

Field Trial of Tiramisu: Crowd-Sourcing Bus Arrival Times to Spur Co-Design

John Zimmerman¹, Anthony Tomasic¹, Charles Garrod², Daisy Yoo¹, Chaya Hiruncharoenvate¹,
Rafae Aziz³, Nikhil Ravi Thiruvengadam⁴, Yun Huang¹, Aaron Steinfeld¹

¹Carnegie Mellon, ²Swarthmore College, ³University of Washington, ⁴Stanford University
johnz@cs.cmu.edu

ABSTRACT

Crowd-sourcing social computing systems represent a new material for HCI designers. However, these systems are difficult to work with and to prototype, because they require a critical mass of participants to investigate social behavior. Service design is an emerging research area that focuses on how customers co-produce the services that they use, and thus it appears to be a great domain to apply this new material. To investigate this relationship, we developed Tiramisu, a transit information system where commuters share GPS traces and submit problem reports. Tiramisu processes incoming traces and generates real-time arrival time predictions for buses. We conducted a field trial with 28 participants. In this paper we report on the results and reflect on the use of field trials to evaluate crowd-sourcing prototypes and on how crowd sourcing can generate co-production between citizens and public services.

Author Keywords

Transit, service design, crowd-sourcing, field trial.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Design, Human Factors

INTRODUCTION

The emergence of crowd-sourcing social computing systems has fundamentally changed computing. The success of services such as Wikipedia [24] (participant generated content), Travelocity [20] (participant generated reviews), Yelp [25] (participant generated location based content and reviews) and reCAPTCHA [23] (participant generated ground truth data for machine learning systems) illustrate just some of the many different forms this new technology can take and the range of problems it can address. Crowd-sourcing represents a “new material” for

HCI designers to play with. However, designers lack proven design processes for creating these types of systems and lack insight into how to address issues including bootstrapping information and initiating and maintaining user participation.

Service design is an emerging discipline uniting operations research and design. Operations research has long focused on redesigning processes to improve efficiency and effectiveness; however, the transition to service design has generated a new focus on customers’ experiences of a service. Interestingly, this transition mirrors the change in HCI from its traditional focus on usability to an emerging focus on user experience. Discourse in service design addresses the differences between products and services as a shift from *value-in-exchange* of goods to *value-in-use* that emerges between a customer and service provider [14]. This new perspective comes from the realization that customers have competence that service providers can use. They characterize this new perspective as *co-creation*, where value emerges through the interactions of customers and service providers [15, 16]. Crowd-sourcing social computing systems provide a rich space to investigate the intersection of HCI and service design.

Our research investigates the application of service design theory and crowd-sourcing technology to improve public services. Our intention is to help citizens who receive public services feel like co-owners. The work follows a research through design approach [27] where we design speculative systems, using methods from HCI and from service design, as a way of learning how to best design these kinds of systems. Our work began with a focus on transit commuters working as sensors in the transit network; reporting problems they experienced as a way of engaging in the “co-design” of this public service. However, fieldwork revealed that commuters had very little interest in reporting problems [26]. Instead, they wanted to know when their bus would arrive. Based on this feedback, we reframed the opportunity for service improvement of transit. We proposed that commuters could engage in co-creation of value by using the GPS enabled mobile phones they carry to generate real-time bus arrival information. Further, we speculated that if commuters engaged in frequent information exchanges with the transit service, then they would be more likely to engage in co-design activities, such as problem reporting.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

Copyright 2011 ACM 978-1-4503-0267-8/11/05...\$10.00.

To further investigate the validity of our hunch and to extend our knowledge of how to effectively design crowd-sourcing systems, we developed Tiramisu (“pick me up” in Italian), a system that allows commuters to share GPS traces and report the “fullness” of a bus they are riding. The interface provides real-time, historic, and scheduled arrival times organized by stop. In addition, it allows commuters to submit reports based on their ongoing service experience. We conducted a field trial with 28 commuters who used Tiramisu for three weeks. In this paper, we provide an overview of related research, details of our prototype interaction design, results of our field trial in terms of participation in providing GPS trace information and problem reports, and reflections on how to more effectively design crowd-sourcing social computing systems.

RELATED WORK

Related work falls into three areas: crowd-sourcing, service design, and similar HCI systems.

Crowd-sourcing and participation (contribution)

Crowd-sourcing systems are difficult to prototype. They resist the tried and true HCI methods of paper prototyping and controlled lab studies, because researchers and developers cannot accurately predict if users will contribute the information needed to make a system valuable to users. In addition, researchers and developers have no tools to help them understand how many users constitute a “critical mass” needed to keep a system working. Interestingly, some early research exploring this challenge has described how researchers can use Amazon.com’s Mechanical Turk as a crowd-sourcing test-bed [10].

While there has been little research on how to improve prototyping of crowd-sourcing systems, there has been a great deal of work on how to increase contribution in social computing systems. Some of the earliest work goes back to investigations of groupware. Researchers noted that many of these systems failed not for cognitive reasons (too difficult to use), but for social reasons; many workers did not enter information into these systems because they did not perceive enough individual value for their efforts [9].

A forthcoming book on the design of online communities reviews the literature on contribution and draws out approaches that system designers can take to improve contribution [11]. This work is also framed in terms of perception of value and it details three specific approaches for increasing people’s contribution to a community:

- **Asking:** People are more likely to contribute if they are asked, and if they are asked specifically/individually.
- **Intrinsic Motivation:** People will contribute if they perceive an intrinsic motivation, such as their own enjoyment in doing the work. In addition, people perceive value in helping others and in helping groups of people they feel an affiliation towards.
- **Rewards:** People will contribute for different kinds of rewards including praise, increased reputation, an increase in privileges, and financial compensation.

