Modelagem Determinística de Tráfegos de Vídeo Comprimido no Padrão MPEG

por

Marcelo Dias de Amorim

UFRJ

Tese submetida para a obtenção do título de

Mestre em Ciências em Engenharia Elétrica

ao Programa de Pós-Graduação de Engenharia Elétrica

da COPPE/UFRJ

Modelagem Determinística de Tráfegos de Vídeo Comprimido no Padrão MPEG

Marcelo Dias de Amorim

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Aprovada por:

Prof. Otto Carlos Muniz Bandeira Duarte, Dr.Ing.

Prof. José Ferreira de Rezende, Dr.

Prof. Weiler Alves Finamore, Ph.D.

RIO DE JANEIRO, RJ - BRASIL

MARÇO DE 1998

AMORIM, MARCELO DIAS DE

Modelagem Determinística de Tráfegos de Vídeo Comprimido no Padrão MPEG [Rio de Janeiro] 1998

XIV, 75 p. 29,7 cm (COPPE/UFRJ, M.Sc., Engenharia Elétrica, 1998)

Tese - Universidade Federal do Rio de Janeiro, COPPE

1. Redes de Alta Velocidade
2. Modelagem Determinística de Tráfegos
3. Controle de Admissão de Conexões

I. COPPE/UFRJ II. Título (Série)

Aos meus familiares e amigos.

Agradecimentos

À minha família.

Ao meu orientador, Professor Otto Carlos Muniz Bandeira Duarte.

Aos demais membros da banca examinadora, José Ferreira de Rezende e Weiler Alves Finamore. Também ao Doutor Hitoshi Nagano que não pôde estar presente.

Aos amigos do GTA: Emerson PS, Belem, Hermes, LuisH, Rubi, Antguria, Andreamb, Adriana, Daniel, Capitão Fred, Bill, Alexszt, Marcos, Ronaldo, Vidal, Sílvia, Marcial, Jonny, Renata, Flávio, Abdalla, Celio, Faerman, Marlos, Ildemar, Spíndola, Sérgio, Gilberto, Iniciações Científica (Paulo, Artur, Rico, Bernardo, Tadeu, Doc, Gardel, Leandro, Luna, Godinho), Mônica, Jorges, Guilherme, Nacao, Galera do Ravel, Lemos, Marcelo Lisboa, Jucele, Ezequiel, Freire, Fábio, Flávio, Profs. Aloysio, Leão e Lanza.

Aos amigos do GEP: Evandro, Tony, Rosas, Marcelo, Pedro, Luis Oscar.

Aos amigos do LPC e LPS: João Marcelo, Salomão, De la Vega, Coutinho, Prof. Vladimir, Castanhol, Prof. Mesquita.

Aos funcionários: Wilson, Wanderlei, Fidel, Paulo Roberto, Bia, Solange, Patrícia, Daisy.

Aos demais amigos: Alexandre, Vaz, Jaca, Ana Paula, Rejane, Rosaldo, Milton, Biba, Cristiano, Charles, Rocha, Naldinho e Esposa, Gerlan, Claus, Siome, Fernando Abelha, Velho, Cagi, Patrícia (Sra Hermes), Carmen (Sra Bill), Claudinha (Sra Rubi), Mauros, Gelson, galera da Letras, Tia Rosa, Toda a Galera da Graduação, Profs da Graduação, Galera do Cavaco, Noël, e mais todo mundo.

Resumo da tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências em Engenharia Elétrica (M.Sc.)

Modelagem Determinística de Tráfegos de Vídeo Comprimido no Padrão MPEG

Marcelo Dias de Amorim

Março de 1998

Orientador: Otto Carlos Muniz Bandeira Duarte
Programa: Engenharia Elétrica

Nas futuras redes de serviços integrados, algumas aplicações serão tão exigentes que a rede deverá ser capaz de oferecer serviços com garantias determinísticas de desempenho. Ao se prover tal serviço, é fundamental que o modelo do tráfego seja o mais acurado possível.

Os modelos determinísticos de tráfegos propostos na literatura são voltados para caracterização de tráfegos VBR em geral. Na modelagem de uma importante classe de tráfegos, o vídeo comprimido no padrão MPEG, estes modelos deixam de capturar algumas características intrínsecas a este padrão. Este trabalho apresenta um estudo das fontes de vídeo MPEG sob diversos aspectos, identificando propriedades que, se inseridas no contexto de um modelo específico de tráfego, poderiam resultar em bons índices de acurácia de parametrização com menos esforço computacional.

Assim, este trabalho propõe dois novos modelos determinísticos para caracterização de tráfegos de vídeo comprimido no padrão MPEG: o modelo GOP-B (*GOP-Based model*) e sua extensão, o modelo XGOP-B (*eXtended GOP-Based model*). As principais vantagens destes modelos são com relação à automatização no processo de extração de parâmetros de tráfegos, realizado de forma muito simples. Os resultados mostram que, particularmente para o modelo XGOP-B, altos índices de acurácia de parametrização podem ser obtidos.

Abstract of the thesis presented to COPPE/UFRJ as partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

Deterministic Modelling of MPEG Compressed Video Traffics

Marcelo Dias de Amorim

March 1998

Supervisor: Otto Carlos Muniz Bandeira Duarte
Department: Electrical Engineering

In the future, integrated services networks will have to support several applications with different quality of service requirements. In some cases, the guarantees will have to be so strong that deterministic services will have to be available. When providing such a service, sources must be modelled by an accurate deterministic traffic model.

Some deterministic traffic models have been proposed in the literature. However, they were created to the modelling of general variable bit rate sources. For a very important sort of traffics, the MPEG compressed video, these models can not capture some intrinsic properties belonging to this type of sources. This work developes a behavioural analysis of MPEG video sources and identifies specific characteristics that, if considered by a deterministic traffic model, could result in high modelling accuracy with less computational efforts.

In this scenario, this work proposes two novel deterministic traffic models for MPEG-compressed video sources: the GOP-Based model (GOP-B) and its extension, the eXtended GOP-Based model (XGOP-B). The main advantage of these models is the automatic parameters extractor, which is not possible to the other models. The parameters are obtained in a very simple way. The results show that, specially to the XGOP-B model, high modelling accuracy can be obtained.

Palavras-chave

Redes de Alta Velocidade

Garantia de Qualidade de Serviço

Modelagem Determinística de Tráfegos

Controle de Admissão de Conexão

Vídeo MPEG

Glossário

QoS :	Qualidade de Serviço (*Quality of Service*);
CAC :	Controle de Admissão de Conexão (*Connection Admission Control*);
ATM :	Modo de Transferência Assíncrono (*Asynchronous Transfer Mode*);
SMG :	Ganho de Multiplexação Estatística (*Statistical Multiplexing Gain*);
CBR :	Taxa Constante de Bits (*Constant Bit Rate*);
UBR :	Taxa Não Definida de Bits (*Unknown Bit Rate*);
ABR :	Taxa Disponível de Bits (*Available Bit Rate*);
VBR-nrt :	Taxa Variável de Bits – não tempo real (*Variable Bit Rate – non real time*);
VBR-rt :	Taxa Variável de Bits – tempo real (*Variable Bit Rate – real time*);
MPEG :	*Moving Picture Expert Group*;
GOP-B :	Modelo Determinístico Baseado no Grupo de Quadros (*GOP-Based model*);
XGOP-B :	GOP-B Estendido (*eXtended GOP-Based model*);
D-BIND :	*Deterministic Bounding INterval Dependent*;
GOP :	Grupo de Quadros (*Group Of Pictures*);
FCFS :	Primeiro a Chegar Primeiro Servido (*First-Come First-Served*);
SP :	Prioridade Estática (*Static Priority*);

Sumário

Resumo **v**

Abstract **vi**

Glossário **viii**

1 Introdução **1**

1.1 Rede de Serviços Integrados 4

1.2 Garantia de Qualidade de Serviço 6

1.3 Modelagem Determinística de Tráfegos 7

1.4 Organização da Tese 9

2 Caracterização Determinística de Tráfegos **10**

2.1 Função de Restrição de Tráfego 12

2.2 Modelo de Alocação pelo Pico 14

2.3 Modelo Fluido (σ, ρ) 16

2.4 Modelo Fluido em Duas Dimensões $(\vec{\sigma}, \vec{\rho})$ 18

2.5 Modelo Discreto $(x_{min}, x_{ave}, I, s_{max})$ 19

2.6 Modelo D-BIND 21

2.7 Discussão 22

3 Modelagem Determinística de Tráfegos de Vídeo MPEG Baseada em Grupo de Quadros 24

3.1 O Padrão MPEG 25

3.2 Análise Microscópica 27

3.3 Modelo GOP-B 28

3.4 Análise Macroscópica 31

3.5 O Modelo XGOP-B 34

4 Análise dos Resultados 38

4.1 Extrator de Parâmetros de Tráfego 41

4.1.1 Extrator (σ, ρ) 41

4.1.2 Extrator $(\vec{\sigma}, \vec{\rho})$ 42

4.1.3 Extrator $(x_{min}, x_{ave}, I, s_{max})$ 43

4.1.4 Extrator D-BIND 43

4.1.5 Extrator GOP-B 44

4.1.6 Extrator XGOP-B 44

4.2 Resultados Obtidos 45

4.3 Influência do fator β 48

5 Conclusões 52

Referências Bibliográficas 56

A CAC Determinístico 61

B Comportamento das Amostras de Vídeo **66**

C Modelagem das Amostras de Vídeo **70**

Lista de Figuras

2.1 Alocação pelo pico. 15

2.2 Função de restrição de tráfego para o modelo de alocação pelo pico. . 15

2.3 Função de restrição de tráfego para o modelo (σ, ρ). 17

2.4 Função de restrição de tráfego para o modelo $(\vec{\sigma}, \vec{\rho})$. 18

2.5 Função de restrição de tráfego para o modelo *Discreto*. 20

2.6 Função de restrição de tráfego para o modelo *D-BIND*. 22

3.1 Codificação interquadros. 26

3.2 Amostra de um vídeo no padrão MPEG. 27

3.3 Função exata para vídeo MPEG. 27

3.4 Exemplos de traços de vídeos. 28

3.5 Função de restrição de tráfego para o modelo *GOP-B*. 30

3.6 Amostras de vídeo MPEG. 31

3.7 Efeito dos pontos de quebra sobre a função exata. 32

3.8 Exemplo de Primeira Derivada. 33

3.9 Exemplo de Segunda Derivada. 33

3.10 Classificação em faixas de uma função exata. 35

3.11 Inclusão de θ como parâmetro de tráfego. 36

3.12 Função de restrição de tráfego para o modelo *XGOP-B*. 37

4.1 Exemplos de traços de vídeos MPEG. 40

4.2 Gráfico de Perda por Aproximação do Modelo. 48

4.3 Curvas de restrição de tráfego para dois trações de vídeo. 49

4.4 Fator β $\times$ número de pontos de quebra. 50

4.5 Fator β $\times$ PAM. 51

A.1 Conceitos: atraso, *backlog* e período ocupado. 62

Lista de Tabelas

4.1 Características dos vídeos utilizados. 39

4.2 Perda por Aproximação do Modelo para cada um dos modelos. 46

4.3 Perda por Aproximação do Modelo para cada um dos modelos. 47

Capítulo 1

Introdução

ANTERIORMENTE , as aplicações em redes de computadores eram caracterizadas basicamente pela transferência de informações sensíveis mais às perdas de dados que às variações observadas no tempo de tranferência dos pacotes. Essas aplicações, como por exemplo correio eletrônico, *login* remoto e transferência de arquivos, continuam existindo, mas agregam constantemente outras aplicações cujas exigências feitas à rede são outras.

A área de telecomunicações é uma das que mais se desenvolvem na atualidade. A evolução na tecnologia dos equipamentos, como integração em *hardware* e comunicações ópticas, bem como a técnica de desenvolvimento de *softwares*, vêm possibilitando um constante avanço nas velocidades de processamento e transmissão de informações por todo o planeta. Em paralelo, o número de usuários de computadores e da Internet, mesmo a nível domiciliar, vem aumentando de forma exponencial. O resultado deste quadro é o crescimento da demanda por melhores serviços e o surgimento de aplicações que se sentem prejudicadas com as políticas tradicionais de utilização da banda-passante da rede.

Este novo conjunto de aplicações inclui aplicações multimídias interativas com vários participantes (vídeo-conferência ou conferência multimídia, telesseminário), visualização científica (reações nucleares, imagens de satélites), visualização médica e telediagnóstico, distribuição doméstica de TV, *home-cinema* (*pay-per-view*), vídeo-

telefonia, tele-educação (vídeo-biblioteca, educação domiciar distribuída), realidade virtual, trabalho cooperativo, transferência de dados a altas velocidades etc. Estima-se ainda que grande parte das aplicações a serem utilizadas no futuro e que serão responsáveis pela maior parte do tráfego veiculado na rede ainda nem foram concebidas.

A maioria das aplicações citadas é caracterizada por introduzir rígidas restrições quanto às garantias de qualidade de serviço (QoS - *Quality of Service*) outrora não prioritárias, tais como vazão, atraso máximo e variação no atraso (*jitter*), além da taxa de perda de pacotes. Acrescido a este fato, o nível de QoS exigido é extremamente variável. Desta forma, a rede precisa prover serviços heterogêneos com diferentes garantias de qualidade de serviço oferecidas por conexão.

Dois tipos de serviços têm sido propostos na literatura para suportarem as novas aplicações [1, 2, 3]: os serviços garantidos e os serviços preditos. Em um modelo de serviço garantido [1, 2], os parâmetros de qualidade de serviço são fornecidos *a priori*, independentemente do comportamento das demais conexões. Em um serviço predito [3], limites de atraso e um ponto de *playback* são fornecidos pelo cliente. Esses limites de atraso e o ponto de *playback* podem variar e a qualidade do serviço pode ser afetada devido às flutuações na rede. As aplicações que utilizam serviços preditos devem poder se adaptar às variações no ponto de *playback* e tolerar momentâneas rupturas na qualidade do serviço.

Existem dois importantes fatores a serem considerados para a escolha entre os dois serviços: a qualidade do serviço e o custo em prover tal serviço [4, 5]. Do ponto de vista do cliente da comunicação, a qualidade do serviço deve ser priorizada, uma vez que limites de desempenho são garantidos e não existem rupturas no serviço. Do ponto de vista da rede, o custo em prover um serviço depende da máxima utilização possível dos recursos da rede. Embora argumente-se que os serviços preditos resultam em melhor aproveitamento da rede por parte dos tráfegos em tempo real, ainda não existe nenhuma comparação quantitativa realizada, uma vez que algoritmos para serviços preditos ainda não estão completamente descritos, especialmente para controle de admissão de conexões.

Apesar da grande diversidade nas características das aplicações nas redes de serviços integrados, sobretudo nas aplicações multimídias, existe um requisito comum com relação ao serviço a ser provido pela rede. As aplicações exigem a garantia sobre os parâmetros de qualidade de serviço contratados.

De forma a prover tal garantia, é fundamental que recursos sejam alocados no caminho de transferência das informações. Um controle refinado sobre os recursos alocados deve atuar de forma que uma solução "pobre" não resulte em uma de duas condições indesejadas. Se recursos forem superalocados para cada conexão, a rede será subutilizada, o que refletirá diretamente no custo da comunicação, uma vez que o usuário será onerado com as despesas relativas à banda-passante não utilizada. Por outro lado, se recursos forem subalocados, não haverá desperdício de recursos e as conexões terão uma certa qualidade de serviço, mas receberão uma quantidade insuficiente de banda-passante, não satisfazendo de forma plena a qualidade de serviço exigida pela aplicação.

Assim, o desafio é alocar recursos de uma maneira eficiente que atinja a mais alta taxa de utilização possível e, ao mesmo tempo, satisfazer os requisitos de qualidade de serviço de cada uma das conexões individuais. Em outras palavras, deve-se simultaneamente prover a garantia de desempenho que as aplicações exigem e a alta taxa de utilização da banda-passante exigida pela rede.

A tarefa de controlar os recursos da rede é realizada pelo controle de admissão de conexão (CAC - *Connection Admission Control*) [6] e pelos mecanismos de policiamento de tráfego [7], responsáveis, respectivamente, pela determinação das conexões que podem ser estabelecidas (aceitas) e pela garantia que fontes abusivas não violarão os parâmetros de tráfego contratados.

As eficiências do controle de admissão de conexões e do policiamento de tráfego dependem principalmente de três componentes [1, 4, 5]: o modelo do tráfego, o escalonamento de pacotes e os testes de controle de admissão. O modelo de tráfego [8, 9, 10, 11, 12, 13, 14, 15, 16] é responsável pela parametrização das características do tráfego, objetivando sempre fazê-la da forma mais acurada possível, a fim de que os mecanismos de controle de admissão não superestimem ou subestimem os recursos

necessários para a conexão. O escalonamento de pacotes [7, 17, 18] assume as tarefas de multiplexação e demultiplexação dos pacotes das várias conexões em um nó da rede. O controle de admissão [1, 2, 4, 5, 6, 19] compreende testes que verificam a viabilidade do estabelecimento de uma nova conexão, dentre eles os testes de capacidade de transmissão, velocidade de processamento de CPU, espaço disponível em *buffers* e atraso máximo sofrido pelos pacotes no caminho fonte-destino.

Os pontos chaves nos esquemas de alocação de recursos são justamente o modelo do tráfego e o controle de admissão de conexões. A eficácia do algoritmo de CAC depende criticamente da escolha do modelo parametrizado que a fonte utiliza para especificar as características do tráfego a ser submetido à rede. Fundamentalmente, sem uma adequada descrição do tráfego, recursos não podem ser eficientemente alocados.

O restante deste capítulo se desenvolverá da seguinte forma. Serão apresentados brevemente os propósitos da rede de serviços integrados. Em seguida, serão situados os requisitos para se prover garantias de qualidade de serviço nas redes de serviços integrados. Dados os conceitos, este capítulo conclui com uma primeira discussão sobre o tema central desta tese, a modelagem determinística de tráfegos, alicerce para se prover garantias determinísticas de qualidade de serviço.

1.1 Rede de Serviços Integrados

Historicamente, redes independentes têm sido desenvolvidas para prover diferentes serviços de naturezas distintas. Por exemplo, a rede telefônica oferece um serviço adequado para transferência de voz. Contudo, não suporta de maneira eficiente tráfegos de vídeo de qualidade razoável, devido à característica variável ao longo do tempo e à exigente qualidade de serviço dos fluxos de vídeo. Assim, foram desenvolvidos sistemas para transporte de vídeo. Da mesma forma, dado que nem as redes telefônicas nem as redes de vídeo são totalmente adequadas para transporte de dados provenientes de aplicações que geram tráfegos em rajada, redes de computadores separadas têm sido desenvolvidas para suportar com eficiência tais fluxos

de dados.

O objetivo de uma rede de serviços integrados é prover múltiplos serviços em uma única infraestrutura de rede. Suportar diversas aplicações em uma única rede traz as vantagens em termos de eficiência, multiplexação estatística e economia de escala. Tais redes podem suportar aplicações com os mais diferentes requisitos de QoS, não sendo limitadas a um restrito número de serviços.

Como solução para a implementação do modo de transferência para as redes digitais de serviços integrados, o ITU-T (*International Telecommunication Union*) propôs o Modo de Transferência Assíncrono (ATM - *Asynchronous Transfer Mode*) [20]. O ATM utiliza uma abordagem orientada a conexão baseada em circuitos virtuais com transferência de pequenos pacotes de tamanho fixo, de 53 octetos, chamados células. O tamanho da célula foi assim escolhido de forma a se evitar excessivos atrasos, de transmissão e de empacotamento, em aplicações como voz. O tamanho fixo é justificado pelo objetivo de se permitir implementações menos expensivas em *hardware*.

Os primeiros componentes para se prover garantias de desempenho são o paradigma orientado a conexão, disciplinas de serviço nos comutadores e roteadores e o controle de admissão de conexões em conjunto com o modelo parametrizado do tráfego. No cenário apresentado por Ferrari [21], de forma a se obter uma conexão com garantias de desempenho, o cliente primeiramente especifica as características de seu tráfego e os requisitos de QoS mínimos suportados. O tráfego é especificado em termos de um modelo de tráfego parametrizado. Seus requisitos de QoS são expressos em termos de parâmetros de desempenho, como vazão, atraso, variação do atraso (*jitter*) e taxa de perda de pacotes. O tráfego a ser submetido e os parâmetros de QoS são então convertidos para parâmetros locais, quando uma série de testes de controle de admissão são realizados em cada nó da rede. Garantindo que requisitos locais de desempenho são obtidos em cada nó, garantias fim-a-fim são também satisfeitas.

Durante a transferência de dados, uma disciplina de serviços garante que a QoS prometida pela rede é realmente provida. Para disciplinas que utilizam a política

de prioridades, a QoS é obtida distinguindo-se fluxos em tempo real ou não, por exemplo transmitindo-se pacotes com limites menores de atraso antes daqueles com limites mais tolerantes.

Finalmente, o paradigma orientado a conexão possibilita uma maneira eficiente de se prover garantias, reservando-se recursos por conexões individuais. Além de serem orientadas a conexão, estas redes se servem da técnica de comutação de pacotes. Caso se optasse pela técnica de comutação de circuitos, recursos seriam desperdiçados se as fontes estivessem temporariamente inativas. Assim, os recursos não utilizados por uma conexão podem ser eventualmente aproveitados pelas demais.

1.2 Garantia de Qualidade de Serviço

O sucesso em se prover garantias de desempenho para uma determinada conexão é fundamentado no modelo de tráfego e no controle de admissão de conexões. Assim, um modelo pouco acurado ou um algoritmo de admissão ineficiente podem resultar em um processo de alocação de recursos de baixa qualidade. Uma conexão pode sofrer com a falta de banda-passante se o tráfego for subestimado. Por outro lado, o usuário pode ser prejudicado pela cobrança dos recursos para ele alocados e não utilizados, se o tráfego for superestimado.

De acordo com Ferrari e Verma [2], dois tipos de serviços garantidos são propostos: serviço determinístico e serviço estatístico. Em um serviço do tipo determinístico, todos os pacotes de uma conexão recebem a garantia de alcançar o destino com o limite da atraso respeitado e sem perda de pacotes devido a estouro de *buffers*. Em um serviço estatístico, garantias de desempenho também são fornecidas. Contudo, estas garantias não são absolutas, uma vez que existe uma probabilidade, por menor que seja, de serem violadas.

Para os serviços determinísticos, a reserva de recursos é feita de acordo com um cenário de "pior caso", onde a sobrealocação (*over booking*) não é possível. Desta forma, apesar do nível de utilização da rede ser limitado, este tipo de serviço é o

mais confiável possível que o cliente pode obter.

