XGOP-B: Um Modelo Estendido com Previsão de Ponto de Quebra para Tráfegos de Vídeo MPEG

Marcelo Dias de Amorim
amorim@gta.ufrj.br

Otto Carlos M. B. Duarte
otto@gta.ufrj.br

Grupo de Teleinformática e Automação – GTA
Universidade Federal do Rio de Janeiro
COPPE/EE - Programa de Engenharia Elétrica
C. P. 68504 - CEP 21945-970 - Rio de Janeiro - RJ - Brasil

Resumo

Nas futuras redes de serviços integrados, algumas aplicações serão tão exigentes com relação à qualidade de serviço que a rede deverá ser capaz de oferecer serviços com garantias determinísticas de desempenho. Ao se prover tal serviço, para que não haja desperdício de recursos, é fundamental que o modelo do tráfego seja o mais acurado possível. Este trabalho propõe o modelo XGOP-B (*eXtended GOP-Based model*) como uma solução extremamente atrativa na caracterização de fontes de vídeo MPEG. Os resultados mostram que, por incorporar propriedades deste padrão em sua estrutura, o modelo XGOP-B resulta em ótimos níveis de acurácia de parametrização.

Abstract

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. This work presents a behavioural analysis of MPEG video streams and proposes the XGOP-B (eXtended GOP-Based model) as a very attractive solution. The results show that, with XGOP-B model, high modelling accuracy can be obtained.

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 transferência dos pacotes. Atualmente, a maioria das novas aplicações introduzem múltiplas e 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. Desta forma, a rede precisa prover serviços heterogêneos com garantias de qualidade de serviço oferecidas para cada fluxo de dados.

Apesar da grande diversidade nas características das aplicações nas redes de serviços integrados, 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 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, as conexões terão uma certa qualidade de serviço, mas receberão uma quantidade insuficiente de banda-passante.

Em [1], dois tipos de serviços garantidos são detalhados: 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.

Independentemente da garantia do serviço ser do tipo estatística ou determinística, o desafio é alocar recursos de uma maneira eficiente que atinja a mais alta taxa de utilização possível da banda-passante 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 requerem 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*) e pelo mecanismo de policiamento de tráfego, responsáveis, respectivamente, pela determinação das conexões que podem ser estabelecidas (aceitas) e pela garantia que fontes abusivas não violarão os parâmetros contratados.

As eficiências do controle de admissão de conexão e do policiamento de tráfego dependem principalmente de três componentes [2, 3, 4]: o modelo do tráfego [1, 5, 6, 7, 8, 9, 10], o escalonamento de pacotes [11, 12] e os testes de controle de admissão [1, 2, 3, 4, 13, 14]. Dentre estes componentes, o modelo de tráfego assume posição estratégica no processo de alocação de recursos. Fundamentalmente, sem uma adequada descrição do tráfego, recursos não podem ser eficientemente alocados [3, 4].

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. Já foram propostos na literatura diversos modelos determinísticos de tráfego [1, 3, 4, 5, 6, 7]. Contudo, todos objetivam a caracterização de tráfegos VBR (*Variable Bit Rate*) 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. Será mostrado que este padrão 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.

Este trabalho introduz um novo modelo determinístico para caracterização de tráfegos de vídeo MPEG, denominado XGOP-B (*eXtended GOP-Based model*). Este modelo tem as vantagens de ser simples na extração dos parâmetros de tráfego e de obter bons resultados com um número reduzido de parâmetros. Estas características são essenciais para que o mecanismo de policiamento de tráfego demande pouco processamento. Deve-se destacar que os parâmetros para o modelo XGOP-B, ao contrário dos demais, são obtidos de forma automatizada e não subjetiva, como será mostrado no restante do trabalho. Para análise e comparação dos modelos estudados e do modelo proposto neste trabalho serão utilizadas diversas amostras de vídeo comprimido no padrão MPEG-1.

2 Caracterização Determinística de Tráfegos

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 [15, 16, 17, 18, 19]. Contudo, muitos clientes que exigem garantias de desempenho possuem tráfegos em rajada, como por exemplo vídeo comprimido [20, 21]. 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 do canal. Neste caso, a rede será subutilizada se a relação $taxa_{pico}/taxa_{média}$ dos tráfegos individuais for alta.

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 prover uma caracterização em *pior caso* da fonte de forma a fornecer 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. Esta seção 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 $(\vec{\sigma}, \vec{\rho})$, aqui identificado como modelo *Fluido em Duas Dimensões*, e o modelo *D-BIND*.

