



Estudo de Caracterização de Tráfego e Alocação de Banda para Redes ATM

Dissertação de Mestrado apresentada à Faculdade de Engenharia Elétrica da Universidade Estadual de Campinas, Departamento de Comunicações, como parte dos requisitos exigidos para a obtenção do título de

→ M
Mestre em Engenharia Elétrica

Autor

Firmiano Ramos Perlingeiro → Autor
Engenheiro Eletrônico pela PUC-RS em 1981

Orientador

Prof. Dr. Lee Luan Ling

Este exemplar corresponde a redação final da tese defendida por Firmiano Ramos Perlingeiro e aprovada pela Comissão Julgada em 16 de julho de 1999.

Banca Examinadora

Lee Luan Ling
Orientador

Prof. Dr. Lee Luan Ling (Presidente)
PhD em Engenharia Elétrica - 1991 - Cornell University - USA

Prof. Dr. Ivanil Sebastião Bonatti
Doutor em Automática - 1981 - Toulouse - França

Prof. Dr. Nelson Luis Saldanha da Fonseca
PhD em Engenharia de Computação - 1994 - Southern California - USA

Campinas, 16 de Julho de 1999

UNICAMP
BIBLIOTECA CENTRAL

UNIDADE BAE/FEEC	
N.º CHAMADA:	
V.	Ex
FRANCO BC/	39658
PROJ.	229/99
Q	D <input checked="" type="checkbox"/>
R. 20	R. B. 11,00
	07-12-99
CPD	

CM-00137473-5

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

P421e

Perlingeiro, Firmiano Ramos

Estudo de caracterização de tráfego e alocação de banda para redes ATM / Firmiano Ramos
Perlingeiro.--Campinas, SP: [s.n.], 1999.

Orientador: Lee Luan Ling

Dissertação (mestrado) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Multiplexação. 2. Telecomunicações - Sistemas de comutação. 3. Telecomunicações - Tráfego. 4. Videoteleconferência. 5. Compressão de dados. 6. Voz. 7. Fractais. 8. Garantia de qualidade. I. Lee, Luan Ling. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título.

Resumo

A crescente competição no mundo dos negócios tem exigido que as empresas racionalizem e otimizem a utilização de seus recursos. Aliado a esse fato, o desenvolvimento de novos recursos computacionais e de telecomunicações nos últimos anos têm promovido uma evolução das aplicações e um crescimento da necessidade de serviços sem precedentes. Ambos os fatos têm causado substanciais alterações nas características de tráfego. Como consequência, a alocação de banda torna-se um importante aspecto no projeto de redes de telecomunicações. As duas maiores abordagens para alocação de banda levam em consideração a probabilidade de perda de células e o atraso das células mas sem considerar a característica auto-similar do tráfego, podendo subestimar a banda necessária.

O objetivo deste trabalho é apresentar um método de caracterização de tráfego facilitando o cálculo da banda necessária para redes ATM. Neste estudo generalizamos o método proposto por Kesidis provendo uma estimativa mais realista considerando a característica auto-similar do tráfego. O resultado é otimizado através de método experimental e combinado com o método de Le Boudec. Neste estudo foram aplicados métodos analíticos, heurísticos e de simulação para validar o método proposto.

Abstract

The increase of economic competition in the world has demanded companies rationalize and optimize the use of their resources. Jointly with this fact, the development of new computational and telecommunication facilities in the recent years have promoted some unprecedented evolution of application and increasing necessity of communication services. Both facts have caused remarkable changes in traffic characteristics. Therefore the bandwidth allocation become an important issue for telecommunication networks design. Cell loss probability and cell latency have been two major approaches to estimate the effective bandwidth of communication network connections. Unfortunately, most of methods underestimated values when traffic is self-similar.

The goal of this work is to present traffic characterization method for ATM networks, facilitating their bandwidth allocation and design. In this study, we generalize the effective bandwidth estimation method proposed by Kesidis providing a more realistic estimate value whether the traffic is self-similar or not. Experimentally the estimated effective bandwidth is optimized to reduce the impact provoked by traffic highly self-similar and the results are combined with the Le Boudec's approach. For this end, analytical, heuristical and simulation methods were used jointly both in design and validation of the proposed method.

Agradecimentos

À minha esposa Belinda, e filhas Gisele, Bruna e Gabriela pelo apoio e compreensão durante a realização deste trabalho.

Ao meu pai e minha mãe que possibilitaram e influenciaram a minha formação profissional e de cidadão.

À *PETROBRAS* pela minha liberação para esta fase de estudos, em especial aos engenheiros Luiz Thadeu M. Filártiga e Luiz Gonzaga Oswald.

Ao meu orientador, Prof. Dr. Lee Luan Ling por ter me aceito como seu orientado, por sua atenção, ensinamentos, e incentivo.

Aos professores do Departamento de Comunicações, Departamento de Telemática e Departamento Automação e Computação, em especial ao Prof. Dr. Dalton Soares Arantes e Prof. Dr. Ivanil Bonatti por seus valiosos ensinamentos.

Aos colegas do DECOM, DT e DCA pelo convívio, colaboração e aprendizado durante este período de estudos, em especial aos colegas Antônio M. Alberti e Ernesto Luiz Andrade Neto pelo apoio a simulação com tráfego de vídeo multiplexado através do SimATM.

À Banca Examinadora meu reconhecimento e respeito.

Conteúdo

	Página
LISTA DE FIGURAS	vii
LISTA DE TABELAS	viii
SIGLAS E ACRÔNIMOS	ix
CAPÍTULO 1	
1.1 Histórico	1
1.2 Introdução Geral	3
CAPÍTULO 2	Redes ATM
2.1 Introdução	9
2.2 Interfaces	10
2.3 Formato da Célula ATM	12
2.4 Conexões na Rede ATM	14
2.5 As Camadas do Protocolo ATM	16
2.6 Parâmetros de Tráfego e Qualidade de Serviço em Redes ATM	19
2.7 Policiamento de Tráfego em Redes ATM	25
2.8 Conclusões	29
CAPÍTULO 3	Modelos de Tráfego
3.1 Introdução	30
3.2 Visão das Escalas de Tempo	30
3.3 Modelos a Nível de Célula	33
3.3.1 Corrente de Renovação	34
3.3.2 Processo de Chegada de Lote Markoviano em Tempo Discreto	36
3.3.3 Filas em Escala de Células	37
3.4 Modelos em Escala de Rajada	39
3.4.1 Processo On/Off	40
3.4.2 Processos de Poisson Markov Modulados	43
3.4.3 Fonte de Tráfego Fluido Markoviano	47
3.4.3.1 Modelo de Fonte Fluída para Voz	49
3.4.3.2 Modelo de Fonte Fluída para Fontes de Taxa Variável	50
3.4.4 Modelos Auto-Regressivos	51
3.4.4.1 Modelo Auto-regressivo Linear (AR)	52
3.4.4.2 Modelo Média Móvel (MA)	53
3.4.4.3 Modelo Auto-Regressivo com Média Móvel - ARMA	53
3.4.4.4 Modelo Auto-Regressivo com Média Móvel Integrada - ARIMA	56
3.4.5 Processos Auto-Similares	57
3.4.5.1 Estimando o parâmetro de Hurst	62

3.4.5.2	Movimento Browniano Fracionário e Ruído Gaussiano Fracionário	63
3.4.5.3	Processo ARIMA Fracionário	65
3.5	Modelos para Escala em Conexão	67
3.6	Conclusões	68
CAPÍTULO 4	Caracterização do Tráfego	
4.1	Introdução	69
4.2	Parâmetros para Caracterização de Tráfego	70
4.3	Caracterização do Tráfego de Dados	72
4.3.1	Ambiente Corporativo	72
4.3.2	Medidas e Resultados obtidos para o Tráfego de Dados	75
4.3.3	Conclusões para o Tráfego de Dados	83
4.4	Caracterização do Tráfego de Vídeo	84
4.4.1	Ambiente de Estudo	84
4.4.2	Análise do Tráfego de Vídeo	86
4.4.3	Conclusões sobre a Caracterização do Tráfego de Vídeo	90
4.5	Conclusões	90
CAPÍTULO 5	Método de Alocação de Banda em Uma Conexão ATM	
5.1	Introdução	92
5.2	O Método de Le Boudec	94
5.3	O Método de Kesidis	97
5.3.1	O Método de Kesidis e o Tráfego Auto-Similar	100
5.4	O Método Proposto para o Cálculo da Banda Efetiva	101
5.4.1	Estudos Experimentais	104
5.4.2	Estudos com Tráfego de Vídeo Multiplexado	107
5.5	Metodologia de Kesidis versus Método Proposto	109
5.6	Metodologia Resultante da Combinação dos Métodos Baseados em Perda e em Atraso	110
5.7	Conclusões	113
CAPÍTULO 6	Conclusão Geral	114
	REFERÊNCIAS BIBLIOGRÁFICAS	115
APÊNDICE A	Teoria dos Grandes Desvios e Banda Efetiva	121
1A.	Introdução	121
2A.	Teorema de Chernoff	122
3A.	Princípio dos Grandes Desvios	123
4A.	Teorema de Sanov	125
5A.	Teorema de Gärtner-Ellis	126

6A. Banda Efetiva	128
APÊNDICE B Listagem dos Arquivos de Tráfego Utilizados	131
APÊNDICE C Método de Cálculo Proposto para Banda Efetiva por Perda	132
APÊNDICE D O Padrão MPEG e AVI	133
1D. Introdução	133
2D. O Padrão MPEG	134
2D1. O MPEG1	134
2D2. O MPEG2	136
2D3. O MPEG4	137
2D4. O MPEG7	138
3D. O Padrão AVI	138
APÊNDICE E SIMULADOR	140
1E. Introdução	140
2E. Simulador Morin	140
APÊNDICE F INSTRUMENTOS UTILIZADOS	142
1F. Introdução	142
2F. DA310	142
3F. DA320	142

Lista de Figuras

	Página
Figura 1 - Ganho de Multiplexagem	4
Figura 2 - Rede ATM e suas Principais Interfaces	11
Figura 3 - Formato das Células	12
Figura 4 - Conexões ATM	14
Figura 5 - Relação <i>SCR</i> e <i>MBS</i>	21
Figura 6 - Modelo de Densidade de Probabilidade do <i>CTD</i>	23
Figura 7 - Algoritmo GCRA	27
Figura 8 - Exemplo de CGRA	28
Figura 9 - Escalas de Tempo para o Tráfego ATM	31
Figura 10 - Multiplexador ATM	37
Figura 11 - Chegadas Aleatórias por Partição	38
Figura 12 - Comportamento da Voz	40
Figura 13 - Modelo de Dois Estados para Fonte de Voz	41
Figura 14 - “ <i>Buffer</i> ” de Acesso	42
Figura 15 - Modelo de <i>N</i> Fontes para Voz	42
Figura 16 - MMPP de Dois Estados	44
Figura 17 - “ <i>Buffer</i> ” de Acesso	49
Figura 18 - Operador de Atraso	54
Figura 19 - Processo ARMA	55
Figura 20 - Processo ARIMA	56
Figura 21 - Cenário Hipotético	59
Figura 22 - Decaimento Exponencial x Hiperbólico	61
Figura 23 - Método das Variâncias	62
Figura 24 - Tráfego Frame-Relay	73
Figura 25 - Tráfego VSAT Regional	74
Figura 26 - Tráfego de Automação Industrial	75
Figura 27 - Taxa de Perda para Tráfego Internet	80
Figura 28 - Taxa de Perda para Tráfego VSAT	81
Figura 29 - Taxa de Perda para Tráfego de Automação	81
Figura 30 - Taxa de Perda para Tráfego Ethernet	82
Figura 31 - Taxa de Perda para Tráfego de Gerência	82
Figura 32 - Taxa de Perda para Tráfego Frame-Relay	83
Figura 33 - Ambiente para Captura do Video	85
Figura 34 - Taxa de Perda para Tráfego de Video MPEG	89
Figura 35 - Taxa de Perda para Tráfego de Video AVI	89
Figura 36 - Taxa Instantânea - Tráfego VBR	93
Figura 37 - Banda Efetiva para um Atraso <i>D</i>	95
Figura 38 - Fonte On/Off	95
Figura 39 - Número de Células no “ <i>Buffer</i> ”	96
Figura 40 - Banda Efetiva Fluido Markoviano On/Off - Kesidis	99
Figura 41 - Tráfego Auto-Similar	101
Figura 42 - Banda Efetiva para $H = 0,6$	104
Figura 43 - Banda Efetiva para $H = 0,7$	105

Figura 44	- Banda Efetiva para $H = 0,8$	105
Figura 45	- Banda Efetiva para $H = 0,9$	106
Figura 46	- Ambiente de simulação vídeo multiplexado	107
Figura 47	- Banda Efetiva Proposta vs simulação SimATM	108
Figura 48	- Comportamento do atraso	109
Figura 49	- Proposto sem Correção x Le Boudec	111
Figura 50	- Proposto com Correção x Le Boudec	112
Figura 1D	- Esquemas Entrelaçado e Progressivo	136
Figura 2D	- Esquemas Entrelaçado e Progressivo	136
Figura 1E	- “ <i>Buffer</i> ” no Comutador ATM	140

Lista de Tabelas

Tabela 1	- Parâmetros ATM Forum e ITU-T	24
Tabela 2	- Eventos nas Escalas de Tempo	31
Tabela 3	- Diferenças entre Processos	63
Tabela 4	- Parâmetro H sem e com Agregação	76
Tabela 5	- Parâmetro PRM sem e com Agregação	76
Tabela 6	- Parâmetros CSQ	77
Tabela 7	- Parâmetros IDC	77
Tabela 8	- Parâmetro MBS	77
Tabela 9	- Tráfego Internet	78
Tabela 10	- Tráfego VSAT	78
Tabela 11	- Tráfego de Automação em Rede Local	78
Tabela 12	- Tráfego de Rede Local Ethernet	79
Tabela 13	- Tráfego de Gerência de Rede	79
Tabela 14	- Tráfego Frame-Relay	79
Tabela 15	- Parâmetro H do Tráfego de Vídeo	86
Tabela 16	- Parâmetro PRM do Tráfego de Vídeo	86
Tabela 17	- Parâmetro CSQ do Tráfego de Vídeo	87
Tabela 18	- Parâmetro IDC do Tráfego de Vídeo	87
Tabela 19	- Parâmetro MBS do Tráfego de Vídeo	87
Tabela 20	- Tráfego de Vídeo AVI	87
Tabela 21	- Tráfego de Vídeo MPEG	88
Tabela 1B	- Listagem dos Arquivos de Tráfego	131

Siglas e Acrônimos

AAL	<i>ATM Adaptation Layer</i>
ABR	<i>Avaiable Bit Rate</i>
ABT	<i>ATM Block Transfer</i>
AR	<i>Autoregressive Process</i>
ARMA	<i>Autoregressive Moving Average Process</i>
ARIMA	<i>Autoregressive Integrated Moving Average Process</i>
ARPA	<i>Advanced Research Project Agency</i>
ATM	<i>Asynchronous Transfer Mode</i>
ABT	<i>ATM Block Tranfer</i>
AVI	<i>Audio Video Interleave</i>
B-ISDN	<i>Broadband Integraded Service Digital Network</i>
BT	<i>Burst Tolerance</i>
BER	<i>Bir Error Rate</i>
CAC	<i>Connection Admission Control</i>
CBR	<i>Constant Bit Rate</i>
CDV	<i>Cell Delay Variation</i>
CDVT	<i>Cell Delay Variation Tolerance</i>
CER	<i>Cell Error Rate</i>
CERN	<i>European Center For Nuclear Research</i>
CLP	<i>Cell Loss Priority</i>
CLR	<i>Cell Loss Ratio</i>
CMR	<i>Cell Misinsertion Ratio</i>
CSQ	<i>Square Coeficient of Variation</i>
CTD	<i>Cell Transfer Delay</i>
CV	<i>Coeficient of Variation</i>
DBR	<i>Deterministic Bit Rate</i>
DCE	<i>Data Communication Equipment</i>
DCT	<i>Discrete Cosine Transform</i>
D-BMAP	<i>Discrete Time Batch Markovian Arrival Process</i>
DMAP	<i>Discrete Time Markovian Arrival Process</i>
DNA	<i>Digital Network Architecture</i>
DSBP	<i>Double Stochastic Bernoulli Process</i>
DSPP	<i>Double Stochastic Poisson Process</i>
DTE	<i>Data Terminal Equipment</i>
FARIMA	<i>Fractional ARIMA</i>
FBM	<i>Fractional Brownian Motion</i>
FDM	<i>Frequency Division Multiplexing</i>
FDDI	<i>Fiber Distributed Data Interface</i>
FGN	<i>Fractional Gaussian Noise</i>
FIFO	<i>First in, First out</i>
FIR	<i>Finite Impulse Response Filter</i>
FR	<i>Frame-Relay</i>
GCRA	<i>Generic Cell Rate Algorithm</i>

GFC	<i>Generic Flow Control</i>
H	<i>Hurst Parameter</i>
HEC	<i>HeaderError Check</i>
IBT	<i>Intrinsic Burst Tolerance</i>
IDC	<i>Index of Dispersion of Counts</i>
IDCT	<i>Inverse Discrete Cosine Transform</i>
IDN	<i>Integrated Data Network</i>
IP	<i>Internet Protocol</i>
ITU-T	<i>International Telecommunication Union - Standardization Sector</i>
LAN	<i>Local Area Network</i>
LANE	<i>Lan Emulation</i>
LCT	<i>Last Compliance Time</i>
MA	<i>Moving Average</i>
MAP	<i>Markov Arrival Process</i>
MBS	<i>Maximum Burst Size</i>
MCR	<i>Minimum Cell rate</i>
MMBP	<i>Markov Modulated Bernoulli Process</i>
MMDP	<i>Markov Modulated Deterministic Process</i>
MMPP	<i>Markov Modulated Poisson Process</i>
MPEG	<i>Motion Picture Experts Group</i>
MPOA	<i>Multiprotocol Over ATM</i>
MPLS	<i>Multiprotocol Label Switching</i>
NNI	<i>Network-to-Network Interface</i>
NPC	<i>Network Parameter Control</i>
OAM	<i>Operations Administration and Maintenance</i>
OSI	<i>Open System Interconnection</i>
P-NNI	<i>Private NNI</i>
PCM	<i>Pulse Code Modulation</i>
PCR	<i>Peak Cell Rate</i>
PMD	<i>Physical Media Dependent</i>
PRM	<i>Peak to Mean Ratio</i>
PT	<i>Payload Type</i>
PDH	<i>Plesiochronous Digital Hierarchy</i>
PDU	<i>Protocol Data Unit</i>
PVC	<i>Permanent Virtual Connection</i>
QoS	<i>Quality of Service</i>
SAR	<i>Segmentation And Reassembly</i>
SDH	<i>Synchronous Digital Hierarchy</i>
SDU	<i>Service Data Unit</i>
SCR	<i>Sustainable Cell rate</i>
SECBR	<i>Severely Errored Cell Block Ratio</i>
SNMP	<i>Simple Network Management Protocol</i>
SMDS	<i>Switched Multimegabit Data Service</i>
SNA	<i>System Network Architecture</i>
SONET	<i>Synchronous Optical Network</i>
S-PVC	<i>Soft Permanent Virtual Connection</i>
STM	<i>Synchronous Transfer Mode</i>

SVC	<i>Switched Virtual Channel</i>
TAT	<i>Theoretical Arrival Time</i>
TC	<i>Transmission Convergence</i>
TCP	<i>Transport Control Protocol</i>
TDM	<i>Time Division Multiplexing</i>
TGD	<i>Teoria dos Grandes Desvios (LDT – Large Deviation Theory)</i>
UBR	<i>Unspecified Bit Rate</i>
UNI	<i>User Network Interface</i>
UPC	<i>Usage Parameter Control</i>
VBR	<i>Variable Bit Rate</i>
VC	<i>Virtual Channel</i>
VCC	<i>Virtual Channel Connection</i>
VCI	<i>Virtual Channel Identifier</i>
VCL	<i>Virtual Channel Link</i>
VPC	<i>Virtual Path Connection</i>
VP	<i>Virtual Path</i>
VPI	<i>Virtual Path Identifier</i>
VPL	<i>Virtual Path Link</i>
VSAT	<i>Very Small Aperture Terminal</i>
WAN	<i>Wide Area Network</i>
WDM	<i>Wavelength-Division Multiplexing</i>
WWW	<i>World Wide Web</i>

Capítulo 1

1.1 Histórico

O cenário atual das comunicações está fundamentado em duas vertentes tecnológicas com finalidades e épocas de surgimento diferentes, mas intimamente relacionadas e com um futuro comum: a telefonia e a comunicação de dados.

A telefonia remonta do final do século XIX, predominando como meio de comunicação entre indivíduos neste século. A telefonia baseia-se na transformação dos sinais de voz (onda sonora) em sinais elétricos (o microfone de carvão foi inventado por Edison em 1875) [41], que eram inicialmente transmitidos diretamente sobre fios de cobre.

A tecnologia telefônica manteve-se analógica até poucos anos atrás, principalmente nos países não desenvolvidos. Embora a técnica de transmissão digital tenha sido concebida a mais de meio século, com a patente do PCM (*Pulse Code Modulation*) em 1938 por Alex H. Reeves do laboratório ITT (*International Telephone and Telegraph Corporation*) de Paris, somente começou efetivamente ser utilizada nos anos 70, nos Estados Unidos e Europa. O primeiro rádio digital implementado na América do Norte, foi fabricado pela Marconi canadense em 1970. Os circuitos digitais foram desenvolvidos para transmitir voz na forma digital [25][79].

A outra vertente tecnológica nasceu mais tarde, com o surgimento de computadores e a necessidade desses sistemas se comunicarem. No início dos anos 60 nasceu o teleprocessamento, onde uns poucos terminais de computadores foram instalados a distância. Nessa mesma época o governo americano, através do Departamento de Defesa, iniciou os estudos para criação de uma rede de computadores estratégicos, ligando centros de pesquisas e universidades. Esse projeto foi coordenado pela ARPA (*Advanced Research Project Agency*) e a rede começou a operar efetivamente em 1972. A rede ARPA proporcionou a origem do conceito

de transmissão de dados baseada em datagramas e o surgimento dos protocolos IP e TCP que são os protocolos base da atual rede Internet [17].

Na década de 70, paralelamente ao desenvolvimento da rede ARPA, foram concebidas algumas soluções proprietárias para a comunicação entre computadores, como arquitetura SNA (*Systems Network Architecture*) e mais tarde a DNA (*Digital Network Architecture*).

Em 1976 o então CCITT (*International Telegraph and Telephone Consultive Committee*), hoje ITU-T, publicou a recomendação X.25, para comunicação de pacotes em redes públicas [50].

Nos anos 80 surgiu a tecnologia de redes locais, fornecendo ambientes heterogêneos e maior possibilidade de distribuição da informação.

A primeira tentativa de integração dos serviços de voz e dados ocorreu em 1972 com a Rede Digital Integrada (IDN- *Integrated Digital Network*), onde foram utilizados os mesmos comutadores e circuitos digitais para diferentes serviços. A evolução da IDN ocorreu em 1984 com o chamado Livro Laranja, Recomendação Série I, que descreve a Rede Digital Integrada de Telefonia que posteriormente evoluiu para RDSI-FE (N-ISDN - *Narrowband Integrated Services Digital Network*)[80][82].

A disponibilidade de circuitos digitais permitiu os usuários transmitirem dados sem necessidade de modems, através de linhas alugadas das concessionárias, com taxas de transmissão na ordem de kbps até dezenas de Mbps, utilizando protocolos de maior desempenho como Frame Relay e SMDS (*Switched Multimegabit Data Service*)[5].

Hoje com a disponibilidade de sistemas óticos e de rádios digitais de alta capacidade de transmissão, viabilizados através das hierarquias digitais síncronas como SONET (*Synchronous Optical Network*) e SDH (*Synchronous Digital Hierarchy*) e mais recentemente a tecnologia WDM (*Wavelength Division Multiple Access*), a tecnologia ATM(*Asynchronous Transfer Mode*) se consolida como o modo de transmissão para a RDSI-FL (*B-ISDN Broadband Integrated Services Digital Network*). E mais, dada a grande escalabilidade do ATM, esta tecnologia está penetrando largamente em ambientes de redes locais, competindo com outras

soluções como o FDDI (*Fiber Distributed Data Interface*), Fast Ethernet e GigaBit Ethernet [25][81].

Outro importante fator na consolidação da tecnologia ATM é a padronização. Organizações como o ATM Forum, criado no início desta década e ITU-T têm emitido inúmeras especificações acelerando o processo de padronização da tecnologia ATM.

1.2 Introdução Geral

Até recentemente o tráfego de voz, dados e vídeo eram suportados por redes determinísticas, ou seja, com alocação de banda fixa em função de partições de tempo (*timeslots*). O dimensionamento da capacidade dos circuitos de comunicação neste tipo de rede é fundamentado principalmente na taxa de pico e na ocupação desses canais.

O advento das redes de alta velocidade, baseado na tecnologia ATM mudou drasticamente esse cenário.

A tecnologia ATM, usada em redes de altas velocidades, emprega comutação de pacotes de tamanho fixo de 53 bytes denominados células, multiplexagem estatística, banda flexível sob demanda, qualidade de serviço (QoS) variada dentre outras características, além de ser orientada à conexão. Isto significa que, para haver qualquer transferência de informação, deverá haver previamente a alocação dos recursos da rede necessários para garantir a qualidade de serviço negociada, de acordo com a classe de serviço em questão [15][53][65]. A multiplexagem estatística proporciona o uso otimizado dos recursos de banda da rede.

O termo multiplexagem (ou multiplexação) significa possibilidade de uso compartilhado do mesmo meio de transmissão[33]. Existem várias técnicas de multiplexagem:

- FDM (*Frequency Division Multiplexing*): utilizada em meios de transmissão analógicos, o espectro de frequências é particionado e a faixa é usada de forma compartilhada de modo permanente (determinístico);

- TDM (*Time Division Multiplexing*): utilizada em meios digitais, um período de tempo é particionado (*timeslots*) e usado de modo compartilhado, também de modo fixo (determinístico); se a disponibilidade dos “*timeslots*” for utilizada de forma estatística, temos como resultado o TDM estatístico.
- Multiplexagem Estatística: utiliza os recursos de acordo com a disponibilidade dos mesmos, não de forma rígida como no caso do FDM ou TDM. Neste caso há necessidade de um estudo estatístico prévio da ocupação desses recursos.

No ATM a multiplexagem estatística é efetuada através de rótulos identificadores das unidades de transporte de informação denominadas células. O ATM será detalhado no Capítulo 2 deste estudo.

A figura 1 compara de forma simbólica a multiplexagem estatística e determinística, destacando a economia no uso dos recursos apresentada na multiplexagem estatística [33].

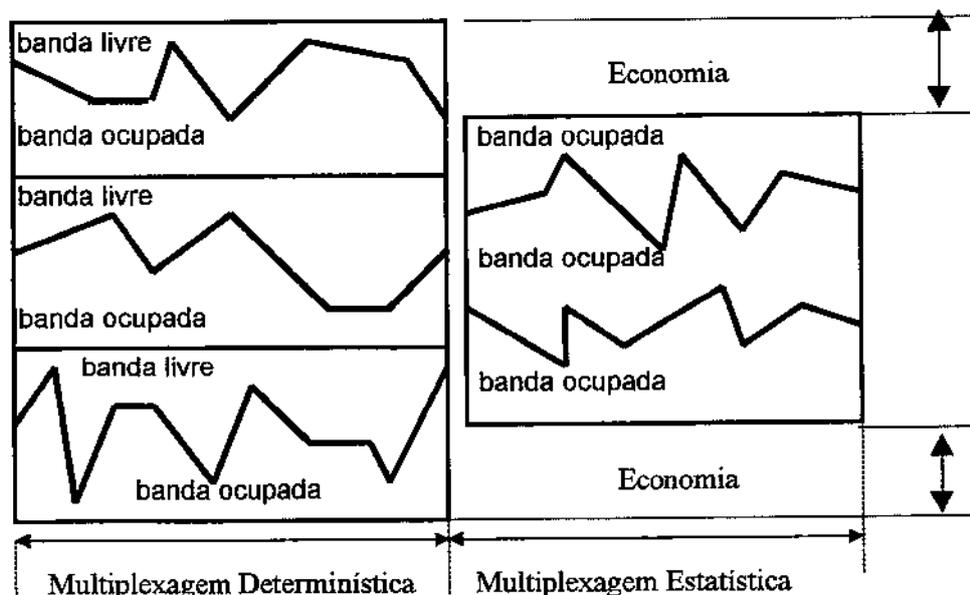


Figura 1 - ganho de multiplexagem

No ATM a administração dos recursos da rede a serem aplicados na multiplexagem estatística é uma das funções do Controle Automático de Admissão (CAC) que reserva e aloca os recursos necessários para uma determinada conexão

ATM. O CAC utiliza os seguintes parâmetros descritores de tráfego: *PCR* (*Peak Cell Rate*), *SCR* (*Sustainable Cell Rate*), *MBS* (*Maximum Burst Size*) e *CDVT* (*Cell Delay Variation Tolerance*) [10][15][53].

Além dos parâmetros de tráfego o CAC usa parâmetros de qualidade de serviço como: *ppCDV* (*peak-to-peak Cell Delay Variation*), *maxCTD* (*maximum Cell Transfer Delay*) e *CLR* (*Cell Loss Rate*), através desses parâmetros e de algoritmos de alocação dinâmica de banda o CAC tenta garantir a qualidade de serviço até o final da conexão [10][53].

Como pode ser constatado, ainda hoje existem diferentes níveis de compreensão sobre a natureza dos diferentes tipos de tráfego, mesmo quando tratados separadamente [65].

Por outro lado, o tráfego de voz foi muito bem caracterizado e explicado por A. K. Erlang no início deste século. O dimensionamento das centrais telefônicas normalmente é efetuado considerando a probabilidade de sucesso de uma chamada realizada por um assinante, no período onde o tráfego é mais intenso [72].

O tráfego de voz pode ser caracterizado por um processo de chegada de chamadas tipo Poissonianas com tempos de espera exponencial negativa, sendo que as chamadas bloqueadas são perdidas [57].

As fontes de tráfego de dados têm sido caracterizadas inicialmente como processos de Poisson ou Bernoulli, tendo uma modelagem aceitável e facilmente compreendidas através da teoria de filas. Ocorre que as características de tráfego, como taxa de chegada, comprimento de pacotes, entre outras, dependem fortemente do tipo de aplicação que está utilizando a rede de comunicação. As aplicações processadas nos sistemas de computações centralizados (*mainframe*) eram limitadas pela capacidade de processamento das máquinas e pelo próprio desenvolvimento tecnológico dos aplicativos e sistemas operacionais. Com o advento das redes locais ocorreu a distribuição da capacidade de processamento e das aplicações, bem como a possibilidade de ambientes hipermídia. Ambientes hipermídia são aqueles com as características do hipertexto, como a estrutura associativa, não linear e não hierárquica que pode ser usada em dados registrados em diferentes tipos de mídia, gerando um tráfego tipo multimídia (voz, dados, imagem, vídeo,...) [13][24]. Todos

esse fatores proporcionaram uma evolução natural das aplicações, tornando-as muito mais complexas. Essa sofisticação das aplicações está gerando uma alteração da natureza do tráfego de dados nas redes de comunicações. Um exemplo disso é a “*World-Wide-Web*” (WWW), que é a concretização de uma rede mundial de computadores interligados, tornando-se um novo paradigma para as comunicações do final deste século e entrada do novo milênio. A WWW foi inicialmente proposta em 1989, pelo CERN (*European Center For Nuclear Research*)[5], hoje expandiu-se de tal forma, que serve aos mais variados objetivos como: comércio eletrônico, informação em geral, educação, entretenimento, etc..

Vários estudos recentes têm demonstrado que os modelos de Poisson não são apropriados para modelagem do novo perfil de tráfego de dados das redes de comunicações atuais. Isso se deve a tráfego com características de rajadas e dependência de longo prazo apresentadas, que são inerentes de tráfego auto-similar ou fractal [84], conceito que será demonstrado mais adiante neste trabalho [35][36][38][44][45][58][78].

Sinais de vídeo em geral, em função de necessidade de elevadas taxas médias de transmissão, têm grande influência no dimensionamento das redes de comunicações. As imagem digitalizadas são em geral representadas por quadros (*frames*), sendo que sua qualidade é determinada pelos números de pixels e níveis de quantização [65].

A forma mais simples de transmissão digital de imagens é enviá-las digitalmente pixel por pixel, resultando na geração de vídeo com taxa constante (*CBR* - *Constant Bit Rate*), porém relativamente elevada. Uma redução dessa taxa pode ser conseguida por meio de codificação da imagem, resultando numa taxa variável (*VBR* - *Variable Bit Rate*) [34]. A codificação neste caso introduz correlações não somente de curto prazo entre células consecutivas, mas também altera estruturas estatísticas em várias escalas de tempo. Note que neste tipo de fonte não é esperado períodos de silêncio e sim grandes variações instantâneas da taxa de células [39].

Os modelos mais usados na literatura para fontes de vídeo são modelos do tipo Markov Auto-Regressivos de estado contínuo ou modelos Markovianos de tempo contínuo e estado discreto[59][64].

Vários estudos têm demonstrado que em função da utilização da multiplexagem estatística do ATM, grandes agregados de tráfego, mesmo formado por fontes não auto-similares, tendem a apresentar um comportamento fractal. O comportamento fractal é caracterizado pela dependência de longo prazo (*Joseph Effect*) e decaimento hiperbólico da função de autocorrelação (*Noah Effect*), em várias escalas de tempo [36][45]. A característica fractal ou auto-similar será apresentada no Capítulo 3.

É importante ressaltar, que estas mudanças no comportamento do tráfego têm tornado a caracterização das fontes de tráfego um fator muito importante para o dimensionamento e análise de desempenho das redes ATM.

A crescente competição no mundo atual, em função da crescente globalização da economia mundial, tem exigido cada vez mais a otimização dos recursos das empresas, incluindo recursos de rede. Aliado a esse fato, o desenvolvimento de novos recursos computacionais e de telecomunicações nos últimos anos têm possibilitado uma evolução das aplicações e das necessidades empresariais. Esses dois fatores têm mudado as características do tráfego principalmente em ambientes corporativos, fazendo com que haja cada vez mais a necessidade de se conhecer essas características, proporcionando um melhor planejamento e dimensionamento das novas redes.

Geralmente os algoritmos de alocação de banda são muito ou pouco rigorosos. A determinação da banda ótima ou realista não é uma tarefa trivial. Para se obter um processo de alocação de banda que otimize os recursos da rede com precisão, há necessidade de se ter um bom conhecimento das características do tráfego que será transportado na rede. Desta maneira o procedimento de caracterização de tráfego é considerado um dos passos essenciais no processo de alocação de banda nas redes emergentes. Assim sendo, o dimensionamento preciso dos recursos de rede para uma determinada conexão ATM tem sido um problema freqüentemente levantado, porém ainda sem uma resposta satisfatória. Vários trabalhos têm sido apresentados, baseados no conceito de banda efetiva, com a finalidade de determinar a banda necessária para um determinado tráfego. Esses estudos têm se fundamentado nas duas abordagens: com a restrição sobre a probabilidade de perda de células ou sobre o tempo de atraso

das células, porém sem levar em consideração um importante aspecto da caracterização do tráfego: a auto-similaridade, que tem forte impacto no desempenho dos comutadores.

Este trabalho tem por finalidade estudar e selecionar diferentes parâmetros que melhor caracterizem os diferentes tipos de tráfego e apresentar um método alternativo para a determinação da Banda Efetiva, considerando a característica auto-similar para cada tipo de tráfego em uma conexão ATM. Nossa proposta generaliza a idéia de Kesidis [28] baseada na probabilidade de perda de células, oriunda da Teoria dos Grandes Desvios, e nos estudos de Le Boudec [40] baseados em atraso. Para tanto utilizamos métodos analíticos, heurísticos e simulação para alcançar uma aproximação mais realista da banda efetiva de uma conexão ATM.

No capítulo 2 são apresentados resumidamente conceitos básicos sobre redes ATM. O capítulo 3 descreve modelos de tráfego difundidos na literatura mais relacionados a esta pesquisa. No capítulo 4 são analisados vários arquivos de tráfego de dados e de vídeo, efetuando-se uma caracterização dos mesmos através de uma seleção de parâmetros de tráfego. No capítulo 5 é proposto um método alternativo para alocação de banda para uma conexão ATM. Finalmente o capítulo 6 é composto pela conclusão geral deste estudo.

Capítulo 2

Redes ATM

2.1 – Introdução

O ATM (*Asynchronous Transfer Mode*) é uma tecnologia de rede para altas taxas de transmissão, baseada na comutação e multiplexação de pacotes de informação de tamanho fixo (53 bytes), denominados células [77]. A comutação nas redes ATM é efetuada no nível 2 do modelo OSI.

O modo de transferência ATM é orientado à conexão, isto quer dizer que antes da transmissão da informação a conexão deve ser estabelecida. O estabelecimento de uma conexão é efetuado atendendo uma série de parâmetros da rede, garantindo o nível de qualidade da conexão (*QoS - Quality of Service*) requerido [53][65].

O termo assíncrono não se refere a transmissão a nível físico em si, que de fato é síncrona, mas refere-se a maneira em que a banda é alocada nas conexões [54]. Outra forma de entender, é em termos de que as células podem ser transmitidas a qualquer tempo da fonte ao destino, em função da natureza da aplicação, e não em espaços de tempo periódicos (*time slot*) como no caso do STM (*Synchronous Transfer Mode - SONET/SDH*)[55].

A multiplexagem faz parte do ATM em função de existirem múltiplas conexões sobre o mesmo meio físico [27][50][77], suportando serviços orientados e não orientados à conexão.

Em função da possibilidade de garantia de Qualidade de Serviço para diferentes tipos de tráfego, a tecnologia ATM está sendo aplicada em diferentes ambientes, como em redes públicas ou privadas de longa distância (*WAN - Wide Area Network*), ou em redes locais (*LAN - Local Area Network*). Neste sentido, o ATM possui uma excelente escalabilidade, tanto em distância quanto em taxa de transmissão.

A aplicação do ATM em LANs pode ser implementada usando-se apenas o ATM ou de forma mista, utilizando-se o ATM em conjunto com as redes locais legadas (*legacy LANs*). Neste caso deve haver uma fase de transição com a finalidade de preservar os investimentos já efetuados na tecnologia de redes locais já tradicionais, como Ethernet, Token Ring e FDDI [5][25][81].

A rede ATM possui a capacidade de transmitir diversos tipos de tráfego, com banda sob demanda e qualidade de serviço diferenciada, viabilizando a redução de custos com gerência e manutenção, com relação a rede determinística, além de possibilitar o uso otimizado dos recursos de rede. Isto permite às companhias de telecomunicações oferecerem diversos tipos de novos serviços, bem como grandes organizações implementarem suas redes corporativas a um menor custo.

Quanto a interconexão das redes legadas através de redes ATM, já existem várias soluções em desenvolvimento, já padronizadas ou disponíveis no mercado como: “*Lan Emulation*”, MPOA (*Multiprotocol Over ATM*), “*IP Switching*” e MPLS (*Multiprotocol Label Switching*), dentre outras [27].

2.2 Interfaces das Redes ATM

Interfaces são dispositivos que têm a finalidade de ligar sistemas diferentes possibilitando a conexão e a interoperabilidade dos mesmos. Geralmente as interfaces devem especificar detalhes mecânicos, elétricos, funcionais e procedurais.

As interfaces obedecem padrões executando associações lógicas e funções, permitindo a troca de informações e interoperabilidade entre diferentes entidades adjacentes. As principais interfaces utilizadas nas redes ATM são a UNI e a NNI.

A UNI (*User-to-Network Interface*) possui as seguintes especificações no ATM Forum: 3.0, 3.1 e 4.0. Esta interface pode ser pública ou privada sendo utilizada para conectar um usuário à rede ATM. O usuário pode ser uma rede ATM privada, que conecta-se a rede pública através de uma UNI pública, ou uma rede local (por exemplo Ethernet) conectada a uma rede ATM privada através de uma UNI privada. O termo usuário ATM representa qualquer dispositivo que faz uso da rede

ATM, via UNI. A principal diferença entre esses dois tipos de UNI é a nível físico. Há também alguma diferença em termos de funcionalidades devido a requerimentos aplicáveis a cada uma das interfaces. Ambas compartilham a mesma camada ATM, mas podem utilizar diferentes meios físicos [5][11][14][15].

