Caracterização e Modelagem de Tráfego Agregado em Redes ATM

Michel Ferreira Abdalla

Setembro de 1994

Resumo

A Rede Digital de Serviços Integrados - Banda Larga (RDSI-BL), baseada no Modo de Transferência Assíncrona (ATM), é capaz de suportar vários serviços de comunicação, como voz, dados e vídeo, cada um possuindo diferentes características de tráfego. A modelagem acurada das fontes de tráfego se faz necessária para uma melhor avaliação de desempenho desta rede. Neste artigo são analisados vários modelos para a caracterização do comportamento do tráfego agregado de várias fontes, sendo descritos modelos para fontes de dados, voz e vídeo. O objetivo principal deste artigo é realizar uma dissertação sobre as diversas modelagens propostas para cada tipo de fonte, analisando suas características e apresentado suas vantagens e desvantagens.

1 Introdução

A necessidade de atender a uma grande variedade de tráfegos e os mais diversos serviços e requisitos de desempenho de uma forma simples conduziu a escolha do Modo de Transferência Assíncrono (ATM) para a implementação das Redes Digital de Serviços Integrados - Banda Larga (RDSI-BL). Estes serviços incluem desde uma simples transferência de dados até serviços como teleconferência, acesso a banco de imagens e interconexão entre redes locais de alta velocidade. Cada tipo de tráfego possui sua própria característica e impõe diferentes requisitos de desempenho, tais como máximo atraso e máxima variação do atraso (Jitter delay) de uma célula, máxima probabilidade de perda e máxima variação da probabilidade de perda de células.

Como a flexibilidade do ATM na alocação dinâmica da banda passante é devido a multiplexação estatística nos seus nós de acesso e de transporte, um grande interesse surgiu na pesquisa da análise de desempenho de um multiplexador estatístico com entradas consistindo na superposição de vários tipos de tráfegos. O interesse mais geral está em calcular parâmetros tais como probabilidade de perda de células e atraso médio das células em função dos descritores de tráfego tais como taxa máxima de transmissão, taxa média de transmissão e tamanho da rajada.

Neste artigo são considerados três tipos de fonte de tráfego: fontes de voz, fontes de dados assíncronos e por último, fontes de vídeo. Na seção 2 é estudada a caracterização de cada um dos tipos de fontes. Na seção 3 são apresentados e analisados vários modelos, para cada um dos tipo de fonte, tais como o Modelo de Fluxo Contínuo [1, 2], o Processo Semi-Markoviano (SMP) [1], as Cadeias de Markov de Tempo Contínuo (CMTC) [1, 3, 4, 5], o Processo de Poisson Modulado por Markov (MMPP) [3, 4, 5, 6, 7] e o Modelo Auto-Regressivo (AR) [4, 5]. Na seção 4 são apresentadas algumas conclusões.

2 Caracterização do Tráfego

Nas redes ATM é esperado suportar uma grande diversidade de conjuntos de aplicações com largas variações nas suas características. Infelizmente, muitas destas aplicações não possuem medições suficientes que permitam satisfatoriamente a sua caracterização de um modo acurado.

As características de fontes de voz vem sendo estudadas por vários anos no contexto de redes telefônicas e são bem conhecidas. As características das fontes de dados, apesar das redes de pacotes já existirem há três décadas, ainda não são bem conhecidas. As características das fontes de vídeo de tráfego constante (fontes CBR) já foram bem estudadas, mas a grande dificuldade está na estima da banda passante necessária para a requerida qualidade de serviço. As fontes de vídeo de tráfego variável (fontes VBR) e de imagens ainda foram pouco estudadas e sua caracterização é muito restrita.

2.1 Voz

O comportamento típico de uma fonte única de voz é mostrado na figura 1 [8]. Nela, pode ser observada que existem dois períodos característicos: o Ativo e o Inativo. A fonte de voz está num período ativo nos momentos de fala. Durante estes períodos, a fonte de voz gera pacotes de tamanho fixo em intervalos regulares. A fonte está num período inativo quando há silêncio. Nestes períodos nenhum pacote é gerado. Em uma conversa normal, os duração dos períodos ativos podem ser aproximados por uma distribuição exponencial, enquanto que a duração dos períodos inativos, apesar de não perfeitamente, aproximam-se de uma distribuição exponencial. O período inativo poderia também ser modelado por um distribuição que seria a mistura do silêncio devido a pausas e do silêncio devido a escuta.

Ao se aproximar o processo de chegada de uma fonte única de voz por uma distribuição exponencial, que é um processo sem memória, significa dizer que o processo de chegada é um processo de renovação, e portanto, não existe correlação entre chegadas sucessivas. Porém vale lembrar que o mesmo é altamente variável (tráfego em rajadas). Como consequência, o tráfego agregado de várias fontes de voz não é um processo de renovação, visto que a taxa de chegada instantânea do tráfego agregado varia de acordo com o número de fontes ativas naquele instante.

Estudando-se agora as propriedades estatísticas do processo de voz, apresentadas

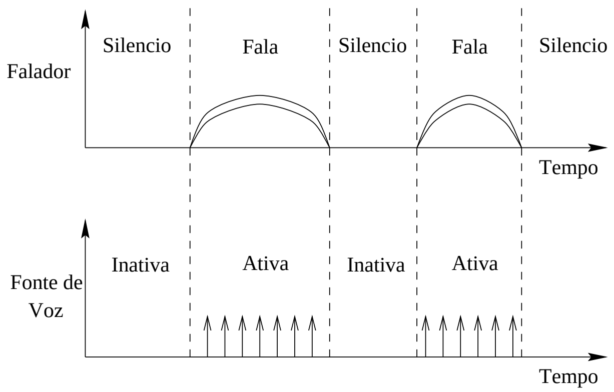

Figura 1: Comportamento típico de uma fonte de voz

em [3], tem-se que a distribuição do tempos entre chegadas é dada por

$$F(t) = \left[(1-\alpha T) + \alpha T \left(1 - e^{-\beta(t-T)}\right)\right] U(T-t) \tag{1}$$

(onde $U(t)$ é a função degrau unitário) e a transformada de Laplace-Stieltjes (LST) é dada por

$$\bar{f}(s) = \int_0^\infty e^{-st} dF(t) = \left[1 - \alpha T + \alpha T \beta/(s+\beta)\right] e^{-st} \tag{2}$$

com a taxa média de chegada de pacotes de uma fonte única dada por

$$\lambda = -1/\bar{f}(0) = 1/\left(T + \alpha T/\beta\right) \tag{3}$$

Esta aproximação corresponde a uma distribuição geométrica do número de pacotes de voz (com média $1/\alpha T$ durante o período de fala, que possui distribuição exponencial com média $1/\alpha$ seguida de uma distribuição exponencial do período de silêncio com média $1/\beta$ Como cada processo de voz é um processo de renovação, os resultados da teoria da renovação podem ser usados para o cálculos dos momentos do número de chegadas num intervalo, que serão usados posteriormente. Assim, seja $N(0,t)$ o número de chegadas de um processo de renovação estacionário no intervalo $(0,t)$ e seja

$$M_r(t) = E\left[N^r(0,t)\right] \tag{4}$$

e seja

$$\bar{M}_r(s) = L\left[M^r(t)\right] \tag{5}$$

onde $L(.)$ denota a transformada de Laplace. Sendo assim, através de (2) e (5), é possível a obtenção de todos os momentos desejados. Com estas equações também pode ser obtido o índice de dispersão para contadores, $I(t)$, definido em [1], que satisfaz

$$\lim_{t\to\infty} I(t) = \lim_{t\to\infty} \frac{\mathrm{var}\,(N(0,t))}{M_r(t)} = \frac{\mathrm{var}(X)}{E^2(X)}\ . \tag{6}$$

Agora considerando a superposição de idênticos n processos independentes de pacotes de voz, denota-se por $N_i(0,t)$ o número de chegadas em $(0,t)$ da fonte i. O número de chegadas do tráfego agregado e seus momentos são dados por

$$N^S(0,t) = \sum_{i=1}^{n} N_i(0,t)\ ; \tag{7}$$

$$M_1^S(0,t) = E\left[N^S(0,t)\right] = nM_1(t)\ ; \tag{8}$$

onde o sobrescrito S denota a superposição.