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 tráfegos 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*) [6] 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

$$\mathcal{G}[t, t + \delta t] \leq \mathcal{F}(\delta t), \qquad t > 0, \tag{1}$$

deve ser satisfeita. A função de restrição de tráfego $\mathcal{F}(\cdot)$ provê um limite invariante no 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 $\mathcal{F}(\cdot)$ exige 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

$$\zeta(\delta t) = sup\{\mathcal{G}[t, t + \delta t]\}, \qquad t \geq 0, \tag{2}$$

denominada contorno empírico (*empirical envelope*) em [3, 4], processo de contorno mínimo (*minimum envelope process*) em [22] e função exata em [7].

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 aproximação superior para a função exata ζ, ou seja,

$$\mathcal{F}(\delta t) \geq \zeta(\delta t), \qquad \forall \delta t. \tag{3}$$

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

$$\zeta(\delta t) - \mathcal{F}(\delta t), \qquad \forall \delta t, \tag{4}$$

deve ser minimizada. A razão principal para se procurar uma solução que se aproxime, ao máximo, da função exata é para que o controle de admissão de conexão opere de forma a maximizar o número de conexões admitidas. Assim, é recomendável que a subtração $\zeta(\delta t) - \mathcal{F}(\delta t)$ seja minimizada.

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

2.2 Modelo Fluido (σ, ρ)

O modelo Fluido, ou (σ, ρ), proposto por Cruz [6, 23], 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.

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 parâmetro T_{pico} (taxa de 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

$$\mathcal{F}_{(\sigma,\rho)}(\delta t) = \sigma + \rho \times \delta t. \tag{5}$$

A figura 1 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.

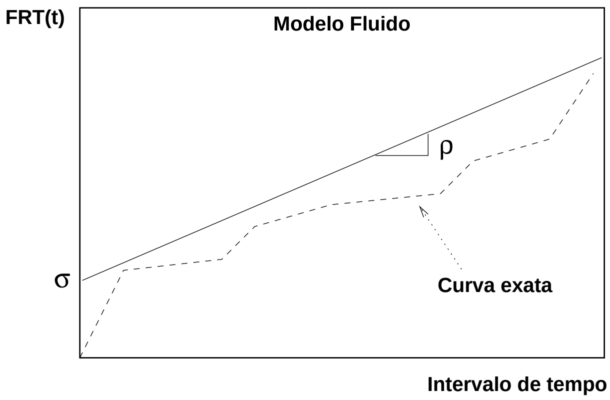


Figura 1: Função de restrição de tráfego para o modelo (σ, ρ).

Uma das maiores vantagens do modelo fluido é a utilização de um número reduzido de parâmetros de tráfego. Contudo, 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.

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

O modelo Fluido em Duas Dimensões [3, 4] 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 e é dada pela equação

$$\mathcal{F}_{(\vec{\sigma},\vec{\rho})}(\delta t) = min_i\{\sigma_i + \rho_i \times \delta t\}. \tag{6}$$

A função de restrição de tráfego dada pela equação 6 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

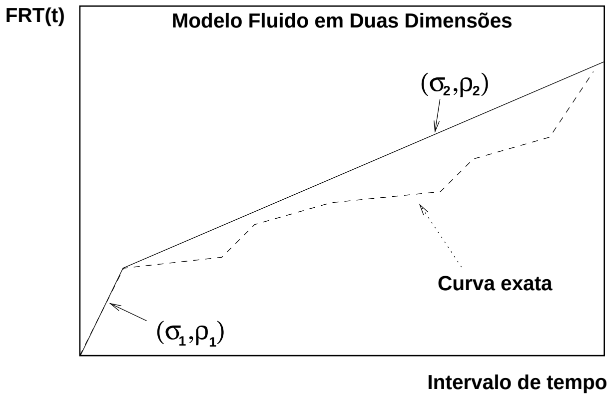


Figura 2: Função de restrição de tráfego para o modelo $(\vec{\sigma}, \vec{\rho})$.

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

$$\zeta(\delta t) = \sigma_k + \rho_k \times \delta t, \tag{7}$$

$$\zeta(t) \leq \sigma_k + \rho_k \times \delta t|_{\forall 0 \leq t \leq \delta t}, \tag{8}$$

$$\zeta(\hat{\delta t}) = \sigma_k + \rho_k \times \hat{\delta t}. \tag{9}$$

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

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

O modelo Discreto [1] 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:

- a utilização do canal $\rightarrow \frac{s_{max}}{x_{min} \times l}$;
- a razão $taxa_{pico}/taxa_{média} \rightarrow \frac{x_{ave}}{x_{min}}$;
- 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

$$\mathcal{F}_{Discreto}(\delta t) = \lfloor \frac{\delta t}{I} \rfloor \frac{I \times s_{max}}{x_{ave}} + min\{\lceil (\frac{\delta t}{I} - \lfloor \frac{\delta t}{I} \rfloor) \frac{I}{x_{min}} \rceil, \frac{I}{x_{ave}}\} \times s_{max}. \tag{10}$$

A curva de restrição de tráfego correspondente à equação 10 está plotada na figura 3. 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.

