2-2014

LiveLabs: Building An In-Situ Real-Time Mobile Experimentation Testbed

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

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

Youngki LEE
Singapore Management University, YOUNGKILEE@smu.edu.sg

Citation

BALAN, Rajesh Krishna; MISRA, Archan; and LEE, Youngki. LiveLabs: Building An In-Situ Real-Time Mobile Experimentation Testbed. (2014). *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications HotMobile '14, Santa Barbara, CA, February 26-27, 2014*.
Available at: https://ink.library.smu.edu.sg/sis_research/2057

LiveLabs: Building An In-Situ Real-Time Mobile Experimentation Testbed

Rajesh Krishna Balan, Archan Misra, and Youngki Lee
School of Information Systems, Singapore Management University
{rajesh,archanm,youngkilee}@smu.edu.sg

ABSTRACT

We present *LiveLabs*, a mobile experimentation testbed that is currently deployed across our university campus with further deployments at a large shopping mall, a commercial airport, and a resort island soon to follow. The key goal of *LiveLabs* is to allow in-situ real-time experimentation of mobile applications and services that require context-specific triggers with real participants on their actual smart phones. We describe how *LiveLabs* works, and then explain the novel R&D required to realise it. We end with a description of the current *LiveLabs* status ($>$ 700 active participants to date) as well as present some key lessons learned.

1. INTRODUCTION

The inability to test novel mobile applications, *context-based* services, or usage patterns with *real* users, using their regular phones in *real-world* environments, remains a key problem for the mobile computing research community. In particular, we have little understanding of how such context inferencing works *at-scale* in the real world, and how consumers react to such context-gathering and the resulting services. Usually, we are limited to controlled user studies, often with a small set of users or restricted to specific campus/office environments, that attempt to capture reality as best as they can, while recognising that such controlled studies do not provide the accuracy and "proof" of real-world deployments.

Real-world testbeds have gained a lot of traction and attention in recent years. Wireless testbeds such as ORBIT [10] and WISEBED [6] allow testing of network protocols and technologies in laboratory settings (as opposed to simulations), whereas TFA [3] provided a city-scale mesh-networking testbed. Data collection campaigns, such as the Nokia Mobile Data Challenge [7] have helped collect rich data sets about the real-world mobile usage patterns; however, such trials do not focus on *in-situ* testing of applications. More recently, we have seen several smartphone-based testbeds, such as NetSense [13] and PhoneLab [14] , which provide open platforms for application testing. While certainly permitting more realistic usage studies, they are presently typically deployed on university campuses with predominantly student participants, and thus do not address issues with creating applications/services in other public spaces (e.g., malls). In some instances, the participation incentives include devices and / or data plans provided by the experimenters, and may thus not accurately capture challenges arising from device diversity and real-world usage artefacts (e.g., how participants would use such services if they had to pay for the data plan).

To overcome such limitations and create a testing substrate that is more diverse (in participant and location types) and reflective of the challenges associated with device diversity, we present our ongoing work in building the LiveLabs Urban Lifestyle Innovation Platform, commonly referred to as *LiveLabs*. *LiveLabs* is a five year research project at the Singapore Management University (SMU), that seeks to turn four real-world public spaces in Singapore into a spatially-distributed testbed, where *context-aware* mobile applications, strategies and interventions can be tested on real people, via their own mobile devices, while they are engaged in regular lifestyle-driven activities. The four environments include the entire SMU university campus (*LiveLabs@SMU*), one of world's busiest airports (*LiveLabs@Changi Airport*), a major multi-story shopping mall that attracts tens of thousands of visitors per day (*LiveLabs@Mall*), , and a large resort island (*LiveLabs@Sentosa*).

We have been working on conceptualising, securing funding, and building and deploying *LiveLabs* for the last three years. As of 20th January 2014, the *LiveLabs@SMU* testbed has become operational with more than 1950 student participants signed up (and over 700 participating actively) across Android and iOS devices. We plan to make *LiveLabs@Changi Airport* operational in the first half of 2014, with *LiveLabs@Mall* and *LiveLabs@Sentosa* going 'live' sometime in the second half of 2014/ first half of 2015.

In the rest of this paper, we will explain the goals of *LiveLabs* and describe the four public spaces it will be deployed at. We then present and motivate three hard research challenges that we need to overcome before LiveLabs can be successful: 1) balancing accuracy and energy overheads in continuous context collection; 2) enabling practical and robust indoor location tracking, and 3) designing and deploying an in-situ mobile experimentation system. Some of these research challenges manifest themselves as major bottlenecks only due to the diversity of *LiveLabs*, in terms of participants, devices and location types. We shall also provide some statistics about the current state of *LiveLabs@SMU* deployment. We hope to involve our research community to *a)* tackle these "hot" research challenges with us collaboratively, and *b)* utilise *LiveLabs* as an international resource for testing novel mobile applications, interventions or services.

