Innovative Technology for Computer Professionals

Computer

MAY 2011

MOBILE COMPUTING

http://www.computer.org

IEEE

IEEE computer society

Computer MAY 2011 Improving Web Accessibility//Mobile Computing//Engineering Robotics Systems Volume 44 Number 5

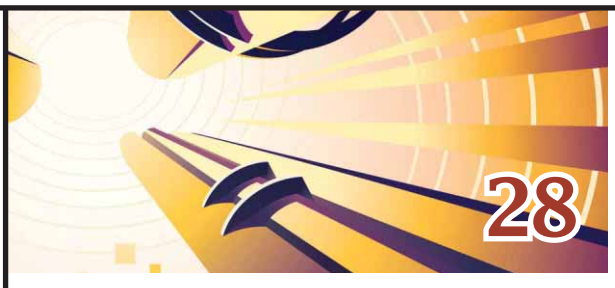
Innovative Technology for Computer Professionals
Computer

Circulation: *Computer* (ISSN 0018-9162) is published monthly by the IEEE Computer Society. **IEEE Headquarters**, Three Park Avenue, 17th Floor, New York, NY 10016-5997; **IEEE Computer Society Publications Office**, 10662 Los Vaqueros Circle, Los Alamitos, CA 90720-1314; voice +1 714 821 8380; fax +1 714 821 4010; **IEEE Computer Society Headquarters**, 2001 L Street NW, Suite 700, Washington, DC 20036. IEEE Computer Society membership includes $19 for a subscription to *Computer* magazine. Nonmember subscription rate available upon request. Single-copy prices: members $20.00; nonmembers $99.00.

Postmaster: Send undelivered copies and address changes to *Computer,* IEEE Membership Processing Dept., 445 Hoes Lane, Piscataway, NJ 08855. Periodicals Postage Paid at New York, New York, and at additional mailing offices. Canadian GST #125634188. Canada Post Corporation (Canadian distribution) publications mail agreement number 40013885. Return undeliverable Canadian addresses to PO Box 122, Niagara Falls, ON L2E 6S8 Canada. Printed in USA.

Innovative Technology for Computer Professionals
Computer
www.computer.org/computer

CONTENTS



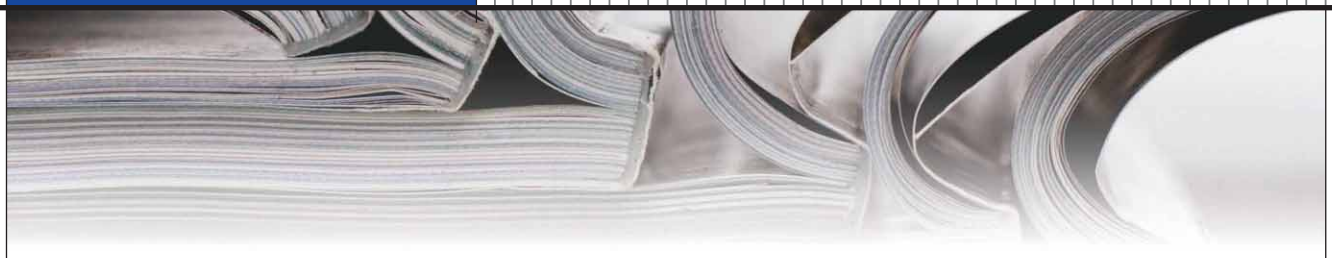
ABOUT THIS ISSUE

IEEE Computer Society: http://computer.org
Computer: http://computer.org/computer
computer@computer.org
IEEE Computer Society Publications Office: +1 714 821 8380

Flagship Publication of the IEEE Computer Society

May 2011, Volume 44, Number 5

ELSEWHERE IN THE CS



Computer Highlights Society Magazines

The IEEE Computer Society offers a lineup of 13 peer-reviewed technical magazines that cover cutting-edge topics in computing including scientific applications, design and test, security, Internet computing, machine intelligence, digital graphics, and computer history. Select articles from recent issues of Computer Society magazines are highlighted below.

IEEE Software

Component-based software engineering poses new challenges for predicting software performance, but it also offers several advantages. In "Facilitating Performance Predictions Using Software Components" in the May/June issue of *Software*, authors Jens Happe, Heiko Koziolek, and Ralf Reussner discuss a simplified system implementation that demonstrates compositional reasoning about software performance. Developers can download and use the open source tool Palladio.

IEEE Intelligent Systems

In "Improving Users' Mental Models of Intelligent Software Tools" in the March/April issue of *Intelligent Systems*, authors Shane T. Mueller and Gary Klein describe the Experiential User Guide, which is designed to address the genuine cognitive challenges users have with the complex, intelligent software tools used in everyday life—commercial navigation systems and Web search algorithms are just two examples. The guide exposes learners to many of the experiences that an expert will have had over time, allowing both novice and expert users to experience a tool's strengths and weaknesses.

IEEE Computer Graphics and Applications

Current colorization based on image segmentation makes it difficult to add or update color reliably and requires considerable user intervention. Authors from the Chinese University of Hong Kong and the Chinese Academy of Sciences describe a new approach that gives similar colors to pixels with similar texture features. The method uses rotation-invariant Gabor filter banks and applies optimization in the feature space. Read more in "Colorization Using the Rotation-Invariant Feature Space" in the March/April issue of *CG&A*.

Computing in Science & Engineering

Supercomputers are usually designed to achieve the highest possible performance. Their architectures have evolved from early custom-design systems to today's clusters of commodity multisocket, multicore systems. Twice a year, the supercomputing community ranks these systems (using the number of 64-bit floating-point operations per second) and produces the Top500 list (www.top500.org) showing the world's 500 highest-performing machines.

In "Trends in High-Performance Computing" in the May/June 2011 issue of *CiSE*, authors Volodymyr Kindratenko and Pedro Trancoso describe the technologies used in today's top-ranked machines and forecast next-generation supercomputing architecture trends.

IEEE Security & Privacy

Data leaks are often the result of usability failures. In healthcare, usability failures risk both patients' health and their identity. In "Usability Failures and Healthcare Data Hemorrhages," in the March/April issue of *S&P*, authors M. Eric Johnson and Nicholas D. Willey of Dartmouth College analyze samples of medical-related files collected from peer-to-peer file-sharing networks. These leaked files contained significant protected health information and demonstrate the risk to patients and institutions. Through interviews and field research, Johnson and Willey document how usability failures lead to such hemorrhages.

IEEE Pervasive Computing

Reducing domestic energy consumption is a hot topic in the pervasive computing and computer-human interaction communities, with a long and varied history. In "Look Back

Published by the IEEE Computer Society

0018-9162/11/$26.00 © 2011 IEEE

before Leaping Forward: Four Decades of Domestic Energy Inquiry," authors Mike Hazas, Adrian Friday, and James Scott give a brief overview of the history and current state of this problem from various perspectives—from pervasive computing feedback-oriented and technology-centric systems, to sociology- and economics-based studies. The article provides an introductory set of references to aid readers in exploring this topic's rich background in depth.

IEEE Internet Computing

The measurement, characterization, and modeling of real workloads are key steps driving the design of cost-effective Internet applications and services. However, some of the most fundamental concepts and methods for characterizing Internet workloads are largely unknown to many Internet and Web practitioners.

Internet Computing's March/April special issue on Internet workloads presents articles that characterize and model different workload types for grid computing, HTTP forward caching, and multimedia services. The articles cover popular applications and draw insights useful for system planning, management, and optimization.

IEEE micro

On-chip networks could become a critical shared resource in many-core systems. If so, devising efficient and fair scheduling strategies is a particularly important and challenging performance component. In "Aérgia: A Network-on-Chip Exploiting Packet Latency Slack" in the January/February issue of *Micro*, researchers from Pennsylvania State University, Carnegie Mellon University, and Microsoft Research describe a novel NoC architecture that uses router prioritization mechanisms based on packet slack—a measure of the lag possible in relation to the effect of its delivery on application performance.

IEEE MultiMedia

"While nearly everyone can agree that pecking out a love note on a tiny mobile phone keypad while simultaneously trying to operate a vehicle is a bad idea, what about the other activities that we perform on a day-to-day basis using the electronic devices either built into or brought into our cars?"

Authors Christian Müller of the German Research Institute for Artificial Intelligence and Garrett Weinberg of Mitsubishi Electric Research Labs start with this question in "Multimodal Input in the Car, Today and Tomorrow," in the January/March issue of *MultiMedia*. They give a brief overview of multimodal theory as it pertains to common in-vehicle tasks and devices before describing the state of the art and suggesting ways to safely broaden in-vehicle system capabilities in the future.

IT Professional

TECHNOLOGY SOLUTIONS FOR THE ENTERPRISE

RFID technologies have been revolutionizing the way we perform asset tracking for more than 30 years, and RFID applications are increasingly appearing in areas such as transportation, banking, healthcare, and security—primarily owing to increased reliability, widespread adoption of international standards, and decreased costs. A passive ultrahigh-frequency tag is expected to cost just a few cents this year, prompting continued RFID growth. In its March/April 2011 special issue, *IT Pro* presents five articles that report recent experiences in developing applications, middleware, and security protocols for applying RFID in real-time systems.

IEEE Design&Test of Computers

Articles in the March/April issue of *Design & Test* highlight a range of problems that IC and system designers face in ensuring high performance under tight power constraints and escalating verification and test costs. In "Customizable Domain-Specific Computing," researchers from the University of California and Rice University show how to learn from nature—the human brain—to design a platform that achieves its efficiencies through customization.

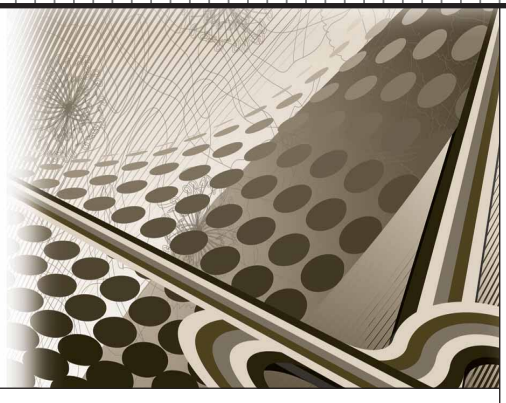
IEEE Annals of the History of Computing

"Kissinger's Computer: National Security Council Computerization, 1969-1972," a feature article in the January-March 2011 issue of *Annals*, recounts the history of Henry Kissinger's information-automation project during the Nixon administration. The project introduced computers into the White House and standardized information management within the US National Security Council.

"Beyond the NSC's day-to-day activities, the information-automation projects had a broader impact on US national security policy …," writes author John Laprise of Northwestern University. "By the end of the Nixon administration, the NSC saw computers as a primary information-management technology that could be applied to the economic spectrum and needed to be controlled to limit the economic strength of communist nations."

Not for All Markets

David Alan Grier, *George Washington University*



Crowdsourcing is one of the more intriguing forms of computation that employs markets.

At times, especially following a major crisis in the nuclear industry or an attack on the global cyberinfrastructure, our office becomes a little media market. Reporters contact us seeking a comment that's provocative, insightful, or counterintuitive, depending upon their tastes. In exchange, they offer a brief moment of fame, a boost to the ego with a personal appearance on Great Day Pocatello! or a quote in an obituary for someone who might once have been famous.

My colleagues and I generally try to approach these transactions rationally, attempting to provide informative answers for educational and policy purposes. However, when the choice is difficult, especially when we're facing deadlines or working on other tasks, we throw the request onto the trading floor and ask if this question is truly worth our time and effort.

QUANTIFYING INFORMATION

During the past couple of weeks, I've been spared requests from the media as the scientific controversy of the moment has nothing to do with computing or information technology. The reporters, tiring of the crisis at the Japanese nuclear power generator, have turned their attention to the more recent announcement by NASA that it would place the retired space shuttles at museums in Washington, D.C., New York, Florida, and California. Somehow, this is a controversy that needs the comments of a scientific policy scholar.

At the bargain price of $28 million plus change, the old shuttle orbiters generated far more demand than the supply on hand. In all, 29 organizations responded to NASA's request for information in the hopes of securing one of the relics. When the announcement was made, 25 of these groups were disappointed, and more than a few of them were willing to express their feelings in an effort to change the decision. For example, organizations in Texas and Ohio provoked their congressional delegations to denounce the "blatant politics" behind the decisions and claim that "NASA ignored the intent of Congress and the interests of taxpayers."

The public evidence suggests that NASA, a giant engineering organization with a well-formed understanding of national politics, approached this decision in a rational, engineered manner. Not only is NASA inclined to trust numerical approaches, its representatives understand that quantification can be used to deflect criticism. "The appeal of numbers," wrote the philosopher Ted Porter, "is especially compelling to bureaucratic officials who lack the mandate of a popular election or divine right." He argued that a decision made by numbers "has at least the appearance of being fair and impersonal."

NASA actually has a 15-page policy document defining space artifacts and describing how they may be loaned to various types of organizations for "public affairs, industrial outreach, and education programs." Hence, it requested a fair amount of information from the organizations that were interested in acquiring a shuttle orbiter.

Much of this information was already in numeric form that directly related to the requesting organization's wealth and popularity. These statistics included attendance figures for the past five years, the budget for the same period, endowments, the local region's population, and the number of hits on the organization's public webpages. Other information could easily be quantified using social science techniques, such as the additional assets that the organization could use or its experience with "artifacts of national significance." Even the most subjective question, which asked how the organization would use the shuttle "to inspire the American public and students in particular," also required the description of a concrete procedure that could be used to "assess, evaluate, and measure these objectives."

With proposals providing quantifiable information, the NASA board could have easily slipped into a familiar procedure. This would have

 Published by the IEEE Computer Society 0018-9162/11/$26.00 © 2011 IEEE

involved creating a matrix with 29 rows—one for each museum—and some appropriate number of columns. The board would have converted much of the information into a five-, seven-, or nine-point scale with the low value representing a poor contribution, the middle value being neutral, and the high value expressing a substantial asset. A quick summation of figures would have identified the proposals with the highest numeric score. Then the board could have determined if this result could be presented to the public as a final judgment.

Such a process, though rationally disciplined, is less connected to the scientific method than it might appear. It's actually part of a market process, in which the parties lay all of their assets on the table and ask if they have sufficient wealth to complete a trade.

Markets include not only quantification but valuation, the act of ranking the importance of goods, services, and ideas. Valuation is often far from scientific. It "is vague and arbitrary," wrote the 19th century economist Jean-Baptiste Say, "when there is no assurance that it will be generally acquiesced in by others." It can also be quite dynamic, as the desires and aspirations of those in the market change.

The world of computation has had an awkward relationship with markets. For computable answers, we rely on logic, mathematics, and probability. In his 1854 master work, *An Investigation of the Laws of Thought*, George Boole argued that "the "mathematical sciences occupy a fundamental place in human knowledge, and that no system of mental culture can be complete or fundamental, which altogether neglects them." We use these tools to participate in, model, and predict the opportunities in markets, but we rarely allow markets to give us an answer to our computational questions beyond the current value of the products we create.

CROWDSOURCING

In the past four or five years, markets have begun to creep into computational processes as a way of answering difficult questions, questions that deal with the efficient use of resources. One of the more intriguing forms of computation that employs markets has come to be known as *crowdsourcing*.

All crowdsourcing applications have four basic elements: a division of labor, computing and communications technology, a crowd of human workers, and a labor market.

Crowdsourcing is generally understood to be a way of using the Internet to employ large numbers of dispersed workers. As an example of a crowdsourcing application, an employer can post a task on a website and invite workers to undertake all or part of it. Sometimes the task requires substantial skill and training, such as a programming assignment or resolving a legal problem. In other cases, it requires only basic human perception, such as the ability to identify faces in a photograph and judge the emotion such faces are expressing.

In general, crowdsourcing is an industry that's attempting to use human beings and machines in large production systems. These systems prove to be useful and efficient when each element does the work it can do best: machines carry out repetitive tasks while people make judgments, handle pattern recognition, and synthesize ideas. Perhaps the most familiar version of crowdsourcing is Wikipedia, which relies on tens of thousands of volunteers to construct an online encyclopedia. Another common example is Amazon's Mechanical Turk, which offers human intelligence tasks or HITs.

At this point, the list of firms interested in this field is long and growing rapidly. It includes companies that handle customer relations, create metadata for photographs and other nontextual items, and gather data on product placement and consumption.

All crowdsourcing applications have four basic elements: a division of labor, computing and communications technology, a crowd of human workers, and a labor market. Sometimes, the market is used merely to recruit workers to the project. In Wikipedia, for example, the market is implicit in the basic wiki interface. Workers bring to that market the material they've written on specific subjects. The transaction is completed when the work is accepted for publication. The worker is compensated by the privilege of having the work posted online, at least temporarily.

A MARKET THEORY OF TRANSLATION SERVICES

In other cases, the labor market is used to extract ideas from the workers as a whole. Recently, I was talking with Kwami, the owner of a firm that offers crowdsourced translation services. At base, his organization allows companies to put documents that they wish to have translated on the Web and pays workers to translate them. His system has a few additional tools that enhance the quality of the work. It provides an initial mechanical translation, offers style rules, and saves translated text for possible reuse. It also allows the employer to gather useful information from the labor market. "The value depends on who's using the final translations," Kwami said. "You get the greatest value when you're translating documents to be used in the field."

Kwami has two types of clients: one type gathers information, the other disperses it. In the first case, the organization draws its

THE KNOWN WORLD

translators from those who gather the information. In the second, it can find translators among the people who actually use the information. "The crowd is most useful in the second case," Kwami explained, as these translators help the employers understand the value of the translated documents.

The second kind of client is a class of firms that sell large, complicated products that come with lots of documentation. Traditionally, the firm would create all of its documents in English and then pay for the translation of a few key documents, such as an overview of the system or an index of key documents. Paying to have all of the documents translated into all possible languages would likely create a lot of text that would remain unused.

In Kwami's system, the translators are drawn from the people who actually work with the product. Generally, they're bilingual customers who are supplementing their income. They're presented with all the base language documents and are allowed to choose the documents they want to translate. According to the theory of markets, the crowd will choose to translate the documents that are of most value to them and, by extension, to those who are supporting, maintaining, and extending the product.

Of course, this market theory of translations has drawbacks that counter the potential advantages. All markets tend to have a notoriously narrow vision. Participants value those things that will help them today rather than something that might be useful next year, such as that rare document containing information that will only be needed in the case of a massive failure. While the defenders of market-based decisions will support these choices by arguing that the market makes the most efficient choices, not one of them desires to face an acute crisis with a procedure manual that needs to be translated from another tongue so that it can be used properly.

So, like all markets, we modify the rules of crowd labor markets to protect ourselves against risk. Some of these modifications can be handled through the pricing mechanism of the market itself. If the organization needs to have certain documents translated first, it can offer premium prices for that work.

However, the standard market mechanisms won't solve all the weaknesses of the labor markets in crowdsourcing. Critics have noted that the dispersed labor pool in crowdsourcing leaves plenty of opportunities for workers to make mistakes, cheat, plagiarize, and even get paid for vacuous submissions. Most companies engaged in crowdsourcing use a complex set of rules for their labor markets to minimize potential problems. "Each piece of data is independently checked by multiple crowd workers," explains the leader of one company. His organization also rates each worker by seeding the market with tasks for which it already knows the result. "Ineffective workers are filtered out," he adds.

Of course, free markets are never truly free. In many spheres of modern life, we've structured markets to minimize both risks and transaction costs: supermarkets, online auctions, even the little media mart that forms in our office whenever science and technology are in the news. The reporters who come to us know that we're unlikely to feed them a quote supplying wrong or embarrassing information. They've restricted their crowd of workers to a mere handful of policy wonks, a handful that they believe will provide them with a good product. C

David Alan Grier *is a senior member of IEEE and an associate professor of International Science and Technology Policy in the Elliott School of International Affairs at George Washington University. Contact him at grier@computer.org.*

32 & 16 YEARS AGO

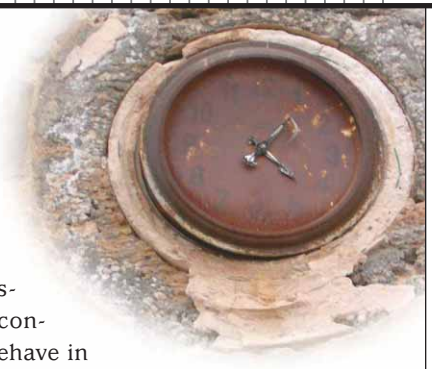
MAY 1979

SPECIAL MESSAGE (p. 4) "*Computer*—a journal which stands between 'transactions' on the one hand and 'trade journals' on the other—is a mechanism for providing technical information. One of its major objectives is to present practical, technically based articles on development projects or programs. Independent submission of such articles is encouraged, but the number received is disappointing. Thus, to present a wide spectrum of recent developments, *Computer* largely relies upon invited papers. This solicitation brings out articles which are not marketing or sales motivated and, consequently, has resulted in a very satisfactory 'technical' content for *Computer*."

STORAGE HIERARCHIES (p. 18) "Sequential access is a major component in accesses to secondary storage devices such as disks. The extension of hierarchy analysis to include secondary storage is important because of the present industry trend toward including disks in virtual memory systems with automatic data transfer procedures. Also, the availability of potential 'gap-filler' technologies has raised interest in cache-like structures for disks. While no longer the predominant mode of secondary-storage access, sequential access is still a major storage-usage component, occurring in scientific processing, word processing, data-flow buffering, and data base activities such as searching, periodic updating and report generating. …"

DATABASE TRENDS (p. 27) "The current status of data base systems can be characterized by two significant trends: rapidly increasing user acceptance, and rapidly improving technology at the logical, physical, and architectural levels. These two trends are closely coupled. Increased user acceptance has come about, in part, because the price of direct access mass storage has fallen by a factor of ten in the last eight years and is currently falling by a factor of two every 30–36 months. The availability of low-cost direct access mass storage has encouraged users to move away from tape-oriented sequential files into structures that support direct access efficiently. …"

A RELATIONAL DATABASE (p. 42) "Perhaps the greatest impediments to the use of a computerized data base management system are installation cost and complexity. At present, installation of these systems requires a staff skilled in telecommunications, operating systems, data management and in applications. In response, our lab designed and implemented System R—an experimental data base management system allowing easy definition of data bases and data base applications without sacrificing the function and performance available in most commercial systems. … System R adopts a relational data model and supports a language called SQL for defining, accessing, and modifying various views of the data base."



CONCURRENT PROGRAMMING (p. 50) "The slightest programming mistake can make a concurrent program behave in an irreproducible, erratic manner that makes program testing impossible. The following describes how this problem was gradually solved by software engineers and computer scientists. This development is seen as an initial hardware challenge followed by a software crisis, a conceptual innovation, and language development which in turn led to formal understanding and hardware refinement. …"

MICROCOMPUTER APPLICATIONS IN JAPAN (pp. 73-74) "Our survey confirms the explosive expansion of microcomputer applications. In Japan, such applications have appeared in a variety of industries, particularly in the process control and consumer electronics areas. Japanese microprocessors are designed to serve as 'onboard intelligence'—specialized hardware in another machine or device not recognizably a computer. As a result, products such as home computers have been slow to appear. …"

MAGNETIC BUBBLES (p. 89) "By the simple tactic of using two distinct types of magnetic bubbles, IBM scientists have increased the information storage density of a conventionally designed magnetic bubble device by a factor of four.

"A team headed by Dr. Ta-Lin Hsu of IBM's San Jose Research Laboratory has fabricated and tested a fully operational experimental device capable of storing 15,000 bubbles at a density of 3.3 million bubbles per square centimeter or 22 million bubbles per square inch. …"

TEACHING DEAF CHILDREN (p. 90) "Computer technology and color are being combined in an experimental learning program which teaches the basics of reading, writing, and other language skills to deaf pre-school and kindergarten-age children."

"The program requires a branching-type educational sequence that includes teaching, assessment, and remediation. The child initially progresses through very simple teaching techniques. As long as no error is committed, a linear progression takes place through remaining sequences. If the child makes a mistake, the computer directs the learner to remedial instruction."

PDFs of the articles and departments from Computer*'s May 1979 and 1995 issues are available through the IEEE Computer Society's website: www.computer.org/computer.*

32 & 16 YEARS AGO

MAY 1995

TELEPHONE COMPANIES (p. 8) "Remember what happened to centralized MIS departments during the PC revolution of the 1980s? Well, the offspring of the Philistines responsible for that are now at the gates of the centralized-PBX manufacturers and soon will be storming the walls of major telecommunications providers. If you're a Microsoftian (rhymes with Martian), think TAPI (telephony API). If you live in a Novellian (rhymes with Orwellian) world, the watchword is TSAPI. But if you're one of those rare individuals who think a standard should be made by a standards group instead of a market-muscle company, call it the dash to SCSA (signal computing system architecture). We call it CTI (computer telephone integration). Regardless of name, the movement toward desktop telephony is afoot, and desktop telephony will do to the telephone equipment manufacturers what Gorbachev did to the FSU (former Soviet Union)."

NETIQUETTE (p. 12) "As more companies enter cyberspace, concern that employees may misbehave on line has prompted many firms to draft guidelines. Among those drafting rules for 'netiquette' (Internet behavior) are Chase Manhattan Bank of New York and Johnson Controls of Milwaukee, Wisconsin."

INTERNET MARKETERS (p. 13) "Although the Internet was originally established for scientific purposes, technical departments have been taking a backseat to sales departments when it comes to on-line interest. A recent US report states that marketing and communication departments are driving Internet gateway projects in user companies more than the IS departments are."

INTERACTIVE TELEVISION (p. 25) "Cable TV and telephone companies, as infrastructure and content providers, are realizing that potential network technologies, coupled with improved computing and compression techniques, will soon profitably deliver interactive services. Recently, entertainment, cable, phone, and computer companies have formed alliances to design a variety of wide-area multimedia infrastructures. Consequently, universities and industrial laboratories have been working intensively to define a suitable architecture, and a number of groups have developed applicable standards. From these activities, we can see the emerging information infrastructure."

MULTIMEDIA SERVERS (p. 40) "Recent advances in computing and communication make on-line access to multimedia information—like books, periodicals, images, video clips, and scientific data—both possible and cost-effective. The architecture for these services consists of multimedia storage servers connected to client sites via high-speed networks … Clients can retrieve multimedia objects from the server for real-time playback. Furthermore, access is interactive because clients can stop, pause, and resume playback and, in some cases, perform fast-forward and rewind operations."

MULTIMEDIA LEARNING (p. 74) "… multimedia, when coupled with other improvements in educational software design, can support effective, quality instruction. We propose an instruction model that moves beyond the 'Tyranny of the Button' and uses intelligent simulation, dynamic links (on-line generation of links based on student behavior), and multimedia composition and creation. …"

INVENTIONS (p. 83) "US patent laws specify that, to remain free from the public domain, two main criteria must be satisfied. First, the invention must not be disclosed to the public and must be kept confidential. Second, the invention cannot be sold, offered for sale, or commercially used in an unrestricted manner."

INTERNET ANARCHY AND CHAOS (p. 87) "… We need to apply our technical expertise to the design of realistic alternative means of running and managing cyberspace. We also should be helping to determine how those alternatives can be sensibly merged with the many legitimate legislative issues—and to identify the appropriate role for the Net's 'business newcomers who now want access to its huge audience,' …"

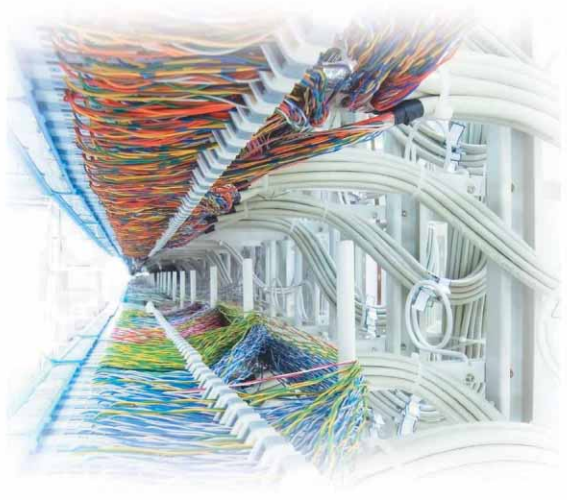
SOFTWARE ENGINEERING (p. 98) "Is the term 'software engineering' a misnomer? That question has long been debated within the computer science, programming, and software engineering community. Naysayers point to the software activity's large trial-and-error component and its notable lack of solid intellectual and ethical underpinnings. On the affirmative side, ACM and the IEEE Computer Society recently joined forces to move software engineering toward professional status."

VLSI TEST AUTOMATION (p. 120) "It's not uncommon to find that a third of the wonderful test vectors cranked out laboriously by designers—or tracked down by smart algorithms—have to be masked out, even though the vectors provide 100 percent single stuck-at-fault coverage for VLSI testing. Either they aren't compact enough; occupy too much pattern memory; don't address bidirectional and tristate conflicts; or overflow scan-data, format-data, or timing-data memory. Worse yet, in many technologies, fault coverage doesn't correlate with defect coverage. Clearly, test automation is a long way from being as usable as front-end tools like synthesis, timing analysis, or simulation."

Editor: Neville Holmes; neville.holmes@utas.edu.au

TECHNOLOGY NEWS

Forecasting in Geologic Time: The Tohoku Quake and Tsunami



Linda World

Japan is a world leader in earthquake and tsunami early-warning systems, but the recent unforeseen magnitude 9.0 earthquake triggered a tsunami that caused many coastal hazard-mitigation structures to fail.

No country is more prepared for natural disasters than Japan. From nationwide emergency-training days to substantial R&D investments in forecast and mitigation technologies, Japan makes a priority of preparedness for earthquakes, volcanoes, tsunamis, and typhoons.

So when the historic 9.0-magnitude Tohoku earthquake ruptured about 60 miles off Honshu island's northeast coast on 11 March 2011, sensors in the world's densest seismic data network, shown in Figure 1, immediately detected and measured its primary (P) wave—the first and fastest of two seismic waves that earthquakes generate. The network forwarded this information to the Japan Meteorological Agency (JMA), which took 8.6 seconds to issue an earthquake alert over the world's first nationwide earthquake early-warning system. After further calculation, it took less than 3 minutes for JMA to issue a tsunami alert.

The slightly slower speed of a quake's secondary (S) waves opens a time gap that earthquake early-warning systems exploit. The lag time is small: P waves typically travel at 3 miles per second through granite, and S waves are about 60 percent slower. However, this gave Tokyo—230 miles south of the epicenter—about 50 seconds to prepare for the shaking that comes with S waves, which are responsible for most of the structural damage and subsequent earthquake deaths.

Although the warning systems worked, the Tohoku quake was 10 times larger than JMA had forecasted for the area, and the damage it caused was significant—estimates range up to US$309 billion. But the tsunami that hit the coastline 30 minutes later was responsible for the staggering loss of life and an ongoing crisis at the Fukushima Daichi nuclear plant.

TSUNAMI WARNING SYSTEMS

Tsunami waves travel much slower than seismic waves—roughly 500 miles per hour in the open ocean. This gives more time to issue warnings, but the waves carry tremendous energy. In this case, the energy originated from water displaced when the fault line between colliding tectonic plates slipped. A slip releases the strain that builds up as one plate moves beneath another. In a special session of the Seismological Society of America's annual meeting in April 2011, California Institute of Technology seismologist Hiroo Kanamori reported that the Tohoku slip ranged between 83 and 164 feet.

That's huge, and the tsunami it generated overwhelmed the preparations of an area familiar with the risk and defended against it with 30-feet-high tsunami seawalls and breakwaters. According to JMA, some waves reached as high as 124 feet in one area, destroying structures engineered to protect shoreline communities. Entire villages were washed away, and more than 27,000 people died.

At the Fukushima Daichi nuclear plant, all three active reactors had shut down with the first earthquake alert, in accordance with the early-

0018-9162/11/$26.00 © 2011 IEEE Published by the IEEE Computer Society

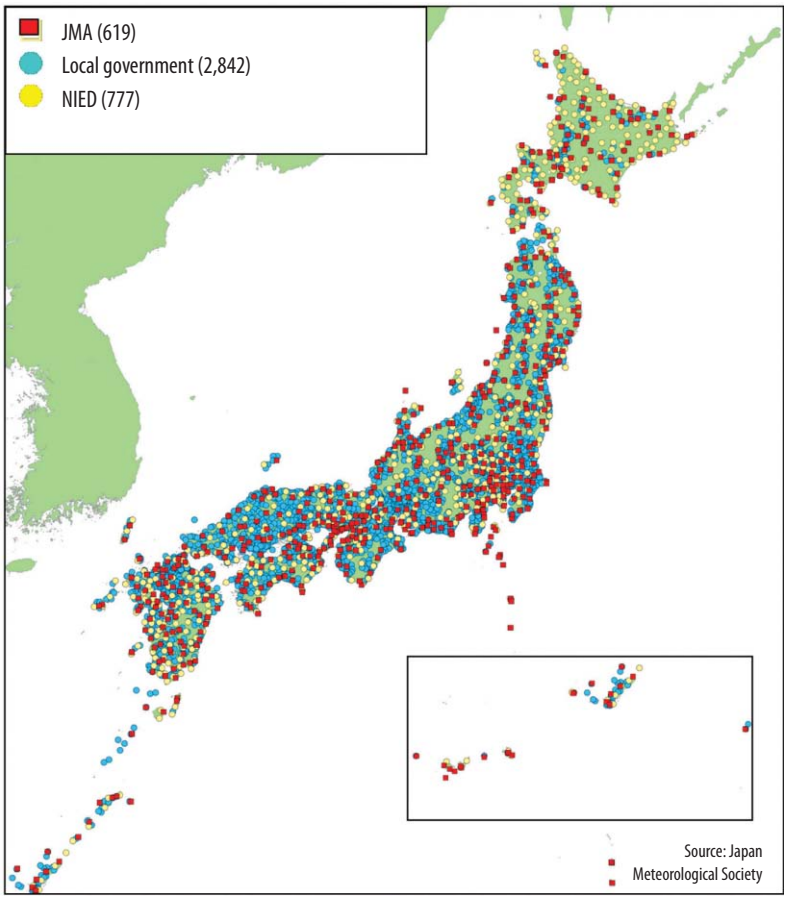


Figure 1. Japan's seismic data network. The Japan Meteorological Agency operates an earthquake observation network comprising about 200 seismographs and 600 seismic intensity meters. It also collects data from over 3,600 seismic intensity meters managed by local governments and the National Research Institute for Earth Science and Disaster Prevention (NIED).

warning system. However, when the tsunami crested at 46 feet, it breached first the 18-foot breakwater and then the 13-foot cliff on which the plant's six reactors were built. The resultant flooding of the generators that power the backup reactor-cooling systems set off the unfolding nuclear crisis.

Mitigation measures

The Tohoku area was still experiencing significant aftershocks more than a month after the initial quake, and more than 100,000 people remain homeless. With the situation at Fukushima still dangerous and the country using rolling blackouts to handle power shortages, it's impossible to fully assess the failures, let alone the successes, of the various investments Japan made in systems to mitigate the hazards of a major earthquake. But the seawalls haven't fared well: many of these structures suffered severe damage or failed, including a state-of-the-art, $1.5 billion breakwater off the coast of Iwate, completed just three years ago.

In reference to the Kamaishi seawall, Patrick Lynett, associate professor in Texas A&M's Department of Civil Engineering, said, "There are currently teams of engineers looking into this and the many other failures of coastal structures in Japan. It may require engineers to revisit the methods used to design structures against the largest tsunamis.

"That said," he continued, "the tsunami that occurred in the local earthquake region was truly enormous and catastrophic, and it would be difficult to completely protect any community from its damaging effects."

Tsunamis in open water

Tsunamis occur more frequently in the Pacific Ocean than anywhere else because of the many large earthquakes that occur along its margin, sometimes called "the Ring of Fire." In 1968, an international cooperative of 26 member nations formed the Pacific Tsunami Warning System (PTWS) to coordinate activities that would improve early-warning systems in the region. The Pacific Tsunami Warning Center (PTWC), which has operated in Hawaii since 1949, is the organization's operational headquarters.

The PTWS uses seismic stations and tide gauges as the first phase of detecting and locating major earthquakes and determining whether they've generated tsunamis. Since 2001, it has refined its detection and monitoring capabilities with the deep-ocean assessment and reporting of tsunamis (DART) system developed by the US National Oceanic and Atmospheric Administration. A sea-level gauge used for real-time data collection, DART consists of a bottom-pressure recorder (BPR) anchored to the ocean floor and a surface buoy that holds the electronics.

The BPR monitors water pressure with a resolution of approximately 1 mm in 15-second averaged samples and transmits the data to the buoy via an acoustic modem. The pressure measurement translates to a measure of the distance to the water's surface, and the 1-mm resolution is necessary because tsunamis cross the deep ocean in shallow but long waves. The length from crest to crest can be 124 miles, while the height from trough to crest might be only a few inches.

The BPR runs an algorithm that generates predicted water height values and compares all new samples with predicted values. If two 15-second water-level values exceed the predicted values, the system enters a tsunami response mode, transmitting data in short intervals to the buoy. From the buoy, the data goes to a satellite that communicates with a control station.

Over the past five years, modeling has also become a primary forecasting tool, according to Brian Shiro, a PTWC geophysicist. "PTWC can now estimate expected tsunami wave heights at coastlines with great enough accuracy to affect decisions on whether certain countries should be in a warning status or not," said Shiro. "Our model predictions for the tsunami that resulted from the Tohoku earthquake were very accurate throughout the Pacific."

This accuracy gave communities across the Pacific Basin several hours' warning of the Tohoku tsunami's arrival time and height.

FORECASTING, NOT PREDICTING

"Over the past 10 years, the death toll from earthquakes and tsunamis approaches 1 million persons worldwide." said John Rundle, professor of physics and geology at the University of California, Davis, and a founding member of the Open Hazards Group (www.openhazards.com). The group maintains a website that offers free and commercial tools for analyzing earthquake risk at a personal level. For example, one tool lets you enter your address and get a probability forecast of an earthquake greater than magnitude 5.0 occurring there within the next year. By offering tools like this, Open Hazards wants to help individuals make informed judgments about the risks they face and begin building "disaster-resilient communities."

Rundle's research includes modeling faults and simulating earthquakes on them. These simulations offer insights into the complex, dynamic behavior manifest in earthquakes—insights that can advance the ability to forecast their real-world behavior. Rundle is careful, however, to distinguish forecasting from prediction: "Prediction is saying, 'The next toss of the coin will come up heads.' Forecasting is saying, 'The probability of the next toss coming up heads is 50 percent.' We calculate probabilities quite accurately."

Ricardo Taborda, a postdoctoral associate in the Computational Physics Lab of Carnegie Mellon's Department of Civil and Environmental Engineering, is working on physics-based simulations of large-scale earthquakes. He also backs away from prediction as a goal. "Mother Nature will continue to have the edge and the potential to surpass our best estimates of what to expect. The recent earthquake in Japan is the proof of that.

"I personally do not think we can anytime soon get closer to fully characterizing what happens at the contact between two tectonic plates, though we can get better at modeling it—something that's not at all trivial. We can also continue to improve the techniques we have—such as physics-based simulations—to answer the question of what would happen if an earthquake of a given magnitude occurs and prepare for that."

Japan has more than 1,500 earthquakes each year and the most recorded tsunamis in the world—16 since 1901. It also has 10 percent of the world's active volcanoes and a typhoon season that runs from May to October. However, earthquakes and tsunamis differ from volcanoes and typhoons in that they give no clear signs that they're about to strike. In some ways, the warning systems for these recurring natural disasters might more accurately be called conterminous- rather than early-warning systems. Whatever you call them, Japan is internationally recognized as a leader in implementing these systems both physically and socially.

The structural engineering and other hazard-mitigation strategies, such as zoning, are global problems. In the years ahead, the data already being gathered from the Tohoku natural disaster will inform models and research that could lead to warning systems that actually precede the event—with a very high degree of probability.

Linda World is a senior editor at the IEEE Computer Society. Contact her at lworld@computer.org.

Editor: Lee Garber, *Computer*; l.garber@computer.org

SPECIAL ISSUE ON EARTHQUAKE ENGINEERING SIMULATION

The upcoming July/August 2011 issue of *Computing in Science & Engineering* features five articles on the state of the art in earthquake engineering simulation:

- Developing an Effective Cyber-infrastructure for Earthquake Engineering: The NEEShub
- Rupture-to-Rafters Simulations: Cyber-Enabled Unification of Science and Engineering for Earthquake Hazard Mitigation
- Numerical Simulation of Complex Tsunami Behavior
- Large-Scale Earthquake Simulation: Computational Seismology and Complex Engineering Systems
- Petascale Computation for Earthquake Engineering

Patrick Lynett, quoted in this technology news story, is coauthor of "Numerical Simulation of Complex Tsunami Behavior." Ricardo Taborda is coauthor "Large- Scale Earthquake Simulation."

http://www.computer.org/cise

NEWS BRIEFS

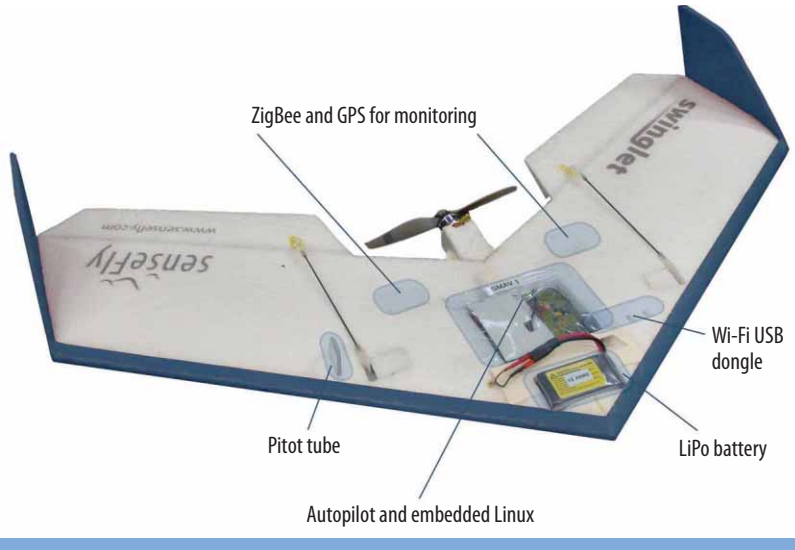
Flying Robots Designed to Form Emergency Network

European academic researchers have developed a constellation of robust, lightweight flying robots using wireless communications that could be employed in mapping, remote sensing, ground searches, and other similar operations.

The scientists are developing technology for a Swarming Micro Air Vehicle Network, said Jean-Christophe Zufferey, group leader for flying robotics research at Ecole Polytechnic Federale de Lausanne's Laboratory of Intelligent Systems. A SMAVNET will consist of a set of robots that can fly at low altitudes, communicate wirelessly with one another, and coordinate their activities. They then form a communications network that could be used by, for example, rescuers in a remote area.

The robots consist of a flying-wing airframe, with neither fuselage nor tail. They are propelled by a single electric motor running on a battery capable of 30 minutes' life. The aircraft have an 80-centimeter wingspan, weigh about 500 grams, and are built with inexpensive, lightweight yet strong polypropylene foam. This lets them land on virtually any terrain without damage, according to Zufferey.

Their airspeeds can range between 8 and 20 meters per second (between 18 and 45 miles per hour), and they can fly as high as several kilometers, although a swarm generally stays below 150 meters to avoid conflicts with general aviation.



Researchers at Ecole Polytechnic Federale de Lausanne's Laboratory of Intelligent Systems are developing flying robots using wireless communications that could fly as a swarm and help with mapping, remote sensing, ground searches, and other similar operations.

Severin Leven, a doctoral student working on the project, said experimental flights with swarms of drones pose a challenge because so many individual aircraft are involved. The robots must be able to fly with minimal or no external human input, but must also coordinate their activities as part of a group.

Once hand launched, Zufferey explained, a robot relies on autopilot to control airspeed, altitude, turn rate, and stabilization. The drones use actuators to control steering.

A small circuit board, running Linux and powered by an Intel XScale PXA270 processor, is programmed with algorithms to provide commands to the autopilot for controlling drones acting as part of a swarm. The robots also have a microcontroller that reads sensor data and computes outputs sent to the actuators.

Swarming algorithms enable the robots to form themselves into an effective flight group, which then functions as a communications network. The drones use the IEEE 802.11n version of Wi-Fi, implemented via dongles, for line-of-sight transmissions to one another over distances of up to 500 meters, explained Zufferey. Users could then communicate with the drones via smartphones or computers.

The drones can use GPS to fly to specified locations to, for example, return to a take-off location, safely land if the battery is drained, or gather with other drones to operate as part of a swarm.

Published by the IEEE Computer Society

Part of the researchers' work was inspired by the way army ants communicate, specifically their practice of laying and maintaining pheromone paths between nests and food sources. To imitate this behavior, the robots can act as either *ants* or *nodes*.

