HCV2_8K	A low speed coding algorithm which only requires a bandwidth of 8 Kbit/s
HDLC	High-level Data Link Control
ISDN	Integrated Services Digital Network
LAPD	Link Access Protocol-D Channel
MCT	Mild Congestion Threshold
MOS	Mean Opinion Score
PBX	Private Branch Exchange
PCME	Packetized Circuit Multiplication Equipment
PVC	Permanent Virtual Circuit
PVP	Packetized Voice Protocol
QOS	Quality of service
SCT	Severe Congestion Threshold

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

Voici les notes d'usage

Voici les notes d'usage

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.

Canada

Congestion Control in Frame Relay Networks with Variable Bit Rate Compressed Voice and Data Traffic

Shu Zhang, M.Eng.

University of Beijing, Science and Technology
Beijing, P.R. China

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 for Electrical Engineering
Department of Systems and Computer Engineering

Carleton University
1125 Colonel By Drive
Ottawa, Ontario, Canada, K1S 5B6

September 1, 1993

© Shu Zhang, 1993

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 0-315-89901-8

Canada

Name SHUO ZHANG

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.

electronics and Electrical

SUBJECT TERM

0544

SUBJECT CODE

U·M·I

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

PHILOSOPHY, RELIGION AND THEOLOGY

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

SOCIAL SCIENCES

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

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 0615
General
International Law and 0616
Relations
Public Administration 0617
Recreation 0814
Social Work 0452
Sociology 0626
General
Criminology and Penology 0627
Demography 0938
Ethnic and Racial Studies 0631
Individual and Family 0628
Studies
Industrial and Labor 0629
Relations
Public and Social Welfare 0630
Social Structure and 0700
Development
Theory and Methods 0344
Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES

Agriculture 0473
General 0285
Agronomy
Animal Culture and 0475
Nutrition
Animal Pathology 0476
Food Science and 0359
Technology
Forestry and Wildlife 0478
Plant Culture 0479
Plant Pathology 0480
Plant Physiology 0817
Range Management 0777
Wood Technology 0746
Biology 0306
General 0287
Anatomy 0308
Biostatistics 0309
Botany 0379
Cell 0329
Ecology 0353
Entomology 0369
Genetics 0793
Limnology 0410
Microbiology 0307
Molecular 0317
Neuroscience 0416
Oceanography 0433
Physiology 0821
Radiation 0778
Veterinary Science 0472
Zoology
Biophysics 0786
General 0760
Medical

Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Polynology 0427
Physical Geography 0358
Physical Oceanography 0410

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 0754
Mental Health 0347
Nursing 0569
Nutrition 0570
Obstetrics and Gynecology 0380
Occupational Health and 0354
Therapy 0381
Ophthalmology
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 0485
General 0749
Agricultural 0486
Analytical 0487
Biochemistry 0488
Inorganic 0738
Nuclear 0490
Organic 0491
Pharmaceutical 0494
Physical 0495
Polymer 0754
Radiation 0405
Mathematics 0605
Physics 0986
General
Acoustics
Astronomy and 0606
Astrophysics
Atmospheric Science 0608
Atomic 0748
Electronics and Electricity 0607
Elementary Particles and 0798
High Energy
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 0537
General 0538
Aerospace 0539
Agricultural 0540
Automotive 0541
Biomedical 0542
Chemical 0543
Civil 0544
Electronics and Electrical 0348
Heat and Thermodynamics 0545
Hydraulic 0546
Industrial 0547
Marine 0794
Materials Science 0548
Mechanical 0743
Metallurgy 0551
Mining 0552
Nuclear 0549
Packaging 0765
Petroleum 0554
Sanitary and Municipal 0790
System Science 0428
Geotechnology
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

**“Congestion Control in Frame Relay Networks with Variable
Bit Rate Compressed Voice and Data Traffic”**

Submitted by

Shu Zhang

in partial fulfillment of the requirements for the
degree of Master of Engineering

Thesis Supervisor

Chairman,
Department of Systems and Computer Engineering

Ottawa-Carleton Institute for Electrical Engineering,
Department of Systems and Computer Engineering
Carleton University
September 1, 1993

Abstract

This thesis proposes an effective congestion control strategy for frame relay networks with combined data and compressed voice traffic. For voice traffic, the control strategy guarantees voice reconstruction with good quality, while for data traffic, it guarantees high transmission reliability (low loss rate). The proposed congestion control procedures include: implementation of voice priority over data; independent congestion detection for voice and data; use of the Consolidated Link Layer Management (CLLM) for congestion notification; source responses to congestion notification including dropping low priority voice frames and reducing the data sending rate; and switch response to the congestion notification such as discarding the Discard Eligible (DE) frames when the voice traffic is severely congested. The quality of voice recovery is enhanced by random setting of the DE bit. Performance of the proposed control scheme was evaluated through simulation. Some modifications of the present frame relay network structure are proposed, without violation of the standard frame relay protocols and without increasing the network complexity substantially.

Acknowledgments

I would like to thank my thesis supervisor Prof. A. R. Kaye for his guidance and encouragement during the course of this research.

I am grateful to the Telecommunications Research Institute of Ontario, Natural Sciences and Engineering Research Council, and Carleton University, for the financial support that made my studies possible.

To:
My Parents and My Family

List of Acronyms

ACT	Absolute Congestion Threshold
ADPCM	Adaptive Differential Pulse Code Modulation
ANSI	American National Standards Institute
BECN	Backward Explicit Congestion Notification
CCITT	International Telegraph and Telephone Consultative Committee
CLLM	Consolidated Link Layer Management
CRC	Cyclic Redundancy Code
DE	Discard Eligible
DLC	Data Link Connection
DLCI	Data Link Connection Identifier
EA	Extension Address
EADPCM	Embedded Adaptive Differential Pulse Code Modulation
EHCV2_16K	A embedded speech coding algorithm which is based on HCV2_8K, requires a variable bandwidth of from 8Kbit/s to 16 Kbit/s, and has a very high MOS
FAD	Frame Assembler Disassembler
FCS	Frame Check Sequence
FECN	Forward Explicit Congestion Notification
FR	Frame Relay

HCV2_8K	A low speed coding algorithm which only requires a bandwidth of 8 Kbit/s
HDLC	High-level Data Link Control
ISDN	Integrated Services Digital Network
LAPD	Link Access Protocol-D Channel
MCT	Mild Congestion Threshold
MOS	Mean Opinion Score
PBX	Private Branch Exchange
PCME	Packetized Circuit Multiplication Equipment
PVC	Permanent Virtual Circuit
PVP	Packetized Voice Protocol
QOS	Quality of service
SCT	Severe Congestion Threshold

TABLE OF CONTENTS

Abstract.....	iii
Acknowledgments.....	iv
Acronyms.....	vi
List of Figures.....	xi
 Chapter 1. Introduction	 1
1.1 Background.....	1
1.2 Previous Research.....	1
1.3 Thesis Objective.....	2
1.4 Thesis Outline	3
 Chapter 2. Frame Relay, Voice Characteristic and Congestion Control	 4
2.1 Frame Relay	4
2.1.1 Protocol.....	4
2.1.2 Frame Structure.....	7
2.1.3 Frame Relay Service.....	10
2.1.4 Strengths and Weaknesses of FR Networks	10
2.2 Compressed Voice Characteristics	11
2.2.1 Delay sensitivity.....	11
2.2.2 Throughput Requirement.....	12
2.2.3 Loss and Loss Pattern	13
2.2.4 Silence Period and Talk Spurt.....	14
2.2.5 Compressed Voice	15
2.3 Congestion Control	17
2.3.1 Implicit Signalling	17
2.3.2 Explicit Signalling	17
2.4 Modelling of the Voice Frame Relay Network.....	19
2.4.1 Network Element	19
2.4.2 Network Configuration	22
2.4.3 Reference Connection.....	22
2.4.4 Traffic Model	23
 Chapter 3. QOS Requirement, Congestion Detection and Notification	 25
3.1 QOS Requirement.....	25
3.2 Buffer Management for Congestion Detection.....	27

3.3 Congestion Status	27
3.4 Buffer Management for Input Queues.	29
3.5 Buffer Management at Output Queues.	30
3.6 The CLLM Notification Mechanism	32
3.7 Buffer Thresholds for Voice as a Function of Output Line Speed	33
3.8 Voice Loss Behavior as the Function of the Buffer Threshold	33
Chapter 4. The Proposed Congestion Control System	35
4.1 Congestion Control for Voice.....	35
4.1.1 Implementation of Voice Priority over Data	35
4.1.2 FAD Response to Explicit Notification	36
4.1.3 Random Setting of DE Bit	36
4.2 Congestion Control for Data Traffic.....	36
4.3 Congestion Control Procedure at Intermediate Nodes.....	37
4.3.1 Reaction to Congestion Status at the Input Queue.....	37
4.3.2 Reaction to Congestion Status at the Priority Output Queue.....	37
4.3.3 Reaction to Congestion Status at the Non-Priority Output Queue	38
4.4 Locating the Network Bottleneck	39
4.5 Performance of the Proposed Congestion Control Scheme	41
4.6 Conclusion	45
Chapter 5. Optimal Parameter Setting and Sensitivity Test	48
5.1 Comparison between BECN and CLLM	48
5.1.1 Comparison between Voice BECN and Voice CLLM	48
5.1.2 Comparison between Data BECN and Data CLLM	51
5.2 Parameter Optimizing	52
5.2.1 Random Setting of the DE Bit	52
5.2.2 Voice Frame Length	53
5.2.3 Data Buffer Size.....	59
5.2.4 MCT	61
5.2.5 SCT	64
5.2.6 CLLM frequency	67
5.3 Sensitivity Tests.....	68
5.3.1 Sensitivity to Access and Output Line Speeds	68
5.3.2 Sensitivity to CLLM Processing Time	70
5.3.3 Sensitivity to Background Traffic Levels	73
5.3.4 Sensitivity to Length of Silence and Talkspurt Periods.....	76
5.3.5 Sensitivity to the Number of Hops.....	76
Chapter 6. Conclusion	79

6.1 Conclusions.....	79
6.2 Suggestion for Further Studies.....	80

List of Figures

Figure 2.1	Functional Architecture of FR Network.....	5
Figure 2.2	Frame Relay Frame Structure.....	8
Figure 2.3	Voice frame structure.....	16
Figure 2.4	The Domain of the explicit Congestion Notification Mechanisms	18
Figure 2.5	Frame Assembler--Disassembler.....	20
Figure 2.6	The queueing model of the FR switch.....	21
Figure 2.7	Networks Configuration	22
Figure 2.8	Reference Connection	23
Figure 3.1	Classification of Congestion Levels.	28
Figure 3.2	Classification of Congestion Levels.	28
Figure 3.3	Relation between queue length and congestion status	29
Figure 3.4	Congestion Status Indicator.....	30
Figure 3.5	Congestion Status in the Input Queue	31
Figure 3.6	Congestion Status in the Output Queue....	31
Figure 4.1	Reaction to the congestion status at input queue.....	38
Figure 4.2	Reaction to the congestion status at the priority output queue.....	39
Figure 4.3	Reaction to the congestion status at the non-priority output queue.....	40
Figure 4.4	Voice throughput vs. voice offered load.....	42
Figure 4.5	Voice delay vs. voice offered load.....	43
Figure 4.6	Voice loss rate vs. voice offered load	44
Figure 4.7	Voice loss burst length vs. voice offered load	45
Figure 4.8	Voice throughput comparison between with voice CLLM and without voice CLLM.....	46
Figure 4.9	Voice queueing delay comparison between with voice CLLM and without voice CLLM.....	46
Figure 4.10	Voice loss rate comparison between with voice CLLM and without voice CLLM	47
Figure 5.1	Difference of voice throughput between voice BECN and voice CLLM	49
Figure 5.2	Difference of voice delay between voice BECN and voice CLLM	50
Figure 5.3	Difference of voice loss rate between voice BECN and voice CLLM.....	50
Figure 5.4	Difference of data throughput between voice BECN and voice CLLM	51
Figure 5.5	Difference of data throughput between data BECN and data CLLM	52
Figure 5.6	The length of voice loss burst vs. the percentage of DE frames	53

Figure 5.7	The voice loss rate vs. the percentage of DE frames.....	54
Figure 5.8	Effect of voice frame length on voice throughput.....	55
Figure 5.9	Effect of variation of voice frame size on Mean queuing delay.....	55
Figure 5.10	Effect of variation of voice frame size on the voice end to end delay	58
Figure 5.11	Effect of variation of voice frame size on voice loss rate	59
Figure 5.12	Effect of variation of voice frame size on loss burst length	60
Figure 5.13	Effect of variation of voice frame size on no loss period length....	60
Figure 5.14	Data Loss Rate as the Function of the Buffer Size.....	61
Figure 5.15	Voice frame loss rate vs. MCT	62
Figure 5.16	Voice queuing delay vs. MCT.....	63
Figure 5.17	Voice throughput vs. MCT	64
Figure 5.18	Length of voice loss burst vs. the SCT.....	65
Figure 5.19	The voice loss rate vs. the SCT	66
Figure 5.21	The voice throughput vs. the SCT	66
Figure 5.20	The voice delay vs. the SCT.....	67
Figure 5.22	Voice throughput vs. intergenerating time of CLLM message.....	68
Figure 5.23	Voice delay vs. intergenerating time of CLLM message.....	69
Figure 5.24	Voice loss burst length vs. intergenerating time of CLLM message	69
Figure 5.25	Voice loss rate vs. intergenerating time of CLLM message	70
Figure 5.26	The voice loss rate vs. the access/output line speed.....	71
Figure 5.27	The voice delay vs. the access/output line speed	71
Figure 5.28	The voice throughput vs. the access/output line speed.....	72
Figure 5.29	Effect of background traffic variation on voice throughput	74
Figure 5.30	Effect of background traffic variation on voice delay	74
Figure 5.31	Effect of background traffic variation on voice loss rate.....	75
Figure 5.32	Processor load vs. voice percentage	75
Figure 5.33	The voice delay vs. number of hops.....	78

Chapter 1. Introduction

1.1 Background

The combination of intelligent end user systems and reliable high speed communication facilities has enabled the development of such “fast packet” technologies as frame relaying and ATM based BISDN to dominate telecommunication in the later 1980’s and early 1990’s. Especially, the attractive features and benefits of frame relaying have resulted in industry implementation preceding standards into the market place [5].