Interestingly, there are negative consequences to haphazardly combining approaches. For example, providing rewards for work people previously performed for intrinsic reasons can reduce their perception of intrinsic value for the activity, thus negatively impacting their contribution [11].

Service Design and Co-design

In the service design community, the concept of “service” has been described as a “perspective” focused on co-creation of value between customers and service providers [6]. This idea of co-creation emerged partially out of the insight that customers can be more than consumers; they can also be a source of competence [15]. As an example, Prahalad and Ramaswamy note that in the software industry, customers commonly share their competence by becoming beta-testers [15]. Interestingly, this insight came in part because the Internet allows companies and customers to engage in dialog and also from the recognition that dialog between the customer and service provider and within communities of customers presents an opportunity for value to emerge [15].

To better understand service design, Kuusisto and Päälylyaho conducted a review that revealed four types of co-production; where consumers participate in the production of the services they consume [12].

- **Consume:** Customer makes use of services and passively co-produces by creating the perception of value.
- **Co-perform:** Customer performs some of the tasks of a service.
- **Co-create:** Consumer uses resources (such as information) from a service to create their own value.
- **Co-design:** Dialog between customers and service providers around the types and form of service desired.

Service designers must take care when working on public services. Often these services function as monopolies and are not driven to improve their service offerings through competition. This is not necessarily a negative quality, as competition can diminish the importance of marginalized communities like the poor and people with disabilities, who are often the intended audience for public services. In fact, many public services exist specifically to address issues of fairness by creating a more level playing field for marginalized communities [3].

Public services have looked to technology as an approach to service improvement. Traditionally they have focused on automation and on improving the speed of service delivery. However, when public services take a more user-centered design approach, they often learn that their customers’ main desires are for different services, not for automated services [2]. In detailing opportunities for public service improvement, researchers have noted seven metrics:

- **Quantity of outputs:** More services offered
- **Quality of outputs:** Speed and reliability
- **Efficiency:** More outputs for less money
- **Equity:** General sense of fairness towards citizens in terms of costs and benefits

- **Outcomes:** Increased usage of a service by the citizens or by a percentage of the citizens
- **Value:** Cost per unit of outcome
- **Consumer satisfaction:** Often a result of several of metrics above [3]

Tiramisu improves “quantity outputs” by providing real-time arrival and bus fullness information, services not currently offered.

Researchers have proposed that co-design, where customers engage in the design of the services with the provider, may be a particularly good approach for service improvement of public services. They noted three distinct beneficial outcomes: services that are more responsive to the changing needs of citizens; increased trust of the government through citizen’s more positive engagement with services; and building of social capital through an increased sense of community [2].

Tiramisu attempts to operationalize the concepts of *co-production*. Specifically, we want transit commuters to *co-perform* in terms of generating bus the arrival information they desire by combining their traces with the transit service’s schedule. We want commuters to engage in *co-design* through problem reporting; indicating issues that effect them enough to be worthy of the transit service’s immediate attention. By conducting a field trial, we have gained insights on how to operationalize these concepts.

HCI systems in support of transit

The HCI research community has engaged in several projects at the intersections of commuting/transit and social computing. Here we discuss three working systems and detail how Tiramisu connects and provides an advance.

Researchers at the University of Washington developed a mobile system called OneBusAway that allows transit commuters to see real-time arrival information [7]. They have had great success at getting lots of participants to use this service and have noted some behavioral effects, such as an improved perception of personal safety and people choosing bus-stops farther from their currently location because the arrival time made it obvious that the bus would get them to their destination sooner. This system is similar to ours in that it focuses on the value of providing real-time bus arrival times. Tiramisu advances this work in that we have bus commuters generate the arrival time information, while their system uses GPS data logged by buses and provided by the transit service.

The Biketastic system was designed to support bike commuters [17]. Researchers conducted a proof of concept trial where participants used the system to record and share GPS traces as a way of describing a specific commute. In addition, cyclists added details of problems or challenges for a trace. This system is similar to Tiramisu in that it supports sharing GPS traces and problem reporting; however, it does not provide real-time information.

Pathfinder focused on investigating the design of systems in support of citizen science [13]. Citizen science is a philosophical approach to the practice of science where citizens in the community take on the role of research assistants, helping to collect information needed for analysis. Pathfinder intended to advance this notion by allowing citizens to participate in making sense of the data instead of limiting their role to data collection. Specifically, citizens shared information, asked questions, and proposed speculative hypotheses on commuting. Tiramisu is fundamentally different in intention from Pathfinder in that it is not designed in support of scientific sense-making by either citizens or professional scientists. Our participants share data and report problems, but they share data in terms of their own self-interest of wanting to know when a bus is coming, and they report problems with the intention of citizenship and civic engagement outside of science.

All three of these projects describe their motivation in terms of increasing sustainability through increased use of public or alternative transportation. Our work has a distinctly different intention of citizens feeling more ownership, responsibility, and control over the public services they use. Fortunately, the use of real-time arrival information has been reported to significantly increase the use of public transportation [4], and thus Tiramisu will likely have a positive impact on sustainability.

Finally, researchers have addressed the expense of commercial AVL systems by simulating a crowd-sourced AVL system for the Chicago transit service [20]. This work differs from Tiramisu in that it attempts to automatically detect when participants are riding a transit vehicle and in that it is a simulation and not a field trail.

INTERACTION DESIGN

Motivated by our fieldwork with transit commuters, discussions with a local transit agency [26], and the concepts of co-production and co-design from service design, we developed a conceptual model illustrating how information exchange between commuters and the transit service could provide a foundation from which co-design could emerge (Figure 1). We focused on crowd-sourcing the vehicle location data, literally appropriating the GPS and communication devices commuters already carry, because the transit system we were working with reported costs of approximately \$70 million to install a commercial Automatic Vehicle Locator system for a mid-sized city.