Basicamente a UNI executa as seguintes funções: conexão física entre um usuário e um comutador, multiplexagem e demultiplexagem das células, endereçamento entre os usuários, sinalização entre o usuário e o comutador.

A NNI (*Network Node Interface* ou *Network-Network Interface*) é utilizada para conectar nós (comutadores) ATM, podendo também ser pública ou privada. A NNI que conecta comutadores dentro de uma rede ATM privada, chama-se P-NNI [66].

A Figura 2 mostra as diferentes interfaces na rede ATM salientando diferentes opções de utilização das interfaces UNI e NNI.

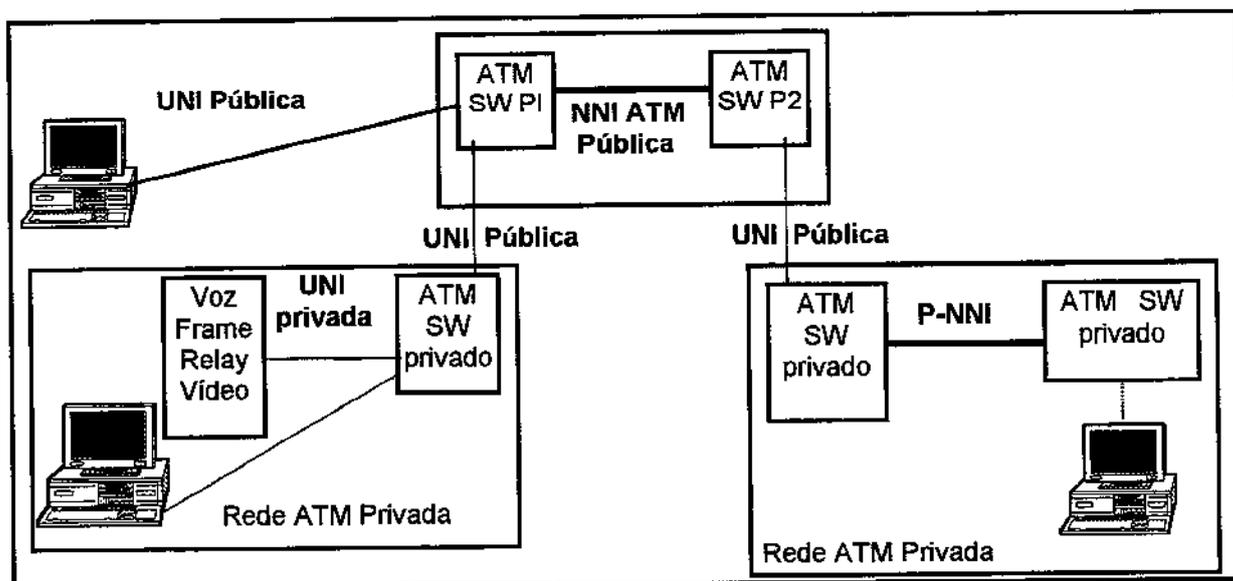


Figura 2 - Rede ATM e suas principais interfaces

Basicamente a NNI executa funções como: conexão física entre comutadores, sinalização entre comutadores, roteamento. É importante observar que em termos de sinalização, enquanto na UNI é assimétrica, na NNI é simétrica. Para maiores detalhes sobre os protocolos de sinalização utilizados na UNI e NNI recomendamos consultar as referências bibliográficas [11][66].

2.3 Formato da Célula ATM

O tamanho de célula escolhido, sendo fixo, permite a minimização da complexidade da comutação e gerenciamento dos “*buffers*”, proporcionando menor tempo de processamento geral da rede.

A figura 3 mostra os dois formatos concebidos para a célula ATM [15][50][53]:

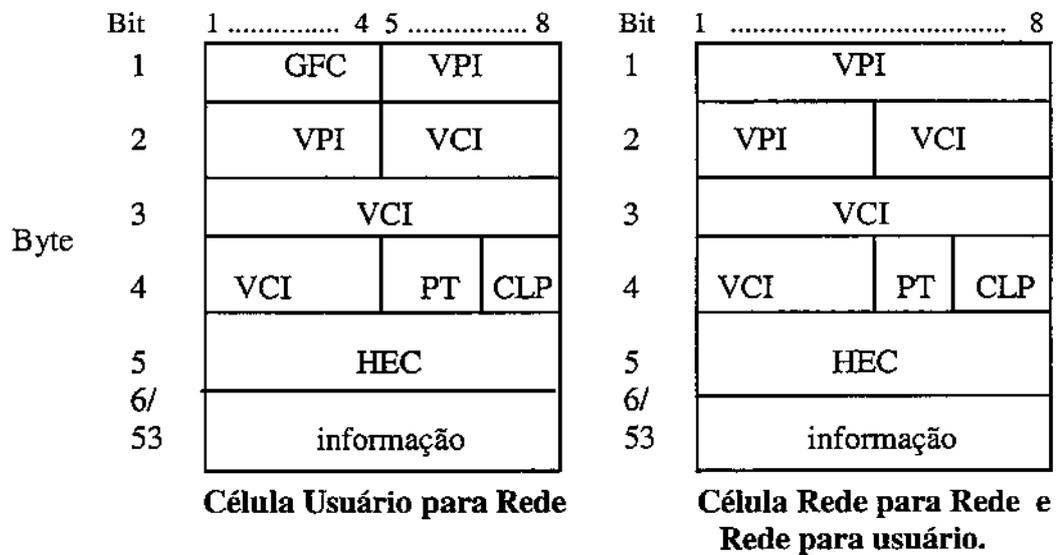


Figura 3 - formato das células

A comunicação nas redes ATM é efetuada por meio de conexões, sendo estabelecidos Canais Virtuais (VC – *Virtual Channel*) para a comunicação fim-a-fim. Somente após toda a conexão estar estabelecida é que se inicia a transmissão da informação. O Canal Virtual (VC), está inserido, no Caminho Virtual (VP – *Virtual Path*), sendo ambos níveis da hierarquia de roteamento ATM.

Conforme mostrado na figura 3, os cinco bytes do cabeçalho são compostos pelos seguintes campos [11][53][54][55][65]:

- O Controle Genérico de Fluxo, GFC (*Generic Flow Control*): é um campo composto dos primeiros quatro bits e são usados para controle de fluxo no acesso do usuário à rede na UNI (*User-Network Interface*). Não há GFC em células no sentido rede-usuário, ou na NNI (*Network-to-Network Interface*). O valor

codificado no campo GFC têm somente significado local no acesso à rede, o mesmo não é transportado através da rede, mas reescrito nos comutadores ATM, passando a fazer parte do VPI.

- VPI e VCI: são o segundo e terceiro campos do cabeçalho que são compostos dos identificadores associados como o conceito de Canal e Caminho Virtuais, respectivamente. Conforme definição da Recomendação I.113 da ITU-T, temos:

- VC (*Virtual Channel*): conceito usado para descrever o transporte unidirecional de células ATM, associado a um identificador de valor único, chamado VCI (*Virtual Channel Identifier*) - 16 bits.
- VP (*Virtual Path*): conceito usado para descrever o transporte unidirecional de células que pertencem a VC's, associado a um único identificador, chamado VPI (*Virtual Path Identifier*) 8 bits na UNI e 12 bits na NNI.

Na verdade VC e VP são níveis de abstração para representar as células multiplexadas em uma determinada conexão, considerando que as células são na realidade um trem de bits, que devem ser identificados como células únicas, possibilitando o roteamento das mesmas através da trajetória da conexão.

- PT (*Payload Type*): é o quarto campo com apenas três bits, é usado para caracterizar o tipo de carga que está sendo transportada, por exemplo se dados de informação ou dados de gerenciamento.
- CLP (*Cell Loss Priority*): é um campo com apenas um bit que possibilita a marcação das células de acordo com a prioridade de descarte. Células marcadas com baixa prioridade (CLP=1) podem ser descartadas pela rede, em caso de congestionamento. A célula marcada com alta prioridade (CLP=0) terá preferência, não sendo descartada.
- HEC (*Head Error Check*): é o último campo do cabeçalho da célula ATM é composto por oito bits. O HEC é utilizado para verificação e correção de um único bit do cabeçalho, verificação de mais de um bit corrompido para

descarte da célula afetada e ainda delimitação da célula (*cell delineation*), identificando o começo e o fim de uma célula no trem de bits. Após o HEC, os demais 48 bytes correspondem a carga de informação (*payload*).

2.4 Conexões na Rede ATM

As conexões na rede ATM podem ser de dois tipos: Permanente ou Comutada. A conexão permanente é chamada de PVC (*Permanent Virtual Connection*) que fica ativa por tempo indeterminado, sendo configurada manualmente através da estação de gerência do sistema. A conexão comutada conhecida como SVC (*Switched Virtual Connection*) é estabelecida automaticamente, sendo requisitada pelo usuário ATM através da UNI, ao contrário da primeira permanece ativa somente durante a utilização do serviço. Os SVC's são iniciados diretamente através dos protocolos das camadas superiores, usando os protocolos de sinalização através da UNI. Os usuários dos serviços de voz, dados ou vídeo, por exemplo, fazem requisições de conexão para um determinado endereço ATM, para uma certa categoria de serviço, para atender um nível de QoS, sendo que a conexão pode ser aceita ou rejeitada [54].

A figura 4 representa os níveis de abstração de canal virtual e de caminho virtual, nota-se que o VC está inserido no VP.

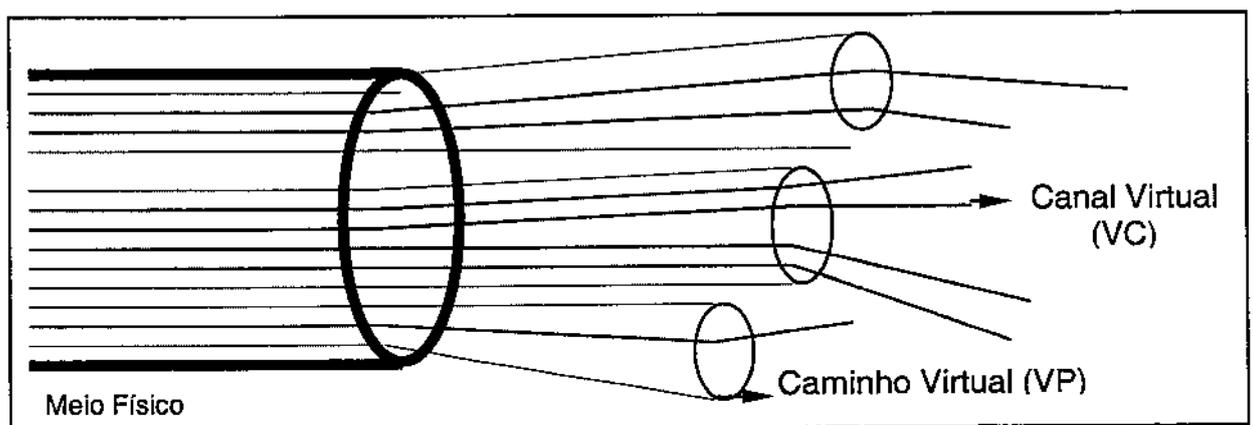


Figura 4 - conexões ATM

O tempo necessário para o estabelecimento de um SVC depende dos seguintes fatores: número de comutadores no percurso da conexão, arquitetura do comutador, as distâncias envolvidas, podendo variar de alguns milissegundos até a um segundo.

As conexões também podem ser classificadas quanto ao número de pontos. Temos a conexão ponto-a-ponto, onde a troca de informações é feita entre apenas dois usuários. Na conexão ponto-multiponto, a troca de informação ocorre entre um usuário chamado raiz e outros múltiplos pontos chamados folha. Este último tipo é também chamado “*multicast*”, pois cópias da informação enviada pela raiz são recebidas pelas folhas [27].

A conexões de Canal Virtual e Caminho Virtual são conexões lógicas a nível da camada ATM. A técnica de VP, considerando que em um VP pode conter vários VC's, proporciona uma otimização dos recursos de controle das conexões e gerência da rede.

A conexão de Canal Virtual VCC (*Virtual Channel Connection*) é uma concatenação de um ou mais VC (*Virtual Channel*), estabelecendo uma comunicação fim-a-fim. Da mesma forma a conexão de Caminho Virtual VPC (*Virtual Path Connection*) é a concatenação de um ou mais VP (*Virtual Path*) [14][47].

O Enlace de Canal Virtual VCL (*Virtual Channel Link*) é usado para descrever o transporte unidirecional de células de um VC. Um VCL inicia onde um VCI é designado até onde ele é trocado ou terminado. Equivalentemente, o Enlace de Caminho Virtual VPL (*Virtual Path Link*) é usado para descrever o transporte de células de um VP, inicia onde um VPI é designado até onde ele é trocado ou terminado.

A célula ATM é transmitida ao longo de um VC de acordo com o rótulo de roteamento (VPI e VCI) contido em seu cabeçalho. O rótulo é trocado (*label swapping*) a cada comutador ao longo do caminho da conexão, que contém uma tabela de mapeamento dos rótulos. As tabelas são configuradas manualmente ou através da sinalização [66].

A colocação dos VPI/VCIs são definidas através da rede durante a fase de estabelecimento da conexão. Cada conjunto VPI/VCI é associado com uma porta

particular do comutador. A comutação pode ocorrer a nível de canal ou a nível de caminho [54].

Uma conexão em uma rede ATM, consiste de uma série de enlaces (*links*) que guiam as células ao longo de um caminho através da rede, baseando-se no valor de VPI e VCI contidos no cabeçalho de cada célula. A relação entre VPI e VCI é hierárquica, ou seja a multiplexagem e comutação é sempre feita primeiramente a nível de VPI e depois a nível de VCI.

O VP é uma conexão lógica semi-permanente, ponto-a-ponto, feita através do uso das funções de controle e gerenciamento da rede. Um VP pode acomodar até 2^{16} VC's (65.536). Este procedimento hierárquico entre VP e VC, simplifica muito as tabelas de roteamento, funções de gerenciamento e controle, seleção do caminho, alocação de banda e o procedimento de estabelecimento da conexão. Logo quando temos uma grande quantidade de conexões entre dois ponto A e B, todos os VC's podem ser acomodados em um único VP, simplificando a tarefa de comutação em cada "switch" do percurso [65].

O VCI é um valor alocado dinamicamente no momento do estabelecimento da conexão (*call setup*) através do canal de sinalização, tendo apenas significado local em cada comutador. Quando a conexão é desfeita, o VCI em um determinado comutador pode ser utilizado em outra conexão.

Geralmente os valores de VCI são únicos para um determinado valor de VPI, e valores de VPI são únicos em relação a um determinado meio físico.

Chamamos Identificador de Conexão (CI - *Connection Identifier*), que é determinado pela tripla VPI, VCI e identificador de enlace físico.

2.5 As Camadas do Protocolo ATM

O protocolo ATM está baseado no modelo de referência para a RDSI-FL, que é composto de plano de usuário, que transporta informações do usuário; plano de controle, responsável por funções como controle de chamadas, conexões e

informações de sinalização; e finalmente o plano de gerenciamento que executa funções de gerência da rede [77][80].

São três camadas no plano de usuário: Camada Física, Camada ATM e Camada de Adaptação. A Camada Física é a mais inferior, seguida da Camada ATM, e acima a Camada de Adaptação. A Camada Física é subdividida em Convergência de Transmissão (TC - *Transmission Convergence*) e Meio Físico (PM- *Physical Medium*). A subcamada TC extrai e insere as células ATM no meio de transmissão. Acima da subcamada TC, está situada a camada ATM que faz multiplexação, comutação e controle. A camada mais alta, é a Camada de Adaptação, sendo subdividida em Segmentação e Remontagem SAR (*Segmentation And Reassembly*) e Subcamada de Convergência CS (*Convergence Sublayer*) [65].

A camada física provê para a camada ATM acesso ao meio físico de transmissão. A Subcamada de Convergência de Transmissão trata com a camada física independente dos aspectos do meio físico em si, a maior parte de suas funções estão envolvidas com a geração e processamento de algum “overhead”. A subcamada PM é dependente do meio físico selecionado, provê informação de tempo para a transmissão dos bits e ainda é responsável pela codificação e decodificação. As interfaces possíveis na camada física podem ser: E-1, DS-3, SDH, SONET, entre outras [55].

Na subcamada de Convergência de Transmissão TC é efetuada a geração e verificação do HEC. Todo o cabeçalho (incluindo o HEC byte) é protegido pela verificação de erro do HEC, corrigindo um único bit errado e detectando múltiplos bits errados. A TC é também responsável pelo embaralhamento e desembaralhamento das células (*Cell Scrambling/Descrambling*), que permite randomização da carga para evitar padrões de bit não variáveis, para melhorar a eficiência do algoritmo de delineação. O delineamento das células (*Cell Delineation*), efetuado também pela TC, permite a identificação do início e final de uma célula. Ainda na TC é realizado o desacoplamento (*Cell Rate Decoupling*) que desacopla ou altera a taxa de células através da inserção e retirada de células vazias.

A Camada ATM é responsável pelo roteamento e comutação da célula ao longo do caminho virtual, realiza multiplexagem/demultiplexagem de células, controle de erros do cabeçalho e translação dos identificadores VPI e/ou VCI (*label swapping*).

A camada ATM é responsável pela colocação e retirada do cabeçalho, exceto o byte que representa o HEC que é responsabilidade da camada física [15][50].

Na estação de origem a camada ATM recebe 48 bytes dos usuários (através da camada de adaptação), adiciona quatro bytes correspondentes a parte do cabeçalho (faltando o HEC que será adicionado na camada física), entregando a célula na camada física para o cálculo do HEC e transmissão. Na estação de destino, a camada ATM recebe as células da camada física, remove o cabeçalho e passa a carga para a camada de adaptação.

Outra importante função da camada ATM é o controle de tráfego e congestionamento, incluindo: controle de admissão (ações feitas na fase de “*setup*” para verificar se a conexão deve ser aceita ou rejeitada); controle de prioridade (CLP) e descarte de células.

Como vimos a camada ATM trata somente com funções relacionadas com o cabeçalho, não se preocupando com o conteúdo da carga. Isto é necessário para que seja possível alcançar alto desempenho a nível de comutação.

A última camada do protocolo ATM é a camada de adaptação AAL (*ATM Adaptation Layer*) que é dependente da classe do tráfego, ou seja, se o tráfego é dependente ou não da informação de tempo, se é orientado ou não à conexão, e ainda se a taxa de transmissão é constante ou variável. A AAL isola a camada ATM das camadas superiores usuárias do protocolo, pelo mapeamento dos protocolos das camadas superiores (PDU - *Protocol Data Unit*) em células [14].

As funções da AAL são agrupadas em duas subcamadas: Convergência (CS) e Segmentação e Remontagem (SAR). A primeira trabalha com funções específicas e comuns de serviços enquanto a última com a segmentação e remontagem dos dados mapeando-os dentro da carga da célula.

Hoje existem as seguintes camadas de adaptação padronizadas:

- AAL 1 - fornece os seguintes serviços: transferência de SDUs com CBR para o destino com usuários em mesma taxa de transmissão; transferência de informações de “clock” entre fonte e destino; transferência de estrutura de informação de fonte e destino; indicação de perda de informação ou informação errada.
- AAL 2 - é utilizada para prover serviços de dados com tráfego de taxa variável. Vídeo e áudio compactado ou comprimido são os exemplos. Os serviços na AAL2 são: transferência de unidade de dados com taxa de bits variável; transferência de informação de tempo entre fonte e destino e indicação de perda de informação ou erro na informação.
- AAL 3/4 - fornece suporte para serviços de transmissão de dados com conexão e sem conexão. Na AAL 3/4 cada bloco de dados que chega a SAR é tratado independentemente, possibilitando múltiplas conexões lógicas SAR utilizar uma única conexão ATM.
- AAL5 - como a AAL 3/4 foi concebida para transmissão de dados, mas AAL 5 foi elaborada para operar de forma mais simples que AAL3/4 reduzindo o “overhead” de processamento de protocolo e transmissão, assegurando adaptabilidade aos protocolos de transporte existentes.

2.6 Parâmetros de Tráfego e Qualidade de Serviço em Redes ATM:

Os parâmetros de tráfego e de qualidade de serviço são utilizados para verificar se uma conexão será ou não aceita, dentro de determinados requisitos. Alguns dos parâmetros de QoS serão negociados entre os usuários da rede e o fornecedor do serviço, já os parâmetros de tráfego não podem ser negociados [10][15].

2.6.1 Parâmetros de Tráfego - o ATM Forum elegeu os seguintes parâmetros como descritores de tráfego quanto a taxa de transmissão e tamanho dos surtos[10][15][53]:

- *PCR (Peak Cell Rate)*: é o valor máximo da taxa de células que pode ser introduzida na rede em uma conexão virtual. O *PCR* é modelado pela taxa de dreno do “*buffer*” (algoritmo balde furado). O *PCR* de uma conexão pode ser visto como o inverso do tempo mínimo entre duas submissões de células na rede.
- *SCR (Sustainable Cell Rate)*: é uma taxa média de células que podem ser introduzida na rede. Este parâmetro é o valor esperado ou requerido da taxa de células ao longo de um intervalo de tempo. A razão *PCR/SCR* é uma forma de se medir o grau de explosividade do tráfego conhecido como coeficiente de rajada do tráfego [5].
- *MBS (Maximum Burst Size)*: é o número máximo de células que podem ser enviadas continuamente na taxa de pico. Ou seja, é o número máximo de células que podem passar transparentemente através do GCRA (*Generic Cell Rate Algorithm*) na taxa de pico. Se as células chegam em grupos igual ao *MBS*, então o intervalo ocioso entre os grupos deve ser suficiente para que a taxa total não exceda o *SCR* [10][15].
- *BT (Burst Tolerance)*: é o período de duração em que a fonte pode submeter tráfego na taxa de pico. O *SCR* é modelado como a taxa de dreno do “*buffer*”.

A relação entre o *BT* e o *MBS* é dada pela equação (1) [15]:

$$BT = (MBS - 1) \left(\frac{1}{SCR} - \frac{1}{PCR} \right) \quad (1)$$

Em termos do algoritmo do balde furado (*Leaky Bucket Algorithm*), que será apresentado na próxima sessão, o parâmetro *CDV (Cell Delay Variation)* define a profundidade do balde para *PCR*. No caso de *SCR* a profundidade do balde é definida pelo parâmetro *BT* (que é proporcional ao *MBS*) [5][10].

A figura 5, mostra a disposição das células em um tráfego tipo CBR, e em um tráfego tipo VBR. Neste último podemos notar a relação do *MBS* com os tempos *T_i* e *T_b*, que representam o intervalo de tempo mínimo entre rajadas e o tempo mínimo na rajada em *PCR*, respectivamente [15].

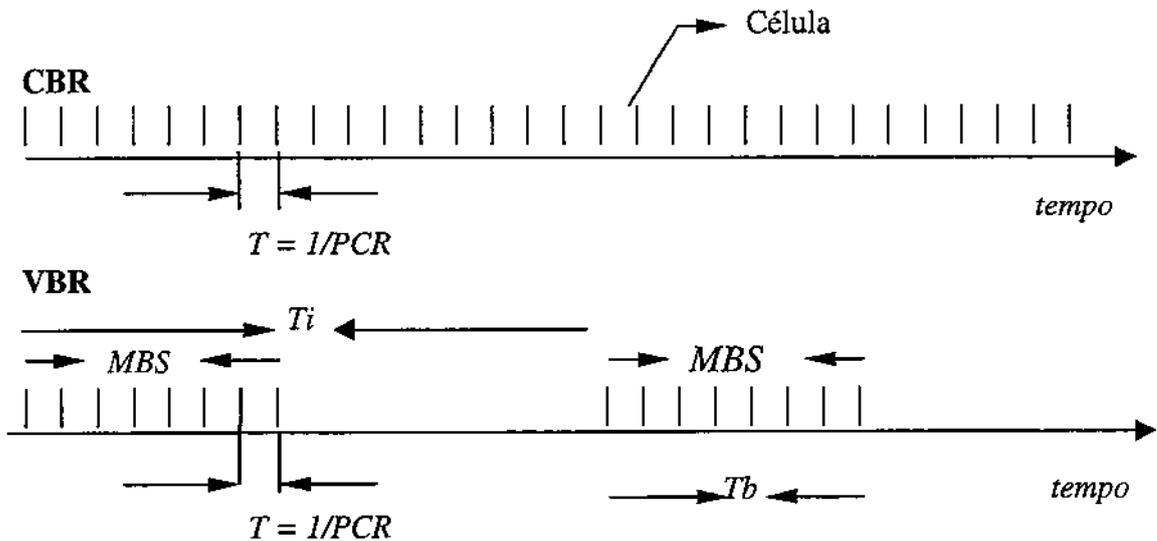


Figura 5: Relação SCR e MBS

Sendo T_i (*Minimum Burst Interarrival Time*) e T_b (*Maximum Burst Duration at PCR*) calculados através das equações (2) e (3):

$$SCR = \frac{MBS}{T_i} \quad (2)$$

$$T_b = (MBS-1)T \leq T_i - T \quad (3)$$

- *MCR (Minimum Cell Rate)*: é a taxa mínima de células que a rede deve garantir para uma conexão, é o número mínimo de células/segundo considerado aceitável pelo usuário da rede.

Quanto a atraso temos o seguinte parâmetro de tráfego:

- *CDVT (Cell Delay Variation Tolerance)*: é o valor máximo tolerado para *CDV (Cell Delay Variation)* que mede quanto de variação estará presente na transmissão das células, este parâmetro será definido no próximo item.

2.6.2 Parâmetros de Qualidade de Serviço - o ATM Forum elegeu os seguintes parâmetros para Qualidade de Serviço [5][10][15][53][65]:

- *CLR (Cell Loss Ratio)*: é a relação entre o número de células perdidas pelo número total de células enviadas. Mede a fração de células transmitidas não usadas, seja por perda ou por se tornarem não utilizáveis em função do retardo.
- *CTD (Cell Transfer Delay)*: é o retardo de transferência de célula, calculado pelo somatório dos tempos devido aos seguintes processamentos: Codificação, Empacotamento, Propagação, Transmissão, Comutação, Filas e Remontagem.
 - retardo de codificação depende do algoritmo de codificação empregado, que é dado pelo tempo gasto para a conversão do sinal analógico em digital.
 - tempo de Empacotamento é o tempo necessário para encher a carga da célula com o sinal digitalizado, que varia com o tipo de AAL a ser utilizada.

Como exemplos de retardo de empacotamento temos: voz na AAL 1 aproximadamente 6 mseg e vídeo MPEG II (MPEG – *Motion Picture Expert Group*, mais detalhes sobre este assunto no Apêndice C) na AAL2 aproximadamente 35 useg [51][65].
 - retardo de propagação é o tempo necessário para o sinal elétrico ou luminoso se propagar pelo meio de transmissão (rádio, fibra ótica ou fio de cobre) da origem ao destino. É calculado pela distância dividida pela velocidade da luz no meio em questão.
 - tempo de transmissão é introduzido pelo canal de transmissão, cujo valor depende da taxa de transmissão do canal de comunicação.
 - retardo de comutação é o tempo necessário para a célula ATM atravessar a matriz de comutação. Esta medida depende fortemente do desempenho do comutador considerado. Em outras palavras, esta medida varia com a arquitetura, algoritmos de roteamento, e políticas de acesso.
 - tempo de fila são gastos nas filas dos dispositivos da rede, incluindo filas nos diversos “*buffers*” internos aos comutadores.
 - retardo de Remontagem (*reassembly*) é o tempo gasto para a retirada da carga da célula e a remontagem dos pacotes da aplicação. Um exemplo desse atraso é o tempo para a camada AAL5 remontar um pacote IP.

O problema é que o retardo de transferência (*CTD*) não é constante, variando com a carga instantânea da rede, causando o *Jitter*.

- *CDV (Cell Delay Variation - Jitter)*: é a variação do retardo fim-a-fim , pode ser definido como a variância do retardo na conexão, mede o grau de uniformidade em que as células estão sendo liberadas, isto é:

$$E\{[CTD - E(CTD)]^2\} \quad (4)$$

A figura 6 mostra a função densidade de probabilidade do retardo de transferência de células [10].

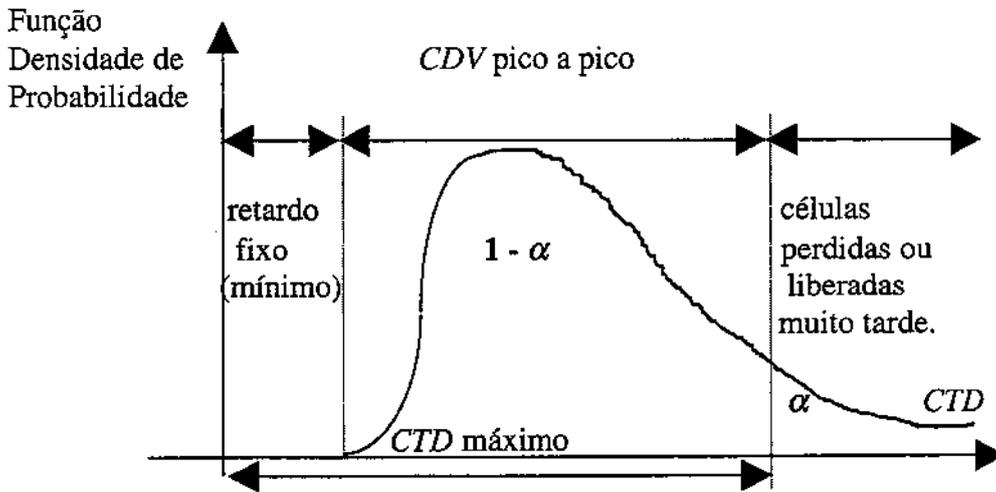


figura 6: Modelo de densidade de probabilidade do CTD

- *CER (Cell Error Ratio)*: é a razão entre o número de células erradas e o número total células enviadas, é a fração de células liberadas com um ou mais bits errados no cabeçalho.
- *CMR (Cell Misinsertion Ratio)*: é a razão entre o número de células inseridas erroneamente pelo intervalo de tempo, significa a taxa de células que são liberadas para destino errado, considerando erros não detectados no cabeçalho.

Existem ainda medidas oriundas da transmissão digital, que podem ser incluídas como parâmetros de qualidade de serviço [53]:

- *BER (Bit Error Ratio)*: é a razão do número de bits errados pelo número total de bits transmitidos. Como os erros podem ocorrer isoladamente ou em rajadas, essa medida deve ser feita em um período de tempo adequado. Sem levar em consideração a confiabilidade, para se medir um $BER = 10^{-4}$, devem ser medidos

pelo menos 10.000 bits. Porém para se ter confiança nos resultados, para o mesmo BER deve-se testar um número de de 1 a 10 milhões de bits [53].

- **SECBR (Severely Errored Cell Block Ratio):** é a razão de blocos de bits errados pelo número de blocos de bits.

A Tabela 1 apresenta sinteticamente os parâmetros de tráfego e QoS do ATM Forum e ITU-T [37].

ATM Forum

CBR	rt-VBR	nrt-VBR	ABR	UBR	tipo
<i>PCR</i> <i>CDVT</i>	<i>PCR</i> <i>CDVT</i>	<i>PCR</i> <i>CDVT</i>	<i>PCR</i> <i>CDVT</i>	<i>PCR</i> <i>CDVT</i>	tráfego
	<i>SCR</i> <i>MBS</i>	<i>SCR</i> <i>MBS</i>			tráfego
			<i>MCR</i>		tráfego
<i>ppCDV</i>	<i>ppCDV</i>				QoS
<i>maxCTD</i>	<i>maxCTD</i>				QoS
<i>CLR</i>	<i>CLR</i>	<i>CLR</i>			QoS

ITU-T

DBR	ABT	SBR	ABR	tipo
<i>PCR</i> <i>CDVT</i>	<i>PCR</i> <i>CDVT</i>	<i>PCR</i> <i>CDVT</i>	<i>PCR</i> <i>CDVT</i>	tráfego
	<i>SCR, IBT e</i> <i>CDVT</i>	<i>SCR, IBT e</i> <i>CDVT</i>		
			<i>MCR</i>	tráfego
<i>CDV</i>	<i>CDV</i>			QoS
<i>CTD</i>				QoS
<i>CLR</i>	<i>CLR</i>	<i>CLR</i>		QoS

Tabela 1: Parâmetros ATM Forum e ITU-T

PCR - Peak Cell Rate
CDVT-Cell Delay Variation Tolerance
CDV - Cell Delay Variation
SCR - Sustainable Cell Rate
maxCTD- Maximum Cell Transfer Delay
ABR- Available Bit Rate
nrt-VBR- non-real-time Variable Bit Rate
DBR- Deterministic Bit Rate

MCR - Minimum Cell Rate
ppCDV-peak-to-peak Cell DelayVariation
CLR - Cell Loss Ratio
MBS- Maximum Burst Size
UBR- Unspecified Bit Rate
ABT- ATM Block Transfer
rt-VBR- real-time Variable Bit Rate
SBR- Statistical Bit Rate

Quanto a negociação dos parâmetros de QoS, temos os seguinte:

Os seguintes parâmetros de QoS são negociados: *ppCDV*, *maxCTD* e *CLR*.

Os seguintes parâmetros de QoS não são negociados: *CER*, *SECBR*, e *CMR*.

A recomendação I.371 do ITU-T especifica somente a taxa de pico do tráfego *PCR*, o ATM Forum adicionou em sua especificação a taxa *SCR* e o *MBS*.

Parâmetros Descritores de Tráfego e de QoS do ATM Forum e ITU-T.

2.7 Policiamento de Tráfego em ATM

Policiamento é o monitoramento do tráfego, efetuado pelo provedor do serviço de rede, com a finalidade de verificar se o usuário está respeitando os níveis de serviço contratados, tal como taxa de transmissão utilizada. Para tanto o provedor de rede faz uso de parâmetros como *UPC (Usage Parameter Control)* e *NPC (Network Parameter Control)* para monitorar e regular o tráfego que está entrando na rede.

O *UPC* é usado para controlar o tráfego na *UNI*, e o *NPC* para controlar o tráfego que passa pela *NNI*. O objetivo é proteger a rede de desvios dos parâmetros negociados, intencionais ou não, que podem afetar adversamente a QoS de outras conexões [5][10][15].

Os mecanismos de *UPC/NPC* devem contemplar: simplicidade na implementação, ser de fácil entendimento aos usuários da rede, permitir uma margem de tolerância, responder rapidamente a violações, aparecer transparente para as fontes que estão conformes. É importante ressaltar que é de responsabilidade do usuário estar em conformidade com os parâmetros negociados no contrato de tráfego.

O *GCRA (Generic Cell Rate Algorithm)* é um algoritmo de policiamento do tipo “*Virtual Scheduling Algorithm*” ou do tipo “*Continuous-state Leaky Bucket*”.

O *GCRA* é usado para definir de maneira operacional a relação entre *PCR (Peak Cell Rate)* e *CDV (Cell Delay Variation)* e a relação entre *SCR (Sustainable Cell Rate)* e *BT (Burst Tolerance)* respectivamente. Ou seja, é usado para verificar os valores de *CDV* e *BT* com relação aos parâmetros de tráfego *PCR* e *SCR* na *UNI* pública ou privada. Para cada chegada de célula o *GCRA* determina se a célula está ou não em conformidade com o contrato de tráfego [10].

O *GCRA* depende somente de dois parâmetros: o incremento *I* e o limite *L*, usando a notação *GCRA(I, L)*.

O algoritmo de agendamento virtual (*Virtual Scheduling Algorithm*) atualiza um tempo de chegada teórico TAT (*Theoretical Arrival Time*) que é o tempo nominal de chegada de células assumindo-se que as mesmas estejam igualmente espaçadas quando a fonte está ativa. Se o tempo real de chegada de uma célula não é “muito” menor que o TAT, em particular se este é maior que a diferença entre TAT e L (TAT-L), então a célula está conforme, caso contrário é não conforme.

Dada uma seqüência de chegada de células em determinados instantes de tempo tal que $\{t_a(k), k \geq 1\}$: a chegada da primeira célula $t_a(1)$, o TAT é inicializado com o tempo corrente $t_a(1)$. Para as células subsequentes, teremos que se na k -ésima célula $t_a(k)$ é maior do que o TAT, a célula é considerada conforme e o TAT é atualizado com o tempo corrente $t_a(k)$ adicionado de um incremento I.

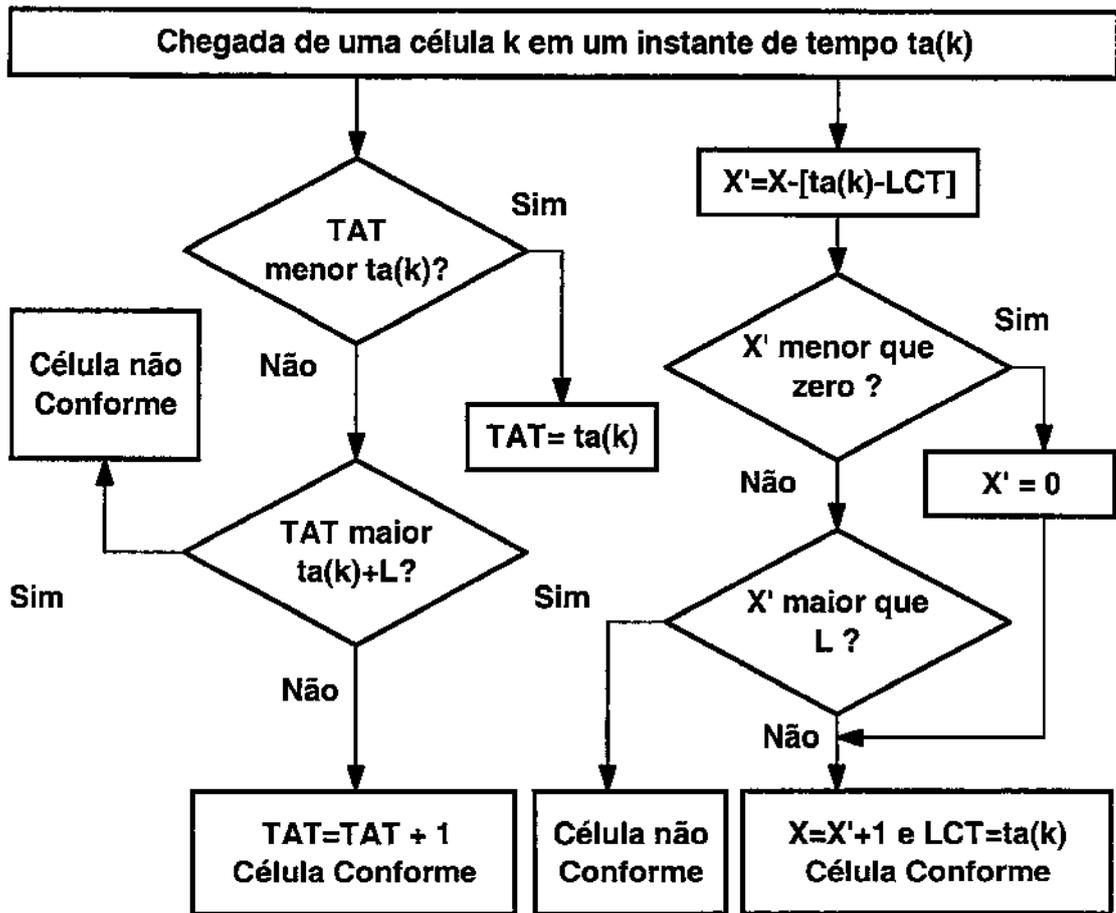
Se na chegada da k -ésima célula $t_a(k)$ é maior ou igual ao TAT - L [$t_a(k)+L$], mas menor que TAT, novamente a célula é considerada conforme e o TAT é incrementado de I. No último caso, se na k -ésima célula, TAT é maior que $t_a(k)+L$, a célula é considerada não conforme e o TAT não é alterado [5][10].

O algoritmo do balde furado de estado contínuo (*Continuous-state Leaky Bucket*) pode ser visto como uma fila (*buffer*) de capacidade finita que drena o conteúdo a uma taxa contínua, ou um contador de valores reais. Se no instante de tempo de chegada de uma célula o conteúdo é menor ou igual ao limite L, a célula é conforme, caso contrário não conforme. A capacidade do “*buffer*” (balde) é L+I.

