Finding the Right Nails: Scenarios for Evaluating Pervasive Systems

John J. Barton and Jeffrey Pierce, IBM Almaden Research Center, San Jose CA. 95120 Position Paper for Common Models and Patterns for Pervasive Computing, 2007.

Effective evaluation aids research progress. Effective evaluation of user-facing computing systems requires metrics determined by analysis of scenarios of use. Therefore our position is that *the critical metrics for evaluating pervasive systems research should be selected from scenarios based on real problems*.

Scenarios are a vital communications tool for pervasive systems design. However, not all scenarios are equally valuable as a source of metrics. Very often our scenarios are motivated primarily by new possibilities created by technology. We'll call these inventor-originated or "hammer" scenarios, since we often refer to our technology as a "hammer looking for a nail." Recently we described an initial framework for semi-quantitative analysis of such hammer scenarios[1]. For system evaluation we need scenarios that are driven by real problems rather than by the technology. These would be realistic scenarios where pervasive computing technology can or could have unique and significant value. In this position paper we characterize problem-motivated "nail" scenarios, connect these kinds of scenarios to ones used for evaluation in other areas of software, and illustrate by example how they can be used to define evaluation metrics for pervasive systems.

Nail Scenarios For Evaluation

Nail scenarios are fundamentally different from hammer scenarios. A nail-scenario problem exists in some form at the time the solution is proposed; hammer scenarios typically speculate on problems of the future (e.g. [4,5]). The adequacy of a solution to a nail scenario problem can be judged by direct comparison to existing solutions; hammer scenarios are invented to illustrate a technology, so they should be judged only against plausibility[1]. We import nail scenarios into pervasive computing as challenges we may solve; we export hammer scenarios as somewhat fuzzy images of a possible future life for computer users.

Despite these differences, both types of scenarios communicate technology ideas by relating them to circumstances people understand. Novel research approaches to both kinds of scenarios can involve lots of "magic" because the research ideas are creating scenarios of possible future usage. Here we are interested in going in the opposite direction: driving research progress with measurements extracted from scenarios.

Our nail scenarios are the kinds of scenarios used in human-computer interaction (HCI) design [6]. However, in HCI design the scenario is used throughout the lifetime of one design [7], while we propose to use scenarios across systems at a single point in time. This is the sense in which scenarios have been used in evaluation of software architecture [8] and in the evaluation of middleware (aka infrastructure) [9]. The previous work provides a solid base for application of scenarios in evaluation.

Among the key findings of past work we highlight two. First there exist cognitive frameworks for organizing and relating scenarios [6,7]. These are somewhat like the frameworks used in software architecture. They enumerate the issues that need to be considered. For example, we need to consider the spectrum of people involved in scenarios, their division of work and organizational roles, and we need to consider the physical environment and the objects as they impact the design. These frameworks help us consider the scenario goals, constraints, and interdependence with other scenarios.

Second, we know from the use of scenarios in evaluating software architectures and for evaluating middleware that multiple scenarios are needed. This requirement should apply to pervasive systems because all three areas have *diffuse value*: the more an architecture, middleware, or pervasive system can be re-used for related or co-located problems, the greater its value. These kinds of technologies have up-front costs: we need to design, develop, deploy, and test systems which may not directly solve a user's problem. Then we begin to benefit from our investment as more and more applications leverage the architecture, middleware, or in our case the pervasive system. Thus we anticipate creating a suite of pervasive scenarios.

Scenarios for Pervasive Systems Evaluation.

The best pervasive computing systems will provide the best user experience while leveraging the key attributes of pervasive systems [2,3]:

- Ubiquity. A defining characteristic of pervasive systems, ubiquity changes use models by allowing computing activity to continue in space and time.
- Invisibility. (calmness, unobtrusiveness, embeddedness); computing should aid without itself being the central subject.
- Physical-virtual connection. Allowing the user experience to depend upon location, orientation, the presence of other people, computers, or objects.
- Heterogeneity with interoperation. Pervasive systems mix different kinds of devices on different kinds of networks.

We propose that evaluation scenarios be constructed by focusing on the pervasive characteristics of nail scenarios to go beyond the standard usability metrics. This will highlight unique features of pervasive systems: improvements in these areas will directly measure progress in our research.

Example: Healthcare Data Access and Entry

To get a concrete sense of the individual metrics that could emerge consider a clinic or hospital equipped with an Electronic Medical Record (EMR) system used by healthcare workers to read, enter, and modify records. Jakob Bardarm [10] describes a client-server PC-based EMR solution in deployment that face many challenges and he suggests ways the pervasive computing technology can help. The set of scenarios outlined by Bardram are all based on observed problems faced by users: these are excellent candidates for problem-oriented nail scenarios in pervasive systems. Each aspect of pervasive systems suggests metrics for progress:

Ubiquity: In the healthcare arena, access to medical records near patients is clearly advantageous, but doctors and nurses are highly mobile and collaborate on the move[10]. An important metric for a pervasive healthcare solution could quantify "availability" of the EMR, say as a fraction of the workers space or time. Ubiquity cannot sacrifice patients' privacy: the availability measure must be subject to security constraints. Medical records are voluminous and include images and videos: several different layers of availability, such as "patient-id only", "vitals only", "full text", and "full image" may be needed to usefully characterize ubiquity in practice.