Initially, they spread out and fly as far as their communication range allows, remaining in contact with neighboring drones. At that point, they become nodes, circling fixed positions and acting as communication highways for ant drones, which use information the nodes provide to determine the best flight paths.

As ant drones fly farther from the grid, they can sometimes become nodes and expand the network, if it strengthens communications. The swarm's purpose is to establish an optimal communications network to, for example, enable first responders to stay in touch with one another in a remote area where communication has been disrupted.

So far, the researchers have built and flown 10 robots at a time. The scientists are commercializing their work via a company they formed, senseFly (www.sensefly.com), which is currently offering small drones for autonomous-mapping and remote-sensing applications.

Mobile Malware on the Rise

Each year since cell phone use became popular, security experts have said malware that affects mobile devices was on the verge of becoming a problem. But each year, that didn't happen. Now, though, it appears that mobile malware has arrived.

Devices such as smartphones can contain valuable information like credit card numbers, usernames, and passwords. According to Vikram Thakur, principle security response manager for security vendor Symantec, such data is a valuable black-market commodity. This makes smartphones potentially desirable targets for cybercriminals.

Another factor driving mobile malware is that wireless devices are becoming more popular than traditional PCs, even for confidential purposes such as banking and other financial transactions. For example, smartphone sales exceeded PC sales for the first time in the fourth quarter of 2010, according to Luis Corrons, director of Panda Security's PandaLabs.

As occurred with PC malware, hackers eventually will figure out how to make money from information collected via mobile malware, encouraging expansion of the practice, predicted Thakur. With this in mind, he noted, hackers have increased proof-of-concept attacks on mobile devices, including spyware-installation and phishing attempts, as well as efforts to circumvent authentication.

There have also been malicious fake wireless-security products, and financially motivated social-engineering-based Trojans have been embedded in mobile games.

Noted Don DeBolt, director of threat research for CA Technologies' Internet Security Business Unit, hackers can use botnets to distribute mobile ransomware, malware that makes data on a system unavailable in an attempt to force a user to pay a ransom for its restoration.

Symantec recently discovered an attempt to create a botnet consisting of devices that run Google's Android operating system, the world's most widely used smartphone OS.

The platforms with the greatest market share are most attractive to hackers and thus face the greatest risk, said CA Technologies senior malware researcher Dinesh Venkatesan.

Recent Android-related incidents highlight mobile devices' vulnerabilities. Several programs available via Google's Android Market for applications early this year appeared to be legitimate software, but hackers had actually added Trojans to them. Google removed the altered programs from the Android Market, but not before they had been downloaded thousands of times. The company subsequently distributed software to eliminate rogue applications from users' devices.

Mobile platforms other than Android have also experienced problems. For example, Trojans have affected devices running the Symbian OS, noted Venkatesan. The SymbOS.Merogo short-message-service worm also targets Symbian, PhoneSnoop spyware affects BlackBerry devices, and the iPhoneOS.Ikee worm was designed for Apple's iOS mobile platform.

Most security vendors are using the same approaches to combat mobile malware that they've used on the PC, such as endpoint scanning, explained Venkatesan. Some infrastructure companies, such as Cisco Systems, are trying to extend security beyond the corporate network's physical perimeter by, for example, enforcing company policies at the router closest to a managed mobile device, he added.

Despite the rise of mobile malicious software, Corrons said, security researchers still see much more new PC malware.

NEWS BRIEFS

Dual-Core Processors to Improve Smartphone Performance

An increasing number of people are now using smartphones as mobile computers, prompting an ongoing drive to improve the devices' performance.

With this in mind, several companies are now providing dual-core chips for smartphones and tablets. Adding more cores, says Nathan Brookwood, Insight 64 analyst, is the most power-efficient way to increase mobile processors' performance. Dual-core processors achieve high performance speeds because they divide up tasks and perform them in parallel—a technique known as symmetric multiprocessing.

Running multithreaded applications in parallel on dual cores can provide a performance boost. Brian Carlson, OMAP 5 product line manager for Texas Instruments' Wireless Business Unit, says his company has measured a 1.6× performance increase using dual-core processors to run Web browsers. Brookwood adds that software developers, led by game developers, have begun redesigning their software to run in parallel.

Dual-core chips also provide scalability for both power and performance, factors that are critical for smartphones' high on-demand performance requirements. They can, for example, either provide additional performance when needed or shut down one of the cores to conserve power.

Smartphones won't use the same types of dual-core processors currently found in PCs and laptops, primarily because mobile chips are significantly smaller. The TI chips, for example, are 12 mm × 12 mm, not the 37 mm × 37 mm or larger chips found in a desktop system. They also run at milliwatts rather than watts.

These devices need ARM-compatible chips to cope with their completely different operational requirements, which include an inability to cool chips and the need for low power. Vendors including Nvidia, Qualcomm, TI, and Motorola are now supplying dual-core ARM chips precisely for this market. The LG Optimus 2X runs Nvidia's dual-core Tegra 2 processor, and TI's OMAP 4 dual-core processors can be found in the LG Optimus 3D as well as in tablets, including the RIM PlayBook. Motorola is using dual-core chips in its ATRIX 4G phone and XOOM tablet. According to Brookwood, Intel anticipates that its Medfield Atom processor, scheduled to ship later this year, will make inroads into the smartphone and tablet markets.

ARM has traditionally emphasized performance per watt at the low-power levels needed for long battery life, Brookwood says, which has allowed the technology to lead in this market segment. Unlike Intel and AMD, ARM licenses its designs to third-party system-on-chip designers, who then combine ARM's intellectual property with other system elements to create custom-tailored SoCs with the size, power, and performance characteristics that smartphone designs require.

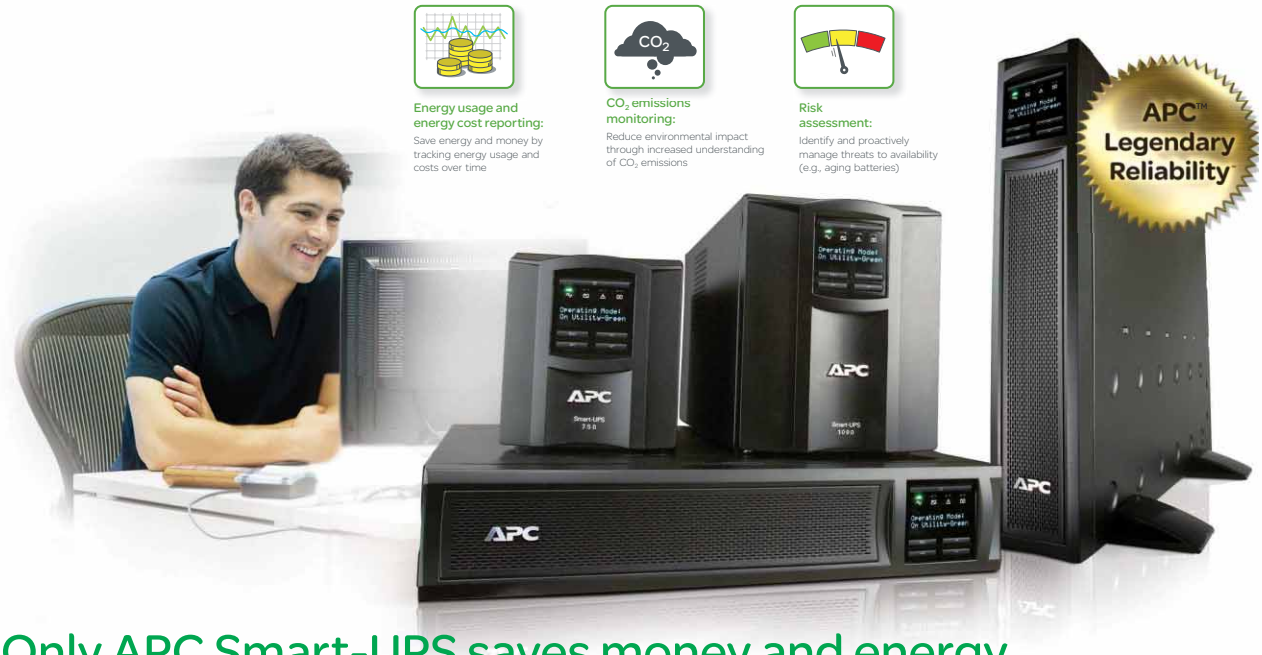
Carlson says the prime issue with using dual-core processors in smartphones is achieving a performance-power balance. These constraints are already familiar to chipmakers and phone designers.

The market opportunities are huge, analysts say. According to a March 2011 report by ABI Research, mobile data plan revenues alone are expected to exceed $102 billion worldwide by 2016, fueled by mobile enterprise customers' smartphone adoption. Brookwood says that while analysts may differ about how smartphone adoption will affect the computer market, they do agree there will be a change.

News Briefs written by ***Linda Dailey Paulson****, a freelance technology writer based in Portland, Oregon. Contact her at ldpaulson@yahoo.com.*

Editor: Lee Garber, *Computer*; l.garber@computer.org

Now, manage both your UPS and your energy proactively



Only APC Smart-UPS saves money and energy without sacrificing availability

Today's more sophisticated server and networking technologies require higher availability. That means you need more sophisticated power protection to keep your business up and running at all times. But that's not all. In today's economy, your UPS must safeguard both your uptime *and* your bottom line. Only APC by Schneider Electric™ helps you meet both of these pressing needs. Specifically, the APC Smart-UPS™ family now boasts models with advanced management capabilities, including the ability to manage your energy in server rooms, retail stores, branch offices, network closets, and other distributed environments.

Intelligent UPS management software

PowerChute™ Business Edition, which comes standard with Smart-UPS 5 kVA and below, enables energy usage and energy cost reporting so you can save energy and money by tracking energy usage and costs over time; CO_2 emissions monitoring to reduce environmental impact through increased understanding; and risk assessment reporting so you can identify and proactively manage threats to availability (e.g., aging batteries).

Best-in-class UPS

Our intelligent, interactive, energy-saving APC Smart-UPS represents the combination of more than 25 years of Legendary Reliability™ with the latest in UPS technology including an easy-to-read, interactive, alphanumeric LCD display to keep you informed of important status, configuration, and diagnostic information; a unique battery life expectancy predictor; and energy-saving design features, like a patent-pending "green" mode. Now, more than ever, every cost matters and performance is critical. That's why you should insist on the more intelligent, more intuitive APC Smart-UPS.

Why Smart-UPS is a smarter solution

Intuitive alphanumeric display

Get detailed UPS and power quality information at a glance – including status, about, and diagnostic log menus in up to five languages.

Configurable interface

Set up and control key UPS parameters and functions using the intuitive navigation keys. On rack/tower convertible models, the display rotates 90 degrees for easy viewing.

Energy savings

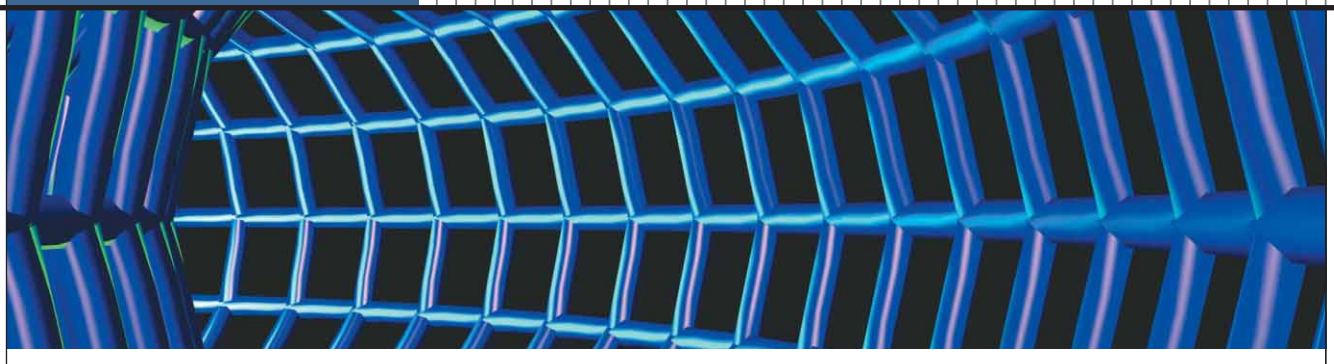
A patent-pending "green" mode achieves online efficiencies greater than 97 percent, reducing heat loss and utility costs.

Download White Paper #24, "Effect of UPS on System Availability," and register to WIN an APC Smart-UPS 1500VA Rack/Tower LCD 120V, a $779 value!

Visit www.apc.com/promo Key Code b606v Call 888-289-APCC x6266

APC™ by Schneider Electric

email: esupport@apc.com • 132 Fairgrounds Road, West Kingston, RI 02892 USA • 998-4144_US



Toward an Equal Opportunity Web: Applications, Standards, and Tools that Increase Accessibility

Lourdes Moreno, Paloma Martínez, Ana Iglesias, and Belén Ruiz-Mezcua
Carlos III University of Madrid

New options offer the promise that those who require assistive technology to access Web content will soon enjoy the full range of interactivity, thereby realizing the equal accessibility goal.

Rich Internet applications (RIAs) and supporting technology, such as JavaScript and Flash, embed components in the user agent or in the browser on the client side, providing more dynamic Web content and more attractive and interactive websites. Through asynchronous client-server communication, RIAs can offer an interactive experience closer to that of a desktop application. However, not all users benefit from this interactivity. Users with disabilities, for example, must access the Web using assistive technologies, such as a screen reader that delivers audio content to the visually impaired. Because RIA technologies are a client-side programming technique, everything loads on that side before the actual interaction. Consequently, the screen reader cannot read the updates, which decreases the user's content accessibility.

Fortunately, with the semantic Web, interface developers should be able to adapt interfaces to meet specific needs, enabling user agents and browsers to understand semantic markup. As the "Road to Web Accessibility" sidebar describes, websites designed in accordance with the Web Content Accessibility Guidelines (WCAG) provide markup semantics that assistive technologies can interpret and successfully relay to the user. Developers can thus use APIs to enable assistive technologies to access dynamic content,[1] adding to techniques that already offer dynamic contextual help through tailored interfaces. More important, interactive technologies might extend learning for disabled users. Users with cognitive and learning disabilities, for example, could experience a wider range of multimodal interaction, since they could tailor the interface to their needs, allowing them to select the most appropriate Web content.

The goal of equal accessibility will not be easy to attain, however. For a fully static webpage, an effective linear presentation is sufficient, but interactive technology involves content updates that preclude a linear approach. The user must be both aware of the update and able to access the new content without disrupting the task in progress. Consequently, assistive technologies must understand the semantics well enough to communicate updates.[2] To assist in promoting accessibility, researchers are exploring development toolkits and frameworks, as well as testing tools to enable faster, easier Web 2.0 application development and to broaden Web accessibility.

 Published by the IEEE Computer Society

ROAD TO WEB ACCESSIBILITY

The myriad perspectives from which to contemplate Web accessibility have yielded a variety of conceptual definitions. According to the Education and Outreach Working Group, web accessibility means that "people with disabilities can use the Web" (www.w3.org/WAI). The International Organization for Standardization's ISO 9241-171 defines it as the "usability of a product, service, environment or facility by people with the widest range of capabilities."

For our purposes, Web accessibility is the development of Web resources that all people can use, regardless of their technical, physical, or cognitive limitations. The number of users affected by website inaccessibility—the digital divide—is growing, largely because users are more frequently encountering Web accessibility obstacles stemming from use context and technological incompatibilities.

Directives, legislation, and standards are attempting to combat this rising inaccessibility, identifying problems and suggesting new, accessible designs. Standards bodies and legislation such as the World Wide Web Consortium (W3C) and Web Accessibility Initiative (WAI) are strong evidence of progress.

The most important WAI component is the Web Content Accessibility Guidelines (WCAG), an official standard in the European Union that is referenced in most legislation worldwide. Other important initiatives include US legislation that requires conformance with Section 508 of the Rehabilitation Act (29 U.S. Code §794d), which describes technical standards related specifically to Web accessibility. Although less extensive, these standards are similar to the WCAG and even map Section 508 to those guidelines.

Over the years, the W3C has published WCAG 1.0 and WCAG 2.0 as W3C recommendations, with the latter becoming official in December 2008. WCAG 2.0 reflects the more interactive nature of Web content and accommodates HTML and Cascading Style Sheets, as well as scripting and non-W3C Web technologies such as RIA.[1]

Although RIA technologies—including JavaScript, Ajax, and Flash—result in more interactive websites, they often impose barriers to accessibility. New specifications can add semantics to a webpage or an application that allow assistive technology to better represent interfaces and interactions to the user. To support this goal, the W3C WAI developed the Accessible Rich Internet Applications (WAI-ARIA) and incorporated ARIA and Flash techniques for WCAG 2.0 (http://www.w3.org/tr/wcag-techs).

The W3C is currently working on HTML 5, aiming to reduce the need for proprietary plug-in-based RIA technologies. HTML 5 will provide Web users and developers with enhanced functionality without using the proprietary technologies that have recently become popular. The new standard, which will replace XHTML standards, should ensure accessibility by fixing the main problems that Web developers are now encountering—such as the need to install plug-ins to access multimedia content and the requirement to separate content and presentation.

HTML 5 will enhance accessibility in several ways: it will add implicit semantic information, be defined on the basis of the document-oriented model, and differentiate content from presentation. HTML 5 will also present better structure than other languages. Div blocks on a webpage will have their own tags—including article, footer, header, and navigation—so that user agents and assistive technologies can recognize them and gather more accurate information. Site authors will be able to embed accessibility multimedia elements natively using <audio> and <video> tags, obviating the need for plug-ins. Although the draft of the new standard that is still under development does not allow the inclusion of HTML 5 elements that provide caption and audio descriptions, it is anticipated that future standards will support adding subtitles and captions using the HTML 5 <video> tag inside webpages.[2] Both user agents and assistive technologies can use this information to enable alternate ways of viewing and navigating a page.

References

1. L.D. Paulson, "W3C Adopts Web-Accessibility Specifications," *Computer*, Feb. 2009, pp. 23-26.
2. S. Pfeiffer and C. Parker, "Accessibility for the HTML5 <video> Element," *Proc. Int'l Cross-Disciplinary Conf. Web Accessibility* (W4A 09), ACM Press, 2009, pp. 98-100.

DEVELOPING ACCESSIBLE RICH APPLICATIONS

Many RIA accessibility challenges arise from the applications' dependency on Ajax (Asynchronous JavaScript and XML) and JavaScript. In addition, RIA toolkits introduce complex user-interface components and dynamically changing content, which is particularly problematic for the keyboard navigation essential to accessibility. In HTML, the focus is only on links and form elements accessible through a keyboard interface; however, in Web 2.0 applications, the focus could also be on span or div elements.

To address accessibility concerns, the W3C Web Accessibility Initiative (WAI) developed specification techniques for RIA technologies. The WAI Accessible Rich Internet Applications (WAI-ARIA) enable keyboard access for all elements by extending the tabindex property applied to any element.

WAI-ARIA provides a navigable page structure or, as Figure 1 shows, an accessible tree widget. The comparison of WCAG 1.0 (without ARIA) and WAI-ARIA semantics shows the severity of Web 2.0 accessibility problems and the inability of some assistive technologies to circumvent these difficulties. However, it also shows the efficiency with which developers can fix these problems by implementing WAI-ARIA recommendations.

The WAI-ARIA suite provides mechanisms to increase Web accessibility, particularly for rich websites and applications with a focus on dynamic content and user interface controls. The addition of semantic data to HTML and XHTML enables assistive technologies to better represent user interface components and dynamic interactions. The suite currently consists of five documents: the WAI-ARIA technical specification, authoring practices, the primer, and the 1.0 user agent implementation guide and roadmap. The documents explain, for example, why JavaScript needs accessibility architecture and relate how the user agent can map such a structure to accessibility frameworks on the native platform.

The *WAI-ARIA technical specification*, a planned W3C Recommendations Web standard that combines the two

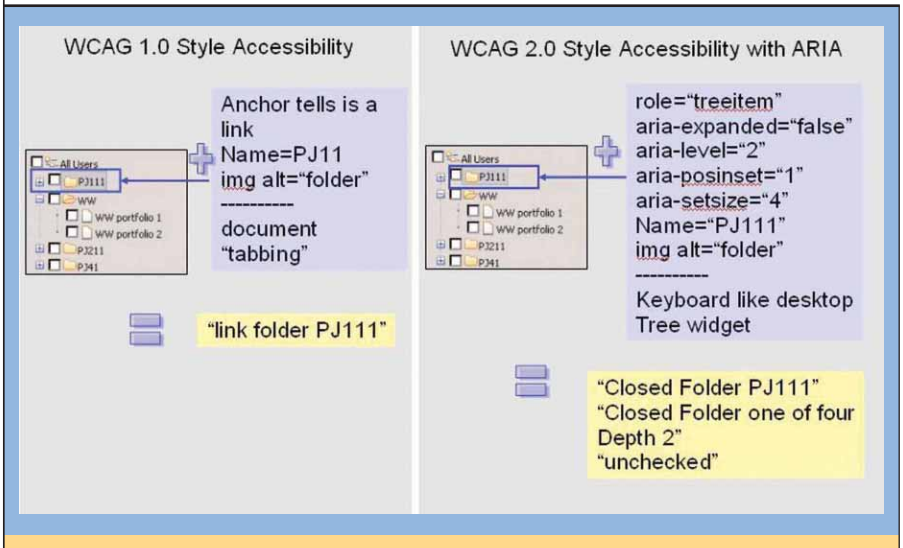


Figure 1. Tree widget using WCAG 1.0 and WCAG 2.0. Within a tree widget, using WCAG 1.0 (left), the screen reader sees only link folder PJ111. Using WCAG 2.0 (right), with WAI-ARIA, the reader sees the information to indicate that the folder is closed (expanded = false) and its depth (aria-level= 2), set position, and set size. Without this information, the user's accessibility to webpage content through the reader is limited. Figure from "WAI-ARIA Primer," W3C Working Draft, 4 Feb. 2008; www.w3.org/TR/2008/WD-wai-aria-primer-20080204.

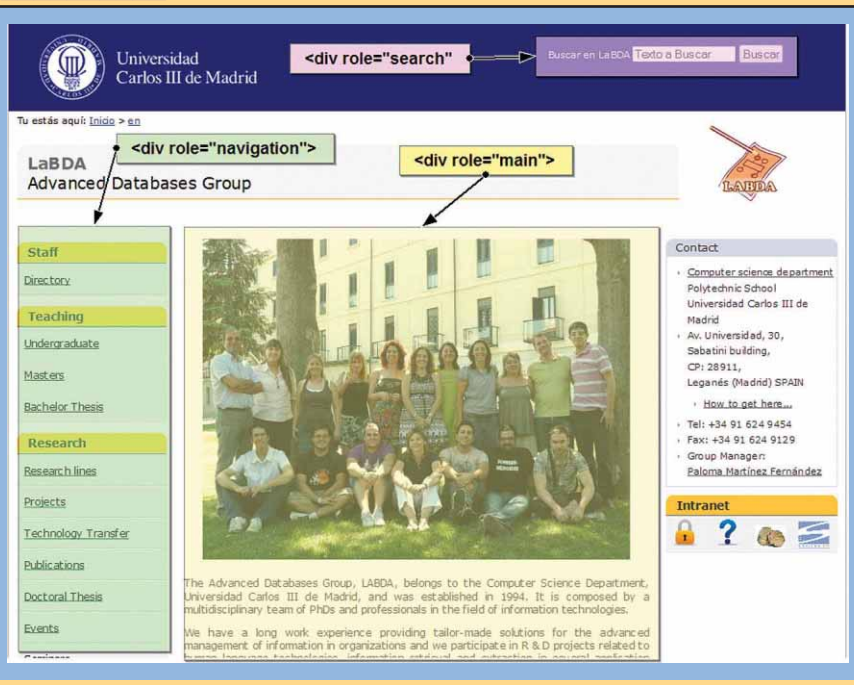


Figure 2. A typical website with regional landmarks indicated. A screen reader aligns keyboard navigation to landmarks, such as navigation, search, and main, allowing the screen reader to provide a visually impaired user accurate information about the page and any updates in the regions the landmarks identify.

previously published WAI-ARIA draft specifications, is intended for developers of Web browsers, assistive technologies, and other user agents. The specification comprises the roles, states, and properties modules. With the WAI-ARIA roles module, the site developer can specify a Web document's functional zones define each component's role, and then use the WAI-ARIA states and properties module to determine each component's characteristics and values.

Applying these modules adds semantics to webpage components and widgets so that assistive technologies can interpret their operation. The result is a guarantee of accessible keyboard navigation, the accessibility of controls and widgets, and accessibility through dynamic content-update notifications.

Figure 2 shows a website with regional landmark roles, which provide a navigable structure within a webpage that a screen reader can access.

Figure 3 shows the HTML code for the regional landmark roles in Figure 2.

Other WAI-ARIA roles, such as those that define page structure, are also useful. The *WAI-ARIA authoring practices* offer guidance on incorporating these additional roles. This document also targets Web developers and provides detailed advice and examples of how to build accessible RIAs.

The *WAI-ARIA primer*, a planned W3C Working Group Note, introduces developers to accessibility-related problems that WAI-ARIA is intended to resolve, as well as to WAI-ARIA's fundamental concepts and technical approach.

The *WAI-ARIA user agent implementation guide*, another planned W3C Working Group Note, describes how browsers and other user agents should support WAI-ARIA. It includes such details as how to expose WAI-ARIA features to platform accessibility APIs.

Finally, the *WAI-ARIA roadmap* defines the path for making rich Internet content more accessible, including steps already taken, future steps, and a timeline. It describes the technologies to map controls, Ajax live regions, and events to accessibility APIs, including cus-

tomized controls for RIAs. It also describes techniques to mark common Web structures, among them menus, primary content, and secondary content.

To understand JavaScript's power, consider the document object model (DOM) node in Figure 4, which is part of a model-view-controller architecture. Without JavaScript, assistive technologies acquire accessibility information only through the HTML element's tag name—and then only the accessibility attributes that the tag can provide. Because the data node (the model) is separate from the user interface node (the view), the user agent manages the document element according to the element's default behavior, and the user agent's default behavior at that element acts as the controller.

With JavaScript, the default user agent's behavior no longer serves as the controller. JavaScript overrides the default user agent behavior at the DOM node, manipulating data, content, and style in response to user interaction events. The result is custom widgets. In this scenario, default accessibility information is no longer valid, so the contract is also invalid. The asterisks in front of the blue text (role, state, actions, and so on) represent potential accessibility errors, as well as gaps in the base markup that result from the author's inability to provide the new semantic data to support the contract.

Both browsers and assistive technologies are providing WAI-ARIA support. All four major browsers have either implemented support or plan to do so. Opera 9.5 and Firefox 1.5+ already include support for ARIA, as does Internet Explorer 9 beta. WebKit, the open source application framework behind Safari, has begun to add support for ARIA.

Assistive technologies that are starting to widely support ARIA include JAWS 7.1+, Window-Eyes 5.5+, NVDA, and Zoomtext 9+. The trends we've observed indicate that this support is likely to increase.

```
<html>
<body>
      ...
      <div role="search">
      The search area
      </div>

      <div role="navigation">
      The navigation area
      </div>

      <div role="main">
      The main content area
      </div>
      ...
</body>
</html>
```

Figure 3. HTML code with regional landmark roles. The roles make navigation easier for a user who must rely on a screen reader or similar assistive technology.

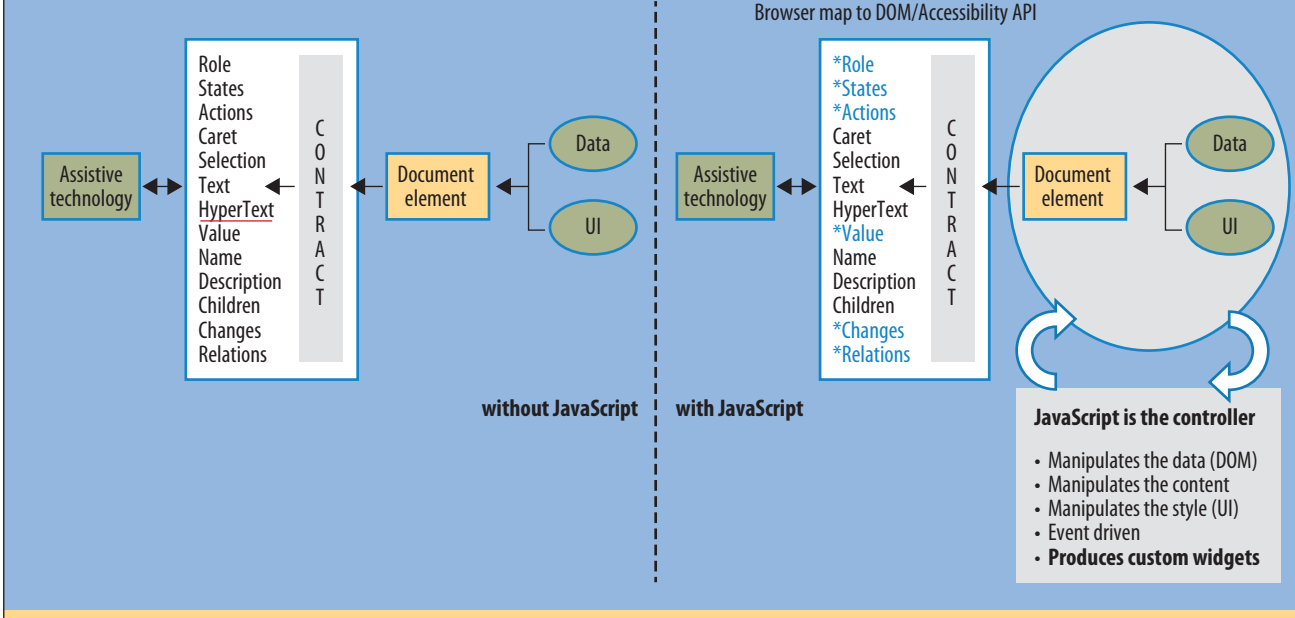


Figure 4. Accessibility interoperability at a DOM node without and with JavaScript. Without JavaScript (left), the data node (Data), which should include semantic information, is separated from the user interface (UI) presentation. Default user agent behavior forms the controller. With JavaScript (right), the same DOM node now has JavaScript as the controller, which overrides the default user agent behavior at the node. Figure from "WAI-ARIA Primer," W3C Working Draft, 4 Feb. 2008; www.w3.org/TR/2008/WD-wai-aria-primer-20080204.

Table 1. Summary of non-ARIA technology for accessible websites.

Technology (base)	License type	Integrated development environment	Compatible browsers	WCAG/ Section 508 support	Accessibility support	Accessibility statement
Flex 3.0 (Flash)	Open source	Adobe Flex Builder	Any browser with Flash player	Section 508 with exceptions	Flash player integration support for MSAA	Section 508 support
Silverlight 4.0 (Windows)	Proprietary software; free license	Visual Studio 2008, Expression Web	IE 6+, Firefox 1.5+, Opera 2.0+, Chrome	Section 508 with exceptions	User interface automation API	Not provided
Java FX 1.3 (Java)	Partially open source	NetBeans for JavaFX	Any browser with JRE and mobiles with JME	Section 508 with exceptions	No accessibility API; not compatible with Java accessibility API	No support

Table 2. Summary of ARIA-based technology for accessible websites.

Technology	License type	Integrated development environment	Compatible browsers	WCAG/ Section 508 support	WAI-ARIA support	Accessibility statement
Dojo 1.4.2	Open source	Aptana or Komodo	IE 6+, Firefox 1.5+, Safari 3.1+, Opera 9.6+, Konkeror 3.5+, older browsers with Dijit	Both	With Dijit (forthcoming)	WCAG/ Section 508
Bindows	Commercial software	Bindows	IE 5.5+, Firefox 1.4+, Netscape 7.1+, K-Meleon, Camino, WebKit 525+	Section 508	Limited (forthcoming)	WCAG/ Section 508
Google Web Toolkit 1.7	Open source	Eclipse and other Java IDEs	IE 6+, any browser with JRE and mobiles with JME	Not specified	Limited (forthcoming)	Not provided
ExtJS 3.2.1	Commercial and open source	Ext Designer	IE 6+, Firefox 1.5+, Safari 3+, Opera 9+	Section 508 extended (forthcoming)	Limited (forthcoming)	Not provided
Fluid Infusion 1.2	Open source	Any JavaScript IDE	Firefox 3+, IE 6+, Opera 9.6+, Safari 3.2+	Not specified	Limited (forthcoming)	Not provided
jQuery User Interface 1.8.1	Open source	Any JavaScript IDE	Firefox 2+, IE 6+, Safari 2.0.2+, Opera 9+	Section 508 extended (forthcoming)	Limited (forthcoming)	Not provided
Yahoo User Interface Library 3.1.1	Open source	Any JavaScript IDE	Firefox 3+, Opera 9.6+, IE 6+, Safari 3.2+	Not specified	Limited (forthcoming)	Not provided
BBC Glow Widgets 1.7.3	Open source	Any JavaScript IDE	IE 6+, Firefox 2+, Chrome 1+, Opera 9+, Safari 2+	Not specified	Limited (forthcoming)	Not provided

ACCESSIBILITY IN WEB PLATFORMS

At present, the Web uses two kinds of platforms.[3] *Non-Ajax platforms* are environments that depend on components embedded in the HTML document, and operation requires additional browser software or a plug-in. *Ajax platforms* are RIA development frameworks based on Ajax technologies. The platforms feature JavaScript implementation and are completely independent of the operating system, since only a browser with Ajax support is necessary.

Tables 1 and 2 summarize the differences between the two platforms, drawing from information we extracted from the websites detailing the various technologies.

Non-Ajax platforms

Flash is the most common RIA technology in the non-Ajax category, with websites being entirely Flash-based or using components, such as calendars, news feeds, and media players.

Sites that are entirely Flash-based present a challenge to accessibility. Because they do not use standards like HTML or Cascading Style Sheets (CSS), responsibility for accessibility falls to the browser and built-in technologies like Microsoft Active Accessibility and accessibility APIs. Problems arise when Web user agents cannot access Flash websites. Developers could offer alternative versions of these sites that any user agent can access, but the alternative sites would have minimal functionality and generally no updates.

Flash components are small programs that any developer can create, and their unknown origin can often mean the lack of an accessibility guarantee before the site author adds the component. If the Flash component does prove inaccessible, the entire site's accessibility suffers. Even if the visible layer is not HTML, the component developer must satisfy all WCAG and Section 508 accessibility requirements.

As a first step in installing these components, the site author should carefully validate the accessibility level and look at possible alternatives if the component doesn't meet requirements. Authors often add Flash animations to their sites, for example, because the animation has a relatively light format and is easy to include, thanks to Flash players.

As is true of other HTML elements on the page, the site author should follow accessibility guidelines, such as Guideline 1.2 of WCAG 2.0 and should offer alternative content to audio tracks (such as captioning), Flash techniques for WCAG 2.0, and other elements where applicable.[4] The WCAG 2.0 suggests adding a group name to the accessible name of a form control, such as a radio button. Figure 5 shows an example of how an author can do this. WCAG 2.0 further states that the suggested Flash technique is applicable for Adobe Flash Professional MX and Adobe Flex.

Technical and best practices documents already recommend making Flash more accessible—in essence by offering alternative content to every Flash object—and these criteria are quite similar to WCAG requirements. Another way to guarantee website accessibility is to provide a logical organization of content that ensures keyboard access and correct tabulation with technical aids.[5]

Silverlight, Microsoft's alternative to Flash, is a plug-in for free Web browsers that adds certain functionalities such as video, vector graphics, and animation display. Like Flash, Silverlight is not an open standard, but multiple browsers and platforms support it, and its functionalities are similar to those of Flash. Although Silverlight is free, it is not open code; however, Microsoft has published parts of the code under a permissive license.

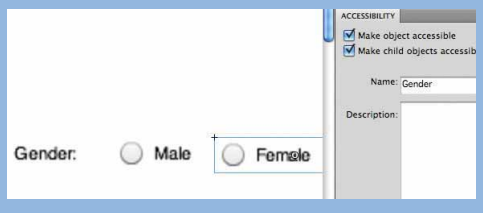


Figure 5. Using the technique suggested in WCAG 2.0, Flash concatenates the group name with each button's individual name to produce "Gender:Female." Figure from Flash Techniques for WCAG 2.0; www.w3.org/WAI/GL/WCAG20-TECHS/Flash.html.

Silverlight presents alternatives for each of its objects and passes any component accessibility concerns to the browser through the platform's accessibility APIs. Because Silverlight uses the Extensible Application Markup Language (XAML) to define the user interface, it can dynamically upload and manipulate XML code with the DOM. Using XAML technology, the developer can separate presentation, content, and events and include metadata with each object at the XHTML code level, thus enabling assistive technologies to interpret content.

Flex Framework applies the same model using component library code to facilitate the development of Flash-based RIAs. Instead of XAML, Flex Framework uses the Multimedia Extensible Markup Language (MXML) to define user interfaces, communicate with application servers, and execute other operations. As with XAML, MXML lets authors separate presentation, content, and events. It also has APIs defined to incorporate properties and accessible metadata in objects, allowing assistive technologies to interpret content.[6]

Flex Framework's principal advantages for accessibility are its compatibility with assistive technologies, such as screen readers, as well as its wide support for Section 508 requirements. According to Adobe, Flex best practices are nearly the same as those for Flash, although Adobe has adapted them to suit MXML and included new characteristics, such as specific recommendations for each accessible Flex component.

JavaFX is a non-Ajax platform that offers very limited accessibility support and is incompatible with Swing's Java Accessibility API. Some independent implementations, such as fxaccessible, use a Swing component to add support for the Java Accessibility API, which in turn allows screen readers to work with the JavaFX-based application.

AJAX PLATFORMS

Ajax-based technologies include but are not limited to asynchronous JavaScript and XML. Their advantages include high upload speed and usability. Google Maps and Gmail are among the numerous Web 2.0 applications that rely on these technologies. On the client side, JavaScript provides full functionality by interacting with the Ajax engine.

TECHNOLOGIES TO ENHANCE ACCESSIBILITY

The following is a list of interface libraries and evaluation and development tools used to promote the development of more-accessible websites. Most of these tools are open source.

- Accessibility Evaluation Toolbar: https://addons.mozilla.org/es-es/firefox/addon/5809
- Adobe Accessibility Best Practices for Flex: www.adobe.com/accessibility/products/flex/best_practices.html
- Adobe Flash Platform: www.adobe.com/flashplatform
- Adobe Flex 3: www.adobe.com/es/products/flex
- ASP.NET Ajax Roadmap: www.asp.net/ajax/documentation/live
- BBC Glow widgets: www.bbc.co.uk/glow/docs/1.5/api/glow.widgets.shtml
- Dojo Toolkit: www.dojotoolkit.org
- Ext JS Cross-Browser Rich Internet Application Framework: www.extjs.com/products/extjs
- Flash Techniques for WCAG 2.0: www.w3.org/tr/wcag20-techs/Flash.html
- Fluid Infusion: http://fluidproject.org/products/infusion
- Firefox Accessibility Extension: http://firefox.cita.uiuc.edu
- Firefox Juicy Studio Accessibility Toolbar: https://addons.mozilla.org/en-US/firefox/addon/juicy-studio-accessibility-tool
- Fxaccessible Accessibility for JavaFX: http://code.google.com/p/fxaccessible
- Google Web Toolkit: http://code.google.com/intl/en/webtoolkit
- jQuery UI: http://jqueryui.com
- MB Technologies, Bindows: www.bindows.net
- Microsoft Silverlight: http://silverlight.net
- The Paciello Group Blog: WAI-ARIA Implementation in JavaScript UI Libraries–updated, www.paciellogroup.com/blog/?p=313
- Web Accessibility Toolbar: www.visionaustralia.org.au/info.aspx?page=614
- Yahoo User Interface Library widgets: http://developer.yahoo.com/yui

Functionalities are grouped into libraries that a JavaScript engine in the user agent or the client browser runs.

Ajax controls HTML documents as well as their updates in a way that is transparent to the user, which is not advantageous for accessibility, since the user might not notice these updates. Ajax operations include, among others, the dynamic modification of Web content with DOM and event capture. Ajax carries out interactions asynchronously using the XMLHttpRequest object for data exchange with the webserver without the need to reload the page. For the presentation layer, Ajax uses W3C standards (HTML, CSS, and XHTML) as well as other embedded components, such as Flash and multimedia content.

Most popular platforms integrate Ajax and, although not all Web user agents implement the XMLHttpRequest object in the same way, most of the better-known user agents generally agree.

Whether they are frameworks, toolkits, or libraries, numerous Ajax-based solutions are available. Two solutions, Dojo and Bindows, deserve special consideration for their support of WAI-ARIA and tagging that permits access for assistive technologies. Additionally, both solutions include components that offer keyboard support, and both provide an accessibility statement. For these reasons, Dojo and Bindows have received special recognition from the W3C for their compliance with the WAI-ARIA specification.

Dojo is an open source toolkit for developing accessible Web applications, and it has the official support of companies like IBM and Sun. Dojo's core widget set, Dijit, has included accessibility support since its 1.0 release, and Dojo claims to be the only fully accessible open source toolkit available for Web 2.0. Because Dijit uses the WAI-ARIA specification, all Dijit widgets are accessible using the latest versions of the JAWS or Window-Eyes screen readers with Firefox 3.

In addition to its WAI-ARIA support, Dojo is compatible with the latest versions of Web browsers and screen readers. It offers the user complete keyboard access for elements, including tabulation, keyboard events, and focus assignments. Dojo also supports visually impaired users by guaranteeing assistive technologies proper access. Its inclusion of alternative themes and texts for multimedia content ensures that assistive technology will be able to detect any content changes and updates and accommodate changes in font sizes—features that enable visually impaired users to adapt Web content to suit their specific needs.[7]

The Bindows Ajax-based solution is the first framework to be officially recognized for its Section 508 compliance. Unlike Dojo, Bindows is not an open source framework, but it does offer many of Dojo's advantages, such as complete keyboard navigation support, the inclusion of alternative texts, and support for notifying users of dynamic content changes. Bindows improves framework accessibility by supporting assistive technologies, such as screen readers and magnifiers, and providing reference material and a module to bring legacy applications up to Section 508 standards.

Table 2 also lists several popular JavaScript user-interface libraries that are adding WAI-ARIA support to their widgets and components; however, their conformance to Section 508 or WCAG guidelines has not yet been documented. The "Technologies to Enhance Accessibility" sidebar provides links for these libraries, most of which are open source:

- *Google Web Toolkit*. This toolkit partially supports WAI-ARIA and fully supports keyboard navigation,

including tabulation, keyboard events, and focus assignments.

- *Ext JS*. Although details are lacking, the literature reports real accessibility improvements.[8] Version 3.0 supports both Section 508 and WAI-ARIA.
- *Fluid Infusion*. This new JavaScript application framework built on top of jQuery lets both developers and users customize user interfaces. Although it does not provide any documentation about Section 508 or WCAG conformance, its developer states that assistive technology, such as screen readers, has been tested.
- *Yahoo User Interface Library widgets*. WAI-ARIA plug-ins exist for many of these widgets. Although Yahoo does not provide documentation about Section 508 or WCAG conformance, it states that most of these widgets would withstand a rigorous Section 508 testing. Yahoo plans to implement more WAI-ARIA support in the upcoming version 3.0.
- *BBC Glow widgets and ASP.Net Ajax Roadmap*. Both the widgets and roadmap have supported WAI-ARIA since mid-2009.

Other technological solutions, such as the combination of Google AxsJAX and the SADIe transcoder, improve accessibility by dynamically injecting WAI-ARIA statements into Web content.[9] AxsJAX is an open source framework that lets developers use high-level patterns developed from the underlying WAI-ARIA markup. The SADIe transcoder uses CSS annotations to generate AxsJAX framework code and insert it into webpages. This approach can, for example, improve page access for visually impaired users who use a screen reader. Similarly, the Hearsay-Dynamo browser[10] lets users with disabilities experience RIA content, even content that the author has annotated according to WAI-ARIA specifications.

ACCESSIBILITY EVALUATION TOOLS

Currently, tools that can evaluate how well a webpage conforms to WAI-ARIA requirements are either still in trials or permit only a limited inspection. Consequently, a complete accessibility evaluation of pages with and without WAI-ARIA support requires manually validating specifications and good practices. Even so, it is worth investigating some of the automatic tools that provide WAI-ARIA testing features, such as the Web Accessibility Toolbar, Firefox Accessibility Extension, and Firefox Juicy Studio Accessibility Toolbar. The "Technologies to Enhance Accessibility" sidebar provides links to these tools.

Although RIA technologies make websites more attractive for many users, they can also raise new accessibility barriers for users with special needs. Because problems arise for keyboard navigation when these applications automatically change pages or content, users who access the Web through assistive technology can feel lost while surfing the Web. To ensure Web accessibility, the Web community has developed new techniques such as WAI-ARIA.