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

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. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the granting agency or Singapore Management University.

2. WHAT IS LIVELABS?

LiveLabs was first conceived in late 2010 as a platform for utilising a large pool of opt-in participants to test innovative, context-aware mobile applications, services and interventions under *naturalistic* settings. After an extended period of building the necessary linkages to industry partners and venue operators, *LiveLabs* was funded by Singapore's Media Development Authority and launched in April 2012, as part of a broad initiative to position the city-state as an organic platform for accelerating innovation in the ICT and digital media sectors. In particular, the various *LiveLabs* testbed environments provide retailers, service providers, applications developers and researchers an ideal opportunity to test mobile and digital media technologies for urban Asian settings, characterised by high population densities and an extremely mobile-savvy population. *LiveLabs* is unique in its goals of *a)* providing experimenters with much deeper, finer-grained, near-real time human context (e.g., location, activity, group dynamics) than currently possible, and b) exposing an experimentation service that frees experimenters from many experimental chores (such as subject selection, privacy enforcement etc.).

2.1 Experiment Scenarios Motivating LiveLabs

The following five experiment scenarios illustrate the capabilities that *LiveLabs* is aiming to provide:

Experiment 1: *A café wants to test an innovative new promotion, whereby a* group *of 5 or more customers receive 30% off their coffee bill if they all coffee together.* To allow this promotion to be selectively delivered only to appropriate groups, *LiveLabs* must be able to detect group sizes in real time. Such a vision requires computation of much richer context, than leveraged upon by current services (such as Xtify [2] and PlaceIQ [1]) that principally enable advertisers to test location & time based delivery of messages.

Experiment 2: *A theme park operator wants to test if providing customised group-based park itineraries (family versus friends etc.) will result in higher customer satisfaction.* To enable this, *LiveLabs* must quickly and correctly identify not just group sizes and the *type*, but also track their history of in-park movement.

Experiment 3: *A movie theatre wants to test if offering discounts to people who have been loitering near the theatre for at least 30 minutes, without buying a movie ticket or sitting down for a meal, will improve sales.* This requires tracking both the location and activities of individual users.

Experiment 4: *The airport duty free store wants to test if offering nationality-specific discounts to passengers who are queuing for immigration control will significantly increase in-store purchase rates. LiveLabs* must be able to determine both an individual's nationality, as well as their physical activity state, in real time.

Experiment 5: *A mobile games company wants to determine the demographics and activity combination that will achieve the highest game satisfaction ratings (e.g., men after a meal, students during class, ladies shopping alone, etc.).* This will require matching demographics data with dynamic activity detection mechanisms.

Today, it is very hard to easily test any of the scenarios listed above. In the rest of this section, we show how *LiveLabs* aims to make this possible by providing experimenters with not just venues and participants, but also sophisticated real-time analytics capability and an easy-to-use experimentation platform that summarises the results of running such experiments.

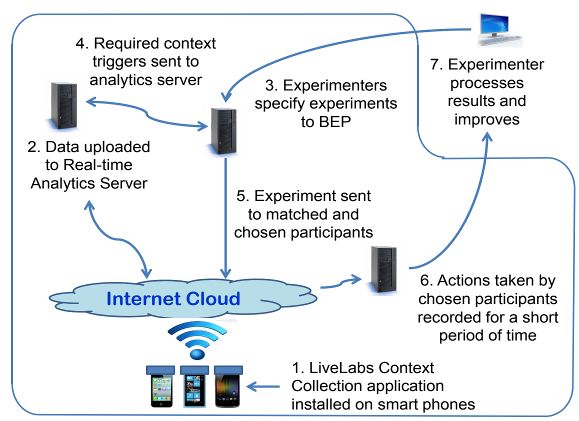
2.2 The Four Real-World LiveLabs Venues

LiveLabs was designed to be large enough to provide experimentation in 5 distinct lifestyle domains (tourism, education, media consumption, shopping, and leisure), with a large enough participant population (we expect at least 30,000 participants in the next 4 years) to provide enough diversity to satisfy real-world experimentation needs of commercial providers. To achieve this, *LiveLabs* has partnered with 4 real-world venues:

1. *LiveLabs@SMU* became operational on August 31st, 2013 and covers the entire SMU downtown campus. The campus comprises 5 academic buildings, each 5 stories high with a basement level connecting all buildings. It will provide up to 7000 student participants [12] and support media consumption and education experiments, as well as allow longitudinal studies related to education and pedagogical outcomes.
2. *LiveLabs@Changi Airport* will be deployed, from early 2014 onwards, at Singapore's main airport. The airport has more than 1,000,000 square meters in space across 3 terminals and serves more than 135,000 people per day [4]. It will allow experiments in the shopping, tourism, and logistics domains.
3. *LiveLabs@Mall* will be deployed, in late 2014, at a large 9 story mall visited by at least 40,000 people every day. *LiveLabs@Mall* will support testing of new pervasive retail apps and services, such as context-aware dynamic promotions and social/group purchasing.
4. *LiveLabs@Sentosa* will be deployed, starting in the second half of 2014, at Singapore's premier resort island. The island occupies more than 5 million square meters and is visited by at least 45,000 visitors per day [11]. This outdoor testbed (the other testbeds are primarily indoors) will allow leisure and tourism-oriented experiments.

Besides the obvious lure of multiple public spaces, *LiveLabs* provides the instrumentation for real-time deep context collection, real-time analytics (group detection, queue detection, etc.) and a unique experiment provisioning layer, which automates many of the tasks in large-scale lifestyle behavioural experimentation (e.g., subject selection, privacy protection of user context).

2.3 How LiveLabs Works



LiveLabs developed components are within the blue outline

Figure 1: How LiveLabs Works

Figure 1 shows the sequence of steps necessary for *LiveLabs* to run experiments. The sequence is:

1. The *LiveLabs* context collector application needs to be installed on participant smartphones. Each participant will have to agree to our IRB clauses and voluntarily opt-in to provide *LiveLabs* with the data needed for operation. We do not just provide an electronic consent form; most of our recruitment to date is via special events, where our staff actively engage each potential participant and individually explain our data collection and usage policies. We currently support iOS 6+, Android 3+, and WP8+ smartphones.

2. The collector application collects sensor and context data from the phone and sends it to our real-time analytics server where it is processed to obtain the required context triggers such as location, current activity, group status etc. To provide a flexible yet power efficient data collection process, the *LiveLabs* collector can be dynamically configured to turn on / off different data collection modules (GPS, accelerometer, etc.) and to change the frequency with which it uploads data to the analytics server.

3. Experimenters wanting to use *LiveLabs* specify their experiments using our Behavioural Experimentation Platform (BEP). Section 3.3 provides more details about the BEP. Note that experiments can specify a wide variety of progressively-richer context predicates that need to be matched by the participants.

4. The BEP validates that the experiment is safe and valid for the participant pool. If it is, it sends the required context triggers to the analytics server. For example, "inform me when you find at least twenty groups of two who just finished watching a movie". The analytics server will keep track of all these context triggers and call back the BEP when the triggers match the current context.

5. When a callback is received with the list of matched participants, the BEP will pick a subset that ensures a valid privacy-preserving experiment and then sends a notification with the experiment details to each selected participant. An experiment could be a discount, a request to run an application, a HTML-5 survey etc.

6. The BEP will monitor the selected participants for a set period of time and record what they did in response to the experiment stimulus. This data is then packaged, in a privacy preserving way, and sent back to the experimenter.

Note: In most market / field trials, the success rate tends to be poor (e.g., a 5% take-up rate is considered good) and there is little visibility into why most users do not react to a stimulus. *LiveLabs's* ability to observe the entire experimental effect (both positive and negative) is a key unique property and selling point.

7. The experimenter processes the results and determines how to change their experiment (if required). They can then either re-run the experiment (with new parameters and constraints), run a new experiment, or declare success.

3. KEY R&D CHALLENGES

Building and operating a large-scale mobile services experimentation testbed, at multiple public spaces and involving members of the general public, poses several difficult challenges, both in classical mobile computing and in mobile-centric experimentation. To be successful, *LiveLabs* is working on innovations in at least three major areas, described next.