Frame relaying is a new form of packet switching technique. Compared with the traditional approach of packet switching such as X.25, the minimum amount of overhead and the simplicity of the network node processing allow the Frame Relay (FR) network to achieve low latency and high throughput. The potential high efficiency of the FR networks are suitable for the various types of new application traffic, although the present application of the frame relay is still to provide data traffic service.

Frame relaying characteristics match some requirements of compressed voice traffic such as small delay and random frame loss. However, the different traffic characteristics of the compressed voice from that of data requires the FR network to have specific processing abilities for compressed voice. The simplification of the FR networks not only achieves high efficiency, but also causes some vulnerability in the control capability. The effort of this present research is to overcome the disadvantages and enjoy the advantages of FR networks in realizing the transport of combined compressed voice and data traffic in a single FR network.

1.2 Previous Research

Before this research, there have been some studies in the area of congestion control in FR networks. Bharant T. Boshi [6] suggested an overall control plan consisting of distributed real-time control, centralized network-management (near real-time) control, and long-term network engineering procedures. Kim-Joan Chen [7] studied adaptive window schemes, employing implicit detection and “stop-duration” based on explicit detection. R. Bhardwaj [9] analyzed a dynamic window congestion control. K. Ab-Hamid and A. G. Waters [46][47][48] proposed a Kalman prediction based congestion avoidance scheme, and compared it with other schemes such

as Bandwidth Management, Explicit Binary Feedback and Stop-Duration. However, they emphasize network performance for providing data service rather than voice service. This thesis intends to make congestion control strategies more practical for both compressed voice transport and data transport.

1.3 Thesis Objective

The objective of this research is to propose an effective congestion control strategy for FR networks with combined data and compressed voice traffic:

1. The quality of the reconstructed voice at the destination end point is evaluated by a mean opinion score (MOS) ranging from 1 to 5 [50]. For compressed voice transport, the control scheme should guarantee that network users achieve reconstructed voice of high quality (high MOS value, e.g. MOS=4.0) without consuming too much of the network's resources (High MOS value, but less bandwidth, so as to provide service to as many users as possible with only limited resources). The network performance is investigated using newly developed codec algorithm EHCv2-16 K [39] (which will be explained in Section 2.2) instead of EADPCM which was widely used before and consumes more bandwidth. Throughput and delay of voice frames are no longer the most significant performance parameters, because introducing voice priority over data guarantees them. The pattern of frame loss becomes more significant because it substantially affects the quality of reconstructed voice. A uniformly random distribution of voice frame loss is preferred to a highly bursty loss distribution.

2. For data transport, the control scheme should guarantee an extremely low frame loss rate, so as to provide higher layer users with good quality services.

In addition to the congestion control schemes for data and voice traffic proposed in this thesis, performance evaluation has been done for an example connection in an FR network. FR networks can be modelled as queueing networks. Due to the complexity of the FR network structure, no analytical method is currently available for performance evaluation. In this research, only some approximate analyses have been done for the input and output queues at the FR switch. Simulations have been run to evaluate the FR network performance. Most simulation results are obtained with an accuracy of 10% at a 95% confidence level, i.e.,

$$Prob (|\theta - \hat{\theta}| < 0.1) \geq 0.95$$

where $\hat{\theta}$ is an estimation of the variable θ obtained from the simulation.

Some performance criteria of the frame loss behavior require an extremely long run of simulations in order to obtain the above accuracy. Due to time limitations, these performances are obtained with a 20% or 25% accuracy.

1.4 Thesis Outline

The remainder of the thesis is organized into five chapters.

Chapter 2 introduces the protocol and functional structure of FR networks, characteristics of compressed voice traffic, and congestion control problems .

Chapter 3 presents the quality of service requirements, and the congestion detection and notification methods for voice and data traffic.

Chapter 4 details the congestion control scheme applied in the FR networks, and shows the high performance of FR networks with combined data and compressed voice traffic.

Chapter 5 discusses the optimal setting of the network parameters and sensitivity to both network and to traffic parameters.

Chapter 6 summarizes the major conclusions from this research, and proposes topics which need further studies.

Chapter 2. Frame Relay, Voice Characteristic and Congestion Control

Because the purpose of this research is to investigate the possibility of implementing both compressed voice and data transportation over frame relay networks, and to propose an effective congestion control strategy for this type of network, it is necessary to present a brief review of the frame relay network, voice characteristics and congestion control mechanisms in this introductory chapter.

2.1 Frame Relay

Frame relay (FR) is a new and simplified technique for transporting packetized data. It is a standard communications protocol specified in CCITT Recommendations I.122 and Q.922 which add relay and routing functions to the data link layer (layer 2 of OSI reference model) [31]. ANSI also issued some standards on frame relay such as ANSI T1.606-1990, ANSI T1.617-1991 and ANSI T1.618-1991 [1] [2] [3].

When a frame is received from a source device, the network looks at the address field to determine with which connection the frame is associated. At the intermediate node, the frames are checked and relayed to the next logical destination. The term “relay” implies that packets are not terminated at each switching node but are relayed or routed to their destination. Only correct frames are relayed to the next logical destination, and frames with incorrect CRCs (Cyclic Redundancy Code) are discarded.

Since frame relay provides for multiplexed service that focuses network resources on getting frames from one end of the network to the other as quickly as possible, it permits dynamic allocation of bandwidth among multiple users at a very high speed. The high speed is especially suitable for such new applications as voice and video, which have stringent requirements on delay limits.

2.1.1 Protocol

Frame relaying is a new ISDN packet model bearer service for data communications which offers link level multiplexing through the Link Access Procedure on the D channel

(LAPD). The current standards for ISDN propose that the LAPD protocol be used for the the signaling channel (D channel). However, in order to take advantage of the higher speed and transmission quality of digital circuits, it has been suggested that the LAPD protocol be used on bearer channels (B channel or H channel) as well for packetized data transfer. The functional architecture of FR network is shown in Figure 2.1. Frame Relay operates completely within the link layer. Based on the LAPD protocol, it statistically multiplexes different users' data streams. These data streams are called Data Link Connections (DLC) and the users' data streams are identified by a Data Link Connection Identifier (DLCI) at call establishment time. Subsequent to call establishment, all data frames carry this DLCI in all data streams. This is unique at the physical channel and hence is purely of local significance between nodes and has no consequence in the global sense. Therefore, the multiplexing function has been pushed down to Layer 2, the

lowest possible protocol layer for statistical multiplexing. The network periphery thus gains intelligence and the network is left with only some core functions. Therefore, the network is capable of providing increased throughput and reduced network delays because it has fewer functions to perform.

Out-of-band call control is another significant characteristic of FR networks. Unlike X.25, in which the inband signalling results in considerable overhead to process, for the FR service there is a logical separation between the channels carrying users' information and the signalling channels carrying messages for control of connections. The former is said to lie in the *user plane* (U-plane) and latter in the *control plane* (C-plane).

The traditional packet switching model, based on X.25, is implemented at all three of the lower layers of the OSI model, while FR concentrates on the core multiplexing functions, implementing some, but not all, of the functions of Level 2 and Level 3 of X.25.

Today's end user devices have more powerful processing capacities and therefore can easily perform additional functions such as link flow control and error recovery by retransmission request. Since the end point devices are capable of performing packet diagnostics and maintenance, the intermediate nodes can concentrate on performing a high speed routing function under the assumption that the digital transmission system is capable of maintaining a high level of packet integrity.

FR offers high performance on packet transportation by minimizing processing at the intermediate switching nodes. Neither flow control nor error recovery service is provided by the FR network. Error recovery is expected to be provided by the end user and flow control is supposed to be provided by a self-healing process initiated at the end user equipment. This is significant in that the designers of these protocols have assumed the presence of a high-quality digital transmission medium requiring only infrequent retransmission of errored frames. Therefore, link-by-link flow control and error recovery is not done by the intermediate node. The processing of layer 3 functionality in the U-plane is completely eliminated. Core functions are limited to rate enforcement, FCS checking, logical channel look-up, switching, and congestion control check; hence part of the layer 2 functions are performed. Thus, protocol functions such as sequence numbering, window rotation, acknowledgment, and supervisory frames are not

performed within the FR network. Stripping so many functions out of the FR network is only possible under the assumption that the digital media provides an almost error-free transmission. The effect of not performing the aforesaid functions is a dramatic increase in throughput because each frame requires much less processing, thereby reducing transportation delays, which are lower than X.25 networks.

The FR network provides the following core functions at the intermediate node using the LAPD protocol:

- frame delimiting, alignment, and transparency provided by the use of High-level Data Link Control (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 neither too long nor too short;
- detection of (but not recovery from) transmission errors; and
- congestion control functions.

2.1.2 Frame Structure

The frame-relay frame format is shown in Figure 2.2. Each frame consists of several fields: flag, address, information and FCS. A brief description of each field follows:

Flag Field. The length of a frame is variable. Frames are delimited by the flag fields at both ends with the unique pattern 01111110. The closing flag may also serve as the opening flag of the next frame.

Address Field. The minimum and default length of the address field is 2 octets. However, the address field may be extended to 3 or 4 octets. To support a larger address range, the 3-octet or 4-octet address fields may be supported at the user-network interface or the network-network interface upon bilateral agreement. The EA bit is used to indicate the end of the address field.

C/R = Command/Response, use is application-specific
 EA = Address Field Extension
 DE = Discard Eligibility Indicator
 FECN = Forward Explicit Congestion Notification
 BECN = Backward Explicit Congestion Notification
 DLCI = Data Link Connection Identifier
 FCS = Frame Check Sequence

Figure 2.2 Frame Relay Frame Structure

The *Data Link Connection Identifier* (DLCI) is used to identify the logical connection multiplexed within the physical channel, with which a frame is associated. It has local significance only.

The C/R bit is application-specific and is not used by the standard FR protocol. It is

conveyed transparently by the DL-CORE protocol between DL-CORE service users. This indicator has been added for possible future expansion of the protocol. It indicates whether the remaining six usable bits of that octet are to be interpreted as the lower DLCI bits or as DL-CORE control bits.

Forward or Backward Explicit Congestion Notification (FECN or BECN) notifies the end user that receives the notification frame that congestion-avoidance procedures should be initiated, where applicable, for traffic in the same or opposite direction of the received frame respectively. If the queue length in a frame handler begins to grow to a dangerous level, then either FECN or BECN bits, or a combination, should be set to try to reduce the flow of frames through that frame handler. These notifications indicate that the frames which the end user transmits on this logical connection have encountered or may encounter congested resources. The choice of FECN or BECN may be determined at configuration time by whether the end users on a given logical connection are prepared to respond to either of the two bits. In any case, the frame handler has some choice as to which logical connections should be alerted to congestion. If congestion is becoming quite serious, all logical connections through this frame handler might be notified. The user's response to a BECN signal is simply to reduce the rate at which frames are transmitted until the signal ceases. The response to an FECN is more complex, since it requires the user to notify its peer user on this connection to restrict the flow of frames. This notification is not supported in the frame-relay protocol and should be done at a higher layer on an end-to-end basis [44]

The DE bit, if used, is set to '1' to indicate a request that a frame should be discarded in preference to other frames in a congestion situation. Setting this bit by the network or the transmitting user is optional. No network will ever clear this bit.

Information Field. The information field contains the user data to be transported by the FR network. The content and format of the information field is of interest only to the originating and receiving entities. Network nodes will never modify the contents of the information field. The maximum size of the FR information field is a system parameter identified as N203. The default maximum information field supported by the network is 262 octets. All other maximum values are negotiated between users and networks and between networks. The minimum information field size is one octet [3].

Frame Check Sequence Field. The FCS is calculated on the address and the information fields to detect any transmission errors in these two fields. The 16-bit *cyclic redundancy check* (CRC) is used for this error check. The generator polynomial for CRC is $X^{16} + X^{12} + X^5 + 1$.

2.1.3 Frame Relay Service

FR offers two communication services: Switched Virtual Circuit (SVC) and Permanent Virtual Circuit (PVC). Since FR is a statistical rather than a fixed time division multiplexed (TDM) service, neither SVC nor PVC is limited to a particular speed. SVC is analogous to a phone call which requires call setup and tear down, while PVC is analogous to a semi-permanent/leased line. In this research, it is assumed that PVC is offered in the FR network.

PVCs are point to point connections between pairs of FR users. Unlike SVCs, PVCs are setup for longer time periods. The call setup and tear down are implemented by entering network management commands at the network control console. PVCs are considered to be a “virtual” connections, because network resources (buffers and bandwidth) are not consumed unless data is being transferred.

2.1.4 Strengths and Weaknesses of FR Networks

Every technology has its strengths and weaknesses relative to application requirements.

