A User-Flocksourced Bus Experiment in Dhaka: New Data Collection Technique with Smartphones

Albert Ching

Massachusetts Institute of Technology, Department of Urban Studies and Planning

77 Massachusetts Avenue, Room 10-401, Cambridge MA 02139, Email: mrching@mit.edu, Phone: 415 490 7389

P. Christopher Zegras*

Massachusetts Institute of Technology, Department of Urban Studies and Planning

77 Massachusetts Avenue, Room 10-403, Cambridge MA 02139, Email: czegras@mit.edu, Phone: 617 452 2433, Fax: 617 258 8081

Stephen Kennedy

Massachusetts Institute of Technology, Department of Urban Studies and Planning

77 Massachusetts Avenue, Room 10-401, Cambridge MA 02139, Email: kennedys@mit.edu, Phone: 678 471 8715

Muntasir Imran Mamun

Kewkradong Bangladesh

C4 Arambag Eastern Housing, Mirpur 7, Dhaka 1216, Bangladesh, Email: muntasir@gmail.com, Phone: +88 019 113 10275

*Corresponding Author
DRAFT November 2012
Do not cite without permissions

ABSTRACT

Flocksourcing, or guided crowdsourcing, is an experimental data collection technique where users become the sensors to generate a large amount of information that might improve a public service like public transport. In an ideal world, users would self-organize to collect data that would improve their own experience—but in a place like Dhaka, the megacapital of Bangladesh, the technology that enables users to become sensors is not (yet) widely available. In this paper, we test the viability of flocksourcing by co-developing and seeding those technologies with a local resident-led flock who targeted a segment of the largely owner-operated bus system in Dhaka. The results from a flock of eight's weeklong data collection efforts demonstrated that flocksourcing can be a viable data collection technique for generating big amounts of data even in some of the most challenging urban environments. While the demonstration helps lay the groundwork showing that the technique *can* be used, questions still remain on whether and how it will be used in a truly sustainable way.

Introduction

Today's 800 million cars will increase to 2-4 billion by 2050, with most growth in Asia, particularly India and China (Zachariadis, 2012). In Beijing, the vehicle fleet has grown by an average 13% per year since 1990 and, in the year before the recently instituted vehicle quota scheme, the fleet grew from 3.5 to 4 million (Wu et al, 2011); in Delhi, forecasts indicate a 500% increase in vehicles by between 1990 and 2030 (Hickman et al., 2011). The laundry list of problems associated with motorization is well known and increasingly plague Asian cities (ADB, 2003). The consequences disproportionately impact the urban poor, who tend to suffer the highest exposure to pollution and accidents and also experience declining public transport service levels as consumers shift to private modes which subsequently clog scarce roadspace. Globally, the transport sector already accounts for approximately 20% of anthropogenic global greenhouse gas emissions and, by most forecasts, will be the most rapidly growing future source (e.g., Timilsina and Shrestha, 2009).

DRAFT

Asia's apparently inevitable motorization stems from development: as income grows, so does demand for private motor vehicles, both cars and two-wheelers (Gakenheimer, 1999; Dargay, 2001), which Zhao (2012) refers to as an "Iron Law." Private vehicles offer functional *and* symbolic utility (Wu, et al., 1999), with the latter being particularly important in the early stages of motorization. Not only do private cars promise a better user experience—convenience, speed, and reliability—relative to the often crowded, hot, uncomfortable alternatives; they also represent an important status symbol and mechanism of middle class formation (Vasconcellos, 1997). Developing Asia has particularly high aspirations for vehicle ownership: a 2005 market research study measuring the relationship between current ownership and future intentions to purchase a vehicle, found the highest aspiration levels in China, Indonesia, India, Thailand, Korea, Hong Kong and the Philippines (Matsuoka, 2005).

Strategies for managing motorization fall into four basic categories: (1) expanding roadway capacity; (2) restraining fleet growth; (3) restricting vehicle use; (4) improving substitutes. How cities respond to motorization now may well dictate their physical and functional futures. Emphasizing (1), roadway supply, encourages motorization and reduces the attractiveness of alternatives; that is, the automobile-oriented response leads to "lock-in": automobile-dependent cities (Barter, 2004). Emphasizing (2), (3), and (4), restraining vehicle ownership and use and investing in public transport, can create an alternative to the auto-motorized city. Barter (2004) suggests that several high income Asian cities (e.g., Singapore, Hong Kong) intervened early to slow their motorization trajectories, which helped maintain strong public transport and a legacy of transit-oriented urban structures. While most large and rapidly developing Asian cities have yet to become "locked in" to automobile dependence, they are quickly approaching the point where their response to traffic saturation may put them firmly

on the auto-dependent path. As Barter (2004) suggests, avoiding auto-dependent entrenchment now is likely easier than reversing it later.

While a few rapidly developing cities, most notably Shanghai (vehicle ownership auctions) and more recently Beijing (vehicle quotas), have adopted policies to slow motorization, such approaches are the exception. Restricting car ownership and use anywhere remains a difficult proposition on a political and institutional capacity level. The more politically attractive option attempts to improve automobile substitutes, primarily focused on "hardware," like urban rail (e.g., Kuala Lumpur, Bangkok) and bus rapid transit (BRT) (e.g., Jakarta, Kunming). While fundamental, hardware alone will not suffice. Developing Asia needs to focus on both the hardware of new infrastructures and the "software" of system management. Towards this end, rapidly growing Asian cities are arriving at the critical historical juncture where they can choose, or not, a path of automobile dependency, with a new tool at hand: the mobile phone.

The mobile-driven mobility intelligence opportunity

The mobile phone has quickly penetrated nearly every corner of Asia at unprecedented speed. In places with the lowest motorization rates (and the greatest opportunity to change course from auto-dependency) like Bangladesh, Pakistan and India, mobile phones are currently 40 to *180 times* more ubiquitous than private motorized vehicles (Fig 2).

Mobile phones represent increasingly powerful tools for collecting, storing, processing, and disseminating data. At one end, even the most basic phones, widely in use in developing Asia today, are equipped with voice and text capabilities already enabling networks of knowledge spanning market prices for fish to emergency alerts for natural disasters. On the other end are smartphones, with data and Internet connectivity and a host of applications and sensors from high resolution cameras to global positioning systems (GPS). Already as powerful

DRAFT

as low-end computers, smartphones and tablets are expected to replace laptops/desktops for large portions of the developing world.

Mobile phones can generate layers of information which can be retrofitted on top of existing infrastructures, hard or soft, in good condition or bad, from roads to social networks. This intelligence layer promises to improve mobility infrastructure and service effectiveness, for users, operators, and regulators. Potential applications include:

(1) Marketing sustainable transport modes;
(2) Providing users with real-time, or almost real-time information that can improve service comfort, accessibility, and safety of shared vehicles;
(3) Providing service operators with real-time, or almost real-time information that can improve efficiency and profitability of shared fleets
(4) Enabling planners and regulators to better monitor and evaluate system performance, precisely and timely.

Landscape of mobiles + mobility in developing Asia

While the technology is increasingly available, mobile phones + mobility mash-ups (combinations) remain rare in developing urban Asia. Field work from July-August, 2011, interviewing entrepreneurs and other stakeholders in 11 South and Southeast Asian cities, revealed most relevant initiatives were focused on services for private cars and taxis (Fig. 3). The up-front time and money costs of such technologies might be leading towards applications for private vehicles (and perceived higher profitability possibilities) and/or in places with larger agglomerations of technical expertise, specifically Bengalaru (India). Fewer innovations appeared in less affluent cities, like Dhaka. Deploying the related technologies across contexts (even different cities in the same country) was not common due to the localized, customized

purpose of the original technology. Irrespective of place, the financial sustainability of the various endeavors remains unclear.

Based on this preliminary field research, we decided to focus on developing a mobile-powered intelligence system for Dhaka's heavily used, highly chaotic bus system, hoping that it could both spark innovation for buses in Dhaka and throughout the region.

Physical and Digital Context

Dhaka Has No Bus Map

Metropolitan Dhaka's population has more than doubled since 1991 and the city remains among the most rapidly growing on the planet, expected to reach 22-25 million people by 2020, which would make it the world's fourth largest (Brinkhoff, 2012; City Mayors, 2012). The city is the economic and political capital of Bangladesh, and somewhat more protected than the other low-lying parts of the densely populated, climatically vulnerable, country. Dhaka's high density and limited space for new transport infrastructure has produced high congestion levels, despite one of the lowest motorization rates in the world (Figure 2). Non-motorized travel dominates; based on the latest available data from 2004, 63% of trips in the city are by foot, 20% by cycle-rickshaw (an estimated 400,000 cycle rickshaws ply the city's streets), 10% by bus, 3% by car, 2% by auto-rickshaw, and 1% by bike (Hossain and Susilo, 2011). Only 2% of Dhaka households own bicycles (Rahman, 2008). Nearly 50% of accidents in Dhaka involve pedestrians (Maniruzzaman and Mitra, 2005); worse, pedestrians account for almost 90% of the traffic deaths in the city (Efroymson, 2012), double the rates in many other large developing Asian cities (Leather, et al., 2011). A limited study of transport infrastructure in the city revealed that only 37% of the observed roads had footpaths on both sides and more than half had no footpaths at all

DRAFT

(Efroymson, 2012).

Despite its high demand, bus service remains fragmented and loosely organized. The government operates less than 5% of the approximately 7,000 buses and minibuses in the city, with more than 60 companies of a range of ownership structures operating on an unknown number of total routes, with weak regulation of fares and service levels (Olsson and Thynell, 2004; Poole, 2011). With support from the Asian Development Bank, plans for a bus rapid transit (BRT) implementation are under development on a corridor with 45 different bus companies and including parts of at least 61 bus routes (ADB, 2012).

