Capítulo 2

Caracterização de Tráfego

2.1 Introdução

As redes ATM foram projetadas para suportar uma grande diversidade de conjuntos de aplicações com grandes variações nas suas características e com os mais diversos requisitos de qualidade de serviço. Infelizmente, muitas destas aplicações não possuem medições suficientes que permitam a sua caracterização de um modo acurado.

Para atender a esta diversidade de aplicações, várias classes de serviço foram definidas: o serviço de taxa binária constante (CBR - *Constant Bit Rate*), o serviço de taxa binária variável (VBR - *Variable Bit Rate*), o serviço de taxa binária indefinida (UBR - *Undefined Bit Rate*) e o serviço de taxa binária disponível (ABR - *Available Bit Rate*). A figura 2.1 ilustra algumas aplicações típicas de cada uma destas classes de serviço. O serviço CBR foi desenvolvido para suportar aplicações com requisitos bem definidos de atraso e que geram tráfego a uma taxa constante, tal como uma emulação de circuito. O serviço VBR também foi desenvolvido para suportar aplicações com requisitos bem definidos de atraso, mas que geram tráfego a uma taxa variável, tal como um vídeo sob demanda. O serviço UBR foi desenvolvido para suportar aplicações que não possuem requisitos definidos de serviço e que estejam dispostas a tolerar qualquer nível de QoS que a rede possar prover, tal como uma transferência de arquivo feita em *background*. A última classe de serviço, a ABR, é adequada para aplicações com requisitos mínimos de vazão e atraso.

A caracterização de uma fonte em um nível mais abstrato define as características de uma fonte de tráfego e seus requisitos de QoS [6]. As características do tráfego de uma aplicação são o conjunto mínimo de parâmetros que um usuário deverá fornecer ao

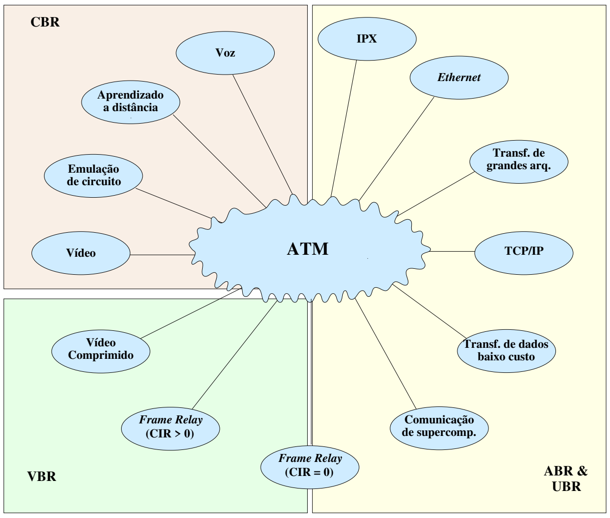


Figura 2.1: Serviços característicos da rede ATM.

gerenciador da rede para o controle efetivo do tráfego e a utilização máxima dos recursos da rede. Por exemplo, caso apenas o valor da taxa de pico fosse fornecido, não restaria outra alternativa a não ser a alocação pelo pico. Porém, para fontes cujas taxas médias de transmissão sejam muito inferiores as suas taxas de pico, esta estratégia representaria um enorme desperdício de banda passante.

Dentre os parâmetros mais utilizados para a descrição de tráfego de uma fonte estão a taxa binária de pico, taxa binária média, índice de rajada e o tamanho médio de uma rajada. Por rajada se entende a quantidade de informação gerada durante o período em que a fonte gera tráfego. Por índice de rajada se entende a razão entre os valores da taxas binárias de pico e média.

Este capítulo tem por objetivo a caracterização de tráfego para diversos tipos de aplicações que se estima encontrar em redes ATM. Com este intuito, cada uma das classes de serviço é vista em detalhes nas seções 2.2 à 2.5. Na seção 2.6, vários parâmetros de QoS associados a estes serviços são apresentados. Por último, na seção 2.7, vários parâmetros descritores de tráfego, utilizados para a descrição das características de uma fonte de tráfego, são apresentados.

2.2 Serviço CBR

Os serviços CBR geram tráfego a uma taxa constante e podem ser descritos apenas pela sua taxa de pico. Como neste caso a taxa binária média é igual a taxa binária de pico, o índice de rajada é igual a 1, ou seja, a fonte permanece no seu estado ativo durante todo o tempo de conexão. Como exemplos típicos deste serviço estão a emulação de circuito, a voz, o vídeo não comprimido e o áudio. Para atender às características de tráfego com taxa constante, criou-se uma camada de adaptação ATM denominada AAL (*ATM Adaption Layer*) tipo 1, que tem como funções a segmentação e remontagem das informações do usuário e o tratamento de condições de erro, perda e atrasos de células, além da manutenção do sincronismo entre a origem e o destino.

A voz CBR em ATM é transmitida usando a modulação por codificação de pulsos (PCM - *Pulse Code Modulation*). A PCM é um método de conversão de sinal analógico em digital através da representação das amostragens da amplitude quantizada em pulsos de código binário.

A banda passante a ser alocada para a voz CBR depende da taxa de amostragem (amostragens/s) e da quantização destas amostras (*bits*/amostras), sendo dada pela multiplicação destas duas.

O áudio CBR é conceitualmente igual a voz. A diferença entre eles está no fato dele requerer uma maior quantização a taxas de amostragem mais altas. Por exemplo, num disco compacto (CD), o áudio é armazenado usando a PCM de 16 *bits* a uma taxa de amostragem de 44.1 kHz dando aproximadamente 700 kbps [6] e o dobro em som estéreo.

O vídeo CBR é um pouco diferente conceitualmente do áudio e da voz, dado que a quantidade de *bits* gerado por quadro varia de acordo com o conteúdo da informação de cada quadro, sendo necessário fazer uma suavização do tráfego gerado. A figura 2.2 representa este processo [6].

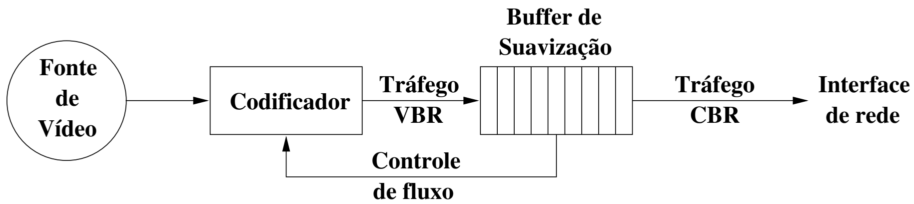


Figura 2.2: Vídeo CBR.

Nela observa-se a existência de um *buffer* de suavização necessário para a regularização da taxa binária na interface de rede. Devido ao tamanho do *buffer* ser finito, o mesmo está sujeito a transbordos, causando assim perda de células. Para evitar isto, é feito um um controle de fluxo sobre o codificador através da diminuição da sua taxa de amostragem e, conseqüentemente, da sua taxa de saída.

