Research in the Large. Using App Stores, Markets, and Other Wide Distribution Channels in Ubicomp Research

Henriette Cramer[1], Mattias Rost, Nicolas Belloni
Mobile Life Centre / SICS
Kista, Sweden
{henriette, rost, nicolas}
@mobilelifecentre.org

Frank Bentley
Motorola Applied Research
Schaumburg, USA
f.bentley@motorola.com

Didier Chincholle
Ericsson Research
Kista, Sweden
didier.chincholle@
ericsson.com

ABSTRACT

The mobile phones that people use in their daily lives now run advanced applications and come equipped with sensors once only available in custom hardware in UbiComp research. At the same time application distribution has become increasingly simple due to the proliferation of app stores and the like. Evaluation and research methods have to be adapted to this new context to get the best data and feedback from wide audiences. However, an overview of successful strategies to overcome research challenges inherent to wide deployment is not yet available. App store platform characteristics, devices, reaching target users, new types of evaluation data and dynamic, heterogeneous usage contexts have to be dealt with. This workshop provides a forum for researchers and developers to exchange experiences and strategies for wide distribution of applications. We aim at building an understanding of the opportunities of various distribution channels and obstacles involved in a research context.

Author Keywords Mobile interaction, app stores, distribution channels, mass evaluation methods, mobile ecosystem

ACM Classification Keywords H.5.2 User Interfaces: Evaluation/methodology

General Terms Experimentation, Human Factors

INTRODUCTION

During recent years, the landscape of means to distribute mobile applications has dramatically changed. Until recently, mobile applications had to be installed manually or downloaded from individual portals, limiting the ability of the research community to test new applications with a wide audience. Often, complex relationships with mobile operators or manufacturers was necessary just to launch a public beta. Since then, the number of distribution channels to reach wide audiences has exploded and all major mobile platforms now offer their own 'app store'. The burden of reaching users for applications has been greatly diminished by these channels, such as the Apple App Store, Android Market, Nokia's Ovi and a host of other platforms. Both lone developers and large research and development teams can now reach wide audiences relatively easily.

At the same time, the possibilities for creating advanced mobile and sensing applications have increased as well, due to the increasing capabilities of both handsets and (web) technologies and the availability of accessible development environments. In the past, development of custom hardware or handing out specific handsets was part of user testing and proving a concept. Increasingly, mobile phones that the general public already carry in their pockets have the hardware and software capabilities we need to test many of our research applications and usage scenarios. This combination of handset improvement and distribution market maturation provides a huge opportunity for both iterative development and gathering research data from a wide audience (and their devices) in everyday contexts of use.

New ways of reaching participants in studies have also appeared, including, for example, Amazon's Mechanical Turk that lets researchers specify tasks that are performed by users over the net [6]. This has the potential of having many small tasks being performed by many users in a short amount of time. Potential users can also be reached through integration with existing services (e.g. through Facebook Connect or promotion via Twitter (re)tweets). Indeed, a selected number of researchers are taking advantage of the opportunities of wide deployments. This offers opportunities for research but also brings up challenging questions. McMillan et al. [7], for example, describe their experiences deploying an app using the iPhone App Store. They describe issues ranging from (self)selection and recruiting of participants, potential difficulties concerning informed consent and moving from quantitative aggregate data to qualitative detail.

We need insight in the practical issues of reaching public audiences for research applications.

