10-2013

LiveLabs: Initial Reflections on Building a Large-scale Mobile Behavioral Experimentation Testbed

Archan MISRA
Singapore Management University, archanm@smu.edu.sg

Rajesh Krishna BALAN
Singapore Management University, rajesh@smu.edu.sg

DOI: https://doi.org/10.1145/2557968.2557975

Published in ACM SIGMOBILE Mobile Computing and Communications Review, Volume 17 Issue 4, October 2013, Pages 47-59.
http://doi.org/10.1145/2557968.2557975

LiveLabs: Initial Reflections on Building a Large-scale Mobile Behavioral Experimentation Testbed

Archan Misra **Rajesh Krishna Balan**
archanm@smu.edu.sg *rajesh@smu.edu.sg*
School of Information Systems
Singapore Management University, Singapore

We believe that, for successful adoption of novel mobile technologies and applications, it is important to be able to test them under real usage patterns, and with real users. To implement this vision, we present our initial effort in building LiveLabs*, a large-scale mobile testbed for in-situ experimentation.* LiveLabs *is unique in two aspects. First,* LiveLabs *operates on a scale much larger than most research testbeds— it is being deployed in four different public spaces in Singapore (a university campus, a shopping mall, an airport and a leisure resort), and is expected to have a pool of over 30,000 opt-in participants. Second,* LiveLabs *not only instruments smartphones and the infrastructure to gather deep individual and collective context, but also provides a unique experimentation platform that automates many aspects of behavioral experimentation, such as subject selection and context-triggered delivery of interventions. We briefly describe some of the research challenges associated with building such a large-scale deep-context collection testbed, as well as the current status of* LiveLabs*. We then share our perspectives on the challenges of setting up and operating such testbeds, with the expectation that our experiences will prove useful to other researchers looking to build similar testbeds elsewhere.*

I. Introduction

The building and operation of large-scale experimental testbeds has gained a lot of prominence in recent years, as exemplified by initiatives for testbed federation such as GENI [7]. Most of the first wave of testbeds, such as ORBIT [14], Kansei [1] and WiseBed [6] had a technology focus— geared to move the study of networking protocols, architectures and algorithms beyond simulations, but still confined to university campuses and using synthetic workloads. More recently, efforts such as PhoneLab [11] and NetSense [15] have begun to open up opportunities for testing mobile technologies and applications using real-world users. We believe that mobile systems and applications research would greatly benefit from more-realistic behavioral testbeds, where prototype applications and services can be tested on *a diverse set of users*, performing *real-world activities*, in *diverse locations.*

Motivated by the desire to support such real-world, mobile-centric behavioral investigations, the LiveLabs Urban Lifestyle Innovation Platform (or *LiveLabs* for short) is a multi-year testbed effort funded by Singapore's Media Development Authority (MDA), situated in multiple urban spaces in Singapore. *LiveLabs*' central goal is not to support technical trials of individual technologies, but instead support experimental investigation of *human adoption and usage* of advanced mobile applications and services. Motivated by the emergence of mobile sensing as an enabler for increased, real-time context awareness, *LiveLabs* especially seeks to enable experimentation with advanced context-aware services.

We believe that *LiveLabs* is unique from other testbeds for three key reasons:

- *Participant Scale: LiveLabs* targets an opt-in participant base of tens of thousands of users (with a notional target of 30,000 participants), restricted not just to specific affinity groups (e.g., students on the SMU campus), but members of the general public, engaged in daily lifestyle activities.
- *Real-life Venues: LiveLabs* encompasses four publicly-accessible venues in Singapore: the entire SMU campus, a major shopping mall, an internationally reputed airport and a highly popular leisure resort. The goal is to experimentally study human interactions and application usage *in the wild*, while they engage in everyday tasks such as shopping, dining or commuting.
- *In-situ Experimentation:* This is perhaps the most interesting capability of *LiveLabs*: it is not just an infrastructure for passively collecting usage data, but for actively performing *in-situ* context-based *interventions* (e.g., offering a special discount to shoppers of a certain group size) , *while the participants are engaged in relevant real-world activities*. *LiveLabs* thus transforms the mobile device from being merely an observer of human context to an enabler of behavioral/sociological experiments on an unprecedented scale.

Building the *LiveLabs* infrastructure itself requires deep research technical innovations in several aspects of mobile computing (such as energy-efficient context collection, real-time stream analytics, privacy preservation etc.). However, *LiveLabs*' emphasis on end-user experimentation with advanced, but real, applications and services implies that the vast majority of industry experimenters using *LiveLabs* are not technology companies, but *consumer services* companies, with interests in verticals such as retail, advertising, social media, leisure and tourism. These companies effectively view *LiveLabs* as a provider of a "dynamic user panel", with an extended set of deep context attributes, that enables them to conduct consumer experiments at a finer granularity than otherwise possible.

In this paper, we shall provide a high-level description of *LiveLabs*, including its goals, the selected venues and the overall system architecture. We shall then briefly describe several of the key research challenges, and provide some high-level insights about our approaches to tackle several of these challenges. However, the sobering reality is that the the bulk of our effort while conceptualizing and launching *LiveLabs* over the past two years has not been about cutting-edge research, but about building and sustaining an operational testbed, including addressing mundane but key issues such as participant sign-up & retention and ensuring reliable system-availability for our participants. The main goal of this paper is therefore to share with the community our experiences in both building up the *LiveLabs* eco-system (especially in our recruitment of willing venue owners & participants) and in setting up and operating the first location of the testbed at our university campus. We believe that our experiences, of both successes and failures, provide "teachable insights" to our mobile systems fraternity, especially as more of us seek to build and operationalize large-scale experimental testbeds.

The rest of the paper is organized as follows. Section II uses a simple motivating example to describe the key capabilities of *LiveLabs*, and then describes the resulting functional components needed to operationalize the testbed. Section III then describes the four different venues for *LiveLabs* and how each venue helps experimental studies for different business domains, as well as shares our experiences and insights in securing appropriate collaborative agreements. Subsequently, Section IV provides a high-level overview of some of the key and practical research areas that need to be addressed to make *LiveLabs* successful. Section V then provides some operational details about the current state of the *LiveLabs* infrastructure, while Section VI then encapsulates our observations and experiences about the journey so far. Finally, after a survey of related work on recent testbed efforts in Section VII, Section VIII concludes the paper.

II. Functional Components of LiveLabs

LiveLabs's high-level goal is to enable realistic testing of context-aware mobile services. To understand the various functional requirements of *LiveLabs*, we first describe a prototypical use case of *LiveLabs*:

> *Imagine a mall-based café, BigBucks, that would like to test an innovative new offering, targeted towards groups of 5 or more customers. BigBucks would like to offer a promotion of "30% off the coffee bill" to such groups, and would especially like to target such groups right after they come out of the movie theater, but also deliver the incentive only if such a group has lingered on outside the movie theater for at least 2 minutes or has joined a competitor's queue with an expected wait time larger than 5 minutes.*

This use case is predicated on BigBucks' belief that it may be more profitable to larger groups of movie-goers rather than individuals, and that it makes sense to target only those groups that are not in a rush, but appear to be debating about what leisure activity they might want to do next. *LiveLabs's* would like to make it easier to experimentally investigate this promotional strategy, by not only providing appropriate subjects, but also detecting the appropriate contextual triggers and delivering the subsequent context-based interventions.

