Challenges in Location-Aware Computing

By Cynthia A. Patterson, Richard R. Muntz, Cherri M. Pancake[1]

Much of the information that underpins the challenges facing society today—such as terrorist activities, global environmental change, and natural disasters—is geospatial in nature. An everyday example of key geospatial data is the location of a cell phone user who has dialed 911 in an emergency. In terrorist situations, the origins and destinations of phone calls, travel patterns of individuals, dispersal patterns of airborne chemicals, assessment of places at risk, and the allocation of resources are all geospatial data. As the volume of geospatial data increases by several orders of magnitude over the coming years, so will the potential for corresponding advances in knowledge of our world and in our ability to react to changes taking place.

This article, which discusses key research challenges in location-aware computing, is excerpted from *IT Roadmap to a Geospatial Future* (SEE NRC REPORTS SIDEBAR). The report outlines an interdisciplinary research roadmap focusing on challenges in location-aware computing, databases and data mining, and human-computer interaction technologies. The convergence (SEE FIGURE 1.1) of advances in these three areas, combined with a sharp increase in the quality and quantity of geospatial information, promises to transform our world. Geospatial data have become of critical importance in areas ranging from crisis management and public health to national security and international commerce. The Just-in-Time Mapping Scenario (SEE SIDEBAR) illustrates the committee's vision of what is possible with integrative, multi-disciplinary research efforts. Diverse technological advances will be needed to achieve this imaginative future.

Advances in databases could improve our ability to integrate geospatial data (which is often of myriad formats, conventions, and semantics) and to represent objects, such as a wildfire, that move and evolve over time, sometimes appearing or disappearing at irregular intervals. Analysis and evaluation of data would benefit from new data mining methods designed to operate on complex, highly dimensioned data representing objects that may be undergoing continuous change. New technologies that support human interaction with geospatial data would enable individuals from geospatial information specialists to policy makers to emergency response personnel to access, visually explore, and construct knowledge from geospatial data and apply the knowledge to critical problems facing both science and society. A brief synopsis from *IT Roadmap to a Geospatial Future* of key research directions in databases, data mining, and human-computer interaction is outlined at the end of the article.

This article focuses on research opportunities in location-aware computing. Advances in this area could have important implications not just for how geospatial data are acquired, but also for how and with what quality they can be delivered, and how mobile and geographically distributed systems are designed. Sensors that record their location and some information about the surrounding environment (e.g., temperature and humidity) are being deployed to monitor the state of roads, buildings, agricultural crops, and many other objects. For example, Smart Dust sensors (devices that combine microelectromechanical sensors with wireless communication, processing, and batteries into a package about a cubic millimeter in size)—currently in their infancy—can be deployed along remote mountain roads to determine the velocity and direction of passing vehicles or can be attached to animals or even insects to record where they travel (Pister, 2002). The data transmitted wirelessly in real time from these sensors increase not only the volume but also the complexity of available geospatial data.

---

[1] Cynthia A. Patterson is a study director for the Computer Science and Telecommunications Board of the National Research Council in Washington, D.C. Richard R. Muntz is a professor of computer science at the University of California at Los Angeles. Cherri M. Pancake is professor and Intel Faculty Fellow in the School of Electrical Engineering and Computer Science at Oregon State University.

