COPPE/UFRJ

Laboratório de Redes de Alta Velocidade -- RAVEL

Modelos de Fontes de Tráfego Para RDSI-FL

Luís Felipe Magalhães de Moraes -- Coordenador

Jorge Roberto Mendes Filho

Relatório Técnico Ravel/02-96

14/10/96

Resumo :

O presente trabalho procura realizar uma apresentação de alguns modelos de fontes de tráfego de voz, dados e vídeo, e sua aplicação à Rede Digital de Serviços Integrados de Faixa Larga. Na introdução são apresentadas particularidades de redes comutadas, critérios para seleção de modelos e algumas características de fontes de tráfego.

No ítem 2 é dada uma visão geral das fontes. A partir do ítem 3, até o ítem 5 são apresentados diversos modelos de fontes usados para caracterização de tráfego de vídeo (ítem 3), voz (ítem 4) e dados/auto-similar (ítem 5).

Modelos de Fontes de Tráfego

1. Introdução [1,10]:

Existem ainda muitas questões a serem resolvidas para que as redes ATM tornem-se realidade. Para projetar, desenvolver e manter o funcionamento apropriado de uma rede de serviços integrados é necessário que as características e as exigências dos diferentes tráfegos transportados por esta rede (voz, dados, vídeo entre outros) estejam bem definidos. Técnicas analíticas, simulações em computadores e projeções feitas a partir da experiência existente são métodos usados para avaliar e comparar os projetos das redes bem como os protocolos que irão fazer parte dela.

Em redes que utilizam a comutação por circuitos, os recursos físicos são alocados estaticamente (por exemplo slots de tempo específicos em um frame TDM) para cada conexão. Durante o estabelecimento da conexão uma rota que liga a fonte ao destino é encontrada e slots livres ao longo do caminho são alocados para esta nova conexão. Caso a rede não disponha de recursos suficientes (no caso slots) a chamada é bloqueada. Redes deste tipo são apropriadas para suportar serviços CBR (constant bit rate) com restrições quanto ao retardo e ao jitter. Entretanto a alocação estatística implica em uma baixa utilização de redes deste tipo quando ela deve suportar tráfego de dados em rajada (neste caso as redes comutadas por pacotes são mais apropriadas).

Nas redes tradicionais de comutação por pacotes os recursos físicos são alocados dinamicamente, em termos de espaço disponível de buffer. Técnicas de feedback ou controle de fluxo por janelas são usadas para assegurar que não heverá "overflow nos buffers". Redes deste tipo são apropriadas a serviços de dados em rajada sem restrições ao retardo e ao jitter.

Em redes ATM a largura de faixa é alocada dinamicamente e virtualmente e um certo grau de serviço deve ser garantido em um contexto probabilístico. Uma estratégia de administração de recursos bem projetada é aquela que faz com que a rede combine vantagens de comutação de circuitos e de pacotes evitando suas desvantagens, o que vem a ser uma tarefa muito difícil. Por exemplo, a alocação dinâmica da banda produz um retardo variável dos pacotes (consideração a ser feita para serviços sensíveis ao jitter, como a voz). Por outro lado, devido à alta velocidade e às longas distâncias mecanismos de controle por janelas resultariam em exigências de tamanho de buffers irreais ou baixa utilização da banda. Uma condição básica para o

funcionamento apropriado de uma rede é o conhecimento da natureza do tráfego que esta irá transportar.

Caracterização da fonte, em um nível mais alto, seria definido como características estatísticas de tráfego da fonte e suas exigências de grau de serviço (QoS). A caracterização de tráfego (descritores de tráfego) de uma aplicação seria um conjunto mínimo de parâmetros (relacionados com as fontes de tráfego) que o usuário declararia à rede para que esta controle o tráfego total submetido pelas fontes e faça o proveito dos recursos disponíveis na rede de maneira eficiente, respeitando as regras definidas para uma rede ATM. Os parâmetros declarados à rede devem ser capazes de servirem como argumentos que caracterizem os processos de geração de células pelas fontes que irão alimentar os multiplex e os comutadores utilizados na rede ATM. De posse destes argumentos a rede passaria a executar mecanismos de controle de admissão de conexões. Alternativamente pode ser especificado apenas o tipo de serviço como declaração implícita de um conjunto de parâmetros de tráfego.

As maiores aplicações da modelagem de fontes de tráfego B-ISDN (vídeo, voz, dados, aplicações multimídias, etc...) estão no controle de admissão de conexões. Podemos definir o controle de admissão como um conjunto de ações executadas pela rede na fase de tentativa de estabelecimento de uma conexão para determinar se esta nova conexão deve ser aceita ou não. A nova conexão será aceita se a rede de comunicação for capaz de suportar esta com um determinado grau de serviço (negociado na tentativa de estabelecimento), e não deteriorar o grau de serviço das conexões já estabelecidas. Por Grau de Serviço ou Qualidade de Serviço (QoS-*Quality of Service*) podemos entender que seja o efeito coletivo do desempenho do serviço e que determina o grau de satisfação do usuário deste serviço. Alguns parâmetros de desempenho são :

- Taxa de perda de células (*CLR- Cell Loss Rate*) : Em se tratando de um ambiente de multiplexação estatística, as células provenientes de várias fontes de tráfego competem por recursos comuns limitados (espaço de buffer no equipamento multiplex). Consequentemente algumas células podem ser perdidas por não encontrarem espaço em buffer. Algumas modalidades de serviço podem tolerar um número moderado de perdas (como serviços de voz), enquanto outras são mais sensíveis à perda de informação (como serviços de transmissão de dados).

- Atraso de transferência de células : Também, neste caso, podemos identificar serviços que são mais sensíveis ao atraso do que outros. Neste contexto podemos

destacar os serviços de voz, uma vez que células deste tipo de serviço devem chegar ao destino dentro de um certo intervalo de tempo, caso contrário serão inúteis. Por outro lado podemos podemos destacar o servço de transmissão de dados que são insensíveis ao atraso. O requisito de atraso restringe o tamanho máximo dos buffers.

- Variação do atraso da célula (*Cell Delay Variation*) : Descreve a variabilidade do atraso de transferência de células. Quando células de várias conexões são multiplexadas, células de uma dada conexão podem ser atrasadas enquanto são inseridas células de uma outra conexão na frente das primeiras.

Como podemos concluir, apartir dos parágrafos acima, caracterização analítica dos processos de geração de células devem ser o mais próximo possível da realidade para que o projeto de redes ATM atendam as espectativas dos usuários e das operadoras de telecomunicações. Quanto mais precisa a caracterização mais precisa serão as estimativas dos parâmetros de desempenho, e o comportamento da rede será mais previsível. As características das fontes de voz têm sido estudadas por várias décadas e são relativamente bem conhecidas. As fontes de vídeo CBR submetem à rede um fluxo contínuo de bits (o que não faz proveito do ganho de multiplexação estatística proposta para redes ATM) e um possível aumento da qualidade exigirá uma taxa maior e consequentemente uma maior banda passante (maiores detalhes serão vistos a seguir). A caracterização de fontes de vídeo VBR e transmissão de imagens VBR são áreas de pesquisa relativamente recentes, e ainda não são conhecidos modelos precisos para preverem o comportamento de tais tráfegos. As redes de pacotes de dados, apesar de serem utilizadas a algumas décadas não tem suas fontes caracterizadas totalmente (para uma dada classe de serviço de dados não são conhecidos comportamentos típicos estatísticos da fonte).

1.1 Critério de Seleção dos Modelos de Tráfego [28] :

A seleção de modelos apropriados para redes ATM é baseada em um conjunto de critérios sumarizados a seguir :

- Proximidade com as fontes de tráfego reais : Um modelo para ser selecionado deve representar da melhor maneira possível a fonte real. A importância deste critério de seleção é mais que óbvia. As principais características estatísticas que influenciarão o comportamento da rede quando alimentada pelas fontes de tráfego devem ser bem representadas pelas estatísticas correspondentes dos modelos. Por exemplo, um importânte parâmetro para fontes de vídeo é a função autocorrelação do

número de células geradas durante períodos sucessivos. Logo um bom modelo para fonte de vídeo deve produzir uma função autocorrelação que aproxime satisfatoriamente uma medida experimenta feita na fonte real.

