1

Introduction and Overview

Information technology (IT) is on the verge of another revolution. Fueled by the increasing capabilities and ever-declining costs of computing and communications devices, IT is being embedded into a growing range of physical devices linked together through networks. These changes are sometimes obvious—pagers and Internet-enabled cell phones, for example—but often IT is buried inside larger (or smaller) systems in ways that are not easily visible to end-users. Audiovisual equipment, home or office appliances, automobiles, aircraft, and buildings themselves all contain growing numbers of microprocessors that are networked together. The range of applications continues to expand with continued research and development. Aircraft manufacturers are already examining the possibility of incorporating processing devices into the wings of aircraft to allow fine-grained control of airflow and, hence, lift and drag; health researchers are investigating microscopic sensors that could traverse the bloodstream, monitoring health conditions and reporting them wirelessly; consumer electronics and information technology companies envision homes filled with intelligent devices that can interact with each other, homeowners, and appliance manufacturers to improve the quality of daily life. The Internet, wireless networking, inexpensive cameras, and automotive telematics can be combined to pass information to millions of commuters in large cities so as to reduce delays, frustration, energy use, and air pollution. Sensor networks can be deployed in large agricultural areas to monitor and report on crop quality and the environment, adjusting irrigation and fertilization as necessary.

To some extent, the emergence of networked systems of embedded computers (EmNets) is simply a natural evolution of the historical trend in computing and communications technologies toward smaller, more powerful information technology devices that have become more ubiquitous (see Box 1.1). As computing has migrated from mainframe computers to minicomputers, personal computers, laptops, and, most recently, palmtop computers and information appliances, it has become more widespread and more a part of everyday life for millions. Meanwhile, embedded computers have been used in automobiles, aerospace engineering, and military applications for quite some time. Advances in networking technologies, including the expansion of the Internet and wireless communications networks, have amplified these trends by making information easier to share and increasing the amount of information that is shared.

At the same time, the shift to EmNets represents a radical departure from this lineage. While most traditional computers tend to interact directly with human operators—typically accepting input through a keyboard and providing output on a visual display—EmNets will interact more directly with the physical world. They will sense their environ-

BOX 1.1
Toward Ubiquitous, Networked Computing

The vision of a world filled with large numbers of computing elements, many of which are hidden inside other objects and networked together, is not new. Trends in the miniaturization of computing and communications elements have been manifested for decades, leading to numerous predictions of computing power being integrated imperceptibly into daily life. One of the leading visionaries, the late Mark Weiser, formerly the chief technologist at the Xerox Palo Alto Research Center (PARC), described in the early 1990s a concept of ubiquitous computing in which computation would blend invisibly into the environment, much as written communication has become so common a part of the physical world that little thought is given to the technology of writing (Weiser, 1991; 1993). Others have elaborated on related themes, coining terms such as pervasive computing (NIST, 1999) and invisible computing (Norman, 1998) to describe the proliferation of information technology into myriad devices and applications. Although differing somewhat in their details, these visions of the future of computing derive from a common set of observations about the rapid pace of innovation in information technology: namely, advances in very-large-scale integrated circuits (VLSI), the increasing bandwidth of wireless and wireline communications media, improvements in wireless communications technologies, and significant efforts in architecture and infrastructure. (See Chapter 2 for a more detailed discussion of enabling technologies.)

ments directly, compute necessary responses, and execute them directly. EmNets will also need to operate in a highly resource-constrained environment. There may be limited power, limited communications bandwidth, limited time, and limited memory. EmNets' heterogeneous components will often be embedded in long-lived structures, thereby making interoperability over time an important issue. All of the above will require new ways of thinking, not just at the input and output ends, but about the very fundamentals of computing and communications. Ways will be needed to ensure that such systems operate reliably, safely, and predictably; that they provide their users with necessary information about their current operating state; and that they can accommodate changes in the overall system configuration or in their operating environment. In addition, EmNets present new opportunities for pervasive, transparent monitoring and information aggregation while at the same time generating a host of privacy and other ethical concerns.[1]

This report identifies and examines research challenges posed by EmNets and provides guidance for addressing them. It addresses fundamental research issues, primarily at the system level, with some attention given to components. The report recognizes that if current technology is applied naively to EmNets, the results could be disastrous. Failures that are all too common today in information technology systems (e.g., security lapses, system outages, safety problems, unanticipated performance) could have even more serious consequences. As such, this report builds on previous work by the Computer Science and Telecommunications Board (CSTB) in the areas of large-scale systems and applications and trustworthy networked information systems (CSTB, 1999; 2000), but in the context of EmNets. It offers recommendations for organizing research and education programs to better ensure that the challenges are being adequately addressed.

EXAMPLES

Characterizing EmNets precisely and uniquely is a challenge. To facilitate this task, the committee decided to introduce three examples, which help to show the variety of systems this report is addressing. Many examples could have been chosen to illustrate EmNets, so those selected

[1] Bill Joy's wide-ranging discussion of robotics, nanotechnology, and genetic engineering and their ethical and social concerns (Joy, 2000) attracted attention because of the author's reputation as a technologist. But only a little imagination is required to link EmNets to scenarios that would call for considering ethical and social issues while the technologies are under development.

should not be seen as canonical in any sense. Moreover, it is virtually a certainty that EmNets will be used in ways that are currently unforeseeable. These examples, which are very distinct applications, should be viewed as representing the potential of EmNet technology. All three combine a number of separable subsystems that would normally be developed independently, preferably with an eye toward interoperation and integration over time. They all offer significant functional and economic incentives for deployment and proliferation. In addition, they exemplify tensions between often opposing forces: complexity and comprehensibility, information aggregation and privacy, and safety and autonomous power.

Notwithstanding all of the above, these examples can be seen as demonstrating, in broad strokes, the potential of EmNets at several different scales. The first example discusses automotive telematics, where the main locus of interaction is a vehicle. The second describes precision agriculture, where the EmNet is distributed over a wide area. The final example incorporates individuals, vehicles, and the surrounding environment into a comprehensive defense systems scenario. A further complication arises that increases the already formidable challenges presented by EmNets when one imagines the experiences of an individual who "joins" and subsequently "leaves" various EmNets while moving through space and time. Whether location- or domain-specific, EmNets will be connected to each other for certain functions, adding yet another level of complexity.