Six government agencies have a stake in regulating Dhaka's bus operations. The Dhaka Transport Coordination Board (DTCB) is the lead agency, tasked to plan and coordinate a long-term city transport strategy. The Bangladesh Road Transport Corporation (BRTC) operates a small share of the buses in the city, the Bangladesh Road Transport Authority (BRTA) regulates public transport vehicles, the Road Transport Committee (RTC) awards route permits, the Dhaka Metropolitan Police (DMP) enforces regulations, and the Dhaka City Corporation (DCC), which split into two bodies in late 2011, is responsible for overall municipal governance and administration (World Bank, 2009). The system suffers from poor quality buses, low service standards, high pollution, over-crowding, on-street competition for passengers, little adherence to schedules, and weak fare control – problems attributable to ineffective regulation and little accountability, causes which have proven to be resistant to political changes (World Bank, 2009).

For Dhaka's millions of bus riders, the results are alarming. Buses are involved in 25% of road accidents in Dhaka (Mannan and Karim, 1999) and reportedly are the major cause of pedestrian fatalities (World Bank, 2009). Vehicle over-crowding affects comfort, safety, security

(Katz and Rahman, 2010), and equity. Rahman (2008), for example, suggests that older adults, women and children often cannot board buses during rush hour, since these groups avoid having to fight the crowding. Enam and Chowdhury (2011) find women are less likely to consider public transport as a travel choice.

The government has tried to improve services and incentivize the private operators through competition, launching its aforementioned bus service in 2001. The BRTC recently put into service several hundred new, high capacity buses, with 300 buses featuring electronic ticketing expected in June 2012 and another 300 planned in the next year (BRTC, 2012). These efforts represent one step toward improving conditions and slowing the erosion of the passenger base. Can better information on Dhaka's buses help?

Despite the promise of limitless data epitomized by the smartphone, urban mobility information and knowledge remain relatively expensive, especially when attempting to measure millions of users on thousands of uncoordinated, shared moving vehicles. The alternative, slow and imprecise analog tools, may partly explains, why, for example, *Dhaka does not have an updated map of its complete bus system*.

Current Bus Data Collection Methods

As Figure 4 illustrates, bus data fall into four general categories: (1) space, (2) time, (3) conditions, and (4) perception. The digitalization of measurement tools, represented by the smartphone, signifies not only a theoretical improvement in *what* can be measured, but also raises new questions about *how* relevant measurements can be made and capitalized upon in new ways, given portability (of the devices and the data generated), precision, and speed of processing and dissemination. Digital sensors can capture objective measures to a level of accuracy that human senses simply cannot (e.g., thousandths of a second). Moreover, mobile

DRAFT

data are “real-time:” capable of being measured and shared almost simultaneously. Earlier digital technologies like video and voice recorders expanded the scope of what *could* be measured; more recent networked digital technologies may be the catalyst for redefining what *is* actually measured, constantly and consistently.

While buses in Dhaka are highly visible on the roads, they have relatively rarely been formally measured. Data collection efforts for buses in Dhaka (and likely in much of the rest of the developing world) can be characterized, to date, as analog and ad hoc. Analog means collected using human observation and, essentially, pen and paper. Since these analog methods tend to be complex and slow, and in competition for scarce resources in a challenging institutional setting, they tend to be ad hoc, or done much less frequently than necessary to remain up-to-date. Dhaka has seen efforts to merge analog and digital approaches. For example, Hasnine (2011) and Katz and Rahman (2010) have attempted to characterize the bus system in Dhaka by supplementing analog tools with digital voice and video recorders to more precisely measure crowding and bus wait times. These studies offer valuable glimpses into the system, yet they remain rare.

In more industrialized contexts, bus data collection methods increasingly employ digital tools resulting in more data actually collected. Automatic-vehicle-location (AVL) technologies that enable real-time tracking of vehicle fleets, through GPS and signpost transmitters, have become common. Originally deployed for fleet control, and operations management, AVL is now increasingly used to generate information services (e.g., bus prediction times) for users. In the USA, urban public transportation systems, primarily owned and operator by the public sector, started providing AVL-based information to users (message boards, web sites, etc.) in the early 2000s. By the late 2000s, some transit agencies began piloting an “open” approach to real-time

data dissemination: publicly releasing bus positions and predictions to enable third party developers to produce smartphone applications (apps) for users. Under this overall model, the agencies are responsible for transforming the data for public use, ensuring data quality, and building the interfaces for public data access. App developers can access these data and transform them for "consumption" by users; these apps are often provided for free or at very low cost (Rojas, 2012). In other settings, transit information remains much more centrally controlled; sometimes this can be due to data "ownership" questions (e.g., in privately operated systems), concerns over regulatory consequences (revealing poor performance), or other worries. Data costs themselves may pose a barrier. AVLs, inclusive of capital, operating and contracting costs, can cost public transit agencies in the USA between $15,000-$60,000 per bus (Parker, 2008), although these costs may be declining. High data costs may decrease the likelihood of sharing and inhibit third party data generation.

Might the smartphone change this technology landscape and provide an opportunity for developing cities to "leapfrog" traditional (i.e., AVL) approaches, just as has happened with mobile telephony more generally (James, 2009)? For one, smartphones, equipped with the same or even better time and location sensors as traditional AVL technology, can potentially be utilized as portable AVL sensors at a fraction of the cost. Importantly, and distinctly from traditional "passive" AVL sensing technologies, smartphones can also be relatively easily converted into counting and surveying devices, enabling the simultaneous collection of data on service dimensions, like crowding, and user characteristics, like demographics and satisfaction. Such smartphone-based surveys carry the added benefit of having precise time and place "stamps."

DRAFT

Furthermore, by potentially putting data collection capabilities into the hands of users themselves, smartphones may lay the foundation for truly intelligent cities. Thomas Erickson (2010), from IBM Watson Research Center, suggests that "tapping human intelligence" and enabling users to sense, analyze and act on urban systems themselves, will increase the likelihood of creating "legitimate" smart cities. The expected increased penetration of smartphones in most developing urban contexts provides an interesting opportunity to not only much more scalably, quickly, accurately, and precisely collect data, but also do so in a manner that keeps the data accessible to the public. Can smartphone-equipped users become the sensors who collect data that improve their own user experience and thus obviate the need for operators or the government to step in? Might crowdsourcing offer the right approach?

The Bus Crowdsourcing Challenge

Despite its seeming ubiquity, crowdsourcing is a recently coined term (Howe, 2006) referring to a "participative online activity" where a "heterogeneous" group of people "voluntarily" undertake a task of "mutual benefit" (Estellés and González, 2012). Promising cost-effectiveness and creativity from its distributed crowd, crowdsourcing is an appropriate data collection method when the scale of the data collection effort can match the scale of data required to be useful. The scale depends on: (1) the size of the potential pool of participants, (2) participants' ease in making a contribution, and (3) the compelling nature of the mission. Perhaps the most successful example of crowdsourcing to date is Wikipedia, which leverages a boundary-less set of volunteers to build a free encyclopedia on the web, at their convenience.

Note that crowdsourcing emerged from a time when desktop computers were the primary data input (and output) devices and precise time- and space-information was still nascent. Geocentric crowdsourcing, or crowdsourcing centered around a specific location, did generate

early applications, such as mapping bicycle lanes (e.g. Cyclopath) and potholes (e.g. FixMyStreet). The proliferation of smartphones and other location-sensing devices, makes geocentric crowdsourcing potentially easier, more precise, and "real-time" – akin to what Robinson et al (2012) refer to as "street computing." A user could geo-trace a bus route map while actually riding a bus; travel times could be measured precisely and communicated instantaneously to other riders.

While precise, 'real-time' geocentric crowdsourcing has become possible with smartphones, generating the most useful information on a bus system stretches the limits of the crowdsourcing model. First, crowdsourcing bus data with smartphones in the field likely requires more effort from the participant than contributing via the comfort of one's home and one's time. Second, the number of potential volunteers who can collect data is limited by space and time (imagine a low-frequency bus line), thus increasing the need for "altruism" among potential members. Thirdly, getting transit riders to be "altruistic" may be a significant hurdle. For example, in perhaps the only documented experiment on crowdsourcing bus data, Steinfeld et al. (2012) developed and deployed smartphone apps on the Pittsburgh (Pennsylvania, USA) bus system to capture real-time bus locations, crowding, and service disruptions. While the technology worked, they found it particularly difficult to motivate transit riders to gather data, especially in the quantities needed to be relevant to users. Finally, in the developing country context, the relative cost of smartphones and their low penetration rates (especially among transit users) remain as barriers.

In theory, then, the advent of smartphones and other location-sensing technologies has markedly expanded the potential to provide precise time- and space-specific information to:

- users, such as service alerts, wait and travel times and crowding levels;

DRAFT

- operators, such as vehicle positions and speeds to help coordinate services and adapt in real-time to emergencies;
- regulators, such as frequencies and stops; and
- planners, such as user demographics, origins/destinations, and satisfaction levels.

The *bus crowdsourcing challenge*, therefore, is not just motivating a sufficient number of participants to collect data on buses, but also collecting the right data at the right time. Generating the most valuable information would require extremely broad coverage or highly targeted coverage, both likely beyond the reach of traditional crowdsourcing efforts.