O gerenciamento do tráfego CBR é bastante fácil, visto que uma banda passante é alocada para a conexão enquanto esta durar, independentemente da fonte estar ou não transmitindo. No entanto, isto pode resultar num grande desperdício de banda passante. Nestes casos, o uso do serviço VBR tende a ser mais apropriado.

2.3 Serviço VBR

O serviço VBR é principalmente usado em aplicações cujas características de tráfego variam bastante, tais como aplicações de vídeo comprimido e *Frame Relay*. O tráfego gerado por estas fontes normalmente alternam entre períodos de atividades e silêncio e/ou tem uma taxa binária que varia continuamente. Em geral, estas fontes apresentam uma razão entre a taxas binárias de pico e média (índice de rajada) muito maior que 1. Assim, a suavização de uma fonte VBR para uma fonte CBR, como na figura 2.2, pode levar a uma subutilização muito grande dos recursos da rede e a uma degradação da qualidade de serviço.

O grande problema que surge na utilização do serviço VBR é como gerenciá-lo de um modo simples e efetivo de modo a fazer um melhor uso dos recursos da rede através do ganho com a multiplexação estatística e garantir a QoS desejada. Dentre os fatores que mais dificultam a oferta deste serviço estão as dificuldades de caracterização dos tráfegos e a falta de conhecimento de como este tráfego é modificado através da rede [7]. Portanto, torna-se fundamental ter um bom conhecimento e caracterização dos diversos tipos de fontes de tráfego que se estima encontrar em redes ATM para a construção de modelos de tráfegos e obtenção dos parâmetros de QoS desejados. Com este objetivo, são analisadas a seguir as fontes de voz, dados assíncronos e vídeo.

2.3.1 Fontes de Voz

O comportamento típico de uma fonte única de voz é mostrado na figura 2.3 [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 nos momentos de silêncio, durante os quais nenhum pacote é gerado. Em uma conversa normal, a duração dos períodos ativos e inativos podem ser aproximados por uma distribuição exponencial, embora a aproximação para o período inativo não seja perfeita. Uma melhor caracterização para o período inativo poderia ser uma distribuição que representasse a mistura do silêncio devido a pausas e com o silêncio devido a escuta. Como normalmente o número de fontes que são multiplexadas em um canal de transmissão é relativamente grande, a qualidade da aproximação do período inativo pela distribuição exponencial é muito boa [9].

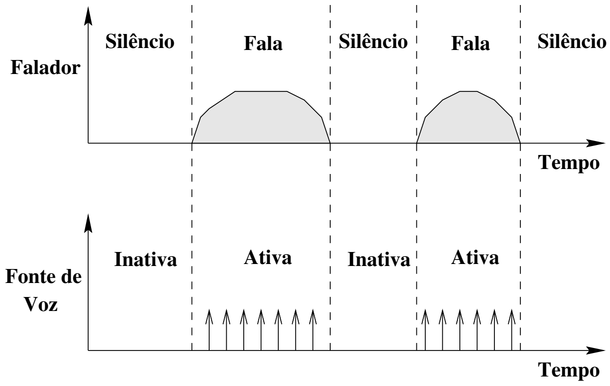


Figura 2.3: Comportamento típico de uma fonte de voz.

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, é importante 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.

As fontes de voz são um caso particular das chamadas fontes em rajadas do tipo silêncio/atividade. Estas últimas também são caracterizadas pela alternância entre períodos de atividade, durante os quais células são geradas a taxa binária de pico, e de silêncio, durante os quais nenhuma célula é gerada. Entretanto, as durações destes períodos não estão necessariamente distribuídos exponencialmente.

Como o objetivo principal é o estudo da multiplexação estatística das fontes de voz, é apresentado na figura 2.4 um modelo para esta superposição. Neste modelo, N fontes independentes de voz são superpostas, cada uma delas gerando pacotes a uma taxa constante. Todos os pacotes gerados alimentam um *buffer* comum de tamanho finito que é servido por um servidor, que representa o canal de transmissão, com uma vazão constante.

Vários modelos analíticos foram apresentados na literatura para a resolução do sistema de fila da figura 2.4 e obtenção dos parâmetros de QoS desejados. Aqui são analisados

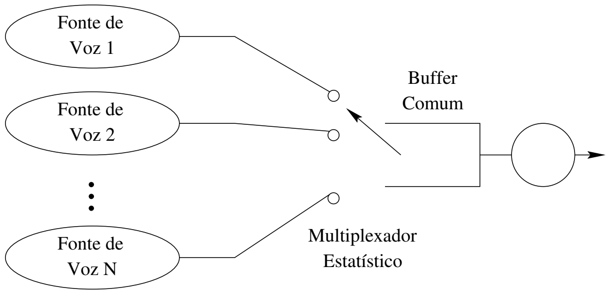


Figura 2.4: Modelo do multiplexador estatístico de fontes de voz.

um modelo de fluxo contínuo, o modelo de chegadas e serviços uniformes (UAS - *Uniform Arrival and Service*); um modelo de um processo Semi-Markoviano, o SMP(*Semi-Markov Process*); as Cadeias de Markov de Tempo Contínuo (CMTC); e um modelo de matrizes geométricas, o MMPP de 2 estados. Eles foram escolhidos por se tratarem de modelos relativamente simples, acurados e amplamente utilizados. Outros modelos podem ser encontrados em [10, 9, 11, 12]. Além disto, para a completa caracterização do tráfego gerado por uma fonte única de voz e pelo seu agregado, algumas de suas propriedades estatísticas também são obtidas.

Propriedades Estatísticas de uma Fonte de Voz

As propriedades estatísticas do processo de voz, apresentadas em [11], que servem como base de comparação com aquelas obtidas nos modelos analíticos, são a média, a razão variância sobre média e o terceiro momento do número de chegadas em um intervalo de tempo.

Sendo T o intervalo fixo entre chegadas de pacotes de voz durante os períodos de fala e $U(t)$ a função degrau unitário, a distribuição do tempos entre chegadas é dada por

$$F(t) = \left[(1 - \alpha T) + \alpha T \left(1 - e^{-\beta(t-T)}\right)\right] U(T - t), \tag{2.1}$$

cuja transformada de Laplace-Stieltjes (LST) é dada por

$$\bar{f}(s) = \int_0^\infty e^{-st} dF(t) = \left[1 - \alpha T + \frac{\alpha T \beta}{s + \beta}\right] e^{-st} \tag{2.2}$$

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

$$\lambda = \frac{-1}{\bar{f}'(0)} = \frac{1}{T + \frac{\alpha T}{\beta}}. \tag{2.3}$$

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 α^{-1}, seguida de uma distribuição exponencial do período de silêncio com média β^{-1}. Alguns valores característicos para estes parâmetros, encontrados em [8], são $\alpha^{-1} \approx 352$ ms, $\beta^{-1} \approx 650$ ms e $T \approx 16$ ms.