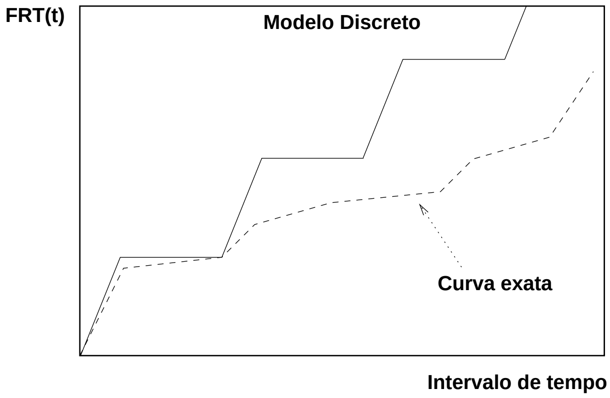


Figura 3: Função de restrição de tráfego para o modelo *Discreto*.

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 dependendo da adequada escolha do parâmetro I.

2.5 Modelo D-BIND

O modelo D-BIND (*Deterministic Bounding INterval Dependent*) [5, 10] 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

$$\frac{\mathcal{G}[t, t+I]}{I} \leq R_j(I), \qquad \forall t, I > 0. \tag{11}$$

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.

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

$$\mathcal{F}_{D-BIND}(\delta t) = \frac{R_k I_k - R_{k-1} I_{k-1}}{I_k - I_{k-1}}(\delta t - I_k) + R_k I_k, \quad I_{k-1} \leq \delta t \leq I_k. \tag{12}$$

A curva de restrição de tráfego é apresentada na figura 4. 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. O maior problema consiste no fato da escolha dos parâmetros não ser feita de forma automatizada.

2.6 Considerações

Os modelos determinísticos apresentados nesta seção foram propostos para caracterização de tráfego VBR em geral. Para a modelagem específica de uma importante classe de tráfegos, o vídeo no padrão MPEG, estes modelos deixam de capturar algumas propriedades intrínsecas a este padrão, o que poderia resultar em um modelo final com um número menor de parâmetros e o alto grau de acurácia desejado. Na seção seguinte será proposto um novo modelo, específico para tráfegos de vídeo comprimido no padrão MPEG,

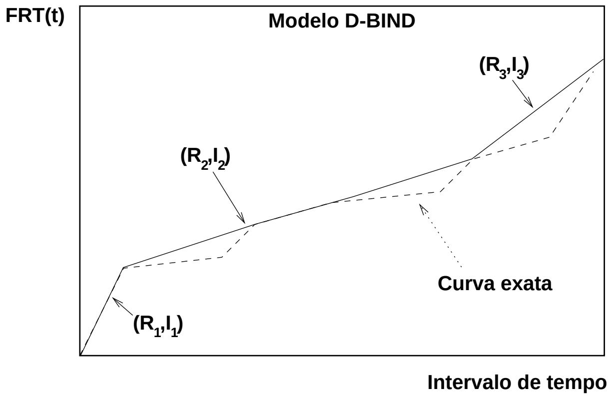


Figura 4: Função de restrição de tráfego para o modelo *D-BIND*.

denominado XGOP-B (*eXtended GOP-Based model*). Inicialmente, será feito 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 à proposição do modelo.

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 na seção anterior foram concebidos para modelagem de tráfegos em taxa variável de bits de uma forma geral.

Será proposto nesta seção um novo modelo específico para caracterização de fontes de vídeo MPEG: o modelo *XGOP-B*. Este modelo tem as vantagens de ser simples na extração dos parâmetros de tráfego e de obter os resultados desejados com um número reduzido de parâmetros, essencial para que os mecanismos de policiamento de tráfego e de controle de admissão de conexão demandem pouco processamento.

3.1 O Padrão MPEG

O algoritmo de codificação MPEG (*Moving Picture Expert Group*) [24] foi desenvolvido, primariamente, para armazenamento de vídeo comprimido em dispositivos digitais. Extensões foram introduzidas no algoritmo de forma a habilitar, além de outras funções, acesso aleatório e rápidas procuras em avanço e retrocesso (*forward/reverse*). Contudo, o padrão é suficientemente flexível para ser utilizado em um grande número de outras aplicações. 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.

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 [25].

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 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 5.

