Crowdsourced Mobile Data Collection: Lessons Learned from a New Study Methodology

Evan Welbourne, Pang Wu, Xuan Bao, Emmanuel Munguia-Tapia
Advanced Software Platform Lab
Future Information Technologies Center
Samsung Research America - Silicon Valley
1732 N 1st, San Jose, CA
{e.welbourne, pang.wu, x.bao, e.tapia}@samsung.com

ABSTRACT

In this paper we explore a scalable data collection methodology that simultaneously achieves low cost and a high degree of control. We use popular online crowdsourcing platforms to recruit 63 subjects for a 90-day data collection that resulted in over 75K hours of data. The total cost of data collection was dramatically lower than for alternative methodologies, with total subject compensation under $3.5K US, and a total of less than 10 hours/week spent by researchers managing the study. At the same time, our methodology enhances control and enables richer study protocols by allowing direct contact with subjects. We were able to conduct surveys, exchange messages, and debug remotely with feedback from subjects. In addition to reporting on study details, we also discuss interesting findings and offer lessons learned.

1. INTRODUCTION

Large mobile data sets are fundamental to research on activity recognition, crowdsensing, data mining, mobility, networking, and social networks among other areas. These data are gathered periodically through mobile data collection studies that recruit 10s or 100s of subjects to carry mobile devices with custom data collection software for a period of several weeks to several years.

However, these studies are high overhead, requiring significant time and budget for recruiting, screening, onboarding, and training subjects as well as for managing subject confidentiality agreements, concerns, equipment, and subject compensation. Recent studies such as the Lausanne Data Collection Campaign [11] and Social fMRI [2] painstakingly recruited and on-boarded subjects in person, compensating them with smartphones, the cost of the mobile subscription, and possibly with cash bonuses for surveys taken. Combined with months of study management tasks, this *locally administered* study methodology brings yearly costs for mobile data collection to well over $70K US for 100 or more subjects. Other studies scale-up at low cost with an *app store* methodology that uses an app store to distribute data collection software to 1000s of subjects [12]. This approach eliminates subject payments but also reduces control. Surveys and interviews are challenging, demographics and retention are unpredictable, and study protocols are inflexible. The cost of app development may also increase dramatically because the only incentive for subjects in this approach is use of the service provided by the app.

HotMobile'14, February 26–27, 2014, Santa Barbara, CA, USA.

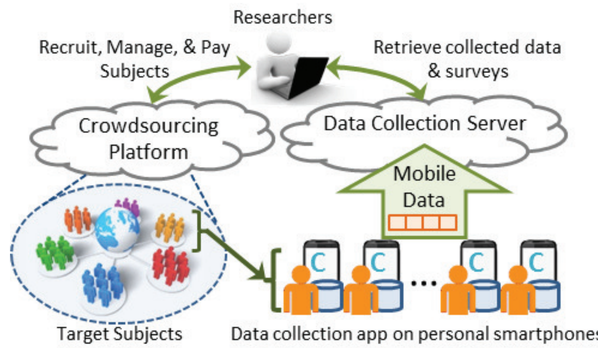

Figure 1: A crowdsourced mobile data collection system uses a crowdsourcing platform to recruit, manage, and pay subjects.

In this paper we explore a crowdsourcing data collection methodology that sits between the locally administered and app store data collection methodologies in terms of control, scale, and cost. As shown in Figure 1, we use popular crowdsourcing platforms to recruit, manage, and pay subjects to use their own smartphones to collect diverse, labeled mobile data, to complete several surveys, and to communicate via messaging systems. Collected data and survey responses are securely uploaded and stored on our server where researchers can access and analyze it throughout the course of the study. With a system prototype and two studies, we make several contributions:

- a review of commercial crowdsourcing that highlights platforms with legal contracts and business models that are amenable to mobile data collection studies;
- a low-cost mobile data collection approach that uses crowdsourcing to collect mobile data from a large number of participants while supporting a rich diversity of data types and maintaining researcher control;
- system prototype with a mobile client, a server, and popular crowdsourcing platforms Elance [8] and ODesk [14];
- detailed results from two pilot studies that show the feasibility and performance of our approach; and,
- lessons learned and guidance for future work on crowdsourced mobile data collection.