- Generalidade : O modelo deve ser o mais geral possível. O ideal seria que o modelo aproximasse uma grande variedade de serviços (como voz, transferência de dados, etc...) , mudando apenas alguns parâmetros dentro do modelo para atingir a todo um espectro de características que existem entre os diferentes serviços.

- Simplicidade : O modelo deve ser descrito por um pequeno número de parâmetros. Além disso estes parâmetros devem ser representativos dos fenômenos físicos de uma maneira mais intuitiva possível com por exemplo a taxa média de geração de bits de uma fonte.

- Facilidade de aproximar a fonte real : O modelo pode representar a fonte real por uma certa seleção de parâmetros que compõem o modelo. Isto é realizado, por exemplo, expressando os momentos de certas variáveis aleatórias relacionadas ao modelo em termos de seus parâmetros, fazendo com que estes parâmetros assumam valores tais que os valores dos momentos resultantes aproximem-se o máximo possível dos momentos do experimento relacionado.

- Tratamento analítico e acurácia : Um modelo ideal deve ser fácil de ser analisado. Por exemplo quando analisamos fontes de tráfego multiplexadas (normalmente para estimar a taxa com que células são perdidas) , a análise exata normalmente é muito complexa de ser feita, a partir das características individuais das fontes. Logo o modelo utilizado para representar o tráfego agregado deverá ser acurado e tratável analiticamente. Para outras estimativas de performance (como por exemplo o retardo médio experimentado por uma célula que chega ao multiplex no mesmo sistema) estas considerações devem ser as mesmas porém o modelo utilizado normalmente é diferente.

- Facilidade de implementação : O modelo deve ser fácil de ser implementado em experimentos, seja através de simulação computacional seja através de geradores de tráfego baseados em hardware.

- Adequação para a modelagem tráfego agregado e/ou tráfego de saída : Normalmente é interessante modelar o tráfego agregado de várias fontes visando a análise do ganho de multiplexação estatística (para estudo de economia de largura de

faixa e buffers). Também é necessária a adequação dos modelos que tentam representar o tráfego que surge como saída de outros links. Isto tem grande importância quando estamos analisando redes com vários nós comutadores. Quanto mais adequados estes modelos, mais precisa será a estimativa do comportamento da rede, informação fundamental para aqueles que operam a rede.

1.2 Principais Características das Fontes de Tráfego :

A caracterização da fonte é necessária para a definição precisa do comportamento de cada fonte particular ; isto também provê uma possível administração da rede com habilidade de manipulação da flexibilidade de vários serviços em termos de aceitação de conexão, negociação de qualidade de serviço, controle de congestionamento e alocação de recursos. Em redes ATM existe uma tendência geral em visualizar a geração de células em uma sucessão de períodos de atividade e de silêncio, com a geração de células sendo feita durante os períodos ativos. Um grupo de células sucessivas que não são interrompidas por um período de silêncio é chamado de rajada. Tendo isso em mente a seguir são representados parâmetros para caracterização da fonte :

p : taxa de pico de chegada de células quando a fonte encontra-se no estado ativo ou quantidade máxima de recurso pedido pela fonte à rede.

m : taxa média de chegada de células ou quantidade média de recurso pedida pela fonte à rede.

b : nível de rajada. É definido como a razão entre a taxa de pico de geração de células e a taxa média de geração de células $(\beta = p / m)$, e pode ser visto como medida de duração do período ativo de uma conexão.

Um conjunto equivalente de parâmetros seria :

R_p : valor de pico da taxa de geração de células.

N : número médio de células dentro de uma rajada.

T : intervalo médio entre chegadas de duas rajadas consecutivas, isto é, intervalo entre o início de uma rajada até o início da próxima rajada.

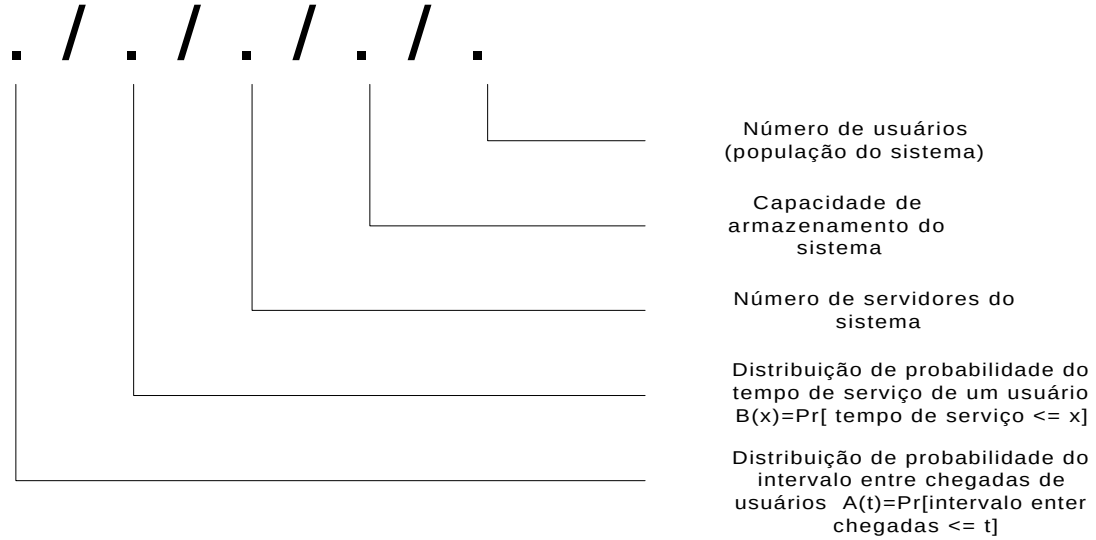
R_{sa} : capacidade de transferência de célula sustentada. É definida como a razão do número máximo de células em uma rajada e o tempo mínimo entre o início de duas rajadas consecutivas.

Passaremos a seguir para uma visão geral dos modelos de fontes, para em seguida estudar algumas de suas aplicações em trabalhos publicados em revistas especializadas .

2. Uma Visão Geral da Modelagem de Tráfego [2] :

No contexto de modelagem de tráfego em redes a ferramenta de maior utilização é a Teoria de Filas, onde o tráfego é oferecido a uma fila ou a uma rede delas e várias medidas de performance são efetuadas, através de metodologias analíticas ou simulações computacionais.

Normalmente usamos a notação de Kendall para formalizar a representação de um modelo de filas. Esta notação está representada a seguir.



O tráfego consiste de chegadas de entidades singulares (células, pacotes, etc...) e pode ser descrito matematicamente através de um processo consistindo de uma sequência de instantes de chegada $T_1, T_2, T_3, ..., T_n, ...$ medidos a partir da origem 0 (zero), por convensão $T_0 = 0$. Duas descrições equivalentes são o Processo de Contagem e o Processo de Intervalos entre Chegadas. O Processo de Contagem $N(t)$ é um processo contínuo onde $N(t)$ é o número de chegadas no intervalo (0,t]. O Processo de Intervalos entre Chegadas é uma sequência aleatória A_n onde

$A_n = T_n - T_{n-1}$ é o tamanho do intervalo que separa a n-ésima chegada da sua antecessora [2].

2.1 Medida de Rajada [2, 12]:

Um tema relacionado a aspectos de tráfego em B-ISDN é a medida de rajada exibida por serviços como vídeo que sofre compressão, tranferência de arquivo, etc...A rajada é observada em um processo de tráfego se os pontos de chegada T_n parecem formar grupos separados (isolados) na escala de tempo; isto é, os intervalos entre chegadas A_n tendem a se apresentar em tamanhos pequenos seguidos de intervalos de tamanhos maiores. Os agrupamentos são causados por características dos serviços e dos protocolos que suportam estes serviços. A variabilidade do processo pode ser obeservada agrupando-se (somando-se) *n* intervalos entre chegadas e comparando-se a variância da série formada com a variância dos intervalos entre chegadas originais. O primeiro valor de variância será maior que *n* vezes o tamanho da segunda medida de variância. Uma outra maneira verificar a existência de rajada é a função de autocorrelação de A_n. Altos valores positivos desta função são consequência do tráfego em rajada. Descreveremos a seguir algumas medidas matemáticas para capturar esta autocorrelação.

A medida mais simples de rajada é razão entre taxa de pico e a taxa média da fonte. Porém esta medida é falha uma vez que é dependente do tamanho do intervalo de tempo utilizado para fazer a medida da taxa. Uma medida mais elaborada seria o coeficiente de variação (c_a) que é a razão entre o desvio padrão e a média dos intervalos entre chegadas .$c_a = \sigma[A_n]/E[A_n]$.