1. Strengths

- a). During the data transfer phase there is no control signal transmitted in the data channel so that the FR traffic is streamlined. The streamlined protocol structure allows the network to achieve high throughput and low latency. This is important for some new applications such as voice and video, which have increasingly stringent delay/throughput requirements.
- b). The application of highly reliable transmission lines simplifies the network function and structure (For example, there is no hop-by-hop error recovery and flow control procedure). This significantly reduces the cost of protocol implementation
- c). A small amount of frame overhead increases network transmission efficiency. Most of

the bandwidth is used to transmit user data and only a small amount is assigned to the transmission of overhead.

2. Weaknesses

- a). FR networks lose some control capability for effectively allocating resource and managing traffic due to the simplicity of the network structure. This makes congestion control difficult.
- b). Simple frame and network structures cause difficulty in voice control. While standard *Packetized Circuit Multiplication Equipment* (PCME) and *Voice Packetization-Packetized Voice Protocols* (PVP) specify many voice processing functions at layer 3, and their packet formats contain the sequence number and time stamp, the entire layer 3 is removed in the FR. Neither a sequence number nor a delay stamp are presented in its frame format.

2.2 Compressed Voice Characteristics

Since FR provides a network service, any type of communication (such as data, voice, video, etc.) can potentially use it as a vehicle for transportation. However, the FR service description specifically states that FR is not particularly suitable for voice and video transportation [45]. In order to implement compressed voice transport in a FR network and still maintain voice reconstruction with good quality, the compressed voice characteristics need to be investigated.

2.2.1 Delay sensitivity

Satisfactory voice conversation imposes strict requirements on end-to-end delay. Voice packets must be received within the maximum allowable network delay if they are to be useful. The exact tolerable delay is subject to debate, but is generally accepted to be in the approximate range of 100-600 ms. The public telephone network, for example, has a maximum specification of 600 ms [30], but normally 200 ms is desirable to avoid affecting speech activity [22] [25].

The total delay d of a voice packet sent through a FR network is composed of the following portions:

$$d = c + t + p + s + q$$

where

c = *codec* delay: the total delay caused by speech buffering and the processing delay of the coder and decoder. It varies according to the voice coding technique used and the length of each voice packet. However, it is a fixed delay which is the same for each packet in the call using a certain voice coding method.

t = transmission time: the total amount of time required for the transmitters to send out a frame into the links. The transmission time in each node is equal to the length of the frame divided by the speed of the output link.

p = propagation delay: the total of the delays taken for a frame to propagate through all the links in the connection. It is equal to the distance between end users divided by the velocity of propagation, which is approximately 2×10^8 m/s [31].

s = switching delay: the total delay for all switches in the connection. The delay in each switch is the amount of time used to move a packet to an output queue, including header processing and error checking. The existing hardware switching architectures are able to switch packets in 1 ms or less, which is small enough to be considered as a constant [32]. Its total value will then be decided by the number of hops in the connection.

q = queueing delay: the amount of time the packet spends in all queues along its transmission path awaiting processing or transmission. q is a function of buffer occupancy and processing and transmission speeds.

Four of the five components of d , namely, codec delay, transmission time, propagation delay, and switching time are fixed in a certain connection in virtual circuit transmission. Their values can be calculated when the connection is established. The queueing delay component is the only significant source of variable delay in the total delay of voice packets.

2.2.2 Throughput Requirement

The current bandwidth for telephone conversations is 300 to 3000 Hz [50] and can be

coded using various types of algorithms. By using PCM, the voice call can be coded into 64 kbits/s stream [50]. If ADPCM is used, the rate of 40kbits/s is required. As compressed coding algorithms are developed, a voice call can be coded into a very low speed bit stream, as will be explained in Section 2.2.3. Not only coded voice bit streams are transmitted in the network, but also some overheads of the voice frames. For EHCV2-16K coding algorithm and 27 octet voice frames, the maximum throughput of one PVC is 20.144 kbits/s (16 kbits/s voice and 25.9% overhead), or 93.26 frames/s. The minimum throughput of one EHVC2-16K [39] voice PVC is 10.07 kbits/s (8 kbits/s voice and 25.9% overhead), or 46.6 frames/s.(refer to Section 2.2.5)

2.2.3 Loss and Loss Pattern

Compressed voice traffic is delay sensitive and tolerant to a limited number of lost voice frames. The voice frame loss rate is defined as the ratio of lost voice frames in a PVC to voice frames sent by the source FAD to the voice PVC. Generally, if the loss rate is 1%, intelligibility is quite acceptable, although degradation may be noticeable.

The loss rate of voice frames alone, however, does not provide complete information on voice frame loss performance. For voice reconstruction at the destination FAD, it does not necessarily mean that the lower the voice frame loss rate, the better the quality of the reconstructed voice. Higher quality may be better achieved under the circumstances of relatively high loss rate (5~10% loss rate) but with a very random loss pattern, than those of a low loss rate with a very bursty loss pattern [51] [52] [53]. To further describe frame loss behavior, the following two criteria need to be defined:

- **Loss burst:** the time interval in which voice codec-PDUs in one voice PVC are dropped consecutively.
- **Period of no voice loss:** the time interval in which there is no lost voice frames in one voice PVC.

In order to obtain a high quality of the reconstructed voice, it is essential to achieve a low loss rate, a short length of loss burst and a long length of the period of no loss. However, these performance criteria sometimes conflict with each other. For example, a better quality of the reconstructed voice can be more readily obtained under the condition of a loss rate of 2% with a

mean loss burst length of 1 frame, rather than with a loss rate of 0.5% with a mean loss burst length of 4 frames.

2.2.4 Silence Period and Talk Spurt

During the period of a voice call, the voice is divided into silence period and talk spurt. Nevertheless, silence periods contain background noise essential to good conversation quality. A large amount of previous research discovered that voice can be described as the following random process: [34] [35] [36]

Let β_s be the mean length of silence period, α_t be the mean length of talkspurt. Usually the talkspurt is slightly shorter than the silence. Assume that

$$\frac{\alpha_t}{\alpha_t + \beta_s} = 44\%$$

That is, 44% of voice traffic is talk spurt and the remainder is silence. Both the length of the silence period and the length of the talk spurt are exponentially distributed. The silence period and the talkspurt occur alternately. The distribution of the silence length is

$$P_s(T \leq t) = 1 - \exp\left(-\frac{1}{\beta_s} t\right) \quad (t > 0)$$

and the distribution of the talkspurt length is

$$P_t(T \leq t) = 1 - \exp\left(-\frac{1}{\alpha_t} t\right) \quad (t > 0)$$

In this research, the basic assumption is that the mean length of the silence is 5 seconds and the mean length of the talk spurt is 4 seconds [34] [35] [36].

The talk spurt contains more information than the silence period. Therefore, the talk spurt is more sensitive to loss. If the speech activity detector is used and voice frames are transmitted

only during the talk spurt, the voice traffic will be bursty. In this research, both talkspurt and silence are coded into the voice frames and transmitted in the same way over the FR networks. Voice transmission without silence requires further studies.

2.2.5 Compressed Voice

In this research, several speech coding algorithms were considered and evaluated in terms of the speech transmission rate, complexity /cost of development, performance based on the Mean Opinion Score(MOS) and expected performance when speech frames are lost due to network congestion. Based on the evaluation results from previous research [37] [38] [39], EHCH2_16K, an embedded (variable rate) coder is recommended as the basis of the work of this thesis. At fixed time intervals, the voice encoder generates a fixed length block, called a codec-PDU. Voice traffic generated by EHCH2_16K is called compressed voice traffic because EHCH2_16K is a low speed coding algorithm which requires a small amount of bandwidth and the voice codec-PDU generated by EHCV2_16K contains little redundancy. Because EHCV2_16K is based on HCV2_8K, HCV2_8K is firstly introduced as follows:

HCV2_8K

HCV2_8K [39] is an 8kb/s algorithm featuring built-in echo cancellation and (inband) transitional signalling. Voice information is collected and a voice codec-PDU is generated at an interval of 20 ms at the speech coder. The length of the voice codec-PDU is 160 bits, the sampling rate is 8 kHz, and the codec delay of a voice codec-PDU is 60 ms. HCV2-8K is a low speed codec algorithm at 8Kb/s, which requires much less bandwidth than the widely used ADPCM algorithm at 32Kb/s. Although the bit rate of HCV2-8K is much lower than that of ADPCM, the quality of reconstructed voice using HCV2-8K (MOS=4.0) is almost the same, since HCV2-8K voice frames contain less redundancy and have almost the same amount of useful information.

EHCV2_16K

EHCV2-16K [39] is an embedded algorithm based on HCV2-8K. Essentially, 8 kb/s of additional excitation information is sent and may be dropped at the source FAD or in the network with a degradation of quality. If the additional 8 kb/s are always dropped, the reconstructed speech is exactly the same as that produced by the HCV2_8K algorithm. Without the frames

dropped, EHCV2-16K has extremely good performance: MOS=4.2. In EHCV2_16K, the high priority 20 bytes are grouped into a codec-PDU and the low priority 20 bytes are grouped into a separate codec-PDU. The code delay is 60 ms.

At the source end equipment, a frame overhead and one or several codec-PDU are grouped into one FR frame. The voice frame structure is shown in Figure 2.3. The first three and

Figure 2.3 Voice frame structure

the last three octets are standard overhead of FR. The information field comprises one octet which is used for the end to end processing such as recognitions of frame priority, and one or several codec-PDU. The sequence number is six bits long. The priority bit (PR) will be set to one if the information field contains the high priority codec-PDU; otherwise, it is set to zero. The silence indicator bit (SI) will be set to one if the codec-PUD in the information field is silence; otherwise, it is set to zero. In this thesis, only the PR bit is used for the purpose of congestion control. Frame

priority recognition is the function of layers above the frame sub-layer at the source end.

2.3 Congestion Control

Congestion is always a possibility in any data communications network. There are concerns about how an FR network should react to overload conditions. The challenge of congestion control is particularly acute for an FR network because of the limited tools available to the frame handlers. As there are no control frames presented within the network for traffic flow management, a frame handler cannot control the flow of frames coming from a subscriber or an adjacent node by using some typical flow control protocols in X.25. Congestion may arise during periods of peak offered loads. Efficient congestion control mechanisms are needed to make the frame relay effective.

Congestion control methods usually include two types of procedures: congestion avoidance procedure, and congestion recovery procedure. In FR networks, congestion avoidance is preferred because the data traffic has a bursty nature, the voice traffic has a real time nature, and flow control and error control are only performed end-to-end. Therefore, this research concentrates on congestion avoidance procedures.

2.3.1 Implicit Signalling

Implicit signaling is used to report dropped frames and is recognized by higher layers. The congestion recovery procedures initiated by this implicit notification operate under the conditions of severe congestion. This is beyond the scope of this research.

2.3.2 Explicit Signalling

Explicit signaling is used to initiate congestion avoidance procedures, of which some are performed at higher layers and some at the frame sub-layer. In the standard of the FR core functions, three types of explicit congestion notification are defined. They are:

- a). Forward explicit congestion notification (FECN).
- b). Backward explicit congestion notification (BECN).
- c). Consolidated link layer management message (CLLM)

b) and c) are the feedback mechanisms used in this research. Congestion actions initiated by a) are only able to be done at a higher layer than the frame sub-layer.

The domain of the explicit congestion notification mechanisms is shown in Figure 2.4. :

Figure 2.4 The Domain of the explicit Congestion Notification Mechanisms

2.3.2.1. FECN

The processor in the node monitors the occupancy of the buffers. When it exceeds the specified threshold, the FECN bit is set in the frames moving toward the destination. The response to an FECN is complex, since it requires the user to notify its peer user to restrict the flow of frames. The core functions in the FR protocol do not support this notification. Therefore, it must be done at a higher layer, such as some other link control protocol above the FR sub-layer or transport layer. This research will concentrate on the other two types of explicit congestion notifications.

2.3.2.2. BECN

The BECN bit in the address field may be set by a congested network to notify the user that congestion avoidance procedures should be initiated when applicable for traffic in the opposite direction of the received frame. It indicates that the frames the source transmits on this logical connection may encounter congested resources. This feedback mechanism is used for the source controlled congestion control mechanism.

2.3.2.3. CLLM.

BECN depends upon two-way traffic in one PVC. But there may be no backward traffic in a PVC or the backward traffic in this PVC may also be congested. Under these circumstances, the increased delay of the congestion notification signal may worsen the congestion status in the network. The FR standard ANSI T1.618-1991 [3] provides another explicit backward congestion notification mechanism, the Consolidated Link Layer Management (CLLM) message. This solution uses a reserved (management) address for all of these messages: "all source quench" messages are consolidated into one, and shuffled around the network. CLLM messages are explicit backward congestion notification messages sent by the congested switch to all transmitting PVCs experiencing congestion on that switch.

The use of the BECN bit, the FECN bit and the CLLM message is in a specific direction. This is shown in Figure 2.4.

Although CLLM message has been defined by CCITT and ANSI, there is no report about their actual implementation in any data networks. One of the most important contributions of this research is the proposal of using CLLM messages for congestion notification, especially for the voice congestion notification.

2.4 Modelling of the Voice Frame Relay Network

This research is based on the system model described in this section. The voice FR network is composed of various network elements.

2.4.1 Network Element

2.4.1.1. FAD

The FAD is a Frame Assembler Disassembler for terminals. In this research it is considered as end user equipment, as shown in Figure 2.5. Through a 64kb/s network access link, the FAD is connected to an FR switch. The functions it performs include:

- Call set-up procedures;

Figure 2.5 Frame Assembler--Disassembler

- Frame size control;
- Multiplexing voice and data frames on an outgoing frame stream;
- Accumulating and framing voice coder output;
- Voice talkspurt synchronization for voice decoder input by setting up and maintaining an Adaptive Buildout Delay during call setup and data-transfer phase as required [41];
- Implementing voice priority over data;
- Inspecting the status of the interface to manage the traffic accessing the FR network.
- Reacting to congestion notification (That is, dropping the low priority voice frames upon receiving congestion notification).