At present, only a handful of solutions are oriented toward guided accessible Web development using RIA technologies. Rather than offer full guided support, existing solutions generally provide only the resources necessary for accessible development. Therefore, the degree of accessibility is largely in the hands of the developer.

The immediate challenge is to build Web 2.0 development frameworks and technologies with a greater degree of WAI-ARIA support so that developers do not have to be experts in what constitutes accessibility. Work must continue in creating tools to support the authoring process in Web applications development. By making it easier to add accessibility, there is a greater opportunity for all users to benefit from the depth and breadth of Web content.

Acknowledgments

This research work is supported by the Research Network MAVIR (S2009/TIC-1542) (www.mavir.net), project TIN2007-67407-C03-01, and the Spanish Center of Captioning and Audio Description (www.cesya.com).

References

1. B. Gibson, "Enabling an Accessible Web 2.0," *Proc. Int'l Cross-Disciplinary Conf. Web Accessibility* (W4A 07), ACM Press, 2007, pp. 1-6.
2. M. Cooper, "Accessibility of Emerging Rich Web Technologies: Web 2.0 and the Semantic Web," *Proc. Int'l Cross-Disciplinary Conf. Web Accessibility* (W4A 07), ACM Press, 2007, pp. 93-98.
3. G. Lawton, "New Ways to Build Rich Internet Applications," *Computer*, Aug. 2008, pp. 10-12.
4. L. Moreno, P. Martínez, and B. Ruiz, "Disability Standards for Multimedia on the Web," *IEEE MultiMedia*, vol. 15, no. 4, 2008, pp. 52-54.
5. B. Regan, "Best Practices for Accessible Flash Design," Macromedia white paper, 2005; www.adobe.com/resources/accessibility/best_practices/best_practices_acc_flash.pdf.
6. R. Tretola, S. Barber, and E. Renaun, *Professional Adobe Flex 2*, John Wiley & Sons, 2007.
7. B. Gibson, "Dojo: An Accessible JavaScript Toolkit," IBM Human Ability and Accessibility Center; www-03.ibm.com/able/resources/dojo.html.
8. Z. Mikovec, J. Vystrcil, and P. Slavik, "Web Toolkits Accessibility Study," *SIGACCESS Accessibility Computing*, June 2009, pp. 3-8.
9. D. Lunn, S. Harper, and S. Bechhofer, "Combining SADIe and AxsJAX to Improve the Accessibility of Web Content," *Proc. Int'l Cross-Disciplinary Conf. Web Accessibility* (W4A 09), ACM Press, 2009, pp. 75-78.

10. Y. Borodin et al., "What's New?: Making Web Page Updates Accessible," *Proc. 10th Int'l ACM SIGACCESS Conf. Computers and Accessibility* (Assets 08), ACM Press, 2008, pp. 145-152.

Lourdes Moreno *is an assistant professor in the Advanced Databases Group in the Computer Science Department at Carlos III University of Madrid and a member of the Spanish Center of Caption and Audio Description. Her research interests include Web accessibility, development of accessible Web applications, and inclusive education. She is a member of the Spanish Association of Human-Computer Interaction and the Mavir consortium. Moreno received a PhD in computer science from Carlos III University of Madrid. Contact her at lmoreno@inf.uc3m.es.*

Paloma Martínez *is a member of the Advanced Databases Group in the Computer Science Department at Carlos III University of Madrid, where she teaches database design and management. Her research interests include Web accessibility and human language technologies, such as multilingual information extraction and retrieval in several domains, question answering, name-entity recognition, and temporal information management. She is a member of the Spanish Society for Natural Language Processing and the Mavir consortium. Martinez received a PhD in computer science from the Polytechnic University of Madrid. Contact her at pmf@inf.uc3m.es.*

Ana Iglesias *is a faculty member in the Computer Science Department at Carlos III University of Madrid. Her research interests include inclusive education, assistive technologies, adaptive intelligent educational systems, natural-language processing and information retrieval, and advanced database technologies. She is a member of the Mavir consortium. Iglesias received a PhD in computer science from Carlos III University of Madrid. Contact her at aiglesia@inf.uc3m.es.*

Belén Ruiz-Mezcua *is an adjunct vice chancellor of research and a lecturer in the Department of Computer Science at Carlos III University of Madrid, where she is the head of the Spanish Center of Caption and Audio Description. Her research interests include speech and speaker recognition, Web accessibility, and accessibility in products, services, applications, and technologies applied to persons with disabilities. Ruiz-Mezcua received a PhD in physics from the Telecommunications School at the Polytechnic University of Madrid. Contact her at bruiz@inf.uc3m.es.*

cn **Selected CS articles and columns are available for free at http://ComputingNow.computer.org.**

IEEE computer society

PURPOSE: The IEEE Computer Society is the world's largest association of computing professionals and is the leading provider of technical information in the field.
MEMBERSHIP: Members receive the monthly magazine *Computer*, discounts, and opportunities to serve (all activities are led by volunteer members). Membership is open to all IEEE members, affiliate society members, and others interested in the computer field.
COMPUTER SOCIETY WEBSITE: www.computer.org
OMBUDSMAN: To check membership status or report a change of address, call the IEEE Member Services toll-free number, +1 800 678 4333 (US) or +1 732 981 0060 (international). Direct all other Computer Society-related questions—magazine delivery or unresolved complaints—to help@computer.org.
CHAPTERS: Regular and student chapters worldwide provide the opportunity to interact with colleagues, hear technical experts, and serve the local professional community.
AVAILABLE INFORMATION: To obtain more information on any of the following, contact Customer Service at +1 714 821 8380 or +1 800 272 6657:

- Membership applications
- Publications catalog
- Draft standards and order forms
- Technical committee list
- Technical committee application
- Chapter start-up procedures
- Student scholarship information
- Volunteer leaders/staff directory
- IEEE senior member grade application (requires 10 years
- practice and significant performance in five of those 10)

PUBLICATIONS AND ACTIVITIES
***Computer*:** The flagship publication of the IEEE Computer Society, *Computer*, publishes peer-reviewed technical content that covers all aspects of computer science, computer engineering, technology, and applications.
Periodicals: The society publishes 13 magazines, 18 transactions, and one letters. Refer to membership application or request information as noted above.
Conference Proceedings & Books: Conference Publishing Services publishes more than 175 titles every year. CS Press publishes books in partnership with John Wiley & Sons.
Standards Working Groups: More than 150 groups produce IEEE standards used throughout the world.
Technical Committees: TCs provide professional interaction in more than 45 technical areas and directly influence computer engineering conferences and publications.
Conferences/Education: The society holds about 200 conferences each year and sponsors many educational activities, including computing science accreditation.
Certifications: The society offers two software developer credentials. For more information, visit www.computer.org/certification.

NEXT BOARD MEETING
23–27 May 2011, Albuquerque, NM, USA

IEEE

revised 10 Feb. 2011

Mobile Computing: Looking to the Future

Bill N. Schilit, *Google*

Innovations in mobile and embedded computing are transforming the way people access information and use network services.

It has been almost 20 years since technology visionary Mark Weiser predicted the proliferation of inch-, foot-, and yard-size computers ("The Computer for the 21st Century," *Scientific American*, Sept. 1991, pp. 94-104). To a large extent, Weiser's predictions have come to fruition, with smartphones (inch), tablets (foot), and Web-enabled televisions (yard) dramatically changing the Internet landscape.

Moreover, the "Internet of devices" is growing much faster than the desktop Internet, and many experts predict that more users will connect to the Internet from mobile and embedded devices than from desktop PCs within the next few years.

Smart devices are becoming pervasive in part because megahertz, megabits per second, and megabytes have become utilities much like water, gas, and electricity. I call these technologies "mega-utilities" because, like their Industrial Age counterparts, their vast economies of scale and broad-based availability have greatly benefited users.

For example, high-speed integrated circuits are making it economical to transform everyday objects including phones, cameras, and TVs into smart devices. Superfast wired and wireless networks make it economical to transmit audio, images, and high-definition videos to these devices. And the mega-storage trend is transforming computers from processors of compact, computer-friendly ASCII data to processors of people-friendly audiovisual data. These mega-utilities have been a major technological driver for the past decade.

IN THIS ISSUE

This issue presents a collection of articles covering innovations in mobile computing. In reviewing the submissions for this special issue, it became clear that more technology drivers are on the horizon, in particular embedded Web browsers and smartphones. Because of their broad-based availability on so many kinds of devices, browsers might well compete with native applications as a platform; they might also enable new interactions in the Internet of Things. Smartphones that flourished in the rich mega-utility ecosystem can now be viewed as their own technology drivers. Among many other benefits, they enable new classes of persuasive applications and help users take advantage of smart spaces.

Traffic from video, music, and e-books now dwarfs HTML, and app stores for custom-native applications are popping up everywhere. However, in "Reports of the Web's Death Are Greatly Exaggerated," the authors argue that the

Published by the IEEE Computer Society

Web will continue to be the premier software platform for the next generation of applications.

Smartphones are not only tools for communication and information but also for persuasion, helping individuals and groups achieve their everyday goals. The authors of "Eco-Feedback on the Go: Motivating Energy Awareness" explore how combining personal phones with principles from human-computer interaction research and psychology can promote energy awareness and help users reduce their carbon footprint.

The common Web browser has had phenomenal success as a viewer for the Internet's information space. Now, a new phase of the Web is developing in which everyday objects such as TVs, thermostats, appliances, and sensors of all types are being connected to the Internet. In "Browsing the Internet of Things with Sentient Visors," the authors explore what the browser will look like in the coming transition from the Internet of Information to the Internet of Things.

When smart devices meet smart spaces, there is an opportunity to merge the personalization afforded by one with the rich human-machine interfaces of the other.

As the authors describe in "Morphing Smartphones into Automotive Application Platforms," Nokia Research is encouraging the adoption of such a fusion between phone and automobile with Terminal Mode, in which the smartphone and car infotainment system work together to present a personalized, networked, high-fidelity application platform.

Innovations in mobile and embedded computing are transforming the way people access information and use network services. The articles included in this special issue offer a look at emerging trends in technology drivers and directions for the smart devices that have become such an integral part of our daily lives.

Bill N. Schilit *is a researcher at Google. Contact him at schilit@computer.org.*

COVER FEATURE

Reports of the Web's Death Are Greatly Exaggerated

Tommi Mikkonen, *Tampere University of Technology, Finland*
Antero Taivalsaari, *Nokia Research Center, Finland*

The development of entirely new types of Web-based software systems built to leverage the vast capabilities of the World Wide Web allows the use of dynamically downloaded applications and services from any type of terminal, including both desktop computers and mobile devices, implying radical changes in the ways people develop, deploy, and use software.

In a 2010 *Wired* magazine article, Chris Anderson and Michael Wolff provocatively claimed that "the (World Wide) Web is dead."[1] They based this claim on two main arguments. The first was that the amount of Internet traffic generated by webpage downloads has decreased dramatically over the years compared to the traffic generated by video and music downloads. The second argument was that users soon will no longer surf webpages with a traditional Web browser because, for the vast majority of Web services such as e-mail, Facebook, Twitter, and Skype, they will prefer custom-built native applications over open, unfettered Web browser access. Anderson and Wolff argued that the trend toward such apps will be even more evident in the mobile device space, in which—according to the authors—the use of Web browsers has already lost the battle against custom-built native apps.

Anderson and Wolff's first argument has been widely refuted in the press and on the Web (www.smallfish-bigpond.com/2010/08/wired). The observation that the amount of text-based Web traffic is insignificant compared to other types of traffic, while literally true, is highly misleading because, at the same time, webpage and Web browser usage has increased dramatically—nearly exponentially, in fact. Consequently, the notion that video and music downloads generate most network traffic is irrelevant and in no way confirms that the Web itself is dead. Furthermore, Web browsers initiate the vast majority of bandwidth-consuming data traffic on the Internet—that is, it can be claimed that to a substantial degree, the music and video downloads are just "subtypes" of the more general Web traffic. The fact that such traffic consumes more bandwidth than text-based HTML, Cascading Style Sheets, and JavaScript code downloads is self-evident and irrelevant from the viewpoint of attempting to prove the Web's demise.

Interestingly, Anderson and Wolff's argument about the transition from open, browser-based Web access to custom-built native apps has generated surprisingly little discussion. While it is tempting to think that the popularity of custom-built native applications for mobile devices would somehow confirm that use of the open Web will fail in the mobile device space,[2] it's likely that the trend toward such custom applications is only temporary. In fact, we believe that the use of open Web applications will eventually surpass the use of custom native applications on mobile devices.

TOWARD THE WEB AS A SOFTWARE PLATFORM

The World Wide Web is arguably the most powerful medium for information-sharing and distribution in the history of humankind. Therefore, it is not surprising that use of the Web has spread to many areas outside its origi-

 Published by the IEEE Computer Society

nal intended use. These days, everyday artifacts such as photos, music, videos, newspapers, and technical documents are widely available on the Web. Online banking and stock trading have become commonplace. Official documents that used to be difficult to access, such as municipal zoning regulations, government budget documents, and tax records, are now readily accessible on the Web. Various industries such as banking, financial services, book and electronics retailing, and music and video distribution have undergone dramatic transformations. Web-based services such as Facebook and Twitter have altered the meaning of social life.

More recently, the Web has begun to transform the political landscape as well, making it easy even for ordinary people to disseminate and discuss political information, and to arrange grassroots campaigns for various causes—much to the chagrin of people in power, who were previously able to centralize and control the flow of information in mass media.

It is quite possible that so far we have seen only the tip of the iceberg with regard to the impact the Web will eventually have on the broader society.

Software system deployment

Given that the majority of useful information is available on the Web, the Web browser has become the most common—or often the only—computer program that most people use. Therefore, although the Web browser was not originally designed to be a software platform, the Web has become more important as a deployment environment for various types of software systems. Systems that previously required substantial investments or distribution costs are increasingly available as services on the Web, as evidenced by the popularity of cloud computing systems and the recent success of companies such as Salesforce (www.salesforce.com).

From the software development viewpoint, the Web offers numerous benefits compared to conventional binary end-user software:

- *Instant worldwide deployment*. Web applications distributed over the public Web require no middlemen, distributors, or app stores. Once a webpage or application is deployed on the Web—whether in Tampere (Finland), Dar-es-Salaam (Tanzania), Kyoto (Japan), or Menlo Park (California)—it is instantly available to everyone connected to the Web.
- *No manual installations or upgrades*. Web applications are generally available simply by entering the correct URL in the Web browser. Unlike binary applications, they do not require manual installation. Furthermore, the user typically runs the latest version; HTML5-based applications also can support offline use without an active network connection.
- *Platform independence*. Web applications running in the Web browser are intended to be platform-independent; from the end user's viewpoint, the Web browser will effectively assume the role of an operating system. Although there are still various compatibility issues between different browsers,[3] these issues are relatively minor in comparison to developing software for currently available operating systems. Libraries such as Dojo (http://dojotoolkit.org) and jQuery (http://jquery.com) are available to hide some of the most annoying differences.
- *Ubiquitous, seamless access to data*. The Web enables data sharing on a scale that exceeds the capabilities of any previous systems. Basically, most of the important public data in the world is already available in digital form on the Web. The availability of such data has given rise to entirely new types of applications and mashups that dynamically combine data in unforeseen ways. Technical details related to data access are irrelevant; service providers are responsible for managing the software infrastructure, access interfaces, and databases, and for ensuring their compatibility.

From the software development viewpoint, the Web offers numerous benefits compared to conventional binary end-user software.

While the ability to share data is the Web's ultimate strength, these other aspects are critical from the perspective of application development and deployment. The capacity for worldwide deployment is incredibly powerful, especially when combined with the notion of "zero-installation software"—software that does not require any manual installation or upgrades. Such software can evolve at a pace that is almost beyond the imagination of conventional binary software developers, who are accustomed to offering "nano releases"—software upgrades occurring on a daily basis or perhaps even hourly.

Web-based software systems

Because the Web is not yet a uniform platform, discussions about Web-based software frequently refer to very different types of systems. Figure 1 illustrates three broad categories of Web-based software systems:

- *browser-based*—systems that use only the software that the standard Web browser offers;
- *plug-in-based*—hybrid systems in which the Web browser uses a custom runtime such as Microsoft Silverlight; and

COVER FEATURE

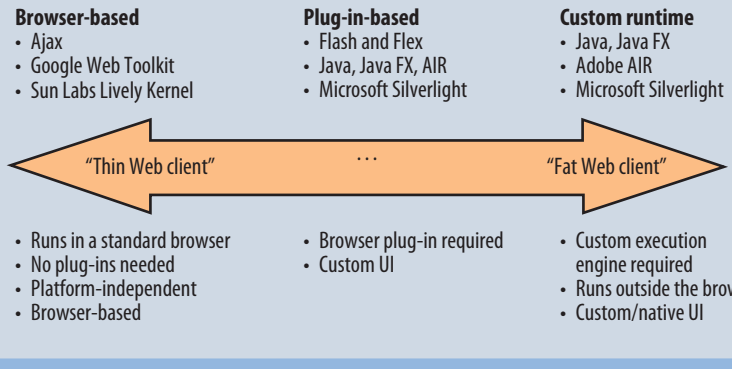


Figure 1. Three broad categories of Web-based software systems.

- *custom runtimes*—systems in which application execution requires a special runtime environment, such as Microsoft's Common Language Runtime (for Microsoft Silverlight) or the Java Runtime Environment; applications run outside the Web browser.

Browser-based systems offer the best application portability across different types of computers and operating systems, but these systems have various limitations related to the lack of low-level platform or device APIs and interactive graphics capabilities. The developers are essentially at the mercy of the limited APIs available in the standard Web browser. Custom runtimes can offer comprehensive graphics capabilities, extensive libraries, and customized development and debugging tools, but at the cost of application portability—basically, such applications do not run in a standard, open Web browser. In many companies, the installation of custom runtimes or additional browser plug-ins is explicitly prohibited for security reasons; this further limits the use of such systems for the development of truly portable Web applications.

A fourth category of Web-connected applications—*custom-built native apps*—is extremely popular today. For example, on Apple's iPhone and iPad devices, as well as on Google's Android devices, users typically apply custom-built native apps rather than a Web browser to access Facebook, Twitter, and many other popular Web services. Although they are not actually Web applications, these apps use the same network protocols to access the backend services as the Web browser does. Developers can use the native graphics libraries to customize the look and feel of such apps to fit the specific needs of the application and the device; these apps can also leverage device-specific features much more comprehensively than a pure Web application could.

The downside of such apps is that they are strictly platform-specific. Apps developed for the iPhone run only on the iPhone, so several different implementations—composed with different platform-specific tools—are needed to run the app on the Android, BlackBerry, Symbian, or other commonly used target platforms.[4] In many cases, a separate app is needed for each version of the target device. Such fragmentation is what effectively killed Sun's (now Oracle's) once highly successful Java ME platform.[5]

Unlike pure Web applications, a native app requires conventional installation. The user usually must download the application binary from a specific location, such as Apple's App Store (http://store.apple.com). To introduce new features, the user typically must download and install a new version or upgrade the application explicitly by device-specific means. This is clumsy and inconvenient for the user because the system might be unavailable while the download and upgrade are in progress.

Open Web application manifesto

Mozilla's recently published Mozilla Manifesto[2] declares that the Web is a global public resource that must remain open, accessible, interoperable, and secure. In line with these principles, the Web must be built around public, open standards rather than using proprietary technologies that serve specific business interests.

Among the four categories of application development technologies, the only one that fulfills the promise of truly open Web applications are browser-based systems—applications that require nothing more than a standard Web browser to run. Because they rely on proprietary technologies, additional plug-in components, or custom runtimes, the other approaches cannot deliver the Web's true flexibility as an application platform.

While native apps can offer considerable gains in terms of performance, usability, and API coverage in the short term, in the long run, convergence will lead to the development of open Web applications that eschew the use of native software—apart from the Web browser or a compatible runtime environment—in favor of pure Web technologies. The rationale is based on the power of qualities such as instant worldwide deployment; no manual installation or upgrades; platform independence; and ubiquitous, seamless access to data, with the ability to flexibly combine data from different sources.

In the near term, there are still numerous challenges in moving toward truly open Web applications. For example, the evolution of the Web browser itself is progressing rather slowly, and many of the APIs needed for developing comprehensive applications simply are not yet available. Further, using the Web in mobile devices poses special challenges.

The Web versus the mobile Web

The software industry is currently experiencing two parallel paradigm shifts—one toward Web-based software and the other toward Web-enabled mobile devices. The resulting convergence of desktop, mobile, and Web application development has already unleashed the development of entirely new types of Web-based software systems.

While the underlying needs to communicate and access information are the same in both the desktop and mobile environments, the ways people consume content and use applications with various terminals and devices in these environments are fundamentally different.

In the mobile space, the time span of users' actions is usually significantly shorter than in the desktop space. Users wish to perform rapid, targeted actions instead of participating in long-lasting sessions; actions must be simple yet focused, and they must be accomplished with ease, using minimal keystrokes or finger presses. Often, these actions are performed while the user is walking, driving a car, or otherwise distracted.

The different usage modalities and smaller screen sizes have a significant impact on application design; generic webpages geared toward laptop or desktop computer users are not usually ideal for mobile use. In addition, performance or network connectivity issues can make Web applications nearly unusable on mobile devices.

Given these challenges, it is not surprising that the mobile Web's evolution diverged from the general Web's evolution early on. The prime evidence of this early divergence is the now infamous Wireless Application Protocol (WAP), which was basically a parallel Web stack—including its own markup language, transport protocols, and browser solution—specifically designed for resource-constrained mobile devices. While the technical and commercial arguments behind WAP seemed sound, the assumption that people would want to use a custom browsing solution on mobile devices—separate from the general Web—turned out to be a huge mistake.

Somewhat surprisingly, companies that currently promote the use of proxy browsers are repeating the same mistake. Proxy browsers preprocess the information from the public Web and transliterate it into formats that fit better on the smaller screens of mobile devices. Although the technical arguments behind such proxy browsers are sound—decreased bandwidth costs, faster downloads, more optimal screen size usage, and so on—in reality, people usually prefer having access to the open, unfettered Web if such a choice is presented to them. However, many currently available low-to-mid-range mobile devices do not offer such a choice.

The WAP experience taught us that no matter how powerful the technical arguments in favor of custom Web runtime solutions specifically for mobile devices might be, the allure of the open Web is so powerful that people generally do not want to be limited by such solutions.

In the early 2000s, the mobile Web developer community expressed significant interest in Web widgets—prepackaged, purpose-built Web applications that can be installed and executed in mobile devices. The difference between Web widgets and general-purpose Web applications is that the installation formats and practices for Web widgets have been predefined specifically to resemble those of traditional (native) applications. Again, although there are sound technical arguments in favor of such solutions, Web widgets actually represent the worst of two worlds: they combine the use of conventional, rigid, explicit installation practices (familiar from binary applications) with languages and tools that usually result in applications that run slower and are often harder to maintain than their native counterparts.

While the technical and commercial arguments behind WAP seemed sound, the assumption that people would want to use a custom browsing solution on mobile devices turned out to be a huge mistake.

LIVELY WEB SYSTEMS FAMILY

During the past six years, we have constructed several systems for studying the use of the Web as a platform for real applications. Two of these systems—Lively Kernel (www.lively-kernel.org) and Lively for Qt (http://lively.cs.tut.fi/qt)—have been made available publicly as open source projects, and they have been used for application development on a wide range of systems from desktops and laptops to mobile devices.[3,6,7]

Lively Kernel, originally developed at Sun Microsystems Laboratories, is a Web application environment that supports desktop-style applications with rich graphics and direct manipulation capabilities, but without the installation or upgrade hassles associated with conventional desktop applications. The system can also function as an integrated development environment, making it self-supporting so that it can improve and extend itself dynamically in the same fashion as Smalltalk programming environments—yet the system requires only a Web browser for its execution. In fact, the entire Lively Kernel system is just a webpage.

Architecturally, Lively Kernel is built around three principal components:

- *JavaScript engine.* We used the JavaScript engine available in the Web browser as a fundamental building block for the system. The system itself, as well as all

Figure 2. Lively Kernel (left) and Lively for Qt (right) running on a Nokia N900 device.

the applications and tools in it, were written entirely in JavaScript.

- *Expressive graphics API.* A fundamental requirement for a modern Web programming environment is the availability of a powerful graphics API that provides support for direct drawing, direct manipulation, and a rich set of predefined graphical widgets. For improved efficiency and development flexibility, the APIs should support a procedural—not just declarative—development style. At the implementation level, the Lively Kernel graphics API is mapped either on the scalable vector graphics (SVG) API or the canvas API available in the Web browser.
- *Asynchronous HTTP networking.* All the networking operations in both systems are performed asynchronously using the Ajax XMLHttpRequest feature.[8] Using asynchronous networking is critical so that the system can perform all networking requests in the background without impairing its interactive response.

Although one of our original goals was to run Lively Kernel on mobile devices, our early attempts to accomplish this failed badly.[6] The primary reason for this failure was that the JavaScript engines and graphics capabilities of the mobile Web browsers available in the mid-2000s simply were not powerful enough for our needs. For these reasons, we started working on another system that would exploit the underlying resources in a more closely coupled fashion.

The resulting system—Lively for Qt—was the logical successor of Lively Kernel in the sense that it implemented a similar, fully interactive, malleable, and lively Web programming environment.[7] However, a key difference was that we built Lively for Qt explicitly to leverage Qt (http://qt.nokia.com), an existing, rich, mature, well-documented application framework. Qt supports a rich set of APIs, widgets, and tools that run on most commercial software platforms, including Mac OS X, Linux, and Windows. More important to us, Nokia had recently acquired Qt, with the intent that Qt libraries would eventually become available on all mobile devices produced by Nokia. This meant that we could rely on Qt libraries being available on Nokia devices without needing to separately download any additional plug-in components or custom libraries.

A similar concept, although independently developed, was later used in Palm's (now HP's) WebOS,[9] which introduced its own set of widgets that have been "baked into" a Web runtime, so that those widgets can be instantiated and used directly from Web applications.

Over the years, our research team members (http://lively.cs.tut.fi) have built numerous applications—games, desktop-style applications, applications that combine local data with Web content, as well as mashups that combine data from numerous websites into a single integrated experience—to demonstrate these systems' capabilities. When using Lively Kernel, the applications appear as individual windows inside the browser, while Lively for Qt applications are usually presented as desktop items or as full-fledged stand-alone applications, comparable to native applications. Figure 2 shows screenshots of some Lively Kernel and Lively for Qt applications running on the Nokia N900 mobile device.

In the mobile space, our research efforts culminated in a prototype GSM phone that we built in collaboration with a start-up company using JavaScript as the primary programming language for all the applications. This meant that any of the phone's applications could be downloaded dynamically over the air, run without explicit installation, and be updated transparently and seamlessly. However, in practice most of the system applications were still stored locally (cached in advance) for faster access.

REMAINING CHALLENGES

The 2010s will be an interesting decade in the software industry. We anticipate a major battle between proprietary native application technologies and the open Web, which will determine the software industry's future for years to

come. Although there will be considerable disagreement about the outcome of this "battle of the decade" in the near term, we would hesitate to bet against the Web in the long run because the distribution model for pure Web apps is so much more powerful. No installation is required, the application suites of computers and mobile devices can be entirely dynamic, and applications and upgrades can be made available worldwide instantly. Further, given the significant impact that the Web has already had on so many industries, it seems obvious that it will eventually revolutionize the distribution of software as well. In the larger scheme of things, the current hoopla around app stores and native apps is likely only a temporary solution.

Some important technologies and initiatives will improve the use of the Web as an application platform.

HTML5

The forthcoming HTML5 standard (www.w3.org/TR/html5) adds numerous features to the existing HTML standards. Although HTML5 is a general-purpose Web standard, many of the new features are aimed squarely at making the Web a better place for desktop-style Web applications. The additions to the earlier versions of the HTML specification include support for automatic application updates (via HTML5 cache manifests), offline applications, local storage, a 2D canvas graphics API, built-in audio and video support, drag-and-drop and context menus, and cross-document messaging. The next generation of HTML5-compliant Web browsers will also support several other important W3C standards, such as Web sockets, Web workers, and file system and geolocation APIs.

WebGL

Developed by Mozilla, the Khronos Group, and a consortium of additional companies including Apple, Google, and Opera, WebGL (www.khronos.org/webgl) is a cross-platform Web standard for a hardware-accelerated 3D graphics API. The main feature that WebGL brings to the Web is the ability to display 3D graphics natively in the Web browser without any plug-in components. Based on OpenGL ES 2.0, WebGL uses GLSL, the OpenGL shading language. WebGL runs in the HTML5 canvas element, and WebGL data is generally accessible through the Web browser's document object model interfaces.

In our opinion, WebGL is an exciting, but still mostly overlooked, standard that will dramatically change the perception of the capabilities of the Web browser specifically, and Web applications more generally. WebGL's most dramatic impact is that it will eliminate the "last safe bastion" of conventional binary applications. Although the majority of end-user software has already started migrating to the Web, so far it has been difficult to convince game developers to take Web-based software seriously. This is partly because suitable development APIs were not available, and partly because until recently the execution speed of Web-based software was inadequate for CPU-hungry gaming applications. However, with the recent introduction of high-performance JavaScript engines such as Google's V8 and Microsoft's Chakra, the situation has changed dramatically.

In combination with HTML5 and other new Web standards, the Web browser will have support for high-performance 2D and 3D graphics, Web sockets, video streaming, audio, Cascading Style Sheets, scalable vector graphics, Web workers, file handling, geolocation, and many other features. When combined with high-performance JavaScript engines, the availability of such capabilities will reset the expectations about what pure Web applications can and cannot do, allowing new types of applications that run in a standard Web browser. Google's Body Browser (http://bodybrowser.googlelabs.com) serves as an example of the new possibilities.

The main feature that WebGL brings to the Web is the ability to display 3D graphics natively in the Web browser without any plug-in components.

Standardization efforts

While HTML5 and related World Wide Web Consortium (W3C) standard activities will play a critical role in turning the Web into a compelling application platform, the feature set that an HTML5-compliant Web browser offers is still somewhat incomplete for real-world applications. For instance, the platform/device APIs that are under definition will still offer only limited access to features that are available in personal computers and mobile devices today.

Further, given that it defines the necessary standards in multiple, separate activities, the current standardization work likely will not yet result in a consistent platform. For example, Microsoft currently has no plans to support WebGL in its IE9 browser. In general, since many vendors have a vested interest in ensuring that their native software development platforms remain competitive, Web platform standards probably will remain one or two steps behind native application platforms.

We predict that another major round of standardization efforts will be necessary within the next five years or so to establish a more complete Web application platform beyond HTML5. A critical goal in that standardization activity will be to more comprehensively "virtualize" the underlying operating system and device capabilities, as well as ensure that the necessary security mechanisms are in place to access all the platform capabilities securely.

Open Web versus one Web?

The open Web concept does not necessarily imply one Web. As envisioned in the W3C Mobile Web Best Practices document,[10] one Web means that the same information and services should be available to users irrespective of the device they are using. While the overall goal is highly desirable, it is important to take into account the different input mechanisms, usage modalities, and screen size limitations on different types of terminals and mobile devices. Consequently, Web standards should not aim at establishing only "lowest common denominator" or "most common denominator" platforms. Instead, they should offer rich capabilities and support a broad variety of input mechanisms so that application developers and service providers can create compelling applications and services for all types of systems and devices. Plenty of interesting challenges still remain in this area.

The World Wide Web's massive popularity is turning the Web browser from a document viewing tool into a general-purpose host platform for various types of services, including desktop-style Web applications. HTML5-style Web applications require no installation or manual upgrades, and they can be deployed instantly worldwide without middlemen or distributors. Conventional binary applications are at a significant disadvantage compared to Web-based software that can be deployed instantly across the planet.

So far, several obstacles have hindered the development and deployment of full-fledged, truly interactive Web applications. These obstacles have been especially apparent in the mobile device space. Emerging standards such as HTML5 and WebGL will eliminate many of the limitations in this area. Although numerous challenges remain, we believe that the transition toward Web-based applications will eventually lead to the end of the binary end-user software era. In the future, the use of conventional binary programs will be limited to system software, while researchers will use Web technologies to develop the vast majority of end-user software—even for mobile devices.

References

1. C. Anderson and M. Wolff, "The Web Is Dead. Long Live the Internet," *Wired*, Sept. 2010, pp. 118-127; 164-166.
2. Mozilla, "The Mozilla Manifesto"; www.mozilla.org/about/manifesto.
3. A. Taivalsaari et al., *Web Browser as an Application Platform: The Lively Kernel Experience*, tech. report TR-2008-175, Sun Microsystems Laboratories, 2008.
4. T. Wasserman, "Software Engineering Issues for Mobile Application Development"; http://works.bepress.com/tony_wasserman/4.
5. R. Riggs et al., *Wireless Devices with the Java 2 Platform*, Micro Edition, 2nd ed., Addison-Wesley, 2003.
6. T. Mikkonen and A. Taivalsaari, "Creating a Mobile Web Application Platform: The Lively Kernel Experiences," *Proc. 24th ACM Symp. Applied Computing* (SAC 09), ACM Press, 2009, pp. 177-184.
7. T. Mikkonen, A. Taivalsaari, and M. Terho, "Lively for Qt: A Platform for Mobile Web Applications," *Proc. 6th Int'l Conf. Mobility Technology, Applications, and Systems* (Mobility 09), ACM Press, 2009, article no. 24.
8. D. Crane, E. Pascarello, and D. James, *Ajax in Action*, Manning Publications, 2005.
9. M. Allen, *Palm WebOS*, O'Reilly Media, 2009.
10. World Wide Web Consortium, *Mobile Web Best Practices* www.w3.org/TR/mobile-bp.

***Tommi Mikkonen**, a professor of computer science at Tampere University of Technology, Finland, pioneered mobile software development in Finland and has arranged numerous courses on software engineering and mobile computing. His research interests include cloud computing, Web programming, embedded systems, and mashup development. Contact him at tommi.mikkonen@tut.fi.*

***Antero Taivalsaari**, a Distinguished Engineer at Nokia Research Center in Helsinki, Finland, played a seminal role in the design of the Java Platform, Micro Edition (Java ME). His research focuses on Web application technologies and Web-based software development, especially for mobile devices. Contact him at antero.taivalsaari@nokia.com.*

Together, Mikkonen and Taivalsaari lead the Lively Web Programming Research Team at Tampere University of Technology. For further information and a full list of research team publications, see http://lively.cs.tut.fi.

IEEE ICDM 2011 Call for Papers

ICDM 2011: The 11th IEEE International Conference on Data Mining

Conference website:
http://icdm2011.cs.ualberta.ca/

Further Information:

- E-mail: ICDM2011Chairs@gmail.com
- *twitter: (@ICDMsocial) (#icdm2011)*
- *Facebook (IEEE ICDM International Conference on Data Mining)*
- *LinkedIn (ICDM 2011)*

The IEEE International Conference on Data Mining (ICDM) has established itself as the world's premier research conference in data mining. The 11th edition of ICDM (ICDM '11) provides a leading forum for presentation of original research results, as well as exchange and dissemination of innovative, practical development experiences. The conference covers all aspects of data mining, including algorithms, software and systems, and applications. In addition, ICDM draws researchers and application developers from a wide range of data mining related areas such as statistics, machine learning, pattern recognition, databases and data warehousing, data visualization, knowledge-based systems, and high performance computing. By promoting novel, high quality research findings, and innovative solutions to challenging data mining problems, the conference seeks to continuously advance the state-of-the-art in data mining. Besides the technical program, the conference will feature invited talks from research and industry leaders, workshops, tutorials, panels, exhibits and demos, a PhD Forum, and the ICDM data mining contest. In addition, there will be several awards including Best Paper Awards and Student Travel Awards.

Topics of Interest:

Data mining foundations: Novel data mining algorithms in traditional areas (such as classification, regression, clustering, probabilistic modeling, pattern discovery, and association analysis), Models and algorithms for new, structured, data types, such as arising in chemistry, biology, environment, and other scientific domains, Developing a unifying theory of data mining, Mining sequences and sequential data, Mining spatial and temporal datasets, Mining textual and unstructured datasets, Distributed data mining, High performance implementations of data mining algorithms.

Mining in emerging domains: Stream Data Mining, Mining moving object data, RFID data, and data from sensor networks, Ubiquitous knowledge discovery, Mining multi-agent data and agent-based data mining, Mining and link analysis in networked settings: web, social and computer networks, and online communities, Mining the semantic web, Data mining in electronic commerce, such as recommendation, sponsored web search, advertising, and marketing tasks.

Methodological aspects and the KDD process: Data pre-processing, data reduction, feature selection, and feature transformation, Quality assessment, interestingness analysis, and post-processing, Statistical foundations for robust and scalable data mining, Handling imbalanced data, Automating the mining process and other process related issues, Dealing with cost sensitive data and loss models, Human-machine interaction and visual data mining, Integration of data warehousing, OLAP and data mining, Data mining query languages, Security and data integrity,

Integrated KDD applications, systems, and experiences: Bioinformatics, computational chemistry, Ecoinformatics, Computational finance, online trading, and analysis of markets, Intrusion detection, fraud prevention, and surveillance, Healthcare, epidemic modeling, and clinical research, Customer relationship management, Telecommunications, network and systems management, Sustainable mobility and intelligent transportation systems

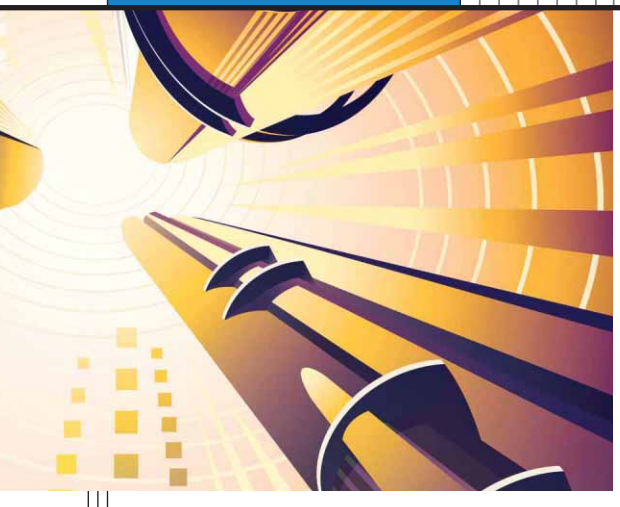
Important Dates

-June 17, 2011: ICDM paper submission (11:59pm Hawaii time)
-June 24, 2011: Demo and Tutorial submission
-August 5, 2011: Workshop paper submission
-September 16, 2011: ICDM paper notifications
-September 23, 2011: Workshop paper notifications

General Co-Chairs:
-Osmar Zaiane, University of Alberta, Canada
-Wei Wang, University of North Carolina at Chapel Hill, USA

Program Chairs:
-Jian Pei, Simon Fraser University, Canada
-Diane Cook, Washington State University, USA

COVER FEATURE



Eco-Feedback on the Go: Motivating Energy Awareness

Anna Spagnolli, Nicola Corradi, and Luciano Gamberini, *University of Padua, Italy*
Eve Hoggan, *University of Helsinki, Finland*
Giulio Jacucci, *Aalto University, Finland*
Cecilia Katzeff, Loove Broms, and Li Jönsson, *Interactive Institute, Sweden*

The EnergyLife mobile interface incorporates lessons from environmental psychology and feedback intervention to relay information from appliance sensors, offering a gaming environment that rewards users for decreased electricity consumption.

Energy saving has been a focus of human-computer interaction (HCI) applications for several years, and interfaces to promote user awareness of energy consumption are appearing on a variety of mobile devices. Commercial off-the-shelf energy-awareness products are now available as well as widgets and meters that promise to support the effective monitoring of household electricity consumption.[1,2] Both commercial products and research prototypes aim to surpass traditional monthly energy bills in providing feedback on consumption through a convenient medium, such as a cell phone or PDA.

As Figure 1 shows, these ecological-feedback technologies offer a range of information and opportunities to encourage rethinking energy consumption. Feedback, typically by device or use category, can include everything from information about instantaneous or historical household consumption to quizzes and tips. Media include desktop displays, mobile devices, or ambient interfaces, which address the periphery of users' attention by embedding information into surrounding objects.[1] An ambient interface could be a refrigerator door that changes color to convey information about an ideal inside temperature, for example.

Although many interfaces provide feedback, we see a distinct lack of attention to lessons from environmental psychology and feedback intervention, such as avoiding monetary measurements and connecting feedback to a savings target like "use 5 percent less energy per year."[2,3] To address design principles based on those ideas, we created EnergyLife, an application for mobile devices that presents information about household electricity consumption for appliances, heating, illumination, and other kinds of electricity consumers in the typical household—information that each device with an energy-aware interface now provides separately.

Our aim is to provide a synergy of actionable feedback without overwhelming the user with information as well as to entice the user to check progress in saving electricity. To meet the second goal, EnergyLife incorporates a gaming environment that rewards users with new features as they achieve certain goals.

In a study of households that used EnergyLife, electricity consumption decreased 5 percent over the previous year's figures (without EnergyLife) during the study's last month. Although EnergyLife is a research prototype, we believe it represents a solid step toward making the average user aware of household consumption and of the actions necessary to reduce that level. Collectively, these reductions could represent a significant decrease in overall electricity use and lead to similar models for other energy types. We are currently conducting another study to refine our results.

Published by the IEEE Computer Society

	Consumption		Granularity			Medium			Advanced features		
	Instant	History	Device	Household	Web/PC	Mobile	Ambient	Tips and quizzes	Community
EnergyLife	*	*	*	*	*	*	*	*	*
7000 Oaks	*			*			*		
UbiGreen	*			*		*			*
Powe Explorer	*			*		*			*
Efergy	*	*		*	*		*		
Navetes	*	*	*	*	*	*	*		
Agilewaves	*	*	*	*	*			*	
Onzo	*	*	*	*	*		*		*
Cent-A-Meter	*			*			*		
Microsoft Hohm		*	*	*	*			*	*
Tendril	*	*		*	*	*		*	*
Energy Joule	*			*			*		
EnergyHub	*	*	*				*		
Wattson	*	*	*		*		*		
Owl meter	*		*		*		*		
Energy Meter	*		*		*		*		
Ewgeco	*	*		*			*		
Google Pow	*	*		*	*				
Greenbox	*	*		*	*				
Manodo	*	*	*		*	*	*		

Figure 1. Features in current eco-feedback technologies. Both research prototypes (light orange) and commercial solutions (dark orange) aim to make users more aware of household energy consumption.

ELEMENTS OF AN ECOLOGICAL FEEDBACK SYSTEM

As Figure 2 shows, EnergyLife is layered atop our BeAware framework,[1] which takes data from a dedicated sensor on the household's fuse box as well as from wireless sensors in appliance plugs. In this way, the framework can measure electricity consumption both from the household overall and from individual appliances. The framework accommodates either off-the-shelf sensors such as Plugwise or wireless sensors that we designed to offer more precision and better connection. At present, BeAware processes data from eight sensors—refrigerator, television, washing machine, microwave, computer, and three generic sensors. The framework could easily accommodate more appliances, but we chose these five because they are common to the users in our evaluation study. The three generic sensors offer additional placement choices, such as coffee machines, stereos, and electric heaters.

Within the BeAware sensing layer is the BeAware base station (not explicitly shown), which processes all household data. It handles communication to all sensors and ambient interfaces, acts as an Internet gateway, and supports both BeAware's 433-MHz wireless network and commercial ZigBee-based networks. The framework's Internet services communicate with the user interfaces directly or through the base station to two embedded ambient interfaces: the light ambient interface (LAIT) and Watt-lite Twist, a flashlight-shaped tabletop projector.

Together these components make up the BeAware framework: the sensors and BeAware base station form the system, which feeds data to the EnergyLife mobile application and to the LAIT desktop and Watt-Lite projector interfaces.

Domain definitions

In our work, *energy* is essentially electric power, although we recognize that a household can use other energy types, according to region and building. A *household* is the smallest organic energy system in a residential sector that purchases electrical devices, monitors their use, and pays the bills to use them. An *electrical device* consumes electricity; it includes both home appliances and other consumption sources such as lights.

Within the household are *users*, who have various roles that involve the use of electrical devices.