Outras medidas que levam em consideração o tempo são mais elaboradas. Uma delas é o Índice de Disperção para Intervalos, explorado a seguir. Considerando-se a variância da soma de *n* variáveis aleatórias (aqui intervalos entre chegadas de dois pacotes consecutivos) temos :

$$\mathrm{var}\left(A_{i+1} + A_{i+2} + \ldots + A_{i+n}\right) = n \cdot \mathrm{var}(A) + 2\sum_{j=1}^{n-1}\sum_{k=1}^{j}\mathrm{cov}\left(A_j, A_{j+k}\right) \qquad (1)$$

Escrevendo var(*A*) consideramos implicitamente o processo tal que o primeiro e o segundo momento da variável *A* seja independente do tempo (bem como sua média E(*A*)). Juntamente consideramos que a autocovariâcia (ou autocorrelação) cov(A_j,A_{j+k}) dependa somente da distância *k* entre as amostras. Essa dependência da

autocovariância é que faz com que a variância da soma dos intervalos entre chegadas seja útil para a descrição do processo de chegadas. A autocovariância assume valores positivos para intervalos entre chegadas que se afastem da média, tanto para mais quanto para menos. Usaremos então a soma da variância, normalizada pelo fator $n \cdot E^2(A)$ como uma medida da variabilidade do processo de chegada de pacotes. A sequência de valores dada por :

$$J_n = \frac{\text{var}\left(A_{i+1} + A_{i+2} + ... + X_{i+n}\right)}{n \cdot E^2(A)} \qquad (2)$$

com *n=1, 2, ...* é chamada de Índice de Disperção para Intervalos (IDI). Podemos observar que J_1 é o coeficiente de variação ao quadrado dos intervalos entre chegadas. Para um processo de Poisson o IDI vale 1 (um) e para um processo de renovação tem um valor constante igual a J_1 para todo *n*. Usando (1) e a definição $\rho_n = \text{cov}\left(A_i, A_{i+n}\right) \Big/ \text{var}(A)$ podemos expressar (2) em função dos coeficientes de correlação:

$$J_n = J_1\left[1 + 2 \cdot \sum_{j=1}^{n-1}\left(1 - \frac{j}{n}\right)\rho_j\right] \qquad (3)$$

O limite em (3) é dado por (considerando-se ainda processo estacionário):

$$\lim_{n \to \infty} J_n = J_1\left[1 + 2 \cdot \sum_{j=1}^{n-1} \rho_j\right] \qquad (4)$$

Devemos observar que se o processo de chegadas não é estacionário as equações (1), (3) e (4) perdem a generalidade mas a equação (2) continua valendo.

Podemos definir um outro índice observando o processo através da contagem de pacotes que chegam em um certo período de tempo, o Índice de Dispersão para Contagem (IDC). Para um intervalo de tempo τ o IDC é definido como a razão entre a variância e o valor esperado do número de chegadas no intervalo $[0,\tau]$, isto é , $I_c = Var[N(\tau)] / E[N(\tau)]$. Este processo foi definido tal que se considerarmos um processo de Poisson o seu valor será 1 (um). Considerando os instantes de tempo discretos e igualmente espaçados (τ_i , $i \geq 0$) mostraremos outra expressão para o IDC. Indicando c_i como o número de chegadas no intervalo $\tau_i - \tau_{i-1}$ nós temos:

$$I_\tau = \frac{\mathrm{var}\left(\sum_{i=1}^{n} c_i\right)}{E\left(\sum_{i=1}^{n} c_i\right)} = \frac{\mathrm{var}(c_\tau)}{E(c_\tau)}\left[1 + 2 \cdot \sum_{j=1}^{n-1}\left(1 - \frac{j}{n}\right)\ \xi_j\right] \qquad (5)$$

onde ξ_j é o coeficiente de correlação de dos c_i 's com distâncias entre amostras de j .I_t não será contante para processos de renovação nos quais contagens em intervalos disjuntos são correlacionadas, salvo alguns casos, como o de Poisson. Pode ser provado que $\lim_{n\to\infty} J_n = \lim_{t\to\infty} I_t$. Podemos observar ainda que IDC e IDI são adimensionais, não dependendo das unidades utilizadas para calculá-las [2, 12].

2.2 Modelos Utilizando Processos de Renovação :

Estes modelos tem como principal característica a simplicidade matemática e os processsos principais utilizados são de Poisson e de Bernouilli. Em um processo de renovação os valores de $\{A_n\}$ são independentes e identicamente distribuídos porém sua distribuição pode ser geral. Os processos de renovação tem a desvantagem de que a função de autocorrelação de $\{A_n\}$ vai aproximando-se de zero a medida que o tempo tende para infinito, isto é, não captura a correlação da sequência. Com algumas exceções, a superposição de processos de renovação é um novo processo de renovação [2].

2.2.1 Processo de Poisson :

Este processo pode ser caracterizado como um processo de renovação em que os intervalos entre chegadas $\{A_n\}$ são exponencialmente distribuídos com parâmetro λ : $\Pr\{A_n \le t\} = 1 - \exp(-\lambda t)$. Equivalentemente este processo pode ser considerado como um processo de contagem satisfazendo : $\Pr\{N(t) = n\} = \exp(-\lambda t)(\lambda t)^n / n!$. O número de chegadas em intervalos disjuntos são independentes. O processo de Poisson apresenta algumas características analíticas importantes como :

- A superposisão de processos de Poisson resulta em um novo processo de Poisson cuja taxa é a soma das taxas individuais .
- É um processo sem memória , o que facilita os problemas de resolução de filas que envolvem este processo [2].

2.2.2 Processo de Bernoulli :

É um processo em tempo discreto análogo ao processo de Poisson. Aqui a probabilidade de uma chegada em um slot de tempo é *p*, independente entre os slots. A probabilidade de que aconteçam *k* chegadas em *n* slots é dada por : $\Pr\{N_k = n\} = \binom{k}{n} p^n (1-p)^{k-n}$, com n entre *0* e *k* . O tempo entre chegadas é geométrico com parametro *p* : $\Pr\{A_n = j\} = p(1-p)^j$, com *j* sendo um inteiro não negativo .

2.3 Modelos de Tráfego Utilizando Cadeias de Markov :

Os modelos de Markov introduzem dependência entre os elementos da sequência $\{A_n\}$. Consequentemente eles podem "capturar" a rajada do tráfego porque a autocorrelação da sequência é diferente de zero.

Consideremos uma cadeia de Markov de parâmetro contínuo $M = \{M(t)\}_{t=0}^{\infty}$ com espaço de estados discreto . Neste caso *M* comporta-se como se segue : a cadeia permanece no estado *i* por um tempo exponencialmente distribuído com parâmetro λ_i , que depende apenas de *i* . A cadeia então muda para o estado *j* com probabilidade p_{ij} , de acordo com a matriz de taxas infinitesimais de probabilidade $P = [p_{ij}]$. Em um modelo simples cada mudança de estado indicaria uma chegada, então os intervalos entre chegadas seriam exponencialmente distribuídos e os parâmetros de taxa de chegadas seriam dependentes dos estados onde a cadeia estava antes da mudança. Isto resulta na dependência entre intervalos entre chegadas. A figura abaixo, figura 1, representa a idéia descrita acima para uma cadeia com *n* estados.