2.4.1.2. FR Switch

The queueing model of the FR switch is shown in Figure 2.6.

Incoming frames are processed in FIFO order by the processor. This involves reading the DLCI, table look-up for the outgoing DLCI and all other FR core functions. The processing is completed when the frame is marked as being on a particular output queue. In each switch there is a pool of memory for all buffering activities. All incoming and outgoing frames share common memory. Upon completion of processing, the frame is not copied in memory but remains in its original location until it has been transmitted. Because the T1 links connecting FR switches in the

Figure 2.6 The queuing model of the FR switch

FR network take much less time to transmit a frame than the low speed access/exit links, only FR switches of high speed can enable users to utilize the large transmission bandwidth of T1 links. In this research, it is assumed that the processing speed is 2150 frames/s independent of frame size. This is equivalent to 465 μ s/frame.

L. Dong [41] has shown that performance for compressed voice traffic will not be satisfied without implementation of the voice priority over data in the FR switch output queues[39]. In the present research, non-preemptive priority service is arranged. An FR switch has two queues per output line instead of one: a priority queue for voice and a non-priority queue for data. The priority queue would be served until empty. Priority service would start after the arrival of a frame at an empty priority queue and after completion of any frame in transmission at the arrival time. Priority service is not a feature of current frame-relay networks and its implementation will be explained later.

2.4.1.3. T1 link and 64kb/s link

FADs are connected to FR switches by 64 kb/s links. FR switches are interconnected by 1.544Mbits/s T1 links. During the 1980's, T1-based private networks became the workhorses of

the corporate backbone. They act primarily as bandwidth managers, providing the functional equivalent of multiple, relatively low-speed private leased transmission lines by using high-speed (e.g. 1.544Mbps/s) dedicated facilities.

2.4.2 Network Configuration

Figure 2.7 Networks Configuration

The network configuration is shown in Figure 2.7.

Private Branch Exchange (PBX) is a call control device which establishes and terminates the PVCs from one geographic location to another.

2.4.3 Reference Connection

To investigate the feasibility of voice transmission in an FR network, a Reference Connection is established to evaluate network performance, as shown in Figure 2.8.

In this model, a voice or data connection is established between the two end users via an FR network. The two FAD are connected via several FR switches and transmission links. Besides the voice and data traffic sent from the source FAD to the destination FAD, there is also voice and

Figure 2.8 Reference Connection

data traffic from other sources at each node. The distance of the first hop is 1000 km, and the second one is 600 km.

2.4.4 Traffic Model

The traffic transmitted from the source FAD to the destination FAD over the reference connection is defined as *reference traffic*, and traffic transmitted over a FR switch, exclusive of reference traffic, is defined as *background traffic* over the FR switch.

It is assumed that the total amount of background traffic is 500 Kbits/s, in which data accounts for 75% of bandwidth and voice accounts for 25%. The maximum volume of reference traffic is 64 Kbits/s. The total maximum volume of traffic on an FR switch is 564 Kbits/s. For reference traffic, the total amount of voice traffic varies as the number of PVCs changes. In a voice PVC, voice frames form a voice frame stream, in which every high priority voice frame is preceded and followed by a low priority frame. The minimum number of reference voice PVCs is one, and the maximum is six. For one reference voice PVC, the minimum bandwidth is 8 Kbits/s and the maximum is 16 Kbits/s. The reference data PVC has a variable bandwidth. As reference voice offered load at the source FAD increases, the sending rate of data traffic at the source FAD

decreases. The maximum sending rate of data at the source FAD is 64Kbits/s (when the amount of reference voice traffic is zero) or 62.5 frames/s. As the amount of reference voice traffic increases, the reference data rate is reduced by the proposed congestion control scheme described in Section 4.2. The total reference traffic amount of voice and data can not exceed the speed of the access transmission links; for example, 64 Kbits/s in the reference connection in this research.

The other traffic parameters in the reference traffic and background traffic model are listed below:

- length of voice frames: 27 octet (1 EHCv2-16K voice codec-PDU which has 20 octets and 7 octets of overhead)
- length of data frames: truncated geometric distribution with mean 128 octets, maximum 268 octets and minimum 7 octets.
- arrival of reference voice traffic at the source FAD is deterministic with mean 46.3 frames/s for each reference voice PVC.
- arrival of reference data traffic at the source FAD has an exponentially distributed interarrival time. The mean arrival rate is dynamically adjusted by the end user on the basis of the network status (see Chapter 4).
- The combined voice and data background traffic entering the FR switch is assumed to have Poisson arrivals because it comes from many other FR switches and FADs, and contains large numbers of voice and data PVCs which have independent arrival distributions.
- The propagation speed is 2×10^8 m/s.
- The processor speed of the FR switch is 2150 frames per second.

Chapter 3. QOS Requirement, Congestion Detection and Notification

This chapter aims to specify the quality of service (QOS) requirement for voice traffic and data traffic, and to provide the congestion control systems with an effective feedback mechanism, including a fast feedback loop and an effective congestion detection method.

3.1 QOS Requirement

In Integrated Services Frame Relay Networks, both packetized voice and data traffic are transmitted over the same physical channels. Packetized voice and data have very different QOS requirements and therefore, require different congestion control strategies.

Voice traffic is delay sensitive, which means that it requires a much lower level of congestion than data traffic. When the network load is heavy enough to make voice delay increase to an alarm threshold, it may still be unnecessary for data traffic to be notified of congestion. In order to satisfy QOS for end users, end to end voice delay over the reference connection is limited within 200 ms. The fixed delays for the reference connection are:

1. Codec delay c : for EHCv2-16K coding algorithm, $c=40+20N_f$ ms, where N_f is the number of EHCv2_16K codec blocks contained in one voice frame. For 27 octet voice frames, $c=60$ ms, and for 47 octet voice frames $c=80$ ms.
2. Transmission delay t : For the given reference connection, the total transmission time includes the time at the access link (64 kbits/s), the time at the two T1 links (1.544 Mbits/s), and the time at the exit link (64 kbits/s). For 27 octet voice frames, $t=7$ ms, and for 47 octet voice frames, $t=12.2$ ms
3. Switching delay s : There are three FR switches in the given reference connection. The processor speed is 2150 frames/s, i.e. the switching delay in each FR switch is 465 μ s. The total switching delay in the reference connection is $s=1.395$ ms.
4. propagation delay p : Since the length of the reference connection is 1600 km and the velocity of propagation is approximately 2×10^8 m/s, the total propagation delay is $p=8$

ms. For each additional hop of 1000 km, the additional propagation delay is 5 ms.

The fixed delay in the reference connection is 76.395 ms. If 10 ms is reserved to achieve a conservative design, the maximum allowable queueing delay will be

$$B_T = 200 - 76 - 10 = 114 \text{ ms}$$

One of the objectives of the congestion control scheme is to limit queueing delay to 114 ms. Any frame experiencing a queueing delay greater than 114 ms will be discarded at the receiver since it will arrive too late to be played out in sequence.

Another objective of the congestion control scheme is to guarantee the minimum required bandwidth for each voice PVC which is multiplexed over the reference connection. For EHC2-16K coding algorithm and 27 octet voice frames, the maximum throughput of one PVC is 20.144 kbits/s (16 kbits/s voice and 25.9% overhead), or 93.26 frames/s. The minimum throughput of one EHVC2-16K voice PVC is 10.07 kbits/s (8 kbits/s voice and 25.9% overhead), or 46.6 frames/s.

The final objective of the congestion control scheme is to allow the network to have a good voice loss behavior. The fact that voice traffic is loss tolerant does not mean that any voice loss rate or loss behavior is acceptable. For EHC2-16K coding algorithm, almost 50% of frames (low priority frames) can be discarded at the source encoder when congestion is detected. In the network, however, a frame loss rate of 5% will degrade the quality of reconstructed voice substantially. On the other hand, average voice frame loss rate alone cannot provide sufficient information for the quality of voice transport. In order to obtain reconstructed voice with good quality at the destination end point, random voice loss pattern is even more important than the low voice loss rate, because the bursty loss pattern in the network will more negatively affect the voice quality. The proposed control scheme guarantees a mean voice frame loss rate of less than 5% and a length of loss burst less than two EHVC2-16K voice codec PDUs.

Data traffic is delay tolerant but loss sensitive. In order to provide high quality service for end users, data frame loss rate should be much lower than voice frame loss rate. Under this condition, only very low data frame loss rate can provide high layers with good quality service. In this research, an objective of the control scheme to obtain a mean data frame loss rate of less than

10^{-6} is assumed.

Because the QOS requirements for voice and for data are different from one another, different congestion detection and notification methods should be used. The existence of separate queues for voice in the FR switches implies that it would be possible and highly desirable to establish separate congestion detection and notification for voice and data traffic.

3.2 Buffer Management for Congestion Detection

No matter what congestion notification mechanism is used, the congestion level has to be measured. Since voice traffic is delay sensitive, the end to end delay should ideally be taken as the congestion level measurement. However, in the FR network, the difficulty of using end to end delay as the feedback information is that there is no way to measure it. Unlike PVP and PCME networks, there is no time stamp in the FR frame structure. In this case, the switch is unable to know the delay that the frame has already experienced. However, the delay in a single switch can be controlled by limiting the queue length, and the buffer occupancies can be taken as the measurement of the degree of congestion for voice traffic.

Data traffic is delay tolerant but loss sensitive. Since losses are usually due to buffer overflow, the data queue length should be limited and buffer occupancies used as the measurement of congestion.

Although buffer occupancies are used for the measurement of congestion level for both voice and data traffic, their purposes are different. For voice traffic, the purpose is to limit the queuing delay, while for data traffic, it is to prevent data frame loss due to buffer overflow.

3.3 Congestion Status

FR networks are designed to handle bursty traffic. During these bursty periods there is a possibility that many calls will be established simultaneously and FR resources will become depleted, causing congestion.

The impact of congestion is degradation in terms of throughput, transit delay, and frame loss. Point A as shown in Figure 3.1 and Figure 3.2 is the point beyond which increasing the offered load results in unacceptable increases in transit delay. As the offered load increases,

transit delay increases nonlinearly, and throughput at transition Point B, shown in Figure 3.1 and

Figure 3.1 Classification of Congestion Levels.

Figure 3.2, will actually decrease. Severe congestion represents a potential loss in the quality of

Figure 3.2 Classification of Congestion Levels.

service to customers in the form of excessive transit delay, frame loss, and throughput decrease.

As analyzed in the previous section, buffer occupancy will be used as measurement of congestion level; the different queue lengths in a FR switch correspond to the different congestion status. As shown in Figure 3.3, there are three buffer thresholds in a queue: mild congestion

Figure 3.3 Relation between queue length and congestion status

threshold (MCT), severe congestion threshold (SCT), and absolute congestion threshold (ACT). When the queue length is shorter than the MCT, it is in the status of no congestion. When the queue length is longer than the MCT but shorter than the SCT, it is in the status of mild congestion. When the queue length is longer than SCT but shorter than ACT, it is in the status of severe congestion. When the queue length exceeds ACT, it is in the status of absolute congestion in which all arriving frames will be discarded. Figure 3.4 shows the state diagram for the transition between various levels of congestion.

3.4 Buffer Management for Input Queues.

At the input queue of an FR switch, the more quickly the processor processes the frame, the less time the frame waits in the input queue, therefore, the input queues are shorter. This means that besides the frame arrival rate, the processor speed is also a decisive factor for the length of the input queue.

BSC is the buffer space consumed per input queue or output queue at an FR switch.

Figure 3.4 Congestion Status Indicator

When a frame arrives at the switch, it is either a voice frame or data frame. It enters the same input queue, but voice frames can not be recognized before they have been processed. Therefore, voice and data frames have the same priority in the input queue. Three thresholds, MCT, SCT and ACT are set to indicate different processor congestion status in a FR switch, as shown in Figure 3.5. The value of the congestion thresholds depend on the processor speed. The faster the processor speed, the larger the values MCT, SCT and ACT can be.

3.5 Buffer Management at Output Queues.

As shown in Figure 3.6, after the processor reads the DLCIs of the frames arriving at the FR switch, it switches the voice frames to the priority output queue and the data frames to the non-priority output queue. Three thresholds, MCT, SCT, and SCT, are also set in the output queue to indicate transmission congestion. Since separate queues exist for voice and data at the output line, as shown in Figure 3.6, different thresholds are set in the priority queue and the non-priority queue to indicate separate congestion statuses for voice and data respectively. Generally, the MCT, SCT and ACT for the priority queue are much smaller the those for the non-priority queue.

Figure 3.5 Congestion Status in the Input Queue

Figure 3.6 Congestion Status in the Output Queue

3.6 The CLLM Notification Mechanism

FR core functions support backward explicit congestion notification (BECN, see Section 2.3.2), which is easy to implement. However, BECN is not always an effective feedback mechanism because when the traffic in the forward direction is congested, there may be no traffic in the opposite direction or the traffic in the opposite direction may also be congested. The result may be that BECN bits can not be fed back to the source FAD in time.

The FR standard ANSI T1.618-1991 provides another explicit backward congestion notification mechanism, known as a consolidated link layer management message (CLLM message). CLLM messages are explicit backward congestion notification messages sent by the congested switch to all transmitting PVCs experiencing congestion on that switch.

