Computer Networks 91 (2015) 244–261

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

# Optimizing network objectives in collaborative content distribution

Xiaoying Zheng [a,*], Ye Xia [b]

[a] Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China
[b] Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL 32611, United States

ARTICLE INFO

*Article history:*
Received 30 May 2014
Revised 1 July 2015
Accepted 21 August 2015
Available online 1 September 2015

*Keywords:*
Content distribution
Peer-to-peer network
Bandwidth allocation
Congestion control
Server selection
Optimization

ABSTRACT

One of the important trends is that the Internet will be used to transfer content on more and more massive scale. Collaborative distribution techniques such as swarming and parallel download have been invented and effectively applied to end-user file-sharing or media-streaming applications, but mostly for improving end-user performance objectives. In this paper, we consider the issues that arise from applying these techniques to content distribution networks for improving network objectives, such as reducing network congestion. In particular, we formulate the problem of how to make many-to-many assignment from the sending nodes to the receivers and allocate bandwidth for every connection, subject to the node capacity and receiving rate constraints. The objective is to minimize the worst link congestion over the network, which is equivalent to maximizing the distribution throughput, or minimizing the distribution time. The optimization framework allows us to jointly consider server load balancing, network congestion control, as well as the requirement of the receivers. We develop a special, diagonally-scaled gradient projection algorithm, which has a faster convergence speed, and hence, better scalability with respect to the network size than a standard subgradient algorithm. We provide both a synchronous algorithm and a more practical asynchronous algorithm.



## 1. Introduction

With the deployment of high-speed access networks such as fiber-to-the-home (FTTH) or its variants, the Internet will be used to transport data on more and more massive scale. The current and future massive content includes high-definition movies and TV programs, large collection of multimedia data, and mountains of all automatically collected/sensed data such as environmental, scientific, or economic data. Beyond that, visionaries are already contemplating 3D super definition (643 Mbps) or 3D ultra definition (2571 Mbps) TV, 3D telepresence, and tele-immersion in virtual worlds. Since the bandwidth mismatch between the network access and network core is expected to be sharply reduced, one can no longer assume virtually unlimited capacity in the future backbone network. Instead, one should not be surprised that, however large, the backbone capacity will be used up by future content. As an evidence, with its early adoption of FTTH, by 2005, Japan already saw 62% of its backbone network traffic being from residential users to users, which was consumed by content downloading or peer-to-peer (P2P) file sharing; the fiber users were responsible for 86% of the inbound traffic; and the traffic was rapidly increasing, by 45% that year [1].

An important networking problem addressed by this paper is how to conduct massive content distribution efficiently in the future network environment where the capacity limitation can equally be at the core or the edge. Our work

* Corresponding author. Tel.: +86 (21)20325135.
*E-mail addresses:* zhengxy@sari.ac.cn, zhengxiaoying@gmail.com (X. Zheng), yx1@cise.ufl.edu (Y. Xia).

http://dx.doi.org/10.1016/j.comnet.2015.08.013
1389-1286/

has applications in the following two-step content distribution process, which is prevalent today and is expected to become more important in the future for reducing wide-area network traffic. In the first step, the content is distributed over infrastructure networks, such as content distribution networks, ISP networks or IPTV networks. In the second step, the end users retrieve the content from one or more nearby content servers. In either step but particularly the first, collaborative distribution techniques are becoming very attractive. By collaborative distribution, we mean different nodes in a distribution session help each other to speed up the distribution or improve other performance measures. A simple form of collaborative distribution is parallel download of a file from multiple nodes, which improves upon the single-server based approach. A more sophisticated form is known as *swarming*, in which each file is broken into many chunks and the nodes (peers) exchange the chunks with each other. One example of swarming is the popular BitTorrent [2]. Although swarming was originally invented in end-system file-sharing applications, it is really a fundamental distribution technique that can be employed by the operators of content distribution networks.