A primeira célula que chega em $t_a(1)$, o conteúdo X do “*buffer*” é “zerado” e o último tempo de conformidade LCT (*Last Compliance Time*) é colocado em $t_a(1)$. Na chegada da k -ésima célula, em $t_a(k)$, provisoriamente o conteúdo do “*bucket*” é colocado no valor X', que é igual ao conteúdo do “*bucket*” X, depois da chegada da última célula conforme menos a quantia que foi drenada do “*bucket*”.

Se X' é menor ou igual ao limite L, e o conteúdo do balde X, é colocado em X mais o incremento, e o LCT é colocado no tempo corrente $t_a(k)$. Mas se X' é maior que o valor limite L, a célula é não conforme e o valor LCT não muda.

A figura 7 apresenta o Algoritmo GCRA em diagrama de blocos, nas duas implementações propostas pelo ATM Forum [10].



TAT - theoretical arrival time
 $t_a(k)$ - tempo de chegada de célula
 I - incremento L - limite
 Na chegada da primeira célula da conexão TAT = $t_a(k)$.
Virtual Scheduling Algorithm

X - valor do contador *leaky bucket*
 X' - variável auxiliar
 LCT - last compliance time
 Na chegada da primeira célula da conexão, X = 0 e o CLT = $t_a(k)$.
Continuous-state Leaky Bucket Algorithm

Figura 7 – Algoritmo GCRA [10]

Como exemplo, suponha um sistema de policiamento que implementa o GCRA baseando-se no “*Virtual Scheduling Algorithm*”, com os seguintes parâmetros: Para uma PCR de 400 cps, suponha que são geradas 500 células sem interrupção passando pelo GCRA(I, L), onde $I = 1/PCR$ e $L = CDVT = 1$ ms. Aplicando o algoritmo apresentado na figura 7, obtemos [37]:

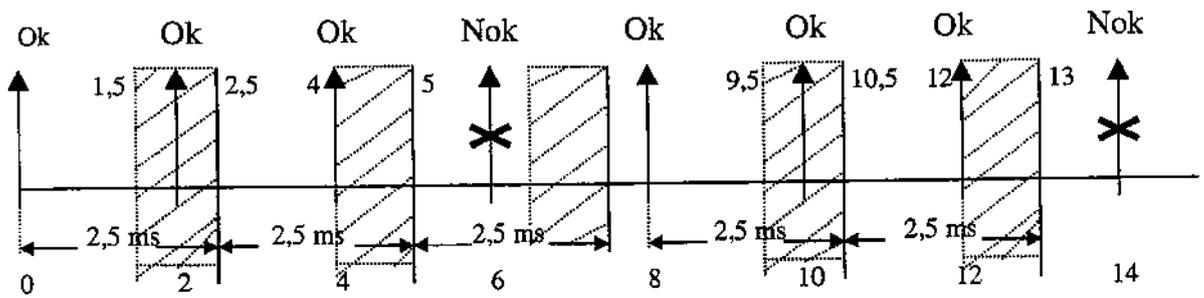


Figura 8: exemplo de GCRA

Na figura 8 observa-se que o tempo teórico de chegada esperado, TAT é representado pelas linhas cheias verticais dado por $1/PCR = 1/400 = 2,5$ ms. O tempo real de chegada das células $t_a(k)$ é representado pelas flechas verticais dados por $1/500 = 2$ ms, tendo como limite de tolerância 1 ms. Observa-se que de cada conjunto de quatro células, apenas três são aceitas, então para um surto de 500 células 375 são aceitas.

O Controle Automático de Admissão (CAC - *Connection Admission Control*) é o conjunto de ações tomadas pela rede na fase de "SETUP", para verificar se uma solicitação de VCC ou VPC pode ou não ser estabelecida. A conexão será aceita se os recursos da rede são suficientes para estabelecer e manter a conexão de acordo com os requisitos de QoS negociados [55].

Para cada requisição de conexão a função CAC deve ser capaz de obter as seguintes informações do contrato de tráfego: valor dos parâmetros dos descritores de tráfego na fonte; QoS requerida; valor do CDVT; definição de conformidade (para a UNI pública).

O CAC fazendo uso das informações obtidas e dos operadores de rede, determina: se a conexão pode ser aceita ou não; parâmetros de tráfego necessários para o controle; roteamento e alocação de recursos da rede.

2.8 Conclusões

Neste capítulo apresentamos resumidamente as principais características das redes ATM, onde salientamos as interfaces mais utilizadas nessas redes, formato das células, o conceito de conexões e seus tipos, as camadas do protocolo e tópicos sobre alguns parâmetros de tráfego e qualidade de serviço, bem como o policiamento de usuários da rede.

As vantagens implementadas nas redes ATM como: multiplexagem estatística, comutação de células a nível de enlace, banda sob demanda, tratamento de tráfego diferenciado de acordo com tipo de tráfego em função dos parâmetros descritores de tráfego, parâmetros de Qualidade de Serviço negociados e facilidade de policiamento em função da implementação de algoritmos como o CGRA, além da excelente escalabilidade, fazem da tecnologia ATM uma forte alternativa para a concretização da Rede Digital de Serviços Integrados Faixa Larga (*B-ISDN*) em ambientes públicos e privados. A disponibilidade de várias soluções de interconexão com outras redes não orientadas à conexão como por exemplo IP, utilizando MPOA e MPLS dentre outras soluções, facilitará a implementação e aumentará o desempenho da Internet e Intranets.

A escalabilidade do ATM e a disponibilidade de soluções a nível de enlace como “*Lan Emulation*”, fazem da tecnologia ATM a terceira geração de redes locais [27][81][83], proporcionando alta vazão com qualidade de serviço garantida para aplicações multimídia, suportando várias classes de serviço com facilidade de interconexão de ambientes LAN e WAN.

Capítulo 3

Modelos de Tráfego

3.1 Introdução

Neste capítulo apresentamos uma breve revisão de alguns modelos de tráfego mais difundidos na literatura, incluindo aqueles efetivamente utilizados neste trabalho. Esses modelos serão considerados em diferentes escalas de tempo e por tipo de tráfego como: voz, dados e vídeo. Também serão explorados os conceitos sobre o fenômeno da auto-similaridade, salientando como ela se manifesta nos diversos tipos de tráfego e como mensurá-la.

3.2 Visão das Escalas de Tempo

A troca de informações entre os usuários das redes de comunicação de faixa larga, incluindo a troca de informações de usuários e de controle, geram um tráfego com comportamento altamente aleatório. Esse processo estocástico pode ser caracterizado por diversas variáveis aleatórias cujos eventos a elas relacionados podem ocorrer em períodos de tempo que variam desde nanosegundos até horas, dependendo do tipo de evento. Assim sendo, no processo de modelamento de tráfego, é interessante verificar e classificar as escalas de tempo em que os eventos ocorrem.

Considerando o antigo modelo de tráfego em telefonia concebido por Erlang, ou seja, uma rede de comunicação de voz comutada por circuito, os processos de chegada de chamadas, de estabelecimento e de término de chamadas, eram representados basicamente em apenas uma única escala de tempo.

Em uma rede ATM, devido as altas taxas de transmissão e de comutação, considerando sua complexidade e diversidade de atividades, por exemplo o tempo de comutação de uma célula, que ocorre na ordem de nanosegundos, obriga-nos a analisar o comportamento do tráfego da rede em diversas escalas de tempo diferentes. Em outras palavras, devemos estudar os eventos que afetam o comportamento de uma rede de comunicação faixa larga em escalas de tempo apropriadas. A Tabela 2 ilustra como seria essa divisão de escala de tempo para as diversas atividades.

	Intervenção Humana		Nível de Rede		
	Planejamento Projeto	O & M Gerência	Nível de Conexão	Nível de Rajada	Nível de Célula
Atividade	desempenho previsão demandas	manutenção configuração	admissão de conexões	multiplexagem grupo de pacotes	comutação de células
Escala de Tempo	semanas meses	minutos horas	segundos minutos	milisegundos	nanosegundos microseg.

Tabela 2: Atividades nas Escalas de Tempo

A divisão e classificação dos vários eventos que ocorrem no tráfego de uma rede faixa larga, foi proposta por Joseph Y. Hui [35][47][55]. Os conjuntos de eventos similares ocorridos ao longo do tempo, podem ser agrupados e descritos por processos estocásticos. Esses agrupamentos podem ser visualizados em três escalas de tempo diferentes, isto é: a nível de Célula, Rajada ou Conexão. As escalas de célula e rajada serão apresentadas neste capítulo, enquanto a escala de conexão será considerada no capítulo 5. A figura 9 ilustra os três níveis de escala de tempo.

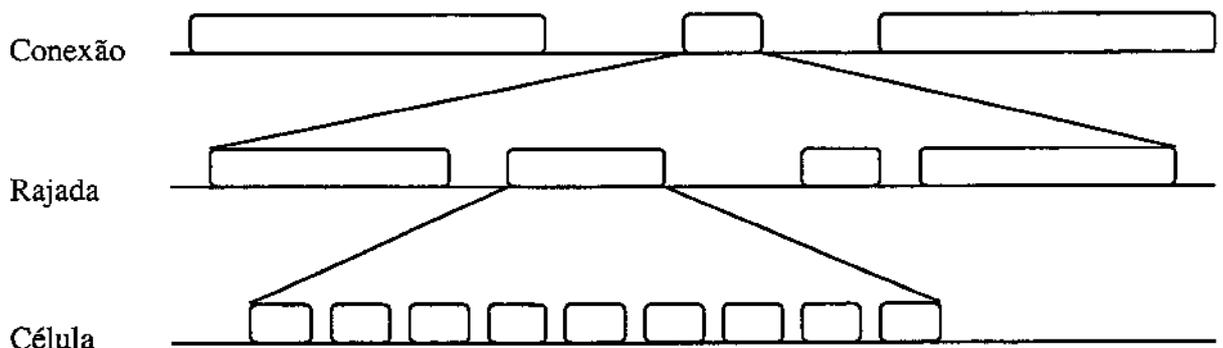


Figura 9: Escalas de Tempo para o tráfego ATM

Um tráfego na escala de tempo a nível de célula consiste de entidades discretas, ou seja, pacotes de tamanho fixo geralmente produzidos pela fonte a uma taxa inferior a taxa de transmissão do enlace de comunicação. Nesta escala de tempo, podemos considerar a maioria das fontes de tráfego, incluindo as VBR, como sendo periódicas localmente. Neste sentido, podemos dizer que a aleatoriedade do tráfego de entrada é oriunda da independência entre fontes locais, e não das flutuações das taxas individuais dessas fontes [47].

Quando o estudo de tráfego é feito na escala de tempo de rajada (*burst*), a granularidade a nível de célula é ignorada, e os processos de entrada são geralmente caracterizados pelas suas taxas instantâneas. Além disso, modelo de fluxo fluido tem sido empregado frequentemente como uma ferramenta natural para a modelagem. Nessa escala de tempo podem aparecer dois tipos de problemas, um provocado pelo excesso da taxa de entrada sobre a taxa de saída, e o outro em função do tipo de multiplexagem adotada no projeto [35][55].

Existem duas abordagens para se efetuar o controle do tráfego em uma rede faixa larga: controle preventivo e controle reativo. O controle preventivo é baseado na descrição de tráfego a priori (*open control loop*), sendo que o policiamento em uma conexão aceita, assegura que o usuário obedecerá os termos do contrato de tráfego. No caso do controle reativo é empregado um “*closed control loop*” permitindo que a taxa de entrada do usuário seja ajustada de acordo com a disponibilidade de banda do estado presente da rede. O controle reativo é indicado para o tráfego tipo ABR (*Avaiable Bit Rate*) proposto pelo ATM Forum [10].

No caso do controle preventivo são possíveis dois métodos de multiplexagem: a multiplexagem por taxa de envelope (*rate envelope multiplexing*) e a multiplexagem de taxa compartilhada (*rate sharing multiplexing*). A multiplexagem por taxa de envelope auxilia a preservar a taxa de bits de uma conexão limitando a admissão, assegurando que seja desprezível a probabilidade de que a soma das taxas de entrada exceda a capacidade do enlace. Por sua vez, a multiplexagem por taxa compartilhada permite um excesso de taxa de entrada, onde o excesso de tráfego é temporariamente armazenado em “*buffers*” de grande capacidade. Este último método permite uma maior utilização da rede, contudo causa atrasos maiores [35].

Usando a opção de multiplexação de taxa de envelope, os “*buffers*” são vistos como filas a nível de escala de células. Neste caso a perda ocorre a nível de rajada. Na opção de taxa compartilhada os “*buffers*” são capazes de absorver pelo menos parte das rajadas em função das flutuações estatísticas do tráfego.

A escala de tempo de conexão ou chamada, é caracterizada pelo tempo de espera das chegadas de chamadas ou das demandas de serviço. Associada com cada conexão está o conceito de capacidade equivalente ou banda efetiva, que é uma medida entre o valor da banda média e valor de pico de banda requerida para uma determinada conexão. O conceito de banda efetiva será largamente explorado neste estudo como critério de alocação de banda para uma determinada conexão.

Um dos aspectos chave para a engenharia de tráfego é a quantificação das características estatísticas relevantes do tráfego para o desempenho da rede. Esses parâmetros estatísticos junto com modelos adequados são necessários para projeto, dimensionamento e avaliação de desempenho de uma rede.

Em intervalos de tempo que correspondem a escala de tempo de célula, isto é, ordem do tempo entre as chegadas de células, o processo de chegada agregado se comporta como um processo de Poisson. Isto é justificado pelo fato que para pequenos intervalos de tempo cada célula é oriunda de diferentes fontes, portanto pode-se considerar que as chegadas são independentes. Para um intervalo de tempo na ordem de rajada, as células que chegam nesse intervalo são geralmente oriundas da mesma fonte, ou seja, correlacionadas, não mais se comportando como um processo Poissoniano.

3.3 Modelos a Nível de Célula

Na escala de tempo a nível de célula, as células consistem de entidades discretas, pacotes de tamanho fixo, geralmente produzidas pela fonte a uma taxa inferior a taxa de transmissão do enlace de comunicação. Nesta escala, a disciplina de filas nos “*buffers*” dos comutadores, em termos da multiplexagem das células, é do tipo FIFO (*First-In-First-Out*).

Nesta sessão serão apresentados alguns modelos a nível de escala de células frequentemente mencionados na literatura, tais como: Corrente de Renovação, processos MAP e D-BMAP, e uma breve descrição dos tipos de filas utilizadas para representar processos em redes ATM.

3.3.1 Corrente de Renovação (*Renewal Stream*):

O tráfego em uma rede de comunicação consiste de entidades discretas que podem ser representadas matematicamente por um processo Ponto. Os tempos de chegadas dessas entidades são representados pelos instantes $T_1, T_2, \dots, T_n, \dots$. Existem duas descrições equivalentes para um processo ponto. A primeira, através do processo de contagem $\{N(t)\}_{t=0}^{\infty}$, que é um processo estocástico não negativo de valores inteiros. A segunda, é através do processo de tempo entre chegadas $\{Y_n\}_{n=1}^{\infty}$, onde $Y_n = T_n - T_{n-1}$, que representa o comprimento do tempo entre as chegadas [76].

O processo de renovação é um caso particular de processo Ponto, onde as chegadas são consideradas independentes. O processo de Renovação é um processo estocástico $\{Y_n(t), t \geq 0\}$ de valores inteiros, não negativos, que registra as ocorrências sucessivas de um evento em um intervalo de tempo $(0, t]$, onde o período de duração de tempo entre os eventos, é uma variável aleatória independente e identicamente distribuída (*i.i.d - Independent Identically Distributed*) [71]. Diz-se que $\{Y_n(t)\}$ é uma Corrente de Renovação se sua distribuição é considerada geral. O processo de Poisson é um caso particular do processo de renovação com o tempo entre chegadas consecutivas distribuído exponencialmente. Infelizmente, com poucas exceções, a superposição de processos de renovação independentes não leva a outro processo de renovação [76]. Mais detalhes sobre processos de renovação e outros processos estocásticos consultar [21][71].

Quando trata-se de fontes individuais de tráfego, os intervalos de tempo entre chegadas podem ser modelados como uma seqüência de variáveis aleatórias independentes e identicamente distribuídas como um processo de renovação.

O caso mais simples entre os processos de renovação é aquele conhecido como corrente periódica estacionária, onde o elemento aleatório é a fase uniformemente distribuída. Normalmente em redes de comunicação reais as distribuições para tempos entre chegadas são de formatos não triviais devido a ampla variação do retardo nas filas dos nós de comutação da rede.

Talvez o processo de renovação mais antigo e mais utilizado seja o processo de Poisson, onde os tempos entre chegadas $\{Y_n\}$ são exponencialmente distribuídos, para taxa de chegadas igual a λ , com função distribuição de probabilidade de $\{Y_n\}$ sendo:

$$P\{Y_n \leq t\} = 1 - \exp(-\lambda t) \quad (5)$$

Como conseqüência a probabilidade de não haver chegadas em um intervalo de tempo t é dada por:

$$P\{Y_n > t\} = \exp(-\lambda t) \quad (6)$$

Uma outra forma de caracterizar um processo de Poisson é através de um processo de contagem $N(t)$, satisfazendo a seguinte expressão:

$$P\{N(t) = n\} = \exp(-\lambda t) \frac{(\lambda t)^n}{n!} \quad (7)$$

onde o número de chegadas em intervalos disjuntos é estatisticamente independente, sendo esta propriedade conhecida como incrementos independentes.

Uma outra propriedade importante do processo de Poisson é que a superposição de fontes independentes, resultará em um novo processo de Poisson com a taxa igual a soma das taxas das componentes. Outra propriedade é a dos incrementos independentes que torna o processo sem memória (*memoryless property*). A propriedade “sem memória” simplifica consideravelmente problemas de filas.

É um fato conhecido que a superposição de um grande número de processos de renovação independentes, até mesmo os periódicos, convergem localmente para um processo de Poisson, isto é, tem-se como resultado um processo de renovação com tempos entre chegadas exponencialmente distribuídos. Embora não se possa modelar uma única fonte ATM como um processo de Poisson, ele é um modelo aplicável para

tráfego agregado com um grande número de fontes relativamente lentas. Esta abordagem está fundamentada na Lei dos Grandes Números, ou seja, a multiplexagem de um número considerável de fontes de tráfego resulta em uma composição de chegadas Poissonianas. Em ATM esta quantidade dita “grande” requer no mínimo vinte fontes multiplexadas [56].

Um outro caso particular de corrente de renovação é o processo de Bernoulli de tempo discreto. Aqui a probabilidade de ocorrer uma chegada em qualquer partição (*slot*) de tempo é p , independente de qualquer outro fator. Assim sendo, o número de chegadas em k partições consecutivas possui uma distribuição binominal:

$$P\{N_k = n\} = \binom{k}{n} p^n \cdot (1-p)^{k-n} \quad (8)$$

onde n é um número pertencente ao conjunto $\{0, 1, \dots, k\}$. Este processo estocástico pode ser caracterizado alternativamente pela distribuição geométrica (com parâmetro p), do tempo entre chegadas consecutivas, ou seja:

$$P\{Y_n = j\} = p(1-p)^j \quad (9)$$

com j um inteiro não negativo.

3.3.2 Processo de Chegada de Lote Markoviano em Tempo Discreto:

Os processos markovianos em lote de tempo discreto (D-BMAP *Discrete Time Batch Markovian Arrival Processes*) são modelos a nível de escala de células, onde a distribuição do número de chegadas em uma partição de tempo é determinada por um processo de estado markoviano. Diferentemente dos modelos de renovação, os modelos de Markov introduzem dependência no processo $\{Y_n\}$. Estes modelos de Markov em tempo particionado (*slotted*) podem ser definidos pelo processo $\{Y_n\}$ em termos da matriz de transição de Markov $P = [p_{ij}]$. Aqui o estado i corresponde a uma partição de tempo inativa separando duas chegadas sucessivas e p_{ij} é a probabilidade de ocorrer uma partição de estado j , dado que a partição prévia foi uma partição de estado i . Note que o número de chegadas pode ser em unidades simples,

em lote de unidades ou qualquer quantidade contínua. Lotes podem ser descritos por uma Cadeia de Markov de estado contínuo [35].

Um caso especial do D-BMAP é o D-MAP (*Discrete time Markovian Arrival Process*), onde o tamanho do lote é igual a uma célula.

Em todos os casos de processos MAP (*Markovian Arrival Process*) a propriedade de Markov introduz dependência no tempo entre chegadas, e no tamanho dos lotes [76].

3.3.3 Filas em Escala de Células:

Nesta seção vamos considerar um multiplexador ATM com várias fontes de tráfego dirigidas para uma única saída. A nível de escala de célula, cada fonte pode ser vista como uma fonte de geração de células de taxa constante, sendo que a somatória das taxas individuais das fontes deve ser inferior a capacidade do enlace de saída do multiplexador. Normalmente o multiplexador não consegue sempre transmitir todas as células recebidas imediatamente, pois várias chegadas podem ocorrer simultaneamente. Assim sendo, as células são armazenadas temporariamente em “*buffers*” evitando as possíveis perdas de células [35]. A Figura 10 representa a entrada das células no “*buffer*”.

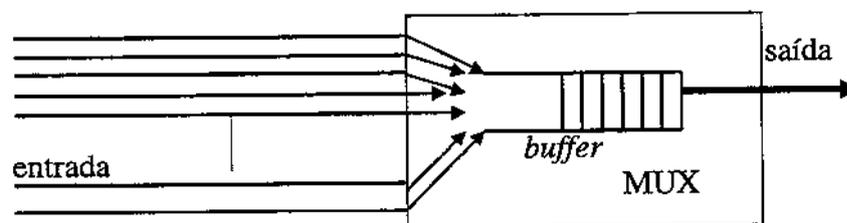


Figura 10: Multiplexador ATM

No modelo representado na Figura 10, pode ocorrer o transbordamento (*overflow*) da fila, se lotes de células que chegarem simultaneamente na mesma partição (*slot*) forem de tamanho superior a capacidade do “*buffer*”.

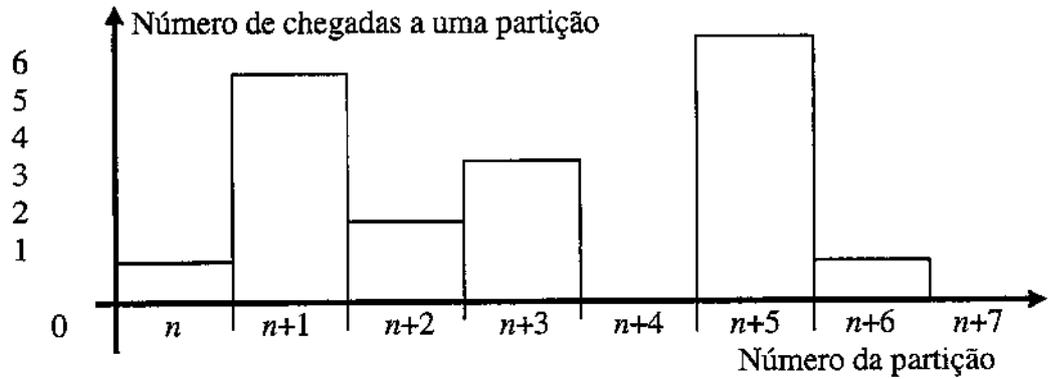


Figura 11: chegadas aleatórias por partição

Por exemplo na figura 11, na partição $n+5$ chegaram 6 células simultaneamente e para não ocorrer transbordamento, o “buffer” deve ter uma capacidade igual ou superior a seis células.

Em geral são usados para modelar o processo de fila única três modelos de filas na escala de célula: $M/D/1$, $ND/D/1$ e $\sum D_i/D/1$ [18]. Note que todos os três modelos possuem apenas um servidor e tempo de atendimento constante. De fato estes modelos são apropriados para pacotes de tamanho fixo. Cabe salientar a diferença entre os modelos $ND/D/1$ e $\sum D_i/D/1$, sendo que:

- $N.D$ - denota a multiplexagem de N processos de eventos determinísticos, onde cada processo possui as mesmas características (mesma taxa de transmissão);
- $\sum D_i$ - denota a multiplexagem de N processos de eventos determinísticos, onde cada processo possui características diferentes (diferentes taxas de transmissão).

A modelagem usando fila $M/D/1$ considera o processo de chegada Poissoniano e tempo entre chegadas exponencial negativa, com tempo de serviço determinístico, com um único servidor. É um modelo razoável para descrever a situação em que as células que estão sendo multiplexadas já passaram por várias multiplexagens prévias em outros nós da rede, fazendo com que o tempo entre chegadas de células tenha perdido qualquer característica determinística inicial. Considerando que as chegadas são independentes, a taxa do agregado é composta pela adição das taxas das fontes individuais [55]. Pode ser verificado que para baixos níveis de utilização a diferença entre os modelos $M/D/1$ e $ND/D/1$ são pouco significativas.

Modelagem usando Fila $\Sigma D_i/D/1$ é a mistura de N fontes periódicas, cada uma caracterizada por sua periodicidade individual de intervalo D_i . O fator de utilização $\rho = \sum I/D_i$ assumido é menor que a unidade.

Para maiores detalhes sobre a aplicação de sistemas de filas consultar [18][35][49] e [73].

3.4 Modelos para Escala em Rajada (*Burst*)

Na escala de Rajada ou Surto é ignorada a natureza discreta das chegadas de células e sim considerada a entrada como um fluido contínuo caracterizado pela taxa instantânea de chegada λ_t .

Nessa escala de tempo estamos interessados em fluxos de tráfego cujos tempos são maiores que os tempos entre chegada de células. Um exemplo típico são as fontes de vídeo, cuja duração do quadro de vídeo pode ser representado nessa escala de tempo.

Em relação a congestionamento, esse fenômeno ocorre quando a taxa de chegada λ_t excede a capacidade do multiplexador levando ao aumento da fila. Existem duas alternativas para solucionar o problema de congestionamento na escala de rajada:

- sistema de perda com “*buffer*” de pequeno tamanho para absorver somente congestionamento a nível de escala de célula;
- sistema de retardo com “*buffer*” de tamanho capaz de absorver congestionamento a nível de rajada.

Na escala de rajada é interessante observar os fenômenos que podem causar modulação da taxa de transmissão de células em pequenas escalas de tempo, tais como:

- empacotamento em protocolo de dados, onde um certo número de células são necessárias para suportar um pacote de dados;
- rajada de dados como resultado de interação em tempo real;

- mudanças de cena em sinais de vídeo VBR com qualidade constante, onde um aumento da taxa é requerido para completar um novo quadro, em vez de pequenas alterações do quadro anterior;
- interação de voz com períodos de atividade e inatividade em função das variações na conversação [55].

Nesta seção vamos apresentar os seguintes modelos na escala de rajada: MMPP, Fluido Markoviano e modelos Auto-Regressivos.

3.4.1 Processo On/Off

Primeiramente apresentamos o modelo On/Off simples, pois o modelo Fluido de Markov é composto pela sobreposição de fontes On/Off.

Um processo On/Off pode ser caracterizado por uma fonte que alterna períodos de atividade (On) e períodos de silêncio (Off), com a taxa de transmissão constante em cada período.

Um exemplo típico do processo On/Off seria uma conversação telefônica, conforme ilustrado na Figura 12. A voz humana consiste de seqüências alternadas de intervalos de falas (*talk spurt*), tipicamente 0,4 a 1,2 segundos de duração seguido de um período de silêncio com duração média de 0,6 a 1,8 segundos. A duração da fala em geral pode ser caracterizada por uma distribuição exponencial, enquanto que os períodos de silêncio não são tão bem representados por essa distribuição [42][56].

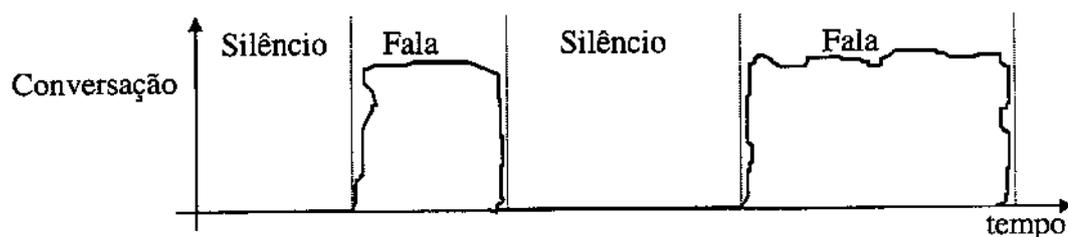


Figura 12: Comportamento da voz

Um processo On/Off pode ser representado via modelo de nascimento e morte. O diagrama de estados típico de um modelo de nascimento e morte é ilustrado na Figura 13, onde o parâmetro β representa a taxa de transição do estado de silêncio

para o estado de fala, e o parâmetro α a taxa de transição do estado de fala para o estado de silêncio. Assim a duração média do período de fala e de silêncio é $1/\alpha$ segundos e $1/\beta$ segundos, respectivamente.

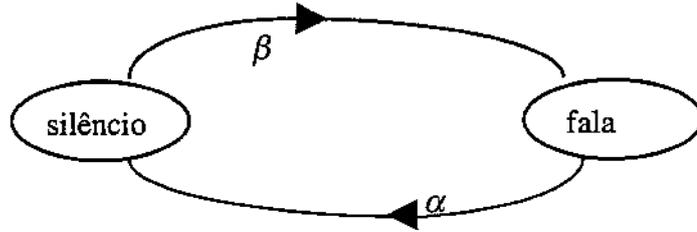


Figura 13: modelo de dois estados para fonte de voz

No estado estacionário, o processo On/Off da Figura 13 pode ser representado através da seguinte equação de equilíbrio:

$$\beta.P\{\text{silêncio}\} = \alpha.P\{\text{fala}\} \quad (10)$$

onde $P\{\text{fala}\}$ e $P\{\text{silêncio}\}$ são probabilidades de período ativo e período inativo respectivamente, satisfazendo a lei da probabilidade total, ou seja:

$$P\{\text{fala}\} + P\{\text{silêncio}\} = 1 \quad (11)$$

Resolvendo o sistema formado pelas suas equações acima temos:

$$P\{\text{fala}\} = \frac{\beta}{\alpha + \beta} \quad (12)$$

$$P\{\text{silêncio}\} = \frac{\alpha}{\alpha + \beta} \quad (13)$$

Adotando valores típicos como $1/\alpha = 0,4$ seg e $1/\beta = 0,6$ seg, teremos que $P\{\text{fala}\} = \beta/\alpha + \beta = 0,4$. Este resultado sugere que para um grande número de usuários uma rede comutada a circuito poderia acomodar um número de usuários 2,5 ($=1/0,4$) vezes maior que o número de circuitos instalados.

Como exemplo podemos citar a codificação PCM de voz (amostragem a cada 125 μ seg, 8 bits por amostra e 256 níveis de quantização). A carga da célula ATM usando a AAL1 é de 47 bytes, ou seja 47 amostras de voz digitalizada. Assim,

durante a fala (período ativo) uma célula ATM é gerada a cada 5,875 mseg (8,47 / 64 kbps), que no período de fala uma cadeia periódica (*stream*) $V = 171$ células ATM geradas a cada segundo ($1/5,875 \text{ mseg} = 170,21$).

A Figura 14 mostra a multiplexagem de N canais de voz.

Naturalmente o requisito básico para que um enlace de capacidade VC possa servir N canais de voz, cada um produzindo V células/segundo, de forma adequada, é $N.V < C$, onde C é um parâmetro adimensional. Considerando o ganho da multiplexagem estatística, temos a seguinte condição melhorada:

$$N \left(\frac{\beta}{\alpha + \beta} \right) < C \quad (14)$$

Assim, o fator de utilização (ρ) do enlace de capacidade VC , por N canais de voz deve obedecer a seguinte restrição:

$$\rho \equiv \frac{N \left(\frac{\beta}{\alpha + \beta} \right)}{C} < 1 \quad (15)$$

A multiplexagem de N canais de voz pode ser vista como um processo estocástico de nascimento e morte de $N+1$ estados, conforme ilustrado na Figura 15.

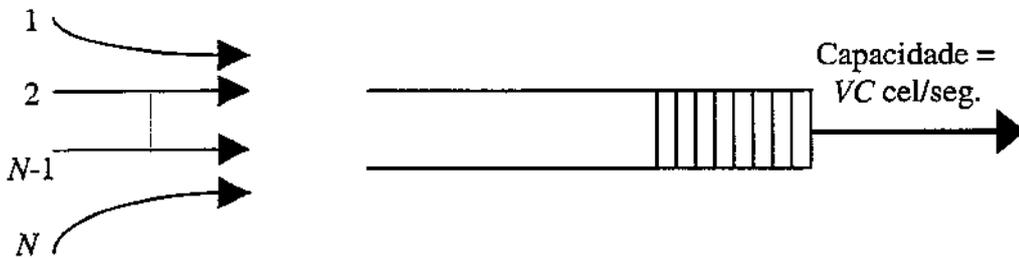


Figura 14: multiplexagem de N canais

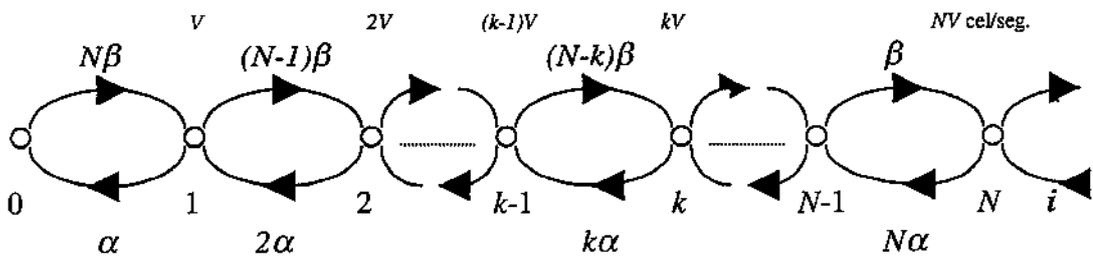


Figura 15: Diagrama de estados para N fontes de voz.

No estado i , teremos i fontes no estado ativo (falando) e a taxa média de células introduzidas no “buffer” é de $i.V$ cel/seg. Como cada fonte gera V cel/seg periodicamente durante a fala, a soma dessas fontes independentes e não sincronizadas pode ser vista como um processo de entrada aleatório com a média de $i.V$ cel/seg.

3.4.2 Processos Poisson Markov Modulados (MMPP - *Markov Modulated Poisson Process*)

O MMPP é um caso especial de processo de chegada Markoviano (MAP - *Markovian Arrival Process*). É um processo estocástico duplo onde a intensidade de chegada $\lambda_k(t)$, do processo de Poisson, é determinada pelo estado de uma cadeia de Markov de tempo contínuo, definida em um número finito de m estados. Quando a fonte se encontra no estado k , gera células aleatoriamente com a taxa média λ_k , segundo um processo de Poisson. Assim sendo, o tempo entre chegadas consecutivas possui uma distribuição exponencial negativa. Dizemos que o processo modulado é de Poisson e a cadeia de Markov é o processo modulante [31][34][56].

Um MMPP pode ser classificado através do número total de estados que a cadeia de Markov modulante possui. Por exemplo, se a cadeia de Markov possui dois estados o processo é denominado MMPP-2, também chamado SPP (*Switched Poisson Process*), como representado na figura 16. Se as duas intensidades das taxas de chegada são iguais, tem-se um processo de Poisson ordinário. No caso de uma das taxas de chegada ser zero, o processo é denominado de IPP (*Interrupted Poisson Process*). O modelo IPP tem sido amplamente usado para modelar o tráfego oriundo de fontes de voz. O estado ON, corresponde a “jorro” (*spurt* - quando a pessoa fala) e o estado OFF corresponde ao silêncio, pausas durante a fala. Uma extensão do processo básico MMPP pode ser considerada uma agregação de fontes de tráfego independentes, cada uma sendo um MMPP modulados por processos individuais M_i de Markov.

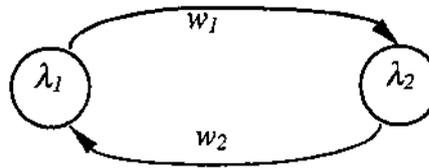


Figura 16: MMPP de dois estado

Um MMPP genérico pode ser definido por duas matrizes Q e Λ de dimensões $m \times m$, onde m é o número de estados ou fases. A matriz Q (16) é chamada matriz de transição de estados da cadeia de Markov modulante e a matriz Λ diagonal (17) contém na sua diagonal as intensidades das taxas de chegada correspondentes aos diferentes estados da cadeia e os demais elementos nulos [34].

$$Q = \begin{bmatrix} -w_{11} & w_{12} & \dots & w_{1m} \\ w_{21} & -w_{22} & \dots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \dots & -w_{mm} \end{bmatrix} \quad (16)$$

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_m \end{bmatrix} \quad (17)$$

Para um MMPP-2 estacionário, considerando o processo de contagem do número de chegadas em um intervalo de tempo de comprimento t , $N(t)$, temos:

$$Q = \begin{bmatrix} -w_1 & w_1 \\ w_2 & -w_2 \end{bmatrix} \quad (18)$$

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \quad (19)$$

O número médio de chegadas no intervalo $[0, t]$ é dado por:

$$E\{N(t)\} = \left[\frac{(\lambda_1 \cdot w_2 + \lambda_2 \cdot w_1)}{(w_1 + w_2)} \right] t \quad (20)$$

O Índice de Dispersão de $N(t)$ é dado por :

$$IDC(t) = \frac{Var\{N(t)\}}{E\{N(t)\}} \quad (21)$$

Considerando a situação em que $t \rightarrow \infty$, temos:

$$IDC = \frac{1 + [2(\lambda_1 + \lambda_2)^2 \cdot w_1 \cdot w_2]}{(w_1 + w_2)^2 \cdot (\lambda_1 w_2 + \lambda_2 w_1)} \quad (22)$$

Note que $1/w_1$ e $1/w_2$, são os tempos médios de permanência nos respectivos estados.

Se a fonte for do tipo On/Off caracterizada pelos seguintes parâmetros:

- $1/\alpha$: duração média em que a fonte se encontra no período ativo;
- $1/\beta$: duração média em que a fonte se encontra no período de silêncio;
- T : tempo médio entre geração de células no período ativo temos.

Considerando $1/\alpha$ e $1/\beta$ exponencialmente distribuídos, o processo de chegada de uma única fonte será um processo de renovação com taxa de chegada λ dado por:

$$\lambda = \frac{\beta}{T(\alpha + \beta)} \quad (23)$$

Calcula-se o quadrado do coeficiente de variação C^2_α , dado por $(Var\{}/E\{})^2$:

$$C^2_\alpha = \frac{[1 - (1 - \alpha T)^2]}{T^2 (\alpha + \beta)^2} \quad (24)$$

O parâmetro “Skewness” S_k , é dado pela razão entre o terceiro momento central e Variância $\}^{3/2}$, referente a distribuição de tempo entre chegada de células [30][34]:

$$S_k = \frac{[2\alpha T (\alpha^2 T^2 - 3\alpha T + 3)]}{[\alpha T (2 - \alpha T)]^{3/2}} \quad (25)$$

A superposição de processos de chegadas de n fontes independentes, não é considerada um processo de renovação, contudo pode ser aproximada por um MMPP de duas fases [30]. Considere w_1 e w_2 as taxas de transição de fase e λ_1 e λ_2 as taxas de chegada, sendo que os parâmetros do MMPP-2 são dados pelas equações (26) e (27):

$$w_1, w_2 = D \left(1 \pm \frac{1}{\sqrt{1 + n\lambda E}} \right) \quad (26)$$

$$\lambda_1, \lambda_2 = \lambda n + F [1 \pm \sqrt{1 + n\lambda E}] \quad (27)$$

com

$$D \equiv \frac{I'(0)}{I(\infty) - 1} \quad (28)$$

$$E \equiv \frac{I'(0)}{F^2} \quad (29)$$

e

$$F \equiv \frac{I'(0)[S(\infty) - 3I(\infty) + 2]}{3[I(\infty) - 1]^2} \quad (30)$$

onde $I(\infty)$, $I'(0)$ e $S(\infty)$ são definidos abaixo. Considere $N(t)$ o número de chegadas no intervalo de $(0, t]$ com média $E\{ \}$, sendo $I(t)$ o índice de dispersão e $S(t)$ o índice de “skewness”, dados por (31) e (32):

$$I(t) \equiv \frac{E\{[N(t) - \lambda t]^2\}}{\lambda t} \quad (31)$$

$$S(t) \equiv \frac{E\{[N(t) - \lambda t]^3\}}{\lambda t} \quad (32)$$

De acordo com os resultados de simulação obtidos por Daigle e Langford [42] o modelo MMPP não é a melhor opção para caracterização do tráfego de voz. Essa conclusão é esperada uma vez que o modelo MMPP assume que as células são geradas a taxas variadas aleatoriamente nos vários estados. Como as fontes de voz produzem uma corrente periódica de células durante o período ativo, é esperado uma imprecisão do modelo MMPP para fontes de voz multiplexadas [56].