The remainder of this paper is organized as follows. In the next Section we discuss crowdsourcing platforms and their suitability for mobile data collection. Then in Section 3 we present our data collection system architecture followed by the design and detailed results from both of our pilot studies in Sections 4 and 5. We discuss lessons learned in Section 6, followed by related work in Section 7. Section 8 concludes by summarizing our findings and highlighting areas for future work.

2. CROWDSOURCING SUBJECTS

Crowdsourcing emerged in the last decade as an efficient way to harness the intelligence and creativity of crowds. More recently, a few researchers sought to apply crowdsourcing to human subjects research [15,19]. In this section we discuss the benefits and challenges of crowdsourcing applied to mobile data collection. We also review popular crowdsourcing platforms with focus on their potential for this type of human subjects research.

2.1 Benefits and Challenges

As noted, crowdsourcing has the potential to dramatically reduce the cost of managing data collection studies. Crowdsourcing platforms streamline recruiting by publishing study details to a large worker pool that may also be searched and filtered for a target subject demographic (e.g., location, occupation) [19]. Further screening through detailed profile matching and contractual agreements is done asynchronously and without in-person meetings. Equipment costs are eliminated because subjects use their own smartphones, and onboarding may be as simple as installing an app and taking online surveys. Unlike the app store methodology, the use of crowdsourcing platforms enables direct communication with workers through messaging and web conference, which can be used to conduct interviews and resolve bugs, disputes, or other issues during a study. Moreover, subjects are efficiently paid with internationally capable payment APIs, thereby ensuring compensation for their data and for participation in other study-related tasks. Finally, most platforms have a competitive worker marketplace that pushes costs down.

Despite its benefits, the crowdsourced approach faces significant challenges. First, the pool of workers may not contain subjects from the target population. For example, certain demographics are unlikely (e.g., wealthy businessmen), and others may be ineligible because they do not own smartphones (e.g., low-income workers). Second, it is difficult to remotely verify compliance with study protocol, and workers may be able to falsify information more easily online than in person. The principle challenge, however, is ensuring that a study is legally and ethically sound. Nations have dramatically different regulations governing collection, management, and retention of personal data. Separate legal contracts and supporting data management systems may be needed for each country from which subjects are recruited. In addition, while crowdsourcing platforms handle subjects' tax information on behalf of study administrators, internal review boards may still have doubts about the confidentiality and security of other sensitive data exchanged through the platform. Similarly, user profiles may prevent subject anonymity when it is desired.

2.2 Crowdsourcing Platforms

There are roughly three categories of paid crowdsourcing: domain-specific platforms, general purpose platforms for short-term jobs, and general purpose platforms for longer-term jobs. All follow a model where requesters post jobs for workers to complete in exchange for an agreed upon payment. However, the first two categories are inappropriate for mobile data collection studies. Domain-specific platforms such as CastingWords [6] have constrained worker demographics (e.g., transcriptionists), and general purpose platforms such as Amazon Mechanical Turk [3] do not support complex contracts or long term jobs in which workers download and install software. General purpose platforms for longer-term jobs do not share these restrictions and as such, they are currently the best match for mobile data collection. We focus on two such platforms in this paper: Elance and ODesk.

Table 1. Platform worker counts by category in Oct 2013.

Category	Elance	ODesk
Total	**2,886 K**	**365 K**
IT and Programming	806 K	216 K
Writing and Translation	805 K	114 K
Administrative Support	667 K	136 K
Design and Multimedia	577 K	96 K
Sales and Marketing	271 K	162 K
Financial and Management	171 K	52 K
Engineering and Manufacturing	125 K	-
Legal	37 K	-

Table 2. Platform worker distribution by region in Oct 2013.