O índice de dispersão do tráfego agregado satisfaz

$$\frac{\mathrm{var}\left[N^S(0,t)\right]}{E\left[N^S(0,t)\right]} = \frac{\mathrm{var}\left[N(0,t)\right]}{E\left[N(0,t)\right]}\ . \tag{9}$$

O terceiro momento para o processo de superposição será

$$\mu_3^{*S}(0,t) = E\left\{\left[N^S(0,t)\right] - E\left[N^S(0,t)\right]^3\right\}, \tag{10}$$

que se reduz para

$$\mu_3^{*S}(0,t) = n\left[M_3(t) - 3M_2(t)M_1(t) + 2M_1^3(t)\right], \tag{11}$$

onde $M_2(t)$ e $M_3(t)$ são obtidos através da transformada inversa de $\overline{M}_2(s)$ e $\overline{M}_3(s)$.

2.2 Dados

Como mencionado anteriormente, o comportamento de uma fonte de dados não foi profundamente estudado, porém para efeitos práticos, é boa a aproximação da geração de dados de uma fonte simples por um processo de chegada de Poisson (para o caso contínuo) ou por um processo geométrico do tempo entre chegadas (para o caso discreto) [5]. Em um caso interativo, uma pacote único pode ser gerado em um instante. Para grandes volumes de dados, como numa transferência de arquivo, um grande número de pacotes pode ser gerado em um instante.

Nas atuais redes de pacotes, os tamanhos dos mesmos podem variar ou não. Entretanto, em redes ATM, a célula tem um tamanho fixo e relativamente pequeno se comparado com o tamanho de um pacote de dado. Podem, com isso, ser geradas várias células para um único pacote de dado.

2.3 Vídeo

Com o desenvolvimento das redes de comunicação, um grande número de aplicações irá utilizar serviços de vídeo. Torna-se, portanto, muito importante a caracterização de suas propriedades para uma concreta predição de seus parâmetros, tais como o atraso devido a multiplexação, o tamanho do *buffer* requerido para a multiplexação e a largura de banda necessária para a transmissão do vídeo.

Sinais de vídeo requerem um grande banda passante. Por exemplo, em aplicações de TV a resolução de quadro de 512 x 512 é transmitido a cada 1/30s, gerando aproximadamente 63 Mbps, se uma esquema PCM de codificação for usado. Entretanto, devido ao fato de que a variação entre quadros sucessivos é pequena, normalmente as fontes de vídeo são comprimidas utilizando-se esquemas de codificação de taxa variável que codificam somente as diferenças significativas entre estes quadros sucessivos. Isto introduz uma grande correlação entre as chegadas de células de quadros sucessivos. De um modo geral, pode-se classificar as fontes de variação nas taxas de bites em mudanças de cenas, movimentos dentro de uma cena e mudanças na sutileza dos detalhes das imagens bidimensionais do vídeo. A tabela 1, descrita em [9], classifica estas variações nas taxas de bites em três classes baseado na escala de tempo.

As variações de longo-intervalo surgem principalmente das mudanças de contexto no vídeo devido a mudanças de cenas. A mudança na taxa de bites muda como um degrau, e as características da variação das taxas de bites seguintes a mudança de cena são completamente diferentes daquelas antes das mudança. A escala de tempo destas mudanças é da ordem de vários segundos.

As variações de curto-intervalo são causadas principalmente pela mudança de contexto da imagem dentro de uma única cena. Estas variações de taxas de bites mudam suavemente e são correlacionadas temporalmente. Estas correlações decaem exponencialmente com o tempo.

As variações intra-quadro surgem principalmente devido ao processamento de blocos dentro de um quadro. Porém, como a maioria dos esquemas de codificação utilizam *buffers* com tamanho suficiente para absorver estas variações, estas se tornam imperceptíveis externamente. A necessidade de se tratar estas variações somente surgirão no caso em que os dados forem entregues diretamente a rede, sem a passagem pelo *buffer*.

Como o principal objetivo é o estudo da multiplexação de várias fontes de vídeo e dos atrasos que poderão surgir, é criado um modelo para este sistema. O modelo utilizado está representado na figura 2 [5]. Neste modelo, N fontes independentes de vídeo de taxas de bites variáveis são suavizadas em pre*buffers* individuais e acumuladas num *buffer* principal (PAD - montador e desmontador de pacotes). A saída destes *buffers* é multiplexada em unidades de células ou pacotes de tamanho fixo. Os pacotes se acumulam num *buffer* comum e são entregues a um canal de alta velocidade segundo uma ordenação FIFO (Primeiro a chegar, Primeiro a sair).

Tabela 1: Classificação das variações das taxas de bites

Tipo	Escala de Tempo	Causas	Características
Variação entre Quadros			
Longo-intervalo	Vários segundos ou mais	Mudanças de cenas	Variação descontínua, diferindo as características estatísticas antes e depois da mudança
Curto-intervalo	Entre 1 período de quadro e alguns segundos	Movimento de pessoas, de câmera e variação de padrões	Variações suaves com correlação temporal, com grandes variações ocasionais devido a movimentação de pessoas e da câmera
Variação Intra-Quadro	Menos que 1 período de quadro	Variações espaciais das características dentro de uma imagem	Variações que tem uma periodicidade devido a escanização de imagens ou processamento de blocos

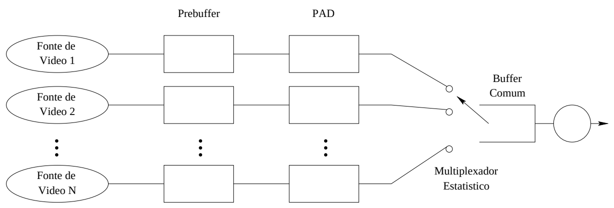

Figura 2: Modelo do multiplexador estatístico de fontes de vídeo

3 Modelagem

3.1 Voz

Um processo de chegada de uma fonte de voz é muito complexo devido a forte correlação existente entre as chegadas. Nesta seção são descritos alguns modelos para análise de uma fonte única de voz e para o seu tráfego agregado. Na seção 3.1.1 é descrito o modelo UAS. Na seção 3.1.2 é descrito o modelo de um processo semi-Markoviano. Na seção 3.1.3 é descrito o modelo das cadeias de Markov de tempo contínuo. Na seção 3.1.4 é descrito o modelo dos processos de Poisson modulados por Markov.

3.1.1 UAS

O UAS [1, 2] é um modelo de fluxo contínuo para o caso de fontes homogêneas. Ele assume que os processos de chegada e de remoção da informação na fila são feitos de modo uniforme.

Cada fonte de voz é modelada como possuindo dois estados, ON e OFF, conforme ilustrado na figura 3. Nele é suposto que os períodos ON e OFF são exponencialmente distribuídos com médias $1/\alpha$ e $1/\beta$ respectivamente. O estado ON corresponde a fase em que a fonte está ativa. Nesta fase, pacotes de voz são gerados regularmente a cada $1/V$ s. O estado OFF corresponde a fase em que a fonte está inativa, na qual nenhum pacote de voz é gerado. Deste modo, o número de fontes ativas do sistema em função do tempo pode ser modelada como uma cadeia de Markov de tempo contínuo como mostrado na figura 4, onde N é o número de fontes de voz no sistema. Este processo é referenciado como processo de fase, sendo denotado por $\wp$.