In this model, transit commuters provide GPS trace data using their phones. The transit service provides planned schedule information. Based on the schedule information and the GPS traces, a computational system generates real-time arrival information.

The transit commuters we interviewed during our fieldwork shared that reporting problems and engaging in improving the transit services offerings was not their responsibility [26]. They felt that the effort to report a problem was simply too high compared to the perceived value. In

working on our conceptual model, we speculated that if transit commuters accessed the arrival information and traced GPS data, then they would often have the system-in-hand. If the system displayed a screen that supports problem reports for the transit service, then, when commuters encountered a problem, having a system-in-hand would reduce the perceived effort of making a report. To capture this relationship in our conceptual model, we placed co-design above the arrival time information exchange.

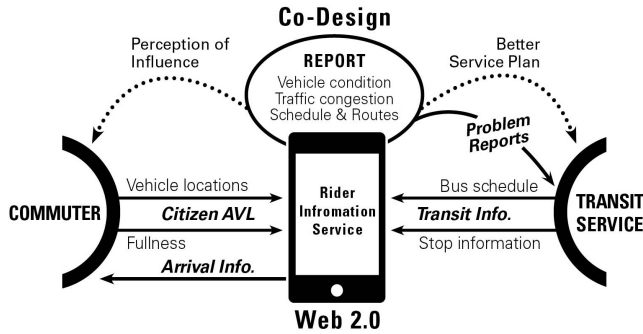


Figure 1. Conceptual model of design problem framing.

We chose to prototype and conduct the field trial using iPhones because of the critical mass of iPhone users at the time of the trial, and because the iPhone has a screen reader that supports blind and low-vision commuters. A production version of the system would need to operate on many smart phone platforms to minimize barriers to participation.

Interface Design

From the “Main Menu,” users can access information on

bus arrival times in three ways: nearby, search, and favorites (Figure 2A). Selecting “Nearby” transitions the interface to the “Main Map” screen, which offers a view similar to Google maps. Nearby stops appear as red pins. Selecting a pin brings up a map “annotation” with the name of the stop and the time of the next bus (Figure 2B).

Selecting the button on the annotation transitions to the “Select Route” screen, which displays a list of upcoming buses for the selected stop (Figure 2C). This screen displays real-time arrival information for a bus when there is a commuter on that bus who is sharing GPS. When this is not available, the system shows historic arrival information, assuming the system has previous trace data for this bus at this time. When neither real-time nor historic arrival information are available, the interface shows the scheduled arrival time. In designing this screen, we chose to combine real-time, historic, and schedule information as a way of addressing the bootstrapping challenge of getting crowd-sourced information into Tiramisu. Our plan is that when this app is first released, commuters will see scheduled information, and as more and more commuters use the system and trace their rides, the schedule will be replaced by real-time and historic information. We also specifically chose not to show the difference between the real-time and the scheduled arrival times. Our intention is to help commuters know when the bus is coming, and not to “shame” the transit service by revealing that their bus may be earlier or later than scheduled.

Real-time and historical data include bus fullness information, something currently not available as an estimate in the schedule provided by the transit service. Our previous fieldwork with commuters revealed their

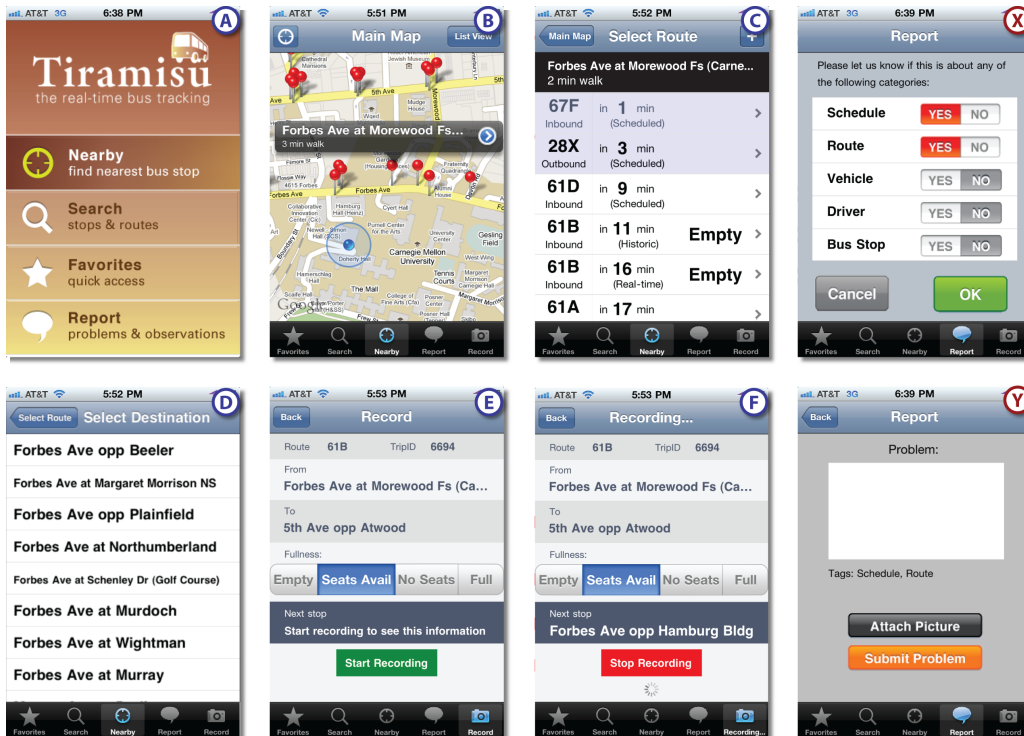


Figure 2. Screens from Tiramisu interface.

Main Menu (A):

Main Map (B): Map with selectable stops based on location.