Geographic Region	Elance	ODesk
North America	42 %	17 %
South Asia	31 %	39 %
East Asia	10 %	23 %
Western Europe	6 %	5 %
Miscellaneous	5 %	6 %
Eastern Europe	3 %	7 %
Australasia	2 %	1 %
Latin America	1 %	2 %

Elance and ODesk collectively retain 3over 3 million diverse, globally distributed workers (see Tables 1 and 2). The hiring process for each platform is similar. First, the requester posts a job description along with any required legal agreement, the estimated timeline, and the budget for payment. Then the requester adds the job to a category such as Admin Support, and tags it with required skills such as Data Entry or Research Support. On ODesk, requesters may optionally limit job post visibility to freelancers in a particular region such as North America. The requester then publishes the job post and may optionally invite any number of workers to apply. Interested workers apply to the job by replying to the post with a *bid* on the job. The bid includes a brief cover letter and the amount of payment he or she would be willing to complete the job for. Bid amounts are public and competition between workers pushes costs down. At any time, the requester may review the list of bids and select those that best match the job. The requester may also view public worker profiles which include their current rating on a 5-star scale, past earnings, and written reviews from past requesters. After accepting one or more workers for the job, the requester can communicate with workers via built-in messaging and video chat systems. On completion of a job milestone the requester can review the worker's progress and pay or propose new payment terms according to the work done. Elance and ODesk both take 10% of all payments to workers, but all other services are free. Any disagreements are settled with a dispute management system. Both platforms also support enterprise APIs and tools that facilitate management of large groups of workers (e.g., recruiting support, bulk payments).

3. SYSTEM ARCHITECTURE

As shown in Figure 2, the mobile data collection system for our pilot studies has three components: the client, the server, and the crowdsourcing platform. The mobile client is an app called Mobility Research (MR) that runs on Android phones. MR is built on the popular open-source tool Funf [9], to which it adds a user agreement as well as a basic questionnaire that asks subjects to categorize their contacts as coworkers, family, friends, or other. It is configured to launch on boot and collect data as shown in Table 3. The battery lasts 24 hours in this configuration on a new

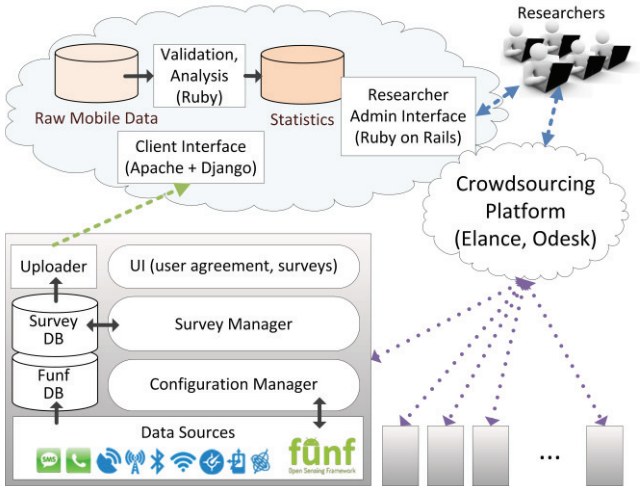

Figure 2: System architecture for the mobile data collection system including a mobile client, a server, and crowdsourcing platform.

Table 3. Collected data types, sample duration and interval.

Data Type	Duration	Interval
Communication logs (Calls, SMS)	-	5 hours
Contacts	-	5 hours
Battery	-	15 min
Apps Installed	-	12 hours
App Usage (launches/closes)	*event-triggered*	
Browser logs (searches, keywords)	-	5 hours
Location: GPS, Wi-Fi, GSM	60 sec	10 min
Radio scans: GSM, Wi-Fi, Bluetooth	0, 30 sec, 30 sec	10 min
Activity level based on motion	5 sec	5 min

Samsung Galaxy S3. To enhance privacy, MR uses a one-way hash on all human readable fields (e.g., app names, contact names, phone numbers, searches, URLs). Users can also turn off all data tracking in MR, or they can turn off individual sensors (e.g., GPS) through the Android interface to stop MR from tracking those sensors. MR uploads collected data to our server periodically whenever the phone is connected to Wi-Fi more than 15 minutes.