Nos serviços estatísticos, ao contrário, os recursos podem ser sobrealocados de uma maneira controlada de forma que não precisam ser reservados com base em um cenário de pior caso. Esta superposição de tráfegos é denominada multiplexação estatística e a métrica que a quantifica é chamada Ganho de Multiplexação Estatística (SMG - *Statistical Multiplexing Gain*). Em outras palavras, o cliente pode alcançar maior utilização da rede compensado pelo fato de poder sofrer com perdas ou atrasos excessivos sobre uma fração dos pacotes. Simplificadamente, o ganho é alcançado explorando-se a propriedade que flutuações estatísticas nas taxas dos fluxos individuais são suavizadas no agregado das conexões.

Deve existir uma forma de se determinar quando um tipo de serviço é mais indicado para uma dada comunicação. Esta escolha é totalmente dependente do tipo de aplicação em questão. Os serviços determinísticos não substituem os serviços estatísticos, nem vice-versa. Se uma aplicação for extremamente rigorosa com relação à qualidade de serviço e exigir que os parâmetros contratados sejam sempre respeitados, mesmo que para isso tenha que custear mais banda-passante, um serviço do tipo determinístico é indicado. Ao mesmo tempo, se uma outra aplicação suportar pequenos índices de degradação no serviço, uma garantia do tipo probabilística se mostra adequada, uma vez que será compensada com um custo mais baixo, por demandar menos recursos.

Apesar de inúmeros trabalhos terem explorado os aspectos estatísticos para garantia de qualidade de serviço, os estudos na área de garantias determinísticas de QoS mostram que ainda existe muita pesquisa a ser desenvolvida.

1.3 Modelagem Determinística de Tráfegos

Como já afirmado anteriormente, as garantias de qualidade de serviço recaem sobre dois princípios básicos: o modelo de tráfego e o controle de admissão de conexões. Este trabalho se concentra na modelagem determinística de tráfegos, em particular tráfegos de vídeo no padrão MPEG.

A caracterização de uma fonte em seu nível mais abstrato define as características de uma fonte e seus requisitos de QoS. As características do tráfego de uma aplicação são o conjunto mínimo de parâmetros que um usuário deverá fornecer ao gerenciador da rede para o controle efetivo do tráfego e a utilização máxima dos recursos da rede.

Para atender à diversidade de aplicações, várias classes de tráfegos foram definidas pelo ATM *Forum*: em taxa constante de bits (CBR - *Constant Bit Rate*), em taxa não definida de bits (UBR - *Undefined Bit Rate*), em taxa disponível de bits (ABR - *Available Bit Rate*) e em taxa variável de bits (VBR - *Variable Bit Rate*), sendo esta última subdividida em VBR-nrt (VBR *non real time*) e VBR-rt (VBR *real time*). A classe CBR foi definida para suportar aplicações com requisitos bem definidos de atraso e que geram tráfego a uma taxa constante, tal como uma emulação de circuitos. A classe ABR é adequada para aplicações com requisitos mínimos de vazão e atraso. A classe UBR foi definida para suportar aplicações que não possuem requisitos definidos de serviço e que estejam dispostas a tolerar qualquer nível de QoS que a rede possa prover, tal como uma transferência de arquivo feita em *background*. A classe de tráfegos VBR, assim como a CBR, também foi definida para suportar aplicações com requisitos bem definidos de atraso, mas que geram tráfego a uma taxa variável, tal como vídeo sob demanda.

A modelagem determinística de tráfegos do tipo VBR-rt introduz desafios únicos. Já foram propostos na literatura diversos modelos determinísticos de tráfego. Contudo, todos objetivam a caracterização de tráfegos VBR em geral. Algumas classes de tráfego, como por exemplo vídeo comprimido no padrão MPEG (*Moving Picture Expert Group*), são tão importantes que uma caracterização por um modelo genérico pode não ser adequada. A generalidade dos modelos estudados mostra que suas respectivas eficiências na caracterização de vídeos no padrão MPEG poderiam ser otimizadas. Será mostrado que o padrão MPEG apresenta características intrínsecas que, se fossem consideradas no momento da concepção de um modelo de tráfego específico, resultariam em uma caracterização com um número menor de parâmetros de tráfego e ainda assim obtendo bons índice de acurácia de parametrização.

Pelos motivos acima descritos, este trabalho introduz dois novos modelos determinísticos para caracterização de tráfegos de vídeo comprimido no padrão MPEG, denominados GOP-B (*GOP-Based model*) e XGOP-B (*eXtended GOP-Based model*). Estes modelos têm as vantagens de serem simples na extração dos parâmetros de tráfego e de obterem excelentes resultados com um número reduzido de parâmetros. Estas características são essenciais para que os mecanismos de controle de admissão de conexão e de policiamento de tráfego demandem pouco processamento. Deve-se destacar que os parâmetros para os modelos GOP-B e XGOP-B, ao contrário dos demais, são obtidos de forma automatizada e não subjetiva, como será mostrado nos próximos capítulos. Para análise e comparação dos modelos estudados e dos modelos propostos neste trabalho serão utilizadas diversas amostras de vídeo comprimido no padrão MPEG-1.

1.4 Organização da Tese

O restante desta tese está organizado da seguinte forma. No capítulo 2 são apresentados os conceitos e requisitos para modelagem determinística de tráfegos. São apresentados também os principais modelos determinísticos de tráfegos propostos na literatura. No capítilo 3 são investigadas as propriedades dos vídeos comprimidos no padrão MPEG. Com base neste estudo, são propostos os dois novos modelos. No capítulo 4, são feitas as devidas análises e comparações dos modelos pela aplicação destes em amostras de tráfegos de vídeo MPEG. Finalmente, são apresentadas as conclusões e propostas de trabalhos futuros no capítulo 5.

Capítulo 2

Caracterização Determinística de Tráfegos

PARA suportar os requisitos de transmissão de mídias contínuas, como áudio e vídeo, as redes de serviços integrados precisam prover às conexões garantias sobre alguns parâmetros de desempenho, tais como: vazão, atraso, variação no atraso e taxa de perda de pacotes. Para as aplicações mais exigentes em termos temporais, a rede deve oferecer garantias determinísticas de atraso. Exemplos de aplicações que podem exigir tais serviços incluem aplicações em vídeo de alta qualidade, vídeo sob demanda, robótica e controle, tele-medicina e algumas aplicações computacionais distribuídas.

A realização de um serviço distribuído envolve uma série de condições a serem respeitadas. Uma rigorosa aproximação para um serviço com limites de atraso determinados deve levar em consideração todos os tipos de atraso aos quais os pacotes podem ser submetidos, incluindo tempo de processamento das informações, atrasos de propagação na rede e atrasos variáveis nos nós da rede devido à multiplexação estatística.

De forma a prover serviço do tipo determinístico, é necessário que recursos sejam alocados de acordo com um cenário de "pior caso". Essencialmente, significa que recursos não podem ser sobrealocados. Desta forma, o controle de admissão de

conexão para um serviço determinístico requer funções limites de comportamento para a fontes individuais e, por conseqüência, deve ser capaz de determinar como o agregado de conexões, quando multiplexadas nos nós da rede, interagem entre si, considerando ainda um cenário de pior caso.

A pergunta chave a ser respondida é: seria possível prover garantias determinísticas de desempenho sem que isso resulte em subutilização dos recursos da rede?

Sabe-se que alcançar altas taxas de aproveitamento dos recursos da rede é uma das mais importantes considerações a serem feitas ao se prover um serviço garantido, seja ele do tipo determinístico ou não. Para prover um serviço garantido, recursos devem ser reservados para as conexões. Em geral, esquemas de reserva de recursos possibilitam a obtenção de altas taxas de utilização da rede para tráfegos suavizados [22, 23, 24, 25, 26]. Contudo, muitos clientes que exigem garantias de desempenho possuem tráfegos em rajada, como por exemplo vídeo comprimido [27, 28]. Se a reserva for feita com base na alocação de recursos pela taxa de pico de cada tráfego, novas requisições serão rejeitadas quando a soma das taxas de pico das conexões já estabelecidas igualar à velocidade máxima do canal. Neste caso, a rede será subutilizada se a relação $taxa_{pico}/taxa_{média}$ dos tráfegos individuais for alta.

É importante se encontrar soluções que possam prover garantias determinísticas de desempenho mesmo quando a soma das taxas de pico das conexões for superior à velocidade do canal. Em [1], são apresentadas metodologias que mostram que garantias determinísticas podem ser providas mesmo nestas condições, e razoável nível de utilização médio da rede pode ser alcançado, mesmo quando os tráfegos são do tipo em rajada.

O principal fundamento para se realizar eficientemente reserva de recursos é a obtenção para o tráfego de um modelo parametrizado o mais acurado possível. No caso de um serviço determinístico, a maneira como a fonte especifica seu tráfego em pior caso é através de um modelo determinístico de tráfego. Esta parametrização é utilizada pelos mecanismos de controle de admissão para realização dos cálculos necessários para estabelecimento da nova conexão requisitada. Independentemente da

qualidade dos algoritmos de admissão de conexão, os resultados serão satisfatórios, ou seja, a banda-passante da rede será eficientemente alocada, se o modelo do tráfego o representar da forma mais acurada possível.

A modelagem de tráfegos para um serviço determinístico tem alguns requisitos fundamentais. Primeiro, o modelo deve ser uma caracterização em *pior caso* da fonte de forma a prover limites superiores de comportamento. Segundo, o modelo precisa ser facilmente *parametrizado*, permitindo assim que a fonte repasse eficientemente à rede as informações sobre o tráfego. Terceiro, o modelo deve caracterizar o tráfego o mais *acuradamente* possível de forma que o algoritmo de admissão de conexão não superestime os recursos destinados ao tráfego. Finalmente, o modelo deve ser *policiável* para que a rede possa evitar que tráfegos mal comportados violem os parâmetros de QoS contratados.

Vários modelos determinísticos de tráfegos VBR têm sido propostos na literatura. Este trabalho apresenta uma análise sobre quatro destes modelos: o modelo $(x_{min}, x_{ave}, I, s_{max})$, ou modelo *Discreto*, o modelo (σ, ρ), também chamado balde-furado ou modelo *Fluido*, o modelo $(\vec{\sigma}, \vec{\rho})$, conhecido por modelo *Fluido em Duas Dimensões*, e o modelo *D-BIND*.

A modelagem determinística de tráfegos VBR é toda fundamentada na chamada *função de restrição de tráfego*, detalhada na seção seguinte. Na seqüência serão apresentados os modelos propostos na literatura.

2.1 Função de Restrição de Tráfego

Uma caracterização é dita determinística quando for capaz de representar o tráfego em todos os casos, ou seja, estando ele em qualquer taxa de transmissão. Esta caracterização permite uma análise de pior caso, provendo limites de vazão e atraso bem determinados. Especificamente, a caracterização determinística de tráfego define o que se chama função determinística de restrição de tráfego. Uma função $\mathcal{F}(\cdot)$ é denominada **função de restrição de tráfego** (*traffic constraint function*) [11] se durante *qualquer* intervalo de tempo de duração δt não forem

gerados pela fonte mais que $\mathcal{F}(\delta t)$ bits. De outro ponto de vista, considerando $\mathcal{G}[t, t+\delta t]$ o número real de bits gerados no intervalo $[t, t+\delta t]$, a equação

$$\mathcal{G}[t, t+\delta t] \leq \mathcal{F}(\delta t), \qquad t > 0, \tag{2.1}$$

deve ser satisfeita. A função de restrição de tráfego $\mathcal{F}(\cdot)$ provê um limite invariante do tempo em torno da função $\mathcal{G}$, de forma que a fonte é limitada para todo intervalo de duração δt.

Como a única restrição sobre a função $\mathcal{F}(\cdot)$ é que ela seja um contorno superior ao processo de geração de bits pela fonte, existe um número infinito de possíveis funções de restrição de tráfego. Um modelo determinístico de tráfego seleciona entre todas as possíveis funções, uma família de soluções com comportamento específico. Por este motivo, a comparação entre as acurácias dos modelos é realizada de fato pela comparação entre as suas respectivas funções de restrição de tráfego.

É importante realçar que, para uma determinada fonte gerando tráfego de acordo com um processo $\mathcal{G}$, a função de restrição de tráfego mais precisa para a modelagem deste tráfego para qualquer intervalo de tempo de duração δt é por definição a função

$$\zeta(\delta t) = sup\{\mathcal{G}[t, t+\delta t]\}, \qquad t \geq 0, \tag{2.2}$$

denominada contorno empírico (*empirical envelope*) em [4, 5], processo de contorno mínimo (*minimum envelope process*) em [29] e função exata em [12]. Por exemplo, se fosse capturada uma amostra de um processo $\mathcal{G}[0, t]$ e fossem realizados todos os cálculos necessários para se obter a mais acurada caracterização determinística para o tráfego, o resultado seria a função ζ.

Contudo, representar o comportamento de tráfegos VBR única e exclusivamente através da função exata $\zeta(\cdot)$ é inconcebível, pela grande quantidade de parâmetros que seriam necessários. Esta escolha violaria um dos princípios da modelagem determinística de tráfegos que diz que o modelo deve ser facilmente parametrizável. Desta forma, para que a função de restrição $\mathcal{F}(\cdot)$ relativa a um modelo seja um

limite superior invariante no tempo em torno do processo $\mathcal{G}$ de geração de bits, ela precisa ser uma aproximção superior para a função exata ζ, ou seja,

$$\mathcal{F}(\delta t) \geq \zeta(\delta t), \qquad \forall \delta t. \tag{2.3}$$

Como a função de restrição de tráfego relativa ao modelo de tráfego deve corresponder a uma curva superior à função exata ζ e sendo esta última o mais alto grau de acurácia de parametrização, o objetivo fundamental passa a ser a concepção de um modelo que resulte em uma função de restrição de tráfego $\mathcal{F}$ a mais próxima possível da função exata ζ. Analiticamente, a expressão

$$\zeta(\delta t) - \mathcal{F}(\delta t), \qquad \forall \delta t, \tag{2.4}$$

deve ser minimizada. Será visto adiante que a expressão 2.4 servirá como base de comparação entre os modelos de tráfego. A razão principal para se procurar uma solução que se aproxime ao máximo da função exata é muito simples. Para que o controle de admissão de conexão opere de forma a maximizar o número de conexões admitidas, é mandatório que a subtração $\zeta(\delta t) - \mathcal{F}(\delta t)$ seja minimizada.

O restante deste capítulo está reservado para a apresentação dos modelos determinísticos de tráfego propostos na literatura.

2.2 Modelo de Alocação pelo Pico

Dentre todas as possibilidades de modelagem de tráfegos VBR, o esquema de alocação pelo pico consiste na forma mais básica possível. O modelo de alocação pelo pico é intrinsecamente determinístico, sendo o mais simples por utilizar apenas um parâmetro para modelagem do tráfego: a taxa de pico T_{pico}.

Esta forma de modelagem é a mais indicada para fontes de tráfego da classe CBR, que geram bits em taxa média constante e igual à taxa de pico. Para caracterização de tráfego em taxa variável de bits (VBR), a modelagem baseada na taxa de pico

pode resultar em um nível de acurácia extremamente baixo, sobretudo se a relação $taxa_{máxima}/taxa_{média}$ do tráfego for alta, como pode ser visto na figura 2.1.

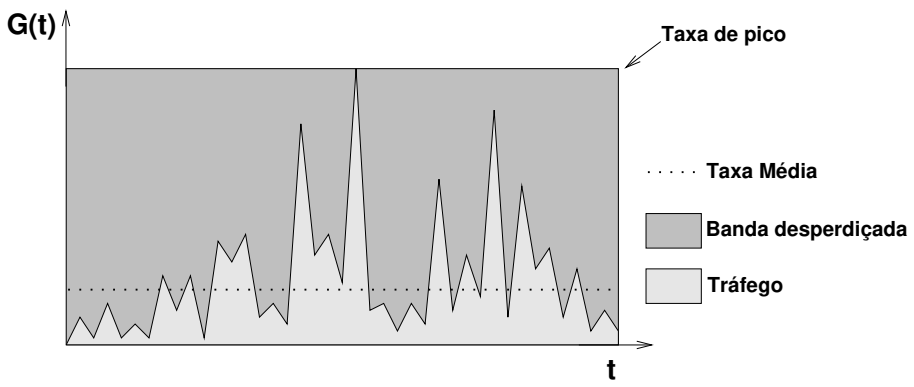


Figura 2.1: Alocação pelo pico.

A função de restrição de tráfego correspondente ao esquema de alocação pelo pico é dada pela equação

$$\mathcal{F}_{pico}(\delta t) = T_{pico} \times \delta t. \tag{2.5}$$

Graficamente, a curva de restrição de tráfego para este modelo é apresentada na figura 2.2. Pode-se observar pelo gráfico que a discrepância entre as duas curvas comprova a ineficácia do modelo de alocação pelo pico.

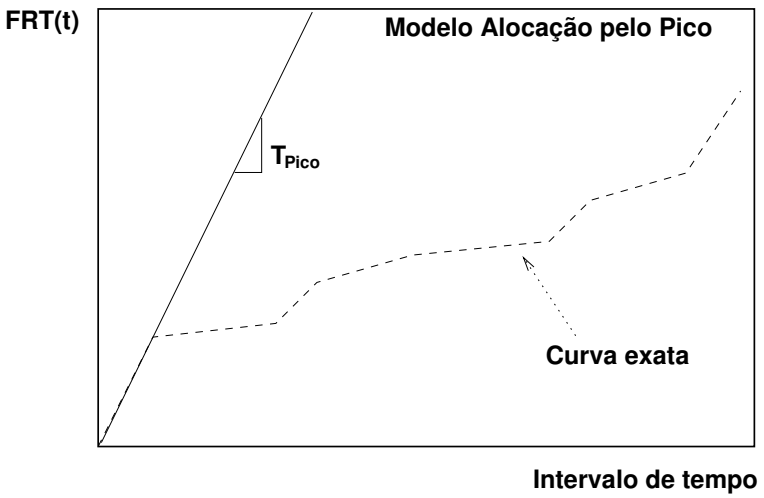


Figura 2.2: Função de restrição de tráfego para o modelo de alocação pelo pico.

2.3 Modelo Fluido (σ, ρ)

O modelo Fluido, ou (σ, ρ), proposto em [11, 30], assume como premissa básica que o tráfego submetido à rede possui características até certo ponto "desconhecidas", mas satisfaz alguns contornos regulares. Estes contornos têm o efeito de limitar as características de rajada do tráfego, procedimento denominado "*burstiness constraints*".

Os limites de rajada considerados atuam na caracterização de tráfegos da seguinte maneira. Dado um número positivo ρ, existe um outro número σ (possivelmente infinito) tal que se o tráfego alimentar um servidor que trabalhe a uma taxa ρ, o máximo número de bits contidos no *buffer* em qualquer instante t não ultrapassa σ.

O modelo (σ, ρ) é por vezes chamado modelo balde-furado (*leaky-bucket*), em analogia ao modelo de policiamento de tráfego de mesmo nome. No algoritmo balde-furado, um contador é mantido para um tráfego sendo submetido à rede. A cada transmissão de um pacote, o valor do contador é comparado a uma constante pré-especificada (número máximo de créditos). Se o contador for menor que esta constante, o pacote é admitido na rede. Caso contrário, o pacote é descartado ou marcado como violador. Enquanto o contador permanecer positivo, ele é decrementado de acordo com uma certa taxa fixa.

Em sua essência, o modelo fluido é uma extensão muito simples ao modelo de alocação pelo pico, a menos do fator σ. Esta variável fica responsável pela modelagem de "instantes de rajada", permitindo que o valor escolhido para ρ seja quase sempre muito inferior ao fator T_{pico} do modelo de alocação pelo pico. Por estes motivos, o parâmetro σ é chamado *fator de rajada* e ρ denominado *fator de vazão*.

Resumindo o comportamento do modelo fluido, a dupla (σ, ρ) exprime que em um intervalo de tempo de duração δt não são gerados pela fonte mais que $\sigma + \rho \times \delta t$ bits. Desta forma, a função de restrição de tráfego correspondente é dada por

$$\mathcal{F}_{(\sigma,\rho)}(\delta t) = \sigma + \rho \times \delta t. \tag{2.6}$$

A figura 2.3 abaixo ilustra um exemplo de função de restrição de tráfego para o modelo (σ, ρ). No mesmo gráfico é apresentada uma possível curva exata ζ.

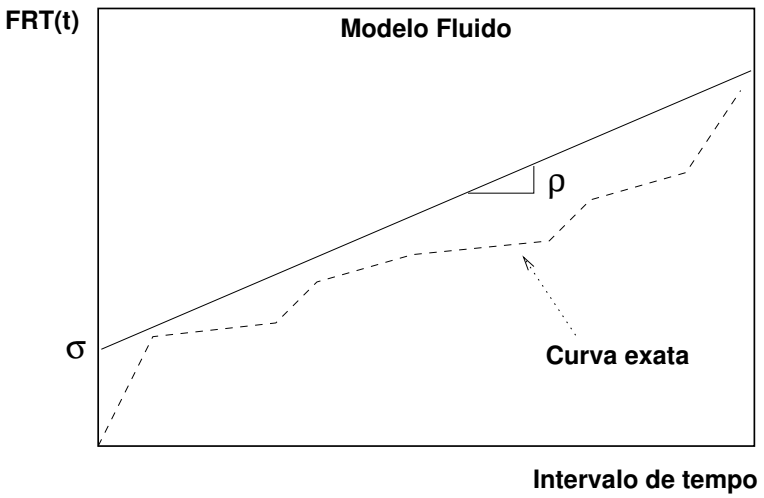


Figura 2.3: Função de restrição de tráfego para o modelo (σ, ρ).

Uma das maiores vantagens do modelo fluido é a utilização de um número reduzido de parâmetros de tráfego. Este fato é extremamente importante para a realização de testes de controle de admissão de conexão com rapidez, visto que o algoritmo utilizado deve manipular apenas dois parâmetros, o que certamente reduz sensivelmente o tempo de computação. Este ganho de desempenho nos cálculos reflete também na redução da complexidade do mecanismo de policiamento de tráfego, o que é mandatório para as redes que operam em altas taxas de transmissão.

Embora este mecanismo seja simples, ele tem dois importantes inconvenientes. Primeiro, as curvas de restrição de tráfego exatas não possuem, na prática, uma taxa de crescimento médio constante. Se a curva relativa ao modelo (σ, ρ) procurar se aproximar da curva exata para pequenos (grandes) valores de intervalo de tempo, se afastará para valores maiores (menores).

