Submission to the CRA Conference on “Grand Research Challenges” in Computer Science and Engineering

June 23–26, 2002 Airlie House, Warrenton, Virginia

Allocating Resources for Millions of Competing and Cooperating Computations

Gaetano Borriello
Department of Computer Science & Engineering
University of Washington
Seattle, WA

Computing is rapidly progressing toward a world of ubiquitous computing devices embedded into virtually every manufactured object. The most numerous of these devices will be wireless sensors that harvest power from their environment. Many will be actuation devices that manipulate the physical world and whose effects will be sensed by the sensor nodes. There will be devices carried by people that will contain their autonomous agents and data stores. Finally, there will be myriad I/O devices both carried and stationary in the environment that will focus on interacting with people through a wide variety of visual, audio, and haptic interfaces.

A “grand challenge” for this world to be realized is resource allocation. There will be dozens of devices per cubic meter, thousands in a home or building, and millions on a large campus. How will these devices coexist? What resources will they be competing for? How can we guarantee an application the resources it needs to carry out its functions? How will such collections of devices evolve over time? What security model will be used to ensure that different uses of the same devices don’t interfere with each other?

The resources in question include communication bandwidth, computation cycles, data storage, as well as, most importantly, user attention. Communication will need to become adaptive in time and space: time so as to minimize power consumption by lowering duty-cycle; and space (that is, range) so that high-densities of devices can be sustained on a temporary or permanent basis. Computation will have to be able to adapt to available power and communication resources by migrating to where it can get its job done most efficiently. This will include splitting computations so that they can be partially done in the sensors (e.g., aggregation of data to minimize transmissions), in the actuators (e.g., feedback control loops for specific effects that meet real-time constraints), and in wired and wireless devices that can provide more plentiful resources to host agents, services, and I/O interactions. Data will have to be able to find its own way through heterogeneous networks of devices and replicate appropriately so as to ensure it will be found in a timely manner by the applications that need it. Finally, all of this has to happen autonomously, proactively, and invisibly to users. There is no way we can expect people to want to know or care for millions of devices. The focus needs to be on providing coherent applications (made possible by evolving collections of devices) that interact with people at levels of abstraction appropriate to the tasks with which they are assisting.

To further complicate matters, all this adaptivity has to happen in a way that is secure both in terms of authorization as well as preventing monopolization of resources. Devices will be in the infrastructure, literally in the concrete, carried by people as personal devices, made available as public resources for all to use within certain guidelines; and owned by entities that will provide them for specific functions in their private spaces. There will no longer be the luxury of deploying new protocols and new run-time environments on a large scale. The sheer number of devices coupled with their being embedded deeply into all objects will mean that they must be made to adapt over time so that evolution can happen incrementally and not in convulsive mass replacements. What applications, on whose behalf, will have the ability to garner communication, computation, and storage resources and where will be a major challenge. Devices and applications will have to negotiate contracts of use autonomously and have methods for enforcing the agreements. Applications will have to be prioritized in a way that will be comprehensible to mobile users entering and leaving widely differing spaces with widely different available resources.

These challenges require us to rethink our basic assumptions in distributed systems. Basic tenets of the field will be challenged including: transparency of local and remote resources, synchronous communication, end-to-end protocol design, etc. We will need to start thinking of our systems as complex organism with intricate communication patterns that govern their overall behavior. We will need to tackle the problems of emerging behavior in systems consisting of millions of autonomous computations and will most likely need to borrow many concepts and principles from biological systems (that offer the only model at the same level of complexity).

The grand challenge will be adaptively allocating resources to mobile computations acting on the behalf of millions of users exploiting even more millions of heterogeneous devices. We will need to solve this problem while providing users with the high level abstractions that will let them govern the overall behavior of their own elements (computations and devices) in a scalable way that will allow people to focus on their tasks and needs.

Biography for Gaetano Borriello

Gaetano Borriello is a faculty member of the University of Washington's Department of Computer Science and Engineering. He is currently on a two-year leave to establish a new research center for Intel Corporation adjacent to the University of Washington campus. His research at UW and the mission of the new laboratory are closely aligned, both focusing on solutions that will make embedded computing devices distraction-free by making them highly decoupled and reconfigurable without human intervention. Dr. Borriello will return to his position at the University of Washington upon completing a 2-year assignment with Intel.

Gaetano Borriello received his BS degree in Electrical Engineering from the Polytechnic Institute of New York (1979), his MS degree in Electrical Engineering from Stanford University (1981), and a Ph.D. in Computer Science from the University of California at Berkeley (1988). He also spent four years at the Xerox Palo Alto Research Center from 1980-84 where he developed the first completely integrated Ethernet controller. He joined the UW faculty in 1988 where he served as Associate Chair from 1998 to 2000. He was awarded an NSF Presidential Young Investigator Award in 1988 and a UW Distinguished Teaching Award in 1995.

His research interests are in the design, development, and deployment of computing systems with particular emphasis on mobile and ubiquitous devices and their application. He has a wide range of interests that can be classified in embedded system design, development environments, user interfaces, and networking. They are unified by the goal of making new computing and communication devices that make life simpler by being as invisible as possible to their owners, being highly specialized and thus highly efficient for the task at hand, and able to exploit their connections to each other and the greater world-wide networks.

His most recent research accomplishments include the development of the Chinook design system for heterogeneous distributed embedded processors which has led to the formation of a company, Consystant Design Technologies, to commercialize the ideas. He has served on numerous conference program committees and will be program chair of the 2002 edition of the Ubiquitous Computing Conference in Goteborg, Sweden. Recently, he participated in a study commissioned by the Computer Science Technology Board of the National Research Council; titled "Embedded, Everywhere" it outlines a research agenda for Networked Systems of Embedded Computers (Em-nets).