Um modelo MMPP de dois estados pode ser usado para modelar uma superposição de processos de renovação representando tráfego agregado de voz e dados. Esta abordagem é plausível porque os dados podem ser modelados por um processo de Poisson com taxa λ_d , e sua superposição com um MMPP-2, sendo o agregado novamente um MMPP com taxas $\lambda_1 + \lambda_d$ e $\lambda_2 + \lambda_d$.

Freqüentemente o tráfego em redes ATM é assumido como fontes do tipo On/Off. A superposição N fontes On/Off homogêneas pode resultar em um processo bastante complexo. No entanto, quando somente um parâmetro é considerado em termos de análise de desempenho, por exemplo, taxa de perda de células, é possível

se aproximar o processo de N fontes homogêneas On/Off por um modelo MMPP-2 sem perda significativa de informação [85].

Outra aplicação do modelo MMPP é na representação de tráfego de vídeo. Neste caso, é conveniente que uma dada fonte de vídeo seja substituída pela versão equivalente quantizada. O tráfego é subdividido em oito níveis que muda de quadro para quadro. Assim sendo, as células de um determinado quadro (quantizadas) enviadas ao multiplexador ATM podem ser vistas como um tráfego de Poisson. Neste caso assumimos que os níveis de transição de quadro para quadro não possui memória. Isto leva a uma cadeia de Markov de tempo discreto de oito estados, com transições entre estados a cada intervalo de quadro. Esta cadeia então modula o processo de Poisson, visto que cada estado é representado por um processo de chegada Poissoniano com taxa de chegada correspondendo ao número de bits ou células a ser gerado no seu respectivo estado [56].

Segundo Yamada [59], a chegada de pacotes de sinais de vídeo que incluem mudanças de cena, podem ser representadas por um processo de Poisson, sendo que a taxa de chegada varia de acordo com a cadeia de Markov. Para isso são assumidas as seguintes premissas:

- uma cena simples é vista como um processo que gera pacotes como Poisson;
- a taxa de transição entre cenas é dada pela cadeia de Markov;
- a alta taxa de bits imediatamente a seguir a uma mudança de cena é considerada como um estado de alto nível ou chegada em grupos.

Note que neste modelo os crescimentos abruptos de taxa ocorrem geralmente nas mudanças de cena [59].

A desvantagem do modelo MMPP é que sua aplicação torna-se difícil quando o número de estados aumenta, em função da solução numérica para as equações de estado.

3.4.3 Fonte de Tráfego Fluido Markoviano

Uma fonte é chamada de fluido Markoviano se sua derivada temporal é função de uma cadeia de Markov de tempo contínuo em um espaço de estados finito. Seja Y_t

a taxa de células produzidas pela fonte quando o fluido de Markov encontra-se no estado i . O Fluido de Markov é o resultado da superposição de vários processos On/Off. Assim a fonte Fluido de Markov possui um número finito de fases que correspondem ao número de fontes On/Off ativas.

A diferença entre as fontes do tipo MMPP e Fluido de Markov está na taxa de geração de células, que no último caso é constante em cada estado. No estado OFF, em cada fonte On/Off o tráfego é cortado, enquanto no estado ON o tráfego chega determinísticamente a uma taxa constante λ . Para simplificar a análise, a duração dos estados ON e OFF são assumidos exponencialmente distribuídos e mutuamente independentes, formando assim um processo de renovação alternado [34][76][85].

Um processo Fluido de Markov também pode ser definido através de duas matrizes de dimensão $m \times m$, sendo Q (33) a matriz de transição de probabilidade e Λ (34) a matriz diagonal das taxas médias em seus respectivos estados.

$$Q = \begin{bmatrix} -w_{11} & w_{12} & \dots & w_{1m} \\ w_{21} & -w_{22} & \dots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \dots & -w_{mm} \end{bmatrix} \quad (33)$$

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_m \end{bmatrix} \quad (34)$$

No caso do Fluido de Markov, o tráfego é considerado como um fluido contínuo, caracterizado por uma taxa de fluxo (como bits/segundo), sendo o mesmo contado por volume simplificando a análise [76]. Este modelo é apropriado para casos onde as unidades individuais são relativamente numerosas. Como exemplo, em ATM, considerando as células de 53 bytes e as altas taxas de transmissão, qualquer impacto em uma determinada célula pode ser considerado desprezível.

No contexto de filas, a manipulação de fluidos em “*buffers*” pode ser considerada fácil. Esta afirmação deve-se aos seguintes fatores: o conceito de tempo de espera simplesmente torna-se o tempo para esvaziar o “*buffer*”; a probabilidade de perda pode ser calculada em termos de volumes de “*overflow*”.

3.4.3.1 Modelo de Fonte Fluida para Voz

Neste modelo é suposto que o número de células geradas durante a fala (*talk spurt*) é tão grande que aparece como um fluxo contínuo de um fluido. Considere a situação da figura 17, onde N é o número de fontes, VC a capacidade do canal em células/segundo; V é o número de células/segundo geradas por cada fonte; e C é um parâmetro adimensional. Esta aproximação é válida se o número de fontes N e a capacidade VC é grande o suficiente, para que a discretização das chegadas e partidas no “*buffer*” a nível de célula seja desprezível. Este é um modelo como mostrado anteriormente a nível de rajada, onde a ocupação do “*buffer*” é uma variável aleatória contínua X [56].

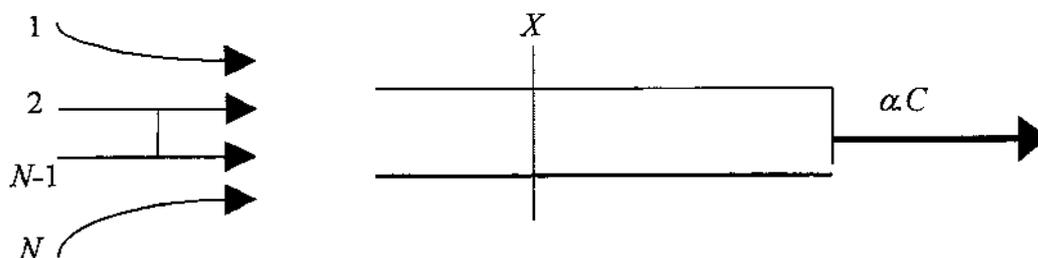


Figura 17: *buffer* de acesso

A variável aleatória X é definida como o número de células que chegam durante o período de fala. Assumindo também que cada fonte de voz gera V cel/seg durante um período $1/\alpha$ segundos, na média X é incrementada por V/α células durante o período de fala. Um sistema com capacidade de VC cel/seg terá uma capacidade equivalente de $VC/(V/\alpha) = \alpha C$ unidades de informação por segundo.

A descobrir a ocupação do “*buffer*”, deve-se achar a distribuição de probabilidade da variável aleatória X . Para tanto segue-se os seguintes passos:

- L é o estado do “buffer” em unidade de células, como X é incrementada por V/α células durante o período de fala, nós temos $L = X.V/\alpha$, como conversão entre unidades de informação e células.
- A probabilidade que a ocupação do “buffer” exceda um número i é dada por $P\{L > i\} = P\{X > \alpha i/V\}$, onde $P\{X > X_0\}$ é a probabilidade de X exceder a X_0 .

Como exemplo tomemos $V = 170$ cel/seg, $1/\alpha = 1,25$ seg ($\alpha = 0,8$) e $V/\alpha = 212,5$ células, temos $P\{L > i\} = P\{X > i/212,5\}$. Se $i = 1.000$ células, obtemos $P\{L > 1.000\} = P\{X > 4,7\}$ [42][56].

Resultados obtidos por Daigle e Langford [42], através de simulação indicam que este modelo fornece boa aproximação para modelagem de tráfego de voz.

3.4.3.2 Modelo de Fonte Fluida para Fontes de Taxa Variável

Em redes ATM o tráfego pode ser gerado através de fontes com taxa variável, isto significa:

- a nível de célula, a taxa pode variar continuamente de modo que o tempo entre células varia, existe correlação entre chegada de células sucessivas;
- a nível de rajada existe a correlação do processo de chegada de células [85].

O modelo de fluxo fluido para a escala de rajada é obtido da equação de distribuição do tempo de espera em vez das equações diferenciais de equilíbrio da distribuição no “buffer”.

Seja $\{W(t), t \geq 0\}$ a quantidade de trabalho que chega no sistema no intervalo de tempo $(-t, 0)$ e V_t o tempo de espera virtual em $-t$. O tempo de espera virtual é definido como o tempo em que um cliente que chega no instante $-t$ deve esperar pelo serviço, que é equivalente a quantia de trabalho a ser realizada no instante $-t$. Definimos $X(t) = W(t) - t$ ser o excesso de trabalho ou a sobrecarga no intervalo $(-t, 0)$. O tempo virtual de espera no instante 0 é dado por

$$V_0 = \sup_{t \geq 0} \{X(t)\} \quad (35)$$

Seja $W(t)$ uma função contínua, considerando que possui a saída a uma taxa constante sempre que seu conteúdo for diferente de zero em V_t . Definimos Λ_t como a taxa de chegada no instante $-t$.

$$\Lambda_t = \frac{dW(t)}{dt} = 1 + \frac{dX(t)}{dt} \quad (36)$$

Seja $v(x)$ a função distribuição complementar de V_0 : $v(x) = P\{V_0 > x\}$. O instante de tempo u é definido o ponto em que $\{X(u) = x \text{ e } V_u = 0\}$. Assim a equação da distribuição complementar do tempo de espera é dada por

$$v(x) = \int_{u>0} \int_{0 \leq \lambda \leq 1} (1-\lambda) \frac{d^2}{dxdt} P\{X(u) \leq x, \Lambda_u \leq \lambda \text{ e } V_u = 0\} d\lambda du \quad (37)$$

O processo fluido da entrada consiste de uma superposição de múltiplas fontes On/Off idênticas e estatisticamente independentes, pode ser derivado da equação (37) tendo forma similar [85].

3.4.4 Modelos Auto-Regressivos

Em uma seqüência de variáveis aleatórias o modelo Auto-Regressivo define a próxima variável aleatória como uma função explícita da prévia. Em termos de série temporal, uma janela de tempo do presente é vista como uma extensão do passado. Este modelo é particularmente interessante na modelagem de vídeo codificado (VBR) que projeta um maior consumo de banda nas redes de comunicação como a de ATM.

A natureza dos quadros de vídeo é tal que quadros sucessivos dentro de uma cena varia visualmente muito pouco. Somente a mudança de cena, ou seja, uma descontinuidade visual, causa mudanças bruscas na taxa de quadros. A literatura sugere, que uma seqüência de taxas de bits compondo uma cena de vídeo seja modelada através de um esquema auto-regressivo e as mudanças bruscas na cena sejam vistas como um mecanismo de modulação. Este mecanismo de modulação tem sido amplamente feito através de uma cadeia de Markov [76].

Nesta sessão são apresentados resumidamente os seguintes modelos auto-regressivos: Modelo Auto-regressivo Linear (AR), Modelo de Média Móvel (MA), Modelo ARMA e finalmente o modelo ARIMA.

3.4.4.1 Modelo Auto-regressivo Linear (AR)

Uma série temporal X_n é considerada como um processo linear se possuir a seguinte representação:

$$X_n = \sum_{j=-\infty}^{\infty} \Psi_j \cdot U_{n-j} \quad , \text{ para } 0 \leq n \leq \infty \quad (38)$$

onde $\{U_n\}$ é o ruído branco com distribuição gaussiana $(0, \sigma^2)$ e $\{\Psi_j\}$ é uma seqüência de constantes reais sob a restrição

$$\sum_{j=-\infty}^{\infty} |\Psi_j| < \infty \quad (39)$$

O modelo Auto-regressivo Linear é representado da seguinte forma:

$$X_n = \phi_0 + \sum_{r=1}^p \phi_r X_{n-r} + \varepsilon_r, \quad n > 0 \quad (40)$$

onde X_0, X_1, \dots, X_{p-1} , são variáveis aleatórias, ϕ_r são constantes reais, e ε_r são variáveis aleatórias iid, chamadas residuais e independentes de X_n 's. Para codificação de vídeo X_{n-r} é a taxa de transmissão do quadro $n-r$ [64].

A equação (40) descreve de forma simples um esquema linear auto-regressivo de ordem p , denotado por $AR(p)$. Os resíduos devem ser menores em magnitude do que X_n [76].

O modelo Auto-Regressivo Linear é aplicável para representar fontes de vídeo com taxa variável (com imagens em movimento) e codec com “buffer” de quadro [34]. De fato o modelo AR é bastante natural para representar correlações temporais “intraframe”, ou seja, correlações dentro de um mesmo quadro. Pode-se mostrar que a correlação da variabilidade das taxas de bits apresenta um decaimento exponencial [59].

A forma recursiva da equação (40), torna claro como gerar aleatoriamente o próximo elemento aleatório na seqüência $\{X_n, 0 \leq n \leq \infty\}$. Esta simplicidade faz o esquema AR um forte candidato para modelagem de tráfego autocorrelacionado, particularmente para VBR vídeo. Modelos mais elaborados podem ser construídos do AR(p) combinado com outros esquemas.

3.4.4.2 Modelo Média Móvel (MA)

Um processo $\{X_n\}$ é dito Média Móvel de ordem q , se:

$$X_n = U_n + \theta_1 U_{n-1} + \theta_2 U_{n-2} + \dots + \theta_q U_{n-q} \quad (41)$$

onde $\{U_n\}$ é o ruído branco e $\theta_1, \dots, \theta_q$ são constantes reais.

Considerando um processo MA(1) caracterizado pela seguinte equação:

$$X_n = U_n + \theta U_{n-1}, \quad n = 0, \pm 1, \pm 2, \dots \quad (42)$$

onde $\{U_n\}$ é o ruído branco com distribuição gaussiana $(0, \sigma^2)$ e θ é uma constante real. Pode-se provar que neste caso $E\{X_n\} = 0$, $E\{X_n^2\} = \sigma^2(1 + \theta^2) < \infty$ e considerando para todo n e $h \geq 0$ a função de Autocovariância e a Autocorrelação são representadas respectivamente:

$$\gamma_X(n+h, n) = \begin{cases} \sigma^2(1 + \theta^2), & \text{se } h = 0 \\ \sigma^2 \theta, & \text{se } h = \pm 1 \\ 0, & \text{se } |h| > 1 \end{cases} \quad (43)$$

$$\rho_X(h) = \begin{cases} 1 & \text{se } h = 0 \\ \theta / (1 + \theta^2), & \text{se } h = \pm 1 \\ 0, & \text{se } |h| > 1 \end{cases} \quad (44)$$

3.4.4.3 Modelo Auto-Regressivo com Média Móvel - ARMA

O modelo ARMA (*Autoregressive Moving Average*) captura dois tipos de correlação que ocorrem em cenas com níveis de atividades não uniformes, sendo uma combinação dos modelos AR(p) e MA(q), denotado por ARMA (p, q). A

primeira refere-se a correlação entre quadros e a segunda entre cada linha de um quadro. Esse processo é denominado de “recorrelação”. Na técnica de codificação de Fornecimento Condicional (CR- *Conditional Replenishment*) ocorre esse efeito de recorrelação que também é exibido pelo processo ARMA [85].

Seja uma série temporal $\{X_n\}$, a mesma é dita um processo ARMA(p, q), se $\{X_n\}$ é estacionária e se para cada n tem-se

$$X_n = - \sum_{k=1}^p \phi_k X_{n-k} + \sum_{k=0}^q \theta_k U_{n-k} \quad (45)$$

ϕ_k e θ_k são seqüências de números reais ou complexos, $\{U_k\}$ é um ruído branco, e ϕ_0 e $\theta_0 = 1$. Representando as seqüências ϕ_k e θ_k via suas transformadas z , temos:

$$\phi(z) = \sum_{k=0}^p \phi_k z^{-k} \quad (\text{transformada } z \text{ da parte AR}) \quad (46)$$

$$\theta(z) = \sum_{k=0}^q \theta_k z^{-k} \quad (\text{transformada } z \text{ da parte MA}) \quad (47)$$

Note que X_{n-k} depende da própria saída do processo e da média móvel do ruído branco. Vale a pena lembrar que no domínio da transformada Z , Z^{-1} corresponde a um atraso de uma unidade da variável X_n , conforme ilustrado na figura 18.

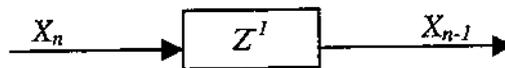


Figura 18: Operador de atraso

Usando B como operador de atraso unitário (*backshift operator*) temos:

$$B.X_n = X_{n-1} \quad (48)$$

No caso de um processo ARMA(1,1) a equação (45) se reduz a:

$$X_n - \phi X_{n-1} = U_n + \theta U_{n-1} \quad (49)$$

onde $\{U_n\}$ é a seqüência de ruído branco. Usando o operador de atraso B , temos:

$$\phi(B) X_n = \theta(B)U_n \quad (50)$$

Note que $\phi(B)$ e $\theta(B)$ podem ser vistos como filtros lineares representados através das seguintes equações

$$\phi(B) = 1 - \phi(B) \quad (51)$$

$$\theta(B) = 1 + \theta(B) \quad (52)$$

A implementação do modelo matemático ARMA é ilustrada na figura 19 através de um filtro transversal de ordem finita e de um filtro recursivo. A ocorrência periódica da autocovariância é executada pela parte AR, enquanto que correlações de diferentes intervalos é feita através da parte MA.

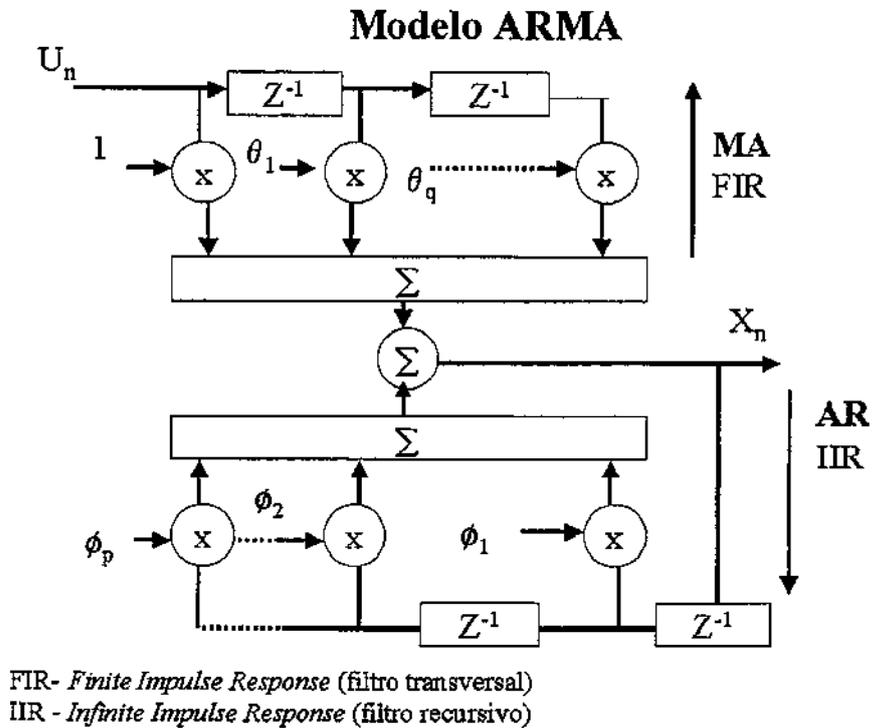


Figura 19: processo ARMA

Outro aspecto a considerar, é que o efeito da correlação torna-se mais evidente, quando a carga ou fator de utilização do canal de comunicação é mais intensa. Nessa situação pequenas mudanças na covariância causam grandes alterações no retardo e na perda de células.

Tem sido observado que o modelo ARMA não é conveniente para ser aplicado em análise de filas [33], mesmo assim os modelos ARMA, bem como o AR e MA são adequados para representar processos com apenas dependência de curto prazo, ou seja, caracterizados por decaimento exponencial ou abrupto.

3.4.4.4 Modelo Auto-Regressivo com Média Móvel Integrada - ARIMA

O modelo ARIMA introduzido por Box e Jenkins em 1970 [36] é baseado no modelo ARMA.

Um processo X_n , este é dito “*Auto-Regressive Integrated Moving Average*” com parâmetros p , d e q , denotado por ARIMA (p,d,q) , sendo d um inteiro não negativo, se for obtido a partir de d integrações de um processo ARMA Y_n , obedecendo a seguinte expressão:

$$X_n = Y_n \cdot \frac{1}{(1-B)^d} \quad (53)$$

Assim X_n é solução da seguinte equação:

$$\phi(B)(1-B)^d X_n = \theta(B)U_n \quad (54)$$

Note que a equação (55) é a operação inversa da equação (53), representando a derivada de ordem d do processo X_n .

$$Y_n = (1-B)^d \cdot X_n \quad (55)$$

Na Figura 20 é representada a seqüência de obtenção do processo ARIMA, onde $\theta(B)$ e $\phi(B)$ são polinômios em B relativos aos modelos AR e MA respectivamente e U_n é uma seqüência de ruído e d a ordem de diferenciação.



Figura 20: processo ARIMA

Generalizando d é inteiro e maior que zero, desenvolvendo $(1-B)^d$ em série, obtemos a seguinte representação [36]:

$$(1-B)^d = \sum_{k=0}^d \binom{d}{k} (-1)^k B^k \quad (56)$$

Uma maneira de se ver o modelo ARIMA é uma sintonia fina da caminhada aleatória (*ramdon walk*). Essa sintonia fina consiste em se adicionar intervalos da série diferenciada e/ou intervalos de previsão de erros na equação de previsão.

Os intervalos da série diferenciada que aparecem na equação de previsão são chamados termos auto-regressivos e os intervalos de previsão de erros são chamados de termos de média mutante (*moving average*). A série temporal que necessita ser diferenciada para se tornar estacionária é dita versão integrada da série, onde:

- p - é o número de termos auto-regressivos;
- d - é o número de diferenças não periódicas;
- q - é o número de intervalos de erros na equação de previsão.

Note que o processo ARMA(p,q) é também um processo ARIMA ($p,0,q$). Se $d \geq 1$, então a série original X_n é não estacionária. Para se obter um processo estacionário, X_n deve ser diferenciada d vezes.

O modelo ARIMA fornece boa representação para certos processos não estacionários. Pode-se mostrar que a função de autocorrelação de processos tipo ARMA e ARIMA decaem exponencialmente.

3.4.5 Processos Auto-Similares

A concepção sobre processo Auto-Similar foi introduzida inicialmente por Kolmogorov em 1941. Esse conceito no entanto só foi usado na década de 60 por Mandelbrot, Wallis e Van Ness quando o mesmo foi empregado na estatística [36].

Por exemplo, uma forma geométrica é dita auto-similar determinística, se a mesma estrutura geométrica for mantida independente da distância que é observada. No contexto dos processos estocásticos, a auto-similaridade pode ser vista através da distribuição de probabilidade do processo envolvido [36].

A manifestação da auto-similaridade nas formas geométricas e em outros fenômenos naturais é conhecida como fractal, exemplos disso incluem: ondas dos oceanos, distribuição de terremotos, formação do relevo e nuvens, etc [84].

Recentes estudos em medidas de tráfego de alta precisão e resolução, têm revelado o fenômeno da auto-similaridade em redes de comunicação [38][44][78]. Esse fato é extremamente interessante e mostra a necessidade de inovação na modelagem, planejamento, dimensionamento e controle das redes. Um dos pioneiros na constatação desse fenômeno em redes de comunicação foi Leland [78] que analisou milhões de quadros de dados de uma rede local Ethernet em um ambiente de P&D. Nesse estudo o tráfego revela as propriedades estatisticamente auto-similares, notavelmente não importando o nível de agregação, o tráfego exhibe estruturas similares através de todas, ou de grande parte, das escalas de tempo. No caso de tráfego de pacotes, a auto-similaridade é manifestada pela ausência de um comprimento determinístico de rajada. Considerando diferentes escalas de tempo variando de poucos milissegundos até minutos e horas, o comportamento em rajadas do tráfego se mantém inalterado, ou seja, não varia em função de diferentes escalas.

Vários outros estudos também têm demonstrado que a presença do fenômeno fractal não ocorre somente no tráfego Ethernet, mas também em fontes de vídeo VBR, incluindo videoconferência e vídeo codificado e também em tráfego Internet, tráfego agregado, dentre outros [1][9][20][22][23][38][44][46][52][58][78].

No capítulo 4 deste estudo, através de medidas em traços de tráfego coletados em uma rede real, foi verificado que esse fenômeno ocorre em diversas situações em redes de telecomunicações.

Como modelos de processos estocásticos auto-similares podemos citar o Ruído Gaussiano Fracionário (FGN) e o processo ARIMA Fracionário que serão apresentados na sessão 3.4.5.2.

Os processos estocásticos auto-similares têm sido definidos de várias formas na literatura. A seguir, faremos uma breve descrição do conceito de processos auto-similares, para maior detalhamento vide as referências [23][36][52][58][83][84].

A figura 21 mostra o cenário onde o processo auto-similar $X(t)$ ocorre nas redes ATM, o maior impacto é causado nos “*buffers*” dos comutadores da rede.

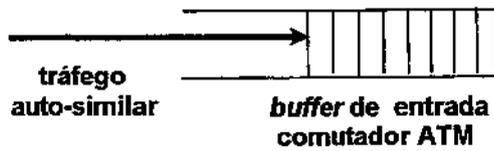


Figura 21: cenário hipotético

Dizemos que um processo é estatisticamente auto-similar em estrito senso se os processos estocásticos $\{X(at), t \geq 0\}$ e $\{a^H \cdot X(t), t \geq 0\}$ são equivalentes estatisticamente para todo $a > 0$ e algum $0 < H < 1$, onde H é chamado parâmetro de auto-similaridade ou parâmetro de Hurst. Da mesma forma dizemos que $\{X(t), t \geq 0\}$ e $\{a^{-H} \cdot X(at), t \geq 0\}$ são auto-similares se eles forem estatisticamente equivalentes. Geralmente $X(t)$ é um processo não estacionário apresentando as seguinte características conforme o valor de H :

- para $\frac{1}{2} < H < 1$, neste caso $\sum_{\tau=-\infty}^{\infty} \rho(\tau) = \infty$, ou seja, a soma da função de correlação tende para infinito decaindo muito lentamente para zero, isto demonstra que o processo possui longa memória;
- para $H = \frac{1}{2}$, o processo é considerado decorrelacionado;
- para $0 < H < \frac{1}{2}$, tem-se $\sum_{\tau=-\infty}^{\infty} \rho(\tau) = 0$, ou seja, a soma da função de correlação é zero, neste caso é dito que o processo possui dependência de curto prazo [36].

A propriedade de dependência de longo prazo em processos auto-similares foi chamada por Mandelbrot de “*Joseph Effect*”, referindo-se aos sete anos de fartura seguidos de sete anos de fome previstos por José no Antigo Testamento.

A fim de se obter tratabilidade matemática assumimos que $X(t)$ é um processo estacionário no sentido amplo (WSS), com média μ , variância σ^2 e função de autocorrelação $\rho(\tau)$, tendo:

$$\rho(\tau) \rightarrow \tau^{-\beta} L(\tau) \text{ com } \tau \rightarrow \infty \quad (57)$$

com $0 < \beta < 1$, $L(\tau)$ varia lentamente a medida que τ tende ao infinito, isto é,

$$\frac{L(zX)}{L(X)} = 1, \quad \forall X > 0 \quad (58)$$

Exemplos de funções com esta propriedade são:

$$L(\tau) = \text{constante} \text{ e } L(\tau) = \log \tau \quad (59)$$

A partir do processo $\{X(t), t \geq 0\}$ constrói-se $X^{(m)} = [X_t^{(m)} : t = 1, 2, 3, \dots]$, $m = 1, 2, 3, \dots$, a série temporal tomando-se médias da série original X sobre blocos de tamanho m não sobrepostos:

$$X_t^{(m)} = 1/m(X_{tm-m+1} + X_{tm-m+2} + \dots + X_{tm}) \quad (60)$$

para $t \geq 0$ e $m = 1, 2, 3, \dots$. A série temporal $X^{(m)}$ estabelece um processo estacionário no sentido amplo com função de autocorrelação $\rho^{(m)}(\tau)$.

Consideremos $X(t)$ como um processo auto-similar de segunda ordem com parâmetro de auto-similaridade $H = 1 - \beta/2$ se o processo agregado $X^{(m)}$ possuir estrutura idêntica de correlação de X , isto é, $\rho^{(m)}(\tau) = \rho(\tau)$ e a variância $Var[X^{(m)}] = \sigma^2 m^{-\beta}$, $\forall m = 1, 2, 3, \dots$. Em outras palavras, X possui auto-similaridade de segunda ordem se o processo agregado $X^{(m)}$ não se distingue de X em termos das estatísticas da primeira e segunda ordem. Uma outra forma equivalente de dizer que X é auto-similar, com parâmetro $H = 1 - \beta/2$, se:

$$\rho^{(m)}(\tau) = 2^{(1-\beta)} \text{ quando } m \rightarrow \infty \quad (61)$$

ou

$$\rho^{(m)}(\tau) \rightarrow (1/2)\delta^2 \tau^{(2-\beta)} \text{ quando } m \rightarrow \infty \quad (62)$$

onde δ^2 refere-se a função operadora de segunda ordem definida como:

$$\delta^2 f(x) = f(x+1) - 2f(x) + f(x-1) \quad (63)$$

Note que para m suficientemente grande, o processo agregado $X^{(m)}$ possui estrutura de correlação fixa determinada somente pelo valor de β .

Uma importante característica da auto-similaridade é que a estrutura do processo agregado não se degenera, isto é, $\rho^{(m)}(\tau) = 2^{(1-\beta)}$ quando $m \rightarrow \infty$. Um exemplo típico de processo auto-similar de segunda ordem é o Ruído Gaussiano Fracionário, com parâmetro $1/2 < H < 1$ [36][58].

Resumindo um processo auto-similar apresenta as seguintes características:

- a função de autocorrelação não é somável, decrescendo hiperbolicamente, indicando dependência de longo prazo;
- a variância de uma amostra média diminui lentamente, isto é, $Var[X^{(m)}] \approx bm^{-\beta}$ quando $m \rightarrow \infty$, com $0 < \beta < 1$, sendo b uma constante finita positiva [1];

- considerando um processo estacionário X , com função de autocorrelação $\rho(\tau)$ e função densidade espectral de potência $f(\lambda)$ que segue uma lei de potência próximo da origem (o espectro de potência diverge próximo a origem), temos: $f(\lambda) \approx c\lambda^{-\alpha}$, quando $\lambda \rightarrow 0$, com $0 < \alpha < 1$, $\alpha = 1 - \beta$ onde $H = 1 - \beta$, sendo c uma constante positiva.

O decaimento lento das variâncias e o decaimento é hiperbólico da função de autocorrelação provocam o “*Noah Effect*”, característico de distribuições com cauda pesada (*heavy-tailed*). Processos com distribuições “*heavy tailed*” possuem alta variabilidade e convergência lenta. A distribuição de Pareto é um exemplo de distribuição “*heavy tailed*”, conforme ilustrado na figura 22 em comparação com a distribuição exponencial.

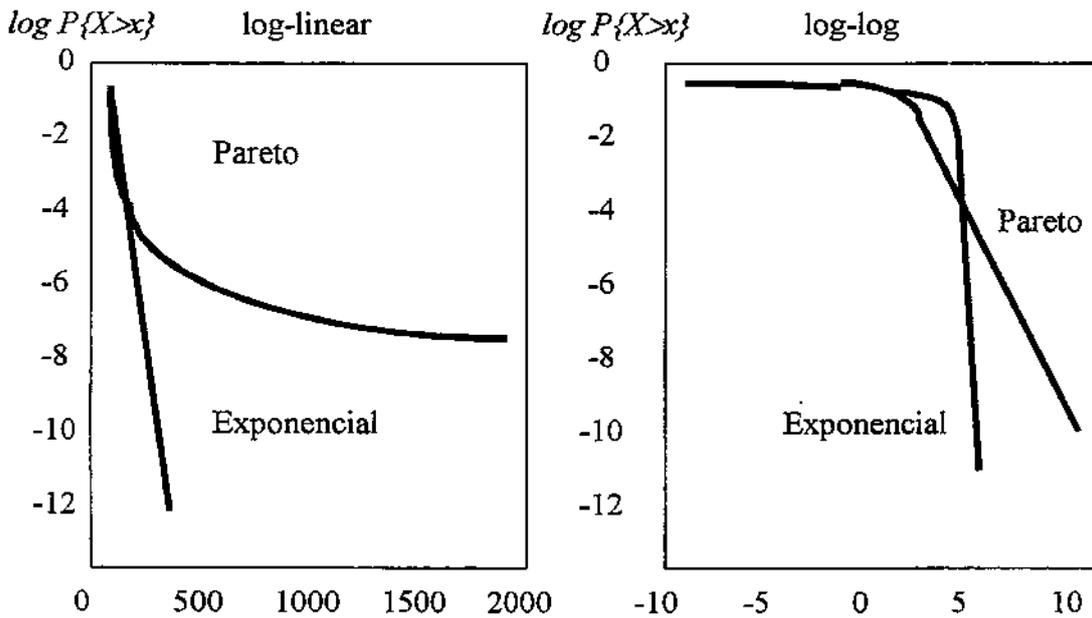


Figura 22: Decaimento Exponencial x Hiperbólico

Note que a distribuição exponencial possui um decaimento exponencial $P\{X > x\} \approx \exp(-x)$ (*light tails*) enquanto que a distribuição de Pareto possui um decaimento hiperbólico $P\{X > x\} \approx x^{-\alpha}$ (*heavy tails*), sendo que para:

- $\alpha > 2$: média finita e variância finita;
- $1 < \alpha \leq 2$: média finita e variância infinita;
- $0 < \alpha \leq 1$: média e variância infinitas.

3.4.5.1 Estimando o parâmetro de Hurst

Vários métodos para estimar o parâmetro H já foram desenvolvidos, como Análise das Variâncias, Análise da Estatísticas R/S, Periodograma, Correlograma, Método de Higuchi, Método de Whittle e através de transformadas “Wavelets”.

Neste trabalho adotamos o Método das Variâncias para a estimação do parâmetro de Hurst [58]. A relação entre a variância e o tamanho do bloco é:

$$Var(X^{(m)}) \sim am^{-\beta} \text{ com } m \rightarrow \infty, \quad (64)$$

equivalentemente

$$\log(Var(X^{(m)})) \sim -\beta \log(m) + \log(a) \text{ com } m \rightarrow \infty. \quad (65)$$

Podemos tomar uma estimativa de β calculando $\log(Var(X^{(m)}))$ para vários valores de m . Note que o valor estimado de β , β' , é obtido do valor negativo da inclinação da reta em (65). Portanto, o valor do parâmetro Hurst estimado $H' = 1 - \beta'/2$ [58]. Como exemplo, é mostrada a Figura 23 onde é estimado o valor de β' que é obtido da inclinação da reta r , para um arquivo de tráfego Ethernet.

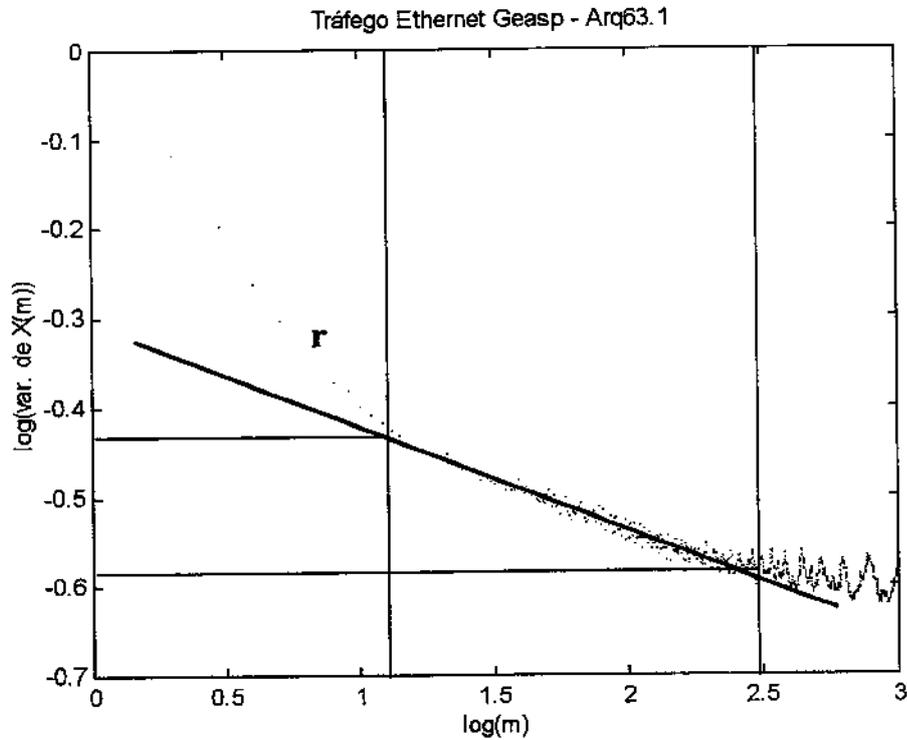


Figura 23: Método das Variâncias

Neste caso $\beta \approx 0,1163$ e como consequência o H estimado $H \approx 0,94$.

Finalmente na Tabela 3, está colocado um resumo das principais diferenças entre um processo estocástico convencional e um processo estocástico auto-similar [45].

<i>Principais Diferenças entre Processo Convencional e Processo Auto-Similar</i>	
<i>Processo Convencional</i>	<i>Processo Auto-Similar</i>
escalas de tempo limitadas	grande número de escalas de tempo
estatísticas bem definidas	3 estatísticas definidas
decaimento rápido da variância	decaimento lento da variância
dependência de curto prazo	dependência de longo prazo
espectro de potência plano próximo a origem	espectro de potência diverge próximo a origem ($1/f$ noise)
dimensão integral	dimensão fractal

Tabela 3: diferenças entre processos

Nas duas próximas seções serão apresentados dois processos auto-similares utilizados na modelagem de tráfego agregado, o Ruído Gaussiano Fracionário e o ARIMA Fracionário.

3.4.5.2 Movimento Browniano Fracionário (fBm - Fractional Brownian Motion) e Ruído Gaussiano Fracionário (FGN - Fractional Gaussian Noise)

Suponha que $\{Y_t\}$ é um processo auto-similar com incrementos estacionários. O processo $\{Y_t\}$ é estatisticamente auto-similar se $\{Y_t(at), t \geq 0\}$ possuir a mesma distribuição finita de $\{a^H Y_t(t), t \geq 0\}$, para todo $a > 0$, sendo H o parâmetro de auto-similaridade. Considere o processo de incrementos estacionários $X_i = Y_i - Y_{i-1}$, ($i = 1, 2, \dots$) com valor esperado igual a zero. Em particular suponha um X_t processo Gaussiano, então sua distribuição é totalmente especificada pela média e covariância. Desta forma para cada valor de $H \in (0,1)$ existe um processo Gaussiano X_i que é o incremento estacionário do processo auto-similar Y_t . O processo de incrementos é chamado de Ruído Gaussiano Fracionário (FGN) e o correspondente processo auto-

similar Y_t é denominado Movimento Browniano Fracionário (fBm) e denotado por $B_H(t)$ [36]. A covariância entre X_i e X_{i+k} para $k > 0$ é dada:

$$\gamma(k) = \text{cov}(X_i, X_{i+k}) = \text{cov}(X_I, X_{k+I}) \quad (66)$$

$$\gamma(k) = \frac{1}{2} \sigma^2 \left[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H} \right] \quad (67)$$

para $k \geq 0$ e $\gamma(k) = \gamma(-k)$ para $k < 0$. A função de correlação é dada pela equação:

$$\rho(k) = \frac{1}{2} \left[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H} \right] \quad (68)$$

para $k \geq 0$ e $\rho(k) = \rho(-k)$ para $k < 0$ [36].