A CLLM message is generated by the network in response to congestion and is transmitted on a layer management DLCI in the U plane. The DLCI address is 1023 (two-octets). CLLM messages are only sent between adjacent nodes. When a CLLM message is received, the per-PVC congestion information is extracted, and new CLLM messages are formatted for transmission to the next hop of the affected PVCs. In this manner, the congestion information is propagated to the source end point of the network for the congested PVCs, and the end users may take appropriate control actions.

The switch processor checks whether the switch node is experiencing congestion once every x seconds. The value of x is to be decided and optimized by simulation in this thesis. When congestion is detected, the processor generates a CLLM message and sends it to the source FAD. The CLLM message generating frequency is defined as $f_M = \frac{1}{x}$.

The CLLM message generating frequency affects the network performance. If f_M is too slow, the feedback information about network congestion will not be able to reach the source FAD, so the network performance will be degraded. If f_M is too high, the feedback traffic will increase the overall network traffic resulting in performance degradation.

The necessity of using CLLM messages for notification of voice congestion and data congestion will be investigated in Chapter 5. The following methods of congestion notification are defined:

Voice BECN: If a voice PVC is congested, the BECN bit of frames in the opposite direction of this PVC is set.

Voice CLLM: If a voice PVC is congested, a CLLM message is sent to the source FAD in the opposite direction of the PVC.

Data BECN: If a data PVC is congested, the BECN bit of data frames in the opposite direction of the PVC is set.

Data CLLM: If a data PVC is congested, a CLLM message is sent to the source FAD in the opposite direction of the PVC.

3.7 Buffer Thresholds for Voice as a Function of Output Line Speed

If the maximum allowable queueing delay in the output queue is D_q and the transmission rate of output line is R_q , the queue length should not exceed $L_q = D_q \times R_q$. If the maximum queueing delay at the last low speed link (64kbits/sec.) is 125ms, then the output queue length can not exceed 1000 bytes. Therefore, we can set the voice ACT to 1000 bytes.

Similarly, for the SCT and MCT, if a queueing delay of longer than 50 ms is considered to be severe congestion, and 25 ms is mild congestion, then SCT and MCT should be 400 bytes and 200 bytes respectively.

The optimum values of MCT, SCT and ACT will be determined by simulation and discussed later.

3.8 Voice Loss Behavior as the Function of the Buffer Threshold

Voice loss behavior is one of the important factors for voice reconstruction. The length of the loss burst is defined as the number of voice frames dropped consecutively, and the period of no voice loss is defined as the time interval in which no voice frames are dropped. The voice SCT and MCT affect the voice loss behavior substantially.

Increasing the SCT will lead to a long voice loss burst and a long period of no voice loss. It is expected to achieve both a short voice loss burst and a long period of no voice loss to obtain

reconstructed voice with good quality.

Chapter 4. The Proposed Congestion Control System

Congestion Control is the joint responsibility of the network and the end users. The relevant network facilities monitor the degree of congestion both at intermediate nodes and the endpoint, while the end users limit the traffic flow entering the network. Some suggestions on the network's and end users' actions will be proposed.

4.1 Congestion Control for Voice

In this thesis, the congestion control procedures for voice traffic comprise the three elements discussed in the following subsections:

- Implementation of Voice Priority over Data.
- FAD Response to Explicit Notification.
- Random Setting of DE Bit

4.1.1 Implementation of Voice Priority over Data

In the standard FR protocol, no prioritization for any type of traffic is specified. Previous research on voice in FR networks [41] proved that implementation of voice priority over data is necessary to guarantee the voice transport over the FR network.

In order to implement different priorities for voice and data, the network should be set up with two PVCs between each pair of communicating FADs requiring integrated service, one for data frames and one for voice frames, so that a voice PVC is identified during the connection setup procedure. During the data transfer phase (in the U-plane), the switch processor reads the DLCI of incoming frames. If the DLCI belongs to the voice PVC, it will be switched to the priority output queue; otherwise, it will be switched to the non-priority output queue. In this research, non-preemptive priority service for voice is implemented; therefore, the priority queue will be served until empty. Priority service will start after the arrival of a frame to an empty priority queue and after completion of any frame in transmission at arrival time.

4.1.2 FAD Response to Explicit Notification

- Upon receiving Congestion Notification Signaling (a BECN bit or CLLM message indicating that the voice connection is congested), the source FAD will drop the low priority voice frames (high priority and low priority voice frames were defined in Section 2.2.5.)
- If the source FAD has not received congestion notification signaling for a certain period of time it stops dropping low priority frames. For example, CLLM procedure is activated once every x seconds; if the source FAD has not received a CLLM message for x seconds, it stops dropping low priority frames. The optimum value of x is investigated in Section 5.2.

4.1.3 Random Setting of DE Bit

The option of setting the DE bit in FR frames provided in the standard FR protocol can be used to improve voice loss behavior when severe congestion occurs in the network. The purpose of random setting of DE bit in this research is to reduce the length of the voice frame loss burst, by randomizing losses at the cost of increasing slightly the voice frame loss rate.

In a voice stream, if all the voice frames have the DE bit clear (DE bit=0), all arriving frames will be excessively delayed when severe congestion occurs. This will lead to a long burst of voice frame losses at the receiver. If a small proportion of voice frames with the DE bit set to one are randomly distributed in the voice stream, then the voice frame loss will also be random, although the voice loss rate may increase slightly. Voice frames with the DE bit set to one are called DE voice frames and ones with it set to zero are called non-DE voice frames. Discarding DE voice frames during congestion will decrease the possibility of buffer overflow and of excessive delay, which is the main cause of long burst of voice frame loss.

The optimal value of the percentage of DE voice frames will be investigated in Section 5.2.1.

4.2 Congestion Control for Data Traffic

Dynamic change of the data frame access rate to the FR network based on received data

congestion notification signaling is used. Assuming that there is always data waiting for transmission at the source FAD, if there is no voice to be transmitted, the data traffic will take all the bandwidth of the reference connection; i.e. 64Kbits/s or 62.5 frames/s. If the source FAD receive consecutively n_I CLLM congestion notification messages, the data frame access rate will be reduced to k_I times of the previous value. If the source FAD has not received CLLM message for y seconds, the data frame access rate will be increased by a factor of k_d times the present data frame access rate. In this research, the value of n_I , y , k_I and k_d have been set at 5, 100 ms, 0.75 and 0.125 respectively. Simulation results show that for a very broad range of the values of n_I (1-10) and y (20-200 ms), the data performance does not show much sensitivity. Therefore, $n_I=5$ and $y=100$ ms were chosen. The values of $k_I=0.75$ and $k_d=0.125$ were suggested in the appendix A of Frame Relay standard document [3], and simulation results also show that these choices give satisfactory data performance.

4.3 Congestion Control Procedure at Intermediate Nodes

The congestion control procedure at the FR switches includes:

- reaction to the congestion status at the input queue;
- reaction to the congestion status at the priority output queue; and
- reaction to the congestion status at the non-priority queue.

4.3.1 Reaction to Congestion Status at the Input Queue

Whenever a frame arrives at a switching node, the processor at the switch checks the buffer space consumed for the input queue. If the input queue is in the state of mild congestion or severe congestion the processor sets the BECN bit and FECN bit of the arriving frame to one. If the input queue enters the state of absolute congestion, the processor discards all newly arriving frames. The reaction to the congestion status at the input queue is shown in Figure 4.1.

4.3.2 Reaction to Congestion Status at the Priority Output Queue

If the priority queue length is shorter than the MCT, an arriving voice frame enters the priority queue. If the priority queue length exceeds the MCT, but does not exceed the SCT, the

Figure 4.1 Reaction to the congestion status at input queue

processor will set the BECN bit of the backward frames and the FECN bit of the forward frames. If the priority queue length exceeds the value of SCT, but does not exceed the value of ACT, the processor discards the arriving DE voice frames, sets the FECN bits of the arriving non-DE voice frames, and sets the BECN bits of the backward arriving voice frames if they are not discarded. If the priority queue length exceeds the ACT, the processor discards newly arriving frames at the priority output queue. The reaction to congestion status in the priority output queue is slightly more complex than the input queue, and is shown in Figure 4.2.

4.3.3 Reaction to Congestion Status at the Non-Priority Output Queue

If the non-priority queue has a length shorter than the MCT, the data frame enters the queue for transmission. If the non-priority queue length exceeds the MCT, but does not exceed the ACT, i.e. either mild or severe congestion occurs, the processor will set the BECN bits of the backward frames and FECN bits of forward frames. If the non-priority queue length exceeds the ACT, the processor will discard newly arriving frames. Unlike the control procedure at the priority output queue, when the non-priority queue length exceeds the SCT, but does not exceed

Figure 4.2 Reaction to the congestion status at the priority output queue

the ACT, the processor will not discard any arriving data frame, because the data traffic is loss sensitive. All the DE bits of data frames are set to zero. The reaction to the congestion status in the non-priority queue is shown in Figure 4.3.

4.4 Locating the Network Bottleneck

For the reference traffic model described, the bottleneck is the switch processor, which has a processing speed of 2150 frames/s, if the output line is a T1 link. Even if all arriving traffic is switched to the same output line, the total amount of traffic in bits/s does not exceed 564 Kbits/s which is less than the capacity of the T1 link. Some other typical circumstances are analyzed

BSC: Buffer space consumed at non-priority output queue

Figure 4.3 Reaction to the congestion status at the non-priority output queue

below:

1. If the frame arrival rate at an output line is 2150 frames/s (the switch processor switches all arriving frames to a single output link at the maximum switch speed), and all arriving frames are voice frames, the arrival rate in bits/s at the output line will be $2150 \times 27 \times 8 = 464.3$ Kbits/s, which is much slower than the T1 link speed. Therefore, the processor forms a bottleneck.

2. If the frame arrival rate at an output line is 2150 frames/s, and all arriving frames are data frames, the arrival rate in bits/s at the output line will be $2150 \times 128 \times 8 = 2.2$ Mbits/s which is higher than the T1 link speed. Therefore, the output T1 link forms a bottleneck. This is the case when the background traffic is greater than 500 Kbits/s.

3. If 25% of the total traffic in bits/s arriving at the output line is voice, and the frame arrival rate at the output line is 2150 frames/s, the number of arriving voice frames per second will be 1316, the number of arriving data frames per second will be 834, and the total traffic will be $1316 \times 27 \times 8 + 834 \times 27 \times 8 = 1.138$ Mbits/s, which is less than the T1 link capacity. The

switch processor is the bottleneck. This is another case where background traffic is greater than 500 Kbits/s.

4. If the output line is 64Kbits/s, and the arriving frames at the output queue are all voice frames, the maximum allowable arrival rate will be $64000 / (27 \times 8) = 296.3$ frames/s, which is slower than the processing speed of the switch processor. If the processor switches more than 296 frames per second to the 64 Kbits/s output line, then the output line will become a bottleneck. If some arriving frames are data frames, the maximum allowable arrival rate at the output queue will be even less than 296 frames/s.

Generally, the low speed transmission link which connects the last FR switch and the destination FAD is the bottleneck of the whole reference connection. The congestion control procedures carried out at the last FR switch and at the source FAD play a more important role than at the other switches.

4.5 Performance of the Proposed Congestion Control Scheme

In this section, the performance of the FR network with the proposed control schemes is investigated to test their effectiveness. The following essential assumptions are made:

- CLLM messages are used for both voice and data congestion notification.
- CLLM message generating time is 5 ms.
- CLLM message generating frequency is 10 times/s.
- 3.5% of voice frames have the DE bit set, and they are uniformly distributed in the voice stream.
- Over the last low speed link there is only reference voice traffic, and there is no background voice traffic.
- At the input queue of a switch, ACT=4000bytes, MCT=2000bytes.
- At the non-priority output queue, ACT=8000bytes, MCT=2000bytes.

Figure 4.4 Voice throughput vs. voice offered load

- At the priority input queue, ACT=1000bytes, SCT=400byte and MCT=100bytes.

As shown in Figure 4.4, the voice throughput increases as the offered voice load increases. When the number of the reference voice PVCs is less than three, the voice throughput increases linearly as the offered load increases since the offered voice traffic is less than the capacity of lowest speed link (4kbits/s). When the number of the reference voice PVCs is greater than four, the offered voice load ($\geq 4 \times 16 \text{ kbits/sec.} = 64 \text{ kbits/s}$) will exceed the capacity of the lowest speed link; therefore, some of the low priority voice frames will be dropped at the source FAD, and a very few frames are dropped in the network. As a result, the voice throughput increases non linearly and tends to saturate.

As shown in Figure 4.5, the mean voice queueing delay increases as the offered voice load increases. When the number of the reference voice PVCs increases from three to four, the offered voice load increases from less than the capacity of the lowest speed link to exceed it and the mean queueing delay increases here very sharply.

Figure 4.5 Voice delay vs. voice offered load

As shown in Figure 4.6, there is no voice frame loss when the number of the reference voice PVCs is less than three. When the number of the reference voice PVCs is greater than three, congestion will result. Some DE voice frames are dropped in the FR networks while some are dropped due to excessive delay. From four to five voice PVCs, the increase of mean voice frame loss rate is substantial because the congestion level increases as the offered load increases. However, from five to six voice PVCs, the mean voice frame loss rate decreases as the voice offered load increases for the following reason: as the voice traffic increases, the data traffic throughput decreases because of the voice priority over the data, and the decreased data traffic reduces the impact of the random factor on voice traffic. Voice frames are of fixed length. Their transmission times are fixed and when the number of voice PVCs is six, most frames transmitted in the reference connection are voice frames. Voice queue length and voice delay vary little. Because the control procedure at the source FAD limits the number of low priority frames entering the network, the voice buffer occupancies do not often exceed the MCT, and voice delay does not often exceed the maximum build out delay. Therefore, the mean voice frame loss rate is lower when the number of voice PVCs is six rather than five.