The server has three components: an HTTP server that accepts data from clients, a batch script that post-processes received data, and an administrative web interface for data review and analysis. The HTTP server is implemented using the Apache web server [5] with the Django framework [7]; it simply stores received data files in directories that correspond to the subjects who generated them. The batch script is written in Ruby and runs periodically to parse, validate, and compute descriptive statistics over received data. The administrative interface is implemented with Ruby on Rails [18] and displays survey results as well as statistics computed by the batch script. Researchers used this interface to review subject data quality and questionnaire responses. The server is hosted on Amazon EC2 [4] for scalability and reliability.

As discussed, we used Elance and ODesk as crowdsourcing platforms in these studies. Users were recruited, screened, managed, and paid using standard, free accounts. In the second study, we also used ODesk's bulk payment API which allows payment of multiple contractors by uploading one spreadsheet with payment information. We also received free guidance from Elance and ODesk sales representatives on how to optimize the wording, categorization, and pricing of job posts.

4. FEASIBILITY STUDY

The system was first tested in a feasibility study using Elance in June of 2013. The goal was to provide a proof-of-concept for crowdsourced mobile data collection as well as to work out any bugs in the mobile data collection system and study methodology.

4.1 Methodology

Only the Elance platform was used in this study. The job post, titled "Research Participants Needed for Smartphone Study", specified that workers download, install, and run MR on their personal Android phone for 30 days in addition to completing a brief on-device survey and a 2-5 minute web survey. The post also contained the user guide for MR which describes how to install and use the app as well as exactly what data would be collected and in what format. The payment amount was fixed at a maximum of $30 US. The Elance sales team advised that the job be categorized as Admin Support > Research. However, the job was visible to all workers and included searchable keywords: Admin Assistant, Data Entry, Research, and Android. Three prerequisites to apply were also listed: workers must be at least 19 years of age, consent to a legal agreement, and reside in the US.

Workers with lowest bids were hired first, after which they were on-boarded with links to download MR and take the web survey. The survey collected demographic information, Elance and ODesk usernames and device IMEIs. MR installation was verified by cross-referencing the IMEI with received client data on the server. The on-device survey labeled subjects' contacts as Colleague, Family, Friend, or Other and could be completed any time during the first week. Subjects could ask questions on any issues or concerns during this initial period. Once on-boarding was complete, subjects were asked to run MR, having their phone connected to Wi-Fi at least 15 minutes each day. After 30 days, subjects uninstalled MR and were paid according to their initial bid amount. The messaging system was used to discuss any problems or concerns that arose during the study.

4.2 Results

Within 3 days of the job post, we received a total of 23 eligible bids, 3 from men and 20 from women. This disparity is likely due to posting under Admin Support, in which a majority of workers are women. Despite a US residency requirement, we also declined 5 bids from non-US applicants. After the first day of recruiting the number of bids was lower than expected; our Elance contact noted that US workers often ignore lower paying jobs (e.g., $30 US). We were able to more effectively recruit workers by *inviting* them to the job, a feature supported by both Elance and ODesk. We invited a total of 65 workers from the Admin Support category; 20 submitted bids and 45 declined or did not respond. Among those who declined, 20 did not own an Android phone, 11 were too busy with other contracts, and 3 were concerned about privacy. Of the workers concerned about privacy, all were female and 2 said they would participate if location tracking was removed.

From the 23 bids, we hired 2 male and 8 female subjects. Most were able to complete surveys, install MR, and upload data within 2 days. Several subjects experienced difficulties during install due to inconsistencies between the MR user guide and their particular version of Android. Others helped us identify and fix a bug in the way our client uploads data. Subjects with hundreds of contacts complained that the on-device survey required them to label all contacts in one time-consuming session. We also noticed that while 8 subjects provided continuous data, 2 turned off tracking for between 30% to 40% of the time.

Table 4. Recruiting results summary for Elance and ODesk.