Ambas as funções $\gamma(k)$ e $\rho(k)$ possuem um comportamento assintótico que segue a fórmula da expansão de Taylor dada pela expressão (69)[36]:

$$g(x) = (1+x)^{2H} - 2 + (1-x)^{2H} \quad (69)$$

Quando k tende ao infinito a expansão de Taylor é igual $H(2H-1)x^2$, portanto para covariância (70) e função de correlação:

$$\gamma(k) = H(2H-1)k^{2H-2} \quad (70)$$

$$\frac{\rho(k)}{[H(2H-1)k^{2H-2}]} \rightarrow 1 \quad (71)$$

A transformada inversa de Fourier da densidade espectral de potência denotada por $f(\lambda)$ é a função de autocorrelação $\gamma(k)$, sendo $f(\lambda)$ dada por:

$$f(\lambda) = c \left(2 \text{sen} \frac{\lambda}{2} \right)^2 \sum_{k=-\infty}^{\infty} \left(\frac{1}{|\lambda + 2\pi k|^{2H+1}} \right) \quad (72)$$

onde c é uma constante e quando $\lambda \rightarrow 0$ $f(\lambda)$ pode ser aproximada por (73):

$$f(\lambda) = c |\lambda|^{1-2H} \quad (73)$$

O movimento Browniano fracionário normalizado, com parâmetro $H \in [1/2, 1)$, é um processo estocástico Y_t , $t \in (-\infty, \infty)$, caracterizado pelas seguintes propriedades:

- Y_t possui incrementos estacionários;
- $Y_0 = Y_t(0) = 0$ e $E\{Y_t\} = 0$ para todo t .
- $E\{Y_t^2\} = |t|^{2H}$ para todo t .
- Y_t é gaussiano, isto é, todas as suas distribuições marginais em termos de dimensão finitas são gaussianas [36][45][58].

A modelagem através do fBm pode ser representada pela equação (74) e fornece três parâmetros: taxa média, “*peakedness*”[12][76] e o parâmetro de Hurst. Considere uma quantidade de tráfego oferecida a um canal de comunicação em um intervalo de tempo $[s, t)$ denotada por $A(s, t)$, sendo que $s, t \in (-\infty, +\infty)$. Podemos representar $A(0, t)$ simplesmente por A_t [35].

$$A_t = mt + \sqrt{am}Y_t \quad (74)$$

- m - taxa média de tráfego ou volume de tráfego no canal de comunicação;
- a - fator de “*peakedness*” mede a magnitude das flutuações em torno da média;
- H - mede a intensidade das correlações entre pacotes.

A variação do “*burstiness*” sobre várias escalas de tempo é modelada por: “*peakedness*”[12] e pelo parâmetro H .

O FGN $\{X_i, i \geq 1\}$ é o processo de incrementos do Movimento Browniano ou uma versão derivada do Movimento Browniano Fracionário (fBm).

$$X_i = Y_t(i+1) - Y_t(i), \quad i \geq 1 \quad (75)$$

3.4.5.3 Processo ARIMA Fracionário - (FARIMA)

O modelo ARIMA Fracionário é uma generalização do ARIMA, que por sua vez, se baseia no modelo ARMA. O FARIMA possui a vantagem de caracterizar o

comportamento de longo prazo bem como o de curto prazo, sendo um modelo usado nos processos auto-similares.

A definição deste modelo foi proposta por Granger e Joyeux em 1980 e Hosking em 1981 [1][36].

O processo FARIMA (p, d, q) gaussiano é amplamente usado para modelar séries temporais. Um processo $\{X_n\}$ é dito um processo ARIMA (p, d, q) fracionário, se $\{X_n\}$ é estacionário e:

$$\phi(B)(1-B)^d X_n = \theta(B)U_n \quad (76)$$

onde U_n é um ruído branco qualquer e $\phi(B)$ e $\theta(B)$ são polinômios de graus p e q , para $-1/2 < d < +1/2$. Note que a diferença entre o ARIMA e o ARIMA Fracionário está no parâmetro d , que no último caso é um número real restrito ao intervalo $(-1/2, 1/2)$.

O intervalo de interesse para processos com dependência de longo prazo é apresentado em $0 < d < +1/2$ [36].

A covariância e função de autocorrelação de um processo FARIMA(0, d , 0) são dadas pelas equações (77) e (77a) respectivamente:

$$\gamma(k) = \sigma^2 \frac{(-1)^k \Gamma(1-2d)}{\Gamma(k-d+1)\Gamma(1-k-d)} \quad (77)$$

$$\rho(k) = \frac{\Gamma(1-d)\Gamma(k+d)}{\Gamma(d)\Gamma(k+1-d)} \quad (77a)$$

Quando k tende para o infinito, temos:

$$\rho(k) \approx \frac{\Gamma(1-d)}{\Gamma(d)} |k|^{2d-1} \quad |k| \rightarrow \infty \quad (77b)$$

onde Γ é a função Gama.

Para o processo FARIMA(p, d, q) genérico, a função de autocorrelação e covariância podem ser obtidas passando-se um processo FARIMA(0, d , 0) através de um filtro linear [36].

Comparando o FGN e FARIMA, pode-se concluir que o primeiro possui a vantagem que muitas relações assintóticas mantêm-se em amostras de tamanho finito, porque o FGN é o incremento do fBm. Por outro lado o FARIMA apresenta a vantagem de possuir particularmente uma densidade espectral de potência simples [58].

Os modelos Auto-Similares fBm e FARIMA são adequados para modelar tráfego agregado. Para uma conexão individual o modelo On/Off é considerado satisfatório.

3.5 Modelos para Escala em Conexão

Nesta escala de tempo, associado está o conceito de banda efetiva requerida para uma determinada conexão, que é a quantidade de banda necessária para estabelecer e manter a conexão enquanto ela existir, dentro dos padrões de Qualidade de Serviço negociada entre o usuário da rede e o provedor.

Existem muitos estudos que demonstram como estimar o valor mais apropriado de banda efetiva. Grande parte desses estudos estão fundamentados em probabilidade de perda de células, como o apresentado por Kesidis [28] ou no retardo considerando sistemas sem perda, como o método proposto por Le Boudec [40].

Este trabalho pretende apresentar uma metodologia alternativa para a estimativa de banda efetiva a ser alocada para uma conexão ATM. Esta metodologia detalhada no capítulo 5, procura combinar métodos analíticos, ferramenta de simulação e métodos heurísticos. Na metodologia proposta neste trabalho, serão combinadas as soluções que consideram atraso e perda de células. Adicionalmente será introduzido o parâmetro de auto-similaridade H ao método de estimação da banda efetiva.

3.6 Conclusões

Neste capítulo apresentamos resumidamente diversos modelos de tráfegos bastante difundidos na literatura, mostrando em que escala de tempo cada modelo está representado e suas respectivas aplicações no modelamento de tráfego de voz, dados e vídeo.

Também foram apresentados os conceitos do fenômeno da Auto-Similaridade, que é uma característica já comprovada em diversos tipos de tráfego das redes emergentes causando importante impacto no dimensionamento e análise de desempenho dessas redes.

Entre modelos auto-similares, foram apresentados o FGN e o FARIMA, que são particularmente importantes modelagem de tráfego agregado.

Também destacamos o tradicional modelo On/Off, Fluido Markoviano e os modelos Auto-Similares, que serão utilizados no trabalho desenvolvido no Capítulo 5 para estimar banda efetiva necessária em uma conexão ATM, para um determinado tipo de tráfego que apresente a característica auto-similar ou não.

Capítulo 4

CARACTERIZAÇÃO DO TRÁFEGO

4.1 Introdução

A crescente competição no mundo dos negócios tem exigido cada vez mais a otimização dos recursos das empresas. Dentre os diversos recursos existentes, a informação tem se tornado preponderante sob os pontos de vista estratégico e de competição. O tratamento adequado da informação requer o uso otimizado de recursos de processamento e de transporte da mesma. Sob o ponto de vista do transporte da informação, a utilização dos recursos de banda disponíveis é de grande importância e possui impacto considerável nos resultados. Outro aspecto a se considerar é que aliado a esse fato, o desenvolvimento de novos recursos computacionais e de telecomunicações nos últimos anos, têm possibilitado uma evolução das aplicações e das necessidades empresariais. Esses dois fatores têm mudado as características do tráfego principalmente em ambientes corporativos, fazendo com que haja cada vez mais necessidade de se conhecer essas características de maneira a proporcionar um melhor planejamento e dimensionamento das novas redes.

Considerando o perfil de tráfego existente nas redes atuais de comunicação, sejam estas públicas ou privadas, o mesmo é composto basicamente de tráfego de voz, tráfego de dados (em redes Frame-Relay, IP, Decnet dentre outras) e tráfego de vídeo e imagens (com ou sem utilização de compressão).

Como o comportamento do tráfego de voz está muito bem estudado, neste capítulo vamos nos concentrar na caracterização do tráfego de dados e tráfego de vídeo. Vale a pena ressaltar que inúmeros estudos têm demonstrado comportamento auto-similar com dependência de longo prazo nestes dois tipos de tráfego.

Os tráfegos de dados coletados para este estudo foi obtido da rede corporativa da *PETROBRAS*, que pode ser considerada de grande porte, abrangendo diferentes tipos de tráfego como: Frame-Relay, Gerência SNMP, Automação, Ethernet, e VSAT.

O tráfego de vídeo foi extraído de arquivos de vídeo comprimido em padrão MPEG e AVI. Para maiores detalhes sobre os padrões MPEG e AVI sugerimos consultar o Apêndice D e referências bibliográficas [1][39][59][74].

Este capítulo tem por finalidade apresentar a caracterização do tráfego de dados e de vídeo, estudando e apresentando quais parâmetros de tráfego que são mais significativos para o estudo de dimensionamento de banda. Os parâmetros de desempenho relevantes serão usados para obter a banda efetiva de uma conexão ATM. A descrição detalhada do cálculo da banda efetiva será apresentada no Capítulo 5.

4.2 Parâmetros para Caracterização de Tráfego

O tráfego de dados e de vídeo com compressão, são classificados como de taxa variável, ou seja, caracterizados pelo conceito de surto ou rajada (*burstiness*), que interfere de forma crucial na análise de tráfego.

Existem diversas formas de medir a intensidade de um surto, neste item apresentamos os seguintes parâmetros alternativos para medida de surto: *PRM* (*Peak to Mean Ratio*), *CSQ* (*Square Coefficient of Variation*), *IDC* (*Index of Dispersion of Counts*), *MBS* (*Maximum Burst Size*) e o parâmetro *H* (*Hurst Parameter*), que são detalhados a seguir.

O primeiro parâmetro analisado, o mais simples, é o chamado coeficiente de surto, dado pela razão entre a taxa de pico e a taxa média, também chamado relação pico pela média (*PRM - Peak to Mean Ratio*)[9][76]. O *PRM* leva em conta somente as propriedades estatísticas de primeira ordem do tráfego não considerando os tempos entre as chegadas de pacotes.

Seja um processo aleatório X , $\{X = X(t), t \in \mathcal{R}, t > 0\}$, que representa a intensidade da taxa de chegada de pacotes no instante t . Seja \bar{X} o valor médio definido pela equação (78) do processo X e X_m , o valor máximo do mesmo processo.

$$\bar{X} = \lim_{T \rightarrow \infty} \frac{1}{T} \int X(t) dt \quad (78)$$

A partir de \bar{X} e X_m o coeficiente de surto desse processo é dado por:

$$\beta = \frac{X_m}{\bar{X}} \quad (79)$$

O segundo parâmetro analisado, é o coeficiente de variabilidade dado pela razão entre o desvio padrão (σ_{Y_n}) e o valor médio (\bar{Y}_n) do tempo entre chegadas consecutivas de pacotes Y_n [76].

$$C_v = \frac{\sigma_{Y_n}}{\bar{Y}_n} \quad (80)$$

O terceiro parâmetro é equivalente ao C_v , é o coeficiente quadrático de variabilidade (*CSQ - Square Coefficient of Variation*). Esse parâmetro é representado como [9]:

$$CSQ = \frac{\sigma_{Y_n}^2}{\bar{Y}_n^2} \quad (81)$$

É importante notar que estes três parâmetros não levam em consideração a dependência temporal.

O Índice de Dispersão, *IDC (Index of Dispersion of Counts)*, representado pela equação (82), que é a razão da variância pela média do processo de contagem $N(\tau)$, leva em conta a dependência temporal no tráfego. Para um determinado intervalo de comprimento τ o índice de dispersão é a razão entre a variância e o número médio de chegadas [9][76].

$$IDC(\tau) = \frac{Var\{N(\tau)\}}{E\{N(\tau)\}} \quad (82)$$

O parâmetro de Hurst, já detalhado no Capítulo 3, mede o grau de auto-similaridade de um processo estocástico $X(t)$, podendo ser observado em várias escalas de tempo. Esse processo é caracterizado pela dependência de longo prazo ou memória longa com forte correlação, decaimento hiperbólico da função de autocorrelação e decaimento lento das variâncias [1][36].

Além dos parâmetros de caracterização de tráfego já apresentados, introduzimos também neste estudo o parâmetro concebido pelo ATM Forum, *MBS (Maximum Burst Size)*[10][15][66], que é equivalente ao *BT (Burst Tolerance)*. A relação entre os dois parâmetros é dada por (83) que é a mesma expressão (1) já apresentada no Capítulo 2. De fato, o parâmetro *BT* é equivalente ao *IBT (Intrinsic Burst Tolerance)* adotado pelo ITU-T. O *MBS* é o número de células que podem ser enviadas continuamente na taxa de pico e o *BT* é o período de tempo em que é permitida a fonte transmitir na taxa de pico.

$$BT = (MBS - 1) \left(\frac{1}{SCR} - \frac{1}{PCR} \right) \quad (83)$$

4.3 Caracterização do Tráfego de Dados

4.3.1 Ambiente Corporativo

Nesta sessão são apresentadas informações resumidas da rede corporativa da *PETROBRAS*, de onde os arquivos tráfego de dados foram capturados.

Basicamente os dados corporativos trafegam em redes locais e na rede de longa distância do tipo Frame-Relay (FR), suportando protocolos como IP e Decnet. Essa rede suporta diversas aplicações corporativas, operacionais e a Intranet da empresa.

A figura 24 mostra um diagrama da rede FR, onde são salientados os nós situados no Rio de Janeiro e em São Paulo, que são interligados por um circuito virtual de 1.344 Kbps. Este circuito virtual suporta o maior tráfego da rede FR com a utilização média de

50%. Na figura 24 a seta mostrada aponta para o nó de São Paulo onde foram capturados os arquivos de tráfego.

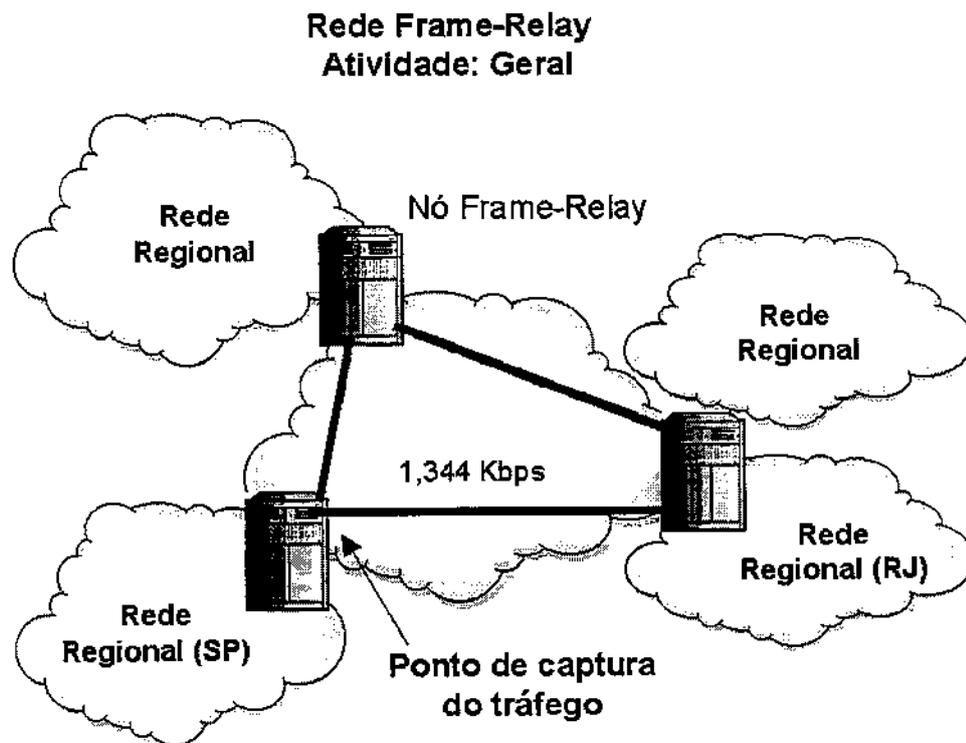


Figura 24: Tráfego Frame-Relay

Os dados de automação da atividade de refino de petróleo são controlados e processados localmente, através de redes locais. O dados de automação da atividade de produção de petróleo são controlados e processados remotamente através da rede corporativa ou através de redes dedicadas. Uma outra possibilidade de controle e processamento de dados de automação remota é por meio de uma rede VSAT como é o caso da atividade de transporte, que é responsável pela operação dos poldutos.

A figura 25 mostra a topologia de uma rede VSAT regional de São Paulo, que é apenas um dos segmentos da rede VSAT nacional. As estações da rede VSAT são interligadas e monitoradas por uma estação mestre regional por meio de circuitos de 9.600 bps, utilizando o protocolo X.25. A utilização média desse segmento da rede VSAT é de

25% para DCE (*Data Communication Equipment*) e de 5% para o DTE (*Data Terminal Equipment*), com picos esporádicos de 90% e 18% respectivamente.

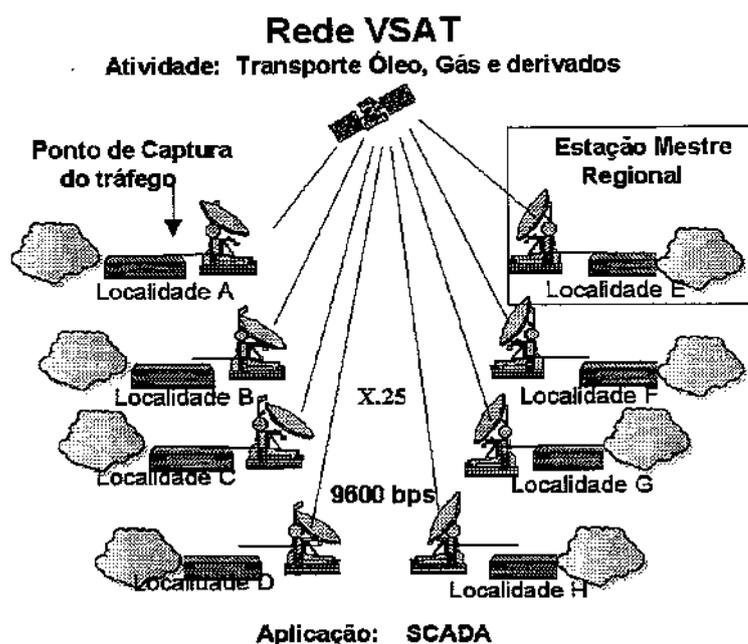


Figura 25: Tráfego VSAT Regional

A automação industrial na atividade de refino é suportada por rede local tipo Ethernet. Um “*switch Ethernet*” separa os diversos segmentos de automação em barramentos de 10 Mbps. A rede em análise possui mais de 40.000 pontos de monitoração de processos. A taxa média de utilização do barramento está em torno de 8%. O detalhamento da topologia da rede de automação industrial é mostrado na figura 26.

Neste estudo também foram analisados algumas amostras de tráfego capturados em rede local tipo Ethernet padrão 802.3 em ambiente administrativo. Essa rede local é composta de “*hubs*” e “*switches*” que distribuídos em um edifício administrativo de 17 andares. Os servidores da rede são ligados em “*Fast Ethernet*” a 100 Mbps.

Outro importante tipo de tráfego analisado foi o tráfego do ambiente *PETROBRAS* com a Internet. A rede corporativa se liga a Internet através de sete “*Firewalls*” localizados em sete diferentes pontos geográficos. O tráfego analisado foi capturado no “*Firewall*” de maior utilização da empresa situado no Rio de Janeiro.

Finalmente foram analisados tráfego de gerência SNMP (*Simple Network Management Protocol*) utilizado na gerência remota de dispositivos de rede como “Hubs”, “Switches” e roteadores. O sistema gerente é composto de plataforma e “software” de gerência de rede, responsável em monitorar e armazenar informações de falhas, configuração e desempenho dos elementos de rede que possuem as informações armazenadas em suas respectivas MIBs (*Management Information Base*). As informações são colhidas através de requisições do gerente e respostas dos agentes (elementos de rede). O tráfego SNMP foi capturado no sistema de gerência localizado em São Paulo.

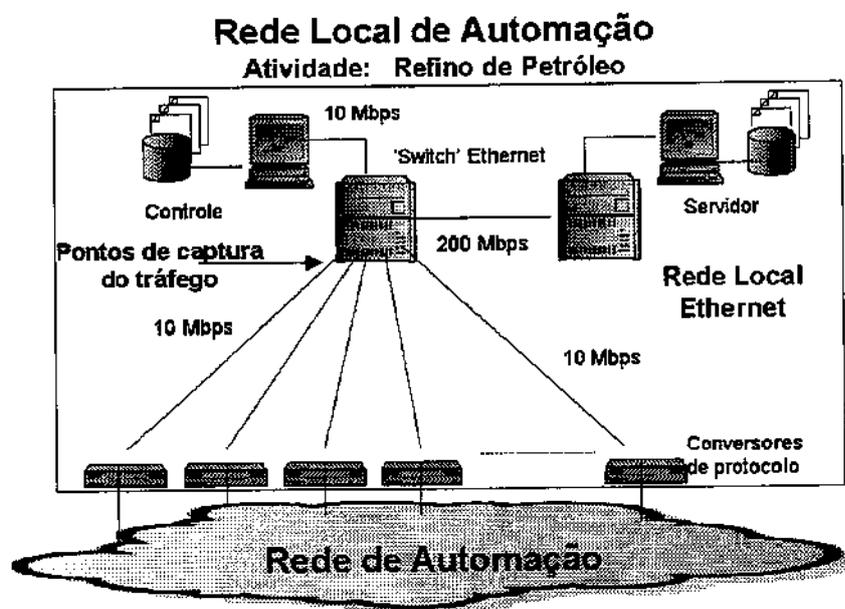


Figura 26: Tráfego de Automação Industrial

4.3.2 Medidas e Resultados dos Parâmetros Tráfego de Dados

Os arquivos de tráfego utilizados neste estudo foram capturados através dos analisadores de protocolos LAN e WAN modelos DA-310 e DA320 da Wandel & Golterman™, mais detalhes sobre a instrumentação utilizada estão disponíveis no Apêndice F.

A tabela 4 apresenta os resultados consolidados das medidas do parâmetro H , para os diversos tipos de tráfego. O método usado para estimação do H foi o método das

variâncias [1][36][44][58][78]. Os resultados mostrados na tabela 4 estão separados em dois tipos: sem agregação e com agregação de 1ms, 5ms, 10ms e 20 ms para os mesmos arquivos de tráfego. A agregação consiste em aglutinar os pacotes que estão dentro de um mesmo intervalo de tempo predeterminado. Esta técnica foi utilizada por Ashok e Wang [9] para análise dos parâmetros H e PRM e serve para confirmar a característica do tráfego em diferentes escalas.

Arquivo de Tráfego	H sem Agregação			H com Agregação			
	Hméd	Hmín	Hmáx	1 ms	5 ms	10 ms	20 ms
Internet	0,65	0,64	0,65	0,66	0,69	0,70	0,69
Gerência	0,67	0,64	0,7	0,66	0,685	-	-
Automação	0,84	0,78	0,89	0,82	0,78	0,80	0,81
VSAT	0,61	0,58	0,67	0,57	0,58	0,61	0,64
Frame-Relay	0,68	0,54	0,80	0,63	0,69	0,76	0,78
Ethernet	0,78	0,64	0,87	0,828	0,856	-	-

Tabela 4: Parâmetro H sem e com agregação

A tabela 5 mostra as medidas de PRM para os diferentes tipos de tráfego, sem e com agregação, respectivamente.

Arquivo de Tráfego	PRM sem Agregação			PRM com Agregação			
	Méd.	Mín.	Máx.	1 ms	5 ms	10 ms	20 ms
Internet	8,147	6,52	9,7	3,867	6,54	7,21	4,98
Gerência	21,51	17,99	25,03	9,205	7,23	6,205	4,56
Automação	28,78	10,04	62,6	5,68	7,84	7,92	7,25
VSAT	289,9	196	401	5,65	6,154	6,24	6,37
Frame-Relay	154,8	99,16	215,5	10,32	6,70	5,47	4,45
Ethernet	4,49	2,78	8,59	4,44	3,82	4,28	4,87

Tabela 5: Parâmetro PRM sem e com agregação

A tabela 6 apresenta os resultados de CSQ (*Square Coeficient of Variation*).

Arquivo de Tráfego	CSQ		
	Méd.	Mín.	Máx.
Internet	2,43	2,26	2,61
Gerência	3,315	2,16	4,47
Automação	2008	1,44	10035
VSAT	144,1	15,48	644,4
Frame-Relay	0,62	0,54	0,74
Ethernet	16,56	11,01	27,38

Tabela 6: Parâmetro CSQ

A tabela 7 apresenta o resultado obtido em termos de *IDC* (*Index of Dispersion of Counts*)

Arquivo de Tráfego	IDC		
	Méd.	Mín.	Máx.
Internet	36,24	13,72	61,28
Gerência	94,48	61,9	127,1
Automação	70,88	22,86	174,5
VSAT	100,9	3,84	469,4
Frame-Relay	1,45	1,34	1,61
Ethernet	496,3	322,2	768,4

Tabela 7: Parâmetro IDC

A tabela 8 apresenta o resultado obtido em termos de *MBS*.

Arquivo de Tráfego	MBS		
	Méd.	Mín.	Máx.
Internet	9	8	11
Gerência	6	6	6
Automação	5	3	9
VSAT	0,16	0,15	0,18
Frame-Relay	9	7	12
Ethernet	15	13	18

Tabela 8: Parâmetro MBS

Observe que encontramos valores de *MBS* menores que a unidade somente no tráfego da rede VSAT. Isto ocorre devido a característica da aplicação de automação e os tempos envolvidos nos processos bem como o tempo de propagação via satélite, tornando o tempo mínimo entre pacotes bastante longo comparado com as outras redes.

Note que os resultados apresentados nas tabelas 4, 5, 6, 7 e 8, são resultados gerais por tipo de tráfego. Estas tabelas mostram os valores médios, máximos e mínimos por tipo de tráfego encontrados.

As tabelas 9, 10, 11, 12, 13 e 14 apresentam os resultados dos parâmetros analisados para cada um dos arquivos capturados por tipo de tráfego.

Arquivo de Tráfego	Internet				
	<i>H</i>	<i>PRM</i>	<i>CSQ</i>	<i>IDC</i>	<i>MBS</i>
Arq12	0,66	6,52	2,61	33,7	11
Arq10	0,65	8,22	2,26	61,2	8
Arq11	0,64	9,70	2,44	13,7	10

Tabela 9: Tráfego Internet

Arquivo de Tráfego	VSAT				
	<i>H</i>	<i>PRM</i>	<i>CSQ</i>	<i>IDC</i>	<i>MBS</i>
Arq14	0,67	196	15,4	5,76	0,18
Arq13	0,59	198	10,0	6,68	0,16
Arq16	0,58	401	26,1	18,8	0,15
Arq17	0,58	308	24,0	469	0,15
Arq15	0,63	346	644	3,84	0,17

Tabela 10: Tráfego VSAT

Arquivo de Tráfego	Automação				
	<i>H</i>	<i>PRM</i>	<i>CSQ</i>	<i>IDC</i>	<i>MBS</i>
Arq20	0,89	10,0	2,47	41,2	8
Arq19	0,80	11,4	10035	37,9	9
Arq22	0,78	62,6	1,44	22,8	3
Arq23	0,81	40,3	1,57	174	4
Arq21	0,80	19,6	1,98	77,2	5

Tabela 11: Tráfego de Automação em Rede Local

Arquivo de Tráfego	Ethernet				
	<i>H</i>	<i>PRM</i>	<i>CSQ</i>	<i>IDC</i>	<i>MBS</i>
Arq62	0,84	5,05	27,4	768	17
Arq63	0,87	3,89	11,0	463	18
Arq64	0,79	8,35	17,4	556	14
Arq65	0,81	6,56	13,6	332	13
Arq66	0,64	2,72	13,5	371	13

Tabela 12: Tráfego de Rede Local Ethernet

Arquivo de Tráfego	Gerência				
	<i>H</i>	<i>PRM</i>	<i>CSQ</i>	<i>IDC</i>	<i>MBS</i>
Arq68	0,64	17,9	2,16	61,9	6
Arq18	0,60	25,0	4,47	127	5

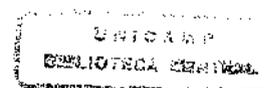
Tabela 13: Tráfego de Gerência de Rede

Arquivo de Tráfego	Frame-Relay				
	<i>H</i>	<i>PRM</i>	<i>CSQ</i>	<i>IDC</i>	<i>MBS</i>
Arq49	0,59	132	1,42	1,63	11
Arq24	0,54	215	1,61	19,5	7

Tabela 14: Tráfego Frame-Relay

O critério de desempenho para avaliar qual parâmetro possui maior relevância, foi a perda de células, que possui grande impacto no dimensionamento dos comutadores.

O método adotado para avaliar os parâmetros de tráfego apresentados neste capítulo foi através de simulação. Usando os arquivos de tráfego reais, foi medida a perda de células, para uma determinada banda e tamanho de “buffer”. Os detalhes sobre os arquivos de tráfego utilizados e do simulador que representa um “buffer” de um comutador ATM, estão disponíveis no Apêndice B e E respectivamente.



Assim sendo, usando os arquivos de tráfego que constam nas tabelas 9, 10, 11, 12, 13 e 14 como entrada, configurando tamanho de “buffer” na faixa de 5 a 10.000 células, assumindo uma banda utilizada de 150.000 bytes/segundo. Obtemos os seguintes resultados apresentados nas figuras 27(Internet), 28(VSAT), 29(Automação), 30(Ethernet), 31(Gerência) e 32(FR), que mostram as taxas de perda obtidas em função do tamanho de “buffer” e banda utilizada, com relação ao respectivo arquivo e tipo de tráfego.

Note que o tamanho de “buffer” e banda passante utilizados, foram aqueles que melhor se adaptaram em função do tamanho de cada arquivo, após vários testes efetuados experimentalmente. Observe que somente a única banda diferente de 150000 bytes/segundo foi a do tráfego VSAT que foi de 1200 bytes/segundo, em função das características desse tipo de tráfego e do tamanho do arquivo capturado.

Convém salientar que as faixas de banda e tamanho de “buffer” utilizadas neste estudo também foram utilizadas por Patrick R. Morin e J. Neilson da Universidade de Carleton, Canadá [43], em estudo de tráfego Ethernet da rede do Bellcore.

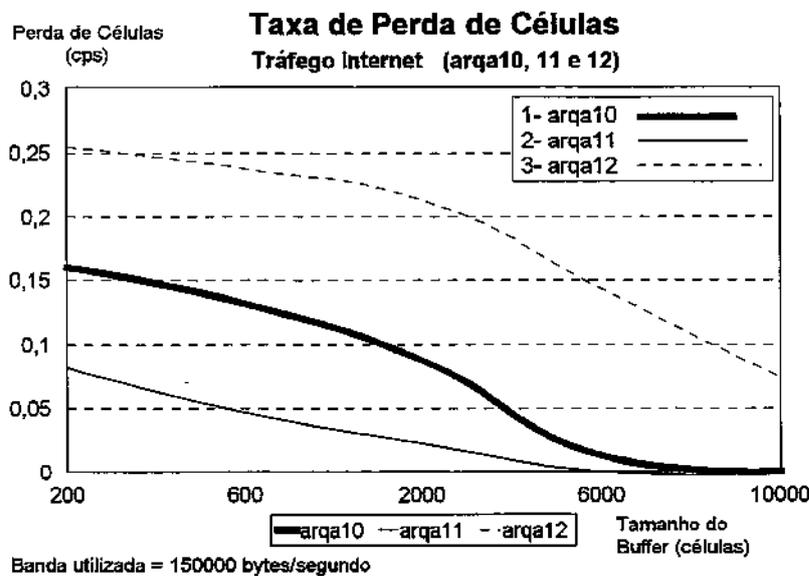


Figura 27: Taxa de Perda para tráfego Internet

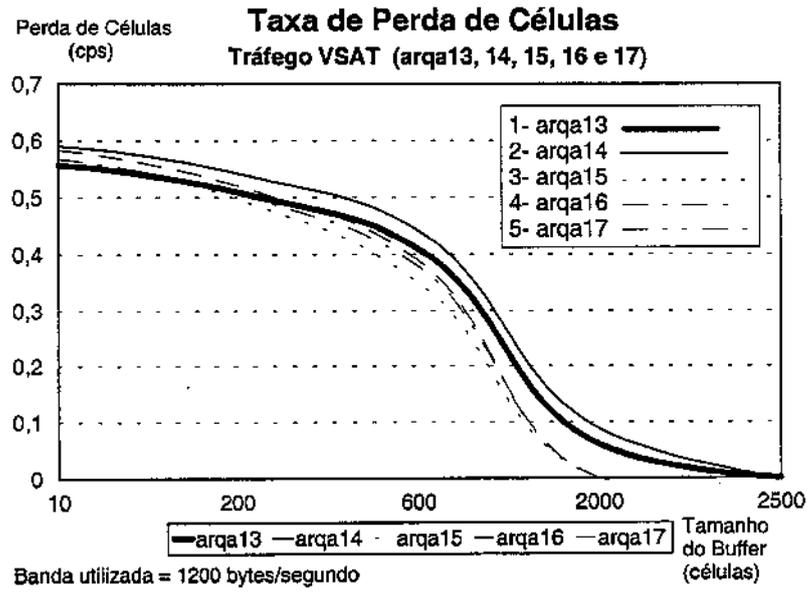


Figura 28: Taxa de Perda para tráfego VSAT

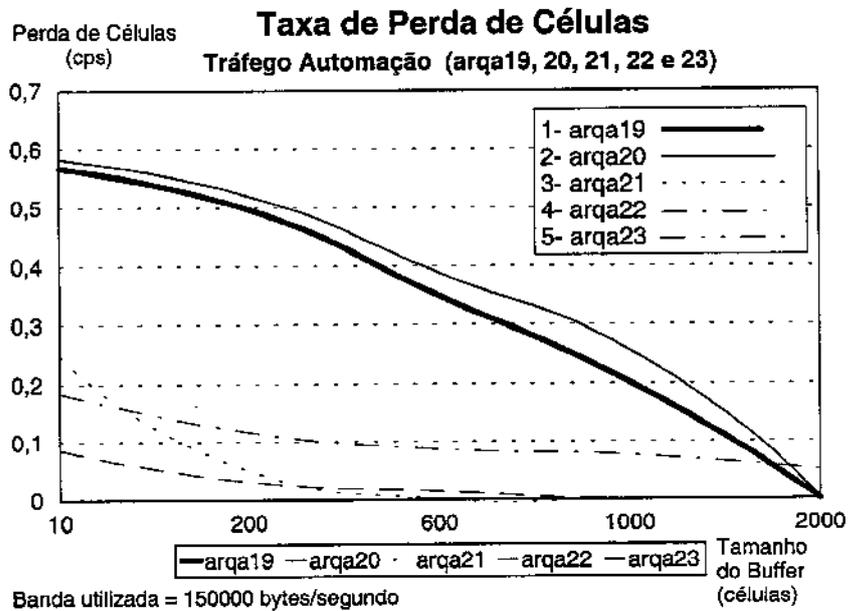


Figura 29: Taxa de Perda para tráfego de Automação

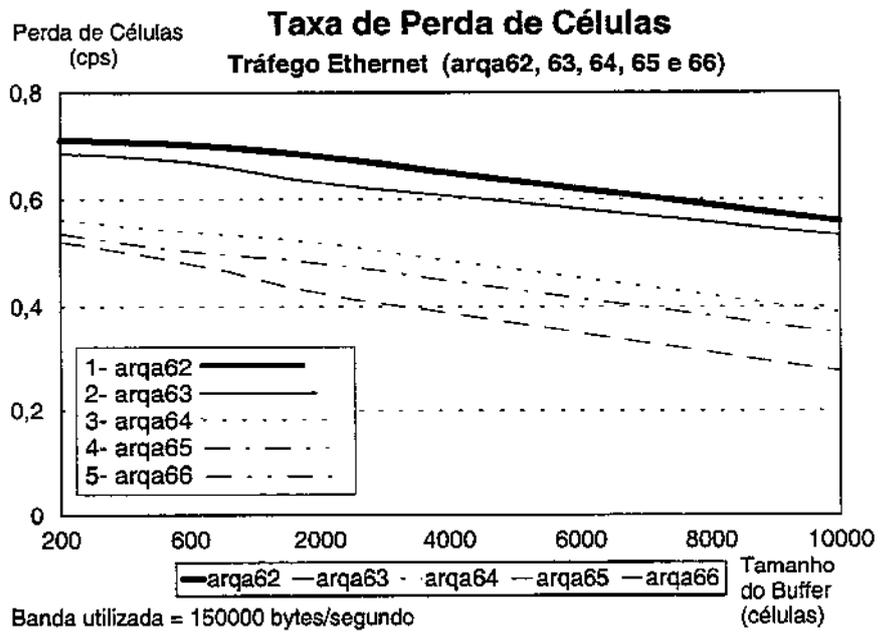


Figura 30: Taxa de Perda para tráfego Ethernet

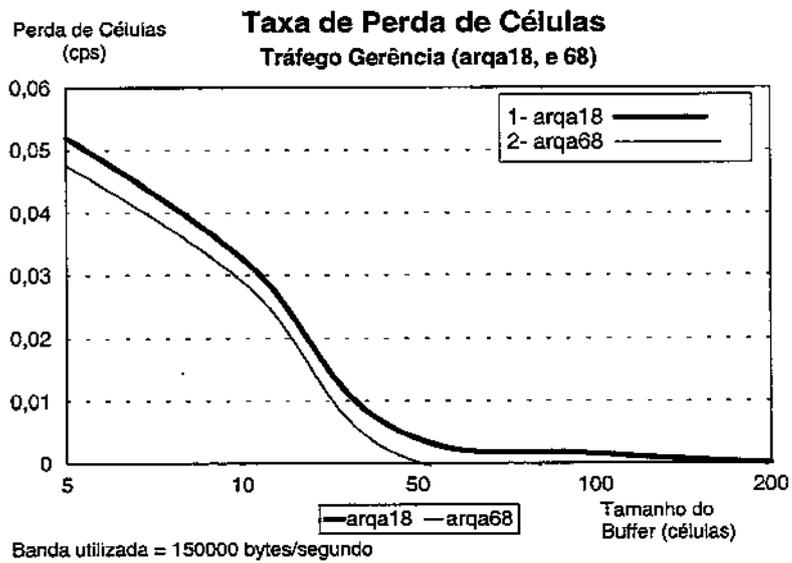


Figura 31: Taxa de Perda para tráfego de Gerência

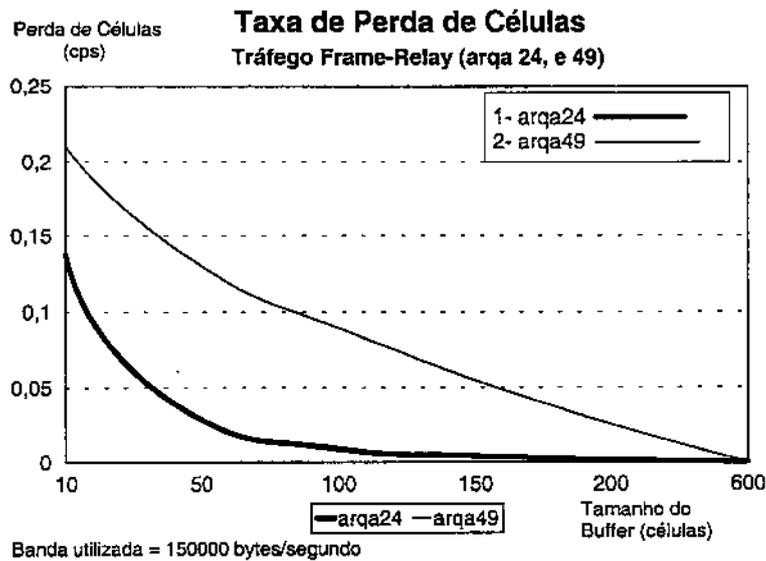


Figura 32: Taxa de Perda para tráfego Frame-Relay

Finalmente os cálculos dos parâmetros de tráfego apresentados neste capítulo, como: *PRM*, *CSQ*, *IDC*, *MBS* e parâmetro *H*, apresentados nas tabelas 9, 10, 11, 12, 13 e 14 foram comparados com os resultados de perda de células obtidos da simulação e apresentados nas figuras 27, 28, 29, 30, 31 e 32.