Interestingly, crowdsourcing arose as a reaction to outsourcing (Howe, 2006), which tasks assignments to a specific group rather than relying on the altruism of a distributed set. In the developing world with relatively inexpensive labor and non-widespread smartphone technology, might another data collection model work, that combines the power of a highly motivated and coordinated group (outsourcing) and a distributed means of contributing information (crowdsourcing)? At least two approaches, distinct from crowdsourcing, exist for using smartphones to increase the amount of publicly available information on Dhaka's buses, which we term: "fleetsourcing" and "flocksourcing" (Fig. 5).

Fleetsourcing

Fleetsourcing is a data collection approach that simply pairs any fleet of shared vehicles with smartphones (i.e., smartphone AVL). Since the device is not built into the buses themselves, one potentially viable approach would be to utilize the assistance of the onboard ticket agent, an employee still common in some systems in the developing world today, including Dhaka. Whereas traditional AVLs are programmed to passively collect data (on time and place),

smartphone AVLs can also be input devices, for collecting additional information like ridership, as well as crowding, traffic conditions, and roadside or other emergencies. Combined with QR codes, the phones could even become alternative fare collection technologies (removing cash from the transaction). While fleetsourcing would likely be significantly cheaper than traditional AVLs, in a place like Dhaka it would remain a challenge since bus operators have little incentive (or financing) to make the necessary investment, which would include: convincing 45+ different bus companies to buy into the idea, procuring and maintaining 6,500 smartphones, and training at least 6,500 ticket agents, who likely have not used a smartphone before. It may be possible to proceed incrementally, with one or two larger-scale operators piloting an approach.

Flocksourcing

As an alternative to fleetsourcing and crowdsourcing, we propose "flocksourcing," an evolution of the crowdsourcing and outsourcing concepts, that attempts to organize and motivate a specific group of participants (a "flock," or guided crowd) to collect data in a distributed way for a specific purpose. Flocksourcing attempts to address the bus crowdsourcing dilemma in developing city contexts by (1) relying more on face-to-face contact and less on online altruism to motivate and organize participants and (2) endowing the technology and the "know-how" to the flock members. Whereas traditional crowdsourcing relies on a pool of largely anonymous online volunteers, flocksourcing (at least in our Dhaka experiment) incorporates social organization to set targets and hold a (usually smaller) group of individuals accountable. This approach adheres to the observation that, despite the increasing popularity of online collaboration, humans are in fact much better at collective action in person due to the power of subtle social cues, which remain difficult to mimic online (Erickson and Kellogg, 2000). While endowing the technology is not a requirement for flocksourcing, in our application it is an

DRAFT

important element, due to the relative scarcity of the most advanced data collection devices. For some information purposes like providing a bus route map for users, successful flocksourcing should be reasonably straightforward, as the need for time- and space-specific details is modest. For others, like period-specific average travel times, the size of the flock and/or the intensity of its deployment would need to be greater (various times of days covered, various days) to ensure accuracy. In Dhaka, travel times vary considerably (how much is unclear due to a lack of data, but Hasnine (2011) provides some indication), not only due to peak morning and evening travel periods but also because different sections of the city shut down depending on the day, in order to minimize congestion. In theory, well-designed flocksourcing could be combined with machine learning to help minimize the amount of data necessary to extract reasonably complete and accurate pictures of the system. Ultimately, the intended product of flocksourcing is not just a large, concentrated sample of data but, perhaps as importantly, a trained and trusted data collection flock. For example, flocks trained to collect data on buses may also be deployed to rapidly collect data during disasters (provided the mobile networks are operational).

The Flocksourcing Experiment

Testing the viability of flocksourcing as a technique for collecting data on the Dhaka bus system required: (1) assuring the technical feasibility of the technology, including accuracy of location sensing, speed of mobile data network, and limitations of locally available hardware; (2) designing the data collection approach to be affordable, understandable, and deployable on Dhaka's buses; (3) understanding the scalability of the data collection system, in terms of the information a flock can generate and its representativeness.

Method

Our flocksourcing experiment was the culmination of 9 months of collaborative research, including an intensive month of technological iteration in January, 2012, and a week of technology deployment in late March, 2012. The local partner in the experiment, Kewkradong Bangladesh, is an adventure-based non-profit organization. The close-knit group of 15-20 volunteers describes themselves informally as "not an NGO or a club where you can join by paying membership dues" but as "out-of-box" thinkers who frequently push the boundaries of what is possible in their city and beyond.

Since smartphones are still relatively new to the Dhaka market and Kewkradong did not have easy access to them, we locally procured a limited set of 25 devices for our pilot. We developed a range of initial mobility-related apps, from vehicle counters to trip trackers to paperless surveys, and these were then tested together in January with Kewkradong at 18 busy intersections across 6 neighborhoods in Dhaka to understand both the technical possibilities and the cultural context. After a week and a half of field testing, we conducted a charette with local university students to devise potential mobile and non-mobile related solutions for the mobility challenge in Dhaka. The result was a decision to focus on buses, due to their city-wide reach, crowding and slowness. An initial app (along with the data collection, analysis and visualization back-end) was designed and then revised about five times in the field and then another 6 to 7 times back in the USA in February, 2012.

The final pilot version of the bus data collection app, called "Share My Bus Dhaka," aimed to test the range of possible metrics that smartphones could collect on buses in Dhaka today, including: (1) bus positions (space and time), (2) bus crowding (conditions), and (3) user demographics and satisfaction of the bus riding experience (perception). We also devised a

DRAFT

cloud-based back end technology to monitor the data that were collected in real-time. The objective was to explore the viability of flocksourcing and the technical feasibility of fleetsourcing. Kewkradong organized a small flock of eight individuals to ride buses for a week in March 2012, focusing on two critical bus lines in the city. The routes were chosen because they were among the most frequently used in the city according to our local partners and third-party sources (Olsson and Thynell, 2004).

Technical feasibility

The experiment needed to test three key technical elements in Dhaka: (1) the accuracy of the location sensing, (2) the speed of the mobile data network, and (3) any performance limitations of locally-available hardware, specifically the screen size and resolution (since the touch screen is the input mechanism for some measurements).

Location Accuracy

Although we did not do a detailed analysis of location sensing accuracy throughout Dhaka, on the two routes we targeted the accuracy of sensed bus locations seemed largely to be within 25 meters of the roads (Fig. 6). At some spots, near the new airport, the sensed locations were quite far from the roads on which the buses traveled. Despite these ranges, for the purposes of tracing bus locations on fixed routes, this level of accuracy seemed sufficient; however, for more flexible route services like on-demand cycle or auto-rickshaws, this level of accuracy may need to be improved.

Mobile Upload Speed

Mobile data upload speeds are a function of network speed and the amount of data being sent.

Despite the relatively competitive mobile market in Dhaka, the current mobile data networks operate on 2 to 2.5G, which is a generation or two behind the most advanced network contexts (e.g., Singapore) beginning to operate 4G. While 2-2.5G might indicate prohibitively slow data speeds, the data packets sent in the pilot were quite small, just a few lines of text. While we had no precise way of measuring the average upload speed, each upload seemed to average around 3 seconds with speeds varying by location and network provider (based on hundreds of field observations). The upload speed mattered because users could not input other data (e.g., passenger counts), while data were being sent. We could not adjust the upload speed, but could adjust the frequency of bus location updates. For our initial pilot, sending location data every 60 seconds, similar to the upload frequency of the Pittsburgh experiment (Steinfeld et al., 2012), seemed to strike the right balance between allowing time to input data and sending enough bus position data to infer metrics like speed.

Local Hardware

Due to a variety of reasons, from cost to usability to scalability in the future, our applications were developed on phones that could be purchased locally. At the time of the January 2012 pilot, the best available Android-powered phone was the Samsung Galaxy Y, designed specifically for the South Asian market, with lower price points (13,000 Taka or $175 USD inclusive of taxes) and a slightly less advanced feature set. These phones were launched in Dhaka a few weeks before the pilot. The main challenge with the phone was the smaller screen, about 25% smaller than the one used to develop the original pilot apps. This meant a smaller touch interface, an important consideration given the high-intensity environments within which users may be entering information. Ultimately, we were able to design for the smaller screen size and found that the other specs, from location sensing to data upload speed, worked comparably

DRAFT

across phones.

Appropriate Design

For our experimental context, appropriate design meant: (1) minimize costs by leveraging free and open global tools, (2) simplify the user interaction by using icons rather than text, and (3) streamline the number of steps required for a user to start collecting data.

Affordable

Minimizing costs required leveraging resources that are relatively cheap in Dhaka—mobile data networks and labor—and open and free tools that are developed for global audiences. The cost structure required upfront investment for hardware, but minimum ongoing investment. For example, the only real ongoing costs were for mobile data, which currently costs about $4 for every GB of data sent, and the cost of each flockmember, which can be as low as $10-$15 per day per person in Dhaka. By comparison, the average GDP per capita in Bangladesh hovers around $2 per day. Otherwise, we used open and free tools, made available in our case by Google and MIT, to develop mobile apps (MIT App Inventor), store data in the clouds and analyze data in real-time (Google fusion tables and Google Maps).

Understandable

One of the biggest challenges for many in the developing world to fully access content on the Internet is low literacy rates, especially in English, since most content on the web and applications are developed first in English (Farrell, et al., 2012). In Bangladesh only 2 percent of the population speaks English (Crystal, 2003). To maximize understandability, we used icon-based design, made easier today due to open online resources like the Noun Project

(thenounproject.com). Using icons resulted in a user-interface (Fig. 7) that required, in our final deployment, minimal training (3 hours) of local users.