Invisibility: Healthcare workers and patients need to communicate directly, aided by and not obstructed by computerized records: the technology needs to "disappear" while being a critical aid. This somewhat vague idea can be broken down in quantitative and semi-quantitative ways. The combination of mobility, shared space, and preservation of privacy conflicts directly with invisibility: security requires login/logout, a disruptive obstruction. The overall login time, the degree of task interruption caused by login for mobile workers, the probability of logout failure, the false-accept and false-reject rate of login technologies are just some of the metrics this in scenario. Subjective metrics include obtrusiveness of the computer technology on the human interaction, sense of trust engendered by the technology for both the worker and patient, robustness to failure, and adaptability to changing requirements.

Physical-virtual integration: Emerging RFID or barcode technologies in healthcare settings connect patients, medicines, or equipment directly to corresponding electronic counterparts. Some metrics, include error rates for identification, average and worst case response times, physical area coverage, object-type coverage, and quality of the information. Different rooms in a healthcare facility have different purposes, allowing location-dependent EMR systems to be evaluated against place-accuracy, specificity of the dependence, adaptability to changing facilities, and so on.

Interoperation: Medical facilities contain many different kinds of electronic equipment, soon to be connected by digital signaling. Interoperation by its nature is situational: a pervasive system that facilitates interoperation of Internet devices may have little value in one setting and help a great deal in another. For quantitative comparisons we may need to focus on specific user tasks assuming that the system allows the interoperation in principle. For example, the connection time for devices can be measured under conditions of use[11].

In addition to evaluation metrics based on specific properties of pervasive systems, some metrics arise from combinations of properties. For example, we can assess the cognitive interference (less invisibility) caused by various kinds of physical-virtual connection technologies or caused by mechanisms supporting interoperation.

Based on past work we expect to create a suite of scenarios related through a framework. In selecting scenarios we need to focus on real problems that would benefit from the primary characteristics of pervasive systems: ubiquity, invisibility, physical-virtual connection, and heterogeneity with interoperation. Likely candidates are scenarios that involve mobile users who collaborate with people locally, interact with local devices, and employ mobile or remote devices. Concrete examples would include healthcare data access and entry. The outcome of this approach would be a set of metrics measuring progress in detailed aspects of pervasive systems. The individual metrics would combine to show value to users in a scenario and the multiple scenarios would combine to show value for the system.

[1] John Barton and Jeff Pierce, “Quantifying Magic In Ubicomp Systems Scenarios”, UbiSys 2006 Workshop.

[2] Mark Weiser, "The Computer for the Twenty-First Century," *Scientific American,* pp. 94-10, September 1991

[3] T. Kindberg and A. Fox. System Software For Ubiquitous Computing. IEEE Pervasive Computing Magazine, 1(1):70--81, January 2002.

[4] David Garlan, Daniel P. Siewiorek, Asim Smailagic, and Peter Steenkiste, “Project Aura: Toward Distraction-Free Pervasive Computing” IEEE Pervasive Computing, Apr, 2003

[5] W. Keith Edwards and Rebecca E. Grinter , “At Home with Ubiquitous Computing: Seven Challenges” Ubicomp 2001, Atlanta, GA

[6] J. M. Carroll “Five reasons for scenario-based design” Interacting with Computers Volume 13, Issue 1 , and “Scenario-based design: envisioning work and technology in system development” 1995

[7] Jakob E. Bardram Scenario-Based Design of Cooperative Systems – Re-designing an Hospital Information System in Denmark.. Group Decision and Negotiation, 9: p. 237-250, 2000.

[8] Kazman, R. Abowd, G. Bass, L. Clements, P Scenario-based analysis of software architecture. Waterloo Univ., Ont.; Software, IEEE Nov 1996 Volume: 13, Issue: 6

[9] WK Edwards, V Bellotti, AK Dey, MW Newman “The challenges of user-centered design and evaluation for infrastructure” - Proceedings of the conference on Human factors in computing

[10] Jakob E. Bardram , “The Trouble with Login – User Authentication and Medical Cooperation” Personal and Ubiquitous Computing, 2005.

[11] John J. Barton Fopefolu Folowosele Stina Nylander Mave Houston Shumin Zhai “Connection Time for Strange Devices” 7th IEEE Workshop on Mobile Computing Systems and Applications April 6-7, 2006