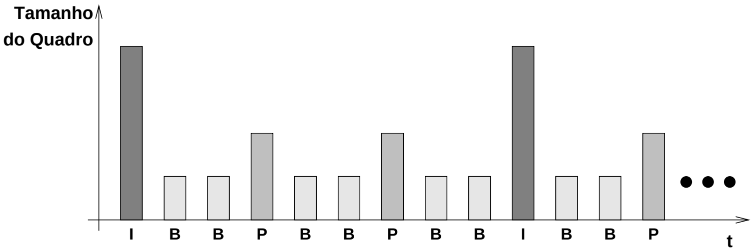


Figura 5: Amostra de um vídeo no padrão MPEG.

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 6, 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.

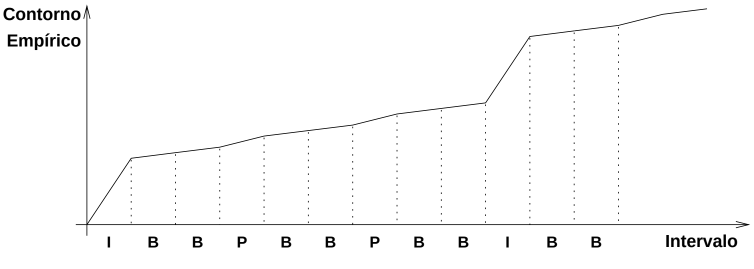


Figura 6: Função exata para vídeo MPEG.

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 7 ilustra dois traços de vídeo com características bem diversas, apesar de terem sido capturadas sob as mesmas condições (padrão do GOP, número de elementos por quadro,

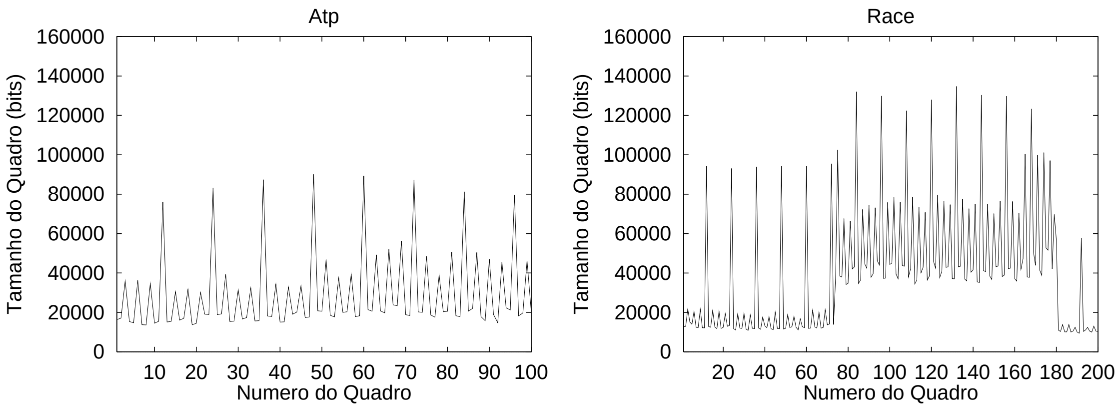


Figura 7: Exemplos de traços de vídeos.

taxa de amostragem etc). Um traço de vídeo é a seqüência ordenada dos tamanhos dos quadros da amostra de vídeo.

Apesar da propriedade acima analisada ser fundamental, ela não é 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.

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

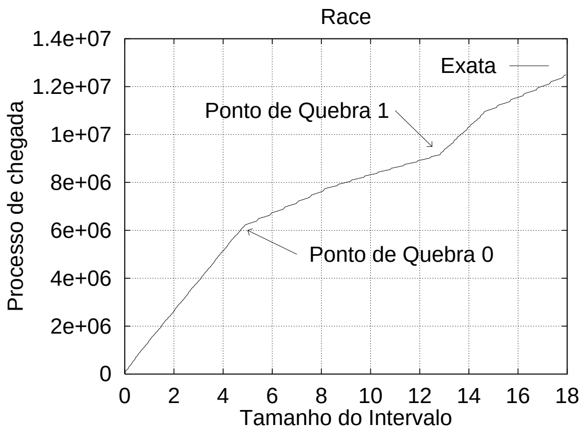


Figura 8: Efeito dos pontos de quebra sobre a função exata.

