The Transcoders' Placement Problem Over Multicast Networks

Thesis Proposal by : Asaf Henig.
Technion - Israel Institute of Technology
Department of Computer Science

Supervisor: Dr. Danny Raz
Department of Computer Science

June 14, 2004

1 Introduction

One of the important functions of the Internet is the capability to deliver live audio and video streams to users at high quality, fast speed and most important - low cost. Delivering live streams across the Internet is becoming more and more popular as computers gradually replace other media and entertainment channels such as radio, television, newspapers etc.
Newly developed standards and encoding technologies, such as AVI and MPEG, have made it possible for users to receive television-quality video with various data rates and various bandwidth consumption. In today's heterogenous environment, end-users use different devices such as laptops, PDAs, smart phones and others. In this environment, sending customers media objects forces the sender to supply the data in different formats to be properly accommodated by the end users' devices. It is reasonable to assume that the multimedia information was not initially encoded in all formats, therefor, in order to suit the clients' devices in terms of CPU availability, screen resolution and others, the multimedia data should somehow be transcoded.

Transcoding is defined as a transformation that converts a media object from one form to another, frequently trading off object fidelity for size [11]. It can be executed at various components in the network such as servers, proxies, and clients. At each component, the cost of the format transformation is different. Placing transcoders at the end-users will preserve the network architecture and will transfer responsibility of the format transformation to the clients. However, this option can be very expensive due to limitation of connection bandwidth and computational power. Putting the responsibility for reformatting at the server's hands can also be an expensive option causing the server to store all formats to each file and wait for clients' requests. If the server chooses to transcode the files only when clients' requests are issued, this can cause a considerable delay in delivering time. We should also remember that servers are usually one of the networks' bottlenecks, especially when media delivering is considered, thus loading them with expensive computational tasks before serving each request will only increase their instability.
The third option is to transcode the media objects at intermediate nodes in the network. This approach was explored extensively in the past [6, 7], under the assumption that intermediate nodes (mostly proxies and even routers in the future) have the capability to transcode requested media objects according to the end-users specifications. Some papers [3, 13] suggested the addition of a "negotiation phase" which is used as a preliminary step before sending any media object. In this phase the source, who wants to send the data, learns about the capabilities of the different clients and decides which format transformations can be used to serve each client. Each one of the options described above regarding where to preform the needed transcoding has its own advantages and disadvantages. In order to be able to decide where to perform the transcoding, it is wise to look at this problem as a simple question of cost.

The cost of sending a media object to users is composed of mainly two elements. First, the transcoding operation, meaning the transformation of formats aiming at supplying the user with his (her) compatible format. Second, the communication cost of transmitting the streams over the network. The cost of the transcoding operation itself is a function of three elements, the two formats that the transformation is dealing with (the origin format and the produced format), and the network node in which the transcoding occurs. Choosing the best network node to preform the transcoding is crucial since the network is very heterogenous even when considering only intermediate nodes. Not all intermediate nodes have the same computational power, thus it is reasonable to assume that a specific transformation would take longer time at a certain node compared to other nodes, meaning, it will cost more.

In general, any data can be delivered from servers to clients in two ways: unicast and multicast. In unicast data is delivered from a specific source to a single destination. Multicast, on the other hand, means that the data is sent from a well defined source to a group of end-users designated by a common multicast group destination. When a certain media object should be delivered to several end-users at a certain point in time, the above mentioned schemes are considered. It can either be delivered via unicast sessions to each end-user separately, or it can be delivered to all clients simultaneously using a multicast connection. Sending the same data to a group of clients using several unicast sessions involves a much simpler communication protocol, but of-course increases the overall delivering time and cost. It also produces a vast workload on the server sending the data. For these reasons, and others, when considering the cost of sending the same data to a group of designated clients, usually sending the data via a multicast session is the preferable choice. The idea of using multicast for media delivery was extensively researched in the past [14, 12].

This thesis aims at studying ways to combine multicasting of media streams along with the ability to transcode the media formats to suit all clients' requirements.

2 Related Works

Efficient proxy placement on the path between the server and its designated clients was studied extensively in the past for unicast sessions [11, 10, 2]. This question was further examined in the specific case of transcoding proxies [3]. In this case, the cost of the communication before and after the transcoding is different. The case of proxy placement for multicast sessions was hardly investigated and to the best of our knowledge the only paper dealing with this case in the context of "Overlay Networks" is [1].

An overlay network is a virtual network designed above the existing internet. The components of such a network interacts mostly with other elements of that network and ignore issues of the infrastructure beneath them. As a result, when the overlay network is sparse, the simulation of real multicast session is bogus, since it is becoming a mixture of long unicast sessions, and thus it loses most of the advantages of a real multicast connection.

The general issue of proxy placements is related to the well known group of theoretical problems called Facility Location. Problems in this group deal with efficiently placing of facilities (in a plane or in a graph) in a way that will best serve potential clients of those facilities. The group of facility location problems includes some of the most well studied optimization problems such as the "k-median" and the "facility location problem" [15, 4]. The problem studied in this thesis, can be seen as another instance of the facility location group of problems, thus, it will contain characteristics similar to other problems in that group.

