PM-1 3 1/2"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISIONSM RESOLUTION TARGET .3

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Yours truly, Votre dévoué

Yours truly, Votre dévoué

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Compressed Voice in Integrated Services Frame Relay Networks: Voice Synchronization and Congestion Control

by

Rukhsana Ansari, B. Eng.

Osmania University, INDIA

A thesis submitted to the
Faculty of Graduate Studies and Research
in partial fulfillment of
the requirements for the degree of
Master of Engineering

Ottawa-Carleton Institute of Electrical Engineering
Faculty of Engineering
Department of Systems and Computer Engineering

Carleton University

November 1, 1995

Ottawa, Ontario

© copyright

1995, R. Ansari

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file / Votre référence

Our file / Notre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-612-08972-X

Name BUKHSANA ANSARI

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

COMPRESSED VOICE IN INTEGRATED SERVICES FRAME

SUBJECT TERM

PC-LAY NETWORKS: VOICE SYNCHRONIZATION AND CONGESTION CONTROL

U-M-I

SUBJECT CODE

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS

Architecture	0729
Art History	0377
Cinema	0900
Dance	0378
Fine Arts	0357
Information Science	0723
Journalism	0391
Library Science	0399
Mass Communications	0708
Music	0413
Speech Communication	0459
Theater	0465

EDUCATION

General	0515
Administration	0514
Adult and Continuing	0516
Agricultural	0517
Art	0273
Bilingual and Multicultural	0282
Business	0688
Community College	0275
Curriculum and Instruction	0727
Early Childhood	0518
Elementary	0524
Finance	0277
Guidance and Counseling	0519
Health	0680
Higher	0745
History of	0520
Home Economics	0278
Industrial	0521
Language and Literature	0279
Mathematics	0280
Music	0522
Philosophy of	0998
Physical	0523

Psychology	0525
Reading	0535
Religious	0527
Sciences	0714
Secondary	0533
Social Sciences	0534
Sociology of	0340
Special	0529
Teacher Training	0530
Technology	0710
Tests and Measurements	0288
Vocational	0747

LANGUAGE, LITERATURE AND LINGUISTICS

Language	
General	0679
Ancient	0289
Linguistics	0290
Modern	0291
Literature	
General	0401
Classical	0294
Comparative	0295
Medieval	0297
Modern	0298
African	0316
American	0591
Asian	0305
Canadian (English)	0352
Canadian (French)	0355
English	0593
Germanic	0311
Latin American	0312
Middle Eastern	0315
Romance	0313
Slavic and East European	0314

PHILOSOPHY, RELIGION AND THEOLOGY

Philosophy	0422
Religion	
General	0318
Biblical Studies	0321
Clergy	0319
History of	0320
Philosophy of	0322
Theology	0469

SOCIAL SCIENCES

American Studies	0323
Anthropology	
Archaeology	0324
Cultural	0326
Physical	0327
Business Administration	
General	0310
Accounting	0272
Banking	0770
Management	0454
Marketing	0338
Canadian Studies	0385
Economics	
General	0501
Agricultural	0503
Commerce Business	0505
Finance	0508
History	0509
Labor	0510
Theory	0511
Folklore	0358
Geography	0366
Gerontology	0351
History	
General	0578

Ancient	0579
Medieval	0581
Modern	0582
Black	0328
African	0331
Asia, Australia and Oceania	0332
Canadian	0334
European	0335
Latin American	0336
Middle Eastern	0333
United States	0337
History of Science	0585
Law	0398
Political Science	
General	0615
International Law and Relations	0616
Public Administration	0617
Recreation	0814
Social Work	0452
Sociology	
General	0626
Criminology and Penology	0627
Demography	0938
Ethnic and Racial Studies	0631
Individual and Family Studies	0628
Industrial and Labor Relations	0629
Public and Social Welfare	0630
Social Structure and Development	0700
Theory and Methods	0344
Transportation	0709
Urban and Regional Planning	0999
Women's Studies	0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES

Agriculture	
General	0473
Agronomy	0285
Animal Culture and Nutrition	0475
Animal Pathology	0476
Food Science and Technology	0359
Forestry and Wildlife	0478
Plant Culture	0479
Plant Pathology	0480
Plant Physiology	0817
Range Management	0777
Wood Technology	0746
Biology	
General	0306
Anatomy	0287
Biostatistics	0308
Botany	0309
Cell	0379
Ecology	0229
Entomology	0353
Genetics	0369
Limnology	0793
Microbiology	0410
Molecular	0307
Neuroscience	0317
Oceanography	0416
Physiology	0433
Radiation	0821
Veterinary Science	0778
Zoology	0472
Biophysics	
General	0786
Medical	0760

EARTH SCIENCES

Biogeochemistry	0425
Geochemistry	0996

Geodesy	0370
Geology	0372
Geophysics	0373
Hydrology	0388
Mineralogy	0411
Paleobotany	0345
Paleocology	0426
Paleontology	0418
Paleozoology	0985
Polynology	0427
Physical Geography	0368
Physical Oceanography	0415

HEALTH AND ENVIRONMENTAL SCIENCES

Environmental Sciences	0768
Health Sciences	
General	0566
Audiology	0300
Chemotherapy	0992
Dentistry	0567
Education	0350
Hospital Management	0769
Human Development	0758
Immunology	0982
Medicine and Surgery	0564
Mental Health	0347
Nursing	0569
Nutrition	0570
Obstetrics and Gynecology	0380
Occupational Health and Therapy	0354
Ophthalmology	0381
Pathology	0571
Pharmacology	0419
Pharmacy	0572
Physical Therapy	0382
Public Health	0573
Radiology	0574
Recreation	0575

Speech Pathology	0460
Toxicology	0383
Home Economics	0386

PHYSICAL SCIENCES

Pure Sciences	
Chemistry	
General	0485
Agricultural	0749
Analytical	0486
Biochemistry	0487
Inorganic	0488
Nuclear	0738
Organic	0490
Pharmaceutical	0491
Physical	0494
Polymer	0495
Radiation	0754
Mathematics	0405
Physics	
General	0605
Acoustics	0986
Astronomy and Astrophysics	0606
Atmospheric Science	0608
Atomic	0748
Electronics and Electricity	0607
Elementary Particles and High Energy	0798
Fluid and Plasma	0759
Molecular	0609
Nuclear	0610
Optics	0752
Radiation	0756
Solid State	0611
Statistics	0463
Applied Sciences	
Applied Mechanics	0346
Computer Science	0984

Engineering	
General	0537
Aerospace	0538
Agricultural	0539
Automotive	0540
Biomedical	0541
Chemical	0542
Civil	0543
Electronics and Electrical	0544
Heat and Thermodynamics	0348
Hydraulic	0545
Industrial	0546
Marine	0547
Materials Science	0794
Mechanical	0548
Metallurgy	0743
Mining	0551
Nuclear	0552
Packaging	0549
Petroleum	0765
Sanitary and Municipal	0554
System Science	0790
Geotechnology	0428
Operations Research	0796
Plastics Technology	0795
Textile Technology	0994

PSYCHOLOGY

General	0621
Behavioral	0384
Clinical	0622
Developmental	0620
Experimental	0623
Industrial	0624
Personality	0625
Physiological	0989
Psychobiology	0349
Psychometrics	0632
Social	0451

The undersigned recommend to the Faculty of Graduate Studies and Research
acceptance of the thesis

**Compressed Voice in Integrated Services Frame Relay Networks:
Voice Synchronization and Congestion Control**

submitted by Rukhsana Ansari, B. Eng.

in partial fulfillment of the requirements for

the degree of Master of Engineering

Thesis Supervisor

Chair, Department of Systems and
Computer Engineering

Carleton University

Abstract

The simplicity of the FR protocol, designed primarily for data communications, makes it difficult to apply existing voice synchronization algorithms requiring network layer processing for voice transmission. This thesis evaluates the performance of an *Adaptive Buildout Delay* mechanism for compressed voice playout in the context of variable delays caused by a mixed voice-data FR network. Segmentation of large LAN packets is recommended at the user-network interface in order to limit delay spikes affecting the target QoS requirements of conversational voice.

A modified adaptive playout algorithm is proposed for improved adaptation, in particular for rapid detection and response to burst losses. A comparative performance evaluation, based on quantitative simulation results and validated by ad hoc qualitative subjective tests using the reference VSELP speech coder, demonstrates significant improvement in voice quality using the modified algorithm.

An adaptive rate control mechanism is proposed for FR congestion control. VBR compressed voice sources at the network periphery dynamically adapt transmission rates in response to explicit congestion notifications from the FR network. The effectiveness of the control strategy is evaluated based on objective measures and is accompanied by informal subjective tests using the reference I-CELP speech coder to confirm the quantitative results.

Acknowledgments

I would like to express my sincere thanks to my thesis supervisor, Prof. A. R. Kaye for helping me achieve a focused and organized approach to research and for his guidance throughout the course of this thesis. I would also like to extend my thanks to Wilfrid P. Leblanc for providing me with the reference voice coders and for taking the time to answer my numerous queries.

I wish to profusely thank my husband, Shad Ansari, for his enduring patience, support and encouragement during this research. Special thanks are due to my family, especially to my mother and brother for their constant support.

I gratefully acknowledge the funding provided by the Telecommunications Research Institute of Ontario.

To my mother

Table of Contents

Acceptance Sheet	ii
Abstract	iii
Acknowledgments.	iv
Table of Contents	vi
List of Tables	x
List of Figures	xi
List of Acronyms	xiv
List of Symbols	xvii
 1 Introduction	 1
1.1 Background	1
1.2 Previous Research.	3
1.3 Motivation and Objectives of this Thesis.	5
1.4 Thesis Organization.	6
 2 Frame Relay and Voice Characteristics	 9
2.1 Frame Relay Bearer Service	9
2.1.1 Frame Format	10
2.1.2 Congestion Management	11
2.2 Voice Characteristics	13
2.2.1 End-to-End Delay and Loss Requirements	14
2.2.2 Delay Composition	16

2.2.3	Voice Synchronization and Random Delay Compensation	17
2.2.4	Reference Speech Coders: VSELP and I-CELP	19
2.3	Adaptation Layer and Priority Service for Compressed Voice	21
2.3.1	Voice Frame Relay Adaptation Layer.	22
2.3.2	Priority Service for Voice	24
3	Voice Synchronization:	
	Adaptive Buildout Delay for Voice in Inter-LAN Traffic	27
3.1	Introduction.	27
3.2	Adaptive Buildout Delay Mechanism.	27
3.2.1	Notation and Terminology	28
3.2.2	Holding-Time Relation to Buildout Delay	29
3.2.3	Holding-Time Adaptation Algorithm	32
3.2.4	Determination of Target Mean Holding Time	35
3.2.5	Adjustments during Speech	35
3.3	Network Simulation.	36
3.3.1	Network Topology	36
3.3.2	Reference Connection.	37
3.3.3	Inter-LAN Data Traffic	38
3.3.4	Quality of Service Requirements and Determination of Target Buildout Delay	40
3.3.4.1	Confidence Intervals of Simulation Results.	41
3.3.5	Simulation Parameters	41
3.3.6	Assumptions.	42
3.3.7	Subjective Test Description	42
3.4	Adaptive Buildout Delay Performance with Delay Spikes	43

3.5	Adaptive Playout with Reduced Delay Spikes	46
3.5.1	Loss Rate Analysis	51
3.5.1.1	Impact of Correlated Losses on Subjective Quality of Reconstructed Speech	54
3.6	Requirement for Modification to the Adaptive Buildout Delay Mechanism .	54
3.7	Summary	55
4	Modified Adaptive Buildout Delay	56
4.1	Introduction	56
4.2	Modified Adaptive Buildout Delay Mechanism	56
4.2.1	Extended Notation Set	57
4.2.2	Modified Holding-Time Adaptation Algorithm.	58
4.2.3	Procedures Invoked by the Modified Playout Algorithm.	60
4.2.4	Inclusion of New Parameters and their Correspondence to Proposed Modifications	63
4.2.5	Improved Voice Performance using Modified Adaptive Buildout Delay.	66
4.3	Multiple Voice Connections.	67
4.3.1	Network Delay Characteristics	70
4.3.2	Comparative Performance of the Original and Modified Adaptive Playout Mechanisms for Multiple Voice Calls	71
4.3.3	Factorial Analysis.	75
4.3.4	Modified Adaptive Buildout Delay with Optimized Holding-Time Adjustments	77
4.4	Comparative Performance of Modified Adaptive Buildout Delay and Packetized Voice Protocol	78

4.5	Re-evaluation of Maximum Permissible Data Segment Size	82
4.6	Summary	89
5	Congestion Control of VBR Compressed Voice	91
5.1	Network Simulation and Reference Connection	91
5.1.1	Voice Buffer Size	93
5.2	Simulation of Network Congestion	94
5.2.1	Simulation Parameters	95
5.3	Voice Performance in Congested Network without Control	96
5.4	Congestion Control	99
5.4.1	Proposed Adaptive Rate Control	99
5.5	Voice Performance with Adaptive Rate Control	101
5.5.1	Effect of varying BECN_clear_threshold.	103
5.5.2	Impact of Smaller Congestion Threshold.	107
5.6	Voice Performance under Varied Congestion Levels	110
5.7	Summary	112
6	Conclusions and Recommendations for Further Study	114
6.1	Conclusions	114
6.2	Recommendations for Further Study	119
	Appendix: Additional Details of the Adaptive Buildout Delay Algorithm	121
	References	123

List of Tables

Table 2.1:	TSM Specification for Received Voice frames.	21
Table 3.1:	Voice Performance without Data Segmentation	43
Table 3.2:	Voice Performance with 500 byte Data Segments	47
Table 4.1:	Factors and Levels selected for Factorial Design	76
Table 4.2:	Voice Performance with 5% 1500 byte Data Packets	87
Table 5.1:	Voice Performance in Congested FR Network.	97
Table 5.2:	Voice Performance with Adaptive Rate Control	102
Table 5.3:	Mean Voice Duration at Transmission Rates for varying BECN_clear_threshold.	104
Table 5.4:	Voice Performance with BECN_clear_threshold = 77 frames.	106
Table 5.5:	Voice Performance for Smaller Values of Congestion Threshold	108
Table 5.6:	Mean Voice Duration at Transmission Rates for varying Link2 Load. .	111

List of Figures

Figure 2.1	Frame Relay Frame Format	11
Figure 2.2	Explicit Congestion Notification	12
Figure 2.3	Frame Relay Network Configuration	23
Figure 2.4	Voice Adaptation Architecture	23
Figure 2.5	Enhanced Voice Frame Format	24
Figure 2.6	Priority Implementation at Access Link	26
Figure 3.1	Frame Relay Network Simulation	36
Figure 3.2	Histogram of Ethernet Packet Size	39
Figure 3.3	Occurrence of Delay Spikes with Unsegmented Data Frames	44
Figure 3.4	CDF of (a) Network Delays (b) Holding Times of Voice Packets with Unsegmented Data Frames	45
Figure 3.5	Voice Packet Loss with Unsegmented Data Frames, measured over 2 s time intervals	46
Figure 3.6	1500 byte Data Packet Segmentation	48
Figure 3.7	Mean Voice Network Delay vs Destination Link Load	49
Figure 3.8	Performance of Adaptive Buildout Delay (a) Mean Loss Rate (b) Mean Variable End-to-End Delay	49
Figure 3.9	Reduced Delay Spikes due to Data Frame Segmentation.	50
Figure 3.10	Comparative CDF of (a) Network Delays (b) Holding Times of Voice Packets	51
Figure 3.11	Voice Packet Loss with Segmented Data Frames, measured over 2 s time intervals	53

Figure 4.1	Immediate Holding-Time Adaptation upon Detection of Losses	64
Figure 4.2	Histogram of Negative Holding Times of Delayed Voice Packets	65
Figure 4.3	CDF of Number of Frames separating Successive Voice Losses	67
Figure 4.4	Comparative Performance of Adaptive Playout Mechanisms (a) Mean Loss Rate (b) Mean Variable End-to-End Delay	68
Figure 4.5	Voice Packet Loss with Modified Algorithm, measured over 2 s time intervals	69
Figure 4.6	Mean Voice Network Delay vs Number of Voice Calls.	71
Figure 4.7	Comparative Performance of Adaptive Playout Mechanisms for Multiple Voice Calls (a) Mean Loss Rate (b) Mean Variable End-to-End Delay	72
Figure 4.8	Voice Packet Loss for Multiple Voice Calls, measured over 2 s time intervals (a) Original Algorithm (b) Modified Algorithm (c) Modified Algorithm with Optimized Holding-Time Adjustments	73
Figure 4.9	Voice Packet Loss for Multiple Voice Calls, measured over 2 s time intervals (a) Original Algorithm (b) Modified Algorithm (c) Modified Algorithm with Optimized Holding-Time Adjustments	74
Figure 4.10	CDF of Holding Times of Delayed Voice Packets for Multiple Voice Calls	75
Figure 4.11	90% Confidence Intervals for Primary and Secondary Effects of Loss Rate	76
Figure 4.12	Comparative Performance of Modified Adaptive Buildout Delay and Packetized Voice Protocol (a) Mean Loss Rate (b) Mean Variable End-to- End Delay	80
Figure 4.13	Voice Packet Loss measured over 2 s time intervals (a) & (c) Time Stamp Method of PVP (b) & (d) Modified Adaptive Buildout Delay	81
Figure 4.14	Mean Voice Network Delay vs Data Segment Size.	83

Figure 4.15	Modified Adaptive Buildout Delay Performance vs Data Segment Size (a) Mean Loss Rate (b) Mean Variable End-to-End Delay	84
Figure 4.16	Modified Adaptive Buildout Delay Performance with 5% 1500 byte Data Packets (a) Mean Loss Rate (b) Mean Variable End-to-End Delay . . .	86
Figure 4.17	Voice Packet Loss with 5% 1500 byte Data Packets, measured over 2 s time intervals	87
Figure 4.18	CDF of Holding Times of Delayed Packets with 5% 1500 byte Data Packets	88
Figure 5.1	Modified Frame Relay Network Simulation	92
Figure 5.2	Full Duplex Low-Speed Access Link Queueing Model	92
Figure 5.3	Average Voice Loss Rate with Network Congestion vs Time	97
Figure 5.4	Voice Packet Loss with Network Congestion, measured over 2 s time intervals	98
Figure 5.5	Generic State Diagram	100
Figure 5.6	Voice Packet Loss with Adaptive Rate Control, measured over 2s time intervals	103
Figure 5.7	Mean Voice Network Delay vs BECN_clear_threshold	104
Figure 5.8	Mean Voice Loss Rate vs BECN_clear_threshold	105
Figure 5.9	Mean Voice Variable End-to-End Delay vs BECN_clear_threshold	105
Figure 5.10	Average Voice Loss Rate with Adaptive Rate Control vs Time	107
Figure 5.11	Voice Packet Loss with Adaptive Rate Control and BECN_clear_threshold=77 frames, measured over 2s time intervals	108
Figure 5.12	Average Voice Loss Rate with Adaptive Rate Control and Smaller Values of Congestion Threshold vs Time	109
Figure 5.13	Fraction of Time spent at the Lower Transmission Rate vs Congestion Threshold	109
Figure 5.14	Fraction of Time spent at 16 kb/s vs Decreasing Link2 Load	112

List of Acronyms

ABD	Adaptive Buildout Delay
ADPCM	Adaptive Differential Pulse Code Modulation
ANSI	American National Standards Institute
ATM	Asynchronous Transfer Mode
BECN	Backward Explicit Congestion Notification
BER	Bit Error Rate
B-ISDN	Broadband Integrated Services Digital Network
BT	Beginning (First) frame of talkspurt
CBR	Constant Bit Rate
CCITT	Consultative Committee for International Telegraphy and Telephony
CELP	Code Excited Linear Prediction
CLLM	Consolidated Link Layer Management
CODEC	Coder/Decoder
C-Plane	Control Plane
CPE	Customer Premises Equipment
CPU	Central Processing Unit
C/R	Command/Response
DE	Discard Eligibility
DLCI	Data Link Connection Identifier
FAD	Frame Assembler Disassembler
FCS	Frame Check Sequence
FECN	Forward Explicit Congestion Notification

FR	Frame Relay
HDLC	High-Level Data Link Control
I-CELP	Integer Code Excited Linear Prediction
IP	Internet Protocol
ISDN	Integrated Services Digital Network
LAN	Local Area Network
LAPD	Link Access Protocol-D Channel
LT	Last frame of talkspurt
MBL	Mean burst length
ND	Network Disk (Protocol)
NFS	Network File System
OSI	Open Systems Interconnection
PBX	Private Branch Exchange
PCM	Pulse Code Modulation
PDU	Protocol Data Unit
PVC	Permanent Virtual Circuit
PVP	Packetized Voice Protocol
QoS	Quality of Service
RQ	Receiver Queue
SEQ	Sequence Number
SVC	Switched Virtual Circuit
T1 (DS1)	24 voice-grade channels, each of 64 kb/s, multiplexed to give 1.544 Mps aggregate connectivity for point-to-point services.
TCP	Transmission Control Protocol
TS	Time Stamp
TSM	Time Scale Modification
TQ	Transmit Queue
U-Plane	User Plane

VBR	Variable Bit Rate
VFRAL	Voice Frame Relay Adaptation Layer
VSELP	Vector Sum Excited Linear Prediction

List of Symbols

c	codec delay
d_f	fixed portion of voice delay
f	framing delay
$\bar{h}_a$	average holding time
h_s	cumulative sum used in computing average holding time
$\bar{h}_T$	mean target holding time
l_s	link speed in kb/s
p	propagation delay
q	network delay
$\bar{q}$	mean network delay
$q(n)$	network delay of the n th frame
$\bar{q}_{max}$	maximum mean network delay
s	switching delay
t	transmission time over low-speed access links
t_{ism}	time increment used to make dynamic adjustment to holding time
w	queueing delay/ waiting time in the transmission queues
A	averaging interval
B	buildout delay
$B(n)$	buildout delay of the n th frame
B_T	target buildout delay/ target variable end-to-end delay
B_a	average buildout delay/ mean variable end-to-end delay
D	end-to-end delay

D_T	target end-to-end delay
E	error function
$H(n)$	Actual holding time of n th frame
$I(n,k)$	source-indicated playout interval between k th and n th frame, $n > k$
$L(n)$	sequence number of n th frame
$L(n-1)$	sequence number of the $(n-1)$ th frame
L_a	mean loss rate
MBL	mean burst length in bits
S_l	Segment length
$T_a(n)$	arrival time of the n th frame at the playout buffer
$T_p(n)$	playout time of the n th frame at the playout buffer
λ_d	mean arrival rate of data burst in bursts/s
λ_v	mean arrival rate of voice in kb/s
ρ	load of the link
ε	minimum deviation of $\bar{h}_a$ for which dynamic adjustment is made
θ_1	congestion threshold
θ_2	absolute threshold

Chapter 1

Introduction

1.1 Background

Evolution of the Broadband-Integrated Services Digital Network (B-ISDN), fuelled by the distributed nature of network processing and an increased demand for advanced computer applications, has resulted in a redefinition of traditional transmission concepts. Broadband technologies now provide a common switching/transport fabric for voice, data and video applications which were historically transmitted over distinct communication channels. The notion of B-ISDN as a technology that will consolidate and augment existing networks, resulting in a unified network infrastructure has lead to the revolutionary vision of a *global information highway*.