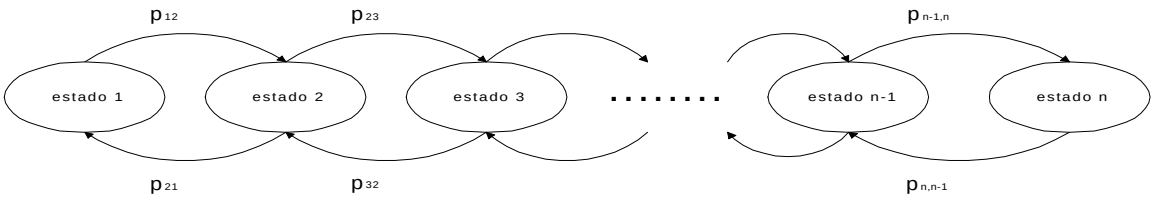


Figura 1

No caso de tempo dividido em slots, o estado *i* representaria *i* slots vazios separando duas chegadas consecutivas e p_{ij} representaria a probabilidade da transição do estado *i* para o estado *j*. As chegadas podem ser unidades singulares, grupos de

unidades ou uma quantidade contínua (uma certa quantidade de carga de trabalho chegando ao sistema).[2]

2.3.1 Modelos de tráfego Modulados por uma Cadeia de Markov [2]:

A idéia aqui é utilizar os estados da cadeia para descrição do "stream" de tráfego enquanto um processo de Markov controla (modula) as leis de probabilidade (uma associada a cada estado). Seja $M = \{M(t)\}_{t=0}^{\infty}$ um processo de Markov em tempo contínuo, com espaço de estados { 1, 2, 3, ..., *m* }. Agora assumimos que enquanto *M* está no estado *k* , a lei que determina o processo de chegadas depende apenas deste estado e isso funciona para todos os estados $1 \le k \le m$. Se ocorre uma transição para outro estado uma nova lei de probabilidade irá reger o processo. A matriz de transição $P = [p_{ij}]$ matém a mesma função anterior, ou seja regular a transição entre os estados da cadeia [2].

2.3.1 Processo de Poisson Modulado por uma Cadeia de Markov (MMPP) [2,8]:

É o modelo mais comum de um processo modulado por uma cadeia de Markov . Neste caso o mecanismo de modulação simplesmente estipula em que estado k de m possíveis (*{1, 2, 3, ..., m }*) as chegadas ocorrem de acordo com um processo de Poisson de taxa λ_k. A medida que o estado em que a cadeia se encontra se altera a taxa do processo de Poisson se altera também.

Como um exemplo do uso do MMPP consideremos uma fonte única de tráfego VBR, cujo processo de geração de pacotes pode ser representado pela quantização em um número finito de taxas exponencialmente distribuídas. Assim cada taxa seria representada por um estado na cadeia de Markov. A matriz de transição entre os estados $Q = [Q_{kj}]$ seria determinada de modo empírico (através de medidas) calculando a fração do tempo em que a cadeia comuta do estado *k* para o estado *j* [2]. Aplicações do MMPP podem ser vistas em [8,9] e será explicada em mais detalhes em 4.3.

2.4 Modelo de Tráfego usando Fluxo de Fluido [2] :

A visão deste tipo de modelo vê o tráfego como um stream de fluido, caracterizado por uma taxa de fluxo (bits / segundo por exemplo). Este modelo é apropriado para casos onde as fontes individuais são numerosas em relação a uma dada escala de tempo. Em outras palavras, uma fonte tem a mesma importância de

uma molécula de água em um fluxo em um cano por exemplo, ou seja, sua contribuição para o fluxo total é infinitesimal. No contexto de B-ISDN, considerando ATM, todas as células tem tamanho fixo de 53 bytes. Considerando-se velocidades de transmissão da ordem de Gigabits o impacto da transmissão de uma célula seria quase desprezível. Por exemplo, contrastando-se o tamanho da célula ATM com um frame de vídeo de alta qualidade comprimido, uma unidade de informação maior que pode conter milhares de células. As chegadas dos frames poderia ser modelada como como chegadas discretas enquanto as chegadas de células poderiam ser modeladas como modelo de fluido.

Uma importante vantagem do modelo de fluxos é observada quando consideramos simulação em computadores. Novamente, considerando-se um cenário ATM, com taxas de transmissão elevadas, a chegada de uma célula a um nó da rede, a granularidade temporal de processamento do evento seria muito pequena e o processamento das chegadas consumiria muito tempo da CPU e recursos de memória. Em contraste, em uma simulação por fluidos, as flutuações de tráfego seriam sinalizadas e aconteceriam mudanças na taxa de fluxo. Em termos de filas a manipulação não seria muito difícil. O tempo de espera na fila seria aquele necessário para esvaziar (servir) o buffer corrente e a probabilidade de perda de células para um buffer finito pode ser calculado em termos de volume de overflow.

2.5 Modelos de Tráfego Autoregressivo :

Os processos de modelagem autoregressivos definem a próxima variável aleatória em uma seqüência como uma função explícita das variáveis aleatórias calculadas anteriormente, dentro de uma janela de um certo tamanho que inicia-se no passado e prolonga-se até o presente . Tal modelagem é utilizada normalmente para representação de fontes de vídeo VBR. A natureza dos quadros de vídeo é tal que quadros sucessivos dentro de uma mesma cena tem uma pequena variabilidade (lembrar que em vídeo de alta qualidade existem aproximadamente 30 quadros por segundo). Somente mudanças de cena podem provocar mudanças bruscas na taxa de bits dos quadros. Logo uma cena de vídeo pode ser modelada por um processo autoregressivo enquanto mudanças de cena podem ser moduladas por um processo modulante (como uma cadeia de Markov por exemplo).

2.5.1 Modelos Autoregressivos Lineares :

Estes modelos tem a seguinte forma :

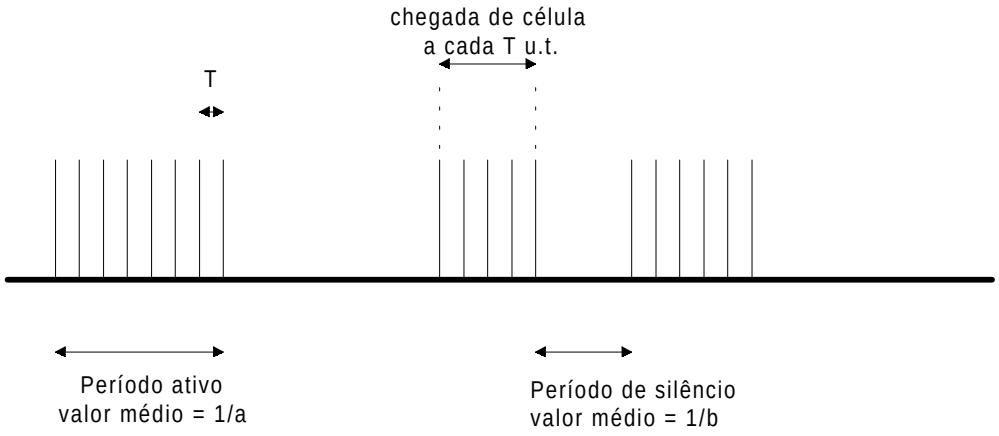
$$X_n = a_0 + \sum_{r=1}^{p} a_r \cdot X_{n-r} + \varepsilon_r, \quad n > 0 \qquad (6)$$

onde $X_0,\ldots,X_{p-1}$ são variáveis aleatórias pré-determinadas, a_r são constantes reais e ε_r são variáveis aleatórias independentes e identicamente distribuídas, chamadas resíduos, que são independentes de X_n . A equação acima descreve a forma mais simples de modelagem autoregressiva, chamada *AR(p)* onde *p* é a ordem deste processo autoregressivo. Podemos observar que as variáveis aleatórias são geradas em função de suas antecedentes dentro de uma seqüência fazendo com que o processo seja indicado para a modelagem de tráfego autocorrelacionado.

2.6 Modelos de fontes ON/OFF :

De acordo com este modelo o fluxo de células de uma fonte de tráfego é modelado como uma sucessão de períodos ativos e períodos de silêncio. A geração de células (ou qualquer outra unidade de informação) ocorre apenas nos períodos de atividade. Normalmente é assumido que os períodos de atividade e de silêncio são independentes entre si e que seus tamanhos são variáveis aleatórias com distribuição exponencial (para tempo contínuo) ou distribuição geométrica (para tempo discreto). Para uma melhor visualização podemos definir :

a^{-1}: valor médio do tamanho do período ativo ;
b^{-1}: valor médio do tamanho do período de silêncio ;
T : intervalo entre geracão de células durante o período ativo ;



As fontes de tráfego do tipo ON/OFF mostradas na figura acima podem ser descritas de duas maneiras alternativas (também equivalentes) através de um conjunto de

parâmetros : (p, m, β, t_{on}) ou (R_p, N, T_i, R_{sa}) que foram apresentados anteriormente. estes parâmetros (juntamente com *a,b,T*) não são independentes uns dos outros. Em particular as seguintes relações são verdadeiras :

$$
\begin{aligned}
& p = 1/T \\
& m = a^{-1} / \left(T \cdot \left(a^{-1} + b^{-1} \right) \right) \\
& \beta = p/m \\
& t_{on} = a^{-1} \\
& R_p = 1/T \\
& N = a^{-1} / T \\
& T_i = a^{-1} + b^{-1} \\
& R_{sa} = R_p
\end{aligned}
\qquad (7)
$$

3. Modelos para Fontes de Vídeo :

3.1 Introdução :

Especula-se que o vídeo digital venha a ser o maior componente de tráfego em redes digitais de faixa larga. Aplicações como video-conferência, videofone, HDTV exigem destas redes uma quantidade relativamente grande de largura de faixa . A instalação de fibras óticas até os usuários em potencial irá encorajar a proliferação de tais serviços, além de outros que surgirão. O uso racional e econômico da banda passante requerida pelos serviços está intimamente ligada e dependente do desenvolvimento de técnicas de compressão de vídeo e dos modelos para as fontes de tráfego de vídeo [3].