Example 1: Automotive Telematics

It should come as no surprise that the modern automobile is already a rolling network of embedded computers. In model year 2001, cars have between 20 and 80 microprocessors controlling everything from the running of the engine to the brake system to the deployment of the airbags. These numbers are expected to grow dramatically over the next several years as automobile manufacturers look for ways to transition electromechanical control systems into electronic control systems. Microprocessors also control the windshield wipers and the door locks and are increasingly used in the entertainment systems. These microprocessors are rarely self-contained; almost all interact with other microprocessors in the automobile through a network, which can be one of half a dozen proprietary or industry-specific designs.

Currently, these networks are highly engineered systems in which each microprocessor and the overall network are carefully designed as a whole. In fact, there are generally two distinct networks in today's cars. The first is the network of safety-critical components, such as those that control the engine and the braking system. The second, often called the telematics system, controls non-safety-critical functions such as the entertainment systems, door locks, and trunk release. These two networks are completely separate, ensuring that the safety-critical portions of the car cannot be compromised by the telematics components.

However, as the complexity of the network and the functionality of the networked elements grows, the ability to approach the networks as single, fully engineered, closed systems is being strained. In particular, a number of forces work against the fully engineered, closed systems approach, including the following:

- *The disparity between the design cycle of the car and the design cycle of the embedded components.* A car takes approximately 5 years to design, and the embedded components are among the first things designed into the car. This has meant that cars contain embedded systems that are significantly less functional than the systems available at the time of the car's manufacture.
- *The desire to allow easy upgrade, either by the manufacturer (in the case of safety-critical components) or third parties (in the case of telematics), over the lifetime of the car.* Such flexibility generates cost savings, as the recall of a part can be tremendously expensive, and also reflects the reality that the lifetime of a car is now 8 to 10 years rather than 3 to 5, so building a post-purchase income flow has become important.
- *The desire to allow owners to integrate their own devices into the auto.* Such devices include personal digital assistants (PDAs) and cellular phones, which can be made more useful (by, for instance, integrating the address book in a PDA with the navigation system in the car) or safer (by, for instance, integrating the cell phone with the speaker system of the car, making the phone hands-free) if such integration is possible.

There is also pressure to break down, to some degree, the strong division between the safety-critical network in the car and the telematics network. Many automobile manufacturers want to move away from the current model of diagnostics to a model of prognostics, which allows them to monitor their products for upcoming faults and allow those faults to be corrected before they happen. For this to be possible, there needs to be a way for the information gathered by the safety-critical parts of the automobile to be sent to the automobile manufacturer. One obvious way of doing this is through the use of automated cell-phone technology (separate from personal use phones) that most cars will have. Currently, however, the cell phone is part of the telematics network of the car, not part of its safety-critical network.

All of these possibilities are taken from current thinking about the network of embedded systems in the car. The outlook for the future complicates the intra-auto network considerably. The major automobile companies plan to change the car from a self-contained network (or pair of networks) into a node in a much larger network. One approach to this is General Motors' immensely successful OnStar offering.[2] OnStar connects the car to the manufacturer, allowing the latter to monitor emergency situations and give on-demand help to the occupants of the car. Not only has this service provided GM with a market differentiator, it has also allowed the company to begin to provide a very profitable subscription service, giving it a revenue stream that is less prone to the fluctuations traditional in the automotive market. The notion of the automobile as a mobile, networked recipient of content is an outgrowth of this seemingly simple beginning.

As envisioned by the automobile companies, the driver of a car will be able to get on-demand directions to anywhere desired, including those locations that are contextually based. From the car's current position, the driver will be able to get directions to the nearest restaurant of a particular type, or the closest automatic teller machine, or an available parking space. The occupants of the car will be able to receive information about the history of the place they are seeing or about its landmarks, or they will be able to get on-demand video or audio stream. The car will be monitored, in real time, to support safe operation, and the driver will be informed of the maintenance needed to keep the car from breaking down. Software upgrades to emission controls or safety systems will be downloadable (obviously at some safe time) to where the car is, making it unnecessary to take the car into the shop. While many of these innovations seem far-fetched, they are in fact being prototyped now;[3] it is likely that new advances and applications will emerge as the technology becomes widely deployed. For example, instrumented vehicles and highways could provide data that would inform a traffic management or control system. Emergency vehicles could be networked to traffic lights to adjust their timing and facilitate passage through crowded areas. Undoubtedly, many new applications of automotive telematics systems connected to larger EmNets are as yet unforeseen.

[2] For more information, see <http://www.onstar.com/>.

[3] A presentation to the Computer Science and Telecommunications Board by Akhtar Jameel of DaimlerChrysler Research in January 2001, "The Future of Vehicle Computing," touched on many of these issues.

Example 2: Precision Agriculture

Incorporating EmNet technology into agriculture can be seen as a logical follow-on to the great advances in crop management over the last several decades. Fertilizers, water supply, and pesticides, among other things, have been experimented with and adjusted in order to learn how best to manage crops and to increase productivity. Even with these adjustments, variations in terrain (soil, elevation, light exposure, microclimates, and so on) can make solutions based on large-scale averages suboptimal, especially for highly sensitive crops such as wine grapes and citrus fruit.

This is where EmNets, in the form of precision agriculture,[4] are beginning to play a role.[5] Precision agriculture features the deployment of sensing and actuation at a much finer and more automated granularity than has been available before. This will allow adjusting water, fertilizer, and pesticides to the minimal levels needed for a particular local area, resulting in better yields, lower costs, and less pollution-causing runoff and emissions. The data collected will be analyzed later on (imagine a viticulturist searching for the best places to cultivate grapes for the next vintage).