Another theoretical problem that has a vast impact on this thesis is the "Steiner Tree

Problem". As explained, the use of multicast connection minimizes the cost of sending the same data to a group of users. This minimization can be accomplished by sending the data via a tree with a minimum cost edges which connects the data sender to all the designated clients. The question of how to build such a tree is the Steiner Tree problem. This problem has many variants which differ in both the exact definition of the tree's cost and the subset of clients we wish to connect to the tree, and was investigated thoroughly for many of these variants [8, 9].

3 Thesis Objctives

In this thesis we concentrate on multicasting media objects from a server to a group of recipients over a network. Each node in the network has a multicast capability, can host a transcoder and can perform transcoding operations. Our goal is to explore the question of placing transcoders in the network in a way that allows users to receive the multimedia data in the required format, at the lowest possible cost. We name this problem the Transcoders' Placement over Multicast networks, or in short, the TPM problem. The TPM is an optimization problem in which we wish to minimize the overall cost of delivering the data from the source to all its clients. The cost of the data transmitting includes both the transcoding cost and the communication cost, as explained before.

Note that in order to minimize the cost of delivering the media in different formats, the routs that are most likely to be chosen can be different from those that would have been chosen by simple layer 3 multicast protocols. It is also possible that minimizing this cost will involve several formatting procedures between a source and a designated client. For example, data originally sent in one format, can be first transcoded to a second format in order to reduce the amount of bits sent over the net. Eventually, before it is received by the client it can be transcode to a third format more compatible to the client's devices' capabilities.

It can be shown that the TPM problem is NP-hard and thus we cannot hope to achieve an optimal solution to it in a reasonable time. Moreover, our initial results, using the work of Fiege [5], indicate that the this problem cannot be approximated better than *ln(n)* where n is the number of designated clients.

In addition to the theoretical study of the TPM problem and its approximation results, this research will include the development of heuristics and running of simulations to test their performance. The heuristics we intend to present will be based on the source ability to learn about its clients devices' capabilities (i.e. which media format best suit each client) and about the capabilities of the intermediate nodes that can host transcoders. The above mentioned information can be gained by a preliminary "negotiation" procedure as we mentioned before. This procedure will supply the source with enough data to calculate (using our hueristics) both the correct placement of the transcoders and the preferable format that will be received by each client.

References

[1] K. Andreev, B. M. Maggs, A. Meyerson, and R. K. Sitaraman. Designing overlay multicast networks for streaming. *ACM Symposium on Parallel Algorithms and*

Architectures archive Proceedings of the fifteenth annual ACM symposium on Parallel algorithms and architectures, pages 149–158, 2003.

[2] Sumi Choi and Yuval Shavitt. Proxy location problems and their generalizations. *23rd International Conference on Distributed Computing Systems Workshops (ICDCSW'03)*, page 898, 2003.

[3] Itai Dabran, Phillip Klein, and Danny Raz. Efficient management of multimedia attachments. 2004. Submited.

[4] Zvi Drezner. *Facility Location: A Survey of Applications and Methods.* Springer Series in Operations Research, 2001.

[5] Uriel Feige. A threshold of ln n for approximating set cover. *J. ACM*, 45(4):634–652, 1998.

[6] Armando Fox and Eric A. Brewer. Reducing www latency and bandwidth requirements via real-time distillation. *Fifth International World Wide Web Conference*, pages 1445 – 1456, 1996.

[7] Richard Han, Pravin Bhagwat, Richard LaMaire, Todd Mummert, Veronique Perret, and Jim Rubas. Dynamic adaptation in an image transcoding proxy for mobile web browsing. *IEEE Mobile Communications Magazine*, pages 8–17, 1998.

[8] Stefan Hougardy and Hans Jurgen Promel. A 1.598 approximation algorithm for the steiner problem in graphs. In *Proceedings of the tenth annual ACM-SIAM symposium on Discrete algorithms*, pages 448–453. Society for Industrial and Applied Mathematics, 1999.

[9] Samir Khuller and An Zhu. The general steiner tree-star problem. *Inf. Process. Lett.*, 84(4):215–220, 2002.

[10] P. Krishnan, D. Raz, and Y. Shavitt. The cache location problem. *IEEE/ACM Transactions on Networking (TON)*, 8:568 – 582, 2000.

[11] Keqiu Li and Hong Shen. Transcoding proxy placement in en-route web caching. *Communication Networks and Services Research Conference (CNSR 2004)*, 2004.

[12] Huadong Ma and Kang G. Shin. Multicast video-on-demand services. *SIGCOMM Comput. Commun. Rev.*, 32(1):31–43, 2002.

[13] Sangheon Pack, EunKyung Paik, and Yanghee Choi. Design of sip transformation server for efficient media negotiation. In *Proceedings of SPIE (ITCOM 2001)*, volume 4522, 2001.

[14] Joseph C. Pasquale, George C. Polyzos, and George Xylomenos. The multimedia multicasting problem. *Multimedia Syst.*, 6(1):43–59, 1998.

[15] David B. Shmoys. Approximation algorithms for facility location problems. In *Proceedings of the Third International Workshop, APPROX 2000*, pages 27–32, September 2000.