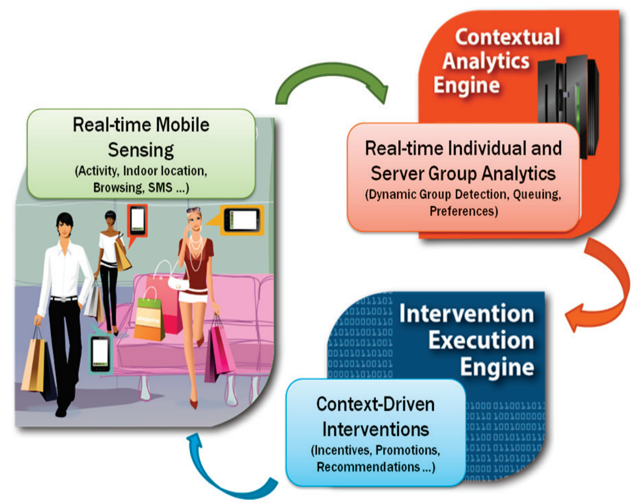


Figure 1: *LiveLabs*: The Sensing-to-Experimentation Loop

To achieve this, *LiveLabs* needs to support 3 key functional capabilities (these are illustrated in Figure 1):

i) *Sense:* Capture data about the real-world activities of individual participants through appropriate mobile sensing. In the example above, such mobile sensing data can provide information about the in-mall location of all the participants, their ongoing activities (such as walking rapidly vs. queuing at another stall).

ii) *Analyze:* Apply deep, near-real time analytics to extract higher level individual and ambient context. In the example above, such analytics would help identify groups of mall visitors, infer the queuing delays at a competitor's store or indicate that a specific group is in a rush.

iii) *Intervene:* Utilize such context to activate appropriate context-aware interventions or application adaptation. In the example above, such intervention would consist of delivering, at the appropriate time, the "30% off promotion" to the mobile devices of group members that satisfy the contextual predicates. Perhaps equally importantly, *LiveLabs* would continue to monitor the outcome of such an intervention–i.e., observing if such a targeted promotion actually causes groups to gravitate towards BigBucks or not.

II.A. Functional Components and Data Flow

Figure 2 shows the corresponding components, and data flow sequence, of *LiveLabs*. These components work together to achieve the in-situ, context-based intervention outlined before in the following manner:

1. The *LiveLabs Context Collector* application executes in the smartphone of each of the participants, collecting both sensor & usage data from each device. For *LiveLabs* to work, each participant must download and install the Context Collector service, and give it the necessary permissions to collect the relevant data feeds. We emphasize that this download mechanism is completely opt-in, and requires the participant to explicit consent to our collection and privacy policies. To provide a flexible yet power efficient data collection process, the *LiveLabs* Context Collector can be dynamically configured to turn on / off different data collection modules (GPS, accelerometer, etc.) and to modify the data upload frequency.

2. The Context Collector then transmits (at appropriate intervals) the collected data (or context derived from such low-level data) to our *Storage and Analytics* server. There, the data is first stored for archival purposes (after being anonymized to remove various sensitive fields) and concurrently streamed in to our *Stream Analytics Engine*. This analytics engine applies a variety of standard and advanced analytics algorithms to derive higher-level context, such as the queuing delays at different locations.

3. Independently, experimenters (such as BigBucks) wanting to use *LiveLabs* specify their experiments using an appropriate experiment specification interface on our *Behavioral Experimentation Platform (BEP)*. More specifically, experiments contain a *trigger* component, describing the contextual predicates that must be satisfied, and an *action* clause, that specifies the specific intervention that must be delivered. Such interventions or experiments could involve, for example, delivering specific promotions/discounts, triggering specific features of an already-installed mobile application or pushing out an HTML5-based survey.

4. The *Intervention Execution Engine* in the BEP then continually monitors the *trigger* specifications of each executing experiment, and subscribes to the Stream Analytics Engine for notifications of matching context. When the context is met, the Analytics Engine notifies the Intervention Execution Engine, which is then responsible for performing the intervention specified in the *action* clause.

5. The BEP continues to then monitor the subjects of the intervention to observe their response to the experiment stimulus. Once sufficient observational fidelity has been achieved (this could, for example, take several weeks), the BEP reports back the results of the experimental intervention to the experimenter, who can then subsequently choose to run modified versions of the same experiment.

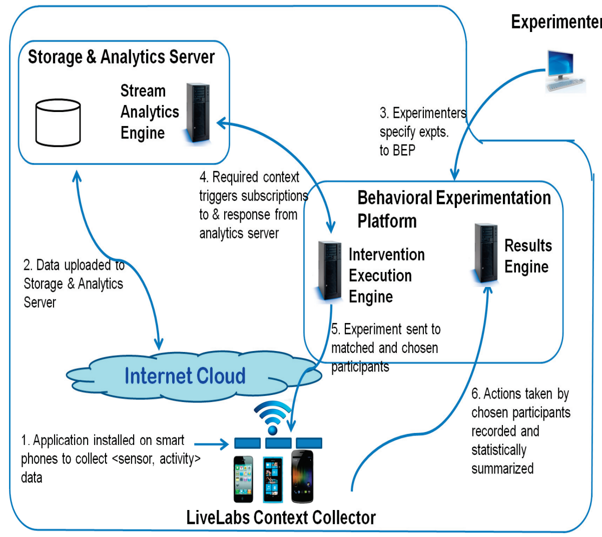


Figure 2: *LiveLabs* components and interaction flow

III. Testbed Venues & Application Domains

LiveLabs was created as part of a larger, national-scale effort by the MDA to position the city-state as a "live observatory", where inhabitants would be voluntarily coopted as participants to enable at-scale, *naturalistic* testing of various innovations in the areas of telecommunications and digital media. Singapore is, in fact, an ideal venue for testing *urban usage* of digital media and mobile computing innovations— it offers many examples of densely-populated public urban spaces (e.g., shopping malls, train stations, convention centers and theme parks), has one of the world's highest penetration of smartphones and a tech-savvy, highly-educated population. Accordingly, from the outset, *LiveLabs's* charter was to create a testbed that extended beyond university campuses to encompass various public urban spaces, where participants would be expected to spend significant portions of their daily lives and which would prove to be compelling for various commercial partners. We spent over 18 months in conceptualizing *LiveLabs* capabilities, and in "recruiting" venue operators and technology partners whose collaboration would be indispensable for both recruiting participants and in operating the testbeds.

LiveLabs's goal is to operate a testbed across four public venues/spaces in Singapore, offering an unprecedented diversity in both the types of testbed locations and the profiles of the participants:

1. *LiveLabs@SMU* refers to the entire SMU campus, located in downtown Singapore and consisting of 5 academic buildings, housing approximately 8000 full-time students spread across 6 schools. Each building is 5 stories high, and the buildings are all connected by an underground, publicly-accessible concourse that houses an assortment of retail outlets, a subway station, student extra-curricular facilities and administrative offices.

2. *LiveLabs@Mall* refers to a large, 9 story mall located close to the SMU campus— besides retail establish-

ments, the mall is integrated with one of Singapore's busiest subway hubs. The mall floors cover an area larger than 1 million sq. ft and are visited by at least 40,000 people every day. Besides typical F&B outlets and retail shops, the mall also houses a large food court, several bank branches and customer service centers.

3. *LiveLabs@Airport* refers to Singapore's main airport, consisting of 3 separate, but inter-connected, terminals. The terminals collectively have an area of over 10,000 sq. ft and serves more than 135,000 passengers daily. Besides passengers, many local inhabitants visit the airport's retail outlets on a regular basis.

4. *LiveLabs@Sentosa* refers to Singapore's premier resort island, spread over more than 45 million sq.ft. It is visited by over 45,000 visitors daily, with a majority being foreign tourists.

It is important to note that, of the four locations, only *LiveLabs@Sentosa* is predominantly outdoors–the others are primarily air-conditioned, pressurized indoor spaces. All three indoor public spaces have Wi-Fi coverage (although of differing quality and with different access restrictions). *LiveLabs@Sentosa* has Wi-Fi hotspots at selected locations throughout the island–this is especially relevant from our standpoint as most of the visitors to *LiveLabs@Sentosa* are foreign tourists, and are thus unlikely to avail of cellular connectivity (due to steep roaming charges).

Figure 3 illustrates these four testbed venues, as well as the applications and service verticals (to be discussed next) primarily associated with each such venue.