Select Route (C): Arrival times for selected stop.

Report (D): Select destination bus stop.

Record (E): Report fullness and share GPS trace info.

Recording (F): Update fullness and stop trace.

Report (X) Select categories.

Report (Y): Input report text and add photo.

unhappiness when they would see a bus they wanted coming towards and then have it drive past them because it was too full to allow more people to board [26]. We assumed that knowing ahead of time that a bus was full might lessen the blow of watching it drive past without stopping. Additionally, we thought commuters could use this information during rush-hour to decide if they should wait a few minutes in order to ride a bus that is less crowded. Finally, we felt fullness would benefit commuters with walkers, wheelchairs, strollers, and large bags.

Clicking on the star in the upper left corner of the “Select Route” screen (Figure 2C, underneath the C label) adds this stop to a commuter’s list of favorites. Additionally, selecting favorites from the main menu (Figure 2A) transitions the interface directly to the “Select Route” screen (Figure 2C).

When commuters decide on the bus they wish to take, they select that bus from the list of arrival times, transitioning the interface to the “Destination” screen (Figure 2D). Here they select from a list of upcoming stops. Making a selection transitions the interface to the “Record” screen (Figure 2E). Once they board the bus, commuters indicate fullness and press the “Start Recording” button to share the fullness rating and the GPS trace from their phone. While tracing, Tiramisu displays the next stop (Figure 2F). At the end of the trip, commuters press a stop recording button (Figure 2F) as they exit.

Selecting “Report” from the “Main Menu” (Figure 2A), or selecting the “Report” tab at the bottom of any screen transitions the interface to the “Report” screen (Figure 2X). This allows commuters to immediately express positive or negative comments about their transit experience. Commuters categorize the report by selecting schedule, route, vehicle, driver, and bus stop. Next, they type a message and possibly provide a photo (Figure 2Y).

System Overview

Figure 3 displays a high-level functional architecture of Tiramisu. A single backend server maintains a client-server arrangement with multiple clients running on iPhones. Clients contain the user interface and a data model that records favorite stops. The server logs recorded (traced) trips, predicts arrival times, logs fullness, logs problem reports, and processes client information query requests.

To initialize the system, a GTFS [8] representation of the transit agency schedule is loaded to the database. Every thirty seconds a real-time model predicts bus arrival time and fullness for buses with trace data within the last 30 seconds. Once a day, the previous month of recorded trips are used to construct a historical model of bus arrival times and fullness. Both arrival-time prediction models first prune outliers and then regress the distance the bus must travel to a stop against the absolute time of the arrival at that stop. This model is robust against a variety of sources of error in the system, such as bad GPS information, dropped trace signals, and user error. Similarly, the real-time and

historical fullness model is based on a simple average of past fullness. Models can be generated based on days of the week, weekends, and holidays. The models predict bus location and fullness approximately 20 minutes into the future.

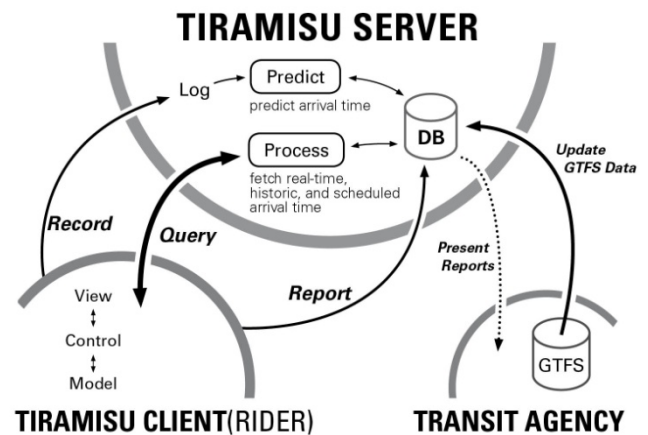


Figure 3. Functional Architecture of Tiramisu.

Client instances issue three types of requests to the server. *Trace messages* contain the client’s GPS coordinates, fullness, departing-stop, destination-stop, bus route, and a trip identifier. *Reports* (containing selected options, text and an optional photo) are written directly to the database. *Informational queries* (e.g., nearest bus stops for the client) are processed against the current state of the database.

FIELD TRIAL

Typically, HCI design processes recommend an iterative process of moving from low to high fidelity prototypes. This reduces risks by addressing critical challenges before making too large of an implementation commitment. Crowd sourcing systems, however, make this iterative approach difficult. The main challenges generally have to do with acceptance of the system and participation in generating the information, and these cannot be easily investigated, with respect to a specific system, through the use of lab studies. These aspects of crowd-sourcing systems require a critical mass of participants to create the social effects of use.

In our case, we needed to know if commuters would actually share GPS traces, since this information would only help commuters downstream of their location. An individual commuter never immediately benefits from the traces they provided. Tiramisu’s interaction design relies on intrinsic motivation to get commuters to share traces and report problems. In addition, during the field trial participants were paid to use the system. While creating the design, we specifically decided to not attempt to amplify people’s contribution using the many approaches recommended in the literature in order to get something close to a baseline for this behavior. We also wanted to know if the co-production of the real-time arrival information would lower the barriers to co-design (problem reporting). Neither of these issues could be assessed in a lab

study using paper prototypes or a roughed out technical system; therefore, we chose to conduct a field trial.

We recruited participants using electronic message systems and flyers at bus shelters. The recruiting materials indicated that participants must be regular commuters and must own an iPhone. Through selective placement of the ads, we targeted people who rode through a specific transit corridor. We wanted to reduce the number of participants needed to create a “critical mass.” We recruited 28 participants: 16 fulltime students, 5 fulltime workers, and 7 part-time students who also worked fulltime.