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 $(0,t)$, seja

$$M_r(t) = E\left[N^r(0,t)\right] \tag{2.4}$$

e seja

$$\overline{M_r}(s) = L\left[M^r(t)\right] \tag{2.5}$$

onde $L(.)$ denota a transformada de Laplace. Assim, é possível a obtenção de todos os momentos desejados através das equações (2.2) e (2.5). Com estas equações também pode ser obtido o índice de dispersão para contadores, $I(t)$, definido em [9], que satisfaz

$$\lim_{t\to\infty} I(t) = \lim_{t\to\infty} \frac{\mathrm{var}\left(N(0,t)\right)}{M_r(t)} = \frac{\mathrm{var}(X)}{E^2(X)}, \tag{2.6}$$

o coeficiente quadrático da variância do tempo entre chegadas de X. Este índice dá a noção da dependência existente entre os intervalos sucessivos entre chegadas. Aplicando-se este resultado em $F(t)$ e $\bar{f}(s)$, tem-se

$$M_1(t) = \frac{t}{T + \dfrac{\alpha T}{\beta}} \tag{2.7}$$

e

$$\lim_{t\to\infty} \frac{\mathrm{var}\left(N(0,t)\right)}{M_1(t)} = \frac{1-(1-\alpha T)^2}{(\alpha T + \beta T)^2}. \tag{2.8}$$

Agora considerando a superposição de n processos de voz idênticos e independentes, 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

$$\begin{aligned} N^S(0,t) &= \sum_{i=1}^{n} N_i(0,t)\ ; && (2.9)\\ M_1^S(0,t) &= E\left[N^S(0,t)\right] = nM_1(t)\ ; && (2.10) \end{aligned}$$

onde o sobrescrito S denota a superposição.

O índice de dispersão do tráfego agregado satisfaz

$$\frac{\mathrm{var}\left[N^S(0,t)\right]}{E\left[N^S(0,t)\right]} = \frac{\mathrm{var}\left[N(0,t)\right]}{E\left[N(0,t)\right]} \ . \tag{2.11}$$

O terceiro momento central para o processo de superposição será

$$\mu_3^{*S}(0,t) = E\left\{\left(\left[N^S(0,t)\right] - E\left[N^S(0,t)\right]\right)^3\right\} \tag{2.12}$$

que se reduz para

$$\mu_3^{*S}(0,t) = n\left[M_3(t) - 3M_2(t)M_1(t) + 2M_1^3(t)\right], \tag{2.13}$$

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

Modelo de Chegadas e Serviços Uniformes (UAS)

O modelo de chegadas e serviços uniformes (UAS) [9, 13] 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 2.5. Nele é suposto que os períodos ON e OFF são exponencialmente distribuídos com médias α^{-1} e β^{-1}, 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 2.6, onde N é o número de fontes de voz no sistema. Este processo é referenciado como processo de fase, sendo denotado por $\wp$.

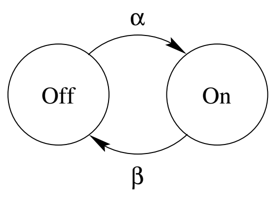


Figura 2.5: Modelo ON/OFF de uma fonte de voz.

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:

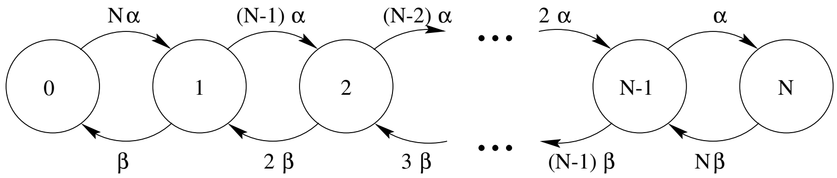


Figura 2.6: Modelo do processo de fases.

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

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 só é colocada no *buffer* de transmissão após a geração completa de um pacote por algumas das fontes ativas, não sendo possível a transmissão de parte de um pacote que ainda não foi completamente recebido. Já no modelo UAS, é possível para a informação ser transmitida enquanto está sendo recebida.

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

$$\begin{aligned}(i-C)\frac{dF_i(b)}{db} &= (N-i+1)\alpha F_{i-1}(b) - \{(N-1)\alpha + \beta i\} F_i(b) + \\ & \quad \beta(i+1)F_{i+1}(b) \ ; \qquad 0 \leq i \leq N\end{aligned} \tag{2.14}$$

ou, equivalentemente

$$D\frac{d}{db}F(b) = MF(b) \ ; \tag{2.15}$$

onde

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

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

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

$$G(b) \triangleq \Pr(\text{tamanho} > b) = 1 - \mathbf{1}' F(b) \ ; \qquad b \geq 0 \ . \tag{2.16}$$

Observe também que

$$F_i(\infty) = \frac{1}{(\alpha+\beta)^N} \begin{pmatrix} N \\ i \end{pmatrix} \alpha^i \ ; \qquad 0 \leq i \leq N \ ,$$

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

É importante observar aqui que C é considerado não inteiro, pois se o fosse, uma das equações diferenciais em (2.14) 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. [14], o conjunto dos auto-valores é obtido através da solução do conjunto de equações de 2^o grau em (2.17).

$$A(k)z^2 + B(k)z + C(k) = 0 \ ; \qquad k = 0, 1, \cdots, N \ ; \tag{2.17}$$

onde

$$\begin{aligned} A(k) &\triangleq (N/2 - k)^2 + (N/2 - C)^2 \ , \\ B(k) &\triangleq 2(1 - \alpha/\beta)(N/2 - k)^2 - N(1 + \alpha/\beta)(N/2 - C) \ , \\ C(k) &\triangleq (\alpha/\beta + 1)^2 \left\{ (N/2)^2 - (N/2 - k)^2 \right\} \ . \end{aligned}$$

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

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

onde

$$\begin{aligned} r_1 &= \left\{ -(z + 1 - \alpha/\beta) + \sqrt{(z + 1 - \alpha/\beta)^2 + 4\alpha/\beta} \right\} /2\alpha/\beta \\ r_2 &= \left\{ -(z + 1 - \alpha/\beta) - \sqrt{(z + 1 - \alpha/\beta)^2 + 4\alpha/\beta} \right\} /2\alpha/\beta \\ k &= \frac{z\beta C - N\alpha + N\alpha r_1}{\alpha(r_1 - r_2)} \ . \end{aligned}$$

E, finalmente, tem-se

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

onde

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

Em [15] é feita uma extensão deste modelo para o caso de filas finitas. Nele, a distribuição de equilíbrio da fila é descrita por um conjunto de equações diferenciais que, juntamente com as equações de contorno que descrevem o comportamento da fila nos seus limites, podem ser resolvidas numericamente para fornecer as distribuições de equilíbrio do atraso e da perda de pacotes.