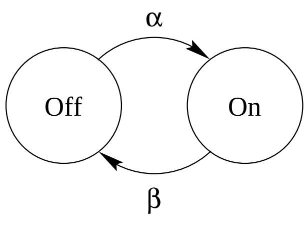

Figura 3: Modelo ON/OFF de uma fonte de voz

Sendo C a capacidade normalizada do canal em relação ao tráfego que uma fonte gera quando se encontra no estado ativo (ou seja, cada pacote de voz é removido pelo servidor a cada $1/VC$ s), e considerando que existem i fontes ativas ($\wp(t) = i$), o comportamento da fila é dado por:

- Se $i < C$, a fila decresce à taxa (C - i)
- Se $i = C$, a fila não se altera
- Se $i > C$, a fila cresce à taxa (C - i).

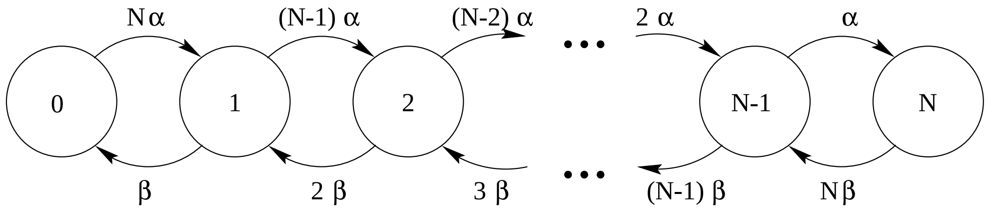

Figura 4: Modelo do processo de fases

As aproximações adotadas no modelo ignoram a existência de variações muito rápidas no conteúdo do *buffer* se comparados com o sistema real. Isto é, no sistema real, a informação não entra num *buffer* de transmissão, não podendo ser transmitida até que uma fonte ativa particular complete a geração de um pacote. Já no UAS, é possível para a informação ser transmitida enquanto está sendo recebida.

De volta a análise matemática, seja B(t) o conteúdo do *buffer* e $\wp(t)$, o número de fontes ativas no instante t. Seja $P_i(t,b) = P\{\wp(t) = i, B(t) \le b\}$ para $0 \le i \le N$, $t \ge 0$, $b \ge 0$. Então, sendo $F_i(b) = \lim_{t\to\infty} P_i(t,b)$, então o conjunto de equações diferenciais simples que descrevem o sistema será

$$\begin{aligned} (i-C)\frac{dF_i(b)}{db} &= (N-i+1)\frac{\alpha}{\beta}F_{i-1}(b) - \left\{(N-1)\frac{\alpha}{\beta}+i\right\}F_i(b) + \\ & \quad (i+1)F_{i+1}(b)\ ; \qquad 0 \le i \le N \end{aligned} \tag{12}$$

ou, equivalentemente

$$D\frac{d}{db}F(b) = MF(b)\ ; \tag{13}$$

onde

$$D = \{-C, 1-C, 2-C, \cdots, N-C\}\ ;$$

$$M = \begin{bmatrix} -n\alpha & \beta & \cdots & 0 & 0 \\ n\alpha & -n\{(n-1)\alpha+\beta\} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -\{\alpha+(n-1)\beta\} & n\beta \\ 0 & 0 & \cdots & \alpha & -n\beta \end{bmatrix} ;$$

Sendo $G(b)$ a probabilidade de que o tamanho do *buffer* seja maior que b, tem-se :

$$G(b) \triangleq Pr(\text{tamanho} > b) = 1 - 1'F(b)\ ; \qquad b \ge 0\ . \tag{14}$$

Observe também que

$$F_i(\infty) = \frac{1}{(\alpha+\beta)^N}\binom{N}{i}\alpha^i\ ; \qquad 0 \le i \le N\ ,$$

onde $F_i(\infty)$ é a probabilidade de i fontes em N estarem ativas.

É importante observar aqui que C é considerado não inteiro (Caso fosse inteiro, uma das equações diferenciais em (12) se degeneraria para uma equação algébrica que poderia ser usada para eliminar uma das componentes desconhecidas de F).

Seja z um auto-valor de $D^{-1}M$ e ϕ, o seu auto-vetor correspondente. Em Anick et al. [9], o conjunto dos auto-valores é obtido através da solução do conjunto de equações de 2^o grau em (15).

$$A(k)z^2 + B(k)z + C(k) = 0 \ ; \qquad k = 0, 1, \cdots, N \ ; \tag{15}$$

onde

$$\begin{aligned}
A(k) &\triangleq (N/2 - k)^2 + (N/2 - C)^2 \ , \\
B(k) &\triangleq 2(\beta - \alpha)(N/2 - k)^2 - N(\beta + \alpha)(N/2 - C) \ , \\
C(k) &\triangleq (\alpha + \beta)^2 \left\{ (N/2)^2 - (N/2 - k)^2 \right\} \ .
\end{aligned}$$

Dado um auto-valor z, o coeficiente ϕ_i do auto-vetor associado ϕ é

$$\phi_i = (-1)^{N-i} \sum_{j=0}^{k} \binom{k}{j} \binom{N-k}{i+j} r_1^{k-j} r_2^{N-k-i+j} \ ; \qquad 0 \le i \le N \ ,$$

onde

$$\begin{aligned}
r_1 &= \left\{ -(z + \beta - \alpha) + \sqrt{(z + \beta - \alpha)^2 + 4\alpha} \right\} / 2\alpha \\
r_2 &= \left\{ -(z + \beta - \alpha) - \sqrt{(z + \beta - \alpha)^2 + 4\alpha} \right\} / 2\alpha \\
k &= \frac{z\beta c - N\alpha + N\alpha r_1}{\alpha(r_1 - r_2)} \ .
\end{aligned}$$

E finalmente tem-se

$$\begin{aligned}
F(b) &= F(\infty) + \sum_{i=0}^{N - \lfloor c \rfloor - 1} exp(z_i b) a_i \phi_i \quad \text{e} \\
G(b) &= - \sum_{i=0}^{N - \lfloor c \rfloor - 1} exp(z_i b) a_i \, (1' \phi_i)
\end{aligned}$$

onde

$$a_j = \left(\frac{\alpha}{\alpha + \beta} \right)^N \prod_{\substack{i=0 \\ j \ne i}}^{N - \lfloor c \rfloor - 1} \frac{z_i}{z_i - z_j} \ ; \qquad 0 \le j \le N - \lfloor c \rfloor - 1 \ .$$

3.1.2 Processos Semi-Markovianos (SMP)

No modelo de processos semi-Markovianos [1], o processo de fases $\wp$ também é utilizado. Neste modelo, as seguintes aproximações são utilizadas, onde $\wp$ denota o número de fontes ativas e C a capacidade do canal.

- Quando $\wp = j = C$, o tamanho da fila não se modifica.
- Quando $\wp = j < C$, então, sempre que a fila não estiver vazia, o tamanho da fila decrescerá de um pacote a cada $1/V(C-j)$ s, a partir do instante em que o processo entrar no fase j.
- Quando $\wp = j > C$, então o tamanho da fila crescerá de um pacote a cada $1/V(C-j)$ s, a partir do instante em que o processo entrar no fase j.

Uma implicação desta aproximação está no fato de que o modelo ignora variações rápidas. Por exemplo, no sistema real sempre que uma fonte de voz gera um pacote durante a transmissão de um outro pacote, o tamanho da fila crescerá de um pacote independente da fase em que o processo de fases estiver, o que não é descrito neste modelo.