Easy to deploy

Perhaps the most difficult task was designing the mobile applications to be usable in the context of riding a bus in Dhaka, which can be quite a chaotic experience. This includes running to catch the bus, struggling for a seat, paying while aboard, shouting for stops, and then alighting safely, often from a still-moving vehicle. In such a high stress context for data collectors, every small step that could be eliminated means fewer chances for error during the data collection process and more data that could be collected. For robustness, we employed techniques like (1) persistent data, such as storing previously used information like user login name and bus number and (2) strictly limiting the onboard survey to nine questions that could be flipped through on one screen.

A Thousand Riders Surveyed

Our original goal was modest: identify the deployment scale required to develop an accurate portrait of the two routes' primary operating characteristics, including speeds, crowding, adherence to routes and their variations across times and days. We also wanted to test the viability of using the technology for carrying out on-board surveys. The plan was for the eight flock members to cover the routes for a week, aiming to achieve a minimum of 120 one-way vehicle rides, in total, and 100 on-board surveys. The flock ultimately carried out 270 rides, recording over 10,000 passenger counts and bus location points and surveying over 1,000 on-board users. Three flocksourcers rode the bus for over 2,000 minutes over the week, equivalent to over 33 hours.

DRAFT

That the flock collected any data was somewhat miraculous. The flocksourcing pilot was delayed by a week due to a violent *hartal*, or political protest, that banned all motorized vehicles, including buses, from the road. The political uncertainty underscored some of the inherent risks that bus commuters, and potential flocksourcers, in Dhaka take. One of the advantages of mobile-driven, real-time data collection techniques is the ability for almost instantaneous monitoring, from anywhere with an Internet connection. In our particular application, with the help of a few online dashboards, the data collection efforts happening in Dhaka could be monitored from 12,500 kilometers away, facilitating quick trans-global feedback. A related benefit is accountability. Location data are difficult to falsify and, with the proper tools, easy to audit. The flock organization, routes ridden, and individual performance could be easily seen (Fig. 8). The location data, linked to surveys and passenger counts, add another layer of information that can help verify and validate data collection and identify anomalies.

The following sections provide a brief first look at the potential revealed by some of the data collected during the experiment. While showing the promise of the approach, these figures should not be extrapolated to represent the entire bus system of Dhaka, or of even the two principal bus lines that were targeted. However, they give a sense of the level of precision that this approach may offer and how the data could be used by various stakeholders in the future.

Geo-coded Bus Routes

As far as we can determine, no existing map exists for Dhaka's extensive bus system (Mydigonto.com has a crude, seemingly incorrect, bus route planner for Dhaka). Bus riders in Dhaka rely instead on a powerful offline social network (people on the street and other riders) for bus information. While our pilot focused on two main lines, No. 36 from Azimpur to Pallabi and No. 27 from Azimpur to Uttara, one can imagine a broader effort to flocksource all the lines in

Dhaka including less well-known routes on the outskirts of the city and the more flexible routes of mini-buses and even cycle rickshaws. The upshot could be the first digital bus map of Dhaka that, layered with predicted travel times, themselves flocksourced, could be of immediate help to users.

Bus (and Potentially Road) Speeds

Despite seemingly pervasively perpetual congestion levels, knowing hourly, daily and weekly variations in travel time can be valuable for short, medium and long term operations and planning interventions. Smartphone-based AVLs on buses may be one way to provide that level of precision in 'real-time' (Fig. 10). If adopted widely, fleetsourced buses could provide traffic data for the major roads in the city, since buses are ubiquitous on those roads. Understanding bus (and potentially road) speeds on a daily basis may help independent operators better time and even coordinate their deployments appropriately. Monitoring road speeds on a longer timescale may help regulators target experiments to improve speeds at key bottlenecks during peak travel times.

Bus Travel Times

For bus riders in Dhaka, the extreme variation in travel times can be a frustration and inconvenience. Similar to speeds, travel times vary by time of day, day of week and route through the city, especially since different parts of the city are closed on different days of the week to alleviate traffic. Our pilot recorded travel times (not including waiting time) on a 20-km commute from Azimpur to Uttara ranging from 43 minutes to 2 hours and 7 minutes (Fig. 11). Knowing even the maximum and minimum travel times may help improve user satisfaction.

DRAFT

Bus Rider Demographics and Commutes

The flock conducted a limited survey of 1,014 riders, not as a probabilistic sample, rather to demonstrate the technological possibilities. Understanding the demographics, trip purposes and characteristics, satisfaction levels, and so on, are fundamental to bus service planning and can help operators and regulators better develop and market services. The more than 1,000 users surveyed represented a user group that is young, male, bus captive, connected and averse to crowding: 85% were between 24-34 years old, 84% were male, 57% ride the bus at least 5 times a week, 100% had a mobile phone (18%, smartphone; 50%, internet-enabled phone), and 23% listed over-crowding as most common complaint. The longest commute measured on this bus line was a 40 km, 2.5 hour one-way journey between Gazipur, a bus stand north of Dhaka, and Azimpur.

Bus Size, Happiness, Crowding, and Female Ridership

Understanding how to keep current bus riders happy may be *the* most valuable use of flocksourced data. An interesting apparent correlation emerged in our sample of data: between user happiness and crowding, a dynamic studied by Katz and Rahman (2010). The flocksourcing data enabled detailed forensics: user happiness (recorded by the flock) seemed to decline when crowding was at its highest, although we found variation across specific companies (the government-owned fleet of bigger, more modern buses, with fewer people standing demonstrated higher happiness levels than the others). BRTC buses also seemed to have a higher share of female riders, possibly associated with crowding levels (Fig. 12).

Conclusion

We proposed and demonstrate Flocksourcing as a technique to rapidly collect bus data in a challenging urban environment like Dhaka. In contrast to existing analog tools that are usually employed in Dhaka, digital smartphones paired with human sensors can quickly and scalably collect multidimensional layers of data in precise detail. The volume and range of data collected in this pilot are promising, especially in a data-sparse environment.

Nonetheless, with the technological feasibility demonstrated, the larger question is towards what end might flocksourcing be technologically appropriate and useful? In the short-term, low smartphone penetration rates may make pure citizen-led efforts prohibitively expensive. We can envision several possibilities of how flocksourcing (and its sister technique fleetsourcing) can be utilized by various stakeholders in Dhaka in the near future (Table 1). In subsequent work we anticipate demonstrating the socio-technical implications of pursuing one or more of these, in Dhaka and/or elsewhere, where a less-motorized path to development may be enabled by mobile-based mobility.

DRAFT

Bibliography

S. Archaya and S. Morichi, "Motorization and Role of Mass Rapid Transit in East Asian Megacities," *IATSS Research* 31:2 (2007), 6-16.

Asian Development Bank (ADB), *Reducing Vehicle Emissions in Asia: Policy Guidelines for Reducing Vehicle Emissions in Asia: A Set of Six Policy Guidelines* (Manila) 2003.

Asian Development Bank (ADB), *Sector Assessment (Summary): Urban Transport*. (Manila: Greater Dhaka Sustainable Urban Transport Project (RRP BAN 42169), 2012) <http://www.adb.org/sites/default/files/linked-docs/42169-013-ban-ssa.pdf> Accessed November 27 2012.

Bangladesh Road Transport Corporation (BRTC), "BRTC added 530 CNG-run AC/Non AC buses from China and Korea for city service under Electronic Ticketing Service," *Latest*

News<http://www.brtc.gov.bd/index.php?option=com_content&view=article&id=110&catid=1&Itemid=57> (BRTC Website, 2012) Accessed November 27 2012.

P. Barter, "Transport, Urban Structure and 'Lock-in' in the Kuala Lumpur Metropolitan Area," *International Development Planning Review* 26: 1 (2004) 1-24.

T. Brinkhoff, "Bangladesh: Dhaka," *City Population* (2012) <http://www.citypopulation.de/php/bangladesh-dhaka.php> Accessed November 27 2012.

City Mayors, *Statistics* (2012) <http://www.citymayors.com/statistics/urban_2020_1.html> Accessed November 27 2012.

D. Crystal, *The Cambridge Encyclopedia of the English Language* (Second ed.) (Cambridge, UK: Cambridge University Press, 2003).

J. Dargay, "The Effect of Income on Car Ownership: Evidence of Asymmetry," *Transportation Research Part A* 35 (2001) 807-821.

D. Efroymson, *Sustainable Urban Transport: Dhaka Bus Rapid Transit Walkability Strategy* (Manila: Asian Development Bank, Technical Assistance Consultant's Report, 2012) <http://www.adb.org/sites/default/files/projdocs/2012/39335-012-reg-tacr-01.pdf> Accessed November 27 2012.

A. Enam and C. Chowdhury, "Methodological Issues in Developing Mode Choice Models for Dhaka City," *Journal of Transportation Research Board* (2011) 84-92.

T. Erickson, "Geocentric Crowdsourcing and Smarter Cities: Enabling Urban Intelligence in Cities and Regions," position paper for the 1st International workshop on ubiquitous crowdsourcing (Copenhagen: UbiComp '10, September 26-29, 2010).

T. Erickson and W. Kellogg, "Social Translucence: An Approach to Designing Systems that Mesh with Social Processes," *Transactions on Computer-Human Interaction* 7:1 (2000) 59-83.

A. Estellés and L. González, "Towards an Integrated Crowdsourcing Definition," *Journal of Information Science* (2012, in press).

R. Farrell, C. Danis, T. Erickson, J. Ellis, J. Christensen, M. Bailey, W. Kellogg, "A Picture and a Thousand Words: Visual Scaffolding for Mobile Communication in Developing Regions," *International Journal of Handheld Computing Research* 1:4 (2010) 81-95.