This paper describes how to improve collaborative distribution techniques for achieving *network* objectives in content distribution networks. Since most of these techniques were designed for the end-user environment, targeting end-user performance objectives, they need considerable modification and improvement before they can be applied to content distribution networks and achieve important network performance objectives, such as low network congestion or high throughput. In particular, one common assumption of the current end-user systems is that the network is *access-limited*. As a result, they do not have built-in congestion control or bandwidth allocation mechanisms that can coordinate the entire distribution session and efficiently cope with internal network congestion. Instead, they either rely on the default TCP congestion control, working independently on each individual connection, or do not have any congestion control at all if UDP is used. As will be demonstrated in the paper, they either cause unnecessarily heavy network congestion at parts of the network (due to poorly balanced network load), or miss the opportunity to achieve a shorter distribution time (or equivalently, a higher throughput) given the same network congestion level. The performance gap between what these systems can accomplish and the best possible can be very wide.

This paper proposes a scheme that makes coordinated bandwidth assignment among different connections in the same distribution session so as to minimize the worst-case network congestion, or equivalently, maximize the distribution throughput. The coordination is achieved through fully distributed algorithms. For ease of discussion, we call the receiving nodes the *clients* and the transmitting nodes the *servers*.[1] The scenario under investigation concerns a set of clients requesting chunks of a file (files) or streaming media from a set of servers. For simplicity, we assume all servers have the same content, but this assumption is not crucial.[2] Each client can make parallel download from multiple servers simultaneously. (See Fig. 1 as an example.) Our problem is to select a subset of the servers for each client and decide the transmission rate from each selected server to the client, so that the clients get their required bandwidth (e.g., for streaming requirement), the servers are not overloaded and the worst-case link congestion in the network is minimized.

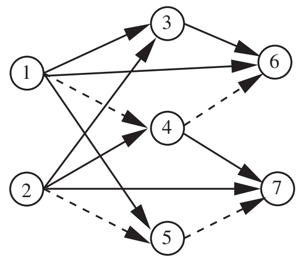

**Fig. 1.** A server–client dependency graph. The solid and dashed lines indicate server-to-client relationship. The solid lines indicate the final server selection results. For instance, node 6 has the server set $S_6 = \{1, 3, 4\}$. It ends up selecting servers 1 and 3. Node 1 has the client set $C_1 = \{3, 4, 5, 6\}$.

Our problem formulation and solution follow the network optimization approach introduced by Kelly et al. [4] and Low and Lapsley [5]. The problem contains a fractional server-selection problem. For instance, a client can get 1/3 of its download from one server and 2/3 from another server. If a connection is assigned a zero or near zero bandwidth, the client essentially has not selected the corresponding server. Our solution to the optimization problem leads to distributed algorithms that combine server assignment with congestion control (or equivalently, bandwidth allocation).

Our contributions are as follows. The results from this paper will be useful for content distribution networks, ISPs and IPTV distribution networks. Up to orders of magnitude improvement in throughput (or reduction in congestion) is possible with our scheme. With respect to network optimization, our solution is a special gradient projection algorithm operating on the primal problem, instead of the subgradient algorithm, which works on the dual problem. For similar network flow problems, the latter is the most frequently used algorithm in the networking literature. For our problem, our experience has shown that the gradient projection algorithm has a faster convergence speed than the subgradient algorithm. The main reason is that the problem of minimizing the worst-case congestion is often ill-conditioned [6]. With the gradient projection algorithm, we are able to overcome this difficulty with diagonal scaling, which tries to emulate the faster Newton's algorithm. For improved practicality, we have also developed an asynchronous version of the algorithm. The correctness (i.e., convergence) of all versions

[1] A node (peer) can be both a client and a server. Every node is a content distribution infrastructure node rather than an end system, although the problem formulation in this paper does apply to end-system P2P file sharing.

[2] Our proposed scheme works the best if appropriate *source coding* is used, such as the Tornado code [3]. With source coding, the file chunks are coded and a receiver can reconstruct the entire file as long as it receives a sufficient number of chunks, irrespective of the identity of the chunks. Each server may contain an arbitrary collection of coded chunks. The nodes exchange coded chunks without the need of knowing what they are.