Available at www.ComputerScienceWeb.com
POWERED BY SCIENCE @ DIRECT•

ELSEVIER

Computer Communications 26 (2003) 2070–2081

computer communications

www.elsevier.com/locate/comcom

Hierarchical Reliable Multicast: Performance Analysis and Optimal Placement of Proxies☆

Sudipto Guha[a], Athina Markopoulou[b,*], Fouad Tobagi[b]

[a]*Computer and Information Science Department, University of Pennsylvania, Philadelphia, PA USA*
[b]*Electrical Engineering Department, Stanford University, Stanford, CA, USA*

Received 18 June 2002; accepted 4 April 2003

Abstract

This paper studies the use of multicast together with proxy nodes for reliably disseminating data from a single source to a large number of receivers. In order to achieve reliability, data must be retransmitted in case of loss either by the source or by special network nodes, called proxies. Each proxy is responsible for reliably delivering the data to a subgroup it is assigned. The multicast tree is partitioned into subgroups that form a hierarchy rooted at the source, hence the term hierarchical reliable multicast. The performance of this approach strongly depends on the topology and the loss characteristics of the underlying tree and the location of proxies. In the first part of the paper, we study the processing and bandwidth performance of such a reliable multicast dissemination given the tree and the placement of proxies. In the second part of the paper, we develop dynamic programming algorithms that give a placement of a fixed number of proxies on an arbitrary tree that minimizes the bandwidth used for reliable transfer. The first algorithm provides an optimal solution to the multicast proxies location problem in polynomial time, in the number of nodes and proxies. The second is an approximation algorithm that gives a solution with cost within a chosen precision from the optimal, in an improved running time. An optimal and an approximate solution are also provided for the proxies location problem if unicast is used for transmissions. Applications of this dynamic programming approach to related problems are discussed.

Keywords: Reliable Multicast; Performance Analysis; Hierarchy; Proxies; Location Problems; Optimization; Dynamic programming; Approximation Algorithm

1. Introduction

Data dissemination applications such as the distribution of newspapers or multimedia component, as well as the updates of stock quotes, software and web caches require reliable data transfer from one source to many receivers of potentially huge number and wide geographical distribution. In this paper, we study the use of multicast for serving such one-to-many applications.

IP multicast naturally fits such applications by constructing the multicast routing tree that allows the source to reach all receivers. The well-known strengths of IP multicast are that it saves network bandwidth by duplicating the packets only where it is necessary and that it allows for dynamic membership in a way transparent to the source. However, IP multicast provides only a best effort delivery, while applications do have reliability requirements.

A large number of *reliable multicast* (RM) protocols have been developed in the last decade mainly for the purpose of ensuring reliability at the transport layer and also for ensuring some congestion control in the network, issues similar to those addressed by TCP. In this paper, we focus on the reliability aspect. Receivers send feedback about whether they received the data packets or not and the source retransmits the packets until all receivers get them.

RM protocols face a number of well-known problems, inherent to the nature of the one-to-many communication

☆ This work has been conducted when the first two authors were Ph.D. students at Stanford. The conference version, which appeared in NGC '00 [17] focused on the analytical metrics for performance evaluation. This journal version, is extended by the algorithms for optimal and near-optimal placement of proxies.

* Corresponding author.
E-mail addresses: amarko@stanfordalumni.org (A. Markopoulou), sudipto@cis.upenn.com (S. Guha).

0140-3664/03/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0140-3664(03)00143-9

model, especially for large numbers of receivers. *Feedback implosion* occurs when large groups send feedback without duplicate suppression, thus unnecessarily wasting bandwidth, overwhelming the source with processing and increasing the latency in the delivery of data. Another inherent problem is the *exposure* [21] (also called the *crying baby problem* [7]): parts of the tree that experience high loss, expose the rest of the tree to unnecessary retransmissions. The *drop-to-zero* problem occurs when receivers with poor capabilities force the rest of the group to adjust at the slowest rate.