As shown in Figure 4.7, the mean length of the voice loss burst is zero when the number of

Figure 4.6 Voice loss rate vs. voice offered load

voice PVCs is less than three, and is 1 when the number of voice PVCs is greater than three. Short voice loss burst is preferable for voice reconstruction at the destinations.

In order to investigate the necessity of the proposed congestion control scheme, network performance with and without it are compared, as shown in Figure 4.8 to Figure 4.10. In those Figures, Curve 1 shows the performance of a network in which the proposed control scheme is implemented, while Curve 2 shows the performance of a network in which neither buffer threshold is set for voice congestion detection, and there are no CLLM messages for voice congestion notification. If the voice traffic load does not exceed the bandwidth of the capacity of the lowest speed link, that is, 64 Kbits/s, the network can work well without a buffer threshold set and CLLM congestion notification for voice traffic: the voice throughput increases as the voice traffic load increases, the mean voice queueing delay is very low, and the voice frame loss rate is zero. However, if the voice traffic load exceeds the capacity of the lowest speed link (the number of voice PVCs over the lowest speed link is greater than 3), the voice throughput decreases to zero, the voice queueing delay is unacceptably large (greater than 135 ms), and the voice frame

Figure 4.7 Voice loss burst length vs. voice offered load

loss rate is one. Therefore, the congestion control scheme proposed in this research is **necessary and effective**.

4.6 Conclusion

Under the assumed conditions, the proposed congestion control scheme is **simple and effective**:

- Demanded voice throughput can be provided when there is no congestion.
- During periods of congestion, low priority voice frames are discarded at the source FAD to guarantee the minimum required bandwidth for each voice PVC.
- Average voice delay is limited to a very low level both under the condition of no congestion and under the condition of congestion.

Figure 4.8 Voice throughput comparison between with voice CLLM and without voice CLLM

Figure 4.9 Voice queuing delay comparison between with voice CLLM and without voice CLLM

Figure 4.10 Voice loss rate comparison between with voice CLLM and without voice CLLM

- Voice frame loss rate is low, and the voice loss behavior is good enough for voice reconstruction at the destination end (mean length of voice loss burst is approximately 1).

The proposed control schemes should not only be effective under the above conditions, but also robust to a range of conditions. This problem will be addressed in the following chapter. Effects on data transmission quality will also be reported.

Chapter 5. Optimal Parameter Setting and Sensitivity Test

This Chapter investigates the problem of comparison between voice BECN and voice CLLM, data BECN and data CLLM optimal parameter setting, structure sensitivity and parameter sensitivity. The optimal parameter setting includes the percentage of DE voice frames in the voice stream, the voice frame length, data buffer size, the MCT, the SCT, and the CLLM frequency. The sensitivity tests include access and output line speed, CLLM processing time, background traffic levels, length of silence and talkspurt periods, and the number of hops.

As analyzed in the previous chapter, the reference connection will be in the worst congestion situation for voice traffic when there are 6 voice PVCs multiplexed over the lowest speed link. Therefore, unless stated, otherwise it is assumed in this chapter that six voice PVCs are multiplexed over the last low speed link to investigate the parameter sensitivity under the worst congestion situation for voice.

5.1 Comparison between BECN and CLLM

5.1.1 Comparison between Voice BECN and Voice CLLM

In Chapter 3, two options were discussed for voice congestion notification: voice BECN and voice CLLM. These two options are compared in this section.

5.1.1.1. Effect on the performance of voice traffic

When voice BECN rather than voice CLLM is used as the voice congestion notification method, the worst case is that there is no backward traffic, or the backward traffic in the reference connection is severely or absolutely congested while the forward traffic is congested. In the case of Figure 5.1 no backward traffic, the voice BECN and the voice CLLM are compared in Figures 5.1 to 5.3. If the voice traffic load is less than the capacity of the lowest speed link (the number of voice PVCs ≤ 3), the voice queue length will be very short due to the voice priority over the data. Therefore, the voice queue length at FR switches would rarely exceed the MCT and never the SCT. Even if the source does not drop any low priority frames, the voice traffic will not be severely congested. When the voice CLLM is used, the source FAD receives the CLLM messages occasionally even if the number of voice PVCs is less than three because the voice queue length

in FR switches occasionally exceeds the MCT. Therefore, the voice throughput is higher if voice BECN is used rather than voice CLLM as shown in Figure 5.1, because the source FAD drops low priority frames upon receiving a CLLM message.

As shown in Figure 5.3, the voice frame loss rate is zero, when the voice traffic load is less than the capacity of the lowest speed link.

Figures 5.1 through 5.3 show that when the voice traffic load exceeds the capacity of the lowest speed link (the number of voice PVCs is greater than 3), the network performance of voice CLLM is superior to that of voice BECN, because when there is no backward traffic, CLLM messages can notify the source FAD of congestion in the FR network. Hence, the source FAD prevents excess traffic accessing the FR network in order to avoid the degradation of the network performance.

Figure 5.1 Difference of voice throughput between voice BECN and voice CLLM

5.1.1.2. Effect on the performance of data traffic

The source FAD is always ready to send data frames, and the mechanism of congestion

Figure 5.2 Difference of voice delay between voice BECN and voice CLLM

Figure 5.3 Difference of voice loss rate between voice BECN and voice CLLM

notification for data is data CLLM instead of data BECN. If there is no voice traffic, the data traffic will be 64 kbits/s, i.e. 62.5 frames/s. As voice traffic increases, data traffic decreases. Figure 5.4 shows the relationship between data throughput and offered voice traffic. As the voice throughput increases, the data throughput decreases much more rapidly if voice BECN is rather than voice CLLM is used.

Figure 5.4 Difference of data throughput between voice BECN and voice CLLM

Based on the above analysis, the conclusion is that using voice CLLM is superior to using voice BECN as congestion notification for voice traffic.

5.1.2 Comparison between Data BECN and Data CLLM

As for voice congestion notification, there are two options for data congestion notification: data BECN and data CLLM, as described in Section 3.6.

Figure 5.5 shows that the data throughput in the reference connection is higher if data BECN is used rather than data CLLM. However, the loss rate of data frames is impractically high, between 20% and 90%, if data BECN is used, while the loss rate of data frames is almost zero

Figure 5.5 Difference of data throughput between data BECN and data CLLM

with data CLLM. In the FR network, BECN depends on backward traffic, which includes both voice and data frames. When the forward data traffic is congested, the BECN bits of backward data frames, instead of backward voice frames, are set to one. Because the backward voice frames have priority over the backward data frames, the backward data frames have to wait in the buffer until the backward voice frames are transmitted. Therefore, the BECN bits of the backward data frames can not be fed back quickly to the source FAD. In this way, control action can not be taken in time to prevent further congestion, leading the data buffer to overflow which resulting a very high loss rate. Consequently, CLLM, rather than BECN, must be implemented as the data congestion notification mechanism.

5.2 Parameter Optimizing

5.2.1 Random Setting of the DE Bit

As proposed in Chapter 4, randomly setting the DE bits for voice frames can change the voice loss behavior. In the standard FR protocol, frames transmitted above the negotiated value of throughput will have their DE bits set equal to one, and DE frames will be discarded during

congestion. However, if all the voice frames above the negotiated value (voice frames above 8 Kbits/s) are discarded in the network, the quality of the reconstructed voice will degrade substantially, and will even be unintelligible. On the other hand, dropping some DE frames reduces the congestion level. Because voice traffic prefers random voice loss, the source FAD randomly sets the DE bits in the voice stream. The simulation results are shown from Figures 5.6

Figure 5.6 The length of voice loss burst vs. the percentage of DE frames

to 5.7. Under the assumed conditions, a low percentage of DE frames does not necessarily imply either low voice frame loss rate or good quality of the reconstructed voice. As the percentage of DE frames increases from 0 to 4%, the mean voice loss burst length decreases from 3.4 to 1.2, and the voice frame loss rate increases from 1.25% to 1.96%. Therefore, if the other conditions are the same, it is preferable to set 4% of the DE frames to avoid a long burst of losses. The mean voice queueing delay also decreases by about 10%, from 52 ms to 48 ms, as the percentage of DE frames increases, because the congestion level is reduced. Throughput of voice frames shows no significant sensitivity to the proportion of DE frames.

5.2.2 Voice Frame Length

The previous results were based on a 27 octet voice frame. Each voice frame contains a 7

Figure 5.7 The voice loss rate vs. the percentage of DE frames

octet overhead and one EHCv2_16K codec-PDU of 20 octets giving a 25.92% overhead on voice bandwidth. To improve the bandwidth efficiency, it is possible to combine more than one codec-PDU into one voice frame of 47 octets (40 octets of voice codec-PDU and 7 octets of overhead) giving a percentage overhead of 14.89%. The codec delay $\epsilon = 40 + 20N_f$, where N_f is the number of EHCv2_16K codec-PDUs contained in one voice frame. The codec delay increases from 60 ms to 80 ms and the transmission time per frame increases approximately 6 ms. Therefore, if the end to end delay is still 200 ms as for the 27 octet frame, the maximum allowable queueing delay will decrease by 22 ms. Hence, the maximum queueing delay is assumed to be 88 ms and all voice frames experiencing a queueing delay greater than 88 ms will be dropped at the destination.

5.2.2.1. Throughput

Figure 5.8 shows that if the other conditions are the same, the longer voice frame leads to higher voice throughput, which is good for the quality of the voice reconstruction.

5.2.2.2. Voice Queueing Delays and Processor Load

Figure 5.8 Effect of voice frame length on voice throughput

Figure 5.9 Effect of variation of voice frame size on Mean queuing delay

Figure 5.9 shows that the total mean voice queueing delay is marginally lower for the longer frame size when the offered load is light and for the shorter frame size when the offered load is heavy. The difference is small enough to be of no significance, however. The differences are explained in the next two paragraphs.

Figure 5.9 shows the difference of queueing delay between the 47 octet voice frames and the 27 octet voice frames. If the number of voice PVCs is less than three, the amount of voice traffic will not exceed the capacity of the lowest speed link (64kb/s, network accessing link and network out going link) and all voice frames arriving at the source FAD will enter the FR network. Under this condition, 47 octet frames results in half the number of voice frames entering the network. Therefore, the load intensity of switch processors is also greatly reduced and the voice queueing delay in the switch input queues is decreased. The increased frame size also slightly decreases the load of the transmission line, because with the same amount of voice traffic in bits/s longer voice frames require less overhead.

When the number of voice PVCs is greater than three, the amount of voice traffic in bits/s exceeds the capacity of the slowest link. Some low priority frames are dropped at the source FAD, and some are lost in the network due to buffer overflow or due to delay. In this case, 47 octet voice frames impose a heavier load on the source FAD than 27 octet voice frames, because the frame arrival rate of the 27 octet frames is higher than that of the 47 octet frames, and the 27 octet frames are dropped more often than the 47 octet. Therefore, the total voice queueing delay of the 47 octet voice frames is higher than that of the 27 octet voice frames.

The detailed effects of frame size on the processor load, on the source FAD load and on the individual mean queueing delays in the various queues of the reference connection are shown in Tables 1 and 22. Table 5.2 shows the relationship between the offered traffic (the number of reference PVCs), processor load and input queueing delay. Table 5.3 shows the relationship between the offered traffic (number of reference PVCs), the source FAD load and the delay in the source FAD load.

Table 1 Relationship between the offered traffic (number of reference PVCs), the switch processor load and the delay in the switch processor

No. of reference voice PVCs	processor load (frame Size=47octets)	processor load (frame Size=27octets)	processor delay(frame Size=47octets) (ms)	processor delay(frame Size=27octets) (ms)
1	.5715	.7665	0.271	0.766
2	.5839	.7902	0.270	0.816
3	.5952	.8105	0.272	0.883
4	.6045	.8272	0.271	0.918
5	.6070	.8339	0.271	0.943
6	.6104	.8431	0.272	0.969

Table 2 Relationship between the offered traffic (number of reference PVCs), the source FAD load and the delay in the source FAD

No. of reference voice PVC	source FAD load (frame Size=47octets)	source FAD load (frame Size=27octets)	delay in source FAD(frame Size=47octets) (ms)	delay in source FAD(frame Size=27octets) (ms)
1	.9467	.9477	9.989	9.928
2	.9460	.9477	12.442	11.100
3	.9599	.9613	15.234	12.364
4	.9837	.9737	29.755	22.279
5	.9873	.9834	33.102	23.853
6	.9924	.9932	33.794	22.649

5.2.2.3. End to End Delay

Considering the end to end delay, excluding build out delay, voice traffic using 47 octet frames has inferior delay performance because the coding delay increases by 20 ms and the transmission delay by 5.2 ms. Therefore, by using 47 octet frames, the end to end delay increases

25.2 ms, although the queueing delay may decrease slightly under the condition of light traffic. Figure 5.10 shows the effect of Variation of the voice frame length on the voice end to end delay excluding build out delay.

Figure 5.10 Effect of variation of voice frame size on the voice end to end delay

5.2.2.4. Voice Frame Loss Rate

The voice frame loss rate of 47 octet voice frames is higher than that of 27 octet voice frames, as shown in Figure 5.11. Connections using 47 octet voice frames require a shorter voice queueing delay than those using 27 octet voice frames, and when the number of voice PVCs is greater than three, 47 octet voice frames experience a longer queueing delay than 27 octet voice frames. The increased voice frame losses are due to discards as a result of excessive delay rather than to buffer overflow. Even if the voice frame loss rates are the same, the longer voice frame option will have a longer mean voice loss burst because longer frames contain two voice codec-PDUs each.