Interwoven with the idea of broadband networks is the concept of “fast packet switching”. Intelligent end-user systems, reliable high-speed digital communication and transmission facilities have been the compelling force behind fast packet networks. Frame Relay and Cell Relay are two divisions of fast packet technologies. Distinguishing features of both these technologies include statistical multiplexing capabilities and streamlined transport leading to lower transit delay and higher throughput. Within the overall context of the broadband hierarchy, cell relay and the associated Asynchronous Transfer Mode (ATM) protocol will be used to develop an integrated environment with multiple access options for varied information types. Frame Relay, in addition to providing cost-effective LAN connectivity, is expected to become a key access option for cell-based transport networks, resulting in seamless interworking over homogeneous and

heterogeneous boundaries [11].

Frame Relay is an ISDN packet-mode bearer service primarily intended for the transport of variable-length, protocol-oriented data units over a connection-oriented path [1]. Originally specified for access speeds of up to 2 Mb/s, viability of the technology has been demonstrated even at 50 Mb/s [2]. Based on the core aspects of LAPD, the FR protocol relies on increased intelligence of present-day network endpoints to effectively eliminate protocol processing done above layer 2 of the OSI model [3, 4, 5]. Characterized by a streamlined protocol, FR networks offer lower latency and higher throughput, compared to packet-switching networks using traditional techniques such as X.25.

The inherent benefits provided by FR in terms of protocol simplicity, efficient bandwidth sharing, lower response time, increased interoperability in a multiprotocol data environment and reduced inter-networking costs have all contributed to make frame relay a better and more cost-effective alternative for LAN interconnections between geographically dispersed sites, as opposed to traditional options such as modems and dial-up phones, point-to-point leased lines, T1 multiplexers, and X.25 services [6, 7, 8]. Although, early commercial impetus was provided by corporate networks, current worldwide deployment of FR networks is significant with over three dozen service providers in North America, more than 18 operating internationally and approximately 90 vendors on the market for FR switch and access equipment [8, 9, 10].

Initially viewed as an interim technology, industry experts envision FR as playing a compatible and complementary role to ATM in the overall context of the information highway. Due in part to the flexibility and ease of software upgrade, FR is expected to be widely used as an interface option for ATM providing seamless migration. Industry standards for Frame Relay/ATM Interworking have been defined and work is underway for standards on service interoperability, supporting FR-to-ATM protocol conversion [11, 12].

Packet switching has been proposed as an effective technology for the integration of

voice and data in a single network [13 - 19]. This technique essentially involves transmission of encoded, packetized speech over a packet switching network, followed by decoding at the receiver end. The principal challenge in supporting real-time voice applications over a packet-switching network is the need to obtain synchronous playout of voice packets in the face of varying network delays. Compensation of variable delays is typically achieved by buffering received packets, and voice playout algorithms differ with respect to the techniques used to estimate the delay to be introduced prior to playout. Packet voice synchronization is a significant problem that has received considerable attention in the literature in the past [13 - 20]. Recent years have witnessed renewed interest in this area [21, 22] with demonstrations that limited amounts of interactive audio can be supported by today's Internet [23].

The objective of voice transmission over FR networks is analogous to the concept of voice-data integration in traditional packet-switching networks, and has similar objectives; primarily, efficient use of leased bandwidth by reallocating it between data and voice on a dynamic basis. Access equipment under user control in customer premises can be used to multiplex traffic according to user preferences, exploiting the bursty nature of data and voice to achieve a reduction in transmission bandwidth.

1.2 Previous Research

Optimized for handling data traffic, the FR protocol was not designed to support real-time, delay-sensitive traffic such as voice. Previously proposed playout mechanisms involve incorporating additional complexity in the FR protocol. The CCITT recommendation G.764, defines a Packetized Voice Protocol (PVP) which outlines use of a time-stamp to record cumulative variable delays experienced by a packet in traversing the network [24]. A similar approach by Montgomery, defined as the *Added Variable Delay* method is outlined in [20]. The use of a time-stamp requires OSI layer 3 processing, functionality which has been eliminated in the FR protocol. Barberis and Pazzaglia suggest

synchronizing the emitter and receiver clock, a requirement that is not provided for in FR networks and most other packet-switching networks. Ramjee et al discuss playout mechanisms [22] all of which allow the receiver to explicitly determine the production time of a packet by use of an *absolute timing* method proposed in [20]. Absolute timing requires the availability or provision of an absolute time reference for synchronization. One can therefore conclude that existing voice synchronization and reconstruction algorithms are inappropriate in the context of FR networks as their applicability is in direct violation of the protocol's intrinsic simplicity.

Unsuitability of current algorithms has lead to exploratory work in adapting FR networks for transmission of delay-sensitive traffic such as voice and video. Dong, Kaye and Mahmoud demonstrated that priority for voice is highly desirable to achieve conversational voice quality when low-speed access links are used for integrated service traffic [25, 26]. A first approach to voice synchronization suitable for FR networks, namely, *Adaptive Buildout Delay*, was also introduced by the same authors. Additional research by S. Zhang investigated use of CLLM as an effective mechanism for congestion control in mixed voice-data FR networks [27, 28]. Early industrial research includes [29], describing a simulation tool and methodology for performance evaluation of multi-service FR networks, and [30] which is concerned with the encapsulation and signalling aspects of voice for transmission over FR networks. FR applications involving video transmission and the feasibility of using network congestion notification messages for rate reduction of video sources have been investigated by C.M. Sharon et al in [31, 32].

Earlier research on voice synchronization in [25, 26] made assumptions with respect to the short size of data frames and network load stabilization, leading to rapid adaptation in a short transient interval and fast convergence towards a steady-state. The assumptions were optimistic when viewed in the context of FR networks primarily used for LAN interconnections, in particular due to the delay variance attributed to the bimodal distribution of ethernet packet size [45 - 49]. Furthermore, performance evaluation of the

algorithm was based only on objective simulation results and therefore an inadequate emphasis on issues affecting quality of reconstructed speech, in particular the impact of correlated losses and the holding-time adjustment step size on the subjective quality of speech. The congestion control mechanism proposed by S. Zhang is based on CLLM messages that have been defined by CCITT and ANSI [5] but currently aren't implemented in FR networks [28]. Effectiveness of the control strategy was validated by simulation results alone.

1.3 Motivation and Objectives of this Thesis

The motivation for this thesis was provided by the observation that so far there has been no published research pertaining to voice transmission in FR networks with emphasis on the subjective quality of speech using voice codecs. Available literature illustrates that performance evaluation was limited to quantitative analysis. Qualitative assessment, based on subjective tests, can assist in identifying problems that might have been neglected in traditional studies, such as audible faults caused by bursty losses.

This thesis presents a consolidated study that is based on quantitative evaluation and supplemented by informal qualitative evaluation of reconstructed voice quality to identify audible faults in played-out speech, interpret objective results and direct design of the adaptive playout mechanism. The study includes:

- Re-evaluation of voice synchronization in the presence of inter-LAN traffic.
- Effect of data frame size on loss rate and end-to-end delay of conversational voice.
- Impact of correlated losses on voice quality.
- Feasibility of using network congestion notifications for congestion control.

The specific objectives of this thesis are to:

- i Present a parametric version of the original *Adaptive Buildout Delay* mechanism as a voice synchronization/playout algorithm for use in a mixed voice-data Frame Relay environment.
- ii Develop a Frame Relay network simulator in a C/UNIX environment to study the performance of the playout algorithm in the context of variable delays caused by inter-LAN Ethernet traffic.
- iii Develop a test environment to permit ad hoc subjective evaluation of reconstructed speech using the VSELP speech coder, a standard low bit rate voice codec producing high quality speech at 8 kb/s.
- iv Propose a modified adaptive playout algorithm for enhanced speech quality after identification of parameters affecting efficiency of the original algorithm.
- v Evaluate the performance of the modified playout algorithm by quantitative simulations and qualitative subjective tests.
- vi Explore the feasibility of implementing a congestion control strategy for variable-bit-rate (VBR) voice sources using explicit network congestion notifications.
- vii Validate the effectiveness of the congestion control mechanism by quantitative simulation results as well as subjective tests using a standard VBR voice coder. Since VSELP is a constant-bit-rate (CBR) speech coder, subjective analysis for the congestion control study uses I-CELP, a VBR speech coder having a dual coding rate capability of 16/8 kb/s.

1.4 Thesis Organization

The remaining chapters of the thesis are organized as follows:

Chapter 2: Reviews fundamentals of Frame Relay bearer service, voice characteristics, voice synchronization, and reference speech coders. Brief descriptions follow

of a voice adaptation layer and priority service for transmission of compressed voice.

Chapter 3: Presents the original *Adaptive Buildout Delay* algorithm as a suitable voice synchronization scheme for FR networks which have no provision for measurement of random delay experienced by packets since there is no common time reference at the transmitter and receiver. The FR network simulation topology, and data traffic model used are described. This chapter investigates the performance of the original adaptive playout algorithm in the presence of inter-LAN ethernet traffic and proposes segmentation of large LAN frames at the source FAD to reduce the impact of delay spikes on QoS requirements of conversational voice. Subsequent voice performance with segmented data frames is studied at various link loads; performance results indicating the need for a modified approach to holding time adaptation.

Chapter 4: A modified playout algorithm is proposed for improved adaptation to variability in network delays, and faster response to delay spikes. A comparative performance evaluation of the original and modified algorithms is presented, inclusive of quantitative results and subjective analysis of reconstructed speech using the reference VSELP coder operating at 8 kb/s.

Performance of the original and modified Adaptive Buildout Delay mechanisms is investigated in the presence of multiple voice calls. A simple factorial analysis is used to evaluate the impact of step-size of the dynamic adjustment on the loss rate and an increased adjustment step is suggested for multiple voice connections.

Other topics covered in this chapter include comparison of modified Adaptive Buildout Delay with the Packetized Voice Protocol, reevaluation of the

maximum permissible data frame segment size and the possibility of transmitting unsegmented frames in light of the improved voice performance using the modified algorithm.

Chapter 5: Investigates the feasibility of controlling FR network congestion by using a closed-loop, feedback based congestion control mechanism. Congestion notifications received from the network, are used for source rate control of variable-bit-rate voice sources. Effectiveness of the congestion control mechanism is evaluated based on simulation results and subjective tests using ICELP, a VBR speech coder operating at 16/8 kb/s. This chapter also studies the effect of various settings of the thresholds incorporated in the adaptive rate control mechanism, on voice performance. Also investigated is voice performance under varying congestion levels.

Chapter 6: Conclusions and Recommendations for further study.

Chapter 2

Frame Relay and Voice Characteristics

2.1 Frame Relay Bearer Service

Frame relay is a high speed, packet-based technology, designed to support transport of variable-length data units over assigned virtual connections. The FR protocol can be viewed as a streamlined version of X.25. With widespread deployment of high-quality, fiber-based communication infrastructure, the error-correction and retransmission capabilities of X.25 equate to redundant network functionality. Streamlining of the FR protocol is accomplished by virtually eliminating all of X.25 network layer (OSI layer 3) functionality, adding a statistical multiplexing capability via individually addressed frames to the data link layer (OSI layer 2), and reducing link layer functionality by removing error correction and retransmission capabilities. “Core functions” performed by the frame sub-layer are based on a subset of the LAPD protocol, defined as follows [5]:

- frame delimiting, alignment, and transparency provided by the use of HDLC flags and zero bit insertion/extraction;
- frame multiplexing/demultiplexing using the address field;
- inspection of the frame to ensure that it consists of an integer number of octets prior to zero bit insertion or following zero bit extraction;
- inspection of the frame to ensure that it is not too long or too short;
- detection of (but not recovery from) transmission errors;
- congestion control functions.

FR was originally developed as an ISDN service, capable of supporting both permanent virtual circuits (PVCs) and switched virtual circuits (SVCs) with provision of a logically separate user plane (U-plane) and control plane (C-plane). Current implementations are based on a non-ISDN platform as well and do not support a separate C-plane for SVC management. Industry efforts in this area have been accelerated following ratification of the Implementation Agreement for FR SVCs by the FR forum [33]. PVCs establish a fixed path between network endpoints in accordance with network administrator specification. Network routing is achieved by means of a Data Link Connection Identifier (DLCI) specified in the frame header, shown in Figure 2.1, whose value defines a logical channel within the underlying physical channel. DLCIs have local significance only and their contents may be dynamically changed at FR nodes. Conceptually, each physical channel is partitioned into multiple PVCs, however, the entire physical bandwidth is available “on-demand” to any active PVC.

2.1.1 Frame Format

Prior to transmission over a FR network, frames must adhere to a frame format that is defined by the FR protocol and illustrated in Figure 2.1. All frames begin and end with a unique, 1 octet flag sequence of 01111110. The default address field length is 2 octets, but may optionally be extended up to 4 octets. The information field is user dependant, but a default maximum to be supported by networks is 262 octets. While a theoretical maximum is defined at 4096 bytes corresponding to maximum integrity of the FCS, higher vendor specific lengths are optional. Segmentation and reassembly of large frames can be accomplished in user equipment, as outlined in [34].

A 16-bit Frame Check Sequence (FCS) is included for error detection of address and information fields. The address field consists of a DLCI, which has a default length of 10 bits and is used to identify the logical channel within a physical channel. DLCIs are assigned on a hop by hop basis, and are examined by the network switches to route the

Figure 2.1 Frame Relay Frame Format

packet onto the appropriate output port. Address field extension is provided by the EA bit. When set to 0, the EA bit signals the presence of a following octet of the address field, and a 1 indicates the finality of the address field. The Command/Response (C/R) bit is application-specific and is not used by the DL-CORE protocol. The FECN, BECN and DE bits used for traffic management purposes in the event of congestion buildup in FR networks are described in greater detail in the following section.

2.1.2 Congestion Management

Congestion buildup in FR networks, as with other packet-switching networks, is a significant problem that has a profound impact on the Quality of service (QOS) parameters such as throughput, end-delay and frame loss. Network congestion typically occurs due to temporary overuse of resources such as bandwidth, memory or CPU. Equipment failure can also contribute to congestion. Congestion control, already needed in traditional networks, acquires a far more critical role in the case of FR network due to the performance objectives of the latter. The FR protocol has defined a set of measures that are to be implemented to achieve congestion control, both in terms of congestion avoidance and congestion recovery.

Congestion avoidance procedures include explicit congestion notifications that are to be used at the onset of congestion to minimize its negative effects on the network and

Figure 2.2 Explicit Congestion Notification

users. Upon detection of congestion, FR nodes are required to set congestion notification bits to alert the end-users of the situation. Forward explicit congestion notification (FECN) bit is set to “1” on frames travelling in the direction of the congesting frames, to indicate to the receiving end-system that the arriving frames have encountered a congested network. Similarly, the Backward explicit congestion notification bit (BECN) may be set to “1” in the traffic travelling in the direction opposite to the congesting frames, to notify the sender to reduce its inbound traffic to the destination. All FR nodes are required to transport the frames without resetting the congestion notification bits. Figure 2.2, adapted from [8], depicts the operation of BECN and FECN.

End users are expected to initiate rate-based strategies to throttle input traffic, in response to network congestion notifications, so as to alleviate congestion. Rate reduction is to continue by the traffic sources until congestion abates and the network indicates the condition by resetting the congestion notification bits. Graceful adaptation to network congestion may be achieved by reducing the input rate in discrete steps to minimize impact on subjective quality. In case of congestion persistence, the discard eligibility (DE) bit may be set by the user as indication to the network that the frame may be dropped. Therefore, in addition to monitoring resources, the network is expected to exercise a selective frame

discard strategy for calls that exceed their negotiated throughput, based on arrival of frames with DE bit set.

Backward congestion notification may alternately be provided by the generation and transport of consolidated link layer management messages (CLLMs) in the user-plane. The end-user response upon receipt of CLLM message is to invoke rate-based reduction techniques similar to those done when a BECN bit indication is received. Generation and transport of CLLM is optional [5] and currently deployed FR networks do not support this control strategy.

Implicit congestion notification occurs by the end-to-end protocol's detection of continued frame loss. User response to implicit notification is far more urgent compared to explicit notification because of immediate perception in the degradation of quality. If effective rate reduction by users in response to notifications proves to be inadequate in the case of severe congestion, the network cannot rely solely on the user's behaviour to control congestion and is not constrained to only discard frames with DE bit set to 1. All arriving frames may be discarded to allow return to normal conditions.

2.2 Voice Characteristics

Integration of voice and data services in FR networks presents technical difficulties arising from the conflicting Quality of Service (QoS) requirements for both these traffic types. The fundamental difference lies in the dissimilar sensitivity of voice and data to delay and loss requirements. The presence of packetized voice, particularly if it is highly compressed, changes the nature of traffic in a FR network because of the short length of the voice frames. Compressed voice transmission in a predominantly data environment necessitates a review of voice characteristics and associated ramifications in terms of synchronization and integration architecture. Review of relevant topics follows in subsequent sections.

2.2.1 End-to-End Delay and Loss Requirements

Traffic performance is usually characterized in terms of the mean end-to-end delay and loss rate. As mentioned earlier, voice has QoS requirements which differ from those of conventional data. Data is extremely loss-sensitive but can tolerate long delays. In direct contrast, voice traffic can accommodate limited amounts of lost or corrupted information but there exists a delay constraint for acceptable conversational voice quality. Stringent delay requirements translate to a bounding of variable network delay in order to guarantee a maximum end-to-end delay limit. Whereas the permissible delay limit may differ based on particular applications, it is generally accepted that a delay bound of 200 ms is desirable for conversational voice quality [15, 17].

Although speech has a higher tolerance to losses than data, satisfactory voice quality requires a restriction on the number of losses as well. Tolerance of losses is dependant on the waveform reconstruction techniques employed at the receiver to camouflage the missing speech segments. Previous studies have shown that silence substitution for missing speech imposes a loss rate limit of 1-2%. Packet repetition (of previously received packets) to fill in the gaps created by losses extends the limit to 5%. Sophisticated reconstruction mechanisms, employing pitch waveform replication, have been demonstrated to extend tolerance to 10% [35, 36].

Voice also has varying sensitivity to loss rates depending on the degree of compression used. Highly compressed voice is more sensitive to losses than uncompressed voice. In low-bit-rate coding schemes, a decoding process indexed to the current time is performed using a previously reconstructed output signal in the decoder so that the degradation by packet loss is greater than that using high-bit-rate schemes such as Pulse Code Modulation (PCM). Moreover, simple substitution techniques used in the reconstruction process have a greater impact on the perceived speech quality compared to the corresponding effect for high-bit-rate coded speech because, the low-bit-rate coded signal is determined sequentially and not solely by the current transmitted codeword [37].

Packet loss may be detected in the receiver by a discontinuity in the sequence numbers of the received packets. Primary factors contributing to the average loss rate include:

- **Transmission errors:** Public FR networks have been designed to deliver a ratio of errored protocol data units (PDUs) to total PDUs $\leq 10^{-13}$ [8]. The high reliability of fiber optics has resulted in the limitation of packet loss caused by the bit error rate (BER) of transmission lines to a negligible minimum.
- **Transit buffer overflows:** The loss rate due to buffer overflows in the transit path of voice traffic can be controlled by engineering the buffer length to an appropriate size. The use of priority service for voice at the transmission link buffers, as will be discussed later, further reduces the loss probability due to buffer overflow. Since it is relatively cheap to provide adequate memory for queues, the losses due to buffer overflow can be limited to a minimum value of 10^{-4} [8]. However, the buffer size may need to be bounded to a certain maximum to restrict the queueing delay.
- **Excessive delay:** Packets arriving later than the target delay limit due to excessive queueing delays experienced in the transit path will be discarded at the receiver. Due to the variable network delays caused by the presence of data traffic, this factor is usually the most significant source of voice frame loss in a FR connection. For the purpose of this research, it is assumed that loss rate is primarily caused by packets discarded at the receiver due to late arrival.

Average loss rate, by itself is an inadequate measure of voice performance. The pattern of losses has significant impact on the perceptual quality of played-out speech. Randomly distributed losses are essential to achieving good quality speech. However, the stochastic nature of packet networks introduces bursty losses. Long periods with almost no losses are interspersed by intervals with high probability of correlated losses. Bursty losses are highly undesirable as they produce audible faults in speech and the effectiveness of waveform substitution techniques is reduced in conditions involving such loss behaviour

[35, 36]. The need to have uncorrelated losses is extremely important in a FR environment that is primarily used for inter-LAN connections characterized by bursty traffic.

2.2.2 Delay Composition

The delay composition of a voice packet from the time of its production at the voice coder until its arrival at the destination, prior to playout consists of several components. Presented in the context of FR networks, these components are however, generic to the delay composition of most packet-switching networks.

Codec delay c - Represents the delay introduced by the speech coder. The actual value depends on the encoding/decoding algorithm used and includes the delay due to computing, processing and buffering of speech at the encoder. Decoding delay is negligible.

Switching delay s - Includes the time spent by the FR switch for header processing, error checking, and for assigning the packet to the appropriate transmit queue.

Framing delay f - Delay incurred due to appending/stripping of headers and trailers to the packet at the Frame Assembler/Disassembler, representing time spent in encapsulating the voice packet in a format suitable for transmission over FR networks.

Propagation delay p - Proportional to the distance spanned between end users in terms of the network links. Calculated on the basis of the approximate velocity of propagation, 2×10^8 m/s [38].

Transmission delay t - Time required to transmit the packet over the output links in the transmission path. Value is determined by dividing the packet length by the transmission speed of the output link.

Queueing delay w - Time spent in the buffers along the transmission path by the packet while it awaits processing or transmission.

Viewed from the perspective of variable and fixed delays; c , s , f and p can be computed for any connection based on the coding scheme used and network topology, and are fixed for the duration of the voice call. The sum of the fixed delays is designated as the total fixed delay, d_f . If a CBR voice coder is used, t , is a fixed component as well. In general the sum of t and w contributes to the variable delay introduced by the network and can be designated as the network delay, q .