Processos Semi-Markovianos (SMP)

No modelo de processos semi-Markovianos [9], 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. Assim,

- 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 $[V(C-j)]^{-1}$ s, a partir do instante em que o processo entrar na fase j; e
- quando $\wp = j > C$, então o tamanho da fila crescerá de um pacote a cada $[V(C-j)]^{-1}$ s, a partir do instante em que o processo entrar na 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. Já no modelo SMP, assume-se que quando a fase do processo é igual ou menor a C, o tamanho da fila não se altera ou descrece, respectivamente, estando, portanto, em desacordo com o sistema real.

Neste modelo, o estado do processo, ilustrado na figura 2.7, é denotado por $(l(t), \wp(t))$, 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 2.7 é um processo semi-Markoviano.

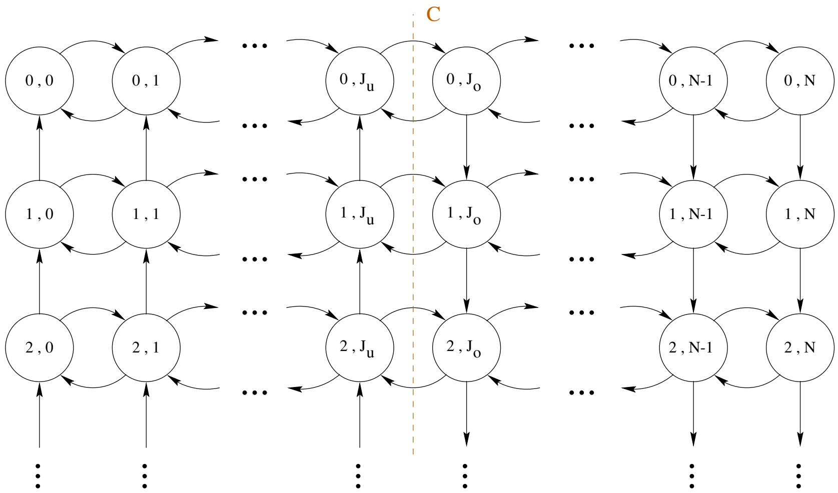


Figura 2.7: Diagrama de estados para o modelo do processo semi-Markoviano.

Seja $p_{i,j} = \lim_{t\to\infty} \Pr\{l(t) = i, \wp(t) = j\}$. 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

$$p_{i,j} = \frac{q_{i,j} m_{i,j}}{\sum_{k=0}^{\infty} \sum_{l=0}^{N} q_{k,l} m_{k,l}} \; ; \qquad 0 \leq j \leq N \; . \tag{2.18}$$

O cálculo dos $q_{i,j}$'s pode ser feito utilizando técnicas de geometria das matrizes [16], e respectivamente, através de (2.18), podem ser calculados os valores de cada $p_{i,j}$. Assim, a probabilidade π_i do tamanho da fila ser igual a i será

$$\pi_i = \Pr(l = i) = \sum_{j=0}^{N} p_{i,j} \; ; \tag{2.19}$$

onde N é o número total de fontes de voz.

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

$$G(b) \triangleq \Pr(\text{tamanho} > b) = 1 - \sum_{i=0}^{b} \pi_i . \qquad b \geq 0 \; . \tag{2.20}$$

Cadeias de Markov de Tempo Contínuo (CMTC)

No modelo de CMTC [9], 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 2.8, 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 2.8. E, finalmente, a probabilidade do tamanho da fila ser igual a i, π_i, será

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

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

$$G(b) \triangleq \Pr(\text{tamanho} > b) = 1 - \sum_{i=0}^{b} \pi_i. \qquad b \geq 0 \ . \tag{2.21}$$

Processos de Poisson modulado por Markov (MMPP)

Neste modelo, o tráfego agregado dos processos de chegada é aproximado por um processo de fluxo mais simples, correlacionado e de não-renovação. O fluxo aproximado é escolhido de modo que algumas características estatísticas combinem exatamente com aquelas da superposição original. Na escolha do fluxo 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 [11, 12, 10, 17, 18], 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 2.9 ilustra este procedimento. O MMPP é parametrizado por uma cadeia de Markov de tempo contínuo de m estados com gerador infinitesimal Q [17] e as m taxas de chegada de Poisson $\lambda_1, \lambda_2, \cdots, \lambda_m$, usando-se a notação

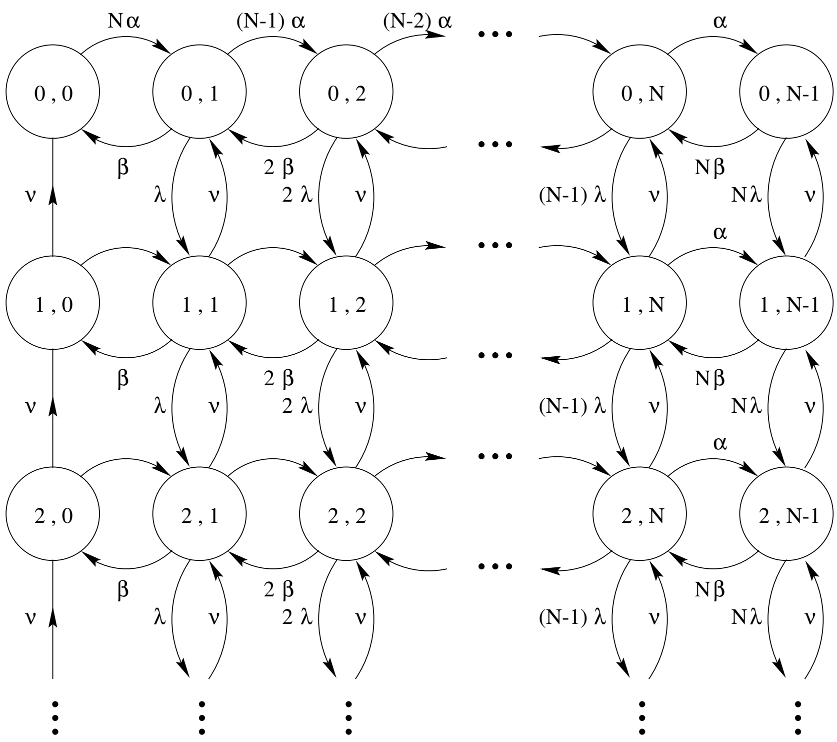


Figura 2.8: Diagrama de estados para o modelo da cadeia de Markov de tempo contínuo.

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

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

Uma importante propriedade do MMPP é que a superposição de MMPP's resulta em uma nova MMPP [17]. 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

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

onde $\oplus$ representa a soma de Kronecker como definida em [17]. 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.