R. Gakenheimer, “Urban Mobility in the Developing World,” *Transportation Research Part A* 33 (1999) 671-689.

M. Hasnine, “Evaluation and Development of Bus Based Public Transport in Dhaka City,” paper presented at 4th Annual Paper Meet and 1st Civil Engineering Congress (Dhaka, December 22-24, 2011).

R. Hickman, O. Ashiru, and D. Bannister, “Transitions to Low Carbon Transport Futures. Strategic Conversations from London and Delhi,” *Journal of Transport Geography* 19: 6 (2011) 1553–1562.

M. Hossain and Y. Susilo, “Rickshaw Use and Social Impacts in Dhaka, Bangladesh,” *Transportation Research Record 2239* (2011) 74-83.

J. Howe, “The Rise of Crowdsourcing,” *Wired* 14:6 (June 2006) <http://www.wired.com/wired/archive/14.06/crowds.html> Accessed December 3 2012.

J. James, “Leapfrogging in Mobile Telephony: A Measure for Comparing Country Performance,” *Technological Forecasting & Social Change* 76 (2009) 991–998.

D. Katz and M. Rahman, “Levels of Overcrowding in Bus System of Dhaka, Bangladesh,” *Transport Research Record 2143*, (2010) 85-91.

J. Leather, H. Fabian, S. Gota, A. Mejia, *Walkability and Pedestrian Facilities in Asian Cities: State and Issues* (Manila: Asian Development Bank, ADB Sustainable Development Working Paper Series, Feb 2011) <http://www.adb.org/sites/default/files/pub/2011/ADB-WP17-Walkability-Pedestrian-Facilities-Asian-Cities.pdf> Accessed November 27 2012.

M. Mannan and M. Karim, “Road Accidents in Metropolitan Dhaka, Bangladesh,” *IATSS Research* 23:2 (1999) 90-98.

K.M. Maniruzzaman and R. Mitra, “Road Accidents in Bangladesh,” *IATSS Research* 29:2 (2005) 71-73.

M. Matsuoka, “Asia Leads Global Car Ownership Aspirations,” (ACNielsen International Research, 2005) <http://kr.en.nielsen.com/pubs/2005_q1_ap_car.shtml> Accessed November 27 2012.

L. Olsson and M. Thynell, *Bangladesh Road Transport Corporation (BRTC) Bus Project in Dhaka* (Stockholm: Swedish International Development Cooperation Agency, Department for Infrastructure and Economic Cooperation, Sida Evaluation 06/38, 2004).

DRAFT

D. Parker, *AVL Systems for Bus Transit: Update A Synthesis of Transit Practice* (Washington, DC: Transit Cooperative Research Program, 2008).

A. Poole, *How-To Notes: Political Economy Assessments at Sector and Project Levels* (Washington, DC: World Bank: Public Sector and Governance Group, March 2011) <http://www.gsdrc.org/docs/open/PE1.pdf> Accessed December 3, 2012.

M.S. Rahman, "Understanding the Linkages of Travel Behavior with Socioeconomic Characteristics and Spatial Environments in Dhaka City and Urban Transport Policy Applications," (Hiroshima: Graduate School for International Development and Cooperation, Hiroshima University, Master's Thesis, Sept. 2008) <http://sr-milan.tripod.com/Master_Thesis.pdf> Accessed November 27 2012.

R. Robinson, R. Rittenbruch, M. Foth, D. Filonik, and S. Viller, "Street Computing: Towards an Integrated Open Data Application Programming Interface (API) for Cities," *Journal of Urban Technology* 19:2 (2012) 1–23.

F. Rojas, *Transit Transparency: Effective Disclosure through Open Data*. Report of Transparency Policy Project (Cambridge: Taubman Center for State and Local Government, Harvard Kennedy School, June 2012) <http://www.transparencypolicy.net/assets/FINAL_UTC_TransitTransparency_8%2028%202012.pdf> Accessed November 27 2012.

A. Steinfeld, J. Zimmerman, A. Tomasic, D. Yoo, R. Dar Aziz, "Mobile Transit Rider Information Via Universal Design and Crowdsourcing," *Transportation Research Record 2217* (2012) 95-102.

G. Timilsina and A. Shrestha, "Transport Sector CO2 Emissions Growth in Asia: Underlying Factors and Policy Options," *Energy Policy* 37 (2009) 4523-4539.

United Nations (UN), *World Urbanization Prospects: The 2003 Revision* (New York: United Nations Department of Economic and Social Affairs, Population Division 2004) <http://www.un.org/esa/population/publications/wup2003/WUP2003Report.pdf> Accessed November 27 2012.

E. Vasconcellos, "The making of the middle-class city: transportation policy in Sao Paulo," *Environment and Planning A* 29:2 (1997) 293-310.

World Bank, "Operationalizing Political Economy: Urban Bus Operations in Dhaka," *South Asia Political Economy and Governance Issues Note No. 1* (Washington, DC: World Bank,

July 2009) <http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2012/06/07/000425962_20120607110827/Rendered/PDF/693780BRI0P1080Bus0Operations0final.pdf> Accessed 3 December 2012.

World Bank, *World Development Indicators and Global Development Finance* (Washington, DC: World DataBank, 2012) <http://databank.worldbank.org/Data/Home.aspx> Accessed November 27 2012.

Y. Wu, R. Wang, Y. Zhou, B. Lin, L. Fu, K. He, J. Hao, "On-Road Vehicle Emission Control in Beijing: Past, Present and Future," *Environmental Science & Technology* 45:1 (2011) 147-153.

G. Wu, T. Yamamoto, R. Kitamura, "Vehicle Ownership Model That Incorporates the Causal Structure Underlying Attitudes Toward Vehicle Ownership," *Transportation Research Record 1676* (1999) 61-67.

T. Zachariadis, "Introduction," Cars and Carbon: Automobiles and European Climate Policy in a Global Context Zachariadis, Ed. (Springer, Dordrecht, 2012) 1-18.

J. Zhao, "Transport—Maker or Breaker of the Cities," lecture in APSC 364 Applied Sustainability, University of British Columbia (Vancouver, March 6, 2012) <http://www.ires.ubc.ca/files/2012/01/Zhao-Transport-Intro-mar6.pdf> Accessed November 27 2012.

DRAFT

Table 1: Summary of potential bus flocksourcing / fleetsourcing products in Dhaka

Bus-info Product	Option A **Bus Route Map**	Option B **Smartphone AVL**	Option C **Bus Rider Surveys**
Key User	Riders	Operators	Regulators
Data collection technique	Flocksourcing (Flock size from 5-30 members)	Fleetsourcing (Fleet size from 1,000-5,000)	Flocksourcing (Flock size from 5-30 members)
Data collection frequency (suggested)	Every 6 months	Everyday	Every 3 months
Metrics collected	Geo-coded Routes, Stops (if any); (Travel times, Crowding)	Live Bus Locations	Satisfaction, Home and Work Locations, Demographics
Approximate Cost	$1000 for base map; $15,000 to include travel times and crowding	$150,000 annually for 1,000 BRTC buses; $750,000+ annually for 4,900 buses owned by Top 45 companies	$1300 for 1,000 survey sample ($600 for subsequent surveys); $9000 for 10,000 survey sample ($6,000 for subsequent surveys)
Time to collect	1 - 4 weeks	N/A	1 - 4 weeks
Add'l benefits	Consumer marketing for buses	Live wait times, traffic (if open); Wi-Fi hot spot	Baseline measurements for experimentation
Outstanding questions	Location Accuracy; Analog or Digital Output	Power + Battery Life; Training	Flock Bias