Neste modelo, o estado do processo, ilustrado na figura 5 é denotado por $(l(t0, \wp(t))$, onde $l(t)$ é o número de pacotes na fila num instante t e $\wp(t)$ é o número de fontes ativas, ou fase do processo, neste mesmo instante. Como a resincronização do processo ocorre nos instantes de transição de fases, a conclusão é que existe uma cadeia de Markov embutidas nestes instantes de mudança de estado das fases. Logo o tempo de permanência provisória de qualquer estado depende somente do próprio estado. Então o processo descrito na figura 5 é um processo semi-Markoviano.

Seja $P_{i,j} = \lim_{t\to\infty} P\{l(t) = i, \wp(t) = j\}$. Seja $q_{i,j}$ a probabilidade de equilíbrio da cadeia de Markov embutida, ou seja, a fração de equilíbrio das transições que entram no estado (i, j) do processo. E seja $m_{i,j}$ o valor esperado do tempo de permanência provisória no estado (i, j), então

$$p_{i,j} = \frac{q_{i,j} m_{i,j}}{\sum_{k=0}^{\infty} \sum_{l=0}^{N} q_{k,l} m_{k,l}} \; ; \qquad 0 \le j \le N \; . \tag{16}$$

O cálculo dos $q_{i,j}$'s pode ser feito utilizando técnicas de geometria das matrizes, e respectivamente, através de (16), podem ser calculados os valores de cada $p_{i,j}$. Assim, a probabilidade π_i do tamanho da fila ser igual a i será

$$\pi_i = P(l = i) = \sum_{j=0}^{N} p_{i,j} \; ; \tag{17}$$

onde N é o número total de fontes de voz.

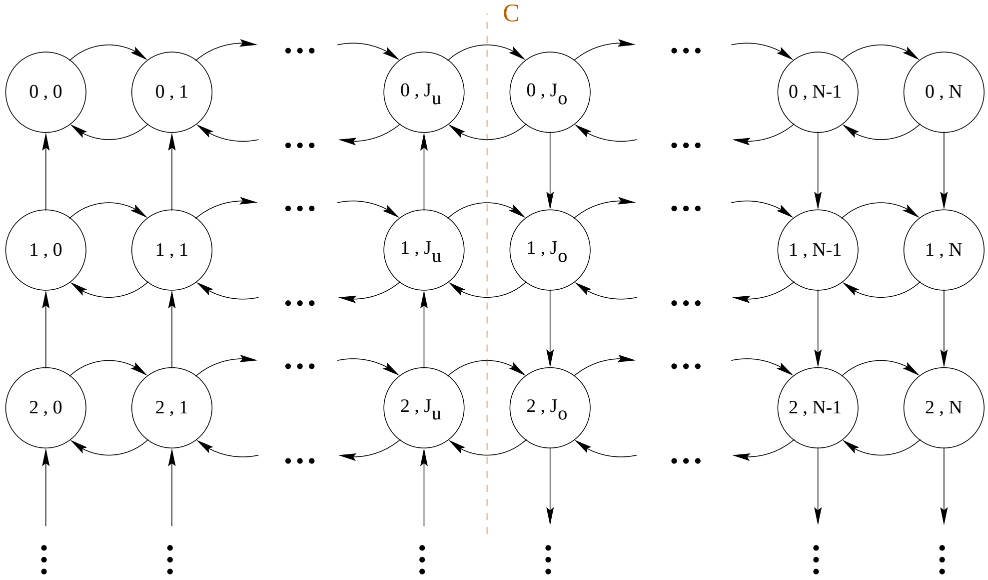

Figura 5: Diagrama de estados para o modelo do processo semi-Markoviano

3.1.3 Cadeias de Markov de Tempo Contínuo (CMTC)

No modelo de CMTC [1], como nos modelos anteriores, cada fonte de voz se torna ativa e inativa de acordo com um processo de renovação alternante em que as durações dos períodos ativo e inativo são exponencialmente distribuídos com parâmetros α e β, respectivamente. Entretanto, a geração de pacotes por uma fonte ativa ocorre de acordo com um processo de Poisson com taxa λ, ao invés de uma taxa constante. Como a superposição de processos de Poisson também um processo de Poisson com taxa igual a soma das taxas de cada processo, segue que quando o processo de fases estiver na fase j, então o processo de chegadas de pacotes ao sistema é Poisson com taxa igual a $j\lambda$. Similarmente o processo de serviço é assumido como exponencialmente distribuído com parâmetro υ. Neste caso, a capacidade do canal, C, correspondente ao número de fontes que saturariam o servidor, será igual a υ/λ.

O diagrama de estados deste modelo é dado na figura 6 onde o estado denotado por $(s, \wp)$ indica que existem s pacotes no sistema e que $\wp$ fontes estão ativas. O vetor $p_{i,j}$ que denota a probabilidade de equilíbrio do sistema estar no estado (i, j), pode ser calculado através das equações de equilíbrio, facilmente obtidas da figura 6. E, finalmente, a probabilidade do tamanho da fila ser igual a i, π_i, será

$$\pi_i = P(s = i + 1) = \sum_{j=0}^{N} p_{i+1,j} \ .$$

3.1.4 Processo de Poisson modulado por Markov (MMPP)

Neste modelo, o tráfego agregado dos processos de chegada é aproximado por um processo de stream mais simples, correlacionado e de não-renovação. O stream aproximado é

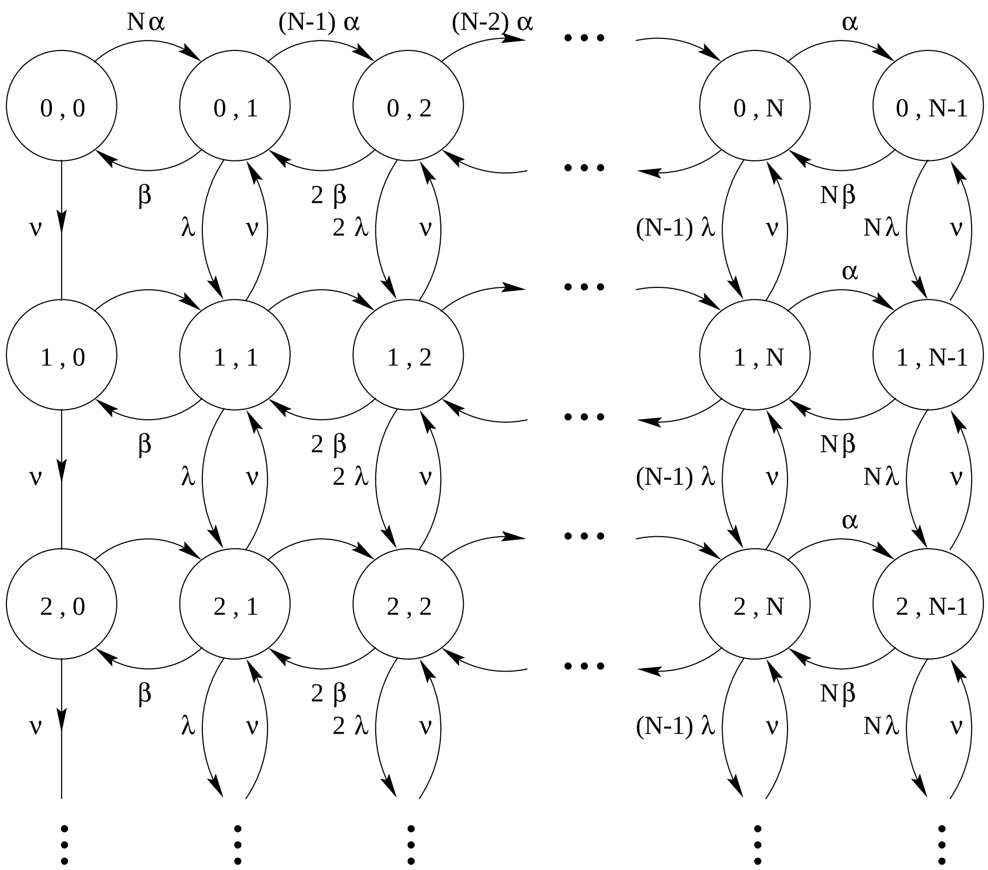

Figura 6: Diagrama de estados para o modelo da cadeia de Markov de tempo contínuo