Consumption and *conservation* are the two facets of energy awareness that acquire meaning through a reference point to discriminate them, such as a consumption

COVER FEATURE

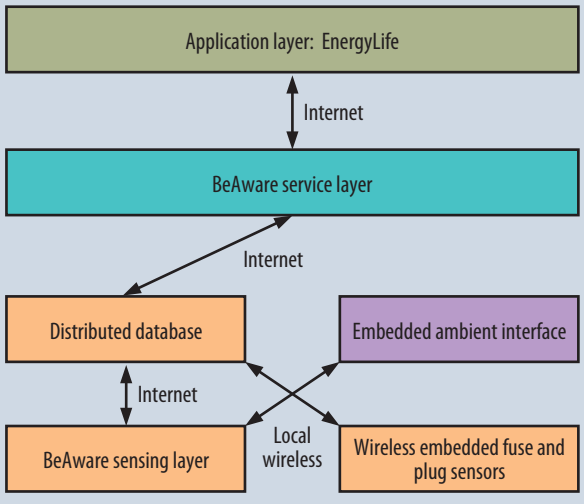


Figure 2. The BeAware framework. EnergyLife is one of several layers within the BeAware framework. Sensors connected to appliances through the BeAware base station (part of the BeAware sensing layer) feed data to ambient interfaces and a database, which in turn feed data to a service layer. EnergyLife uses this data to provide user feedback through EnergyLife. Ambient interfaces use the data to organize abstractions such as consumption pie graphs.

level. In our case, the reference point is a household's weekly consumption. The *baseline* reference point is the average electricity consumption during the calendar week prior to the current day. The baseline also helps highlight any significant consumption difference relative to the previous week about which the user might be unaware. Knowing that difference can help all household members plan for next week's consumption to compensate for excesses in the current week.

Feedback is information about the consequences of household actions that involve electricity consumption. Two basic types of feedback are consumption/saving information and smart advice tips.

Finally, *awareness* is the knowledge that users acquire about how and why to reduce waste by operating devices more efficiently, including hints that show a larger environmental impact. Awareness is a condition for conservation, so awareness tools are a necessary counterpart to feedback provision. EnergyLife provides this awareness by offering tips and quizzes.

Feedback and system requirements

Researchers from the behavioral sciences have studied how to provide feedback to persuade users to change behavior,[3,4] studying how feedback relates to behavior and the best format for providing feedback. Studies have found, for example, that it is easier to persuade users by addressing specific rather than general behaviors, since each behavior has its own reasons and constraints. Similarly, effective feedback should be actionable, demonstrating a way to fill the gap between current and desired actions. Thus, system designers should tailor feedback to the users' specific behaviors and profile. The goal must be constantly clear to the user, be tailored to the specific household, and provide instructions for how to achieve the goal.

Too much information might lead to overload and drop-outs,[3] so the displayed information should be self-explanatory; the best strategy is to have the interface present information in successive detail levels, not all at once.

Feedback format is also a critical design principle. Studies show that financial savings is not a long-term motivator relative to behavioral efficiency measures. Thus, designers should opt for historical feedback—insights that compare current and past distance to the goal—over feedback that compares a user's achievements with those of others. Because each household has its own constraints, such a comparison risks being irrelevant, flattering, or frustrating.

Other feedback and system requirements include

- *Nonintrusiveness*. The system should not disrupt everyday family habits by requiring additional measures and should not require substantial deviation from the user's intended activity.
- *Intuitiveness*. System interaction should encourage the correct user input without requiring a manual.
- *Energy efficiency*. The feedback system itself should be energy-efficient.
- *Privacy*. Devices that personalize feedback should address the possible risks to privacy and protect the user against them.
- *Sustained involvement*. The interface should evolve and reward improvements to keep the user enticed and motivated after curiosity about the system wanes.

Design principles

On the basis of these requirements, we evolved four main design principles for EnergyLife and the two ambient interfaces. The first was to provide near-real-time feedback of electricity consumption related to a reference point—the baseline overview of past consumption behavior. This comparison influences or informs the user's choice of future consumption behaviors. In addition, EnergyLife measures consumption in watts to show the effects of small behaviors, such as keeping a device on standby, and gives the cost of energy consumption to the environment when these behaviors accumulate, such as telling the user how many trees must absorb the carbon dioxide that an appliance has released. Awareness tools such as tips and quizzes related to the monitored devices direct users' efforts by reminding them of desirable conservation practices.

Another design principle recognizes the need to support engagement for sustained interaction and social incentives. To this end, we structured the EnergyLife game in several levels, offering players challenges and opportunities for knowledge building. Our aim was to allow long-term engagement with the application and to have users learn its features gradually. EnergyLife accomplishes this aim through game levels.

The third design principle was to ensure the interface's usability through intuitive and coherent navigation metaphors. In the mobile applications, we incorporated this principle through touch-screen technology using a card carousel that anchors the other interface elements. The ambient interface promotes usability by embedding consumption information into daily routines through familiar objects. For example, when a household member turns on the lights, they might dim slightly to indicate an increase in overall household electricity consumption.

Finally, EnergyLife acknowledges the actors responsible for electrical conservation in the household and relates them within the larger household nucleus. EnergyLife supports competition, discussion, and reflection inside the household by interlacing individual and family achievements and providing community tools. In this way, it aims to facilitate a shared process of awareness building by considering electricity consumption from a social viewpoint. If feedback follows requirements, such as being actionable, intuitive, tailored, and nonintrusive, users should be able to easily absorb the game into their daily life and routines.

COMMUNICATING CONSUMPTION

Persuasive eco-technology games like Eco Island[5] and PowerAgent[6] rely on self-reporting or aggregate measures and focus on competition among households. EnergyLife differs from these games in two ways. First, we customized the application[7] by acknowledging the household as the target and identifying the smaller, related units in which it is articulated. Second, EnergyLife uses real consumption data and feeds it automatically into the system, offering timely feedback that reflects any nuance of changed behavior.

Appliance cards

Figure 3 shows the EnergyLife mobile interface, which consists of a touch screen displaying a 3D card carousel. Each of eight cards represents an electrical appliance that a sensor is monitoring, and a household card shows overall consumption. Each appliance card shows that appliance's percentage consumption or savings relative to its average. Users can "flip" a card to see additional information and functionality for that appliance.

The front of the card displays a picture of the appliance, its current electricity consumption, and how much electricity the appliance is saving. The menu on the back of each

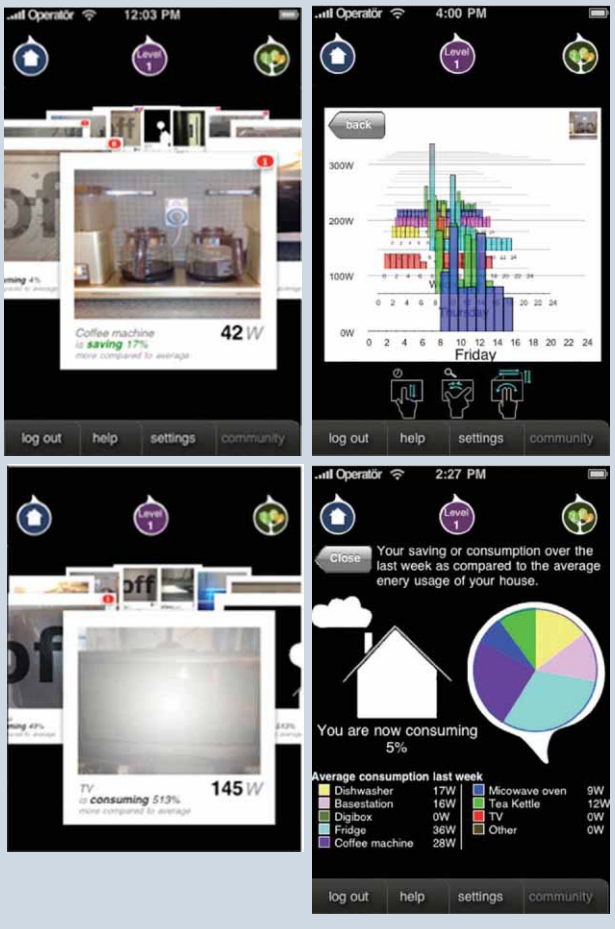
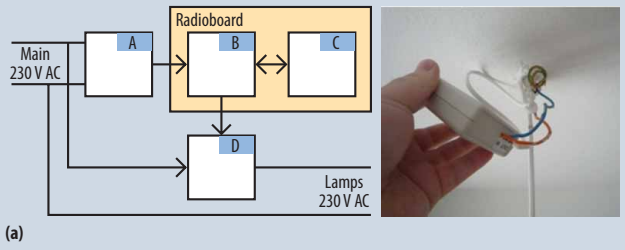


Figure 3. EnergyLife interface views. The main interface is EnergyLife's card carousel (upper left), which provides consumption history (upper right) and signals overconsumption through blinking cards (lower left). Users can also see a consumption breakdown per device. At present, eight sensor cards are available, each representing an electrical appliance; along with a ninth card (lower right) that reflects the household average on the basis of data from the dedicated sensor in the fuse box. The "average" in the card text refers to the average for that appliance. The small red dots on the upper right corner of some of the cards indicates that the user can read advice or take a quiz.

card lets users access advice, a quiz, or consumption history for that device. A household card shows consumption for the entire household. Users also see consumption history in the form of graphs (top right in Figure 3). Each row represents one day, and each column represents consumption for one hour. Users can browse the previous day's data using multitouch interaction with the graph itself.

To help the user conserve energy and gain awareness, the application provides advice tips and quizzes at certain times. Random advice appears once daily, taken from a tips repository. Smart advice is triggered by a condition connected to wasteful behavior, such as a television being

COVER FEATURE



(a)

(b)

(c)

Figure 4. Ambient interfaces. (a) The light ambient interface (LAIT) microcontroller circuit, shown as a diagram and completed controller, interfaces with an installed light and feeds information to the BeAware base station via Bluetooth. (b) The Watt-lite Twist resembles a large flashlight that the user can twist to the desired kilowatt-hours of consumption. (c) During the measurement period, the flashlight projects consumption progress, starting at the left with a white circle (the target amount of kilowatt-hours) or, which disappears as energy is consumed, indicating consumption breakdown by device. When the target amount of kilowatt-hours has been consumed, the white circle disappears, leaving a darkened shadow of a circle instead.

on for days. The implication is that someone forgot to turn it off and it has been running while no one is watching.

In advanced levels, advice tips are triggered for specific situations and give tailored feedback, such as "28 trees had to absorb the carbon dioxide your PC produced this week. Try to reduce its use by turning it off when you don't use it." The calculation of 28 trees stems from the exact number of kW consumed during the week by the monitored desktop PC and is triggered after this consumption exceeds a certain value.

Earning awareness points

EnergyLife users progress through game levels by earning awareness points. At Level 0, the application only processes measurements from the sensors to calculate the starting baseline. From the next level, users must acquire a certain number of awareness points, either by saving electricity for a specified time, reading advice, or correctly answering quiz questions. At subsequent levels, EnergyLife provides context-aware advice tailored to use parameters for that appliance, and unlocks a game community in which users can interact with EnergyLife users from other households and gain points for each message to improve their position with respect to the other players.

Unlike a direct competition between households, our approach uses gaming behavior to motivate individual change. Competition is to gain points in a game, not explicitly to save utility cost, although that is certainly a by-product.

Ambient interface components

The EnergyLife game includes specific goals and detailed feedback that require a reasonable amount of user effort and voluntary interaction. Two ambient interface components complement the game by sending messages to and receiving information from the BeAware base station, which they use to control consumption. Their aim is to alert users of consumption states, particularly overconsumption, and to indicate when it would be beneficial to look at the mobile application. Relative to the card carousel, which users voluntarily select, the components provide constant representations of electrical consumption and savings in a more abstract form.

Light ambient interface. Figure 4a shows the LAIT controller, which controls fixed household lighting and lamps to cue the user about the state of electricity consumption. The BeAware base station monitors incoming messages and requests from the LAIT service layer, and controls individual lamps in the household. A small printed-circuit board transformer and a low-power regulator power the LAIT.

If the household is on target, the lamp will turn on as it usually does. If the household is off target, the light will slowly brighten, from dark to full illumination.

Watt-lite Twist. As Figure 4b shows, the Watt-lite Twist is essentially an oversized flashlight that projects household electricity consumption as a pie chart during a specified number of kilowatt-hours, as Figure 4c shows.

We wanted Watt-lite Twist to provide a more intuitive representation than LAIT because studies show many people do not fundamentally grasp concepts used in

communicating energy. For example, an interview of 14 households transitioning to smart meters revealed only a vague understanding of a smart meter's specific functions. Indeed, none of those interviewed understood the concept of kilowatt-hour except as "something I pay for."[8]

Results from another study, this one of households using the Energy Aware Clock, also showed a failure to grasp the multifaceted concept of a kilowatt-hour.[9] One participant understood that appliance type affects the kilowatt-hours consumed, but she did not realize that a consumption calculation had to include the time and frequency of use.

With those studies in mind, we designed Watt-lite Twist to make it easier to understand the kilowatt-hour concept and how it relates to the electricity consumed in various activities using household appliances.

At the heart of the Watt-lite Twist is the Dingoo, a handheld game console modified with a Linux OpenWRT solution. A miniature laser projector attaches to the Dingoo, which responds to the movement of the front of Watt-lite Twist and starts the projection. Watt-lite Twist communicates with the BeAware base stations via Bluetooth. The station in turn communicates with the appliance sensors so that it can feed the Dingoo information about the electricity that individual appliances are consuming.

Starting Watt-lite Twist by winding its front part to a specific kilowatt-hour or fraction thereof gives users a feel for the duration of a kilowatt-hour—somewhat like starting an egg timer. Users can choose to twist the dial to anywhere from the minimum 0.1 kilowatt-hour to the maximum of 2 kilowatt-hours.

As the progression in Figure 4c shows, the first picture projected after the wind-up is a white circle. A black slice enlarges the allotted time. By sliding a switch at the top of Watt-lite Twist, the user can project a color pie chart showing the proportion of energy that each appliance is consuming, information that is based on communication with the BeAware base station. When the consumption of the target kilowatt-hour amount is complete, Watt-lite Twist issues an audio alert signal, and the projection shows a darkened circle with the time taken to achieve this target.

REMAINING CHALLENGES

In creating the components of our BeAware framework, we identified several open issues related to designing energy awareness systems: pervasive sensing, the need for multiple modes of interaction, the potential of pattern tracking in a particular context, and the tradeoffs of combining platforms.

Pervasive sensing

Integrating individual sources of information about household electricity consumption is a significant usability challenge. An effective monitoring system should provide an overview by highlighting power-eaters and tracking consumption routines, yet remain intuitive to use. The range of device types in a single household can result in large amounts of information, which can be confusing and overwhelming to view. To mitigate this complexity, feedback must be as simple as possible, such as having device categories. The lights category, for example, could consolidate information only about individual lights in a household.

Measuring power consumption to conserve energy is a way to recognize user activities.[10] A system can measure power for the entire house,[11,12] or for selected sockets.[13] Substantial challenges remain in fingerprinting electrical devices—characterizing electrical attributes, detecting use, and recognizing any use patterns—and breaking down overall household consumption according to different categories or devices. A breakdown using only one sensor seems to be unrealistic for the moment, as does the deployment of sensors for each electrical device. We chose a mixed approach for EnergyLife, using one sensor for the entire house to enable at least some fingerprinting for larger loads, as well as inserting sensors in key plugs for more targeted sensing.

Substantial challenges remain in fingerprinting electrical devices and breaking down overall household consumption according to different categories or devices.

Other challenges include choosing the best communication medium from current wireless options, such as ZigBee and low-power Bluetooth. Designers must carefully evaluate parameters such as reach, bandwidth, and scalability, since competing solutions remain despite federated efforts to propose standards.

Multimodality

Household energy-awareness systems should not disrupt the inhabitants' everyday routines. Offering nonintrusive feedback that does not tax the user is an important design challenge in mobile energy-awareness applications because mobile device size continues to shrink while the amount of important energy-consumption information is increasing. This combination poses a temptation to overload the display.

Another challenge stems from the dynamics of mobile device use. A user can be in a variety of use scenarios, but in each case, vision is the primary sense used to accomplish a task. However, many of these scenarios impose no restrictions on also using other senses. For

example, a user searching the Web for a restaurant address can still hear an alarm to remember to turn off the coffee machine.

In energy awareness, multimodal interaction is a promising way to reduce display complexity. Devices like the iPhone have built-in vibration motors and audio speakers that can provide energy-awareness information without requiring the user to look at the device. Moreover, a simple vibration might use less energy than displaying the same information on the device's screen.

Context awareness

Detailed energy sensing is an opportunity to track user activities. Additional information from other sensors (temperature, lights, mobile phone presence and use) can provide data to recognize events and situations and trigger different system responses. Building on this smart advice idea, systems can track household energy-use patterns by identifying use profiles and their effect on energy conservation.

Future work in context awareness should include developing catalogs, models, and classifications of meaningful situations along with recognition algorithms, taking into account that households differ significantly in their energy-use habits.

Combined platforms

Ubiquitous computing allows system interaction through a variety of modalities and interaction platforms. Mobile devices and ambient (embedded) interfaces can provide energy-efficient user interaction. Smart phones are on most of the time, and placing sensors on installed lights doesn't require large amounts of additional energy. Rather, the main challenge is how to synchronize platform components. Ambient interfaces that require embedded computing pose additional challenges such as energy-efficient connectivity and scalability, for example.

Our ambient interface uses the same communication platform as the wireless sensors, and the latest sensor we developed uses the same component to track the use of lights and communicate the information to users. Our vision is that all devices, including power plugs, will have electricity sensors and the ability to communicate with each other.

Mobile and embedded computing is already pervasive, so a logical next step is to exploit the technology for the betterment of society or the environment. EnergyLife harnesses the ubiquitous nature of mobile devices to address energy awareness by helping users understand how their energy-consumption habits contribute to or degrade environmental health. By providing a shared awareness of energy consumption, real-time feedback, and a gaming environment that helps sustain interaction, EnergyLife overcomes many of the problems encountered in other energy-awareness system designs.

EnergyLife is not just a concept. We tested it several times before releasing the first prototype.[3] We then installed the working system in five research-team and eight lay-user households in northern and southern Europe, all of which completed a longitudinal three-month trial. A second trial with another 10 households is in progress, and we anticipate results by August 2011.

A preliminary analysis of data from the first working-system trial shows decreased consumption from the trial's start to end, and most participants agreed that they had changed their energy-consumption habits. Comparing consumption during the trial's last month with the same period in the previous year (which we inferred from energy bills) also shows a 5 percent decrease in overall household consumption.

Naturally, issues remain, such as pervasive sensing and user empowerment, as well as the multimodal integration of different displays with different feedback modalities. However, our preliminary studies with EnergyLife are encouraging. The benefits of creating mobile and embedded energy-awareness systems are countless, but, undeniably, the most important contribution is the shared awareness of how individual actions influence environmental health. C

Acknowledgments

This work was cofunded by the European Union as EU FP7/ICT-2007.6.3, project 224557; http://energyawareness.eu. All authors contributed equally to this article.

References

1. C. Björkskog et al., "A Framework for Residential Services on Energy Awareness," *Proc. 12th ACM Int'l Conf. Ubiquitous Computing* (UbiComp 10), ACM Press, 2010, pp. 361-362.
2. J. Froehlich, L. Findlater, and J. Landay, "The Design of Eco-Feedback Technology," *Proc. Conf. Computer-Human Interaction* (CHI 10), ACM Press, 2010, pp. 1999-2008.
3. G. Jacucci et al., "Designing Effective Feedback of Electricity Consumption for Mobile User Interfaces," *PsychNology J.* vol. 7, no. 3, 2009, pp. 265-289.
4. C. Fischer, "Feedback on Household Electricity Consumption: A Tool for Saving Energy?" *Energy Efficiency*, vol. 1, no. 1, 2008, pp. 79-104.
5. T. Nakajima et al., "Reflecting Human Behavior to Motivate Desirable Lifestyle," *Proc. Conf. Designing Interactive Systems* (DIS 08), ACM Press, 2008, pp. 405-414.
6. A. Gustafsson, M. Bång, and M. Svahn, "Power Explorer: A Casual Game Style for Encouraging Long-Term Behavior Change among Teenagers," *Proc. 5th Conf. Advances in Computer Entertainment Technology* (ACE 09), ACM Press, 2009, pp. 182-189.
7. H.A. He, S. Greenberg, and E.M. Huang, "One Size Does Not Fit All: Applying the Transtheoretical Model to Energy

Feedback Technology Design," *Proc. 28th Conf. Human Factors in Computing Systems* (CHI 10), ACM Press, 2010, pp. 927-936.
8. I. Lindstedt et al., "Vad Hände? Intervjustudie Om Kundernas Förväntan På De Nya Elmätarna Och Deras Kunskaper Om El Och Energi" ["What Happened? An Interview Study about Customers' Expectation for New Meters and Their Knowledge about Electricity and Energy"], 2008 (in Swedish); http://www.elforsk.se/rapporter/?rid=08_76.
9. L. Broms et al., "Coffee Maker Patterns and the Design of Energy Feedback Artifacts," *Proc. Conf. Designing Interactive Systems* (DIS 10), ACM Press, 2010, pp. 93-102.
10. J. Fogarty, C. Au, and S.E. Hudson, "Sensing from the Basement: A Feasibility Study of Unobtrusive and Low-Cost Home Activity Recognition," *Proc. 19th ACM Symp. User-Interface Software and Technology* (UIST 06), ACM Press, 2006, pp. 91-100.
11. S. Patel et al., "At the Flick of a Switch: Detecting and Classifying Unique Electrical Events on the Residential Power Line," *Proc. 9th ACM Int'l Conf. Ubiquitous Computing* (UbiComp 07), Springer-Verlag, 2007, pp. 271-288.
12. G. Hart, "Nonintrusive Appliance Load Monitoring," *Proc. IEEE*, Dec. 1992, pp. 1870-1891.
13. X. Jiang et al., "Design and Implementation of a High-Fidelity AC Metering Network," *Proc. Int'l Conf. Information Processing in Sensor Networks* (IPSN 09), ACM Press, 2009, pp. 253-264.

Anna Spagnolli *is an assistant professor and scientific coordinator of the Human Technologies Laboratories in the Department of General Psychology at the University of Padua, Italy. Her research interests include human-computer interaction and presence in mediated environments, particularly the use of qualitative methods. Spagnolli received a PhD in personality and social psychology from the University of Padua. She is a member of the ACM; the International Society for Gesture Studies; the International Association of CyberPsychology, Training, and Rehabilitation; the International Society for Presence Research; and the International Society of Conversation Analysis. Contact her at anna.spagnolli@unipd.it.*

Nicola Corradi *is a PhD student in cognitive science at the University of Padua. His research interests include human-computer interaction, energy saving, and cognitive neuroscience. Corradi received an MS in psychology from the University of Padua. Contact him at nicola.corradi@unipd.it.*

Luciano Gamberini *is an associate professor at the University of Padua, where he is also Rector's Delegate for the International Ranking System and director of the Human Technology Laboratories. His research interests include social and cognitive ergonomics and mediated communication. Gamberini received a PhD in experimental psychology from the University of Padua. Contact him at luciano.gamberini@unipd.it.*

Eve Hoggan *is a postdoctoral researcher at the University of Helsinki and at the Helsinki Institute for Information Technology, Finland. Her research interests include human-computer interaction, multimodal interaction, and novel interaction techniques. Hoggan received a PhD in cross-modal audio and tactile interaction from the University of Glasgow, UK. She is a member of IEEE and the ACM. Contact her at eve.hoggan@helsinki.fi.*

Giulio Jacucci *is coordinator of the BeAware project (www.energyawareness.eu) at Aalto University, Finland, a professor in the Computer Science Department at the University of Helsinki, and director of the Network Society Research Program at the Helsinki Institute for Information Technology. His research interests include human-computer interaction in ubiquitous computing with a focus on multimodality and surface and adaptive computing. Jacucci received a PhD in information processing science from the University of Oulu, Finland. Contact him at giulio.jacucci@helsinki.fi.*

Cecilia Katzeff *is scientific director and senior researcher in the Department of Energy Design at the Interactive Institute, Sweden. Her research interests include human behavior in relation to interactive media; ubiquitous computing, particularly as related to energy use; and, more recently, the design of digital artifacts and user studies to encourage energy awareness and energy-conserving behavior in the home and workplace. Katzeff received a PhD in psychology from the University of Stockholm. Contact her at cecilia.katzeff@tii.se.*

Loove Broms *is an interaction designer in the Department of Energy Design at the Interactive Institute and a PhD candidate in interaction design at the Royal Institute of Technology, Stockholm. His research interests include the use of interaction design in increasing energy-use awareness, particularly in influencing a user's energy-use habits, and the design of artifacts and environments that reflect the present and engage users. Contact him at loove.broms@tii.se.*

Li Jönsson *is an interaction/product designer in the Department of Energy Design at the Interactive Institute and a PhD candidate in design at the Danish Design School, Copenhagen, Denmark. Her research interests include finding new ways to engage people in potentially serious issues by rethinking how products embody values. Contact her at ljo@dkds.dk.*

COVER FEATURE

Browsing the Internet of Things with Sentient Visors

J. Antonio Garcia-Macias, Jorge Alvarez-Lozano, Paul Estrada-Martinez, and Edgardo Aviles-Lopez
CICESE Research Center, Mexico

Unlike the traditional Internet, the emerging Internet of Things constitutes a mix of virtual and physical entities. A proposed IoT browser enables the exploration of augmented spaces by identifying smart objects, discovering any services they might provide, and interacting with them.

Before the advent of the World Wide Web there were computing systems that, with varying degrees of success, let users navigate the Internet's information space. Archie, for example, allowed access to data repositories via the File Transfer Protocol, while Gopher provided menu-based browsing of topics. The Web, however, accomplished what its predecessors could not: it unified information naming and access through URLs and allowed information browsing through a consistent paradigm (hypermedia) as well as a consistent interface (the browser).

Web 2.0 went beyond organizing and searching information to providing facilities for user-generated content and social networking. More recently, Tim Berners-Lee's seminal vision of having machine-readable metadata to allow agents and other software entities to access the Web more intelligently is becoming realized in the Semantic Web.

A fourth stage in the Web's evolution is taking shape as everyday objects are being connected to the Internet; this includes various types of sensors, household appliances, ambient devices, actuators, and so on. In this Internet of Things (IoT), developers augment common objects with technology—for instance, an embedded microcontroller and radio frequency unit—to provide computing and communication capabilities, thereby creating "smart objects."

We believe that it is possible to extend the current paradigm of a universal information browser to the IoT. However, any such browser must consider the important differences between the Internet and the IoT, which encompasses a combination of virtual and physical entities rather than solely Web documents. These differences led us to propose a design space for IoT browsers called *sentient visors* and to implement a prototype system.

BROWSING THE INTERNET OF THINGS

As the IoT becomes a reality, some questions arise: How do users navigate through the myriad information generated not only by people but also by machines? Is the paradigm of a universal information browser still valid? If so, what characteristics should such a browser have? What is an appropriate model for interacting with smart objects in our surroundings?

To answer these questions, it is important to compare the IoT with the traditional Internet. Table 1 highlights the key differences.

Proximity

Where a Web document resides—on a server down the hall or across the world—is usually irrelevant. Web caches might include the geographic location of information, but in general users do not care about this information. In contrast, people might want to know what smart objects in their vicinity are telling them, so the locality of generated information in the IoT becomes relevant.

Published by the IEEE Computer Society

Table 1. The Internet versus the Internet of Things.

Characteristic	Internet	Internet of Things
Proximity	Irrelevant	Local objects generate information
Physicality	Physical objects can be connected	Physical objects should be connected
Interactivity	Limited to two-dimensional GUI	Physical objects plus information can add other dimensions to interactions
Semantics	Necessary for better searching	Necessary for better searching and interactions
Service discovery	Web service discovery	Web service discovery plus location

Physicality

Connecting devices to the Internet is not new. Vending and coffee machines have been networked since the 1970s, and for quite some time telemetry equipment has generated data accessible on the Web via CGI (common gateway interface) and other scripts. However, the IoT is all about connecting objects and augmenting them with computing and communication capabilities, so physical entities play as important a role as virtual ones (software).

Interactivity

Although most Web browsers can render user interfaces that go beyond the simple interactive forms of plain HTML—for instance by using Ajax, Java, or Flash technologies—they retain traditional GUI metaphors to facilitate data entry and manipulation. In the IoT, both physical space and information are part of the browsing experience, adding new dimensions to interactions. Examples include direct manipulation of objects and taking advantage of proximity to perform semantic zooms.

Semantics

The Semantic Web exploits the semantics associated with data to improve searches and interactions between applications. The IoT goes further, using semantics about smart objects to identify their properties (metadata), provide appropriate interactions, and infer context about them. This in turn is critical to constructing proactive, context-aware systems.

Service discovery

An essential component in the Semantic Web is software agents that automatically discover and use one another's services. In the IoT, physical entities can also be abstracted to provide services, but their physical location has special relevance as a user might be interested in services provided by nearby objects.

SENTIENT VISORS

Because the IoT comprises physical as well as virtual space, an IoT browser should be aware of its surroundings, capturing and interpreting information emitted by smart objects. Sentient visors enable the exploration of augmented spaces by identifying smart objects, discovering any services they might provide, and interacting with them.

For more than three years we have been exploring the sentient visor design space and its supporting infrastructure.[1-4] Our design methodology has included the creation of domain-specific prototypes (such as ambient assisted-living systems), topics for focus groups to provide feedback, and user interviews and questionnaires. These experiences, combined with analysis of previous research and input from colleagues in the field, have led us to identify six design dimensions, shown in Figure 1, that should be considered when developing sentient visors.

Table 2 summarizes the design dimensions, with sample values and examples. An instantiation of a sentient visor should include all dimensions, but visors can differ depending on the values they have along the dimensions. For example, for the multimodal information dimension, one visor might render images on top of a video stream, while another might also provide audio output. In the case of the mixed-reality user interface dimension, one browser might feature an augmented reality (AR) display, while another offers augmented virtuality capabilities using real-world entities in virtual worlds, possibly via a head-mounted display.

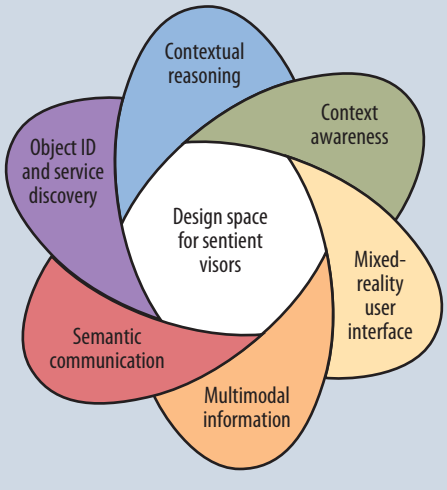


Figure 1. Six design dimensions should be considered when developing sentient visors.

Table 2. Design dimensions for sentient visors.

Dimension	Some values	Examples
Contextual awareness	Tagging	"This patient needs further analysis"
	Presenting information and services	Food chart with nutritional values Representation options Indoor guide
	Automatic execution of a service	Concealing/revealing information according to a user's role Opening doors
Object identification and service discovery	Visual identification	Bar codes Quick response (QR) codes Object recognition
	Electronic signals identification	Radio frequency identification (RFID) Near field communication (NFC) Global Positioning System (GPS)
	Service discovery	Directory-based Ad hoc
Contextual reasoning	Thematic	"Medication intake" "Watering pot"
	Spatial	"Outside" "Meeting room" "Near the stairs"
	Temporal	"Afternoon" "From 10 to 11 am" "Next 5 minutes"
Semantic communication	Representation	Data format Engineering units Ranges
	Visualization	Text Graphics Contextual choices
	Interaction	Choose interaction according to discovered object's functions and capabilities
	Self-discovery	Device the visor is running on User profile
Mixed-reality user interface	Ranging from augmented reality (AR) to augmented virtuality	Virtual guide with AR on smartphone Engine assembly guide using a head-mounted display
Multimodal information	Input	Buttons Multitouch screen Direct object manipulation Voice
	Output	AR display Voice Actuator

Contextual awareness

The primary purpose of sentient visors is to allow browsing information spaces that constitute a mix of virtual and physical entities. A visor must obtain contextual information gathered by accelerometers, microphones, and other sensors available in mobile devices to determine if any entities are augmented with information, what this information means, and how to interpret and represent it. Contextual awareness is thus an important design feature of sentient visors.

Anind Dey[5] has proposed three categories of features that a context-aware application should support: presentation of information and services to a user, automatic execution of a service for a user, and tagging context to information to support its retrieval.

Object identification and service discovery

To identify smart objects within the environment and the services they provide, sentient visors should reside in smartphones, tablet computers, and other physical devices

incorporating video cameras, graphics processing abilities, data communication interfaces, and possibly expansion interfaces such as USB ports to connect new hardware capable of sensing.

A smart object will typically have some kind of tag such as radio frequency identification (RFID), a barcode, or a quick response (QR) code to enable its identification. A more sophisticated sentient visor could also exploit object recognition technologies.

Once it has identified a smart object, a sentient visor must be able to discover what kind of service it provides—for instance, temperature readings or location determination. Thus, service discovery is an important design consideration as well as object identification.

Contextual reasoning

Identifying a smart object and discovering the services it provides are necessary first steps to interact with it, but context-aware systems should also be anticipatory and proactive. At present, browser interaction commonly occurs via direct text input, touch gestures, and clickable areas. However, a sentient visor must also react to more subtle inputs such as pointing to places, the current time, the user's gender, and an object's owner that are not important for a traditional browser. Sentient visors should therefore support contextual reasoning.

Semantic communication

Users can obtain primary context from data captured by sensors, but obtaining derived context to receive relevant information or services requires making inferences about the data. Semantic communication with smart objects allows a sentient visor to go beyond simple object identification and service discovery to determine what the object is trying to convey. This enables better inferences and thus richer interactions. Semantic information can include, for example, the type and format of readings made by the object, the visual representations it uses (for example, pictographs or scientific plots), and characteristics of the services it provides.

Mixed-reality user interface

To navigate the IoT's integrated physical and virtual space, users with little technical knowledge or experience with different types of mobile devices need a mixed-reality interface.[6] The AR paradigm is ideal, as an AR device can superimpose visual information from a smart object at which it is pointed on top of a video feed of the surrounding environment. As the "Using AR Displays to Navigate Augmented Environments" sidebar describes, researchers have developed many AR systems to augment the user's perception of the real world.

USING AR DISPLAYS TO NAVIGATE AUGMENTED ENVIRONMENTS

Augmented reality has been used since the mid-1990s to navigate physical spaces and access data about objects in those spaces. Early AR systems such as a prototype developed at Columbia University[1] and NaviCam[2] enabled a user to explore the environment and see information about objects in the line of sight. However, these systems could not find and use services provided by those objects, nor could they infer, for example, when to play an audio stream rather than display an animated graphic according to user needs and contextual data.

Joseph Newman, David Ingram, and Andy Hopper[3] designed an indoor system to provide context-sensitive visualization using head-mounted displays or mobile devices such as PDAs. Their prototype relied on an in-building positioning system to determine the location of smart objects, persons, and other entities and construct a software model of reality.

Alex Wright[4] proposed using next-generation smartphones to run context-aware applications that could track users' behavior, anticipate their intentions, and provide information about the immediate environment. He envisioned these applications using AR as the primary means for user interaction and visualization.

Recently introduced smartphone-based AR browsers such as Layar (www.layar.com), Nokia Point & Find (http://pointandfind.nokia.com), and Wikitude (www.wikitude.org) cannot be considered sentient visors as they do not comply with all of the design requirements. Most notably, these browsers do not have contextual reasoning capabilities, their context awareness is very limited (primarily reduced to location awareness), and they completely lack semantic communication with objects in the environment.

Some recent projects are using AR to navigate the IoT. For instance, Antto Ajanki and colleagues[5] have developed an AR platform that infers relevant information about objects, people, and the environment based on gaze and speech signals. However, it differs from the concept of sentient visors in that it does not use service abstractions for objects, it does not consider location information, and there are no semantics to support making inferences.

References

1. S. Feiner et al., "A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment," *Proc. 1st Int'l Symp. Wearable Computers* (ISWC 97), IEEE Press, 1997, pp. 74-81.
2. J. Rekimoto and K. Nagao, "The World through the Computer: Computer Augmented Interaction with Real World Environments," *Proc. 8th Ann. ACM Symp. User Interface and Software Technology* (UIST 95), ACM Press, 1995, pp. 29-36.
3. J. Newman, D. Ingram, and A. Hopper, "Augmented Reality in a Wide Area Sentient Environment," *Proc. IEEE and ACM Int'l Symp. Augmented Reality* (ISAR 01), IEEE CS Press, 2001, pp. 77-86.
4. A. Wright, "Get Smart," *Comm. ACM*, Jan. 2009, pp. 15-16.
5. A. Ajanki et al., "An Augmented Reality Interface to Contextual Information," *J. Virtual Reality*, 15 Dec. 2010, pp. 1-13.

Multimodal information

Given the variable behavior and environments that could be present in different scenarios, sentient visors need to manage multimodal information, adapting input and output to user preferences. For instance, if a user is visually impaired, the visor could use

Figure 2. An augmented pot indicates that the plant within needs watering through an iconic depiction (left) and by showing soil temperature and humidity values (right).

auditory, instead of only visual, means to communicate information.

SENTIENT VISOR PROTOTYPE

We have developed UbiVisor, a prototype sentient visor system that includes a browser and the supporting infrastructure. To illustrate its use, we consider a simple but representative application in which a plant pot is equipped with a small device that senses ambient temperature and soil humidity and wirelessly transmits the results.

A person with a UbiVisor-equipped device receives signals when passing by an augmented pot and is alerted if any of the sensed parameters lie outside the appropriate value ranges. For example, if the soil is too dry, the system sends an alert indicating that the humidity is below the accepted threshold; if the user points the device at the pot, a thirsty face icon superimposed on top of the pot visually represents this situation, as the left side of Figure 2 shows.

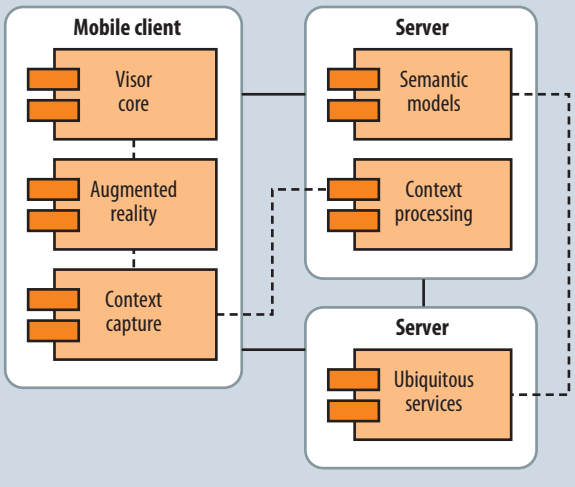


Figure 3. UbiVisor system architecture.

In this example, the icon adequately conveys the plant's need for water. However, what if, as the right side of Figure 2 shows, the user wants to know the exact soil humidity or temperature, or wants to choose among different forms of representing the information—for example, a bar thermometer, gauge, or some other type of meter?

This is where semantics come into play, as metadata annotations let software agents share, reuse, understand, and make inferences about contextual information. Enabling this behavior requires developing a knowledge representation model that defines objects, concepts, relations, and other entities in the domain of interest. The UbiVisor prototype currently uses Resource Description Framework (RDF) ontologies, but we are considering other options.

Continuing with the plant pot example, the sentient visor's discovery module could be hard-coded to scan the right radio frequencies, decode the data packets being emitted, identify the sending object, and interpret the information by converting it to temperature and humidity values. However, anyone wishing to interact with the augmented pot must carry this hard-coded version of the visor. This situation is far from optimal; it would be preferable to have a visor capable of scanning the radio waves and interpreting whatever the object is sending, even if it is a new type of object. As in the case of displaying information, semantics holds the key to this problem.

SYSTEM ARCHITECTURE

Figure 3 shows the general UbiVisor architecture, which is currently implemented in Java. The system uses the Processing language (http://processing.org) to render 3D graphics, the Jena framework (http://jena.sourceforge.net) to manipulate RDF documents, and the ZXing barcode

image processing library (http://code.google.com/p/zxing) to decode QR codes.

Mobile client

Residing on a mobile device, the *context capture* module discovers nearby smart objects abstracted as services; it also manages contextual information such as user and device profiles, location information, object identifiers, and other variables useful for determining context. The context capture module identifies an object via an RFID or a QR code. The decoded information contains a uniform resource identifier (URI) pointing to an RDF resource; this resource uses ontologies and rules that are useful for interpreting the meaning of the information sent by the smart object, as well as for rendering the information in the chosen format on the device's display.

Also on the mobile client, the *augmented reality* module receives a semantic model indicating the type of graphic element to be rendered and a URI containing any image to be displayed. Overlaying images on a video feed is the most common way to augment reality, but additional modalities include using audio or other sensorial outputs. The *visor core* module initiates and coordinates both the AR and context capture modules.

The UbiVisor software within the mobile client has all the necessary elements to detect the presence of smart objects, initiate their identification, and interact with them, including rendering information using AR. However, a sentient visor is essentially a distributed system that involves several networked elements. For instance, once the system detects a QR code tag, it obtains a URL and consults an external server to obtain the associated resource. Communication between networked entities is conducted via HTTP using the Representational State Transfer (REST) software architecture, an approach proposed by others researching the IoT.[7]

Servers

An external server contains a *semantic models* module and a *context processing* module. Both modules collaborate to receive any information sent from the context capture module on the mobile client, create a semantic model, and, based on this model, make inferences about the type of information to send to the client. User and device profiles based on the Composite Capability/Preference Profiles (CC/PP) model help in making inferences.

In some situations, the system might need to invoke other external servers to use third-party services such as Google Maps. We have developed the UbiSOA platform, which provides basic services for ubiquitous computing environments, such as indoor localization, access to sensing platforms, and interaction with RFID infrastructures. Detailed information about the project as well as source code are available at www.ubisoa.net.

We are considering several aspects to further develop our sentient visor system. Privacy is an important feature that is often overlooked in the IoT and in ubicomp systems in general. Not all smart objects should disclose information to everyone, so there should be mechanisms to restrict disclosure to selected entities—for example, through profiles.

Semantics play a key role in sentient visors, and we are implementing RDF-based Semantic Web technologies in UbiVisor. However, these technologies have not been widely adopted because Web publishers find it cumbersome to augment their content with semantics; in addition, the World Wide Web Consortium (W3C) standardization process has been slow. Consequently, alternative approaches such as microformats have started to emerge and gain adopters.

We are currently considering the RDFa schema, which embeds RDF annotations within an XHTML document in manner similar to microformats, enabling a compliant user agent to automatically extract an RDF serialization. However, whatever standard prevails does not impact the long-term prospects of our proposed system, as the design space does not require using a particular semantic technology.

Another fundamental issue worth examining is the browser-centered paradigm, which has its limitations. It might not be appropriate or even practical to always carry a physical device with a sentient visor, and sometimes it might be more effective to visualize information on public displays, especially in the case of shared information.

Some smart objects could take the form of ambient devices that convey subtle but sufficient information, eliminating the need for another entity to visualize their data. Also, some applications like environmental monitoring might lend themselves better to presenting data from collections of objects—for example, the average temperature and humidity of an entire field. In these cases, it is not necessary for the user to be physically present in the place where the data is being gathered. In addition, some objects, although networked, could have simple, specialized displays that users could consult when they are nearby, such as an HVAC system with a panel that displays current settings via a small LCD.

It is also worth noting that browsing through 2D pages and using backward/forward functions do not translate adequately to the IoT's dynamic physical environments. For the same reason, the Web's stateless nature poses important challenges. Accessing virtual entities, such as Web pages, is also very different from accessing physical entities such as actuators.

References

1. E. Aviles-Lopez et al., "Taking Care of Our Elders through Augmented Spaces," *Proc. Latin American Web Congress* (LA-WEB 09), IEEE CS Press, 2009, pp. 16-21.

2. E. Aviles-Lopez and J.A. Garcia-Macias, "Managing Primary Context through Web Services," *Proc. 2008 IEEE 24th Int'l Conf. Data Engineering Workshop* (ICDEW 08), IEEE CS Press, pp. 164-167.
3. E. Aviles-Lopez and J.A. Garcia-Macias, "Providing Service-Oriented Abstractions for the Wireless Sensor Grid," *Proc. 2nd Int'l Conf. Advances in Grid and Pervasive Computing* (GPC 07), LNCS 4459, Springer, 2007, pp. 710-715.
4. E. Aviles-Lopez and J.A. Garcia-Macias, "TinySOA: A Service-Oriented Architecture for Wireless Sensor Networks," *Service-Oriented Computing and Applications*, June 2009, pp. 99-108.
5. A.K. Dey, "Understanding and Using Context," *Personal and Ubiquitous Computing*, Feb. 2001, pp. 4-7.
6. P. Milgram and A.F. Kishino, "A Taxonomy of Mixed Reality Visual Displays," *IEICE Trans. Information and Systems*, Dec. 1994, pp. 1321-1329.
7. D. Guinard et al., "From the Internet of Things to the Web of Things: Resource Oriented Architecture and Best Practices," to appear in *Architecting the Internet of Things*, D. Uckelmann, M. Harrison, and F. Michahelles, eds., Springer, 2011.