Participants visited our lab where they took a pre-survey detailing their typical transit use and probing on their history of reporting transit issues. In addition, we loaded the Tiramisu software on their iPhone. They then left the lab and used the system for two weeks.

Near the end of two weeks of usage, we schedule an appointment for participants to return to our lab for a post survey and for payment. When they returned, we gave them the TAM 3 (Technology Acceptance Model) [5, 22]. This instrument measures perceptions of usefulness, ease of use, enjoyment in use, output quality, results from use, and behavioral intent (likelihood of use in the future). Responses use a 7-point Likert scale ranging from 1 = “extremely unlikely,” to 4 = “neither,” to 7 = “extremely likely.” We use this instrument to get a baseline perception of the user experience and possible acceptance of Tiramisu.

While at our lab, we connected their phones; downloading the favorites they created and the impressions they experienced. The term “impression” indicates that an arrival time was displayed. For example, Figure 2C shows the “Select Route” interface screen. This screen has 5 visible impressions: 1 real-time, 1 historic, and 3 scheduled. We also conducted a semi-structured interview to find out more about their experience with Tiramisu to get arrival information, to provide traces, and to report problems. Each participant was paid \$20. They were told they could continue to use Tiramisu, but that we would no longer pay.

Participants did not all start using Tiramisu on the same day. Instead, they continued to enroll over the 38 days our trial ran (Figure 4). For our analysis, we normalized the data based on an individual participant’s first day, and we only included their first 21 days of use in calculating the averages. Because of the slow startup of use, participants who joined later in the trial had a higher likelihood of encountering real-time and historic data.

Throughout the trial, Tiramisu logged when and where individuals accessed bus arrival information, when and where they traced, and any problems that they reported.

During the field trial we encountered many typical challenges not usually faced in lab studies. These include a schedule change instituted by the transit agency affecting a small number of bus routes, broken and upgraded phones that required participants to return to our lab to have the

software reloaded, one participant moving outside of our targeted corridor, and participants taking vacations.

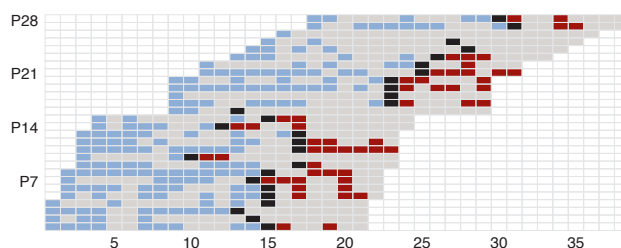


Figure 4. Ramp up of participants using the system for 21 days each over the course of the 38-day field trial. Blue indicates use of the system on a “paid” day. Black indicates the day participants were paid. Red indicates use of the system by participants after they were paid.

FINDINGS

Findings are reported in terms of **participation**, sharing GPS traces; **bootstrapping**, commuters access to real-time, historic information, and fullness information; **co-design**, contributions of problem reports; and **user experience**.

Participation

Participants provided traces (the participation rate) for approximately 56% of their sessions. A “session” happens anytime a participant accessed the arrival time information on their phone. We infer an untraced-session when participants do not start a trace within 15 minutes of a session. We calculated the participation rate by dividing sessions that contain a trace against total sessions. The participation rate is only an estimate of traces to actual “rides”. We had no way to detect cases when participants boarded a bus without using Tiramisu or when they accessed the arrival time information and then chose not to ride a bus, and thus did not trace. The highest contributing participant traced 96% of their sessions and the lowest contributing participant traced 4% of their sessions.

Figure 5 shows the participation rate over time, normalized to the number of days a participant has been in the trial. Across all participants, sharing traces decreased over the course of the trial as indicated by the regression equation ($y = -0.0044x + 0.6038$) with $R^2 = 0.0551$.

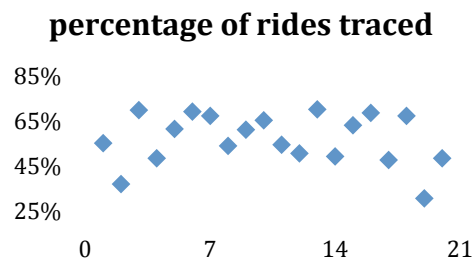


Figure 5. Change in percentage of tracing over the course of the field trial in days (14-paid followed by 7-unpaid days).

There were differences in average tracing rates for different user groups. Fulltime workers traced 68% of rides; students

traced 48% of rides; and participants who were both fulltime workers and part-time students traced 59% of rides.

Bootstrapping

Participants viewed 16,263 arrival impressions. These were captured when we synced their phones in our lab at the point of payment; black marks on Figure 4. Across all participants, 5.9% of the impressions they viewed showed real-time arrival times and 7.2% showed historic arrival times. Together, 13% (2,132 out of 16,263) of all impressions were created by data participants supplied and 87% came from the schedule.

In the post-survey, we asked each participant about the three types of arrival times: real-time, historic, and scheduled. Interestingly, three participants could not claim a specific preference between these three. In addition, four participants rarely if ever had the experience of having real-time or historic data available for a bus they were taking.

In general, participants really liked the real-time data, when it was available. Twenty-one participants stated that this was their favorite kind of arrival time. P28 elaborated on the difference between the printed paper schedule and the real-time: *“Although the scheduled time was [indicating] 20-minutes later, the real time showed 2-minutes later. And the bus did come in 2-minutes.”*

Two participants expressed a preference for historic data. P27: *“It was most useful when a route I was planning on taking had a historical data time with it. It was then I noticed that the historical time was closer to actual time than the published schedule.”* However, 11 participants did not trust historic data. P7: *“I least trust historical data because the system is new and has not necessarily collected enough data to be useful, although this will presumably become more reliable as the amount of data increases.”* One participant expressed a desire to see the scheduled data instead of historic. P18: *“I prefer scheduled time [to historic] because if the scheduled time is wrong, at least I can get angry at Port Authority for being inefficient.”*

Twelve participants expressed distrust of scheduled data. They stated that in their experience, the buses never arrived as scheduled, and that they were always late.

