Abstract:
The growing traffic of WWW related services requires the development of efficient protocols for reducing traffic, balancing load, and improving service time. One way of achieving these effects is via caching or replication. Studies like [Bestavros et al, 1995] show that simple demand-driven caching is not enough, and that aggressive caching policies have to be adopted. One such policy is prefetching.
In an earlier paper, the potential of speculation (server-initiated prefetching) in distributed information systems (such as the WWW) was investigated and shown to be effective in reducing service time and server load. This speculation was based on statistical information that the server maintains for each document it serves. In this paper we study the performance of a client-initiated prefetching protocol, whereby speculation is based on past user-specific access patterns. We propose a technique whereby the history of a user is analyzed to predict his/her future accesses. Our technique does not make a distinction between embed links and traversal links. In particular, embedded links are treated as a special case of traversal links with the probability of traversal being 1. We show that performance gains are possible to obtain by identifying common access patterns. Our study was conducted using client-based traces obtained from our departmental labs over a period of 100 days [Cunha et al, 1995].