escolhido de modo que as características estatísticas combinem exatamente com aquelas da superposição original. Na escolha do stream aproximado, procura-se obter um tratamento analítico simplificado, ao mesmo tempo que se deseja versatilidade. O modelo aqui apresentado é o MMPP de m estados [3, 4, 5, 6, 7], que é um processo de Poisson duplamente estocástico, onde a taxa de chegada do processo de Poisson varia de acordo com o estado de uma cadeia de Markov de m estados, independente do processo de chegada. A figura 7 ilustra este procedimento. O MMPP é parametrizado por uma cadeia de Markov de tempo contínuo de m estados com gerador infinitesimal Q [6] e as m taxas de chegada de Poisson $\lambda_1, \lambda_2, \cdots, \lambda_m$, usando-se a notação

$$Q = \begin{bmatrix} -r_1 & r_{12} & \cdots & r_{1m} \\ r_{21} & -r_2 & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & -r_m \end{bmatrix} , \qquad r_i = \sum_{\substack{j=1 \\ j \neq i}}^{m} r_{i,j} \ ,$$

$$\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_m)^T \qquad \text{e} \qquad \lambda = (\lambda_1, \lambda_2, \cdots, \lambda_m)^T \ .$$

Uma importante propriedade do MMPP é que a superposição de MMPP's é novamente uma MMPP [6]. O gerador infinitesimal Q e a matriz de Taxas Λ do MMPP composto são calculados através dos geradores Q_i e das Matrizes de Taxas Λ_i de cada MMPP individual, assim

$$\begin{aligned} Q &= Q_1 \oplus Q_2 \oplus \cdots \oplus Q_m \ , \\ \Lambda &= \Lambda_1 \oplus \Lambda_2 \oplus \cdots \oplus \Lambda_m \ , \end{aligned}$$

onde $\oplus$ representa a soma de Kronecker como definida em [6]. Se Q_i e Λ_i são matrizes $k_i \times k_i$, então Q e Λ são matrizes $k \times k$, onde $k = \prod_{i=1}^{m} k_i$.

O poder de representação deste modelo está diretamente relacionado com o número de estados da cadeia de Markov m. Porém, com o aumento da dimensão do MMPP (isto é, m), o mesmo pode se tornar computacionalmente proibitivo.

O modelo que será analisado possui 2 estados, onde os tempos médios de permanência nos estados 1 e 2 são $1/r_1$ e $1/r_2$, respectivamente. Quando a cadeia estiver no estado j ($j = 1, 2$), o processo de chegada é Poisson com taxa igual a λ_j.

A escolha dos quatro parâmetros do MMPP de 2 estados é feita de modo a combinar com as seguintes características do tráfego agregado original:

1. A taxa média de chegadas;
2. A razão pico-sobre-média do número de chegadas em $(0, t_1)$;
3. A razão variância-sobre-média de longo-intervalo do número de chegadas; e

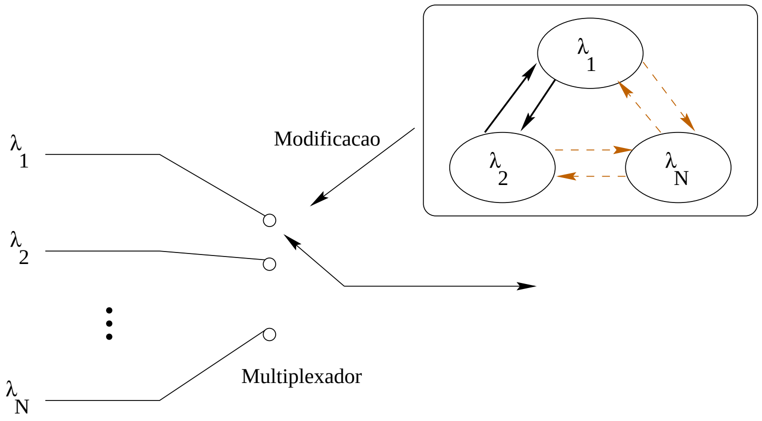

Figura 7: Modelo do processo de Poisson modulado por Markov

4. O terceiro momento do número de chegadas em $(0, t_2)$.

Estas quantidades já foram obtidas na seção 2.1 para o processo do tráfego agregado das fontes de voz. São elas respectivamente $M_i^s(t)$ (1), $\text{var}\left[N^s(0,t_1)\right]/E\left[N^s(0,t_1)\right]$ (2), $\lim_{t\to\infty}\left(\text{var}\left[N^s(0,t)\right]/M_1^s(t)\right)$ (3) e $\mu_3^{*s}(0,t_2)$ (4).

Usando a função de geração de probabilidade do número de chegadas em um intervalo [2],

$$g(z,t) = \pi exp\left\{\left[Q + (z+1)\Lambda\right]t\right\} e \ , \tag{18}$$

onde, para o MMPP de 2 estados,

$$\pi = \tfrac{1}{r_1+r_2}\begin{pmatrix} r_1 & r_2 \end{pmatrix} \ , \qquad e = \begin{pmatrix} 1 & 1 \end{pmatrix}^T \ ,$$

$$Q = \begin{bmatrix} -r_1 & r_1 \\ r_2 & -r_2 \end{bmatrix} \quad \text{e} \quad \Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \ ,$$

são obtidas expressões exatas para estas quantidades no Apêndice A. Combinando estas quantidades com as do tráfego agregado original são obtidos os parâmetros do MMPP de 2 estados r_1, r_2, λ_1 e λ_2 [1].

Em [6] são apresentados algoritmos para a solução da fila MMPP/G/1 e obtenção dos parâmetros desejados.

3.2 Dados

Como já foi descrito na seção 2.2, a geração de pacotes de dados por uma fonte única pode ser modelada, ou por um processo de chegadas de Poisson (caso contínuo), ou por um processo geométrico entre chegadas (caso discreto).

3.3 Vídeo

O tráfego de vídeo pode ser classificado de acordo com as variações de suas taxas de bites. Nas seções 3.3.1 e 3.3.2 serão analisados modelos para variações de curto-intervalo. Os modelos de variação das taxas de bites dentro de uma cena são úteis para fontes de vídeo com poucas mudanças, como por exemplo em vídeos de teleconferência. Como já foi discutido, estes tipos de fontes de vídeo apresentam uma distribuição das taxas de bites bem comportada (não variam subitamente) e alta autocorrelação intra-quadro. Nas seções 3.3.3 e 3.3.4 serão analisados modelos para variações de longo-intervalo.