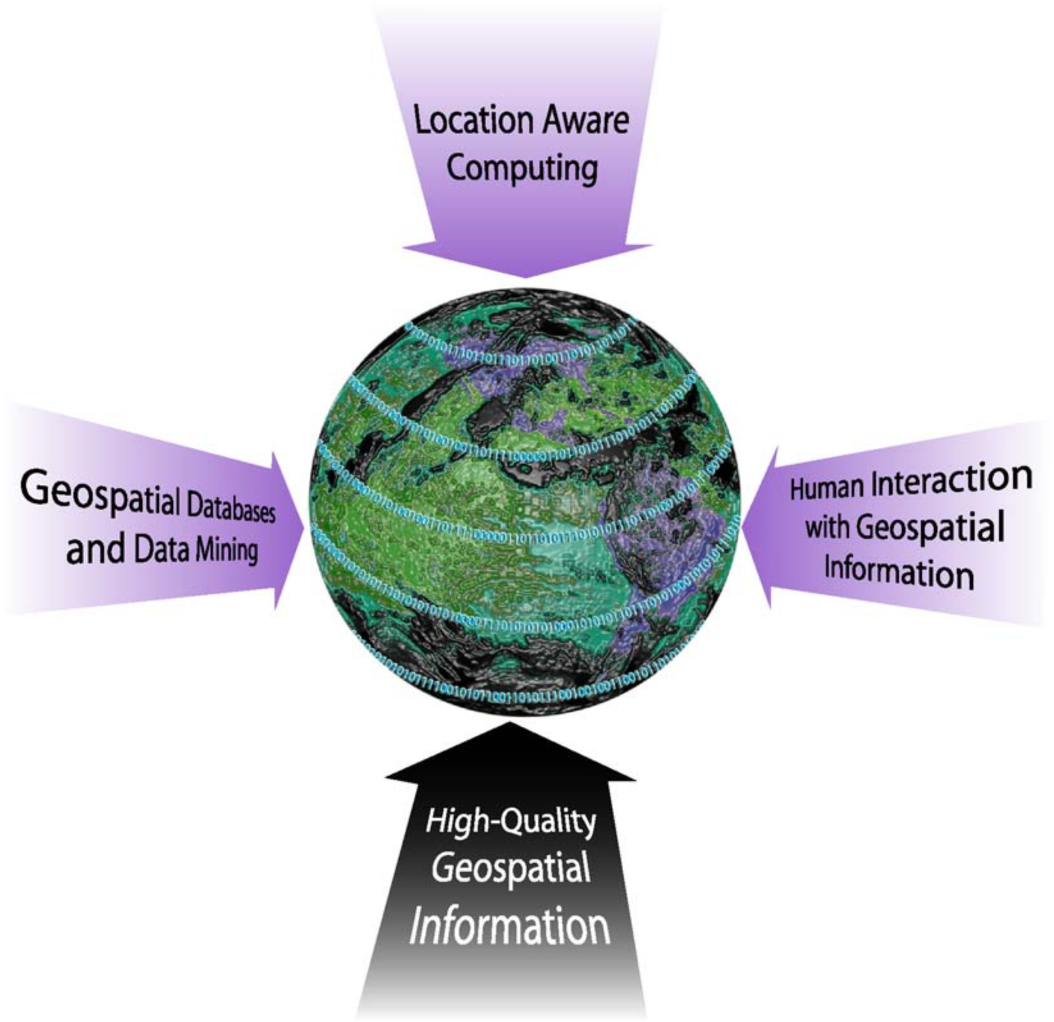

FIGURE 1.1 The convergence of four independent technological advances has the potential to transform the world.

ORIGIN AND BACKGROUND OF THE STUDY

In response to a request from the National Science Foundation and the National Aeronautics and Space Administration, the Computer Science and Telecommunications Board (CSTB) of the National Research Council convened an expert committee (SEE COMMITTEE SIDEBAR) to explore opportunities and directions for increased interaction between the geospatial and computer science research communities. The Environmental Protection Agency (Office of Research and Development) became an additional sponsor after the project began.

The committee met in July 2001 to plan a 2-day workshop that was held in October 2001 and met again in January 2002 to plan the structure and format of this summary report.

The objective of the workshop was to illuminate directions for future research that would enhance the performance, accessibility, and usability of geospatial information. An overarching goal was to foster greater computer science research interest in the challenges presented by proliferating geospatial information.

The workshop participants (listed in Appendix B of the report), like the committee members, included experts from multiple disciplines and experts knowledgeable about applications in specific domains. The selection of workshop participants was weighted slightly more toward computer science in an effort to engage that community more broadly in the problems raised by geospatial data. Workshop participants were divided into breakout groups to outline the current technology trends with respect to geospatial applications, identify and explore the current shortfalls, and propose promising research directions within location-aware computing, databases and data mining, and human interaction technologies.

The workshop demonstrated the value of assembling a diverse group of experts embodying many complementary perspectives. It also demonstrated how differently people in diverse disciplines—or people with different subspecialties within a given discipline—perceive, analyze, and discuss the needs of the research and development communities. That recognition implies that the workshop should be seen as part of a process of interdisciplinary convening and exchange that should continue.

The content of *IT Roadmap to a Geospatial Future* reflects the issues identified at the workshop—in plenary presentations, position papers submitted by several of the participants, and group discussions—and during subsequent deliberations by the committee. The report's contribution lies in its integration of a very diverse set of perspectives to illuminate promising directions for research, with an emphasis on directions that cross-disciplinary boundaries.

OVERVIEW OF LOCATION-AWARE COMPUTING

The evolution of mobile computing, location sensing, and wireless networking has created a new class of computing: *location-aware computing*. Mobile computing is commonly associated with small form-factor devices such as PDAs and untethered (wireless) connectivity. Such devices provide access to information processing and communication capabilities but do not necessarily have any awareness of the context in which they operate. "Context-aware" computing describes the special capability of an information infrastructure to recognize and react to that real-world context. Context, in this sense, includes any number of factors, including: user identity; current physical location; weather conditions; time of day, date, or season; and whether the user is asleep or awake, driving or walking. The most critical aspects of context are location and identity. Location-aware computing systems respond to a user's location, either spontaneously (e.g., warning of a nearby hazard) or when activated by a user request (e.g., is it going to rain in the next hour?).

Hardware for Mobile Computing

Over the past few years, lightweight and compact laptops and handheld computers have come into extensive use; wearable computers are beginning to make an impact as well. Where progress has been slowest is in the integration of mobile hardware into systems that seamlessly bridge a user's desktop, his or her activities while mobile, and the Internet. Four fundamental issues complicate the design and implementation of such systems (Satyanarayanan, 1996). Mobile elements are resource-poor relative to static elements, due to limitations on power, size, and weight. Mobile communications is also inherently vulnerable to security violations, since data are transmitted through open airspaces. Further, wireless connectivity is highly variable in performance and reliability. Finally, mobile elements must rely on limited energy sources.

It is important to note that these issues are not artifacts of current technology but are intrinsic to mobility. Collectively, they complicate the design of mobile computing systems. Consequently, although significant research progress has been made, the design and implementation of mobile computing systems (and, therefore, location-aware computing systems) remain problematic.

Location-Sensing

The Global Positioning System (GPS) is the most widely known location-sensing system today. Using time-of-flight information derived from radio signals broadcast by a constellation of satellites in earth orbit, GPS makes it possible for a relatively cheap receiver (on the order of $100 today) to deduce its latitude, longitude, and altitude to an accuracy of a few meters (Hightower and Borriello, 2001). The expensive satellite infrastructure is maintained by the U.S. Department of Defense, but many civilian users benefit from the investment. Indeed, there has been a veritable explosion of GPS-based services for the consumer market over the past few years.

