple runs of the route and combines them to produce a model with greater resolution than a single run. Thus, more runs of the same route contribute to overall accuracy of the prediction.

In order to provide accurate estimates of arrival times, a route model must also reflect the timing characteristics of buses whose status is currently being queried. The extent to which this is true depends on multiple factors. On a bus system such as the one in Seattle, where there are set stops, and drivers are expected to arrive at stops at given times, variance can be caused by factors such as weather or unusually high traffic. In Bishkek, although marshrutkas follow set routes, there is no schedule of any sort. Thus, we would expect that there might be a wider variance in timings for buses in Bishkek. Collecting timing data from more bus runs and using an algorithm that incorporates time of day and day of week would help to improve accuracy in this case.

#### 2) Prediction Accuracy Results

We deployed \*boxes on portions of two Metro bus routes in Seattle and collected data for four runs on each of these routes for a total of eight runs. The distance traveled was 2 miles on one route and 2.8 miles on the second. Each run consisted of an average of 11 minutes of collected data. While this running time is shorter than the length of routes we would expect to use the system with, it allowed us to collect enough data to test and provide a proof of concept.

To examine the accuracy of our preliminary route prediction algorithm, we did the following. For each route, we chose one to three of the four runs to create a model. The remaining runs, which did not become part of the model, were treated as a bus



in motion and we ran our algorithm to generate predictions about those runs. These predictions were made for five to six simulated user locations per route, and when the bus was anywhere from one to nine minutes away from that simulated user. We provided the algorithm with the model, a time to treat as the current time of the system, calculated by subtracting the expected prediction value from the actual arrival time, and a set of location updates prior to this time, accounting for location message delays (see VI. B). We compared the prediction generated to the actual timing data gathered for that run for a total of 547 measurements of error.

The error in a prediction is the absolute value of the difference between the actual time it took for a bus to arrive at the destination and the time predicted by the model. The mean error from our predictions was approximately 94.9 seconds, with a standard deviation of 83.2 seconds. As seen in Figure 10, the amount of error in the prediction increases with the amount of time the bus actually is away from the user. The error can be understood as a proportion of the actual time to arrival of the bus, with the mean proportion equal to approximately 37.9% with a standard deviation of 32.2%. For example, if the bus will actually arrive at a destination in 10 minutes, we can expect the prediction of this arrival time to be, on average about 3 minutes 45 seconds off.

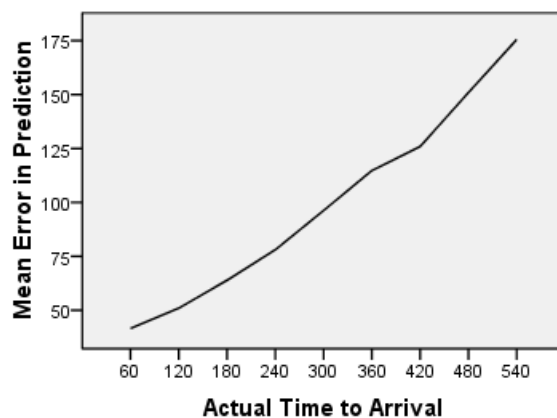
In addition to the accuracy of our model, the amount of time required for SMS messages to travel through the network, both those giving location updates and those from user interaction, also affects the accuracy and usefulness of bus arrival time predictions. We examine these factors below.

## B. Message Latency

### 1) \*box to Server Latency

The accuracy of a particular prediction depends on both the accuracy of the route model and on receiving a recent, and thus accurate, location update from the bus whose status is currently being queried. Ideally, the server would have instantaneous information about the exact location of a moving bus at the point it receives a user query about that bus. In practice, the server must base its prediction on the last location update message received from that bus. Both the configured \*box frequency of sending updates and any delay in that message getting from the \*box to the server affects accuracy of information about the current bus location.

Latency of SMS messages can be affected by a variety of factors, such as cell phone network congestion, signal coverage for the \*box or the server, or the need for messages to cross networks. We recommend that the \*box and the server be on the same cell phone network, as they were in our implementation, to minimize this delay.



**Figure 10 Mean error in the predicted arrival time, plotted by the actual time to arrival (both axes in seconds).**

We examined 381 messages sent from the \*box to the server to measure this latency. The average latency, measured from the time it was sent by the \*box to the time it was received by the server, is 19.96 seconds, with a range of 10 to 40 seconds (standard deviation of 10.3 seconds). We consider this an acceptable latency for our particular application. Although these measurements were taken in Seattle, fieldwork over the years has demonstrated to the research team that cell coverage and SMS latency in Bishkek is consistent with that experienced in Seattle.

### 2) User Query Latency

Queries return information to users in the form of an estimated time (e.g. 5:30pm) that a bus is expected to arrive at their location. The usefulness of this information depends upon both its accuracy and how quickly it is returned to the user.

Factors affecting query processing latency include: server load, cell phone network load, and system limitations on rate of SMS message sends. Based on our measurements of message latency in the previous section, we would predict the roundtrip time for the user to receive a response to their query to be the latency of two SMS messages (assuming time spent at the server is negligible), or approximately 40 seconds. Considering that the typical usage scenario would return predictions on buses several minutes away, we consider this latency to be acceptable.

## VII. CONCLUSIONS AND FUTURE WORK

The \*bus system is a direct response to an articulated problem in the developing world, and it takes into close consideration technology usage patterns and technology infrastructure available in such resource-constrained environments.

The system recognizes that information is valuable, and that the lack of information inhibits certain kinds of growth and efficiencies. However, there are not always central authorities or institutions with the resources or motivation to supply information infrastructure for citizens, and the \*bus system provides a mechanism by which decentralized operators of a

transportation system can create information resources for themselves. It is also a system that uses mobile technology in a way that does not require the direct participation of the mobile provider, thus cutting out one potential barrier to adoption. Finally, \*bus enables community use of a single GPS system to allow for user-generated geo-coding of the environment.