2.2.3 Voice Synchronization and Random Delay Compensation

One of the most significant problems in supporting real-time applications such as voice over packet-switching networks involves the reconstruction of a continuous stream of voice from a set of packets that have experienced random network delays. Delay jitter introduced by the network can be smoothed by employing a playout buffer at the receiving end to delay the initiation of periodic playout. Voice synchronization involves choosing a target delay that has to be introduced by the receiver prior to playout. Intrinsic to voice synchronization, irrespective of the reconstruction strategy employed, is the concept of the "buildout delay". Buildout delay is defined as the maximum allowable variable delay in the transmission path for which compensation will be applied at the receiver. It is set at the receiver to the maximum limit of end-to-end delay D , minus the fixed delay d_f :

$$B = D - d_f \quad (1)$$

Setting the value for the buildout delay involves a trade-off between the delay that an audio application is willing to tolerate and packet loss that ensues as a result of late arrivals. Choosing a large value leads to lower losses but the end-to-end delay constraint may be violated, thereby affecting speech quality. However, a smaller value can increase the probability of buffer underflow and hence a higher loss rate. Having set the target buildout delay, an appropriate playout strategy is required to determine the playout time for each incoming packet. Existing algorithms achieve this goal by either explicitly determining the production time of each packet, or equivalently by estimating the actual

transport delay experienced by each packet. The first packet of a conversation is held in the playout buffer, for a holding time, h , based on the buildout delay in conjunction with the delay estimate for the network delay. Subsequent packets are played out depending on their relative timing with respect to first packet. When, the i th packet arrives having experienced a network delay of q_i , then the holding time h_i is:

$$h_i = B - q_i \quad (2)$$

Additionally, the buildout delay introduced may either be fixed for the length of the audio call or be recalculated at the beginning of each talkspurt. Adaptive variation of the delay is achieved by contracting or expanding silence periods by small amounts. Compression or expansion of silence by small amounts is not noticeable in the reconstructed speech [20].

The absence of an absolute time reference presents a major challenge in achieving voice synchronization. As mentioned in the earlier chapter, current implementations such as the PVP and other algorithms discussed in recent papers cope with this problem in one of the following ways. PVP uses cumulative delay stamps to record the actual delay encountered in each node to compute the playout time of each packet [24]. Absolute timing forms the basis of the adaptive algorithms discussed in [22]. Each packet carries an indication of its production time that allows the receiver to determine unambiguously when the packet was produced. Barberis and Pazzaglia propose inclusion of a similar time stamp in the packet header to achieve synchronization [18]. None of the existing approaches can be applied to FR networks without modifying the switches in such a way as to destroy their simplicity. OSI layer 3 processing required for the PVP and the provision of an absolute time reference implicit in the other approaches are prerequisites that do not adhere to the FR protocol simplicity.

This thesis presents an adaptive mechanism, that is explained in greater detail in subsequent chapters, to achieve voice synchronization in FR networks without making any underlying assumptions of a common time reference or proposing modification to the FR

protocol. The algorithm essentially involves holding the voice packets in the playout buffer for a target holding time and adaptively varying this holding time based on computation of the average holding time, and comparison of this value with the target value to obtain an error function that is used for dynamic adjustments to buildout delay. The holding time is directly observable and is related to the buildout delay. Dynamic adjustments to the holding time translate to adjustments in the buildout delay and these are achieved by what is defined as “time scale modification” (TSM) of speech samples. Time scale modification involves expansion or contraction of silence or speech segments based on the error function computed.

2.2.4 Reference Speech Coders: VSELP and I-CELP

Speech coding algorithms at 64 kb/s (G.711, PCM) and 32 kb/s (G.721, ADPCM) which are international CCITT standards are in ubiquitous use in both public and private speech telecommunications. The bit rates offered by these algorithms suffice for most applications but recent developments in mobile and packet communications have generated considerable interest in the use of low bit rate speech coding techniques [39 - 43]. The concepts of “bandwidth on demand” and statistical multiplexing inherent in the FR protocol provide compelling motivation for the use of highly compressed voice. Additional motivations for low bit rate coding are the need to minimize transmission costs, provide cost-efficient storage, and the demand to transmit over channels of limited capacity. Also to be noted is the fact that, given the widespread use of low-speed access links to FR networks [44], low bit rate speech coding implies that a larger number of voice connections can be accommodated over a given link.

In view of the low bit rate and high quality speech requirements for FR networks, the thesis evaluates voice performance using low-bit rate variants of the Code Excited Linear Prediction (CELP) technique. The following description of the coders is limited to a high-level view of the parameters involved. Detailed codec structure, frame formats and speech

processing techniques employed by the coders are not reviewed as they are not the focus of this research.

Vector Sum Excited Linear Prediction (VSELP) falls into the class of speech coders known as Code Excited Linear Prediction [39]. The VSELP speech coder is a low bit rate coder operating at 8 kb/s, designed to produce high quality speech with reasonably low computational complexity and is characterized by a robustness to channel errors. It has been chosen by the Telecommunications Industry Association for the North American IS-54 digital cellular standard. Encoded voice samples, called here codec-Protocol Data Units (PDUs), are generated at 20 ms intervals, at a sampling frequency of 8 KHz, with a frame length of 20 bytes, each codec-PDU comprising four 5 ms subframes. Encoding delay for generation of a single codec-PDU is 60 ms [41]. VSELP is the reference coder used to assess the qualitative performance of the proposed adaptive buildout delay mechanism in the thesis.

Integer-CELP is a CELP variant used as the reference coder for studying the effectiveness of the proposed congestion control mechanism in the thesis. Subjective evaluation of the congestion control mechanism requires use of a VBR voice coder to study the impact of reduction of coding rate in response to congestion notifications from the network. Since the VSELP coder is limited in its operation to a single coding rate an embedded, VBR I-CELP coder is used for the purpose of this study, operating at 16 kb/s, but capable of lowering the rate to 8 kb/s in response to external triggers [42]. When operating at 16 kb/s, the coder generates 40 byte codec-PDUs at 20 ms intervals, involving an encoding delay of 60 ms. The 20 ms codec-PDU is subdivided into 5 ms subframes. Operation at 8 kb/s reduces the speech quality of the coding but may increase the subjectively perceived quality at the receiver due to the reduction of congestion and losses. Rate reduction is achieved by suppressing the low-significance blocks of the codec-PDUs, resulting in a 20 octet voice packet. The coding delay and codec-PDU generation time remain the same as before but the codec-PDU length is reduced to 20 bytes.

Expansion	Contraction
20 ms $\rightarrow$ 25 ms	20 ms $\rightarrow$ 15 ms
20 ms $\rightarrow$ 30 ms	20 ms $\rightarrow$ 10 ms
20 ms $\rightarrow$ 35 ms	20 ms $\rightarrow$ 5 ms
20 ms $\rightarrow$ 40 ms	20 ms $\rightarrow$ 0 ms

Table 2.1: TSM Specification for Received Voice frames

Software versions of both the reference coders used in the study are not equipped with speech activity detection mechanisms and therefore silence periods of voice conversations are transmitted as well. Time-scale-modification (TSM) of the playout lengths of received voice frames is necessary for network delay compensation. The coders have provisions to expand or contract the length a 20 ms codec-PDU to one of the set of permissible values specified in Table 2.1. This is achieved by replication or truncation of the codebook excitation sequence [42].

TSM is incorporated in the VSELP codec format by appending a trailer specifying the type of modifications. The value of the trailer is chosen from a set of numerical values: $\{\pm 1, \pm 2, \pm 3, \pm 4\}$, each value proportional to the actual size of the increment/decrement, with expansions involving positive set elements and contractions referring to negative set elements. TSM is achieved in the I-CELP coder by duplication of subframes proportional to the size of expansion required and deletion of subframes proportional to size of contraction specified.

2.3 Adaptation Layer and Priority Service for Compressed Voice

Since the FR protocol was designed for the transport of data traffic, it has no provision for the maintenance of synchronization between end-users transmitting isochronous traffic with strict quality of service requirements. The following sections summarize the requirements for a voice adaptation layer in the end-user equipment and for provision of

priority service for real-time traffic.

2.3.1 Voice Frame Relay Adaptation Layer

A voice frame relay adaptation layer (VFRAL) was proposed in [26] for the purpose of adapting voice to a form suitable for transmission over FR networks. The VFRAL does not modify or enhance the existing FR protocol. It is implemented in the end-user equipment and requires no change of the FR protocol and switches. Frame relay access equipment typically refers to the customer premises equipment (CPE) used to interface with the frame relay switching equipment. Access equipment may be bridges, routers, hosts, packet switches or specialized Frame Assembler Disassemblers (FADs). Throughout this thesis, FADs are assumed to be the access equipment used. The adaptation layer can be integral to the FAD, or may be implemented in a separate device interfaced to the FAD. A sample frame relay configuration and the associated voice adaptation architecture is illustrated in Figure 2.3 and Figure 2.4 respectively. The primary functions of the Voice Frame Adaptation Layer (VFRAL) include:

- Accumulation, framing and encapsulation of the voice coder output
- Voice synchronization at the receiver and reconstruction of received speech using the playout algorithm.
- Reception of congestion notifications from the network and forwarding these notifications as adaptive rate control signals to the control input of voice coder. (For the purpose of chapter 5 only)

Codec-PDUs output by the voice coder are sent to the adaptation layer. Prior to the 6-octet overhead appended to the codec-PDU by the frame sub layer, the VFRAL attaches a 1-octet header to the codec-PDU. An enhanced frame format that results from the cumulative 7-octet overhead is illustrated in Figure 2.5. The standard frame format is supplemented by additional fields that are to be used by the playout algorithm to achieve synchronization:

Figure 2.3 Frame Relay Network Configuration

Figure 2.4 Voice Adaptation Architecture

Figure 2.5 Enhanced Voice Frame Format

- **Sequence Number:** is a wrap-around 6-bit counter used to determine the sequential order of the periodic voice frames. The sequence number is also essential for detection of frame loss, identified by a discontinuity in the sequence number.
- **BT and LT:** denote the start and end of a talkspurt. These fields are included for use with voice coders capable of detecting silence periods. They may also be used in conjunction with the DE bit of the frame relay sublayer, at the FR switches, for discarding silent frames upon congestion detection.

2.3.2 Priority Service for Voice

Although FR standards provide for access speeds upto 2 Mb/s, cost control is often a

prime motivation for many users to limit access links to lower speeds. Access links from end-user equipment to the first FR switch may be at low speeds, operating at the ISDN basic rate of 64 kb/s or the common North American rate of 56 kb/s. A recent survey of North American frame relay subscribers indicated that 64% of access links speeds used, operate at 56/64 Kb/s or lower [44]. Transport of variable-length data units permissible by the FR standard, allows for transmission of large-sized data frames such as 1500 byte frames specified in the IEEE 802.3 standard [50]. When voice and data traffic streams are multiplexed over low-speed links, the size of data frames has a great effect on the waiting time of voice frames. Transmission of large-sized data frames over low-speed links can cause excessive queueing delay to the delay sensitive voice streams if only a single QoS category is supported, as is the case in current FR networks.

It has been shown in earlier work [25, 26] that good voice performance in such networks, at high total link loads, requires the use of priority transmission for voice frames, at least the low-speed links in order to eliminate load control in the destination access link. Implementation of priority in a FR switch requires the use of separate transmission queues for the two classes of traffic. The FR protocol does not provide for priority indication, however internal virtual circuits are used in FR networks so that all frames on a particular connection follow the same route through the network. This is implemented by inserting a DLCI in each frame. The DLCI is valid only on a single data-link between switching nodes and the connection is maintained by means of a DLCI translation table in each node on the route of a connection.

Minor modification to the FR switch software can accommodate priority transmission. The DLCI table in a node could be modified to indicate not only the outgoing link for the frame but also the outgoing priority queue. This can be achieved by an indication in the optional field of the SETUP message for the connection and would not require any modification of the standard frame format.

Non-preemptive priority implementation at the output link of the source FAD which is

Figure 2.6 Priority Implementation at Access Link

under user control, and in the last FR switch that drives the destination access link, both operating at 64 kb/s, is achieved by use of two output queues per outgoing link instead of one: a priority queue for voice and a non-priority queue for data. Figure 2.6 illustrates the modeling of the low-speed access links by means of a single-server, dual queue model.

Chapter 3

Voice Synchronization: Adaptive Buildout Delay for Voice in Inter-LAN Traffic

3.1 Introduction

Earlier work by Dong, Kaye and Mahmoud [25, 26] specified a first approach to voice synchronization suitable for FR networks called *Adaptive Buildout Delay*. This chapter presents the original playout mechanism in an algorithmic form and re-evaluates its performance in the context of inter-LAN ethernet traffic [51]. The impact of data frame size on voice performance is investigated and segmentation of large LAN frames is proposed at the user-network interface to satisfy the QoS requirements of conversational voice applications.

3.2 Adaptive Buildout Delay Mechanism

This section presents the original adaptive buildout delay mechanism which does not require the use of a measured and recorded value of the queueing delay experienced by each packet as required by the PVP. The algorithm requires no network layer processing in the switching nodes and makes no underlying assumptions of availability of a common time reference between the sender and receiver. The playout mechanism is based on the idea, that given some initial estimate of the queueing delay for a particular network configuration, the holding time of voice packets at the receiver playout buffer can be adaptively adjusted at periodic intervals, based on the measured value of the mean holding

time, to effectively smooth out the random delay observed in packet arrival times. Dynamic adjustment of holding time is achieved by Time-scale-Modification (TSM) of packet playout lengths at the receiver, and packets are played out at the same uniform rate at which they were generated by the source coder.

3.2.1 Notation and Terminology

The following notation is required to describe the adaptive buildout delay mechanism, and includes parameters referenced directly or indirectly in the algorithm:

Directly Referenced Notation:

$T_a(n)$: Arrival time of the n th frame.

$T_p(n)$: Playout time of the n th frame.

playout_length(n): Playout length of the n th frame (measured in ms).

$H(n)$: Actual holding time of the n th frame.

$\bar{h}_T$: Initial, preset value of target mean holding time: $\bar{h}_T = B_T - \bar{q}$.

$\bar{h}_a$: Average holding time of frames measured over a certain number of frames.

A : Averaging interval (measured in terms of number of frames).

h_s : A cumulative sum used in computing $\bar{h}_a$.

E : Error function: $\bar{h}_T - \bar{h}_a$.

$I(n,k)$: Source-indicated playout interval between the k th and n th frame, $n > k$.

t_{ism} : Time increment used to make dynamic adjustments to target mean holding time.

ϵ : Minimum deviation of $\bar{h}_a$ for which a dynamic adjustment will be made to $H(n)$.

Indirectly Referenced Notation:

$\bar{q}$:	Mean network delay.
d_f :	Fixed delay.
D_T :	Target end-to-end delay.
B_T :	Target Buildout delay: $B_T = D_T - d_f$

3.2.2 Holding-Time Relation to Buildout Delay

The adaptive buildout delay algorithm eliminates delay jitter by buffering received voice packets, observing the actual holding times of these packets to determine a short-term average holding time and dynamically varying the playout length of packets periodically based on an error criterion to coerce the average holding time towards a desired target value. The holding-time paradigm is based on the arrival process of voice frames and relates the holding times of frames to the actual buildout delay.

Buildout delay has been defined in Section 2.2.3 as the maximum allowable variable delay in the transmission path for which compensation may be applied and is considered to be the target buildout delay, B_T , for the algorithm. The adaptive mechanism defines a target mean holding time, $\bar{h}_T$, which is set based on an initial estimate of the mean network delay, $\bar{q}$, in conjunction with the target value of buildout delay:

$$\bar{h}_T = B_T - \bar{q} \quad (3.1)$$

The first packet of voice conversation is held in the playout buffer for a holding time equal to $\bar{h}_T$. Subsequent packets are played relative to their original relationship to $T_p(1)$, if they arrive before their scheduled playout time, else they are discarded. Thus, for the n th packet:

$$T_p(n) = T_p(n-1) + I(n, n-1) \quad (3.2)$$

where $I(n, n-1)$ is a deterministic time interval between successively played-out voice packets. Since the source voice coder generates voice packets at a uniform rate, $I(n, n-1)$,

indicates the periodic relationship that exists between successive voice packets. Viewed in equivalent terms:

$$T_p(n) = T_a(n) + H(n) \quad (3.3)$$

where, the n th packet is actually held for a time:

$$H(n) = B(n) - q(n) \quad (3.4)$$

$H(n)$ represents the time interval between the arrival time of a voice packet at the destination and the time when it is played out. $H(n)$ is directly measurable, depends on its network delay, $q(n)$, and is related to the actual buildout delay, $B(n)$. Since the playout-time of the n th packet is scheduled based on the previously received packet, Eq. 3.3 implies that the actual holding time, $H(n)$, depends on the arrival time of the packet which in turn is dependant on the network delay, $q(n)$, experienced by the packet in the transmission path as seen from Eq. 3.4. It is thus observed that the network delay has a direct impact on the actual holding time of a voice packet. Smaller holding times are reflective of larger network delays and vice-versa. Thus although the actual network delay experienced by packets isn't observable, the algorithm relies on observation of the actual holding times to reflect network conditions.

The mean holding time (i.e., the mean of the actual holding times), $\bar{h}_A$, can be estimated at periodic averaging intervals, A , as $\bar{h}_a$, and is related to the average of the actual buildout delay, B_a , as follows:

$$\bar{h}_a = B_a - \bar{q} \quad (3.5)$$

The computation of the error function:

$$E = \bar{h}_T - \bar{h}_a \quad (3.6)$$

determines the deviation of the average holding time from the target value. Given a value of $\bar{h}_T$, the error function is a directly measurable parameter. The magnitude of the error function is indicative of the extent of deviation of the prevailing network conditions from the target mean, $\bar{q}$. A positive error indicates that the actual holding times of arriving

packets is below the target value, which implies that the current mean network delay is greater than the mean value, $\bar{q}$, used in the computation of $\bar{h}_T$. Negative values of the error function are reflective of longer holding times and, correspondingly, network delays less than the mean value. Viewed in the context of network conditions, the magnitude of the error function also serves as an indicator of the average buildout delay divergence from the target value. Combining Eq. 3.1 and Eq. 3.5, determines the deviation of the average buildout delay from the target value:

$$\bar{h}_T - \bar{h}_a = B_T - B_a \quad (3.7)$$

The magnitude and sign of the error function determines the size of the adjustment that has to be made to the playout time of the current frame. If the absolute value of the error function is greater than a target error, ϵ , then the holding time of the subsequent packet is increased or decreased by expanding or compressing the playout length of the current packet in small amounts, so as to avoid too serious a distortion of reconstructed speech. The size of the adjustment, t_{ism} , may also be limited by the set of valid TSM values allowable by the voice decoder. Any change to the playout length of the n th packet, determined by the sign of the error function, $\text{sgn}(E)$, and the allowable step size or TSM value, t_{ism} :

$$\text{playout_length}(n) = \text{playout_length}(n) + \text{sgn}(E)t_{ism} \quad (3.8)$$

translates to a change in the actual playout time of $n+1$ th packet as it will be offset by an amount equivalent to $\pm t_{ism}$:

$$T_p(n+1) = T_p(n) + I(n+1,n) + \text{sgn}(E)t_{ism} \quad (3.9)$$

This implies a change in holding time of the $n+1$ th packet and in effect a change to the buildout delay of the $n+1$ th packet, as deduced from Eq. 3.3 and Eq. 3.4. Hence the nomenclature for the algorithm.

Dynamic adjustments of the playout time of voice packets are necessary to achieve adaptive variation of the actual holding time, translating to adaptive variation of the

buildout delay. Absence of dynamic adjustments would mean that voice synchronization was determined by the arrival of the first packet. If that packet had experienced a delay significantly different from the average delay, subsequent packets would not be played out either due to the lack of buffer space or because they arrived too late.

Since $\bar{h}_a$ is only an estimate of the actual mean holding time, $\bar{h}_A$, the length of the averaging interval, A , determines the accuracy of the estimate. The number of frames used should be large enough for the estimate, $\bar{h}_a$, to be statistically significant, but as small as possible so that the adjustment can be made as early as possible. Too short an interval may lead to an inaccurate estimate and hence an inappropriate adjustment step. Too long an averaging interval may result in accuracy obtained at the expense of the loss and delay constraints being violated due to delay in dynamic adjustment.

Adjustments to the holding time are continued at regular averaging intervals as long as $\bar{h}_a$ varies significantly from $\bar{h}_T$, which essentially means as long as voice packets experience network delays that vary significantly from the mean network delay estimate used in computation of $\bar{h}_T$. Stabilization in existing network conditions results in the buildout delay approaching the target buildout delay within a target error, ϵ :

$$|\bar{h}_T - \bar{h}_a| \leq \epsilon \quad (3.10)$$

and the system completes a period of adjustment to prevailing network conditions. However, as conditions continue to change in the network, the target error will again be exceeded and adjustments will occur in a continuous adaptive process. Since bursty traffic typically results in network delays changing rapidly within short time intervals, it is assumed that the adaptive process will be a continuous one, characterized by discrete adjustments at periodic time intervals.

3.2.3 Holding-Time Adaptation Algorithm

The algorithmic version of the adaptive buildout delay mechanism is presented below:

Step 1:

Set $h_s := 0$

Set $\text{frame_count} := 0$

Step 2:

When the first frame arrives with sequence number n :

Record $n := \text{sequence number of new frame}$

Record $T_a(n)$

Increment frame_count by 1

Set $T_p(n) := T_a(n) + \bar{h}_T$

Compute $H(n) := T_p(n) - T_a(n)$

Play out frame n at $T_p(n)$

Set $h_s := h_s + H(n)$

Set $k := n$

Step 3:

When the next frame after frame k arrives with sequence number n :

/ normally $n = k + 1$, but a frame could have been lost in transmission */*

Record $n := \text{sequence number of new frame}$

Record $T_a(n)$

Increment frame_count by 1

Compute $I(n,k)$ */* See (1) in Appendix */*

Set $T_p(n) := T_p(k) + I(n,k)$ */* See (2) in Appendix */*

Compute $H(n) := T_p(n) - T_a(n)$

```

IF frame_count = A

    Compute  $\bar{h}_a := h_s / A$ 

    Compute  $E := \bar{h}_T - \bar{h}_a$ 

    IF abs(E) >  $\epsilon$ 

        /* Modify playout length of current (nth) frame based on the error function */

        Set playout_length(n) := playout_length(n) + sgn(E)ttsm

    ENDIF

    /* Reset all counters used in computation of  $\bar{h}_a$  */

    Set frame_count := 1

    Set  $h_s := 0$ 

ENDIF

/* If current frame arrived before its scheduled playout time, then place it at
appropriate position in playout buffer */

IF  $T_a(n) < T_p(n)$ 

    Playout frame n at  $T_p(n)$ 

/* Current frame arrived after its scheduled playout time, hence discard frame */

ELSE

    Discard frame

ENDIF

Set  $h_s := h_s + H(n)$ 

Set k := n

GOTO Step 3

```


3.2.4 Determination of Target Mean Holding Time

In the above description of the algorithm, the target mean holding time, $\bar{h}_T$, and consequently the setting of the holding time for the first packet, was obtained supposing that the mean network delay, $\bar{q}$, is known. A major contribution to this value is the delay experienced at the low-speed access links to the FR network. The actual value of the mean network delay varies, depending on the actual level of the link utilization. However, since the playout buffer is unaware of network conditions, $\bar{h}_T$ must be set based on the value of the maximum mean network delay expected for a given network configuration. Assuming that the FR switches are interconnected by links of T1 speeds or higher, a reasonable estimate of the maximum mean network delay, $\bar{q}_{max}$, can be made *a priori* based on the network configuration and the expected maximum load on the low-speed access links. This value of $\bar{q}_{max}$, can be used in place of q in Eq. 3.1 for the computation of $\bar{h}_T$:

$$\bar{h}_T = B_T - \bar{q}_{max} \quad (3.11)$$

This value of $\bar{h}_T$, determined *a priori*, can then be used as the actual holding time of the first packet arriving on a connection. The performance of the algorithm in terms of the mean values of end-to-end delay and loss rate obtained, is conditional on the choice of $\bar{h}_T$, and a trade-off is involved in the setting of $\bar{h}_T$. Higher values of $\bar{h}_T$ may result in lower loss rates at the expense of the end-to-end delay limit being violated. On the other hand, a small value of $\bar{h}_T$ can result in greater number of packets being lost as a result of late arrivals.

