Efficient Group Communication with Guaranteed Quality of Service*

Clemens Szyperski Giorgio Ventre†

The Tenet Group
International Computer Science Institute
and
Computer Science Division
University of California, Berkeley
1947 Center Street, Suite 600
Berkeley, CA, 94704

ABSTRACT *A specific class of multimedia applications is expected to be of importance for future communication networks: Multi-Party Interactive Multimedia (MIM). Based on the isolation and characterization of MIM applications, concrete network support requirements are derived in this paper. The varying degree of connectivity, the vastly different sizes in terms of participants and the reliance on a guaranteed Quality of Service make MIM support a difficult problem. Starting with the definition of multimedia communication abstractions, principles of solutions are sketched. For an important subclass of applications a particularly efficient and practicable alternative implementation based on half-duplex channels is introduced. Finally, interfaces at both the transport and network layers are considered.*

1 INTRODUCTION

In the area of distributed systems, two events can be considered as among the major sources of the research activity on high-speed networking. The first one is the increasing demand in terms of communication bandwidth due to the efforts to make multimedia technology available in distributed environments. The second event is the presence of more stringent requirements in terms of performance and reliability of the communication system due to the isochronous nature of the new media involved. The concept of reliability itself is intended now in a broader sense, and includes the capability of a network to guarantee certain performance parameters for the whole duration of a communication.

Among multimedia applications, those based on multi-party interactive multimedia (MIM) [7], show an inherent complexity, stemming from the different media they use and the large number of participants they might involve. Due to their origin from interpersonal communication models, multi-party interactive applications have been, in general, rather vaguely defined.

Other proposals have been presented in the literature for the classification of MIM applications from the viewpoint of the support they require from the communication subsystem [1, 6, 8, 10, 11, 12, 13]. In this paper we propose a different approach to such classification based on the identification of a set of basic functional characteristics that this class of multimedia applications exhibits. The first goal of this research is to show that such characteristics can be effectively exploited to reduce the amount of resources required to provide real-time communication. The second goal is to still accomplish this result with a guaranteed quality of service according to the Tenet approach to real-time networking [2]. According to the Tenet approach means, first of all, support of a contract between network client and network provider; in this contract, the client specifies its traffic model and its Quality of Service (QoS) requirements, and the network, if it accepts, establishes a channel that guarantees the QoS as long as the client sticks to its traffic model. Secondly, for the network to be able to keep its contracts, it needs to protect well-behaved channels from misbehaving ones. Thirdly, the traffic model and the QoS requirements are to be specified in a general way, not pertaining to any specific class of traffic.

The solution we propose is based on a new abstraction, called the *Half-Duplex Real-Time Channel*. This solution reduces the complexity of the creation of network support for common MIM applications, and decreases the amount of resources to be reserved in the network.

We start, in section 2, from examples of multi-party applications to extract a set of attributes we believe to be the most significant for the support of these applications by means of computer-based tools. We focus our attention on those attributes related to the logical aspects of the information exchange, and make little assumptions on possible implementations in a real-time communication network.

In section 3 we broaden our scope by considering the size of a communication event in terms of the number of participants. Based on that, section 4 follows with basic approaches, and section 5 introduces the half-duplex real-time channel. Section 6 contains the required admission tests for the Tenet protocol suite, and section 7 proposes actual client service interfaces.

---

*This research is 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.

†On sabbatical leave from Università degli Studi di Napoli "Federico II", Napoli, Italy.

2 Model of Interaction

Interpersonal communication models show differences in the way data is exchanged among involved participants. In order to offer computer-based multimedia tools implementing these and other models, we need to analyze how such differences influence the kind of communication service these tools require from the network. This can be done by considering, as the prevailing aspect, the *model of interaction* criterion, defined as the aspect that covers how interaction may occur among participants of distributed applications. For example, we are interested in applications based on multi-party interactive multimedia (MIM). Depending on the ability of a participant in such applications to produce and transmit multimedia data, we can identify three different kinds of interaction: *Dynamic*, *Static*, and *Controlled.*

We will present our conjectures by considering a computer supported multimedia environment where participants to a multi-party communication are geographically distributed entities (*nodes*), interacting through communication abstractions (*channels*) over a real-time internetwork.

A *dynamic* interaction is required for applications in which the amount of information that can be exchanged among all the participants is almost equivalent and the characteristics of the applications do not allow the determination of a particular subset of participants that will be the main source of the data flow. Since the coexistence of multiple flows of information should be, in these cases, permitted, support for simultaneous communications among the nodes should be provided by the network, in spite of any arbitration or control mechanisms that might be imposed by the users.

A *static* interaction is required in all applications where the information flows from a statically determined subset of nodes, or where the amount of data flowing from the other nodes does not justify the creation of a dedicated, real-time communication channel. Typical applications requesting such kind of interaction are those generally referred to as *telepresence* [5]. An example of this kind of application is *teleprogramming* [4], where a user receives a number of different media from remote devices, sensors or computers, typically to control or manipulate a remote apparatus, according to a many-to-one information flow.

A *controlled* interaction is required when the application specifically permits the identification of a dynamically variable subset of participants that contributes to the data flow. The subset may depend on the specific attributes of the participants. For example, in a conference, there exist subsets including the conference speakers and the conference managers respectively. Both sets are expected to contribute to the data flow more than the attendees. The identification of the subset of senders can also depend on application-specific rules and agreements.

Table 1 lists some interesting multi-party applications. The entries are meant to indicate the dominant cases, but should not be taken as discriminative in the mathematical sense.

Type	***Description***	***Model of Interaction***	***Data Flow***
Course Lecture	Long term setup, lecture with managed discussion	Controlled	1 → N
Seminar	Special event version of Course Lecture	Controlled	1 → N
Meeting	All participants actively send/receive data	Dynamic	N → N
Public Hearing	Small subset of senders many receivers	Static	M → N (M << N)
Conference Panel/Debate	Moderated Questions & Answers from a subset of participants	Controlled	M → N (M << N)
CSCW	All participants actively send/receive data	Dynamic	N → N
Telepresence	Many senders, one receiver	Static	N → 1
Broadcasting Services	One sender, many receivers	Static	1 → N
General Distributed Computing	All participants actively send/receive data	Dynamic	N → N
Phased Distributed Computing	All participants actively send/receive data	Controlled	N → N

Table 1: Characteristics of Selected MIM Applications.

3 Aspects of Size

So far we considered purely functional aspects of MIM applications and no limits of any sort were assumed for the capability of a real-time communication system to provide the services required by these applications.

We believe, however, that the important issue of size of MIM applications should be taken into account [10]. By *size* we refer to the expected bound on the number of participants. In particular, we are interested in how support from a real-time communication subsystem has to cope with this aspect.

Indeed, the maximum number of participants that a particular application may support depends on when it becomes infeasible either for the human user or for the service provider to add another participant. For example, from a logical viewpoint, the dynamic model of interaction of some MIM applications might require a fully interconnected communication structure. This requirement seems to dictate a small bound on the maximum number of participants, since the connection complexity grows quadratically, and the information to be consumed by each participant grows linearly with the number of participan