Adaptation to changing environments will be a crucial component in EmNets used for precision agriculture. Sensors and actuators can be used to very precisely control the concentrations of fertilizer in the soil, based on information gathered from the soil itself, the ambient temperature, and other relevant environmental factors. While there are models for how much fertilizer and water are needed for crops under various conditions, those models are imperfect, mainly because not enough accurate data have been collected across diverse agricultural systems. EmNets can provide that data. Incorporating feedback into the system through the use of sensors, actuators, and adaptation will allow a more fine-grained analysis that could adjust flow rate and duration in a way that is informed by local soil conditions and temperature. One can imagine the use of such precise information in particularly sensitive crops. Sensors that are able to monitor the crop itself (sugar levels in grapes, for example) to provide location-specific data could prove very effective. EmNets will need to be adaptive, multimodal, and able to learn over time in order to solve the problems described above.

Information gathered by sensor networks in a field could be used to

[4] For more information on precision agriculture, see BANR (1998).

[5] See Li and Wang (2000) for a description of a wireless sensor network for precision agriculture.

guide planting for maximum yields, in addition to monitoring and reporting on the status of the crops. A future application of EmNets might be to deploy sensors for the early detection of bacterial development in crops or viral contamination in livestock. Another application might be to employ EmNets to monitor flows of contaminants from neighboring areas and send alerts when necessary.

EmNets are also being extended to livestock management. Current computerized feeding systems for dairy cattle, for example, can adjust feed and vitamins for individual animals. Networked sensors, including swallowable sensors, to monitor amounts of food eaten, activity/exercise, and vital signs will provide valuable health information about individual animals and the state of the herd as a whole.

These systems are moderately engineered (along a spectrum from highly engineered to ad hoc), but the need to work under a wide range of unpredictable environmental conditions, as well as to interact with farm vehicles and new elements of the system as they become available, argues for adaptability within the EmNets at multiple time scales.

Example 3: Defense Systems

EmNet applications to defense systems include battlespace surveillance, monitoring the condition and location of materiel and vehicles, monitoring the health status of personnel, and making information accessible to individuals in the field.[6] As efficiency and speed of deployment become more important, the requirements for network access to assets and information become more important too. Each of these application areas is discussed briefly below.

Distributed EmNets in the battlespace will provide seismic, acoustic, magnetic, and imaging tactical information. EmNets can be dispersed by airdrop, inserted by artillery, and/or individually placed by a team securing a building. Military forces are expected to exploit EmNet battlespace surveillance systems to provide capabilities for battlefield shaping and force protection. Battlespace shaping capabilities restrict the movement of an opponent or constrain its advance or retreat. EmNets can provide the critical threat-identification information that enables remote engagement of targets and the halting or redirection of opponent forces. Force protection capabilities provide security on the battlefield and act as a force multiplier. EmNets enable a new force-protection capability by providing threat identification and early warning of an infiltration or

[6] EmNet research in these areas will probably prove particularly relevant for DARPA's Future Combat Systems program. See <http://www.darpa.mil/fcs/index.html>.

threatening advance. Force protection may be implemented by distributing EmNets around a protective perimeter or deploying them in advance of maneuvering troop formations. EmNets may allow a small force to operate with the security of a larger force by exploiting densely distributed, autonomous EmNet detection networks.

EmNets offer a new approach to battlespace surveillance. In the past, battlespace sensor systems were large and required large teams for deployment. As expensive assets, they were deployed only sparsely. EmNets, in contrast, involve less expensive, even disposable, devices that may be deployed in large numbers with a high spatial density. This allows the typical EmNet sensor to detect stronger signals from threats than the signals detected by more sparsely distributed sensors, facilitating a response to those threats. Because they are closer to the targets they need to detect, EmNets also engage fewer threats within their area of regard, simplifying signal identification and data association. EmNets can exploit their networking capabilities to cooperatively identify and track the motion of threats.

EmNets in battlespace situations must be highly interoperable and able to accept data from and provide data to other systems. Data from various kinds of sensor platforms (airborne, vehicle-mounted, ground-based, and so on) will need to be integrated and processed. Combining locally derived information with information from remote locations will be important, enabling updates to situational descriptions on a very short time scale. In addition to accruing and processing the data, EmNets will need to make such data readily accessible to personnel, requiring good user interfaces. Such dissemination might involve airborne relays or satellite communications, making communications another major challenge for EmNets in the application. These communications will need to remain secure while resisting jamming, detection, and interception. Challenges are also faced in the implementation of distributed computing for EmNets that must operate at low energy dissipation while maintaining a network for exchanging the appropriate threat signal characteristics.

In addition to battlespace shaping and force protection, EmNets will also be used for asset management. Defense forces rely on diverse vehicles, weapons, and equipment that require a mission-critical, high level of availability.[7] EmNets enable distributed, condition-based monitoring for detecting wear and faults in vehicle chassis systems and vehicle power trains. Applications include wheeled and tracked land vehicles and rotary- and fixed-wing aircraft. Prototype EmNet networks have ap-

[7] Large quantities of equipment in many locations create significant logistical challenges that may also benefit from the use of EmNets.

peared in condition-based monitoring onboard Navy ships for power plant monitoring. EmNet condition-monitoring applications require compact, low-power devices that measure and locally evaluate vibration and temperature signatures from rotating and reciprocating equipment. EmNet monitoring also applies to battle damage assessment and fire safety. The challenge of battlespace monitoring for EmNets includes the implementation of low-power, compact devices capable of both high-performance sensing and signal processing, along with networking, self-configuration, adaptation, and collaborative sensing, to exploit the distributed processing capabilities. All are needed to achieve unattended, robust, long-lived systems.

EmNets will also be applied in more tightly coupled systems, such as smart materials and structures. Collections of sensors and actuators on airplane and submarine hulls will enable new modes and efficiencies of operation by adjusting the physical properties of the surfaces to environmental and task conditions. In addition to developing the requisite MEMS components, this application will require many of the developments described in this report, from computational models to distributed coordination and safety evaluation.

EmNets also appear in health status monitoring of personnel. An important emerging requirement is for technologies that provide troops with personal location capabilities to enable security within a platoon and that monitor health, detect injury, and provide notification of injury. Here, EmNets must be wearable and integrated into existing or dedicated networks. The technologies may also be used to detect the use of biological or chemical warfare agents. Challenges include the need for security and low-power operation and the support of multiple biomedical sensor channels. Ultimately, the combination of EmNets for surveillance, condition monitoring, and personnel health status will enable a new tasking, control, and safety capability accessible at multiple command levels.