LiveLabs@ Application Domain & (*Example Services*)	SMU	Malls	Airport	Sentosa
Retail & Advertising (*Context-aware Promotions*)		P	P	
Leisure & Tourism (*Dynamic Itineraries*)				P
Rich Media (*HD Mobile Gaming*)	P			
Logistics & Operations (*Smart Passenger Alerts*)			P	P
Education & Pedagogy (*Interactive Classrooms*)	P			
Behavioral Social Sciences (*Content Recommendation*)	P			

P: Primary focus of testbed

Figure 3: LiveLabs Venues and Key Verticals

III.A. Key Application Domains

These four venues are not just geographically distributed across Singapore, but were carefully chosen to offer interesting experimentation possibilities for at least five distinct domains:

1. *Retail & Advertising:* Both *LiveLabs@Mall* and *LiveLabs@Airport* have a lot of retail establishments and offer the ability to test new forms of pervasive retail applications and services. Candidate examples include the ability to test the efficacy of context-aware customized incentives (e.g., "give a buy-2-get-1-free discount to a set of 3 friends shopping together") or product recommendations (e.g., "highlight the clothing products on the aisle that have seen the highest visitor footfalls in the last 4 hours").

2. *Leisure & Tourism:* The *LiveLabs@Sentosa* testbed is especially well suited for testing new types of context-aware services that improve a visitor's experience at a leisure resort. Candidate examples include the ability to offer adaptive itineraries (e.g., "suggest a different order of attractions, based on real time estimates of queuing delays at these attractions" or investigate dynamic pricing and incentive strategies (e.g., "offer a family group a special discount at a family-oriented attraction, if it is currently uncrowded").

3. *Rich Media Consumption:* The *LiveLabs@SMU* testbed will be specially instrumented to offer a high-bandwidth, usage-adaptive wireless network that enables testing of various types of rich-bandwidth applications, under high occupancy densities. Unlike the other locations that will principally utilize existing network deployments, *LiveLabs@SMU* will include experimental access network prototypes, such as TV Whitespace and small-cell (i.e., femtocellular) deployments. Such an infrastructure allows experimenters to test advanced forms of mobile media consumption (e.g.,"understand the performance of Wi-Fi networks for supporting live HD video streaming to 50 students in a single lecture hall") as well as real-time content distribution (e.g., "investigate peer-to-peer strategies for low-latency, interactive mobile gaming").

4. *Pedagogy:* Besides an advanced, experimental network substrate, *LiveLabs@SMU* also offers a unique set of participants–undergraduate and post-graduate students at a tertiary institution, interacting with one another within a relatively compact, urban milieu. This offers an ideal platform for testing newer forms of interactive teaching (e.g., "set up dynamic groups within a classroom for collaborative problem-solving during lectures"), as well as the efficacy of pedagogical interventions (e.g., "testing the impact, on eventual grades, of real-time reminders issued to students who are observed to be spending too much of their time in study areas interacting in overly large groups").

5. *Operational Logistics: LiveLabs@Airport* is not just a major retail destination–it is also a functional airport, with highly critical operational challenges such as guiding passengers to their flight gates, managing the utilization of airplane parking bays and ensuring passenger adherence to safety and security guidelines. The availability of real-time context about passenger and staff movement and behavior inside the

terminals will allow experimental study of new strategies for passenger coordination (e.g., "dynamically re-prioritizing those passengers at security queues whose flight is observed to have commenced boarding") and security monitoring (e.g., "identifying passengers or staff who may have strayed back into secure areas without going through appropriate screening").

6. *Social Sciences:* While *LiveLabs* was principally intended to serve as an accelerator for translating mobile technologies to commercial use, we subsequently realized that *LiveLabs* also serves as a unique *social observatory*, providing us with hard-to-obtain insights into the interaction patterns among individuals in *both* the digital and physical world. This capability turns out to be of deep interest to our academic colleagues in the disciplines of computational social science and behavioral sciences, as it enables them to study and model properties such as content propagation (e.g., "how do social media messages percolate in the physical world") and community behavior (e.g., "how do online communications correlate with physical world interactions"). This promises to be a game changer for these disciplines, as they can now migrate their observational and experimental strategies from traditional simulated laboratory settings to the real world.

We should emphasize that these are just high-level characterizations of the principal functions of each testbed. In reality, each location is diverse enough to permit experimental studies beyond this narrow categorization. For example, *LiveLabs@SMU* has a rich mix of retail operations as well (e.g., high-end fashion stores, foodcourts, travel agencies, pharmacies, etc.) and is already being used to conduct studies in the effectiveness of on-campus advertising and promotions. Likewise, *LiveLabs@Sentosa* also has a significant number of retail and F&B partners (thus supporting novel retail & advertising services), and also operates several inter-island bus services (thus enabling investigations for improved operational logistics).

III.B. The Art of Securing Venue Partners

The scope of *LiveLabs* may appear to be ambitious, and it would truly be an understatement to say that the process of recruiting venue partners was a long and complicated one! We now summarize some of the key insights and experiences we gained through this process.

Focus on simple capability demonstrators: There is often a mismatch between the "big picture" vision that we academics espouse and the operational realities and problems faced by the venue operators. In our initial interactions, we could feel a disconnect between our abstract, conceptual vision and the shorter-term, tangible aspirations of the venue onwers. Moreover, in all 3 external (i.e., non-SMU) venues, information technology is viewed as an enabler of business capabilities. Accordingly, to establish a successful collaborative relationship, it was important to provide evidence of not just our technical capabilities, but of the relevance of such capabilities to the core *business needs* of the venue operator. We discovered that the most effective way to gain their confidence and acquiescence was to invest some effort (even prior to the formal signing of collaboration contracts) to demonstrate simple, but high-impact, capabilities of our research group. For example, both *LiveLabs@Mall* and *LiveLabs@Airport* became convinced of our capabilities when we demonstrated real-working prototypes of indoor location tracking, not just in the SMU campus, but in their own environments and using their own Wi-Fi infrastructures. This demonstration proved far more effective in securing executive blessing for our collaboration that multiple conceptual pitches on advanced concepts, such as real-time queue detection or dynamic group detection.

Data sharing: Beyond our own campus, the other three testbed are all semi-public spaces–while lay members of the public can usually enter significant parts of each venue, the venues themselves are under the control of the venue owners. In practice, this dichotomy led to some interesting discussions about the ownership rights on data generated by the *LiveLabs* participants. While we initially felt that we (i.e., SMU) would be the owners of the data (as we were the ones signing up the participants), the venue owners felt equally strongly about their data ownership rights (as the data about participant interactions was occurring on their premises). In the end, we came up with a solution involving *joint ownership* of the data, with either side free to use the data for their own purposes (of course, within the limits consented to by each participant).

Control on Experimental Interventions: *LiveLabs's* vision of empowering a wide swathe of companies, academics and industry researchers with the ability to test experimental services and interventions on participants raised two additional concerns for our venue partners. The first was reputational risk: even though *LiveLabs* is fundamentally an SMU project, the venue owner feared loss of reputation (as a primary project partner) if some of the services or experiments were sub-standard or not viewed favorably by the participants. Secondly, there could be situations where our experiments, promotions, etc. create business conflicts–for example, an anchor tenant in the mall could get upset if *LiveLabs@Mall* participants were routinely being exposed to promotions from a competing retailer anchored in another mall. To address these concerns, we have had to set up a small operations committee for each venue partner, such that this operations committee will be responsible for vetting and approving individual experimenters and specific interventions in that venue.

IV. Key Research Challenges

To provide the functionality described earlier, *LiveLabs* requires cutting edge research in a number of domains. In this section, we briefly summarize some of the key research initiatives.

Energy-efficient Continuous Context Sensing: As explained earlier (Section II), a key benefit of using *LiveLabs* is the ability to conduct mobile experiments that use very specific contextual triggers (such as "A couple that just finished eating") to launch the experiment. To enable these triggers, we built the *LiveLabs* context collection application (Section II.A) that is installed, with explicit permission, on each participant's phone. This application will run in the background and record the participants' actions, activities, and context.