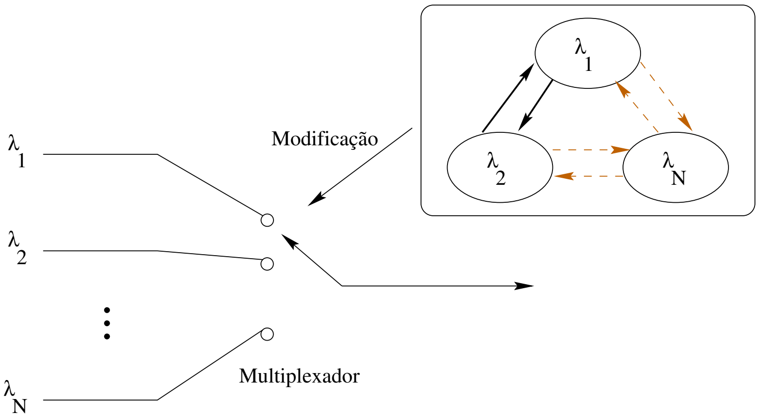


Figura 2.9: Modelo do processo de Poisson modulado por Markov.

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 da variância pela taxa média do número de chegadas em $(0, t_1)$;
3. a razão da variância pela taxa média de longo-intervalo do número de chegadas; e
4. o terceiro momento do número de chegadas em $(0, t_2)$.

Estas propriedades estatísticas já foram obtidas nesta mesma seção para o processo do tráfego agregado das fontes de voz. São elas respectivamente $M_1^s(t)$ (eq. 2.10), $\text{var}\left[N^s(0,t_1)\right]/E\left[N^s(0,t_1)\right]$ (eq. 2.11), $\lim_{t\to\infty}\left(\text{var}\left[N^s(0,t)\right]/M_1^s(t)\right)$ (eq. 2.11) e $\mu_3^{*s}(0,t_2)$ (eq. 2.12).

Usando a função de geração de probabilidade do número de chegadas em um intervalo [9],

$$g(z,t) = \pi \exp\left\{\left[Q + (z+1)\Lambda\right]t\right\} e \ , \tag{2.22}$$

onde, para o MMPP de 2 estados,

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

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

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 [9].

Calculados os parâmetros do MMPP que modela o tráfego agregado original, o multiplexador estatístico pode ser modelado por um sistema de fila do tipo MMPP/D/1/K, onde K representa o tamanho máximo da fila mais 1. Em [17], são apresentados algoritmos para a solução da fila MMPP/G/1 e obtenção dos parâmetros estatísticos desejados, tais como a distribuição do tempo de espera na fila, o atraso médio na fila e a distribuição do tamanho da fila. Com estes parâmetros, a probabilidade de transbordo da fila do sistema real pode então ser aproximada pela probabilidade do tamanho da fila ser maior ou igual a K.

2.3.2 Fontes de 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 banda passante 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 1/30 s, gerando aproximadamente 63 Mbps, se uma esquema PCM de codificação for usado. Entretanto, devido ao fato da variação entre quadros sucessivos ser usualmente 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 binárias em mudanças de cenas, movimentos dentro de uma cena e mudanças na sutileza dos detalhes das imagens bidimensionais do vídeo. A tabela 2.1, descrita em [9], classifica estas variações nas taxas binárias 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 binária muda como um degrau, e as características da variação das taxas binárias 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 binárias 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 *buffer*s 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*.

As fontes de vídeo são um caso particular das chamadas fontes VBR. Estas fontes, como as fontes de vídeo, possuem uma variação contínua da taxa de transmissão. Entretanto, a curva que descreve esta variação de taxas é genérica e pode assumir qualquer forma, não estritamente a de uma fonte de vídeo.

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 2.10 [10]. Neste modelo, N fontes independentes de vídeo de taxas binárias variáveis são suavizadas em *pré-buffers* individuais e acumuladas num *buffer* principal, o montador e desmontador de pacotes (PAD - *Packet Assembly and Desassembly*). A saída destes *buffer*s 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 (*First In, First Out* - Primeiro a chegar, Primeiro a sair).

Vários modelos para o tráfego de vídeo foram propostos na literatura para a resolução do sistema de fila da figura 2.10. Aqui são analisados alguns dos mais importantes. São eles os processos auto-regressivo, as cadeias de Markov de tempo contínuo, os processos de Markov e os processos de Poisson modulados por Markov. Os dois primeiros são modelos para variações de curto-intervalo, onde apenas são consideradas as variações das taxas

Tabela 2.1: Classificação das variações das taxas binárias.

Tipo	Escala de Tempo	Causas	Características
Variação entre Quadros de Longo-intervalo	Vários segundos ou mais	Mudanças de cenas	Variação descontínua, diferindo as características estatísticas antes e depois da mudança
Variação entre Quadros de 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

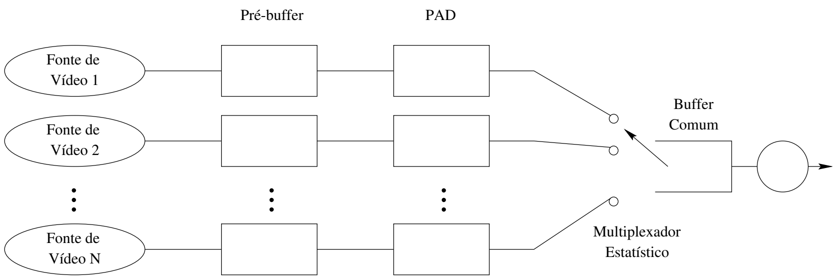


Figura 2.10: Modelo do multiplexador estatístico de fontes de vídeo.

binárias dentro de uma cena. Estes modelos são úteis para fontes de vídeo com poucas mudanças, como por exemplo em vídeos de teleconferência. Como já mencionado, estes tipos de fontes de vídeo apresentam uma distribuição das taxas binárias bem comportada (não variam subitamente) e alta autocorrelação intra-quadro. Os dois últimos são modelos para variações de longo-intervalo.

Processo Auto-Regressivo

Nesta seção, uma fonte única de vídeo é aproximada por um modelo de processo auto-regressivo [12, 10]. O processo Auto-Regressivo (AR), como o próprio nome já diz, é um processo no qual a taxa atual de transmissão é uma função dos seus valores passados, sendo ideal para os casos onde existe uma grande correlação entre as taxas sucessivas de amostragens como no caso do vídeo.

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 binária 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.

Em [19], é mostrado que a taxa de decaimento da função auto-covariância é dominada por uma exponencial simples. Sendo assim, o processo AR de 1^{a} ordem é suficiente para os fins de engenharia. Para o modelo de um processo AR de 1^{a} ordem, tem-se

$$\lambda(n) = a_1 \lambda(n-1) + b\omega(n) \ . \tag{2.23}$$

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 a autocovariância $C(n)$ e o valor médio $E(\lambda)$ $(E(\lambda) = \lim_{n\to\infty} E(\lambda_n))$ do processo AR com os valores reais medidos. Os valores de $E(\lambda)$ e $C(n)$ para o processo AR de 1^{a} ordem são dados por:

$$E(\lambda) = \tfrac{b}{1-a_1}\eta \ ; \qquad C(n) = \tfrac{b^2}{1-a_1^2} a_1^n \ ; \qquad n \geq 0 \ . \tag{2.24}$$

Este modelo provê uma aproximação um tanto acurada da variação da taxa binária 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. A grande vantagem deste modelo está na sua simplicidade, sendo bastante útil para fins de simulação.