3.3.1 Processo Auto-Regressivo (AR)

Nesta seção, uma fonte única de vídeo é aproximada por um modelo de processo Auto-Regressivo [4, 5]. A definição de um processo AR é a seguinte:

$$\lambda(n) = \sum_{m=1}^{M} a_m \lambda(n-m) + b\omega(n) \ ,$$

onde λ_n representa a taxa de bites da fonte durante o n-ésimo quadro; M é a ordem do modelo; ω_n é o processo randômico gaussiano; e a_m $(m = 1, 2, \ldots, M)$ e b são coeficientes. Para um modelo de Markov de um processo AR de 1ª ordem, tem-se

$$\lambda(n) = a_1 \lambda(n-1) + b\omega(n) \ . \tag{19}$$

O modelo de 1ª ordem é suficiente para os propósitos de engenharia. Assumindo que $\omega(n)$ tem média η e variância 1, e que $|a_1|$ é menor que 1, os valores dos coeficientes a_1 e b são determinados combinando o valor médio $E(\lambda)$ $(E(\lambda) = \lim_{n\to\infty} E(\lambda_n))$ e autocovariância $C(n)$ do processo AR com os reais valores medidos. Os valores de $E(\lambda)$ e $C(n)$ para o processo AR de 1ª ordem são dados por:

$$E(\lambda) = \tfrac{b}{1-a_1}\eta \ ; \qquad C(n) = \tfrac{b^2}{1-a_1^2} a_1^n \qquad n \geq 0 \tag{20}$$

Este modelo provê uma aproximação um tanto acurada da variação da taxa de bites para uma fonte única de vídeo sem mudanças de cena. Entretanto, a análise do modelo de fila com a entrada sendo um processo AR pode ser muito complexa e apenas viável para fins de simulação.

3.3.2 Processo de Markov de tempo contínuo

Maglaris et al. demonstraram um método baseado num processo de Markov de tempo contínuo e espaço discreto [4, 5]. Nesta aproximação, o processo $\lambda(t)$, que descreve a taxa de bites de uma fonte de vídeo no instante t, é quantizada somente em valores discretos e que são amostrados em instantes randômicos de um processo de Poisson no domínio do tempo. A qualidade da aproximação pode ser melhorada reduzindo-se a quantidade de informação em um degrau de quantização e aumentando-se a taxa de amostragem.

Para modelar estas fontes de vídeo, deve-se observar que as variações das taxas de bites dentro de uma cena são suaves, não ocorrendo saltos súbitos. Com isto, modela-se estas fontes como na figura 8, onde só existem transições entre estados vizinhos (modelo de Markov de nascimento e morte). Neste modelo existem $M+1$ estados, e o tamanho do degrau de quantização de estados é A bites por quadro (constante). Nos estados mais à esquerda, as taxas de bites são mais baixas, e aqueles mais à direita, as taxas são mais altas. A probabilidade de uma transição ocorrer para uma taxa de bites mais alta será maior para aqueles estados de mais baixa taxa de bites, decrescendo conforme a aumento da taxa de bites. Já a probabilidade de uma transição ocorrer para uma taxa de bites mais baixa será maior para aqueles estados de mais alta taxa de bites, decrescendo conforme a diminuição da taxa de bites.

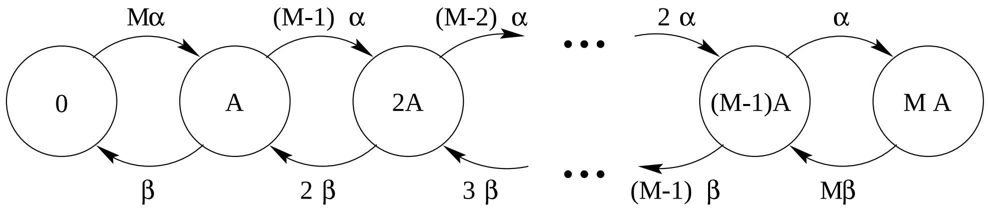

Figura 8: Diagrama de transição de estados para um modelo de fonte agregada

Este modelo pode aproximar tanto uma fonte única de vídeo, como também o tráfego agregado de várias fontes de vídeo.

O modelo apresentado na figura 8 pode ser decomposto na superposição de M independentes e idênticas minifontes ON-OFF, com comportamento similar ao modelo da fonte de voz apresentado 3.1.1, como ilustrado na figura 9. O estado do processo de chegada agregado pode então ser representada pelo número de minifontes que estejam no estado ON (ativo).

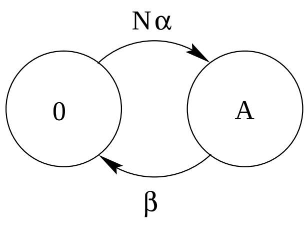

Figura 9: Modelo da minifonte

Para determinar os valores do degrau de quantização A, e as taxas de transição α e β, os valores de $E\left(\overline{\lambda_n}\right)$, variância $\overline{C_n}(0)$ e a função de autocovariância $\overline{C_n}(\tau)$ e do processo $\lambda(t)$ são combinados com os dados reais medidos. Seus valores são dados por

$$\begin{aligned} & E(\overline{\lambda_n}) = MA\tfrac{\alpha}{\alpha+\beta} \ ; \qquad \overline{C_n}(0) = MA^2\tfrac{\alpha}{\alpha+\beta}\left(1-\tfrac{\beta}{\alpha+\beta}\right) \ ; \\ & \overline{C_n}(\tau) = \overline{C_n}(0)e^{-(\alpha+\beta)\tau} \ ; \end{aligned} \tag{21}$$

O número de níveis de quantização M é escolhido arbitrariamente, porém deve ser grande o suficiente para cobrir todas as taxas de bites prováveis.

3.3.3 Processos de Markov

Usando o modelo de Markov [4, 5], é possível aproximar as variações dentro de uma cena mais precisamente do que nos modelos descritos anteriormente. Neste modelo são levados em consideração tanto as variações de curto-intervalo quanto as de longo-intervalo.

Este modelo é construído como uma combinação de duas taxas de bites básicas: a taxa de bites alta A_h e taxa de bites baixa A_l. A taxa de bites A_l corresponde a taxa de bites A do modelo do processo de Markov sem mudanças de cena. Este modelo supõe que um vídeo apresenta cenas com movimentos lentos, assim como cenas com movimentos rápidos. A figura 10 representa este modelo. Nela, as cenas com movimentos lentos estão representadas por A_l, e as cenas com movimentos rápidos por A_h.

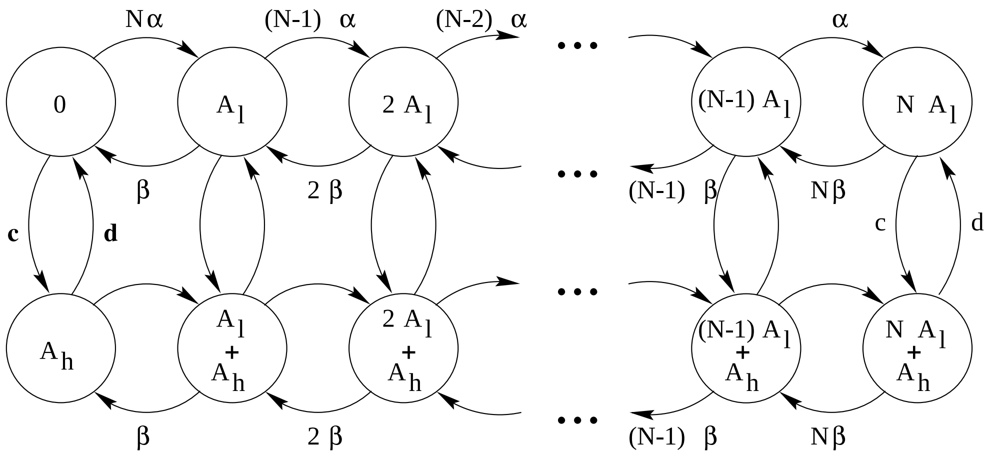

Figura 10: Diagrama de transição de estados da taxa para um modelo de font e de dois níveis de atividade

Os parâmetros deste modelo são determinados assim. Seja $c : d$ a razão entre o tempo gasto no nível de alta atividade (A_h) e o tempo gasto no período de baixa atividade (A_l). A razão do período de alta atividade q é derivada das medidas das fontes de vídeo reais como $q = c/(c + d)$. A taxa média geral λ_{all}, a razão média γ (razão da taxa média no nível alto e da taxa média no nível baixo) e as autocovariâncias $C(0)$ e $C(\tau)$ são dados por

$$\lambda_{all} = MpA_l + qA_h \tag{22}$$

onde

$$\begin{aligned} p &= a/(a+b) \; ; \\ \gamma &= (MpA_l + A_h)/MpA_l \; ; \\ C(0) &= Mp(1-p)A_l^2 \; ; \end{aligned}$$

$$C(\tau) \; = \; C(0)exp(-(a+b)\tau) \; ;$$

O único parâmetro livre é o número de degraus de quantização M.

