Available online at www.sciencedirect.com
SCIENCE DIRECT®
The Journal of Systems and Software 75 (2005) 305–318
The Journal of Systems and Software
www.elsevier.com/locate/jss

An agent based adaptive bandwidth allocation scheme for multimedia applications

S.S. Manvi, P. Venkataram *

Protocol Engineering and Technology Unit, Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560 012, India

Received 10 January 2003; received in revised form 6 August 2003
Available online 24 March 2004

Abstract

Bandwidth allocation for multimedia applications in case of network congestion and failure poses technical challenges due to bursty and delay sensitive nature of the applications. The growth of multimedia services on Internet and the development of agent technology have made us to investigate new techniques for resolving the bandwidth issues in multimedia communications. Agent technology is emerging as a flexible promising solution for network resource management and QoS (Quality of Service) control in a distributed environment. In this paper, we propose an adaptive bandwidth allocation scheme for multimedia applications by deploying the static and mobile agents. It is a run-time allocation scheme that functions at the network nodes. This technique adaptively finds an alternate patchup route for every congested/failed link and reallocates the bandwidth for the affected multimedia applications. The designed method has been tested (analytical and simulation) with various network sizes and conditions. The results are presented to assess the performance and effectiveness of the approach. This work also demonstrates some of the benefits of the agent based schemes in providing flexibility, adaptability, software reusability, and maintainability.

Keywords: Mobile agents; Bandwidth allocation; Alternate routing; Multimedia applications

1. Introduction

Multimedia applications are delay sensitive because of its real-time requirements and consume significant amount of network bandwidth due to huge data involvement. Therefore, a network should be capable of offering guaranteed bandwidth, bounded delay and real-time delivery to provide acceptable quality of service (QoS) to them. Bandwidth is an important QoS parameter around which other QoS parameters (queuing delays, loss, etc.) revolve. Initial bandwidth allocation does not guarantee bandwidth for the life-time of an application since the network is often subjected to congestion and failures. Hence, there is a need for developing adaptive bandwidth allocation strategies at the transport level to cope up with these problems. Several transport level allocation schemes namely, compression, variable rate decoding, multicasting, statistical multiplexing and temporal smoothing are used to reduce the bandwidth requirements (Lu, 2000; Krunz, 1999). Some of the bandwidth reduction techniques are based on temporal and spatial resolution which uses frame/packet dropping and low resolution data, respectively.

The bandwidth allocation schemes must minimize packet losses, delays, minimize renegotiations and maximize network bandwidth utilization. Several static and adaptive bandwidth allocation schemes are used to support multimedia applications. Static allocation schemes such as peak rate allocation are simple and lead to under-utilization of the network bandwidth. To overcome under-utilization of the bandwidth, several statistical allocation schemes such as equivalent capacity, heavy traffic approximations, critical and optimal bandwidth allocation schemes are proposed (Perros and Elsayed, 1996; Feng et al., 1997).

Some of the adaptive bandwidth allocation schemes readjust the bandwidth according to implicit or

* Corresponding author.
E-mail addresses: sunil@protocol.ece.iisc.ernet.in (S.S. Manvi), pallapa@ece.iisc.ernet.in (P. Venkataram).
URL: http://pet.ece.iisc.ernet.in/pallapa.

0164-1212/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jss.2003.09.031

explicit feedback about the network status from the intermediate nodes and the receivers (Hou et al., 2000; Busse et al., 1996). Feedback information indicates congestion level on the links of a path (congestion can be a measure of buffer overflow, packet delays, and offered load). Other factors that cause adaptive allocation are pricing policies, application priorities, queuing mechanisms for scheduling bandwidth, etc. (Dasilva, 2000; Congestion, http site; Bar-Noy et al., 1995). Even-though, these schemes optimally allocate the bandwidth over a path, but they do not optimally utilize the network bandwidth.