3.1 Real-Time Deep Contextual Analytics

A key capability of *LiveLabs* is the use of mobile sensing software, implemented on participants' personal devices, to collect and generate *detailed real-time* (or near real-time) individual and group context, and make such contextual attributes available to experimenters in a controlled way. While there has been a lot of work (e.g., [9]) to address the well-known energy challenges of continuous mobile sensing and data collection, we believe that *LiveLabs* imposes a set of unique challenges:

Unified context collection: Most mobile sensing implementations tune the adaptation of sensor data collection for a specific purpose–e.g., accelerometer samples for activity recognition or GPS data for capturing movement trajectories. *LiveLabs*, however, is intended to serve a wide diversity of experimentation needs, possibly concurrently! For example, different experiments may require accelerometer data for both inertial tracking-based indoor localisation, as well as queue detection. Such dual requirements may

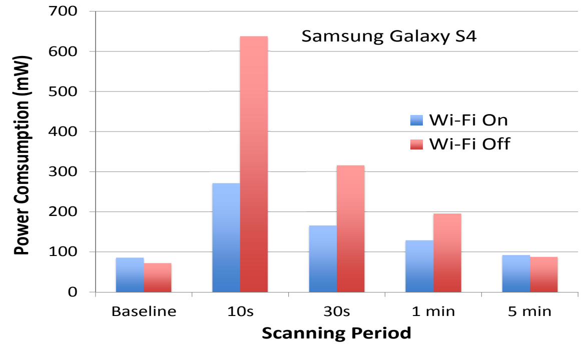


Figure 2: Power vs. Wi-Fi Scan Interval

restrict the possible system-level adaptations (e.g., sampling frequency), thereby limiting the extent of energy savings possible.

Support for multi-platform, consumer devices: In our current testbed, we have a roughly equal fraction of iOS and Android users (with WP8 devices soon to appear). As is well known, these platforms differ significantly in the amount of sensor data and application events that may be collected in the user space. As *LiveLabs* collects data from the personal device of an individual participant, research workarounds, such as "rooting the phone", are not permissible. Accordingly, the quantity and types of context that can be sensed may vary significantly across participants, and also continues to evolve over time, as new versions of iOS and Android get rolled out. This has implications for both the quality and types of context collected–for example, fine-grained Wi-Fi based location is possible for Android but not for iOS devices. As an illustrative consequence, for *Experiment 1* described earlier, this implies that *group detection* cannot rely on universal ability of fine-grained location data, but should be practically possible even in case when members of the group have a mix of Android and iOS devices.

Dependence on Individual Usage Patterns: In many situations, the energy overhead is not well-known *a priori*, but depends on the individual participant's device usage pattern. For example, we observed (Figure 2) that the energy overhead of performing Wi-Fi fingerprinting based location tracking on a smartphone is much lower when the Wi-Fi radio is already on, compared to the case where the Wi-Fi is intermittently activated to collect scan measurements. Accordingly, different users will have different perceptions of the "energy intensity" of our monitoring software, depending on whether or not they stay connected to the SMU Wi-Fi network.

Generate Useful Real-Time Analytics Output: LiveLabs must not only collect the individual device-level context, but apply advanced analytics on these incoming streams to infer a variety of interesting individual and collective context. For example, to address scenarios similar to *Experiment 4*, we have developed a prototype system to infer queuing activities in real-world locations (e.g., at cafes and movie theatre counters) from accelerometer data. Similarly, to address Experiment 1, we have developed techniques to determine the composition of shopper groups in malls from the sensor data of multiple devices. These stream analytics also present significant system challenges as they require performing computation-intensive operations (such as clustering or dynamic time warping) on tens of thousands of streams in real time.

We are exploring a few key ideas and research directions to address these challenges. To conserve sensing energy, we have the *LiveLabs* server infrastructure periodically contact each participant phone and adjust its collection policy (driven by the current experiment needs). In particular, we note that several use cases require fine-grained data only for short durations (e.g., *Experiment 4* requires real-time capture and analysis of accelerometer samples *only when* the user is in the neighbourhood of the immigration area). Accordingly, the *LiveLabs* server employs a default policy of collecting coarse-grained data over longer durations, and *triggers* fine-grained, energy-intensive data collection only when an experiment's context requires it. To potentially further reduce the energy overheads, we are also investigating the usefulness of a probabilistic query framework. In this approach, we can use correlation in the context across multiple individuals (e.g., inferring that person B is likely to be in the cafeteria, from the observation that A is in the cafeteria) to pro-actively infer B's context, without actually activating the necessary sensors. Finally, we are engaging with leading industry research firms to implement our analytics algorithms on high-performance stream processing engines.