J. Antonio Garcia-Macias *is a researcher in the Computer Science Department at CICESE Research Center, Ensenada, Baja California, Mexico. His research interests include ubiquitous computing, human-computer interaction, and computer networks. Garcia-Macias received a PhD in computer science from Institut National Polytechnique de Grenoble, France. He is a member of the IEEE Computer and IEEE Communication societies and the ACM. Contact him at jagm@cicese.mx.*

Jorge Alvarez-Lozano *is a PhD student in the Computer Science Department at CICESE. His research interests include participatory sensing, opportunistic communications, and social networking. Alvarez-Lozano received an MSc in computer science from CICESE. Contact him at jalvarez@cicese.mx.*

Paul Estrada-Martinez *is an MSc student in the Computer Science Department at CICESE. His research interests include semantic communications and new user interfaces for the Internet of Things. Contact him at pestrada@cicese.mx.*

Edgardo Aviles-Lopez *is a PhD candidate in the Computer Science Department at CICESE. His research interests include frameworks for building ubiquitous computing systems, user interface design, and sensor networking. Aviles-Lopez received an MSc in computer science from CICESE. Contact him at avilesl@cicese.mx.*

COVER FEATURE

Morphing Smartphones into Automotive Application Platforms

Raja Bose, Jörg Brakensiek, Keun-Young Park, and Jonathan Lester, *Nokia Research Center Palo Alto*

The Terminal Mode technology integrates smartphones into in-vehicle infotainment systems and transforms them into automotive application platforms on the fly, allowing drivers to safely access and interact with mobile applications.

The combination of the application store phenomenon and rapidly evolving hardware and software capabilities has transformed smartphones into the electronic equivalent of a Swiss Army knife. Smartphones provide services ranging from navigation to entertainment, and mobile applications make that content accessible anytime, anywhere.

However, automobiles remain the one environment where most people have seen little change in accessibility due to safety concerns. Many have resorted to the dangerous habit of using their handheld mobile devices while driving, sparking new laws aimed at reducing the number of distracted drivers. But as societies have become more industrialized, the amount of time people spend in their vehicles has increased sharply.[1] Even if users are willing to install expensive in-vehicle infotainment (IVI) systems or head units, they're usually restricted to a limited number of applications provided by the car manufacturer because of the closed and specialized nature of existing platforms.

Nokia developed the Terminal Mode technology (www.nokia.com/terminalmode) in collaboration with the Consumer Electronics for Automotive Working group (www.ce4a.org) to address this challenge by allowing smartphones to self-integrate into IVI systems, transforming them from mobile application platforms into automotive application platforms. This technology also opens up new opportunities to develop and deploy mobile applications that are customized for in-vehicle use while maintaining a safe user experience.

TODAY'S OPTIONS

As Table 1 shows, there are currently several ways for drivers to access services and applications inside their cars.

Car docks

Docks are an inexpensive and popular solution for mounting a smartphone on a car's dashboard or windshield so that the driver can use the phone's touch screen or keypad to access maps, play music, or send and receive messages. Although this option gives drivers access to their favorite applications, such usage is fraught with risks. A mobile device's form factor, interface, and interaction mechanisms are not designed to minimize driver distraction or to be used beyond arm's length. Furthermore, smartphones don't have reliable mechanisms for enforcing local driver distraction laws, such as, for example, blocking access to video content while the vehicle is in motion.

Another popular option is to mount a personal navigation device (PND) such as those manufactured by Garmin and TomTom on a car dock. However, because they focus primarily on navigation services, PNDs have less versatility compared to smartphone platforms in terms of application capabilities.

In either case, little or no interoperability exists between a smartphone or PND mounted on a dock and a car head unit, which provides driver-friendly UI controls for accessing applications and services.

0018-9162/11/$26.00 © 2011 IEEE Published by the IEEE Computer Society

Table 1. Comparison of application platforms available for in-vehicle use.

Feature	Terminal Mode	Car dock	Head unit	Bluetooth	Apple's iPod Out	Fiat's Blue&Me	Ford's SYNC	Kia's UVO
Support for legacy mobile applications and services	Yes	Yes	No	Limited to audio-based applications	Yes	Limited to phone call and audio	No	No
API support for mobile applications designed for in-vehicle use	Ongoing; designed to work across multiple vehicle platforms from multiple manufacturers	No	No	Limited to audio-based applications	No	Limited to phone call and audio; works in Fiat vehicles only	Yes; works in Ford vehicles only	Limited to phone and audio; works in Kia vehicles only
Provides application context information	Yes	No	No	Limited to audio content	No	Limited to audio content	No	No
Downloadable applications	Yes	Yes	No	Limited to audio content	Yes	Limited to audio content	Yes	Limited to audio content
Attestation of approved devices	Yes	No	N/A	N/A	Yes, but requires custom Apple ASIC inside head unit	No	No	No
Attestation of approved applications and content	Ongoing	No	Yes; self-certified by manufacturer	N/A	No	No	No	No
Ensures safety and prevents driver distraction	Yes	No	Yes	No	No	No	Yes	Yes
Driver-friendly display and controls	Yes	No	Yes	Limited to audio I/O	Yes	Limited to audio I/O	Limited interaction capability and output limited to low-resolution displays	Yes, but supports limited screen sizes
Standardized solution	Yes	No	No	Yes	No	No	No	No
Access to car data and services	Yes for car audio system; access to other car services and sensors ongoing	No	Yes	Access to car audio system only	No	Access to car audio system only	Yes	Yes
Connectivity to smartphone	USB, Wi-Fi, Bluetooth	N/A	N/A	Bluetooth	Proprietary wired	Bluetooth	USB, Bluetooth	USB, Bluetooth

Car head units

BMW's iDrive,[2] Audi's MMI (http://microsites.audiusa.com/brandsite/us/tech/mmi.html), Ford's SYNC (www.fordvehicles.com/technology/sync), and after-market head units such as those available from Alpine (www.alpine-usa.com/product/category/av-head-units-head-units) have been the primary automotive application platforms to date. On the plus side, these options provide driver-friendly input and output mechanisms such as multifunctional knobs and sunlight-readable displays that are designed to minimize driver distraction. But on the negative side,

- head units provide access to a smaller set of car-friendly applications compared to smartphones;
- the automotive industry has a much longer product development cycle compared to the consumer electronics industry, meaning that the technology embedded in a head unit is already "obsolete" by the time it's introduced;
- consumers typically use their vehicles for an average of 8 to 12 years, implying that they're saddled with head units that are not only expensive but also out of date in terms of hardware and software long before the end of a car's life cycle; and
- head unit applications are almost exclusively developed by vehicle manufacturers and their partners, who simply can't match the speed and variety of new and updated applications provided by third-party smartphone developer ecosystems.

Vertical integration of smartphones and car head units

Most of today's head units can handle phone calls through the Bluetooth Hands-Free Profile Protocol, which lets drivers dial and receive calls while driving without touching the phone. Head units also utilize the USB Media Transfer Protocol to stream music files from mobile devices such as smartphones and portable music players.

Several recent efforts have attempted to vertically integrate smartphones with head units—for example, Ford's SYNC system integrates with Google Maps so that a smartphone can send map directions to the head unit's navigation software, and BMW's iDrive head units can import e-mail messages and media content from BlackBerry devices. However, these integration efforts are subject to some limitations because they still depend on applications running natively on the head unit.

Vehicle manufacturers are making some progress in working with third-party vendors and partners, but they still encounter challenges. For example, Ford released the SYNC API, which lets users interact with smartphone applications through the SYNC head unit's I/O mechanisms, to select partners, but this API works only in Ford vehicles because it's proprietary to the SYNC platform. Developers who want their smartphone applications to work with other head units still must rewrite them for each vehicle manufacturer.

The mobile device side of the equation encounters challenges as well. Apple's iPod Out technology lets users manipulate a head unit's display and controls to communicate with iPod and iPhone devices, but it requires proprietary communication hardware and doesn't provide features for preventing driver distraction or enforcing regulations about the use of specific applications or content.

Jan Sonnenberg[3,4] described a system in which a combination of HTML5 interfaces and Web services can display and control mobile device applications via head units. Although this concept is broadly in the same class as Terminal Mode, it lacks some critical components and features. For example, using HTML5 to define user interfaces and Web services for business logic drastically reduces the number of existing applications that can immediately benefit from it and doesn't allow the use of nonlegacy applications in the car. Moreover, the system Sonnenberg described doesn't offer concrete mechanisms for ensuring user safety and security, such as attestation of devices, applications, and content. This is especially important because the head unit is actually running code from a foreign mobile device. The system also lacks a mechanism through which the head unit can impose rules to meet local driver safety regulations.

Terminal Mode gives smartphone users access to their favorite applications and services inside the car while also ensuring a seamless and safe user experience.

WHY TERMINAL MODE?

Terminal Mode was conceived and developed with the goal of giving smartphone users access to their favorite applications and services inside the car, while also ensuring a seamless and safe user experience. Various factors motivated this work.

User safety and experience

In an automotive environment, driver and passenger safety is always of paramount importance, but people now also demand an excellent user experience. Smartphones provide access to a wide variety of quality applications and services, but they are not equipped to handle critical issues such as driver distraction, whereas car head units ensure user safety but are limited by outdated platforms. Hence,

COVER FEATURE

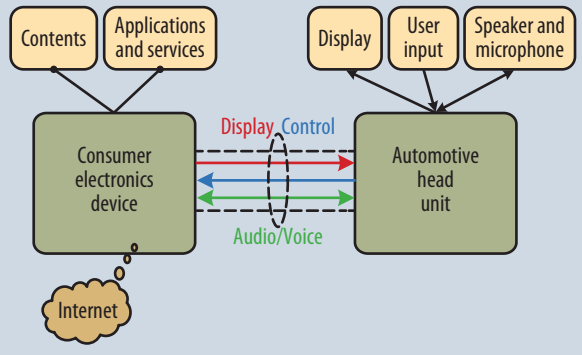


Figure 1. Terminal Mode concept. The consumer electronics device (smartphone) is the application platform for the automotive environment, and the car head unit is responsible for user input and output.

we designed our solution so that it could give users access to applications and services with a high-quality experience, yet support all necessary mechanisms for ensuring safety, including legal compliance with local regulations.

Seamless plug-and-play interoperability

To ensure a high-quality user experience, the transition from stand-alone smartphone usage to an interoperability mode must be seamless. Terminal Mode was designed from the ground up to provide capabilities that automate the setup and teardown processes for enabling interoperability as users enter and leave the vehicle environment without interrupting any task they might currently be engaged in.

Furthermore, Terminal Mode enables APIs to facilitate the development of smartphone applications specifically tailored for in-car use that can work across a large number of automotive platforms. It also provides platform-agnostic mechanisms that smartphone applications can use to access vehicle data and services, such as in-car GPS and onboard diagnostics data. However, to provide the user with immediate access to popular legacy applications and services, we designed specific abstraction layers into Terminal Mode so that approved legacy applications can also be used safely within a car without having to be rewritten.

Standardization

We had to consider every aspect of our design with the goal of ensuring open industry standardization across both the mobile device and automotive industries. We didn't want our solution to be tied to specific hardware or software platforms, yet it had to provide the same high quality of user experience across vehicles from a wide range of automotive manufacturers.

We also wanted to use standardized hardware and software interfaces and protocols that have withstood the test of time. Automotive electronics typically have a much longer life cycle compared to consumer electronics devices, yet they're already notoriously out of date upon introduction. As a rule of thumb, any technology introduced into a car must be maintainable for at least the next 15 years. Hence, we avoided the use of proprietary technologies and protocols in our solution.

Security via authentication and attestation

To ensure driver and passenger safety as well as compliance with legal regulations, protocols must be in place to ensure that car head units can authenticate and trust smartphone software, applications, services, and content. Terminal Mode provides core mechanisms that not only allow authentication and attestation but also enable the car head unit to dynamically specify safety rules that need to be enforced based on specific context such as geographical location and vehicle status.

THE TERMINAL MODE CONCEPT

As Figure 1 shows, in Terminal Mode, the smartphone is the application platform for the automotive environment, and the car head unit is responsible for user input and output. The smartphone hosts and executes all applications and services and also acts as a communications gateway to the cloud. The head unit provides the physical I/O mechanisms through which the driver or passenger accesses the smartphone.

The user interface hardware can include output devices such as one or more displays, audio playback, and haptic feedback systems as well as input devices such as touch screens, buttons, multifunctional knobs, and microphones. By treating the smartphone like an application platform, Terminal Mode makes it much easier to add new capabilities to the car head unit by simply upgrading the smartphone.

Figure 2a shows an example of a smartphone connected via Terminal Mode, and Figure 2b shows the use of the Music Player application via Terminal Mode. The user controls the music through the head unit's touch screen, which displays the Music Player user interface running on the smartphone, as the music is played through the car's audio system.

A major advantage of the Terminal Mode approach is that it provides stable hardware interfaces and software abstraction layers to automatically tailor the user's smartphone experience to the environment at any given time. For example, consider a use case involving a navigation application. When the user is outside the car, she uses the navigation application with the phone's I/O mechanisms; when she's driving, she interacts with the same navigation application through the car's built-in head unit display and driver-friendly controls. Hence, she gets the same experience as when using a navigation application running natively on the head unit but without any of the associated disadvantages.

Figure 2. Music Player interface. (a) Smartphone connected via Terminal Mode and (b) head unit touch screen displaying the Music Player application.

THE TERMINAL MODE STACK

As Figure 3 shows, the Terminal Mode stack consists of several components with responsibilities ranging from connectivity to security and safety.

Connectivity

The connectivity component handles the smartphone's physical and network connection to the car head unit. To ensure that head units aren't exposed to proprietary interfaces and protocols, this component provides several standardized choices.

For the physical and link layer, head units have the option of communicating with the smartphone via USB or wireless LAN (WLAN); Terminal Mode also supports Bluetooth connectivity for audio input and output. For USB connectivity, Terminal Mode requires both the smartphone and head unit to support the USB Communications Device Class/Network Control Model,[5] which enables Ethernet packet transfer. USB not only provides high bandwidth but also charges the smartphone while it's connected. Although WLAN-based connections offer the convenience of not having to physically connect the smartphone to the head unit, they currently suffer from lower bandwidth and significantly drain the phone's battery. However, with the rapid growth in wireless charging technologies and higher bandwidth promised in upcoming versions, WLAN is likely to become the preferred connectivity solution in the future.

Terminal Mode requires mandatory support for IP-based networking. IP was chosen as a suitable networking abstraction because it supports several protocols for transporting multiple types of content streams over a single physical or link layer. If there is a wired connection between the phone and vehicle and the user opts for different hardware connectivity solutions for each type of content stream—for example, HDMI for video and USB for control information—it would be necessary to use a proprietary composite physical connector. This is undesirable for automotive manufacturers due to compatibility and cost issues.

Service discovery and configuration

Terminal Mode uses the UPnP 1.0 protocol (www.upnp.org) to enable automatic discovery. The smartphone has a UPnP server that advertises the applications available through Terminal Mode upon establishment of a connection to a head unit. The car head unit has a UPnP Control Point that listens for advertisements or proactively searches for specific services. Terminal Mode uses two

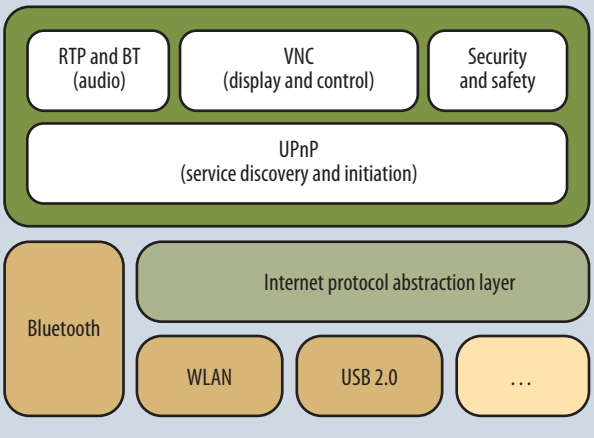


Figure 3. Terminal Mode stack. The various components have responsibilities ranging from connectivity to security and safety.

Figure 4. Native phone versus car mode user interface: (a) smartphone application menu, (b) customized menu generated by in-vehicle infotainment system.

UPnP services for service discovery and configuration—TmApplicationServer and TmClientProfile.

The TmApplicationServer service enables the head unit to discover and control smartphone applications available through Terminal Mode. The head unit can access lists of applications, launch or terminate a specific application, and query or get notification about an application's status. As Figure 4 shows, instead of simply replicating the smartphone's menus, the TmApplicationServer service lets the car head unit generate a user interface with its own look and feel for accessing and controlling applications. This not only helps automotive manufacturers preserve the look and feel of their head unit interfaces but also provides the user with the same experience as with applications running natively on the head unit.

The TmClientProfile service helps the head unit notify the smartphone about its capabilities, such as the ability to initiate a Bluetooth connection, and also enables it to set preferences, including safety rules. The service does this by configuring different client profiles on the smartphone that provide a stand-alone runtime session while adhering to head-unit-specified settings and rules.

When the head unit uses the TmApplicationServer service to launch applications, it can specify which client profile must be used to run the application. Each client profile on the smartphone has its own independent display and control buffer, meaning that the driver and passengers each feel that they have exclusive access to the smartphone. For example, suppose a car head unit has two display/control units—one for the driver and one for backseat passengers. For the client profile corresponding to the driver display, the head unit can specify rules that prohibit Internet browsing while the car is in motion, but the second client profile corresponding to the passenger display can allow it. A passenger can surf the Internet, but the driver can't access that content even though it resides and is running on the same smartphone device.

Remote user interface and control

Terminal Mode transfers the output from smartphone applications to the car head unit, and it transfers the user input from the head unit to the smartphone. Currently, Terminal Mode supports the following types of I/O, but new types such as haptic and gesture-based control can be included in the future without affecting compatibility with existing applications or head units:

- *input*—touch, including touch screens and track pads; nontouch (buttons and multifunctional knobs); and audio or voice; and
- *output*—video display and audio playback.

For remote display and control, Terminal Mode uses the Virtual Network Computing (VNC) Protocol.[6] The VNC server resides on the smartphone, and its client resides on the car head unit. After the head unit launches an application on the smartphone using the TmApplicationServer service, the application's GUI transfers to the head unit via VNC. Similarly, when the user provides input through the head unit controls, VNC transfers it to the smartphone.

The application's regular phone UI isn't expected to be transferred to the head unit, but, as Figure 5 shows, when the phone switches to Terminal Mode, applications are notified of the change and can adapt their regular UI to a driver-friendly one. Moreover, because VNC transfers content from the phone frame buffer, it provides a useful abstraction that allows exporting application UIs without requiring explicit application-level support. It also allows the smartphone to use different virtual frame buffers for multiple simultaneous user sessions, each corresponding to a different client profile.

To provide a safe, high-quality user experience, Terminal Mode specifies certain extensions to the standard VNC protocol that provide the following functionalities:

- support for scaling and rotation of the smartphone frame buffer and other modifications according to head unit's preferences;

Figure 5. Remote display and control: (a) smartphone navigation UI, and (b) driver-friendly navigation UI.

- mapping the head unit's set of possible input events to those supported on the smartphone, which is especially important if the head unit and the smartphone have different input modalities, such as a touch-based smartphone connected to a non-touch head unit;
- enabling or disabling certain input events—for example, if the user enters an address via the smartphone's navigation software using the head unit's speller knob, the smartphone can indicate the next possible subset of valid characters, cutting down the number of choices and reducing driver distraction;
- cursor detection and a keyboard trigger so that when the smartphone detects a text field in focus, it can send a keyboard trigger message to the head unit along with the cursor location, enabling the head unit to automatically pop up a virtual keyboard and focus on a specific portion so that the user can enter text with minimal distraction; and
- application category and context information from the smartphone so that the head unit can block undesirable or unknown content and applications from being displayed and accessed.

For audio input and output, Terminal Mode supports Bluetooth and Real-Time Protocol (RTP)-based streaming.[7] The car head unit is responsible for providing audio playback and getting the user's voice input, whereas the smartphone is responsible for transmitting audio content and receiving or processing audio input. For Bluetooth, Terminal Mode requires both the smartphone and head unit to support the Hands-Free Profile and the Advanced Audio Distribution Profile. For RTP, Terminal Mode uses IP networking.

To ensure smooth interoperability between Bluetooth and RTP components, Terminal Mode specifies rules for audio link selection and hand-off and specifically addresses cases in which audio link contention might occur. This ensures that someone who is already using the phone to make a call isn't interrupted while transitioning into Terminal Mode, or, if a phone call comes in during the use of turn-by-turn voice navigation, the navigation's audio output is temporarily suspended while the user takes the call.

Security and safety

In an automotive environment, user security and safety are crucial. Terminal Mode recognizes the following potential hazards and provides mechanisms to combat them:

- man-in-the-middle attack,
- smartphone connected to a noncertified head unit that claims to be Terminal Mode-compliant,
- car head unit connected to noncertified smartphone that claims to be Terminal Mode-compliant, and
- using unapproved applications and content while driving.

Terminal Mode provides device attestation mechanisms so that only certified compliant smartphones and head units can interoperate with each other. It specifies a Device Attestation Protocol (DAP) based on standard X.509 certificates[8] and attestation procedures standardized by the Trusted Computing Group (www.trustedcomputinggroup.org/resources/mobile_phone_work_group_mobile_trusted_module_specification_version_10).

To implement application and content certification, a trusted smartphone can provide device attestation information to the head unit, along with a trust level value. This trust level indicates whether the smartphone manufacturer, a trusted third-party authority, the application vendor, or the user has certified an application. Based on this trust level, the application category, and the content category, the head unit can decide whether to provide access to a specific application or not. Automotive manufacturers use the trust level to deliver new third-party applications for cars that are already on the market, thereby ensuring that the user can obtain and use new applications inside a car without compromising safety or security.

To enable car head units to control the type of content users can access, the smartphone indicates to the head unit the categories of the current display contents being transmitted and their associated locations on the displayed UI. The head unit can then either block certain portions of the display corresponding to the banned categories or ask the smartphone to do so.

To ensure driver safety, the TmClientProfile UPnP service lets the head unit specify content rules that the smartphone must enforce. Examples include rules specifying the minimum font size to ensure proper readability of content with minimum distraction as well as rules prohibiting certain types of applications such as Internet browsers. Car head units are customized for their local markets, and they are programmed with the legal regulations that need to be enforced. Future head units will also be able to download new rules based on the vehicle's current location. Because Terminal Mode provides mechanisms for the car head unit to specify such rules, it allows the dynamic specification and enforcement of local legal regulations to ensure a safe driving experience without requiring the smartphone to be updated. The same smartphone applications continue to provide a safe driving experience via Terminal Mode smoothly and seamlessly when users travel to different geographic areas.

Terminal Mode opens up the formerly closed niche automotive application space, allowing it to benefit from robust mobile developer ecosystems and their well-established application distribution infrastructure. Developers and service providers can use this completely new category of mobile applications without having to program using unfamiliar development platforms or technologies.

This technology is designed to work across a variety of vehicles from a wide range of manufacturers and is proposed as an industry standard.[8] Partners from both the mobile and automotive industries, which includes the world's major consumer electronics and automotive manufacturers, have come together to establish the Car Connectivity Consortium with Nokia (www.terminalmode.org) for developing and adopting Terminal Mode, ensuring interoperability not only across multiple vehicle models but also across multiple smartphone models. The consortium is also working to establish application certification processes that can be tied to existing Terminal Mode application attestation mechanisms.

Terminal Mode enhances the value of car head units by making them more upgradable and relevant in the long run. The response from both the mobile and automotive industries has been overwhelmingly positive, with several Terminal Mode-equipped smartphones and automotive products slated for worldwide launch in 2011 and beyond. The response from consumers, too, is strong because Terminal Mode provides a powerful incentive to buy head units without concern about their becoming obsolete. When a customer upgrades a smartphone, the car head unit will also automatically be upgraded, without having to wait for the vehicle or head unit manufacturer to release a new version. Furthermore, customers will be able to access their smartphone applications, services, and data safely and legally with a high-quality user experience.

In collaboration with the Car Connectivity Consortium, Nokia is already working on future versions of Terminal Mode that will include video-streaming mechanisms for use in passenger-focused head unit setups, vehicle-based services hosted on head units with access to onboard diagnostic data and other sensor data, and new platform-independent APIs for development of mobile applications and services targeted specifically for in-car use. With these key developments in progress, the future definitely looks promising for the development of mobile applications and services that can provide a quality in-car user experience without compromising safety.

Acknowledgments

We thank the Volkswagen Electronics Research Laboratory Palo Alto, especially Marc Leidenroth, Sinan Akay, and Jonathan Lee, for their help and collaboration.

References

1. US Census Bureau, "American Community Survey," 2009; www.census.gov/acs/www.
2. B. Niedermaier et al., "The New BMW iDrive: Applied Processes and Methods to Assure High Usability," *Proc. 2nd Int'l Conf. Digital Human Modeling* (ICDHM 09), LNCS 5620, Springer, 2009, pp. 443-452.
3. J. Sonnenberg, "Service and User Interface Transfer from Nomadic Devices to Car Infotainment Systems," *Proc. 2nd Int'l Conf. Automotive User Interfaces and Interactive Vehicular Applications* (AutomotiveUI 10), AMC Press, 2010, pp. 162-165.
4. J. Sonnenberg, "A Distributed In-Vehicle Service Architecture Using Dynamically Created Web Services," *Proc. 14th Int'l Symp. Consumer Electronics* (ISCE 10), IEEE Press, 2010, pp. 1-5.
5. USB Implementers Forum, "Universal Serial Bus Communications Devices Class Subclass Specifications for Network Control Model Devices, Revision 1.0," 2009; http://e-tools.info/project/documents/18001_19000/18516/ncm10.pdf.
6. T. Richardson et al., "Virtual Network Computing," *IEEE Internet Computing*, vol. 2, no. 1, 1998, pp. 33-38.
7. H. Schulzrinne et al., *RTP: A Transport Protocol for Real-Time Applications*, RFC 3550, IETF, 2003; www.ietf.org/rfc/rfc3550.txt.
8. D. Cooper et al., *Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile*, RFC 5280, IETF, 2008; http://tools.ietf.org/html/rfc5280.

Raja Bose *is a senior researcher at Nokia Research Center Palo Alto. His research interests focus on the theme of mobile interoperability, including creating new devices, systems, and services that provide innovative user experiences by leveraging interoperability of mobile devices with other smart environments. Bose received a PhD in computer engineering from the University of Florida. Contact him at raja.bose@nokia.com.*

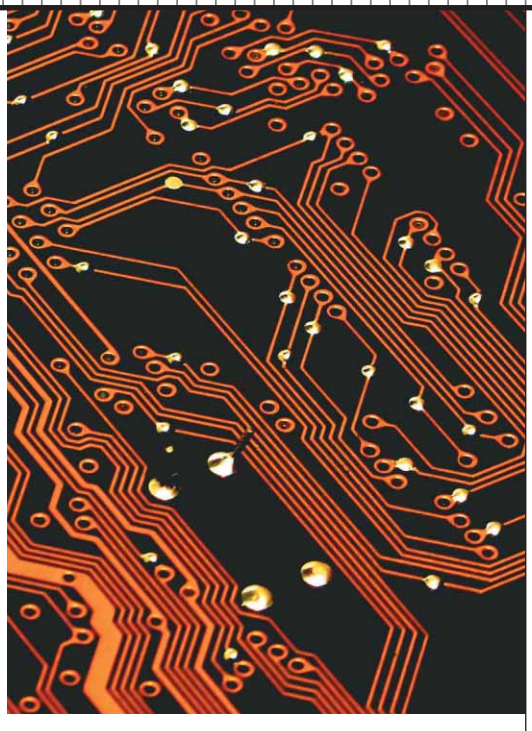
Jörg Brakensiek*, a research leader at Nokia Research Center Palo Alto, is editor in chief of the Terminal Mode specification. His research interests include ubiquitous mobile device interoperability, with a specific focus on the automotive domain. Brakensiek received a Dr. Ing. in electrical engineering from the University of Dortmund, Germany. Contact him at jorg.brakensiek@nokia.com.*

Keun-Young Park *is a senior researcher at Nokia Research Center Palo Alto. His research interests focus on the connectivity and interoperability of mobile devices with other close proximity intelligent devices, including in-vehicle infotainment systems. Park received an MS in aerospace engineering from Seoul National University. Contact him at keun-young.park@nokia.com.*

Jonathan Lester *is a senior researcher at Nokia Research Center Palo Alto. His research interests include the application and development of sensing technologies and embedded systems and the creation of novel sensing devices. Lester received a PhD in electrical engineering from the University of Washington. Contact him at jonathan.lester@nokia.com.*

RESEARCH FEATURE

Engineering Heterogeneous Robotics Systems: A Software-Architecture-Based Approach



Nenad Medvidovic, Hossein Tajalli, Joshua Garcia, and Ivo Krka
University of Southern California
Yuriy Brun, *University of Washington*
George Edwards, *Blue Cell Software*

RoboPrism, a framework that supports software-architecture-based development of robotic systems, is accessible to nonexperts in robotics, deals effectively with heterogeneity in distributed and mobile robotics systems, and facilitates adaptation in complex, dynamic environments.

Robotics systems exhibit characteristics that argue for a software engineering focus, including

- a high degree of heterogeneity among constituent subsystems,
- strict operational requirements dictated by real-time interactions with the physical world, and
- system complexity that extends beyond a single engineer's ability to grasp.

In fact, developers have increasingly applied software engineering to robotics systems, as reflected in a recent special issue of *IEEE Robotics and Automation*[1] and in the formation of the *Journal of Software Engineering and Robotics*.[2]

Despite these developments, it is still common for an engineering team to develop the decision-making and control apparatus of a robotics system from scratch, only to discover that it is too difficult to separate this software from the rest of the system and reuse it. They are thus forced to develop a new system from scratch.

The use of common robotics software libraries, such as Player and CLARAty, only partially alleviates this problem. Although these libraries consist of robotics-specific middleware that provides a low-level robot framework and helps with specific advanced features such as distributed communication and code mobility, the existing solutions provide no guidance or support for faithfully preserving the design-time structure of robotics systems.

Furthermore, relying on a given library results in applications that aren't easily ported to robot platforms that do not already support the library. Likewise, engineers must devise solutions for dealing with requirements (such as the dynamic loading of components) that the chosen technology does not natively support.

Recent approaches have adopted an explicit software engineering perspective for building robotics systems,

 Published by the IEEE Computer Society

resulting in reusable design and implementation frameworks. However, these approaches tend to neglect critical software engineering issues, including

- exploration of the design space and of the effective software design solutions within that space, needed both for the initial system design and subsequent dynamic runtime adaptations;
- modeling the distributed software-intensive system that is deployed on a set of robot (and possibly traditional) platforms as opposed to modeling robotic algorithms;
- analysis of the system models for key properties before constructing and deploying the system and during dynamic adaptations;
- traceability of the design-time artifacts, such as components and connectors to implementation constructs; and
- support for heterogeneous development and deployment platforms.

Our approach aims to remedy these shortcomings. The basis of this work is software architecture,[3] a set of principal design decisions about a software-intensive system embodied in the system's components (operational entities that perform computation), connectors (entities that facilitate interaction and coordination among components), and configurations (assemblies of components and connectors into system-specific topologies). Our approach uses a robotics system's architectural basis to address the five problem areas.

ROBOTICS THROUGH THE PRISM OF SOFTWARE ARCHITECTURE

We propose a novel architectural style that supports guided exploration of design alternatives for a dynamically adaptive robotic system and uses a rigorous system modeling and analysis framework. It also uses implementation and deployment middleware with the explicit architecture traceability support that's necessary for heterogeneous settings. In the process, our work can make the development of robotics software more accessible to nonexperts in robotics, reduce the time and effort required to create and maintain robotics software, and improve the exchange of design solutions among robotics engineers.

Our approach to engineering robotics software adapts and applies three important software architecture concepts:[3]

- *architectural design abstractions*, enabling the creation of reusable, adaptive, and hierarchical components and systems;
- *architectural modeling and analysis*, allowing early, integrated, and continuous (re)evaluation of system behaviors and properties; and
- *architectural middleware*, permitting system implementation, deployment, monitoring, and runtime (self-)adaptation in highly dynamic, mobile, and heterogeneous environments.

The "Scenario for a Heterogeneous Robotics System" sidebar illustrates a typical multirobot application scenario that calls for a software engineering approach.

Design abstractions

A significant focus of software engineering research has been to codify design abstractions, which engineers use to represent and reason about complex systems at a high level. To this end, software architecture researchers have developed a canonical set of architectural design constructs: components, connectors, communication ports, interfaces (or services), events, and configurations. Furthermore, the uses of these constructs, prescribed via design heuristics or constraints, result in architectural styles (such as client-server or peer to peer) that are key design principles in software engineering. These constructs and principles have been highly useful in practice.

Implementation and deployment middleware provides the explicit architecture traceability support that's necessary for heterogeneous settings.

In traditional software, layering implies that components at a given layer invoke the services of components at the layer below. In contrast, components at a given layer in the *adaptive-layered* style monitor, manage, and adapt components at the layer below.[4]

The bottom layer in an adaptive-layered system is the application layer. Components in this layer implement functionality that achieves the application goals. An adaptive-layered architecture can have an arbitrary number of metalayers. Components in these layers—*collectors*, *analyzers*, and *admins*—are designed to handle operations that deal with monitoring, analysis, and adaptation. Collectors monitor lower-layer components, analyzers evaluate adaptation policies or plans based on monitored data, and admins perform adaptations. This approach ensures the separation of application-level from metalevel functionality, while allowing the system a high degree of autonomy.

We used an adaptive-layered style to realize different adaptive software systems.[4] In recent work, we leveraged this approach to design the Plan-Based Layered Architecture for Software Model-Driven Adaptation (PLASMA).[5] As Figure 1 shows, PLASMA employs three adaptive layers. Application-level components reside in the bottom layer. The middle layer—called the adaptation layer—monitors,

SCENARIO FOR A HETEROGENEOUS ROBOTICS SYSTEM

Consider the following scenario. A convoy of mobile robots must assemble autonomously and follow a leader robot along a prespecified path given as a series of waypoints, as Figure A shows. These robots collect and process data from onboard sensors and stationary sensor nodes deployed at various locations within the environment. As they traverse the path, the robots encounter several base stations, which can assess the robots' state, allow a robot to dock and recharge its battery, transfer data to and from the robot, and even release software updates to the robot.

Robots can collaborate by exchanging data as well as computational components (such as mobile code). They can also run onboard analyses to track their own health. For example, a robot with a depleted battery can minimize its remote communication or its onboard computation. Robots also need to adapt to changing environmental conditions, such as GPS signal loss or low visibility. Finally, the mission's goal might change at runtime from, for example, following the leader to mapping an unknown terrain. Overall, the robots, sensors, and base stations are a distributed, decentralized, and heterogeneous computing environment that must be capable of dynamic adaptation.

Such a scenario involves several technical challenges; some—such as developing effective algorithms to achieve the robot-following behavior—are clearly robotics specific. However, we argue that a majority of the remaining technical challenges fall within software engineering, and that software engineering provides the appropriate abstractions, methods, techniques, and tools to address such problems. This position has, in fact, been increasingly recognized by researchers who have tried to construct robotic systems using model-driven development and reusable domain-specific middleware platforms. In fact, several software engineering researchers have recently targeted their techniques toward dynamically adaptive robotics systems.[1-3]

Figure A. Convoy of four robots following a leader. The red robot leaves the group to charge its battery at a base station.

References

1. J.C. Georgas and R.N. Taylor, "Policy-Based Self-Adaptive Architectures: A Feasibility Study in the Robotics Domain," *Proc. 2008 Int'l Workshop Software Eng. for Adaptive and Self-Managing Systems* (SEAMS 08), ACM Press, 2008, pp. 105-112.
2. D. Sykes et al., "From Goals to Components: A Combined Approach to Self-Management," *Proc. 2008 Int'l Workshop Software Eng. for Adaptive and Self-Managing Systems* (SEAMS 08), ACM Press, 2008, pp. 1-8.
3. H. Tajalli et al., "PLASMA: A Plan-Based Layered Architecture for Software Model-Driven Adaptation," *Proc. 25th IEEE/ACM Int'l Conf. Automated Software Eng.* (ASE 10), IEEE CS Press, 2010, pp. 467-476.

manages, and adapts components in the application layer. The top layer (planning) manages the adaptation layer and the generation of plans based on user-supplied goals and component specifications. The planning layer defines both the target architecture for the application layer (in the adaptation plan) and the actions for the application layer to carry out (in the application plan). The planning layer can respond to changing system requirements or operational environments by regenerating plans.

This three-layer architecture offers a high degree of autonomy and enforces a clear separation of concerns, whereby each layer provides a different form of adaptation capability. To use the adaptation capabilities, an architect must provide an architectural description of the system components and application goals. Alternatively, an architect can use only the application layer when developing a nonadaptive system.

Modeling and analysis

Our approach to engineering robotics software employs architectural models and analyses to inform and direct design decisions related to dynamic planning and adaptation.

First, we use architecture models specified in the Software Architecture Description and Evaluation Language (SADEL)[6] to automatically generate models needed for planning. A SADEL model specifying the functional interfaces of application components helps determine the actions available to the system and the effects of those actions on the environment. A SADEL model specifying the management interfaces of components (such as deploy, suspend, connect, and so on) helps determine how the adaptation layer can manipulate components to achieve a goal.

Second, we implemented tools that let engineers experiment with different

- system design decisions with respect to nonfunctional properties,
- policies for triggering dynamic replanning, and
- options for redeploying software components.

These tools are extensions to the Extensible ToolChain for Evaluation of Architectural Models (XTEAM) modeling and analysis toolset.[7] XTEAM provides an editing environment for specifying architecture models,

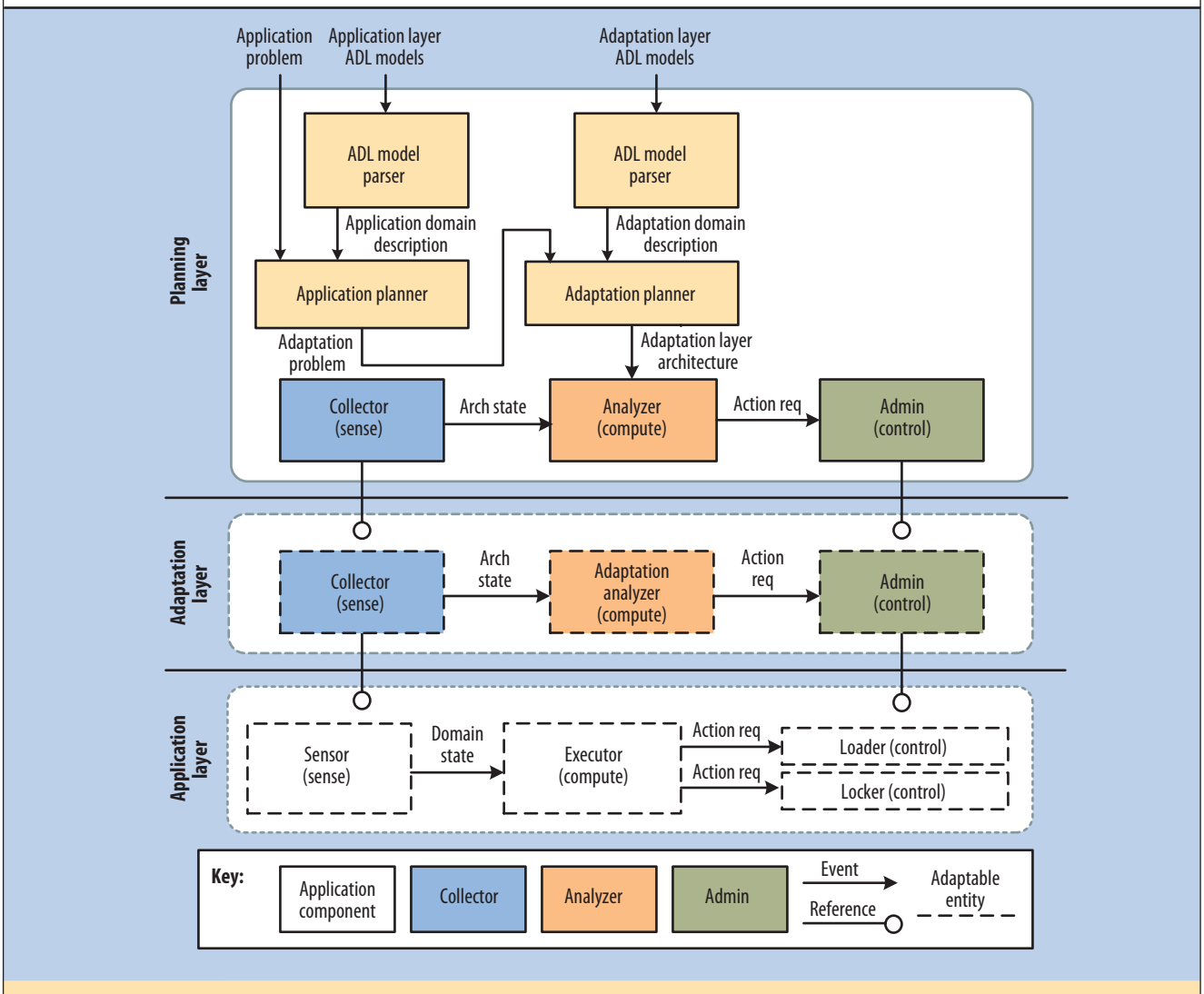


Figure 1. PLASMA adaptive-layered architecture.

a simulation generator for generating a discrete-event simulation of a system, and a code generator. Developers use the discrete-event simulation to observe the system's dynamic behavior under different operational conditions, assumptions, and constraints. XTEAM natively includes facilities for

- representing a software system architecture's structure and behavior in a formal model;
- attaching properties to model elements to capture parameters needed for various analyses; and
- analyzing simulations generated from models with respect to performance, reliability, and energy efficiency.

Engineers can use XTEAM to determine the impact of different replanning and redeployment strategies and to establish varying policies on system performance, reliability, and power efficiency.

Middleware

The existing robotics libraries and frameworks, although useful in many settings, are not always effective middleware platforms for developing robot-based software systems. This is particularly the case for systems distributed across multiple, heterogeneous platforms. Instead, we have developed and modified a layered middleware solution, RoboPrism, that alleviates these shortcomings by

- providing the necessary low-level abstractions for interfacing with the underlying operating system, network, and hardware;
- incorporating different robotics libraries, as appropriate;
- implementing software systems in terms of constructs (component, connector, event, port, style, and so on) that directly mirror architectural-design-level concepts;

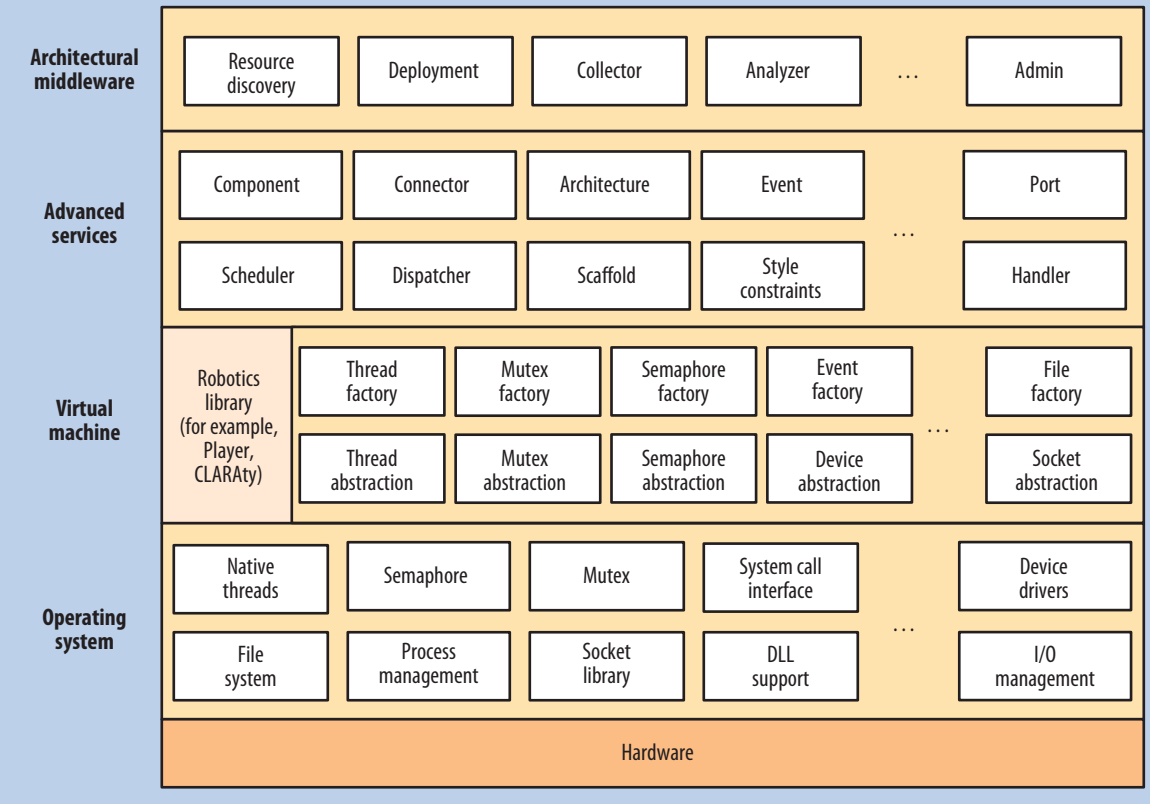


Figure 2. RoboPrism, a layered architectural middleware platform. Researchers have successfully integrated different robotics libraries within RoboPrism's virtual machine layer.