In discussing the addition of fullness information, several participants expressed an appreciation for this new information. P5 described a situation where she used the system to adapt her evening commute. While using the system she saw a bus scheduled to arrive 30 minutes later than her usual bus, and the historic data showed this to typically be empty. She shared that her usual bus never had any seating. She left work a little later and the bus did in fact have many seats available.

Co-design

In the pre-survey, 3 participants answered they had previously reported problems to a transit authority within their lifetime. Over the course of our trial, 14 participants submitted 22 problem reports: 9 related to transit

infrastructure, 9 related to schedule information, and 4 reporting that the Tiramisu software had crashed.

Reports on infrastructure generally addressed vehicle conditions. Examples include: *“Broken seat”* (this report included a photo of the broken seat), *“The bus died!”*, and *“The AC barely works and the inside of the bus smells like diesel fuel. It also seems to have troubles going up hills and the panel covering electronics above the driver’s head is about to fall off it appears.”* Interestingly, one report contained both positive and negative observations: *“The [annunciator] and message board are not indicating the next stop. The bus was nice and cool though.”*

Several reports expressed frustration about late and missing buses. *“Late, not on time.”*, *“I waited 1 hour for a 500 bus from Oakland to Highland park.”*, *“No 61A, B or D has been by in the last 30 mins.”* In two reports, riders attempting to explain arrival time problems: *“Traffic jam at Murray at Hazelwood in 61D.”* and *“Driver eliminated upper loop in Trafford which is part of the route.”*

Participants submitted 9 reports regarding errors with the arrival list. Typical comments included: *“Boarded 61D. Bus is full. It was not displayed in the arrival list.”* *“I got on a 67E that was about 10 minutes late. It had fallen off the list of possible buses”*, *“67 not listed. Bus routes on 67 line should be updated in Tiramisu to reflect recent schedule changes.”* There are four possible causes for these breakdowns. First, as mentioned in a message, a bus could deviate from the schedule enough to fall off our interface. Second, some drivers pull over to stops they have not been assigned, because commuters push the stop request button or the driver spots a waiting commuter who signals the bus. Third, there were a few errors in the schedule we received. Fourth, there was a slight schedule change during the trial and we did not incorporate this into Tiramisu.

User Experiences

In the post-use interviews, participants shared that they liked using Tiramisu. Several participants even expressed concern when we cabled their phone to our computer, fearing we would delete the software. They intended to keep using Tiramisu even without a financial reward. In fact, 26 of the 28 participants asked us to email them if a commercial version becomes available.

Results from the TAM 3 are captured in Table 1. Various analyses impression values as independent variables revealed no significant differences. These included direct analyses of impression subtypes (real-time, historic, and scheduled) as well as analyses where the independent variables described the percentage of impressions containing value-added data viewed by each participant. Since none of the comparisons generated a statistically significant result, we simply report the average rating for each category across all participants.

Participants shared how Tiramisu influenced their commuting practices. P23: *“Sometimes I don’t need to look up the schedule but [with Tiramisu] I felt forced to look it*

up ...” P31: “[Normally,] I chat with my friends and get on whatever bus that comes. Didn’t really pay attention to when exactly the bus was coming ... [With Tiramisu,] I often checked the [arrival] time.”

Table 1. Average TAM 3 ratings across all participants.

Measure	Ave	STD
Utility	5.51	1.25
Ease of use	5.56	1.55
Enjoyment	5.68	1.08
Output quality	4.82	1.60
Results of use	5.59	1.03
Behavioral Intent	5.90	1.44

In speculating on new features, two participants inquired about reporting availability for bikes on the bike rack mounted on the front of many buses. Two participants mentioned that Tiramisu would be improved if it could notify users of upcoming stops. Finally, one participant suggested that Tiramisu should automatically stop recording at the appropriate time.

Participants made many comments in terms of their phone. Several expressed concern about how much battery Tiramisu requires when sharing traces. P23: “It sucks up your battery.” P14 (who had an hour and a half commute): “[Tiramisu] uses 30% of the battery from home to work.” In addition, participants expressed concern for having to explicitly stop tracing. Several noticed hours after they had exited the bus that their phone still indicated that it was “recording.” Finally, several participants complained about a lack of multi-tasking on the iPhone and their desire to access other applications during their commute.

DISCUSSION

Following our analysis, we reflected on our findings. Below we highlight key insights in terms of advancing the design of Tiramisu, understanding the connection between social computing technology and service design, and prototyping social computing system via field trials. These insights are organized around the themes of participation, bootstrapping, co-design, and user experience.

Participation

During the field trial, participants traced more than half of their rides. This contribution rate seemed surprisingly high to us; however, we had no way of predicting a baseline for participation prior to the field trial. It is difficult to claim that this is a good rate or even a “good enough” rate of participation for a crowd-sourcing social computing system to survive and flourish.

While participation seems higher than expected, it did have a slight downward trend over the course of the trial. We see three possible reasons for this. First, there was most likely a novelty effect when the app was new on the phone, and over time this effect began to diminish. Second, the earliest participants in our trial as well as participants who lived outside the main transit corridor used by other participants experienced fewer impressions of real-time and historic

data, and this lack of value over time may have reduced their participation. Third, we paid participants to use Tiramisu for two weeks, and after the payment, they may have felt less obligated to trace their rides.

Surprisingly, students traced fewer rides than workers, and their traces were generally less valuable in terms of minutes of GPS trace, as their trips tended to be shorter; they generally lived close to school. This finding is important for others who wish to conduct field trials of these types of systems. It is important to predict which participants will receive the most value and also which participants have the most valuable information to contribute.