Como mencionado anteriormente, existe uma infinidade de pares (σ, ρ) que podem modelar um certo tráfego. Unido ao problema descrito anteriormente, há ainda este compromisso na determinação dos parâmetros. Uma escolha mal dimensionada pode refletir negativamente no nível de acurácia de parametrização do modelo.

2.4 Modelo Fluido em Duas Dimensões $(\vec{\sigma}, \vec{\rho})$

O modelo Fluido em Duas Dimensões [4, 5] é uma extensão ao modelo (σ, ρ) apresentado na seção anterior, onde, no lugar de se definir uma dupla, são definidos dois vetores $(\vec{\sigma}, \vec{\rho})$. Para um intervalo de tempo de duração δt não podem ser transmitidos pela fonte mais que $min_i\{\sigma_i + \rho_i \times \delta t\}$ bits. Desta forma, o modelo é capaz de distinguir mais acuradamente "níveis de rajada" inerentes ao tráfego, se comparado ao modelo (σ, ρ).

A função de restrição de tráfego correspondente ao modelo fluido em duas dimensões está plotada na figura 2.4 e é dada pela equação

$$\mathcal{F}_{(\vec{\sigma},\vec{\rho})}(\delta t) = min_i\{\sigma_i + \rho_i \times \delta t\}. \tag{2.7}$$

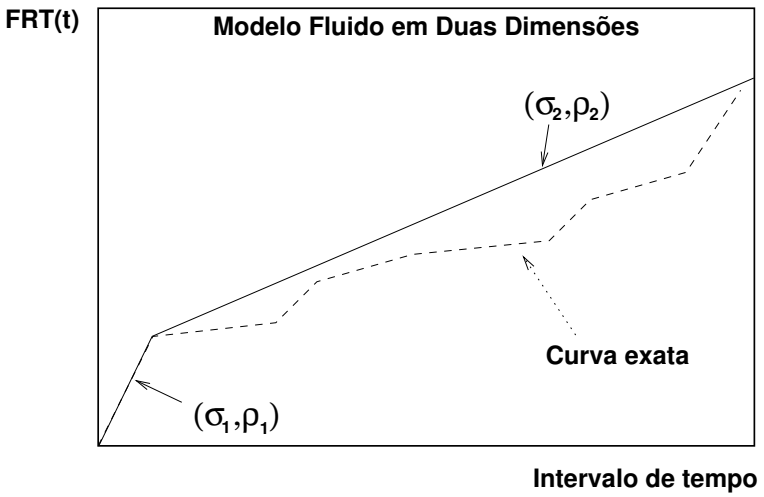


Figura 2.4: Função de restrição de tráfego para o modelo $(\vec{\sigma}, \vec{\rho})$.

A função de restrição de tráfego dada pela equação 2.7 tem a característica principal de ser linear por partes e côncava. Como, na prática, a curva de restrição de tráfego correspondente à função exata ζ não mantém sua taxa de crescimento médio constante, o modelo $(\vec{\sigma}, \vec{\rho})$ é capaz de se aproximar da função exata com maior flexibilidade, em comparação ao modelo (σ, ρ). Por outro lado, este modelo é mais complexo, uma vez que requer um maior número de parâmetros.

A determinação dos parâmetros para o modelo $(\vec{\sigma}, \vec{\rho})$ é feita da seguinte forma. Para um intervalo de tempo δt dado, um par (σ_k, ρ_k) e um novo intervalo $\hat{\delta t}$ são escolhidos tal que

$$\zeta(\delta t) = \sigma_k + \rho_k \times \delta t, \tag{2.8}$$

$$\zeta(t) \leq \sigma_k + \rho_k \times \delta t|_{\forall 0 \leq t \leq \delta t}, \tag{2.9}$$

$$\zeta(\hat{\delta t}) = \sigma_k + \rho_k \times \hat{\delta t}. \tag{2.10}$$

O procedimento acima é repetido até que $\hat{\delta t} = 0$.

2.5 Modelo Discreto $(x_{min}, x_{ave}, I, s_{max})$

O modelo Discreto se encaixa no contexto dos algoritmos de gerenciamento de recursos do projeto Tenet [2] da Universidade da Califórnia, em Berkeley. Estes algoritmos são baseados em uma abstração chamada canal de tempo real (*real-time channel*) [2]. Um canal de tempo real é uma conexão de rede associada a um tráfego e a alguns parâmetros de desempenho. Estes parâmetros são utilizados pelo cliente para especificação das características do tráfego e constituem os parâmetros do modelo *Discreto*.

Este modelo utiliza quatro parâmetros para a caracterização do tráfego $(x_{min}, x_{ave}, I, s_{max})$, onde x_{min} é o intervalo mínimo entre as chegadas de dois pacotes consecutivos, x_{ave} o máximo intervalo médio entre as chegadas de pacotes em um intervalo de tempo qualquer de duração I e s_{max} o tamanho máximo de um pacote. Dispondo-se destes parâmetros e mais a velocidade l do canal, pode-se extrair alguns indicadores importantes, tais como:

- utilização do canal $\rightarrow \frac{s_{max}}{x_{min} \times l}$;

- a razão taxa_de_pico / taxa_média $\rightarrow \frac{x_{ave}}{x_{min}}$;
- o máximo comprimento da rajada $\rightarrow \frac{x_{min} \times I}{x_{ave}}$.

A função de restrição de tráfego para o modelo *Discreto* é expressa por

$$\mathcal{F}_{Discreto}(\delta t) = \lfloor \frac{\delta t}{I} \rfloor \frac{I \times s_{max}}{x_{ave}} + min\{\lceil (\frac{\delta t}{I} - \lfloor \frac{\delta t}{I} \rfloor) \frac{I}{x_{min}} \rceil, \frac{I}{x_{ave}}\} \times s_{max}. \qquad (2.11)$$

A curva de restrição de tráfego correspondente à equação 2.11 está plotada na figura 2.5. Analisando-se tanto o gráfico quanto a equação, pode-se observar que a curva do modelo *Discreto*, assim como do modelo (σ, ρ), tem sua taxa média de crescimento constante, o que pode ser indesejável em alguns casos.

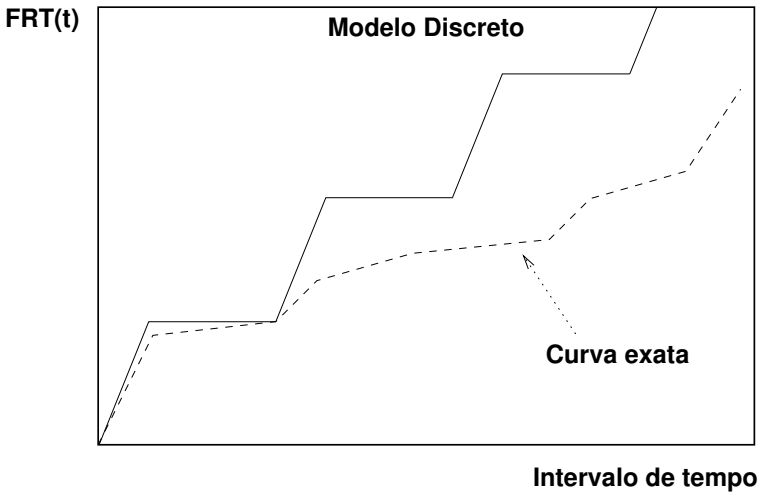


Figura 2.5: Função de restrição de tráfego para o modelo *Discreto*.

Na determinação dos parâmetros para o modelo *Discreto*, o algoritmo apresenta um grau de liberdade na escolha do valor do intervalo de tempo I. Assim, os resultados obtidos (ou seja, as acurácias de parametrização) serão melhores ou piores de acordo com a adequada escolha do parâmetro I.

2.6 Modelo D-BIND

O modelo **D-BIND** (*Deterministic Bounding INterval Dependent*) [8, 15, 31] é definido da seguinte forma. Uma fonte j pode ser descrita pela curva $R_j(I)$ onde $R_j(I)$ é a taxa máxima de geração de bits pela fonte em um intervalo de tempo de duração I de forma que

$$\frac{\mathcal{G}[t, t+I]}{I} \leq R_j(I), \qquad \forall t, I > 0. \tag{2.12}$$

Assim, a fonte é deterministicamente restrita a transmitir não mais que $b_j(\delta t) = \delta t \times R_j(\delta t)$ bits durante qualquer intervalo de duração δt.

Várias observações podem ser feitas com relação a esta caracterização. Dentre elas, para uma dada fonte, a tendência geral da curva $R(I)$ é se aproximar da taxa de pico da fonte para valores de I pequenos, enquanto que para intervalos maiores $R(I)$ se aproxima da taxa média de longo termo, definida empiricamente por $\lim_{t\to\infty} \mathcal{G}[0,t]/t$.

A representação da função $R(I)$ é feita de acordo com o seguinte procedimento. O modelo *D-BIND* consiste em N pares taxa-intervalo da forma $\{(R_n, I_n)|_{n=1,2,\ldots,N}\}$. Pela interpolação entre os pares (R_n, I_n), o modelo *D-BIND* resulta em uma curva de restrição de tráfego linear por partes, cuja função é dada pela equação

$$\mathcal{F}_{D-BIND}(\delta t) = \frac{R_k I_k - R_{k-1} I_{k-1}}{I_k - I_{k-1}}(\delta t - I_k) + R_k I_k, \quad I_{k-1} \leq \delta t \leq I_k. \tag{2.13}$$

A curva de restrição de tráfego é apresentada na figura 2.6. O modelo fluido em duas dimensões difere do modelo *D-BIND* basicamente pelo fato deste último resultar em uma curva de restrição não necessariamente côncava, o que possibilita uma modelagem mais acurada do tráfego alvo.

Como será visto no capítulo 4, dentre os modelos neste capítulo apresentados, o modelo *D-BIND* resulta em maior acurácia de parametrização, devido justamente à característica linear por partes não necessariamente côncava.

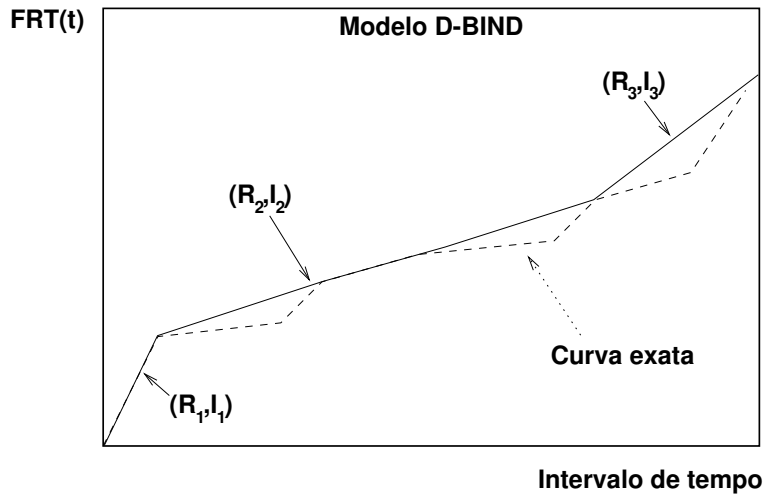


Figura 2.6: Função de restrição de tráfego para o modelo *D-BIND*.

2.7 Discussão

Foram apresentados neste capítulo quatro modelos determinísticos para tráfegos VBR propostos na literatura, além do modelo de alocação pelo pico. Estes modelos, para determinação dos parâmetros de tráfego, se fundamentam na definição da função de restrição de tráfego. Esta função representa um determinado tráfego considerando, no lugar do eixo do tempo, seu comportamento de pior caso para valores de **intervalo de tempo**.

Foi visto também que a função de restrição de tráfego mais precisa para caracterização de uma fonte de dados é a função exata ζ. O problema desta representação é com relação a sua parametrização, uma vez que seu comportamento, na prática, demanda um número muito grande de parâmetros, inviabilizando sua utilização.

A grande diversidade de aplicações emergentes implica em uma de duas condições concernentes à modelagem de tráfegos. Ou o modelo proposto assume características de generalidade, sendo aplicável a todos os tipos de tráfego, ou o modelo se restringe a algum tipo específico de aplicação. Se ele for geral, a precisão será comprometida em nome da versatilidade. Se for específico, poderá não ser largamente utilizado, mas resultará em ganho de acurácia de parametrização. Dependendo da importância da aplicação, pode ser justificada a escolha de um modelo específico.

No atual contexto da multimídia, o vídeo comprimido tem importância ímpar. Dentre os padrões de compressão, o MPEG tem mostrado maior grau de aceitação por parte dos desenvolvedores de sistemas, resultando em uma área de pesquisa onde são aplicados grandes esforços. Por estes motivos, a concepção de um modelo de tráfego para caracterização de tráfegos de vídeo MPEG se mostra extremamente interessante.

Os modelos determinísticos apresentados neste capítulo foram propostos para caracterização de tráfego VBR em geral. No capítulo seguinte, serão introduzidos dois novos modelos, específicos para tráfegos de vídeo comprimido no padrão MPEG, chamados GOP-B (*GOP-Based model*) e XGOP-B (*eXtended GOP-Based model*).

Capítulo 3

Modelagem Determinística de Tráfegos de Vídeo MPEG Baseada em Grupo de Quadros

A GENERALIDADE de um modelo de tráfego pode trazer a vantagem de torná-lo largamente utilizado pelas aplicações, mas pode resultar por outro lado em baixos níveis de acurácia de parametrização. Os modelos determinísticos apresentados no capítulo anterior foram concebidos para modelagem de tráfegos em taxa variável de bits de uma forma geral.

Serão propostos neste capítulo dois novos modelos específicos para caracterização de fontes de vídeo MPEG: o modelos *GOP-B* e sua extensão, o modelo *XGOP-B*. Estes modelos têm as vantagens de serem simples na extração dos parâmetros de tráfego e de obterem os resultados desejados com um número reduzido de parâmetros, essencial para que os mecanismos de policiamento de tráfego e de controle de admissão de conexões demandem pouco processamento.

Inicialmente, será introduzido um estudo a respeito do padrão MPEG, direcionado para os objetivos deste trabalho. De acordo com as propriedades observadas, algumas premissas serão ressaltadas para servirem de fundamento para a proposição dos modelos.

3.1 O Padrão MPEG

O algoritmo de codificação MPEG (*Moving Picture Expert Group*) [32] foi desenvolvido primariamente para armazenamento de vídeo comprimido em mídias digitais. Extensões foram introduzidas no algoritmo de forma a habilitar, além de outras funções, acesso randômico e rápidas procuras em avanço e reverso (*forward/reverse*). Contudo, o padrão é suficientemente flexível para ser utilizado em um grande número de outras aplicações que envolvam vídeo. Aplicações recentes envolvendo o padrão MPEG têm surgido para uma grande variedade de serviços, desde aqueles voltados para estações multimídias até outros como TV de alta definição.

O algoritmo MPEG alcança compressão de dados em três estágios: redução da banda-passante (*bandwidth reduction*), compressão com perda subjetivamente adaptada (*subjectively adapted lossy compression*) e um estágio final de compressão sem perdas (*final stage of lossless compression*). O primeiro estágio consiste basicamente em levar a resolução da fonte para a taxa em bits desejada e reduzir a resolução da crominância para um valor satisfatório. O segundo estágio, o algoritmo de compressão propriamente dito, elimina as redundâncias temporal e espacial pela observação de elementos repetitivos dentro de um quadro ou entre quadros consecutivos. O terceiro e último estágio mapeiam a informação em uma sintaxe combinando códigos de tamanho fixo e variável.

A seqüência de entrada de um codificador MPEG consiste em uma série de quadros, cada um contendo um vetor bidimensional de elementos de figura, chamados *pixels* ou ainda *pels* (*picture elements*). O número de quadros por segundo, assim como o número de linhas por quadro e *pels* por linha, depende de padrões nacionais. Para cada *pel*, tanto as informações de luminância quanto as de crominância são armazenadas. O algoritmo de compressão é utilizado para reduzir a taxa em bits por segundo do vídeo antes de sua transmissão pelo sistema de comunicação.

Existem várias versões para o padrão MPEG, dentre elas o MPEG-1, o MPEG-2 e o MPEG-4, além do MPEG-7, que deve estar totalmente especificado por volta do ano 2002 [33]. As técnicas apresentadas neste trabalho podem ser aplicadas a

qualquer versão do padrão MPEG que mantiver as propriedades analisadas.

Um codificador MPEG gera três tipos de quadros:

- quadros I (*Intra-pictures*), que utilizam compressão intra-quadros, baseada na transformada discreta do cosseno;
- quadros P (*Predicted-pictures*), que utilizam um algoritmo de codificação similar ao utilizado para os quadros I, além de compensação de movimento em relação a quadros I e P anteriores;
- quadros B (*Interpolated-pictures*), para predição bidirecional, similares aos quadros P, mas que utilizam compensação de movimento com base tanto em um quadro passado quanto em um quadro de referência futuro.

A figura 3.1 ilustra graficamente o relacionamento de causalidade entre os quadros.

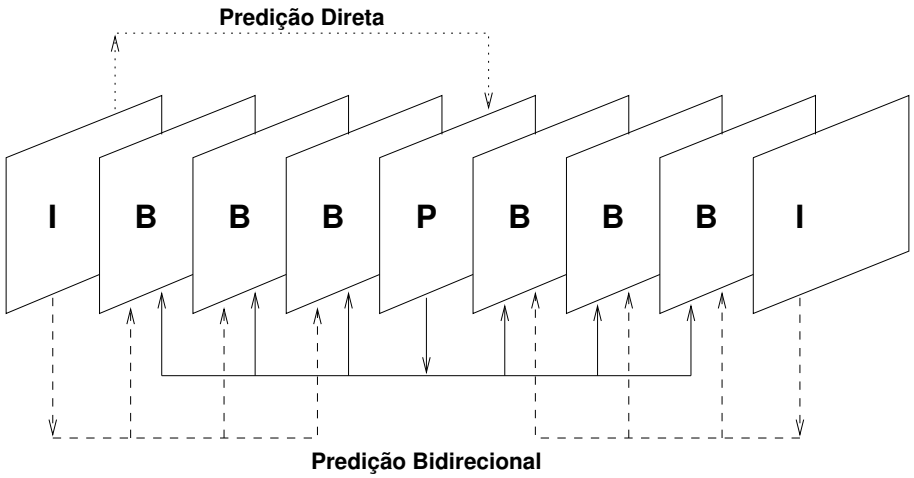


Figura 3.1: Codificação interquadros.

A característica da taxa variável de bits de um vídeo MPEG vem da diferença entre os tamanhos dos sucessivos quadros da seqüência, causada pela alternância entre intra-quadros (maiores) e inter-quadros (menores), como ilustra a figura 3.2.

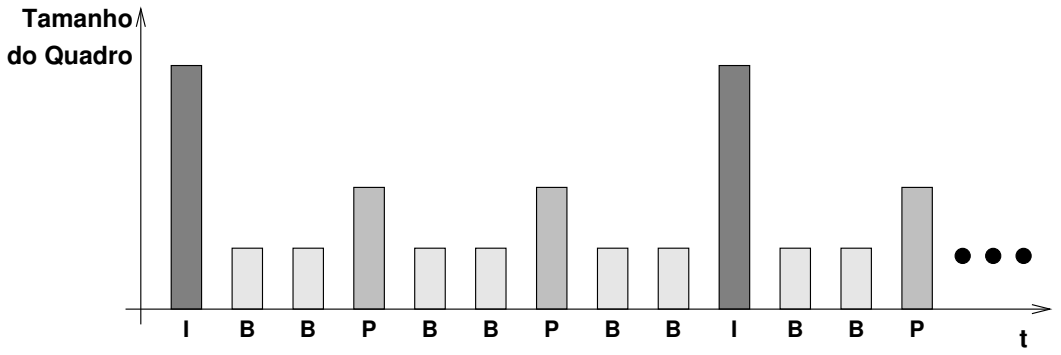


Figura 3.2: Amostra de um vídeo no padrão MPEG.

3.2 Análise Microscópica

Um vídeo MPEG é cíclico no que diz respeito à seqüência de formatos de quadros usada. Cada ciclo é caracterizado por um GOP (*Group of Pictures*) contendo um quadro I (inicial) e alguns quadros P e B de acordo com uma seqüência pré-determinada, como por exemplo IBBPBBPBB.

A grande maioria dos vídeos MPEG tem os quadros do tipo I bem maiores que os quadros P e B. Em termos de geração de bits, isto corresponde a surtos de informação no início de cada GOP. A função exata ζ básica de um vídeo MPEG está ilustrada na figura 3.3, onde se pode notar um "salto" na função a cada geração de um quadro I. Em um nível de abstração mais elevado, um vídeo MPEG é, então, gerado basicamente em uma taxa de bits baixa (com pequenas variações) mais os surtos devidos aos quadros I em instantes **definidos** de tempo.

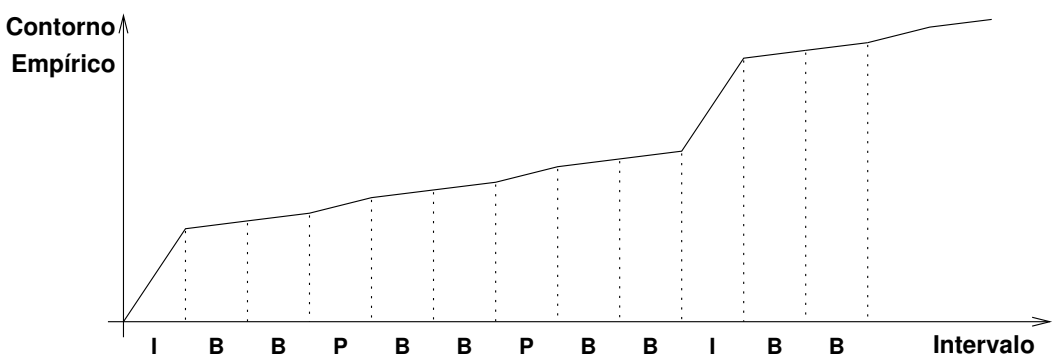


Figura 3.3: Função exata para vídeo MPEG.

Estas características são observadas durante todo o tempo de geração de um vídeo MPEG. Porém, na grande maioria das vezes, as taxas de compressão alcançadas pelo codificador variam muito ao longo do tempo. O resultado pode ser uma seqüência de quadros extremamente irregular no que se refere à métrica número de bits por quadro. A figura 3.4 ilustra quatro traços[1] de vídeo com características bem diversas, apesar de terem sido capturadas sob as mesmas condições (padrão do GOP, número de elementos por quadro, taxa de amostragem etc).