Although certainly important, GPS is not a universally applicable location-sensing mechanism, for several reasons. It does not work indoors, particularly in steel-framed buildings, and its resolution of a few meters is not adequate for many applications. GPS uses an absolute coordinate system, whereas some applications (e.g., guidance systems for robotic equipment) require coordinates relative to specific objects. In addition, the specialized components needed for GPS impose weight, cost, and energy consumption requirements that are problematic for mobile hardware.

Consequently, a number of other mechanisms for location sensing have been developed (e.g., active badges, e911, and Cricket), which vary significantly in their capabilities and infrastructure requirements. System costs vary as well, reflecting different trade-offs among device portability, device expense, and infrastructure needs. For applications involving mobile objects, orientation sensing (determining the direction an object faces) is also important, and this continues to be an active area of research.

Wireless Communications

The past decade has seen explosive growth in the deployment of wireless communication technologies. Although voice communication (cell phones) has been the primary driver, there also has been substantial growth in data communication technologies. The IEEE 802.11 family of wireless LAN technologies is now widely embraced, with many vendors offering hardware that supports it. Bluetooth is another standard, backed by a growing number of hardware and software vendors; although it offers no bandwidth advantage relative to 802.11, it has been designed to be cheap to produce and frugal in its power demands.

Infrared wireless communication is the lowest-cost wireless technology available today, primarily because it is the mass-market technology used in TV remote controls. Most laptops, many handheld computers, and some peripheral devices such as printers are manufactured today with built-in support for IrDA. Infrared wireless communication must be by line of sight, with range limited to a few feet. It is also affected adversely by high levels of ambient light, such as prevail outdoors during daylight hours.

It is difficult to foresee what new wireless technologies will emerge in the future. Power consumption clearly will be an important factor for untethered devices, such as mobile computers, PDAs, and Smart Dust. In addition, it is clear that advances will be constrained by trade-offs among four factors: frequency, bandwidth, range, and density of wired infrastructure. Devices operating at a higher frequency could have greater bandwidth but would require major advances in high-frequency VLSI design. Advances also will be constrained by policy decisions on frequency usage (spectrum allocation) by the Federal Communications Commission. Range is fundamentally related to transmission power, but generating high power at high frequency always has been a difficult technical challenge.

The standard solution to limited range and frequency allocation, coupled with line-of-sight constraints, is to use a wired infrastructure with base stations that define cells of wireless coverage around them. This is the basis of the now-widespread cell phone technology and wireless LAN technologies such as 802.11. Wired infrastructures impose significant costs for conditioning the environment, with density (and hence cost) increasing with bandwidth. Cheap, high-bandwidth, low-power, and ubiquitous wireless coverage will not be attained easily, so location-aware computing systems will have to be designed to cope with hard realities. This is not a short-term annoyance but a core, long-term requirement of successful system architectures.

RESEARCH CHALLENGES IN LOCATION-SENSING INFRASTRUCTURE

Advances in the technologies for data acquisition and data access are enabling more and more applications of geospatial data and location-based services. The Global Positioning System and other localization technologies, wireless communication, and mobile computing are key components. Although progress has been made in these areas, a significant amount of additional research in the development of a location-sensing infrastructure is needed before location-aware computing can become commercially viable.

Technology-Independent Location Sensing

No single location-sensing technology is likely to become dominant; there are simply too many dimensions along which location-sensing mechanisms can vary (Hightower and Borriello, 2001) CYNTHIA: I THINK WE STILL NEED TO REFERENCE IT HERE, SINCE WE BORROWED FROM THEIR IDEAS . Examples include indoor vs. outdoor use, accuracy, precision, energy usage, and the extent to which there is potential loss of privacy for users of the technology. As a result, the choice of location-sensing technology is likely to depend on the usage context, and various technologies are likely to coexist well into the future.

These considerations argue for research into the creation of a technology-independent, high-level software application programming interface (API) for location sensing. The operating system interface is the most obvious level for this API, although the middleware level also might be feasible. By freeing application writers from the specifics of location-sensing technologies, an API will encourage the creation of long-lived applications. It also can encourage the creation of new location-aware applications by helping to amortize development efforts. Further, by lowering the barriers to adoption, it can stimulate new location-sensing technologies.

Although the design and validation of such an API remain an open research problem, certain attributes are already clear. It would be beneficial to have an API that is based on an open standard rather than one that is proprietary to one company or a consortium of companies. It must mask technology-dependent attributes of the underlying technology. It should allow specification of desired accuracy and discovery of actual accuracy. It should be capable of dynamically combining location information from multiple sources in a manner that is transparent to applications.

Calibrating Location Sensing Technologies

Location-sensing technologies are expensive to deploy today. The share of system costs incurred by the infrastructure vs. that borne by end users varies significantly in current technologies. With GPS, for example, the end-user cost is relatively small but the cost of the satellite infrastructure is enormous, whereas the split is more balanced in an active badge system. Moreover, although hardware costs are likely to decline as volume increases, many technologies incur hidden costs that are much harder to reduce. An approach based on signal strength maps, for instance, requires creating the maps in every location where the system is deployed. The maps must be re-created whenever the physical topology of the location is modified in any significant way.

The growth of location-aware computing will be hindered as long as the costs of deploying and managing location-sensing systems remain high. Fundamental research on techniques to rapidly calibrate an environment for specific location-sensing technologies would reduce these costs. Another example of a hidden (albeit necessary) cost is the need to monitor and audit the release of location information to guard against privacy lawsuits.

