

ACM Computing Surveys 28(4es), December 1996, <http://www.acm.org/pubs/citations/journals/surveys/1996-28-4es/a215-pasquale/>. Copyright © 1996 by the Association for Computing Machinery, Inc. See the [permissions statement](#) below. This article derives from a position statement prepared for the [Workshop on Strategic Directions in Computing Research](#).

Towards Internet Computing

Joseph Pasquale

Dept. of Computer Science and Engineering, U. C. San Diego
La Jolla, CA 92093-0114, USA
pasquale@cs.ucsd.edu, <http://www-cse.ucsd.edu/~pasquale/>

Abstract: Internet computing is highly network-integrated computing on existing networks of computers, done on the grand scale of the Internet. Supporting Internet computing will require advances in network communications architectures. I outline three research directions that I believe are critical: resource sharing based on market-economic principles; communication of agents (i.e. where the basic type of message is a program and not simply data); and provisions for ranges of qualities of service.

Categories and Subject Descriptors: C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design - *network communications, store and forward networks*; C.2.4 [**Computer-Communication Networks**]: Distributed Systems - *network operating systems*; D.4.4 [**Operating Systems**]: Communications Management - *network communication*; D.4.6 [**Operating Systems**]: Security and Protection - *authentication*; D.4.7 [**Operating Systems**]: Organization and Design - *distributed systems*;

General Terms: Design, Management, Performance.

Additional Key Words and Phrases: market-based resource allocation, agent-based computing, quality of service.

1 Internet Computing

Today's large networks offer, for the most part, a simple message-based communication service between computing nodes. I would like to briefly put forth what I think is needed to create much larger-scale and more highly network-integrated computing environments out of existing networks of computers, such as the Internet. The promise of a "network computing" system, an idea that is certainly not new but is now gaining popularity over more traditional ones like high-performance multiprocessors and supercomputers, is increased scalable cumulative power, more efficient (and economical) use of existing resources, and more effective system management. Ultimately, large economies of scale are achievable by the ability to

arbitrarily deliver and redirect the shared power of the network to any user, in contrast to today's systems where power is essentially statically distributed and preallocated, and not shared. On the scale of the Internet, we call this *Internet Computing*.

Much work has been done in network operating systems and distributed computing that support the idea of Internet Computing. However, to realize this idea on a truly grand scale, I see three areas of networking research as being pivotal.

2 Resource-sharing based on market-economic principles

The prospect of harnessing a significant fraction of a network to execute a distributed application is tantalizing, but how would these resources be made available? They will generally be owned by different organizations that will not necessarily volunteer to make them freely available, even when they are not being used. To provide the incentives for large-scale resource sharing, resources need to be buyable and sellable, with the possibility of contracts that (at least in theory) can be enforced. This requires flexible communication protocols for establishing contracts. A set of protocols are needed that range from ones that are very lightweight, for small "micro-value" transactions, to those that are more heavyweight, where enforceability (perhaps at a later time) is more important than efficiency.

This will support the creation of a market economy of network-accessible resources, including those that are specifically part of the network such as link bandwidth and switch/router processing and memory capacities. These resources would be purchased on demand, in very short periods of time, and for very short (and larger when necessary) periods of time. Users will rely more on buying the power in the network, only when and for as long as they need it, rather than solely on what is locally available. This will be more economical and offer the user more potential power.

3 Communication (and remote execution) of foreign software agents

Today's most popular Internet application, browsing the World Wide Web, is based on the computing paradigm whereby data (and more recently, programs) are retrieved from remote sites and processed (executed) locally. However, this model assumes a significant amount of local computing and storage power, and lots of bandwidth for moving large data items. If processing, memory, and storage can be purchased remotely, it makes more sense to run the computations near the data, alleviating the user's requirement for local power and getting better performance by reducing communication time and costs. In other words, I am arguing that buying remote compute (and memory and storage) resources, on demand, will be cheaper than buying local power as well as the bandwidth needed to obtain data objects. This makes even more sense when you consider that in the future, the typical scenario will be that what the user has locally is some small cheap low-power wireless computing/communication device, e.g. telephone or personal digital assistant, and so retrieving and absorbing large data objects and processing them locally is the wrong model.

The alternative model of doing the computation remotely, ideally near the data, requires a more flexible communication (and computation) architecture, e.g. one that supports the dissemination of "agents". An agent is simply a software module that (1) is executable, (2) is mobile (over computing nodes of a network), (3) can maintain state across movements, and (4) can be authenticated (as to who it is an agent for). The last requirement is especially important since an agent must have purchasing power to be a player in a market

economy as described above, and will carry digital certificates representing funds.

In this communication model, the basic type of message is not (simple) data but a program, i.e. an agent programmed to buy computational (and communication) resources as it moves about. I expect that applications will often involve the dissemination of many agents as part of a distributed computation. A simple but illustrative example is a remote video browser that generates two agents. One agent runs at the remote video server, buying the necessary processing, memory, and I/O resources at the server to (1) obtain and send the video frames to the user, and (2) support adaptivity to trade off bandwidth and processing. Since network communication bandwidth may not be guaranteed, the remote agent will work with the other agent, which runs locally, to insure smooth-motion video despite fluctuations in bandwidth and delay by adapting the frame rate, the level of compression, the quality of resolution, the amount of buffering, or some other parameter to control performance. The client simply receives frames from the local agent and displays them. This is an example of two disseminated agents that negotiate an application-specific end-to-end function on behalf of the client. One might also imagine a set of agents running on a set of nodes along the path from server to client, implementing some type of negotiated application-specific hop-by-hop function (e.g. distributed phased filtering).

4 Providing a range of qualities of service

A market economy will not work if buyers don't know the quality of what they are buying or sellers can't insure the quality of what they are selling. In networking, one would like various types of assurances regarding bandwidth, (end-to-end) delay, etc., and be willing to pay based on the strength of the assurance. These assurances may range from weak hints to strong guarantees. In some cases, like mobile devices in a wireless network environment, weak assurances may be the best one can expect.

This points to two research directions: (1) how to provide a range of qualities of service, including those that are at least predictable (but not necessarily guaranteed), and (2) how to provide adaptivity tools for applications that require more than best-effort services in environments that support services whose qualities are less than guaranteed. Given that control over resources, including those pertaining to the network, will be decentralized, it is clear that one cannot hope for a world of purely guaranteed service as has been the case for telephone networks; indeed, this will not be economical, and will in many cases be unnecessary. And yet, there is certainly a place for guarantees where they can be provided, especially in a market economy where they are worth implementing despite their high cost if there is a market.

5 Conclusions

Supporting Internet Computing will require advances in network communications architectures. I have outlined three research directions that I believe are critical: resource sharing based on market-economic principles; communication of agents (i.e., where the basic type of message is a program and not simply data); and provisions for ranges of qualities of service. In fact, these directions have dimensions other than just that of communication (which is not surprising, given the trend toward higher levels of integration between computation, communication, and storage). For example, operating systems must provide finer control over physical resources if stronger assurances for quality of service are required; they must also provide more controlled execution environments if foreign agents are to be allowed to execute while limiting their potential damage. At the application level, new methods are required for building programs that are more adaptive to changes in underlying resources during their executions. Advances in these areas

in any one dimension should provide insight on how to achieve the same advances in the other dimensions.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Publications Dept, ACM Inc., fax +1 (212) 869-0481, or permissions@acm.org.

Last modified: Mon Feb 24 14:36:44 EST 1997

Joseph Pasquale <pasquale@cse.ucsd.edu>