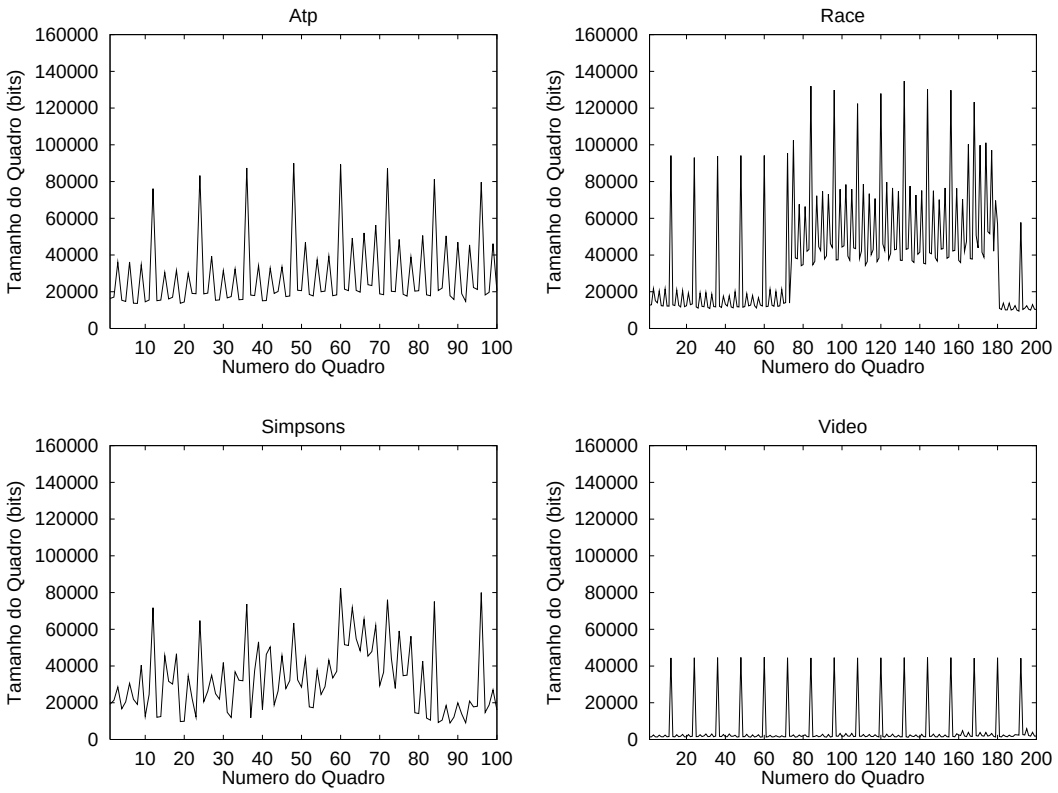


Figura 3.4: Exemplos de traços de vídeos.

3.3 Modelo GOP-B

Foi visto na seção anterior que a diferença de tamanho entre os quadros I, P e B refletem em uma curva de restrição de tráfego com características bem particulares. Na quase totalidade das amostras de vídeo, os intra-quadros possuem dimensões

[1] Um traço de vídeo é a seqüência ordenada dos tamanhos dos quadros da amostra de vídeo.

bem superiores àquelas dos inter-quadros. Unido a este fato, uma seqüência de vídeo MPEG pode ser subdividida em unidades chamadas grupo de quadros (GOP). Um GOP é caracterizado por um número pré-definido de quadros, sendo que o primeiro deles é obrigatoriamente do tipo I e os demais são quadros P e B ordenados de forma regular, por exemplo IBBPBBPBB. Este formato é repetido sucessivamente durante todo o tempo de geração de quadros pela fonte.

Conhecendo-se o padrão do GOP, a taxa de amostragem e o instante inicial de geração do vídeo, pode-se determinar exatamente qual tipo de quadro está sendo produzido em determinado instante de tempo. Com as propriedades naturais de um vídeo MPEG mantidas, o reflexo observado na função de restrição de tráfego consiste em uma curva monotonicamente crescente contendo variações bruscas (saltos) pela presença de intra-quadros.

O modelo GOP-B [13, 14] se baseia justamente neste princípio, o dos vídeos MPEG apresentarem, na taxa de geração de bits, surtos de informação devidos às presenças dos quadros do tipo I. Este modelo utiliza um par (η, ν) como parâmetros de tráfego. O fator η, chamado *fator de impulsos*, modela os surtos (impulsos) causados pela geração de intra-quadros. Este fator deve ser adicionado à função de restrição de tráfego em intervalos múltiplos do tamanho do GOP. O fator ν, denominado *fator médio*, parametriza o crescimento médio da curva, dado pela presença dos quadros P e B.

A função de restrição de tráfego relativa ao modelo *GOP-B* é composta, então, por duas parcelas. Uma diretamente proporcional ao intervalo de tempo e outra discreta só computada para valores de intervalo de tempo proporcionais ao período de geração de um GOP completo. A função de restrição é representada pela equação

$$\mathcal{F}_{GOP-B}(\delta t) = \lceil \frac{\delta t + T}{T \times D} \rceil \eta + \nu \times \delta t, \tag{3.1}$$

onde T é o inverso da taxa de geração de quadros por segundo e D é a distância[2]

[2]Por exemplo, se um determinado padrão de GOP for IBBPBB, a distância entre quadros I será 6.

entre quadros I.

A curva de restrição de tráfego correspondente à função apresentada na equação 3.1 é mostrada na figura 3.5.

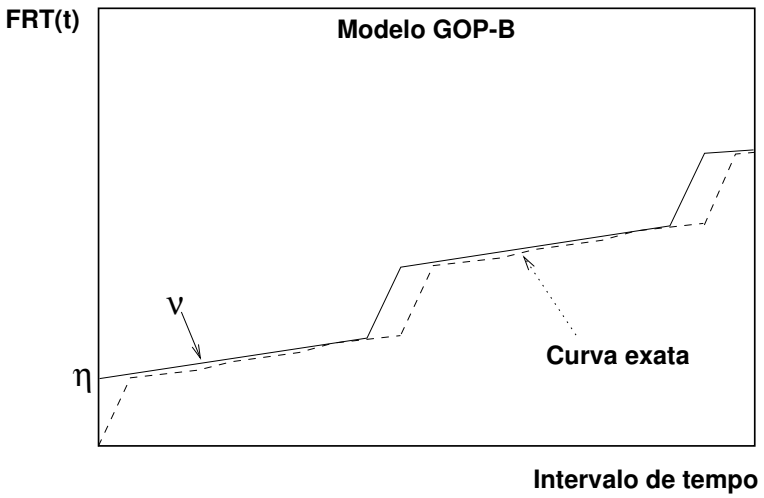


Figura 3.5: Função de restrição de tráfego para o modelo *GOP-B*.

Uma observação deve ser ressaltada quando, na primeira parcela da função de restrição de tráfego, é adicionado ao numerador da fração uma unidade de T. Esta decisão pode ser justificada pelo ganho de acurácia de parametrização obtido ao se realizar esta soma. Adicionando-se uma unidade de T, a curva de restrição é deslocada T segundos para a esquerda. Ao se fazer este deslocamento, permite-se nos cálculos que a inclinação média da curva, dada por ν, seja reduzida. Alguns testes foram realizados e constatou-se que perdas de acurácia de parametrização são obtidas muito mais pelo aumento de ν que por outro motivo.

O modelo *GOP-B* é dito baseado no grupo de quadros porque a unidade básica utilizada no algoritmo de extração de parâmetros de tráfego é o GOP. Enquanto os demais modelos (com excessão de *Discreto*) precisam conhecer *a priori* o valor da função ζ para todos os intervalos de tempo, o modelo *GOP-B* exige o valor do contorno empírico somente para o intervalo de um GOP. Como será visto no capítulo 4, o algoritmo para extração de parâmetros do modelo *GOP-B* é bem mais simples que os demais. O fato de não exigir o cálculo da função exata ζ torna a utilização do modelo *GOP-B* bastante atrativa.

3.4 Análise Macroscópica

As primeiras observações feitas a respeito das propriedades específicas do padrão MPEG mostram que uma seqüência de quadros de uma amostra de vídeo segue um padrão de codificação das imagens que se repete periodicamente ao longo do tempo, chamado GOP. O ponto chave desta análise se concentra no fato do primeiro quadro de cada GOP ser, geralmente, pelo tipo de codificação realizada, bem maior que os demais quadros da seqüência. O reflexo observado por este comportamento sobre a curva de restrição de tráfego foi ilustrado na figura 3.3. De acordo com a figura, a taxa de crescimento da curva se mantém com pequenas variações, pela geração de quadros P e B, mas apresenta surtos de informação para intervalos de tempo proporcionais ao período de um GOP, quando são produzidos intra-quadros.

Apesar desta propriedade ser fundamental, ela não é a única. Seria, sim, se os tráfegos de vídeo MPEG mantivessem uma certa uniformidade no seqüenciamento dos quadros. Mas, na prática, as cenas variam constantemente e, por conseqüência, as taxas de compressão alcançadas pelo codificador apresentam grandes variações ao longo do tempo. A figura 3.6 ilustra duas amostras de vídeo, uma bastante regular e outra apresentando variações nítidas na taxa de compressão dos dados para diferentes quadros.

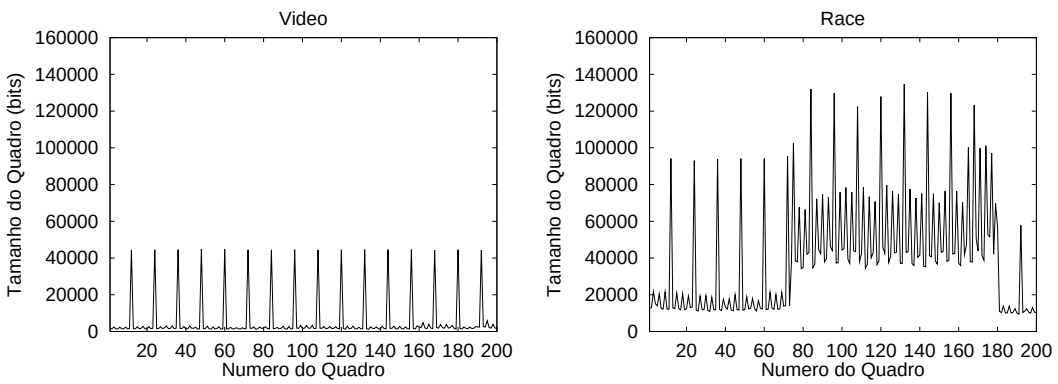


Figura 3.6: Amostras de vídeo MPEG.

Em termos da função de restrição de tráfego, o reflexo observado pelo mau comportamento do vídeo a longo prazo é a ocorrência de variações "bruscas" na taxa média de crescimento da curva. Esse efeito pode ser melhor notado com auxílio

da figura 3.7. Esses pontos onde a curva sofre deflexões acentuadas são chamados **pontos de quebra** (ξ).

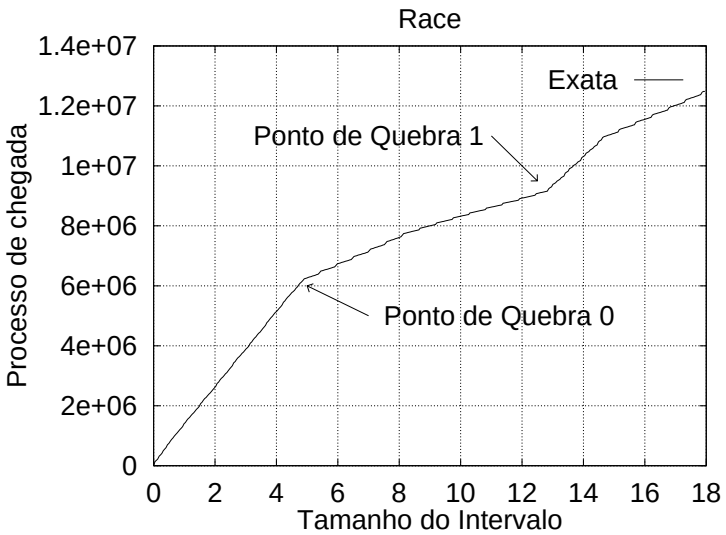


Figura 3.7: Efeito dos pontos de quebra sobre a função exata.

Para os modelos (σ, ρ), *Discreto* e *GOP-B*, a aparição de pontos de quebra é o principal fator responsável pela perda de acurácia. Uma das propriedades comuns a estes modelos é a constância na taxa média de crescimento da curva de restrição de tráfego. Com a ocorrência de um ponto de quebra negativo (aquele que causa redução na taxa média de crescimento da curva), a função de restrição de tráfego diverge da função exata gerando imprecisões.

A existência de pontos de quebra é onerosa para o modelo $(\vec{\sigma}, \vec{\rho})$ quando estes forem do tipo positivo (que causa aumento na taxa média de crescimento da curva). A característica de concavidade intrínseca ao modelo $(\vec{\sigma}, \vec{\rho})$ leva ao distanciamento a função de restrição de tráfego da função exata em regiões próximas a estes pontos. Já para o modelo *D-BIND* o mesmo não ocorre. No entanto, os pontos de quebra requerem mais pares (R_k, I_k), o que pode aumentar significativamente o número de variáveis a serem manipuladas em etapas cruciais, como nos testes de admissão de conexão e policiamento de tráfego.

Foi visto que, dependendo do modelo escolhido, a existência de pontos de quebra pode levar a imprecisões na modelagem do tráfego para diferentes comprimentos de

intervalo de cálculo. Contudo, a identificação dos pontos assinalados no figura 3.7 foi feita de forma intuitiva. É preciso que se formule matematicamente um método de identificação.

O primeiro passo consiste na determinação de uma função correspondente às sucessivas inclinações da função exata. Esta função armazena a taxa de geração de bits pela fonte por unidade de tempo. As inclinações são particularmente calculadas com intervalos proporcionais ao tamanho do GOP. Esta função é chamada **Primeira Derivada** e está ilustrada na figura 3.8.

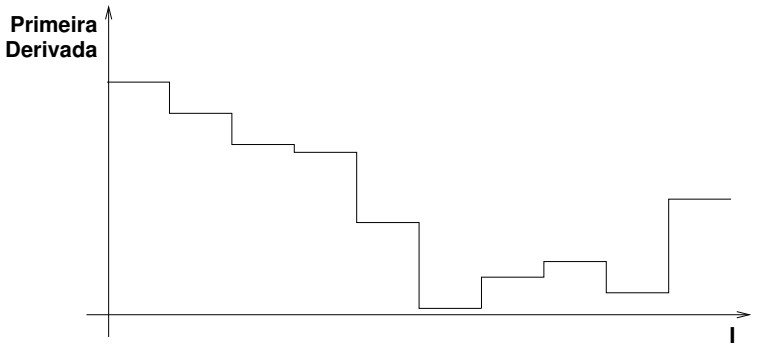


Figura 3.8: Exemplo de Primeira Derivada.

A próxima etapa consiste em se obter, a partir da curva de Primeira Derivada, uma função discreta correspondente somente às variações da primeira derivada. Esta nova função, denominada **Segunda Derivada**, é apresentada no gráfico da figura 3.9 ilustram exemplos de primeira e segunda derivadas.

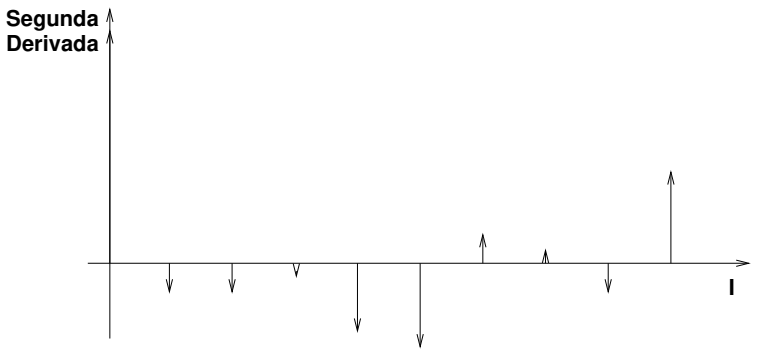


Figura 3.9: Exemplo de Segunda Derivada.

Os impulsos plotados na figura 3.9 assinalam os instantes onde as variações de

inclinação ocorrem. Contudo, resta ainda determinar, dentre estes pontos, quais devem ser assumidos como pontos de quebra. A escolha será determinada pelo nível de qualidade do modelo desejado pelo usuário. Se mais pontos forem escolhidos, o modelo será mais acurado pois serão determinados mais parâmetros de tráfego. Por outro lado, se forem selecionados menos pontos, o modelo será mais simples em termos do número de parâmetros, mas em contrapartida apresentará menor acurácia de parametrização. A técnica adotada para solucionar este problema consiste em se oferecer para o módulo extrator de parâmetros de tráfego uma variável β denominada **fator de qualidade do modelo**. Os pontos de quebra são selecionados de acordo com as proposições a seguir:

Proposição 1 *Iniciando-se com uma inclinação, será identificado um ponto de quebra se esta inclinação variar negativamente ou positivamente de um fator β, com $\beta > 0$.*

Proposição 2 *É identificado um ponto de quebra se o somatório de uma seqüência ininterrupta de segundas derivadas variar negativamente ou positivamente, em relação ao valor da primeira inclinação desta seqüência, de um fator β, com $\beta > 0$.*

3.5 O Modelo XGOP-B

O modelo *GOP-B* assume em sua estrutura as principais características intrínsecas ao padrão MPEG. Contudo, a propriedade de que as funções de restrição de tráfego exatas apresentam pontos de quebra não é considerada, apesar de sua notória importância.

A alteração introduzida pelo modelo XGOP-B (*eXtended GOP-Based model*) consiste na adequação da inclinação média da função de restrição de tráfego a um novo valor induzido pela existência de um ponto de quebra. Assim, a função de restrição de tráfego é subdividida em "faixas" de valores de intervalo de tempo limitadas pelos pontos de quebra obtidos a partir do procedimento descrito na seção 3.4. A figura 3.10 ilustra a classificação de uma função.

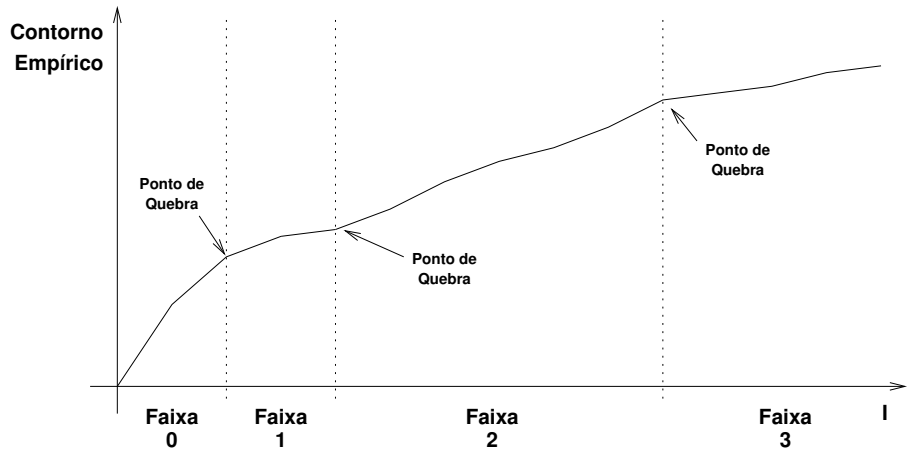


Figura 3.10: Classificação em faixas de uma função exata.

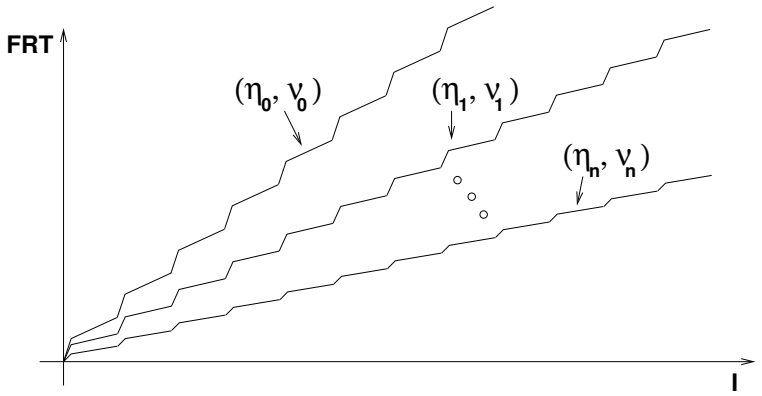
Agora, o problema introduzido pelas duas propriedades volta a se resumir em apenas uma, aplicada separadamente em cada faixa. Em outras palavras, o objetivo principal continua sendo obter uma dupla (η, ν) que parametrize o tráfego. Entretanto, agora serão determinadas diversas duplas (η_0, ν_0), (η_1, ν_1), ..., (η_n, ν_n) para cada uma das faixas delimitadas pelos pontos de quebra.

Se forem fornecidos como parâmetros de tráfego somente os dois vetores $(\vec{\eta}, \vec{\nu})$, o resultado será uma gama de curvas de restrição de tráfego que não terão nenhum significado especial, como pode ser visualizado na figura 3.11(a). Assim, além dos pares (η_k, ν_k), são incluídos como parâmetros de tráfego um outro vetor $\vec{\theta}$ que armazena os valores do *offsets* das curvas dadas por cada par (η_k, ν_k), como ilustra o gráfico da figura 3.11(b), e o vetor $\vec{\xi}$ de pontos de quebra.

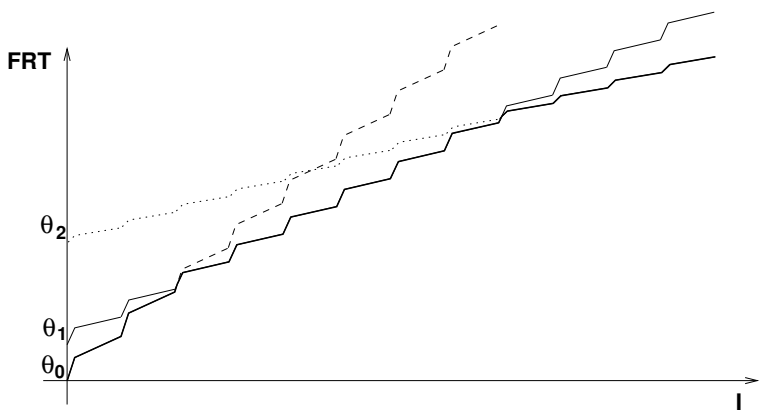
O modelo *XGOP-B* tem, então, como parâmetros de tráfego os vetores $(\vec{\theta}, \vec{\eta}, \vec{\nu}, \vec{\xi})$. A função de restrição de tráfego do modelo *XGOP-B* é uma simples extensão à função relativa ao modelo *GOP-B*, sendo dada pela equação

$$\mathcal{F}_{XGOP-B}(\delta t) = \theta_k + \lceil \frac{\delta t + T}{T \times D} \rceil \times \eta_k + \nu_k \times \delta t, \quad \xi_{k-1} \leq \delta t < \xi_k. \tag{3.2}$$

A figura 3.12 ilustra o comportamento gráfico da função de restrição de tráfego para o modelo *XGOP-B*.