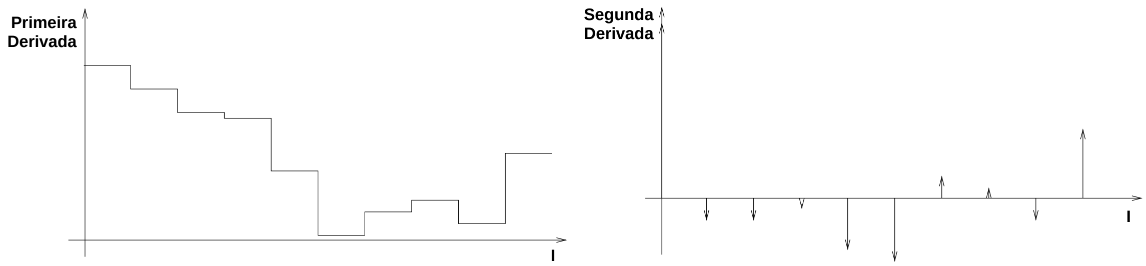
Para os modelos (σ, ρ) e *Discreto*, a aparição de pontos de quebra é extremamente indesejada. Uma das propriedades comuns a estes modelos é a constância na taxa média de crescimento da curva de restrição de tráfego. Com a ocorrência de um ponto de quebra negativo (como será visto, aquele que causa redução 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*, 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. Estas características são muito importantes pois a decisão de aceitação (ou não) de uma conexão deve ser rápida e o policiamento deve ser simples para atender altas velocidades.

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 na figura 8 foi feita de forma intuitiva. É preciso formular 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**.

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 é denominada **Segunda Derivada**. Os gráficos da figura 9 ilustram exemplos de primeira e segunda derivadas.



(a) (b)

Figura 9: Exemplo de funções de (a) Primeira e (b) Segunda Derivadas.

Os impulsos plotados no gráfico 9(b) 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 de um fator β, com $\beta > 0$, em relação ao valor da primeira inclinação desta seqüência.*

3.2 O Modelo XGOP-B

O modelo XGOP-B tem como fundamento a definição do modelo GOP-B (*GOP-Based model*) [8, 9]. O modelo GOP-B se baseia justamente na primeira propriedade analisada, que os vídeos MPEG apresentam, 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

$$\mathcal{F}_{GOP-B}(\delta t) = \lceil \frac{\delta t + T}{T \times D} \rceil \times \eta + \nu \times \delta t, \tag{13}$$

onde T é o inverso da taxa de geração de quadros por segundo e D é a distância entre quadros I. Por exemplo, se um determinado padrão de GOP for IBBPBB, a distância entre quadros I será 6.

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, o que certamente reduz sensivelmente o tempo de cálculo dos parâmetros.

O modelo *GOP-B* assume em sua estrutura uma das 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 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, como ilustra a figura 10.

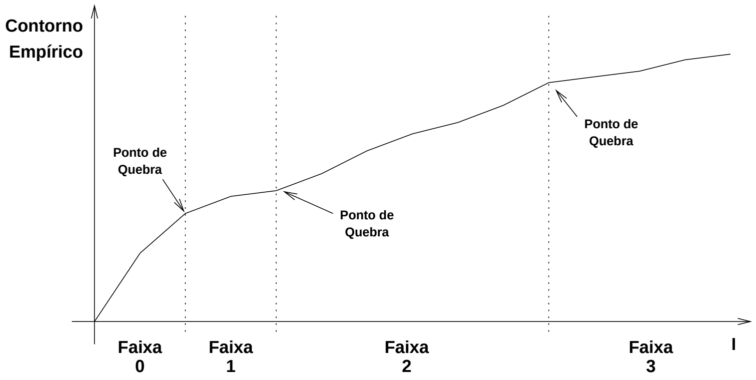
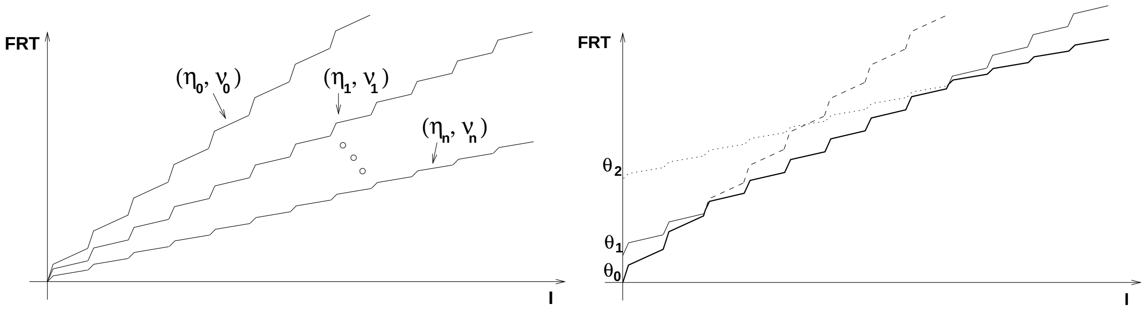


Figura 10: Classificação em faixas de uma função exata.