Two very different approaches are conceivable. One is to develop modeling and analysis techniques, predictive algorithms, tools for optimizing the deployment of infrastructure, and self-configuring technologies that could eliminate or minimize the need for human intervention and calibration. Some of the existing research on modeling the propagation of wireless signals may be relevant, but it will need to be substantially extended and refined. A different approach is to retain physical calibration but to develop automation techniques to speed the process. An intriguing possibility is the use of mobile robots for calibration. For example, rather than having a human engineer sample signal strengths, it might be possible to program a robot to construct a signal strength map. To further speed the process, multiple mobile robots might exploit parallelism; the results of robotic research in planning and team coordination are relevant here.

Opportunistic Data Acquisition

Entities that combine location-sensing technologies with other types of sensors are growing in number. Many cars today, for example, are equipped with both GPS and an antilock brake system (ABS). Adding wireless transmission would complete the elements necessary for the automated collection of road surface conditions. For example, during a snowstorm, road maintenance personnel might wish to monitor road conditions to determine how best to allocate their resources. Every time an ABS detected the onset of wheel lockup (e.g., due to icy conditions), its GPS coordinates could be transmitted to a regional data collection site. Many ABS activations over a short period of time might signal icy conditions on a segment of road. Road maintenance personnel, using data mining and visualization software, could identify the problem locations and direct salt trucks to the needed areas. Real-time deployment of resources to needed areas could prevent accidents, conserve labor, reduce the use of salt, and so on.

The key sensing capability (in this case, antilock brakes) is of value in and of itself, but adding locational, wireless data gathering capabilities amplifies the value of the primary capability. This is referred to as *opportunistic data acquisition*. If we are to take advantage of such opportunities in the future, an investment must be made in research that explores appropriate techniques for data acquisition and redistribution. Some of the challenges to be addressed are scalability, mobile sensor sources, appropriate information-sharing policies, and mechanisms for preserving privacy without sacrificing functionality. Research also will be needed on how location-sensing systems might be designed to reasonably exploit new data acquisition opportunities as they arise.

End-to-End Control of Location Information

Some location technologies, such as cell-based location sensing, expose the location of the user to the sensing infrastructure. Today, cell-phone providers know where you are to within the resolution of a cell. There is already some concern about this loss of privacy, and the concern will grow in intensity as the use of location-aware computing grows. There is an inherent tension between privacy and the transparent use of location information: indeed, the more seamless and easy to use that a location-aware application is, the fewer the cues to remind users that their locations are being monitored. Historical location information can be analyzed to obtain insights into a user's typical movements. On the other hand, the authentication of locations is difficult, because it requires that both the identity of the user and his or her current location be established.

There is a need for system design techniques capable of providing end-to-end control of location information that include research on system layering to control the exposure of location information, as well as efficient auditing mechanisms for recording such exposure. Fine-grained access control mechanisms—permitting the precise release of location information to just the right parties under the right circumstances—are required. The use of location information to enforce security policies (e.g., company policy may dictate that a user's laptop should not be able to access sensitive files from outside a specific building) also should be explored. Research is needed in protocols and mechanisms for authenticating and certifying the location of an individual at any given time. Finally, user interface techniques must be developed that remind users that their locations are being monitored and alert them when the trustworthiness of the entity performing that monitoring changes.

Test Bed for Experimental Research in Location-Aware Computing

A complex relationship exists between commercial support for location-aware computing and deployment of the needed research. In the long term, it is clear that location-aware computing should be a commercial activity. However, to get to that point, a considerable amount of research, deployment, and experience will be necessary. It is not sufficient for the research community to develop the concepts, algorithms, and architectures and then leave it to industry to pick up and carry on. Rather, the research community—both computer scientists and scientific users of information technology—has the opportunity to lead by example. There is value in the research community being engaged in the initial phases of actual use of location-aware computing so that it can explore new application paradigms enabled by this technology and conduct the high-risk, high-payoff experiments in the use of the technology.

A key obstacle to research progress in location-aware computing is the lack of adequate large-scale experimental infrastructure. Such infrastructure is essential for empirical validation of concepts, techniques, and architectures for location-aware computing. The government can act as a catalyst by funding the creation and maintenance of a national test bed for experimental research in location-aware computing.[2] Broad access to a national test bed could stimulate research in the development of a rich, open, location-aware computing infrastructure (e.g., standard protocols, APIs, platform-independent capability descriptions, scalability, reconciliation of conflicting information from different network nodes, adaptive resource management, static-mobile load balancing, and mediation of requests).

Where feasible and cost-effective, such a test bed should leverage the physical infrastructure of the Internet. One strategy would be to use the Internet for low-level data transport but to overlay experimental location sensing and routing functionality on top of it. Part of this effort can include the creation and dissemination of benchmarks and testing methodologies for location-aware systems and applications. These artifacts can become part of the discourse of the research community and help forge a common basis for evaluation of ideas in the field.

The test bed also can serve as the focal point for standardization efforts, through collaboration between researchers and industry groups. Because it would encourage location-aware applications, a large-scale test bed could speed up the commercialization of research results.

RESEARCH CHALLENGES IN ADAPTIVE RESOURCE MANAGEMENT

The ability to obtain information on demand, wherever a user happens to be, cannot be realized without new technologies and methods specifically accommodating user mobility. The relative resource poverty of mobile elements, as well as their lower levels of security and robustness, argues for reliance on static servers. At the same time, the need to cope with unreliable, low-performance networks and to be sensitive to power consumption argues for self-reliance.

Any viable approach to mobile computing must strike a balance between autonomy and interdependence. The balance cannot be static; as the circumstances of a mobile client change, the system must be able to react, dynamically reassigning the responsibilities of client and server. In other words, mobile clients must be adaptive. The need for adaptation complicates many fundamental aspects of location-aware computing systems.