3.2.5 Adjustments during Speech

The original adaptive playout mechanism outlined in [25, 26] assumed that holding time adjustments are made only during silence periods. In view of the reference coders used here, which do not have the ability to detect silence periods but can change the playout duration of voice segments, and the expected variability in observed network delays due to bursty data traffic, for the purpose of this study adjustments are made during talkspurts as

Figure 3.1 Frame Relay Network Simulation

well. Due to the discrete and periodic nature of the adjustments, with the overall TSM below 1% (Section 3.3.6), it was anticipated the impact of TSM would be minimal on subjective quality of reconstructed speech.

3.3 Network Simulation

A software simulator of a mixed voice-data, corporate frame relay network was developed in a UNIX environment, using the "C" programming language, to study the performance of the Adaptive Buildout Delay algorithm.

3.3.1 Network Topology

The network topology used in the simulation study is illustrated in Figure 3.1. Two corporate branch offices, each equipped with a private branch exchange (PBX), representing customer premises equipment (CPE), are connected to a corporate frame relay network via full duplex, low-speed access links (Link1/Link2), each operating at 64 kb/s. Each PBX incorporates a Frame Assembler/Disassembler (FAD), which performs all functions related to assembling and disassembling of frames prior to transmission over the

network or after reception at the receiving end.

A common, receiver queue (RQ) is implemented in the FAD for the incoming voice and data traffic. The framed input is directed to the appropriate transmission queue (TQ) for transmission over the access link. The FR network comprises a single node (switch) characterized by a common receiver queue (RQ) for the CPU and separate transmission queues (TQ) for each output link. Voice and data frames are indistinguishable at the RQ, but are switched to the appropriate transmit queue at the destination links. Separate buffers are used for voice and data traffic at all the transmit queues (TQ), to implement a non-preemptive priority mechanism as described in Section 2.3.2.

A typical Frame Relay enterprise network usually consists of multiple FR switches interconnected by high-speed T1 lines, operating at 1.544 Mb/s. The inter-nodal trunks of such FR networks, based on T1 lines offer relatively low transmission delays, compared to the queueing delay experienced by packets in the low-speed access links. Incorporating additional nodes interconnected by high-speed trunks has minimal impact, in terms of contribution to overall end-to-end delay [26]. Therefore it is assumed that the performance of the algorithm is virtually unaffected by additional FR switches, connected by high-speed output lines. Consequently a simple network configuration, characterized by a single FR node connected by low-speed fractional T1 lines, is used for the simulation study.

3.3.2 Reference Connection

The reference connection comprises a compressed voice call, originating from Branch1, directed to Branch2 via Link1 and Link2. InterLAN ethernet traffic is multiplexed onto the reference connection by data calls L1/L2. Additional background traffic used to load the network switch is provided by a notional branch office traffic, voice call V2 and data traffic, L3. The composition ratio of the integrated voice-data stream V2/L3 is set such that voice accounts for 25% of the total bandwidth (frames/s) and data accounts for the remaining 75%.

The reference coder used in this study is the VSELP coder, operating at 8 kb/s, described in Section 2.2.4. Voice calls, V1 and V2, represent constant-bit-rate (CBR) traffic from the voice coder, characterized by 20 octet packets, separated by 20 ms interarrival times. Voice packets from the VSELP coder are encapsulated by the FAD with a 1 octet VFRAL overhead, followed by a 6 octet header inclusive of the DLCI, resulting in a cumulative 7 octets overhead, and then placed in the transmit queue. Traffic model used to characterize the interLAN ethernet traffic, L1/L2/L3, is described in the following section.

3.3.3 Inter-LAN Data Traffic

Since LAN interconnection is a major application of frame relay networks, it seems reasonable to assume that data traffic used in the simulation model must capture the traffic properties of the output of a typical LAN. From the earliest study of Ethernet traffic by Shoch and Hupp [45], to a more recent one by Gusella [46], a bimodal packet size distribution has been reported. Although the mean of the packet lengths was different in the two studies due to the difference in the maximum packet sizes of the two Ethernets - 554 bytes for the 2.94 Mb/s experimental Ethernet of [45] and 1500 bytes for the 10 Mb/s Ethernet in [46], the packet size distribution was dominated by two types of packet lengths.

To confirm the above evidence, an empirical histogram was constructed from data collected over a 10 Mbps local-area research network to illustrate the bimodal nature of the packet-size distribution. The histogram plotted in Figure 3.2 indicates that LAN traffic is dominated by short size packets, between 64-245 bytes and packets of the maximum length, 1500 bytes, as specified by the IEEE 802.3 standard [50]. Packet sizes in the vicinity of 64 bytes are generated by the conventional character and acknowledgment traffic of the Transmission Control Protocol (TCP). Packet sizes clustered around 1500 bytes are the result of two major protocols: Paging traffic generated by the workstation's

Figure 3.2 Histogram of Ethernet Packet Size

virtual memory, handled by the Network Disk (ND) protocol; and file access traffic of a remote file server using the Network File System (NFS) protocol. Although the actual percentages of packet lengths may vary depending on the LAN configuration and protocol implementation, it has been observed that a substantial portion of ethernet traffic is dominated by 1500 byte packets.

The request-response protocols, both ND and NFS generate interarrival times whose distributions have pronounced modes. The Poisson model, often used in simulation studies does not capture the correlation in packet arrival times and consequently the burstiness of ethernet traffic [46, 47]. A batch (burst) Poisson model, characterized by simultaneous arrival of packets within a batch, and exponentially distributed batch interarrival times, has been shown to be an acceptable approximation of the LAN traffic arrival process [47, 48]. The traffic model used for interLAN ethernet traffic in the simulation study is characterized by a batch Poisson model, similar to the one implemented in [49]. A constant

batch size of 8 packets is used. Within a burst, packets lengths are determined in accordance with the probability distribution of packet sizes shown in Figure 3.2. LAN packets arriving at the source FAD are framed with a 6 octet overhead, without segmentation, and switched to the appropriate transmit queue of the access link.

3.3.4 Quality of Service Requirements and Determination of Target Buildout Delay

The performance of the algorithm is evaluated quantitatively in terms of the average values of loss rate (L_a) and variable end-to-end delay (B_a) obtained for a given network load. The mean variable end-to-end delay is synonymous with the average buildout delay, defined previously. Based on the packet-repetition strategy employed by the VSELP coder to minimize speech distortion in case of lost frames, an average voice packet loss rate less than 2% is assumed as the target quality of service required for a voice connection [43]. The maximum permissible value of the end-to-end delay for a voice conversation over the reference connection, termed D_T in Section 3.2.3, is set at: $D_T = 185$ ms.

The target buildout delay is computed based on the end-to-end delay constraint, D_T , and the fixed delay, d_f , for the network configuration used in the simulation model.

$$B_T = D_T - d_f \quad (3.12)$$

The total fixed delay is a cumulative sum of all fixed delays, outlined in Section 2.2.2: $d_f = c + s + p + f$. The following values of fixed delays are imposed on each voice packet that is transmitted over the reference connection:

Codec delay c : 60 ms

Switching delay s : 0.465 ms (FR Processor speed: 2150 frames/s)

Framing delay f : 2 ms (1 ms at each FAD)

Propagation delay p : 8 ms

The sum of all the above values of fixed delays, results in a total fixed delay, $d_f =$

70.465 ms. B_T , determined using Eq. 3.12, and rounded to the nearest millisecond is set at 114 ms.

3.3.4.1 Confidence Intervals of Simulation Results

The performance results presented in this thesis have been obtained with a 95% confidence level and satisfy:

$$Prob(|\theta - \hat{\theta}| < \epsilon) \geq 0.95$$

ϵ is the accuracy criterion, where it is desired to estimate the performance metric θ by $\hat{\theta}$. [52]. ϵ is set at 1 ms for the mean variable end-to-end delay, B_a , and, 0.25% for the mean loss rate, L_a .

3.3.5 Simulation Parameters

Parameters used in the simulation study are classified according to parameters related to the FR network simulation and parameters specific to the adaptive buildout delay algorithm.

Network and Traffic Parameters:

- load of Link1: 0.6
- load of FR nodal processor: 0.9
- load of Link2: 0.7
- length of voice frames: 27 bytes
- length of data frames: bimodal distribution
- batch (burst) size: 8 packets
- inter-batch time: exponentially distributed inter-batch time, with mean = $1/\lambda_d$ s where λ_d is given by $\lambda_d = \frac{(\rho \times l_s - \lambda_v)}{MBL}$ in bursts/s

Adaptive Buildout Delay Parameters:

- A: 100 packets (2s)
- ϵ : 5 ms
- t_{ism} : chosen from the set of allowable TSM values specified by the VSELP coder:

$$t_{ism} \in \pm\{5, 10, 15, 20\} \text{ ms}$$

3.3.6 Assumptions

- The buffer size of the output queues in the FR switch is infinite, consequently no voice packets are lost in transit due to buffer overflow in FR nodes.
- The playout buffer size is infinite, therefore there are no losses due to buffer overflow. Loss rate obtained from the simulations is only due to late arrival of voice packets at the playout buffer.
- Holding time adjustments are made during silent periods as well as talkspurts because the reference coders are not equipped with speech activity detection mechanisms. However, since the adaptive buildout delay mechanism is characterized by discrete and non-contiguous frame adjustment, with adjustments made only if required at 100 frame (2s) intervals, translating to a TSM of <1%, the perceived impact on conversational speech quality is minimal.

3.3.7 Subjective Test Description

Subjective evaluation of speech quality was not based on formal methods employing subjective criteria such as the Mean Opinion Score (MOS). Ad hoc qualitative tests were used as a means of detecting problems such as audible faults in played-out speech as well as a corollary mechanism to interpret and confirm quantitative results. Informal listening tests were performed using speech files of 30-60 sec duration (1500-3000 frames). Each file contained speech segments of both male and female speakers and was constructed by

Performance metrics	Mean Value	Confidence Interval
Mean Network Delay ($\bar{q}$)	101.02 ms	[99.96 <= $\bar{q}$ <= 102.08]
Mean Variable End-to-End Delay (B_a)	111.18 ms	[110.11 <= B_a <= 112.23]
Mean Loss Rate (L_a)	41.41%	[41.15 <= L_a <= 41.67]

Table 3.1: Voice Performance without Data Segmentation

concatenating smaller files containing 200-300 frames. Subjective evaluation included assessment of reconstructed speech by a single listener to determine discernible distortion such as pops and clicks, as well as unintelligible and muffled speech due to losses or TSM.

3.4 Adaptive Buildout Delay Performance with Delay Spikes

Twenty independent steady state simulations were run using the simulation parameters summarized previously, to investigate the performance of the playout algorithm in the presence of unsegmented, large LAN frames. The mean network delay obtained from the simulations for the given network conditions is used as $\bar{q}_{max}$ in Eq. 3.11 to determine the target mean holding time, $\bar{h}_T$, which is set at $\bar{h}_T: 114 - 101 = 13$ ms. Average values of loss rate and variable end-to-end delay (buildout delay) obtained using the preset value of $\bar{h}_T$ are tabulated in Table 3.1.

Table 3.1 shows that the measured value of mean network delay for voice frames is extremely high. This is primarily due to “spikes” in queueing delay experienced by voice frames accumulating in the priority queue during the transmission of 1500 byte data frames over low-speed access links of 64 kb/s. The transmission of a 1500 byte frame over a 64 kb/s link takes approximately 190 ms and this leads to large waiting times for voice frames buffered in the priority queue, resulting in packets arriving later than their scheduled playout time at the receiver playout buffer. Figure 3.3, adapted from a trace of network delays, illustrates the occurrence of delay spikes in the network delay of voice

Figure 3.3 Occurrence of Delay Spikes with Unsegmented Data Frames

packets. A *spike* is characterized by a sudden, large increase in network delay, followed by a series of packets arriving almost simultaneously, leading to the completion of the spike. Whereas under normal network conditions, the average arrival rate of packets at the playout buffer is 10 packets/200 ms; the spikes observed in the FR simulation typically constitute 10 packets arriving within a time span of about 30 ms.

Spikes of a similar nature, though larger magnitude in terms of the network delay scale, have been observed in the network delay of audio packets transmitted point-to-point between several Internet sites [22]. It is noted that, although non-preemptive priority at the access links of the reference network configuration, reduces the magnitude of the delay spikes, it does not prevent the recurrence of the spikes as seen from Figure 3.3. Delay spikes recur until the buffered data burst comprising large LAN packets at the low-priority queue of the access links has been completely transmitted.

Additionally, large-sized data frames cause significant variance in the arrival times of

Figure 3.4 CDF of (a) Network Delays (b) Holding Times of Voice Packets with Unsegmented Data Frames

Negative values of holding time indicate “virtual” holding times of packets discarded due to late arrival

voice packets at the playout buffer. The cdf of actual network delays experienced by voice frames in the transmission path, plotted in Figure 3.4 (a) shows that 40% of packets suffered network delays that are far greater than the maximum permissible variable delay limit, specified by $B_T = 114$ ms. In view of the large queueing delays, the value of $\bar{h}_T$ used for the adaptive mechanism is insufficient to compensate for the random variation in delays. As a result, far too many packets are discarded due to late arrival at the playout buffer. Negative holding times are indicative of packets that have arrived after their scheduled playout time and are therefore discarded. The cdf of actual holding times of packets at the playout buffer, plotted in Figure 3.4 (b), shows that more than 40% of the observed holding times are negative, translating to an average loss rate of the same magnitude.

Subjective quality of played-out voice evaluated using the VSELP coder confirms the quantitative performance results. Since voice packets are transmitted at periodic intervals by the source codec, near simultaneous arrival of voice packets observed during the delay spikes effectively means that subsequent voice packets arrive at the playout buffer having

Figure 3.5 Voice Packet Loss with Unsegmented Data Frames, measured over 2 s time intervals

Link1 Load = 0.6, Link2 Load = 0.7, FR Processor Load = 0.9, $\bar{h}_T = 13$ ms

experienced similar but decreasing delays, causing correlated losses. Correlated losses cause audible faults in reconstructed speech. Played-out speech is characterized by relatively short periods of acceptable quality, followed by longer periods of extreme speech distortion.

Figure 3.5 shows the number of lost packets, measured over 2 s intervals, which may be interpreted as the loss rate during averaging intervals used by the playout algorithm (length of the averaging interval being 100 packets). It is observed from the graph that substantial number of averaging intervals (2 s) witnessed loss rates much greater than the average loss rate of 40%. Periods such as these cause severe degradation in speech quality.

3.5 Adaptive Playout with Reduced Delay Spikes

It is evident that, in order to minimize the impact of correlated losses on reconstructed speech and to achieve an end-to-end delay which is within the permissible limits, the

Link2	Mean Variable	Mean Loss Rate
Load	End-to-End Delay (ms)	(%)
70%	106.05 [105.42 $\leq B_a \leq$ 106.68]	2.51 [2.42 $\leq L_a \leq$ 2.60]
80%	109.64 [109.24 $\leq B_a \leq$ 110.04]	2.53 [2.42 $\leq L_a \leq$ 2.64]
90%	115.05 [114.53 $\leq B_a \leq$ 115.57]	2.64 [2.53 $\leq L_a \leq$ 2.75]

Table 3.2: Voice Performance with 500 byte Data Segments

maximum delay of voice due to data frames has to be limited. An Enterprise Network manager has two options: data frames can be segmented before entering the FR network, or, if the network is so equipped, preemptive priority could be used for voice over *long* data frames. Either of these approaches has a similar effect on voice quality though each has a different effect on data performance and congestion. In this study, we use the segmentation approach but the results would be similar for the corresponding preemptive priority approach.

Segmentation of large data frames can be supported by the FAD, at the boundary of the FR network, as discussed in [34]. Large data frames are segmented into smaller frames by the FAD, encapsulated by a FR header and then switched to the data queue for transmission over the low-speed access link, Link 1. The choice of segment size involves a trade-off between the increased overhead caused by smaller data frames and increased queueing delay caused by large data frames.

Initial experimentation involved segmenting 1500 byte packets into 500 byte segments. The segment size was chosen based on the observation that the mean packet length of the bimodal LAN packet distribution used in this study is 515 bytes. In addition, previous studies have used a packet size in the range of 500-590 bytes to model LAN traffic [47, 54]. Voice performance was investigated using $\bar{h}_T = 63$ ms, based on the observed value of $\bar{q}_{max} = 51.85$ ms, obtained when Link2 was subjected to an extreme load of 90%. Performance results obtained from simulations are tabulated in Table 3.2.

Figure 3.6 1500 byte Data Packet Segmentation

The measured values indicate that maximum permissible bounds of delay and loss rate are breached at high loadings.

In order to meet the delay and loss rate objectives at high loadings, a smaller segment size of 400 bytes was chosen, resulting in segmentation as illustrated in Figure 3.6. Each 1500 byte frame was partitioned into three 400 byte frames and a single 300 byte data frame. Simulations were conducted to determine the impact of a smaller data frame size on the overall voice performance. Voice performance was studied for varying destination link (Link2) loads by increasing the arrival rate of inter-LAN ethernet traffic at the link.

A monotonic increase was observed in the mean network delay as the load on the destination link was increased as illustrated in Figure 3.7. Performance of the adaptive mechanism was investigated using $\bar{h}_T = 71$ ms, based on $\bar{q}_{max} = 42$ ms (Link2 load = 90%). Figure 3.8 (a) shows that the average loss rate is well below the target limit of 2% at all levels of Link2 utilization. Furthermore, this target QoS was realized without any violation of the maximum permissible delay bound, as seen from Figure 3.8 (b). Data frame segmentation has effectively reduced the magnitude of the spike in queueing delay, resulting in smaller network delays as well as reduced delay variance. Figure 3.9 illustrates the reduced delay spikes for two different values of Link2 load.

The cdf of network delays in Plot 2 of Figure 3.10 (a), illustrates the effective decrease in network delays as a result of data frame segmentation. Figure 3.4 (a) which

Figure 3.7 Mean Voice Network Delay vs Destination Link Load
Link1 Load = 0.6, FR Processor Load = 0.9

Figure 3.8 Performance of Adaptive Buildout Delay (a) Mean Loss Rate (b) Mean Variable End-to-End Delay
Link1 Load = 0.6, FR Processor Load = 0.9, $\bar{h}_T = 71$ ms

Figure 3.9 Reduced Delay Spikes due to Data Frame Segmentation

(a) Link1 Load = 0.6, Link2 Load = 0.6, FR Processor Load = 0.9, $\bar{h}_T = 71\text{ms}$

(b) Link1 Load = 0.6, Link2 Load = 0.9, FR Processor Load = 0.9, $\bar{h}_T = 71\text{ms}$

graphed the delay distribution of voice packets obtained using 1500 byte LAN packets has been reproduced as Plot 1 in Figure 3.10 (a) to illustrate the comparison. In contrast to Plot 1 of Figure 3.10 (a), which shows that a large proportion of packets experienced delays far greater than even the maximum permissible variable delay limit, it is seen from Plot 2 that a very small percentage of packets have experienced network delays in violation of the delay bound. Therefore fewer voice packets are discarded at the playout buffer due to late arrival. The small percentage of negative holding times observed in Plot 2 of Figure 3.10 (b) is characteristic of the low loss rate due to reduced delay spikes. This can be compared to Plot 1 which depicts the significantly greater number of packets discarded due to late

Figure 3.10 Comparative CDF of (a) Network Delays (b) Holding Times of Voice Packets

1: Unsegmented Data Frames 2: Segmented Data Frames

Negative values of holding time indicate "virtual" holding times of packets discarded due to late arrival

arrival when unsegmented 1500 byte LAN packets were transmitted.

3.5.1 Loss Rate Analysis

Analyzing the quantitative results, one observes that although the mean variable end-to-end delay increases monotonically as the load on the destination link is increased, the

loss rate characteristics are non-monotonic. The non-monotonic characteristics can be evaluated in terms of the factors influencing the loss rate.

Factors which have a direct influence on the loss rate are:

- **Delay Spikes:** Although data frame segmentation resulted in reduced delay spikes, it is observed from Figure 3.9 that the frequent recurrence of spikes continues until the buffered burst of segmented data frames is completely transmitted. Though smaller in magnitude than the corresponding spike caused by a 1500 byte data packet, the reduced delay spike is nevertheless sufficient to cause substantial delay leading to losses. In particular, quick succession of such spikes causes packet losses to occur in bursts or in close proximity of the previous loss.
- **Holding-Time Reduction:** Averaging intervals characterized by the error condition: $\bar{h}_T - \bar{h}_a \leq -15$ ms. due to small values of mean network delay, result in equivalent holding-time reduction of 20 ms caused by the maximum size of delay adjustment. Compressions of packet playout lengths by 20 ms essentially translate to voice packet losses, contributing to the effective mean loss rate.

Factors exerting an indirect influence on loss rate:

- **Target Holding-Time, $\bar{h}_T$:** The value of $\bar{h}_T$ used at all levels of link utilization is based solely on the maximum network delay, $\bar{q}_{max}$, observed for link load of 90%. It is evident that at link loads lower than 90%, the value of $\bar{h}_T$ is smaller than what would have been the accurate setting, had the corresponding mean network delay been used. This lower setting of $\bar{h}_T$ leads to higher values of the error function, contributing to compressions of larger magnitude that would have otherwise been required.
- **Variability in observed holding-times during successive averaging intervals:** The burstiness of data traffic implies that an averaging interval characterized by $\bar{h}_a \gg \bar{h}_T$ may subsequently be followed by another characterized by $\bar{h}_a \ll \bar{h}_T$. Compression of packet playout length in view of the large value of average holding time of the first

Figure 3.11 Voice Packet Loss with Segmented Data Frames, measured over 2 s time intervals

(a) Link1 Load = 0.6, Link2 Load = 0.3, FR Processor Load = 0.9, $\bar{h}_T = 71$ ms

(b) Link1 Load = 0.6, Link2 Load = 0.9, FR Processor Load = 0.9, $\bar{h}_T = 71$ ms

(c) Link1 Load = 0.85, Link2 Load = 0.85, FR Processor Load = 0.9, $\bar{h}_T = 65$ ms

interval may be inappropriate for the succeeding interval as this causes an equivalent holding-time reduction of the same magnitude for the subsequent packets. Since the subsequent interval is characterized by delay spikes as implied by the latter condition, holding-time reduction of a large magnitude may result in losses due to late arrivals.