Category	Elance	ODesk
Total rounds of mass invites	6	3
Total invited	545	300
Total bids	52	55
Bids within 1 / 5 / 10 days of invite	37 / 46 / 52	43 / 50 / 55
Male / Female bids	6 / 46	11 / 44
Total subjects recruited	33	30

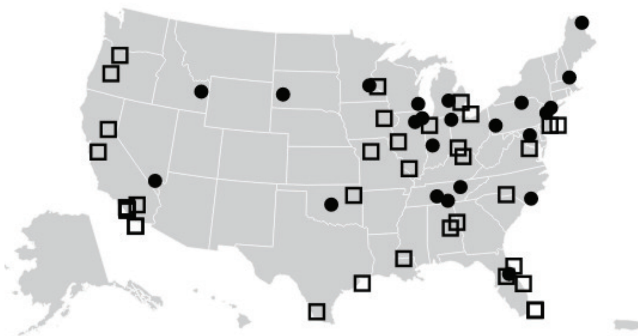

Figure 3: Locations of recruited subjects. Squares are workers from Elance and circles are workers from ODesk.

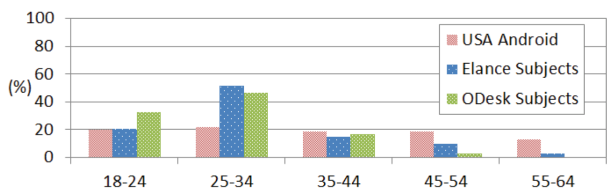

Figure 4: Age distribution over US Android smartphone owners as well as subjects recruited from Elance and ODesk.

5. PERFORMANCE STUDY

Encouraged by the results of the feasibility study, we refined our system and methodology to conduct a second pilot study on the performance of the crowdsourcing methodology. In particular, we assessed our approach in terms of recruiting speed and diversity of recruited subjects, data quality, management overhead, and cost. This study was conducted between July and October of 2013.

5.1 Methodology

The methodology was adjusted based on the feasibility study results. First, the job title was changed to "Android Smartphone Data Collection" to emphasize the Android requirement. We also set the maximum payment to $90 US to attract more workers. However, the study duration was increased to 90 days as well. Moreover, to compensate for partial data from subjects that turn off tracking, payment was pro-rated based on the amount of data received. For example, if a worker bid $90 but uploaded only 45 days (1080 hours) of data, the final payment would be $45. During recruiting, more effort was put into invitations, and the MR user guide was clarified to facilitate on-boarding. The on-device survey was also updated to allow incremental contact labeling. Finally, a web survey on privacy was added.

5.2 Results

In 90 days, 63 crowdsourced subjects provided over 75K hours of diverse mobile data. The total subject compensation was less than $3.5K US, and researchers spent under 10 hours/week managing the study. We present and analyze these results in detail here.

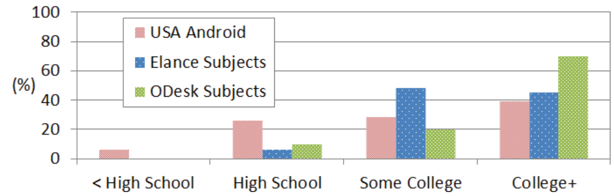

Figure 5: Education distribution over US Android smartphone owners as well as subjects recruited from Elance and ODesk.

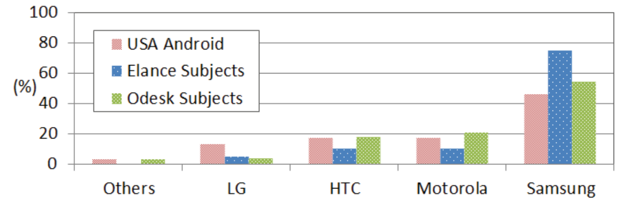

Figure 6: Distribution of smartphone manufacturers.

(a) Elance Categories **(b) ODesk Skills**

Figure 7: Word clouds for subject worker category and skills.

5.2.1 Recruiting

The recruiting performance is summarized in Table 4. Overall, we received 107 bids from which we recruited 66 subjects. Bids were in response to several rounds of invites which had response rates of 10% and 18% for Elance and ODesk respectively. Over 70% of bids arrived within 24 hours of an invite. This was a dramatic return on the 3-5 minutes spent sending each batch of 100+ invites using Elance and ODesk invitee recommendations. There was also a steep drop-off in bids 24 hours after invites were sent; this underscores the role of invites for efficient recruiting.