4.3.3 Conclusões sobre a Caracterização do Tráfego de Dados

Observando-se as variações dos valores calculados dos parâmetros de tráfego obtidos e apresentados nas tabelas 9, 10, 11, 12, 13 e 14 e comparando o comportamento dos resultados da simulação apresentados nas figuras 27, 28, 29, 30, 31 e 32 para os respectivos arquivos de tráfego, nota-se que os arquivos que apresentam maior taxa de perda de células são aqueles que possuem maior parâmetro *H* e *MBS* na grande maioria dos casos. Os parâmetros como *PRM*, *CSQ* e *IDC* não acompanham de forma regular a variação da taxa de perda obtida na simulação.

Nas figuras 28 e 29, observa-se que a taxa de perda varia com o tamanho do “buffer” de forma não regular. Especificamente na figura 28 para tamanho de “buffer” até aproximadamente 240 células o arqa17 apresenta maior perda do que o arqa13, mas para tamanho de “buffer” entre 240 e 2.500 células o arqa13 apresenta maior perda em relação

ao arqa17. Note ainda que no caso da figura 28 foi utilizada uma banda de 1.200 bytes/segundo. A figura 29 mostra o mesmo fenômeno entre os arquivos arqa21 e arqa22.

Em geral o tráfego de dados corporativo, representado pelos arquivos analisados, apresentaram um parâmetro H maior que 0,5, mais precisamente situado entre 0,54 e 0,87, que implica em tráfego com comportamento auto-similar.

O tráfego Frame-Relay coletado a 18 meses atrás [22] apresentou um parâmetro H com valor próximo a 0,5. O fator preponderante causador desse aumento é muito provavelmente o grande crescimento da utilização da Intranet na empresa nos últimos meses.

Os parâmetros PRM , CSQ e IDC apresentam variações abruptas e não regulares, sem uma relação coerente com as variações apresentadas pelos parâmetros H e MBS por tipo de tráfego e taxa de perda de células obtida na simulação.

De acordo com os resultados apresentados neste estudo recomendamos o uso dos parâmetros H e MBS para caracterizar o tráfego de dados em ambientes corporativos.

4.4 Caracterização de Tráfego de Vídeo

4.4.1 Ambiente de Estudo

Até recentemente o principal cenário do tráfego de vídeo com compressão em redes públicas ou privadas, foi a videoconferência utilizando canais determinísticos baseados em TDM, em modo “unicast” ou “multicast”. Este cenário está gradualmente mudando, com a utilização do ATM em ambientes LAN e WAN, para videoconferência baseadas em VCC VBR em “desktop” ou não.

Particularmente, uma aplicação popular conhecida como vídeo sob demanda, engloba a situação onde uma estação cliente assiste a um vídeo em tempo real através da rede. Note que neste caso, o vídeo está armazenado em um servidor remoto. Outro exemplo desta natureza seria uma aplicação na medicina, onde imagens de uma cirurgia

são transmitidas para uma outra localidade, possibilitando intervenção da equipe médica remota. A principal diferença da aplicação de vídeo sob demanda para a aplicação médica, sob o enfoque de rede, é que a primeira é uma comunicação unidirecional enquanto a segunda bidirecional.

Neste trabalho implementamos a primeira situação, com a finalidade de coletar arquivos de tráfego vídeo com compressão que estão sendo transmitidos na rede em tempo real. A figura 33 mostra em digrama de blocos o “*setup*” do ambiente para captura de vídeo que foi utilizado. O ambiente é composto por uma estação cliente e uma estação servidora, ambas com sistema operacional Windows95™. A estação cliente possui “*software*” capaz de abrir os arquivos de vídeo em padrões MPEG (*Moving Pictures Expert Group*) e AVI (*Audio and Video Interleave*) armazenados no servidor. Para informações sobre os padrões de vídeo comprimido utilizados, consultar o Apêndice C. A conexão entre as estações é feita através de um barramento padrão Ethernet IEEE 802.3 fornecido por um “*Hub*” marca Digital™ modelo DecHub. Os arquivos de tráfego de vídeo foram capturados através de um analisador de protocolos LAN modelo DA320 da Wandel & Golterman™, mais detalhes sobre os instrumentos utilizados vide o Apêndice F.

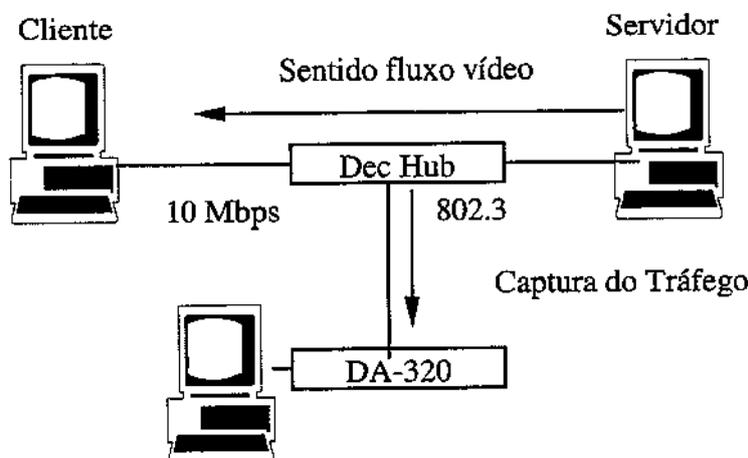


Figura 33: Ambiente para Captura de Vídeo

4.4.2 Análise do Tráfego de Vídeo

Os arquivos de tráfego de vídeo padrão MPEG e AVI foram obtidos de “sites” da Internet e fornecidos pelo Centro de Computação da UNICAMP. Os arquivos possuem tamanhos variando de 4Mbytes até 45 Mbytes, com cenas variadas em termos de movimentação. É importante ressaltar que a captura de amostras de vídeo dos arquivos estava limitada pela capacidade do “buffer” do analisador Wandel & Golterman™ que é de aproximadamente 3 Mbytes.

A tabela 15 apresenta os resultados gerais consolidados das medidas do parâmetro H (valores mínimo, máximo e médio), para os dois tipos de vídeo com compressão analisados. Igualmente ao estudo do tráfego de dados, estimamos o parâmetro H do tráfego de vídeo através do método das variâncias [1][36][44][58][78].

Arquivo de Tráfego	H		
	Méd.	Mín.	Máx.
AVI	0,552	0,5	0,66
MPEG	0,768	0,7	0,84

Tabela 15: Parâmetro H

As tabelas 16, 17, 18 e 19 apresentam os resultados gerais consolidados das medidas dos parâmetros PRM , CSQ , IDC e MBS respectivamente, mostrando os valores médio, mínimo, máximo e desvio padrão, para os dois tipos de compressão analisados.

Arquivo de Tráfego	PRM		
	Méd.	Mín.	Máx.
AVI	2,7812	1,71	3,606
MPEG	2,6634	1,13	6,94

Tabela 16: Parâmetro PRM

Arquivo de Tráfego	CSQ		
	Méd.	Mín.	Máx.
AVI	109,15	95,72	126,45
MPEG	148,26	16,96	377,14

Tabela 17: Parâmetro CSQ

Arquivo de Tráfego	IDC		
	Méd.	Mín.	Máx.
AVI	29,232	14,48	48,26
MPEG	18,119	0,94	57,61

Tabela 18: Parâmetro IDC

Arquivo de Tráfego	MBS		
	Méd.	Mín.	Máx.
AVI	27	17	43
MPEG	115	32	184

Tabela 19: Parâmetro MBS

As tabelas 20 e 21 apresentam os resultados dos parâmetros analisados para cada arquivo por tipo de tráfego.

Arquivo de Tráfego	AVI				
	<i>H</i>	<i>PRM</i>	<i>CSQ</i>	<i>IDC</i>	<i>MBS</i>
Arq83	0,5	3,606	116,78	48,26	17
Arq87	0,5	2,07	103,23	25,06	35
Arq90	0,5	3,63	95,72	18,66	17
Arq93	0,66	1,71	126,45	14,48	43
Arq94	0,60	2,89	103,57	39,70	22

Tabela 20: Tráfego de Vídeo Padrão AVI

Arquivo de Tráfego	MPEG				
	<i>H</i>	<i>PRM</i>	<i>CSQ</i>	<i>IDC</i>	<i>MBS</i>
Arq70	0,75	1,17	377,14	0,94	169
Arq71	0,72	1,13	197,61	1,44	184
Arq74	0,83	6,94	18,31	28,93	73
Arq75	0,84	1,827	16,96	1,678	177
Arq95	0,70	2,25	131,32	57,61	32

Tabela 21: Tráfego de Vídeo Padrão MPEG

O método adotado para avaliar os parâmetros de tráfego de vídeo, foi idêntico ao método aplicado ao tráfego de dados, ou seja, através de simulação. Os detalhes sobre os arquivos de tráfego utilizados e do simulador, estão disponíveis no Apêndice B e E respectivamente.

Os parâmetros de tráfego apresentados neste capítulo, como: *PRM*, *CSQ*, *IDC*, *MBS* e parâmetro *H*, apresentados nas tabelas 20 e 21 foram comparados com os resultados de perda de células obtidos da simulação e apresentados nas figuras 34 e 35 que mostram a taxa de perda obtida por respectivo arquivo e tipo de compressão.

Os valores de tamanho de “buffer” utilizados foram na faixa de 200 a 30.000 células. Para essa simulação também a banda utilizada foi de 150.000 bytes/segundo.

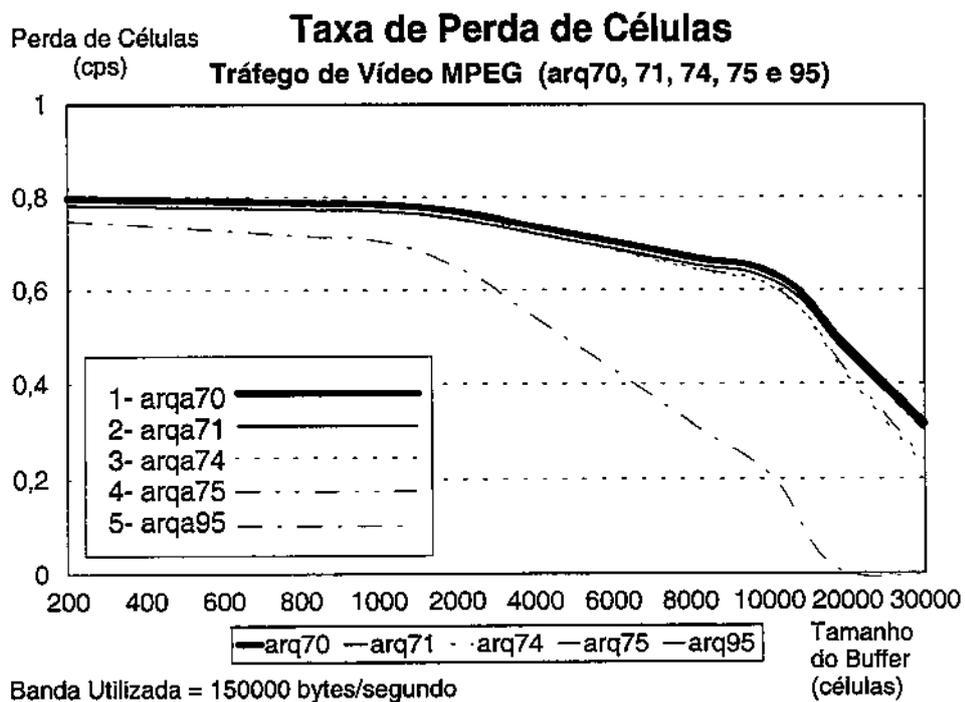


figura 34: Taxa de Perda para tráfego de Vídeo MPEG

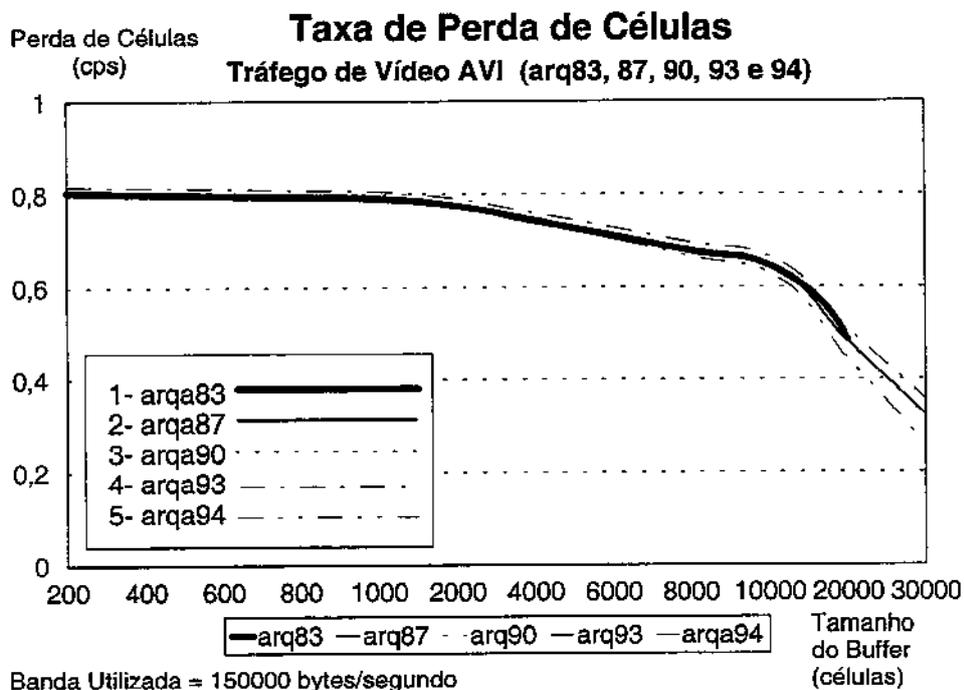


figura 35: Taxa de Perda para tráfego de Vídeo AVI

4.4.3 Conclusões sobre a Caracterização do Tráfego de Vídeo

Observando-se os valores calculados dos parâmetros de tráfego das tabelas 20 e 21 e comparando com os resultados da simulação apresentados nas figuras 34 e 35 para os respectivos arquivos de tráfego, nota-se que os arquivos que apresentam maior taxa de perda de células são aqueles que possuem maior parâmetro H e maior MBS , assim como ocorreu para o tráfego de dados. Os parâmetros como PRM , CSQ e IDC não acompanham de forma regular a variação da taxa de perda obtida na simulação.

De acordo com o gráfico apresentado na figura 34, para os arquivos de tráfego MPEG, os arquivos 70 e 71 apresentaram a maior perda com parâmetros H de 0,75 e 0,72 e MBS de 169 e 184 respectivamente.

Na figura 35, para os arquivos de tráfego AVI, o arquivo 93 apresentou maior perda, com parâmetros H de 0,66 e MBS de 43.

Os arquivos com compressão MPEG apresentam maior taxa de compressão que os arquivos AVI, como resultado observamos medidas de H e MBS superiores para os arquivos MPEG.

Os parâmetros PRM , CSQ e IDC , da mesma forma do resultado obtido para o tráfego de dados, apresentaram variações não regulares, sem uma relação coerente com as variações apresentadas pelos parâmetros H e MBS por tipo de tráfego e taxa de perda de células obtida na simulação.

4.5 Conclusão

Neste capítulo apresentamos medidas de alguns parâmetros presentes na literatura para análise de desempenho de tráfego. Esses parâmetros foram testados em vários arquivos de tráfego reais de dados e de vídeo. Nossa conclusão baseada em resultados obtidos em simulação é que os parâmetros que mostraram-se mais significativos em

termos de impacto na perda de células, tanto no caso de dados bem como de vídeo, foram os parâmetros H e MBS .

O parâmetro MBS é utilizado tanto nos métodos de perda bem como de atraso (Kesidis e Le Boudec respectivamente) e o parâmetro H será introduzido no método proposto. Esses métodos de alocação de banda estão descritos no Capítulo 5 deste estudo.

Capítulo 5

MÉTODO DE ALOCAÇÃO DE BANDA EM UMA CONEXÃO ATM

5.1 Introdução

O dimensionamento preciso de recursos de rede para uma determinada conexão é um fator fundamental nas redes ATM. Vários trabalhos têm sido apresentados, baseados no conceito de banda efetiva, com a finalidade de determinar a banda necessária para um determinado tráfego. Esses estudos têm se fundamentado na probabilidade de perda de células ou no tempo de atraso das células, contudo na grande maioria deles, sem levar em consideração um importante aspecto da caracterização do tráfego: a auto-similaridade, que tem forte impacto no desempenho dos comutadores.

Este estudo apresenta um método alternativo para a determinação da Banda Efetiva, considerando a característica auto-similar para cada tipo de tráfego, em uma conexão ATM, combinando os critérios de perda e atraso de células.

Nossa proposta generaliza a idéia de Kesidis baseada na probabilidade de perda de células, oriunda da Teoria dos Grandes Desvios. Para esse fim, efetuamos a caracterização do tráfego, comparamos e combinamos esses resultados com o método de atraso empregado por Le Boudec, usando métodos analíticos, heurísticos e simulação para alcançar uma aproximação mais realista da banda efetiva de uma conexão ATM.

O dimensionamento e garantia de largura de banda necessária para uma conexão é uma das principais funções do controle de admissão. Quando o tráfego é CBR (*Constant Bit Rate*), basta se alocar a taxa de pico; quando o tráfego é UBR (*Unspecified Bit Rate*), é alocada a banda disponível sem nenhuma garantia de QoS (*Quality of Service*); mas quando o tráfego é VBR (*Variable Bit Rate*) deve ser alocada uma banda situada entre a taxa média e a taxa de pico, que é denominada banda efetiva, conforme mostrado na

figura 36. Assim cada conexão VBR possui sua banda efetiva garantida durante o período de duração da mesma, sendo que [8]:

$$\sum_{i=1}^N c_i \cdot n_i \leq C \quad (84)$$

onde C é a capacidade do canal de comunicação, c_i é a banda efetiva de uma dada conexão do tipo i e n é o número de conexões do tipo i . Porém, estudo apresentado por Choudhury e Lucantoni [26] mostra que a equação (84) pode ter a banda superestimada quando muitas das n fontes possuem comportamento mais para rajada do que para Poisson. Assim, a determinação mais realista possível dessa banda efetiva, principalmente para muitas fontes não é uma tarefa trivial.

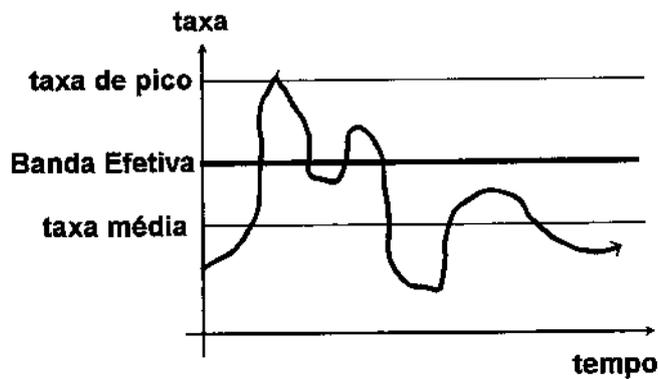


Figura 36: taxa instantânea – tráfego VBR

Existem vários métodos para estimar a banda efetiva para uma conexão VBR. Dentre os quais podemos destacar o método da banda efetiva determinística proposto por Le Boudec [40], que calcula o retardo de pior caso considerando um sistema sem perda. Este método está baseado em um trabalho prévio apresentado por Cruz [68][69].

Uma outra abordagem para determinar a banda efetiva bastante conhecida é o método de Kesidis [28], baseado em probabilidade de perda admissível derivada da Teoria dos Grandes Desvios. Infelizmente este método nem sempre fornece uma boa aproximação quando o tráfego possui característica auto-similar (vide a figura 41).

Neste trabalho generalizamos o método de Kesidis no cálculo da banda efetiva incorporando o parâmetro de auto-similaridade, evitando portanto a perda excessiva em

determinados tipos de tráfego cujos comportamentos são auto-similares e possuem dependências ao longo prazo.

5.2 O Método de Le Boudec

Neste método, em vez de se utilizar o modelo aleatório para representar o tráfego que está entrando na rede, se assume que o mesmo é “desconhecido”, mas com certa regularidade restrita. A restrição refere-se ao efeito de limitar a característica de rajada do tráfego (*burstiness*) em função do policiamento da rede [68].

Inicialmente assumimos uma conexão servindo a um fluxo de tráfego a uma taxa C . Esse fluxo de tráfego de chegada é restringido pela curva α denominada curva de chegada que caracteriza o valor mínimo da taxa C , requerido para um dado α . Na figura 37 é mostrada a curva de chegada α que representa o número de células que chegam no decorrer de um intervalo de tempo, dentro do limite imposto pela restrição ao fluxo de tráfego de entrada exercida pelo policiamento. Neste método é imposta a restrição de um atraso fixo D para o fluxo.

Considerando-se $\alpha(t)$ o número de células geradas pela fonte no intervalo $[0, t]$, sendo que $c_D(\alpha)$ é chamada banda efetiva determinística correspondente a curva α para uma restrição de atraso D . Pode se provar que um sistema de fila, com “buffer” infinito, com taxa de serviço constante C , garante um retardo fixo D para um fluxo com uma curva de chegada α se [40]:

$$c_D(\alpha) = \max_{t \geq 0} \frac{\alpha(t)}{t + D} \quad (85)$$

Se a curva de chegada $\alpha(t)$ é diferenciável, $c_D(\alpha)$ é dada pela inclinação da tangente a curva de chegada, conforme representada na figura 37.

O modelo de banda efetiva determinística não considera perda, portanto o “buffer” deve ter a capacidade b tal que [34][37]:

$$b = \sum_i c_D(\alpha)_i \cdot D_i \quad (86)$$

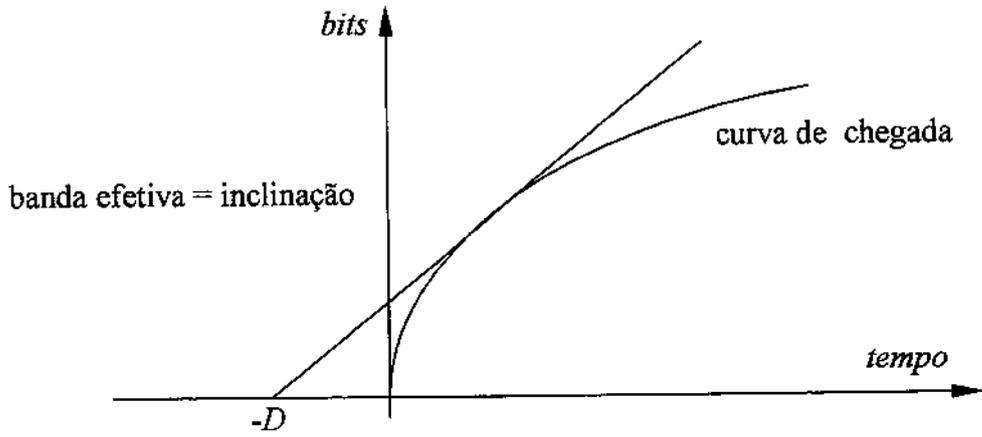


Figura 37: banda efetiva para um atraso D

Considerando p a taxa de pico (PCR), r a taxa média (SCR) e comprimento máximo da rajada L , teremos [34][37]:

Para fonte de taxa constante: $c_D = \max_t \frac{pt}{t+D} \rightarrow p$ (87)

Para fonte de taxa variável: $c_D = \frac{p \cdot T_{on}}{T_{on} + D}$ (88)

onde: $T_{on} = \frac{L}{p}$ (89)

$r = \frac{T_{on}}{T_{on} + T_{off}} p$ (90)

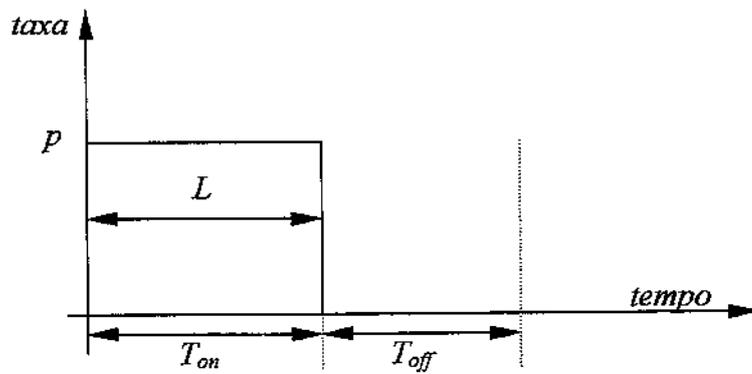


Figura 38: fonte On/Off

O enchimento e esvaziamento do “buffer” a uma taxa c na ocorrência da rajada é mostrado na figura 39.

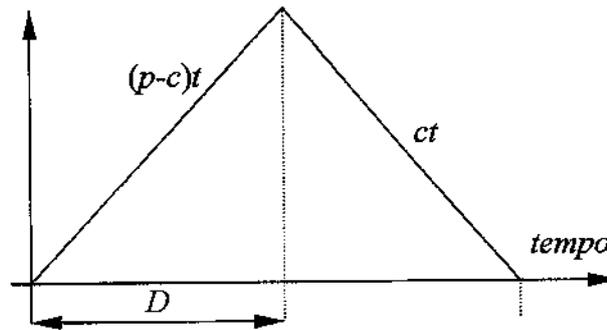


Figura 39: número de células no “buffer”

Observa-se que o enchimento do “buffer” obedece a função $(p-c)t$. O “buffer” é esvaziado a taxa c em um tempo máximo D , sendo que o número de células no mesmo ao final da rajada é $(p-c)T_{on}$ [37].

Definimos SCR (*Sustainable Cell Rate*) como [40]:

$$SCR = \limsup_{t \rightarrow \infty} \frac{\alpha(t)}{t} \quad (91)$$

e definimos PCR (*Peak Cell Rate*) como [40]:

$$PCR = \sup_{t \geq 0} \frac{\alpha(t)}{t} \quad (92)$$

então $SCR \leq c_D(\alpha) \leq PCR$ para qualquer atraso D . Além disso se $\alpha(t)$ é côncava temos que:

$$\lim_{D \rightarrow \infty} c_D(\alpha) = SCR \quad (93)$$

Desta forma a banda efetiva é dada pelo máximo de:

$$c_D = \max \left\{ SCR, \frac{MBS}{\frac{MBS}{PCR} + D} \right\} \quad (94)$$

$$\text{com: } D = \min \left\{ \max CTD - t_{prop} - \frac{C}{53 \times 8}; ppCDV \right\} \quad (95)$$

onde:

- t_{prop} é o tempo de propagação no enlace;
- $\max CTD$ é o atraso máximo permitido;
- $ppCDV$ é a máxima variação do atraso (“*Jitter*” máximo).

5.3 O Método de Kesidis

O método de Kesidis [28] baseia-se na Teoria dos Grandes Desvios (*Large Deviation Theory*), onde para pequenas probabilidades de perdas assume-se que cada fonte transmite a uma taxa constante chamada Banda Efetiva. Para maiores detalhes sobre a Teoria dos Grandes Desvios e sua relação com o conceito de banda efetiva, consultar o Apêndice A e referências bibliográficas [2][35][61][63]. O trabalho de Kesidis apresenta cálculo da banda efetiva para diversos tipos de fontes de tráfego, dentre as quais fontes de tráfego do tipo Fluidos Markovianos.

Sejam X o número de células no “*buffer*” e b a capacidade do “*buffer*”, adota-se o seguinte critério de desempenho:

$$P\{X > b\} \leq \exp(-b\delta) \quad (96)$$

A chave do conceito de banda efetiva está no regime assintótico para grandes tamanhos de “*buffer*”, com a probabilidade de perda caindo exponencialmente com o tamanho de b , com o expoente dado por δ , isto é:

$$\lim_{b \rightarrow \infty} \frac{1}{b} \log P(b) = -\delta \quad (97)$$

Para cada fonte do tipo i , $A_i(t)$ é o número de células geradas pela fonte no intervalo de tempo $[0, t]$. Assim, a função logarítmica da geradora de momentos de $A_i(t)$ é definida por:

$$h_i(\delta) = \lim_{t \rightarrow \infty} \frac{1}{t} \log E \exp(A_i(t)\delta) \quad (98)$$

Assumindo-se que as fontes são estacionárias, ergódicas e satisfazem o teorema de Gärtner-Ellis. Isto significa que a função assintótica logarítmica geradora de momentos de A_i existe e é finita para todo real δ e h_i diferenciável. Verifica-se que h_i é convexa, positiva e crescente para $\delta > 0$ [61]. Para maior detalhamento vide Apêndice A.

Seja H_i é a transformada de Legendre de h_i , isto é:

$$H_i(M_i) = \sup_{\delta \in \mathfrak{R}} \{ \delta M_i - h_i(\delta) \} \quad (99)$$

onde δ é um número real finito. Define-se a banda efetiva da fonte tipo i por

$$c_i(\delta) = h_i(\delta) / \delta \quad (100)$$

Para uma fonte do tipo Fluido Markoviano, por exemplo, a banda efetiva da fonte é dada por [34][61]:

$$c = \frac{1}{\delta} \text{eigmax}[M + \delta \cdot N] \quad (101)$$

onde M é a matriz com taxas de transição entre estados e N é a matriz diagonal das taxas de cada estado i .

As seguintes equações foram derivadas e usadas para determinação da banda efetiva para fonte Fluido Markoviano On/Off [23][28][37]:

$$c = \alpha + \sqrt{\alpha^2 + \beta} \quad (102)$$

onde $p = PCR$ e:

$$\alpha = \frac{1}{2\delta} \left(p\delta - \frac{1}{T_{on}} - \frac{1}{T_{off}} \right) \quad (103)$$

$$\beta = \frac{P}{\delta T_{off}} \quad (104)$$

$$T_{on} = \frac{MBS}{PCR} \quad (105)$$

$$T_{off} = MBS \left(\frac{1}{SCR} - \frac{1}{PCR} \right) \quad (106)$$

$$\exp(-b\delta) = CLR \quad (107)$$

Fontes de tráfego tipo Fluido Markoviano são de especial interesse neste estudo, pois são modelos adotados no método proposto, no próximo tópico.

A figura 40 mostra a banda efetiva calculada pelo método de Kesidis variando tamanho de “buffer” de 2 até 100.000 células para uma perda admissível de células na ordem de 10^{-6} . Note que o cálculo da banda efetiva através do método de Kesidis utiliza apenas parâmetros como *PCR* (*Peak Cell Rate*), *SCR* (*Sustainable Cell Rate*) e *MBS* (*Maximum Burst Size*), considerando uma perda admissível de células, sem levar em consideração o comportamento do tráfego, auto-similar. Mostraremos a seguir que o método de Kesidis pode subestimar a banda efetiva para determinados tamanhos de “buffers” quando o tráfego possui a característica auto-similar.

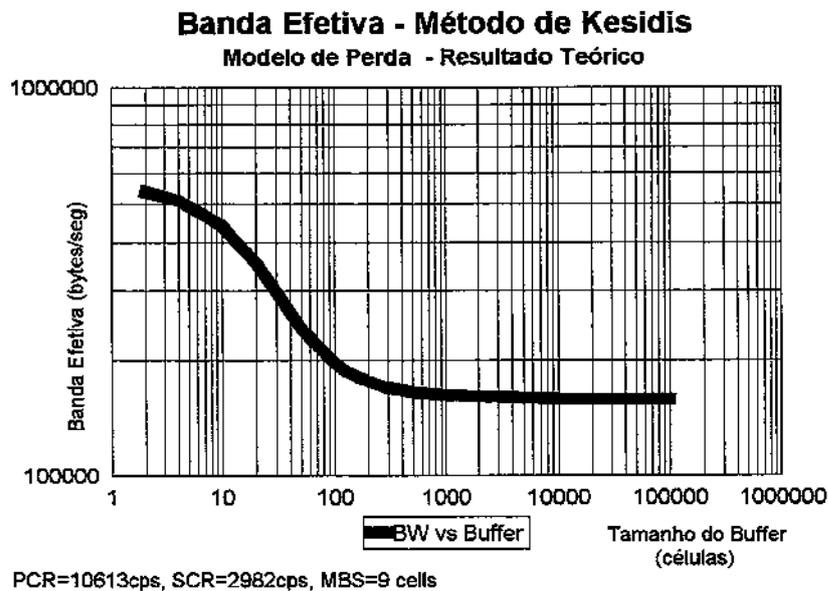


Figura 40: Banda efetiva da fonte Fluido Markoviano *On/Off* usando o método de Kesidis.

5.3.1 O Método de Kesidis e o Tráfego Auto-Similar

Empregamos o método de Kesidis (96) para calcular a banda efetiva para cinco arquivos de tráfego sintético [20], quatro deles com características auto-similar e um poissoniano, de onde foram obtidos os seguintes parâmetros de tráfego: parâmetro de Hurst, 0,5; 0,6; 0,7; 0,8 e 0,9, $PCR = 10613$ cps; $SCR = 2982$ cps e $MBS = 9$ células. Usando estes parâmetros foi calculada a banda efetiva necessária para cada fonte de tráfego considerando uma perda não superior a 10^{-6} .

Os resultados analíticos obtidos, referentes a banda efetiva e tamanho de “buffer”, alimentaram um simulador (maiores detalhes sobre o simulador utilizado consultar o Apêndice E e referências [43][44]).

A figura 41 mostra os resultados na simulação com arquivos de tráfego com H distintos. Analisando a taxa de perda de células em função do tamanho de “buffer” utilizado, podemos concluir:

- método de Kesidis não está adequado para dimensionar a banda efetiva para “buffer” de tamanho reduzido mesmo para tráfego Poissoniano. Isto é, a taxa de perda está acima do valor permitido, 10^{-6} quando a capacidade do “buffer” é menor do que 370 células (vide curva com $H = 0,5$).
- método de Kesidis não está adequado para dimensionar a banda efetiva para tráfego auto-similar, já que a taxa de perda se torna excessiva com o aumento do nível de auto-similaridade. Para reduzir a perda é necessária a utilização de “buffers” com capacidade superior a calculada pelo método de Kesidis.

O aumento do tamanho do “buffer” melhora o desempenho em termos de perda de células, contudo degrada o desempenho em termos de atraso.

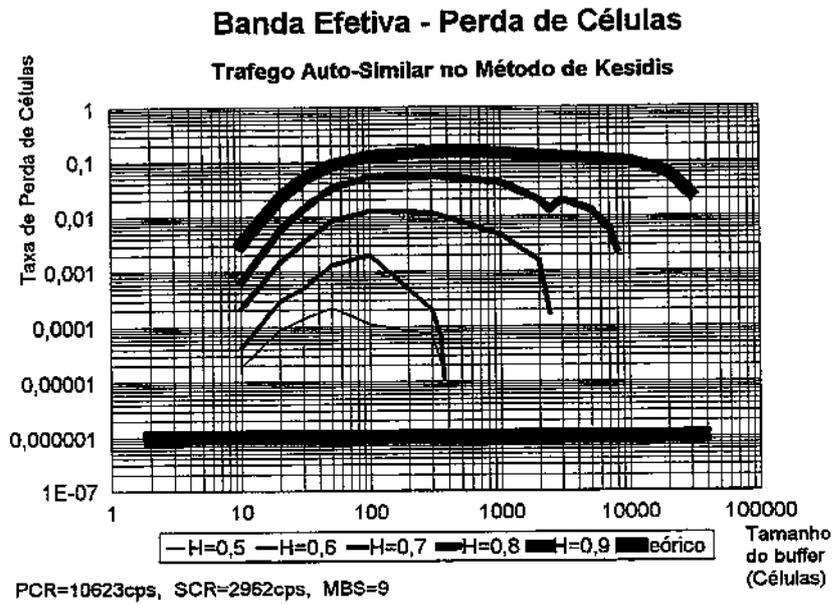


Figura 41: Tráfego Auto-Similar

5.4 O Método Proposto para o Cálculo da Banda Efetiva

O método proposto pretende ser mais preciso na determinação da banda efetiva de uma conexão ATM, mantendo qualidade de serviço exigida. Isso só é possível por que o critério de perda admissível de células é conservador. Para convalidar nossa abordagem, os resultados analíticos são comparados com os resultados de simulação. Além disso, um método heurístico foi usado para otimizar o cálculo de banda efetiva.

Considerando-se a característica fractal de vários tipos de fontes de tráfego, podemos assumir, segundo Wang [45], que tráfego agregado pode ser modelado por um processo do tipo FBM (*Fractional Brownian Motion*) ou pelo processo F-ARIMA (*Fractional Auto-Regressive Integrated Moving Average*). No entanto, para uma única conexão o tráfego pode ser modelado através de uma fonte *On/Off*.

Nosso objetivo é determinar a banda efetiva necessária para uma única conexão com a fonte apresentando um comportamento fractal, para tanto generalizamos o modelo de perdas proposto por Kesidis baseado na Teoria dos Grandes Desvios [28].

Na verdade este método pode ser aplicado para várias conexões multiplexadas, calculando-se a banda efetiva de cada conexão individual e usando-se a equação (84).

Contudo, segundo Choudhury e Lucantoni [26], a qualidade dessa aproximação deteriorará na medida que o número de fontes multiplexadas aumenta. Portanto recomendamos este método para obter a banda efetiva de uma conexão.

Segundo Glynn e Whitt [63], um sistema de fila de um único servidor, em estado estacionário, com carga X (ou tempo de espera), tem-se:

$$1/b \log P(X>b) \rightarrow \delta \text{ com } b \rightarrow \infty, \text{ para } \delta > 0, \quad (108)$$

δ é chamada de taxa de decaimento assintótica. Em alguns casos a equação (108) atende a seguinte condição:

$$P(X>b) \rightarrow \alpha^* \cdot \exp[-b \cdot \delta] \text{ com } b \rightarrow \infty, \quad (109)$$

onde α^* é chamada de constante assintótica. Em geral a equação (109) é adequada para um valor de b não muito pequeno, e ainda, para alguns propósitos é adotado $\alpha^* \approx 1$.

A generalização do resultado de Glynn e Whitt foi efetuada por Duffield e O'Connell [60][61][63], mostrando que quando $P(X>b)$ é apropriadamente escalonada, satisfará o seguinte princípio da Teoria dos Grandes Desvios:

$$\lim_{b \rightarrow \infty} b^{-2(1-H)} \ln P(X > b) = -a^{-2(1-H)} (a + C)^2 / 2 \quad (110)$$

onde $a = C/H-C$, com $0,5 < H < 1$, e C a taxa de serviço. A equação (110) pode ser aproximada por [60]:

$$P\{X > b\} \leq \exp[-\delta b^{2(1-H)}] \quad (111)$$

onde

$$\delta = -a^{-2(1-H)} (a+C)^2 / 2 > 0 \quad (112)$$

que é a equação aplicada para fontes de tráfego com dependência de longo prazo.

Contudo esta aproximação não é válida para todos os valores de b , para b muito pequeno pode ocorrer imprecisão.