Current Status: We currently have deployed our Context Collector software over both iOS and Android platforms. By default, the Android version runs as a background service and can collect sensor data, phone events, activity events and location (both indoor and outdoor) continuously; it currently uploads the data over a secure channel every 3 hours; as mentioned, the Context Collector server can, *on-demand*, dynamically instruct such clients to collect and upload higher-fidelity sensor traces much more frequently. The iOS version can collect sensor data, a limited set of phone events and location; moreover, as iOS only allows a background App to run once every 10 minutes, the context is collected intermittently. The collected data is received by the *LiveLabs* server, where the data is first anonymised (by applying one-way hashes to the sensitive fields), before being stored for use by the Analytics component.

3.2 Practical Indoor Location

Given that 3 out of the 4 *LiveLabs* testbed locations are predominantly indoor spaces, determining the indoor location of an individual (or a designated group) is clearly an important piece of context. In spite of significant recent research on low-overhead indoor location tracking [15, 5], we believe that building a practically deployable, large-scale indoor location system is still an unsolved problem, especially in Asian urban settings characterised by extremely high and *variable* occupancy densities.

Currently, we have deployed an operational fingerprinting-based solution at *LiveLabs@SMU*. Additionally, we have tested existing solutions at two other locations: *LiveLabs@Changi Airport* and *LiveLabs@Mall*. During these tests, existing location tracking solutions often came up short for several reasons:

Changing Environmental Conditions: Fingerprinting-based solutions are known to not perform well when the underlying propagation environment changes dynamically. At both *LiveLabs@Changi Airport* and *LiveLabs@Mall*, such environment changes are common and occur at both: i) medium time-scales, caused by changes in the layout (e.g., the mall has a different layout of temporary bargain bins and display carts every week), and ii) shorter time-scales, due to changing densities of occupants (e.g., the peak occupancy level of the mall is 3x higher than non-peak). Manual fingerprint *maintenance* is not feasible as the timescales are too small; neither is it possible to build *a-priori* fingerprint models for specific times of the day, as the occupancy levels are not predictable.

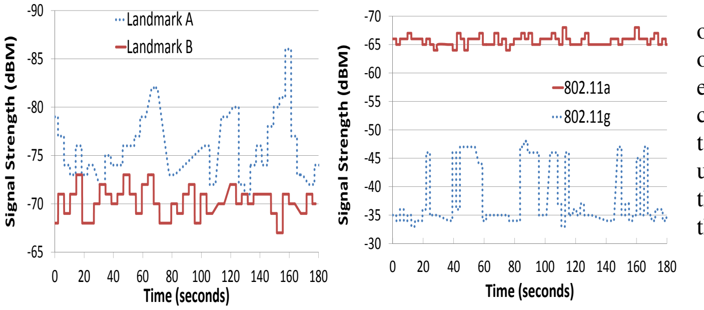
Support for Heterogeneous Devices: Most RF-based location tracking research has focused on client-side localisation — where a mobile device's RSSI scans are used to compute its location. While such RSSI support is available on Android, corresponding APIs are not available, without root access, for other smartphone OSs such as iOS, or WP8. We thus cannot rely on client-based localisation as our participant base uses a mix of iOS and Android devices.

Heavy Tail in RF Fluctuations: Localisation research prototypes and results often focus on "representative" results, such as the median (or 90th percentile) error. In practice, we have observed very high variance (up to $\pm$30dBm) in Wi-Fi RSSI readings, not everywhere, but in an appreciable fraction of the indoor locations at *LiveLabs@SMU* and *LiveLabs@Mall*, even at low occupancy density and without any ostensible environmental changes. As a consequence, location errors may be within acceptable limits 80-90% of the time, but may be significantly higher for non-negligible periods (10-20%). Figure 3a shows the RSSI signal strength of an AP measured by the same Android phone at two nearby locations at *LiveLabs@SMU*. As shown, the readings are fairly stable at landmark B, but show a fluctuation of over -16 dBm at a nearby landmark A.

Limitations of Commercial Deployments: Many research prototypes provide compelling evidence that location accuracy can be increased by augmenting the underlying infrastructure (e.g., adding UWB or ultrasound beacons). Moreover, many research prototypes are tested in campus or office environments, where the Wi-Fi deployment is well-planned and well-provisioned or during low-load off-peak hours. In our experience with the various *LiveLabs* venues, Wi-Fi deployments may not be as extensive or carefully architected. More importantly, modifying the existing installed infrastructure (by either adding new APs, or installing various new sensor or beacons) is a very challenging process that often takes several months to materialise, for a variety of concerns such as logistics (e.g., adding additional power outlets), aesthetics (upgrades are only possible during planned renovation periods), and security concerns (e.g., at the airport). In practice, we found that sparse Wi-Fi AP deployments are common and results in non-obvious trade-offs between the technology used, accuracy, and coverage. For example, we empirically observed (Figure 3b.) that 802.11a channels exhibited greater RSSI stability than corresponding 802.11b channels — albeit at the cost of reduced range and coverage.

