**IDEA GROUP PUBLISHING**

701 E. Chocolate Avenue, Hershey PA 17033-1117, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITB7314

## Chapter XIII

# Playout Control Mechanisms for Speech Transmission over the Internet: Algorithms and Performance Results

Marco Roccetti
University of Bologna, Italy



Sophisticated applications of Internet multimedia conferencing will become increasingly important only if their users will perceive the quality of the communications as sufficiently good. The result of extensive experiments has shown that audio is frequently perceived as one of the most important components of multimedia communications. Unfortunately, the actual architecture of the Internet is not a good environment for real-time audio communications, since very high transmission delay and transmission delay variance (known as jitter) may be experienced that impair human conversations. Hence, in the absence of network support to provide guarantees of quality to users of Internet voice software, an alternative to coping with problems caused by delay and delay jitter is to use adaptive control mechanisms. These mechanisms are based on the idea of using a voice reconstruction buffer at the receiver in order to add artificial delay to the audio stream to smooth out the jitter. In this chapter, we describe three different control mechanisms that are able to dynamically adapt the audio application to the network conditions so as to minimize the impact of delay jitter (and packet loss). We also present a set of performance results we have gathered from an extensive experimentation with an Internet audio tool we have designed and developed in order to conduct voice-based audio conversations over the Internet.

This chapter appears in the book, Multimedia Networking: Technology, Management and Applications by Syed Mahbubur Rahman.

# INTRODUCTION

The value of integrating real-time and traditional data services onto a common network is well known. Telecommunication companies have proposed ATM as a standard for upgrading the Internet to provide both real-time and data services. In contrast, it has been empirically demonstrated that real-time multimedia services may be added to traditional IP networks that were originally designed for data transmission only. However, since the IP community has not considered the provision of Quality of Service guarantees with the same intensity as the ATM community, the current Internet service model offers a flat, classless, best-effort service to its users. This is not a good environment for real-time audio transmissions since the delay between the arrival of subsequent packets becomes dependent on the traffic condition of the network. It is a common experience that very high packet delay and packet delay variance (known as jitter) may be experienced over many congested Internet links. As a consequence, audio packet loss percentages, due to the effective loss and damage of packets as well as belated arrivals, may be very large. Unfortunately, user studies demonstrate that, among the many possible metrics that influence the user perception of audio, probably the most important factors that disrupt the user perception of audio are represented by the packet audio playout delay and the packet loss rate.

With the term *playout delay*, we refer to the total amount of time that is experienced by the audio packets of a given talkspurt from the instant they are generated at the source and the instant they are played out at the destination. Summarizing, such a playout delay consists of: i) the *collection* time needed for the transmitter to collect audio samples and to prepare them for transmission, ii) the *transmission* time needed for the transmission of audio packets from the source to the destination over the underlying transport network, and finally iii) the *buffering* time, that is the amount of time that a packet spends queued in the destination buffer before it is played out. Thus, in the absence of network support to provide guarantees of quality to users of Internet voice software, an interesting alternative to coping with problems caused by jitter and high packet loss is to use adaptive control mechanisms. Typically, these mechanisms are based on the idea to use a voice reconstruction buffer at the receiver in order to add artificial delay to the audio stream to smooth out the jitter. Clearly, the longer the scheduled playout delay, the more likely it is that an audio packet will arrive at the destination before its scheduled playout deadline has expired. However, too long playout delays can significantly impair human interactive conversations.

In this chapter, we will describe the most characterizing features of a number of those control mechanisms that try to dynamically adapt the audio application to the network conditions so as to minimize the impact of delay jitter and packet loss. We will report, also, on a set of performance results we have gathered from an extensive experimentation with an Internet audio tool we have designed and developed in order to conduct unicast, voice-based audio conversations over the Internet.

The remainder of this chapter is organized as follows. In the next section, we present a brief overview on digital audio coding techniques that aims at describing the state of the art of the audio compression methods for human conversational speech. In the third section, three important algorithms are detailed that are typically used for transmitting human speech over the Internet. In the fourth section, performance figures derived from several experiments are discussed that illustrate the adequacy of those mechanisms in transmitting speech across the Internet. Finally, the conclusions complete the chapter in the last section.

## Related Content

Simulation-Based Comparison of TCP and TCP-Friendly Protocols
Gábor Hosszú (2009). *Encyclopedia of Multimedia Technology and Networking, Second Edition (pp. 1307-1315).*
www.irma-international.org/chapter/simulation-based-comparison-tcp-tcp/17550

A Practice Perspective on Transforming Mobile Work
Riikka Vuokko (2011). *Handbook of Research on Mobility and Computing: Evolving Technologies and Ubiquitous Impacts (pp. 1119-1131).*
www.irma-international.org/chapter/practice-perspective-transforming-mobile-work/50643

Learning Full-Sentence Co-Related Verb Argument Preferences from Web Corpora
Hiram Calvo, Kentaro Inuiand Yuji Matsumoto (2012). *Quantitative Semantics and Soft Computing Methods for the Web: Perspectives and Applications (pp. 137-162).*
www.irma-international.org/chapter/learning-full-sentence-related-verb/60119

Ethical Reasoning and Reflection as Supported by Single-Player Videogames
Jose P. Zagal (2011). *Designing Games for Ethics: Models, Techniques and Frameworks (pp. 19-35).*
www.irma-international.org/chapter/ethical-reasoning-reflection-supported-single/50729

Buffer Control Techniques for QoS Provisioning in Wireless Networks
Michael M. Markouand Christos G. Panayiotou (2009). *Handbook of Research on Wireless Multimedia: Quality of Service and Solutions (pp. 157-182).*
www.irma-international.org/chapter/buffer-control-techniques-qos-provisioning/22023