A variabilidade da quantidade de informação gerada pode ser vista de duas formas : variações intraframe e variações interframe (variações de intervalo longo e variações de intervalo curto).

As variações de intervalo longo surgem devido às mudanças bruscas de cena. Esta mudança pode provocar uma variação brusca na taxa de geração de bits da fonte. Com o surgimento de uma nova cena mais informação deve ser gerada porque houve uma mudança muito rápida no conteúdo dos quadros.

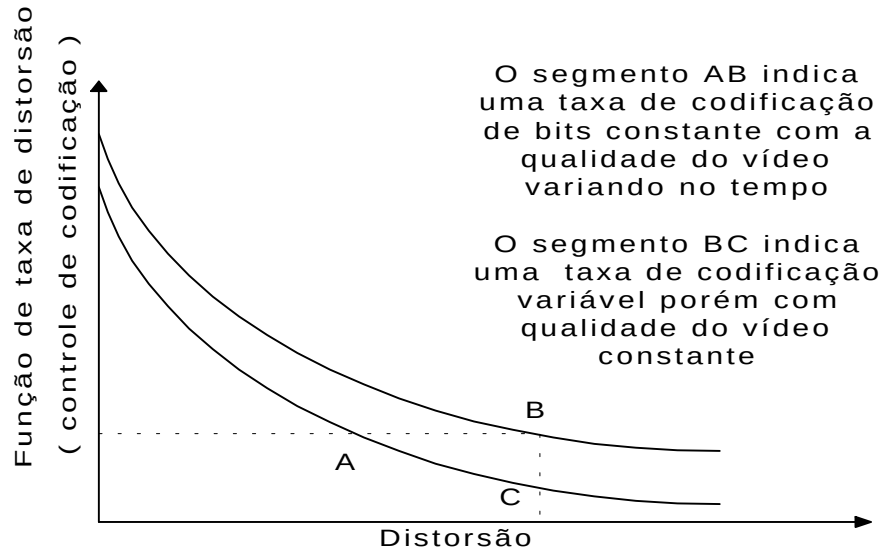
As variações de intervalo curto são provocadas pelas mudanças nas imagens que compõem uma única cena. Estas variações ocorrem de maneira mais suave. As variações intraframe surgem devido ao processamento em separado dos blocos dentro de um frame.

3.2 Características do Sinal de Vídeo :

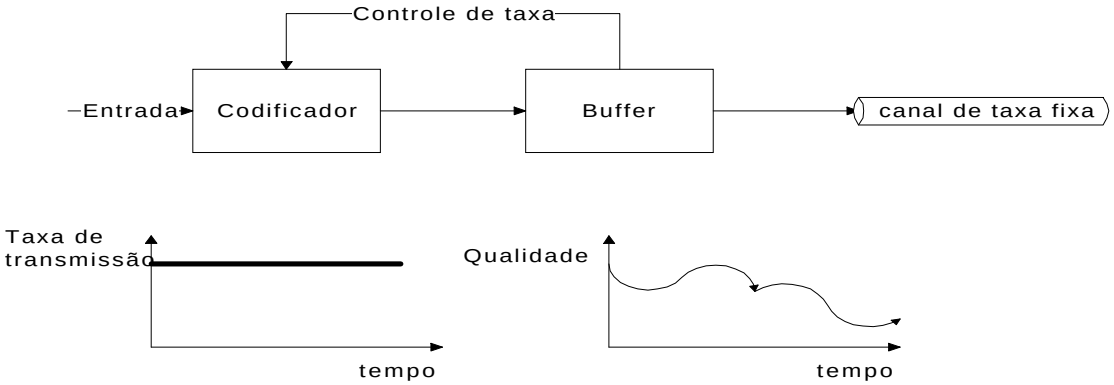
Para fazer uso do ganho de multiplexação estatística usando-se tráfego VBR é necessário ter total domínio das características estatíticas dos sinais de vídeo em redes de pacotes. O sinal original de vídeo apresenta uma grande correlação, que pode variar no tempo, entre imagens consecutivas. O grau de compressão do sinal de vídeo depende da correlação associada entre estas imagens (frames). Logo para um nível fixo de distorsão, a quantidade de informação produzida após a codificação irá variar ao longo do tempo, de acordo com o conteúdo do sinal de vídeo. As características estatísticas do tráfego gerado pelo vídeo podem ser obtidas observando-se a variabilidade da quantidade de informação gerada, bem como as características do codificador que gera esta informação.

3.2 Esquemas de Codificação [3] :

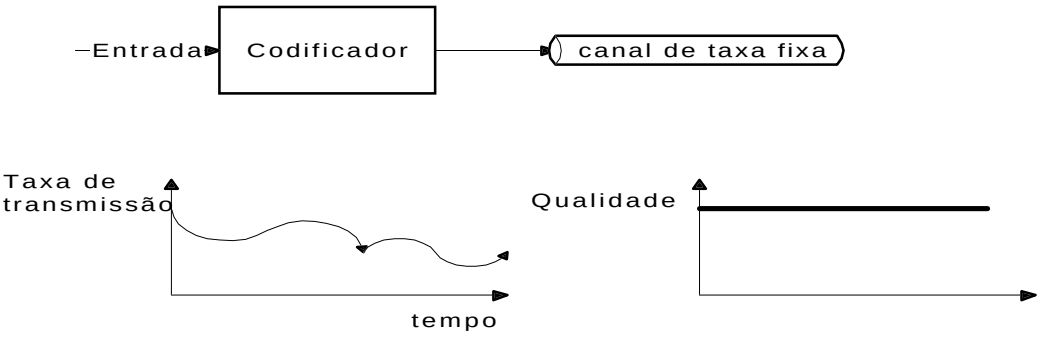
Atualmente os codificadores, que utilizam canais com taxa de transmissão fixa, contém buffer para suavizar as variações da taxa que ocorrem no sinal comprimido, tentando manter a qualidade do vídeo constante (fazendo com que não haja perda de informação). Se a quantidade de informação armazenada no buffer ultrapassa um certo limite o codificador operará em taxas prejudicando a qualidade da imagem. O controle de codificação é feito ajustando-se a distribuição de bits utilizado na compressão (por exemplo mapa de bits na DCT - Discrete Cossine Transform, quanto mais bits utilizados no mapa maior a qualidade da imagem). As restrições quanto ao retardo impõem limite sobre o tamanho do buffer para prevenir a qualidade do serviço. Este compromisso com o tamanho do buffer degrada a qualidade do serviço. Logo teremos um compromisso entre a taxa de codificação (controle de codificação) e qualidade da imagem (distorsão) codificada que pode ser representada no gráfico 1, abaixo [3]:



A figura 2, abaixo, apresenta a idéia do esquema do codificador descrito acima :



Caso haja disposição de um canal com taxa de transmissão variável a situação é diferente. Se a maior taxa que pode ser gerada pelo codificador (entrada) é menor que a taxa máxima do canal não há a necessidade de buffer. Observamos que, se os parâmetros que descrevem a fonte são tratados no estabelecimento da interface entre rede e usuário, não há a necessidade de controle da taxa do codificador e a transmissão do sinal em uma taxa ideal (podendo variar no tempo) é possível e a qualidade do vídeo pode ser mantida constante. A figura 3, abaixo, apresenta a idéia descrita acima :