Cadeias de Markov de Tempo Contínuo

Maglaris et al. [19] demonstraram um método baseado numa cadeia de Markov de tempo contínuo e espaço discreto [10, 12]. Tempo contínuo porque as mudanças de estados ocorrem continuamente no tempo e espaço discreto porque suas taxas binárias assumem apenas valores discretos. Nesta aproximação, o processo $\lambda(t)$, que descreve a taxa binária do agregado de fontes de vídeo no instante t, é quantificada somente em valores discretos (níveis) e que são amostrados em instantes randômicos de um processo de Poisson no domínio do tempo. A quantidade de níveis utilizados está diretamente relacionado com a qualidade da aproximação do modelo. Por exemplo, a utilização de apenas dois níveis equivaleria a aproximar esta fonte por uma fonte em rajada, variando apenas entre a taxa máxima e a nula. Esta qualidade também pode ser melhorada através do aumento do valor da taxa de amostragem.

Para modelar estas fontes de vídeo, deve-se observar que as variações das taxas binárias dentro de uma cena são suaves, não ocorrendo saltos súbitos. Com isto, modela-se estas fontes como na figura 2.11, 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 *bits* por quadro (constante). Nos estados mais à esquerda, as taxas binárias são mais baixas, e naqueles mais à direita, as taxas são mais altas. A probabilidade de uma transição ocorrer para uma taxa binária mais alta será maior para aqueles estados de taxa binária mais baixa, decrescendo conforme a aumento da taxa binária. Já a probabilidade de uma transição ocorrer para uma taxa binária mais baixa será maior para aqueles estados de taxa binária mais alta, decrescendo conforme a diminuição da taxa binária.

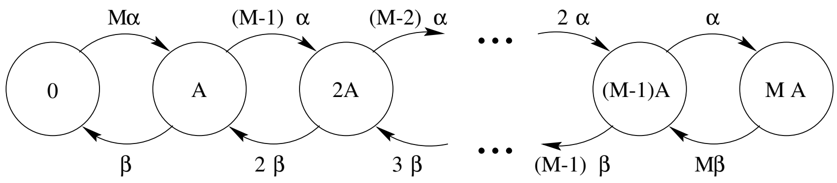


Figura 2.11: Diagrama de transição de estados para um modelo de fonte agregada.

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

Maglaris et al. [19] observaram que o modelo apresentado na figura 2.11 pode ser decomposto na superposição de M minifontes independentes e idênticas que alternam entre os estados de envio de 0 *bits*/quadro e A *bits*/quadro.

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

$$E(\overline{\lambda_n}) = MA\frac{\alpha}{\alpha+\beta}\ ; \qquad \overline{C_n}(0) = MA^2\frac{\alpha}{\alpha+\beta}\left(1-\frac{\alpha}{\alpha+\beta}\right)\ ; \qquad \overline{C_n}(\tau) = \overline{C_n}(0)e^{-(\alpha+\beta)\tau}\ ; \tag{2.25}$$

O número de níveis de quantização M é escolhido arbitrariamente, porém deve ser grande o suficiente para cobrir todas as taxas binárias prováveis para uma melhor caracterização do tráfego do agregado de conexões. Em [19], é mostrado que um valor de M igual a 20 N, onde N representa o número de fontes multiplexadas, produz resultados analíticos bastante acurados.

Processos de Markov

Usando o modelo de processos de Markov [12, 10], 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 binárias básicas: a taxa binária alta A_h e taxa binária baixa A_l. A taxa binária A_l corresponde a taxa binária 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 2.12 representa este modelo. Nela, as cenas com movimentos lentos estão representadas por A_l, e as cenas com movimentos rápidos por A_h.

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

$$\lambda_{\text{all}} = MpA_l + qA_h \tag{2.26}$$

onde

$$p \;=\; a/(a+b)\ ;$$

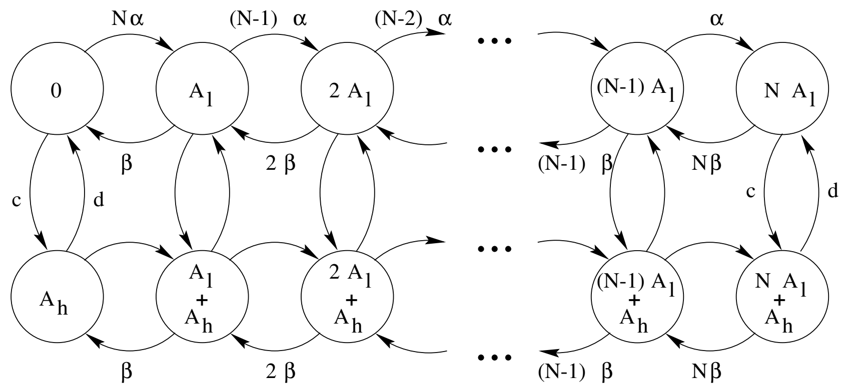


Figura 2.12: Diagrama de transição de estados da taxa para um modelo de fonte de dois níveis de atividade.

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

O único parâmetro livre é o número de degraus de quantização M. Como no modelo CMTC, o valor de M deve ser escolhido de modo a cobrir todas as taxas binárias possíveis.

A multiplexação de várias fontes pode ser modelada similarmente com a estrutura apresentada para o caso de uma fonte única. A figura 2.13 descreve o diagrama de transição de estados da taxa da modelagem da multiplexação de N fontes de vídeo.

As características da multiplexação são analisadas de modo similar ao descrito na seção 2.3.1 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 2.14 (a) e N minifontes ON-OFF independentes e idênticas, como na figura 2.14 (b).

Processos 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 [10]. 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;

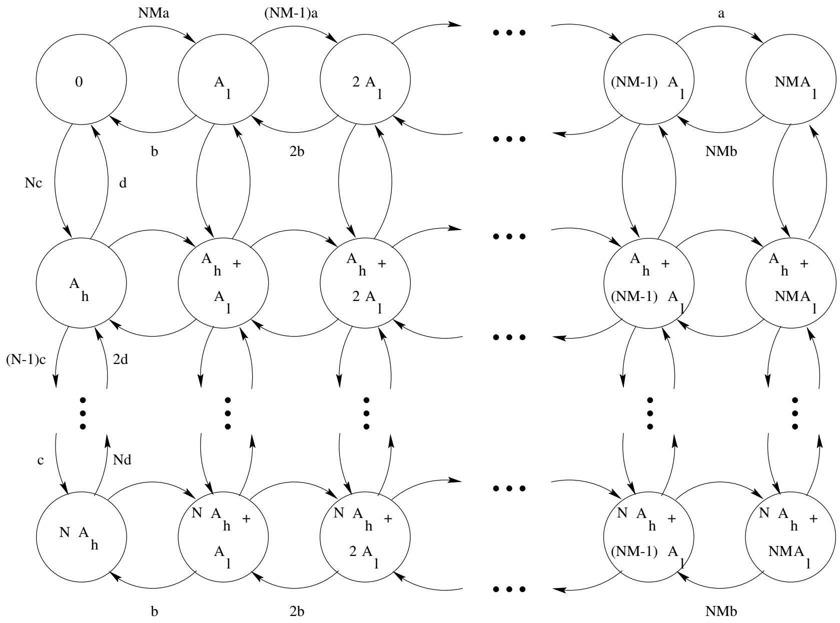
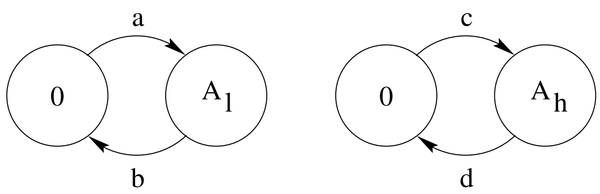