There was substantial diversity among recruited subjects. Figure 3 shows that the geographic distribution of subjects from both platforms roughly matches population density in the US. Figures 4, 5, and 6 summarize the distribution of age, education level, and smartphones by manufacturer among subjects. The plots also compare subject distributions with distributions over all US Android smartphone owners [13,20]. The word clouds in Figure 7 show the most common worker categories and skills for subjects.

Though the subject sample size is small, the data show biases toward certain demographics. Among bidding workers there was a strong bias toward women. This trend may stem from the platform-based invitations which only recommend workers in the job category, Admin Support, which has significantly more female than male workers. However, some work has also shown that women are more likely both to work on a crowdsourcing platform, and to participate in human subjects research [19]. The plots also show that the distributions for recruited subjects are similar to US distributions, but with more college degrees and more people in the 25-34 year old age group. However, there are workers in most demographic categories and it may be possible to shape the subject population through careful recruiting.

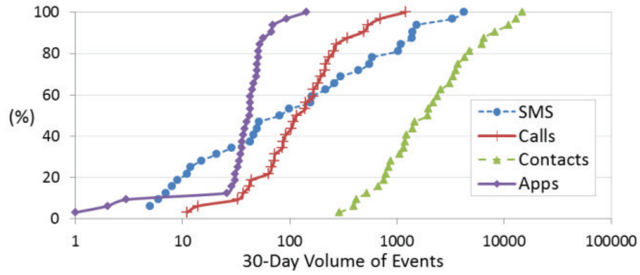

Figure 8: CDFs for number of Contacts as well as volume of SMS, Call, and App Usage events over all subjects for 30 days.

5.2.2 Data Volume

We received more than 20 GB and over 75K hours of data from the 63 subjects each month. The average number of hours logged per day was 18. Figure 8 shows CDFs for the number of data events for a 30 day period across all subjects, the x-axis is log-scale. It is notable that the volume of events varies dramatically across users. An analysis of subject mobility patterns revealed that subjects visited an average of 7 places each week, with a standard deviation of 2. This level of mobility is not inconsistent with findings from other mobility studies [11] and suggests that this crowd worker population is not unusually stationary.

All received data were correctly formatted and no data was lost due to file corruption or system failure. However, 19 subjects uploaded incomplete data streams with long gaps throughout. In 3 cases where gaps lasted several days, we found that the subject could not connect to Wi-Fi due to travel or other circumstances (e.g., power outage from a blizzard in South Dakota), but missing data were recovered after connectivity was reestablished. Another subject with a 2010 Android model had an out-of-memory error when uploading a large file, but this data was recovered after patching the client. Most gaps lasted a few hours and were caused when subjects powered off their phone or disabled MR tracking. Subjects who turned off their phone explained that it was to accelerate charging. Those who disabled tracking explained that they did so either for privacy reasons or to conserve battery.

Given the variation in data volume among subjects, it is natural to consider whether volume can be predicted by a worker's profile. Correlation was computed between data volume (i.e., full, sparse) and profile characteristics such as 1-5 star rating, number of jobs completed, hours worked on other jobs, and total payments earned on the platform. No significant correlation could be found between these attributes and uploaded data volume. In fact, the most reliable predictor of volume, with a correlation of 0.36, was the amount of data uploaded in the first week of study.

5.2.3 Cost

Budgetary issues constrained hiring to the 63 lowest bids. The average bid among this group for Elance was $69.71 ($0.77/day) with standard deviation of $11.83; the average bid for ODesk was $79 ($0.88/day) with standard deviation of $11.32. Median bids were $72 for Elance and $81 for ODesk. However, since payments were pro-rated by the amount of data actually received, the actual cost was much lower for subjects with low data volumes. In particular, we paid only 69% of a subject's bid on average. This translates to a total cost of about $3200 for a 3-month data collection with 63 subjects, a significant savings compared to the cost of a study at the same scale using the locally administered study approach. It may also be possible to reduce costs by lowering the suggested payment on which bids are based.