One successful approach that deals with the above problems is the hierarchical proxies approach, which we call *hierarchical reliable multicast*(*HRM*). This approach partitions the multicast delivery tree into subgroups that form a hierarchy rooted at the source. Each subgroup has a representative, called proxy, which keeps copies of data packets, collects the feedback from the receivers in the subgroup and locally retransmits the packets, if needed. Feedback implosion is limited because each proxy handles a smaller size of subgroup. Limiting the feedback and the retransmissions locally saves bandwidth and limits the exposure of the good parts of the tree. The recovery latency is reduced as repairs come from a proxy located close to the point of loss. Large numbers of receivers may join the group in this hierarchical scalable way.

The performance of such hierarchical schemes, strongly depends on the underlying tree topology and its loss characteristics, and on the selection of proxies. In this paper, we study two problems: the performance evaluation of a tree given a placement of proxies and the optimal placement of proxies.

The structure of the rest of the paper is as follows. In Section 2 we discuss work in this area and how our work relates to it. In Section 3 we present our model for HRM. The performance evaluation of an entire multicast tree is reduced to evaluating the performance of every subgroup, in Section 4. The performance measures considered are the work at the source (Section 4.1) and the network bandwidth (Section 4.2) needed for reliable transfer. Section 5 studies the optimal placement of a fixed number of proxies on the multicast tree in order to minimize a total bandwidth measure, for multicast and unicast transmissions in Sections 5.1 and 5.2, respectively. In Section 5.1(1), we give an algorithm that provides an optimal placement for the multicast problem. In Section 5.2(2), we give an approximation algorithm for the multicast problem too. An example of the approximation algorithm using a realistic topology and MBONE measurements is given in Section 5.3(3). Also algorithms to find both the optimal and an approximation are given in Section 5.2 for the unicast case. Possible applications of the dynamic programming approach to related problems are discussed in Section 5.3. Section 6 concludes the paper.

2. Related work and contributions

There are three bodies of work related to this paper: (i) implementation work that motivates the study (ii) performance evaluation work related to the first problem and (iii) optimization algorithms for location problems, related to the second problem we are studying.

As discussed in Section 3, there are many applications [9,29] and transport protocols [3,6,7,15,16,21,23,28] that follow the model of HRM by using local error recovery combined with hierarchy. They address performance evaluation by means of simulation and they place their proxies using heuristics. Our work is of analytical nature and can be used for the assessment of these schemes and for optimizing their hierarchical structure. On the other hand, measurements [27] and realistic network scenarios [6], are the input to our algorithms.

The first problem we study is performance evaluation in terms of work at the source and bandwidth. Previous work on performance analysis of reliable multicast has focused on the work at the source. The $E[M]$ measure of 3 Section 5.1 was defined and analytically calculated in Ref. [1] and ever since it has been used in most performance analysis work on RM: [6,10,12,18,19,20,26,]. Its calculation is computationally intense, so approximations [18] or simulations [6] are often used. We extend the understanding of this measure and we apply it on a realistic example. The second measure $E[T]$ captures the bandwidth used to reach all receivers at least once. So far, it has been only counted by simulation [15] or approximated by an upper bound which is tight in simplified star topologies [10,12,19]. In Section 4.2 we calculate it analytically based on our previous work [17]. We give closed formulas in special cases and a recursive method for the general case.

The second problem we study is the optimal placement of a fixed number of proxies (k) in order to minimize the total bandwidth needed to reach all (n) receivers at least once in an expected sense. Similar problems that look for optimal placement of facilities on graphs, known in general as *location problems* and in particular as the *K-median* problem, have been studied in length, [2,4,11,14,24] for unicast data delivery. To the best of our knowledge, the location problems for multicast transmission had not been solved so far. If unicast transmissions are used, the cost of a subtree depends only on the subtree itself. The multicast case is fundamentally different in that the cost of a subtree depends on the transmissions destined to nodes not only inside but also outside of the subtree itself; each multicast transmission is heard by all nodes whether they need it or not (*exposure* problem). We dealt with special cases, i.e. chains and uniform trees, in previous work [17].

In this paper, we provide two algorithms that solve the location problem for an arbitrary multicast tree. They both follow a dynamic programming approach, similarly to