(a)



(b)

Figura 3.11: Inclusão de θ como parâmetro de tráfego.

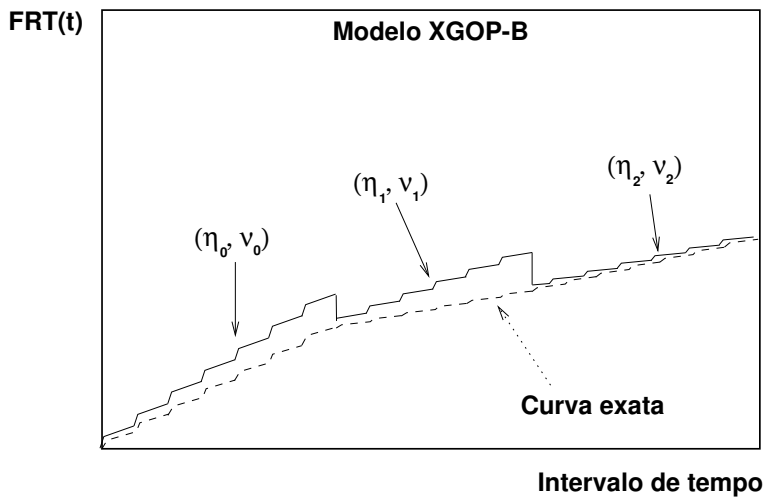


Figura 3.12: Função de restrição de tráfego para o modelo *XGOP-B*.

Capítulo 4

Análise dos Resultados

NESTE capítulo serão apresentados os resultados obtidos da comparação dos modelos estudados e propostos em termos de acurácia de parametrização.

Para realização das medidas são utilizadas 21 amostras de vídeo codificadas no padrão MPEG-1 [34]. As seqüências foram obtidas de filmes, eventos esportivos e shows apresentados na TV. Todas as amostras foram codificadas pelo UC Berkeley MPEG-1 Software Encoder [35]. A plataforma utilizada para o processamento e codificação do vídeo foi uma Sparc 20 da Sun e para a captura das seqüências utilizou-se a placa de captura SunVideo SBus. A tabela 4.1 apresenta as 21 amostras com as respectivas taxas média e de pico.

As amostras possuem as seguintes características em comum:

- padrão do GOP: IBBPBBPBBPBB (distância $D = 12$);
- definição: 384 x 288 pels;
- 12 bits de informação de cor;
- número de quadros: 40.000 (aproximadamente 30 minutos);
- taxa de amostragem: 22 quadros/s.

No contexto deste trabalho, as amostras de vídeo serão particularmente chama-

Tabela 4.1: Características dos vídeos utilizados.

Vídeo	*Taxa* Média (Mbps)	*Taxa* de Pico (Mbps)	*Vídeo*	*Taxa* Média (Mbps)	*Taxa* de Pico (Mbps)
1. asterix	0.49	3.24	12. race	0.68	4.45
2. atp	0.48	4.20	13. sbowl	0.52	3.10
3. bond	0.53	5.38	14. simpsons	0.41	5.29
4. dino	0.29	2.63	15. soccer1	0.60	4.12
5. lambs	0.16	2.95	16. soccer2	0.55	4.19
6. mbean	0.39	5.04	17. star	0.20	2.75
7. movie2	0.31	3.80	18. talk1	0.32	2.35
8. mtv1	0.54	5.04	19. talk2	0.39	2.92
9. mtv2	0.44	5.53	20. terminator	0.24	1.75
10. news1	0.36	4.23	21. video	0.02	1.02
11. news2	0.34	4.18			

das traços de vídeo. Um traço de vídeo é uma lista que contém a seqüência dos tamanhos dos quadros. A figura 4.1 ilustra graficamente traços de alguns dos vídeos selecionados. É interessante observar que mesmo sendo capturados sob as mesmas condições o resultado final se mostra bastante diversificado.

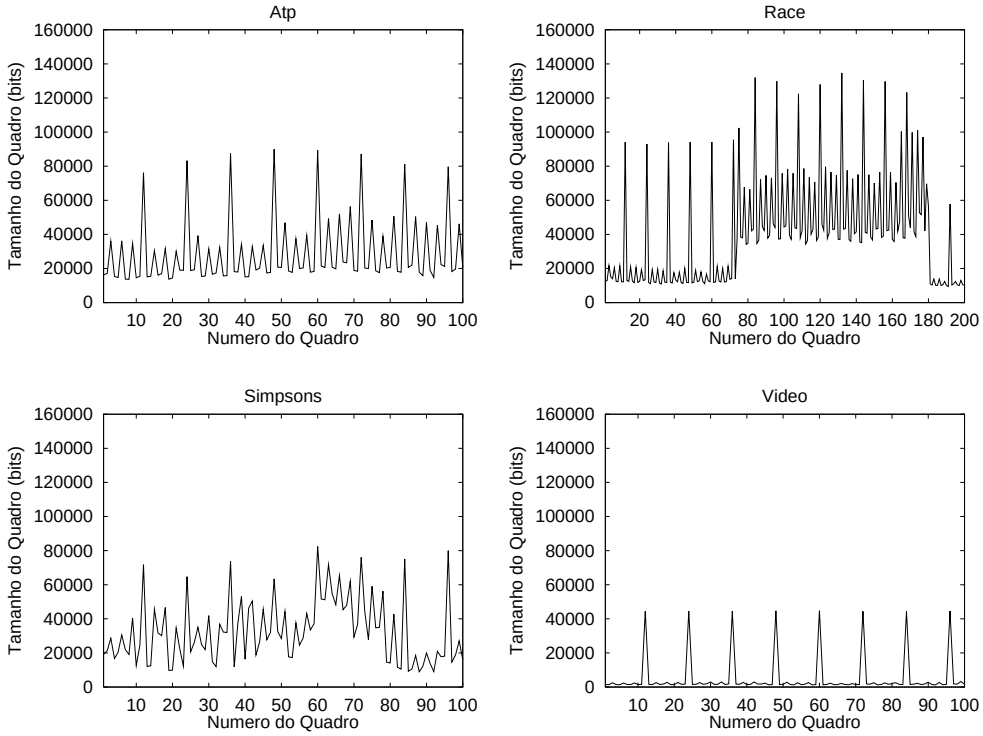


Figura 4.1: Exemplos de traços de vídeos MPEG.

A medida de comparação dos modelos utiliza o conceito de **Perda por Aproximação do Modelo** (PAM) [12], que indica a porcentagem de perda de precisão do modelo em comparação ao resultado mais preciso possível, a função exata ζ. Para cada curva de restrição de tráfego, a PAM correspondente é obtida pela diferença das áreas sob esta curva e a curva exata, como formulado na equação

$$PAM = \left(\frac{S_{\mathcal{F}}}{S_{\zeta}} - 1 \right) \times 100\%, \tag{4.1}$$

onde $S_{\mathcal{F}}$ é a área sob a curva de restrição de tráfego e S_{ζ} é a área sob a curva exata. Quanto menor for o valor de PAM, mais acurada será uma determinada caracterização. Se $PAM = 0\%$, as duas curvas se encontram e, conseqüentemente,

obtem-se o máximo em acurácia, visto que o contorno empírico é a curva mais precisa possível.

4.1 Extrator de Parâmetros de Tráfego

Com excessão do modelo $(\vec{\sigma}, \vec{\rho})$, os demais modelos propostos na literatura não apresentam uma solução ou metodologia para obtenção, a partir dos traços de um vídeo, dos parâmetros de tráfego. Para o modelo (σ, ρ), por exemplo, existe uma infinidade de possíveis soluções para σ e ρ, e o nível de acurácia de parametrização é diretamente relacionado à adequada escolha destes parâmetros. Para o modelo $(x_{min}, x_{ave}, I, s_{max})$, o problema se concentra na escolha do intervalo de tempo I. Se a opção for mal dimensionada, a imprecisão se reflete também em x_{ave}. Na seção 4.2 serão apresentadas soluções para diferentes valores de I, e será visto que a acurácia de parametrização é realmente afetada por I. Com relação ao modelo D-BIND, o resultado obtido também é mais ou menos satisfatório dependendo da adequada escolha dos pares (R_k, I_k).

Pelos motivos acima descritos, foram propostos e implementados extratores de parâmetros de tráfego para cada um dos modelos estudados, além dos modelos GOP-B e XGOP-B.

4.1.1 Extrator (σ, ρ)

O extrator para o modelo Fluido recebe como entrada o traço do vídeo e retorna os valores dos parâmetros (σ, ρ). Como foi dito anteriormente, existe uma infinidade de soluções para σ e ρ, e o extrator de parâmetros proposto calcula o melhor valor para o par (σ, ρ). Os cálculos são realizados de forma que o valor de PAM seja minimizado, ou seja, levando-se em conta a minimização da expressão $S_{\mathcal{F}} - S_{\zeta}$.

Modelo Fluido:

Procedure (σ, ρ) $(\mathcal{G})$

ζ= Calcula_Exata($\mathcal{G}$)

$\rho = \frac{\zeta(t+\delta t)-\zeta(t)}{\delta t}|_{max[\delta t]\ \ e\ \ \forall u, \sigma+\rho u \geq \zeta(u)}$

$\sigma = (\zeta(t+\delta t) - \rho \times (t+\delta t))|_{max[\delta t]\ \ e\ \ \forall u, \sigma+\rho u \geq \zeta(u)}$

End

4.1.2 Extrator $(\vec{\sigma}, \vec{\rho})$

O modelo Fluido em Duas Dimensões é o único modelo proposto na literatura que apresenta um algoritmo extrator de parâmetros de tráfego, descrito por Knightly et al em [5]. O procedimento foi descrito na seção 2.4. O extrator recebe como entrada, além do traço do vídeo, o valor do intervalo de tempo máximo de validade do modelo (Υ).

Modelo Fluido em Duas Dimensões:

Procedure $(\vec{\sigma}, \vec{\rho})$ $(\mathcal{G}, \Upsilon)$

ζ = Calcula_Exata($\mathcal{G}$)

n=0

$\tau = \Upsilon$

While $\tau > 0$ Do

n=n+1

$\sigma_n = max\left\{\frac{1}{\tau - t}(\tau\zeta(t) - t\zeta(\tau))\right\}$

$\rho_n = \frac{\zeta(\tau)-\sigma_n}{\tau}$

$\tau = min\{t | \sigma_n + \rho_n \times t) = \zeta(t)\}$

End

End

4.1.3 Extrator $(x_{min}, x_{ave}, I, s_{max})$

O procedimento de extração dos parâmetros para o modelo Discreto recebe como entrada o valor do parâmetro I, o traço do vídeo e o inverso da taxa de amostragem do vídeo (T).

Modelo Discreto:

```
Procedure Discreto (G,I,T)
  ζ = Calcula_Exata(G)
  x_min = min{ Ts_max / (ζ(t+T)−ζ(t)) })
  x_ave = min{ Is_max / (ζ(t+I)−ζ(t)) }
End
```

4.1.4 Extrator D-BIND

O modelo D-BIND não introduz qualquer condição para escolha dos parâmetros de tráfego. Knightly [36] propõe a escolha entre 4 e 8 pares (R_k, I_k). Contudo, fica a critério do usuário esta decisão, o que pode introduzir imprecisões desnecessárias. Com o intuito de automatizar ao máximo o processo de determinação dos parâmetros e de se evitar a subjetividade na escolha dos valores, foi introduzido no algoritmo aqui proposto uma pequena restrição. Na realização dos cálculos, o máximo valor permitido para os intervalos I_k é I_{max}. Desta forma, para cada intervalo I_k é aplicado algoritmo similar ao apresentado para o modelo $(\vec{\sigma}, \vec{\rho})$, reduzindo a complexidade e permitindo cálculo automático dos parâmetros.

Modelo D-BIND:

```
Procedure D-BIND (G , I)
  ζ = Calcula_Exata(G)
  n = 0
  τ = 0
  While !Fim Do
```

```
    n=n+1
    t = {τ ≤ t ≤ τ + I_max | max ( (ζ(t)−ζ(τ))/(t−τ) | ) }
    R_n = { ζ(t)/t }
    I_n = t
    τ = t
  End
End
```

4.1.5 Extrator GOP-B

Para o algoritmo de obtenção de parâmetros para o modelo GOP-B são fornecidos, além do traço do vídeo, o tamanho do GOP (distância entre quadros I) e a inversa da taxa de amostragem do vídeo (T). Estes argumentos são importantes para que o extrator tenha a localização exata do tipo de quadro que está sendo processado e em que instante de tempo. É importante observar que o modelo GOP-B só precisa calcular a função exata para intervalos de tempo múltiplos de D (tamanho do GOP).

Modelo GOP-B:

```
Procedure GOP-B (G, D, T)
  η = max_Quadro(G);
  ζ(D) = Calcula_Exata(G, D);
  ν = (ζ(D)−η)/(T×D);
End
```

4.1.6 Extrator XGOP-B

O algoritmo extrator de parâmetros para o modelo *XGOP-B* recebe como argumentos o traço do vídeo, a distância entre quadros I, o inverso da taxa de amostragem e o fator de qualidade do modelo β. Para cada faixa delimitada por dois pontos de quebra, o algoritmo aplicado é semelhante àquele do modelo *GOP-B*.

Modelo GOP-B:

```
Procedure XGOP-B (𝒢, D, T, β)
  η0 = max_Quadro(𝒢);
  ζ(D) = Calcula_Exata(𝒢, D);
  ν0 = (ζ(D)−η0)/(T×D);
  θ0 = 0
  CalculaDerivada1;
  CalculaDerivada2;
  DeterminaPontosQuebra
  for (i=1;i≤#PontosQuebra;i++)
       ηi = max_impulso_intervalo;
       νi = max_inclinação_intervalo;
       θi = Calcula_offsets;
End
```

4.2 Resultados Obtidos

Esta seção apresenta os valores de perda por aproximação de cada um dos modelos para os 21 traços de vídeo. A tabela 4.2 mostra os índices de PAM para as amostras de vídeo utilizando os modelos propostos na literatura. Para o modelo Discreto foram realizados os cálculos com três valores diferentes para o parâmetro I: 540ms, 1080ms e 1640ms, que equivalem, respectivamente, a uma, duas e três vezes o tamanho de um GOP. A tabela 4.3 apresenta os resultados para os modelos GOP-B e XGOP-B. Para este último, realizou-se os cálculos com valores de fator de qualidade do modelo, β, iguais a $0.5, 0.1$ e 0.03. Em 4.3 será feita uma análise do efeito desta variável no nível de acurácia do modelo.

Alguns dos resultados apresentados nas tabelas 4.2 e 4.3 estão plotados na figura 4.2. Para o modelo *XGOP-B*, a curva corresponde ao fator de qualidade do modelo igual a 0.5. É notável a grande diferença entre os níveis de acurácia dos modelos. Entretando, deve ser levado em conta também que existe um compromisso

Tabela 4.2: Perda por Aproximação do Modelo para cada um dos modelos.

	Perda por Aproximação do Modelo (%)					
Vídeo	(σ, ρ)	$(\vec{\sigma}, \vec{\rho})$	Discreto I=540ms	Discreto I=1080ms	Discreto I=1620ms	D-BIND
1. asterix	23.21	3.96	49.21	54.27	54.94	0.32
2. atp	9.31	1.02	17.76	10.07	21.54	1.02
3. bond	44.02	2.16	31.75	26.24	29.88	1.48
4. dino	15.30	2.17	48.78	44.92	31.67	1.71
5. lambs	30.56	3.42	20.31	23.46	31.62	2.23
6. mbean	17.69	1.26	33.40	34.25	37.87	1.00
7. movie2	6.96	1.55	76.96	60.40	60.75	1.49
8. mtv1	13.42	1.43	75.44	67.47	61.43	1.43
9. mtv2	9.20	1.77	30.14	36.27	35.45	1.32
10. news1	23.35	2.43	42.85	45.38	45.19	1.07
11. news2	7.44	2.56	33.67	38.29	40.58	2.47
12. race	48.95	0.64	24.35	22.03	23.48	0.60
13. sbowl	4.51	1.03	19.50	22.54	25.30	0.96
14. simpsons	7.91	1.03	55.76	44.29	36.62	0.95
15. soccer1	4.62	2.90	28.30	28.97	31.88	1.10
16. soccer2	7.21	2.20	39.58	29.19	25.03	1.53
17. star	18.19	1.89	22.59	26.18	30.78	1.88
18. talk1	14.92	2.99	17.52	22.73	27.33	1.15
19. talk2	3.09	2.31	25.28	23.20	24.57	2.05
20. terminator	9.26	2.59	41.86	35.41	38.87	1.17
21. video	6.79	5.42	13.49	16.07	26.02	5.18

Tabela 4.3: Perda por Aproximação do Modelo para cada um dos modelos.

	Perda por Aproximação do Modelo (%)			
Vídeo	GOP-B	XGOP-B $\beta = 0.5$	XGOP-B $\beta = 0.1$	XGOP-B $\beta = 0.03$
1. asterix	44.83	5.41	1.76	1.67
2. atp	15.74	15.74	6.17	1.96
3. bond	32.61	20.37	4.09	0.41
4. dino	59.22	7.41	2.34	1.89
5. lambs	18.36	8.51	4.77	1.63
6. mbean	27.75	6.90	1.97	1.48
7. movie2	78.51	17.92	2.25	1.61
8. mtv1	76.55	7.97	1.99	1.30
9. mtv2	24.79	6.23	1.71	1.48
10. news1	46.70	15.48	2.51	1.37
11. news2	37.94	37.94	2.31	1.58
12. race	18.71	14.58	6.96	2.80
13. sbowl	18.15	18.15	3.69	2.12
14. simpsons	55.62	16.89	1.54	1.25
15. soccer1	27.07	27.07	2.75	1.61
16. soccer2	42.11	42.11	3.34	1.22
17. star	18.90	18.90	2.92	1.50
18. talk1	14.59	10.18	2.62	1.57
19. talk2	21.64	21.64	8.47	1.68
20. terminator	37.62	15.95	2.56	2.20
21. video	11.62	11.62	3.36	1.93

entre o resultado obtido e o número de parâmetros necessários para se alcançar este valor.

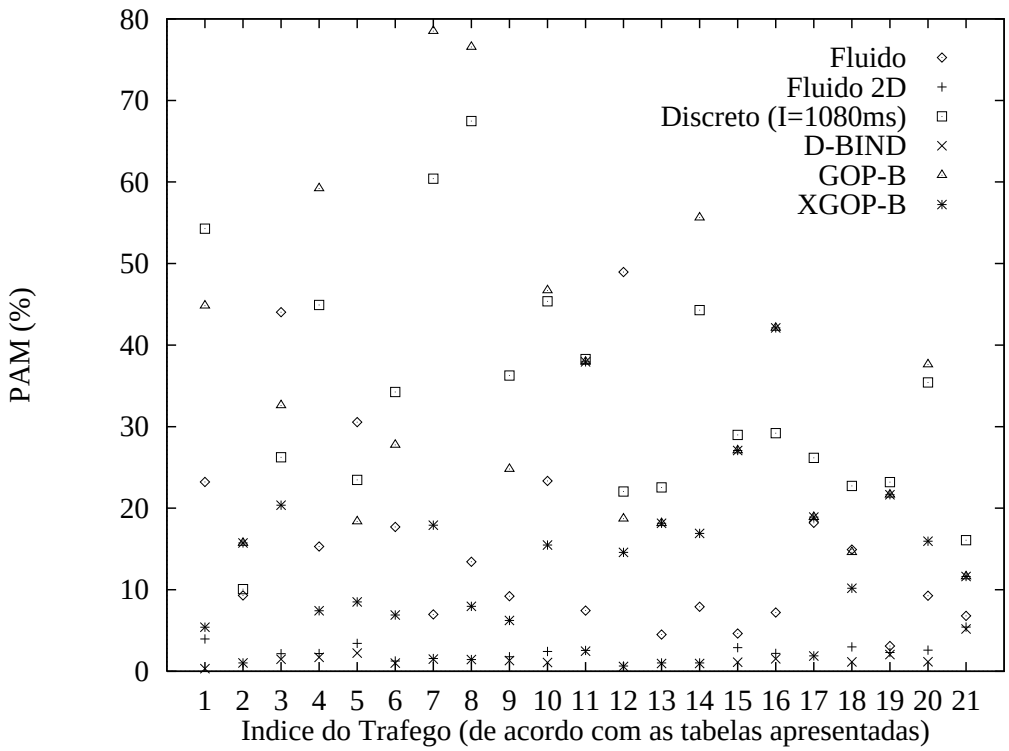


Figura 4.2: Gráfico de Perda por Aproximação do Modelo.

O comportamento das curvas de restrição de tráfego de cada um dos modelos pode ser visualizado na figura 4.3. São apresentados os resultados para dois traços de vídeo, um extraído do desenho animado *Asterix* e outro oriundo de um noticiário de TV.

4.3 Influência do fator β

Um dos resultados para o modelo *XGOP-B* foi obtido com o fator de qualidade do modelo igual a 0.5. Para este valor, pode-se observar que para alguns traços de vídeo a PAM para o modelo XGOP-B é igual à PAM para o modelo GOP-B. Isso ocorre porque para estas amostras não foram identificados nenhum ponto de quebra.