Interpreting the loss rate in terms of the factors outlined above, the increasing trend in the loss rate at moderate loads (50%-60%) is directly related to the maximum size of the delay adjustment, lower setting of $\bar{h}_T$ and the slow response of the algorithm to burst losses due to delay spikes. At higher loads, since the value of the observed mean network delay is closer to q_{max} used in computing $\bar{h}_T$, the variability in holding times is reduced

and fewer compressions of 20 ms magnitude are necessary. Ensuing losses are primarily due to delay spikes leading to the decreasing trend in loss rate characteristics.

3.5.1.1 Impact of Correlated Losses on Subjective Quality of Reconstructed Speech

Subjective analysis of reconstructed voice quality demonstrated that, even though the mean loss rate is well below the allowable maximum of 2%, played-out speech is characterized, as before, by bursty losses corresponding to the delay spikes of Figure 3.9. It is observed that the loss rate at various levels of network load is characterized by frequent peaks that correspond to bursty losses. Figure 3.11 shows the number of packet losses measured over 2 s averaging intervals. It is observed from the figure that peaks are shown to occur in the loss rate even at a low link load of 30%. Correlated losses cause audible degradation in the quality of reconstructed speech. Since the reference coder employs a “most recently received correct packet” repetition strategy to camouflage the losses, it is evident that reconstructed speech may be altered if losses occur in close vicinity. If the loss condition persists, repeated voice packets are also gradually attenuated until speech is completely muted.

3.6 Requirement for Modification to the Adaptive Buildout Delay Mechanism

Although the target numerically specified QoS for voice has been achieved, there is clearly a need for improvement in the adaptation to delay spikes. The playout algorithm responds too slowly to delay spikes and may establish a new holding time that is inappropriate for the subsequent averaging interval. Indication of losses, provided by the error condition: $\bar{h}_T - \bar{h}_a > 20$ ms at the end of an averaging interval leads to adaptation of holding-time by time-scale-expansion of packet playout length. Holding-time adaptation accomplished at periodic intervals, is adequate when network conditions result in mean delays close to the

value of $\bar{q}_{max}$ used in computing $\bar{h}_T$. However, delay spikes are characterized by large network delays and this reactive approach to holding-time adjustment is too slow to prevent the occurrence of further losses. Furthermore, dynamic adjustment based on short-term observation of holding times may be inappropriate for succeeding intervals due to the variability in network delays.

The inadequacy of the adaptive buildout delay (ABD) mechanism in coping with delay spikes is evident from the perceived speech distortion due to correlated losses. A modified adaptive playout algorithm is presented in the next chapter for improved adaptation to network delay variance.

3.7 Summary

Performance of the original adaptive buildout delay mechanism was investigated in the context of variable delays caused by inter-LAN ethernet traffic and limitation of delay spikes due to large LAN packets has been shown to be desirable to achieve the target QoS for conversational voice. A maximum segment size of 400 bytes allowed the voice loss rate objective to be met at high loadings, with no violation of the permissible end-to-end delay bound.

Identification of factors affecting the voice loss rate brought to light the slow response of the holding-time procedure to burst losses at all levels of last link utilization and the increased number of lost frames due to late arrivals as a result of holding-time reduction by the maximum adjustment step. Subjective tests demonstrated speech distortion due to correlated losses, highlighting the need for a modified playout algorithm, explained in greater detail in the subsequent chapter.

Chapter 4

Modified Adaptive Buildout Delay

4.1 Introduction

It was shown in the previous chapter that correlated losses corresponding to delay spikes caused by segmented LAN frames resulted in below optimal voice quality. A modified playout algorithm is proposed in this chapter for improved adaptation to variability in network delays, in particular for faster response to delay spikes [51]. The performance of the original and modified playout algorithms is investigated for single and multiple voice connections. Also presented in this chapter is a comparative performance evaluation of the modified ABD and the buildout delay procedure of the PVP. Finally, in light of the improved voice performance obtained using the modified algorithm, the maximum permissible data segment size is reevaluated and the possibility of transmitting a small percentage of unsegmented LAN frames is also investigated.

4.2 Modified Adaptive Buildout Delay Mechanism

The modified holding-time algorithm presented in this chapter includes enhancements to the original ABD algorithm to obtain a faster response to delay spikes and improved adaptation to variability in network delays. The fundamental precept adhered to is that holding-time adjustment should be periodic when packets arrive early due to low network delays experienced in transit but the adaptation should be accelerated in case of correlated late arrivals due to delay spikes. A conceptually similar approach has been advocated by Ramjee et al in [22] for playout delay estimation. The authors have compared the

performance of four different algorithms for audio applications in WANs and performance results indicate that those algorithms which adjust rapidly to changing delays achieve a lower loss rate due to late arrivals. Three of the four algorithms use delay estimates (for dynamic adaptation of playout delay between talkspurts) from the cumulative delay observations of the talkspurt. However, the fourth algorithm which provides the best performance in terms of the loss rate is one which immediately adjusts the delay estimate once a sudden large increase in network delay is observed followed by a series of packets arriving almost simultaneously.

4.2.1 Extended Notation Set

The following notation is required for the modified algorithm in addition to the terminology of the original algorithm, outlined in Section 3.2.1:

local_lost_frame_count:	number of lost frames within an averaging interval.
lost_frame_type:	variable used to classify lost frames as being consecutive or otherwise.
L(n):	sequence number of current lost frame.
L(n-1):	sequence number of previous lost frame.
cum_increase:	cumulative increase that has been made to the holding time of voice packets within an averaging interval, A.
loss_distance:	distance between successive frame losses, measured in terms of number of frames.
loss_threshold:	limit on the number of losses after which immediate change to holding time is required.
proximity_threshold.	maximum number of frames separating two losses such that the second loss can be categorized as having occurred in close proximity of the first one.

4.2.2 Modified Holding-Time Adaptation Algorithm

Step 1:

Set $h_s := 0$

Set frame_count := 0

Step 2:

When the first frame arrives with sequence number n:

Record $n :=$ sequence number of new frame

Record $T_a(n)$

Increment frame_count by 1

Set $T_p(n) := T_a(n) + \bar{h}_T$

Compute $H(n) := T_p(n) - T_a(n)$

Play out frame n at $T_p(n)$

Set $h_s := h_s + H(n)$

Set $k := n$

Step 3:

When the next frame after frame k arrives with sequence number n:

/* normally $n = k + 1$, but a frame could have been lost in transmission */

Record $n :=$ sequence number of new frame

Record $T_a(n)$

Increment frame_count by 1

Compute $I(n,k)$ /* See (1) and (3) in Appendix */

Set $T_p(n) := T_p(k) + I(n,k)$ /* See (2) in Appendix */

Compute $H(n) := T_p(n) - T_a(n)$

IF frame_count = A

 Compute $\bar{h}_a := h_s / A$

 Compute $E := \bar{h}_T - \bar{h}_a$

 IF $\text{abs}(E) > \epsilon$

 /* Modify playout length of current (*n*th) frame based on the error function */

 Call Time_Scale_Modification

 Set $\text{playout_length}(n) := \text{playout_length}(n) + \text{sgn}(E)t_{ism}$

 ENDIF

 /* Reset all counters and variables used in computation of $\bar{h}_a$ and for immediate holding time adaptation */

 Set frame_count := 1

 Set $h_s := 0$

 Set $L(n) := L(n-1) := 0$

 Set local_lost_frame_count := 0

 Set cum_increase := 0

 ENDIF

/* If current frame arrived before its scheduled playout time, then place it at appropriate position in playout buffer */

IF $T_a(n) < T_p(n)$

 Playout frame *n* at $T_p(n)$

/* Current frame arrived after its scheduled playout time, invoke immediate holding time adaptation procedure to: determine if the number of losses exceed the loss threshold and if so enumerate the size of the dynamic adjustment needed to holding-time based on the nature of losses, then discard frame */

ELSE

IF ($1 < \text{frame_count} < A$) (See (4) in Appendix)

Call Immediate_Holding_Time_Adaptation

ENDIF

ELSE /* frame_count = 1 */

Increment local_lost_frame_count by 1

Discard frame

ENDIF

Set $h_s := h_s + H(n)$

Set $k := n$

GOTO Step 3

4.2.3 Procedures Invoked by the Modified Playout Algorithm

1. Time_Scale_Modification

/* This procedure determines the required t_{ism} from the set of allowable TSM values based on the magnitude of E. Maximum adjustment step: 15 ms */

IF $E < 0$

$t_{ism} \in \{5, 10, 15\} \text{ ms}$

ELSE

/* Time-scale expansion only if there has been no previous TSM in the current averaging interval */

IF (cum_increase == 0)

$t_{tsm} \in \{5, 10, 15\}$ ms

ENDIF

ENDIF

2. Determine_Lost_Frame_Type

/*This procedure computes the loss distance between frames and categorizes frames as being consecutive or otherwise based on the value of the loss distance */

Compute loss_distance := L(n) - L(n-1)

IF (loss_distance == 1)

lost_frame_type := CONSECUTIVE

ELSEIF (loss_distance <= proximity_threshold)

lost_frame_type := PROXIMITY

ELSE

lost_frame_type := FAR_APART

ENDIF

3. Immediate_Holding_Time_Adaptation

/* This procedure determines if the loss threshold has been breached and specifies the TSM value based on the loss statistics */

Record sequence number of lost frame L(n) := n

Increment local_lost_frame_count by 1

IF (local_lost_frame_count >= loss_threshold)

/* local_lost_frame_count will be equal to loss_threshold if it is assumed that there
no losses due to voice buffer overflow in the network */

Call Determine_Lost_Frame_Type

IF ((lost_frame_type == CONSECUTIVE)

OR (lost_frame_type == PROXIMITY))

IF (cum_increase < 15)

expansion_step := 15

ELSE

expansion_step := 5

ENDIF

ELSE

expansion_step ∈ {5, 10, 15}

ENDIF

Reset local_lost_frame_count := 0

ELSE

IF $H(n) \leq -1.5$

expansion_step := 5

ENDIF

Set playout_length(n) := playout_length(n) + expansion_step

Set cum_increase := cum_increase + expansion_step

Set $L(n-1) := L(n)$

4.2.4 Inclusion of New Parameters and their Correspondence to Proposed Modifications

Modifications to the original algorithm include a limit on the maximum size of dynamic adjustment of holding time and incorporation of immediate change of holding-time in case of bursty losses. The limitation of holding time adjustment was included to reduce the effects of the lower setting of $\bar{h}_T$, and variability in observed holding times during successive averaging intervals, on the effective mean loss rate, especially at lower network link loads. The maximum size of dynamic adjustment was limited to 15 ms based on the observation that a large proportion of delayed voice packets, using the previous value of 20 ms arrived at the playout buffer less than 5 ms after their scheduled playout time. The histogram of holding times of delayed packets plotted in Figure 4.2 shows that, even for a nominal Link2 utilization of 50%, a significant percentage of packets were delayed by 5 ms. Similar distributions were observed for various Link2 loads. The choice of 15 ms was also dictated by the fact that although a smaller maximum size may achieve a lower loss rate, this could be at the expense of the end-to-end delay bound, particularly at high link loadings.

The modified algorithm monitors the number of losses within an averaging interval and, when a preset loss threshold is breached, immediate expansion is made to the holding-time. The flow chart of Figure 4.1 demonstrates the holding-time adjustment process upon detection of losses. Two new parameters, *loss_threshold* and *proximity_threshold*, have been introduced for detection of losses and determination of the magnitude of dynamic adjustment to be made to the holding time. *Loss_threshold* sets a limit on the number of losses allowable in an interval after which immediate adjustment of holding time is required. *Proximity threshold* defines the maximum number of packets separating successive losses, such that the second loss can be said to have occurred in close vicinity

Figure 4.1 Immediate Holding-Time Adaptation upon Detection of Losses

of the first loss. The value of both the thresholds is determined by the nature of losses for a given voice-data traffic composition and the QoS requirements of voice applications.

As illustrated in Figure 4.1, the expansion step size is chosen based on the distance between successive packet losses. The size of the expansion step is also dictated by the impact that TSM of packet playout length has on the subjective quality of reconstructed speech. Initial expansion in case of successive losses is by the maximum adjustment step. It was anticipated (and verified subsequently by the performance results of Section 4.2.5)

Figure 4.2 Histogram of Negative Holding Times of Delayed Voice Packets

Negative values of holding time indicate “virtual” holding times of packets discarded due to late arrival

Link1 Load = 0.6, Link2 Load = 0.5, FR Processor Load = 0.9, $\bar{h}_T = 71$ ms

that this would effectively prevent occurrence of further losses due to late arrivals until completion of the delay spike. However, extreme network delays may translate to delay spikes occurring in rapid succession. Holding-time, in such cases, is further adapted by the minimum allowable expansion step to prevent audible impact on speech.

In addition to holding time adaptation when the loss count exceeds the preset loss threshold, an incremental adjustment is made by the minimum allowable expansion step if any voice packet has arrived at the destination, delayed beyond its scheduled playout time by an amount of time, chosen as 15 ms for the current simulation topology. This adjustment is included to accelerate the response to large delay spikes with minimal impact on reconstructed speech. It is anticipated that adjustments of this kind will be minimal. The delay limit of 15 ms is set, based on the holding time distribution of delayed packets observed at various link loads, when the maximum data frame size is 400 bytes. Figure 4.2 illustrates this observation for a particular link load. If a larger segment size is used, the delay limit may be increased based on the delay distribution of packets.

Alternately, holding-time adaptation may be limited to immediate adjustments only when the loss-threshold is breached.

The parametric nature of the proposed modifications provides the end-user with flexibility in setting the thresholds based on the network configuration, expected traffic composition and the QoS requirements of the voice application.

4.2.5 Improved Voice Performance using Modified Adaptive Buildout Delay

Performance of the modified algorithm was investigated using the network simulation parameters of Section 3.3.5 and the following additional algorithmic parameters: `loss_threshold` = 2 packets, `proximity_threshold` = 10 packets. The loss threshold was set to be equal to the overall voice loss rate objective of 2%. The proximity threshold was set based on the observation that 85% of successive losses occurred within 10 frames of the previous loss. Figure 4.3 shows the cumulative distribution function of successive losses for a particular link load. Similar distributions have been observed for various values of destination link loadings. These threshold values have been set for the reference connection in the simulation study based on the voice-data bandwidth ratio used for performance evaluation and may be varied depending on the network traffic composition.

Simulation results show that the mean loss rates for all levels of link utilization were reduced by more than 50%. Figure 4.4 (a) illustrates the comparison between the performances of the original and modified algorithms in terms of the loss rate. This reduction in loss rate was achieved at the expense of a very small increase in mean variable end-to-end delay, of the order of less than 1 ms, as shown in Figure 4.4 (b). Improvement in voice performance can be attributed primarily to the ability of the modified algorithm to detect and respond quickly to burst losses. In addition, smaller adjustments to holding time during periods of low network delays results in improved adaptation to lower mean network delays and to variability in network delays, especially at lower link utilizations.

Figure 4.3 CDF of Number of Frames separating Successive Voice Losses

Link1 Load = 0.6, Link2 Load = 0.6, FR Processor Load = 0.9, $\bar{h}_T = 71$ ms

This is evident from the slow decrease of the slope of the loss rate characteristics as the destination link load is increased from moderate to high values. The earlier algorithm resulted in a steeper slope due to frequent holding time adjustments of the maximum size at lower link loads, where the mean network delay is lower than the value of $\bar{q}_{max}$ used in computation of $\bar{h}_T$.

Subjective tests showed a marked improvement in voice performance. Audible faults in speech are minimal as the modified algorithm has effectively contained the number of multiple losses within an averaging interval. Figure 4.5 illustrates the reduction in the number of lost frames. Peaks observed in Figure 3.11 have been reduced and the number of losses within an interval is rarely greater than 4 packets.

4.3 Multiple Voice Connections

Circuit-switching inhibits the reassignment of bandwidth from one class of traffic to another even as the demand for different classes of traffic varies throughout the day and

Figure 4.4 Comparative Performance of Adaptive Playout Mechanisms (a) Mean Loss Rate (b) Mean Variable End-to-End Delay

1: Original Algorithm 2: Modified Algorithm

Link1 Load = 0.6, FR Processor Load = 0.9, $\bar{h}_T = 71$ ms

the week. In contrast, FR networks offer the potential for such flexibility of bandwidth management and this is one their most attractive features to a manger of an Enterprise Network. The previous sections explored the impact of varying the percentage of data

Figure 4.5 Voice Packet Loss with Modified Algorithm, measured over 2 s time intervals

(a) Link1 Load = 0.6, Link2 Load = 0.3, FR Processor Load = 0.9, $\bar{h}_T = 71$ ms

(b) Link1 Load = 0.6, Link2 Load = 0.9, FR Processor Load = 0.9, $\bar{h}_T = 71$ ms

(c) Link1 Load = 0.85, Link2 Load = 0.85, FR Processor Load = 0.9, $\bar{h}_T = 65$ ms

traffic while the reference connection comprised a single voice call between the corporate branch offices. Performance results and subjective tests demonstrated that the target QoS requirements were met even at high loadings. However, when more than one voice call traverses the same network path from the source to the destination, the performance of the reference call may be significantly altered. This section investigates the impact of variation of voice traffic and the resulting performance of the adaptive playout algorithms when the reference connection comprises multiple active voice connections. Variation of voice traffic is achieved by increasing the number of active voice connections and decreasing the percentage of data traffic.

The network simulation model was modified only to the extent that multiple voice calls from Branch1 to Branch2, co-existing simultaneously, are accommodated. The effective bandwidth of each call of the reference VSELP coder being 10.8 kb/s (inclusive of FR overhead), the maximum number of calls that can be transmitted over a 64 kb/s link is five. The modified reference connection comprises multiple calls, ranging from one to five. Inter-LAN ethernet traffic is multiplexed onto the connection as done earlier. Due to the periodic nature of voice traffic, the queueing delay experienced by voice frames at the access link buffer is directly proportional to the phase difference between the calls. If the calls are phased apart significantly, the impact on waiting time of frames is minimal. Likewise, randomly distributed start times for the calls will typically result in smaller queueing delays. To assess the maximum possible impact due to additional voice calls, a worst case scenario was simulated, wherein all calls arrive almost simultaneously at the link buffer. The last call of the reference connection, is used for performance evaluation of the adaptive playout mechanisms.

4.3.1 Network Delay Characteristics

Simulations were conducted for a varying number of voice connections with the integrated load on the low-speed access links fixed at 85%, which is the maximum link utilization for five voice calls. The mean network delay obtained as the number of voice calls is varied from one to five is plotted in Figure 4.6. The non-monotonic nature of the network delay characteristics can be interpreted in terms of the voice-data traffic composition at the low-speed access links. As the number of calls is increased from one to three data accounts for a greater proportion of the link bandwidth in bits/s, contributing to 80% of the total load for one call and 60% for three calls. Spikes observed in the queueing delay of the reference call due to the data frame in service are compounded by the presence of additional voice frames in the priority buffer. This results in the increasing trend observed in the network delay, as the calls are increased from 1 to 3. As the number of calls are further increased, the trend is reversed, with voice accounting for 80% and 99% of the net bandwidth at 4

Figure 4.6 Mean Voice Network Delay vs Number of Voice Calls

Link1 Load = 0.85, Link2 Load = 0.85, FR Processor Load = 0.9

and 5 calls respectively. As the number of voice connections is increased, the random data traffic is effectively replaced by periodic voice traffic. The probability of encountering data frames in service is reduced and queueing delay is primarily due to voice frames, transmission time of each frame over the 64 kb/s link being 3.375 ms. Lower queueing delays lead to decreasing values of mean network delay.

4.3.2 Comparative Performance of the Original and Modified Adaptive Playout Mechanism for Multiple Voice Calls

Performance of the playout mechanisms was investigated, using $\bar{h}_T = 63\text{ms}$, based on the observed maximum value of mean network delay, $\bar{q}_{max} = 51\text{ms}$. Mean values of loss rate and variable end-to-end delay obtained for the original playout algorithm are shown in Plot 1 of Figure 4.7 (a) and (b) respectively. The increasing trend of the loss rate characteristics as the number of active calls is increased from 1 to 3, is due to the increasing value of network delay, and therefore greater number of losses due to late arrivals. A sudden, large increase in loss rate as the number of calls is further increased to 4, the value approaching target limit of 2%, is largely due to two factors: compression of playout lengths by 20 ms

Figure 4.7 Comparative Performance of Adaptive Payout Mechanisms for Multiple Voice Calls (a) Mean Loss Rate (b) Mean Variable End-to-End Delay

1: Original Algorithm 2: Modified Algorithm 3: Modified Algorithm with optimized holding-time adjustments

Link1 Load = 0.85, Link2 Load = 0.85, FR Processor Load = 0.9, $\bar{h}_T = 63$ ms

caused by large value of $\bar{h}_a$ (implying low $\bar{q}$) and inability of the original algorithm in adapting to variability in network delays between successive averaging intervals. The sharp decrease in the loss and end-to-end delay characteristics for the maximum number of calls is due to the corresponding decrease in the mean network delay.

Figure 4.8 Voice Packet Loss for Multiple Voice Calls, measured over 2 s time intervals (a) Original Algorithm (b) Modified Algorithm (c) Modified Algorithm with Optimized Holding-Time Adjustments

Link1 Load = 0.85, Link2 Load = 0.85, FR Processor Load = 0.9, $\bar{h}_T = 63$ ms

Number of Voice Calls = 3

As observed earlier for a single voice call, reconstructed speech for multiple voice connections is also characterized by bursty losses. However, the peaks observed in number of losses measured over averaging intervals are far more frequent and of a greater magnitude. This is illustrated in Figure 4.8 (a) and Figure 4.9 (a). Qualitative impact of multiple losses, assessed using the VSELP coder demonstrates poor speech quality, especially during intervals wherein the number of lost frames is greater than 20 frames.

Performance results for the modified playout algorithm indicate a vast improvement in voice quality. Quantitative simulation results demonstrate a 65% reduction in the observed mean loss rate for calls ranging from 1 to 5. As before, reduced loss rates were achieved with minimal increase in variable end-to-end delay. Plot 2 of Figure 4.7 (a) & (b) illustrates the comparative performance of the two playout algorithms. Immediate adjustment to holding time upon detection of delay spikes prevents successive losses

Figure 4.9 Voice Packet Loss for Multiple Voice Calls, measured over 2 s time intervals
(a) Original Algorithm (b) Modified Algorithm (c) Modified Algorithm with Optimized Holding-Time Adjustments

Link1 Load = 0.85, Link2 Load = 0.85, FR Processor Load = 0.9, $\bar{h}_T = 63$ ms

Number of Voice Calls = 4

resulting in fewer peaks in mean loss rate as seen from Figure 4.8 (b) and Figure 4.9 (b).