- providing an extensible collection of advanced, meta-level services, such as resource discovery or dynamic replanning and self-adaptation components;
- enabling the management and adaptation of the meta-level components and services to provide an adaptive layered system; and
- achieving the preceding without imposing unacceptable resource costs (in terms of memory, CPU, or network) on the resulting systems.

The resulting middleware is an adaptation of the Prism-MW middleware platform (http://sunset.usc.edu/~softarch/Prism) developed for embedded systems. Prism-MW focuses primarily on the architectural middleware layer in Figure 2. It relies on substrates (the virtual machine layer in Figure 2), such as the JVM for the Java version. Extensive measurements indicate that Prism-MW introduces less than 5 percent overhead for advanced services (deployment, mobility, disconnected operation, and monitoring),[8] which is acceptable for the architectural traceability that benefits analysis, maintenance, and reuse. Furthermore, providing these additional architectural abstractions does not impose a noticeable performance penalty.[8]

Using the RoboPrism platform yields several important benefits. First, systems designed according to RoboPrism insulate application software developers from reliance on the underlying robotics libraries, if any: the architectural middleware layer exports a single interface to application developers.

Second, RoboPrism allows the implementation of applications in multiple programming languages: the architecture construct bounds an address space, while specialized first-class connectors carry out interaction across address spaces.

Third, RoboPrism provides meta-architectures, which contain specialized metacomponents (admin, collector, and analyzer) that enable adaptive-layered applications. In such applications, components on each host are separated into distinct Prism-MW meta-architectures corresponding to each layer. Separating layers into distinct architectures enforces and guarantees the following architectural constraints:

- components in different layers only interact through prescribed mechanisms, and
- each meta-architecture only manages and adapts the architecture in the layer immediately below it.

Moreover, this separation insulates components in each layer from failures and adaptations in other layers, thus supporting a high degree of autonomy.

EXPERIENCE

We have investigated these concepts in the context of two scenarios using the iRobot Create platform. Our investigations also used the eBox3854 embedded PC running Linux, laptops running Windows XP and Vista, and Compaq iPAQ PDAs running Linux and Windows CE.

To dock iRobots and charge their batteries during scenario execution, we used the iRobot Home Bases. Creative Webcam and Logitech QuickCam cameras (controlled via the Java Media Framework, or JMF) provided visual information that enabled robot following, and SunSpot Java-based sensors provided the ability to manually control robot movement through accelerometers.

We relied on three options for controlling the iRobots: the Player and Create Open Interface libraries, both of which are in C, as well as our custom iRobot driver in Java. This, in turn, let us use two versions of RoboPrism: the Java version running on JamVM and the GNU C++ version running on a virtual machine developed by Bosch RTC. The 2.0.5 version of Player is compatible with JavaClient2, offering two options for interacting with iRobots for each version of RoboPrism. This highly heterogeneous environment has proven appropriate for validating the benefits of our approach. The "Hardware and Software Sources" sidebar provides links to websites providing more information about these tools.

HARDWARE AND SOFTWARE SOURCES

- iRobot Create, www.irobot.com/home/index.jsp
- eBox3854 embedded PC, www.microcomputersystems.com/eBox.htm
- iRobot Home Base, http://store.irobot.com/product/index.jsp?productId=2814855
- Java Media Framework (JMF), http://java.sun.com/javase/technologies/desktop/media/jmf
- SunSpot sensors, www.sunspotworld.com
- Create Open Interface library, http://code.google.com/p/libcreateoi

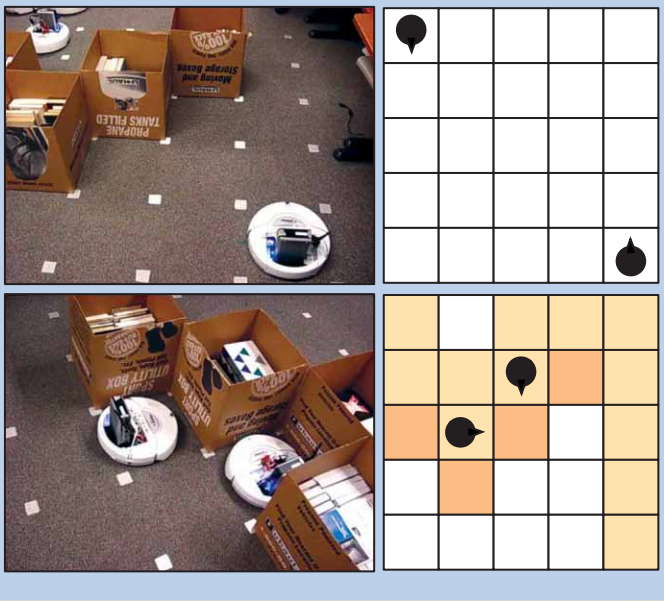


Figure 3. Two remote-controlled robots map out a 5 × 5 grid with unknown obstacles. The initial configuration, indicated by the blank map containing only the robots' positions and orientations, appears at the top. An intermediate configuration, with a majority of the grid traversed and four obstacles found, appears at the bottom.

Environment exploration scenario

Our initial scenario involved exploring and mapping an unknown environment with randomly placed obstacles, as Figure 3 shows. We designed, modeled, and implemented this scenario using the Java version of RoboPrism. Five teams of two or three graduate students worked on this scenario during a 10-week, two-part project. Only one student had prior robotics or embedded-systems experience; four other students had previously been exposed to PrismMW, the precursor to RoboPrism. The project was initiated before, but completed after, we obtained the iRobots. The project's objective was to investigate whether an explicit focus on software architecture and the use of architectural middleware could

- reduce the initial development effort and subsequent modification of a robotics system for nonexperts in robotics,
- facilitate traceability (that is, preserve the designed architecture in the implementation),
- enhance exchange of design solutions, and
- alleviate heterogeneity challenges.

The project's first part involved developing a simulated environment exploration system, in which the robots were "virtual"—simulated in a GUI. The virtual robots had to run on a host other than the host from which they were controlled. Like real robots, they had to move in the requested direction and report any obstacles found so that the students could construct a map of the environment.

The project's second part involved replacing the virtual robots with the iRobots. Students had to do so without altering the application's architecture: all changes to the components running on the (initially virtual and then real) robots had to be contained entirely inside the components.

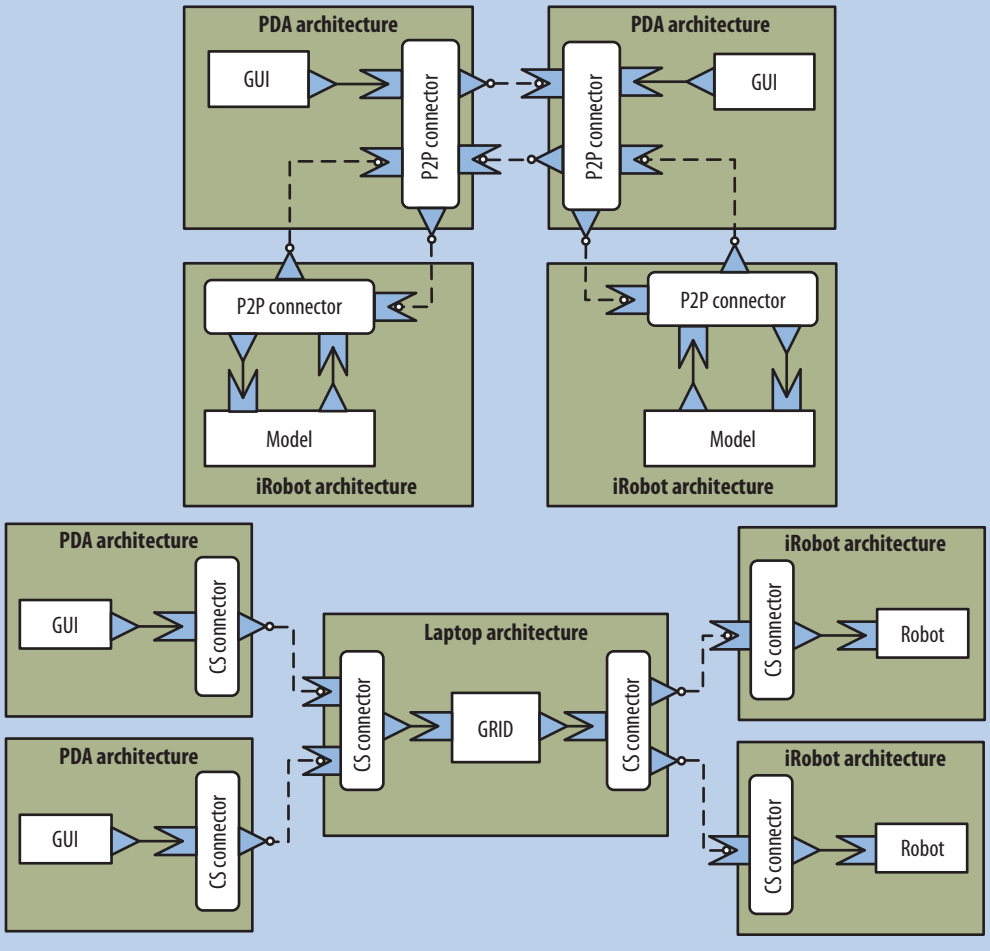


Figure 4. Two architectures for the environment exploration scenario, each relying on a different style: peer-to-peer (top) and client-server (bottom).

All five teams succeeded in preserving their architectures during the migration to iRobots. This success implies better maintainability of the resulting systems because the implementations preserve the designed architectures, avoiding architectural drift.

Two teams experienced difficulty controlling the iRobots' movement while trying to accurately map the unknown environment. The primary difficulties arose from their unfamiliarity with programming robots, the iRobot Create platform, and the Player library. The five resulting applications had similar functionalities with minor variations in numbers of PDAs used and the navigation algorithm's degree of automation. However, because our approach does not mandate a particular architecture for a system, engineers can explore and decide on the architecture that best fits their design decisions and objectives. As a result, the five architectures were substantially different in terms of the system decomposition into components and connectors, interfaces, interactions via events, and deployment onto the hardware nodes.

For example, Figure 4 shows two architectures that emerged from this project. The peer-to-peer solution in the top diagram will likely scale well and remain tolerant to host failures. On the other hand, it could experience data consistency problems if the events sent by peers are dropped or arrive and are processed in the incorrect order. The client-server solution in the bottom diagram has a central grid component that ensures a consistent global view of the system and avoids synchronization problems. At the same time, the grid component represents a single point of failure and might also become a performance bottleneck.

While this system's heterogeneity otherwise might have posed a serious problem in migrating from a Java GUI-based back end to the iRobots running Player, the application's use of RoboPrism greatly reduced such problems. In particular, the middleware allowed seamless communication among components regardless of the hardware platform that housed them. Demonstrating code portability and modularity, the students were able to easily wrap the robot

control libraries and use them inside the components they had developed in the first part of the project.

The software design and implementation support let users with little domain expertise rapidly develop distributed, user-friendly robotics applications. The explicit focus on software architecture facilitated easy communication and exchange of high-level design solutions. Furthermore, we reused several modules from these systems in later research.

To evaluate the reduction of effort from using our approach, we measured the source lines of code (SLOC) and development effort estimates for the students' application code; Table 1 shows these results. We estimated effort using the Cocomo II software project cost estimation model,[9] and intend these numbers to indicate the complexity of the students' application code. Cocomo II suggests that two- to three-person teams with no personnel turnover would have required 3.8 to 8.8 months to produce the respective amounts of code. In contrast, the students in this project completed their work much more quickly; on average, they expended about four weeks of concerted programming effort. Although a more definitive conclusion would require further investigation, these numbers are suggestive of RoboPrism's effectiveness.

Table 1. Source lines of code and development effort estimates for the student projects.

Code base	No. of team members	SLOC	Development effort estimate (person-months)
1	3	1,600	3.8
2	2	2,700	6.7
3	2	2,900	7.5
4	2	1,700	4.2
5	2	3,400	8.8

Robot-following scenario

We designed and implemented several variations of the robot-following scenario. In the first set of scenarios, designed in tandem with an industrial collaborator, we manually designed adaptation policies and coded them in metalevel components, according to the adaptive-layered style. In the second set of scenarios, we leveraged PLASMA[5] to automatically design the adaptation plans as well as the application architecture.

Adaptive-layered implementations. In this scenario, the leader robot follows a line drawn on the floor using infrared sensors. Other robots use a camera to observe the color of and follow the robot in front of them. A robot also can follow an infrared signal emitted from the robot in front of it. A robot uses the infrared mechanism when it doesn't have a camera or its camera malfunctions. Along the way, robots encounter base stations and SunSpot sensors; they can choose to dock with the base stations to recharge their batteries, exchange data with SunSpots, or perform software updates. Robots dock and update software through autonomous control components.

Researchers also can use SunSpots as remote controllers to correct the orientation of an iRobot when it loses sight of the robot in front of it. When a robot leaves the convoy, it notifies the robot immediately behind it, and the remaining robots adjust their leader-follower roles to maintain the organization. A robot can rejoin the convoy when it sees the trailing robot's color. Researchers can issue commands from laptops and iPAQ, and they can receive feedback about the robots' progress and energy consumption.

To enable this functionality, we designed several components, including LineFollower, ColorFollower, IRFollower, SunSpotController, and SunspotReader. We also designed metalevel RoboPrism components to directly support runtime monitoring, analyses, and the system's dynamic adaptation. These components monitor and adapt the system's architecture in anticipated situations. For example, a monitor component detects camera failures and initiates an adaptation plan, which in turn replaces the ColorFollower component with an IRFollower component. These RoboPrism components organize the application and metalevel components into a two-layer adaptive-layered architecture.

We designed the adaptation policies captured within the metalevel components and refined them using XTEAM models. First, we used the rate of battery drain during different operational modes such as camera following, infrared following, and so on to determine appropriate thresholds at which to trigger recharging. Second, XTEAM analyses determined that we could not deploy all the follower components simultaneously due to the robots' limited available memory, necessitating component redeployment when hardware or software faults trigger adaptation policies.

This scenario demonstrates several benefits of our approach, including

- modeling and nonfunctional property analysis for adaptive systems,
- heterogeneity support;
- traceability, reuse, and modularity; and
- runtime architectural analysis and adaptability.

We first designed the entire system by exploring appropriate decompositions into components and connectors, as well as different candidate architectural styles. Then, we modeled the resulting design and analyzed it using XTEAM for completeness, consistency, and nonfunctional characteristics. We then transferred the model directly to the system implementation via RoboPrism's native support for architectural constructs. This allowed us to create a modular architecture that exhibited desired properties.

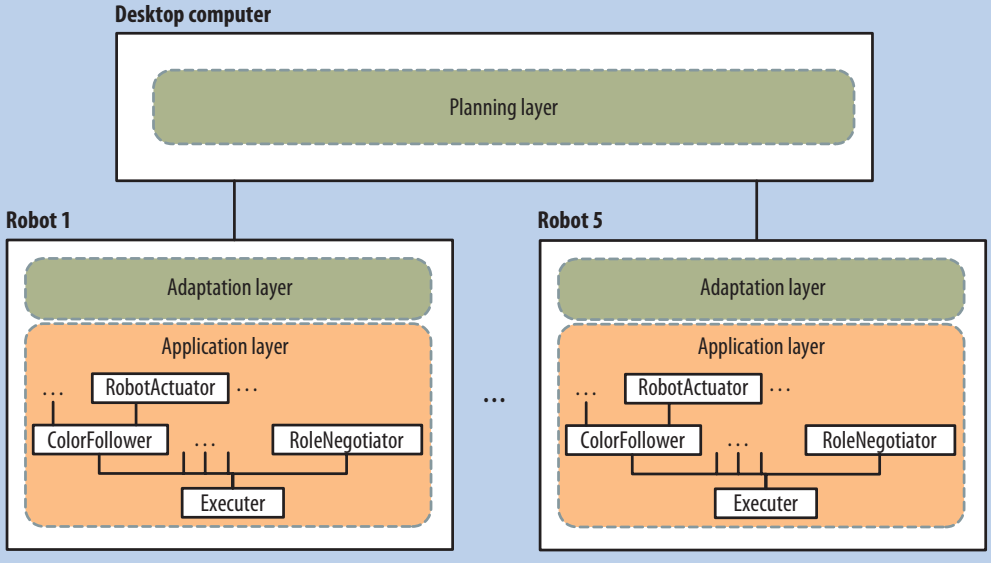


Figure 5. Deployment view of the system architecture for the robotics scenario.

PLASMA implementation. The three-layer PLASMA architecture follows the adaptive-layered style supported by RoboPrism, enabling a high degree of separation, modularity, and multilayer adaptation. To provide a high degree of autonomy, PLASMA relies on architect-generated SADEL models of the components in the application and adaptation layers. We transform the SADEL models into state transition models to use in adaptation planning.

PLASMA constructs separate plans for the application and adaptation layers. The application plans control the application behavior to achieve system goals. Similarly, the adaptation plans control the behavior of the adaptation layer—setting and adapting the application layer's architecture. To assess the benefits of PLASMA's dynamic adaptation support, we implemented a variant of the robotics scenario.

While transferring the robotics scenario to PLASMA, we successfully reused most of the application components from the scenario implementation; this further validated the reusability of our approach. In the PLASMA version, the leader robot follows a path defined by a series of spatial coordinates called waypoints. Initially, we provided PLASMA with the SADEL models of 15 application components. One component developed for the PLASMA scenario, RoleNegotiator, implements a distributed negotiation to assign a role (leader or follower) to all robots in the convoy. The negotiation protocol ensures that it assigns only one robot the leader role. Only the leader uses waypoint following; followers use other types of following.

PLASMA reduces the burden on the system architect by automatically generating adaptation plans, which the architect would otherwise specify manually. In PLASMA, the architect only provides the application's goal.

In our scenario, each robot's goal is to follow the robot in front of it and avoid obstacles. The PLASMA planning layer, deployed on a laptop, generates application and adaptation plans. The planning layer also automatically generates and compiles implementation code for the adaptation analyzer and executor components that perform the adaptation.

PLASMA then deploys compiled binaries of all required components (application components, adaptation analyzer, collectors, and so on) and instantiates an identical adaptation layer on each robot. The adaptation layer on each robot instantiates the application layer, and the Executor begins executing the application plan, in which the first step is role negotiation. Figure 5 shows an instance of this architecture's deployment.

Automatically generated application and adaptation plans support different types of system adaptations under different circumstances. As a result, the system architect need not predict and plan for all adaptations. The application plan automatically handles basic adaptations. For example, if a robot is using a camera for following and the area becomes dark, the Executor can use an application plan to automatically switch to GPS or infrared following. More powerful adaptations require dynamic replanning.

Consider the case in which robots must recharge their batteries using docking stations along the route. To satisfy this requirement, we specified new SADEL models for the BatteryMonitor and StationDocker components. We also specified a new application goal that defines the acceptable battery power threshold, and then initiated replanning. PLASMA computed new plans and regenerated and rede-

ployed the Adaptation Analyzer and Executor, along with the other required application components.

Another adaptation occurs when a component fails and the application removes it from the set of available components. In this case, replanning adapts the application with a new application plan that does not use the removed component.

The automatically generated plans in these scenarios ranged from 790 to 4,390 state actions. Each state action specifies the behavior required in a specific state (for example, the invocation of a particular operation). Manually specifying policies of this size would be tedious, cumbersome, and error prone. By automating the process, PLASMA removes this burden, letting architects focus on their primary task—architectural description. For example, modifying the application goal in the case of battery recharging only requires specifying two additional SADEL models (55 lines of architectural description) and a new problem description (a single line change), along with the implementations of the two components.

Software architecture provides critical abstractions, techniques, and tools for designing and organizing software systems, and is particularly important in the case of complex heterogeneous systems that might need future extension or modification. To make it easier to use software architectural concepts in robotics, we have created three tools: XTEAM to automate system modeling and analysis; RoboPrism to give architectural abstractions first-class status in system implementations and allow dynamic analysis and redeployment of the system; and PLASMA to dynamically generate complex adaptation plans.

In our future research, we intend to expand the boundaries of using software engineering and software architecture concepts in the context of robotics systems. The recent improvements in the area of domain-specific modeling languages can facilitate flexible modeling of robotics applications in different domains, while preserving compatibility with existing analysis tools.[10] Further, we plan to enhance our adaptive framework with runtime reasoning about nonfunctional properties in an environment that has notable resource constraints. We believe that these enhancements will make robotics systems more accessible, reproducible, reusable, and adaptable to changes in their runtime environment.

References

1. D. Brugali and E. Prassler, "Software Engineering for Robotics," *IEEE Robotics and Automation Magazine*, Mar. 2009, pp. 9, 15.
2. D. Brugali, "From the Editor-in-Chief: A New Research Community, a New Journal," *J. Software Eng. for Robotics*, Jan. 2010, pp. 1-2.
3. R.N. Taylor, N. Medvidovic, and E.M. Dashofy, *Software Architecture: Foundations, Theory, and Practice*, John Wiley & Sons, 2009.
4. G. Edwards et al., "Architecture-Driven Self-Adaptation and Self-Management in Robotics Systems," *Proc. Int'l Workshop Software Eng. for Adaptive and Self-Managing Systems* (SEASS 09), IEEE CS Press, 2009, pp. 142-151.
5. H. Tajalli et al., "PLASMA: A Plan-Based Layered Architecture for Software Model-Driven Adaptation," *Proc. 25th IEEE/ACM Int'l Conf. Automated Software Eng.* (ASE 10), IEEE CS Press, 2010, pp. 467-476.
6. N. Medvidovic, D.S. Rosenblum, and R.N. Taylor, "A Language and Environment for Architecture-Based Software Development and Evolution," *Proc. 21st Int'l Conf. Software Eng.* (ICSE 99), IEEE CS Press, 1999, pp. 44-53.
7. G. Edwards and N. Medvidovic, "A Methodology and Framework for Creating Domain-Specific Development Infrastructures," *Proc. 23rd IEEE/ACM Int'l Conf. Automated Software Eng.* (ASE 08), IEEE CS Press, 2008, pp. 168-177.
8. S. Malek, M.M. Rakic, and N. Medvidovic, "A Style-Aware Architectural Middleware for Resource Constrained, Distributed Systems," *IEEE Trans. Software Eng.*, Mar. 2005, pp. 256-272.
9. B. Boehm et al., "Cost Models for Future Software Life Cycle Processes: COCOMO 2.0," *Ann. Software Eng.*, Dec. 1995, pp. 57-94.
10. G. Edwards, "Automated Synthesis of Domain-Specific Model Interpreters," doctoral dissertation, Dept. Computer Science, Univ. of Southern California, 2010.

Nenad Medvidovic *is a professor in the Department of Computer Science at the University of Southern California and director of the USC Center for Systems and Software Engineering. Medvidovic received a PhD in information and computer science from the University of California, Irvine. Contact him at neno@usc.edu.*

Hossein Tajalli *is a PhD student in the Department of Computer Science at the University of Southern California, where he is a member of the Software Architecture Research Group in the Center for Systems and Software Engineering. Tajalli received an MS in electrical engineering from the University of Tehran, Iran. Contact him at tajalli@usc.edu.*

Joshua Garcia *is a PhD student in the Department of Computer Science at the University of Southern California. He received an MS in computer science from the University of Southern California. Contact him at joshuaga@usc.edu.*

Ivo Krka *is a PhD student in the Department of Computer Science at the University of Southern California, where he is a USC Provost's Fellow. He received an MS in computer science from USC and an MEng in computing from the University of Zagreb. Krka is a member of IEEE and ACM Sigsoft. Contact him at krka@usc.edu.*

Yuriy Brun *is an NSF CRA postdoctoral Computing Innovation Fellow at the University of Washington. He received a PhD in computer science from the University of Southern California. Brun is a member of the ACM and ACM Sigsoft. Contact him at brun@cs.washington.edu.*

George Edwards *is the chief scientist at Blue Cell Software in Los Angeles. He received a PhD in computer science from the University of Southern California. Contact him at george@bluecellsoftware.com.*

CAREER OPPORTUNITIES

invent

Hewlett-Packard Company has an opening for a

Thermal Engineer

Houston, TX

Reference: HOUPMA2

Reqs: Master's in Mechanical Eng or rltd & 3 years exp. Coursework/exp. must include: Research exp in the field of thermal sciences; Particle Image Velocimetry (PIV) technique & liquid cooling; Fluid Mechanics, Heat Transfer, & Thermodynamics; & Computational Fluid Dynamics.

List full name, address & email on resume. Ref to Job# HOUPMA2. Please send resumes with job number to Hewlett-Packard Company, H1-6E-28, 5400 Legacy Drive, Plano, TX 75024. No phone calls please. Must be legally authorized to work in the U.S. without sponsorship. EOE.

UF UNIVERSITY of FLORIDA

The Foundation for The Gator Nation

Intel / Charles E. Young Endowed Chair in Information Technology

The College of Engineering invites nominations and applications for the endowed position of Intel/Charles E. Young Leadership Chair in Information Technology; established to honor Dr. Charles E. Young, who served as president of the University of Florida from 1999-2004. The ideal candidate will be a worldwide recognized scientific leader in fundamental and applied IT research with a proven record of excellence in research, education, funded large-scale activities and team building. This is a full professor level position, tenured in the Dept. of Electrical and Computer Engineering. Candidates must possess a Ph.D. in Electrical and Computer Engineering or related discipline and have a demonstrated record of exemplary achievement and leadership in their field of research.

The search committee will begin reviewing applications immediately and will continue to receive applications until the position is filled. To apply please send a pdf of a curriculum vitae, statement of research interests, and contact information for five references via e-mail to renato@acis.ufl.edu.

The University of Florida is an equal employment opportunity employer. Women and minorities are encouraged to apply. The "government in the sunshine" laws of Florida apply to the search.

SR. JAVA DEVELOPER. Judge Technical Staffing, Conshohocken PA seeks individual to work as part of ICG (Institutional Client Group) team, involved in dvlpmt & maintenance of core tech components used by other s/ware applics w/in ICG group in Hudson County NJ & other unanticipated locs. Utilize Java technologies, EAI tools, message oriented systems such as TIBCO EMS & open source tools. Travel reqd. Qualified apps must have BS + 5 yrs Java dvlpmt exp at least 1 yr w/top tier fin'l services firm. Will accept MS + 3 yrs. Submit resume & refs to D. Lock, Judge Group, 300 West Conshohocken State Rd., Ste. 300, Conshohocken, PA 19428.

ALLIED TELESIS has an IT SalesForce Administrator and Developer position available in San Jose, CA. Manage users and data, conduct training and drive use, and create and manage ad hoc reports and executive dashboards. BS in CE, CS, EE or related + 2 yrs in related exp. Mail resumes to Allied Telesis, Inc., 3200 N.

Seagate®

Seagate US LLC, leader in data storage technology, has professional engineering opportunities at various levels in **Bloomington, MN: Development Engineer (DEVENG11):** develop lithographic processes used in the formation of leading edge magnetic recording heads; **Sr. Engineer (SRENG16):** develop materials and devices for application towards advanced recording heads; **Senior Engineer (SRENG14):** research and develop novel slider air bearing design paradigms to enable future data storage technologies; **Senior Engineer (SRENG15):** investigate new metrology concepts to support the design and development of next generation recording head technology; **Senior Engineer, Development (DEVENG12):** serve as primary contact for all cost model requests and ROI analysis for the Normandale Slider/HGA R&D and MR&D organizations; and **Staff Developer - Web Applications (SDWA01):** manage and maintain applications deployed in Weblogic application servers that host R&D applications.

In Shakopee, MN: Firmware Engineers (FENGR21 and FENGR22): design and implement disc drive controller firmware for an embedded ARM microprocessor; **Firmware Test Development Engineer (FTDE01):** support Seagate's enterprise Solid State Drive products; **Servo-Mechanical Engineer (SME01):** work on large-scale technology integration projects; **and Staff Engineer (SENG11):** develop firmware for Solid State Drive.

In Fremont, CA: Senior Engineer (SRENG17): perform research and development related to film structure and magnetic materials.

In Schaumburg, IL: Sr. Data Recovery Engineer (DRE01): evaluate and determine if and how data can be recovered from a wide variety of recording and storage media, including magnetic and optical media, file and mail systems.

In Oklahoma City, OK: Database Administrator (DA01): administer database applications, including TIBCO Administrator, BusinessWorks, and Foundation and Business Connect.

Also in Scotts Valley, CA: Senior Analyst (A02): translate business goals and objectives into measurement and tagging strategy.

If interested, Send resumes to 10321 West Reno Avenue, M/S OKM010, Oklahoma City, OK 73127 (Must reference job code).

Published by the IEEE Computer Society 0018-9162/11/$25.00 © 2011 IEEE

JUNIPER NETWORKS

Juniper Networks, Inc. is recruiting for our Sunnyvale, CA office:

Software Engineer #19125: Design and implement software for company EX series switches. Review and provide feedback for software written y other engineers.

Test Engineer #20055: Write and execute test plans for features given. File problem reports, verify and close the reports.

Software Engineer #19742: Work on all aspects related to video compression, processing and communication. Work involves topics related to video coding performance optimization to improve the speed and visual quality.

Test Engineer Staff #21103: Work on company data center network products qualification and deployment. Provide support and interface with multiple internal teams (such as PLM, TME and Engineering teams), customers and partners to develop and implement joint qualification plans and procedures.

Security Engineer Staff #6853: Responsible for feature development for VoIP application-level gateway/firewall.

Software Engineer #14320: Design, document, develop & debug S/W for next generation highly scalable networking platforms.

Software Engineer #18555: Participate on a project team of engineers involved in the development of software for routers. Define, design, develop, test, debug, release, enhance and maintain device driver, platform software and routing protocol related to software.

Software Engineer #18070: Design & develop mobile signaling protocols for the GGSN, PDN Gateway & Serving Gateway as part of the Mobile Edge Solution. Req's: MS+0.

Software Engineer #8233: Design and develop JUNOS software on network routing platforms.

Software Engineer #6999: Design, develop and maintain features related to IPSEC VPN for Junos.

Technical Support Engineer #10370: Document and reproduce customer problems and coordinate fixees by engineering department. Manage critical customer accounts and escalation requests.

Juniper Networks, Inc. is recruiting for our Westford, MA office:

ASIC Engineer #5719: Develop software programs used for the verification of Application Specific Integrated Circuits (ASIC).

Technical Publication Engineer Staff #3709: Lead high-profile Technical Publications strategic initiatives such as customer usability programs, information architecture, and optimization of XML tools.

Juniper Networks, Inc. is recruiting for our Herndon, VA office:

Sr. Systems Engineer #19919: Responsible for assisting Account Executives & Partners w/ the technical sales activities in driving managed security services and hosted services development. Req's 10 yrs related security & networking knwlge on Managed Services or BS+8.

Mail resume with job code # to Attn: MS A1.2.1.435
Juniper Networks
1194 N. Mathilda Avenue
Sunnyvale, CA 94089

First St., San Jose, CA 95134. Attn: Sr. Human Resources Generalist. Must reference job title and job code # MS0001 to be considered.

SENIOR CONSULTANT sought by Red Planet Consulting, Inc. in Boulder, CO. Work focusing on dsgn, dvlpmt & implmt'n of s/ware applics built up on GE, Smallworld Geographic Information GIS platform. Min. Req BS Geoinformatics Engg. Resumes to 1107 12th St., Ste. 300, Boulder, CO 80302.

UNIVERSITY OF SOUTH CAROLINA, Computer Science and Engineering, Faculty Positions. The College of Engineering and Computing at the University of South Carolina seeks to fill one senior and two junior tenure-track positions in the Departments of Computer Science and Engineering, Electrical Engineering, and Mechanical Engineering. The departments were ranked in the top group nationally according to recent NRC metrics. This cluster will establish a national center of excellence in safety-critical systems, especially for aerospace applications, with particular focus on software engineering, computational modeling, sensing and control, and condition-based maintenance. Requirements include a PhD in a relevant field, a record of exceptional accomplishments, and strong potential for research and teaching. For details, see http://www.c.sc.edu.employment.html. Applications containing CV, research and teaching plans, and contact information for three references should be submitted to ClusterSearch@cec.sc.edu. Screening of applicants will begin May 1, 2011 and continue until the positions are filled. AA/EOE.

PROGRAMMER ANALYST - dsgn, dvlp, test & implmt applic s/w utilizing knowl of & exp w/PL/SQL, ASM, RAC. Oracle Applications 11i/R12, XML Publisher, Discoverer, Hyperion, Essbasereporting, OA framework, Workflow, AME, BPEL, SOA, DAC, ETL, OBIEE, BIAPPS, HTML, .Net Framework, AJAX, Java. Reqs MS Comp Sci, Eng or rel. Mail resumes to Nitya Software Solutions Inc., 9690 South 300, Ste. 319, Salt Lake City, Utah 84070.

FREEBALANCE INC. seeks a VP Sales, Latin America for its Wash, DC HQ. Will have overall respons for bus dev & sales activities rltd to govt accountability s/ware solutions in Latin Amer. Candidates must poss Bachelor or equiv deg in Comp Sci, Info Sys, Sys Eng or rltd fld & min 12 yrs strategic enterprise s/ware sales exp in Latin Amer region, incl mgmt of complex sales agreements. Permitted to work out of a home office located w/ in Southeastern region of U.S. Up to 70% int'l trvl to Latin Amer req'd. F/T employ-

(Continued on page 77)

Microsoft

Microsoft Corporation currently has the following openings available:

Aliso Viejo, CA
Software Development Engineers, Software Development Engineers in Test (includes Lead, Senior and Principal levels): Responsible for developing computer software applications, systems or services. http://bit.ly/MSJobs-SDE

Boulder, CO
Software Development Engineers, Software Development Engineers in Test (includes Lead, Senior and Principal levels): Responsible for developing computer software applications, systems or services. http://bit.ly/MSJobs-SDE

Cambridge, MA
Software Development Engineers, Software Development Engineers in Test (includes Lead, Senior and Principal levels): Responsible for developing computer software applications, systems or services. http://bit.ly/MSJobs-SDE

Program Managers (includes Lead, Senior, Principal and Group levels): Coordinate program development of computer software applications, systems or services. http://bit.ly/MSJobs-ProgMgr

Charlotte, NC
Support Engineers / Escalation Engineers (includes Senior levels): Provide technical support on issues experienced with Microsoft technologies. http://bit.ly/MSJobs-Support

Chevy Chase, MD
Software Development Engineers, Software Development Engineers in Test (includes Lead, Senior and Principal levels): Responsible for developing computer software applications, systems or services. http://bit.ly/MSJobs-SDE

Chicago, IL
Engagement Manager: Responsible for managing complex computer software products and all aspects of client engagements. http://www.jobs-microsoft.com/job/go/1199103/

Fargo, ND
Software Development Engineers, Software Development Engineers in Test (includes Lead, Senior and Principal levels): Responsible for developing computer software applications, systems or services. http://bit.ly/MSJobs-SDE

Support Engineers / Escalation Engineers (includes Senior levels): Provide technical support on issues experienced with Microsoft technologies. http://bit.ly/MSJobs-Support

Irvine, CA
Premier Field Engineer: Provide technical solutions and support on issues experienced with Microsoft technologies. http://bit.ly/MSJobs-Support

Irving, TX
Support Engineers / Escalation Engineers (includes Senior levels): Provide technical support on issues experienced with Microsoft technologies. http://bit.ly/MSJobs-Support

Technical Account Manager: Manage the delivery of premier support services to an assigned group of corporate customers. http://www.jobs-microsoft.com/job/go/1198981/

Iselin, NJ
Premier Field Engineer: Provide technical solutions and support on issues experienced with Microsoft technologies. http://bit.ly/MSJobs-Support

Premier Field Engineer: Provide customers with reliable technical solutions to complex integration problems associated with business solutions, technical support to enterprise customers, partners, internal staff and/or others on mission critical issues experienced with Microsoft technologies. http://www.jobs-microsoft.com/job/go/1199179

Madison, WI
Software Development Engineers, Software Development Engineers in Test (includes Lead, Senior and Principal levels): Responsible for developing computer software applications, systems or services. http://bit.ly/MSJobs-SDE

Mountain View, CA & Palo Alto, CA
Software Development Engineers, Software Development Engineers in Test (includes Lead, Senior and Principal levels): Responsible for developing computer software applications, systems or services. http://bit.ly/MSJobs-SDE

Program Managers (includes Lead, Senior, Principal and Group levels): Coordinate program development of computer software applications, systems or services. http://bit.ly/MSJobs-ProgMgr

Reliability Engineer: Coordinate program development of computer hardware applications, systems or services. http://www.jobs-microsoft.com/job/go/1194206/

New York, NY
Sales Engineers: Provide sales technical support, responses to proposals, and/or evangelism to partners and/or customers. http://bit.ly/MSJobs-SalesEng

Software Development Engineers, Software Development Engineers in Test (includes Lead, Senior and Principal levels): Responsible for developing computer software applications, systems or services. http://bit.ly/MSJobs-SDE

Senior UX Design Lead: Oversee a unified user experience for Bing across all major user touch-points and platforms. http://www.jobs-microsoft.com/job/go/1199334/

Account Manager: Provide world-class scaled service to a portfolio of advertisers and agencies. http://www.jobs-microsoft.com/job/go/1199422/

Plano, TX
Engagement Manager: Utilize technical knowledge in software applications to coordinate the deploying of computer software products and services. http://www.jobs-microsoft.com/job/go/1199150/

Redmond, WA
Software Development Engineers, Software Development Engineers in Test (includes Lead, Senior and Principal levels): Responsible for developing computer software applications, systems or services. http://bit.ly/MSJobs-SDE

Program Managers (includes Lead, Senior, Principal and Group levels): Coordinate program development of computer software applications, systems or services. http://bit.ly/MSJobs-ProgMgr

Database Engineers: Administer production SQL Server or other databases or systems, including performance tuning, capacity planning analysis and troubleshooting. http://bit.ly/MSJobs-DBSys

Hardware Engineer: Design, implement, and test computer hardware. http://bit.ly/MSJobs-Hardware

International Project Engineers/Managers and Localization Engineers: Ensure the successful localization of software components for foreign markets. http://bit.ly/MSJobs-Localization

Support Engineers / Escalation Engineers (includes Senior levels): Provide technical support on issues experienced with Microsoft technologies. http://bit.ly/MSJobs-Support

Service Engineers, Service Operations Engineers, and Systems/Operations Engineers (includes Lead and Senior levels): Plan, architect, deploy or support complex client/server or database software systems. http://bit.ly/MSJobs-SysOps

Technical Writers and Editors: Research, write, edit and develop online, Internet-based or print media documentation for software or hardware products or technologies. http://bit.ly/MSJobs-TechWriting

User Experience Designers: Develop user interface and user interaction designs, prototypes and/or concepts for business productivity, entertainment or other software or hardware applications. http://bit.ly/MSJobs-UX

User Experience Researchers: Design and conduct usability studies to evaluate software user interface prototypes, products and/or services. http://bit.ly/MSJobs-UX

Animation Director: Manage Animation Team to ensure animation goals are met according to schedule and budget. http://www.jobs-microsoft.com/job/go/1199408/

Applied Researcher: Responsible for conducting research and leading research collaborations that yield new insights, theories, analysis, data, algorithms, and prototypes and that advance the state-of-the-art of computer science and engineering, as well as general scientific knowledge. http://www.jobs-microsoft.com/job/null/1051739/

CRM Custom Success Manager: Responsible for the delivery of design, planning, and implementation services that provide IT solutions to customers and partners. http://www.jobs-microsoft.com/job/go/1199109/

Development Lead: Lead the design, implementation and/or test of computer software applications, systems or services. http://www.jobs-microsoft.com/job/go/1087097/

General Manager: Lead the Server and Tools Online (STO) division that designs, builds and runs some of the largest and most popular developer and IT pro websites. http://www.jobs-microsoft.com/job/go/1199421/

Global Launch Manager: Use Operations, Supply Chain Management, and Process Improvement to define and drive improvements in processes that constitute global launch of new products/programs at Microsoft. http://www.jobs-microsoft.com/job/null/1199247/

Regional Executive Producer: Work on components of the design, implementation, and expansion of a portal structure across multiple geographic markets in different languages, including the design and production management of home pages and channel pages. http://www.jobs-microsoft.com/job/null/1199194/

SE Service Design Engineer: Responsible for analyzing, reviewing and resolving all aspects of the design, development and implementation of computer software products and services to deploy successful product releases. http://www.jobs-microsoft.com/job/go/1199174

Senior Development Manager: Design, implement and/or test computer software applications. http://www.jobs-microsoft.com/job/go/1199186

Senior Games Designer: Evaluate video game design and coordinate game development on issues including scope, design, trial experience, and others. http://www.jobs-microsoft.com/job/go/1199211/

Solution Manager 3: Analyze business and systems practices and processes and coordinate between business and engineering teams to deliver IT solutions that add efficiencies and enhancements to business processes. http://www.jobs-microsoft.com/job/null/1199162/

Spam Analyst 2: Responsible for, as a member of the FOPE team, triaging spam and false positive emails on a daily basis. http://www.jobs-microsoft.com/job/go/1202063/

Technical Artist: Responsible for the creation of visual concepts, effects, animation, or other visual images using computers and electronic tools. http://www.jobs-microsoft.com/job/go/1027255/

Traffic Quality Engineer: Provide technical support to enterprise customers, partners, internal staff and/or others on mission critical issues experienced with Microsoft Technologies. www.jobs-microsoft.com/job/go/1199114/

San Francisco, CA
Program Managers (includes Lead, Senior, Principal and Group levels): Coordinate program development of computer software applications, systems or services. http://bit.ly/MSJobs-ProgMgr

Southfield, MI
Premier Field Engineer: Provide technical solutions and support on issues experienced with Microsoft technologies. http://bit.ly/MSJobs-Support

Multiple job openings are available at various levels for each of these categories. To view all opportunities, detailed job descriptions and minimum requirements, and to apply to specific job opportunities, visit the website address listed for each job category or visit www.jobs-microsoft.com. Microsoft is an equal opportunity employer and supports workplace diversity.

Hewlett-Packard Company has an opening for a

Sr. Software Developer

Ft. Collins, CO
Reference: FTCVTH2

Reqs: Bachelor's & 5 yrs exp with software dev using C#, VB, HTML/DHTML & C++; and exp working w/ cross-functional teams; IIS; Windows Forms, ActiveX, GUI design; HTTP, DNS; UPnP AV and/or DLNA; Web servers/hosting & HTML; Windows 7, Vista. List full name, address & email address on resume. Send resume & refer to Job# FTCVTH2

Please send resumes with job number to Hewlett-Packard Company, H1-6E-28, 5400 Legacy Drive, Plano, TX 75024. No phone calls please. Must be legally authorized to work in the U.S. without sponsorship. EOE.

Nokia Inc.

has the following open position in

Sunnyvale, CA

SENIOR ENGINEER

Duties include UI development for the Media Player, TNE & Web Video Projects involving Series 60, Symbian OS or mobile platforms in software development & TDMA/CDMA/GPRS/3G, mobile phone using RTOS, Platform SDK, OOAD & UML & Rational Rose & in-circuit debugger; work with embedded C/C++, Java, XML & related processes and tools; & other duties/skills required. [Job ID: NOK-11SV-SR]

Mail resume to: Nokia Recruiter,
3575 Lone Star Circle, Ste. 434
Ft. Worth, TX 76177
& note specific Job ID#.

ment offered w/comp sal & benefits. Resume to Carol Tracey at hr@freebalance.com & ref job #801.

EXECUTIVE PROJECT MANAGER needed w/mgmt exp in the electric power utility industry. Mail resume to: Open System Int'l, Attn: H.R., 3600 Holly Ln N #40, Minneapolis, MN 55447.

PROGRAMMER ANALYST: Analyze, dsgn, dvlp, modify & implmt s/ware/systems applics in client/server environment using MS Visio, MS Excel, MS Access, Java, J2EE, Oracle, MS Word, Rational Clear Quest, XML, SQL, RAD, Spring, Struts, JSF, Web Services, CVS, Tivoli Identity & Access Manager, WebSphere Application Server, Windows, UNIX. Reqs MS Comp Sci, Eng or rel. Mail resumes to Professionals LLC 1000 N West St, Ste # 1283, Wilmington, DE 19801.

