



INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Appendix I
Rec. G.764
(11/95)

GENERAL ASPECTS OF DIGITAL TRANSMISSION SYSTEMS

PACKETIZATION GUIDE

Appendix I to ITU-T Recommendation G.764

(Previously "CCITT Recommendation")

FOREWORD

The ITU-T (Telecommunication Standardization Sector) is a permanent organ of the International Telecommunication Union (ITU). The ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Conference (WTSC), which meets every four years, establishes the topics for study by the ITU-T Study Groups which, in their turn, produce Recommendations on these topics.

The approval of Recommendations by the Members of the ITU-T is covered by the procedure laid down in WTSC Resolution No. 1 (Helsinki, March 1-12, 1993).

Appendix I to ITU-T Recommendation G.764 was prepared by ITU-T Study Group 15 (1993-1996) and was approved under the WTSC Resolution No. 1 procedure on the 13th of November 1995.

NOTE

In this Appendix, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

© ITU 1996

All rights reserved. No part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the ITU.

CONTENTS

	<i>Page</i>
I.1 Introduction	1
I.2 Historical background.....	1
I.3 Summary of design issues	2
I.4 Reconstitution of speech signals.....	2
I.5 Delay equalization	2
I.5.1 Blind delay	3
I.5.2 Absolute time stamp.....	3
I.5.3 Relative time stamp.....	4
I.6 Robustness to errors.....	4
I.6.1 Sensitivity of speech to bit errors.....	5
I.6.2 CRC calculation over part of the frame	5
I.6.3 CRC calculation over whole frames	5
I.7 Congestion control.....	5
I.7.1 Global control	6
I.7.2 Local control	6
I.8 Packet loss	8
I.8.1 Case of speech.....	8
I.8.2 Case of voiceband data	9
I.8.3 Fill-in strategies (speech).....	9
I.9 Choice of packet size	10
I.9.1 Speech considerations.....	10
I.9.2 Bit error considerations.....	10
I.9.3 Integrated traffic.....	10
I.9.4 Recommendation G.764.....	11
I.10 Compression issues.....	11
I.10.1 Speech coding algorithms	11
I.10.2 Digital speech interpolation	11
I.11 Channel-oriented signalling.....	12
I.12 Extensions.....	12
I.12.1 Facsimile.....	12
I.12.2 Speech/video synchronization.....	13
I.12.3 Interface between the PSTN and LANs.....	13
I.12.4 Extension to new algorithms.....	14
I.13 Summary.....	14
References	14

Page blanche

p. ii



Documents
Document Sharing Hub

Appendix I to Recommendation G.764

Packetization Guide

(Geneva, 1995)

(This appendix does not form an integral part of this Recommendation)

I.1 Introduction

This appendix summarizes the current views on speech packetization in Study Group 15 of the ITU-T during the study period 1992-1996 and in CCITT SG XV during the study period 1988-1992. These views may change in the future.

This appendix does not address the following topics:

- 1) the different treatments within the network that may be accorded to packets depending on their assignment to priority classes, in general; however, it is agreed that the network should give priority to speech over digital data to reduce delay and delay variability, and to decouple bursty traffic from real-time traffic;
- 2) quality of service for different traffic classes.

The purpose of this appendix is to:

- explain the various issues that affect the packetization of speech;
- provide an overview of the techniques and considerations in the transport of packetized speech, that are used in Recommendation G.764;
- disseminate information on the various topics of concern to designers, implementors of packetized speech equipment and to the service providers that use them.

I.2 Historical background

Traditionally voice services have been implemented in the Public Switched Telephone Network (PSTN) (also denoted as Wide Area Network or WAN) using a circuit-oriented approach. The growth of packet transport techniques [e.g. Recommendation X.25/X.75, Internet, Wideband Packet Technology, Frame Relay and the Asynchronous Transfer Mode (ATM)] has stimulated research in new techniques for the transport of speech.

Packetized systems can exploit the bursty nature of traffic to multiplex different types of traffic (e.g. voice, data, video) of many users so that they can share transmission bandwidth and switching resources dynamically. Packetization facilitates the integration of the different types of traffic to allow more efficient utilization of the available bandwidth and switching resources. Packetization offers more flexibility than circuit-oriented approaches because the packet header contains the necessary control information that identifies, for example, the type of traffic and, where appropriate, the coding scheme.

Work in speech packetization began in the CCITT since the middle of the 1984-1988 study period in Working Party XVIII/8 and continued in Working Party XV/2 during the 1989-1992 study period. It is now continuing, in the ITU-T for wideband packet and ATM networks. The goal is to provide a uniform basis for speech packetization, with or without speech compression and speech interpolation, to facilitate the interworking of equipment from various vendors in telecommunications applications.

The work in the CCITT has led to the Voice Packetization Protocol of Recommendation G.764 and its extensions in Recommendation G.765. These protocols are compatible at the link layer with the ISDN protocols LAPD and LAPF specified in Recommendations Q.921 and Q.922, respectively.