Subjective voice tests validated the improvement in voice quality due to limitation of correlated losses. However, an occasional blurring was observed in played out speech. This was shown to be related to the size of the holding-time adjustment. The modified algorithm uses an incremental approach in making adjustments to the holding-time to minimize speech distortion. During periods witnessing persistent losses caused by recurrence of delay spikes, frequent adaptation may follow due to the minimum size of delay adjustment, which affects the perceived voice quality. The incremental approach was adequate for the case of a single voice call, when successive losses were less frequent. However, in the present case, a larger delay adjustment is required if the loss condition persists.

Observation of the holding times of delayed frames indicated that a significant number

Figure 4.10 CDF of Holding Times of Delayed Voice Packets for Multiple Voice Calls

Negative values of holding time indicate “virtual” holding times of packets discarded due to late arrival

Link1 Load = 0.85, Link2 Load = 0.85, FR Processor Load = 0.9, $\bar{h}_T = 63$ ms
Number of Voice Calls = 4

of voice packets constituting a large delay spike have arrived at the playout buffer much later than 5 ms. Figure 4.10 shows that 70% of all late arrivals have been delayed by a value greater than 5 ms. It may be inferred from the previous observations that the size of the dynamic adjustment needs to be increased to achieve better voice quality. A simple factorial analysis was performed to determine if expansion step size can be increased to achieve lower loss rates. The factorial design is explained in greater detail in the subsequent section.

4.3.3 Factorial Analysis

Factorial analysis is a statistical design technique, which quantifies the effect of various decision variables, called factors, on the system response variables [52]. A 2^k factorial design characterizes the effect of k factors, each of which have 2 alternatives or levels [53],

Figure 4.11 90% Confidence Intervals for Primary and Secondary Effects of Loss Rate
Number of voice calls = 4

Factor	Symbol	Level 1 (ms)	Level 2 (ms)
Expansion step for packets delayed by a time value greater than 15 ms	E1	5	15
Expansion step when loss_threshold is breached	E2	15: if cum_increase < 15 5: if cum_increase >= 15	15

Table 4.1: Factors and Levels selected for Factorial Design

determines if the factors involved have a statistically significant impact on performance and, if so, sorts out the factors in order of impact. Effects are derived as linear combination of the observations obtained from steady-state experiments conducted for each factor-level combination. To quantify experimental errors, the 2^k design can be repeated r times, resulting in $2^k r$ observations. The statistical significance of effects ($e_1, e_2, \dots$) and their combined effects ($e_{12}, \dots$), is characterized by the confidence interval of effects computed from variance of effects. Any effect whose confidence interval includes zero is not

PM-1 3 1/2"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISIONSM RESOLUTION TARGETS

statistically significant.

Mean loss rate was chosen as the specific system response variable for this factorial design. Table 4.1 illustrates the 2 factors along with the corresponding levels chosen for each factor selected for the design analysis. Twelve steady-state simulations were performed for each one of the multiple voice calls ranging from 1 to 5, using 3 replications for each factor-level combination. Results obtained for the experiments are summarized as follows:

- Effects computed for mean loss rate for network conditions involving 1, 2 and 5 voice calls did not result in a significant impact as measured by the 90% confidence interval of effects.
- 60% of variability in the loss rate obtained for 3 voice calls is due to E1, which has a significant bearing on the loss rate. The confidence intervals for E2 and E12 include zero, proving that these factors do not have a significant impact on the loss rate.
- The effects for 4 voice calls are illustrated in Figure 4.11. Both E1 (33% variability) and E2 (51% variability) have a significant impact on loss rate as observed from the confidence intervals; whereas the combined effect does not have a significant bearing on loss rate.

4.3.4 Modified Adaptive Buildout Delay with Optimized Holding-Time Adjustments

The factorial analysis indicates that larger holding-time adjustments, corresponding to Level 2 of factors, E1 and E2, have a significant influence on the average loss rate only when the number of calls is 3 or 4. This is consistent with the previous observation that larger number of adjustments due smaller adjustment step size caused occasional blurring in reconstructed speech for network conditions involving 3 and 4 voice calls. Conclusions drawn from the factorial analysis suggest possibility of improved voice quality with larger

adjustment steps. To study the impact of a larger adjustment step, simulations were conducted with the adjustment step set at Level 2. The measured loss rate with inclusion of larger adjustment was less than the earlier value by a small amount. Plot 3 of Figure 4.7 (a) and (b) illustrates the comparison for loss rate and end-to-end delay using the earlier adjustment steps of 5 ms and the optimized adjustments of 15 ms. Although the quantitative results do not point to a significant improvement in voice performance, subjective tests demonstrate greater clarity in played-out speech. Larger holding-time adjustments upon detection of bursty losses prevent frequent adaptation, resulting in improved voice quality. Furthermore the number of losses within short term intervals is limited as well, as seen from Figure 4.8 (c) and Figure 4.9 (c).

It can be concluded on the basis of the performance evaluation of the modified algorithm that the holding time adjustment size depends on the network traffic composition and the loss statistics. Therefore it is recommended that the choice of expansion step for playout lengths should be determined based on the expected voice-data traffic mix for a particular corporate network and the required subjective quality of reconstructed speech.

4.4 Comparative Performance of Modified Adaptive Buildout Delay and Packetized Voice Protocol

The CCITT Recommendation G.764 on Packetized Voice Protocols [24] specifies a buildout delay procedure to compensate for the variable delay experienced by voice packets within the packet-switching network. A time stamp (TS) is incorporated in each packet to record the cumulative queueing delays experienced by the packet while traversing the network with a resolution of 1 ms. The sequence number (SEQ) of a packet, in conjunction with the TS, allows the removal of network delay jitter.

The buildout delay is a system variable that defines at each end the maximum

allowable variable delay in the transmission path. Its purpose is to mask the variability in the delay that each packet may experience so that all packets are played at regular intervals thereby achieving packet voice synchronization. The delay bound as specified by the buildout delay variable is 199 ms and packets that experience delays beyond this value are discarded. The first packet of a talkspurt is played out based on the TS value. The additional duration that the packet is held in the buffer prior to playout is given by:

$$(\text{Buildout delay}) - (\text{TS value}) = (\text{ms to hold before playout}) \quad (3.1)$$

Succeeding packets that are received in sequence are played out without a gap in their original time relationship to the preceding packets. Packets that arrive before their scheduled playout time are placed in the proper sequence in the buffer. Packets that arrive after their scheduled playout time are discarded.

Simulations were conducted to study the performance of the time-stamp method of the PVP, using network simulation parameters similar to those used for the adaptive buildout delay mechanism. The buildout delay was set at 114 ms, which is the target buildout delay, B_T , used for the adaptive playout mechanisms discussed earlier. Simulation results indicate that the loss rate increases as the destination link is subjected to progressively higher loads. Increasing values of link load translate to a corresponding increase in mean network delay, $\bar{q}$. Therefore, it is observed that, if the PVP buildout delay procedure is employed, voice connections characterized by smaller values of $\bar{q}$ result in lower loss rates. Figure 4.12 (a) illustrates the comparative performance with respect to the loss rate of the two playout mechanisms. Although the ABD algorithm results in slightly higher loss rates at smaller link loads it is seen from the graph that, at high link loadings, the loss rate is comparable to that obtained by the PVP resulting in a lower value at the maximum load of 90%. In all cases the loss rate is sufficiently low that differences are not significant.

The major performance advantage of the ABD is that connections with shorter mean network delays have the advantage of smaller end-to-end delay or, correspondingly, shorter buildout delay. The PVP is characterized by a fixed buildout delay of 114 ms for all

Figure 4.12 Comparative Performance of Modified Adaptive Buildout Delay and Packetized Voice Protocol (a) Mean Loss Rate (b) Mean Variable End-to-End Delay

1: Time Stamp method of PVP 2: Modified Adaptive Buildout Delay

Link1 Load = 0.6, FR Processor Load = 0.9

levels of link utilization, as seen from Graph 1 of Figure 4.12 (b). Graph 2 of Figure 4.12 (b), shows that the delay values at low-medium link loads are smaller than the

Figure 4.13 Voice Packet Loss measured over 2 s time intervals (a) & (c) Time Stamp Method of PVP (b) & (d) Modified Adaptive Buildout Delay

(a) & (c) Link1 Load = 0.6, Link2 Load = 0.3, FR Processor Load = 0.9
 (b) & (d) Link1 Load = 0.6, Link2 Load = 0.9, FR Processor Load = 0.9

corresponding PVP values by a factor of 10-15%. The ABD acts adaptively to reduce the mean end-to-end delay to its minimum acceptable value, resulting in an average buildout delay, B_a , which is always smaller than the target buildout delay, B_T , by an amount dependant on the mean network delay, $\bar{q}$.

Although the quantitative evaluation indicated the target QoS requirements are met using the PVP, it is observed that the buildout delay procedure is unable to adaptively alter the holding time so as to contain the number of losses within short-term intervals. This can be seen from Figure 4.13. Whereas, the number of losses using the modified adaptive buildout delay mechanism hardly exceed 4 packets, the corresponding values using the PVP are far greater, thereby affecting the perceived quality of reconstructed speech. Qualitative evaluation using the VSELP coder indicated that the performance of the two

mechanisms is comparable at lower link loads, but the adaptive playout mechanism produces better speech quality at higher link loads.

To summarize, it may be stated that apart from the significantly smaller buildout delay obtained for comparable loss rates, the overriding advantage of the ABD is that it requires no OSI layer 3 processing in the network. The time-stamp method of the PVP necessitates the same requirement for voice synchronization in a packet-switched network.

4.5 Re-evaluation of Maximum Permissible Data Segment Size

Performance evaluation of the original adaptive buildout delay mechanism has demonstrated that limitation of delay spikes due to large LAN packets is desirable to achieve the target QoS for conversational voice. It was also shown that a maximum segment size of 400 bytes allowed the voice loss rate objective of 2% to be met at high link loadings. The implication of data frame segmentation for an enterprise network is that the network should be configured to accept frames only up to the maximum size specified earlier. If a PVC-based carrier provided, public FR network is deployed, although the segmentation can be accommodated at the source FAD, which is under user-control, it may not be possible to enforce the same at other user-network interfaces.

Lower loss rates obtained using the modified playout algorithm indicate a possibility of increased data frame segment size. This section re-evaluates the maximum permissible data segment size in view of the improved voice performance to determine if:

- Longer segment lengths can be transmitted over Link 2.
- A limited percentage of unsegmented 1500 byte frames may be transmitted over Link2.

In order to re-evaluate the maximum permissible data frame size, the simulation model was modified to accommodate larger data segment sizes for the destination link, ranging from 500-1400 bytes. With the exception of the 500 byte segment, all other

Figure 4.14 Mean Voice Network Delay vs Data Segment Size

Link1 Load = 0.6, Link2 Load = 0.7, FR Processor Load = 0.9

segment lengths, S_l , are used to partition the large LAN packet into two smaller frames, of size S_l and $(1500 - S_l)$ bytes. The modified playout algorithm used the following threshold values: `loss_threshold` = 8 packets, `proximity_threshold` = 10 packets. Loss threshold was set at a higher value, in anticipation of greater number of losses due to increased data frame size. Also, immediate adaptation was limited to holding-time adjustment only when the loss threshold is breached.

Simulation results indicate that except for a slight decrease at 700/800 bytes, the mean network delay characteristics, as illustrated in Figure 4.14, exhibit an increasing trend for increasing values of data segment size. The decrease in mean network delay is due to the smaller spikes observed in network delay for 700/800 byte segmentation as compared to 600/900 byte segmentation. Performance results using the adaptive playout algorithm indicate that the loss rate objective of 2% is breached by all data frames longer than 500 bytes, as shown in plot 1 of Figure 4.15 (a). The end-to-end delay rapidly increases beyond the maximum permissible delay bound, for segment sizes greater than 1000 bytes, as illustrated in Figure 4.15 (b). The repeated occurrence of a large number of correlated losses due to delay spikes at longer segment lengths, cause frequent adaptation of holding

Figure 4.15 Modified Adaptive Buildout Delay Performance vs Data Segment Size
(a) Mean Loss Rate (b) Mean Variable End-to-End Delay

- 1: Link1 Load = 0.6, FR Processor Load = 0.9, Loss_threshold = 8 packets
 2: Link1 Load = 0.6, FR Processor Load = 0.9, Loss_threshold = 4 packets
 3: Link1 Load = 0.6, FR Processor Load = 0.9, Loss_threshold = 16 packets

time, undermining the effectiveness of the modified algorithm in achieving the target QoS.

The performance of the algorithm was also investigated to determine its sensitivity to the loss threshold. The loss threshold was set at a smaller limit in the 500-1000 byte segment length range to determine if the mean loss rate may be lowered with minimal increase in end-to-end delay. Similarly, a higher setting of the threshold was chosen in the

1000-1400 byte range to determine if the mean variable end-to-end delay limit can be achieved for small increases in loss rate. Plot 2 of Figure 4.15 (a) demonstrates that although the loss rates are lower than their corresponding values in Plot 1, the target loss rate objective of 2% was still not met, even with a minor violation of the delay bound shown in Plot 2 of Figure 4.15 (b). Plot 3 of Figure 4.15 (a) and (b) illustrates that the higher setting of the threshold resulted in minimal lowering of the end-to-end delay at the expense of large increase in loss rate.

Quantitative evaluation of the modified algorithm with increased segment size has demonstrated that, given the bimodal nature of ethernet packet size distribution, the maximum segment size can be increased by only 25%, reaffirming the need to keep data frame size below a certain maximum in order to achieve the target QOS requirements of conversational voice. In light of the performance results, it is evident that unsegmented 1500 byte frames cannot be transmitted if the data traffic composition used in the reference connection follows a bimodal distribution, with significant percentage of large LAN packets. To investigate the possibility of transmission of unsegmented LAN frames, the data traffic model was modified such that 1500 byte packets accounted for only 5% of the total data bandwidth (in frames/s) at Link2.

Parameters used for the adaptive playout mechanism were set to: $\bar{h}_T = 60$ ms, based on $\bar{q}_{max} = 53$ ms at 90% Link2 load, $loss_threshold = 8$ packets, $proximity_threshold = 10$ packets. Performance metrics obtained from simulations indicate that the loss rate, as shown in Figure 4.16 (a), increases monotonically as the load is increased, with the maximum permissible loss rate limit being breached at low link loads. The mean end-to-end delay characteristics show a similar trend; the delay bound was exceeded during moderate high link loadings, as depicted in Figure 4.16 (b). Although the objective simulation results indicate that the loss rate and end-to-end delay are below the permissible limits at low link utilizations, subjective tests did not reflect target voice quality. Qualitative evaluation of voice demonstrated that the perceived quality of reconstructed

Figure 4.16 Modified Adaptive Buildout Delay Performance with 5% 1500 byte Data Packets (a) Mean Loss Rate (b) Mean Variable End-to-End Delay

Link1 Load = 0.6, FR Processor Load = 0.9, $\bar{h}_T = 60$ ms

Percentage of 1500 byte packets = 5%

speech was substantially affected by the burst losses even though they are infrequent. Voice packet losses measured over the averaging intervals are characterized by large peaks corresponding to delay spikes as seen in Figure 4.17.

The holding time distribution of packets at the playout buffer indicates that a significant proportion of packets discarded due to late arrival have arrived at the buffer

Figure 4.17 Voice Packet Loss with 5% 1500 byte Data Packets, measured over 2 s time intervals

(a) Link1 Load = 0.6, Link2 Load = 0.2, FR Processor Load = 0.9, $\bar{h}_T = 60$ ms

(a) Link1 Load = 0.6, Link2 Load = 0.2, FR Processor Load = 0.9, $\bar{h}_T = 85$ ms

Link 2 Load	L_a (%)	B_a (ms)
0.2	0.38 [0.34 $\leq L_a \leq$ 0.42]	110.17 [109.90 $\leq B_a \leq$ 110.44]
0.3	1.45 [1.32 $\leq L_a \leq$ 1.58]	117.09 [116.32 $\leq B_a \leq$ 117.86]

Table 4.2: Voice Performance with 5% 1500 byte Data Packets

more than 15 ms after their scheduled playout times. The cdf of holding times of delayed packets, plotted in Figure 4.18 (a), shows that, even at a low link load of 20%, almost 80% of all delayed packets have arrived later than 15 ms. The percentage is higher for higher link loads as seen from Figure 4.18 (b). Even with the number of unsegmented LAN frames confined to low maximum value of 5% the variability in network delays due to the delay spikes is too large to be effectively compensated by immediate holding time adjustments, hence the higher loss rates even at low link loads.

The performance of the algorithm was also studied using a larger value of $\bar{h}_T = 85$ ms

Figure 4.18 CDF of Holding Times of Delayed Packets with 5% 1500 byte Data Packets

- (a) Link 1 Load = 0.6, Link 2 Load = 0.2, FR Processor Load = 0.9, $\bar{h}_T = 60$ ms
 (b) Link 1 Load = 0.6, Link 2 Load = 0.9, FR Processor Load = 0.9, $\bar{h}_T = 60$ ms

to obtain lower loss rates at low link loadings. However, performance metrics obtained from simulations, tabulated in Table 4.2, indicate only a minimal improvement in voice performance. Furthermore, correlated losses persist, and peaks are still observed in loss rate over short term time intervals, as illustrated in Figure 4.17 (b). Subjective analysis of the speech quality clearly indicates that in spite of reducing the percentage of large sized LAN packets, and restricting the load on the destination link, satisfactory conversational voice quality is not achieved.

In view of the above findings, it is concluded that, given the stringent QoS requirements of conversational voice, the FR network has to be configured to accept shorter data frames or otherwise restrict the maximum delay of a voice packet in order to minimize the impact of variable delays and correlated losses on reconstructed speech quality.

4.6 Summary

A modified playout algorithm was proposed for improved adaptation to variable network delays. Modifications to the original algorithm included limiting the size of dynamic adjustment of the playout length of a voice packet, and detecting and responding quickly to burst losses. The parameter set was extended to include additional parameters for detection of losses and determination of expansion step size in case of correlated losses. Additionally it was shown that the thresholds incorporated in the extended parameter set may be determined based on the network traffic composition and loss rate characteristics, to achieve immediate holding time adaptation with minimal impact on reconstructed speech quality. Lower average loss rates were observed at various link loads with minimal increase in mean variable end-to-end delay. Subjective tests, using the reference VSELP coder, also demonstrated a marked improvement in voice quality.

Performance of the modified algorithm was investigated for multiple voice calls, and optimized holding time adjustments were suggested in view of the larger network delays and frequent adaptation. A comparative performance evaluation of the modified playout algorithm and the time-stamp method used by the PVP, demonstrated the improved ability of the former algorithm in adapting to burst losses, compared to the latter. Additionally it was observed that whereas the buildout delay using the time-stamp method was constant at levels of link utilization, the ABD offers the advantage of shorter buildout delay for voice connections with smaller mean network delay.

Re-evaluation of the maximum permissible data frame size, in light of the improved voice quality using the modified algorithm, demonstrated that the data segment size can be increased to 500 bytes; the resulting voice performance metrics being within the allowable loss and delay limits. However, performance results negated the possibility of transmission of unsegmented 1500 byte LAN packets, despite reduction in percentage of the large LAN packets and restriction of last link load. Therefore it is strongly recommended that large

LAN packets should be segmented into smaller frames prior to transmission over low-speed access links to reduce the occurrence of delay spikes and the corresponding impact of correlated losses on voice quality.

Chapter 5

Congestion Control of VBR Compressed Voice

This chapter evaluates the impact of network congestion on voice performance and investigates the feasibility of implementing a congestion control mechanism for voice traffic based on rate-reduction by variable-bit-rate (VBR) voice sources in response to explicit network congestion notifications. A dynamic rate control mechanism is proposed to relieve congestion caused by statistical variations of network traffic at the low-speed access links. Performance evaluation of the proposed mechanism includes quantitative simulation results as well as qualitative subjective tests using the reference I-CELP speech coder. The I-CELP coder is used in the subjective evaluation because of its dual-rate capability.

5.1 Network Simulation and Reference Connection

The network simulation model of the previous chapter has been modified to include additional network traffic, introduced to create an environment conducive to congestion. Referring to Figure 5.1, the reference connection comprises three compressed voice calls, V1, V2, and V3 from Branch 1, arriving almost simultaneously at the source FAD, destined to Branch 2 via Link1 and Link2. A fourth call, V4, from a notional branch office terminates on the same destination link as the reference connection. V1, which is the reference call, operates at 16 kb/s and is the last to arrive at the source FAD. V2, V3 and V4 operate at 8 kb/s. Inter-LAN ethernet traffic is multiplexed onto the reference

Figure 5.1 Modified Frame Relay Network Simulation

Figure 5.2 Full Duplex Low-Speed Access Link Queueing Model

connection by L1/L2. Similar to the earlier model, additional background traffic, V5/L3 is used to load the FR network switch. Inter-LAN ethernet traffic from Branch2 directed to Branch1, is provided by data stream L4. L4 represents reverse traffic, i.e; traffic travelling in the direction opposite to the frames of the reference connection.

The transmit queues of the full, duplex, low-speed access links have been modeled using the queueing model illustrated in Figure 5.2, which has been adapted from [54]. For the purpose of the congestion control, to be discussed in subsequent sections, it is assumed

that the transmit queues are managed based on the occupancy thresholds in the corresponding buffers. Two threshold levels are assumed to be provided: Congestion Threshold (θ_1) and Absolute Threshold (θ_2). The following actions are taken as the thresholds are reached in the transmit queues:

- θ_1 : Congestion notification is invoked to inform peripheral voice sources.
- θ_2 : All arriving frames are discarded and congestion notification is continued.

Congestion reaction upon breaching of the thresholds is described in greater detail in Section 5.4.1.

The reference coder used in this study is a variant of the Code-Excited Linear Prediction (CELP) technique, namely the Integer-CELP [41, 42]. I-CELP is an embedded VBR speech coder, capable of reducing its transmission rate from 16 kb/s to 8 kb/s and vice-versa in response to external triggers. Transmission at 16 kb/s involves generation of 40 octet codec-PDUs at 20 ms intervals. Suppression of low-significance blocks of the codec-PDUs, in response to network congestion notifications, results in 20-octet voice packets during periods of congestion. After FR encapsulation, the resulting frame size is 47 octets and 27 octets at 16 kb/s and 8 kb/s respectively. Inclusive of the FR overhead, the operational rate of a 16 kb/s voice call is 18.8 kb/s.

Traffic model used for data traffic L1, L2 and L4 is similar to the batch Poisson model, discussed in the earlier chapter; 1500 byte packets being partitioned into smaller frames as determined by the 400 byte segment size. However, a larger batch size, as well as a larger segment size of 500 bytes is used for L2.