Como no modelo *GOP-B*, 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 no gráfico 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 dos *offsets* das curvas dadas por cada par (η_k, ν_k), como ilustra o gráfico 11(b), e o vetor $\vec{\xi}$ de pontos de quebra.



(a) (b)

Figura 11: Inclusão de θ como parâmetro de tráfego.

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

$$\mathcal{F}_{XGOP-B}(\delta t) = \theta_k + \lceil \frac{\delta t + T}{T \times D} \rceil \times \eta_k + \nu_k \times \delta t, \qquad \xi_{k-1} \leq \delta t < \xi_k. \tag{14}$$

A figura 12 ilustra o comportamento gráfico da função de restrição de tráfego para o modelo *XGOP-B*.

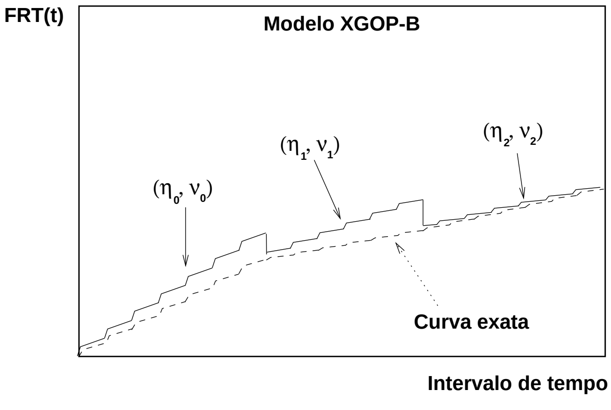


Figura 12: Função de restrição de tráfego para o modelo *XGOP-B*.

4 Análise dos Resultados

Nesta seção 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 [26]. As seqüências foram obtidas de filmes, eventos esportivos e shows apresentados na TV. As amostras possuem as seguintes características em comum:

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

A medida de comparação dos modelos utiliza o conceito de **Perda por Aproximação do Modelo** (PAM) [7], 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

$$PAM = \left(\frac{S_{\mathcal{F}}}{S_{\zeta}} - 1 \right) \times 100\%, \tag{15}$$

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

A tabela 1 mostra os índices de PAM para as amostras de vídeo exploradas. Para o modelo *Discreto* foram realizados os cálculos com dois valores diferentes para o parâmetro I, 540ms e 1080ms, que equivalem, respectivamente, a uma e duas vezes o tamanho de um GOP. Para o modelo *XGOP-B*, fixou-se o valor do fator de qualidade do modelo em $\beta = 0.1$. Mais adiante será feita uma análise do efeito desta variável no nível de acurácia do modelo.

Os mesmos resultados apresentados na tabela 1 estão plotados na figura 13. É notável a grande diferença entre os níveis de acurácia dos modelos. Entretanto, deve ser levado em conta que existe um compromisso entre o resultado obtido e o número de parâmetros necessários para se alcançar este valor.

Na figura 14 são apresentadas as curvas de restrição de tráfego para dois traços de vídeo, um extraído do desenho animado *Asterix* e outro oriundo de um noticiário de TV.

Alguns dos resultados obtidos podem não ser satisfatórios. O procedimento a se adotar consiste em se ajustar o valor de β até que os resultados desejados sejam alcançados. Quanto menor for o valor de β, possivelmente mais pontos de quebra serão identificados e, conseqüentemente, mais parâmetros de tráfego poderão ser obtidos, resultando em maior acurácia de parametrização.

Para os mesmos traços de vídeo, a figura 15 ilustra os efeitos do ajuste de β sobre o número de pontos de quebra e sobre o nível de acurácia do modelo.

Tabela 1: Perda por Aproximação do Modelo para cada um dos modelos.