Cyber Foraging and Infostations

Location sensing can provide the basis for novel techniques of resource management, potentially improving the user experience. For example, location awareness can be used to guide a mobile user from a bandwidth-impoverished to a bandwidth-rich environment Satyanarayanan (2001). This is an example of an emerging technique, *cyber foraging*, which temporarily extends the resources of a mobile computer by pointing to remote resources that are found opportunistically. For instance, suppose a user waiting at a busy airport gate needs to send several important documents before boarding the plane. Given the large number of users at this gate surfing the Web, the system knows that the amount of bandwidth available is insufficient to send all of the documents before the plane leaves. The system notifies the user (based on

---

[2] The government played a similar role in the development of today's Internet. The Computer Science and Telecommunications Board found that "rapid growth in the past decade was also based on the acceptance of an evolving set of common standards that enabled scaling up, competition, and interoperation. The development of the Internet suite of protocols, along with the establishment of processes for evolving them, is perhaps the most widely recognized example. A significant portion of these technologies and standards resulted directly from ongoing, farsighted government investment by a number of research agencies. Indeed, without this investment, it could be argued that the Internet phenomenon would not have come into existence—it was by no means an inevitable development" (CSTB, 2002).

current flight schedules) that sufficient bandwidth is available at a gate only a few minutes away. The user walks to the designated gate area, sends the important documents, and returns to the original gate in time to board the flight.

This example illustrates how cyber foraging can enhance the necessarily limited capabilities of portable or wearable computers by guiding users to locations where additional resources (power, bandwidth, etc.) are available. Another approach, proposed by researchers at Rutgers University (Goodman et al., 1997; Frenkiel et al., 2000), is the use of *infostations* to provide high-bandwidth connections for mobile devices. A network of frequent, short-range infostations could provide low-cost, low-power access to information services such as large files (e.g., books, videos), location-dependent information (e.g., maps), and remote information (e.g., for military personnel in the field).

Many research problems must be addressed before the use of surrogates (infostations or the compute- and data-staging servers used in cyber foraging) can be accomplished invisibly and seamlessly. Mechanisms are needed for discovering and selecting surrogates and negotiating their use. The computational, bandwidth, and power requirements of applications must be characterized in platform-independent ways. Techniques must be developed to ensure and verify an adequate level of trust in a surrogate, and practical boundaries must be established for what constitutes useful levels of trust. In addition, the shared use of surrogates leads directly to questions of load balancing and scalability. For example, it is not clear if admission control, best effort, or some new approach is best for surrogate allocation, or what the implications of these alternatives are for scalability or for the provisioning of fixed infrastructures to avoid overloads during peak demand.

Additional investigation is needed to establish reliable and cost-effective techniques for monitoring mobile resources, discovering resources as they come in and out of service, partitioning and off-loading computation, and staging data. Because energy is a particularly critical resource in mobile computing, the research investment should include location- and orientation-sensing techniques that adapt to battery state. The nature of the connections—they are brief, intermittent, uncertain, and unpredictable—requires new strategies and algorithms for caching and prefetching data. For example, when an intensive computation that accesses a large volume of data needs to be performed, the mobile computer can ship the computation to a surrogate. If a user is stationary, it might be possible to complete the session with one surrogate. However, if a user is passing quickly through an area (in a vehicle, for instance), the session may have to span multiple surrogates (with data cached at multiple locations). In addition, it may be desirable to have the surrogate stage data ahead of time in anticipation of the user's arrival in a given location. The long-term research goal is to develop the design principles and implementation techniques needed for well-engineered mobile computing environments, such as those suggested by cyber foraging and infostations.

Tracking and Predicting Location

The ability to manage information about the availability of devices based on their location is an enabling technology for many of the applications discussed in *IT Roadmap to a Geospatial Future*. In the case of mobile devices, the situation is complicated by the fact that location is transient (Wolfson, 2001). Suppose an object starts at the corner of 57th Street and Eighth Avenue in New York City at 7:00 AM and heads for the intersection of Oak and State streets in Chicago. A trajectory can be constructed using an electronic map geocoded with distance and travel-time information for every road section. Given that the trip has a starting time and assuming that speed is constant, we can compute the time at which the object will arrive at the beginning of each straight-line segment on the path. This trajectory information gives the route of the moving object, along with the time at which it will be at each point on the route. It is stored on a remote server and revised according to location updates from the moving object and real-time traffic conditions obtained from traffic Web sites.

The location of a moving object is inherently imprecise, because the object does not necessarily move in straight lines or at constant speed; therefore, the location stored in the database cannot always be identical to the actual location of the object. Assuming that one can control the amount of uncertainty in the system, how should it be determined? Obviously, lowering the uncertainty would come at a cost. For example, if a moving object transmits its location to a location database every *x* minutes or every *x* miles, then reducing *x* would decrease the uncertainty in the system but increase bandwidth consumption and location-update processing cost.

Recently developed methods can predict the location of a moving object at future times, based on the fact that objects often have some degree of regularity in their motion. A typical example is the home-office-home pattern. If this pattern can be detected, location prediction is relatively easy for rush hour. Note that patterns are only partially periodic (i.e., sometimes only part of the motion repeats). For example, Joe usually may travel from home to work along a fixed route between 7:00 and 8:00 every workday morning and back home between 5:00 and 6:00 PM, but he may do other things and go other places during the rest of the day. Further, the patterns are not necessarily repeated perfectly; the home-to-work trajectory may be different from one day to the next because of different traffic conditions. Lastly, the motion can have multiple periodic cycles (Joe may go fishing every Saturday and every other Sunday). The goal of future location prediction research should be to detect motion patterns that can be partially periodic, are not perfectly repeated, and have multiple periodic cycles and to use this information to estimate where an object is most likely to be. Protocols and mechanisms to authenticate and certify the location of an individual at any given time will be required, as will adaptation techniques for handling situations when location information becomes stale, is unavailable, or is deliberately withheld. Precursors of such technology include radio frequency identification tags embedded in retail products (such as clothing) and used for public transportation systems. Query language extensions will be required to allow applications to refer to future events and to support automatic, user-request triggers, such as “car navigator should inform the driver when any hospital is within 10 miles.”