3.3 Modelos para fonte de vídeo sem mudança abrupta de cena [3]:

Nos modelos que serão apresentados a seguir utilizaremos um multiplex estatístico onde fontes independentes dividem um canal cuja capacidade é menor que a soma das taxas de pico individuais (fontes). As fontes, gerando 30 frames por segundo, compartilham um buffer comum, após um armazenamento em pré-buffers (para pré-suavização da fonte de tráfego). O algoritmo de compressão codifica e transmite a diferença entre os níveis de pixel de frames subsequentes, se esta ultrapassa um certo limiar. O multiplex monta os pacotes a serem transmitidos e identifica-os a partir da fonte para que na demultiplexação cheguem aos seus destinos corretos. Os pacotes são armazenados em filas FIFO (first-in-first-out), sendo transmitidos na mesma ordem que são montados. A figura 4, abaixo, especifica a idéia descrita acima :

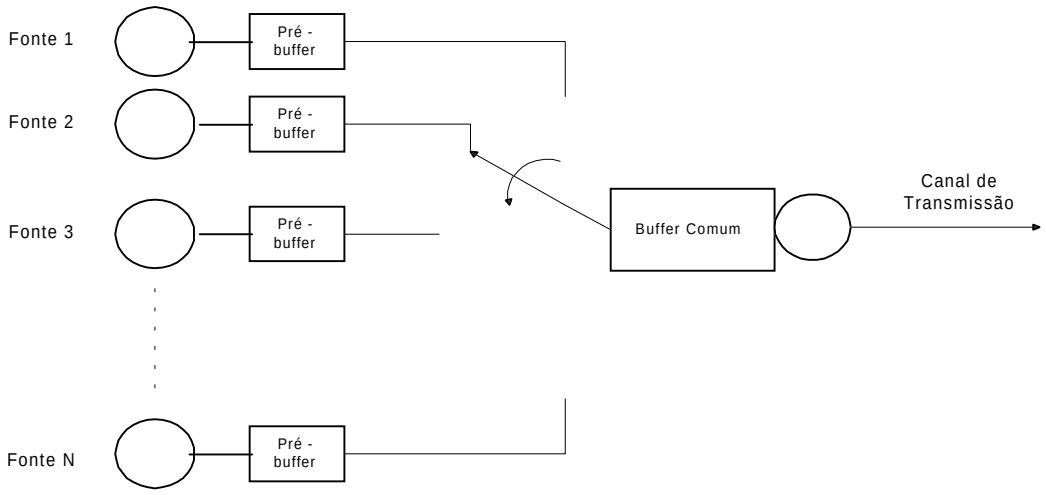


figura 4

A comutação por pacotes é uma extensão natural da multiplexação estatística, com dados de cada fonte sendo segmentados em pequenos pacotes que são armazenados e retransmitidos de comutador a comutador até que cheguem ao seu destino. Tanto a multiplexação estatística como a comutação por pacotes introduzem retardos variados na entrega dos dados devido aos estágios onde há "bufferização". Além disso há perda de pacotes devido ao overflow dos buffers. Logo é muito importante selecionar parâmetros e velocidades de linhas para minimizar estes efeitos.

3.3.1 Caracterização da Fonte de Tráfego a partir de Resultados Experimentais :

Para caracterizar estatisticamente o buffer (a fila) precisamos de um modelo da taxa da fonte codificada. A taxa irá depender do algoritmo de compressão e da

natureza da imagem. Para um vídeo que não exibe movimentos abruptos, como videofone, esperamos uma taxa instantânea com pouca variabilidade e que exibe grande correlação entre os frames.

Um parâmetro importante a ser avaliado seria o número de pixels (N_p) , que combinados em grupo, acarretariam em um stream de N_b bits no pré-buffer. A taxa média seria N_b / N_p bits/pixel . Quanto maior N_p maiores retardos e tamanhos de pré-buffer. Fazendo N_p igual ao número de pixels no frame (aproximadamente 250.000) a taxa da fonte, $\lambda(t)$, em bits/pixel , irá variar depender somente da variação de atividade da sequência de frames.

Medidas experimentais apresentadas em [3] mostraram que para 300 frames (10 segundos de videofone) a taxa média foi de $\mu = 0.52$ bits/pixel e desvio padrão de $\sigma = 0.23$ bits/pixel, com a distribuição das taxas aproximando-se de uma normal. Também foi medida a função de autocovariância da sequência de taxas [$C(\tau) = E\{\lambda(t) \cdot \lambda(\tau + t)\} - \mu^2$], que foi aproximada por uma exponencial $C(\tau) = \sigma^2 \cdot \exp(-a\tau)$ com a=3.9s^{-1}. Este comportamento exponencial foi verificado em outros experimentos citados em [3].

3.3.2 Modelo Autoregressivo de Markov :

Modelaremos a taxa do codificador por um processo estocástico em tempo discreto e estado contínuo. Seja $\lambda(n)$ a taxa de uma fonte durante o n-ésimo frame. O modelo autorregressivo é dado por :(similar ao visto em 2.5)

$$\lambda(n) = a \cdot \lambda(n-1) + b \cdot \omega(n), \qquad (8)$$

onde $\omega(n)$ é uma sequência de variáveis aleatórias Gaussianas independentes e *a* e *b* são constantes. Assumiremos que $\omega(n)$ tem média η e variância igual a 1 e que $|a|<1$ para que o processo seja estacionário para altos valores de n .O valor esperado de λ e a autocovariância discreta C(n) são dadas por [4]:

$$E(\lambda) = \frac{b}{(1-a)}\eta, \quad (9)$$

$$C(n) = \frac{b^2}{1-a^2}a^n, n \geq 0, \quad (10) \quad .$$

A autocovariância, como dito anteriormente, pode ser aproximada por uma exponencial. A distribuição de probabilidade estacionária de λ pode ser aproximada

por uma Gaussiana com média $E(\lambda)$ e variância C(0) como mostra o histograma em [fig. 4 de 3]. Dos dados medidos temos que :

$$E(\lambda)=0.52 \text{ bits / pixel}$$
$$C(n)\approx 0.0536\ (e^{-0.13})^{n} \text{ (bits/pixel)}^2 \quad (11)$$

A autocovariância discreta C(n) é obtida da aproximação experimental $C(\tau) = 0.0536 \cdot \exp(-3.9\tau)$ por amostragem com n/τ = 30 frames/seg. A partir daí temos dados suficientes para calcurarmos os valores de $a\approx 0.8781$, $b\approx 0.1108$ e $\eta\approx$ 0.572 a partir de (9) e (10) . Valores mais precisos poderiam ser obtidos se for utilizado um modelo de grau mais elevado (incluindo-se $\lambda(n-k)$, $k\geq 1$, em (1)), contudo seria achado um somatório de exponenciais em que apenas uma seria dominante sobre as outras. Um trabalho mais complexo ainda utilizando modelos auto-regressivos pode ser encontrado em [22].

3.3.3 Modelagem do Multiplex por um Processo de Markov de parâmetro Contínuo e Estados Discretos :

O modelo apresentado em 3.3.2 é fácil de ser simulado mas não pode ser levado a uma análise de fila de trato razoável. O modelo de Markov com estados discretos possibilita um tratamento analítico mais simples. Aqui a taxa agregada das fontes será discretizada em finitos níveis (estados). As transições entre os estados ocorrem com taxas exponenciais que dependem dos níveis correntes. Os níveis podem ser obtidos por amostragem do processo contínuo em intervalos de tempo aleatórios durante todo o processo, e quantizados em seguida. A aproximação pode ser melhorada diminuindo intervalo entre os níveis de quantização e aumentando a taxa de amostragem. Consideraremos N fontes, cada uma com taxa $\lambda(t)$, taxa média $E(\lambda)$ e autocovariância $C(\tau)\approx C(0)e^{-a\tau}$ no estado estacionário. Para a taxa agregada λ_N teremos :

$$E(\lambda_N) = N \cdot E(\lambda)$$
$$C_N(\tau) = N \cdot C(\tau) \quad , (12)$$

Existe um número enorme de escolhas de processos de Markov que poderiam modelar as equações em (12). A escolha deveria ser tomada observando-se a saída do codificador e atentando-se para a complexidade do modelo. Como não estamos considerando imagens com mudanças abruptas (grandes variações de taxas) de cena, um processo nascimento e morte (transições permitidas somente entre estados

adjacentes ou níveis de quantização) seria apropriado. Considera-se a tendência da taxa aumentar quando estamos em níveis mais baixos e diminuir quando o processo encontra-se em níveis mais altos . Para modelar o processo utilizaremos a figura 5, representada abaixo :