I.3 Summary of design issues

Using the OSI protocol stack as a reference, the major design considerations in devising a speech packetization protocol are:

- 1) Layer 1 (physical layer) – The issue is whether the physical interface will conform to that of public telephone networks (Recommendation G.703/G.704) or to other local area networks, such as IEEE 802.2, 802.3 or 802.9, etc.
- 2) Layer 2 (link layer) – Some of the issues are:
 - a) whether the logical layer will be compatible with ISDN (LAPD/LAPF) protocols or will have the same structure as those for LANs;
 - b) how to deal with the loss of frames;
 - c) robustness to errors.
- 3) Layer 3 (procedures to deal with digitized voice and voiceband data traffic) – Issues are:
 - a) the delay variability for speech packets;
 - b) the transport of channel-associated signalling.
- 4) Higher-layer issues involve the speech coder and the type of compression used.

A speech packetization protocol has the following requirements:

- The speech must be reconstructed at the receiving end from packets arriving at irregular intervals (or, in some architectures, out of order).
- The protocol must be robust against line errors.
- It must offer an easy method for congestion control in the network.
- It must specify procedures at the terminating end to recover from packet loss or excessive delay.
- It must carry channel-associated signalling.
- If digital speech interpolation is used to eliminate silence intervals, it shall specify the level at which noise is re-injected at the terminating end.

I.4 Reconstitution of speech signals

To achieve good speech quality, the terminating end must reconstitute a continuous speech stream and play it out at regular intervals despite varying packet arrival times. This involves two aspects:

- 1) preserving the relative timing of information within one speech burst; and
- 2) delay equalization.

In Recommendation G.764, a packet sequence number is used to encode the relative timing of the information within one speech burst. The first packet of a speech burst always has the sequence number of 0; subsequent packets in the same burst have the numbers from 1 to 15, rolling back to 1. The terminating endpoints use the packet sequence number to:

- 1) determine the first packet of a speech burst; and
- 2) to detect packet loss. The determination of the first packet is useful for delay equalization and may be needed for some speech coding algorithms such as Recommendation G.728. Delay equalization is discussed in the next subclause.

I.5 Delay equalization

Delays in packet communication consist of two components: a fixed delay and a variable delay [1]. The fixed delay arises from signal propagation on the transmission links, and from fixed processing delays in the originating and terminating endpoints and within the network. The effects of variations in the propagation delay for a given path are assumed to be negligible.

For speech packetization, the fixed processing delays consist of the following components:

- 1) packetization delay during which the speech samples are buffered for further processing;
- 2) if digital speech interpolation is used to remove silent intervals, the hang-over time of the speech detector [2];
- 3) the end-to-end algorithmic delay due to the encoding and decoding of speech – this delay depends on the coding scheme; for example, it is 125 μ s for PCM, 250 μ s for the Adaptive Differential Pulse Coded Modulation (ADPCM) algorithms of Recommendations G.726 and G.727 while it is less than 2 ms for the Low-Delay Code Excited Linear Predictor (LD-CELP) algorithm of Recommendation G.728;
- 4) any added delay at the terminating end to mask the timing jitter resulting from the variability in the delay – this added delay is denoted as build-out.

Variable delays result primarily from the queueing and processing of packets. They depend on the characteristics of the route of each packet: the number of hops (nodes), the type and speed of each link and the traffic intensity.

Speech traffic requires low and uniform delay. Recommendation G. 114 (1993) discusses the effect of end-to-end delays on the quality of a conversation. Delay variations that may, be acceptable for digital data transmission, usually affect the gaps between words and syllables and are troublesome for conversational speech. Available data suggest that variation in the interval between speech bursts should be less than 200 ms to avoid subjective degradation of speech quality. In applications such as videotelephony where both audio and video information are transmitted and must remain synchronized, the effect of the variable delay on this synchronization must also be taken into account.

There are several methods to mask the variability of the delay in the network. These techniques include:

- 1) blind delay;
- 2) absolute time stamp; and
- 3) relative time stamp.

The effect of all these methods is to increase the effective end-to-end buffering delay and, therefore, the total end-to-end delay.

I.5.1 Blind delay

In the blind delay method, a fixed buffer delay is always added at the terminating end at the first packet of a speech burst. This delay corresponds to the maximum variable delay expected.

The advantage of the blind delay scheme is its simplicity which makes it a good candidate when the transmission speech is such that the variable delays are of the order of a fraction of a millisecond (e.g. local area networks or broadband networks at 150 Mbit/s). In these cases, a fixed build-out delay on the order of 10 ms will be adequate to eliminate end-to-end delay jitter [3].

Over long haul connections in the PSTN, the scheme may require too large a delay that the total end-to-end delay would exceed the performance limits for network delays specified in Recommendation G. 114. For example, if the first packet has already experienced the worst case delay variation, the total variable delay to be added will be twice the worst case value [1]. Because this approach does not mask the delay variability completely, the gaps between the words may be varying which may degrade the subjective quality of speech. Voiceband traffic, included demodulated facsimile, may be perturbed; total delays larger than 500 ms cause premature disconnections of facsimile calls, especially when echo is present [4].

I.5.2 Absolute time stamp

This is the method used in datagram networks. The packet header includes a field for a time stamp that represents real time. The time stamp has a resolution sufficient to allow accurate detection of packet jitter, and to cover the worst case transit time of a packet across the network. Thus, packets that arrive out of sequence may be correctly ordered and buffered at the receiver using time stamp information [5]. This datagram scheme also requires clock synchronization between the transmitter and receiver so that the delay of each incoming packet can be compared to the previous ones, assuming that the network delay is fixed.