However, one key challenge we face is balancing between energy usage, accuracy, and latency when running the Context Collector. For example, if we want to know the user's activity accurately, and with a latency less than 1sec, we will need to sample the accelerometer at the highest sampling rate *and* transfer either the raw sensor data or processed results to the *LiveLabs* server immediately. Both of these actions can consume a large amount of battery energy. On the other hand, if we want to conserve the battery as much as possible, we are limited in what we can collect (almost nothing).

The heterogeneity of real-world devices used by our participants poses an additional challenge. In addition to having to support iOS, Android, and Windows Phone devices, we also have to handle multiple OS versions within each device family (iOS 6 and iOS 7 for example). This results in data skews where the amount of contextual data collected can change quite significantly depending on the phone used by the participant.

To address these concerns, we have designed, built, and deployed our first version of a scalable context collection infrastructure. This has two main parts; a) the front-end context collection application that runs on each participant's phone, and b) a back-end server infrastructure that dynamically adjusts the collection criteria on each participant's phone. We have different context collection applications for each OS type. However, all of the applications support the following features: a) ability to retrieve latest policies from the backend server, and b) ability to use the policies to change the sensors and context modules that are active — both in terms of which sensors / context modules and what frequency / polling rate to use (e.g., "sample the accelerometer at 100Hz every 10s for 1s each", "do a Wi-Fi scan for 10s every time you see this set of cell tower IDs", "scan the call logs for new items every 3rd time the user receives a call"). These dynamic policies allow us to change (for example, to satisfy a specific experiment) the energy, accuracy, and latency tradeoffs of any individual participant's context collection module in real-time.

Currently, we are testing the effectiveness of our context collectors in supporting a rich set of experiments without significantly impacting the user's phone experience (battery life etc.). Our future plans include building more sophisticated duty cycling and policy mechanisms as well as integrating some of the collaborative / cloud sensing techniques (e.g., CoMon [10]) that have recently been reported.

Practical Indoor Location: Participant location is clearly a key piece of contextual information needed for many mobile experiments. Given that three our of the four *LiveLabs* venues are primarily indoors, this requires an accurate indoor location tracking system. This is a very active field that has spanned numerous research solutions that use various techniques such as fingerprints (RADAR [2], HORUS [17], etc.) and models (Zee [13] etc.), as well as many commercial systems (Google Indoor Location [8], Cisco Location Appliance [5] etc.). However, when testing various existing location tracking solutions at *LiveLabs@Airport* and *LiveLabs@Mall* , we discovered, to our surprise, that we were unable to support *reliable*, *all encompassing*, *accurate* and *energy-efficient* indoor location tracking.

In particular, the limitations we found were that existing solutions tend to work only on certain types of devices; in particular, solutions that supported any participant-carried smartphone device (iOS, Android etc.) were rare. In addition, the accuracy of some of the solutions was poor as they were designed for environments that were fairly stable in terms of environmental (e.g. layout of building) and load changes (i.e., density of people). In both *LiveLabs@Mall* and *LiveLabs@Airport*, it is not uncommon for the people density to change by an order of magnitude within a few hours (and not be consistent across days and hours) and for the layout of the environment to change on the order of days (due to changes in floor displays, temporary exhibits, store movements, etc.). Both of these effects greatly impact the accuracy of current solutions as they depend on some level of stability (for their fingerprints and models). In addition, the randomness of the changes in these environments prevents the use of multiple fingerprints or models as a possible solution. Finally, many of the solutions provide excellent average or best case performance, but at the expense of a much worse "tail" performance. Venue operatos, at the mall and airport, were quick to point out their preference for a solution with worse average case performance, as compared with another solution, if the deviation between average and worst-case performance was small. In particular, we learned that users who see big differences in their location accuracy are more likely to complain than users who see worse but consistent performance.

Currently, we have built a location tracking prototype that works for any participant device, using server-side tracking mechanisms, and deployed it operationally at *LiveLabs@SMU*. We plan to extend our deployment to *LiveLabs@Airport* and *LiveLabs@Mall* in the next few months.

Real-Time User Analytics: To enable useful forms of context-triggered experimentation, a key goal in *LiveLabs* is to derive meaningful high-level context (about individual, groups and the public spaces they inhabit) from the underlying data provided by the Context Collector and location tracking systems. This requires advances in both real-time analytics algorithms, and in the use of scalable stream-processing platforms that can apply such algorithms over data streams from tens of thousands of participants.

For example, how can we figure out what the current spending preferences of a participant are? This type of real-time contextual state information is crucial for dynamic experiments. For example, it would not be useful to send a "buy coffee" discount to a participant who is not interested in coffee (as possibly evidenced by her having walked obliviously past three prior cafés). As another example, knowing who a participant is currently with and their relationship to that person (e.g. spouse, colleague, friend, etc.) can greatly impact the type of experiment that would be applicable (e.g., a "buy 1 get 1 free" is better suited to participants in pairs versus single participants, and a "holiday for two" promotion would be better for couples versus friends).

Some of our current research efforts in the area of such mobile analytics are presently focused on detecting if participants are moving in groups, and if so, identifying the relationship among the group members. In addition, we are building solutions that can detect a) if participants are in a queue, and b) how long they are likely to remain in that queue. Both of these solutions are useful for dynamic experiments and will be tested at the various *LiveLabs* venues soon.

Efficient, Easy, & Effective Experimentation Support: As stated in the Introduction and Section II, the ability to perform real-world in-situ interventions is a key differentiator for *LiveLabs*. However, providing this experimentation capability requires a number of key research innovations. These innovations straddle three different areas:

1. **User Interface**: *LiveLabs* is designed to be used by non-technical experimenters (e.g., sales managers, marketing executives etc.). Hence, there must be an intuitive, yet sufficiently powerful way for these non-technical experimenters to easily specify the kinds of experiments that they are interested in. This requires developing novel user interfaces that can capture the rich experiment space (e.g. "send a $5 off coffee coupon to couples who have just finished shopping for home products."). We have currently built an initial user interface prototype for specifying experiments as well as another prototype for specifying promotions. We plan to engage with experimenters and improve both prototypes based on their feedback.

2. **Picking a Statistically Valid Experiment Group**: The experiments run by *LiveLabs* will be in public spaces with "uncontrolled" participants. This is both a strong benefit as well as a problem for the experimental component of *LiveLabs*. The benefit is that the results will be generated from real participants engaging in real activities in real environments. The disadvantage is that it requires a lot more work to eliminate the experimental bias inherent in these sorts of real uncontrolled environments. In particular, some of the questions we are addressing include "How many participants do you need to accurately test any particular question?", "How do we control the bias in an experiment when setting up a control group is extremely difficult due to the uncontrolled nature of the environment?", and "How do you represent the bias and errors inherent in this system to the experimenter so that they can understand how *significant* their results are?". All of these questions have possible solutions in the experimental psychology and marketing research spaces. We are currently investigating and testing some of these solutions.

3. **Analytics-Preserving Privacy Support**: One of the most compelling experimental features of *LiveLabs* is the ability for experimenters to gain insights about the context and actions of every participant involved in an experiment, irrespective of whether the response was positive or negative. A key challenge here is to develop mechanisms to share such meaningful insights with the experimenter, while preserving the privacy preferences of each individual participant. We believe that this tension, *between the accuracy requirements of analytics outcomes and the privacy preferences of participants*, is a fundamental research challenge that will become progressively more critical across many applications of mobile and urban computing. To address this, we have just initiated work on developing a framework that will support reasoning about these tradeoffs in a more structured way.

Usage-Adaptive Wireless Networks: Our early survey and discussions indicated a widespread desire by our partners to conduct experiments that involved high-bandwidth multimedia content — i.e., experiments that sent participants videos, audio, and / or highly interactive HTML-5 layouts that had many multimedia components. However, supporting such high-bandwidth multimedia content in *LiveLabs* is challenging, as all our test venues have very high occupancy densities (at least 40,000 people visit *LiveLabs@Mall*, *LiveLabs@Airport*, and *LiveLabs@Sentosa* daily). Hence, the wireless infrastructure at these venues is not able to support many concurrent high-bandwidth data flows.