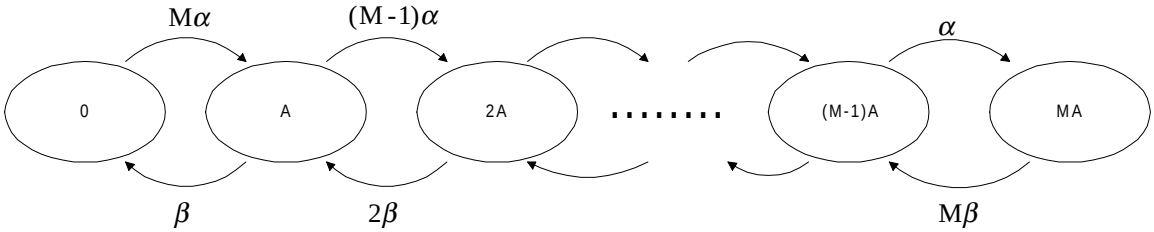


Figura 5

Onde : **A** representa o degrau de quantização (bits/pixel) e **M** +1 níveis possíveis (0, A, …, MA).As taxas exponenciais de transição $r_{i,j}$ do estado iA para o estado jA são dadas por :

$$\begin{aligned} r_{i,i+1} &= (M-i)\alpha \quad , \; i < M \\ r_{i,i-1} &= i\beta \quad i > 0 \\ r_{i,i} &= 0 \\ r_{i,j} &= 0 \quad |i-j| > 1. \end{aligned} \qquad (13)$$

Pode ser mostrado [5 pag. 107] que :

$$\begin{aligned} &P\{\lambda_{\mathrm{N}}(t) = kA\} = \binom{M}{k} p^k (1-p)^{M-k}, \;\; p = \frac{\alpha}{\alpha+\beta} \\ &E(\lambda_{\mathrm{N}}) = MAp \\ &C_{\mathrm{N}}(0) = MA^2 p(1-p) \\ &C_{\mathrm{N}}(\tau) = C_{\mathrm{N}}(0) e^{-(\alpha+\beta)\tau} \end{aligned} \qquad (14)$$

Os parâmetros do modelo *M*, *A*, α e β são obtidos através das equações (6)-(9) e dos dados medidos. Uma vez tendo caracterizado totalmente a fonte de informação podemos utilizá-la como fonte de tráfego em um sistema de filas.

3.3.4 Análise da fila utilizando o modelo de 3.3.3 :

Consideremos uma fila sendo alimentada por uma fonte de taxa $\lambda_N(t)$ bits/seg. Esta fonte pode assumir valores discretos (*0, A, 2A, ..., MA*). Seja $r_{i,j}$ a taxa exponencial de transição do nível discreto da taxa *i* para o nível *j* . A taxa de serviço é de *c* bits/seg . Denotaremos o tamanho da fila por *q(t).* A descrição completa do sistema de fila requer um estado bidimensional { $q(t),\lambda_N(t)$ }. As estatísticas de um estado podem ser descritas por :

$$P_i(t,x) = P\{\lambda_N(t) = iA, q(t) \le x\} \quad (15)$$

Para um sistema estável ($\rho = E(\lambda_N)/c < 1$) o sistema atinge o estado estacionário com a distribuição limite $\lim_{t\to\infty} P_i(x,t) = F_i(x)$, podendo ser descrito pela seguinte equação diferencial :

$$(iA-c)\frac{dF_i(x)}{dx} = \sum_{j\neq i} r_{j,i}F_j(x) - F_i(x)\sum_{j\neq i} r_{i,j}\ ,\ 0 \le i \le M \qquad (16)$$
$$F_i(x) = 0 \quad x < 0$$
$$F_i(0) = 0 \quad \text{para } iA > c$$

A metodologia de solução da equação pode ser vista em [3]. Para o caso do modelo apresentado na figura 5 teremos :

$$(iA-c)\frac{dF_i(x)}{dx} = (M-i+1)\alpha F_{i-1}(x) + (i+1)\beta F_{i+1}(x) - [i\beta + (M-i)\alpha]F_i(x),\ 0 < i < M \quad (17)$$

3.3.5 Conclusões e alguns resultados [3]:

Para a comparação dos modelos apresentados aqui usaremos uma função que representa a fração de dados que chega ao sistema quando este atinge um certo limiar $F(x) = P\{q(t) > x\}$. Para um dado x_0 esta função pode representar a probabilidade de perda de pacotes. Para o modelo apresentado em 3.3.2 a perda de pacotes mostrou-se insensível aos seus tamanhos, utilizando sempre um fator de utilização de 0.8. Gráficos apresentados em [3] também sugerem a insensibilidade do comportamento das filas em relação a distribuição de probabilidade (para 3.3.2 distribuição Gausiana e para 3.3.3 distribuição Binomial) utilizado para caracterizar o estado estacionário. Utilizando $M = 20 \cdot N$ os resultados obtidos para os modelos de 3.3.2 e 3.3.3 são bastante próximos (apesar do esforço computacional ser maior em 3.3.2) . Ainda é apresentado um gráfico em [3] onde os resultados de probabilidade de perda

,aumentando-se M (M =10 seria razoável) , permanecem praticamente os mesmos. Também é observado que a medida que aumentamos o número de fontes sendo multiplexadas (modelo 3.3.3) , mantendo o fator de utilização constante , a probabilidade de perda de pacotes diminui progressivamente, demonstrando que a multiplexação estatística para codificadores de vídeo VBR não irá prejudicar qualidade da imagem recebida. A variação estatística oferecida pelas fontes de tráfego é "compensada" pela multiplexação estatística . Isto pode ser pensado ainda como uma consequência da Lei dos Grandes Números que diz que a medida que o número de fontes aumenta, a taxa agregada tende para a média das fontes e a probabilidade de "bufferização" ou atraso de pacotes diminui além de um certo limite.

3.3.6 Cenas com Variações Rápidas e Lentas : Novas Considerações [6]

Nos itens anteriores consideramos somente tráfegos que possuiam pequenas variações de taxas. Aqui consideraremos a situação de integração de serviços não similares como por exemplo videofone e televisão. Ainda consideramos a codificação variável interframe, formando frames que são divididos em pacotes para posteriormente serem multiplexados. O esquema da figura 4 ainda é válido. Consideraremos taxas na ordem de megabits / seg e pacotes menores que kilobits, possibilitando o descarte da natureza discreta do pacote, tratando-os como fluxo contínuo de bits.

Para redes que integram tráfegos de vídeo diferentes devemos considerar dois tipos diferentes de correlação que regem o processo de geração de dados (taxas em que os dados são gerados) . A correlação de intervalo curto e queda rápida, correspondendo a níveis de atividade uniformes, que perdura por algumas centenas de milisegundos. A correlação de intervalo longo de queda lenta corresponde a mudanças repentinas no grau de atividade da cena (mudanças rápidas em cenas de TV por exemplo) e perdura por alguns segundos. Os modelos apresentados nos itens anteriores capturavam apenas o primeiro tipo de correlação.

O modelo utilizado será ainda um processo de Markov, com o diagrama de transição de estados mostrado na figura 6, representada a seguir.

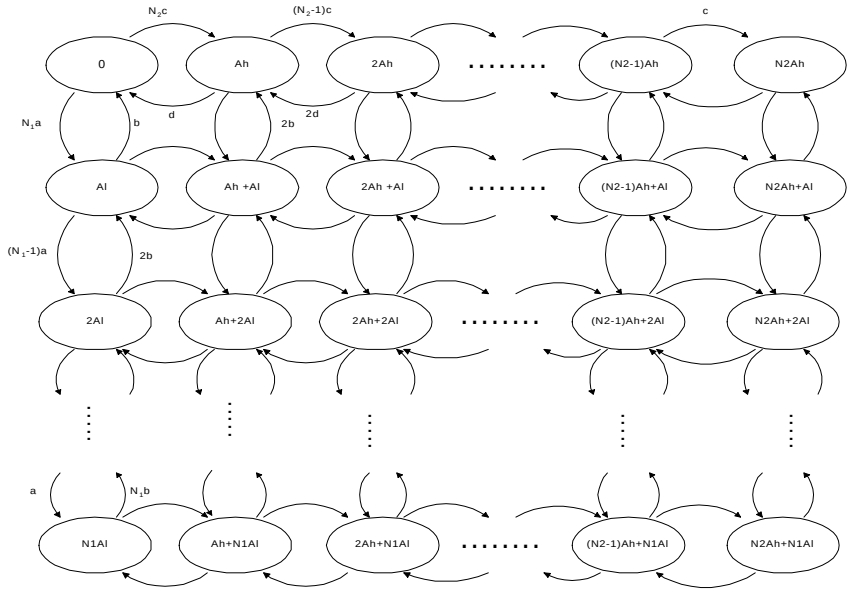


figura 6

O modelo representa uma fonte cujos valores de taxa variam , assumindo valores fixos diferentes em estados diferentes. As taxas correspondentes são as de saída do pré-buffer. As taxas possíveis são combinadas de dois níveis básicos : uma taxa alta *Ah* e uma taxa baixa *Al* , existindo N_1+1 diferentes valores para taxas baixas e N_2+1 diferentes valores para taxas altas.