A multiplexação de várias fontes pode ser modelada similarmente com a estrutura apresentada para o caso de uma fonte única. A figura 11 descreve o diagrama de transição de estados da taxa da modelagem da multiplexação de N fontes de vídeo.

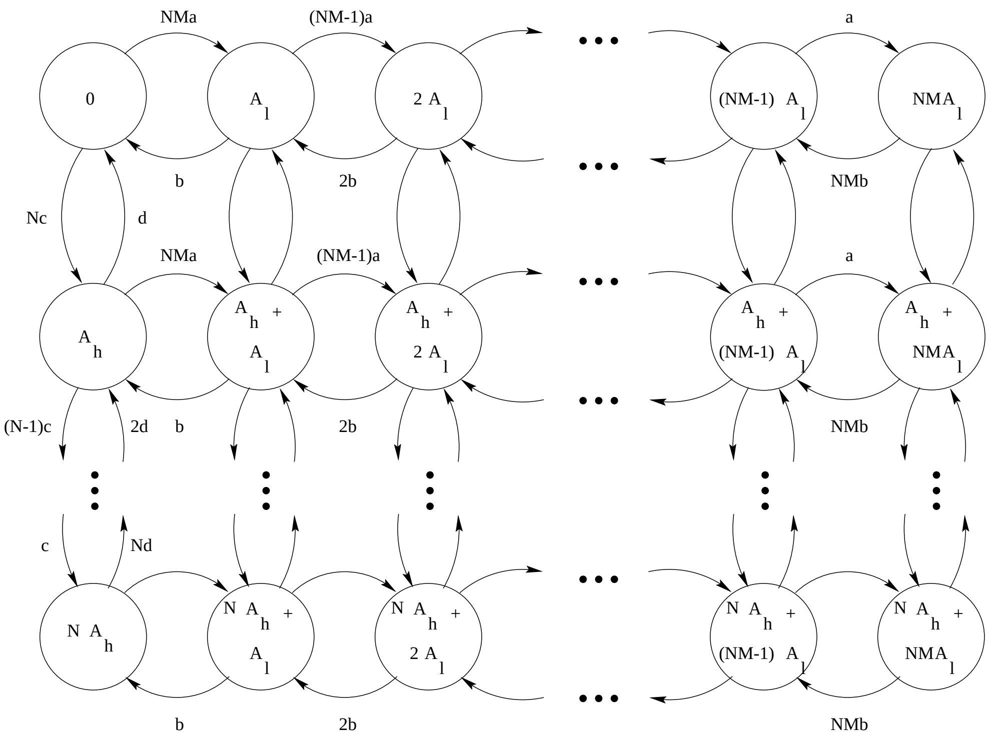

Figura 11: Diagrama de transição de estados da taxa para o modelo de fonte s agregadas (N fontes de vídeo multiplexadas)

As características da multiplexação são analisadas de modo similar ao descrito na seção 3.3.1.2, sendo decompostos na superposição de processos mais simples. Este processo pode ser pensado como sendo a superposição de $N \times M$ minifontes ON-OFF independentes e idênticas, como na figura 12 e N minifontes ON-OFF independentes e idênticas, como na figura 13.

3.3.4 Processo de Poisson modulado por Markov (MMPP)

O MMPP usado anteriormente na análise da multiplexação de pacotes de voz também pode ser usado para a análise de sinais de vídeo [5]. Este modelo é utilizado em tais sinais de vídeo segundo as seguintes hipóteses

1. Uma cena única é vista como um processo único que gera pacotes com um processo de Poisson.

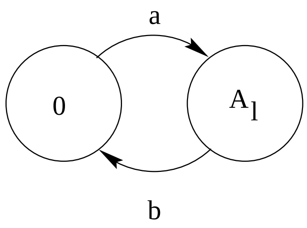

Figura 12: Modelo da minifonte a

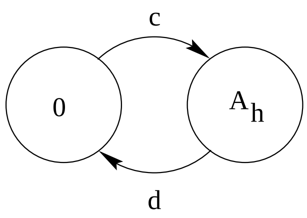

Figura 13: Modelo da minifonte b

2. As taxas de transições entre cenas são dadas por uma cadeia de Markov.

3. As altas taxas de bites seguintes a uma mudança de cena é considerada como sendo causada tanto por um estado de taxa alta, como também devido a chegadas em grupo.

A grande vantagem deste modelo é que ela atenta para modelar o aumento rápido das taxas de bites que ocorrem numa mudança de cena.

Vários métodos podem ser usados para modelar estes crescimentos rápidos. Dentre eles, podemos citar:

a. um modelo baseado em estados de alto nível que seguem chegadas de Poisson, com razões de chegada extremamente altas.

b. um modelo baseado em chegadas de grupo de 1 ordem.

c. um modelo baseado em estados de alto nível com chegadas de Poisson de grupos.

Os parâmetros requeridos por este modelo incluem o intervalo médio de mudança de cena (todos os modelos), o tempo médio de persistência e a razão de chegadas do estado de alto nível (modelos *a* e *c*), e o tamanho do grupo em chegadas de grupo (modelos *b* e *c*). Os valores destes parâmetros são derivados de medidas reais.

4 Conclusão

Dentre as técnicas propostas para o modo de transferência das Redes Digitais de Serviço Integrado de Banda Larga (RDSI-BL), o conceito de ATM é considerado como a mais promissora técnica de transferência devido a sua flexibilidade e eficiência. Porém, a importância do estudo de alocação de recursos para as RDSI-BL não reside apenas no fato de possibilitar um maior aproveitamento da capacidade do canal. Uma vez descoberto o quanto de recurso é solicitado por uma fonte, torna-se mais fácil a decisão relativa à aceitação ou não de uma nova chamada.

É necessário, então, o meio mais adequado de se calcular esta capacidade. Para que isto seja possível, deve-se não só obter um número razoável de resultados para cada modelo, mas também fazer uma comparação entre eles.

Vários modelos para cada tipo de fonte foram apresentados e resultados foram obtidas para cada modelo.

Os Sistemas de Fluxos Contínuos (SFF) são uma aproximação de fluxo contínuo para o estado de um multiplexador, isto é, a informação flui uniformemente através da fila. A modelagem do SFF consiste em conjunto de fontes modeladas por um processo Markoviano que assume dois estados, ON e OFF, uniformemente distribuídos e independentes. O estado ON corresponde a fonte ativa e o OFF, a fonte inativa.

Os modelos de Processos Semi-Markovianos (SMP) modelam aproximadamente os tempos constantes entre chegadas e os tempos de serviço das células quando as fontes são idênticas. O sistema é representado por uma cadeia de Markov bidimensional cujos estados denotam o número de fontes ativas e o número de células no sistema em um dado instante de tempo. As CMTC são menos acuradas que os SMP, visto que elas assumem que a geração de células por uma fonte ativa ocorre de acordo com um processo de Poisson.