5.1.1 Voice Buffer Size

The end-to-end delay constraint of conversational voice, necessitates a limit on the network delay, particularly during periods of network congestion. Congestion control action for data traffic is usually invoked by the threat of buffer overflow in a node as a

buffer approaches its maximum capacity. For voice the corresponding threat occurs when the delay in a buffer approaches the target delay bound, resulting in the voice frame being discarded upon late arrival at the receiver playout buffer. The network delay experienced by a voice frame may be bounded by appropriate sizing of the voice buffer at the low-speed access links.

From previous experiments, it was observed that the maximum voice buffer occupancy was 20 frames in the worst case scenario involving multiple voice calls, operating at 8 kb/s. This value of 20 frames translates to 4320 bits, or a maximum queueing delay of 67.5ms, which is approximately 60% of the maximum permissible variable end-to-end delay limit, $B_T = 114$ ms. This value of 4320 bits is used to size the voice buffer at the access links, Link1 and Link2. Larger buffer sizes for real-time traffic typically lead to inefficient utilization of network resources. A larger buffer size implies higher value of network delay, especially in view of network congestion, leading to both higher end-to-end delay for the voice connection as well as increased loss rate due to late arrivals at the playout buffer.

5.2 Simulation of Network Congestion

The previous section outlined an environment conducive to creating congestion in the FR network. This section includes description of how the network was driven to a state of congestion. Congestion, which can seriously degrade the QoS provided by the network, occurs when the demand for one or more of the network resources exceeds the maximum capacity that can be provided. Bandwidth, or frame stream congestion, occurs when the combined capacity all network traffic at a particular link is greater than the available bandwidth of the network link. Memory, or nodal congestion, occurs when the total instantaneous traffic buffered at a particular node exceeds the global memory capacity of the switch. Processor CPU congestion is characterized by the packet arrival rate being far greater than the switching capacity of the processor. Each of these congestion scenarios is

characterized by a particular network bottleneck. For the purpose of this study, it is assumed that the low-speed access links, in particular the destination link is the bottleneck. Congestion is thus limited to bandwidth or frame stream congestion.

Both the low-speed access links were driven to saturation by subjecting them to extreme loads of 100%. Voice traffic accounted for approximately 65% of the total load offered to Link1, the remaining 35% accounted for by data traffic. At Link2, percentage contribution of voice increased to 80%, because of the background voice call, V4. Congestion was specifically achieved at Link2 by introduction of background traffic, V4 and L2. The batch Poisson model used for L2, differed from the earlier model as follows:

- Large 1500 byte LAN packets, were partitioned using 500 byte segments.
- Batch size for the large LAN packets was increased to 12 packets, resulting in a modified batch size of 36 frames, each frame of size 500 bytes

The increased burstiness of data traffic, impacted by the increased batch and frame size, combined with multiple voice calls in progress resulted in the network state fluctuating between periods of moderate utilization to that of extreme utilization leading to congestion. In addition, congestion that results from the network traffic is dynamic and repetitive in nature. Voice buffer at Link2 reaches maximum occupancy levels during the transmission of large data frames, subsequently followed by rapid depletion due to the short transmission time of voice frames. This phenomenon is repetitive in the presence of a buffered burst of data frames, leading to increasing queue lengths in the priority buffer. Congestion is acute and persists with simultaneous arrival of batches comprising large LAN packets at both the low-speed links.

5.2.1 Simulation Parameters

Parameters used in the network simulation:

Network and Traffic Parameters:

- Link1 load: 100%
- Link2 Load: 100%
- FR nodal processor load: 0.9
- Link2 load due to reverse traffic: 0.65
- Voice buffer size: 4320 bits

Adaptive Buildout Delay Parameters:

- A: 100 packets (2s)
- ϵ : 5 ms
- $t_{ism} \in \pm\{5, 10, 15\}$ ms
- $\bar{h}_T$: 45 ms
- loss_threshold: 2 packets
- proximity_threshold: 10 packets

5.3 Voice Performance in Congested Network without Control

Previous experiments have demonstrated that for a worst case condition of 4 voice calls, which resulted in a high loss rate using the adaptive playout algorithm, the mean network delay was 50 ms. It was also observed from simulations involving larger segment sizes, that the algorithm is unable to provide acceptable voice quality for $\bar{q}_{max}$ greater than 50 ms, highest loss rates being recorded at 70 ms. $\bar{h}_T$ was set at 45 ms, based on this value of $\bar{q}_{max} = 70$ ms. To minimize frequent adaptation in view of the anticipated losses, the modified playout algorithm used larger holding time adjustments as determined from the factorial analysis for multiple voice calls.

The impact of congestion on voice quality can be quantified by the values of performance metrics obtained from the simulations. Performance measures obtained are

Figure 5.3 Average Voice Loss Rate with Network Congestion vs Time

Link1 Load = 1.0, Link2 Load = 1.0, FR Processor Load = 0.9, $\bar{h}_T = 45$ ms

Performance metrics	Mean Value	Confidence Interval
Mean Network Delay ($\bar{q}$)	75.38 ms	[74.72 <= $\bar{q}$ <= 76.04]
Mean Variable End-to-End Delay (B_a)	125.38 ms	[124.68 <= B_a <= 126.04]
Mean Loss Rate (L_a)	3.16%	[3.07 <= L_a <= 3.25]

Table 5.1: Voice Performance in Congested FR Network

tabulated in Table 5.1 which show that the values of measured loss rate and end-to-end delay have exceeded the constraints imposed on them. It is also observed that the actual mean network delay was approximately 76 ms, translating to $\bar{h}_T$ of 38 ms. Higher average loss rate will be recorded for this smaller value of $\bar{h}_T$.

Primary factors contributing to the loss rate include losses due to buffer overflow at the voice buffer of Link2 as well as losses due to late arrivals at receiver playout buffer. Figure 5.3 illustrates the average loss rate characteristics for the duration of the reference voice call. The target loss objective of 2% is breached relatively early, and the loss rate

Figure 5.4 Voice Packet Loss with Network Congestion, measured over 2 s time intervals

Link1 Load = 1.0, Link2 Load = 1.0, FR processor load = 0.9, $\bar{h}_T = 45\text{ms}$

remains above the maximum permissible limit for the remainder of the call. Immediate adaptation, even with larger holding-time adjustments has been inadequate to compensate for the large network delays caused by network congestion. Consequently, the end-to-end delay limit has been exceeded.

The impact of congestion on reconstructed voice quality can be inferred from the number of losses during short term averaging intervals. Figure 5.4 demonstrates that most of the averaging intervals recorded a loss rate substantially higher than the maximum of 2%, even with the use of the modified algorithm. Packet discarding due to buffer overflow at the destination link buffer, combined with burst losses at the playout buffer due to correlated late arrivals, result in perceptible degradation of voice quality. Subjective tests using the reference I-CELP coder, corroborated the quantitative performance results. Observed speech quality deteriorates during periods recording losses in the range of 10-18 packets. Reconstructed speech is characterized by speech distortion and unintelligible

speech due to missing segments in voice conversation.

5.4 Congestion Control

Congestion control mechanisms typically involve traffic management policies to ensure packet delivery during periods of network overload, and flow control strategies aimed at achieving rapid congestion abatement. Congestion control strategies can be broadly classified as:

- **Preventive Control:** Avoid long delays and throughput degradation by adapting to changing network conditions through means such as adaptive routing
- **Reactive Controls:** Take effect after onset of congestion and resort to drastic measures such as packet discarding to prevent further deterioration of network conditions.

5.4.1 Proposed Adaptive Rate Control

The proposed congestion control mechanism has elements of both preventive and reactive controls. The transmit queues in the low-speed access link for high priority voice traffic are monitored to detect congestion buildup. Control is implemented in response to feedback received from network regarding impending congestion, by regulating traffic flow into the network. Key elements of the scheme include:

1. Detection of impending congestion by monitoring occupancy of destination link voice buffer.
2. Explicit congestion notification to voice sources at network periphery.
3. Use of adaptive rate control to limit traffic admission.

The generic state diagram of the proposed congestion control scheme is illustrated in Figure 5.5 and the algorithm can be in any of the three states depending on the status of congestion in the network and the voice source reaction to this congestion. The significance of the states and the events that trigger a state change are now described.

Figure 5.5 Generic State Diagram

- Normal:** Voice sources are allowed to operate at the highest available coding rate in the absence of any network congestion. The FR switch processor monitors the voice buffer occupancy of Link2 to determine if the congestion threshold, θ_1 , or the absolute threshold θ_2 , has been exceeded. With reference to Figure 5.5, occupancy levels, determined by θ_1 , are indicative of congestion onset. When θ_1 is exceeded, the BECN bit is set on all traffic returning to Branch1. Setting of the BECN bit continues until buffer occupancy falls below the congestion threshold. θ_2 being exceeded suggests deterioration of network conditions and all arriving frames are dropped. Explicit congestion notification is continued by the FR switch.
- Lower Coding Rate:** When a frame is received at Branch1 with the BECN bit set, rate control is initiated for all sources originating from that branch which are operating at the higher coding rate. Fine granularity may be achieved for a varied range of operational rates, however, since the only available rates for the reference coder are 16 kb/s and 8 kb/s the voice transmission rate is reduced from the higher to the lower rate upon receipt of a frame with the BECN bit set.

- **Monitoring at Source:** Transmission is continued at the lower rate, until a predefined number of reverse traffic frames are received with the BECN bit cleared. This corresponds to the monitoring phase of Figure 5.5. A non-negative counter is maintained at the source FAD to keep track of arriving frames with BECN bit cleared. When the counter reaches a predetermined threshold, `BECN_clear_threshold`, congestion is assumed to have abated and the codec is triggered, via the rate control input, to resume transmission at the higher coding rate. The process of rate restoration is consistent with the congestion control mechanism of TCP [55]; characterized by rapid rate decrease in response to congestion and slow rate increase as congestion abates.

The use of buffer thresholds in congestion detection is not new to congestion control in FR networks. Previous work has demonstrated that congestion detection for delay sensitive voice traffic can be based on setting thresholds in the voice queues, that are dependant on the delay in the queue [27, 28]. This results in thresholds of a much lower value than are appropriate for the data queue, where the criterion for setting thresholds is based on prevention of buffer overflow and frame loss. Congestion control strategies similar in concept to the proposed scheme have been suggested for data traffic [54] and video traffic [31, 32] in FR networks.

5.5 Voice Performance with Adaptive Rate Control

Network simulation parameters used to evaluate the effectiveness of the proposed scheme included:

- Absolute Threshold, θ_2 : 4320 bits.
- Congestion Threshold, θ_1 : $\theta_1 = 0.75 * \theta_2 = 3240$ bits.
- `BECN_clear_threshold`: 17 frames.

The `BECN_clear_threshold` was set, based on the observation that the average arrival

Performance metrics	Mean Value	Confidence Interval
Mean Network Delay ($\bar{q}$)	66.06 ms	[65.21 $\leq \bar{q} \leq$ 66.91]
Mean Variable End-to-End Delay (B_a)	115.10 ms	[114.25 $\leq B_a \leq$ 115.95]
Mean Loss Rate (L_a)	1.27%	[1.23 $\leq L_a \leq$ 1.31]

Table 5.2: Voice Performance with Adaptive Rate Control

rate of reverse traffic at Branch1 from Branch2 is 17 frames/s. Simulation results, tabulated in Table 5.2 show significant improvement in voice performance. The average loss rate was reduced by approximately 60% of its original value, with a minimal breach in the end-to-end delay constraint. Early detection of impending congestion, followed subsequently by immediate lowering of transmission rate by the source codec, in response to received congestion notifications, has prevented the voice buffer of Link2 from reaching maximum occupancy levels, thereby reducing the losses due to buffer overflow. Subjective tests demonstrated substantial improvement in voice quality, as observed from the reduction in number of successively lost frames. Figure 5.6 illustrates the reduced number of lost frames, measured over 2s time intervals.

However, despite improvement in voice quality, the following observations were made with respect to played-out speech:

- Impact of losses on speech quality during periods having losses of 8-11 frames.
- Frequent oscillations between the two transmission rates of 16 kb/s and 8 kb/s, the average duration of voice at any one rate being close to 2s.

The fluctuations observed between the transmission rates in view of the queue length transients caused by the repeated nature of network congestion, suggest that the value of the BECN_clear_threshold is inadequate, leading to the source codec switching to the higher transmission rate too early. The oscillations in coding rates affect the efficiency of the control and may lead to additional losses. An optimal value of the threshold is required, to reduce the oscillations between the two rates. The following section studies the impact

Figure 5.6 Voice Packet Loss with Adaptive Rate Control, measured over 2s time intervals

$$\bar{h}_T = 45 \text{ ms, BECN_clear_threshold} = 17 \text{ frames}$$

of variation of the BECN_clear_threshold on voice performance, in order to derive an optimal threshold value via simulations.

5.5.1 Effect of varying BECN_clear_threshold

In view of the oscillations between the dual transmission rates, simulations were conducted to determine if an increase in the threshold value has any significant impact on voice quality. Decreasing trends were observed for all performance metrics as the value of the threshold was increased from 17-107 frames. Figure 5.7, Figure 5.8 and Figure 5.9 graph the change in the observed value of $\bar{q}$, L_a and B_a respectively, for varying threshold values. Average time spent at the transmission rates is tabulated in Table 5.3. It is observed from the table that the mean voice duration at any one of the transmission rates increases as the value of the threshold is increased, the implication of this being that there are fewer oscillations between the two rates.

Smaller values of loss rate and end-to-end delay clearly point to an improvement in

Figure 5.7 Mean Voice Network Delay vs BECN_clear_threshold

BECN_clear_threshold (frames)	Mean Voice Duration 8 kb/s	Mean Voice Duration 16 kb/s
	(s)	(s)
17	2.0	2.2
27	3.3	2.8
37	4.8	3.5
47	6.3	4.2
57	7.8	4.6
67	9.6	5.3
77	11.5	5.7
87	13.4	6.3
97	15.7	6.9
107	16.9	7.1

Table 5.3: Mean Voice Duration at Transmission Rates for varying BECN_clear_threshold

voice performance for larger values of the threshold. Furthermore, it is observed that these lower values of loss and delay are not at the expense of larger variation in

Figure 5.8 Mean Voice Loss Rate vs BECN_clear_threshold

Figure 5.9 Mean Voice Variable End-to-End Delay vs BECN_clear_threshold

transmission rates, as can be seen from the larger mean voice durations of Table 5.3, implying fewer fluctuations in transmission rates. Narrowing the choice to any one particular value of the threshold from the wide range of available values, is subject to the

Performance metrics	Mean Value	Confidence Interval
Mean Network Delay ($\bar{q}$)	61.71 ms	[60.88 <= $\bar{q}$ <= 62.54]
Mean Variable End-to-End Delay (B_a)	110.73 ms	[109.81 <= B_a <= 111.65]
Mean Loss Rate (L_a)	1.03%	[1.00 <= L_a <= 1.06]

Table 5.4: Voice Performance with BECN_clear_threshold = 77 frames

ensuing voice quality for that particular option. The slope of the loss-rate characteristics, observed in Figure 5.8, provides the first step in restricting the available options to the range of 57-107 frames.

Information from the graphs indicates that whereas the average values of network and end-to-end delay show a continued decrease for the entire range of threshold values, a slowly decreasing slope is observed for the mean loss rate in the BECN_clear_threshold range of 57-107 frames. The inference drawn from this is that increasing the value of the threshold beyond 57 frames has a negligible effect on voice performance in terms of the overall average loss rate. Once again, the choice of a particular value in the range is dictated by the resulting voice quality. Preliminary voice tests indicated that a choice of 77 frames resulted in acceptable voice quality. Therefore the BECN_clear_threshold was set at this value of 77 frames.

Performance metrics obtained using BECN_clear_threshold set at a value of 77 frames are tabulated in Table 5.4. Observed values of end-to-end delay and loss rate are below the maximum allowable limits. The average loss rate characteristics, plotted in Figure 5.10, demonstrate the effectiveness of the congestion control mechanism, in achieving the target QoS for the entire duration of the voice call. The average loss rate was reduced by 67% of the former value recorded during congestion, translating to a 7% improvement over the earlier observed value with BECN_clear_threshold set at 17 frames. Lowering of the loss rate is due to longer periods of operation at 8 kb/s, as a result of a cautious ramp up to the higher coding rate. Although operation at 8 kb/s implies lower voice quality compared to 16 kb/s, this approach is better than the former one, which caused frequent adaptations

Figure 5.10 Average Voice Loss Rate with Adaptive Rate Control vs Time

BECN_clear_threshold = 77 frames, $\bar{h}_T = 45$ ms

between the dual rates. The quality of voice, observed from subjective tests demonstrated significant improvement due to fewer oscillations as well as fewer losses within averaging intervals. Loss rates measured over 2 s intervals are shown in Figure 5.11. The improvement in performance is clearly visible when this graph is compared to the earlier graphs of Figure 5.4 and Figure 5.6.

5.5.2 Impact of Smaller Congestion Threshold

The results of the previous sections were obtained with θ_1 set at 75% of θ_2 . In order to determine whether the value of the congestion threshold is too close to the value absolute threshold, for it to be an effective indicator of congestion onset, performance of the adaptive rate control mechanism was investigated using smaller values of θ_1 . θ_1 was set at 50% and 60% of the absolute threshold value. Referring to Table 5.5, it is observed that for smaller values of θ_1 , the loss rate and end-to-end delay have been reduced by small amounts. This is due to earlier adaptation by the source codec to the lower transmission rate of 8 kb/s, in response to the early congestion notification received from the FR

Figure 5.11 Voice Packet Loss with Adaptive Rate Control and BECN_clear_threshold = 77 frames, measured over 2s time intervals

$$\bar{h}_T = 45 \text{ ms, BECN_clear_threshold} = 77 \text{ frames}$$

θ_1	B_a (ms)	L_a (%)
2160 bits	108.25 [107.29 <= B_a <= 109.21]	0.95 [0.92 <= L_a <= 0.98]
2592 bits	108.88 [107.94 <= B_a <= 109.82]	0.94 [0.92 <= L_a <= 0.96]

Table 5.5: Voice Performance for Smaller Values of Congestion Threshold

network.

It is also observed that voice performance does not differ significantly at either of the threshold values. Loss rate characteristics of Figure 5.12 illustrate that the observed average loss rates are similar for the reference voice call; minimal lowering of loss rate is observed for the smaller of the two threshold values. Although an incremental improvement is observed in terms of the lower loss rate, with respect to the earlier value of θ_1 , this is at the expense of the voice quality of played-out speech. The fraction of time (with respect to the total call duration) spent at the lower coding rate of 8 kb/s is much higher for smaller values of the congestion threshold. This is demonstrated in Figure 5.13,

Figure 5.12 Average Voice Loss Rate with Adaptive Rate Control and Smaller Values of Congestion Threshold vs Time

1: $\theta_1 = 0.5 * \theta_2$ 2: $\theta_1 = 0.6 * \theta_2$

Figure 5.13 Fraction of Time spent at the Lower Transmission Rate vs Congestion Threshold

which shows that percentage of time spent at 8 kb/s when $\theta_1 = 0.5 * \theta_2$, is approximately 15% larger than the corresponding value at $\theta_1 = 0.75 * \theta_2$.

In view of the above observations, it is concluded that $\theta_1 = 3240$ bits suffices as an adequate indication of congestion buildup. Smaller values of the threshold lead may lead to longer periods of operation at the lower coding rate, and hence lower average operational rates for voice than is required for the target QoS.

5.6 Voice Performance under Varied Congestion Levels

The robustness of the proposed adaptive rate control mechanism was validated by the performance results of the earlier sections. It was observed that even when the network was subjected to extreme congestion conditions of 100% link utilization, the control mechanism effectively adapted the source codec rate to alleviate congestion. Referring to Table 5.3 and Figure 5.13, it is seen that the average amount of time spent at the higher transmission rate of 16 kb/s was less than the corresponding time spent at the lower rate of 8 kb/s. This observation is consistent with the extreme level of congestion in the network. The dynamic and prolonged nature of congestion leads to longer periods of operation at 8 kb/s.

One can intuitively predict that as the load on the links is reduced, time spent at the higher rate will increase. Having proved the effectiveness of the control mechanism in extreme conditions, its performance was tested in more normal conditions to validate the above mentioned reasoning. The network was subject to varying levels of congestion by varying the load on the destination link, Link2, while the load on Link1 was fixed at 80%. The mean duration of the voice call at the two transmission rates for different link loads is tabulated in Table 5.6. Initially the only load on Link 2 is due to the reference connection comprising 3 voice calls V1, V2, and V3. Since the network has not been subjected to significant congestion, the source coder is allowed to operate at 16 kb/s for the entire call

Link 2 load	Mean Voice Duration	Mean Voice Duration
	8 kb/s (s)	16 kb/s (s)
0.63	-	length of voice call
0.7	5.5	90.78
0.8	5.6	37.6
0.9	6.8	17.18
1.0	7	4.8

Table 5.6: Mean Voice Duration at Transmission Rates for varying Link2 Load

duration. As the network is subjected to increasing levels of congestion by multiplexing additional data traffic and a background voice call, V4, the average time duration at 16 kb/s decreases.

Plot 1 of Figure 5.14 illustrates the increasing trend observed in the time spent at the higher coding rate as the load on Link2 is progressively decreased. The percentage of time (with respect to total call duration) spent at 16 kb/s is 40% when Link2 is subjected to an extreme load of 100%. However, as the level of congestion is decreased, at 70% load, the time value has increased to 95% of the total call duration. Lower values of Link2 load means that fewer frames of reverse traffic are received at Branch1 with their BECN bit set, hence smaller number of adaptations by the source codec to the lower transmission rate. The increased length of voice operation at the higher transmission rate for smaller values of Link2 load, clearly indicates the ability of the control mechanism to operate under various congestion levels.

The performance of the algorithm was also tested for varying congestion levels at smaller values of θ_1 . Plots 2 and 3 of Figure 5.14 indicate that the fraction of time spent at 16 kb/s for various Link2 loads is less than the corresponding values for Plot 1. Furthermore, the increase in time spent at the higher coding rate is much lower for smaller values of θ_1 , the implication of this being that, for smaller values of the congestion

Figure 5.14 Fraction of Time spent at 16 kb/s vs Decreasing Link2 Load

1: $\theta_1 = 3240$ bits 2: $\theta_1 = 2592$ bits 3: $\theta_1 = 2160$ bits

Link1 Load = 0.8, BECN_clear_threshold = 77 frames, $\bar{h}_T = 45$ ms

threshold, the voice call spends more time than required at the lower coding rate. These results reinforce the inference of the previous section that setting the congestion threshold to a smaller value may result in lower voice quality than is required for that particular level of congestion.

5.7 Summary

This chapter illustrated the impact that network congestion has on the QoS requirements of conversational voice. A congestion control mechanism was proposed for voice traffic, involving rate reduction in conjunction with feedback received from the network. The effectiveness of the proposed mechanism was validated by simulation results and subjective voice tests. Also investigated in this chapter is the sensitivity of the control mechanism to various thresholds and optimal threshold values were recommended for use

with the reference connection. Finally, the adaptive rate control mechanism was tested under varying levels of congestion and it was shown that it met the objective of well controlled reduction in voice quality as congestion increased.