The other key cost of a data collection study is management overhead for researchers. In our approach, researchers had to spend time recruiting, hiring, managing, and paying users. Our team tracked precise hours spent on the Elance and ODesk websites using a simple timer app. The combined total time spent by researchers administering the study each week was between 5 and 7 hours. Days of additional time spent debugging and refining the system are not included in this calculation. As noted in Section 2.1, equipment costs are eliminated with our approach.

6. LESSONS LEARNED

The study results show that crowdsourced mobile data collection is both feasible and cost effective, but the studies also revealed several interesting insights and areas for improvement. In this section we discuss these lessons learned.

6.1.1 Review Legal Agreements Early and Often

Any crowdsourced study must comply with both the subject agreement and the crowdsourcing platform's terms of service. We found that changes in study protocol (e.g., increased duration, pro-rated payment) often required accompanying changes to the subject agreement. Accidentally on-boarding a subject with the wrong agreement could legally bind researchers to carry out the study differently with different subjects. In-depth understanding of platform terms of service is also mandatory. Our initial prototype used Amazon Mechanical Turk for crowdsourcing until the job was banned for violating the terms of service.

6.1.2 Invest Significant Resources in Recruiting

As highlighted in Section 5.2, recruiting can be done efficiently with crowdsourcing, but must also be done carefully to shape the subject demographics. As a follow-up experiment, we hired a 5 star admin support contractor for $8/hour from ODesk to recruit and on-board a subject population with 7 men and 7 women, each having a different skill set. The contractor was able to recruit and on-board the subjects over 3 days with less than 4 hours of work.

6.1.3 Adapt Payments to Each Subject

The study prices the mobile data in Table 3 at $1/day, but does so based only on conjecture and budget allocations. The real price is unknown and may vary dramatically by user and situation. As such, it is most efficient to pay users based on how much they value their data and how much they contribute. The bidding mechanic provides one way to adjust price toward a practical value, but bid amounts are still based on the suggested job price, and refer to aggregate data. It may be possible to implement dynamic payment schemes with fine-grained data pricing. For example, if a subject bid a rate of $0.80/day, then data containing location could be worth an additional 5% ($0.04/day); if data is labeled by the subject then it could be worth an additional 10% ($0.08/day). Then a subject could dynamically trade-off privacy and payment by opting in or out of data streams as their situation varies. Another insight is that data collected while subjects sleep may not be worth as much as data from waking hours.

6.1.4 Design to Fail Gracefully

Bugs in software and study protocol are common even after pilot studies. Some bugs didn't surface until the client was deployed on diverse hardware, making it difficult to debug remotely. We found the Application Crash Reports for Android (ACRA) [1] library for remote crash logging and reporting to be exceptionally useful in this situation. However, two bugs forced us to request that all subjects update software by hand which was time consuming. A better alternative would be to include an auto-update feature that can push bug fixes. Subject contracts should also allow a low cost and low risk way to settle disputes with subjects. Two subjects were asked to leave the study after providing almost no data, but there was no clause in the contract to support this situation.

6.1.5 Maintain Worker Relationships
It is common to follow-up with subjects after a study to conduct additional surveys and interviews, or invite them to another study. As such, it is useful to keep in touch with workers that provided good data or that belong to a target demographic. Both Elance and ODesk workers can be bookmarked for re-hiring. However, future studies may achieve faster, more targeted recruiting by building a separate database on worker metadata (e.g., data quality, demographics, phone model). In addition, since workers decide to apply partly based on employer reputation [15,19], it is critical to manage relationships and request feedback at the end of a study.

6.1.6 Secure Data and Support Privacy
Data security is a concern in all human subjects research. In both studies, data was transmitted securely to the server, from which it was moved to secure storage once a week. There were no records of security breaches during the study, but there was an incident in which a subject attempted to falsify data. This event was detected as a combination of repeated timestamps and an invalid IMEI. The fraudulent data was removed and the subject was warned, but allowed to continue in the study.