Vídeo	**Perda por Aproximação do Modelo (%)**						
	(σ, ρ)	$(\vec{\sigma}, \vec{\rho})$	Discreto I=540ms	Discreto I=1080ms	D-BIND	GOP-B	XGOP-B
1. asterix	23.21	3.96	49.21	54.27	0.32	44.83	1.76
2. atp	9.31	1.02	17.76	10.07	1.02	15.74	6.17
3. bond	44.02	2.16	31.75	26.24	1.48	32.61	4.09
4. dino	15.30	2.17	48.78	44.92	1.71	59.22	2.34
5. lambs	30.56	3.42	20.31	23.46	2.23	18.36	4.77
6. mbean	17.69	1.26	33.40	34.25	1.00	27.75	1.97
7. movie2	6.96	1.55	76.96	60.40	1.49	78.51	2.25
8. mtv1	13.42	1.43	75.44	67.47	1.43	76.55	1.99
9. mtv2	9.20	1.77	30.14	36.27	1.32	24.79	1.71
10. news1	23.35	2.43	42.85	45.38	1.07	46.70	2.51
11. news2	7.44	2.56	33.67	38.29	2.47	37.94	2.31
12. race	48.95	0.64	24.35	22.03	0.60	18.71	6.96
13. sbowl	4.51	1.03	19.50	22.54	0.96	18.15	3.69
14. simpsons	7.91	1.03	55.76	44.29	0.95	55.62	1.54
15. soccer1	4.62	2.90	28.30	28.97	1.10	27.07	2.75
16. soccer2	7.21	2.20	39.58	29.19	1.53	42.11	3.34
17. star	18.19	1.89	22.59	26.18	1.88	18.90	2.92
18. talk1	14.92	2.99	17.52	22.73	1.15	14.59	2.62
19. talk2	3.09	2.31	25.28	23.20	2.05	21.64	8.47
20. terminator	9.26	2.59	41.86	35.41	1.17	37.62	2.56
21. video	6.79	5.42	13.49	16.07	5.18	11.62	3.36

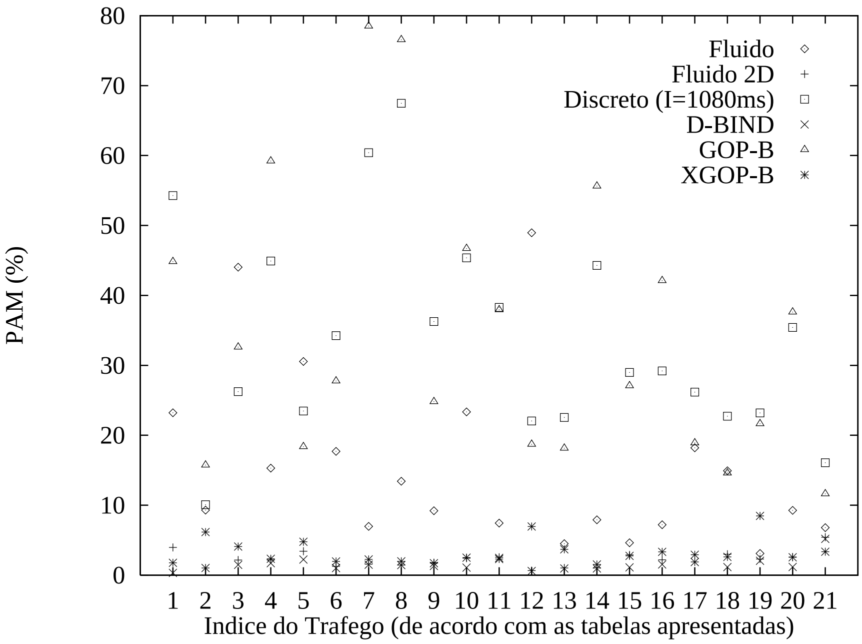


Figura 13: Gráfico de Perda por Aproximação do Modelo.