To address this challenge, our research is looking at methods to improve the *context-aware* and *coordinated* usage of the available wireless spectrum across multiple co-existing broadband wireless access technologies. Possible approaches include using the 40 GHz and higher wireless bands, dynamically moving participants between the 2.4 GHz and 5.5 GHz Wi-Fi bands, LTE cellular channels and Bluetooth, as well as looking at other kinds of wireless technologies (White Spaces, FM, Wi-Fi Direct, etc.) that can help provide more flexible *reuse* of existing spectral allocations. In addition, we are also researching methods (such as Cyber Foraging [3] and other data offloading techniques) to migrate participants as quickly as possible off the congested wireless spectrums (Wi-Fi or cellular) and onto a wired high-bandwidth backbone. As an initial effort in this area, we have deployed a small 3G femtocellular testbed at *LiveLabs@SMU* and are investigating methods to selectively offload participants from the cellular network to a wired access backbone.

V. Current Testbed Status

We now describe the current status of our *LiveLabs@SMU* deployment on the SMU campus. *LiveLabs@SMU* became operational on August 31, 2013, when we began to actively recruit student participants on the SMU campus. The initial deployment of *LiveLabs@SMU* consists of Context Collector services available for both Android3+ and iOS6+ devices, a server-side location tracking service that tracks on-

Context Type	**Android**	**iOS**
Location	GPS & Wi-Fi scans	Location API
Phone States	Attached celltowers, battery levels, network statistics, website visited, calendar, settings, contacts, media state	Battery level calendar, contacts, settings, media states
Phone Events	Apps installed, Apps in use, Screen interactions, call & texting events	App state
Activity	accelerometer	accelerometer (only in foreground)
Raw Sensors	accel, compass, gyro, barometer, light	accel, compass, gyro, light (only in foreground)

Table 1: Context Collected At Present

campus locations of all devices that connect to the SMU Wi-Fi infrastructure, and an initial mobile student application, called *SMUddy* (also available for the Android3+ and iOS6+ platforms). We now provide a few more technical details of this testbed and some statistics on participant signup and activity.

V.A. Platforms supported and context collected

Our Context Collector platforms run as a background service in the user space on both Android and iOS platforms. For Android, the Collector platform can be configured to collect sensor data, phone events, client-side indoor location coordinates, etc.— Table 1 details the various key events that are collected. For iOS, a background application gets a foreground processing timeslice once every 10 minutes; at that point, our collector can collect sensor data, limited location information and a much smaller set of phone events (see Table 1). The collection policies of the Context Collector (what subset of sensors and events to collect, and frequently) are configurable by the *LiveLabs* server infrastructure, through a set of XML-based policies. Such policy-based collection allows us the flexibility of continuously adapting the granularity and fidelity of the sensor collection (with corresponding impact on the power drain) for each individual sensor. The current version of the Context Collector does not upload the collected data continuously; while the upload interval is configurable, it is currently set to 3 hours by default (implying a total of 8 collection episodes a day). Moreover, uploading happens only if the participant is on the SMU campus and connected to the SMU Wi-Fi network, so as to alleviate concerns about cellular data usage costs.

The collected data is first stored in textual format (multiple .csv files) on the smartphone's storage, from where it is zipped and then encrypted before being transferred over a secure HTTP interface to the backend repository. At the server end, the data files are unzipped and extracted, and then the fields of each file are anonymized by a special anonymization service before being stored in appropriate DB tables. The anonymization service hashes various device and usage specific identifies (e.g., IMEI numbers, email addresses, etc.) to ensure that even the day-to-day research staff of *LiveLabs* do not get access to sensitive personal context data. As one consequence of this "delayed sending" policy, we note that our measurements suggest that the storage of intermediate context data on the flash storage of the mobile device consumes significant energy, in many cases exceeding even the sensing energy costs. As our context collection becomes increasingly more real-time, we will need to revisit this energy overhead issue in greater detail.

V.B. Participant Signup & Related Statistics:

As of October 23, *LiveLabs@SMU* has 670 total registered participants, with approx. 120-150 users uploading data and using the *SMUddy* App (described shortly). Our students were recruited from all full-time undergrads at SMU, with a bias towards freshmen (as we wanted participants who are likely to offer us the greatest quantity of longitudinal data). Overall, the participants are currently evenly divided between Android (42%) and ioS (48%). As part of the sign-up process, each participant is required to specify additional demographic detail (such as gender and primary school of study), as well as a variety of demographic attributes. Our participants are fairly distributed across all 6 schools at SMU, with understandably larger numbers from SIS (as we had the greatest advertising outreach in our school) and the School of Business (which has the largest pool of undergraduate) students. Figure 4 shows the number (and year-wise breakdown) of the student participants recruited thus far in *LiveLabs*.

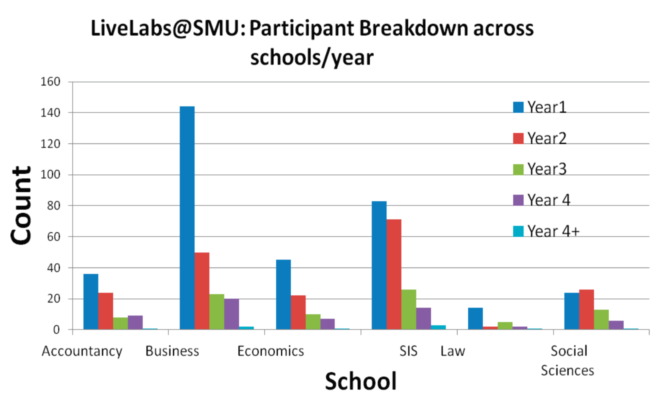


Figure 4: LiveLabs participant statistics

V.C. The SMUddy App

As we shall see (later, in Section VI), we found out that the most effective way of securing voluntary participation by our students was by making available mobile Apps that they would really use. Accordingly, as part of the current *LiveLabs@SMU* deployment, we have built and re-

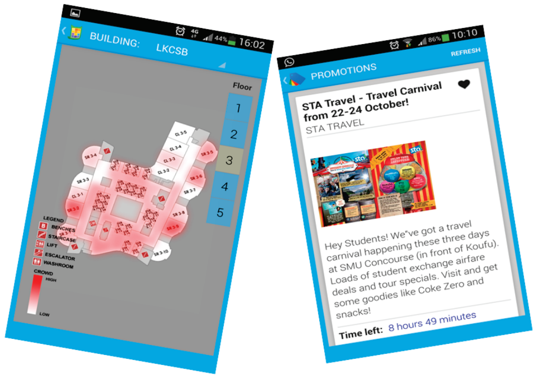


Figure 5: Screenshots of SMUddy Application

leased the first version of an App, called SMU-Buddy (or *SMUddy*). *SMUddy* utilizes the large-scale, near-real time location tracking of all on-campus & Wi-Fi connected devices to offer three key features (see Figure 5 for sample screenshots from the *SMUddy* App):

- *Real-time Occupancy Heatmaps:* We compute occupancy heatmaps, based on the number of mobile devices located in different parts of the SMU buildings and the nominal capacity of the area. Currently, given the 2-3 min frequency of location computation, we specify occupancy levels at the granularity of sections of a building floor, instead of room-level. This feature has proven to be very popular with our students, who utilize this feature extensively to find relatively quiet or unoccupied parts of the campus to hold impromptu project group meetings or consultations.

- *Location-Aware Messaging:* This features allows participants to exchange short messages with their buddies, with the added flexibility of allowing location & time-based predicates for message delivery. The two most common uses of this are to broadcast a message to all buddies by location (e..g., sending a "going for lunch" IM to all friends who are currently located on the same floor) or to mimic digital post-its (e.g., delivering a "remember to bring the exam paper along" IM to a colleague only when she steps out of her office.)

- *Promotions:* This feature allows participants to avail of special discounts and promotions offered by vendors and retailers on and around the SMU campus. We can presently support simple forms of proximity-based (e.g., "show the Subway promotion only when the participant is within 5 meters of the Subway store") or demographics-based notifications. However, thus far, all our retailers have preferred to blast their promotions to all participants (possibly because our campus is small enough).

