NORTHWESTERN
UNIVERSITY

Computer Science Lecture Series

Efficient and Adaptive Replication Using Content Clustering

Mr. Yan Chen
University of California at Berkeley

12:30-1:30pm
Wednesday, March 12, 2003
Room 381 - Computer Science Dept.

Abstract:

With the astounding growth of the World Wide Web, content distribution networks (CDNs) emerged to improve content delivery performance by replicating content to multiple locations in the Internet and having users get data from the nearby data repository. However, existing CDNs have three problems:
a) inefficient replication wastes bandwidth and storage, due to lack of knowledge on replica locations;
b) coherence of replicas can not be supported; c) network monitoring is ad hoc and unscalable.

To address these challenges, I propose an innovative peer-to-peer CDN infrastructure, SCAN (Scalable Content Access Network). First, given millions of objects, SCAN replicates them in clusters to build replica directories for replica sharing. Secondly, it self-organizes the replicas into an application-level multicast tree for disseminating updates. Finally, it uses a scalable overlay network monitoring system for network distance, congestion and failure estimation. In this talk, I will focus on the first feature.

With replica directories, SCAN achieves similar performance with only 4 - 5% of replication and update traffic compared to the conventional CDNs. To reduce the computation and management overhead, several types of Web content locality for clustering are explored. Simulations using various topologies and several large Web server traces show that clustering-based replication reduces the overhead by two orders of magnitude with little impact on performance. Further, to adapt to dynamic users' access patterns, I propose online incremental clustering that adaptively adds new documents (even before being accessed) to the existing content clusters. This technique is especially useful in improving document availability during flash crowds.

For a complete calendar, see: www.cs.northwestern.edu; click on ‘CS Seminars’	**Join our e-mail list** for notices of upcoming Computer Science lectures: send an e-mail with the word ‘subscribe’ in the subject line to: cs_seminar@cs.northwestern.edu