Resource-aware Mobile Programs

Anurag Acharya, M.Ranganathan, Shamik Sharma and Joel Saltz
Department of Computer Science
University of Maryland, College Park 20742
{acha,ranga,shamik,saltz}@cs.umd.edu

Mobile programs are expected to operate in an environment where the availability of resources changes dynamically and is not under the control of the user. For example, network characteristics change based on traffic; server availability changes based on usage and administrative decisions. To be effective and efficient in such an environment, mobile programs need to be able to monitor the level and quality of resources available and to adapt to significant changes.

As a part of the "Green" project (on resource-aware programs), we are seeking answers to the following four questions:

- How rapidly does the environment change and how stable the changes are? Also, by how much does the environment have to change for it to significantly impact the operation of a mobile program?
- What is the cost of monitoring the environment?
- What is the benefit of adapting to changes in the environment?
- How should mobile programs be structured to be able to adapt to changes in their environment?

We are addressing these questions in an empirical manner - by building and studying resource-aware mobile programs. We have developed an infrastructure for developing adaptive mobile programs as well as a distributed resource monitor. We are currently in the process of implementing and studying a set of adaptive mobile programs.

Based on early experience with a network-latency-sensitive application and an extensive study of end-to-end Internet latency, we have developed three working hypotheses. First, a resource-aware placement of components of a distributed program can provide significant performance gains over a resource-oblivious placement. Second, effective mobility decisions can be based on coarse-grained monitoring. Finally, a simple and cheap distributed resource monitoring scheme can provide sufficient information for effective mobility decisions. We found that latency-sensitive programs that run for at most an hour or so gain little from continuous monitoring and that most of the gains for these programs can be achieved by on-demand determination of resource levels. On-demand monitoring would also be suitable for resources that are more expensive to monitor, for example, network bandwidth.

We are taking two directions in our current research. First, we are working on developing cheap-to-compute measures for other resources, in particular, network bandwidth and server availability. Second, we are developing several other mobile programs. We are presently working on a bandwidth-aware multi-site data-combination program. This program monitors the network bandwidth between several data-source sites and dynamically moves data-combination sites in response to significant changes in bandwidth. Preliminary experience indicates that a bandwidth-aware placement can provide significant performance gains over a bandwidth-oblivious placement. We are also in the process of conducting a study of end-to-end Internet bandwidth between a set of sites around the world. Our other target applications include custom combination of periodically generated large-volume datasets (such as weather information) and resource-aware pre-fetching for web clients.