RESEARCH CHALLENGES IN MOBILE INFORMATION USE

As the world and its inhabitants are increasingly “wired,” individuals traveling through and between places have real-time access to an increasing variety of information, much of it geospatial in nature. Freeing users from desktop computers and physical network connections will bring geospatial information into a full range of real-world contexts, revolutionizing how humans interact with the world around them. Imagine, for example, the ability to call up place-specific information about nearby medical services, to plan emergency evacuation routes during a crisis, or to coordinate the field collection of data on vector-borne diseases, using just a handheld device.

Portable Display Technologies

Underlying the goal of “geospatial everywhere” is the ability to obtain information on demand, wherever the user happens to be. This will necessitate the development of technologies and methods specifically accommodating user mobility. Traditional visual representation methods, developed for desktop (or larger) displays, are not effective in most mobile situations, where display screens are small and local storage and bandwidth capacities will always be more limited than in desktop systems. Research is needed to develop context-sensitive representations of geospatial information and to accommodate data subject to continual updating from multiple sources.

Although the available technologies provide limited visual representations of geospatial information in field settings, visual display remains the most efficient and effective method of geospatial access for sighted users. Accordingly, it makes sense to invest in the development of portable, lightweight display technologies, such as electronic paper, foldable displays, handheld projectors (which can be pointed at any convenient surface), and augmented reality glasses. To exploit these technologies, we also must invest in appropriate interaction paradigms, such as voice- and gesture-based interfaces applied to PDA-like devices.

Because the geographical context will be somewhat constrained, it may be possible to devise more natural interfaces. For instance, because the system will know where the user is located when a request is made, the spatial language of gestures or sketching movements may be interpreted more literally. Integrating two-dimensional (or three-dimensional) mobile displays, which support natural mechanisms for interacting with maplike representations and augmented reality methods and technologies, poses a range of technology and HCI challenges.

Supporting the acquisition and use of geoinformation from the field also will require attention to interaction issues associated with database access and knowledge discovery. Both efficient rendering and efficient transmission of geospatial representations are essential. A long history of research on map generalization provides an important conceptual base for meeting this challenge, but that research does not deal with real-time generation of dynamically changing representations. Rather, coordinated research drawing on both computer science (efficient algorithms) and cartography (understanding of the geospatial information abstraction process) is required. Intelligent mechanisms for transmitting data, such as context-sensitive data organization and caching, also must be developed.

Mobile Augmented Reality

Augmented reality technologies use virtual information representations (visual, aural, or other) to enhance human perception. Surveillance camera images that make crime perpetrators more recognizable is a simple example. Mobile augmented reality does this dynamically while the user moves through an environment. Heads-up displays, for instance, have been used to help jet-fighter pilots find their targets and to assist civilian drivers see objects in the road ahead when visibility is poor.

Because mobile augmented reality requires both detailed geospatial databases describing the fixed world and location-aware computing support to match the location of the user with that description, it is a classic example of a spatiotemporal application of geospatial information. As the geodata infrastructure expands, such applications will become increasingly important. Consider, for example, what it might mean in terms of human life if firefighters could look at a burning building and see (as a transparent layer superimposed over the building) a representation of the activities on each floor (retail space on the first floor, a fitness center on the second, offices for the next five, and apartments above).

Mobile augmented reality imposes constraints on interaction and display that go well beyond those already discussed. One issue is how the system should determine which aspects of reality to augment with which components of information. Real-world point-and-click offers one approach. Building on the desktop graphical user interface (GUI) metaphor, it allows users to interact with objects (in this case, integrated real/virtual objects) using a pointer device such as a gyro mouse or a laser pointer, effectively blurring the distinction between the physical and virtual worlds. The goal is to be able to point a wand, display-enabled goggles, or other mobile device at an object (e.g., a bus stop, building, or tree) and gain access to a wide range of information about that object (the time when the next bus will arrive, the location of the closest stop on an emergency snow route, or the best connections for getting to the airport by noon). Many research problems must be solved before this capability becomes reality. The stab line indicated by the device must be interpreted very precisely, which means that the orientation as well as the location of the wand must be determined. Identifying the object of interest will, in most cases, require a fusion of information from several sources, including a comprehensive 3D model of the locale, dynamic information from sensors if the object pointed to is mobile, and information derived from analysis of the image captured by the camera at the wand's tip. The human interface must not only help the user specify what type of information is desired, but also help him or her to disambiguate the target of the request. For example, if the wand is pointed at a building, the user must be guided to clarify which portion of the building is of interest (general information about the building, the floor plan for the second floor, the building address, etc.).

These examples of mobile augmented reality all deal with enhancing human vision. Research here could also yield significant benefits for sight-impaired individuals, helping them overcome many obstacles to freedom of movement. High-resolution geospatial data could deliver key information about the immediate environment to mobile users, through sounds or tactile feedback. Similar techniques could be used to augment human hearing. Research investments in this area not only could make it possible for users to hear sounds outside their normal perceptual range or to mitigate hearing deficiencies but also could provide added sensory input in situations where vision already is fully engaged.