As mentioned, we did not choose a specific strategy to increase participation with Tiramisu. In looking at theories on contribution, we see at least three ways to proceed. We could give commuters an individual score (points) for each minute or for each stop-to-stop segment they trace. We could add this scoring and then also generate competition between groups, such as different businesses competing. Finally, we could start scoring and then charge users points (such as the number of minutes of traces they share) in order for them to access the valuable real-time arrival times.

Bootstrapping

During the field trial, 28 participants created 13% of the data they experienced: impressions of real-time and historic arrival information. Once again, this seems higher than expected and again we did not have a good method of calculating how good this is or how good “good enough” is in terms of crowd-sourcing data. In particular, participation rates depend on the initial distribution of participants, which in turn affects participation rates, which in turn affects the distribution of participants, leading to a feedback-based system that is difficult to accurately model without measurements from the field.

Our design strongly benefited from the fact that we could start with a static dataset—in this case the transit schedule—which users would find to have some value in use. Tiramisu then allows users to improve this dataset by replacing instances of the static data with crowd-sourced data. This structure allowed for a slow rollout of the application. We did not need to generate an initial set of crowd-sourced data to bootstrap the system. In deciding on whether to field test a crowd-sourcing social computing system, we speculate that this model of augmenting valuable static data with more valuable crowd-sourced data decreases the risk of abandonment of the system when working with a small set of users.

Extending the transit schedule with fullness information also appeared to improve the value-in-use of Tiramisu for participants. In one case a participant was able to adjust her schedule to ride a much less crowded bus. Access to this information could potentially increase the flexibility of commuters by letting them see into situations they have not yet experienced. In the long-term, a feature like fullness might allow commuters to adjust their schedules, resulting

in better load sharing across all the commuters. This information helps commuters improve their personal experience as well as others' experience with commuting.

The conflicting findings on historic arrival times surprised us. We assumed that participants would view historic as more valuable. In the next iteration of Tiramisu, we will consider the issue of trust in intelligent systems [28].

Co-design

One of the main hunches driving our design was that co-design in terms of problem reporting would increase due to participants having a phone-in-hand as they accessed the arrival time information and provided traces. The change in behavior, from 3 previous problem reports from 28 participants, to 22 problem reports from 14 of 28 participants provides some confidence that this hunch could pay off. Interestingly, the problem-reporting rate of our participants appeared to be about equal with the reported problem reporting rates of commuters with disabilities, a segment of public service users who may have higher motivation to participate in co-design due to a higher-level of dependence on this service [19].

In terms of the design, reporting problems seem like a shallow level of co-design, with a narrow focus on complaints as a measure of an issues importance. In advancing this work, it will be important to investigate if this idea of phone-in-hand would also result in increased participation of commuters in regards to more strategic planning issues, such as route and schedule changes or issues relating to taxes and budgets.

It was interesting to see participants submit reports that explained why buses were delayed; possibly in the hope the transit service could use this information to make changes. However, this action raises the possibility that commuters might be willing to share this information with each other. In reconsidering our design, we see an opportunity for commuters to add comments about an individual bus or a stop that are immediately pushed out into the system, sort of like a very localized "tweet" without the need for a special language for vehicle, stop, or route information.

In thinking about having the "system-in-hand," it is impossible for us to tell if problem-reporting contributions are linked to tracing contributions or simply to the ability to access arrival information via a mobile phone. An important advance to service design research would be a clearer investigation of the effect of commuters producing traces versus a system that simply provides this information in a mobile form. Little research has investigated how the activity of co-production changes a user's perception of the relationship they have with a service provider [1]. This analysis would get more to the heart of the co-production and co-design concepts in service design.

One concern that surfaced in our previous fieldwork from both commuters and the transit service was a lack of trust in the other commuters. Both groups felt that some commuters

would maliciously produce "bad" information in terms of traces and problem reports. We did not see this effect in our field trial; however, this absence is not surprising since participants knew their behaviors were being monitored. This result offers some insight that a way to address the perceived issue of malicious behavior might be to force users of the system to give up their anonymity. This design decision is an interesting question in terms of privacy and location based technologies and in terms of governments monitoring the behavior of citizens. It may be possible to convey accountability without revealing identity [18]. This is clearly a topic for deeper investigation.

User Experience

In terms of user experience and Tiramisu, our TAM scores were quite high. Our inability to tease out more specific trends in the data suggests a ceiling effect combined with very low overall variability. Scores were uniformly high for almost all TAM indices, thereby leaving little room for differentiation on the upper end of the scale. This indicates an increased likelihood of acceptance for Tiramisu and for other systems that follow a similar design. However, this speculation is based on results of a self-selected group who volunteered for this trial. Nevertheless, these results give us confidence that the basic framing, features, and usability of the design are "good enough" to push out for a larger scale field deployment and data collection effort.

Comments from the interviews indicate that co-design of the transit service might also include co-design of Tiramisu. Participants had design ideas for new features that emerged from their use; however, our report function did not make these ideas easy to share. This desire provides further evidence of a strong connection between web 2.0 technology and opportunities for co-design.

CONCLUSIONS

Crowd-sourcing social computing systems are difficult to prototype. Iterative design techniques like testing with paper prototypes cannot investigate the critical risks of bootstrapping and contribution. In our research, we addressed this challenge by conducting a field trial of Tiramisu; a system that allows bus commuters to share GPS traces using their smart phones. Tiramisu then uses these traces to generate real-time arrival estimates of buses commuters can access on their phones. Tiramisu's design integrates theories from service design on co-production and co-design, with the intention of getting commuters to engage more deeply with the transit service and begin to feel they can influence its service offerings.

Our field trial demonstrates that this concept works in practice. Participants were successful at tracing information and at using the real-time and historic estimates to improve their commuting experience. In addition, we noted high participation rates for explicitly sharing location traces and for reporting problems encountered with the transit service.