5.2.2.5. Loss Burst Length

Figure 5.12 shows that if the number of reference PVCs is greater than three, the voice

Figure 5.11 Effect of variation of voice frame size on voice loss rate

loss burst length in a voice PVC using 47 octet frames is two codec-PDUs, while the voice loss burst length in a voice PVC using 27 octet is one codec-PDU. Therefore, shorter voice frames have better loss behavior than longer ones.

5.2.2.6. No Loss Period Length

Both 47 octet frames voice connection and 27 octet frames voice connection have a long mean length of no loss periods, as shown in Figure 5.13, which are good for the voice reconstruction at destination FAD.

5.2.3 Data Buffer Size

As voice traffic increases, the delay of data traffic will increase because voice traffic has priority over data traffic. This is especially obvious at the switch connected to the last low speed link. The relationship between the mean data frame loss rate and the buffer size for data traffic is obtained by the simulation as shown in Figure 5.14. In these simulations, all the parameters are the same as those described in the traffic model in Chapter 2, except the buffer size for data at the

Figure 5.12 Effect of variation of voice frame size on loss burst length

Figure 5.13 Effect of variation of voice frame size on no loss period length

Figure 5.14 Data Loss Rate as the Function of the Buffer Size

output queue which varies from 2500 bytes to 4000 bytes. The voice traffic also affects the loss rate of data frames. The heavier the voice traffic, the higher the data frame loss rate will be, because the voice traffic has priority to use the limited bandwidth and data queue longer. From Figure 5.14, the data frame loss rate is much higher when the number of the voice PVCs is six than when the number of the voice PVCs is one. When the number of the voice PVCs is six, if the buffer size of the data traffic is 2500 bytes, the loss rate approaches 10^{-1} ; when the buffer size is 4000, the loss rate is approximately 10^{-4} . According to the tendency shown in Curve 2 of Figure 5.14, a buffer size of 5000 bytes should be chosen for a loss rate less than 10^{-6} or 8000 bytes for a loss rate less than 10^{-9} .

5.2.4 MCT

If all six voice PVCs on the last low speed link are reference PVCs, the last link will have the same traffic load as the access link. The congestion level at the last link will not be worse than at the access link and congestion information at the access link will be crucial. The control procedure at the source FAD will take control action based on the congestion information from

the access link. In this case, even without feeding back the congestion information from the last link to the source FAD, the congestion state of the reference connection can be controlled very well, because the congestion information from the access link is sufficient to reflect the entire congestion state over the reference connection. The source FAD will discard the low priority voice frames if the queue length at the source FAD exceeds the MCT, and the buffer threshold setting at the FR switch connected to the last link does not have a significant effect on the network performance. Therefore, the network performance shows no significant sensitivity to the value of the MCT in this case.

A more interesting case occurs when five reference voice PVCs and one background voice PVC from a different connection are multiplexed on the last low speed link. The last link has different congestion states from the access link, and the buffer threshold setting will have a significant effect on the network performance. The SCT is constant at 400 bytes while the MCT is varied.

The mean voice frame loss rate increases as the MCT increases below 200 bytes, as shown in Figure 5.15, because as the MCT increases, the source FAD drops low priority voice frames

Figure 5.15 Voice frame loss rate vs. MCT

less frequently and the queue length in the FR network increases. Therefore, the potential that the queue length exceeds the SCT increases. For MCT greater than 200 bytes, the mean voice frame loss rate does not change as the MCT increases, because as it approaches the SCT, the SCT becomes the decisive factor. As the MCT increases, more frames enter the network, the congestion level is elevated and the mean voice queueing delay increases, as shown in Figure 5.16.

Figure 5.16 Voice queueing delay vs. MCT

The voice queueing delay does not change much as the MCT changes above 200 bytes, as MCT approaches SCT.

The voice throughput increases as the MCT increases when the MCT is less than 100 bytes, as shown in Figure 5.17. For the MCT greater than 100 bytes, the voice throughput decreases as the MCT increases, because the large value of the MCT allows too many voice frames to enter the FR network and the congestion level rises as the value of the MCT continues to increase. When the MCT exceeds 200 bytes, the voice throughput does not continue to decrease because the MCT approaches the SCT.

Figure 5.17 Voice throughput vs. MCT

The mean voice loss burst length shows no sensitivity to the value of the MCT as it defines only the mild congestion states. However, long bursts of voice frame losses occur in the severe and absolute congestion state. Therefore, the MCT setting does not affect the mean length of voice frame losses.

MCT=50~100 bytes appears to be the best choice for low voice loss rate, short voice frame loss burst, small voice queueing delay and high voice throughput.

5.2.5 SCT

If only one of the six PVCs over the last link is a background PVC, the simulation results do not show significant differences in network performance as the SCT changes. If four of the six voice PVCs over the last link are reference voice PVCs and two of them are background PVCs, the difference of the network performance becomes significant as the SCT varies. Let the MCT=50 bytes and the SCT vary from 100 bytes to 900 bytes. Figures 5.19 through 5.21 show the effect of change of the SCT for voice on the network performance. From Figure 5.18, the

Figure 5.18 Length of voice loss burst vs. the SCT

mean length of the voice loss burst increases as the SCT increases. As the SCT increases, the network will begin to drop DE voice frames at the higher congestion level when the queue length is already very long, causing the mean loss burst length to increase. Therefore, the SCT must be restricted to guarantee the quality of reconstructed voice. From Figure 5.19, when the SCT is less than 600 bytes, the mean voice frame loss rate decreases as the SCT increases, but when the SCT is greater than 600 bytes, the mean voice frame loss rate increases as the SCT increases, because a large number of frames are discarded due to excess delay. From Figure 5.20, the voice delay increases as the SCT increases, because a larger value of the SCT will make the queue longer. From Figure 5.21, the voice throughput increases as the SCT increases, because a larger value of SCT allows more voice frames to enter the FR network.

In order to obtain good quality of reconstructed voice, short voice loss burst, low voice loss rate, small voice delay and high voice throughput should be achieved. However, as shown above, for setting the optimum value for the SCT, there is a conflict between the voice loss burst length and the voice frame loss rate and another between voice throughput and voice queueing delay. The simulation results show that, overall, the mean voice frame loss rate is very low (<5%)

Figure 5.19 The voice loss rate vs. the SCT

Figure 5.21 The voice throughput vs. the SCT

Figure 5.20 The voice delay vs. the SCT

and the voice throughput is high enough (>58 frames/s per voice PVC, e.g. 12.7 kbits/s per voice PVC). The key problem is to choose a suitable value of the SCT to get acceptable voice loss burst length and voice loss rate. In order to limit the voice burst length to less than two voice codec-PDUs, and limit the voice delay to less than 36 ms, the SCT should be less than 200 bits.

5.2.6 CLLM frequency

The traffic model used in this section is the same as that stated in Chapter 2, except that five voice PVCs on the last link are reference PVCs and one is background. Figures 5.22 to 5.25 show the effect of the CLLM frequency on the voice throughput, voice queueing delay, voice frame loss rate and voice loss burst length respectively. If the intergenerating time of the CLLM message is longer than 100 ms, voice throughput decreases sharply, and voice queueing delay, voice frame loss rate and length of loss bursty increase dramatically as the CLLM message intergenerating time increases. If the intergenerating time of the CLLM message is very long, the source FAD will have already sent many voice frames before receiving the congestion notification signal causing performance degradation. On the other hand, an inappropriately short

intergenerating time of the CLLM message also leads to performance degradation because the high CLLM message generating frequency will impose heavy additional traffic to the FR network. Figures 5.22 through 5.25 show that when the intergenerating time of CLLM message

Figure 5.22 Voice throughput vs. intergenerating time of CLLM message

is between 50 ms and 100 ms the voice throughput reaches the maximum value, and the voice queueing delay, the voice frame loss rate and the voice loss burst length almost reach the minimum value. Therefore, under the given conditions, 100 ms is chosen as the optimal value of the intergenerating time of the CLLM message.

5.3 Sensitivity Tests

5.3.1 Sensitivity to Access and Output Line Speeds

Although the speeds of FR switches and T1 links also affect voice delay, throughput, and loss behavior in the reference connection, the speed of the low speed access and exit links is the most significant factor affecting network performance. Simulation shows that the mean voice loss burst length is approximately one as the line speed varies from 60 Kbits/s to 100 Kbits/s, showing that the proposed control scheme is effective over a wide range of link speeds. Figures 5.26 to

Figure 5.23 Voice delay vs. intergenerating time of CLLM message

Figure 5.24 Voice loss burst length vs. intergenerating time of CLLM message

Figure 5.25 Voice loss rate vs. intergenerating time of CLLM message

5.28 show that as the line speed increases, the mean voice frame loss rate and the mean voice queueing delay decrease, and the voice throughput increases. Increased speed of access and exit links will broaden the bottleneck of the reference connection, so that the congestion level in it will be reduced. Therefore, increase of line speed is very significant for improvement of network performance.

5.3.2 Sensitivity to CLLM Processing Time

When congestion occurs in a FR switch, a CLLM message is generated, sent to the transmitting PVC experiencing the congestion and relayed backward to the source FAD using this PVC. The processor at the FR switch is responsible for generating the CLLM message; the CLLM processing time is defined as the time that it takes the processor to generate a CLLM message or to relay a CLLM message from one node to the next one (refer to Section 4.5). When the incoming traffic at the FR switch is heavy, the load of the processor will also be heavy. Therefore, it is necessary to investigate the sensitivity of the results to CLLM time. It is assumed that the CLLM procedure is activated every 100 ms and the CLLM processing time is varied from 1 to 5

Figure 5.26 The voice loss rate vs. the access/output line speed

Figure 5.27 The voice delay vs. the access/output line speed

Figure 5.28 The voice throughput vs. the access/output line speed

ms. The simulation results showed no change of network performance over this range. Results are shown in Figures 4.4 to 4.7 in Section 4.5.

If the congestion information can be fed back efficiently, the source FAD will be able to effectively limit the voice traffic accessing the network, and the network will not experience frequent congestion. Under this condition, CLLM messages are not generated at a high frequency. Therefore, an increase of the CLLM processing time does not cause a significant increase of the processor load and the network performance is not affected substantially. Table 3 shows the effect of CLLM time change on the processor load factor. Assuming the number of reference voice PVCs is six, the processor load factor increases only from 0.8448 to 0.8871. Therefore, the CLLM mechanism is an effective congestion notification method for voice traffic, with robustness to a wide range of CLLM processing times.

Table 3 Effect of CLLM time change on the processor load

CLLM time (ms)	Processor load factor
1	.8448
2	.8512
3	.8675
4	.8792
5	.8871

5.3.3 Sensitivity to Background Traffic Levels

Both the amount of total background traffic and the percentage of voice traffic will affect the reference voice traffic

Curve 1 in Figures 5.29 to 5.31 shows that when the percentage of voice is lower than 70%, the performance does not change significantly if the amount of total background traffic is 300 Kbits/s. If the percentage of voice is higher than 70%, the performance of the voice traffic degrades dramatically: the throughput decreases substantially, the voice queueing delay soars sharply, and the voice frame loss rate and the loss burst length increase to an unacceptable degree. The reason is that the increase of the voice percentage will increase the traffic load on the processor, since the mean length of the voice frames is much shorter than the data frames. In the intermediate nodes, even if the total value of traffic does not exceed the capacity of the output T1 link, the processor can become the bottleneck. Figure 5.32 shows the relationship between the processor utilization and the voice percentage when the total value of background traffic is 500 Kbits/s. The percentage of voice traffic should be controlled.

As the total value of traffic increases, the processor load also increases. Under this circumstance, even a small voice percentage increase will dramatically increase the processor load. The allowable percentage of voice also become lower, as shown in from Figures 5.29 through 5.31.

Figure 5.29 Effect of background traffic variation on voice throughput

Figure 5.30 Effect of background traffic variation on voice delay

Figure 5.31 Effect of background traffic variation on voice loss rate

Figure 5.32 Processor load vs. voice percentage

5.3.4 Sensitivity to Length of Silence and Talkspurt Periods

Different speakers generate silence gaps and talkspurt with different distributions of silence length and talkspurt length. The performance of networks with the proposed control scheme should be robust over a wide range of these lengths.

As defined in Chapter 2, β_s is the mean length of silence gap and α_t is the mean length of talkspurt. Usually the talkspurt is slightly shorter than the silence.

For the range of silence lengths, from $\beta_s=200$ ms to $\beta_s=9$ s, the same simulation results as shown in Figures 4.4 to 4.7 are obtained. The reason that the length of talkspurt does not change network performance is that by using EHCv2_16K codec algorithm, both talk spurt and silence contain two types of frames, low priority frames that can be dropped during the congestion, and high priority frames that should be always transmitted. In the proposed control scheme during congestion, the source FAD drops low priority frames instead of only silence frames.

5.3.5 Sensitivity to the Number of Hops

An increased number of hops in the reference connection could have an effect on network performance, because the length of the feedback control loop increases.

Simulation shows that an increased number of hops does not have a significant effect on the voice throughput and the voice frame loss rate, but there is a small effect on the voice queueing delay, as shown in Figure 5.33

The increased number of hops means that the number of switches and the number of transmission trunk increase. Therefore, the end-to-end delay will increase because the processing delay at the switch, the transmission delay and the propagation delay increase. The increased number of hops also increases the number of queues in the reference connection; therefore, the voice queueing delay also increases.