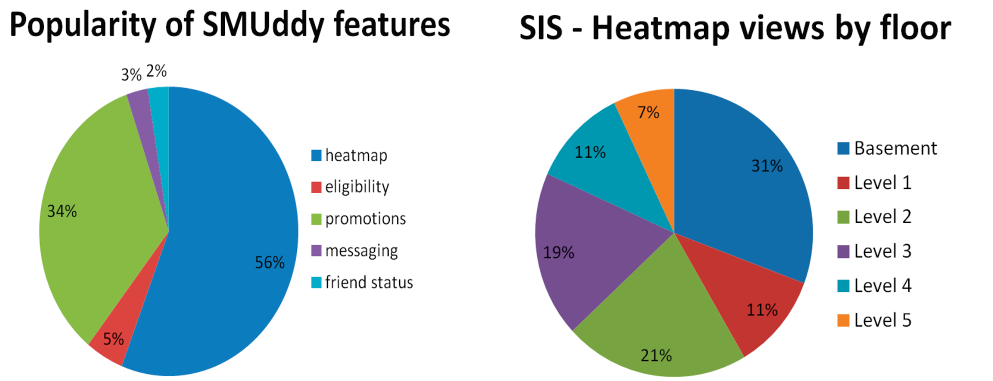
Figure 6 shows two different statistics about the usage of *SMUddy* by our current participant pool. Figure 6.a shows the relative popularity of our two features; we can see that heatmaps are by far the most heavily used feature, reflecting the paucity of meeting spaces on our crowded campus. Figure 6.b further dissects heatmap usage, showing the floor-wise breakdown of the heatmap viewing requests made by participants for the SIS building. Not surprisingly, the largest fraction of queries pertain to the underground (basement-level) concourse area, where students congregate for both academic discussions and extra-curricular activities.

VI. Key Insights and Lessons Learned

Through the process of planning, implementing, launching and maintaining *LiveLabs*, we have learned many practical lessons that we would like to share with our readers. These include:

Diversity of Real-world Deployments: *LiveLabs* has to address challenges of diversity in *i) Device types*: As participants use their own primary phone, we have to support not only multiple OS platforms (chiefly, Android, iOS and WP8), but also multiple devices and different OS versions and *ii) Venue Characteristics:* While they are all indoor spaces, *LiveLabs@SMU*, *LiveLabs@Mall* and *LiveLabs@Airport* all differ in their physical characteristics, such as layout, density & diversity of the Wi-Fi infrastructure, and the absolute quantity, density, and temporal variation of the number of people at the location. These have deep implications for both the day-to-day operational characteristics of the testbed and the types (and accuracy) of context that we can get at each venue. In particular, these differences imply that it is hard, if not impossible, to offer the same level of context accuracy across venues (or even at different locations within the same venue). Accordingly, *LiveLabs's* experimentation component has to explicitly incorporate such context variation while specifying experiment constraints— for example, permit queuing-based interventions in floors 3 & 5 of a mall, but not on floors 2 & 4.

As mentioned before, our applications operate in user space and require no kernel modifications on participant smartphones. Android offers more flexibility in sensing various forms of context, whereas successive versions of iOS provide decreasing levels of access to useful individual context. While building the Collector software, we have had to work very carefully to maintain backwards compatibility, to ensure that we continue supporting existing users who may not always update to the most recent OS version. In several instances, a new version of an OS also required us to make fundamental changes in our data collection and anonymization software. For example, in iOS 6 and prior versions, the OS allowed programmatic retrieval of the device's *IMEI* number. Accordingly, a hash of this IMEI was used to represent a randomized user in the *LiveLabs* infrastructure. However, the recently released iOS 7 no longer supports the retrieval of an IMEI; as a consequence, we have had to devise workarounds and now use the smartphone's Wi-Fi MAC address (obtained through a user-assisted process and then appropriately hashed) as the per-user unique identifier. However, this also necessitated changes in the Context Collector and Analytics Engine, as special bridging code is now needed to map previous IMEI-



a. **Relative popularity of *SMUddy* Features** b. **Heatmap views by floor of SIS building**

Figure 6: *SMUddy* usage statistics

based user identifiers with newer MAC-based user IDs.

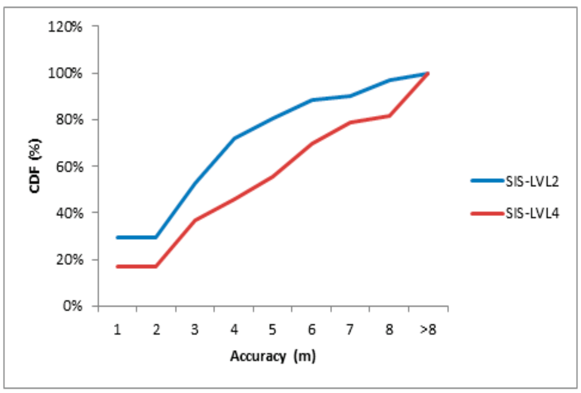


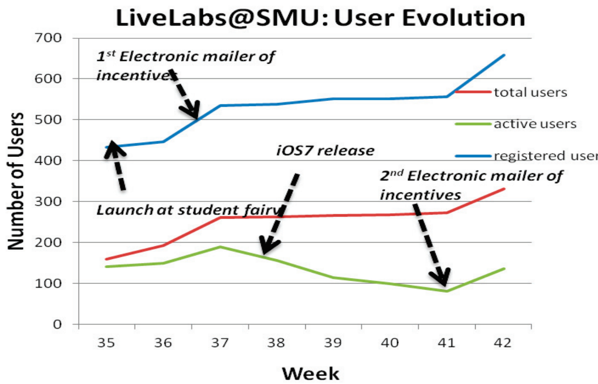
Figure 7: Wi-Fi Location Accuracy in *LiveLabs@SMU*

Differences in the characteristics of the indoor spaces have also tangibly translated into different location accuracy metrics. These differences are the result of differences in the characteristics of the Wi-Fi infrastructure, such as the density of APs (*LiveLabs@SMU* has a much denser deployment of APs compared to *LiveLabs@Mall*), the versions of Wi-Fi controller software used (although both venues use Wi-Fi infrastructure supplied by the same vendor, *LiveLabs@SMU* 's software version provides location updates once every 2-3 minutes, as compared to an update interval of as low as 1 second at *LiveLabs@Mall*) and the site-specific operational settings (e.g., *LiveLabs@SMU* APs advertise as many as 7 virtual SSIDs, while such AP virtualization is absent at *LiveLabs@Mall*). All these differences greatly affect the location accuracy possible at each venue. Interestingly, these accuracy variations occur not just across different venues, but also across different floors of the same venue. As an illustration, Figure 7 shows a CDF plot of the location accuracy achieved with our initial implementation of RADAR, in the SIS building at SMU, but across two floors (level 2 and level 4). Even with similar Wi-Fi AP deployments, the accuracy is much higher at level 2, as opposed to level 4.

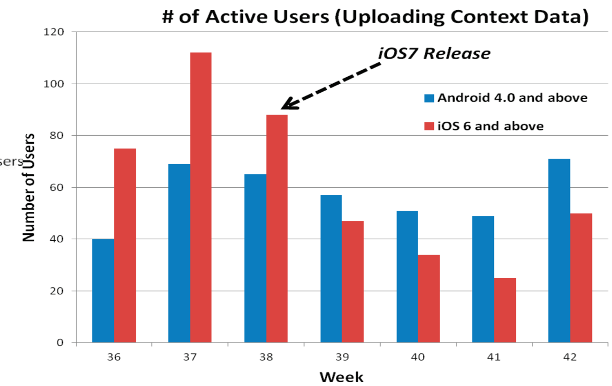
The Vital Role of Energy vs. Accuracy Tradeoffs: Based on empirical feedback, we learned that participants were most sensitive to the perceived energy drain of their phone's battery. To allay this concern, we adopted a policy of restricting the Context Collector overhead to be no more than 10% of the phone's battery over a single day. However, our experience with the *LiveLabs* Context Collector reveals that, with present-day technologies, it is simply not possible to support continuous, real-time context monitoring without significantly using more than 10% of the battery. Staying under the maximum battery usage limit will thus require trading off *how long* a particular context will be monitored, *at what accuracy* and *the frequency* of uploads to the *LiveLabs* server. As a secondary consequence of the need to manage the energy overhead carefully, we have been very conservative in our introduction of additional forms of sensing and sensor data processing on the smartphone.