OTHER RESEARCH CHALLENGES

The earlier parts of this article focused on the location-aware computing aspects of the *IT Roadmap to a Geospatial Future* report. Here we briefly highlight some of the issues and recommendations from the other two themes of the report (SEE FIGURE 1.1).

In the spatial-temporal databases there is a need for more comprehensive data models and query languages, particularly those capable of representing geometric objects that move and evolve continuously in space and time. Algorithmic issues include propagation of error due to discrete representations, I/O efficient algorithms for massive data sets, and kinetic data structures for evolving data. Development of ontologies for geospatial data is critical. The varied and complex nature of semantics of geospatial data came to the fore as a major cross-cutting issue in all areas. Tools for developing and maintaining ontologies as well as mapping between ontologies are an important area for research. There are numerous research opportunities in mining of spatiotemporal data. High dimensionality, uncertainly, and spatial or temporal correlations are problems that are not dealt with adequately in traditional data mining. Novel methods for mining complex geospatial objects that evolve over time and space are needed.

The presentation and interaction with users of large volumes of complex geospatial data present a range of research problems. A key barrier has been the lack of a comprehensive framework for understanding human interaction with geospatial information. The vast variety of geospatial phenomena and the disparate views of the phenomena by different disciplines complicate the problem. Issues of display technology and communication technology to support augmented reality are discussed in the report. The use of other interaction modalites (e.g., sound and touch) should also be explored.

The availability of massive amounts of high quality data as well as advances in computing and communications suggest the possibility of powerful new applications in science, commerce, environment, and government—but many research challenges remain. Research in geospatial information systems is inherently multidisciplinary and research efforts are likely to be most successful if conducted by a team that combines expertise in applications as well as information technology.

CONCLUSIONS

The rapid evolution of mobile computing, location sensing, and wireless networking has engendered a new paradigm for computing, where information services are able to recognize and adapt to the location of individual users. Location-aware computing can be activated by user requests about the immediate surroundings, or spontaneously triggered by events that are important to a particular user in a particular context. Potential applications span virtually every aspect of everyday life, including our ability to respond appropriately to natural and human-induced hazards.

There are still a number of research challenges that will have to be addressed before location-aware computing can become mainstream. Chief among these are developments in: viable infrastructure for location-sensing, including a national test bed; adaptive methods and mechanisms for mobile clients; viable support for privacy protection; and appropriate technologies for representing and displaying geospatial information “everywhere,” even when communication and display capabilities are severely limited. The investments will need to be significant, but the payoffs will be pervasive as well as critical to future well-being.

NATIONAL RESEARCH COUNCIL REPORTS

The National Academy of Sciences was chartered by Congress and signed by President Abraham Lincoln in 1863 to advise the government in scientific and technical matters. This institution expanded to include the National Research Council in 1916, the National Academy of Engineering in 1964, and the Institute of Medicine in 1970. These non-profit organizations, collectively called the National Academies, provide a public service by working outside the framework of government to ensure independent advice on matters of science, technology, and medicine. They enlist committees of the nation's top scientists, engineers, and other experts—all of whom volunteer their time to study specific concerns. The committee nomination and selection process assures that committees are well balanced and free from conflicts of interest, and can provide the range of expertise required. Supervisory boards and commissions oversee the study process to ensure its integrity. Reports are released only after outside reviewers, anonymous to the committee that prepared the report, agree that the findings are fully supported by the evidence presented. *IT Roadmap to a Geospatial Future* is the final report of the Computer Science and Telecommunications Board's Committee on Intersections Between Geospatial Information and Information Technology (Computer Science and Telecommunications Board, National Research Council, *IT Roadmap to a Geospatial Future*, National Academies Press, Washington, D.C., 2003). The report is available online at <http://www.cstb.org/project_geospatial.html>.

COMMITTEE ON INTERSECTIONS BETWEEN GEOSPATIAL INFORMATION AND INFORMATION TECHNOLOGY

The NRC convenes panels of experts to serve as volunteers to address critical national issues and give unbiased advice to the federal government and the public. Nominations to serve on NRC committees, which are solicited from a wide range of experts, are reviewed and approved by the National Academies. The role of this committee was not only to organize the workshop but also to sift through many inputs to distill key themes, ideas, and recommendations for its final report, *IT Roadmap to a Geospatial Future*.

RICHARD R. MUNTZ, University of California, Los Angeles, *Chair*
TOM BARCLAY, Microsoft Research
JEFF DOZIER, University of California, Santa Barbara
CHRISTOS FALOUTSOS, Carnegie Mellon University
ALAN M. MACEACHREN, Pennsylvania State University
JOANNE L. MARTIN, IBM.com e-business Solutions, IGS Global Web Solutions
CHERRI M. PANCAKE, Oregon State University
MAHADEV SATYANARAYANAN (SATYA), Carnegie Mellon University and Intel Research Pittsburgh

Staff

CYNTHIA A. PATTERSON, Study Director and Program Officer
MARJORY S. BLUMENTHAL, Director
MARGARET HUYNH, Senior Project Assistant

JUST-IN-TIME MAPPING SCENARIO

The committee created several scenarios to illustrate how additional research could enhance the accessibility and usability of geospatial information. The "Just-in-Time Mapping" scenario shows how future technology could be harnessed to manage situations, such as the aftermath of a tragedy, in which human lives are at risk:

> A devastating earthquake, "the Big One," has hit downtown San Francisco. A huge complex of skyscrapers built on reclaimed land has caved in. It is feared that thousands of people are trapped in the rubble. Emergency personnel have little time in which to rescue them. Although cranes and heavy earthmoving equipment have been put in place with amazing speed, it is not clear how the excavation should proceed. With unstable interior spaces and broken gas and electric lines, it is not clear how to excavate in a way that is fast yet will not further injure survivors. Time is ticking away and with it, hopes for survival.
>
> With few options left, the disaster-relief director decides to use an experimental, robot-based, just-in-time three-dimensional (3D) mapping capability that was developed after the September 11, 2001, World Trade Center calamity. Thousands of small mobile robots ("mapants") burrow into the rubble. Each robot is equipped with location-sensing ability as well as with visual, toxic gas, and other sensors. The key to the speed of the just-in-time mapping application is the enormous parallelism made possible by the huge number of mapants. To conserve energy and to enable communication through the rubble (which has large concentrations of steel), the robots use ad hoc wireless communication to share data with one another and with high-powered computers located outside the rubble. The computers perform planning tasks and assist the mapants with compute-intensive tasks such as image recognition and visualization of the map as it is constructed.
>
> Although early trials of this approach have been promising, it still is considered highly experimental. This is its first use in a real-world event. After an initial planning phase, the mapants are let loose. Each has its own mission but also is cognizant that this is a team effort. The thousands of mapants organize themselves according to the planned strategy, burrowing and climbing as needed. They possess sufficient autonomy to handle unexpected situations. Once a mapant has reached a designated region, it explores that region and reports on what it senses. The input from these mapants is combined to produce a 3D map showing (with centimeter accuracy) the location of potential survivors, fires, dangerous gases, and other critical information. Human experts monitor the progress of the mapants, review early maps, direct the robots to areas of interest, evaluate dangers, and select strategies for mapping refinements. As the mapping of top layers of the rubble is completed, the mapants move deeper and excavation of the just-mapped region begins. Within 48 hours, many survivors are rescued who might have perished were it not for the assistance of the mapants.

Clearly this scenario is science fiction today. Yet we might question why it appears to be so futuristic, since many of the component technologies—such as miniature robots, ad hoc wireless communication, and strategy planning—are active areas of research today. The answer is that the situation represented in this scenario would considerably stretch each of these technologies and, more importantly, would require that they be integrated in ways that have never before been attempted. Following are some examples of the scientific and engineering challenges that will have to be overcome:

- Engineering of small, autonomous mobile robots capable of burrowing, climbing, and so forth;
- Planning and coordination of scalable, ad hoc robot diffusion;
- Centimeter-accuracy location sensing without supporting infrastructure;
- Robot-to-robot wireless communication through steel-filled rubble;

- Communication infrastructure that is robust and pervasive, even during emergencies;
- Real-time analysis and integration of heterogeneous data (detection of fire, dangerous chemical leaks, or life signs) from distributed sources;
- Real-time map construction and refinement; and
- Map-based user interfaces that allow coordinated teams of human experts to quickly and easily direct the mapants and analyze the results.

REFERENCES

Frenkiel, Richard, B.R. Badrinath, Joan Borras, and Roy Yates. 2000. “The Infostations Challenge: Balancing Cost and Ubiquity in Delivering Wireless Data.” *IEEE Personal Communications*, April, pp. 66-71.

Goodman, D.J., J. Borras, N.B. Mandayam, and R.D. Yates. 1997. “Infostations: A New System Model for Data and Messaging Services.” In *Proceedings of the IEEE Vehicular Technology Conference*, 97(2):969-973.

Hightower, Jeffrey, and Gaetano Borriello. 2001. “Location Systems for Ubiquitous Computing.” *IEEE Computer*, 33(8):57-66.

Pister, Kris. 2002. “Smart Dust: Autonomous Sensing and Communication in a Cubic Millimeter.” Available online at <http://robotics.eecs.berkeley.edu/~pister/SmartDust/>.

Satyanarayanan, M. 1996. “Fundamental Challenges of Mobile Computing.” In *Proceedings of the 15th ACM Conference on Principles of Distributed Computing*, February.

Satyanarayanan, M. 2001. “Pervasive Computing: Vision and Challenges.” *IEEE Personal Communications*, 8(4), August.

Wolfson, Ouri. 2001. “The Opportunities and Challenges of Location Information Management,” prepared for the National Research Council workshop on Intersections Between Geospatial Information and Information Technology.

[[ADDITIONAL REFS WE COULD ADD – THOUGHTS?]]

Golledge, Reginald G., Roberta L. Klatzky, Jack M. Loomis, Jon Speigle, and Jerome Tietz. 1998. “A Geographical Information System for a GPS-Based Personal Guidance System.” *International Journal of Geographical Information Science*, 12(7):727-750.

Haartsen, J.C. 2000. “The Bluetooth Radio System.” *IEEE Personal Communications*, 7(1), February, pp. 28-36.

Institute of Electrical and Electronic Engineers, Inc. (IEEE). 1997. “Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications.” IEEE Std 802.11-1997.

Noble, B., M. Satyanarayanan, D. Narayanan, E. Tilton, J. Flinn, and K. Walker. 1997. “Agile Application-Aware Adaptation for Mobile Computing.” In *Proceedings of the 16th ACM Symposium on Operating Systems Principles*, October.

Priyantha, N.B., A.K.L. Miu, H. Balakrishnan, and S. Teller. 2001. “The Cricket Compass for Context-Aware Mobile Applications.” In *Proceedings of the Seventh Annual International Conference on Mobile Computing and Networkin*g, July.

Rappaport, Theodore S. 1996. *Wireless Communications: Principles and Practice*. Upper Saddle River, N.J.: Prentice Hall.

Williams, S. 2000. “IrDA: Past, Present and Future.” *IEEE Personal Communications*, 7(1), February, pp. 11-19.