Figura 2.13: Diagrama de transição de estados da taxa para o modelo de fontes agregadas (N fontes de vídeo multiplexadas).



(a) Minifonte a. (b) Minifonte b.

Figura 2.14: Modelos das minifontes.

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

3. as altas taxas binárias 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 binárias que ocorrem numa mudança de cena.

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

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^a 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.

2.3.3 Fontes de Dados

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) [10]. Em um caso interativo, um 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.

2.3.4 Resumo

Nos modelos UAS, CMTC e SMP para fontes de voz, todas as fontes são consideraradas independentes e possuem processos de chegadas representados por um processo de renovação, que alterna entre os períodos ativos e inativos. A diferença entre eles está na modelagem do processo chegadas durante o período ativo e como o serviço é realizado para fins de tratamento analítico. Já o modelo MMPP não considera o processo de chegada de uma única fonte de forma isolada, mas sim o do tráfego agregado como um todo.

À excessão do modelo UAS com filas infinitas, nenhum dos modelos apresentados possuem fórmulas fechadas para o cálculo dos parâmetros desejados, tal como a taxa de perda de células, que só podem ser obtidos através de resolução numérica. Além disto, em alguns destes modelos, como o MMPP, o esforço computacional envolvido é muito grande.

Numa comparação feita entre os resultados obtidos nos modelos UAS, CMTC e SMP, com os resultados obtidos por simulação, Daigle e Langford [9] relatam que tanto o modelo SMP quanto o UAS mostraram-se bastante acurados, com um melhor desempenho do modelo SMP. Entretanto, o esforço computacional requerido pelo modelo UAS é bem menor.

Numa análise semelhante feita com o modelo MMPP, Heffes e Lucantoni [11] relatam que o modelo MMPP se mostrou bastante acurado numa comparação feita entre os seus resultados e aqueles obtidos por simulação. Entretanto, a resolução deste modelo envolve esforços computacionais bem maiores do que os modelos anteriores citados.

Em todos os modelos usados para fontes de vídeo, as fontes são consideradas independentes e apresentam uma grande correlação entre chegadas sucessivas. Os modelos AR e CMTC são adequados para fontes de vídeo com variações entre quadros de curto-intervalo (como por exemplo, um videofone), enquanto que os modelos de processos de Markov e o MMPP servem também para fontes com variações de longo-intervalo (como por exemplo, um filme, onde ocorrem diversas mudanças de cena).

A análise do processo de chegada de um modelo AR para fins analíticos é muito complexa e, como conseqüência, o seu uso é apenas viável para fins de simulação. Entretanto, como mostrado por Maglaris et al. [19] numa comparação com uma seqüência de vídeo real, esta aproximação produz resultados bastante acurados.

O modelo CMTC permite uma análise bem simples para o processo de chegada de uma fonte de vídeo e de seu agregado. Sua principal vantagem está no pequeno esforço computacional envolvido, já que seus resultados podem ser obtidos de modo similar ao modelo UAS, usado para fontes de voz. Além disto, como mostrado no capítulo 3, seus resultados são bastante acurados.

Os modelos de processos de Markov e o MMPP atentam para fontes de vídeo com mudanças de cenas, mas necessitam de uma análise bem mais complexa do que os modelos AR e CMTC. Enquanto nos processos de Markov, é difícil a determinação de seus parâmetros de modo a combinar com os de uma fonte de vídeo genérica, mais resultados são necessários para a validação do modelo MMPP.

Neste primeiro trabalho do GTA voltado para o Controle de Admissão de Conexão (CAC) em redes ATM, optou-se por utilizar somente os modelos UAS para fontes de voz e CMTC para fontes de vídeo na resolução analítica e o modelo AR na simulação de fontes de vídeo, pelos mesmos se tratarem de método relativamente simples. Esta opção foi feita com base em comparações já feitas na literatura e por acreditar que a perda ocasionada é pequena. Além disto, como é mostrado no capítulo 3, estes modelos apresentam uma ótima conformidade com os resultados obtidos por simulação.

2.4 Serviço ABR

A maior parte das aplicações comunicação de dados podem suportar variações no atraso de pacotes, mas são sensitivas quanto a perda destes. Para a redução desta perda, há uma necessidade para que a rede informe ao usuário um provável congestionamento que pode resultar em perda, através de uma realimentação com o transmissor.

O serviço ABR foi desenvolvido para implementar esta realimentação e, assim, suportar aplicações que não possam efetivamente caracterizar seus tráfegos no estabelecimento da conexão, mas que possam adaptar-se a um protocolo de controle de fluxo. O serviço ABR parte da premissa que se a fonte se comporta de um certo modo ao controle de fluxo, então nenhum pacote será perdido. Por exemplo, um usuário pode saber que a sua aplicação funciona perfeitamente em uma rede local *Ethernet* de 10 Mbps e pessimamente através de um modem de 9.6 kbps. Porém, este mesmo usuário teria dificuldades em definir qual o melhor valor que serviria tanto como uma garantia da qualidade de serviço quanto um limite de banda passante para esta aplicação, como seria necessário para a utilização do serviço CBR. O serviço ABR é apropriado para estes tipos de aplicações [20].

A principal motivação, entretanto, para a criação do serviço ABR é o suporte econômico de tráfego de dados, onde um pacote é segmentado em várias células e a perda de qualquer uma delas causaria a retransmissão de todo o pacote por um protocolo de uma camada superior. Alguns estudos mostram que o descarte indiscriminado de células ATM em estados de congestionamento poderia levar ao colapso a vazão de aplicações de transmissão de dados assíncronos [20, 21].