Current Status: Given these observations, we posit that building and deploying a solution that provides *store-level* accuracy (quantitatively, accuracies of 3-5 meters, as typical mall store fronts have a width of 8 meters) consistently is not a trivial task. To support a more universal list of clients, we have currently implemented a *server-side* location tracking mechanism, based on the RADAR algorithm, where we use *uplink* RSSI measurements (reported by the APs) of transmissions from the phone to locate individual devices. To retrieve such measurements, we have built vendor-specific query interfaces for the Wi-Fi controllers in each testbed venue.

Our location system has been deployed across the entire *LiveLabs@SMU* campus and captures (with appropriate device id anonymisation) the location of every device that connects to our campus Wi-Fi. One of the most important advantages of server-side location tracking is its *energy efficiency* as it imposes no additional client device energy overhead. However, server-side location tracking is not trivial either, and must address challenges such as: i) *device heterogeneity*: different users have different mobile devices, which transmit at different power levels and thus require device-specific calibration; ii) *advanced power management*: mobile devices routinely adjust their transmission power for data packets and the transmission frequency for management packets (e.g., the PROBE_REQUEST packets used to search for available APs), based on both the quality of the channel and the residual battery capacity, implying a level of unpredictability about the accuracy of location tracking. Moreover, manual maintenance of such fingerprint maps is unsustainable, given the continually changing environments. Accordingly, an important research goal is to develop an automated algorithm for fingerprint map evolution, that progressively incorporates new measurements from *LiveLabs* clients and dynamically updates the fingerprint map, to track evolutionary changes in the environment (at both medium and shorter time scales).



a. **Variation across 2 Landmarks** b. **Variation: 802.11a vs. 802.11g**

Figure 3: RSSI Variation at *LiveLabs@SMU*

3.3 Privacy-Conscious, Useful Experimentation

As described in Section 2.3, the BEP is a key component of *LiveLabs*, using contextual triggers to actuate specified experiments/ interventions on matching subjects. While the function of selecting subjects for experimental "treatment" seems, ostensibly, to be a problem for experimentalists and statisticians, the fact that the subjects are selected based on *personal context* collected from *energy-constrained* mobile devices leads to several challenges that straddle mobile systems design and experimental statistics:

1. Experiment Specification Complexity: Experiments consist of two components — a *trigger* that specifies a set of contextual predicates that must be matched, and an *action* that specifies the resulting experimental treatment. While there are complex ontological models for context specification, such specifications would clearly be lost on the non-technical personnel (e.g., sales managers) who might be our primary experimenter base. A key challenge is thus to develop an intuitive, yet sufficiently expressive, GUI interface that allow experimenters to perform differentiated experimentation.

2. Participant Numbers and Control Conditions: Unlike laboratory experiments, *LiveLabs* has little control over the participants who satisfy an experiment's contextual constraints. Hence, the BEP must permit some form of soft-matching, especially when context inferences are uncertain. For example, if several participants have a location granularity of $\pm$5meter, and can thus be either inside or outside a store, does the BEP deliver a discount coupon that is supposed to be dispatched "when a shopper enters a store" to all, none or some of these participants? Making this harder is the realisation that setting up a *control* group in real environments like *LiveLabs* is hard. E.g., assume that we observe that "only 20% of participants in front of a store enter it after receiving a store-specific discount" — we may not be able to create a corresponding control group, as clearly we should avoid spamming individuals who are not in front of the store with this message.

3. Balance Result Validation vs. Privacy: One of *LiveLabs's* key privacy features is that experimenters can test with *LiveLabs* participants without gaining the participants' personal context details. However, in certain situations, implicit privacy exposure is unavoidable – e.g., if a participant actually redeems a context-based discount, the experimenter will be aware that she has satisfied the corresponding trigger. There is thus a possible privacy cost associated with every context-based experimentation. More intriguingly, an overly aggressive experiment specification (e.g., "check if a person stands for 5 secs at any point in the mall") may cause the *LiveLabs* server to require deep context from an individual for a long period of time, thereby potentially draining the energy of the mobile device. Consequently, experiments often have a non-obvious energy cost as well. Thus, while data collection must be detailed enough to satisfy the quality requirements of analytics algorithms, such data collection comes with two costs: i) the energy overhead on the mobile device of the participant, and ii) the likely exposure of fine-grained and sensitive behavioural context information. To eventually translate the technologies developed by *LiveLabs* to real commercial environments, it is important that we as a community tackle this problem of systematically modelling such costs (e.g., using metrics such as differential privacy), so that users can view their context data as a virtual currency ([8]) and trade off between the triad of: energy overhead, privacy and analytics accuracy.