Next steps for \*bus include testing in Kyrgyzstan in early 2009. Our goal is to test both intra-city transportation in Bishkek, and inter-city transportation elsewhere, perhaps in another country. Although the \*bus system has been largely built out with the concept of urban transport in mind, the development impact for rural areas is conceivably even greater. In areas where distances are long, roads limited, and vehicles intermittent, a system such as \*bus could have dramatic implications.

Future work on \*bus includes examining possibilities such as incorporating some way to indicate the capacity of buses or allowing communication between drivers and riders, particularly for longer inter-city routes. A richer set of queries may be developed that allows users to subscribe to a notification system, so that they only need to send one SMS message to the server in order to receive several bus arrival time messages. Another potential expansion is to allow users to share tagged location data. Further investigation into making the system more scalable by using multiple phones per server for high density areas and by running stress tests is needed. Lastly, because the current system is based on SMS usage which generally requires a certain literacy level, further work could explore the possibility of voice interfaces.

Whether moving goods to market or people to medical clinics, transportation is a crucial resource for any community. Allowing individuals to find ways to maximize such shared resources is the goal of \*bus. Our approach has been a method of problem-solving that is bottom-up, that conceptualizes a technology device as shared among the community, and that recognizes the importance of information in people's everyday lives. Our goal with this paper has also been to bring to the fore the importance of transportation for both economic and human development – an area of ICT for development that has not yet received much attention.

#### ACKNOWLEDGMENTS

We would like to thank members of the original \*bus and MySMS teams of the 2007-2008 CSE477 at UW: Martin Hecko, Naasir Ramji, Jordan Walke, Brad Campbell, Ruibo Li, and David St. Hilaire. We would also like to thank CAICT team members Cynthia Putnam, Emma Rose, Rebecca Walton, Erica Johnson, Mark Licata, and also Professor Medina Aitieva, our local research team, and research participants. We thank Aidai Seidakmatova, Shairbek Juraev, and Natalie Linnell for their contributions. A special thanks to Bruce Hemingway and Jesse Doshier for the design and construction of the hardware.

- [1] G. Weisbrod, "Models to predict the economic development impact of transportation projects: historical experience and new applications," *The Annals of Regional Science*, Volume 42, Number 3 / September, 2008.
- [2] S. Démurger, "Infrastructure Development and Economic Growth: An Explanation for Regional Disparities in China?" *Journal of Comparative Economics*, Volume 29, Issue 1, March 2001, pp. 95-117.
- [3] D. Parikesit and K. Czuczman, "Transport, the missing link? A catalyst for achieving the MDGs," *id21insights*, #63, July 2006, <http://www.id21.org/insights/insights63/insights63.pdf>
- [4] "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\\_TransportsRole.pdf](http://siteresources.worldbank.org/INTWDR2004/Resources/22572_TransportsRole.pdf)
- [5] A. Bradbury, "Transport, mobility and social capital in developing countries," *Proc. of the Institution of Civil Engineers, Engineering Sustainability 159*, June 2006 Issue ES2, pp. 79-86.
- [6] NextBus, Last accessed: 22 Sept. 2008. <http://www.nextbus.com/>
- [7] Metro Tracker, Last accessed: 22 Sept. 2008. <http://transit.metrokc.gov/olttools/tracker.html>
- [8] MyBus, Last accessed: 22 Sept. 2008. <http://mybus.org/>
- [9] The World Factbook, CIA, Last accessed: 22 Sept. 2008. <https://www.cia.gov/library/publications/the-world-factbook/>
- [10] Pew Internet and American Life Project, "A typology of Information and Communication Technology Users", May 2006. Available: [http://www.pewinternet.org/PPF/r/213/report\\_display.asp](http://www.pewinternet.org/PPF/r/213/report_display.asp)
- [11] J. Donner, "Research Approaches to Mobile Use in the Developing World: A Review of the Literature". *The Information Society* 24(3), 2008.
- [12] "Improving Rural Mass Transit in South Africa - Team Smile," Last accessed: 22 Sept. 2008. <http://blogs.technet.com/unlimitedpotential/archive/2008/07/03/improving-rural-mass-transit-information-in-south-africa-team-smile.aspx>
- [13] MSR India SMS Toolkit, Last accessed: 22 Sept. 2008. <http://www.codeplex.com/sms toolkit>
- [14] FrontlineSMS. Company Website, Last accessed: 22 Sept. 2008. <http://www.frontlinesms.com/>
- [15] 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.
- [16] B. Kolko, E. Rose, and E. Johnson, "Communication as information-seeking: the case for mobile social software for developing regions," Proceedings of the 16th international conference on World Wide Web, 2007.
- [17] P. Javid and T. Parikh, "Augmenting Rural Supply Chains with a Location-Enhanced Mobile Information System," in *Information and Communication Technologies for Development 2006*, May 25-26, 2006, Berkeley, CA.
- [18] M. Chen et al., "Practical Metropolitan-Scale Positioning for GSM Phones," in *UbiComp 2006: Ubiquitous Computing*, Springer Berlin / Heidelberg, 2006.
- [19] MySMS, Last accessed: 22 Sept. 2008. <http://mysms.sourceforge.net/>
- [20] SMSLib, Last accessed: 22 Sept. 2008, <http://smslib.org/>
- [21] A. Karbassi and M. Barth, "Vehicle route prediction and time of arrival estimation techniques for improved transportation system management," in Proc. of *Intelligent Vehicles Symposium, 2003*. IEEE, June 2003, pp. 511-516.