Many challenges remain in scaling user studies from small evaluations to full-fledged deployment in the wild. Research,

[1] Cramer's work was carried out during the tenure of an ERCIM "Alain Bensoussan" Fellowship Programme

for instance, might now require promotion and business plans. As researchers in academia, it is not ideal to 'only pay a dollar' for participation, if you get two million participants. Simply 'throwing an application out there' and hoping to get 'something' back is hardly a satisfying strategy. Platform compatibility, handset fragmentation, evaluation methods, user feedback and research methods need to be explored and adapted to this new context. As a community however, we do not yet have a deep understanding of how these channels can be used, nor have we developed the research strategies and methods needed to fully grasp these opportunities. Beyond practical issues, we have also not yet engaged in serious discussions on the consequences of using these methods for our research from both validity and ethical standpoints. Which research questions are best answered through these large deployments? How can researchers validate the integrity of collected data? Understanding the issues involved and a deeper insight in these platforms is currently needed to make the most of this opportunity. This workshop aims at providing a platform for researchers from both academia and industry to discuss these opportunities and challenges from their own perspectives and previous experiences. Together we will investigate which strategies work (and which do not) and the opportunities that wide deployment brings.

MOTIVATION

App stores have spurred a revolution for distribution opportunities. (Academic) research to a certain extent appears to lag behind, with user studies in venues like UbiComp traditionally focusing on small-scale deployments and limited (lab) experiments. Although some of the systems developed in the community have been deployed to the general public, the majority of them have been confined to research sites and limited groups of test users. However, not only are industry research teams distributing their applications to the masses, the first steps of wide deployment of research prototypes and gathering of real world data are already being taken [e.g. 4, 7, 3, 8]. Such new possibilities bring about huge opportunities, but also world-scale challenges. Data is being collected from app store deployments, but no serious discussion has occurred on the validity of this data compared to existing controlled research methods. We do not yet know the implications and possible side-effects of wide distribution methods. Fundamental questions are raised whether these are actually appropriate strategies. We hope that this workshop can serve as a venue to begin these discussions.

Opportunities

Wide deployment and rapid handset development has enabled the development of new and advanced services. Research concepts can now in theory be turned into actual applications used by 'people in the wild' within a matter of hours. Mobile applications can now also be used to collect large amounts of data from their users and their devices and can target users in specific locations of interest or provide a large overview of location-based data [2, 4, 5].

Lab evaluation, or limited field deployments, does not take into account the rich possibilities of use that mobile systems provide. They also do not expose the ecosystem of network operators, device manufacturers, other users, developers and distribution and payment models, which all ultimately contribute to the user's experience or detract from it [9]. Factors such as steps needed in installing applications, paying for services, integration with other applications, media, and data on the phone and policies of service providers cannot all be considered in small, local studies. Trying out a mobile system while sitting in a lab is considerably different than using the latest application onto your mobile phone while at a coffee shop in New York or out with friends. Beyond considering the impact of the mobile ecosystem on users, wide distribution also offers chances for iterative development that takes advantage of feedback and ideas of users whom researchers historically may not have had access to. Wide deployment is also crucial for investigating the social impact of applications. While reaching a certain mass of users is vital for most social applications, the effects of mass usage of services and emergent social practices also are hard (if not impossible) to explore in smaller scale studies. Finally, wide distribution allows for learning about varied contexts of use and accuracy of sensors in a variety of environments that are not feasible for controlled experiments. The new distribution channels offer a great help in reaching the large and varied groups of end-users needed to gain such crucial insight. Not taking advantage of the available distribution channels and app stores would be missing an important opportunity. But great community effort is required to overcome the new challenges that accompany the opportunities.

The challenges we are facing

Getting the users and data we need

Wide deployment enables collection of large amounts of data, but we need to make sure we actually get the data we need. Relatively few ways are available to limit deployments when using open distribution channels, possibly resulting in a 'all or nothing' tradeoff. The choice for a specific platform might also mean reaching a specific type of people (to invoke some stereotypes: more wealthy users or design-oriented users on the iPhone platforms, geeks for the Android market). On some platforms, releases can be restricted to specific countries; the chosen locales an application is released in will affect results. How do we 'recruit' an appropriate mix of users for deployments if we cannot fully control - or even know - who downloads our applications? With the growing number of available applications, developers also need to deal with promoting their application and making sure it reaches potential end-users. Will we only reach early adopters with unlimited data plans who are trying all new applications? Who is most likely to provide feedback? Associating demographic data with specific feedback or usage logs is a difficult problem. Also, how should an application be presented to set the appropriate expectations? Presenting an application as a research

prototype might infer a negative bias on the application, where as presenting it as a commercial application might build too high expectations. Is presenting the application as beta version of a commercial service the solution? What kind of expectations does that infer? How does the way an application is presented in promotion, PR and buzz affect research results?