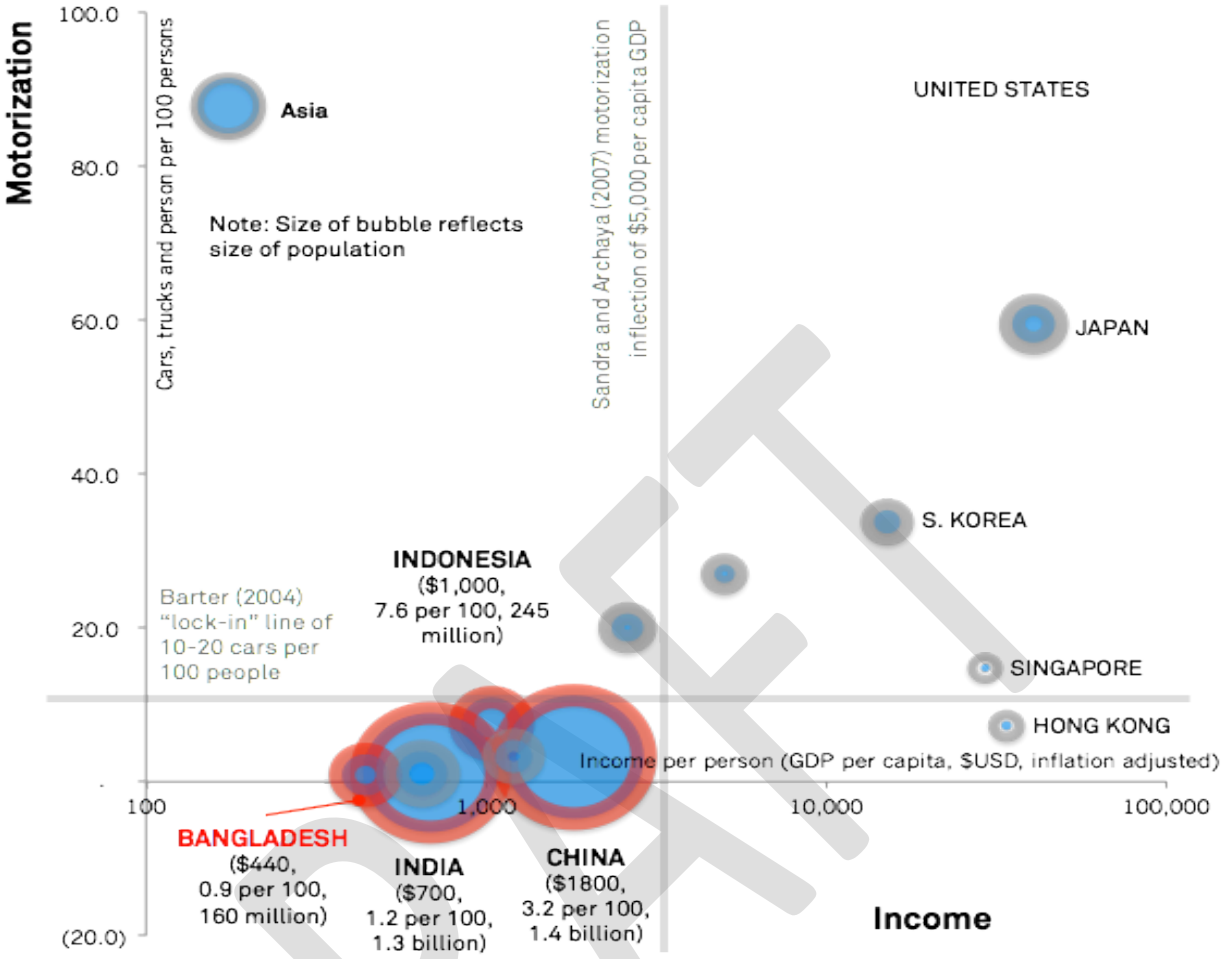

Figure 1: Motorization's "Iron Law," "Lock In," and Developing Asia

Sources: Archaya and Morichi, 2007; UN, 2004.

DRAFT

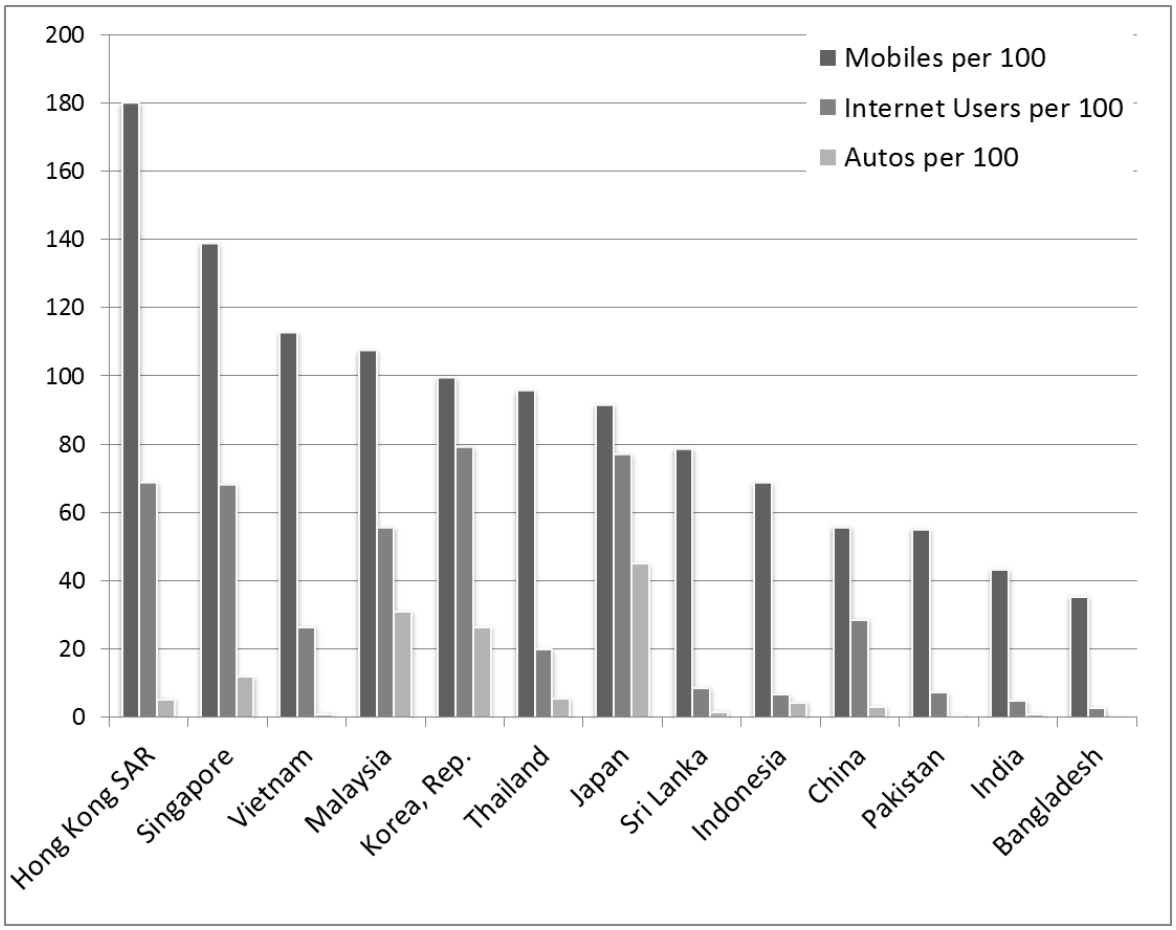

Figure 2: Mobile phone and internet penetration and motorization rates across Asia (2009)

Sources/Notes: World Bank, 2012; passenger cars in Thailand for 2006 and Vietnam for 2007.

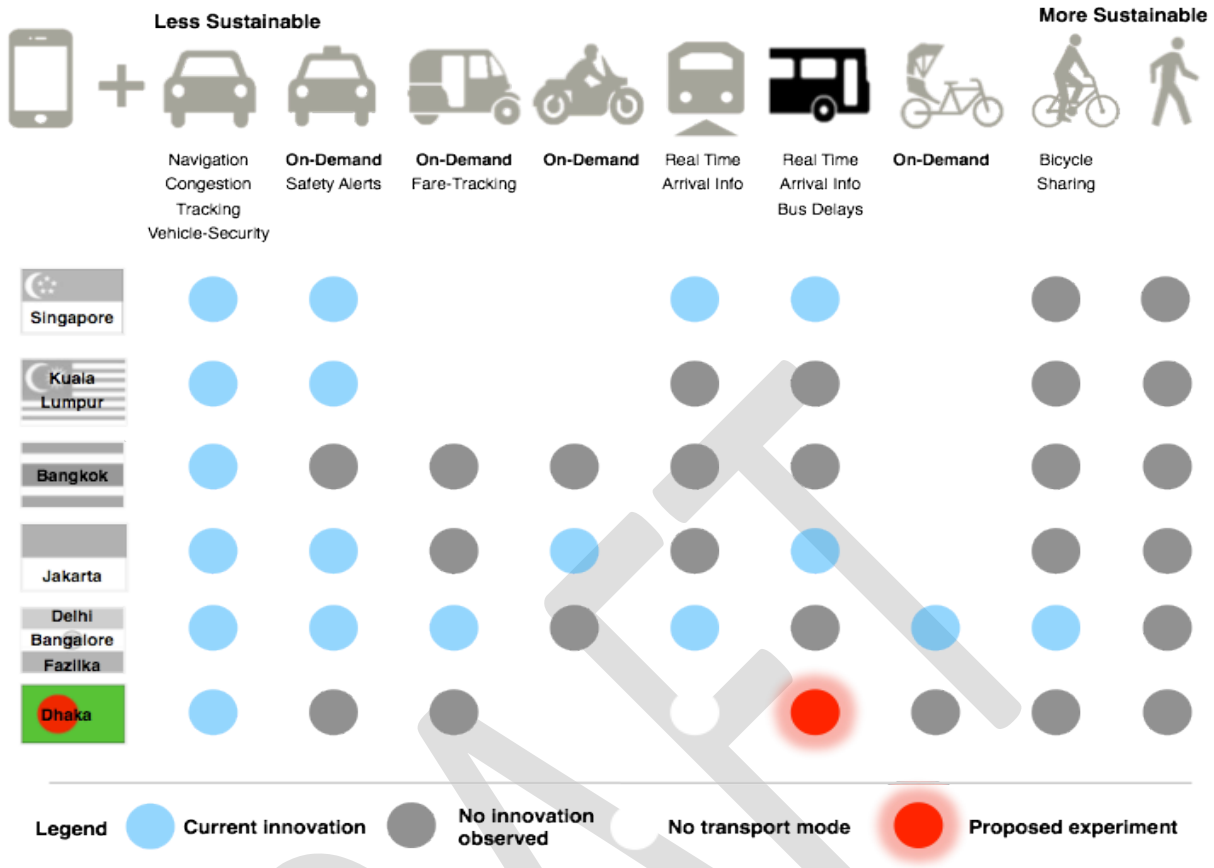

Figure 3: Mobile-driven innovations by transport modes in select cities in developing Asia

