Network Support For Multimedia

A Discussion of the Tenet Approach*

Domenico Ferrari, Anindo Banerjea, and Hui Zhang
Tenet Group
Computer Science Division
University of California at Berkeley
&
The International Computer Science Institute,
Berkeley, CA

TR-92-072
November 1992

Abstract

Multimedia communication can be supported in an integrated-services network in the general framework of realtime communication. The Tenet Group has devised an approach that provides some initial solutions to the realtime communication problem. This paper attempts to identify the principles behind these solutions. We also describe a suite of protocols, and their implementations in several environments, that embody these principles, and work in progress that will lead towards more complete solutions.

*This research was supported by the National Science Foundation and the Defense Advanced Research Projects Agency (DARPA) under Cooperative Agreement NCR-8919038 with the Corporation for National Research Initiatives, by AT&T Bell Laboratories, Digital Equipment Corporation, Hitachi, Ltd., Hitachi America, Ltd., Pacific Bell, the University of California under a MICRO grant, and the International Computer Science Institute. The views and conclusions contained in this document are those of the authors, and should not be interpreted as representing official policies, either expressed or implied, of the U.S. Government or any of the sponsoring organizations.

1 Introduction

High speed networking has introduced opportunities for new multimedia applications such as video conferencing, scientific visualization and medical imaging. These applications have stringent network performance requirements in terms of parameters such as bandwidth, delay, delay jitter, loss rate or some combination of these. In the area of computer networking, multimedia applications can be considered a subset of the larger class of realtime applications, which include realtime distributed computations, remote control systems, and interactive applications. We shall therefore address the problem of realtime communication rather than the more restricted one of multimedia communication (see also §3.2). While circuit-switching networks can provide realtime support, the belief is now widespread that future integrated-services networks will use packet-switching techniques to achieve a more efficient usage of network resources. Since current packet-switching networks cannot guarantee the required levels of performance to their clients, the problem of how to support multimedia and other realtime applications in a packet-switching environment has become a pressing research issue.

The Tenet Group at the University of California at Berkeley and the International Computer Science Institute has been working since 1988 to provide a practical solution to the problem of realtime communication. The initial goal of the work was to devise and specify in all its details a set of algorithms that, when implemented in a network, would enable the network to offer a realtime service to its clients. Such a set of algorithms was called a *realtime communication scheme*, or a *scheme* for brevity. When the initial scheme (to be called Scheme 0 in this paper) was completed, the group was offered the opportunity to perform realtime communication experiments on the Xunet 2 wide-area networking testbed [FKK+92]. This, however, required several extensions to Scheme 0; the resulting scheme (Scheme 1 in the sequel) has been embodied in a suite (Suite 1) of realtime protocols designed by the Tenet Group.

The design of these schemes and suite required, as is the case with all designs, a number of compromises and restrictive assumptions. However, the group's previous publications paid much more attention to the features of the schemes and suite than to the distinction between the principles of the approach and the design compromises. This has unfortunately misled many discussions of our work. In this paper, we try to identify the principles that have guided, since the beginning, the Tenet Group's work, so that the debate can be focused on the crucial issues rather than on less important or unimportant considerations. These principles can be regarded as constituting and characterizing the *Tenet approach* to realtime communication. Figure 1 shows the relationships among the concepts mentioned in this section.

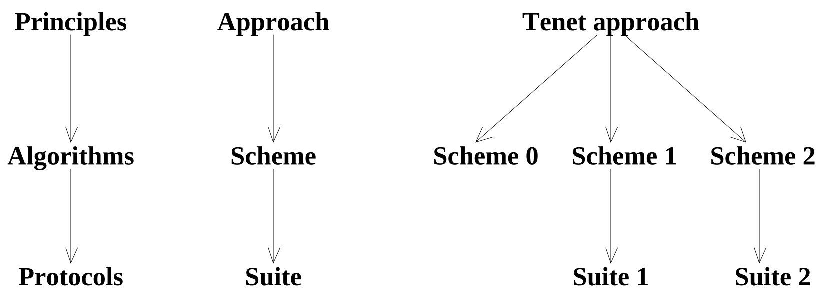

Figure 1: Components of the Tenet approach.

A large number of schemes can be constructed that are consistent with the Tenet approach. Some schemes may obey only a subset of the principles of the Tenet approach if the corresponding restrictions are deemed necessary to make the scheme's construction manageable.

We summarize the main features of Scheme 0 and Scheme 1 in Section 2 as a background for the following discussion. In Section 3, we first clarify the meanings of some of the terms we have used, and identify the most important compromises we made in our designs; then we discuss the principles that constitute the Tenet approach to realtime communication. Section 4 describes the design of Suite 1 and its relationships to both Scheme 1 and the networking testbeds for which it has been implemented. Some of the novel aspects of Scheme 2, now being designed by the Tenet Group, are outlined in Section 5. Section 6 briefly examines related work from the viewpoint of the Tenet principles. Section 7 concludes the paper.

2 The initial Tenet scheme and its extensions

2.1 The initial scheme

The initial Tenet scheme (Scheme 0) is described in detail in [Fer89], [FV90a], [FV90b], and [VZF91]. The network type it assumes has general topology, but consists of identical switches and identical hosts. Thus, Scheme 0 applies to a homogeneous network; internetworking is not supported.

The communication abstraction on which Scheme 0 is based is the *realtime channel*. For the purposes of Scheme 0 a *realtime channel* is defined as a simplex unicast end-to-end connection with performance guarantees and restrictions on traffic. A realtime channel is associated with a set of nodes and links within the network, through which the realtime data packets for the connection pass, and which we call the *route* of the channel. To satisfy the performance guarantees, we reserve resources in the nodes on the route, and perform admission control tests during the process of establishing the channel.

2.1.1 Service definition

A realtime communication service is defined by the interface it offers to its clients [Fer90]. The interface of a service based on Scheme 0 allows the clients to specify their loads and their performance needs, and can be thought of as leading to a contract between each client and the network.

The traffic specification to be given by the client includes parameters expressing the peak and average load on the network, and an indication of the burstiness of the load. These parameters, which are allowed to take values from a continuous range over the set of positive integers, are:

- Minimum packet inter-arrival time x_min
- Average packet inter-arrival time x_ave
- Averaging interval I
- Maximum packet size s_max

The performance parameters by which clients describe their requirements are:

- Delay bound D
- Delay violation probability bound Z
- Buffer overflow probability bound W
- Delay jitter bound J

Again, each of these parameters is allowed to take values from a continuous range. The delay violation probability bound (Z) allows the client to specify the hardness of the guarantee on the delay bound. For example, if Z is one, then the guarantee is *deterministic*, otherwise it is *statistical*. A guarantee on bandwidth is obtained from the traffic specifications, because the network promises