Different platforms provide different analytics tools and ways of gathering feedback. Applications in app stores potentially gather a large corpus of user comments. The problem from a research perspective however is how valid and reliable these comments actually are. Which are representing the general public, and which are 'nerd-rage'? How do we analyze such feedback? Feedback could be very valuable for iterative development and including new features. Do we need to update our applications and deal with version management issues? When are we getting caught in a consumerist cycle instead of doing research? How can we gain more reliable user feedback that is useful for research? We need both strategies for large deployment and triangulation with for example smaller ethnographic-style studies [1].

Technical development issues
Wide deployment means having to deal with a wide variety of brands, device types, hardware limitations and plat-forms. Developers need to decide whether they make a choice for a specific platform or reaching a wider audience and supporting multiple. For instance, the experiences of Zhai et al. [11] in bringing a mobile research prototype to the market via wide distribution, illustrate a number of challenges and opportunities that the choice for wide deployment via a particular platform (e.g. iPhone) bring. Even while not specifically focusing on these issues, Zhai et al. touch on, for example, having to work around development challenges posed by the iPhone's SDK and Apple approval process. In their case this only took a week, but less positive experiences are reported [10].

The specific choice of a platform will result in tradeoffs for both developers and end-users. From the user's perspective usability of markets and app stores themselves can differ in, for example, the effort and ease in downloading and installing a new application. For developers, the investment in having to learn about a specific platform, the tools provided, the ease with which applications can be developed and uploaded are of considerable importance. The choice of a platform has consequences on flexibility in research and development (e.g. uploading and updating).

Miluzzo et al., [8] provide a comparison between platforms on some of these issues, but such overviews are quickly out of date and keeping up requires quite some effort. In addition, the choice of one platform means having to deal with specific brands and phones. It is virtually impossible to foresee all potential compatibility issues and needs for changes and updates that may occur during deployment, making planning more difficult. Stress tests will also be necessary to make sure research infrastructures can deal with large numbers of users and data traffic.

Beyond choice of platform, wide deployment introduces other deep technical issues. How should mobile applications be instrumented to better collect meaningful usage data? And how should this data be analyzed and used? How robust should the applications be in dealing with a wide variety of networks and different sensor hardware on different platforms? Compass and accelerometer calibration issues on different manufacturer's Android phones are one recent example of this issue.

Apps under construction
Releasing an application to a wide audience via established distribution channels raises expectations. An application presented as a research prototype will invoke different reactions than an application presented as a finished product. At what stage is a research prototype actually ready for wide distribution? What caveats or labels should be posted along with the application? Putting apps out there also bring both chances and added complexities in dealing with exposure. If an application under development gains great momentum, possible effects on developers' and researchers' reputations need to be considered. Controlling feedback and publicity, both positive and negative, is an illusion. Research involves risks and 'mistakes'. How will these brush off on other prototypes we might want to release? How does it impact evaluations and the data we gather? Can specific strategies help gain the most of user feedback, while avoiding potential disasters?

Monetary Concerns
With wide deployment potentially come other expenses not present in small field or lab studies. With an increase of users potentially comes large server operating costs, costs of promotion in app stores, and additional tech support for users who may be experiencing issues. How do we handle these costs? (How) should users be compensated?

Playing by 'the rules'
Procedures, regulations and ethics for research in app stores and across countries are unclear. How do we deal with informed consent if we use data from users of freely available applications? Are disclaimers and notices during installation sufficient? Which legal issues do we have to deal with? Henderson and Ben Abdesslem [5] point out the differences between countries and cultures in terms of ethics, regulation and practices; wide distribution has made knowledge on cross-cultural research even more crucial.