No apêndice B são mostrados graficamente os resultados da modelagem dos vídeo

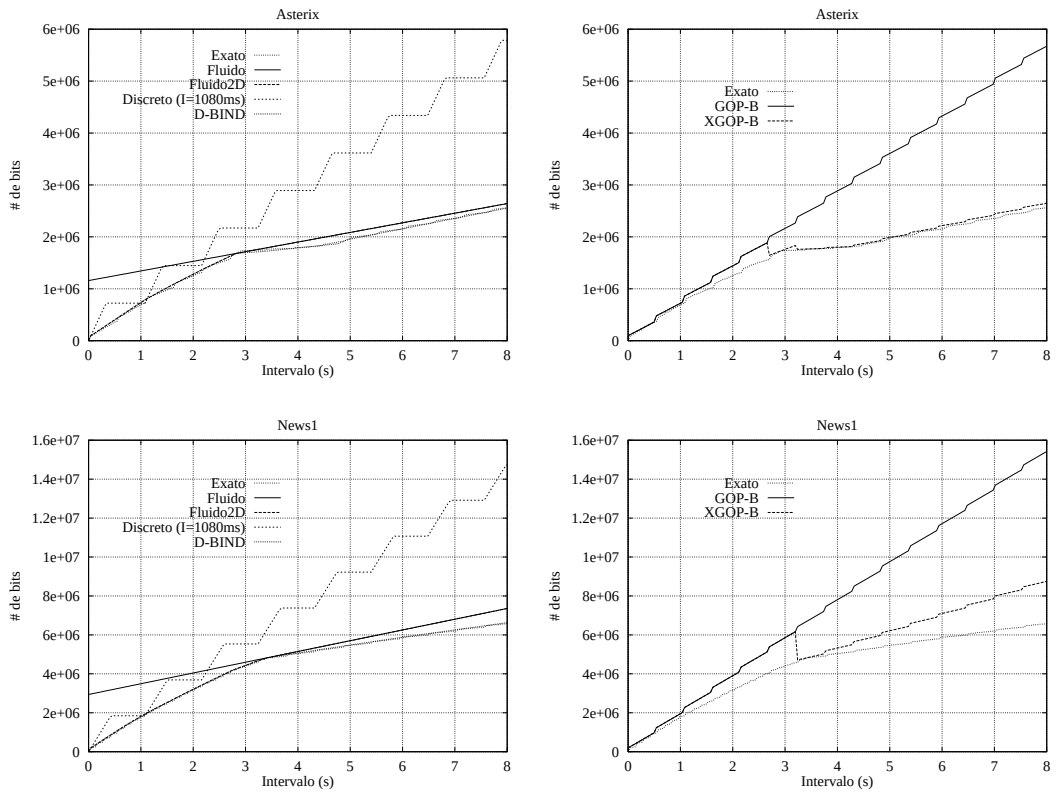


Figura 4.3: Curvas de restrição de tráfego para dois traçõe de vídeo.

para todos os modelos apresentados. Os casos onde os modelos GOP-B e XGOP-B agem com o mesmo nível de acurácia correspondem às amostras de vídeo de comportamento mais "suave", ou seja, àquelas com maior linearidade na taxa de crescimento da curva de restrição de tráfego.

No caso da acurácia para $\beta = 0.5$ não ser satisfatório, pode-se ajustar o valor de β até que os resultados desejados sejam alcançados. Quanto menor for o valor de β mais pontos de quebra serão identificados e, conseqüentemente, mais parâmetros de tráfego serão obtidos, resultando em maior acurácia de parametrização. Quanto mais pontos de quebra forem considerados, maior será o número de parâmetros do modelo e, conseqüentemente, mais complexos são o controle de admissão de conexão e o policiamento do tráfego.

Serão investigados agora os efeitos do ajuste de β sobre o nível de acurácia do modelo e sobre o número de pontos de quebra. A figura 4.4 ilustra a relação entre o fator β e o número de pontos de quebra para um intervalo máximo igual a $10 \times D$,

ou seja trinta vezes o tamanho de um GOP.

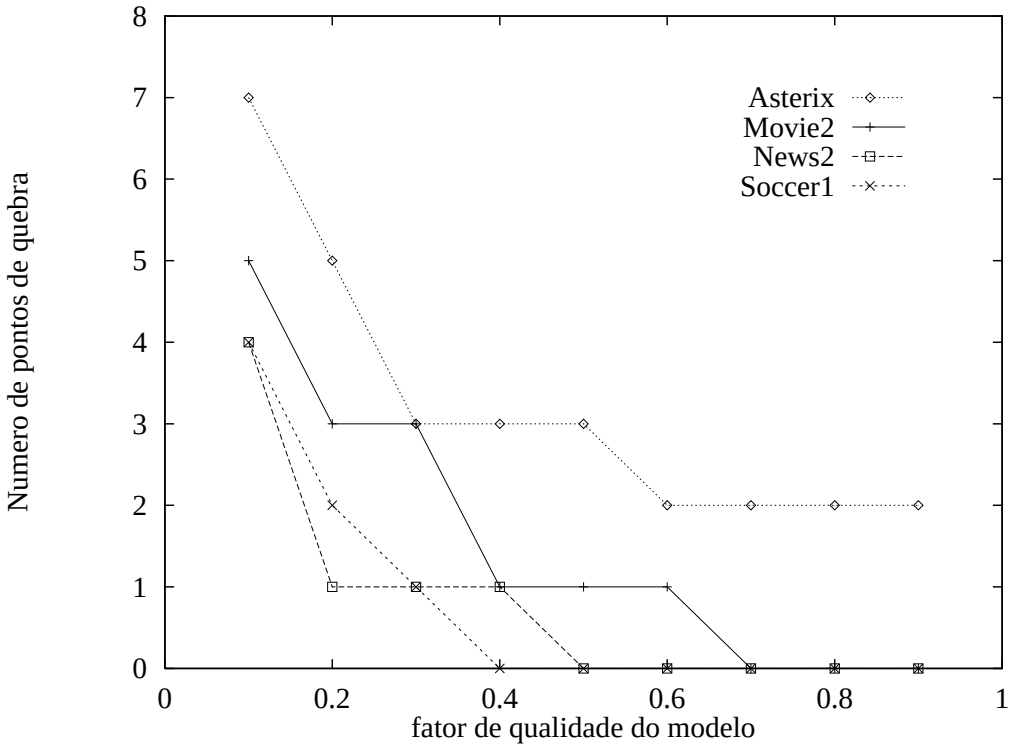


Figura 4.4: Fator β $\times$ número de pontos de quebra.

Com os mesmos traços de vídeo utilizados na figura 4.4, o gráfico ilustrado a seguir (figura 4.5) apresenta os níveis de acurácia dos modelos, representados por seus respectivos valores de PAM, também em função do fator de qualidade do modelo.

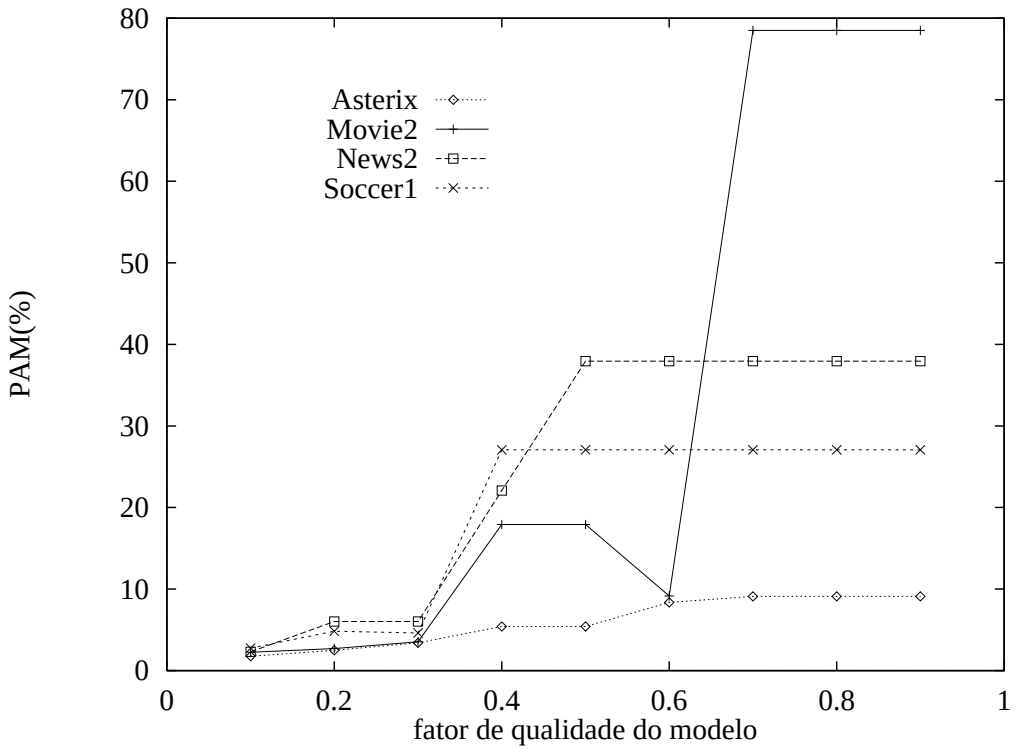


Figura 4.5: Fator β $\times$ PAM.

Capítulo 5

Conclusões

CADA vez mais, as aplicações multimídias introduzem restrições em termos de garantia de qualidade de serviço diferenciadas daquelas impostas até então. Algumas aplicações serão tão exigentes que a rede deve ser capaz de suportá-las baseando-se em garantias deteriminísticas de desempenho, ou seja, os parâmetros de QoS contratados serão sempre respeitados, independentemente da influência de outras fontes de tráfego.

Nas redes de serviços integrados, dois componentes são o fundamento para se prover um serviço garantido: o modelo de tráfego e o controle de admissão de conexões. A sintonia entre estes componentes se faz tão necessária que uma escolha inadequada pode resultar em condições indesejadas. Se, em conjunto, o modelo do tráfego e o mecanismo de CAC forem super-restritivos, ou seja, se desnecesariamente rejeitam conexões que poderiam ser admitidas na rede sem violarem parâmetros de QoS de outras conexões, então a rede será subutilizada, uma situação dispendiosa dado os altos custos dos recursos. Por outro lado, se o modelo de tráfego e o mecanismo de CAC não forem restritivos o suficiente, admitindo conexões além do número máximo suportado, violações dos parâmetros de QoS provavelmente ocorrerão, podendo ocorrer congestionamentos na rede.

Para aplicações tolerarem momentâneas degradações na qualidade do serviço oferecido, garantias estatísticas podem ser perfeitamente aplicadas em prol do ganho de

multiplexação estatística. Contudo, muitas aplicações são tão exigentes em termos da qualidade do serviço que garantias determinísticas devem ser suportadas.

O fundamento para se prover garantias determinísticas de desempenho recai principalmente no modelo determinístico de tráfego. Vários trabalhos apresentados na literatura propõem soluções para caracterização determinística de tráfegos. Dentre eles, foram destacados nesta tese os modelos de alocação pelo pico, o modelo (σ, ρ), o modelo $(\vec{\sigma}, \vec{\rho})$, o modelo *Discreto* e o modelo *D-BIND*. O ponto chave neste estudo consiste no fato destes modelos terem sido concebidos para a modelagem de tráfegos em taxa variável de bits em geral. Alguns tipos de tráfego são tão representativos para as redes de serviços integrados que a caracterização destes tráfegos através de modelos genéricos pode ser inadequada. Neste contexto se encaixa o vídeo comprimido, em particular o vídeo MPEG, que tem se mostrado como um dos padrões *de facto* nas comunicações multimídias.

Foi realizado neste trabalho um estudo aprofundado sobre o comportamento temporal dos vídeos codificados segundo o padrão MPEG. Desta análise foram extraídas duas propriedades fundamentais. A primeira se deve à observação que, na geração de um vídeo MPEG, ocorrem surtos de informação no início de cada grupo de quadros (GOP). Isto se deve à geração, nestes instantes, de intraquadros (tipo I), que são geralmente bem maiores que os interquadros (tipos P e B). O reflexo desta propriedade na função de restrição de tráfego corresponde a um comportamento muito particular. Baseado neste princípio foi proposto o modelo GOP-B (*GOP-Based model*). Este modelo utiliza apenas dois parâmetros de tráfego e alcança níveis de acurácia de parametrização da ordem do modelo *Discreto*, que por sua vez utiliza quatro parâmetros.

A primeira propriedade observada é fundamental, mas incompleta para se definir o comportamento global de uma amostra de vídeo MPEG. As cenas de uma amostra de vídeo, na prática, variam constantemente e, por conseqüência, as taxas de compressão alcançadas pelo codificador apresentam grandes variações ao longo do tempo. Neste caso, a taxa média de crescimento da função de restrição de tráfego apresenta desvios acentuados, correspondendo à presença de pontos de quebra. Esta

segunda propriedade levou à proposta de uma extensão ao modelo GOP-B, o modelo XGOP-B (*eXtended GOP-Based Model*). Este modelo segue o princípio do modelo GOP-B, mas incorpora em sua estrutura a propriedade dos pontos de quebra.

A medida de comparação entre os modelos utiliza o conceito de Perda por Aproximação do Modelo. Este índice quantifica a diferença entre as áreas sob a curva de contorno empírico, isto é, a de máxima acurácia, e sob as curvas de restrição de tráfego de cada um dos modelos.

Os resultados mostram que os modelos (σ, ρ), *Discreto* e *GOP-B* são prejudicados pelo fato de terem a taxa média de crescimento da curva constante, ou seja, por não considerarem que as fontes analisadas apresentam diferentes níveis de rajada para diferentes comprimentos de intervalo de tempo. Entretanto, o modelo *GOP-B* apresenta uma vantagem sobre os demais. A solução na determinação de seus parâmetros é única, ou seja, os parâmetros são obtidos de forma automática. No caso do modelo (σ, ρ), existe uma infinidade de possíveis soluções, o que requer que o usuário especifique de alguma forma qual curva será utilizada para a modelagem de um tráfego. Se a escolha for mal dimensionada, certamente recursos podem ser sobrealocados desnecessariamente. Para o modelo *Discreto*, o parâmetro I também deve ser especificado *a priori*. Mais uma vez, uma opção mal calculada pode se refletir em menores índices de acurácia de parametrização.

Os modelos $(\vec{\sigma}, \vec{\rho})$, *D-BIND* e *XGOP-B*, ao contrário dos demais, permitem que a taxa média de crescimento da curva de restrição de tráfego possa ser diversificada para diferentes valores de intervalo de tempo. Entretanto, os modelos $(\vec{\sigma}, \vec{\rho})$ e *D-BIND* exigem o cálculo do contorno empírico para todos os valores de intervalo de tempo. Apesar de resultarem em em bons índices de acurácia de parametrização, o cálculo do contorno empírico introduz uma carga extra de processamento. O modelo *D-BIND*, assim como os modelos (σ, ρ) e *Discreto*, também não permite que seus parâmetros sejam obtidos automaticamente. A escolha dos pares (R_k, I_k) é feita de forma objetiva, e uma escolha mal ponderada pode refletir em queda no nível de acurácia do modelo. O modelo *XGOP-B*, por sua vez, resulta em alta acurácia de parametrização e ainda permite que o grau de acurácia seja ajustado através do

fator de qualidade do modelo. Nesta tese foram realizados os cálculos de extração dos parâmetros para diversos valores para o fator de qualidade do modelo. Os resultados obtidos apontam a utilização do modelo *XGOP-B* como opção extremamente atrativa na caracterização determinística de tráfegos de vídeo MPEG.

Como sugestões para trabalhos futuros, pode-se destacar:

- obter resultados com relação ao número de conexões admissíveis para diversas fontes de vídeo aplicando-se os modelos GOP-B e XGOP-B;
- aplicar as mesmas técnicas apresentadas no estudo do comportamento de fontes de vídeo codificadas em outras versões do padrão MPEG;
- verificar a influência do fator de qualidade do modelo β na otimização da utilização da banda-passante da rede;
- verificar o *overhead* introduzido pelo aumento no número de parâmetros sobre o processamento dos mecanismos de controle de admissão de conexão;
- realizar medidas, também em termos de número de conexões admissíveis ou índice de utilização da banda-passante, para diversos algoritmos de escalonamento de pacotes, como FCFS (*First-come First-served*), RCSP (*Rate Controlled Static Priority*), EDF (*Earliest Deadline First*) e RPQ (*Rotating Priority Queues*).

Referências Bibliográficas

[1] H. Zhang e D. Ferrari, "Improving utilization for deterministic service in multimedia communication", *1994 International Conference on Multimedia Computing and Systems*, pp. 295–304, maio de 1994.

[2] D. Ferrari e D. Verma, "A scheme for real-time channel stablishment in wide-area networks", *IEEE Journal on Selected Areas in Communications*, pp. 368–379, abril de 1990.

[3] D. Clark, S. Shenker e L. Zhang, "Supporting real-time applications in an integrated services packet network: Architecture and mechanism", in *ACM SIGCOMM'92*, pp. 14–26, 1992.

[4] D. E. Wrege, E. W. Knightly, H. Zhang e J. Liebeherr, "Deterministic delay bounds for VBR video in packet-switching networks: Fundamental limits and practical tradeoffs", *IEEE/ACM Transactions on Networking*, pp. 352–362, junho de 1996.

[5] E. W. Knightly, D. E. Wrege, J. Liebeherrand e H. Zhang, "Fundamental limits and tradeoffs of providing deterministic guarantees to VBR video traffic", in *ACM Sigmetrics*, Ottawa, Canadá, maio de 1995.

[6] M. F. Abdalla, "Análise de mecanismos de controle de admissão de conexão para redes ATM", Tese de Mestrado, Universidade Federal do Rio de Janeiro, PEE/COPPE/UFRJ, Rio de Janeiro, setembro de 1996.

[7] M. Faerman, "Monitoração e policiamento de tráfego em sistemas de comunicação multimídia", Tese de Mestrado, Universidade Federal do Rio de Janeiro, PEE/COPPE/UFRJ, Rio de Janeiro, agosto de 1995.

[8] E. Knightly e H. Zhang, "Traffic characterization and switch utilization using a deterministic bounding interval dependent traffic model", in *IEEE INFOCOM'95*, Boston, MA, abril de 1995.

[9] E. W. Knightly, R. F. Mines e H. Zhang, "Deterministic characterization and network utilizations for several distributed real-time applications", in *IEEE WORDS'94*, Dana Point, CA, outubro de 1994.

[10] M. Krunz e H. Hughes, "A traffic model for MPEG-coded VBR streams", in *Proceedings of ACM SIGMETRICS'95*, pp. 47–55, 1995.

[11] R. L. Cruz, "A calculus for network delay, part I: Network elements in isolation", *IEEE Transactions on Information Theory*, pp. 114–121, janeiro de 1991.

[12] M. D. de Amorim e O. C. M. B. Duarte, "Eficiência de modelos determinísticos de tráfego na representação de fontes de vídeo MPEG", *XXIV Seminário Integrado de Software e Hardware - SEMISH'97*, agosto de 1997.

[13] M. D. de Amorim e O. C. M. B. Duarte, "Um novo modelo determinístico para tráfegos de vídeo MPEG baseado em grupo de quadros", *XV Simpósio Brasileiro de Telecomunicações - SBT'97*, setembro de 1997.

[14] M. D. de Amorim e O. C. M. B. Duarte, "Um modelo determinístico para caracterização de tráfegos de vídeo MPEG", *II Seminário Franco-Brasileiro em Sistemas Informáticos Distribuídos - SFBSID'97*, novembro de 1997.

[15] E. W. Knightly e H. Zhang, "D-BIND: An acurate traffic model for providing QoS guarantees to VBR traffic", *IEEE/ACM Transactions on Networking*, vol. 5, no. 2, pp. 219–231, abril de 1997.

[16] M. Krunz e S. Tripathi, "On the characterization of VBR MPEG streams", in *Performance Evaluation Review (Proceedings of ACM SIGMETRICS'97 Conference)*, pp. Vol. 25, No. 1, pp. 192–202, 1997.

[17] S. El-Henaoui e S. Tohme, "UPC parameters for real-time VBR MPEG traffic applying dynamic bandwidth allocation", *ICC'96*, junho de 1996.

[18] C. V. N. de Albuquerque, "Arquitetura de implementação de alto desempenho para sistemas de comunicação multimídia", Tese de Mestrado, Universidade Federal do Rio de Janeiro, PEE/COPPE/UFRJ, Rio de Janeiro, abril de 1995.

[19] J. Liebeherr, D. E. Wrege e D. Ferrari, "Exact admission control for networks with a bounded delay service", *IEEE/ACM Transactions on Networking*, pp. 885–901, dezembro de 1996.

[20] CCITT, "Recommendation I.121: Broadband aspects of ISDN". 1988.

[21] D. Ferrari, "Client requirements for real-time communication services", in *IEEE Communications Magazine*, pp. 28(11):65–72, novembro de 1990.

[22] E. W. Knightly e P. Rossaro, "Effects of smoothing on end-to-end performance guarantees for VBR video", in *International Symposium on Multimedia Communications and Video Coding*, Brooklyn, NY, outubro de 1995.

[23] C.-Y. Hsu e A. Ortega, "Joint encoder and VBR channel optimization with buffer and leaky bucket constraints", in *International Symposium on Multimedia Communications and Video Coding*, Brooklyn, NY, outubro de 1995.

[24] E. Knightly e P. Rossaro, "Improving QoS through traffic smoothing", in *IFIP IWQoS'96*, Paris, France, março de 1996.

[25] D. Saha, S. Mukherjee e S. K. Tripathi, "Muiti-rate traffic shaping and end-to-end performance guarantees in ATM", in *Proceedings of 1994 International Conference on Network Protocols*, Boston, Massachussets, pp. 188–195, outubro de 1994.

[26] M. D. de Amorim e O. C. Duarte, "Efeitos da utilização de filas FIFO na suavização de tráfegos VBR", relatório técnico, Grupo de Teleinformática e Automação da COPPE/UFRJ, setembro de 1996.

[27] P. Pancha e M. E. Zarki, "A look at the MPEG video coding standard for variable bit rate video transmission", *IEEE INFOCOM'92*, maio de 1992.

[28] M. Krunz e S. Tripathi, "Exploiting the temporal structure of MPEG video for the reduction of bandwidth requirements", in *IEEE INFOCOM'97*, abril de 1997.

[29] C. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks", in *IEEE Transactions on Automatic Control*, pp. 39(5):913–931, maio de 1994.

[30] R. L. Cruz, "A calculus for network delay, part II: Network analysis", *IEEE Transactions on Information Theory*, pp. 121–141, janeiro de 1991.

[31] E. W. Knightly, "H-BIND: A new approach to providing statistical performance guarantees to VBR traffic", in *IEEE INFOCOM'96*, San Francisco, CA, março de 1996.

[32] D. L. Gall, "MPEG: A video compression standard for multimedia applications", *Communications of the ACM*, pp. 305–313, abril de 1991.

[33] P. C. S. Vidal, "Evolução do padrão MPEG", relatório técnico, Grupo de Teleinformática e Automação da COPPE/UFRJ, dezembro de 1997.

[34] O. Rose, "Statistical properties of MPEG video traffic and their impact on traffic modelling in ATM systems", *Proceedings of the 20th Annual Conference on Local Computer Networks*, 1995.

[35] K. L. Gong, "Berkeley MPEG-1 video encoder, user's guide", relatório técnico, University of California, Berkeley, Computer Science Division - EECS, março de 1994.