Para calcular a banda efetiva para uma conexão ATM através do método proposto, é também usado o modelo de fonte On/Off. Após a caracterização do tráfego, isto é, de posse dos parâmetros de tráfego com PCR , SCR , MBS e H , é aplicada a equação (111) para uma dada taxa de perda de células admissível. A seqüência de aplicação do método está detalhada no Apêndice C.

Finalmente, após a banda efetiva ter sido calculada é aplicado um fator de otimização de banda. A introdução do fator de otimização, têm por finalidade obter uma aproximação mais realista da banda efetiva a ser determinada, para diferentes tamanhos de “buffers” e diferentes valores de parâmetro H . O fator de otimização proposto é uma função dependente de “ b ” e “ H ” e foi obtido empiricamente.

De fato o fator de otimização pode ser aprimorado através da aplicação exaustiva do método para um grande número de arquivos de tráfego de características diferenciadas.

Experimentalmente verificamos que para “buffer” de até 300 células o fator de otimização $f_o = \text{constante} = 1,07$, pode ser adotado. Para “buffer” maior que 300 células, f_o é uma função de b e H , dada por:

$$f_o = \frac{1}{2} \arcsen h(H \ln b) \quad (113)$$

resultando como banda efetiva otimizada:

$$c_{CO} = f_o^{-1} \cdot c_{SO} \quad (114)$$

onde: c_{CO} - banda efetiva com otimização e c_{SO} - banda efetiva sem otimização, e finalmente a banda efetiva a ser adotada é dada por:

$$c_P = \max \{ c_{CO}, SCR \} \quad (115)$$

5.4.1 Estudos Experimentais

Neste estudo foram utilizados cinco arquivos de tráfego sintético [20], cada um com diferentes parâmetros de Hurst, 0,6; 0,7; 0,8 e 0,9, e parâmetros de tráfego como $PCR = 10613$ cps; $SCR = 2982$ cps e $MBS = 9$ células. Usando estes parâmetros foi calculada a banda efetiva necessária para cada fonte de tráfego considerando uma perda não superior a 10^{-6} , baseado no método de Kesidis (96) e no método proposto neste trabalho (111).

Os resultados analíticos obtidos, referentes a banda efetiva e tamanho de “buffer”, alimentaram um simulador (maiores detalhes sobre o simulador utilizado consultar o Apêndice E e referências [43][44]), baseado na biblioteca de simulação Parasol, fornecendo como saída a perda real para cada tipo de tráfego.

Cada uma das figuras 42, 43, 44 e 45 compara como a banda efetiva varia com o tamanho de “buffer”, considerando uma perda de 10^{-6} , para o método de Kesidis, o método proposto sem otimização, o método proposto com fator de otimização e os resultados obtidos através da simulação para um determinado tráfego.

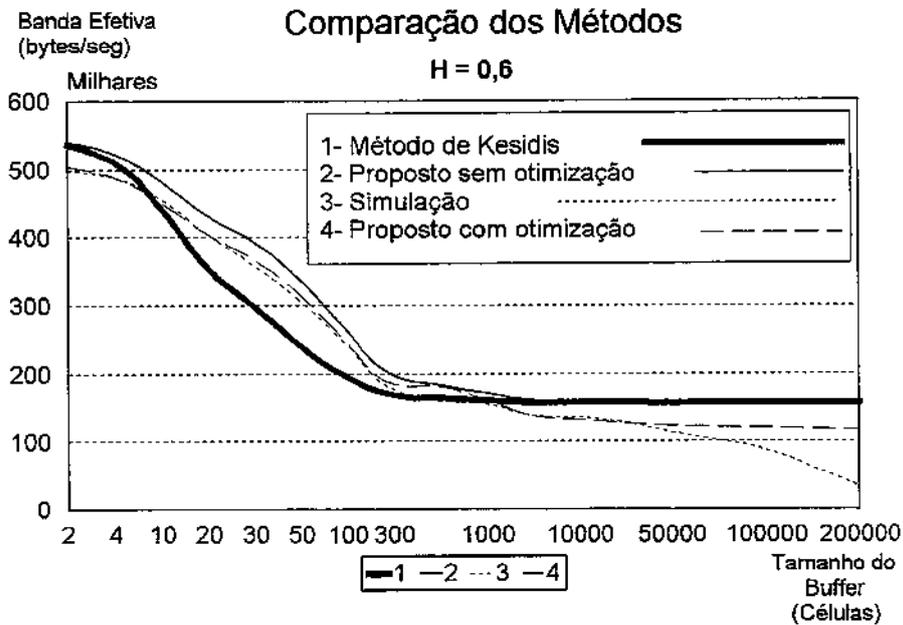


Figura 42: Banda Efetiva para H=0,6

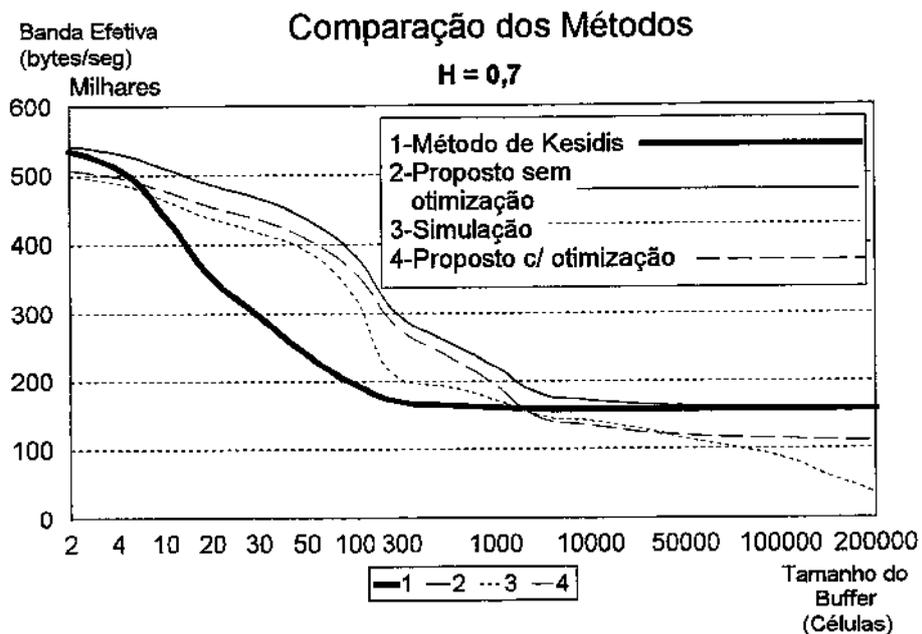


Figura 43: Banda Efetiva para $H=0,7$

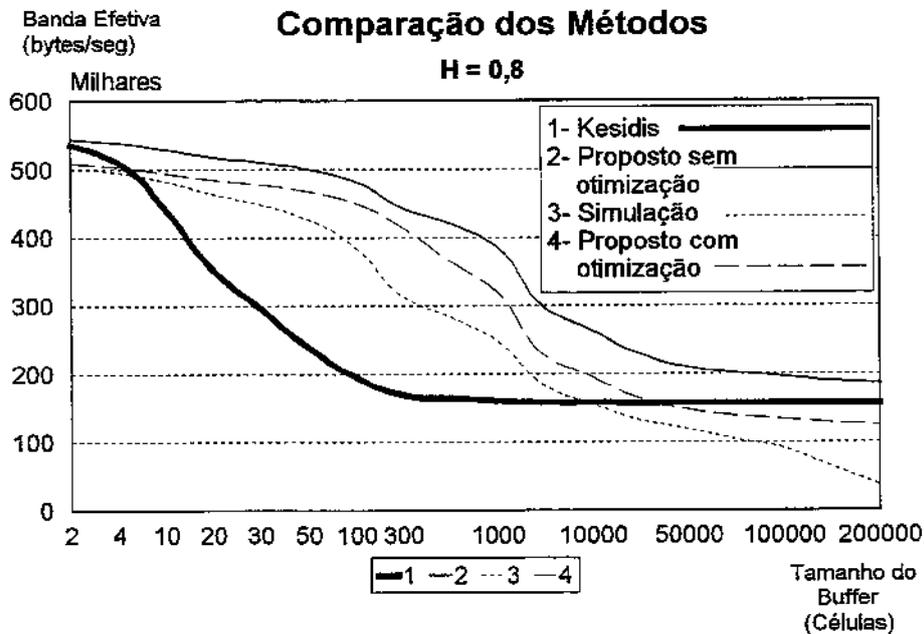


Figura 44: Banda Efetiva para $H=0,8$

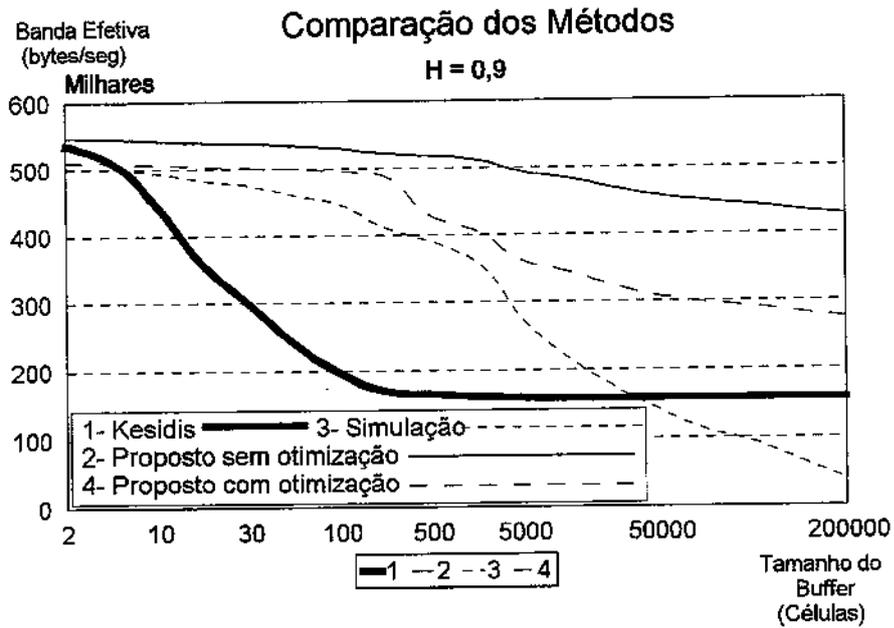


Figura 45: Banda Efetiva para $H=0,9$

Observamos experimentalmente que para os métodos propostos, com e sem otimização, não há perda de células, tendo como consequência a qualidade de serviço mantida. Contudo para $H = 0,8$ e $0,9$, nota-se que há uma reserva de banda em excesso em relação ao resultado da simulação, tornando-se um critério muito exigente (vide figuras 44 e 45).

Para minimizar esse fato, através da análise dos resultados analíticos e da simulação, foi aplicado o fator de otimização para reduzir a banda efetiva calculada. Note que este fator de otimização não é único, pois ele varia largamente com H e do tamanho do “buffer”. Obviamente quando o tamanho do “buffer” tende ao infinito, a banda efetiva tende ao valor mínimo, mas o atraso introduzido pode ser excessivamente grande.

5.4.2 Estudos com Tráfego de Vídeo Multiplexado

Para convalidar o método proposto por perda, apresentamos os resultados da simulação de duas conexões multiplexadas em um comutador. Utilizamos para esse fim um arquivo de tráfego de vídeo AVI (arq94) e um arquivo de vídeo MPEG (arq95), mais informações sobre os arquivos utilizados consultar o Apêndice B.

Em termos de atrasos, consideremos valores típicos da aplicação de videoconferência, que é uma aplicação em tempo real, ou seja, $\max CTD = 150$ msec e $ppCDV = 20$ msec.

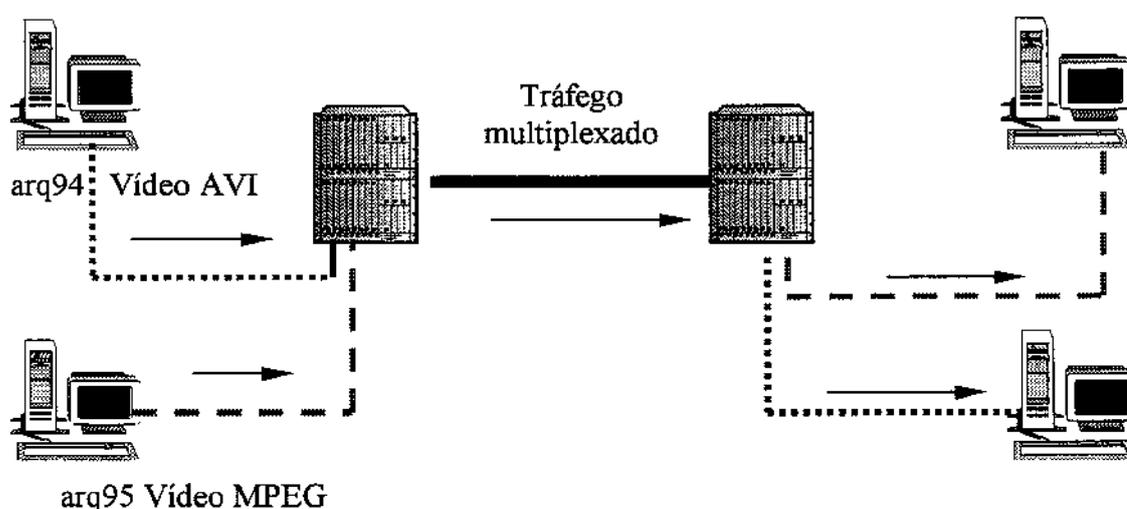


Figura 46: Ambiente simulado para tráfego de vídeo multiplexado
Videoconferência em “desktop”

A ferramenta utilizada para simulação da multiplexagem do tráfego de vídeo foi o SimATM [87] e [88], desenvolvido pela UNICAMP/FEEC/DECOM.

O tráfego de vídeo tipo AVI (arq94) apresentou os seguintes parâmetros de tráfego $PCR= 2133333$ bytes/seg, $SCR= 736620$ bytes/seg, $MBS= 23$ e $H= 0,6$. O tráfego de vídeo tipo MPEG (arq95) apresentou os seguintes parâmetros de tráfego $PCR= 1209523$ bytes/seg, $SCR= 537485$ bytes/seg, $MBS= 23$ e $H= 0,7$.

A Figura 47 apresenta os resultados de banda efetiva calculada pelo método proposto versus os resultados obtidos na simulação.

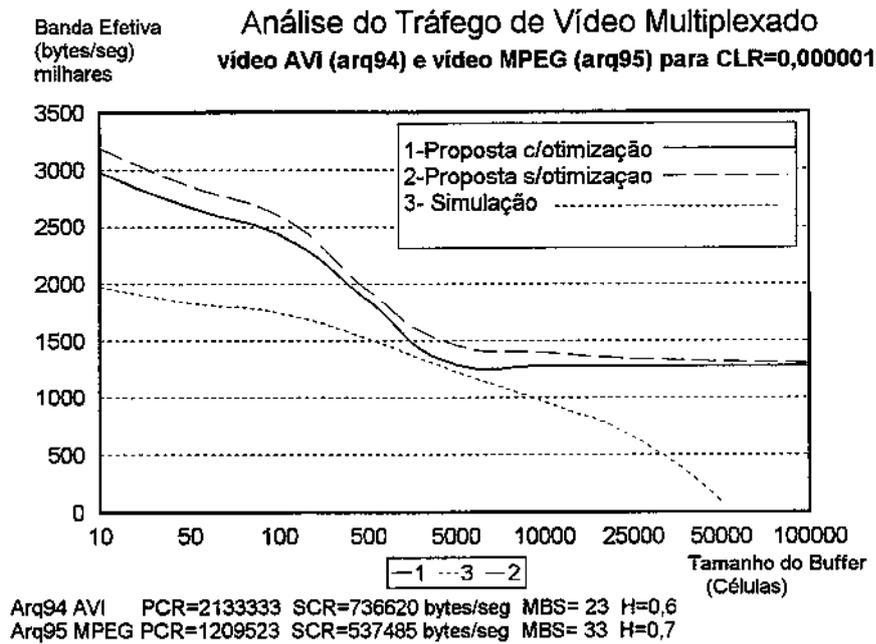


Figura 47: Banda Efetiva Método Proposto vs Simulação

Através da observação da Figura 47, pode-se concluir que a banda calculada pelo método proposto, já aplicado o fator de otimização, atendeu para o tráfego de vídeo comprimido multiplexado, para a taxa de perda admissível de 10^{-6} .

A Figura 48 apresenta o comportamento do atraso das células do tráfego multiplexado no “buffer” de saída do comutador. O resultado do ppCDV nesse “buffer” foi de aproximadamente 35 μ seg. Como os valores típicos da aplicação de videoconferência são de maxCTD = 150 msec e ppCDV = 20 msec, conclui-se que o valor encontrado de ppCDV para um “buffer” está aceitável. Neste caso a banda efetiva calculada atende ao requisito atraso, bem como perda de células.

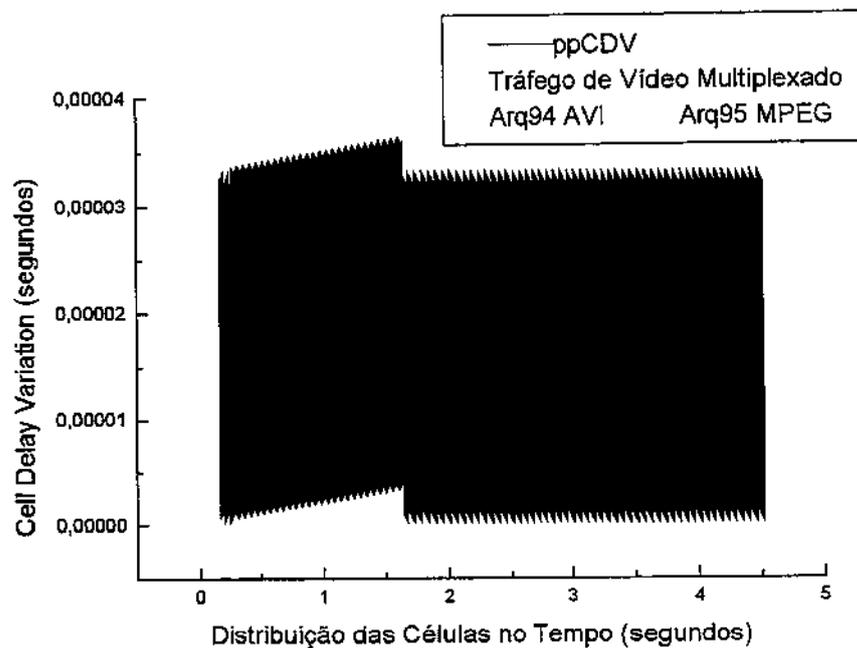


Figura 48: Comportamento do atraso

Baseado nos resultados obtidos na simulação do tráfego de vídeo multiplexado, constatamos que a banda efetiva calculada pelo método proposto de perda atendeu aos dois critérios analisados.

5.5 Método de Kesidis versus Método Proposto

O método de Kesidis não leva em consideração a característica de auto-similaridade inerente ao tráfego com dependência de longo prazo. No cálculo da banda efetiva são considerados somente os parâmetros *PCR*, *SCR*, *CLR* e *MBS*, onde este último não é suficiente para representar o tráfego em rajada com a devida precisão.

Na figura 41, o resultado da simulação para uma taxa de perda admissível de 10^{-6} , mostra claramente o resultado do impacto de tráfegos com diferentes parâmetros de Hurst na perda de células.

O método proposto introduz o parâmetro *H* no cálculo da banda efetiva, abrangendo desde tráfego Poissoniano até ao tráfego altamente auto-similar.

Como já foi mencionado, para pequenos tamanhos de “*buffer*”, ambos os métodos introduzem alguma imprecisão, visto que a consideração é assintótica.

O método proposto gera um resultado bastante conservador, alocando banda suficiente para garantir a Qualidade de Serviço. Contudo, de acordo com os resultados obtidos através de simulação, para grandes tamanhos de “*buffer*”, há uma alocação de banda em excesso. Para melhorar a aproximação do método foi introduzido um fator de otimização para duas faixas de tamanho de “*buffer*”: uma função constante igual a 1,07 e uma função hiperbólica dada pela equação (113).

5.6 Metodologia Resultante da Combinação dos Métodos Baseados em Perda e em Atraso

A simples comparação do método de perda (Proposto/Kesidis) com o método baseado em atraso (Le Boudec) não é uma tarefa trivial, uma vez que para o primeiro caso o atraso não é considerado. No segundo caso, o tamanho do “*buffer*” não tem importância uma vez que o mesmo é considerado de tamanho infinito, logo não há perda de células. Além disso, o método de Le Boudec pressupõe uma restrição na característica de rajada do tráfego, através da ação de um policiamento.

Para efeito do nosso trabalho, vamos supor a situação da existência de um policiamento de rede bastante relaxado, partindo-se do princípio que o provedor tenha banda suficiente disponível a oferecer e queira aumentar o faturamento e o cliente deseje transmitir o tráfego com a máxima qualidade de serviço possível.

Na verdade cada um dos métodos apresenta vantagens dependendo da aplicação a ser utilizada. Por exemplo, para tráfego em tempo real onde o fator atraso é crítico o método de Le Boudec atende. Para tráfego Poissoniano o método de Kesidis é satisfatório e finalmente quando o tráfego possui característica auto-similar o método proposto satisfaz.

Nosso objetivo neste momento é criar um critério de alocação de banda combinando os métodos apresentados considerando a classificação das classes de

serviços da ITU-T. Estamos interessados, para este estudo, em atender tráfego com as seguintes características:

- orientado ou não à conexão;
- com taxa de transmissão variável (VBR);
- com dependência da variável tempo ou não.

Para tanto apresentamos a seguir, as figuras 49 e 50, onde são mostrados os resultados do método proposto, com e sem otimização, comparando os mesmos com resultados obtidos através do método de Le Boudec, considerando o “buffer” de tamanho fixo com capacidade de 200.000 células. Os parâmetros utilizados são os mesmos dos arquivos de tráfego utilizados na simulação e no cálculo do método proposto, ou seja: $PCR = 10.613$ cps, $SCR = 2983$ cps e $MBS = 9$ células, considerando $maxCTD$ entre 20 a 150 msec, tempo de propagação entre 5 a 10 msec e $ppCDV$ entre 1 a 50 msec. Os valores utilizados para $maxCTD$ e $ppCDV$ estão dentro da faixa para aplicações em tempo real, como videoconferência, disponível na literatura.

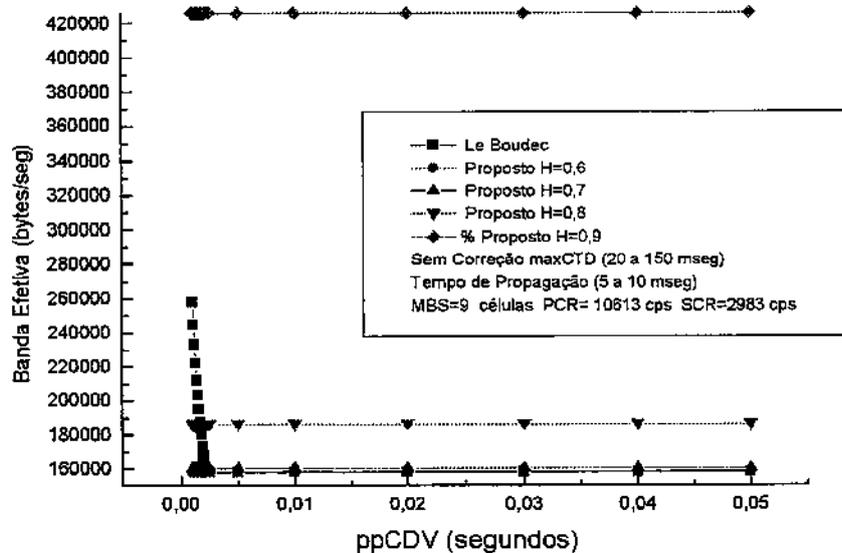


Figura 49: Proposto sem Correção x Le Boudec

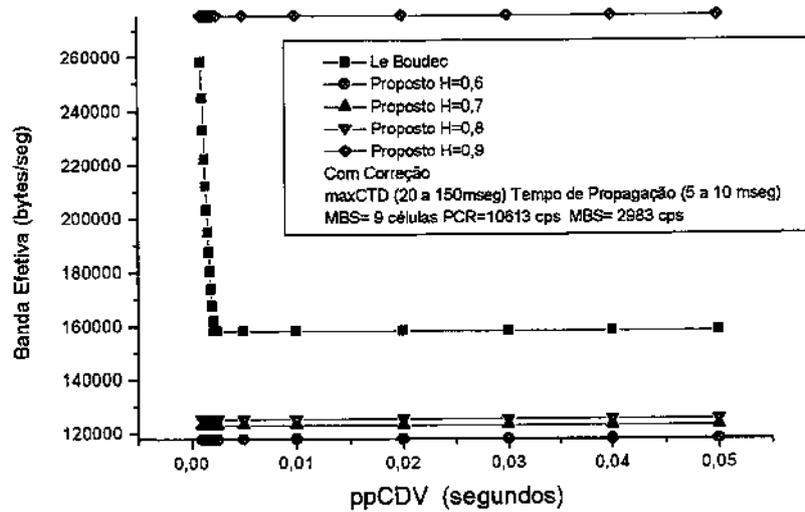


Figura 50: Proposto com Correção x Le Boudec

Através das figuras 49 e 50 pode ser verificado que para o método de Le Boudec com uma ppCDV acima de 5 mseg a banda efetiva permanece inalterada em 158099 bytes/seg.

A grande diferença entre os resultados dos dois métodos, ocorre no tráfego com $H=0,9$, onde o método proposto torna-se significativamente mais conservador.

Um exemplo típico de tráfego VBR em tempo real, com auto-similaridade é a aplicação videoconferência, que tipicamente utiliza $maxCTD = 150$ mseg e $ppCDV = 20$ mseg.

Diante do exposto, nossa proposta é aplicar o método proposto com otimização para qualquer tipo de tráfego orientado ou não à conexão, VBR sem dependência de tempo, ou seja calcular diretamente c_p (115).

Para o tráfego orientado ou não à conexão, VBR com dependência de tempo, surgem duas situações:

1. para tráfego não auto-similar, $H = 0,5$, aplica-se diretamente o método de Le Boudec, obtendo-se a banda efetiva determinística, através de c_D (94).
2. para tráfego auto-similar, $0,5 < H < 1$, calcular a banda efetiva através dos dois métodos, ou seja aplica-se Le Boudec e o método proposto, tomando como resultado final a maior (116), ou seja:

$$c = \max \{ c_D, c_P \} \quad (116)$$

5.7 Conclusões

Neste capítulo mostramos as duas principais abordagens para se estimar a banda efetiva de uma conexão ATM, representados pelos métodos de Le Boudec e Kesidis.

Generalizando a idéia de Kesidis, fundamentada na Teoria dos Grandes Desvios, introduzimos o parâmetro de auto-similaridade no cálculo da banda efetiva. Utilizamos métodos analíticos, heurísticos e simulação para validar a nossa proposta.

Finalmente propomos uma metodologia para cálculo da banda efetiva, que atende as diversas classes de serviço da ITU-T e do ATM Forum, através da combinação dos métodos apresentados.

Capítulo 6

6.1 Conclusão Geral

Estamos assistindo uma verdadeira transformação no que tange a utilização dos recursos da informação neste final de século. Neste aspecto salientamos o elevado grau de complexidade do novo perfil de tráfego que está surgindo e que dominará o cenário das redes emergentes de comunicação. Neste cenário a tecnologia ATM surge como um dos principais atores para garantir desempenho, qualidade de serviço e otimização de recursos de rede para as diferentes classes de serviço existentes.

Neste estudo apresentamos uma breve visão das principais características das redes ATM, procurando destacar os parâmetros de tráfego e qualidade de serviço utilizados nestas redes.

Com a finalidade de desenvolver um método alternativo para alocação de banda para uma conexão ATM, apresentamos os diversos modelos de tráfego aplicados na atualidade e difundidos na literatura existente, incluindo o modelo aplicado na metodologia proposta.

Outro importante aspecto incluído neste trabalho foi a caracterização de tráfego, apresentando diversos parâmetros de desempenho de tráfego e mostrando quais possuem maior impacto e importância no dimensionamento da banda. Para tanto foram coletadas amostras de tráfego real de diversos tipos de redes de comunicação em diferentes tipos de aplicações.

Finalmente apresentamos as duas principais abordagens para estimação da banda efetiva necessária para uma conexão. Neste tópico apresentamos um método alternativo para alocação de banda em uma conexão ATM, baseado no método de perda mas considerando a característica auto-similar do tráfego. Através da combinação dos métodos apresentados, foi proposta uma metodologia para o dimensionamento de banda efetiva em redes ATM.

Bibliografia

- [1] Albanita G.Dantas, Marcelo M.Carvalho, Dalton Soares Arantes, "Self-Similar Models for Variable Bit Rate in ATM Networks", Technical Report, Pub FEEC 027, UNICAMP, 1997.
- [2] Alan Weiss, "An Introduction to Large Deviations for Communication Networks", IEEE Journal on Selected Areas in Communications, Vol.13, No.6, August 1995.
- [3] Andrea Baiocchi, Nicola B.Melazzi, Marco Listanti, Aldo Roveri, Roberto Winkler, "Loss Performance Analysis on an ATM Multiplexer Loaded with High-Speed On/Off Sources, IEEE Journal os Selected Areas In Communications, vol.9, April 1991.
- [4] Andrea Baiocchi, Nicola B.Melazzi, "Steady State Analysis of the MMPP/G/1/K Queue, IEEE Transactions on Communications, Vol 41, No.4 April 1993.
- [5] Andrew S. Tanenbaum, "Computer Networks", Prentice Hall, 1996.
- [6] Anthony S. Acampora, "An Introduction to Broadband Networks", Plenum Press, 1994.
- [7] Anwar I.Elwalid, Debasis Mitra, "Effective Bandwidth of General Markovian Traffic Sources and Admission Control of High Speed Networks", IEEE/ACM Transactions On Networking, Vol 1, No.3, June 1993.
- [8] Arthur Berger, Ward Whitt. "Extending the Effective Bandwidth Concept to Networks with Priority Classes", IEEE Communications Magazine, August 1998.
- [9] Ashok Erramilli, Jonathan L. Wang, "Monitoring Packet Traffic Level", IEEE Globecom, 1994.
- [10] ATM Forum, "Traffic Management Specification", version 4.0, AF-TM-0056.000, Abril 1996.
- [11] ATM Forum, "User-Network Interface Specification", version 3.1, 1995.
- [12] Brian L.Mark, David L.Jagerman, G.Ramamurthy, "Peakedness Measures for Traffic Characterization in High-Speed Networks", NEC.
- [13] C.Miguel Tobar Toledo, Ivan Luiz Marques Ricarte, "Sobre Tipos de Ligações em Modelos de Dados para Hiperímídia", Instituto de Informática e FEEC-UNICAMP.
- [14] Daniel Minoli, "ATM & Cell Relay For Corporate Environment", McGraw-Hill, 1994.
- [15] David E. McDysan, Darren L. Spohn, "ATM Theory andApplication", McGraw-Hill, 1994.

- [16] David N.C. Tse, Robert G. Gallager, John N. Tsitsiklis, "Statistical Multiplexing of Multiple Time-Scale Markov Streams", LIDS-P-2275, MIT, November 1994.
- [17] Douglas E.Commer, David L.Stevens, "Internetworking with TCP/IP – Volume III", Prentice Hall, 1997.
- [18] Donald Gross, Carl M. Harris, "Fundamentals of Queueing Theory", John Wiley & Sons, 1974.
- [19] Edwin R. Coover, "ATM Switches", Artech House, 1998.
- [20] Ernesto Luiz de Andrade Neto, Antônio Alberti, Dalton S.Arantes, Leonardo Mendes. "A Realistic Model for Self-Similar Ethernet Lan Traffic in Sim-ATM", SBT/IEEE ITS 1998.
- [21] Emanuel Parzen, "Stochastic Process", Holden-Day Inc, 1976.
- [22] Firmiano R. Perlingeiro, Lee Luan Ling "Is The Ethernet Lan Traffic Over Frame-Relay in ATM Networks Self-Similar ? ", SBT/IEEE ITS 1998.
- [23] Frank Kelly, . "Notes on Effective Bandwidths ", University of Cambridge, 1995.
- [24] Frank G. Halasz, "Reflections on NoteCards: Seven Issues For The Next Generation of Hypermedia Systems", ACM, 1988.
- [25] Fred Halsall, "Data Communications, Computer Networks and Open Systems", Addison Wesley, 1996.
- [26] Gagan L.Choudhry, David Lucantoni, . "Squeezing the Most Out of ATM", IEEETrans. On Communications, Febuary 1996.
- [27] George C. Sackett, Cristopher Y. Metz, "ATM and Multiprotocol Networking", McGraw-Hill, 1996.
- [28] George Kesidis, Jean Walrand, Cheng-Shang Chang, "Effective Bandwidth for Multiclass Markov Fluids and Other ATM Sources". IEEE Trans.Networking, Vol 1, Agosto, 1993.
- [29] G.Veciana, J.Walrand, G.Kesidis, "Resource Management in Wide-Area ATM Networks Using Effective Bandwidth", IEEE 1995.
- [30] Haruo Akimaru, Takashi Okuda, Kazunori Nagai, "A Simplified Performance Evaluation for Burst Multiclass Traffic in ATM Networks", IEEE Transactions on Communication Vol.42, May 1994.
- [31] Harry Heffes, David M.Lucantoni, "A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance", IEEE Journal on Selected Areas in Communications, September 1986.
- [32] Hwei Hsu, "Probability, Random Variables, & Random Processes", McGraw Hill, 1996.

- [33] Hiroshi Saito, "Teletraffic Technologies in ATM Networks", Artech House, 1994.
- [34] Ivanil S.Bonatti, Amaury Krueel Budri, . "Modelos de Tráfego para B-ISDN ", December 1994.
- [35] James Roberts, Ugo Mocci, Jorma Virtamo, "Broadband Network Teletraffic - Final Report of Action COST 242", Springer, 1996.
- [36] Jan Beran, "Statistics for Long-Memory Process", Chapman & Hall, 1994.
- [37] Jaudelice C. Oliveira, . "Dimensionamento de Enlaces em Redes de Telecomunicações", MSc Thesis, UNICAMP, 1998.
- [38] Javier Aracil, Richard Edell, Pravin Varaiya, "A Phenomenological Approach to Internet Traffic Self-Similarity", UCB.
- [39] Jean-Pierre Leduc, "Digital Moving Pictures-Coding and Transmission on ATM Networks", Elsevier, 1994.
- [40] Jean-Yves Le Boudec, . "Network Calculus Made Easy ", Technical Report EPFL-DI, December, 1996.
- [41] John M. Griffiths, "ISDN Explained - Worldwide Network and Applications Technology", John Wiley & Sons, 1990.
- [42] John N. Daigle, Joseph D.Langford, "Models For Analysis of Packet Voice Communication Systems", IEEE Journal on Selected Areas in Communications. September 1986.
- [43] J.Neilson, "Parasol User's Manual", Version 3.1, Carleton University, School of Computer Science, Ottawa, Canada, pp 3-37, 1995.
- [44] J.Neilson, P.R.Morin. "The Impact of Self-Similarity on Network Analysis", Carleton University, School of Computer Science, Ottawa, Canada, 1995.
- [45] Jonathan L. Wang. "Analysis, Modeling and Management of Self-Similar Traffic", SBT/IEEE ITS 1998.
- [46] Judith L.Jerkins, Jonathan L. Wang, "A measurement Analysis of ATM Cell-Level Aggregate Traffic", IEEE Globecom, November 1997.
- [47] J.M. Pitts, J.A. Schormans, "Introduction ATM Design and Performance", John Wiley & Sons, 1997.
- [48] Lenord J.Forys, Ashok Erramilli, Jonathan L. Wang, "New Traffic Analysis and Engineering Methods for Emerging Technologies", IEEE Globecom, November 1995.
- [49] Leonard Kleinrock, "Queueing Systems", John Wiley & Sons, 1975.
- [50] Luiz F.G.Souares, G.Lemos, S.Colcher, "Redes de Computadores das LANs, MANs e WANs às Redes ATM", Editora Campus, 1995.

- [51] Mallikarjun Tatipamula, Bhunip Khasnabish, "Multimedia Communications Networks – Technologies and Services", Artech House, 1998.
- [52] Mark E. Crovella, Murad S. Taqqu, A. Bestavros, "Heavy-Tailed Probabilities Distributions in the World Wide Web", Boston University, 1996.
- [53] Martin De Prycker, "Asynchronous Transfer Mode Solution For Broadband ISDN", Prentice-Hall, 1995.
- [54] M. Thomas Chen, Stephen S. Liu, "ATM Switching Systems", Artech House, 1995.
- [55] Mike Sexton, Andy Reid, "Broadband Networking: ATM, SDH, and SONET", Artech House, 1997.
- [56] Mischa Schwartz, "Broadband Integrated Networks", Prentice Hall, 1996.
- [57] Michel Daoud Yacoub, "Foundations of Mobile Radio Engineering", CRC Press Inc - 1993.
- [58] Murad S. Taqqu, Vadim Teverovsky, Walter Willinger, "Estimators For Long-Range Dependence: An Empirical Study. Fractals. Vol 3. World Scientific.
- [59] Naohisa Ohta, "Packet Video – Modeling and Signal Processing", Artech House, 1994.
- [60] Nol Rananand, "Traffic Modeling and Performance Evaluation for ATM Networks: Short and Long Range Dependent Models", PhD Thesis, University of Maryland, 1996.
- [61] N.G. Duffield, O'Connell. "Large Deviations and Overflow Probabilities for the General Single-Server Queue, with applications ", DIAS-STP, 1993.
- [62] N.G. Duffield, J.T. Lewis, O'Connell, Raymond Russel, Fergal Toomey, "Entropy of ATM Traffic Streams: A Tool for Estimating QoS Parameters ", IEEE Journal on Selected Areas in Communications, Vol 13, No 4, August 1995.
- [63] P. Glynn, W. Whitt. . "Logarithmic Asymptotics for Steady-State Tail Probabilities in a Single-Server Queue ", 1993.
- [64] Peter J. Brockwell, Richard A. Davis, "Introduction to Time Series and Forecasting", Springer, 1996.
- [65] Raif O. Onvural, "Asynchronous Transfer Mode Networks Performance Issues", Artech House, 1995.
- [66] Raif O. Onvural, R. Cherukuri, "Signaling in ATM Networks", Artech House, 1997.
- [67] Ramesh Hagarajan, James F. Kurose, Don Towsley, "Approximation for Computing Packet Loss in Finite-Buffered Voice Multiplexers", University of Massachusetts, June 1990.

- [68] Rene L. Cruz, "A Calculus for Network Delay, Part I: Network Elements in Isolation", IEEE Transactions on Information Theory, Vol.37, No.1, January 1991.
- [69] Rene L. Cruz, "A Calculus for Network Delay, Part II: Network Elements in Isolation", IEEE Transactions on Information Theory, Vol.37, No.1, January 1991.
- [70] Roch Guérin, Hamid Ahmadi, M.Naghshineh, . "Equivalent Capacity and Its Applications to Bandwidth Allocation in High Speed Networks ", IEEE J.S.A in Communication, September, 1991.
- [71] Samuel Karlin, Howard M. Taylor, "A First Course in Stochastic Processes", Academic Press, 1975.
- [72] Siemens AG, "Teoria do Tráfego Telefônico", Volume 1, Editora Edgard Blucher Ltda 1975.
- [73] Thomas G. Robertazzi, "Computer Networks and Systems - Queueing Theory and Performance Evaluation", Springer-Verlag, 1994.
- [74] Vasudev Bhaskaran, Konstantinos Konstantinides, "Image and Video Compression Standards", Kluwer Academic Publishers, 1998.
- [75] V.Paxton, S.Floyd, "Wide-Area Traffic: The Failure of Poisson Modeling", Lawrence Berkeley Laboratory, University of California, 1994.
- [76] Victor Frost, Benjamin Melamed, "Traffic Modeling for Telecommunication Networks", IEEE Communication Magazine, Março 1994.
- [77] Walter J. Goralski, "Introduction to ATM Networking", McGraw-Hill, 1995.
- [78] W.E. Leland, W.Willinger, D.V.Wilson, "On The Self-Similar Nature of Ethernet Traffic (Extended Version)", IEEE, pp 1-15, 1994.
- [79] Wayne Tomasi, Vincent F.Alisouskas, "Telecommunications Voice/Data with Fiber Optic Applications", Prentice-Hall, 1988.
- [80] William Stallings, "ISDN and Broadband ISDN with Frame Relay and ATM", Prentice-Hall, 1995.
- [81] William Stallings, "Local & Metropolitan Area Networks", Prentice-Hall, 1997.
- [82] William Stallings, "Data and Computer Communications", Prentice-Hall, 1996.
- [83] William Stallings, "High-Speed Networks – TCP/IP and ATM Design Principles" , Prentice-Hall, 1998.
- [84] Y.Fisher, "Fractal Image Compression, Theory and Application", Spring-Verlag, 1994.
- [85] Yong-Qing Lu, David W.Petr, Victor S.Frost, "Survey of Source Modeling Techniques for ATM Networks", Technical Report TISL-10230-1, University of Kansas, September, 1993.