Due to volatile nature of the Internet traffic, it may not be possible to allocate the required amount of bandwidth online over a path, since the traffic on it may be overloaded. In such a case, it is beneficial to consider other lightly loaded paths and reallocate the bandwidth for efficient utilization. Thus it is required to take into account the joint problems of rerouting and bandwidth reallocation in case of network congestion/failures by exploiting the uneven load distribution in the network.

The possible factors influencing the rerouting and adaptive bandwidth allocation are: *link or node failures*; *variations of traffic load leading to congestion*; *and policy changes by administrator*. Some of the techniques which use rerouting and adaptive bandwidth allocation are as follows. Self healing and backup virtual path schemes are proposed for restoration of the network during failures by adaptively allocating the bandwidth on the alternate routes (Kawamura et al., 1994; Anderson et al., 1994; Hou, 1997; Wipusitwarakun et al., 1998). In Yahara et al. (2000), the authors consider a new self healing scheme that reallocates bandwidth for differentiated bandwidth requirements of the applications during rerouting under link/node failure situations.

Some of the other works done on adaptive bandwidth allocation and rerouting under traffic load variations are presented in (Yamashita et al., 1995; Gupta et al., 1998; Wong et al., 2000; Chu and Tsang, 1995; Gupta and Gandhi, 1994; Karol and Shaikh, 1991; Orda et al., 1992). They mainly use either of the following principles: bumping packets, least loaded paths, minimum delay routes, partition policy with reservation on a link, Markov decision process (MDP) and path rearrangement.

Separated and unified schemes of bandwidth allocation and routing are analyzed in (Wong et al., 1996). In separated scheme, a virtual path can carry only applications' of same class whereas in the unified scheme, a path can carry data from the different classes. Source based adaptive bandwidth allocation techniques using agents are suggested in (Manvi and Venkataram, 2000; De Meer et al., 2000; Guedes et al., 1998). The work given in (Lipperts and Kreller, 1999) uses mobile agents to learn the load on the network and update the routing tables to have load balance in a network. But, they do not consider the joint problem of bandwidth allocation and rerouting based on traffic load variations and link failures.

We observe from the literature that none of the existing schemes comprehensively deal with the distributed adaptive bandwidth allocation and rerouting in Internet by considering both congestion and link failures. And, also the traditional allocation schemes lack extensibility, customizability, software reuse, maintainability and flexibility which is needed in current Internet software development process (Griss and Pour, 2001). Agent technology seems to provide a solution to deal with these issues and facilitate network programmability, although technology is still in its infancy and several problems are to be resolved while using mobile agents. The problems are: security to agents from hosts and vice versa; routers' support for agents; coordination and communication protocols among agents and fault tolerance. Nevertheless, an agent sent across a network may allow the applications and the service providers to share the resources more effectively, as they can be controlled at the application level. However we assume that the mentioned problems will be rectified at the earliest.

In this paper, we propose an agent based adaptive bandwidth allocation and rerouting of multimedia applications in the case of network congestion/failures (as well as network approaching towards congestion). It is a run-time allocation scheme, does not look into initial start-up bandwidth allocation, and assumes the availability of application's start-up bandwidth allocation information in every node of the network. The scheme uses partial topology information (first and second degree neighbors of the node which detected the link problem) to compute the disjoint alternate routes to patchup a congested/failed link. The discussion of alternate routing issues is beyond the scope of this paper.

Agent technology has been used in the past for allocating the bandwidth over a given path without considering the rerouting. In our scheme, we propose a novel idea of using agent technology for performing dynamic bandwidth allocation along with rerouting in the network nodes to facilitate the smooth running of multimedia applications under event of congestion or failures. The mobile agents in our work not only perform bandwidth allocation for a single application, but also, attempts to allocate the bandwidth for all the applications passing through its originator node (agent creator) by creating a set of alternate routes for those applications.

In the proposed scheme, mobile agents at a node are triggered as and when either a link congestion or failure or a link approaching towards congestion are detected. The scheme triggers the mobile agents on disjoint alternate route (local patchup route) to capture and reserve the bandwidth for the affected applications in both directions and reroutes the accepted applications. The scheme per-