Finally, making all of the information described above—along with other dynamic, mission-specific information—readily accessible to the warfighter is a task for which EmNets as described throughout this report will be well suited. Vast amounts of information are available in battlespaces that, put to use, could increase the survivability and effectiveness of warfighters. For example, sensors and wireless communications could be used to keep track of the exact location of team members and enemies. Providing warfighters with data on asset locations and readiness, team members' health and capabilities, and overall battlespace information in an accessible, manageable fashion could greatly increase their capabilities and effectiveness.

UNDERSTANDING NETWORKED SYSTEMS OF EMBEDDED COMPUTERS

With the above examples as starting points, this section describes some of the features of EmNets and issues related to them that should be kept in mind when developing a research agenda. Without attempting a rigid definition of networked systems of embedded computers, this report discusses systems with the following general characteristics:

- *Multiple interacting nodes.* EmNets involve the interaction of more than two embedded computing elements or nodes. The systems of greatest interest are those in which the number of interacting elements is very large (for example, on the order of thousands of nodes).
- *Embedded in control systems operating without human intervention.* EmNets are intended to operate largely without human intervention. Although they may provide information to human operators and require some degree of supervisory control, they are often part of an automated control loop (that is, the system adjusts itself when necessary and directs component behavior), and they tend to interact more directly with their environment than traditional computing systems and to assume a high degree of autonomy. Computation can be local (at the nodes/elements) or centralized or somewhere in between, with localized or regional levels of hierarchical control. In any case, they tend to be tightly coupled to the physical world. They are therefore usually located close to the elements they monitor or control, and they operate in real time.
- *Purpose other than general computing and communications.* The computing elements in EmNets are themselves components of larger systems whose primary purposes are other than general-purpose computing or communications. The elements do not form a general-purpose computer even though particular components of the system may be general purpose. The individual computing elements help to monitor and control the local system, acquiring information from a variety of sensors, implementing changes through a variety of actuators, making decisions locally, and/or possibly relaying processed information to decision makers.
- *Natural or engineered contexts*. EmNets may be incorporated into either natural or engineered systems. The EmNets themselves are engineered, but they may be deployed in a natural system such as the local environment to provide information for scientists, urban planners, or military commanders. They may also be deployed as part of a larger engineered structure such as an aircraft or building.

Within systems that meet these criteria there are useful distinctions to be made. In particular, the following dichotomies characterizing how

EmNets, their requirements, and the applicable technical solutions differ will often be referred to:

- *Energy-constrained nodes versus non-energy-constrained nodes.* Energy-constrained devices are those that are not tethered to an easily replenishable energy source and have a small form factor (size, shape, and total volume), as well as those that exist where heat dissipation is a negative factor. Small form factor implies a fundamental limit on battery size, which in turn sets a fundamental limit on the number of bits that can be processed and/or communicated by the device during its entire lifetime. Other energy sources can be exploited in some cases, but in the general case components will rely on traditional battery technology for the foreseeable future. In this context, energy is the one system resource that is *not* easily renewable. Memory can be reclaimed and bandwidth-consuming data can be delayed to a time when congestion has dissipated, but once a unit of energy has been used, it cannot usually be replenished without intervention beyond the scope of what software can accomplish. When energy is a constraint, communication is often the major consumer of the energy. This, in turn, will have significant influence on the way systems are designed.
- *Fixed topology versus flexible topology.* Virtually all the systems considered here must continue to operate in the presence of node arrival, departure, and failure. That is, configuration will not remain constant throughout a system's lifetime. However, some of the systems are dominated by a fixed topology, whereas others are dominated by a flexible and variable topology that changes significantly during the course of regular operation. A fixed topology facilitates testing and repeatable deployment. Flexible topology introduces a new dimension of variability under which a system's performance must be verified.
- *Safety-critical applications versus non-safety-critical applications.* Some of the systems described will be used in safety-critical applications. When these systems malfunction, property can be damaged irreversibly and people harmed. The implications for designing and engineering such systems are fundamentally different from those for systems in which malfunction produces only degraded speed or visual quality, or even economic harm. Further, many EmNets will utilize general networking protocols. These protocols were originally precluded for safety-critical environments such as aircraft, but newer tools and techniques are starting to emerge and could be greatly enhanced by appropriate research.
- *Highly engineered versus unconstrained, ad hoc systems.* Some EmNets are highly engineered systems, such as those used in ships and aircraft to perform particular functions, like monitoring and controlling the performance of the engine. These are more traditional applications of embedded

computing, and they have been the subject of considerable engineering design work. They must, in general, meet strict criteria for system performance, reliability, and safety. They are highly constrained in that system elements are determined during the design and implementation of the system and the configuration of the system is fully controlled. The addition of networking into such systems allows the embedded computing devices to be remotely upgraded (e.g., new code can be downloaded to them to provide new or improved capabilities) or to relay information to a centralized source (e.g., for monitoring performance or use of resources). It also allows information to be shared among embedded devices to aid in local (and global) decision making. Other EmNets are unconstrained, ad hoc systems that have limited a priori system design and limited (or no) control over the overall system configuration, such as in sensor networks deployed in battlefield situations or in public smart spaces.[8] New elements can be introduced into such systems by a number of actors/ participants, and the systems will automatically reconfigure. Such systems can be expected to have a high degree of heterogeneity in the computing elements they contain and a dynamic structure as elements enter and leave the network. A particular challenge is ensuring that the overall system can meet global levels of performance as components are added to or removed from the system. There are, of course, EmNets that fall between the highly engineered and completely ad hoc categories.