[36] E. W. Knightly, *Traffic Models and Admission Control for Integrated Services Networks*. PhD thesis, University of California at Berkeley, 1996.

[37] H. Zhang, "Service disciplines for guaranteed performance service in packet-switching networks", in *IEEE*, pp. 83(10):1374–1399, 1995.

[38] D. Lindley, "On the theory of queues with a single server", in *Cambridge Philosophical Society*, pp. 48:277–289, 1952.

[39] A. Banerjea e S. Keshav, "Queueing delays in rate controlled networks", in *IEEE INFOCOM'93*, pp. 547–556, 1993.

[40] A. Parekh e R. Gallager, "A generalized processor sharing approach to flow control – the single node case", in *IEEE INFOCOM'93*, pp. 521–530, 1993.

[41] H. Zhang e D. Ferrari, "Rate-controlled service disciplines", in *Journal of High Speed Networks*, pp. 3(4):389–412, 1994.

Apêndice A

CAC Determinístico

AS condições de controle de admissão determinístico recaem nas técnicas de análise de atrasos apresentadas em [1, 11, 29, 36], representadas na figura A.1. O eixo horizontal representa o tempo e o eixo vertical bits. A curva superior representa o número total de bits que chegaram na fila de transmissão até o instante τ e a curva inferior representa o número total de bits transmitidos até o instante τ. A diferença entre as duas curvas é o número de bits na fila, chamada função de registro (*backlog function*). Quando a função de registro retorna a zero (as duas curvas se encontram), não há nenhum bit na fila e um período ocupado está encerrado. A chave para esta análise é a seguinte: se a curva superior for um contorno de pior caso sobre as chegadas de pacotes de todas as fontes, então o atraso máximo pode ser expresso em função das duas curvas. Duas observações podem ser feitas [11, 36]: o máximo período ocupado prevê um limite superior de atraso para qualquer escalonador de pacotes do tipo trabalho conservativo [37]; o máximo *backlog* dividido pela velocidade do canal provê um limite superior de atraso para um escalonador do tipo FCFS (*First-Come First-Served*) [37].

A função de restrição de tráfego provê o contorno desejado sobre a chegada de pacotes em um intervalo qualquer de duração t. Assim, com o agregado das respectivas funções de restrição de tráfego $\mathcal{F}(t)$ das fontes individuais formando a curva superior da figura A.1, condições de controle de admissão para atraso e vazão determinísticos podem ser obtidas.

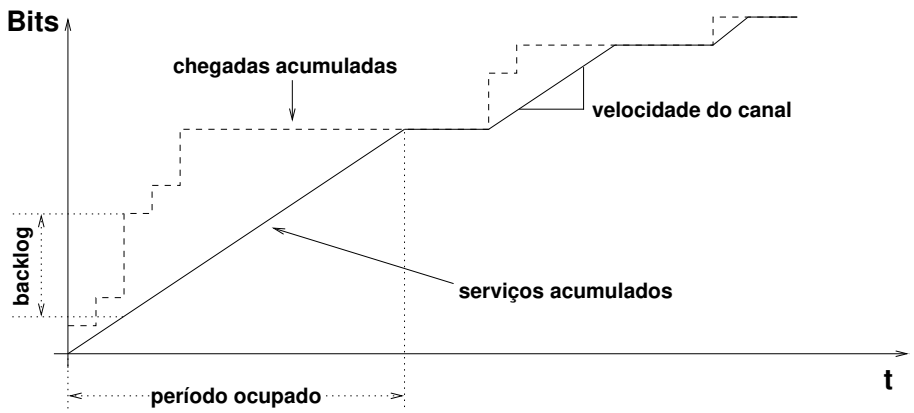


Figura A.1: Conceitos: atraso, *backlog* e período ocupado.

Formalmente, o teorema seguinte provê um limite superior absoluto de atraso para um escalonador FCFS.

Teorema 1 *Considere um escalonador que serve pacotes com a ordem primeiro que chega primeiro servido (FCFS). Para $j = 1, 2, ..., N$ conexões multiplexadas contornadas por suas respectivas funções de restrição de tráfego $\mathcal{F}_j(t)$, e com velocidade do canal l, um limite determinístico de atraso para todas as conexões é dado por*

$$d = \frac{1}{l} max_{t \geq 0} \left\{ \sum_{j=1}^{N} \mathcal{F}_j(t) - l \times t \right\}. \tag{A.1}$$

Prova: A evolução no tempo do comportamento do multiplexador pode ser expressa pela equação de Lindley [38]:

$$q(s + \delta) = \Big(q(s) + \sum_{j} A_j[s, s + \delta] - l \times \delta \Big)^+, \tag{A.2}$$

onde $q(s)$ é a função de registro (ou ocupação da fila) em um instante s, $(x)^+ = max(0, x)$ e $A_j[s, s + \delta]$ representa as chegadas reais de uma conexão j no intervalo $[s, s + \delta]$.

Da equação A.2 segue que

$$q(s) = max_{\tau \leq s} \left\{ \sum_{j=1}^{N} A_j[\tau, s] - l \times (s - \tau) \right\}. \quad \text{(A.3)}$$

Para a disciplina de serviço FCFS, um limite superior no tempo que um pacote pode ser atrasado no multiplexador é dado por

$$d = max_s \frac{q(s)}{l} \quad \text{(A.4)}$$

ou a máxima ocupação da fila durante todo o tempo s dividida pela velocidade do canal. Assim, temos

$$d = \frac{1}{l} max_s max_{\tau \leq s} \left\{ \sum_{j=1}^{N} A_j[\tau, s] - l \times (s - \tau) \right\}. \quad \text{(A.5)}$$

Em seguida, utilizamos

$$max_{\tau \leq s} \left\{ \sum_{j=1}^{N} A_j[\tau, s] \right\} \leq \sum_{j=1}^{N} max_{\tau \leq s} A_j[\tau, s] \quad \text{(A.6)}$$

e

$$A_j[\tau, s] \leq \mathcal{F}_j(s - \tau) \quad \text{(A.7)}$$

para $s \geq \tau$.

Combinando-se as equações A.6 e A.7, a equação A.1 segue. □.

As técnicas utilizadas na prova acima são baseadas nos resultados obtidos por Cruz em [11]. Limites de atraso para outras disciplinas de serviço também podem ser obtidas. Por exemplo, limites de atraso para escalonadores como *Hierarchical-Round-Robin*, *Static-Priority*, *Earliest-Due-Date*, *Packetized Generalized Processor Sharing* e *Rotating Priority Queues* também podem ser obtidos a partir das funções de restrição de tráfego [11, 19, 37, 39, 40]. Por exemplo, o teorema seguinte é apresentado e provado em [41] para obtenção de limites de atraso em um escalonador *Static Priority*.

Teorema 2 *Assuma que um escalonador SP (Static Priority) tenha n níveis prioridade. Seja C_q o conjunto de conexões no nível q, e a j-ésima conexão em C_q satisfaz a função de restrição de tráfego $b_{q,j}(\cdot)$. Sendo a velocidade do canal l e o tamanho máximo de um pacote $\bar{s}$, o máximo atraso sofrido por qualquer pacote em um nível k é limitado por d_k, onde*

$$d_k = max\{t \geq 0 | b'_k(t) \geq l \times t\} \tag{A.8}$$

e $b'_k(\alpha)$ é definido para todo α como

$$b'_k(\alpha) = max_{\beta \geq 0}\left\{\bar{s} + \sum_{j \in C_k} b_{k,j}(\beta) + \sum_{q=1}^{k-1}\sum_{j \in C_q} b_{q,j}(\alpha + \beta) - l \times \beta\right\}. \tag{A.9}$$

Os teoremas A.1 e A.2 podem ser usados como testes de CAC para escalonadores dos tipos FCFS e SP, verificando se um grupo de conexões podem ser multiplexadas de forma que cada pacote de cada conexão possa ser entregue dentro dos limites de atraso oferecidos ao cliente da rede. Assim, os testes de CAC e os limites de atraso das equações A.1 e A.9 podem ser vistos também em termos do número máximo de conexões admissíveis para uma determinada qualidade de serviço. Por exemplo, para um escalonador do tipo FCFS, a equação A.1 pode ser reescrita de forma a expressar o número máximo de conexões admissíveis como função do limite de atraso:

$$N(d) = max\left\{n | \frac{1}{l} max_{t \geq 0}\left\{\sum_{j=1}^{n} b_j(t) - l \times t\right\} \leq d\right\}. \tag{A.10}$$

Obtido o número máximo de conexões admissíveis, a utilização média do multiplexador pode ser calculada como

$$U = \sum_{j=1}^{N} \frac{R_\infty}{l}, \tag{A.11}$$

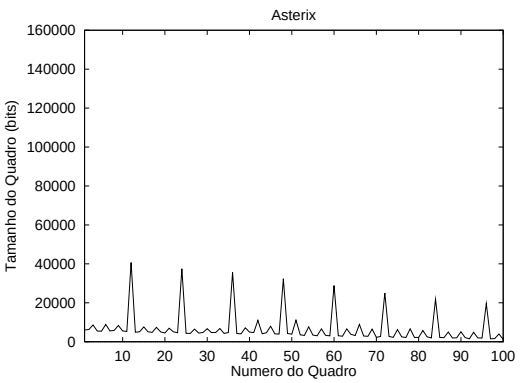
onde R_∞ é a taxa média de longo termo da conexão j.

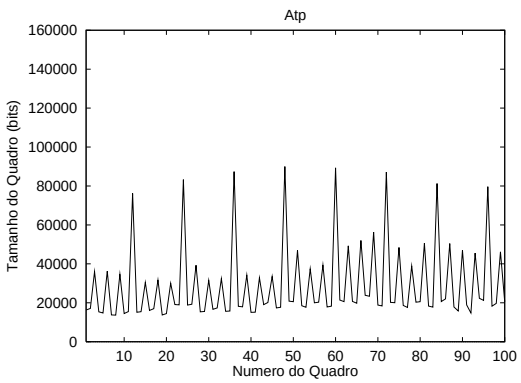
Fica claro a partir da equação A.11 que a utilização média do multiplexador aumenta com o número de conexões admitidas na rede. O mais importante é que, como pode ser observado na equação A.10, quanto mais acurada for a função de restrição de tráfego, maior será o número de conexões admissíveis e, por conseqüência, maior nível de utilização da rede será obtido.

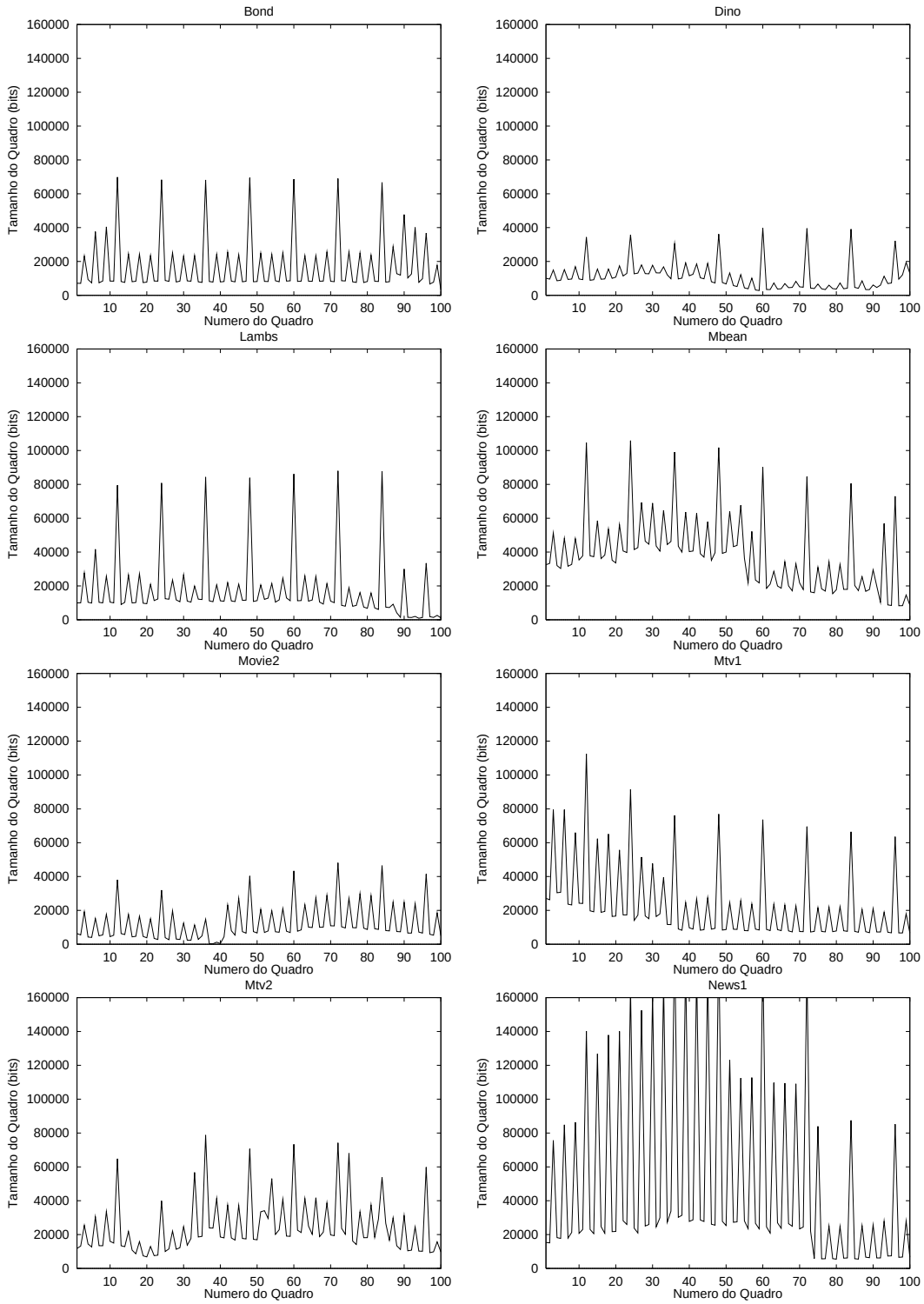
Apêndice B

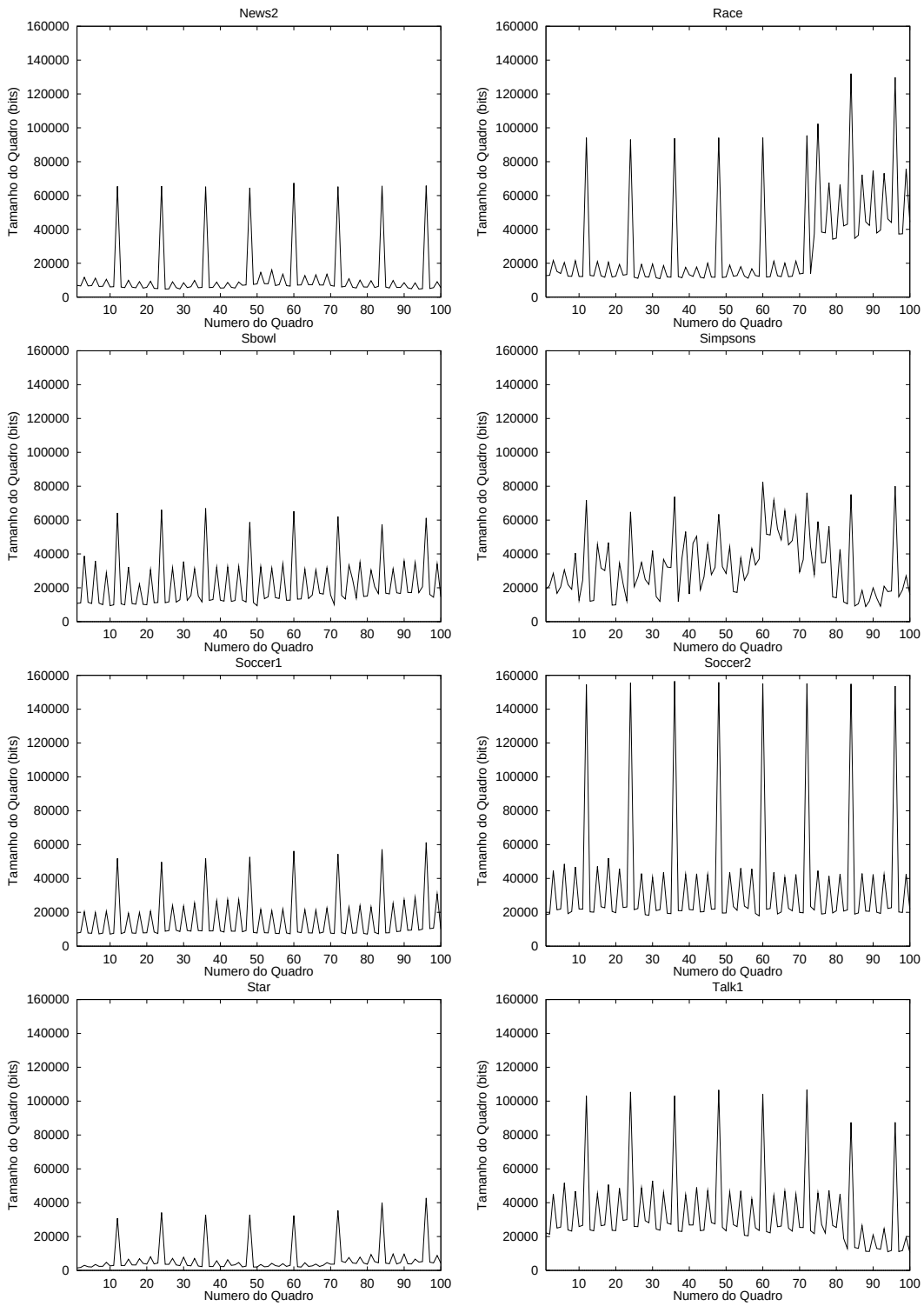
Comportamento das Amostras de Vídeo

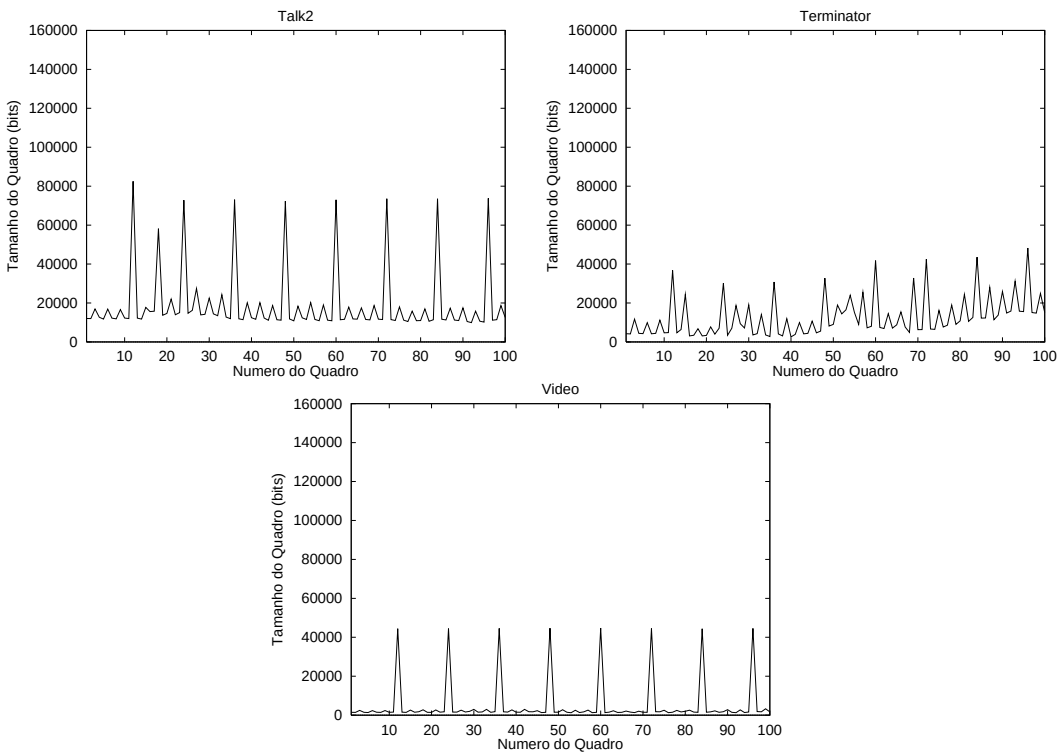
ESTE apêndice apresenta os traços dos vídeos explorados neste trabalho.