O MMPP descreve um processo de Poisson cuja taxa de chegada varia de acordo com o estado de uma Cadeia de Markov de Tempo Contínuo independente do processo de chegada. Sua principal aplicação está na modelagem de processos de chegada correlacionados e de não-renovação. A modelagem será mais acurada quanto maior o número de estados da cadeia de Markov, podendo, contudo, tornar-se analiticamente inviável de se tratar.

O modelo Auto-Regressivo (AR) é usado para descrever processos cujas chegadas de células guardam uma grande correlação entre si. A taxa de chegada atual é função das anteriores e de um processo randômico gaussiano.

Apêndice A

Neste apêndice são obtidas expressões exatas do número esperado de chegadas no intervalo $(0, t)$, sua variância e seu terceiro momento para um processo de Poisson modulado por Markov (MMPP) de 2 estados [1]. O MMPP de 2 estados é um processo duplamente estocástico, onde a taxa do processo de chegada varia de acordo com o estado de uma cadeia de Markov de 2 estados. Sejam $1/r_i$ o tempo médio de permanência provisória no estado i ($i = 1, 2$) e λ_i a respectiva taxa de chegada do processo de Poisson para este estado. Seja π o vetor das probabilidades de equilíbrio da cadeia de Markov. Assim, a função de geração de probabilidade do número de chegadas em um intervalo de comprimento t é

$$g(z,t) = \pi exp\{[Q + (z+1)\Lambda]\, t\}\, e \ , \tag{23}$$

onde, para o MMPP de 2 estados,

$$\pi = \tfrac{1}{r_1+r_2}\begin{pmatrix} r_1 & r_2 \end{pmatrix} \ , \qquad e = \begin{pmatrix} 1 & 1 \end{pmatrix}^T \ ,$$

$$Q = \begin{bmatrix} -r_1 & r_1 \\ r_2 & -r_2 \end{bmatrix} \qquad \text{e} \qquad \Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \ ,$$

Definindo N_t como o número de chegadas em $(0,t)$ e $\overline{N_t} = E[N_t]$, tem-se por [6] que

$$\overline{N_t} = \pi\lambda t = \frac{\lambda_1 r_2 + \lambda_2 r_1}{r_1 + r_2} t \ ; \tag{24}$$

onde

$$\lambda = \begin{bmatrix} \lambda_1 & \lambda_2 \end{bmatrix} \ .$$

A razão variância-sobre-média de N_t, dada em [1], é

$$\frac{\text{var}[N_t]}{N_t} = 1 + \frac{2(\lambda_1-\lambda_2)^2 r_1 r_2}{(r_1+r_2)^2(\lambda_1 r_2 + \lambda_2 r_1)} - \frac{2(\lambda_1-\lambda_2)^2 r_1 r_2}{(r_1+r_2)^3(\lambda_1 r_2 + \lambda_2 r_1)t}\left(1 - e^{-(r_1+r_2)t}\right) \tag{25}$$

O terceiro momento de N_t é dado por

$$E\left[\left(N_t - \overline{N_t}\right)^3\right] = \left.\frac{\delta^3 g(z,t)}{\delta z^3}\right|_{z=1} - 3\overline{N_t}(\overline{N_t}-1)\frac{\text{var}(N_t)}{\overline{N_t}} - \overline{N_t}(\overline{N_t}-1)(\overline{N_t}-2) \tag{26}$$

Para obter a terceira derivada requerida em (26), deve-se lembrar que

$$g^{(3)}(1,t) = \left.\frac{\delta^3 g(z,t)}{\delta z^3}\right|_{z=1} = L^{-1}\left[\left.\frac{\delta^3 \tilde{g}(z,s)}{\delta z^3}\right|_{z=1}\right] \tag{27}$$

onde

$$\tilde{g}(z,s) = \pi\,[sI - Q - (z-1)\Lambda]^{-1}\, e \tag{28}$$

é a transformada de Laplace de $g(z,t)$ e L^{-1} denota a transformada inversa de Laplace. Assim, tem-se

$$\begin{aligned} g^{(3)}(1,t) &= L^{-1}\left\{6\pi\left[(sI-Q)^{-1}\Lambda\right]^3 (sI-Q)^{-1} e\right\} \\ &= L^{-1}\left\{\frac{6}{s^2}\pi\Lambda\left[(sI-Q)^{-1}\Lambda\right]^2 e\right\} \end{aligned} \tag{29}$$

que após a inversão resulta em

$$g^{(3)}(1,t) = \frac{6}{r_1+r_2}\left[\frac{A_{11}}{6}t^3 + \frac{A_{21}}{2}t^2 + A_{31}t + A_{12}te^{-(r_1+r_2)t} + A_{41}\left(1-e^{-(r_1+r_2)t}\right)\right] \quad (30)$$

onde

$$\begin{aligned}
A_{11} &= \frac{(\lambda_1 r_2 + \lambda_2 r_1)^3}{(r_1+r_2)^2}\,, \\
A_{21} &= \frac{2r_1 r_2(\lambda_1-\lambda_2)^2(\lambda_1 r_2+\lambda_2 r_1)}{(r_1+r_2)^3}\,, \\
A_{31} &= \frac{r_1 r_2(\lambda_1-\lambda_2)^2[\lambda_1 r_1+\lambda_2 r_2 - 2(\lambda_1 r_2+\lambda_2 r_1)]}{(r_1+r_2)^4}\,, \\
A_{41} &= \frac{-2r_1 r_2(\lambda_1-\lambda_2)^3(r_1-r_2)}{(r_1+r_2)^5}\,,
\end{aligned}$$

e finalmente

$$A_{21} = \frac{r_1 r_2(\lambda_1-\lambda_2)^2(\lambda_1 r_1+\lambda_2 r_2)}{(r_1+r_2)^4}\,.$$

Referências

[1] J. N. Daigle and J. D. Langford. Models for analysis of a packet voice communications systems. *IEEE Journal on Selected Areas in Communications*, 4(6):847–855, September 1986.

[2] Roberta D. Cavalheira and José Augusto Suruagy Monteiro. Modelos de fluxo contínuo para alocação de capacidades em redes ATM. In 12° *Simpósio Brasileiro de Redes de Computadores*, pages 305–320, Curitiba, Brazil, May 1994.

[3] H. Heffes and D. M. Lucantoni. A markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance. *IEEE Journal on Selected Areas in Communications*, 4(6):856–868, September 1986.

[4] J. J. Bae and Tatsuya Suda. Survey of traffic control schemes and protocols in ATM networks. *Proceeding of the IEEE*, 79(2):170–189, February 1991.

[5] Naohisa Ohta. *Packet Video: Modeling and Signal Processing.* Artech House, Boston, London, first edition, 1994.

[6] W. Fischer and K. Meier-Hellstern. The markov-modulated poisson process (MMPP) cookbook. *Performance Evaluation*, 18:149–171, March 1991.

[7] F. Yegenoglu and B. Jabbari. Characterization and modeling of aggregate traffic for finite buffer statistical multiplexer. *Computer Networks and ISDN Systems*, 26(9):1169–1186, May 1994.

[8] K. Sriram and W. Whitt. Characterizing superposition arrival processes in packet multiplexers for voice and data. *IEEE Journal on Selected Areas in Communications*, 4(6):833–846, September 1986.

[9] D. Anick, D. Mitrani, and M. Shondi. Sthocastic theory of a data handling system with multiples sources. *Bell Systems Technical Journal*, 61(8):1871–1894, October 1982.