HOW EMNETS DIFFER FROM TRADITIONAL SYSTEMS

EmNets are a composite technology, built as aggregations of software and hardware elements. Any given part of a network of embedded computers will look familiar to technologists: the networking constraints will find partial solutions in today's literature; the software controlling the nodes will start out as a variant on today's real-time control code; the hardware at the nodes will be developed from today's best microcontrollers, MEMS sensing devices, and interconnect transceivers. However, as the rest of this report makes clear, *incremental improvement to today's solutions will not suffice to realize the full potential of EmNets.*

The development of packet-switched networks was in a similar nascent period in the late 1960s and early 1970s. Few at the time could have predicted the development of this basic technology into today's Internet,

[8] Smart spaces are home or work environments containing information appliances, embedded computers, sensors, cameras, and microphones that allow people to perform tasks efficiently by offering access to information and assistance from computing technology through a variety of input devices and by monitoring on the part of the space itself.

a world-encompassing, ubiquitous communication network that has already eclipsed the telegraph and telephone in the variety of activities and services it supports. By the 1990s, its processing, routing, and interconnection aspects were becoming well understood. The extrapolation to Web sites, search engines, portals, and so on was by no means obvious, even to people working in related fields. The power, universality, and potential of EmNets will stem from combining these components into a system that is more than the sum of its parts. The dangers and difficulties will likewise emerge once the components have been combined, but they will not be immediately visible from any particular piece.

While many of the solutions found for EmNets might apply to other kinds of systems to one degree or another, what is unique about the problems posed by EmNets is the set of constraints on their solutions, several of which are discussed below. While one or even more of these constraints might be present for a traditional system, the combination is what poses one of the largest research challenges for the development of EmNets. More specifically, EmNets present the challenge of building large systems that are

- *tightly coupled to the physical world* and each other in a
- *resource-constrained environment* that will
- *persist for long periods of time* while consisting of
- *many interacting components* and being
- *used and interacted with by nonexpert users*.

Research needs to turn, as it did at the corresponding time for packet-switched networks, to developing the appropriate models, abstractions, and methodologies that will make it possible to build these systems on a large scale, for a wide variety of uses, by a necessarily large collection of people. These factors are elaborated on below.

EmNets Are Tightly Coupled to the Physical World

As noted previously, a major distinguishing characteristic of EmNets is that they interact strongly with the physical world. One EmNet might control all of the major systems of a large battle cruiser. Another might control tens of thousands of actuators based on tens of thousands of sensors to maximize the efficiency of a farm (BANR, 1998). They sense the physical world (e.g., its temperature, air quality, soil factors, or engine vibrations), they communicate and process those sensory data, and in real time they cause physical actions to be taken. Each node of an EmNet might be responsible for, say, one square meter of a farm. In the event of a one-node failure, data from geographical neighbor nodes might be interpolated, so that the affected square meter of farmland does not go unattended until repairs can be made. Accordingly, the precise geolocation of that node is important in a way that is seldom true of today's networks.

An EmNet (hypothetically) controlling a ship will necessarily be held to a much higher standard of performance and trustworthiness than, say, a traditional local area network (LAN) in an office whose primary function is to provide intra-organizational communications capability. If such a LAN goes down, productivity is lost and users become disgruntled. The loss of a ship's control at an inopportune time due to failures in an EmNet physically coupled to critical control mechanisms could result in a collision. This physical coupling of many EmNets means that safety considerations play a paramount role.

EmNets' tight coupling to the physical world also raises issues of usability. Individuals interacting with EmNets are not likely to think of themselves as interacting with a computer or computational device but rather with the objects to which EmNets are coupled (e.g., a sprinkler system as opposed to a digitally controlled irrigation device.) This has broad ramifications for usability research and for safety, reliability, and security as well.

EmNet Nodes Are Often Resource-Constrained

EmNet nodes are likely to be untethered so that they can be deployed in very close proximity to, or even embedded within, the physical systems they are designed to support. This factor places important constraints on the EmNet nodes, organization, system policies, and hardware. Untethered and/or mobile computing elements are usually battery operated, or perhaps they are very low power and run from solar panels. The limited amount of raw power available will have a substantial effect on all aspects of EmNets, from the amount of computation that can be performed on a local physical sensing node to how much bandwidth can be achieved, across what distance, by the EmNet node input/output links (e.g., radio). EmNet nodes may also have important physical constraints, such as allowable thermal dissipation or radio bandwidth limits. For example, an EmNet consisting of a large set of detectors deployed over an area of countryside will have to limit overall radio transmissions in order to avoid massive interference with other EmNets, normal communications traffic, and local regulations. EmNets that include sensors carried by the human body will have to be thermally cool to be practical. There are other kinds of resource constraints aside from power. EmNet components may have limited memory and/or bandwidth available to them. Energy constraints may limit the amount of storage available. Such resource limitations place constraints on the amount of computation and communication that can be accomplished.

EmNets' Long Lifetimes

The artifacts within which EmNets are embedded will undoubtedly have very long lifetimes compared with the lifetimes of the rapidly changing technologies that support the EmNets. Just as it has taken many years to upgrade the basic telephone wiring systems to homes, despite growing demand for bandwidth, EmNets deployed in buildings, on farms, or in the countryside will face this same problem. The longevity of EmNets will thus have to be taken into account during design, as the basic technology will continue to evolve and the previously deployed system will eventually have to interoperate with the new technologies. As networked, embedded devices are scattered throughout the environment, their useful technological life will be determined by Moore's law. Older devices may consume too large a share of valuable resources, so mechanisms for identifying, locating, and replacing or upgrading them will be necessary. The upgradability of today's computing systems is a marketing feature, but for EmNets it is a basic requirement.

The uses to which EmNets will be put may vary considerably over time. A system may have components that are used to measure physical properties and provide raw data that will be elaborated by other components or other systems. It is not always possible in advance to predict what the data will be used for.[9] A change in the application, or in the overall computing structure, may take place while the system and its components persist. In addition, it is very unlikely that entire EmNets will be replaced; instead, individual components may be replaced, upgraded, or decommissioned from time to time. The system lifetime is likely to far exceed the component lifetime.

Complicating long-term planning, EmNets will have to interface with a wide variety of sensors, network gateways, displays, actuators, power sources, antennas, and other EmNets. This heterogeneity, which is itself a major challenge to designing economical EmNets, is multiplied by the longevity requirement. Good interface standards will play a part in solving hardware interconnectivity, but striking a good compromise among cost, performance, and feature set has always been problematic. Solving

[9] As an example, consider city buses with sensors that can provide information about their location. This information could also be used to turn the buses themselves into sensors for traffic congestion. Such technology is being developed in several localities (see, for example, <http://www.gcn.com/archives/sl/1998/July/1B.htm>).

the analogous problems in the software domain may be even more difficult.