Bond
Dino
Lambs
Mbean
Movie2
Mtv1
Mtv2
News1
Tamanho do Quadro (bits)
Numero do Quadro
0
20000
40000
60000
80000
100000
120000
140000
160000
10
20
30
40
50
60
70
80
90
100

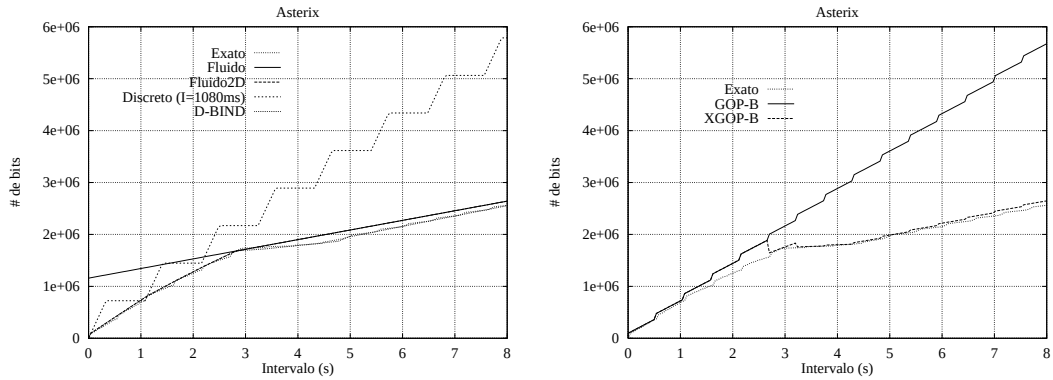

News2
Race
Sbowl
Simpsons
Soccer1
Soccer2
Star
Talk1
Tamanho do Quadro (bits)
Numero do Quadro
0
20000
40000
60000
80000
100000
120000
140000
160000
10
20
30
40
50
60
70
80
90
100

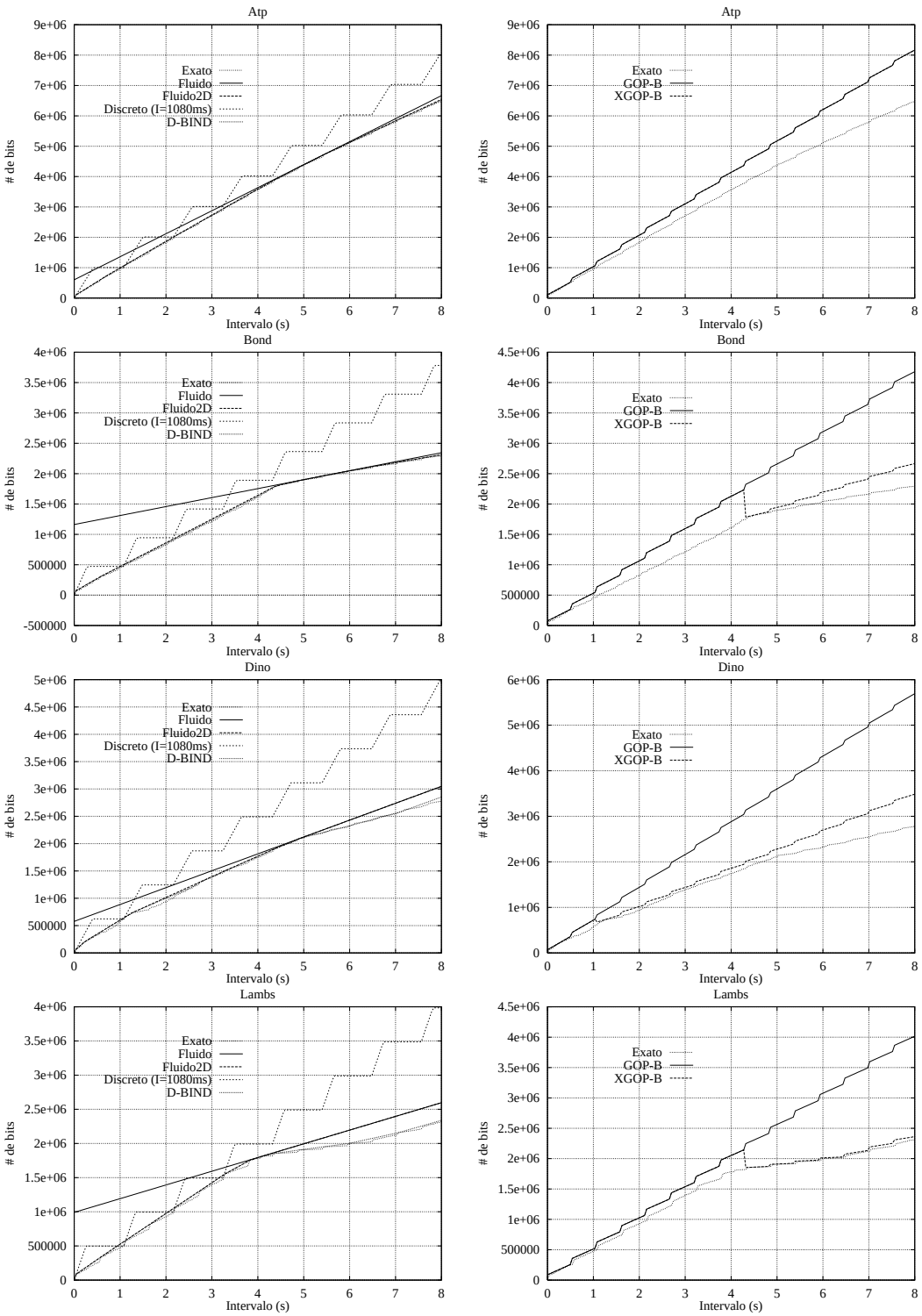

Talk2
Tamanho do Quadro (bits)
Numero do Quadro
Terminator
Tamanho do Quadro (bits)
Numero do Quadro
Video
Tamanho do Quadro (bits)
Numero do Quadro

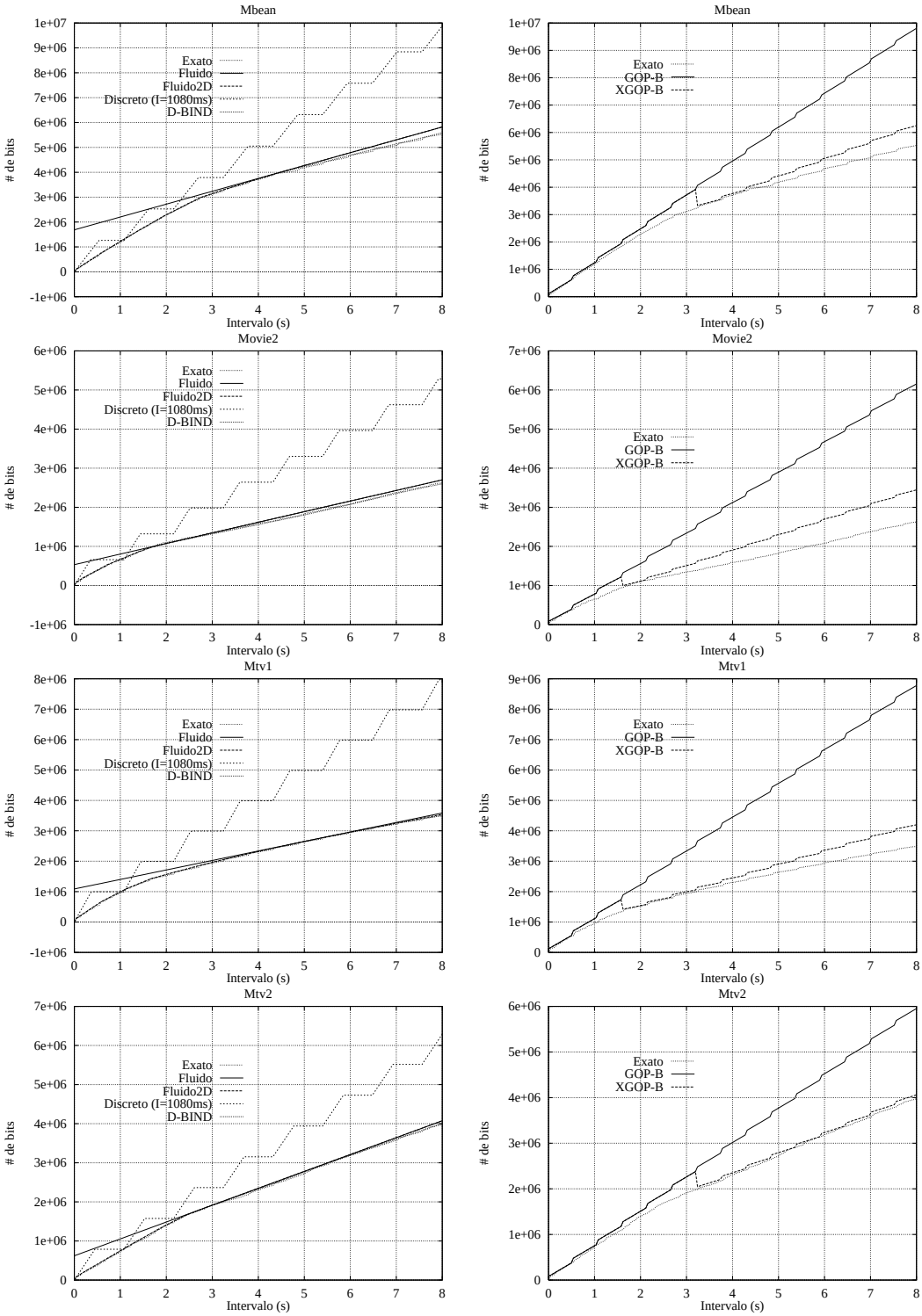
Apêndice C

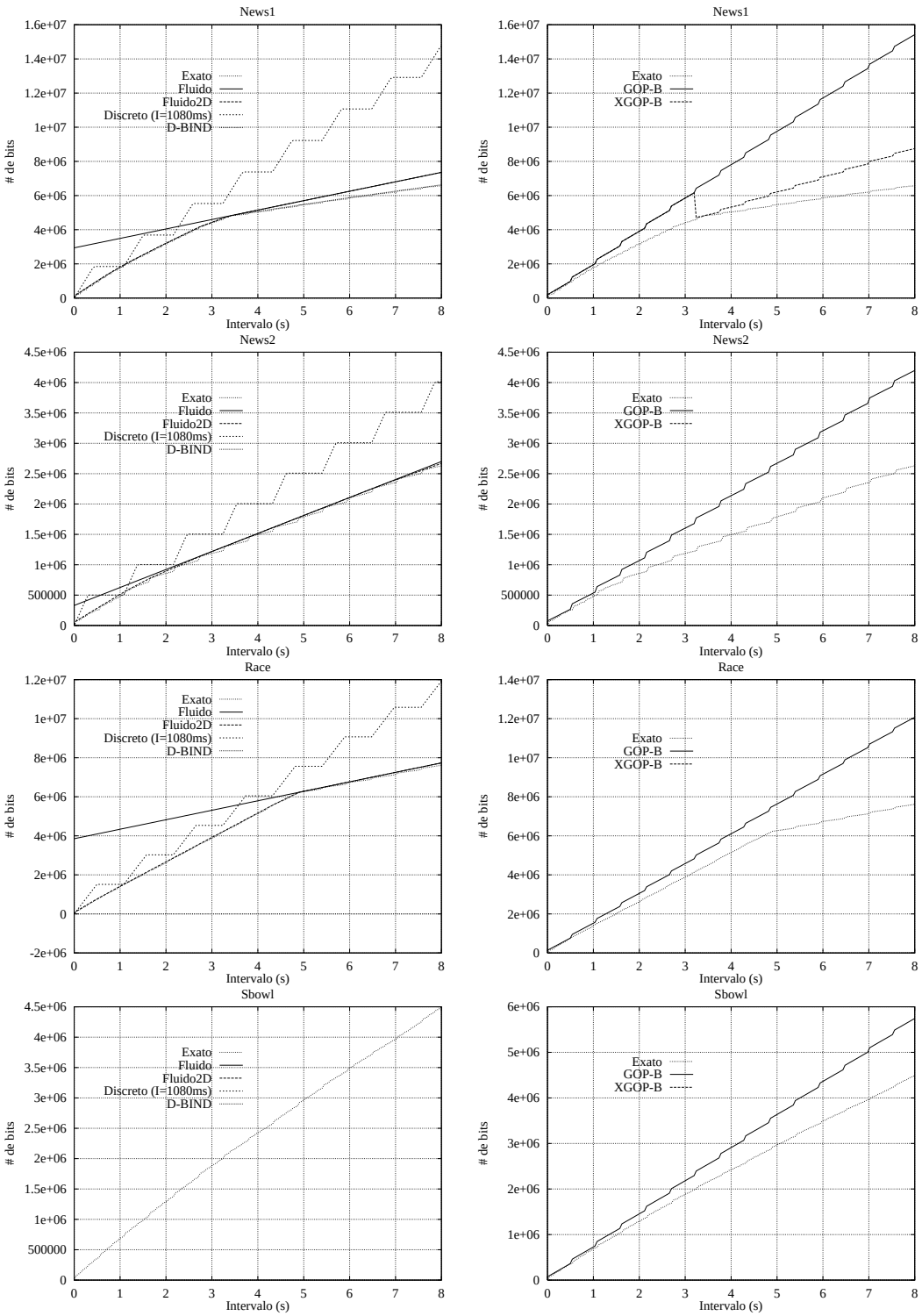
Modelagem das Amostras de Vídeo

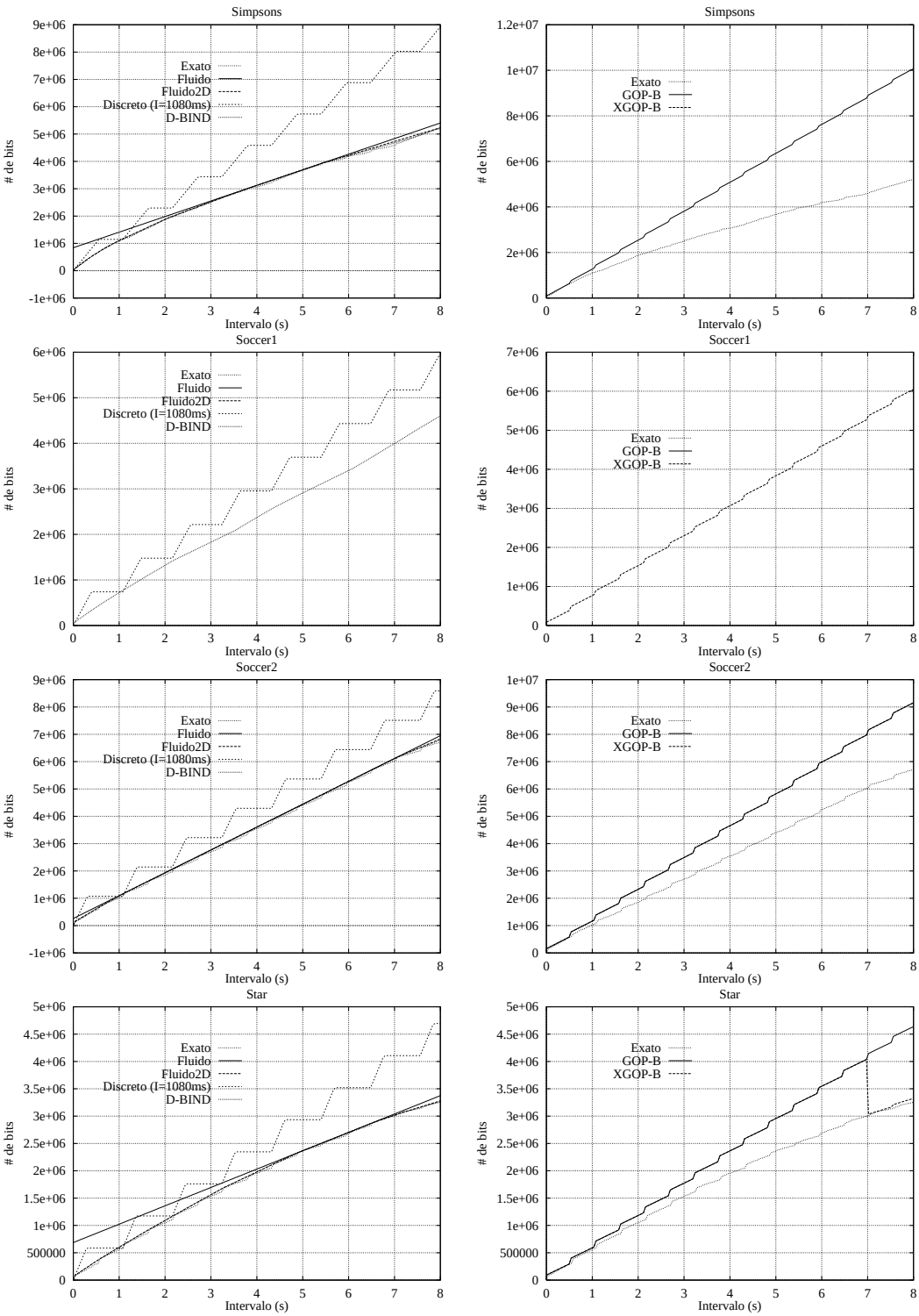
ESTE apêndice apresenta os resultados da modelagem dos traços de vídeo analisados através de suas respectivas funções de restrição de tráfego para cada um dos modelos.

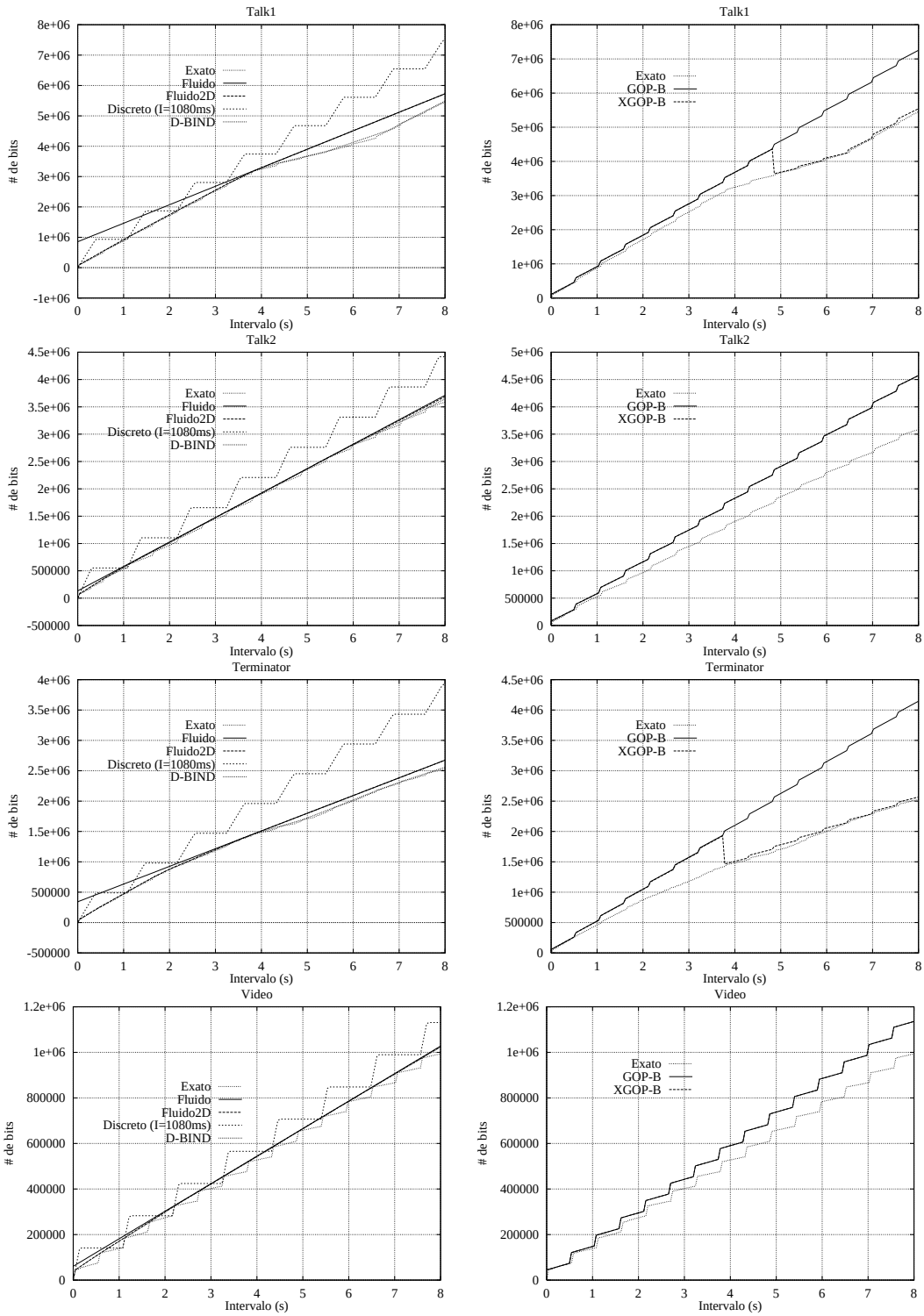



Atp
Exato
Fluido
Fluido2D
Discreto (I=1080ms)
D-BIND
GQP-B
XGQP-B
de bits
Intervalo (s)
Bond
Dino
Lambs


Mbean
Exato
Fluido
Fluido2D
Discreto (I=1080ms)
D-BIND
de bits
Intervalo (s)
Mbean
Exato
GOP-B
XGOP-B
de bits
Intervalo (s)
Movie2
Exato
Fluido
Fluido2D
Discreto (I=1080ms)
D-BIND
de bits
Intervalo (s)
Movie2
Exato
GOP-B
XGOP-B
de bits
Intervalo (s)
Mtv1
Exato
Fluido
Fluido2D
Discreto (I=1080ms)
D-BIND
de bits
Intervalo (s)
Mtv1
Exato
GOP-B
XGOP-B
de bits
Intervalo (s)
Mtv2
Exato
Fluido
Fluido2D
Discreto (I=1080ms)
D-BIND
de bits
Intervalo (s)
Mtv2
Exato
GOP-B
XGOP-B
de bits
Intervalo (s)


News1
de bits
Intervalo (s)
Exato
Fluido
Fluido2D
Discreto (I=1080ms)
D-BIND
News1
de bits
Intervalo (s)
Exato
GQP-B
XGQP-B
News2
de bits
Intervalo (s)
Exato
Fluido
Fluido2D
Discreto (I=1080ms)
D-BIND
News2
de bits
Intervalo (s)
Exato
GQP-B
XGQP-B
Race
de bits
Intervalo (s)
Exato
Fluido
Fluido2D
Discreto (I=1080ms)
D-BIND
Race
de bits
Intervalo (s)
Exato
GQP-B
XGQP-B
Sbowl
de bits
Intervalo (s)
Exato
Fluido
Fluido2D
Discreto (I=1080ms)
D-BIND
Sbowl
de bits
Intervalo (s)
Exato
GQP-B
XGQP-B


Simpsons
Exato
Fluido
Fluido2D
Discreto (I=1080ms)
D-BIND
de bits
Intervalo (s)
Simpsons
Exato
GOP-B
XGOP-B
de bits
Intervalo (s)
Soccer1
Exato
Fluido
Fluido2D
Discreto (I=1080ms)
D-BIND
de bits
Intervalo (s)
Soccer1
Exato
GOP-B
XGOP-B
de bits
Intervalo (s)
Soccer2
Exato
Fluido
Fluido2D
Discreto (I=1080ms)
D-BIND
de bits
Intervalo (s)
Soccer2
Exato
GOP-B
XGOP-B
de bits
Intervalo (s)
Star
Exato
Fluido
Fluido2D
Discreto (I=1080ms)
D-BIND
de bits
Intervalo (s)
Star
Exato
GOP-B
XGOP-B
de bits
Intervalo (s)


Talk1
de bits
Intervalo (s)
Exato
Fluido
Fluido2D
Discreto (I=1080ms)
D-BIND
Talk1
de bits
Intervalo (s)
Exato
GOP-B
XGOP-B
Talk2
de bits
Intervalo (s)
Exato
Fluido
Fluido2D
Discreto (I=1080ms)
D-BIND
Talk2
de bits
Intervalo (s)
Exato
GOP-B
XGOP-B
Terminator
de bits
Intervalo (s)
Exato
Fluido
Fluido2D
Discreto (I=1080ms)
D-BIND
Terminator
de bits
Intervalo (s)
Exato
GOP-B
XGOP-B
Video
de bits
Intervalo (s)
Exato
Fluido
Fluido2D
Discreto (I=1080ms)
D-BIND
Video
de bits
Intervalo (s)
Exato
GOP-B
XGOP-B