Moreover, our default model is now to use *server-side* indoor location tracking, where the Wi-Fi infrastructure is responsible for localizing a mobile device, based on measurements made by the family of APs. While such a location tracking method is not as accurate as the rapid-scanning based client-side solutions, it has a tremendous advantage— it imposes no additional overhead on a transmitting mobile device. To support experiments that require much finer-grained and more frequent location determination (e.g., delivering a coupon as soon as a participant enters a specific store), we now use such coarse-grained, server-side localization as a *trigger* for short bursts of energy-intensive, more accurate Wi-Fi based location tracking. Moreover, our *LiveLabs* server policies explicitly restrict experiments from requiring fine-grained location tracking for extended periods of time.

Judicious Combination of Mobile & Infrastructure Sensing: The *LiveLabs* vision was founded on the vision of accelerating the use of mobile sensing to collect individual and crowd-level context in indoor public spaces. While this vision is undoubtedly appealing, our experience suggests that there are many practical cases where mobile sensing is simply not effective (in terms of either cost or complexity) as simple infrastructure-based sensing alternatives. As a practical example, our SMUddy application uses our live location tracking of mobile devices on the SMU campus to indicate the occupancy of rooms and benches. Independent of the accuracy of any specific location tracking technology, this approach often gives rise to false negatives (indicating that a room or space is empty when it is actually occupied), simply because the occupants do not have *any*



a. **Total participant evolution**



b. **Active participants by OS type**

Figure 8: Variation in Registered & Active Participants at *LiveLabs@SMU*

mobile device being actively used at that point. However, simple infrastructural solutions (such as infra-red motion sensors or low-end cameras) can detect such context much more accurately and at fairly low costs.

Based on our experiences at both the mall and airport venues, we have realized that public spaces are increasingly being equipped with a wide variety of embedded sensors (e.g., motion detectors, security cameras, RFID readers, temperature sensors etc.). Researchers in our community often tend to ignore the proliferation of such devices, and focus purely on maximizing the observational potential of mobile devices. Looking towards the future, we however believe that judiciously *fusing* the ambient context of embedded infrastructural sensors with the personalized context of mobile devices will prove to be the most effective and economical solution for large-scale real-time context sensing. As a simple illustration of this broader concept, we are now actively exploring the use of simple embedded sensing solutions (e.g., smart pressure strips or people-counting cameras) to provide real-time occupancy statistics on our SMU campus.

Signing Up Participants Is Easy, Retaining Them Is Much Harder! Much of the initial effort in getting *LiveLabs@SMU* launched focused on reaching out to participants and getting them signed up. To achieve this, we spent a significant amount of effort (definitely over 3 person months) in increasing the visibility of the testbed, by emails to the whole SMU community, by actively seeking out speaking engagements at various freshman orientation camps and events and finally by setting up booths during major SMU student events. Such publicity has proved to be quite effective–after each round of publicity, we have seen significant increases in our set of *registered* (people who have visited the *LiveLabs* sign-up Website and agreed to our IRB terms and conditions), *downloaded* (people who have downloaded our monitoring service on their smartphone) and *active* (those who regularly upload data to our server infrastructure) users.

Figure 8.a plots the time evolution of the users, and indicates our key publicity (and external) events. However, we notice that in several instances, the number of active users begins to drop steadily after the initial uptick. Some of the most significant causes include mobile OS upgrades that cause our existing software to become non-operational or simple human errors (e.g., forgetting the password) that, at least temporarily, disrupt the continuity of context collection. In particular, from Figure 8.b, we can observe a sharp drop in the number of iOS active users during week 38–this was due to the recent launch of iOS 7, which necessitated modifications in our *LiveLabs* data collection software.

We observed that the most active participants in *LiveLabs* were those users who actively used *SMUddy*, our first location-based *LiveLabs* App. In particular, we observed a very steady usage of SMUddy's real-time heatmaps feature across all weeks. Accordingly, we realize that the participant base will hold steady only if we are able *to offer popular and sticky applications*. As a consequence, we have now significantly modified our recruitment tactic by focusing on building multiple Apps rapidly, with plans to roll out these Apps on a significantly shorter timeline (e.g., a new App every month or so). We urge readers interesting in operating similar testbeds to take note of this lesson, and to plan a fairly aggressive App rollout in advance, ahead of the actual signup process.

Research and Production — Chalk and Cheese: While *LiveLabs* has an ambitious research agenda, it is also a major engineering effort (at least for a university research group). Initially, as faculty leads, we ran the project team like a regular research group, using Ph.D. students and research-minded software engineers to develop and maintain the *LiveLabs* codebase. Over the past year of project execution, we have learned that this approach is simply not sustainable or practical. There are multiple problems with our original approach: i) most Ph.D. students produce code that is undocumented, or very difficult to maintain; ii) periodic research conference deadlines meant that operational code frequently took a backseat and was often hastily done, and iii) researchers often focused on moving on to the "next cool" idea and did not have the patience and attention to detail to respect the rigors of a proper software engineering process. This occurred in spite of our attempts, as the faculty leads, to adhere to some basic software engineering practices (such as maintenance of a release schedule, maintenance of bug reports and occasional code reviews).

To overcome these issues, we have migrated to a clearly-articulated triple track strategy. The first track is the production track that consists of full-time professional programmers who are responsible for developing, maintaining, and documenting all of the production code. We also encourage the use of standard software frameworks (e.g., CodeIgniter) for most of the "conventional codebase", and explicitly appointed a project manager to devise and maintain schedules. We have clearly separated our testing and "live" production environment, and instituted explicit email & phone-based helpline services to assist potential *LiveLabs* participants with any operational problems. The second track consists of the research team (post-docs, students, etc.) who focus on developing the technologies that will be needed 3-12 months down the road. Once a technology is mature enough, the research team will hand it over to the production track to harden, deploy, and maintain. Finally, the third track consists of business development, accounting, and marketing and is run by full-time non-technical professional marketing, accounting, and media personnel. This track is responsible for meeting with client companies who want to use *LiveLabs* (this alone can consume multiple hours every day!), managing the relationships with existing *LiveLabs* partners, managing all aspects of funding, and devising, producing,and administering various media and publicity tasks related to *LiveLabs*.

While some of these observations may seem banal and obvious, we were truly surprised by the overheads (in terms of time and effort) that we had to incur to deal with scores of minor issues. In particular, we found ourselves frequently pulled in different directions to handle the production, research, and partner management requirements of *LiveLabs* — and occasionally ended up not doing a great job in any of them due to massive amounts of context switching. Humbled by our own experiences, we urge our colleagues to aggressively budget for such full-time engineering resources when formulating the overall testbed budget, especially if it requires interactions with real-world participants and external venues.

VII. Related Work

An increasing appreciation of the limitations of pure simulation-based performance studies has led to a strong emphasis on testbed building and maintenance over the last decade. Many of these testbeds focused on enabling *technology* testing and verification. For example, PlanetLab [12] was a distributed testbed that proved to be indispensable for testing (largely wired) Internet technologies, such as overlay routing and content distribution. The ORBIT testbed [14] provided several hundred programmable wireless nodes, that allowed verification of both wireless technologies and emulation-based studies of mobility effects. Kansei [1] and WiseBed [6] focused on at-scale testing of wireless sensor networks protocols, offering an experimentation interface over several hundred sensor nodes. While these testbeds help evaluate the impact of realistic wireless environments, they are primarily housed in university campuses and are not designed to experimentally test *real-world* human usage of such technologies.