SENIOR SAP ABAP DEVELOPER. Responsible for maintenance, enhancement and production support of SAP applications utilizing Advanced Business Application Programming (ABAP) language, SAP ABAP development, ABAP workbench, Data Dictionary Debug, SQL tracing tools, SAP Workflow, and

(Continued on page 79)

HP Enterprise Services, LLC is accepting resumes for the following positions:

Services Information Developer

Monona, WI (Ref. # ESMOSID11)
Nashville, TN (Ref. #ESNASSID21)
Folsom, CA (Ref. #ESFOLSID21)
Washington, DC (Ref. #ESDCSID11)
Oklahoma City, OK (Ref. #ESOKSID21)

Conceptualize, design, develop, unit-test, configure, or implement portions of new or enhanced (upgrades or conversions) business and technical software solutions through application of appropriate standard software development life cycle methodologies and processes.

Technology Consultant

Washington, DC. (Ref. #ESDCTC11)

Provide technology consulting to customers and internal project teams. Provide technical support and/or leadership in creation and delivery of technology solutions designed to meet customers' business needs and, consequently, for understanding customers' businesses.

Information Systems Architect

Pontiac, MI (Ref# ESPONISA21)

Deliver business value by architecting effective information systems solutions that address business problems, needs and opportunities, in a manner consistent with the company's strategic and business goals.

ITO Service Delivery Consultant

Folsom, CA (Ref. #ESFOLITO11)

Provide expertise for IT infrastructure, application infrastructure, and related services throughout the lifecycle of a deal in accordance with contractually established terms and conditions and established technical standards.

Presales EDS Consultant

Plano, TX. (Ref. #ESPLANNG1)

Create, qualify, develop, and sell company solutions to customers.

Mail resume to HP Enterprise Services, LLC, 5400 Legacy Drive, MS H1-6F-61, Plano, TX 75024. Resume must include Ref. #, full name, email address & mailing address. No phone calls please. Must be legally authorized to work in the U.S. without sponsorship. EOE.

invent

Hewlett-Packard Company has an opening for a

Software Designer V

Cupertino, CA
Reference: CUPVSU2

Reqs: Master's + 2 yrs exp. & exp w/ Enterprise App Devlp; Java/J2EE architecture; Objected Oriented Designs & Sw Devlp Best Practices; & WebServices, XML, HTML, Struts, JUnit, Mockito, JBoss, Eclipse IDE, Subversion, ANT, Log4J, Linux & Windows. List full name, address & email address on resume.

Please send resumes referencing Job# CUPVSU2 to Hewlett-Packard Company, H1-6E-28, 5400 Legacy Drive, Plano, TX 75024. No phone calls please. Must be legally authorized to work in the U.S. without sponsorship. EOE.

invent

Hewlett-Packard Company has an opening for an

Software Designer VI

Cupertino, CA
Reference: CUPPMU2

Reqs: Exp with developing e-business solutions, and all phases of s/w dev lifecycle; Java, C; Expertise in full J2EE stack and design patterns; Web Services (SOAP/RESTful) and Axis Toolkit; Weblogic, Websphere & Tomcat; Rational Clearcase, CVS, Sub Version; Multi-threaded prog using VC++ & Java 1.5 Concurrency API; Agile methodology. List full name, address & email address on resume. Send resume & refer to Job# CUPPMU2.

Please send resumes with job number to Hewlett-Packard Company, H1-6E-28, 5400 Legacy Drive, Plano, TX 75024. No phone calls please. Must be legally authorized to work in the U.S. without sponsorship. EOE.

invent

Hewlett-Packard Company has an opening for a

Systems/Software Engineer

Cupertino, CA
Reference: CUPIDE2

Requires exp w/ backup/recovery techs; SAN; Databases; Storage devices & disc rays; & Network mgmt methodologies. List full name, address & email address on resume.

Please send resumes referencing Job# CUPIDE2 to Hewlett-Packard Company, H1-6E-28, 5400 Legacy Drive, Plano, TX 75024. No phone calls please. Must be legally authorized to work in the U.S. without sponsorship. EOE.

Positions at the Institute for Defense Analyses Center for Computing Sciences

The Institute for Defense Analyses Center for Computing Sciences (IDA/CCS) is looking for outstanding researchers to address difficult computing problems vital to the nation's security. IDA/CCS is an independent, applied research center sponsored by the National Security Agency (NSA). Emphasis areas for IDA/CCS technical staff include high-performance computing, cryptography, and network security. Members of the technical staff come from a diverse variety of backgrounds, including computer science, computer architecture, computer/electrical engineering, information processing, and the mathematical sciences; most have Ph.D.s. Special attention is paid to the design, prototyping, evaluation, and effective use of new computational algorithms, tools, paradigms, and hardware directly relevant to the NSA mission. Stable funding provides for a vibrant research environment, and an atmosphere of intellectual inquiry free of administrative burdens.

The center is equipped with a very large variety of hardware and software. The latest developments in high-end computing are heavily used and projects routinely challenge the capability of the most advanced algorithms and architectures. IDA/CCS research staff members have always been at the forefront of computing, as evidenced by lasting, visible contributions to areas as varied as multi-threaded architectures (e.g., Horizon), novel computing systems (e.g., FPGA-based Splash and Splash-2, Processing-In-Memory chips), design and implementation of operating systems (e.g., the Linux kernel), and programming language design and implementation for high-performance computing systems (e.g., Universal Parallel C and Cinquecento).

IDA/CCS research staff work on complex topics often engaging multidisciplinary teams; candidates should demonstrate depth in a particular field as well as a broad understanding of computational issues and technology. Because the problems of interest are continually evolving, IDA/CCS recruitment focuses on self-motivation, strength of background, and talent, rather than specific expertise.

Located in a modern research park in the Maryland suburbs of Washington, DC, IDA/CCS offers a competitive salary, an excellent benefits package, and a superior professional working environment. U.S. citizenship and a Department of Defense TSSI clearance (with polygraph) are required. IDA/CCS will sponsor this clearance for those selected. The Institute for Defense Analyses is proud to be an equal opportunity employer.

Please send responses or inquiries to:

Dawn Porter
Administrative Manager
IDA Center for Computing Sciences
17100 Science Drive
Bowie, MD 20715-4300
dawn@super.org

invent

Hewlett-Packard Company has an opening for an

Product Marketing Manager

Cupertino, CA
Reference: CUPGTU2

Reqs: MS in Bus., CS or rltd. & 6 yrs exp as a Prod Mktg Mgr or rltd. Exp with Enterp. SW prod mktg, Prod Mgmt. exp. w/ bus. apps (ie: SAP, Oracle, Siebel or PeopleSoft); Knowledge of QA processes, testing, test automation, SW dvlpmt lifecycle methodologies; Knowledge of HP SW prods or prod knowledge of HP SW competitors, specifically around Qlty Mgmt. List full name, address & email address on resume.

Please send resume referencing Job# CUPGTU2 to Hewlett-Packard Company, H1-6E-28, 5400 Legacy Drive, Plano, TX 75024. No phone calls please. Must be legally authorized to work in the U.S. without sponsorship. EOE.

ADS Alliance Data Systems, Inc
is accepting resumes for the following position in:

Arlington, VA

SENIOR SOFTWARE ENGINEER

Exp. in the software industry; develop organizational digital marketing platforms, services, & products, maintenance of legacy systems, work with object-oriented core Java & C++/C# mid-tier & system level development; service-oriented software system & application integration; & other duties/skills required. [Job ID: AD-VA11-SRE]

Mail resume to J. Fuson-HR, Alliance Data Systems, 4445 Lake Forest Dr., Ste. 200, Cincinnati, OH 45242 & note Job ID#.

ADS Alliance Data Systems, Inc is accepting resumes for the following position in:

New York, NY

TECHNICAL PROJECT MANAGER

Duties include being responsible for the coordination, management & timely completion of the company's platform operations, which include working with Macro level with Windows (Exchange, SharePoint, Active Directory, clustering, IIS), Unix/Linux (Apache 2, RHEL, FreeBSD, SendMail, IPTables, BIND), network administration, Oracle/MySQL/SQL database administration, load balancing; & other duties/skills required. [Job ID: AD-11NY-TPM]

Mail resume to J. Fuson-HR, Alliance Data, 4445 Lake Forest Dr., Ste. 200, Cincinnati, OH 45242 & note Job#.

SAP Scheduling. Participate in defining/documenting technical system requirements. Assist with troubleshooting and application performance issues and work collaboratively with business partners, provide input to projects-related decisions, and solve technical and complex applications problems. Send applications to: M.Gagne, Mailstation F110, Massachusetts Mutual Life Insurance Company, 1295 State Street, Springfield, MA 01111; Please Reference Job ID: 50306757.

SIEMENS PLM SOFTWARE INC. has an opening in Cincinnati, OH for Solutions Architect to provide Solutions Architect Consulting Srvcs for Teamcenter PLM implementation. Requires Bachelor's & 5 yrs exp. in Teamcenter PLM. Send resumes to PLMCareers@ugs.com. Job code UGS90 must be referenced in email subject line. EOE.

SIEMENS PLM SOFTWARE INC. has an opening in San Jose, CA for Software Product Consultant to provide technical implementation & consultation in architecture, customization, configuration, installation of Siemens PLM Teamcenter suite of PLM products. Send resumes to PLMCareers@ugs.com. Job code UGS87 must be referenced in email subject line. EOE.

hp®

invent

Hewlett-Packard State & Local Enterprise Services, Inc.
is accepting resumes for the following positions:

IT Developer/Engineer

Sacramento, CA. (Ref. #SLSACITDE11)

Research, design, develop, configure, integrate, test, and maintain existing and new business applications and/or information systems solutions including databases through the integration of technical and business requirements.

Information Testing

Tulsa, OK. (Ref. #SLTULASA1)

Design, develop and execute all testing-related activities on applications, infrastructure or hardware components of IT solutions, which include both third party software and internally developed applications and infrastructure.

Services Information Developer

Vancouver, WA. (Ref. #SLVANSID21)

Conceptualize, design, develop, unit-test, configure, or implement portions of new or enhanced (upgrades or conversions) business and technical software solutions through application of appropriate standard software development life cycle methodologies and processes.

Mail resume to Hewlett-Packard State & Local Enterprise Services, Inc., 5400 Legacy Drive, MS H1-6F-61, Plano, TX 75024. Resume must include Ref. #, full name, email address & mailing address. No phone calls please. Must be legally authorized to work in the U.S. without sponsorship. EOE.

invent

Hewlett-Packard Company has an opening for an

Technical Consultant IV

Bellevue, WA
Reference: BELPAL2

Reqs: Bachelor's & 10 yrs exp, and exp with installing, testing and trouble-shooting Telco products; HP OpenCall and/or related products; Software pre-sales consulting and presenting; Expertise in Telco Networks; and strong experience with AT&T wireless operations, networking & engineering. List full name, address & email address on resume. Send resume & refer to Job# BELPAL2.

Please send resumes with job number to Hewlett-Packard Company, H1-6E-28, 5400 Legacy Drive, Plano, TX 75024. No phone calls please. Must be legally authorized to work in the U.S. without sponsorship. EOE.

invent

Hewlett-Packard Company has an opening for an

Systems/Software Engineer

Houston, TX
Reference: HOUSRA2

Reqs: Bachelor's + 5 yrs exp. & exp w/ C or C++ devt exp on real time, multi-threaded interrupt driven SW. System ROM or storage FW QA exp; expert knowledge (devt & app) in multiple operating systems; must possess expertise in multiple programming languages & tech such as C, C++, Perl. Develop & debug Kernel-mode driver. Must be able to develop storage test tool that operates as Linux Kernel driver. List full name, address & email address on resume.

Please send resumes ref Job# HOUSRA2 to Hewlett-Packard Company, H1-6E-28, 5400 Legacy Drive, Plano, TX 75024. No phone calls pls. Must be legally authorized to work in the U.S. w/o sponsorship. EOE.

CLASSIFIED LINE AD SUBMISSION DETAILS: Rates are $400.00 per column inch ($500 minimum). Eight lines per column inch and average five typeset words per line. Free online listing on careers.computer.org with print ad. Send copy at least one month prior to publication date to: Marian Anderson, Classified Advertising, Computer Magazine, 10662 Los Vaqueros Circle, Los Alamitos, CA 90720-1314; (714) 821-8380; fax (714) 821-4010. Email: manderson@computer.org.

invent

Hewlett-Packard Company is accepting resumes for the following positions:

Mechanical/Hardware Engineer

San Diego, CA. (Ref. #SDMHE21)

Responsible for static and dynamic design and analysis of mechanical systems and equipment.

Systems/Software Engineer

Roseville, CA. (Ref. #ROSSSE31)

Conduct or participate in multidisciplinary research and collaborate with equipment designers and/or hardware engineers in the design, development, and utilization of electronic data processing systems software. Design, develop, troubleshoot, and debug software programs.

Engineering Program Manager

Boise, ID (Ref. #BOIEPM11)

Responsible for engineering programs through the coordination of multiple projects across multiple disciplines including, but not limited to, the different engineering job families. Manage the life cycle of the program ensuring that the deadlines are achieved, and day-to-day functions of the program and related projects take place as required.

IT Developer/Engineer

Houston, TX (Ref. #HOUITDE21)

Research, design, develop, configure, integrate, test, and maintain existing and new business applications and/or information systems solutions including databases through the integration of technical and business requirements.

Mail resume to Hewlett-Packard Company, 5400 Legacy Drive, MS H1-6F-61, TX 75024. Resume must include Ref. #, full name, email address & mailing address. No phone calls please. Must be legally authorized to work in the U.S. without sponsorship. EOE.

Apple is looking for qualified individuals for following 40/hr/wk positions. To apply, mail your resume to 1 Infinite Loop 84-REL, Attn: AEC Staffing-LJ, Cupertino, CA 95014 with Ref # and copy of ad. Job site & interview, Cupertino, CA. Principals only. EOE.

Video Coding & Processing Engineer [Ref#7077907]

Develop video coding and processing. Req.'s Master's degree, or foreign equivalent, in Electrical or Computer Engineering, Information Systems Engineering and Communications Engineering or related & One (1) year professional experience in job offered. Must have academic background or professional experience with: C/C++, Video Coding algorithm and H264 Standard, Video Codec development and optimization, Video/Image processing development and optimization, Intel/X86 SSE and Shell scripting.

Senior Network Engineer [Ref#7138193]

Responsible for looking across all network infrastructures. Requires Bachelors' degree, or foreign equivalent, in Electronics and Communications Engineering, or related Eight (8) years professional experience in job offered. Professional experience must be post-baccalaureate and progressive in nature. Also exp. with Cisco Routers and Switches; Routing Protocols/OSPF and BGP; Firewall/ ASA/FWSM; IP Addressing/Subnetting; Etherchannels VLAN and Layer 2; Firewall Architecture and DMZ; Load Balancing; and Cisco product Line 6500/4000.

SAP CRM ISA Developer [Ref#7140163]

Define, design and implement cutting-edge software and modify existing software to add features. Requires Master's degree, or foreign equivalent, in Computer Science, or related and 3 years professional experience in job offered or in a related occupation, including developing applications using SAP ISA framework; applications development and support using Java, J2EE, Struts and SAP NetWeaver application server; development of high available and heavy traffic web applications, SAP component based programming using NetWeaver Developer Studio IDE and SAP DTR; customizing SAP CRM web applications including b2b, useradmin, shopadmin and programming with SAP Indexing engine/TREX; SAP CRM ISA functional configuration and SAP CRM ISA related ABAP programming; and implementation of IPC Java user exits for Pricing and Variant Configuration.

Software Engineer - MobileMe [Ref#7151618]

Provide 24x7 production support for MobileMe services and back-end applications. Requires Bachelor's degree, or foreign equivalent, in Computer Science and Engineering, or related and 5 years professional experience in job offered or in a related occupation. Professional experience must be post-baccalaureate and progressive in nature. Also experience with software development using client-server architecture; UNIX concepts and strong exposure to Mac OSX; UNIX command line; scripting languages including Korn-Shell and Perl; Oracle RDBMS & PL/SQL; and Java and web technologies.

Software Engineer, Applications [Ref#7285236]

Analyze, design, code, inspect, debug, and test new software solutions in the security area with emphasis on C/C++ development. Requires Master's degree, or foreign equivalent, in Computer Science and Engineering, or related and 3 years professional experience in job offered or in a related occupation. Also experience with C/C++; algorithms; data structures, computer architecture, computer security; object oriented programming; operating systems; and embedded software development.

Hardware Development Engineer [Ref#7203090]

Perform NAND usage modeling and system specification definition for NAND. Requires Master's degree, or foreign equivalent, in Electrical Engineering, or related degree and 3 years professional experience in job offered or in a related occupation, including NAND flash memory system design, project management; design and release of consumer products, high volume manufacturing; NAND flash memory system design including NAND technology, NAND architectural design and NAND device; high volume manufacturing of consumer electronics; and with NAND flash memory.

Hardware Develop Engineer 3 [Ref#7359495]

Design antenna for mobile communication products. Requires Master's degree, or foreign equivalent, in Electromatnetics, Antennas, Communications, Physics, Electronic Engineering, or related and one (1) year professional experience in job offered or in a related occupation including 10-20% of international is needed to support manufacturing of Apple's mobile devices at contract manufacture factory. Also experience with Electromagnetics; Antenna Design; Antenna and RF Test and Measurement; Electrical/Electronics Engineering: principles of how to test and measure antennas.

Apple is looking for qualified individuals for following 40/hr/wk positions. To apply, mail your resume to 1 Infinite Loop 84-REL, Attn: AEC Staffing-LJ, Cupertino, CA 95014 with Ref # and copy of ad. Job site & interview, Cupertino, CA. Principals only. EOE.

Senior Macbook Hardware Design Engineer [Ref#7057727]

Design and develop Intel CPU based MacBook computers and their sub-components from concept to mass production. Requires Bachelor's degree, or foreign equivalent, in Electrical Engineering, or related field and 5 years professional experience in job offered or in a related occupation. Professional experience must be post-baccalaureate and progressive in nature. Also experience with digital and analog high-speed circuits; advanced characterization equipment including oscilloscopes; design, evaluation and characterization; PCB design best practices; layout inspection; and graphics card design and characterization.

iPod/iPhone Display Electrical Engineer [Ref#7056617]

Design and develop mobile display electronics to support liquid crystal displays development in system integration. Requires Ph.D. degree, or foreign equivalent, in Electrical Engineering, or related degree, including digital circuit design; FPGA HDL design; algorithms; analog circuit analysis; semiconductor physics; optoelectronic devices; and C/C++ programming.

Software Engineer, Applications [Ref#7056804]

Enhance features to streamline and automate day to day ad network operations. Requires Bachelor's degree, or foreign equivalent, in Computer Science, Electrical Engineering, Mathematics, or related field and 4 years professional experience in job offered or in a related occupation, including Unix type environments; relational databases; MySQL; Oracle; Java fundamentals; developing server type systems using Java; scripting languages including Python; domain knowledge of online and mobile ad networks; and analyzing large data sets.

Wireless System Engineer [Ref#7057883]

Support the design, integration, implementation, testing, and qualification of wireless subsystems of Mac computers. Requires Master's degree, or foreign equivalent, in Engineering, or related degree and academic background or professional experience with: communications systems; radio frequency and networking, and advanced wireless communication technology (MIMO, OFDM); wireless and hardware system design, implementation, and prototyping; electrical hardware engineering and digital circuit design; computer architecture design; and programming, data analysis and simulation skills.

ASIC Design Engineer [Ref#7057940]

Perform synthesis and Static Timing Analysis for blocks and full-chip using leading industry standard tools. Requires Bachelor's degree, or foreign equivalent, in Electronics and Communications Engineering, or related degree and 7 years professional experience in job offered or in a related occupation. Professional experience must be post-baccalaureate and progressive in nature. Also requires experience with static timing analysis; timing budgeting; synthesis; physical design including placement; clock tree synthesis, routing and timing & noise closure; and formal verification expertise.

Senior Supplier Quality Engineer [Ref#6995014]

Develop specific Product Quality Plan appropriate to program and commodity and coordinate all quality engineering functions for a program. Requires Master's degree, or foreign equivalent, in Mechanical Engineering, or related and 6 years professional experience in job offered or in a related occupation, including mechanical quality engineering, supporting manufacturing environments; tooling; fixtures; jigs; stamping; grinding; polishing; CNC; Milling; Injection Molding; Plastics; enclosures, casings; developing product quality plans; driving corrective actions and failure analysis efforts; mechanical quality and reliability concepts; general manufacturing operations; project management; and DFx and DOE principles.

RF Systems Engineer [Ref#7077345]

Work closely with RF Design Engineers, Electrical Engineers, SW Engineers, ATE Engineers, and Antenna Engineers to define, refine and perform RF evaluation for communication RF systems, including RF co-existence evaluation. Requires Master's degree, or foreign equivalent, in Electrical Engineering, or related and 1 year professional experience in job offered or in a related occupation, including Bluetooth RF standard requirements and test tools; RF lab equipment; power meters; spectrum analyzers; vector signal generators/analyzers; RF wireless systems and microwave devices and systems; digital signal processing; modem design; Bluetooth protocol; TCL/Perl Scripting; and VLSI system design.

Grid Pioneer Foster Named 2011 Kanai Award Winner

The University of Chicago's Ian T. Foster has been named winner of the 2011 IEEE Computer Society Tsutomu Kanai Award for his accomplishments in grid computing. The primary focus of Foster's research has been the acceleration of discovery in a networked world. In partnership with many others, notably Carl Kesselman and Steven Tuecke, Foster developed and promulgated the concepts and methods that underpin grid computing. These methods allow computing to be delivered reliably and securely on demand, as a service, and permit the formation and operation of virtual organizations linking people and resources worldwide.

Foster won the Kanai Award "for pioneering research in grid computing, integrating geographically distributed instruments, computers, and data."

Foster is the Arthur Holly Compton Distinguished Service Professor of Computer Science at the University of Chicago and an Argonne Distinguished Fellow at Argonne National Laboratory. He is also the director of the Computation Institute, a joint effort of Argonne and the University of Chicago.

Foster received a BS from the University of Canterbury, New Zealand, and a PhD from Imperial College, United Kingdom, both in computer science. He is a fellow of the American Association for the Advancement of Science, Association for Computing Machinery, and British Computer Society.

Ian T. Foster led the creation of Globus Online

His other awards include the Global Information Infrastructure Next-Generation Award, the British Computer Society's Lovelace Medal, *R&D Magazine*'s Innovator of the Year, and honorary doctorates from the University of Canterbury, New Zealand, and CINVESTAV, Mexico. Foster also cofounded Univa, a company established to deliver grid and cloud computing solutions.

TSUTOMU KANAI AWARD

The Tsutomu Kanai Award recognizes major contributions to state-of-the-art distributed computing systems and their applications. Established by the IEEE Computer Society in 1997 through a generous endowment from Hitachi, the award is named in honor of Tsutomu Kanai, who served as Hitachi's president for 30 years.

The Kanai award consists of a crystal model, certificate, and $10,000 honorarium. Read more about the Kanai and other Computer Society awards at www.computer.org/portal/web/awards/kanai.

Astrin Receives Hans Karlsson Award

Communication engineer Arthur W. Astrin, whose innovations and efforts contributed to the birth and development of the Wi-Fi industry, was recently named winner of the 2011 IEEE Computer Society Hans Karlsson Award.

Astrin received the award "for leadership and diplomatic skills applied to LAN/MAN wireless personal area network standards; mediating rivalry of competing corporate entities and personal aspirations by promoting the value of IEEE wireless standards-based approaches."

Astrin has held technical and management positions at Apple, IBM, Siemens, ROLM, Memorex, and Citicorp. At Apple, he boosted the Wi-Fi industry by developing AirPort, the first consumer-oriented, wireless networking solution for PCs. He also worked toward creating industry compatibility in testing compliance for the IEEE 802.11 Wi-Fi standard. He has taught communication and computer engineering at San Jose State University and the University of California, Berkeley.

Arthur W. Astrin chairs Body Area Network Task Group 6 for the IEEE 802.15 standard.

In 1967, Astrin built the first "bit-slice" architecture computer using transistor-transistor logic technology and received an award from the US Navy's Admiral Grace Hopper. He holds seven US patents and received a master's degree in mathematics from the University of California, San Diego, and a PhD in communication engineering from the University of California, Los Angeles.

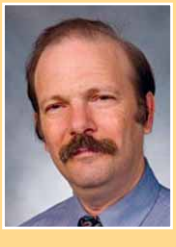
HANS KARLSSON AWARD

Established in 1992 in memory of Hans Karlsson, chairman and father of the IEEE 1301 family of standards, the Karlsson Award is one of the Computer Society's highest honors. The award recognizes outstanding skills and dedication to diplomacy, team facilitation, and joint achievement in the development or promotion of standards in the computer industry where individual aspirations, corporate competition, and organizational rivalry could otherwise be counter to the benefit of society. Learn more about the Karlsson Award and other Computer Society awards at www.computer.org/portal/web/awards/karlsson.

Vardi Receives Goode Award

Rice University computer scientist Moshe Vardi has been named winner of the IEEE Computer Society's 2011 Harry H. Goode Award.

Vardi, Rice's Karen Ostrum George Professor in Computational Engineering and director of Rice's Ken Kennedy Institute for Information Technology, is a noted logician and member of the National Academy of Engineering. He also serves as editor in chief of the Association of Computing Machinery's *Communications of the ACM*. Vardi was honored for his "fundamental and lasting contributions to the development of logic as a unifying foundational framework and a tool for modeling computational systems."



Moshe Vardi is a noted logician and member of the National Academy of Engineering.

Logic, which is sometimes called "the calculus of computer science," is fundamental to research areas such as artificial intelligence, computational complexity, distributed computing, database systems, design verification, programming languages, and software engineering. Using logic as a framework, Vardi has performed research in intelligent databases, multiagent systems, and automated reasoning.

He was honored with the 2010 Outstanding Contribution to ACM Award for his leadership, including the organization of an influential 2006 report on overseas job out-

sourcing in the software industry. The report dispelled some myths about software offshoring and reinforced the case that computing plays a fundamental role in defining success in the global economy.

Vardi earned a doctorate from the Hebrew University of Jerusalem and is the author or coauthor of approximately 400 articles and two books. Vardi's other awards include the 2010 Distinguished Service Award from the Computing Research Association, the 2000 Goedel Prize for outstanding papers in the area of theoretical computer sciences, and the 2008 ACM Presidential Award. He is a member of the American Academy of Arts and Science, the European Academy of Sciences, and the Academia Europea. Vardi is also a Guggenheim Fellow and a fellow of IEEE, the ACM, the American Association for the Advancement of Science, and the Association for the Advancement of Artificial Intelligence.

HARRY H. GOODE AWARD

The Goode Award was established to recognize achievement in the information-processing field—either a single contribution of theory, design, or technique of outstanding significance or the accumulation of important contributions on theory or practice over an extended period.

Further information about the Goode Award, including a list of past recipients, can be found at: www.computer.org/portal/web/awards/harrygoode.

CS&E Undergraduate Numbers Up

Total enrollments among US computer science undergraduates increased 10 percent in 2010, according to the Computing Research Association's annual Taulbee Survey. This is the third straight year of increases in total enrollment and may indicate that the post "dot-com crash" decline in undergraduate computing program enrollments is over.

The CRA, of which the IEEE Computer Society is a member, conducts the survey each year to document trends in student enrollment, degree production, employment of graduates, and faculty salaries in PhD-granting departments of computer science, computer engineering, and information sciences in the United States and Canada.

SPECIFIC SURVEY FINDINGS

Overall, bachelor's degree production in computer science, computer engineering, and information sciences departments in 2010 rose nearly 11 percent compared to 2009. Bachelor's degree production in computer science departments was up more than 9 percent. The increases in new students observed during each of the past two years have resulted in increased degree production, a turnaround from the past several years of declining bachelor's degree production.

The 2010 Taulbee Survey also found that PhD production in computing programs held steady in 2009-2010. Among CRA member schools, the share of bachelor's degrees in computer science granted to females rose to 13.8 percent in 2010, an increase of 2.5 percentage points over 2009. The share of bachelor's degrees in computer science granted to minority students held nearly steady at 10.3 percent in 2010.

The full report, which also includes information about faculty size, demographics and salaries, graduate student support, and research expenditures, will be available soon on the CRA website.

COMPUTING RESEARCH ASSOCIATION

The CRA is an association of more than 200 North American academic departments of computer science, computer engineering, and related fields; laboratories and centers in industry, government, and academia engaging in basic computing research; and affiliated professional societies.

The Taulbee Survey is named after the University of Pittsburgh's Orrin E. Taulbee, who conducted the surveys from 1974 to 1984 for the Computer Science Board (the predecessor organization to the Computing Research Association). Learn more about the CRA and its programs at www.cra.org.

CALLS FOR ARTICLES FOR *COMPUTER*

Computer seeks submissions for a September 2011 special issue on security and privacy in an online world.

We are heading for a world in which convenience goods have unprecedented computing power and are permanently connected to the Internet. This not only changes our way of living, but also the way we approach security and privacy. In such an online world, many traditional security mechanisms no longer work, mainly because it is increasingly difficult if not impossible to separate the trusted inside from the untrusted outside.

Against this background, there is a large gap between the general knowledge and wisdom in computer security and privacy and what is really needed in practice. This special issue of *Computer* will explore alternatives to the general knowledge and wisdom in computer security and privacy, particularly with regard to their applicability to an online world.

Appropriate topics of interest include security challenges and privacy concerns related to an online world; the future of perimeter protection; economic models for security and privacy; and next-generation security technologies.

Articles are due by **1 June**. Visit www.computer.org/portal/web/computingnow/cocfp9 to view the complete call for papers.

CALLS FOR ARTICLES FOR IEEE CS PUBLICATIONS

IEEE Internet Computing plans a May/June 2012 special issue on infrastructures for online social networking services.

The proliferation of rich social media, online communities, and collectively produced knowledge resources has accelerated the convergence of technological and social networks, resulting in a dynamic ecosystem of online social networking services, environments, and applications. OSN sites' success is reshaping the Internet's structure, design, and utility. Moreover, this trend is creating numerous challenges and opportunities for the development, deployment, management, and operation of scalable, secure, interoperable OSN infrastructures that can sustain a cycle of innovative application development, improved end-user experience, high-quality service provision, privacy protection, and healthy market expansion.

IC's guest editors seek recent research results in systems, software, and services that provide novel ubiquitous, scalable, secure, and trustworthy OSN infrastructures.

Articles are due by **1 September**. Visit www.computer.org/portal/web/computingnow/iccfp3 to view the complete call for papers.

IT Professional invites submissions for a January/February 2012 special issue on cloud computing.

Cloud computing has become a scalable-services consumption and delivery platform for services computing. Its technical foundations include service-oriented architectures and virtualizations of hardware and software. It aims to share resources among cloud-service consumers, cloud partners, and cloud vendors in cloud value chains.

The guest editors seek submissions on all aspects of cloud computing, including infrastructure, software, or application clouds; service-oriented architectures in cloud computing; virtualization of hardware or software resources; and cloud computing consulting methods.

Articles are due by **1 June**. Visit www.computer.org/portal/web/computingnow/itcfp6 to view the complete call for papers.

SUBMISSION INSTRUCTIONS

The Call and Calendar section lists conferences, symposia, and workshops that the IEEE Computer Society sponsors or cooperates in presenting.

Visit www.computer.org/conferences for instructions on how to submit conference or call listings as well as a more complete listing of upcoming computer-related conferences.

CALENDAR

JUNE 2011

19-25 June: CVPR 2011, IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, Colorado Springs, Colorado; http://cvpr2011.org

Published by the IEEE Computer Society

EVENTS IN 2011

June

19-25 ..... CVPR 2011
21-24 ..... ICDCS 2011

July

11-15 ..... ICME 2011
18-22 ..... COMPSAC 2011
18-22 ..... SAINT 2011
25-27 ..... ARITH 2011

August

1 ..... Hot Chips 2011
15-18 ..... ICGSE 2011
25-27 ..... NCA 2011

SRDS 2011

The Symposium on Reliable Distributed Systems is a forum that brings together researchers and practitioners interested in distributed systems design, development, and evaluation, with particular emphasis on reliability, availability, safety, security, trust, and real-time applications.

Conference organizers have solicited original research papers and practical experience reports that deal with the design, development, and experimental results of operational systems.

SRDS 2011 is sponsored by the IEEE Computer Society Technical Committee on Distributed Processing in collaboration with Spain's Universidad Politécnica de Madrid. SRDS 2011 takes place 4-7 October in Madrid. Visit http://lsd.ls.fi.upm.es:11888/srds for complete conference details.

21-24 June: ICDCS 2011, 31st Int'l Conf. on Distributed Computing Systems, Minneapolis; www.seas.gwu.edu/~cheng/ICDCS2011

JULY 2011

11-15 July: ICME 2011, IEEE Int'l Conf. on Multimedia and Expo, Barcelona, Spain; www.icme2011.org

18-22 July: COMPSAC 2011, IEEE Int'l Computer Software and Applications Conf., Munich; http://compsac.cs.iastate.edu/

18-22 July: SAINT 2011, Symp. on Applications and the Internet (with COMPSAC), Munich; www.saintconference.org

25-27 July: ARITH 2011, Symp. on Computer Arithmetic, Tubingen, Germany; www.ac.usc.es/arith20

AUGUST 2011

1 Aug: Hot Chips 2011, 23rd Symposium on High-Performance Chips, Palo Alto, California; www.hotchips.org

15-18 Aug: ICGSE 2011, IEEE Int'l Conf. on Global Software Eng., Helsinki, Finland; http://icgse2011.soberit.hut.fi

25-27 Aug: NCA 2011, 10th IEEE Int'l Symp. on Network Computing and Applications, Cambridge, Mass.; www.ieee-nca.org

SEPTEMBER 2011

11-14 Sept: ASAP 2011, 22nd IEEE Int'l Conf. on Application-Specific Systems, Architectures, and Processors, Santa Monica, California; http://asap-conference.org

14-16 Sept: ECOWS 2011, 9th IEEE European Conf. on Web Services, Lugano, Switzerland; http://ecows2011.inf.usi.ch

18-23 Sept: ITC 2011, Int'l Test Conf., Anaheim, California; www.itctestweek.org

25 Sept-1 Oct: ICSM 2011, 27th IEEE Int'l Conf. on Software Maintenance, Williamsburg, Virginia; www.cs.wm.edu/icsm2011

26-30 Sept: Cluster 2011, IEEE Int'l Conf. on Cluster Computing, Austin, Texas; www.tacc.utexas.edu/ieee/2011/index.php

OCTOBER 2011

4-7 Oct: LCN 2011, 36th IEEE Conf. on Local Computer Networks, Bonn, Germany; www.ieeelcn.org/index.html

4-7 Oct: SRDS 2011, 30th IEEE Int'l Symp. on Reliable Distributed Systems, Madrid, Spain; http://lsd.ls.fi.upm.es/srds2011

10-14 Oct: PACT 2011, 20th Int'l Conf. on Parallel Architectures and Compilation Techniques, Galveston, Texas; http://pactconf.org

22-25 Oct: FOCS 2011, 52nd IEEE Symp. on Foundations of Computer Science, Palm Springs, California; www.cs.ucr.edu/~marek/FOCS11

The Developer's Perspective

Christopher L. Huntley
Fairfield University

A compelling tech story has five elements: people, projects, problems, products, and practices.

Systems development remains one of the most potentially rewarding professions, directly or indirectly relating to many of the top careers in Money/PayScale.com's annual rating of the best jobs in America (http://money.cnn.com/magazines/moneymag/bestjobs). In 2009, systems engineer was ranked number one, followed by IT project manager (5), software developer (12), software product manager (16), and IT business analyst (17). In 2010, software architect topped the list, followed by database administrator (7), information security engineer (17), software engineering/development director (18), and information technology manager (20).

As the economy continues to recover, recruiting top talent in technical corridors such as Silicon Valley, Boston, and Northern Virginia has become extremely competitive. A recent *New York Times* article quoted one young developer as saying that "recruiting in Silicon Valley is more competitive and intense and furious than college football recruiting of high school athletes" (C.C. Miller and J. Wortham, "Silicon Valley Hiring Perks: Meals, iPads and a Cubicle for Spot," 25 Mar. 2011). Companies are going beyond the usual high salaries, stock options, and lifestyle perks to offer more strategic enticements like providing assistance to acquire venture capital for new start-ups, even start-ups that might someday be competitors.

Despite this apparent demand for engineering talent, interest in technical careers in the US has sharply declined in recent years. Many of our most promising young people are choosing medicine, law, finance—in fact, just about anything other than engineering or computer science.

The problem has become so pronounced that President Obama made it the centerpiece of his 2011 State of the Union address. He argued that to "win the future" we must re-create the conditions for technical innovation in the US, a traditional strength that has been challenged by global competition in the past few years. While other countries produce more talented engineers and entrepreneurs, the US share is dropping precipitously.

What is driving young people in the US to other professions? And why are so many elsewhere in the world becoming engineers?

A PERFECT STORM

When I ask engineering friends from the 1990s tech boom generation why they became interested in technology, many cite Tracy Kidder's *The Soul of a New Machine* (Little, Brown and Company, 1981) as being formative in their career choices. In addition to being the winner of the 1982 Pulitzer Prize, it's also a collection of wonderful engineering war stories.

The book chronicles the development of a new minicomputer at Data General in the 1970s. The characters were compelling and had their own unique perspectives, and the projects they undertook along the way were complex and challenging. Descriptions of the various parts of the computer itself captivated budding engineers by offering visions of the technology they might someday help to develop. The failures in the book were also interesting, showing how professionals did their work even when things didn't go exactly right.

In his introduction to the book, Kidder related a parable about a ship caught in a summer squall that threatened to capsize it. Like any good sea story, at the center of it was a hero, Tom West, a volunteer crew member who seemed to enjoy the

 Published by the IEEE Computer Society

weather and get more cheerful as the boat got whipped around by the rising storm. By storm's end, West was at the helm while the captain rested in the cabin. West hadn't slept for four days. When discussing the adventure back in port, he explained his ability to stay focused and energized by saying, "I build computers."

Imagine the impact this story had on young engineers, including myself, who were just starting to plan their working lives. Building computers was the kind of work that made you resourceful, productive, and powerful.

Another book published a generation later at the height of the late 1990s tech boom, *The New New Thing: A Silicon Valley Story* by Michael Lewis (Penguin, 2001), also proved to be a page-turner among enterprising young engineers. This time, however, the engineer at the center of the book, Jim Clark, was more intent on building companies than building great technology.

Clark's measure of success wasn't a superior product but the billions he made before the game was over. It's telling that he also worked on a boat and had troubles at sea. However, Clark had one of the largest yachts in the world, and his strategy for coping with bad weather was to use his enormous wealth to get the best technology (and engineers) that money could buy.

While Lewis's story was certainly entertaining, it didn't speak to aspiring engineers in the way that Kidder's book did. Business had supplanted engineering as the focus. After reading the book, young people were more likely to want to be entrepreneurs than engineers.

If we are to continue to attract new talent to systems development, we need to recapture some of the magic of *The Soul of a New Machine*. The stories we seek need not be Pulitzer Prize-winning novels. In fact, given the nature of media these days, provocative blog-style commentaries and interviews with people doing interesting work may be preferable anyway.

ELEMENTS OF A GOOD TECH STORY

A compelling tech story has five elements: people, projects, problems, products, and practices.

People

First and foremost, the story must capture more than just the technical and logistical details of project X. Stereotypes aside, engineering is an inherently human activity that serves the needs of people using technology.

Ask developers about their work, and they will invariably describe multiple projects in varying states of maturity.

Whenever I read an article about a programmers' group meeting somewhere, I always enjoy the photos of attendees gathered around a table to participate in a friendly discussion. I imagine the stories they must be telling and what I could take from them. While I don't necessarily want to know details of their personal lives, I like reading about their work and future plans.

I likewise enjoy watching developers interact with users, whether through the bug database or, even better, through video of user experience tests to see how people react to new problems and opportunities. I can't help but think through these problems on my own, and seeing others work out solutions is both interesting and instructive.

Projects

Ask developers about their work, and they will invariably describe multiple projects in varying states of maturity. Nobody seems to work on just one thing; there are always at least three others in the hopper, some official and some personal.

Interesting projects run the gamut from rough concepts and theories yet to be teased out, to products under active development, to consulting work on other people's efforts. All of these can be compelling, provided the account goes beyond deadlines and status reports. After all, a story is more than just a plotline.

Problems

Every epic tale has some sort of conflict or other dynamic element that guides and embellishes the narrative. Without it there aren't any heroes or villains, no story arc to follow from beginning to end. Developer stories are no different.

Sometimes the problems belong to the clients, who often don't even recognize a problem until it's "miraculously" solved for them. Other times the problems involve design of the product itself, with the developer puzzling over how to meet a given requirement with the technology at hand. Still other times the problem is how the project team interacts with management or perhaps one another: How do you get X people to be productive and sufficiently motivated to complete the project? How do you explain to management that a requirement is infeasible or even counter-productive despite it showing up on a competing product's feature list?

Products

Toolies love their products. When I first stumbled upon open source software directories like Freshmeat or SourceForge years ago, I remember spending days browsing through the various projects to see what everyone was working on. I loved it that there might actually be five or more different ways to collect e-mail from my disparate accounts or that I could choose from among seven or eight different text editors. Even the oper-

ating system wars were interesting, although I found myself on both sides of the battle lines.

From a developer's perspective, there are probably two kinds of products that tell a good story. The first is a product that solves a challenging problem. These sorts of products are more of general interest, although sometimes analysis of the problem can be engaging and instructive. The second and more compelling kind of product is one that solves a problem the developer currently has or expects to have in the near future. After all, wasn't scratching our own itches part of the magic of open source development?

Practices

Despite what managers (or professors) might say, developers are usually pretty quick to learn new things. They wade through a multitude of issues every day and take away something that makes them better developers. Sometimes what they learn is technical, but often it's about how to manage their work with others or even by themselves.

It's from this sort of independent learning process and the sharing of black arts that all methodology, good or bad, derives. Consider, for example, the progression from the waterfall models of 40 years ago to today's agile/lean development practices.

When early developers were asked to deliver software on schedule, they found that having a plan was useful and that it was usually easier to explain the plan to management if it involved a fixed set of steps or stages to track progress. After struggling to make this work for several decades, developers discovered it was sometimes possible to adopt a more agile/iterative approach as long as the overall project risk was reduced.

This heavy-to-agile progression itself makes a good story, as I learned from Robert C. Martin at one of his professional seminars a decade ago. Martin had a room full of engineers enraptured as he talked about how programmers in the old days would construct a plausible project plan instead of an accurate one that would account for the inevitable dead ends and disasters that could delay development by weeks or more. He also described the pressures to get the plan "right" even when the engineers knew that any plan was fiction. The punch line of the story, of course, was the Agile Manifesto (http://agilemanifesto.org), which has colored so much of the systems development discussion over the past 10 years. Martin got our attention by telling us a good story and then used it to make us better engineers. We need more of that.

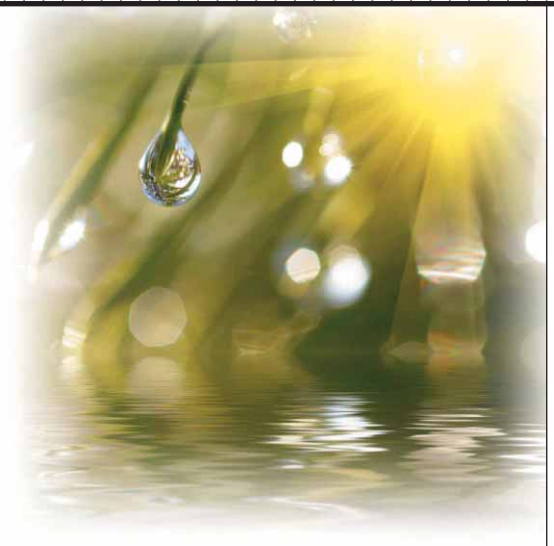
This new bimonthly *Computer* column will include stories, commentaries, and interviews that are of particular interest to developers. Whenever possible, it will provide firsthand accounts, in the engineer's own words.

As column editor, I intend to seek out interesting blogs, podcasts, and other sources for material, but I also encourage developers to provide feedback and contribute a good story or two. After all, who doesn't enjoy a good tech story?

***Christopher L. Huntley**, In Development column editor, is a professor and chair of the Department of Information Systems and Operations Management at the Charles F. Dolan School of Business at Fairfield University. Contact him at chuntley@mail.fairfield.edu.*

GREEN IT

Saving the World, One Server at a Time, Together



Parthasarathy Ranganathan and Jichuan Chang,
HP Labs

Achieving the next order-of-magnitude improvement in power management requires a holistic approach to how we study and reason about energy efficiency and how we optimize system design across traditional implementation boundaries.

From mobile devices to warehouse-size datacenters, power and energy are important system design considerations, impacting electricity consumption, the ability to cost-effectively cool the system, and overall reliability. For example, the annual electricity cost to power and cool cloud datacenters can be several million dollars, exceeding even the hardware cost.