Acknowledgements

The Rehabilitation Engineering Research Center on Accessible Public Transportation (RERC-APT) is funded by grant number H133E080019 from the United States Department of Education through the National Institute on Disability and Rehabilitation Research. In addition, this work was partially supported by Google and by the Hillman Foundation through a grant from the Traffic21 project at Carnegie Mellon.

REFERENCES

1. Bendapudi, N. and Leone, R. P. Psychological Implications of Customer Participation in Co-Production. *Journal of Marketing*, 67, 1 (2003), 14-28.
2. Bradwell, P. and Marr, S. *Making the Most of Collaboration an International Survey of Public Service Co-Design*, DEMOS Report 23. (2008), DEMOS, in association with PriceWaterhouseCoopers (PWC) Public Sector Research Centre, London.
3. Boyne, G. A. Sources of Public Service Improvement: A Critical Review and Research Agenda. *Journal of Public Administration Research and Theory*, 13, 3 (July 2003), 367-394.
4. Casey, C. Real-time information: Now arriving. *Metro Magazine*, <http://www.metro-magazine.com/Article/Story/2003/04/Real-Time-Information-Now-Arriving.aspx> (Accessed Sept 11, 2009).
5. Davis, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly* (Sept. 1989), 319–339.
6. Edvardsson, B., Gustafsson, A. and Roos, I. Service Portraits in Service Research: A Critical Review. *International Journal of Service Industry Management*, 16, 1 (2005), 107-121.
7. Ferris, B., Watkins, K. and Borning, A. Onebusaway: Results from Providing Real-Time Arrival Information for Public Transit. In *Proceedings CHI*, (2010) ACM Press, 1807-1816.
8. Google GTFS: http://code.google.com/transit/spec/transit_feed_specification.html
9. Grudin, J. Why Groupware Applications Fail: Problems in Design and Evaluation. *Information Technology & People*, 4, 3 (1989), 245-245.
10. Kittur, A., Chi, E. H. and Suh, B. Crowdsourcing User Studies with Mechanical Turk. In *Proceedings of CHI*, (2008) ACM Press, 453-456.
11. Kraut, R. E. and Resnick, R. Encouraging Contribution to Online Communities. In Kraut, R. E., Resnick, P., Kiesler, S., Riedl, J., Konstan, J. and Chen, Y. (Eds): *Designing From Theory: Using the Social Sciences as the Basis for Building Online Communities*, (forthcoming).
12. Kuusisto, A. and Päällysaho, S. *Customer Role in Service Production and Innovation — Looking for Directions for Future Research, Report 195*, (2008) Lappeenranta University of Technology, http://akseli.tekes.fi/opencms/opencms/OhjelmaPortaali/ohjelmat/Serve/fi/Dokumenttiarkisto/Viestinta_ja_aktivointi/Julkaisut/Kuusisto_Paeaellysaho_raportti.pdf. (Accessed: August 29, 2009).
13. Luther, K., Counts, S., Stecher, K. B., Hoff, A. and Johns, P. Pathfinder: An Online Collaboration Environment for Citizen Scientists. In *Proceedings of CHI*, (2009) ACM Press, 239-248.
14. Ng, I. C. L., Maull, R. S. and Smith, L. Embedding the New Discipline of Service Science. In *The Science of Service Systems*. H. Demirkan, J. H. Spohrer and V. Krishna (eds). Springer, 2009.
15. Prahalad, C. K. and Ramaswamy, V. Co-Opting Customer Competence. *Harvard Business Review*, 78, 1 (2000), 79-90.
16. Prahalad, C. K. and Ramaswamy, V. Co-Creating Unique Value with Customers. *Strategy & Leadership*, 32, 3 (2004), 4-9.
17. Reddy, S., Shilton, K., Denisov, G., Cenizal, C., Estrin, D. and Srivastava, M. Biketastic: Sensing and Mapping for Better Biking. In *Proceedings of CHI*, ACM Press (2010), 1817-1820.
18. Steinfeld, A. (2010). Ethics and policy implications for inclusive intelligent transportation systems. *Second International Symposium on Quality of Life Technology*.
19. Steinfeld, A., Aziz, R., Von Dehsen, L., Park, S. Y., Maisel, J., & Steinfeld, E. The value and acceptance of citizen science to promote transit accessibility. *Journal of Technology and Disability*, 22, 1-2 (2010), 73-81.
20. Thiagarajan, A., Biagioni, J., Gerlich, T., Eriksson, J. Cooperative Transit Tracking using Smart-phones. In *Proceedings of SenSys*, (2010) ACM Press, 85-98.
21. Travelocity: <http://www.travelocity.com/>
22. Venkatesh, V., and Bala, H. Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences*, 39, 2 (May 2008), 273–315.
23. Von Ahn, L., Maurer, B., McMillen, C., Abraham, D. and Blum, M. Recaptcha: Human-Based Character Recognition Via Web Security Measures. *Science*, 321, 5895 (2008), 1465.
24. Wikipedia: <http://www.wikipedia.org/>
25. Yelp: <http://www.yelp.com>
26. Yoo, D., Zimmerman, J., Steinfeld, A., Tomasic, A., (2010): Understanding the space for co-design in riders' interactions with a transit service. In *Proceedings of CHI*, (2010) ACM Press, 1797-1806.
27. Zimmerman, J., Forlizzi, J. and Evenson, S. Research through Design as a Method for Interaction Design Research in HCI. In *Proceedings CHI*, (2007) ACM Press, 493-502.
28. Zimmerman, J. and Kurapati, K. Exposing Profiles to Build Trust in a Recommender. In *Ex. Abs. of CHI*. ACM Press (2002), 608-609.