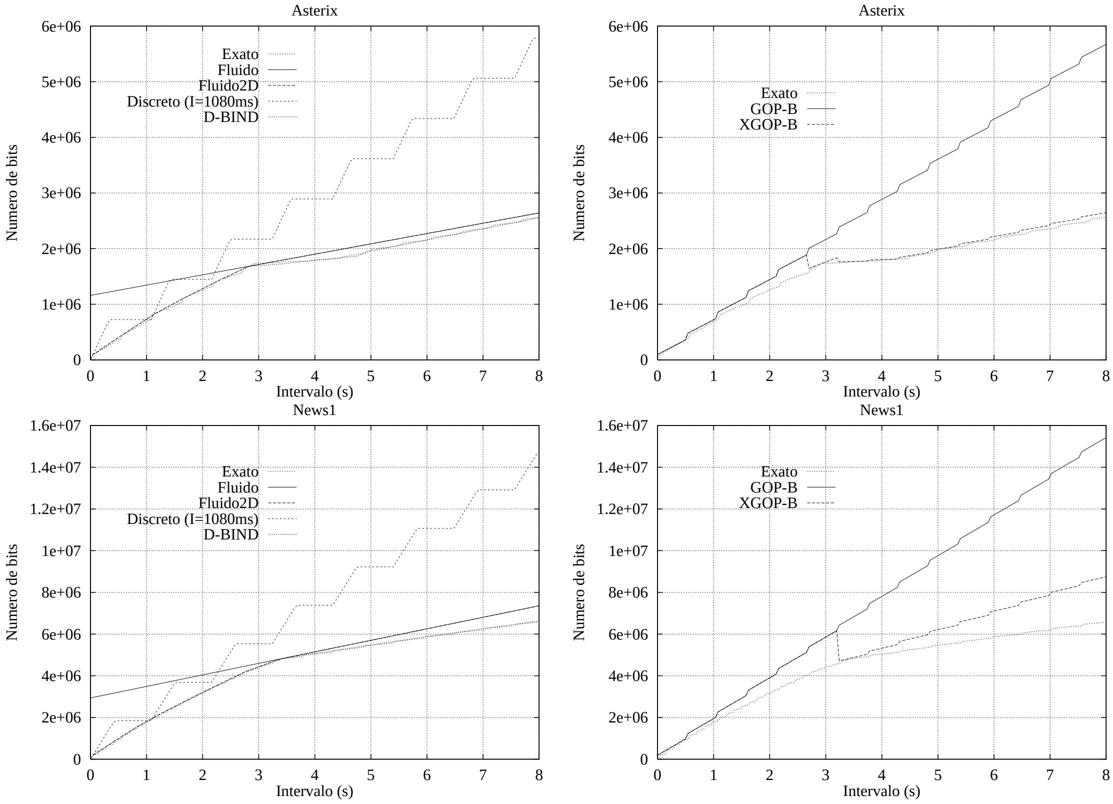


Figura 14: Curvas de restrição de tráfego para dois traços de vídeo.

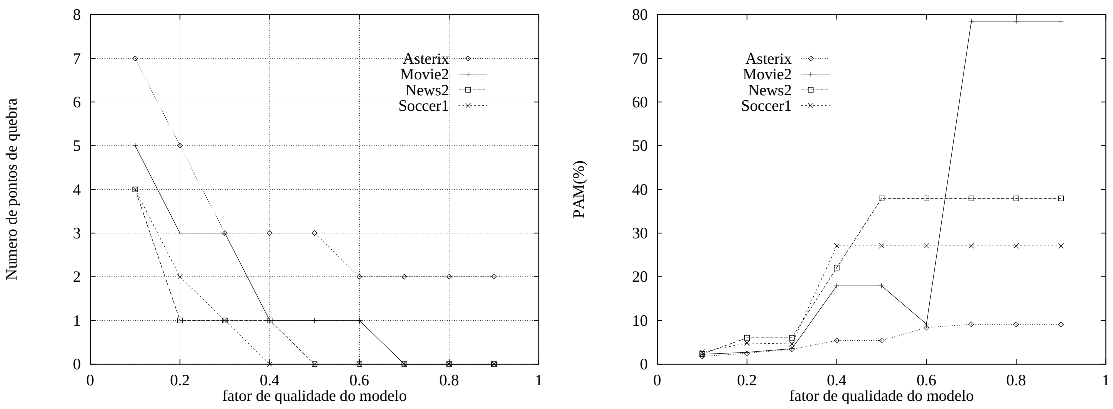


Figura 15: Relações '$\beta \times$ número de pontos de quebra' e '$\beta \times$ PAM'.

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 determinísticas 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 conexão. 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 cara 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.

Neste trabalho foi apresentado um estudo sobre os principais modelos determinísticos de tráfegos. Estes modelos, por terem sido propostos para caracterização de tráfegos VBR em geral, deixam de capturar algumas características intrínsecas a uma importante classe de tráfegos: o vídeo comprimido no padrão MPEG. Desta forma, foi descrita uma análise em dois níveis sobre o comportamento das fontes de tráfego de vídeo MPEG: no nível microscópico, pela análise sobre a geração de cada quadro, e no nível macroscópico, pela análise sobre cada unidade de grupo de quadros (GOP).

Todo este processo resultou na proposta do modelo XGOP-B para caracterização determinística de tráfegos de vídeo MPEG. Os resultados mostraram que altos índices de acurácia de parametrização podem ser obtidos, inclusive podendo ser melhorados pelo ajuste do parâmetro β, chamado fator de qualidade do modelo.

O mais importante, entretanto, consiste no fato do modelo XGOP-B ter seus parâmetros calculados de forma automática, sem intervenção externa, ao contrário dos modelos inicialmente analisados. Esta característica é extremamente importante, uma vez que uma escolha mal dimensionada pode resultar em perda de acurácia de parametrização.

Agradecimentos

Este trabalho recebeu apoio financeiro da UFRJ, FUJB, PROTEM-CC, CNPq e CAPES.

Referências

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] E. Knightly e H. Zhang, "Traffic characterization and switch utilization using a deterministic bounding interval dependent traffic model", in *IEEE INFOCOM'95*, Boston, MA, pp. 1137–1145, abril de 1995.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

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

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

[18] D. Saha, S. Mukherjee e S. K. Tripathi, "Multi-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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

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

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

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

[26] 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.