This emphasis on power efficiency is already visibly impacting computing system designs, the recent shift to multicore processors being a notable example. There is also growing recognition that true energy-optimized designs must go beyond computing equipment to include the supporting infrastructure. For example, in many datacenters, every watt of power the server consumes can lead to an additional half-watt of power consumption by the power delivery and cooling infrastructure. Consequently, there is a new focus on total energy consumption across the entire IT infrastructure.

Environmental sustainability—minimizing the environmental impact of natural resource consumption or the production of unwanted carbon emissions—is also emerging as a central issue. Industry analyst firm IDC estimates that the electricity consumption associated with IT systems is responsible for 2 percent of total global carbon emissions, exceeding that of the entire aviation industry. But more importantly, IT is increasingly being used to address the remaining 98 percent of carbon emissions from other industries—for example, video conferencing to avoid travel and cloud services to economize logistics.

In response to these trends, major IT vendors including Hewlett-Packard, IBM, and Intel are investing heavily in sustainability. Government agencies such as the US Environmental Protection Agency are also becoming involved by legislating and incentivizing sustainable solutions. Google engineers have even estimated the carbon footprint of a single search query.

While there has been much progress on power and energy management, achieving the next order-of-magnitude improvement requires rethinking how we study and reason about energy efficiency to optimize system design. This involves crossing traditional design boundaries to look at all design levels, from chips to datacenters; at all aspects of computing science, from hardware to integration to applications; and at research from other disciplines, including materials science, mechanical engineering, and environmental engineering. Understanding and optimizing energy holistically can yield rich dividends.

UNDERSTANDING ENERGY HOLISTICALLY

It will be critical for green IT researchers to consider both the demand side and supply side of energy consumption of future systems. The demand side of energy consumption refers to the traditional metric of operational energy consumption, and

0018-9162/11/$26.00 © 2011 IEEE Published by the IEEE Computer Society

green IT researchers already study this. However, the supply side—the energy used to make a system—has already become a dominant component in mobile devices such as smartphones and tablets, and it is fast becoming important in servers as well.

Life-cycle analysis (LCA) uses an end-to-end approach to assess a system's environmental impact across various stages in its life cycle, including raw materials extraction, manufacturing, transportation, operation, and disposal. Taking LCA a step further, recent green IT studies have examined the thermodynamic notion of *exergy*—the amount of available energy in a system—as a possible metric. Studying exergy lets system designers reason about renewable energy and material recycling. It also enables using a large body of public information on exergy associated with materials usage and common thermal, chemical, metallurgical, manufacturing, and transportation processes.

Such holistic energy metrics can help system designers identify instances in which the most efficient design for operational energy might not produce the most sustainable solution. It also helps them discover new optimizations that address the energy used to make the system (J. Chang et al., "Green Server Design: Beyond Operational Energy to Sustainability," *Proc. 2010 Int'l Conf. Power Aware Computing and Systems*, Usenix Assoc., 2010; www.usenix.org/events/hotpower10/tech/full_papers/Chang.pdf). This is a relatively nascent research area, however, and more sustainability metrics are needed.

Similarly, a systematic approach is required to understand common sources of energy inefficiencies and ways to remedy them. A recent study identified four key aspects of common design practices that lead to energy inefficiencies and proposed a framework to systematically reason about energy efficiency improvements (P. Ranganathan, "Recipe for Efficiency: Principles of Power-Aware Computing," *Comm. ACM*, Jan. 2010, pp. 60-67). In addition to well-studied approaches such as energy proportionality, this framework recommended less common optimizations such as cross-layer designs for improved energy savings, trading off uncommon-case efficiency to improve common-case efficiency, and spending power to save power.

Again, more research is needed in developing such frameworks. Indeed, the 2009 SciPM Workshop (http://scipm.cs.vt.edu), sponsored by the National Science Foundation, posited that a new science of power management can play a key role across the broader computer science community.

Green IT researchers must consider both the demand side and supply side of energy consumption of future systems.

OPTIMIZING ENERGY HOLISTICALLY

It is also important to optimize energy holistically. Scheduling across an ensemble of systems or system components, cooling-aware scheduling (such as temperature-aware workload placement), and exchanging information across the network stack can all increase energy efficiency. Three examples illustrate the advantages of holistic optimization.

No 'power struggles'

Recent power management solutions for datacenters have been multifaceted, addressing different dimensions—for example, average power reduction to lower electricity costs; peak power management to reduce cooling and power delivery infrastructure costs; local resource management; distributed resource scheduling; and virtual machine migration.

Other dimensions include where the power management solution is implemented—at the chip, server, cluster, or datacenter level or in hardware, software, or firmware—and specific properties of the "knobs" used to control power, such as processor voltage scaling at fine time granularities and turning systems off at coarser time granularities.

The behavior that emerges from this collection of optimizations might or might not be globally optimal, stable, or even correct. Individual solutions can, however, synergize to better address the dynamic and diverse nature of typical datacenter workloads.

A 2008 collaborative effort involving computer scientists, thermomechanical engineers, and control engineering experts identified one such coordination architecture to allow individual power management agents to learn and react to the effects of other agents (R. Raghavendra et al., "No 'Power Struggles': Coordinated Multilevel Power Management for the Data Center," *Proc. 13th Int'l Conf. Architectural Support for Programming Languages and Operating Systems*, ACM Press, 2008, pp. 48-59).

The proposed architecture achieved 65 percent energy savings with minimal performance loss compared to a state-of-the-art baseline, and resulted in more energy efficiency than if the solutions were individually deployed. It also enabled formal mathematical analysis of stability and provided better flexibility to respond to dynamic changes in the controllers and system environments.

Subsequent studies have demonstrated similar benefits from other types of cross-layer coordination—for example, across the applications and systems layers.

Dematerialized datacenters

A recent project on sustainability-aware datacenter design brought together computer engineers, environmental engineers, and mechanical engineers (J. Meza et al., "Lifecycle-Based Data Center Design," *Proc. ASME 2010 Int'l Mechanical Eng. Congress and Exposition*, ASME, 2010; www.hpl.hp.com/techreports/2010/HPL-2010-117.pdf). The researchers found that many intuitive efforts to apply the principles of sustainability—reduce, reuse, recycle—often lead to nonintuitive tradeoffs between materials usage, system power consumption, cooling infrastructure efficiency, and performance.

Consider, for example, dematerialization. Reducing the amount of physical materials, such as sheet metal, in building systems improves sustainability but can impact cooling efficiency by changing air flow patterns. Likewise, using an optical backplane instead of electrical signaling to reduce material can affect the network topology, and the altered system power budget may require use of low-power servers.

Consequently, sustainability-aware datacenter designs should iterate across the system architecture, cooling infrastructure, and physical organization and packaging. The researchers concluded that such a holistic codesign can achieve 40 percent improvement.

From microprocessors to nanostores

A cover feature in the January 2011 issue of *Computer* discussed a cross-disciplinary effort by researchers in device technologies, computer architecture, and system software to manage future data-centric workloads (P. Ranganathan, "From Microprocessors to Nanostores: Rethinking Data-Centric Systems," pp. 39-48). The article examined traditional designs using deep memory/storage hierarchies and argued that the emergence of new nonvolatile memory technologies, such as memristors and phase-change RAM, offers a unique opportunity to rearchitect systems.

For example, *nanostores*—single-chip systems with 3D-stacked nonvolatile memory with a layer of compute cores and a network interface—eliminate nonessential cache layers and provide balanced energy-efficient computation colocated with the data. Nanostore-based distributed system designs promise up to an order of magnitude more energy efficiency, but these benefits are achievable only by designing the technology, architecture, and software layers holistically.

In particular, reexamining decades-old assumptions about magnetic disk latencies and redesigning the software stack to leverage the persistent byte-addressable memory interface of nonvolatile memory can eliminate significant complexity and associated energy inefficiencies.

The results from these studies show promise, but we have only scratched the surface of what is possible. Several decades ago, Richard Feynman estimated the theoretical limits on the power costs for information transfer to be 10^{18} bit operations per second for one watt of power. That translates to thousands of cloud datacenters (each with the computational power of a Google "warehouse computer") packed into a single cell phone. While this bound is theoretical, it points to the tremendous unrealized opportunities to improve energy efficiency and promote green IT.

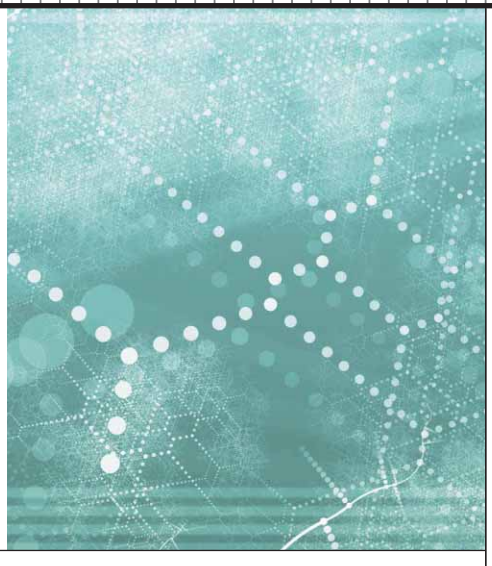
Researchers are just starting to explore cross-disciplinary innovation in energy and power management. Other rich opportunities for optimization include innovative hardware-software codesigns (for example, new interfaces and abstractions for vertically integrated solutions) and unconventional system designs (for example, biologically inspired "brain" computing). By continuing to cut across traditional design boundaries—across different layers of the design, across different fields in computer science, and across different disciplines—we can do even better. The future is bright and exciting, and green.

Parthasarathy Ranganathan *is a distinguished technologist at HP Labs, where he is the principal investigator for the data-centric datacenter project. Contact him at partha.ranganathan@hp.com.*

Jichuan Chang *is a senior research scientist at HP Labs. Contact him at jichuan.chang@hp.com.*

Editor: Kirk W. Cameron, Dept. of Computer Science, Virginia Tech; greenit@computer.org

Matching Forensic Sketches and Mug Shots to Apprehend Criminals



Anil K. Jain and Brendan Klare,
Michigan State University

Emerging face recognition technology can use forensic sketches to help identify criminal suspects.

Dating back to the late 1800s, law enforcement agencies have used forensic sketches to help identify unknown culprits. A forensic sketch is a drawing by a specially trained artist of a suspect's face that is based solely on a verbal description by a witness or victim. Figure 1a shows a modern example of one such sketch of Jack the Ripper, the famous serial killer who terrorized London's Whitechapel district in 1888. Forensic sketches are a valuable tool for visually representing a subject when no other medium exists for capturing his or her face—for example, a surveillance camera.

Even to this day, forensic artists routinely generate sketches to help authorities apprehend criminals—it's not uncommon to see a forensic sketch in a newspaper or on the nightly news. In fact, the method for using forensic sketches has remained unchanged for decades: disseminate the sketch to relevant law enforcement agencies and media outlets with the hope that an officer or citizen will recognize the individual portrayed.

However, this approach is both expensive and time-consuming. Given the often egregious nature of crimes committed by culprits depicted in forensic sketches—including murder, terrorism, sexual assault, and armed robbery—failing to quickly capture them can have severe consequences. Improving forensic sketch recognition would greatly increase public safety.

Under the broad umbrella of biometric recognition, a new paradigm has emerged for identifying suspects using forensic sketches. A sketch can be converted to a digital image and then automatically matched against mug shots (arrestee photos) and other face images in a database—for example, drivers' license photos—to help make an identification. This automated approach, enabled by progress in computer vision and machine learning algorithms, offers a valuable resource to authorities seeking to accurately and quickly capture dangerous criminals.

AUTOMATIC FACE RECOGNITION

During the past decade, the accuracy of automatic face recognition systems increased dramatically, as benchmarked in a series of assessments by the National Institute of Standards and Technology (P.J. Phillips, "Improving Face Recognition Technology," *Computer*, Mar. 2011, pp. 84-86). However, these systems are highly sensitive to variations in facial pose and expression as well as environmental factors such as ambient illumination.

Face images acquired from mug shots and ID cards such as drivers' licenses and passports are generally well suited for automatic face recognition because the acquisition conditions and subject's cooperation minimize pose, expression, and lighting variations. This has prompted approximately half of all US department of motor vehicle (DMV) agencies to implement face recognition technology. While the primary use of face recognition by DMVs is to prevent the issuance of multiple drivers' licenses to the same person (deduplication), legislation in many states also permits law enforcement agencies to access the DMV face database.

Between DMV and mug shot databases, law enforcement agencies often have access to digital images of a large percentage of their jurisdiction's population. Unfortunately, commercial off-the-shelf (COTS) face recognition systems that have high

 Published by the IEEE Computer Society

levels of success matching two face photos perform poorly in matching forensic sketches to mug shots.

IMPROVING SKETCH RECOGNITION

Matching a hand-drawn sketch to a photograph is difficult because

- a subject's facial appearance may change between the time the photo was taken and the sketch is made, and
- incomplete and inconsistent witness descriptions can result in inaccurate facial depictions.

Improving forensic sketch recognition requires a feature extractor and matcher specifically designed for this task. At Michigan State University, we recently developed such a system, which demonstrated a four times improvement in accuracy over a leading COTS face recognition system (B. Klare, Z. Li, and A.K. Jain, "Matching Forensic Sketches to Mug Shot Photos," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Mar. 2011, pp. 639-646).

As Figure 2 shows, our system first partitions sketch and face images into N slices. It then computes scale-invariant feature transform (SIFT) and multiscale local binary pattern (MLBP) descriptors for each slice, which remain stable between sketches and photos. Next, it uses local-feature-based discriminant analysis (LFDA) to extract the most salient features for each slice. Finally, the system measures the similarity between feature vectors to match sketches with photos.

The LFDA system's accuracy is further improved using subjects' demographic information such as race, gender, age, and height. This information is generally stored in DMV and face databases and often available from witness accounts.

Despite the significant improvement in matching accuracy achieved by our automatic sketch recognition system, the difficult nature of matching sketches and photos requires human verification. The routine procedure in forensic face recognition, similar to that for latent fingerprint matching, is for a forensic examiner to vet the top N retrieved matches ($N < 100$) to determine whether a true match exists. Thus, for a given sketch, our system narrows down possible matches from a face database containing thousands of images to around 100 likely subjects for further human investigation.

Figure 3 shows three successful matches our LFDA system made between a forensic sketch and a mug

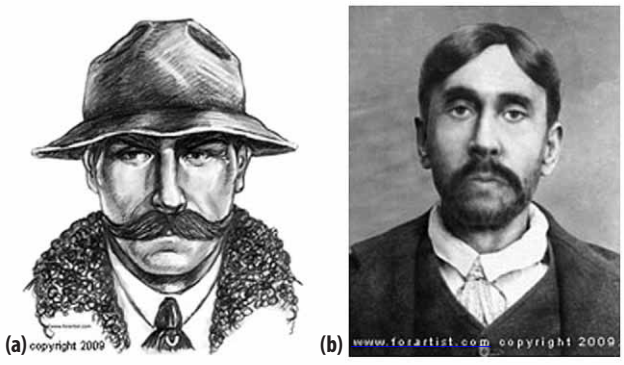


Figure 1. Applying sketch recognition technology to a 120-year-old series of crimes. (a) Example of a forensic sketch of serial killer Jack the Ripper based on a witness account. (b) Age-regressed mug shot of a perpetrator who committed similar murders in New York in the early 1900s. Source: www.forartist.com.

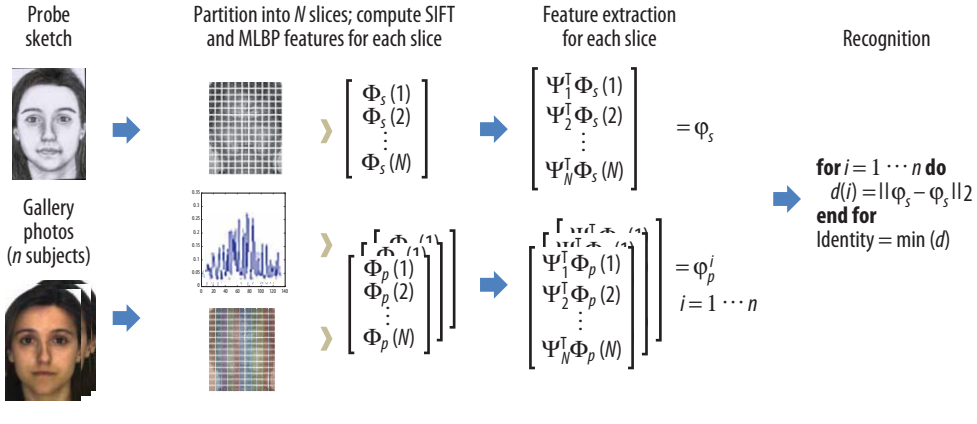


Figure 2. Automatic forensic sketch recognition system. The system, which uses local-feature-based discriminant analysis (LFDA) to match sketches and gallery photos, demonstrated a four times improvement in accuracy over a commercial off-the-shelf (COTS) face recognition system.

Method	Rank	Method	Rank	Method	Rank
LFDA	1	LFDA	1	LFDA	1
COTS	320	COTS	299	COTS	2,131

Figure 3. Examples of three successful matches the LFDA system made between a forensic sketch (top row) and a mug shot (bottom row) from a database of 10,000 face images. In each case, the COTS face recognition system failed to retrieve the correct photo within the top 100 matches.

shot from a database of 10,000 face images. In each case, the COTS face recognition system failed to retrieve the correct photo within the top 100 matches.

Automatic sketch recognition can be used in other ways to assist criminal investigations. For example, as Figure 4 shows, forensic artists can draw high-quality sketches of suspects captured in low-resolution surveillance video, and investigators can then use the system to match these sketches to mug shots in their database.

Investigators also could use the system to match forensic sketches to mug shots subjected to manipulation such as age progression and regression. For example, Figure 1b shows an age-regressed mug shot of a perpetrator who committed murders in New York in the early 1900s and had facial features resembling the forensic sketch of Jack the Ripper.

Law enforcement's need for a system that matches sketches to photos has prompted continued research to increase automatic face recognition accuracy. With a prototype automatic sketch recognition system soon to be deployed in the field, this technology will improve a process that has helped identify suspects for well over a century.

Anil K. Jain is a University Distinguished Professor in the Departments of Computer Science and Engineering and Electrical and Computer Engineering at Michigan State University. Contact him at jain@cse.msu.edu.

Brendan Klare is a PhD student in the Department of Computer Science and Engineering at Michigan State University. Contact him at klarebre@msu.edu.

Editor: Karl Ricanek Jr., director of the Face Aging Group at the University of North Carolina Wilmington; ricanekk@uncw.edu

Figure 4. Images from a recent Los Angeles Police Department investigation in which forensic artists generated high-quality sketches from low-quality video. The first two suspects were then identified from the sketches. Source: www.lacrimestoppers.org/wanteds.aspx.

SOCIAL COMPUTING

Folksonomy Formation

Shilad Sen, *Macalester College*
John Riedl, *University of Minnesota*

The combination of eager taggers and intelligent algorithms that nurture effective tagging behavior promises a bright future for tagging systems.

Tags—words or short phrases attached to items on the Internet—organize the social Web. We rely on tags to organize, find, and interpret videos on YouTube, images on Flickr, and webpages on Delicious. Unlike expert-maintained taxonomies such as the Dewey Decimal Classification, tags evolve organically based on the collective action of individual users.

The collection of tags in a tagging system, called a folksonomy, can vary widely from site to site. For example, users can tag books on both Amazon and LibraryThing.

A LibraryThing user searching for "conservatism" would find that 18 users have applied this tag to Jonah Goldberg's *Liberal Fascism: The Secret History of the American Left, from Mussolini to the Politics of Meaning* (Doubleday, 2007). The user might form an impression of the book based on its other popular tags: "politics" (95 users), "history" (76), and "fascism" (31). However, a search on Amazon would reveal very different tags, as Figure 1 shows. Amazon taggers have most often described the book as "wingnut welfare" (399) and "propaganda" (307).

Across Amazon's folksonomy, many users seem to choose tags to entertain the community. For example, 44 users have tagged pop sensation Lindsay Lohan's album *Speak* with "pig vomit," while five users have tagged L. Ron Hubbard's book *Dianetics: The Modern Science of Mental Health* with "keeping America stupid." In contrast, LibraryThing taggers focus on conveying the objective themes and metadata of books—even books that elicit strong reactions from readers.

The differences between LibraryThing and Amazon's folksonomies suggest that tagging is nondeterministic: similar tagging systems can spawn distinct folksonomies. What causes folksonomies to be so different?

OPPORTUNITIES AND CHALLENGES

The open nature of folksonomies provides both opportunities and challenges to designers. On the positive side, folksonomies scale. More than a million users of LibraryThing have used tags to catalog 61 million books in six years. In comparison, experts at the Library of Congress have catalogued only 33 million books in 211 years.

Figure 1. Tags for the book *Liberal Fascism* by Jonah Goldberg as created by LibraryThing taggers (left) and Amazon taggers (right). Amazon users seem to choose tags intended to entertain the community, while LibraryThing users choose tags describing factual thematic elements of the book.

0018-9162/11/$26.00 © 2011 IEEE Published by the IEEE Computer Society

	Organization	Communication
Self	Browse, search (for example, labeling items by genre so I can browse them later)	Memory (for example, labeling DVDs by the number of stars I give the movie)
Social	Search, self-promotion (for example, labeling books to help others find them)	Description, self-expression (for example, labeling items to communicate, such as "wingnut welfare")

Figure 2. Simplified version of Morgan Ames and Mor Naaman's model of users' motivations for tagging. Each cell describes the motivation and provides an example.

Because algorithms that can reliably annotate multimedia such as images and videos with keywords have not yet been developed, tags are crucial in making these items searchable. In fact, Flickr recognizes its 20 million unique tags as "the bread and butter of what makes our search work so beautifully" (http://blog.flickr.net/en/2008/01/16/many-hands-make-light-work).

Relying on a folksonomy rather than a controlled vocabulary, however, does expose a social website to risks. Since the folksonomy is under the control of the crowd, it could evolve in ways that the site's owner did not anticipate—as with *Liberal Fascism* on Amazon, for instance.

Even small initial differences in folksonomies can be important. Clay Shirky argues that folksonomies evolve organically to match the concepts important to a community precisely because the community itself creates the tags (www.shirky.com/writings/ontology_overrated.html). As Shirky points out, people using the tag "queer politics" and people talking about "the homosexual agenda" represent very different populations.

A term that becomes popular early in the life of a site can shape the user experience forever. Given the volatility of folksonomy evolution, how can a designer encourage beneficial tags?

WHY USERS TAG

Why do millions of website users apply billions of tags to items ranging from pictures to movies to blog posts, all for free?

User motivations for tagging

In "Why We Tag: Motivations for Annotation in Mobile and Online Media" (*Proc. 2007 Conf. Human Factors in Computing Systems*, ACM Press, 2007, pp. 471-480), Morgan Ames and Mor Naaman produced an elegant two-by-two model. As the simplified version in Figure 2 shows, their model divides tagging according to who the tag is intended for (social or self), and what the tag is intended to accomplish (organization or communication). These categories are more of a continuum than truly independent categories. For example, tagging some movies "watch again!" has attributes of both organization and communication.

Several years ago, we were members of a team in the University of Minnesota's GroupLens research group (www.grouplens.org) who surveyed users to find out why they tagged (S. Sen et al., "tagging, communities, vocabulary, evolution," *Proc. 2006 20th Anniversary Conf. Computer Supported Cooperative Work*, ACM Press, 2006, pp. 181-190). In addition to the reasons Ames and Naaman gave, respondents to this survey emphasized "fun" and "giving back to the community." This last reason is important, because it could nurture folksonomies that most benefit their communities.

Experts suggest that tagging systems work because of a sort of reverse tragedy of the commons. Individual taggers work for selfish reasons, labeling items for their own purposes—the "self" row in Figure 2. Meanwhile, the tags they add accidentally create social good for others, serving social purposes as in the bottom row in Figure 2.

In practice, though, the tags users add for selfish reasons are sometimes useless to others. For example, one common early tag in the MovieLens movie recommender website (www.movielens.org) that GroupLens maintains was "Erlend's DVDs." Presumably this tag was useful to Erlend—but to no one else. If users tag for explicitly social reasons, there is hope they can avoid these personal tags and create more communal value.

Personal preference versus community influence

Given this understanding of why people tag, we can begin to develop an abstract model of the forces that play a role in their decisions to use a particular tag, as Figure 3 shows.

Users enter a tagging system with an internal preference for a particular vocabulary. This preference comes broadly from their prior experience with language, and more narrowly from their familiarity with other tagging systems.

As users randomly apply tags in the new system, they develop distinct, self-reinforcing tagging habits. Over time, however, they synchronize their habits to the style of the folksonomy and in turn contribute more tags to the emerging style. Although GroupLens survey respondents denied being influenced by social pressure, field studies confirmed that they were.

One interesting question is whether tagging systems "intention-

ally" create different vocabularies to separate themselves from other similar sites. Robert Axelrod speculates in *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration* (Princeton Univ. Press, 1997) that the rapid evolution of human languages suggests that there must be an evolutionary benefit to subcultures using distinct languages.

Studying social evolution in human cultures is generally difficult, but perhaps folksonomies offer a simple laboratory for exploring language evolution.

STATISTICAL DISTRIBUTION OF TAGS

Not all tags are equal. For example, more than 27,000 Amazon users have applied the tag "paranormal romance," but only one has applied "varied ingredients." The wide differences in tag popularity suggest structure behind tag choice.

Scott Golder and Bernardo Huberman studied the distribution of tags in del.icio.us to shed light on this structure ("The Structure of Collaborative Tagging Systems," *J. Information Science*, Apr. 2006, pp. 198-208). They found that after a URL has received about 100 tag applications (many different users can apply a single tag to the same item), the distribution of the number of applications of the tags for that URL changes relatively little.

This insight led the researchers to model a URL's tags using a "rich get richer" statistical model called Polya's urn. The urn initially contains one red ball and one white ball (each ball represents a tag application). A ball is randomly withdrawn from the urn, and is replaced along with a third ball of the same color representing another application of that tag. Over time the proportion of balls of each color converges to a nearly fixed value—though the specific final proportions will vary widely if the entire process is repeated.

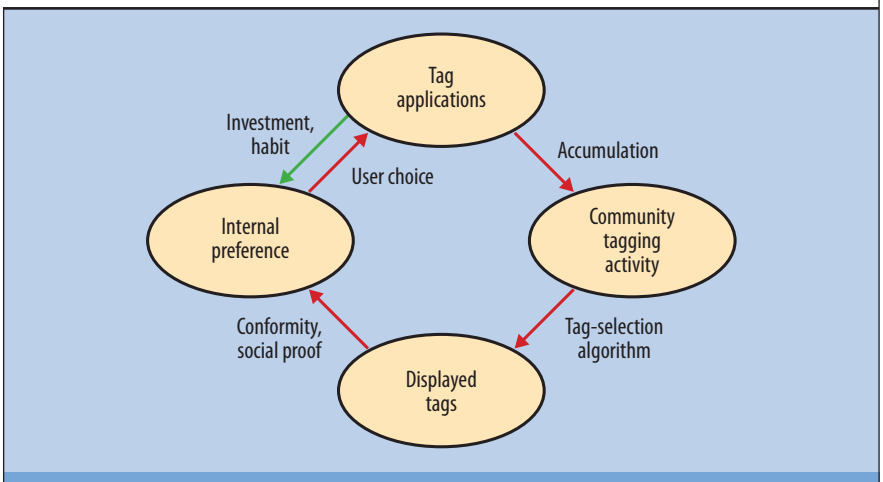


Figure 3. Relationship between internal preference and community influence. Arrows indicate social and technical processes.

Golder and Huberman offered two explanations for the stable proportions of tags for a particular URL: imitation and shared knowledge.

First, users could be imitating others' tags. However, the researchers pointed to evidence suggesting that imitation is not the only force at work. Even rare tags not commonly displayed for a URL maintain consistent proportions over time. Users cannot see these tags, so some process besides social influence must also affect folksonomy development.

A second explanation is that the choice of tags is related to shared knowledge. When taggers come to the site, they bring their own skills, experience, and personal vocabulary. Their tags could reflect this prior knowledge, which many taggers might share.

THE POWER OF SOCIAL INFLUENCE

The GroupLens team wondered how strongly these two forces affect taggers, and in 2005 researchers added tags to MovieLens to conduct an experiment.

A field study tested the effects on folksonomy evolution of four tagging interfaces, which used different algorithms to choose tags to display for a movie:

- *unshared*—users see tags they create, but nobody else's.
- *shared*—users see tags created in their group; tags are ordered randomly.
- *shared-pop*—users see tags created in their group; tags are ordered by popularity.
- *shared-rec*—users see tags created in their group, and the system infers additional tags from movies rated similarly; tags are ordered by popularity.

Over the course of 100 days, 635 users created 11,443 tag applications. To study patterns in the types of tags users created, the GroupLens team categorized each tag as factual (for example, "martial arts" or "time travel"), subjective (for example, "funny" or "overrated"), or personal (for example, "in Netflix queue" or "my dvds").

Figure 4 shows the distribution of tags over time in each of the four experimental groups. The red line represents factual tags; the blue line, subjective tags; and the green line, personal tags. Overall, different experimental groups created different types of tags. The unshared group is relatively balanced among the three classes, subjective tags dominate the shared group, and factual tags dominate the shared-pop and shared-rec groups.

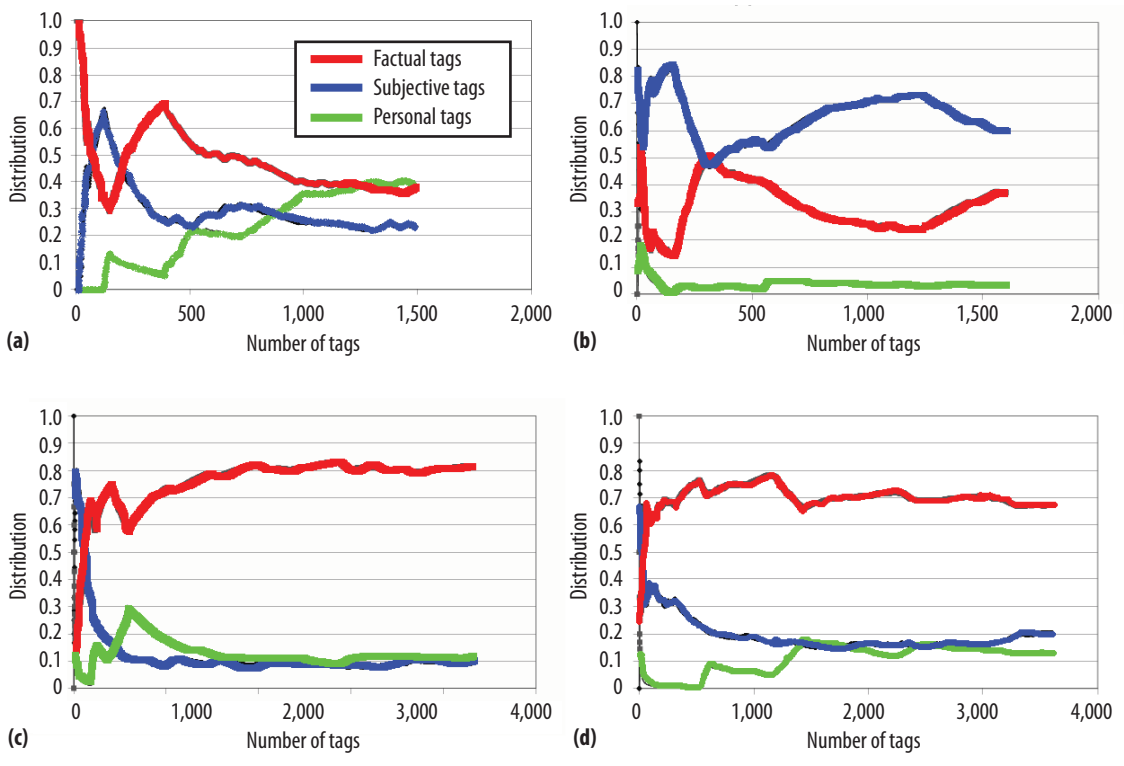


Figure 4. Distribution of MovieLens tags over time in four experimental groups: (a) unshared, (b) shared, (c) shared-pop, and (d) shared-rec. Source: S. Sen et al., "tagging, communities, vocabulary, evolution," *Proc. 2006 20th Anniversary Conf. Computer Supported Cooperative Work*, ACM Press, 2006, pp. 181-190.

The tag-selection algorithms could have influenced these outcomes. In the unshared group, users did not see others' tags, so there was no incentive for them to create unselfish tags. The shared-rec and shared-pop groups both favored popular tags, and they tended to use factual tags more often during the experiment than subjective or personal tags.

The study findings support the folksonomy differences observed on Amazon and LibraryThing—even relatively similar systems can produce very different tags.

To better understand why users create the tags they do, the GroupLens researchers used the personal/factual/subjective category system to measure how similar users' tags were to tags they had applied in the past as well as those they had seen from the community. Beginning taggers appeared to be more heavily influenced by community tagging norms, while experienced taggers' behavior reflected their investment in a personal tag vocabulary.

These findings suggest a social influence feedback loop: users in a community perceive a particular tagging norm, they adjust their behavior to be more consistent with that norm, and the new tags they create strengthen that norm.

Other researchers studying online social influence have observed similar effects. For example, an influential field experiment by Matthew Salganik, Peter Dodds, and Duncan Watts analyzed users' ratings of music ("Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market," *Science*, 10 Feb. 2006, pp. 854-856). The researchers found that when the salience of community norms is increased (for example, by highlighting the average rating), both the inequality in music popularity and the music rankings' volatility increase.

SHAPING FOLKSONOMIES

It is becoming increasingly clear that social reinforcement can lead to strong preferences within a community, including in important areas like the choice of a tagging vocabulary. Once set, these preferences are difficult to change, but early in their life cycle they are vulnerable to even small influences that can dramatically change their course. For example, the MovieLens study revealed that small changes in a user interface can apparently lead to big changes in how users perceive a tagging vocabulary, and thus how the language evolves.

Designers can use this understanding to reinforce a socially desirable folksonomy, steering users toward tags that are more useful to other users.

One challenge in creating such systems is understanding what makes one tag more useful than

another. Toward that end, the GroupLens team is developing algorithms that, based on ratings of previous tags, predict how users will rate a new tag.

The strongest predictor of tag quality so far is the tagger's reputation: taggers who have applied high-quality tags in the past will tend to apply high-quality tags in the future.

Richer models are emerging that do not focus on the value of individual tags but on the folksonomy's overall value in meeting users' needs. For example, Edi Chi and Tom Mytkowicz studied the information-theoretic efficiency of tagging systems and found that current recommender algorithms nudge users toward tags that are already popular ("Understanding the Efficiency of Social Tagging Systems Using Information Theory," *Proc. 19th ACM Conf. Hypertext and Hypermedia*, ACM Press, 2008, pp. 81-88). Future algorithms will balance popularity against search value to encourage tags that help people find what they want as efficiently as possible.

Other interactions could make folksonomies more useful to individual users as well as the community. For example, algorithms could alert users when a particular tag seems to have two different meanings based on its co-occurrence with other tags. The tagging system could encourage the user to disambiguate these meanings, perhaps by suggesting a less ambiguous tag.

The combination of eager taggers and intelligent algorithms that nurture effective tagging behavior promises a bright future for tagging systems.

Shilad Sen *is an assistant professor of computer science at Macalester College, St. Paul, Minnesota, and a former member of the GroupLens research group. Contact him at ssen@macalester.edu.*

John Riedl*, Social Computing column editor, is a professor in the Department of Computer Science and Engineering at the University of Minnesota, where he coleads the GroupLens team. Contact him at riedl@cs.umn.edu.*

Continued from page 104

```
<a href=http://en.wikipedia.org/wiki/Anoa>anoa</a> (buffalo)
```
(a)

```
<def name="wp">
<a href=http://en.wikipedia.org/wiki/&1;&3;>&2;&3;</a> 
</def>
```
(b)

```
<wp A a "noa"/>(buffalo)
```
(c)

Figure 1. HTML macrocoding: (a) bare code, (b) macro definition, (c) macrocoded example.

those infrequent occasions when it's actually needed. I wish browser programs would automatically try putting a "www" in front of a domain address that isn't given with one and that fails first up.

Anyway, it's often convenient to have short URLs. I use tinyurl.com to get short ones for these essays—that way I can comfortably put in more of them. There's now a variety of other websites that provide a similar service. However, some problems have arisen (tinyurl.com/EcTnyURL), and the extra step in getting to the site behind the short URL is being used to gather data about Web users (tinyurl.com/ScAmBtLy).

But even these short URLs aren't very helpful to anyone putting together an HTML file composed almost entirely of URLs. This is the kind of file I like to use to make my own indexes to the Web. For example, I've put together a word index file to help people understand the answers to a nine-letter word game popular in Australian newspapers (eprint.utas.edu.au/10468). The problem is that different newspapers publish puzzles that seem to be based on different dictionaries, and it can be frustrating trying to find what uncommon words might mean. My help file is comprised almost entirely of lines like that in Figure 1a.

Clearly the repetition is highly redundant, which makes the file harder to update. Using a URL shortening method like tinyurl.com provides doesn't help much, so I tried to find a macro definition facility (tinyurl.com/WpMcrDf) in HTML. There doesn't seem to be one, which is ironic as HTML comes from SGML, which comes from GML, which was implemented using the macro definition facility of Script/VS (tinyurl.com/WpScrVS).

HTML needs a macro definition facility like that shown in Figure 1.

Figure 1b shows the definition of a macro named wp, and the body of the definition gives the text to be used, with provision for the text of three variable fields to be filled in given by &1; &2; and &3;. Figure 1c shows the coding used to produce the effect of Figure 1a.

While a facility like this could be elaborated, it would be tremendously useful even if it could only provide what this example shows. The benefit wouldn't be just for people like me who put HTML files together. For instance, it could be used to provide skeleton HTML files that teachers could build on to give pupils simple access to Web data so that they could better understand lesson content and to put together homework.

Macro definition has much wider application than HTML programming, as it not only reduces redundancy but also allows better approaches to helping programmers and users generally (The Profession, July 2002, p.110, Blunder 4).

USERS

Users are stashing personal data—pictures and videos—in ever-increasing quantities. Enterprises like Google and security companies with surveillance cameras are stashing community data in an even larger fashion. Then there are the electronic books and cloud computing.

All this digital stashing poses various questions, but the most significant one concerns how long all these data files are good for. Traditional books and pictures last for centuries. Files on magnetic tapes, discs, and the various data sticks won't last anything like that length of time (tinyurl.com/ScAmSgFe). Who will be their curator?

The Internet

Data doesn't just get filed. It is also transmitted across the Internet and displayed on users' terminals. But there's an awful lot of transmitted data that's quite redundant. Internet traffic is obese.

The obesity of spam and phishing is well known, although the Internet loading is minuscule compared to other kinds of traffic. However, the interesting aspects of these activities is that they're only rewarding to the senders: sending out millions of spam or phishing e-mails to catch a small proportion of gullible receivers is very inexpensive. Control of such activities is complex and questionable (The Profession, Apr. 2005, pp. 88, 86-87), but why don't governments require the telecommunications companies to charge the senders for all the transmission costs involved rather than the poor users? After all, the sender puts the postage stamps on old-fashioned letters.

There's another form of obesity in e-mail: the advertisements. They take up space on the screen, and the data involved often far exceeds the text of the messages proper, especially now that movie ads are so common. If e-mail providers were to offer ad-free services for a reasonable monthly fee

or traffic charge, I wonder how many would take it up? I would, for one.

The World Wide Web

Nowadays, e-mail is a relatively minor use of the Internet. The browsers that support e-mail services like Yahoo and Gmail are used much more often for getting data from sites on the World Wide Web.

Commercial websites also get the browsers to run ads, and these ads seem to be getting more and more obese in taking up screen space and in slowing down loads. Some websites, noticeably newspapers, have the courtesy to give the user the option of running video content or not, but I would like a browser that downloaded only the first screen of a video and gave me the choice of running it or not. Or better, the choice of whether to have any downloaded at all. After all, most browsers give you some control over cookies, so why not over video data?

One kind of website that properly attracts admiration is the kind that provides high-quality data. The best are noncommercial ones such as gutenberg.org, which provides thousands of books that can be downloaded free of charge. One that I use a lot in my own word-game index is Wikipedia, sustained as an international cooperative project.

Such websites are very large, but they're not obese as they have relatively little redundancy. They concentrate on using their data to convey as much information as feasible and providing as good information as possible. But there's a social problem lurking in the shadow of such sites.

Many teachers are now adopting the Web and its search engines as the focus for their students. All the knowledge we need, these teachers claim, is available on the Web, so there's no reason for them to impart knowledge to their students other than teaching them how to get from the Web what they might need from time to time. This attitude fallaciously equates data to knowledge. It also equates knowledge to understanding.

Data is the conventional representation of facts and ideas; information is the meaning that people give to data (The Profession, May 2001, pp. 96, 94-95). Knowledge is the experience that enables people to give meaning to data. Understanding is the skill that enables people to use knowledge personally and socially.

Data obesity is all too likely to flood the classroom and the workplace, blocking the acquisition of the knowledge and understanding needed to realize the available information.

There are several professional responsibilities in respect to preventing data obesity.

First, programs should be kept as simple and useable as possible by reducing redundancy.

Second, both the sender and the receiver should be given reasonable control of the transmission of data.

Third, the procedures governing the use of data, especially in education, should foster the development of understanding and the acquisition of knowledge.

After all, digital technology is there for people to exploit, and the better they understand it, the more benefit they will be able get from it. C

Neville Holmes is an honorary research associate at the University of Tasmania's School of Computing and Information Systems. Contact him at neville.holmes@utas.edu.au.

Aspects of Data Obesity

Neville Holmes,
University of Tasmania

Data obesity is a consequence of data gluttony and vice versa.

For this column a year ago, I wrote an impulsive rant titled "The Rise and Rise of Digital Gluttony" (May 2010, pp. 96, 94-95). Since then, that increase has, if anything, accelerated.

Digital gluttony is seen as justifying data obesity, and the obesity grows with the gluttony. Digital technology enables the storage of ever more data, and has done so for decades. In the 1960s, this was jokingly but truly said to be a corollary to Parkinson's law: data expands to fill the space available.

As a parallel to the double dactyl (tinyurl.com/ykyk4qh) included in the 2010 column, consider the following:

> Higgledy piggledy,
> Data obesity,
> Digital gluttony,
> Out of control.
>
> Data technology's
> Parkinsonality
> Looms astronomically
> Like a black hole.

The "developed" world seems to welcome the obesity. The author of a recent *New York Times* article, "Remapping Computer Circuitry to Avert Impending Bottlenecks" (tinyurl.com/4l9qc57), seems delighted that "the amount of data consumed by computers is growing even more quickly than the increase in computer performance" and that computing professionals look like they're beating the problem. Incidentally, this article springs from a January 2011 *Computer* article by Parthasarathy Ranganathan ("From Microprocessors to Nanostores: Rethinking Data-Centric Systems," pp. 39-48), and also quotes from an article by Peter Kogge in *IEEE Spectrum*'s February 2011 issue ("Next-Generation Supercomputers," pp. 44-50).

It's disturbing that discussion of digital obesity's adverse technical effects seems muted, though the personal effects of digital gluttony are starting to get the attention they deserve. For instance, see the excellent review by Jim Harris of Nicholas Carr's book *The Shallows: How the Internet Is Changing the Way We Think, Read and Remember* (Atlantic, 2010) in the online London Review of Books (tinyurl.com/NkCrRvJH).

The following revisit to the topic focuses on remediable aspects relevant to the computing profession and, as the phrase "data obesity" suggests, to a few of the various aspects of the data involved rather than of the machinery processing the data. Incidentally, "Family Court Philosopher" Glenn Campbell, an early user of the term "data obesity," seems to use it to refer more to what I would call digital gluttony (tinyurl.com/GnCbDtOb).

PROGRAMMERS

Popular programs tend to grow and grow. Growth can be curbed by simplifying. Simplicity can come through various measures such as rigorous design, separating distinct functions, and removing redundancy.

Much programming nowadays is for the World Wide Web in one form or another, and I quite enjoy putting vanilla HTML programs together using a simple text editor. Some of these are for bullet point presentations (see eprints.utas.edu.au/1301 for an example), and some are for keeping track of online data.

However, when I look at their source code, most HTML files run by a browser appear to be highly redundant. For instance, one niggling example is in the URLs themselves.

Tim Berners-Lee reportedly said that "the two forward slashes that form part of every web address are entirely useless" (tinyurl.com/tblmstks), and he's the one who put them there in the first place. By the way, "forward slash" and "reverse slash" are annoyingly verbose and frequently heard. I prefer "slash" and slosh."

Another aspect of Web addresses is that the usual "www" prefix is not only awkward to say but seems often to be redundant, and I normally leave it out. But this is a pain on

Continued on page 102

 Published by the IEEE Computer Society