Chapter 6

Conclusions and Recommendations for Further Study

6.1 Conclusions

This thesis has presented a consolidated study of voice transmission over FR networks, having investigated two of the principal challenges that have to be overcome in order to have satisfactory conversational voice quality: voice synchronization/reconstruction in the face of stochastic network delays and congestion control in the presence of network congestion. Although the performance evaluation was limited to a specific FR network configuration, the proposed voice synchronization algorithm is generic and can be applied to any packet-switching network which has no provision for network layer processing by intermediate switches and in which no absolute time reference is maintained between the sender and receiver. Similarly, the adaptive rate control scheme can be applied to any network which has provision for monitoring the queue lengths of buffers and generation of explicit congestion notifications.

Previous research [25, 26] specified a first approach to voice synchronization suitable for FR networks, namely the *Adaptive Buildout Delay*, designed to absorb network delay jitter by periodic, adaptive variation of the holding time of voice packets at the playout buffer. This thesis outlines an algorithmic version of the original playout mechanism and reevaluates its performance under bursty data traffic with a significant percentage of large frames. Performance evaluation of the adaptive algorithm in the context of inter-LAN

ethernet traffic has shown that large LAN packets created by paging and file transfer traffic caused excessive queueing delays and significant variability in the network delays of voice packets. In particular, delay spikes led to correlated losses in the buffer. Subjective analysis showed severe degradation of reconstructed speech quality due to the extreme value of loss rate and the occurrence of burst losses.

Limitation of delay spikes, shown to be desirable to achieve the target QoS for conversational voice, may be accomplished by limiting the maximum size of data frames. Segmentation of large LAN packets was proposed at the source FAD prior to transmission over the low-speed access links. The simulation study showed that a maximum segment size of 400 bytes would allow the voice loss rate objective to be met at high loadings. Performance results indicated a vast improvement due to data frame segmentation. The measured values of L_a and B_a were well below the maximum limits of 2% and 114 ms respectively, for all levels of destination link utilization.

On the basis of the specific improvement in voice performance observed for FR networks, it can be inferred that limitation of data frame size, accomplished by frame segmentation or otherwise, is a generic concept that can be applied to any packet-switching network characterized by: a mixed voice-data environment, small size of voice frames, variable length data frames and constrained by access links operating at 64 kb/s or less. The maximum frame segment size of 400 bytes can serve as a benchmark value for network traffic composition similar in kind to that used in the simulation.

An analysis of loss rates obtained using the original playout algorithm indicated that the increasing trend in loss rates at moderate link utilizations is directly related to the maximum size of the delay adjustment, lower setting of $\bar{h}_T$, and the slow response of the algorithm to burst losses due to delay spikes. The original algorithm responds too slowly to spikes and may establish a new holding time that is inappropriate for the subsequent averaging interval.

A modified holding-time algorithm was proposed for faster response to delay spikes and improved adaptation to variability in network delays. Modifications to the original algorithm included a limit on the maximum size of dynamic adjustment and monitoring of losses within an averaging interval to determine if a preset loss threshold has been breached. Immediate adaptation of holding-time is incorporated as a fast response to burst losses as well as a preventive measure in anticipation of future successive losses. Additionally, the immediate holding-time adaptation procedure is characterized by an expansion step size based on the nature of losses, as determined by the proximity_threshold, and an incremental adjustment approach in order to minimize the impact of TSM on the reconstructed speech quality.

A 50% reduction was observed in average loss rates for various loadings with minimal increase in mean variable end-to-end delay. Subjective tests showed a marked improvement in voice performance with minimal audible faults in speech. Based on the successful functioning of the modified adaptive playout algorithm in the context of variable delays caused by inter-LAN ethernet traffic, it may be concluded that the algorithm can be used as an effective voice synchronization mechanism in other packet-switching networks which have to accommodate the real-time constraints of voice applications in the absence of an absolute time reference between the transmitter and receiver. The parametric nature of the proposed modifications provides the end-user with flexibility in setting the thresholds based on the network configuration, expected traffic composition, voice loss distribution and the QoS requirements of the voice application.

Larger, optimized, holding-time adjustments were included in the adaptation algorithm when the reference connection comprised multiple active voice calls in view of the larger number of losses within short-term time intervals and frequent adaptation due to minimum size of holding-time adjustment.

A comparative performance analysis of the time-stamp method of the PVP and the modified ABD indicated that the average loss rates were comparable for both the methods

at all link loads. However, unlike the modified ABD, the PVP was unable to adapt to the delay variance caused by delay spikes, resulting in a higher number of correlated losses that affected the speech quality. Furthermore, it was observed that, in contrast to the characteristic fixed buildout delay for all link loadings using the time-stamp method, using the ABD, connections with shorter mean network delay have the advantage of a shorter buildout delay. The PVP approach achieves a lower loss-rate than the maximum acceptable rate at the cost of greater than necessary delay, which may be a more important consideration in voice quality for some connections.

Re-evaluation of the maximum size of the data frames, in view of the improved performance with the modified algorithm, demonstrated that the segment size may be increased to 500 bytes while still achieving the target QoS. Larger data segment sizes undermine the effectiveness of the algorithm in responding to burst losses caused by delay spikes, reaffirming the earlier observation that, given the bimodal nature of ethernet packet size distribution, it is necessary to bound the queueing delay caused by large frames so as to accommodate the stringent constraints of conversational voice.

Qualitative evaluation of speech indicated that despite limitation of large LAN frames to a maximum of 5%, and restriction of last link load, significant voice distortion was observed. Therefore, it is concluded that even a small number of unsegmented LAN frames can significantly impair voice conversation, necessitating strict enforcement of data frame segmentation at the user-network interface.

A dynamic rate control mechanism was implemented using reduction of VBR voice sources from the higher to lower coding rate in response to explicit congestion notifications from the FR network. The effectiveness of the proposed control strategy was validated by quantitative results, demonstrating a 60% reduction in the average loss rate and improvement in voice quality compared to an uncontrolled network. An optimal value of the BECN_clear_threshold was derived from simulations to prevent premature adaptation to the higher codec transmission rate and to limit the frequent oscillations

between the dual transmission rates that were observed due to transients in queue lengths caused by repeated periods of congestion. The congestion control scheme was evaluated using the two available coding rates. Clearly a finer granularity of control could be obtained with multiple coding rates.

The sensitivity of the control strategy to various settings of the congestion threshold, θ_1 , was also examined. It was found that a θ_1 value equal to 75% of the absolute threshold served as an adequate indicator of congestion onset. Smaller values of θ_1 typically lead to longer periods of operation at the lower rate of 8 kb/s than is otherwise required to achieve the target QoS. This value of $\theta_1 = 0.75 * \theta_2$ can therefore serve as a guideline for network congestion studies involving a small sized buffer for priority voice traffic. The adaptive rate control scheme was also shown to be robust under various levels of network congestion and it was shown that the mean voice duration followed the intuitive trend of longer periods of operation at the higher coding rate of 16 kb/s as the level of network congestion was reduced. The implication of this is that, the adaptive rate control algorithm was able to handle increasing network congestion gracefully, resulting in a controlled reduction of voice quality.

6.2 Recommendations for Further Study

- Subjective analysis of speech quality, using the adaptive playout algorithm and congestion control scheme, was based on informal listening tests performed using a limited number of voice files. A more formal qualitative analysis could be based on subjective ratings by a substantial number of listeners, using the Mean Opinion Score (MOS) scale. A score of 3.5 on the MOS scale can denote “communications quality”. This level of quality is characterized by a level of speech degradation that is easily detectable but not bad enough to impede natural telephone communication [40].

- Real-time implementation of the reference speech coders can be used to determine the impact of delays on speech quality.
- The research assumed that the size of the playout buffer was infinite and losses due to receiver buffer overflow weren't considered. The accuracy of the performance evaluation could be extended by limiting the size of the playout buffer and observing losses that may occur due to overflow caused by voice frames arriving early after having experienced small queueing delays in the network.
- The thesis has demonstrated the suitability of the modified adaptive buildout delay mechanism for use with narrowband telephonic speech, given the importance of applications involving conversational voice quality in the telecommunications industry. Emerging applications such as digital audio broadcasting, multimedia/hypermedia and satellite TV have led to the use of wideband monophonic and stereophonic audio in current commercial technology and the corresponding development of high quality audio compression schemes.

The MPEG Layer III audio coder is capable of near transparent quality of monophonic CD audio signals at 64 kb/s [56] and excellent stereophonic quality at 128-160 kb/s [57]. The modified playout algorithm could also be used for wideband speech and audio with appropriate setting of the algorithmic parameters and thresholds, taking into account the use of higher speed access links and subjective improvements in represented speech quality.

- The modified Adaptive Buildout Delay mechanism could also be tested as an effective voice synchronization scheme for packetized audio applications in Wide-Area Networks, specifically the Internet. Internet voice traces can be used to obtain voice loss statistics and the ability of the modified algorithm to respond to bursty losses may be evaluated. The target loss and end-to-end delay requirements will, of course, have to

be modified to accommodate larger delays and greater number of losses to be expected in the Internet.

- A conditional “maximum delay preemptive priority” could be implemented at the low-speed links, in which preemption of data frames would be initiated if a voice packet was delayed beyond a certain maximum permissible delay. This would be analogous to the delay incurred due to the maximum data segment size obtained from this simulation study. Performance results could be compared to those obtained using data frame segmentation.
- An additional interesting result would be the effect on data throughput of using this conditional preemption mechanism at the low-speed link buffers.
- Use of the FECN bit to indicate temporary queue buildup in the priority buffer due to transmission of large LAN frames needs to be investigated. A FECN_counter could be maintained at the receiver and its value could be used in conjunction with observation of $\bar{h}_q$ and burst losses to achieve holding-time adaptation.
- The effectiveness of the rate control scheme could be investigated under a extended variety of network conditions, such as: congestion caused by node or link failure, congestion caused by voice traffic alone, impact of additional FR switches in the path of the reference connection, monitoring the output queue lengths of all the voice buffers in the FR switches to determine congestion onset and setting the BECN bit on reverse traffic, to inform VBR voice sources at network periphery.

Appendix: Additional Details of the Adaptive Buildout Delay Algorithm

- (1) $I(n,k) = (n-k) * 20 \text{ ms}$ for reference speech coders
- (2) If playout length of the previous frame (kth frame) has been altered due to TSM, then, the playout time of the current frame (nth frame) is calculated taking into the account the time increment, t_{tsm} , that was used to make dynamic adjustment to the target mean holding time. Therefore the effective playout time of the current frame is offset by an amount equivalent to $\pm t_{tsm}$:
 - Set $T_p(n) := T_p(k) + \text{sgn}(E)t_{tsm} + I(n,k)$
- (3) If $n - k > 1$, indicating frame loss due to buffer overflow then if modified ABD is employed:
 - $\text{local_lost_frame_count} := \text{local_lost_frame_count} + (n - k - 1)$
- (4) The condition $(1 < \text{frame_count} < A)$, translates to $(\text{frame_count} \neq 1)$, since frame_count is reset to 1 at the end of an averaging interval. $(\text{frame_count} = 1)$ suggests that the current frame is the last frame of an averaging interval, whose playout length may have to modified based on the TSM value specified by the Time Scale Modification procedure.

In order to prevent a conflicting situation that may arise if the specified TSM value is overridden, the Immediate Holding Time Adaptation procedure is invoked only when $\text{frame_count} \neq 1$, i.e; only when a frame has been lost within an averaging interval. Otherwise a mismatch may occur between the theoretical playout length determined by the algorithm, which includes the cumulative increase due to TSM procedure as well as Immediate Holding Time Adaptation procedure, and the time period for which the voice packet is actually played out by the voice decoder,

which is determined only by the final trailer specification in the voice frame.

When `frame_count = 1`, in the event of occurrence of a late arrival, the only step taken is incrementing the `local_lost_frame_count` followed by discarding the voice frame.

References

- [1] ANSI T1.606, "Integrated Services Digital Network (ISDN) - Architectural Framework and Service Description for Frame Relaying Bearer Service", 1990.
- [2] "Frame Relay: Networks for Tomorrow and Today", The Frame Relay Forum, World Wide Web: [http:// frame-relay.indiana.edu](http://frame-relay.indiana.edu), 1994.
- [3] W. S. Lai, "Frame Relaying Service: An Overview", *Proc. Infocom '89*, Ottawa, Canada, pp. 668-673.
- [4] ANSI T1.617, "Integrated Services Digital Network (ISDN) - Signaling Specification for Frame Relay Bearer Service for Digital Subscriber Signaling System Number 1 (DSS1)", Oct. 1991.
- [5] ANSI T1.618, "Integrated Services Digital Network (ISDN) - Core Aspects of Frame Protocol for use with Frame Relay Bearer Service", Oct. 1991.
- [6] J. Lamont, J. Doak and M. Hui, "LAN Interconnection via Frame Relaying", *Proc. Infocom '89*, Ottawa, Canada, pp. 686-690.
- [7] R. J. Cheruki, and J. F. Derby, "Frame Relay: Protocols and Private Network Applications", *Proc. Infocom '89*, Ottawa, Canada, pp. 676-684.
- [8] D. Minoli, "*Enterprise Networking*", Artech House, Boston, 1993.
- [9] "Frame Relay Special Report: 1994 and Beyond", The Frame Relay Forum, World Wide Web: [http:// frame-relay.indiana.edu](http://frame-relay.indiana.edu), 1994.
- [10] "Frame Relay Forum Members and Products", The Frame Relay Forum, World Wide Web: [http:// frame-relay.indiana.edu](http://frame-relay.indiana.edu), 1994.
- [11] "When Frame Relay Meets ATM", The Frame Relay Forum, World Wide Web: [http:// frame-relay.indiana.edu](http://frame-relay.indiana.edu), 1994.

- [12] "Frame Relay/ATM PVC Network Interworking Implementation Agreement", Document No. FRF.5, The Frame Relay Forum, World Wide Web: [http:// frame-relay.indiana.edu](http://frame-relay.indiana.edu), 1994.
- [13] G. J. Coviello, "Comparative Discussion of Circuit vs. Packet-Switched Voice", *IEEE Trans. Commun.*, Vol. COM-27, No. 8, pp. 1153- 1160, Aug. 1979.
- [14] W. L. Hoberecht, "A Layered Network Protocol for Packet Voice and Data Integration", *IEEE J. Sel. Areas Commun.*, Vol. SAC-1, No. 6, pp. 1006-1013, Dec. 1983.
- [15] J. G. Gruber, "Delay Related Issues in Integrated Voice and Data Networks", *IEEE Trans. Commun.*, Vol. COM-29, No. 6, pp. 786-800, June 1981.
- [16] J. G. Gruber and L. Strawczynski, "Subjective Effects of Variable Delay and Speech Clipping in Dynamically Managed Voice Systems", *IEEE Trans. Commun.*, Vol. COM-33, No. 8, pp. 801-808, Aug. 1985.
- [17] T. Suda, H. Miyahara, and T. Hasegawa, "Performance Evaluation of a Packetized Voice System - Simulation Study", *IEEE Trans. Commun.*, Vol. COM-32, No. 1, pp. 97-102, Jan. 1984.
- [18] G. Barberiz and D. Pazzaglia, "Analysis and Optimal Design of a Packet-Voice Receiver", *IEEE Trans. Commun.*, Vol. COM-28, No. 2, pp. 217-222, Feb. 1980.
- [19] J. S. Turner, and L. F. Wyatt, "A Packet Network Architecture for Integrated Services", *Proc. IEEE Globecom'83*, June. 1983, pp. 45-50.
- [20] W. A. Montgomery, "Techniques for Packet Voice Synchronization", *IEEE J. Sel. In Areas Commun.*, Vol. SAC-1, No. 6, pp. 1022-1028, Dec. 1983.
- [21] I. Dalgic, W. Chien, and F. A. Tobagi, "Evaluation of 10Base-T and 100Base-T Ethernet Carrying Video, Audio and Data Traffic", *Proc. IEEE Infocom '94*, Toronto, Canada, pp. 1094-1102.
- [22] R. Ramjee, J. Kurose, D. Towsley and H. Schulzrinne, "Adaptive Playout Mechanisms for Packetized Audio Applications in Wide-Area Networks", *Proc. IEEE Infocom '94*, Toronto, Canada, pp. 680-688.

- [23] S. Casner and S. Deering, "First IETF Internet Audiocast", *ACM Computer Communication Review*, Vol. 22, No. 3, pp. 92-97, July 1992.
- [24] International Telegraph and Telephone Consultative Committee, *CCITT Recommendation G. 764, "Voice Packetization - Packetized Voice Protocols"*, Aug. 1990.
- [25] L. Dong, A. R. Kaye, and S.A. Mahmoud, "Transmission of Compressed Voice over Integrated Services Frame Relay Networks: Priority Service and Adaptive Buildout Delay", *IEE Proc. Commun.*, 141(4): 265-274. Aug. 1994.
- [26] L. Dong, "Compressed Voice in Integrated Services Frame Relay Networks", M. Eng thesis, Carleton University, Dec. 1992.
- [27] A. R. Kaye and S. Zhang, "Congestion Control in Integrated Voice-Data Frame Relay Networks and the Case for Embedded Coding", *Proc. IEEE Globecom '94*, San Francisco, CA, Dec. 1994, pp. 1565-1570.
- [28] S. Zhang, "Congestion Control in Frame Relay Networks with Variable Bit Rate Compressed Voice and Data Traffic", M. Eng thesis, Carleton University, Sept. 1993.
- [29] S. Mahant, M. H. Sherif and K. Sriram, "Simulation of Multiservice Frame Relay Operation", AT&T Document T1Y1.2/91-056, Oct. 1991.
- [30] B. Bergman and T. Hatala, "Voice Over Frame Relay Data Networks", Frame Relay Forum/94-017R1, June 1994.
- [31] C. M. Sharon, M. Devetsikiotis, I. Lambadaris, and A. R. Kaye, "Rate Control of VBR H.261 Video on Frame Relay Networks", *IEEE ICC '95*, Seattle, Washington, June 1995, pp. 1443-1447.
- [32] C. M. Sharon, "Compressed Video in Integrated Services Frame Relay Networks", M. Eng thesis, Carleton University, Dec. 1994.
- [33] "Switched Virtual Circuit Implementation Agreement", Document No. FRF.4, The Frame Relay Forum, World Wide Web: <http://frame-relay.indiana.edu>, 1994.
- [34] T. Bradley, C. Brown and A. Malis, "Multiprotocol Interconnect over Frame Relay", RFC 1490, 1993.

- [35] D. J. Goodman, G. B. Lockhart, O. J. Wasem, and W. C. Wong, "Waveform Substitution Techniques for Recovering Missing Speech Segments in Packet Voice Communications", *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. 34, no. 6, pp. 1440-1448, Dec. 1986.
- [36] O. J. Wasem, D. J. Goodman, C. A. Dvorak, and H. G. Page, "The Effect of Waveform Substitution on the Quality of PCM Packet Communications", *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. 36, no. 3, pp. 342-348, March 1988.
- [37] J. Suzuki and M. Taka, "Missing Packet Recovery Techniques for Low-Bit-Rate Coded Speech", *IEEE J. Sel. Areas in Comm.*, Vol. 7, no. 5, pp. 707-717, June. 1989.
- [38] W. Stallings, "Data and Computer Communications", fourth edition, New York: Macmillan, 1994.
- [39] I. A. Gerson and M. A. Jasiuk, "Vector Sum Excited Linear Prediction (VSELP) Speech Coding at 8 kbps", *Proc. ICASSP*, pp. 461-464, 1990.
- [40] N. S. Jayant, "High-Quality Coding of Telephone Speech and Wideband Audio", *IEEE Communications Magazine*, pp. 10-20, Jan. 1990.
- [41] W. P. LeBlanc, "CELP Speech Coding at Low to Medium Bit Rates", Ph.D thesis, Carleton University, October 1992.
- [42] T. W. Leung, "Voice Frame Reconstruction Methods for CELP Speech Coders", M. Eng thesis, Carleton University, Sept. 1993.
- [43] T. W. Leung, W. P. Leblanc and S. A. Mahmoud, "Voice Frame Reconstruction Methods for CELP Speech coders", *Proc. Wireless '93*, Calgary, Canada, July 1993.
- [44] S. Taylor, "Frame Relay 1995: Carrier and User Profile Update", The Frame Relay Forum, World Wide Web: <http://frame-relay.indiana.edu>, 1994.
- [45] J. F. Shoch and J. A. Hupp, "Measured performance of an Ethernet local network", *Commun. ACM*, Vol. 23, No. 12, pp. 711-721, Dec. 1980.
- [46] R. Gusella, "A Measurement study of Diskless Workstation Traffic on an Ethernet", *IEEE Trans. Commun.*, Vol. COM-38, No. 9, pp. 1557-1568, Sept. 1990.

- [47] H. J. Fowler and W.E. Leland, "Local Area Traffic Characteristics, with Implications for Broadband Network Congestion Management", *IEEE J. On Sel. Areas in Comm.*, Vol. 9, no. 7, pp. 1139-1149, Sept. 1991.
- [48] N. Yin and M. G. Hluchyj "Simple Models for Statistically Multiplexed Data Traffic in Cell Relay Networks," *Proc. IEEE Globecom '93*, Vol. 2, pp. 824-829.
- [49] W. E. Leland and D. V. Wilson, "High Time-Resolution Measurement and Analysis of LAN traffic: Implications for LAN Interconnection," *Proc. IEEE Infocom '91*, Bal Harbour, FL, April. 1991, pp. 1360-1366.
- [50] J. Walrand, "Communication Networks: A First Course", Boston: Richard D. Irwin Inc., 1991.
- [51] R. Ansari and A. R. Kaye, "Compressed Voice in Integrated Services Frame Relay Networks: Voice Synchronization," *Proc. Canadian Conference on Electrical and Computer Engineering*, Montreal, Canada, Sept. 5-8, 1995.
- [52] J. Banks, and J. S. Carson, "Discrete-Event System Simulation", Prentice-Hall, 1984.
- [53] R. Jain, "The Art Of Computer Systems Performance Analysis", John Wiley & Sons Inc., 1991.
- [54] A. J. Vernon, T. Jawor, S. Rabie et al, "Congestion Control Strategies for Mixed X.25/Frame Relay Networks", *IEEE ICC '94*, New Orleans, May 1994, pp. 1621-1628.
- [55] V. Jacobson, "Congestion Avoidance and Control", *Proc. SIGCOMM'88*, California, LA, pp. 314-329.
- [56] S. Boland and M. Deriche, "High Quality Audio Coding Using Multipulse LPC and Wavelet Decomposition", *Proc. ICASSP*, pp. 3067-3069, 1995.
- [57] J. Princen and J. D. Johnston, "Audio Coding with Signal Adaptive Filterbanks", *Proc. ICASSP*, pp. 3071-3073, 1995.

END

0 5-0 6-9/6

FIN