Sources: compiled from authors' fieldwork, July/August 2011

DRAFT

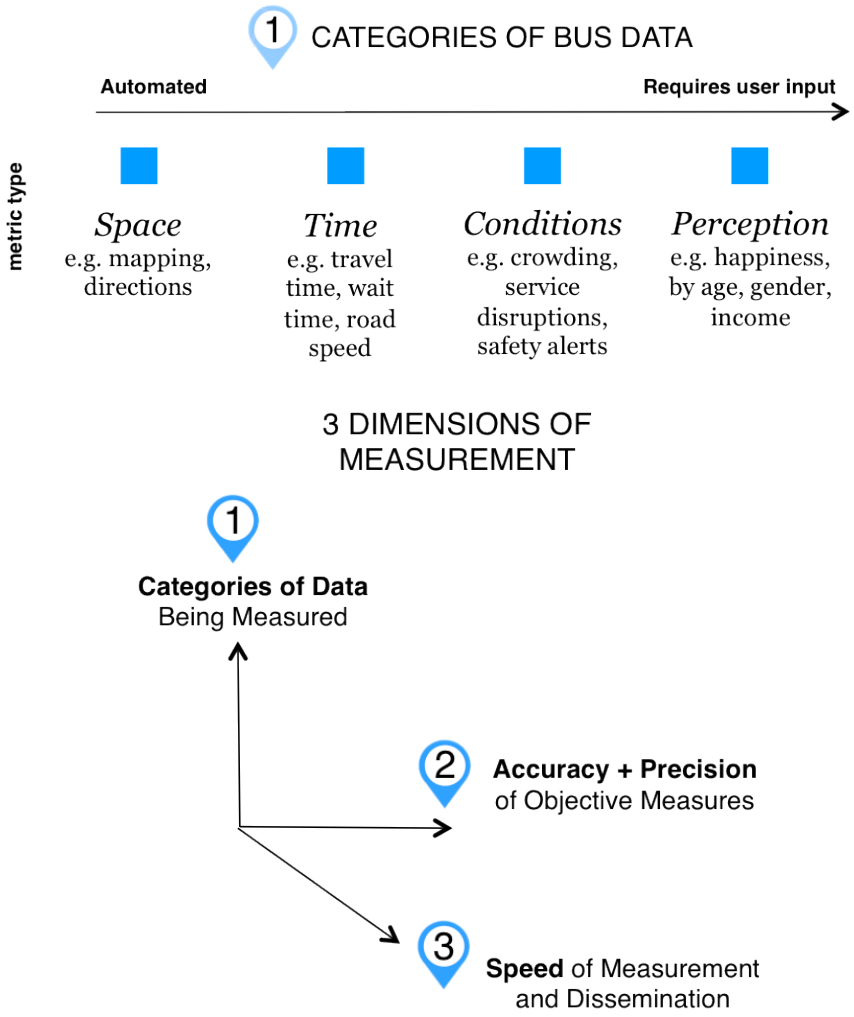

Figure 4: Categories of bus data and dimensions of measurement

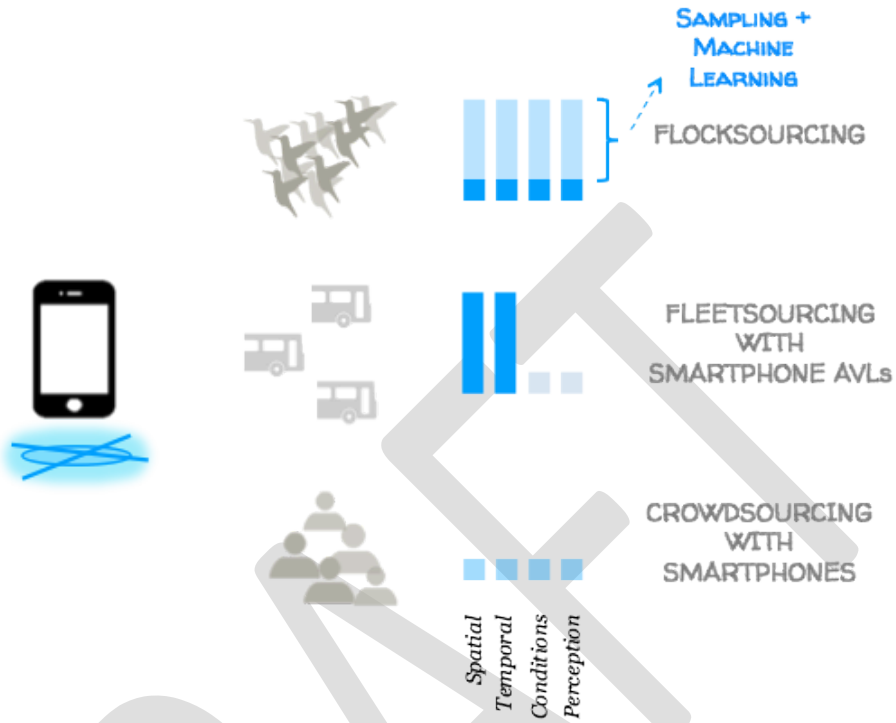

Figure 5: Bus data collection options with smartphones

DRAFT

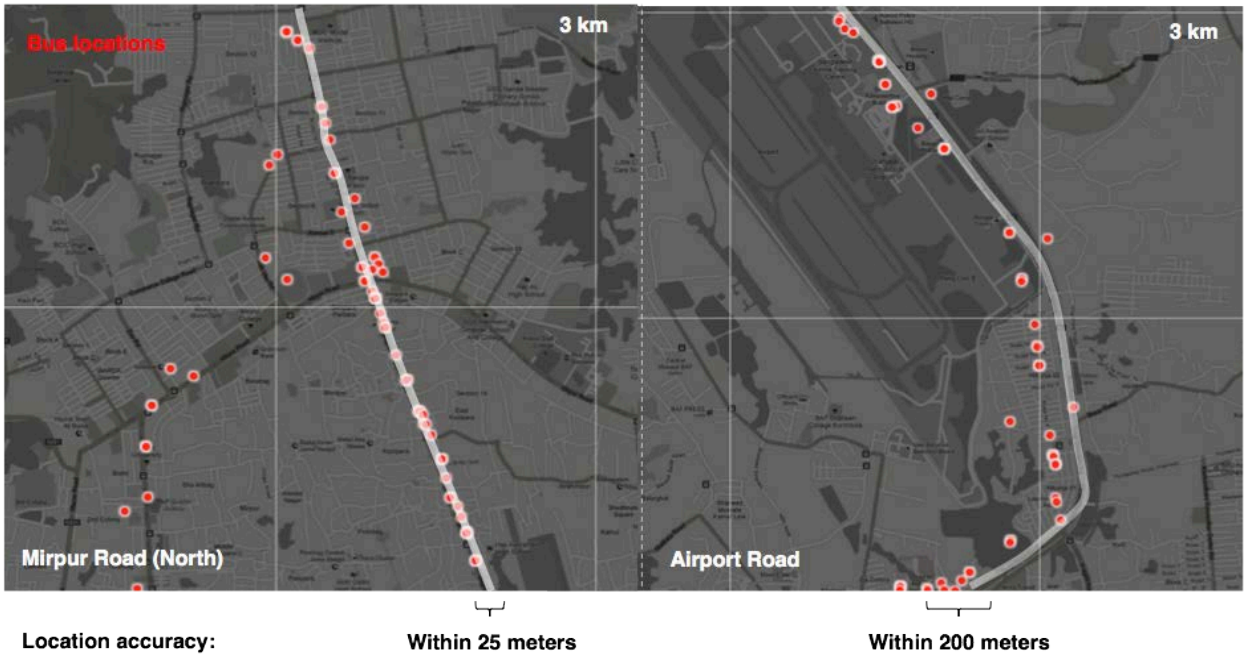

Figure 6: Sensed bus locations vs. roads positions in Dhaka

Figure 7: Mobile user interface of Share My Bus Dhaka: bus position input (left), counting passengers including females (middle), 9-question rider survey (right)