As noted previously, privacy concerns played a significant role in our study. Tens of invited workers declined for privacy reasons, and subjects periodically disabled tracking to protect privacy. The web-based privacy survey gave more insight. Subjects reported that the most sensitive types of data were location, SMS, Calls, and browser logs; app usage was ranked less sensitive. Indeed, subjects said they would sell 30 days of app usage data for under $2, while 25% wanted over $4/month for browser logs. Subjects also wanted to review collected data before and after upload.

6.1.7 Know the Limitations
Our methodology has several clear limitations. First, no custom hardware can be deployed and no in-person technical support is possible. In addition, while the recruited population seems to be fairly diverse, studies with very specific target demographics (e.g., seniors that practice yoga, groups of 10 friends or more) may be difficult or impossible to find. This does not preclude externally recruiting such subjects and then managing and paying them through a crowdsourcing system however.

7. RELATED WORK
Mobile data collection is fundamental to many research areas and is a challenging activity for researchers. Recently successful studies have approached the task through large-scale local administration [2,11,16] and through app store based distribution [12]. These approaches offer opposite extremes in terms of study control, cost, and scale. The crowdsourced approach offers an alternative middle ground. Many studies have also been conducted to study worker demographics on crowdsourcing platforms and even to assess the feasibility of recruiting workers for human subjects research [10,19,20]. However, none of this work examines mobile phone based research or data collection.

8. CONCLUSION
We explored a new mobile data collection methodology that sits between the locally administered and app store-based study methodologies in terms of control, scalability, and cost. We showed that this approach is feasible. Moreover, by implementing a prototype and conducting two pilot studies, we recruited 63 subjects for 90 days to collect over 75K hours of data. The total cost was dramatically lower than alternate methodologies, with total subject compensation less than $3.5K US, and less than 10 hours/week spent by researchers managing the study. We also discussed lessons learned and shared insight on how to optimize future crowdsourced mobile data collection.

9. ACKNOWLEDGEMENTS
Thanks to Annie Luo, Zhixian Yan, and Jun Yang for their excellent adaptation of the Funf software for our client. Thanks also to the sales teams at Elance and ODesk for their support.

10. REFERENCES
[1] ACRA https://github.com/ACRA Oct 2013

[2] Aharony, N. et al. Social fMRI: Investigating and shaping social mechanisms in the real world. Pervasive and Mobile Computing 7 643-659 (2011)

[3] Amazon Mechanical Turk. http://www.mturk.com Oct 2013

[4] Amazon EC2. http://aws.amazon.com/ec2/ Oct 2013

[5] Apache. http://httpd.apache.org/ Oct 2013

[6] CastingWords. https://castingwords.com/ Oct 2013

[7] Django. https://www.djangoproject.com/ Oct 2013

[8] Elance. http://www.elance.com Oct 2013

[9] Funf Open Sensing Framework. http://funf.org/ Oct 2013

[10] Kocchar, S. et al. The Anatomy of a Large-Scale Human Computation Engine. KDD-HCOMP '10, 2010

[11] Laurila, J.K. The Mobile Data Challenge: Big Data for Mobile Computing Research. Mobile Data Challenge Workshop in conjunction with Pervasive '12, 2012

[12] Miluzzo, E. et al. Research in the App Store Era: Experiences from the CenceMe App Deployment on the iPhone. Ubicomp '10, 2010

[13] Nielsen. Who's Winning the U.S. Smartphone Market? http://www.nielsen.com/us/en/newswire.html Aug 2013

[14] ODesk. http://www.odesk.com Oct 2013

[15] Paolacci, G. et al. Running experiments on Amazon Mechanical Turk. Judgment and Decision Making, 5(5), 2010

[16] PhoneLab. http://www.phone-lab.org Jan 2014

[17] Ra, M. et al. Medusa: A Programming Framework for Crowd-Sensing Applications. MobiSys'12, 2012

[18] Ruby on Rails. http://rubyonrails.org/ Oct 2013

[19] Schmidt, L.A. Crowdsourcing for Human Subjects Research. CrowdConf 2010, 2010

[20] Smith, A. Smartphone Ownership – 2013 Update. Pew Research Center Report, Jun 2013