EmNet Size and Scale Are Significant

Networked systems of embedded computers can grow extremely large. It is easy to imagine deploying sensor technology with which one could sense various conditions within buildings or the environment; such networks might embody thousands or tens of thousands of nodes. In fact, building control systems with tens of thousands of nodes already exist.[10] Networking many of these systems would yield systems of millions of nodes.[11] Economics will allow such large systems to be built, and demand will come from many sources, ranging from environmental researchers to government regulators to the general public. Military applications and battlespace EmNets are also inherently large, encompassing millions of nodes in a three-dimensional space anywhere from the seabed to satellites in space.

Scale matters—systems designed to work properly at one size will often fail at a larger (or even a smaller) size. In systems the size of the EmNets being contemplated here, it is very reasonable to expect that many of the networking, software, and hardware solutions known at present will be unsuitable, or even dangerous. EmNets are particularly vulnerable in this regard, because they appear at first glance to be reasonable extrapolations of current technology. The committee fears that they will be built naively in exactly that way and, worse, that they may even appear to work as desired for a time. The ability to predict accurately how complex engineered systems will behave, especially under unusual or boundary conditions, is limited at best. EmNets will stretch the ability to analyze system behavior beyond current capabilities, making it likely that such systems will exhibit emergent, or unexpected, behaviors.[12]

[10] See for example, products made by the Echelon Corporation, <http://www.echelon.com/>.

[11] With just a little more imagination, systems of billions of nodes can be conjured.

[12] Emergent behavior is often described as behavior of a whole that seems more organized and purposeful than that of its component parts. This notion often arises in the context of complex systems, where there are many pieces interacting with one another such that the study of individual pieces in isolation is insufficient to predict the behavior of the entire system (Rapaport, 2000).

EmNet Users Are Not System Experts

EmNets will increasingly be used by people who have little or no systems training. Modern aircraft cockpits have extensive computer-based systems with which the pilot must interface. Even with extensive training, pilots (who are expert users of the systems they operate) make errors a disturbing share of the time.[13] An EmNet that requires extensive user training will have failed in its fundamental promise—computing systems must adapt to users, not the other way around. Yet combining extremely complicated systems with casual or inexperienced users is a potential recipe for disaster. If history is a guide, such users will drive the system into operating conditions that were never considered by the system designers, they will misunderstand what the system is trying to tell them about its own health, and they will put themselves inadvertently at risk by trusting the EmNet when it is no longer trustworthy. An additional complicating factor is that people will less often interact with EmNets per se than with the devices and objects within which EmNet components are embedded. People's expectations of objects in their environment are likely to be very different from their expectations of explicitly computational or communication devices such as PCs or cell phones. The computer industry has a very poor record overall of designing effective user interfaces, much less interfaces that, if misunderstood, can still prevent danger to the users themselves (CSTB, 1997; Laurel and Mountford, 1990; Norman, 1998). Designing for casual interaction (as opposed to explicit use) is arguably an even larger challenge. The change of attitude required of the system designers is profound and infrastructural, and attitudes will need to be quite different from the attitudes that created today's successful networks.

WHY A NEW RESEARCH AGENDA?

This report explores how the characteristics of EmNets demand new kinds of research. It examines the different kinds of applications and configurations in which EmNets may be deployed and identifies technical challenges that have not heretofore been addressed by the research community or resolved in a way that is amenable to EmNets. The report

[13] The software in high-tech avionics systems is extremely complex, and most training programs now concentrate on teaching pilots how to use the automation but not necessarily how the automation works. Existing training material is based on a proceduralized, operational model with little attention to causality or the structure of the underlying system. In fact, there have been suggestions that a limiting factor in aircraft automation design may be the level of complexity a pilot's mind can maintain and readily access (Billings, 1996).

attempts to be as far reaching as possible, identifying research challenges in a broad range of areas. The goal is not to specify particular technologies or solutions that need to be developed but to articulate fundamental, underlying research problems that need to be addressed. The areas identified are therefore candidates for fundamental exploratory research that will try as much to understand the problems as to solve them.

To the extent that EmNets represent a continuation of longstanding progress in IT, it is reasonable to ask why special consideration needs to be paid to the research needs for EmNets. In a broad sense, the potential impact of EmNets themselves is justification for an EmNet-specific national research agenda. But as described previously, EmNets present unique technological challenges as well. Research into developing and understanding these systems is vital, for the reasons outlined below.

As EmNets mature and extend into even more areas of society, research will be needed into ways of thinking about designing systems. One can envision systems that are self-monitoring and self-healing—that is, systems that provide active agents to monitor possible problems (as well as their own health) and take appropriate actions, such as to defend against denial-of-service attacks or attempted injection of malicious code. At the same time, continued advances will be needed in enabling technologies. Research will also be needed (1) to make EmNets easy to construct, (2) to make EmNets self-configuring and adaptive, (3) to ensure their performance and safety, and (4) to make them easy to use. These research areas involve system-level issues that arise from the interconnection of large numbers of long-lived information processing devices managed by users who are likely to be experts in a particular application domain but not necessarily in EmNet technology. These users will need to know not just whether the system is working or has failed, they also need to know how close to its safety margins or how healthy the system is so they can make intelligent decisions on whether to use it or take it offline and repair it. While work has progressed in many of these areas over the past decade, it has not generally occurred in the context of embedded computing. Clearly, a number of familiar topics will need to be reexamined, and new topics will need to be addressed.

The potential benefits of EmNets are accompanied by risks that may be exacerbated by the EmNets' very pervasiveness and by the fact that they may be invisible to most who interact with them. The creation and distribution of vast amounts of information about people creates privacy concerns. As EmNets become increasingly critical to our communication, transportation, power distribution, and health-care infrastructures, failures and security breaches will be increasingly dangerous. By the time EmNets are broadly deployed, it will be too late to call them back easily. Therefore, it is critical that we study these systems now, in order to mitigate the risks as much as possible and maximize the benefits.