- [86] F.C.Alves, “Agregador de Tráfego csv”, www.lrprc.unicamp.br
- [87] Antônio M. Alberti, Ernesto Luiz Andrade Neto, “SimATM”
Tese de Mestrado, UNICAMP/FEEC/DECOM, 1998, www.mc21.unicamp.br
- [88] Ernesto Luiz Andrade Neto, “Simulador ATM”, Projeto FAPESP 96/12372-2
UNICAMP/FEEC/DECOM.
- [89] Patrick Billingsley, “Probability and Measure”, John Wiley & Sons, 1986.
- [90] Walter Rudin, “Principles of Mathematical Analysis”, McGraw-Hill, 1976.
- [91] Tay Vaughan, “Multimedia Making It Work”, Osborne McGraw-Hill, 1993.
- [92] “Multimedia Developer’s Guide”, Sams Publishing, 1994.
- [93] <http://www.mpeg.org/>
- [94] <http://www.ccuac.unicamp.br/>
- [95] <http://www.rahul.net/jfm/avi.html>

Apêndice A

Teoria dos Grandes Desvios e Banda Efetiva

1A. Introdução

O método de Kesidis para o cálculo da banda efetiva [28], apresentado no Capítulo 5, utiliza a probabilidade de perda de células admissível para uma determinada conexão ATM.

As redes ATM são suportadas por meios de transmissão digitais de alta confiabilidade, são orientadas à conexão e o CAC (*Connection Admission Control*) somente aceita conexões que pode garantir a QoS (*Quality of Service*) negociada para uma determinada CoS (*Class of Service*) [5][25][53][55][65]. Assim sendo, podemos considerar a perda de células como um evento raro. A Teoria dos Grandes Desvios é uma metodologia poderosa para se obter estimativas razoavelmente precisas de probabilidades de eventos raros [35].

Este Apêndice tem por finalidade apresentar resumidamente os conceitos básicos da Teoria dos Grandes Desvios relacionando a mesma com o conceito de banda efetiva baseado no critério de perda de células limitada.

Este estudo está baseado no trabalho de Alan Weiss intitulado “*An Introduction to Large Deviations for Communication Networks*” [2], e no trabalho de George Kesidis, Jean Walrand e Cheng-Shang Chang chamado “*Effective Bandwidth for Multiclass Markov Fluids and Other ATM Sources*” [28].

Neste apêndice vamos apresentar resumidamente os conceitos básicos da Teoria dos Grandes Desvios para facilitar o entendimento do conceito de banda efetiva baseado em perda, incluindo: o teorema de Chernoff, o princípio da Teoria dos Grandes Desvios, teorema de Sanov, teorema de Gärtner-Ellis e a relação da Teoria dos Grandes Desvios com o conceito de banda efetiva baseado em perda.

Adotaremos o termo “TGD” para representar de forma abreviada a Teoria dos Grandes Desvios.

De forma simples dizemos que o cálculo da probabilidade de um evento raro, consiste em atribuir um custo para cada amostra de caminho do processo estocástico que pode causar a ocorrência deste evento, achando-se então o caminho de menor custo [2]. Tal probabilidade é dada por:

$$P\{\text{evento}\} \approx \exp(-n.\text{custo}) \quad (1a)$$

onde n é um parâmetro assintótico, frequentemente representando o tamanho do sistema considerado ou intervalo de tempo de observação. Aqui, consideramos eventos raros em termos de amostras de caminho.

Através do exposto, nota-se que a TGD é aplicável a certos tipos de eventos raros.

2A. Teorema de Chernoff

O teorema de Chernoff também conhecido como teorema de Cramér responde uma das questões fundamentais da TGD: considerando as variáveis aleatórias X_i i.i.d., com média $EX_i < \infty$, qual é a probabilidade:

$$P \left\{ \frac{X_1 + \dots + X_n}{n} \geq a \right\} \text{ e onde } EX_i < a \quad (2a)$$

Definimos $M(\theta)$ como a função geradora de momentos de X_i e $L(.)$ é uma função não negativa (colocando $\theta = 0$ em 5a) e convexa. A transformada aplicada em $\log M$ é conhecida como transformada convexa, transformada de Fenchel, transformada de Legendre ou transformada de Cramér. Sem perder a generalidade, assumimos que o operador de logaritmo seja neperiano. Assim,

$$M(\theta) = Ee^{\theta x_i} \quad (3a)$$

$$L(a) = - \log \left(\inf_{\theta} e^{-\theta a} M(\theta) \right) \quad (4a)$$

Aplicando álgebra simples em (4a), obtemos:

$$L(a) = \sup_{\theta} (\theta a - \log M(\theta)) \quad (5a)$$

O supremo em (5a) é alcançado no ponto θ^* no interior do intervalo onde $M(\theta)$ é finito. Se $M(\theta)$ é diferenciável em θ^* , então:

$$L(a) = - \log E e^{\theta^*(X_1 - a)} \quad (6a)$$

TEOREMA 1: [2] Considere uma seqüência de variáveis aleatórias X_1, X_2, X_3, \dots i.i.d., para cada $a > EX_1$ e um inteiro positivo n , temos:

$$P\{X_1 + X_2 + \dots + X_n \geq na\} \leq e^{-nL(a)} \quad (7a)$$

Assuma que $M(\theta) < \infty$ para θ em alguma vizinhança de 0 e que se mantém para algum θ^* dentro da vizinhança. Então para cada $\varepsilon > 0$ existe um inteiro n_0 tal que $n > n_0$.

$$P\{X_1 + X_2 + \dots + X_n \geq na\} \geq e^{-n(L(a) + \varepsilon)} \quad (8a)$$

As equações (6a) e (7a), implicam em:

$$P\{X_1 + X_2 + \dots + X_n \geq na\} = e^{-nL(a) + o(n)} \quad (9a)$$

onde $o(n)$ é a notação de Landau [89], isto é $o(n) \rightarrow 0$ quando $n \rightarrow \infty$.

3A. Princípio dos Grandes Desvios

Generalizando o teorema 1, considerando que a média pode não existir, temos:

TEOREMA 2: [2] Considere as variáveis aleatórias X_1, X_2, X_3, \dots i.i.d., a função L em (8a) é convexa e semicontínua à esquerda. Para qualquer conjunto F real, teremos:

$$\lim_{n \rightarrow \infty} \sup \frac{1}{n} \log P \left\{ \frac{X_1 + \dots + X_n}{n} \in F \right\} \leq - \inf_{a \in F} L(a) \quad (10a)$$

e para qualquer conjunto aberto G , teremos:

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P \left\{ \frac{X_1 + \dots + X_n}{n} \in G \right\} \geq -\inf_{a \in G} L(a) \quad (11a)$$

Uma declaração mais geral para ilustrar o princípio dos grandes desvios é apresentada através da definição 1. Porém vamos inicialmente especificar funções e parâmetros envolvidos:

- considerando uma seqüência de variáveis aleatórias Z_1, Z_2, Z_3, \dots , ou processos sobre um espaço de estado X .
- a função real I sobre X conhecida como função taxa satisfaz os seguintes requisitos:
 - $I(x) \geq 0$;
 - I é semicontínua inferiormente, isto significa que se Y_1, Y_2, Y_3, \dots é uma seqüência tal que $Y_n \rightarrow Y$ tal como $n \rightarrow \infty$, então $\liminf_n I(Y_n) \geq I(Y)$.
 - Se o conjunto $\{x : I(x) \leq a\}$ é compacto [90] para cada real a , então I é dita uma boa função de taxa.

DEFINIÇÃO 1: [2] Dizemos que Z_1, Z_2, Z_3, \dots satisfazem o princípio dos grandes desvios com uma função taxa I , se para cada conjunto fechado $C \subset X$, nós obtemos:

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P \{Z_n \in C\} \leq -\inf_{x \in C} I(x) \quad (12a)$$

Para cada conjunto aberto $G \subset X$, nós temos:

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P \{Z_n \in G\} \geq -\inf_{x \in G} I(x) \quad (13a)$$

4A. Teorema de Sanov

O teorema de Sanov é mais genérico, não considera somente a média amostral.

$$\frac{S_n}{n} = \frac{X_1 + X_2 + \dots + X_n}{n} \quad (14a)$$

Podemos escrever a média da amostra S_n como:

$$S_n = n \int_{-\infty}^{\infty} t dF_n(t) \quad (15a)$$

com:

$$F_n(t) \equiv \frac{1}{n} \sum_{i=1}^n 1[X_i(t) \leq t] \quad (16a)$$

onde $1(\cdot)$ é a função indicadora.

Sejam X_1, X_2, \dots, X_n , variáveis aleatórias i.i.d, com função distribuição F e a distribuição $\mu(\cdot)$ definida por $F(y) = \mu((-\infty, y])$.

Considere F_n representa a distribuição empírica da amostra, e μ_n corresponde a medida empírica para um conjunto qualquer $A \subset \mathcal{R}$ [2] dada por

$$\mu_n(A) = \int 1[t \in A] dF_n(t) \quad (17a)$$

Sabemos que para cada y , a Lei dos Grandes Números (aplicada a variáveis aleatórias i.i.d. $1[X_i \leq y]$) produz $F_n(y) \rightarrow F(y)$ com $n \rightarrow \infty$. Assim podemos escrever

$$\frac{X_1 + \dots + X_n}{n} = \int y dF_n(y) = \int y d\mu_n(y) \quad (18a)$$

Definimos o conjunto de medidas de probabilidades ν sobre o conjunto dos \mathcal{R} , por [2]:

$$S_a = \left\{ \nu : \int y d\nu(y) \geq a \right\} \quad (19a)$$

Podemos declarar:

$$P\{X_1 + \dots + X_n \geq na\} = P\{\mu_n \in S_a\} \quad (20a)$$

Seja p' o espaço de medida de probabilidade sobre \mathfrak{R} . Podemos perguntar quão próximo $\mu_n \approx \nu$, ou tentar estimar a probabilidade $P\{d(\mu_n, \nu) < \varepsilon\}$ para um ν fixo.

Considerando uma coleção de variáveis aleatórias X_1, X_2, \dots, X_n , o teorema de Sanov se aplica em qualquer caso em que dependa somente dos valores tomados da coleção X_1, X_2, \dots, X_n , mas independa da sua ordem.

Considere uma seqüência $\mu_1, \mu_2, \mu_3, \dots$, para cada conjunto $C \subset P$, onde P é o espaço de medida de probabilidade em \mathfrak{R} , teremos:

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P\{\mu_n \in C\} \leq -\inf_{\nu \in C} I(\nu) \quad (21a)$$

e para cada conjunto aberto $G \subset P$, teremos:

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P\{\mu_n \in G\} \leq -\inf_{\nu \in G} I(\nu) \quad (22a)$$

O teorema de Sanov estima a probabilidade de se estabelecer o princípio da TGD para μ_n .

5A. Teorema de Gärtner-Ellis

Este teorema prova os limites da TGD em \mathfrak{R}^d usando as propriedades da função geradora de momentos. Este método é bastante aplicado quando as variáveis aleatórias de interesse não são i.i.d. de fato, mas aproximadamente [2].

Dado uma seqüência de variáveis aleatórias, S_1, S_2, \dots definimos:

$$\phi(\theta) \equiv \frac{1}{n} \log E\{exp(\theta S_n)\} \quad (23a)$$

Estimaremos as assíntotas de $\{S_n\}$. Note que se X_1, X_2, \dots são i.i.d. e

$$S_n \equiv \sum_{i=1}^n X_i, \quad (24a)$$

$$\phi(\theta) \equiv \frac{1}{n} \log E\left\{ \prod_{i=1}^n exp(\theta X_i) \right\} = \frac{1}{n} \log \prod_{i=1}^n E\{exp(\theta X_i)\} = \log M(\theta) \quad (25a)$$

DEFINIÇÃO 3: [2] o domínio efetivo da função f é $D_f \equiv \{x : f(x) < \infty\}$.

A seguinte suposição é feita para que as variáveis aleatórias sejam consideradas quase i.i.d.:

- o limite $\lim_n \phi_n(\theta) \equiv \phi(\theta)$ existe pontualmente;
- permitimos que o limite para qualquer ϕ_n assuma o valor $+\infty$. No caso i.i.d., $\phi_n = \phi$.

DEFINIÇÃO 4: [2] $I(x) \equiv \sup_{\theta} (\theta x - \phi(\theta))$

Note que esta definição está de acordo com a definição de L para o caso i.i.d.

O Teorema de Gärtner-Ellis é apresentado para os limites superior e inferior conforme os teoremas 3 e 4 respectivamente.

TEOREMA 3: [2] o limite superior (*upper bound*), sob as considerações assumidas, para $-\infty < a < b < \infty$, se $[a, b] \cap D_f \neq \emptyset$, então:

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P \left\{ \frac{S_n}{n} \in (a, b) \right\} \leq - \inf_{x \in [a, b]} I(x) \quad (26a)$$

Para o limite inferior são necessárias maiores restrições, como: ϕ é diferenciável sobre $D\phi$. Note que $D\phi$ não é aberto, temos que tomar cuidado ao próximo da fronteira.

TEOREMA 4: para o limite inferior (*lower bound*), faça $-\infty < a < b < \infty$. Assuma que para qualquer $v \in (a, b)$ existe θ_v tal que $\phi'(\theta_v) = v$. Então:

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P \left\{ \frac{S_n}{n} \in (a, b) \right\} \geq - \inf_{x \in (a, b)} I(x) \quad (27a)$$

considerando que ϕ' é contínua, $D\phi = \mathfrak{R}$ e se:

- $\limsup_{\theta \rightarrow \infty} \phi'(\theta) \geq b$ e
- $\liminf_{\theta \rightarrow -\infty} \phi'(\theta) \leq a$.

Se $D\phi \neq \mathcal{R}$, estas condições devem permanecer quando θ se aproxima da fronteira de $D\phi$.

6A. Banda Efetiva

Estamos interessados no caso onde a perda de células é um evento raro (na ordem de 10^{-4} a 10^{-6}). O cálculo da banda efetiva para um determinado tráfego está baseada na aproximação da probabilidade de perda de células $P(b)$ considerada na TGD, quando o tamanho do “buffer” for suficientemente grande [16].

A chave do cálculo da banda efetiva baseada em perda está no regime assintótico para grandes tamanhos de “buffer”, com a probabilidade de perda diminuindo exponencialmente com o tamanho de b , com o expoente dado por δ , isto é:

$$\lim_{b \rightarrow \infty} \frac{1}{b} \log P(b) = -\delta \quad (28a)$$

$$\lim_{b \rightarrow \infty} \log P(b) = -\delta b \quad (29a)$$

Considerando somente o comportamento assintótico de (29a), ou seja, $b \rightarrow \infty$:

$$P(b) \approx \exp(-\delta b) \quad (30a)$$

Comparando a equação (1a) proposta por Weiss [2] e a equação (30a), identificamos a seguinte correspondência:

$$n \Leftrightarrow b \text{ (tamanho do “buffer”)} \text{ e } \textit{custo} \Leftrightarrow \delta \text{ (que é a função taxa)}$$

Assumindo que as fontes de tráfego são estacionárias e ergódicas. Faça o número de chegadas da fonte tipo i no intervalo de tempo $[0, t]$ denotado por $A_i(t)$ [28][29].

Considere um “buffer” de tamanho infinito com uma taxa de serviço c , compartilhado por N_i fontes de tráfego do tipo i , $i = 1, 2, \dots, k$, assume-se que todas as fontes são independentes.

Seja a taxa de uma fonte i , para todo M_i maior que a taxa média de células produzidas pela fonte de tipo i , assuma que a probabilidade dessa fonte produzir $M_i T$ células em um intervalo de tempo de duração T é aproximadamente

$\exp(-TH_i(M_i))$, onde H_i é convexa e não negativa (transformada de Legendre ou convexa aplicada a função geradora de momentos). Em função da independência das fontes e da convexidade da função H_i , para k fontes do tipo i , podemos escrever:

$$\exp\left(-T \sum_{i=1}^k N_i H_i(M_i)\right) \quad (31a)$$

Portanto a probabilidade de que iniciando-se com o “buffer” vazio, a ocupação do mesmo exceda b , considerando-se que as fontes do tipo i produzam uma taxa $N_i M_i$, é dada por:

$$\exp\left(-b \frac{\sum N_i H_i(M_i)}{\sum N_i M_i - c}\right) \quad (32a)$$

Na verdade, $T = b/(\sum N_i M_i - c)$ é o tempo que a ocupação do “buffer” alcança b quando a taxa do agregado é $\sum N_i M_i$.

Portanto a probabilidade que a ocupação do “buffer” seja b , antes de voltar a ficar vazio é:

$$P\{X > b\} = \exp\left(-b \inf_{\sum N_i M_i = c} \frac{\sum N_i H_i(M_i)}{\sum N_i M_i - c}\right) \quad (33a)$$

Note que, observando-se a equação (31a), temos:

$$\delta = \left(\inf_{\sum N_i M_i = c} \frac{\sum N_i H_i(M_i)}{\sum N_i M_i - c} \right) \quad (34a)$$

ou

$$\delta = \inf_{M_i \in A_i(a)} \frac{H_i(M_i)}{M_i - a} = I(a) \quad (35a)$$

onde $A_i(a) = \{M_i: a < M_i < \Gamma_i\}$.

Do teorema de Gärtner-Ellis, no limite inferior, temos:

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P \left\{ \frac{S_n}{n} \in (a, b) \right\} \geq - \inf_{x \in (a, b)} I(x) \quad (36a)$$

Podemos escrever que:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P \left\{ \frac{S_n}{n} \in (a, b) \right\} \geq - I(x) \quad (37a)$$

Para o intervalo de $[0, t]$, em termos de ocupação de “buffer”:

$$\lim_{b \rightarrow \infty} \frac{1}{b} \log P \{X > b\} \geq - I(a) \quad (38a)$$

onde pela TGD $I(a)$ representa a função taxa, sendo $I(a) = \delta$.

Assumindo que A_i satisfaz as condições do teorema de Gärtner-Ellis, apresentadas acima, significa que a função assintótica logarítmica geradora de momentos de A_i existe e é finita para todo real δ e h_i diferenciável. Verifica-se que h_i é convexa, positiva e crescente para $\delta > 0$ [61].

$$h_i(\delta) = \lim_{t \rightarrow \infty} \frac{1}{t} \log E \exp(A_i(t)\delta) \quad (39a)$$

Seja H_i a transformada de Legendre de h_i , onde:

$$H_i(M_i) = \sup_{\delta \in \mathfrak{R}} \{ \delta M_i - h_i(\delta) \} \quad (40a)$$

Para uma fonte única do tipo i , para $\delta > 0$, definimos $\alpha_i(\delta)$ como sendo a banda efetiva para o tráfego tipo i , tal que [28]:

$$I_i(a) = \inf_{M_i \in A_i(a)} \frac{H_i(M_i)}{M_i - a} = \delta \quad (41a)$$

$$\alpha_i(\delta) = \frac{1}{I_i(\delta)} = \frac{h_i(\delta)}{\delta} \quad (42a)$$

Apêndice B

Listagem dos Arquivos de Tráfego Utilizados

Nome do Arquivo	Tipo de Tráfego	Número de Pacotes	Número de Mbytes
Arq10	Internet	7.405	3,412275
Arq11	Internet	8.880	3,346974
Arq12	Internet	8.109	3,379928
Arq13	VSAT	16.698	0,362921
Arq14	VSAT	15.379	0,348409
Arq15	VSAT	11.039	0,249599
Arq16	VSAT	11.876	0,275196
Arq17	VSAT	12.621	0,275162
Arq18	Gerência SNMP	17.405	2,946143
Arq19	Automação	10.027	3,292994
Arq20	Automação	9.568	3,313541
Arq21	Automação	11.181	3,245995
Arq22	Automação	15.542	1,712897
Arq23	Automação	22.897	2,728754
Arq24	Frame-Relay	31.477	6,298732
Arq49	Frame-Relay	6.600	3,189631
Arq62	Ethernet	6.948	3,435125
Arq63	Ethernet	7.201	4,400121
Arq64	Ethernet	9.745	3,302504
Arq65	Ethernet	7.222	3,0112441
Arq66	Ethernet	9.740	3,310317
Arq68	Gerência SNMP	18.563	2,887034
Arq70	Vídeo MPEG	2.697	3,625337
Arq71	Vídeo MPEG	2.742	3,624099
Arq74	Vídeo MPEG	2.257	3,008474
Arq75	Vídeo MPEG	2.331	3,106685
Arq83	Vídeo AVI	3.531	3,58456
Arq87	Vídeo AVI	3.527	3,584515
Arq90	Vídeo AVI	3.551	3,583538
Arq93	Vídeo AVI	2.741	3,624511
Arq94	Vídeo AVI	2.250	3,06125
Arq95	Vídeo MPEG	1.101	1,008893

Tabela 1B: Listagem de Arquivos de Tráfego

Apêndice C

Método de Cálculo Proposto para Banda Efetiva por Perda

1C. Introdução

Este apêndice mostra o algoritmo utilizado para calcular a banda efetiva de uma conexão ATM mencionado no Capítulo 5. Os resultados são mostrados através dos gráficos 42, 43, 44, 45 e 47, também no Capítulo 5.

2C. Passos de Execução

A seqüência de aplicação do método proposto é a seguinte:

1. Inicialmente coletar o tráfego na rede;
2. De posse do arquivo de tráfego calcular os seguintes parâmetros de tráfego: PCR , SCR , MBS e H ;
3. Conhecendo-se a rede a ser dimensionada determinar o tamanho do “buffer” disponível para a conexão;
4. Especificar a perda admissível para a conexão em função da aplicação, utilizar o valor do parâmetro H obtido no passo 2, introduzir os valores na equação (111) para obter δ ;
5. De posse do valor de δ , usar nas equações (103), (104), (105) e (106) para obter os parâmetros α , β , T_{On} e T_{Off} .
6. Calcular a banda efetiva usando a equação (102);
7. Finalmente, utilizar o fator de otimização, obtendo-se a banda efetiva final para uma conexão ATM, através das equações (113), (114) e (115).

Apêndice D

Os Padrões de Compressão de Vídeo: MPEG e AVI

1D. Introdução

Os arquivos de tráfego de vídeo comprimido utilizados neste trabalho foram obtidos em “sites” da Internet [93] ou fornecidos pelo Centro de Computação da UNICAMP [94].

Neste apêndice são fornecidas algumas informações básicas sobre os padrões de compressão de vídeo MPEG e o AVI. Maiores informações sobre o assunto recomendamos consultar as referências bibliográficas [74], [91], [92], [93] e [95].

2D. O Padrão MPEG

MPEG é o nome dado a uma família de padrões adotados internacionalmente para codificar informações de áudio e vídeo em um formato digital comprimido. Essa família inclui os padrões MPEG1, MPEG2, MPEG4 e MPEG7.

Em 1988 em resposta a crescente necessidade de uma padronização comum para o formato de codificação e armazenamento de sinais de vídeo digital, a ISO (*International Organization for Standardisation*) estabeleceu um grupo conhecido como “*Moving Pictures Expert Group*” (MPEG), com a missão de desenvolver padrões para representação de imagens em movimento codificadas, associadas a informação de áudio sobre uma mídia de armazenamento digital.

Na primeira fase deste processo de padronização, completada em 1991, foi emitido o padrão MPEG1, com a taxa de 1,5 Mbps para vídeo e áudio associados, denominado padrão ISO/IEC 111172 (IEC - *International Electrotechnical Commission*) [74].

Em 1990, o grupo MPEG estabeleceu o padrão MPEG2, denominado pela ISO/IEC 13818, adotado pela ITU-T (*International Telecommunication Union*) como H.262.

Em 1994 iniciou-se o esforço no sentido de se estabelecer o padrão MPEG4, hoje conhecido como ISO/IEC 14496.

Atualmente, o padrão MPEG7 está em fase de desenvolvimento [93].

2D.1 O MPEG-1

É um algoritmo de codificação de vídeo com um esquema de compressão com perda que pode ser aplicado a uma grande quantidade de formatos de entrada e aplicações.

O objetivo do MPEG1 foi desenvolver um algoritmo de compressão para ser utilizado em CD-ROM ou em linhas telefônicas de baixa velocidade. A intenção do grupo era alcançar qualidade a nível de “*full motion and full screen*”, equivalente VHS.

O MPEG1 inicia com uma seqüência de vídeo com baixa resolução, de 352 x 240 pixels por quadro a 30 quadros/segundo. As imagens em cores são convertidas para o espaço YUV, sendo os dois canais de crominância (U e V) reduzidos a 120 a 176 pixels [93].

O esquema básico do MPEG é prever os movimentos de quadro para quadro na direção temporal, e então usar DCT (*Discrete Cosine Transform*) para organizar os arranjos estatísticos da informação no espaço.

Os quadros são divididos em blocos, geralmente de 8 x 8 pixels, sendo que o mecanismo DTC é aplicado individualmente sobre os blocos. A DTC converte a intensidade dos pixels em frequências equivalentes. O resultado é uma série de números que representam cada detalhe do bloco em pixels. A compressão ocorre através da eliminação de um certo número de pontos da série, que são imperceptíveis ao olho humano. Este tipo de algoritmo é considerado por perda (*lossy compression*), pois a compressão ocorre pela eliminação de informação considerada redundante. A predição do movimento é feita sobre o canal de luminância em blocos de 16 x 16 pixels por quadro a ser codificado. Os coeficientes da DTC são quantizados. Finalmente as informações obtidas, como os coeficientes da DTC, vetores e movimento e parâmetros de quantização são codificados pelo método de Huffman [74], [93].

Esta técnica permite o descarte de 99,5 % da informação sem afetar seriamente a qualidade da imagem. Como resultado o MPEG1 pode chegar a taxas de compressão superiores a 180:1.

O padrão ISO/IEC 11172 é dividido nas seguintes partes:

- ISO 11172-1 descreve o sistema, sincronização e multiplexagem do áudio e vídeo;
- ISO 11172-2 descreve a compressão de sinais de vídeo não intercalado, é detalhada a sintaxe (elementos do cabeçalho e do trem de bits) e a semântica (algoritmo dizendo o que fazer com os bits).
- ISO 11172-3 descreve a compressão dos sinais de áudio;
- ISO 11172-4 descreve procedimentos de testes de conformidade das características do vídeo codificado, processo de decodificação e outros requerimentos.

As funções de compressão dentro do MPEG incluem o seguinte:

- redução da taxa de amostragem nos domínios espacial e temporal para os componentes de luminância e crominância;
- DCT (*Discrete Cosine Transform*) baseada em blocos para intraquadros e interquadros;
- Compensação do movimento em blocos, para interquadros baseados em predição e interpolação;
- Compressão dos vetores de movimento sem perda e coeficientes quantizados DCT com codificação de Huffman.
- MPEG1 não reconhece fontes de vídeo entrelaçadas, usando o termo “*picture*” em vez de quadro. No vídeo entrelaçado cada quadro é composto de dois campos, ao passo que no vídeo progressivo não há a noção de campos. Observe na figura 1C.

A sintaxe do MPEG1 permite tamanhos de “*pictures*” acima de 4.095 x 4.095 pixels.

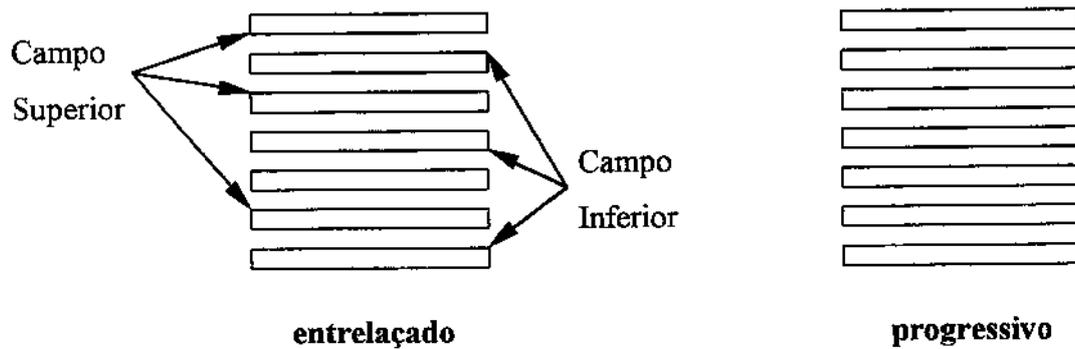


Figura 1D: Esquemas entrelaçado e progressivo representação no domínio espacial

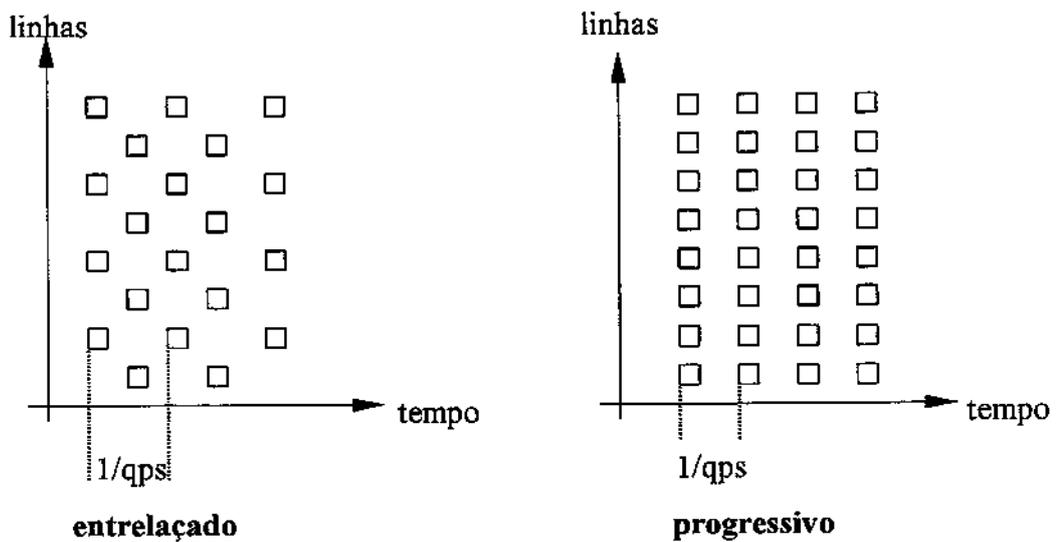


Figura 2D: Esquemas entrelaçado e progressivo representação no domínio temporal

2D.2 O MPEG-2

É a segunda fase do MPEG, com a missão de definir um padrão genérico que poderia ser utilizado em uma vasta classe de aplicações, como suporte a taxas fechadas de 5 Mbps para vídeo de qualidade NTSC/PAL, ou ainda 10 Mbps com vídeo de qualidade próxima a de estúdio.

O MPEG2 não foi criado para substituir o MPEG1, e sim porque o MPEG1 não atende aos requisitos de qualidade de vídeo para radiodifusão, ou seja a qualidade do MPEG1 não

é aceitável para vídeo destinado a entretenimento. Este tipo de vídeo em padrões como o NTSC e PAL, necessitam 720 x 480 e 704 x 576 pixels por quadro respectivamente. Contudo, a maior dificuldade é que o vídeo para radiodifusão é intercalado, ou seja, cada quadro possui dois campos entrelaçados, esse tipo de vídeo é incompatível com MPEG1. MPEG2 pode representar vídeo intercalado ou progressivo, operando a taxas de 6 Mbps, atendendo aos requisitos de radiodifusão, permitindo ainda a multiplexagem de seis a dez canais dentro de uma banda destinada a um canal analógico [91], [92] e [93].

Requisitos impostos ao MPEG2:

- compatibilidade com o MPEG1;
- boa qualidade de imagem;
- flexibilidade quanto ao formato de entrada;
- capacidade de acesso aleatória;
- escalabilidade;
- baixo retardo.

O MPEG2 suporta vídeo entrelaçado, esquemas de subamostragem para 4:2:0; 4:2:2 e 4:4:4, apresenta novas opções para compressão de movimento melhorando a qualidade e nova sintaxe.

2D.3 O MPEG-4

Este padrão é destinado a programas de áudio e vídeo que requeiram baixas taxas de codificação através de novos algoritmos. As dimensões da amostragem são superiores a 174x144x 10 Hz, com taxa de codificação entre 4,8 a 64 kbps [93].

As aplicações alvo do MPEG4 são: comunicação móvel interativa com multimídia, videofones, correio eletrônico com multimídia, base de dados interativas com multimídia, jogos, dentre outras. Para tanto, o MPEG4 deve atender aos seguintes requisitos:

- operar em uma grande gama de taxas, de poucos kbps até na ordem de Mbps;
- suporte a identificação e gerência de IPR (*Intellectual Property Right*);

- possibilidade de “*hyperlink*” e interação simultânea com múltiplas fontes de informação através da Web;
- capacidade de tratar simultaneamente com informações reais ou sintéticas, em tempo real ou não;
- aderência ao conceito de objetos, pois a representação da informação é feita a nível de pedaços individuais denominados AVOs (*Audio-Visual Objects*);

2D.4 O MPEG-7

Este grupo está em desenvolvimento, tendo como objetivo criar um padrão para busca, filtragem, gerência e processamento de informação multimídia. O MPEG7 se tornará um Comitê “*Draft*” em Outubro do ano 2000 e é esperado que se torne um padrão “*Draft*” internacional em Julho de 2001 [93].

3D. O Padrão AVI

O AVI (*Audio Video Interleave*) definido pela Microsoft™ é um caso especial de RIFF (*Resource Interchange File Format*). Os arquivos tipo RIFF são um formato de propósito geral para intercâmbio de recursos de dados multimídia inventado pela Electronic Arts em 1984 e definido pela antiga aliança entre Microsoft™ e IBM™. O AVI é o mais comum formato de áudio e vídeo utilizados em microcomputadores, na realidade é um padrão de “fato” [91], [92].

A Microsoft™ criou um sistema parcialmente unificado para tratar multimídia em ambientes Win16™ e W32™. Este sistema consiste de uma interface de controle de alto nível denominada MCI (*Media Control Interface*) que é formada pela API (*Application Program Interface*) e seus “*drivers*” associados.

Este sistema multimídia do Windows™ possui algumas APIs de baixo nível, como a API tipo WAVE para tratar do áudio, que é anterior ao AVI.

Para o sistema NT 4.0™, o vídeo para Windows™ é implementado como uma coleção de DLL's de 32 bits. No Microsoft 32-bit como “*Common Object File Format*” ou formato COFF. Normalmente esses arquivos estão localizados no diretório \WINNT\SYSTEM32 onde o Windows NT™ armazena a maior parte dos arquivos DLL's.

Os arquivos são os seguintes [91], [92] e [95]:

MSVFW32.DLL	(<i>Microsoft™ Video for Windows DLL - NT 4.0</i>)
AVIFIL32.DLL	(<i>AVIFILE API for Reading and Writing AVI Files and Streams</i>)
AVICAP32.DLL	(<i>AVI Capture Window Class</i>)
MCI32.DLL	(<i>Video for Windows™ MCI Driver</i>)
MSACM32.DRV	(<i>Microsoft™ Audio Compression Manager</i>)
MSACM32.DLL	(<i>more Microsoft™ Audio Compression Manager</i>)
MSRLE32.DLL	(<i>Microsoft™ RLE Video Codec</i>)
IR32_32.DLL	(<i>Intel™ Indeo 3.2 Video Codec</i>)
MSVIDC32.DLL	(<i>Microsoft™ Video 1 Codec</i>)
ICCVID.DLL	(<i>Cinepak for Windows 32™ - Radius</i>)

Mais informações consultar o “*site*”: <http://www.microsoft.com/>

Apêndice E

Simulador

1E. Introdução

Neste estudo foram utilizados dois simuladores. O primeiro, projetado e implementado por Patrick R. Morin e J. Neilson da Universidade de Carleton, Canadá [43], foi usado para obter a banda efetiva simulada para uma única conexão ATM, apresentada no Capítulo 5. O segundo, projetado e implementado por Antônio M. Alberti e Ernesto Luiz Andrade Neto [87] e [88] da UNICAMP/FEEC/DECOM, foi utilizado para convalidar o método proposto por perda no estudo de tráfego multiplexado, também no Capítulo 5. Para maiores informações sobre este simulador consultar www.mc21.unicamp.br.

2E. Simulador Morin

Este simulador foi escrito em linguagem C e baseado na biblioteca de simulação Parasol. Este simulador foi construído com objetivo de avaliar o impacto de tráfego auto-similar no desempenho de comutadores ATM. O comutador ATM é modelado através de uma fila com disciplina FIFO (*First-in-First-out*), com “buffer” finito, conforme ilustrado na Figura 1E.

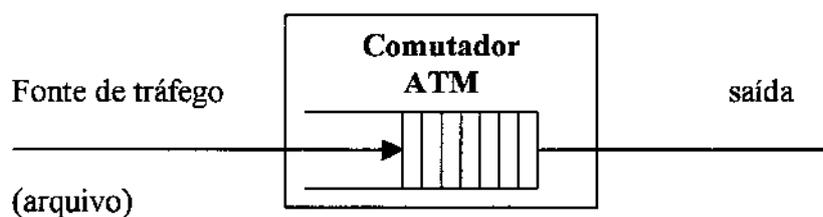


figura 1E “Buffer” no comutador ATM

As fontes de tráfego são organizadas em arquivos, cada um composto de uma seqüência de números inteiros x_0, x_1, x_2, \dots , onde x_i , representa o número de bytes enviados em intervalos sucessivos a cada 10 mseg. Esses pacotes de dados são convertidos em células ATM de 48 bytes de carga, que em seguida são enviadas de modo uniforme até entrada do comutador ATM.

O simulador monitora a taxa de perda de células, a utilização ou carga do “*buffer*” e o retardo (latência) das células provocado pelo “*buffer*”, que é dada pelo intervalo de tempo em que a célula acessa e deixa o “*buffer*”.

Para usar simulador são necessários os seguintes passos:

- os arquivos de tráfego disponíveis em formato ASCII como entrada;
- os dados necessários para configurar o simulador são: nome do arquivo, tempo de simulação, banda e tamanho do “*buffer*” a ser utilizado;
- os dados de saída do simulador são: atraso das células, perda de células e utilização do “*buffer*”.

Apêndice F

Instrumentos Utilizados para Captura do Tráfego

1F. Introdução

Dois analisadores de protocolo da Wandel & Goltermann™ foram utilizados para captura dos arquivos de tráfego apresentados no Capítulo 4. Em ambientes WAN, como rede VSAT, rede Frame-Relay, rede IP foi utilizado o modelo DA310 e ambientes LAN como Ethernet o modelo DA-320.

Para maiores informações sobre os referidos instrumentos de medição indicamos consultar o “*site*” da WG na Internet: <http://germany.wg.com/products/domino/>

2F. Wandel & Goltermann™ DA-310 Domino WAN

O analisador DA-310 destinado a monitorar e analisar ambientes WAN. Possui as seguintes interfaces RS-232/V.24, V.35, RS-449/V.36, RS-530, and V.11), T1, 2Mbit/s-ETSI e RDSI básica. Este analisador possui capacidade de decodificar mais de 300 protocolos, como X.25 and Frame-Relay. Possui funcionalidade BERT.

3F. Wandel & Goltermann™ DA320 Domino LAN

O analisador DA-320 é destinado à monitoração e análise de ambientes, como Ethernet e Token Ring com capacidade de filtrar 3 camadas de endereços de rede, decodificando mais de 300 protocolos. O DA-320 possui capacidade para gerar tráfego Ethernet e Token Ring (4 and 16 Mbit/s).