By increasing one hop, the processing delay increase $465\mu s$. The transmission delay of voice frames increases $27 \times 8 / (1.544 \times 10^6) = 139.9 \times 10^{-6} s = 139.9\mu s$

It is assumed that one hop has the length of 1000 km; thus, the increased propagation

delay will be $1000 \times 10^3 / 2 \times 10^8 = 0.005s = 5ms$

The total fixed part of the end-to-end delay is $5ms + 465\mu s + 139.9\mu s = 5.6ms$.

Besides the above fixed portion of increased end-to-end delay, there is also a variable portion of increased delay, i.e. queueing delay at the input and output queues in an additional switch. Because the output line is a T1 link which has a very fast transmission rate, the queueing delay at the output queue is very small. If there are 6 EHCV2-16K voice PVCs over the reference connection (worst voice congestion level), the queueing delay at the output queue is approximately 83 μs . The queueing delay at the input queue is approximately 1.45ms, which is relatively larger than that at the output queue.

The input queueing delay and the propagation delay are dominant factors of increased end-to-end delay. When the number of hops increases by one, the end-to-end delay increases approximately by 7.13ms if the number of voice PVCs is six.

Although the end-to-end delay increases in relation to the number of hops increased, the proposed control scheme still works well. The increased delay in the feedback control loop is small enough and the congestion notification can still reach the source end within a short time. Certainly, the number of hops should be limited to prevent the end to end delay from exceeding the maximum allowable amount. As the number of hops increases, the voice loss rate and the length of voice loss burst increase because the feedback time of the CLLM notification increases. When the number of hops exceeds eight, the voice frame loss rate will be greater than 6% and the length of voice loss burst will be greater than three voice codec-PDUs. Therefore, the maximum number of hops should be limited to seven.

Figure 5.33 The voice delay vs. number of hops

Chapter 6. Conclusion

6.1 Conclusions

To exploit the advantages of the FR network such as the simplicity of the protocol structure, its high performance and possibilities for new applications, and to overcome its disadvantages such as limited sources for traffic management, this research proposed an efficient congestion control scheme in the FR network carrying both compressed voice and data traffic. As well, substantial performance evaluation has been done. The proposed control scheme builds on some previous results:

- Compressed voice traffic, which is generated by an embedded codec algorithm comprises two types of voice frames: high priority frames, which should be guaranteed to be transmitted, and low priority frames which can be dropped during the period of congestion [39].

- Voice traffic is prioritized over data traffic on a PVC basis to guarantee voice traffic to achieve minimum delay [38][41].

The principal contribution of this thesis are as follows:

- It is highly possible to implement voice transport in FR networks with some modifications to the nodes and to network management procedures, and these modification do not violate the FR standard. The modifications are to prioritize voice traffic over data traffic and to notify separately voice congestion and data congestion.

- An efficient congestion notification mechanism is essential to guarantee effective control action. Consolidate Link Layer Management (CLLM) is a more effective explicit congestion notification method than Backward Explicit Congestion Notification (BECN), because if there is no traffic in the opposite direction, the congestion notification signal can be relayed to the source when a CLLM message is used, but not when BECN is used.

- In order to satisfy the specific characteristics of voice traffic, separate congestion notification has to be used for voice traffic from the data traffic, because the voice traffic has a more stringent requirement on the frame delay.

- Buffer occupancy in the FR switches should be used to detect the congestion level. The different buffer threshold values MCT, SCT and ACT represent the different congestion levels of mild congestion, severe congestion, and absolute congestion respectively and are different for the voice and data buffers.

- The access and exit low speed link is the bottleneck of the reference connection. Therefore, the CLLM message generation procedure is strictly necessary only at the FR switch which is connected to the destination FAD, and not at the switches which are only connected to other FR switches by the high speed T1 link.

- The source FAD discards low priority frames upon receiving a CLLM or BECN congestion notification signal to prevent excessive traffic from entering the FR networks which would worsen the congestion status.

- Voice DE frames are randomly distributed in the voice stream, and intermediate nodes discard the DE frames during periods of congestion to reduce the traffic load in the FR network so that the congestion status can be alleviated.

- The proposed control scheme is robust over a wide range of variations of the network parameters and network structure. These variations include the silence and talkspurt lengths, voice frame length, pattern of background traffic in the network and the number of hops in the reference connection.

- Optimal settings have been determined for the most significant operational parameters

6.2 Suggestion for Further Studies

In addition to the situations which have been checked in Chapter 5, some other cases need further studies such as:

- The MCT and SCT settings over a wider range of the number of background voice PVCs multiplexed on the last low speed link.
- The network performance under the circumstance that the speed of the access link is different from that of the output link.

- The network performance under the circumstance that all silence frames are low priority frames.
- The network performance under the circumstance that the volume of background traffic on each FR switch is different.
- The backward explicit congestion notification mechanism under the condition that the backward traffic in the reference connection is severely or absolutely congested.
- The short-term performance of the control scheme under sudden, dynamic change in background load conditions.
- Methods of call acceptance control for both voice and data.

In addition for the compressed voice and traditional data traffic, separate congestion notification mechanisms and traffic control schemes for some newly developed application traffic such as LAN to LAN application, facsimile, video, and telemetry traffic need to be extended.

References

- [1] ANSI T1.606, "Integrated Services Digital Network (ISDN) - Architectural Framework and Service Description for Frame Relaying Bearer Service", 1990
- [2] ANSI T1.617, "Integrated Services Digital Network (ISDN) - Digital Subscriber Signalling System No.1 (DSS1) - Signalling Specification for Frame Relay Bearer Service", 1991
- [3] ANSI T1.618, "Integrated Services Digital Network (ISDN) - Core Aspects of Frame Protocol for use with Frame Relay Bearer Service", June, 1991
- [4] CCITT Draft Recommendation G.764, "Voice Packetization - Packetized Voice Protocols", Aug. 1990
- [5] M. Woollacot, "An interim technology (Frame Relay)". *Commun. Networks*, Vol. 8, no. 12, pp. 27-28, 30, December 1991.
- [6] B.T. Doshi, and Nguyen, H. O., "Congestion Control in ISDN Frame-Relay Networks", *AT&T Technical Report*, Vol. 67, Iss. 6, pp.35-46, 1988
- [7] Kim-Joan Chen and K. M. Rege, "A comparative Performance Study of Various Congestion Controls for ISDN Frame-Relay Networks", *Proc. INFOCOM'89*, pp 674-675.
- [8] M. Irfan Ali, "Frame Relay in Public Networks". *IEEE Commun. Mag.*, pp.72-78, March 1992.
- [9] R. Bhardwaj, and B. Jabbari, "ISDN Frame Relay and Its Congestion Control", *Intel J. of Satellite Commun.*, Vol. 9, No. 5, pp. 363-380, Sept.-Oct. 1991
- [10] T. G. Zerbic, "Considering the Past and Anticipating the Future for Private Data Networks". *IEEE Commun. Mag.*, pp.36-46, March 1992.
- [11] J.J. Brosemer and D.J. Enright, "Virtual Networks: Past, Present and Future". *IEEE Commun. Mag.*, pp.80-85, March 1992.
- [12] S.A. Taylor, "Frame Transport Systems". *IEEE Commun. Mag.*, pp.66-70, March 1992.
- [13] M. Rahnema, "Frame Relay and the Fast Packet Switching Concepts and Issues". *IEEE Network Mag.*, pp.18-23, July 1991.
- [14] K. J. Chen, K. K. Y. Ho, and V. R. Saksena, "Analysis and Design of a Highly Reliable Transport Architecture for ISDN Frame-Relay Networks", *IEEE J. Sel. Areas Commun.*, Vol. 7, No. 8, pp.1231-1241 Oct. 1989

2 of/de 2

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS
STANDARD REFERENCE MATERIAL 1010a
(ANSI and ISO TEST CHART No. 2)

- [15] R. J. Cherukuri and J. H. Derby, "Frame Relay: Protocols and Private Network Applications", *Proc. INFOCOM'89*, pp.676-685.
- [16] W. S. Lai, "Frame Relaying Service: An Overview", *Proc. INFOCOM'89*, pp.668-673.
- [17] A. Platt and M. J. Morse, "Traffic Management in Frame Relay Networks", *Computer Networks and ISDN Systems*, Vol. 9, No.1, pp.305-316, Jan. 1992
- [18] S. Civalar, "A Failure Detection Method for Integrated Voice/Data Packet Networks", *IEEE Trans. Commun.*, Vol. COM-39, No. 3, pp. 356-359, March 1991
- [19] S. Mahant, M. H. Sherif, and K. Sriram, "Simulation of Multiservice Frame Relay Operation", *AT&T Document T1Y1.2/91-056*, Oct. 1991
- [20] G. J. Coviello, "Comparative Discussion of Circuit- vs Packet-Switched Voice", *IEEE Trans. Commun.*, Vol. COM-27, No. 8, pp. 1153-1160, Aug. 1979
- [21] S. Dravida, and K. Sriram, "End-to-end Performance Models for Variable Bit Rate Voice over Tandem Links in Packet networks", *IEEE J. Sel. Areas Commun.*, Vol. SAC-7, No. 5, pp. 718-728, June, 1989
- [22] 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
- [23] 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, June, 1983.
- [24] K. Sriram, and D. M. Lucantoni, "Traffic Smoothing Effects of Bit Dropping in a Packet Voice Multiplexer", *IEEE Trans. Commun.*, Vol. 37, No. 7, pp. 703-712, July 1989
- [25] 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
- [26] G. Barberis and D. Pazzaglia, "Analysis and Optimal Design of a Packet-Voice Receiver", *IEEE Trans. Commun.*, Vol. COM-28, No. 2, Feb. 1980
- [27] E. C. Russell, "Building Simulation Models with SIMSCRIPT II.5", CACI Products Company, 1990
- [28] "SIMSCRIPT II.5 Programming Language", CACI Products Company, 1983
- [29] J. Banks, and J. S. Carson, "Discrete-Event System Simulation", Prentice-Hall, 1984
- [30] T. M. Chen, "Dynamic Priority Protocol for Packet Voice", *IEEE J. Sel. Areas Commun.*, Vol. SAC-7, No.5, pp. 623-643, June 1989

- [31] W. Stallings, "Data and Computer Communications", Third edition, New York: McMillan, 1991
- [32] J. S. Turner, and L. F. Wyatt, "A packet Network Architecture for Integrated Services", *Proc. GLOBECOM'83*, Vol. 1
- [33] L. Kleinrock, "Queueing Systems, Volume I: Theory", John Wiley & Sons, 1975
- [34] P. T. Brady, "A Technique for Investigating On-Off Patterns of Speech", *Bell Syst. Tech. J.*, Vol. 44, pp. 1-22, Jan. 1965
- [35] P.T. Brady, "A Statistical Analysis of On-off Patterns of Speech", *Bell Syst. Tech. J.*, Vol. 47, pp.73-91, Jan. 1968
- [36] P. T. Brady, "A Model for Generating On-Off Pattern in Two-way Communications", *Bell Syst. Tech. J.*, Vol. 48, 2445-2472, Sept. 1969
- [37] A. R. Kaye, "Private Communication: Modelling of Frame Relay Networks", Internal Document, August 14, 1992, Carleton University
- [38] A. R. Kaye, "Compressed Voice over Frame Relay: Protocols and Architecture", Internal Document, August 14, 1992, Carleton University
- [39] S. A. Mahmoud, A.R. Kaye, and W. P. LeBlanc, "Transmission of Codec Speech over the Frame Relay Protocol", Internal Document, August 14, 1992
- [40] W. P. LeBlanc, "CELP Speech Coding at Low to Medium Bit Rates", Ph.D thesis, Carleton University, October, 1992
- [41] L. Dong, "Compressed Voice in Integrated Services Frame Relay Networks", Master's thesis, Carleton University, November, 1992
- [42] CCITT Recommendation G. 721, "32 Kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)", Study Group XVIII - Report R 26(C), Aug. 1986.
- [43] W. A. Montgomery, "Techniques for Packet Voice Synchronization", *IEEE J. Sel. Areas Commun.*, Vol. SAC-1, No. 6, pp.1022-1028, 1983
- [44] W. Stallings, "ISDN and Broadband ISDN", second edition, New York: Macmillan, 1992
- [45] A. K. Roy, "Frame Relay and Its Transportation Dilemma", *Proc. of the 1991 Singapore International Conference on Networks, Towards Network Globalisation.* pp. 393-400.
- [46] K. Ab-Hamid and A.G. Waters, "Prediction-based Congestion Avoidance Scheme for Frame relay Networks", *Proc. of 9th UK teletraffic Symposium.* University of Surrey. 8-10 April, 1992

- [47] A.G. Waters and K. Ab-Hamid, "Congestion Control for Frame Relay Networks". *IEE Colloquium on "Frame Relay"* , 22 April, 1992
- [48] K. Ab-Hamid and A.G. Waters, "Kalman Prediction Method for Congestion Avoidance in ISDN Frame Relay Networks" *Proc. of Phoenix Conference on Computer and Communications*, pp.565-571, March 27-30, 1991.
- [49] M. Nassehi, "Window Flow Control in Frame-Relay Networks", *IEEE GLOBECOM'88*, Vol.3, pp. 1784-1790.
- [50] C. Rowden, *Speech Processing*, McGraw-Hill, 1991
- [51] O.G. Jaffe, "Reconstruction of Missing Packets of PCM and ADPCM Encoded Speech", Master's Thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, June 1986.
- [52] 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, pp. 1440-1448, Dec. 1986.
- [53] O.J. Wasem, D.J. Goodman, C.A. Dvorak and H.C. Page, "The Effect of Waveform Substitutions on the Quality of PCM packet Communications", *IEEE Trans. on Acoustics, Speech and Signal Processing*, Vol.36, pp. 342-348, March 1988.

END

2 1 - 0 7 - 9 4

FIN