As this report documents, the technological research issues that are important to EmNets are not unique in and of themselves. Issues of scalability, adaptation, reliability, safety, and performance have all been faced to some extent by other IT systems and have been addressed by research in the more general computing and information technology arenas. What differentiates EmNets and necessitates a new research agenda is that the solutions that have been worked out in areas for more general computing and information technology systems will not work for EmNets. Existing solutions often make a number of assumptions—among them: that energy is readily available, that there is sufficient computing power to allow various layers of abstraction, that the computational elements are generally in static relationships with respect to the physical world, that bandwidth is not terribly constrained, that the computational elements are expensive and therefore rarely duplicated, and that the computational elements are the entities that need to be identified—that simply do not hold for EmNets. While EmNets have many characteristics that distinguish them from traditional systems, it is very likely that the techniques developed to realize EmNets will have enormous positive impact on the design of traditional systems as well; a key example is techniques for self-configuration (see Chapter 3).

It is important to note that networked systems of embedded computers will be and are being implemented, even without the benefit of additional research. Some of these may actually succeed, and others may appear to have succeeded, at least for a time. However, if the maximum benefits are to be gained from EmNet technology at minimum overall risk, much research is needed. It is extremely important that the research community take the lead in this area if there is to be any hope of significant impact. Once systems are established, it is incredibly difficult to upgrade or update them, as has been the case with PCs and the Internet. Designing and deploying them well initially will probably be more cost-effective in the long term, and if the research community can, in a timely fashion, articulate a notion of what is more correct, efficient, secure, safe, reliable, and so on, companies may well adopt it. Once they are deployed, though, history suggests that it will not be possible to effect significant changes or upgrades. It is therefore critical to start addressing the challenges presented by EmNets. Specific research recommendations are provided throughout the remainder of this report.

WHAT THIS REPORT DOES NOT DO

This report is intended to be broad and comprehensive, but there are several topics it does not, by design, treat in depth. These include sensor and actuator technologies that might be used as elements within an EmNet (especially within a sensor network); ethical and policy issues associated with different applications of EmNets and the use of the data they might collect; particular issues of commercialization and market acceptance; and stand-alone (as opposed to networked) embedded systems. These are all extremely important issues—in fact, each is worth its own separate study—that could not be given full consideration here in light of the charge to the committee.

Advanced Sensors and Actuators

The inexorable march of silicon-based technology is making possible the design and deployment of extremely inexpensive, highly capable, low-power sensors (Saffo, 1997). Advances in MEMS technology have already made it feasible to sense odors, vibration, acceleration, pressure, temperature, and many other physical phenomena in ways that will be extraordinarily useful across a wide range of human endeavors. New sensors for sound, visible light, infrared, and extremely low light, combined with ever faster and cheaper digital signal processors, will make large-scale system sensing practical and commonplace. Likewise, new MEMS-based actuators, such as micromotors, will allow EmNets to affect the world in unprecedented ways. The implications of these improving sensor technologies are profound, and this report explores many of them, but the technology of the sensors themselves is largely outside its scope.

Public Policy Issues

There are few, if any, ethically neutral technologies. Powerful technologies such as computing, especially on the scale addressed in this report, have the potential to be utterly pervasive in people's lives. These technologies will be deployed with the best of intentions, but as with all previous technologies, an array of forces will come to bear on them that can be only partially anticipated. These forces will bring a corresponding array of ethical, legal, and policy issues.

The committee believes that the issues will be profound and important. They will require consideration at all levels during the conception, design, deployment, and use of large EmNets. This report can offer no a priori prescription for the ethical, legal, and policy questions posed by EmNets, so its focus has been purposely restricted to technological issues and implications. However, the policy issues are numerous, important, and evident in many contexts. Privacy may be at much greater risk than at any previous time in history, security is a pressing concern when one's attackers can be physically anywhere, and system reliability will become paramount when these new systems have supplanted previous tried-and-true (and simpler) solutions such as telephones, home security systems, agriculture management, and industrial automation. Other issues that will undoubtedly arise concern intellectual property (to whom does the data collected by EmNets belong?), liability (who is responsible when systems fail?), and the "digital divide" (who will have access to what kinds of systems?). There is also an important sense in which the committee believes the technology will permit the easy accretion of large systems—that is, that smaller, self-contained systems will be combined in an ad hoc manner to create much larger systems. The difficulties of engineering a system that is, by definition, unplanned pale in comparison with grappling with its ethical implications.

The reader should not misconstrue the focus on technology in this report to mean the authors believe the policy implications are trivial or benign. The truth is, the committee believes they deserve far more attention than can be given here if the basic task of exploring the technology itself is also to be fulfilled. Powerful technologies can be used for good or ill (or both). EmNets qualify as powerful technology by any definition. The ethical, legal, and policy issues must be addressed during the design and use stages of these systems. In this report the committee raises these issues when they seem particularly pertinent to the discussion in order to draw attention to some of the far-reaching implications of this technology. However, a more in-depth analysis of public policy issues is urgently needed that would lead to appropriate recommendations for solving likely problems.

Commercialization Issues, Standards, Business Models

Deploying very large numbers of anything is unavoidably an exercise in both technology and economics. The technology must be inexpensive enough for large numbers of people to be able to afford it, yet it must be powerful enough to solve some need. And ultimately, there must be enough profit in the venture for the purveyor of the technology to develop products and support them. It is by no means a given that the best technology will prevail, and if there is no economic benefit (or too high a perceived risk, particularly of consequential damages), no vendors may wish to participate. For the purposes of this report, the committee assumed that the technology will be associated with large markets but that part of the research and development challenge may relate to lowering costs for a given level of performance or quality. One area of uncertainty about EmNet markets relates to instances where an EmNet may have a broad public benefit that cannot be easily captured by one or more vendors. Sensors that collect data on individual exposure to toxins whose aggregation could identify the source of the pollution and its distribution patterns are an example of an application with primarily public benefit, and as in other instances of environmental technology deployment, the chief customer (or motivator of purchases by others) may be one or more governmental units. The environment, which is an area where there is an understanding of the economics and a government framework in place, may embrace relevant EmNets as it has embraced other technologies. For public-benefit EmNets that constitute new applications domains, the way forward may be less clear and market development more uncertain. By contrast, for EmNets with inherent commercial value (such as smart office buildings), the committee expects significant markets to develop.