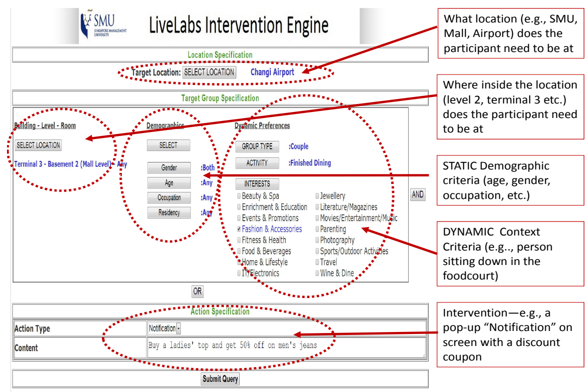


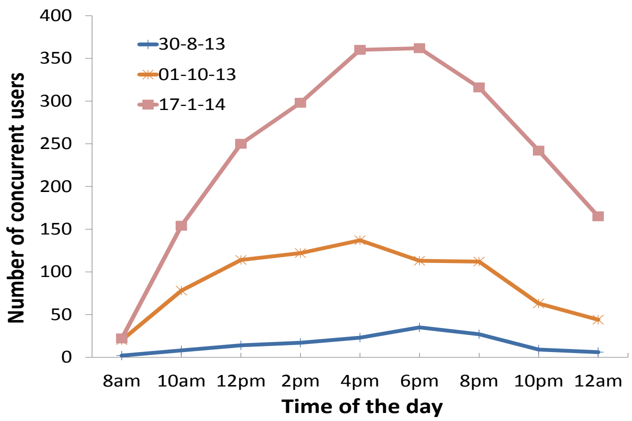
Figure 4: UI For Specifying Experiments

Current Status: Currently, we have an initial UI prototype, shown in Figure 4, that allows experimenters to specify their constraints, such as deliver a specific targeted discount to a set of students in *LiveLabs@SMU* who are moving around in groups of two, as a set of logical rules on a set of predefined context predicates, that can be chained together using explicit AND operators. We also currently have a Web-based application portal that retailers are using to advertise promotions to *LiveLabs@SMU* participants.

4. CURRENT STATUS & DATA COLLECTED

LiveLabs@SMU became operational on August 31st 2013, and as of January 20th 2013, it has 1,954 registered users (52.7 % on iOS and 45% on Android) with 741 active participants (i.e., the *LiveLabs* context collector is running properly on their phones). Furthermore, our participant pool is growing every day and is expected to reach a few thousand active users by the first half of 2014 as we ramp up at SMU (about 10% of university population is currently active user) and deploy *LiveLabs* at other venues. Participants in *LiveLabs@SMU* are incented to join *LiveLabs* via two different incentives: *a)* Financial, such as a monthly rebate off their phone bill and lucky draws, and *b)* Novel context-based Apps, that address participant needs. In particular, on the urban and highly-crowded SMU campus, we have found real-time access to occupancy "heatmaps" (of study areas and benches) to be a major draw in getting participants to sign up.

To enable real-time experiments, *LiveLabs* will require a certain number of users to be present at the same location — both to achieve the minimum amount for a specific experiment, and to provide enough users to run multiple independent experiments in parallel. Figure 5 shows the number of concurrent *LiveLabs@SMU* users on the SMU campus. For a recent day(January 17th 2014), the average number of concurrent users was $\approx$240 with up to 360 concurrent users between 4 PM and 6PM . These numbers are quite encouraging and already allow a number of interesting experiments to be run. In terms of demographics, *LiveLabs@SMU* is highly skewed, as expected, towards the 18 to 24 year old age group (only about 6% of the users are outside this age group) with 53 % of the users self-reporting themselves as heavy phone users.



Each line is a different day (date given in legend)

Figure 5: No. of concurrent *LiveLabs* users over a day

5. THREE KEY LESSONS LEARNED

Addressing 'Worst-Case' Behaviour is Important: From proof-of-concept location tracking deployments at *LiveLabs@Changi Airport* and *LiveLabs@Mall* , we learned that venue operators seriously care about the worst-case performance. In particular, a system that has worse average performance, but provides stable service, is better than an alternative that has better average performance but an appreciable performance drop-off (e.g. long CDF tail). Venue operators do not want their users to see, for example, highly accurate (say 2-3 meter granularity) location tracking at most times, but suffer accuracy degradation (dropping to 6-8 meter granularity) during highly congested peak periods. Even a 1% drop-off can translate to thousands of affected users in a dense environment (all our venues are very dense!) and lead to appreciable customer dissatisfaction. This lesson motivates us to explore *adaptive* indoor location methods instead of re-using existing solutions.

