CCIT Report #377 May 2002

Optimal content location in overlay networks with updates

Israel Cidon
Department of Electrical Engineering
Technion - Israel Institute of Technology
Email: cidon@ee.technion.ac.il

Oren Unger
Department of Electrical Engineering, Technion
and Zoran Microelectronics
Email: unger@tx.technion.ac.il

Abstract

A major challenge for organizations and application service providers (ASP) is to provide high quality services over the network to geographically dispersed consumers at a reasonable cost. As the performance of the Global Internet and Corporate Intranets is unpredictable, such providers employ content delivery networks (CDN) and overlay networks to bring content and applications closer to their service consumers. The overlay network is composed of a set of distributed servers which serve local consumers and are updated and synchronized with the most recent information.

The architecture of overlay networks should encourage high-performance, high-scalability and reduced costs. This becomes more crucial as despite technological advances, communication, storage costs as well as service latencies grow with the exploding amounts of data exchanged and with the size and span of the overlay network. Examples for objects which incur high traffic, storage and update costs are interactive media objects as well as large design materials and software repositories. For that end, multicast methodologies can be used to deliver content from regional servers to end users, as well as for the timely and economical synchronization of content among the distributed servers. Another important architectural problem is the efficient allocation of objects to servers to minimize storage, delivery and update costs while maintaining the expected quality of service.

In this work, we suggest an IP multicast based architecture and address the optimal allocation and replication of objects that are both consumed and updated. Our model network includes consumers which are served using IP multicast or IP unicast transmissions and media sources (that may be also consumers) which update the objects within the servers using multicast communication. General costs are associated with distribution (download) and update traffic as well as the storage of objects in the servers. Optimal object allocation algorithms for tree networks are presented with complexities of $O(N)$ and $O(N^2)$ in case of multicast and unicast distribution respectively.

To our knowledge, the model of multicast distribution combined with multicast updates has not been analytically dealt before, despite its popularity in the industry.

I. Introduction

Recent years have witnessed tremendous activity and development in the area of content and services distribution. Geographically dispersed consumers and organizations demand higher throughput and lower response time for accessing distributed content, outsourced applications and managed services. In order to enable high quality and reliable end-user services despite unpredictable Internet and Intranet conditions, organization and applications service providers (ASPs) employ content delivery networks (CDN) and overlay networks. These networks bring content and applications closer to their consumers, overcoming slow backbone paths, network congestions and physical latencies. Multiple vendors such as Cisco[1], Akamai[2] and Digital Fountain[3] offer CDN services and overlay technologies. Recently, more collaborative models such as distributed storage and peer-to-peer computational models require both consumption and modification of the content by multiple, geographically distributed users[5][6].

Naturally, organizations and ASPs try to optimize the overall cost of the overlay network mainly in terms of storage and communication costs. Efficient allocation of information objects to the overlay network servers reduces the operational cost and improves the overall performance. This becomes more crucial as the scale of services extend to a large number of users over international operation where communication and storage costs as well as network latencies are high. The optimization problem becomes more difficult as the service becomes dynamic and needs to be changed, updated and synchronized frequently.

Our model network includes consumers, which are served from servers using IP unicast or IP multicast communication. The network also includes media sources (that may also be consumers) that update and modify the objects within