Standards are expected to be important for EmNets because of the fundamental concern about interoperability and the variety of other kinds of interfaces. A dominant producer—and, like other products, most IT products seem to have a small number of major producers once their markets mature—may drive a de facto standard. Alternatively, various groups—industry groups concerned with specific enabling technologies, applications domains that may work through trade associations or focused consortia, or groups such as those convened under the auspices of the National Institute of Standards and Technology (NIST) or even the Internet Engineering Task Force (IETF)—may work to develop standards that may or may not be open. However, it is not a purpose of this report to attempt to identify such standards.

Stand-alone Embedded Systems and Other Networked Information Systems

This report emphasizes the characteristics of EmNets that stem from the embedded, physically coupled aspects of the nodes in combination with the networked aspects of these systems. There are still many research challenges for stand-alone embedded systems, and indeed any progress there will have an important impact on networked embedded systems. Networking allows innumerable new kinds of interactions. It also provides an ability to coordinate across multiple, heterogeneous devices and make use of information gathered by geographically distant actuation devices. In this report, the committee focuses explicitly on networked systems of embedded computing devices, while acknowledging that many of the issues that arise with stand-alone systems will be relevant in the networked arena as well.

While the research recommendations and discussion in this report can and should be seen as part of a larger networking research agenda, the emphasis here is on EmNets that are purposefully built to perform specific sets of tasks, as opposed to ad hoc interconnections of PDAs and laptops for general-purpose application support. Large-scale societal IT systems, such as financial systems, are not included. These systems are engineered, like EmNets, and they have processors and networking capabilities embedded in the fabric of their operation. They are not considered in this study because the computing elements are generally not embedded in devices that have an apparent purpose other than computing and communications. Cellular telephone systems are a particularly interesting case for definitional purposes. They are clearly engineered systems, and they clearly involve embedded processors. They are also, by their very nature, networked, power-constrained, and mobile—as the cell phone moves around in the physical world, real-time handoffs are made between the various transceiver towers so as to keep the user continuously connected to a given phone call. Cellular telephony can provide a number of valuable lessons for the design and operation of EmNets, but there are also circumstances specific to cell phones that the committee believes will cause some of its solutions to be inapplicable to the kinds of EmNets anticipated here. This report tries to carefully distinguish the aspects of cell phone technology that are relevant to EmNets from those that are not.

ORGANIZATION OF THIS REPORT

The remainder of this report elaborates on the themes introduced in this chapter. The report can be read as a progression from very concrete issues involving component technologies such as chips and wireless communications all the way to the abstract computational models that will be used to reason about these systems. Chapter 2 examines several enabling technologies without which EmNets as they are described here would not as easily or as flexibly come to pass. It discusses component technologies used to construct EmNets. Readers who are interested in learning about the larger systems issues related to EmNets should feel free to move directly into Chapter 3, which explores self-configuration and adaptive coordination as these concepts pertain to EmNets and how EmNets organize themselves and respond to changes within the environment and the system. In other words, Chapter 3 examines how the component technologies in Chapter 2 should be arranged to form an EmNet and what kinds of technologies will be needed to achieve this. Chapter 4 moves up another level and examines the features that EmNets will need to have. It explores trustworthiness of EmNets, including the issues of safety, reliability, security, privacy, and usability. Chapter 5 examines the need for better kinds of abstractions and computational models to describe and analyze EmNets that incorporate the features described previously. Finally, Chapter 6 considers the current research infrastructure and how it could be adjusted to better address the challenges that EmNets present.

It outlines several broad areas in which research is needed and makes recommendations to various federal funding agencies.

REFERENCES

Billings, Charles E. 1996. *Aviation Automation: The Search for a Human-Centered Approach.* Mahwah, N.J.: Erlbaum.

Board on Agriculture and Natural Resources (BANR), National Research Council. 1998. *Precision Agriculture in the 21st Century: Geospatial and Information Technologies in Crop Management.* Washington, D.C.: National Academy Press.

Computer Science and Telecommunications Board (CSTB), National Research Council. 1997. *More Than Screen Deep: Toward Every-Citizen Interfaces to the Nation's Information Infrastructure.* Washington, D.C.: National Academy Press.

CSTB, National Research Council. 1999*. Trust in Cyberspace.* Washington, D.C.: National Academy Press.

CSTB, National Research Council. 2000. *Making IT Better: Expanding Information Technology Research to Meet Society's Needs*. Washington, D.C.: National Academy Press.

Joy, Bill. 2000. "Why the future doesn't need us." *Wired*, 8.04. Available online at <http://www.wired.com/wired/archive/8.04/joy.html>.

Laurel, Brenda, and S. Joy Mountford, eds. 1990. *Art of Human-Computer Interface Design.* New York, N.Y.: Addison-Wesley.

Li, Y., and R. Wang. 2000. "Precision agriculture: Smart farm stations." IEEE 802 plenary meeting tutorials, document no. 00362r0P802-15_LRSG-Precision-Agriculture-Smart-Farm-Stations.ppt.

National Institute of Standards and Technology (NIST). 1999. *Testing and Standards for Pervasive Computing.* Gaithersburg, Md.: Information Technology Laboratory, NIST.

Norman, Donald. 1998. *The Invisible Computer*. Cambridge, Mass.: MIT Press.

Rapaport, D.C. 2000. *Computer Simulation Studies in Condensed Matter Physics*. Volume XIII, D.P. Landau et al., eds. New York: Springer-Verlag.

Saffo, Paul. 1997. "Sensors: The next wave of infotech innovation." *1997 Ten-Year Forecast.* Menlo Park, Calif.: Institute for the Future.

Weiser, Mark. 1991. "The computer for the 21st century." *Scientific American* (September): 94-104.

Weiser, Mark. 1993. "Some computer science issues in ubiquitous computing." *Communications of the ACM* 36(7):75-83.