Panel:
The Role of End-to-End Quality of Service in Distributed Computing

Chair:
Peter Steenkiste
Department of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Panelists:
Dilip Kandlur, IBM T.J. Watson Research Center
Gurudatta Parulkar, Washington University
Andreas Polze, Humboldt University of Berlin
John Zinky, BBN Systems and Technologies

Real-time operating systems that can guarantee the availability of resources (CPU cycles, memory, ..) to applications with real-time guarantees have been an active area of research for a long time, and more recently, the development of “integrated services networks” that can provide a wide range of services, including services with hard guarantees, has been a significant research effort. Combining results from these two areas can provide distributed applications with end-to-end quality of service guarantees. Guaranteed services hold great promise for distributed computing, since they may make it possible to develop distributed applications that combine resources connected by (shared) networks, while maintaining high efficiency and good response times. Moreover, by dedicating a set of resources to an application, guaranteed services may greatly simplify the develop of distributed applications over a shared set of networks and endpoint resources.

However, many technical challenges have to be overcome before guaranteed end-to-end services are a widely available and effective tool for distributed computing applications. For example, while many people have looked at applying service guarantees to continuous media applications or periodic application with strict deadlines, there is much less experience in the use of guaranteed services in the more general area of distributed computing. More research is needed to identify appropriate service definitions and programming models. Once defined, these services and models have to be implemented in terms of the services supported for individual system elements. It is also useful to look beyond services that provide hard guarantees. Relying heavily on guaranteed services may result in low resource utilization since providing strong guarantees often involves overallocating resources. This
may make guaranteed services expensive and thus unattractive. This raises the issue whether service
classes can be defined that allow more efficient resource utilization while providing a useful service to
applications. Finally, there is an increasing interest in “network-aware” or “system-aware” applications,
i.e. applications that can adapt in intelligent ways to the status of the network or the system in general.
The availability of different service classes raises the question whether it is possible to define and
implement services that support this class of applications effectively.

The goal of the panel is to gain insight into what type of end-to-end services will be of value to
distributed computing applications and to identify what technology is missing if we want to support
these services efficiently. Panelists will review relevant techniques in the areas of end-point resource
management, network services, and middleware, and will then address a number of open research prob-
lems. Questions include: what are appropriate service models for distributed computing applications?
Do the current real-time operating systems and integrated services networks adequately address the
needs of distributed computing applications? What are the cost/benefit tradeoffs in using guaranteed
services instead of existing "best effort" services?