O serviço ABR pode ser descrito pelos seguintes parâmetros: Taxa mínima de células (MCR - *Minimum Cell Rate*), Taxa de pico de células (PCR - *Peak Cell Rate*) e Taxa disponível de células (ACR - *Available Cell Rate*). Os parâmetros MCR e PCR são fornecidos pelo usuário, enquanto que o parâmetro ACR é fornecido pela rede. O parâmetro ACR é variável e indica a taxa atual de transmissão de células para a conexão fornecida pela rede. Seu valor estára sempre compreendido entre os valores de MCR e PCR de acordo com o estado da rede.

2.5 Serviço UBR

O serviço UBR não oferece nenhuma garantia de QoS e não requer nenhum conhecimento anterior das características do tráfego. Neste tipo de serviço, o usuário deve estar disposto a tolerar qualquer capacidade e taxa de perda de células que a rede possa prover no instante em que a célula é transmitida pela rede [22]. A QoS efetiva pode ser gerenciada através de regras de engenharia. O único mecanismo de realimentação usado é a notificação de congestionamento que poderá ser usado pelos sistemas finais para se adaptarem ao estado de congestionamento.

2.6 Parâmetros de Qualidade de Serviço

O gerenciamento de tráfego da camada ATM deve suportar um conjunto de classes de QoS suficientes para todos os serviços previsíveis em RDSI-FL [3]. Cada classe de serviço tem associado a ela uma característica de tráfego e um determinado nível de QoS. A classe de QoS define os objetivos de desempenho a serem garantidos para cada conexão dentro deste serviço.

O nível de QoS é uma função dos seguintes parâmetros de QoS da camada ATM:

- Taxa de perda de células,

- Atraso de transferência de célula e
- Variação do atraso de célula.

2.6.1 Taxa de Perda de Células

A taxa de perda de células (CLR - *Cell Loss Ratio*) para uma conexão ATM é definida como a razão entre o número de células perdidas ou descartadas e o número total de células transmitidas.

A perda de células em uma conexão ATM pode afetar o serviço de diversas maneiras. Em uma conexão de áudio ou voz, ela será notada como uma quebra na continuidade da apresentação. Numa conexão de vídeo, ela pode representar a perda de uma parte de um quadro ou mesmo de todo ele. Em aplicações de dados, a perda de uma célula pode causar a retransmissão pelos protocolos das camadas superiores dos pacotes danificados. Além disto, se a causa da perda de células é devido a um estado de congestionamento, estas retransmissões podem levar a um quadro ainda mais caótico.

Para cada um dos casos mencionados, deverá ser dado um tratamento específico por parte da aplicação. No Grupo de Teleinformática e Automação (GTA) da UFRJ, estes problemas têm sido alvo de muitas pesquisas e trabalhos. Para uma melhor referência, ver [23, 24, 25].

2.6.2 Atraso de Transferência de Célula

O atraso de transferência de célula (CTD - *Cell Tranfer Delay*) representa o tempo gasto na transmissão de uma célula entre os endereços fonte e destino. O CTD em uma rede ATM pode ser causado por vários fatores. Dentre estes, estão os atrasos de codificação, empacotamento, propagação, transmissão, comutação, fila e remontagem [6]. Um baixo valor de CTD indica uma QoS mais alta.

2.6.3 Variação do Atraso de Célula

A variação do atraso de célula (CDV - *Cell Delay Variation*) descreve a variação no atraso de transferência de células. Existem dois modos possíveis para a sua medição: em um único ponto ou em dois pontos de referência. A medição de um ponto do CVD dá a medida de variação no padrão de células em um único ponto, sendo igual a diferença entre o tempo real de chegada de uma célula e o seu tempo previsto. A medição em dois pontos usa como pontos de referência a fonte e o destino e é igual a diferença entre o valor real do CDV observado entre estes dois pontos e o seu valor calculado, causado por estes dois pontos.

Esta variação pode ser causada por diversos fatores, dentre os quais encontram-se a multiplexação de células de duas ou mais conexões, a multiplexação na camada de adaptação (AAL) e a multiplexação com as células de manutenção (fluxo OAM - *Operation and Maintenance*).

2.7 Descritores de Tráfego

Os parâmetros de um tráfego descrevem as características de tráfego de uma conexão ATM. Um parâmetro de tráfego é a especificação de um aspecto particular do tráfego, qualitativo ou quantitativo. Estes parâmetros podem descrever, por exemplo, a taxa de pico de células ou até mesmo o tipo da fonte (telefone, videofone etc.). O descritor de tráfego ATM é uma lista genérica de parâmetros de tráfego que podem ser usados para capturar as características de tráfego intrínsecas de uma conexão ATM [3]. Os descritores de tráfego são necessários para assegurar uma alocação de recursos apropriada e garantir a QoS através de uma rede ATM.

A descrição das características de tráfego que uma dada conexão possa oferecer deve ser fornecida pelo usuário na fase de estabelecimento de conexão.

O descritor de tráfego ATM é composto pelo descritor de tráfego da fonte e pela tolerância à variação do atraso de célula.

2.7.1 Descritor de Tráfego da Fonte

O descritor de tráfego da fonte é o conjunto de parâmetros de tráfego, pertecente ao descritor de tráfego ATM usado na fase de estabelecimento de conexão, que captura as características de tráfego intrínsecas da conexão pedida pela fonte. Os parâmetros atualmente definidos são:

- Taxa de Pico de Células (PCR): representa a taxa máxima do tráfego que pode ser submetido pela fonte na conexão ATM. O inverso do PCR representa a tempo mínimo teórico entre chegadas de células em uma fila;

- Taxa Média de Célula (SCR - *Sustainable Cell Rate*): representa a taxa média de envio de células em um intervalo definido de tempo. O inverso da SCR representa o tempo médio entre chegadas de células com respeito a velocidade do canal;
- Taxa Mínima de Células (MCR) e Taxa Disponível de Célula (ACR): são usados pelo serviço ABR. A MCR representa a taxa mínima de transmissão de células garantida pela rede para a fonte. A ACR representa a taxa atual de transmissão de células fornecida pela rede. Seu valor é variável, mas nunca é inferior ao valor da MCR, estando sendo compreendida entre os valores da MCR e da PCR; e
- Tamanho Máximo da Rajada (MBR - *Maximum Burst Size*): representa o número máximo de células que pode ser transmitido a taxa de pico de células pela fonte enquanto de acordo com a taxa média de células.

2.7.2 Tolerância à Variação do Atraso de Célula

A tolerância à variação do atraso de célula (CDVT - *Cell Delay Variation Tolerance*) é um descritor de rede que fornece a medida de variação (*jitter*) no padrão do tempo entre partidas de células numa conexão. Esta variação ocorre normalmente devido a multiplexação estatística de várias conexões através do canal de transmissão.

Durante o estabelecimento da conexão, a CDVT é adicionada ao descritor de tráfego da fonte o para formar o descritor de tráfego da conexão.

A CDVT é usado juntamente com a PCR e a SCR para assegurar que células que foram geradas pela fonte num intervalo apropriado, mas que sofreram uma variação no atraso devido aos efeitos da multiplexação, continuem sendo vistas como células em conformidade com os seus descritores.