There have also been several R&D efforts in building testbeds that operate on neighborhood or city-scale, thus enabling more realistic studies of technologies based on human-generated data workloads. For example, TFA [4] deployed a city-scale mesh networking testbed in Houston, whereas the WOMEN [16] project deployed and tested a mobile mesh network testbed in both indoor and outdoor environments in Trento. Unlike *LiveLabs*, these projects focus more on improving wireless connectivity and bandwidth, and not on testing human interaction with context-aware mobile services.

The field of mobile computing and mobile sensing has benefited from several recent efforts to collect large-scale traces of real-world smartphone usage data. Examples of such data collection efforts include Nokia's Mobile Data Challenge (MDC) [9], a data collection campaign involving close to 200 participants over an entire year and Microsoft's GeoLife dataset [18], that contains GPS trajectories of over 150 users observed over a period of 2 years. These datasets principally focus on gathering city-scale mobility data; in contrast, *LiveLabs* focuses not just on gathering fine-grained context data inside indoor public spaces, but also provides the ability to perform *in-situ* interventions in real time.

We have also witnessed the emergence of several smartphone-based testbeds recently. The PhoneLab testbed [11] is probably closest in spirit to ours–it seeks the participation of several hundred university students and offers an open, programmable platform for testing mobile technologies and applications. Like *LiveLabs*, the NetSense project [15] uses special phone-embedded software to collect extensive metadata about phone usage events (e.g., calls, emails etc.), with the goal of better understanding social interconnections and interactions among people. We believe *LiveLabs's* differentiator lies not just in our focus on a wider variety of public urban spaces (beyond just university campuses), but also in our ability to make deep context data available in near-real time to an integrated platform for *in-situ* behavioral studies.

VIII. Conclusion

This paper described our journey thus far with *LiveLabs*, starting from the vision of a large-scale *real world* testbed for helping test advanced, context-aware mobile services and applications to the deployment and operation of a testbed on the SMU campus. Our goal is to explore and understand the extent to which context-aware mobile computing can become a reality, by experimentally studying both the limits on gathering such context and how regular consumers, engaged in daily lifestyle-driven activities, interact with such context-aware services, offerings and applications. To make this vision possible, we have spent significant effort in expanding the scope of *LiveLabs* beyond just our SMU campus to include three other major publicly accessible locations in Singapore: a shopping mall, a premier airport and a major resort facility. The diversity of these locations enables realistic experimentation in several business domains, such as retail, leisure & tourism and operational logistics.

At present, the *LiveLabs@SMU* testbed is operational, with over 500 participants signed-up and being offered a variety of services (leveraging upon our capability to track indoor location across all SMU academic buildings). We believe that our experiences to date have two important takeaways for our mobile systems fraternity. First, building and operating a 24X7 testbed in multiple public spaces is a major engineering and logistics effort, that needs careful planning, staffing and scheduling in ways that most university research faculty are not naturally inclined to perform or trained to manage. Second, we have discovered that many of the solutions for apparently well-researched and "solved" technical problems, such as indoor location tracking and mobile sensing, do not work as well as anticipated in densely occupied public spaces. More specifically, in our view, building solutions that are *robust* to wide variations in environments and consumer usage patterns remains an open challenge.

We emphasize that *LiveLabs* is intended to be an open testbed, for use by the international and Singapore-based research community. Accordingly, as the testbed evolves, we solicit participation for testing applications and services, as well as invite collaboration for building some of the foundational technologies.

IX. Acknowledgment

This research is supported by the Singapore National Research Foundation under its IDM Futures Funding Initiative and administered by the Interactive Digital Media Programme Office, Media Development Authority.

References

[1] Arora, A., Ertin, E., Ramnath, R., Nesterenko, M., and Leal, W. Kansei: A high-fidelity sensing testbed. *IEEE Internet Computing*, volume 10, 2006.

[2] Bahl, P. and Venkat N, P. RADAR: An in-building RF-based user location and tracking system. *Infocom*, Tel-Aviv, Israel, Mar. 2000.

[3] Balan, R., Flinn, J., Satyanarayanan, M., Sinnamohideen, S., and i Yang, H. The case for cyber foraging. *ACM SIGOPS European Workshop*, Saint-Emilion, France, July 2002.

[4] Bejarano, O., Miskovic, S., Aryafar, E., and Knightly, E. TFA: A large scale urban mesh network for social and network research. *ACM Workshop on Wireless of the Students, by the students, for the Students*, New York City, NY, Sept. 2010.

[5] Cisco. *Cisco Wireless Location Appliance*. `http://www.cisco.com/en/US/products/ps6386/`.

[6] Coulson, G., Porter, B., Chatzigiannakis, I., Koninis, C., Fischer, S., Pfisterer, D., Bimschas, D., Braun, T., Hurni, P., Anwander, M., Wagenknecht, G., Fekete, S. P., Kröller, A., and Baumgartner, T. Flexible experimentation in wireless sensor networks. *Commun. ACM*, 55:82–90, Jan. 2012.

[7] Duerig, J., Ricci, R., Stoller, L., Strum, M., Wong, G., Carpenter, C., Fei, Z., Griffioen, J., Nasir, H., Reed, J., and Wu, X. Getting started with geni: a user tutorial. *SIGCOMM Comput. Commun. Rev.*, 42(1), 2012.

[8] Google Inc. *Google Maps Indoor*. `http://maps.google.com/help/maps/indoormaps/`.

[9] Laurila, J. K., Gatica-Perez, D., Aad, I., Blom, J., Bornet, O., Do, T., Dousse, O., Eberle, J., and Miettinen, M. The mobile data challenge: Big data for mobile computing research. Newcastle, UK, 2012.

[10] Lee, Y., Ju, Y., Min, C., Kang, S., Hwang, I., and Song, J. Comon: cooperative ambience monitoring platform with continuity and benefit awareness. *MobiSys*, Lake District, UK, June 2012.

[11] Nandugudi, A., Maiti, A., Bulut, F., Batra, S., Ki, Y., Challen, G., Demirbas, M., Ko, S., Kosar, T., and Qiao, C. Participant behavior in phonelab. *Third International Conference on the Analysis of Mobile Phone Datasets (NetMob)*, 2013.

[12] Peterson, L., Bavier, A., Fiuczynski, M. E., and Muir, S. Experiences building planetlab. *Proceedings of the 7th symposium on Operating systems design and implementation*, OSDI '06, 2006.

[13] Rai, A., Chintalapudi, K. K., Venkat N, P., and Sen, R. Zee : Zero-effort crowdsourcing for indoor localization. *Mobicom*, Istanbul, Turkey, Aug. 2012.

[14] Raychaudhuri, D., Seskar, I., Ott, M., Ganu, S., Ramachandran, K., Kremo, H., Siracusa, R., Liu, H., and Singh, M. Overview of the ORBIT radio grid testbed for evaluation of next-generation wireless network protocols. *IEEE Wireless Communications & Networking Conference*, New Orleans, LA, Mar. 2005.

[15] Striegel, A., Liu, S., Meng, L., Poellabauer, C., Hachen, D., and Lizardo, O. Lessons learned from the NetSense smartphone study. *ACM HotPlanet workshop*, Hong Kong, China, 2013.

[16] Vincenzi, M., Tomasi, R., Tacconi, D., Kliazovich, D., and Granelli, F. Lessons learned from the NetSense smartphone study. 6^{th} *International Conference on Testbeds and Research Infrastructures for the Development of Networks & Communities (TRIDENTCOM*, Berlin, Germany, 2010.

[17] Youssef, M. and Agrawal, A. The horus WLAN location determination system. *MobiSys*, Seattle, WA, June 2005.

[18] Zheng, Y., Zhang, L., Xie, X., and Ma, W.-Y. Mining interesting locations and travel sequences from gps trajectories. *Proceedings of the 18th international conference on World wide web*, WWW '09, 2009.