Privacy turned out to be secondary to energy concerns: On the SMU campus, we were surprised when our *LiveLabs@SMU* participants did not raise many privacy related questions. Instead the energy consumed by the *LiveLabs* service application was a bigger concern than privacy! This surprised us as we expected the participation rate to be determined by the privacy policies put in place. Instead, we needed to focus on making our applications as energy efficient as possible. When asked about the lack of privacy concerns, some participants replied that privacy was a lower concern as they lived in a densely populated urban city where their activities were already known by numerous people. Clearly, however, this phenomenon may be different, once we deploy in public venues, such as *LiveLabs@Changi Airport* and *LiveLabs@Mall*.

Operational issues are paramount: As academics, we assumed that the operational aspects of *LiveLabs* would be easy — a very bad assumption!! First, we learned that negotiating with real companies takes a long time, even when both sides want to work together, as approvals need to be obtained from multiple places within the organisation. We found that it takes at least 8 months to sign a new industry partner with some negotiations taking years.

Second, running a research lab with production requirements requires fairly fundamental operating model changes. For example, the standard "just-in-time" scheduling method does not work very well and every piece of code needs to be rigorously checked for bugs before being released to end users (a task that is rarely done in research). In particular, any software released to end users needs to be supported from that point onwards, no matter how buggy that code is, as not all users will upgrade their software even when asked. Thus you either have to support older software versions or risk losing users and receiving bad feedback. To overcome these production challenges, we partitioned the lab into two parts; 1) a research arm that focuses on solutions that are needed 6 to 12 months down the road, and 2), a production arm, with a full-time project manager, that hardens and operationalises ready-to-use research solutions, builds any required non-research system (for managing participants etc.), and provides support to our participants.

6. CONCLUSION

In this paper, we have described the *LiveLabs* Urban Lifestyle Innovation Platform — a novel real-world testbed that aims to allow testing of mobile solutions, services, and interventions in real environments with real people on real phones. To enable this, there are a number of key research challenges that need to be overcome and these challenges as well as some of the solutions have been described in this paper. In addition, we have also discussed some of the practical challenges we faced in making *LiveLabs* a reality.

7. REFERENCES

[1] *PlaceIQ*. http://www.placeiq.com/.

[2] *Xtify, Mobile Customer Engagement*. http://www.xtify.com.

[3] Bejarano, O., Miskovic, S., Aryafar, E., and Knightly, E. TFA: A large scale urban mesh network for social and network research. *Proc. of ACM S^3 Workshop*, 2010.

[4] Changi Airport Group. *Facts & Statistics*. http://www.changiairport.com/our-business/about-changi-airport/facts-statistics.

[5] Chintalapudi, K. K., Iyer, A. P., and Padmanabhan, V. Indoor localization without the pain. *Proc. of Mobicom*, 2010.

[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. *CACM*, 55:82–90, Jan. 2012.

[7] 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. *Proc. of Nokia Mobile Data Challenge Workshop*, 2012.

[8] Leman-Langlois, S. Privacy as currency: crime, information and control in cyberspace. *Technocrime: technology, crime and social control*, pages 112–138, July 2008.

[9] Lu, H., Yang, J., Liu, Z., Lane, N., Choudhury, T., and Campbell, A. The jigsaw continuous sensing engine for mobile phone applications. *Proc. of SenSys*, 2010.

[10] 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. *Proc. of IEEE WCNC*, 2005.

[11] Sentosa Development Corporation. *Sentosa 40 Years*. http://www.sentosa.gov.sg/wp-content/uploads/Media-Release_Sentosas-40th-Anniversary_Aug-2012_FINAL.pdf.

[12] Singapore Management University. *Quick Facts*. http://www.smu.edu.sg/smu/about/university-information/quick-facts.

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

[14] SUNY Buffalo. *PhoneLab - A Programmable Smartphone Testbed*. http://www.phone-lab.org.

[15] Wang, H., Sen, S., Elgohary, A., Farid, M., Youssef, M., and Choudhury, R. R. No need to war-drive: unsupervised indoor localization. *Proc. of MobiSys*, 2012.