WORKSHOP GOALS
Wide deployment and evaluation on a grand scale are challenges both industry and academia are wrestling with. We need to make the most of the opportunities that are currently presenting themselves. Practical strategies, tips and tricks are needed to overcome these challenges, along with insight in how these issues impact on our research results. We aim to build on the participants' experiences on available mobile distribution platforms to create an overview of the

strategies and methods that have worked (and those that did not) and the effects of various ecosystem features on research and evaluation results.

This workshop features contributions from developers and researchers from both industry and academia. We especially have sought contributions on experiences in deploying mobile applications and evaluations at a larger scale.

Both papers on past experiences and position papers outlining challenges and proposed strategies have been accepted. Workshop themes include:

- Creative uses of data from the wild collected through wide distribution and experiences in wide distribution of research prototype applications

- Strategies for research and development using app stores, markets and other wide distribution platforms. How to recruit a suitable sample and analyze usage logs, and feedback on publicly released applications in for example app stores.

- How to combine methods, such as augmenting large-scale user studies with smaller ethnographic style studies.

- Effects of differences between distribution platforms on evaluation and research results

- Differences in experiences based on ecosystem properties (distribution, payment, business models, service providers etc). Which aspects are relevant?

New developments in the landscape of app stores (such as developments for other devices than phones only, or evaluations of app stores themselves) and views on whether we are heading towards a new 'app store bubble' will be part of the discussion at the workshop.

Workshop contributions are two to four page papers addressing previous experiences and for example case studies, or position papers outlining challenges not yet addressed. A program committee consisting of both industry and academia partners has reviewed the solicited contributions.

Accepted papers can be found via the workshop's website at large.mobilelifecentre.org.

REFERENCES

1. Ames, M. and Naaman, M. 2007. Why we tag: motivations for annotation in mobile and online media. *Proc. CHI'07*, 971-980.

2. Anderson, I., Maitland, J., Sherwood, S., Barkhuus, L., Chalmers, M., Hall, M., Brown, B., Muller, H. 2007. Shakra: tracking and sharing daily activity levels with unaugmented mobile phones. *Mobile Networks and Applications.* 12, 185-199.

3. Campbell, A., and Choudhury, T. 2009. Toward Societal Scale Sensing using Mobile Phones. White paper in *Proc. NSF Workshop on Future Directions in Network Sensing Systems*.

4. Fraunhofer AICOS, mover.projects.fraunhofer.pt

5. Henderson, T. and Ben Abdesslem, F. 2009. Scaling measurement experiments to planet-scale: ethical, regulatory and cultural considerations. *Proc. HotPlanet '09*.

6. Kittur, A., Chi, E., Suh, B. 2008. Crowdsourcing User Studies With Mechanical Turk, *Proc. CHI'08*, 453-456.

7. McMillan, D., Morrison, A., Brown, O., Hall, M. Chalmers, M. 2010. Further into the Wild: Running Worldwide Trials of Mobile Systems, *Proc. Pervasive'10*, 210-227.

8. Miluzzo, E. et al. 2008. Evaluating the iPhone as a mobile platform for people-centric sensing applications, *Proc. UrbanSense'08*.

9. RuiXue, X., Rost, M., and Holmquist, L. E. 2010. Business models in the mobile ecosystem. *Proc. of Global Mobility Roundtable*, Athens, Greece.

10. Wired, 2009 Apple's appalling approach to iPhone app approvals. http://bit.ly/7maDL

11. Zhai, S., Kristensson, P.O., Gong, P., Greiner, M., Peng, S.A., Liu, M.L., Dunnigan, A. 2009. Shapewriter on the iPhone: from the laboratory to the real world. *Proc. CHI'09*, 2667-2670.