DRAFT

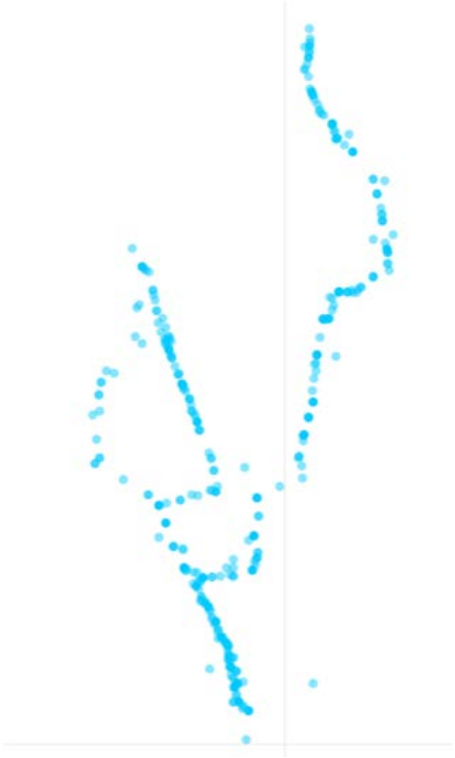

Figure 8: Locations of 1,014 riders surveyed onboard buses in Dhaka

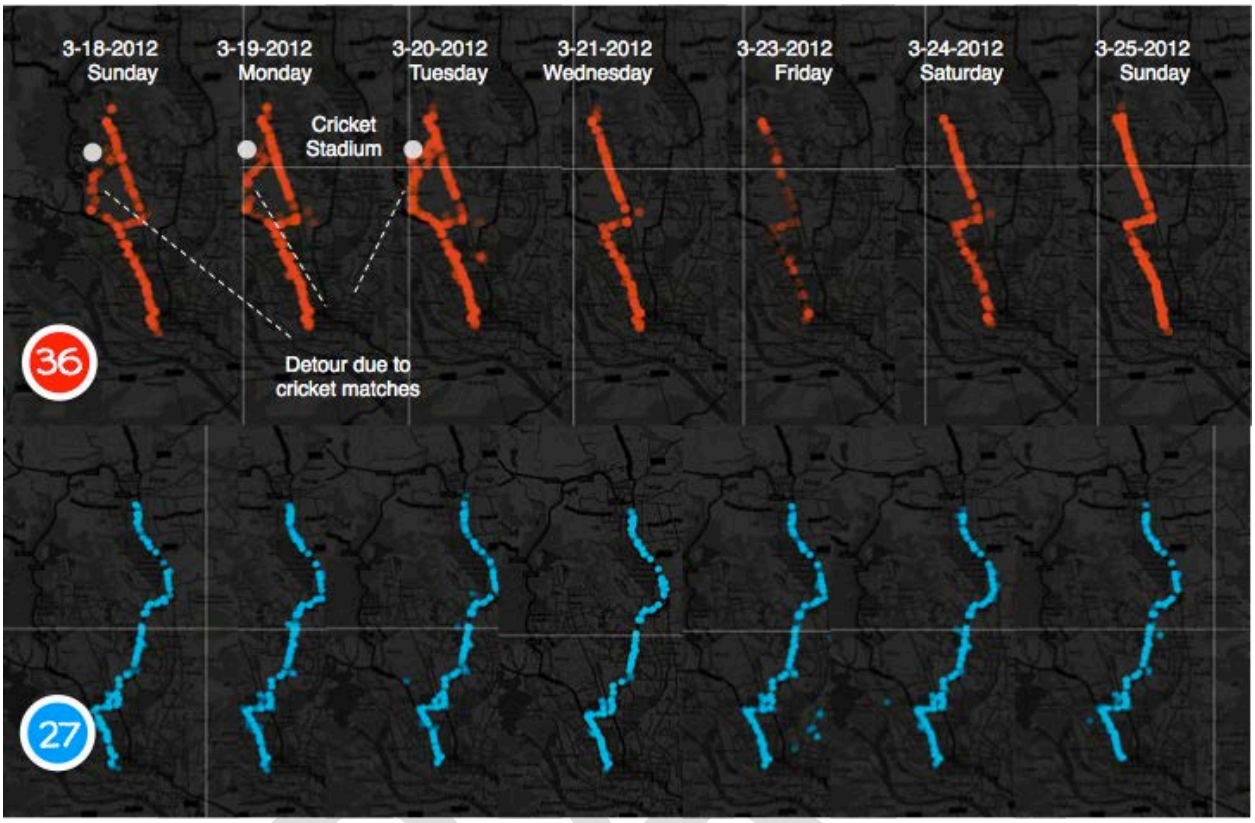

Figure 9: Bus traces by day and by line (March 2012)

DRAFT

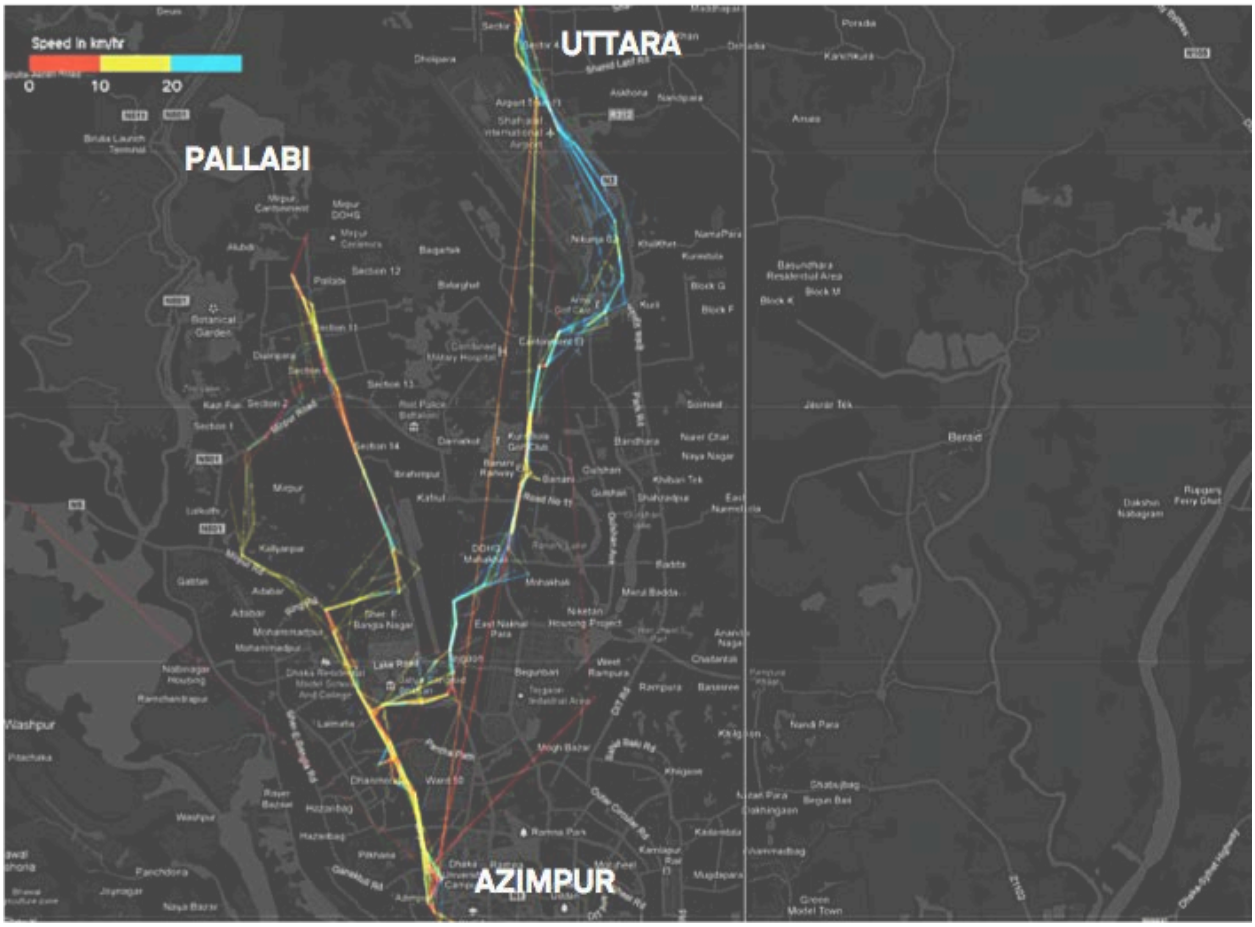

Figure 10: Bus speed map (March 2012)

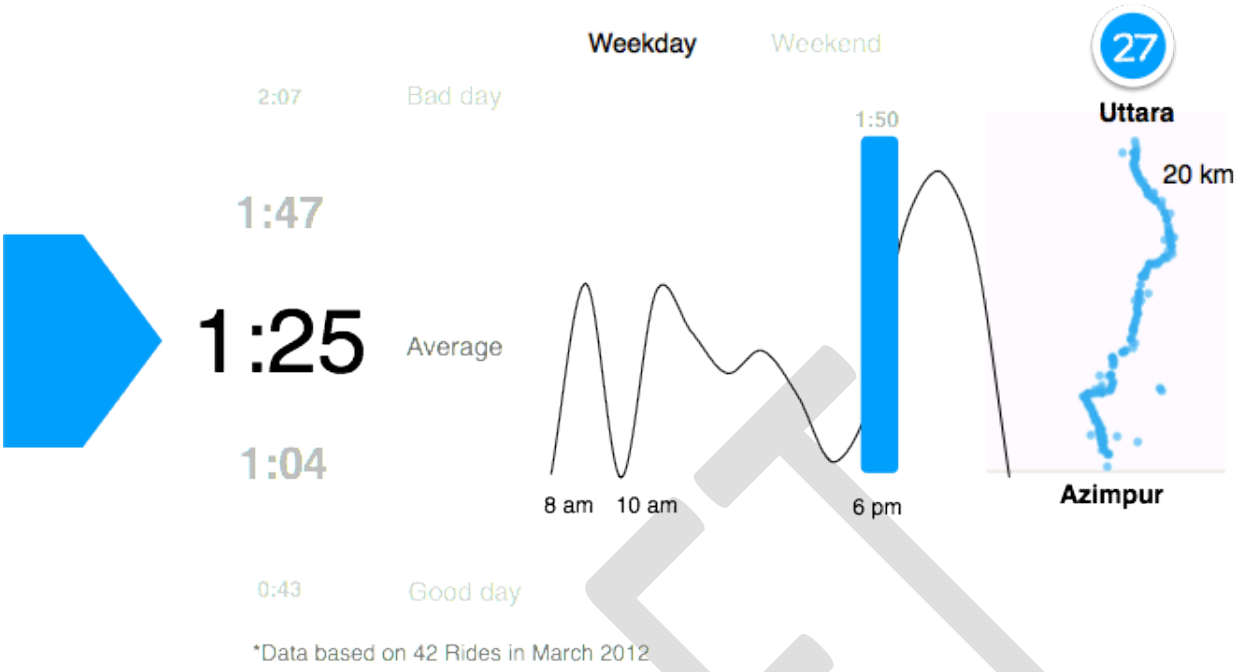

Figure 11: Estimated one-way travel times for the 12 km route of Bus No. 27 from Azimpur to Azimpur (March 2012)

DRAFT

Figure 12: Bus size, rider happiness, crowding and female ridership on 2 main bus companies serving the No. 27 route, BRTC and VIP (March 2012)