Quando múltiplas fontes são multiplexadas a taxa de bits agregada pode ser modelada usando-se a mesma estrutura utilizada para a fonte individual (os valores em cada estado serão diferentes). Para este modelo as fontes podem ter comportamentos estatísticos diferentes. Para modelar a taxa agregada a única restrição será de que o comportamento da autocovariância das fontes individuais será aproximado por duas constantes de tempo (uma rápida e outra lenta). Para o modelo de fontes sem mudanças abruptas, apresentado em 3.3.1, apenas uma constante de tempo foi utilizada na caracterização da autocovariância da taxa da fonte.

Passamos agora a descrever a análise de performance da fila em um multiplex onde a taxa instantânea de chegada (fluxo de fluido) é representada por um certo estado no esquema da figura 6. Seja μ a taxa, fixa , de serviço do multiplex e *q(t)* o tamanho instantâneo da fila. A seguinte equação pode representar o comportamento deste sistema :

$$\frac{dF_{i,j}(x)}{dx} = \frac{(N_1 - i + 1)\,a}{\lambda_{i,j} - \mu} F_{i-1,j}(x)$$
$$- \frac{ib + jd + (N_1 - i)a + (N_2 - j)c}{\lambda_{i,j} - \mu} F_{i,j}(x)$$
$$+ \frac{(i+1)b}{\lambda_{i,j} - \mu} F_{i+1,j}(x)$$
$$+ \frac{(N_2 - j + 1)c}{\lambda_{i,j} - \mu} F_{i,j-1}(x)$$
$$+ \frac{(j+1)d}{\lambda_{i,j} - \mu} F_{i,j+1}(x) \qquad (18)$$

Onde $F_{i,j}(x)$ é a probabilidade do processo estar no estado (i , j) e o tamanho da fila no multiplex ser menor ou igual a x . $\lambda_{i,j} = iA_l + jA_h$. A solução para a equação (13) pode ser vista com maiores detalhes em [6].

Como exemplo veremos o modelo usado para representar uma fonte de vídeofone. A figura 7 ,abaixo, representa o modelo.

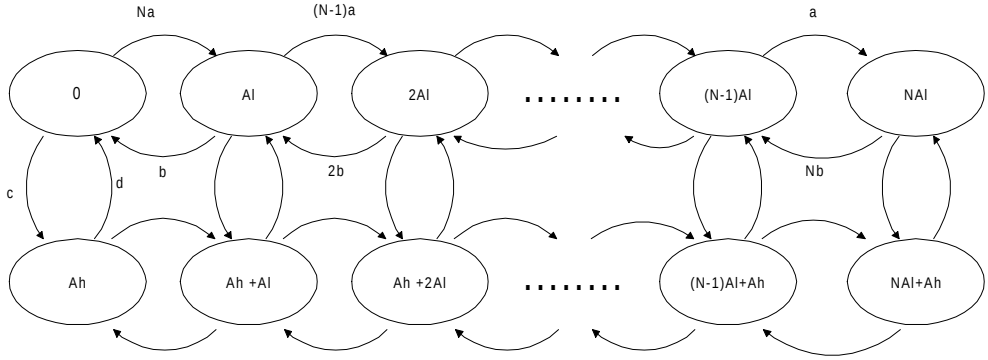


figura 7

A fração do tempo em que o processo permanece no nível de alta atividade e o tempo médio permanecido neste nível serão usados para "amarrar" c e d. Definimos γ como a razão entre a taxa média gerada no nível de alta atividade e a taxa média gerada no nível de baixa atividade. Casando os valores da taxa média ($\bar{\lambda}$), de γ e das estatísticas de segunda ordem em apenas um nível de atividade (função autocovariância $C(\tau)$), determinamos os parâmetros a , b , A_l e A_h .N indica o número de níveis de quantização em cada nível de atividade (no caso acima temos dois níveis). As equações utilizadas para a determinação de a , b , A_l e A_h são :

$$C(\tau) = C(0)e^{-(a+b)\tau} \qquad (19)$$

$$C(0) = Np(1-p)A_l^2 \quad \text{onde} \quad p = \frac{a}{a+b} \qquad (20)$$

$$\gamma = \frac{NpA_l + A_h}{NpA_l} \qquad (21)$$

$$\overline{\lambda} = NpA_l + qA_h \quad \text{onde} \quad q = \frac{c}{c+d} \qquad (22)$$

Para determinação dos parâmetros devemos seguir a seguinte ordem dos procedimentos : da medida real dos dados, a fração do tempo gasto em no nível de alta atividade é dado por *q* e o tempo médio gasto no nível de alta atividade é dado por *1/d*. Isso fixa *c* e *d* . Casando a autocovariância da taxa real da fonte [$C(\tau) = E\{\lambda(t)\cdot\lambda(\tau + t)\} - \overline{\lambda}^2$], γ medido e $\overline{\lambda}$ medido , obtemos os parâmetros desejados com a ajuda de (19)-(22). Quando M fontes são multiplexadas o processo pode ser representado por um processo com a mesma forma da figura 6 com N_2 = M e N_1=MN onde N é de livre escolha.

Experimentalmente, de acordo com os resultados em [6], a taxa média de uma fonte individual foi de 3.9 Mbits/seg, variância de 3.015 $Mbits^2/seg^2$ e o expoente da correlação de termo curto 3.9/s. Para os parâmetros da correlação de longo termo a escolha foi *c = d* (a pessoa fica falando ou ouvindo, na média, por intervalos de tempo relativamente iguais). A taxa de serviço é variada de acordo com uma certa utilização escolhida. γ também pode ser variado (através de N). A medida que o número de fontes multiplexadas aumenta a probabilidade de perda diminui (mantendo-se a utilização constante). Para atingir uma mesma probabilidade de perda de pacotes com uma utilização mais alta (em relação a uma mais baixa, mantendo γ constante) ou com um γ mais alto (em relação a um mais baixo, mantendo a utilização constante) deveremos multiplexar um número maior de fontes [6].

3.4 Modelo para fonte de vídeo baseado no histograma [20] [21]:

Em geral a modelagem ideal seria aquela capaz de lidar com uma grande quantidade de seqüências independentemente de alguns parâmetros como conteúdo de cena e algoritmo de compressão adotado. Cosideraremos neste item que a taxa de bits de uma determinada fonte estão de forma que seja um certo número de bits é gerado para cada frame dentro de uma seqüência de frames. Dentro de um mesmo frame podemos fazer várias considerações (modos) sobre como as células distribuiem-se dentro dele. Ao início de um frame a fonte apresenta um número *N* de células que devem ser transmitidas durante o próximo periodo de frame $1/f$.

Os modos Poisson, uniforme e determinístico são modos mais suaves. Fontes neste modo estão sempre ligadas e sua geração de células varia frame a frame. Ainda podemos considerar o modo rajada em que quando a fonte está ligada há uma geração de bits a uma taxa constante de λ_p .

No modo Poisson a fonte gera células com intervalo de chegadas entre elas exponencialmente distribuídos (logo o numero de células gerado durante um período de frame não é exato mas é distribuido de acordo com um processo de Poisson). No modo uniforme o numero de células gerado dentro de um frame e exato e o intervo de chegadas entre as células tem uma distribuição aproximada de Poisson. Finalmente a fonte no modo determinístico gera células com espaçamento deterministico de tal modo que para aquele frame a última célula é gerada de modo que seu fim coincida com o fim do frame. A tabela a seguir sumariza a idéia acima :

Modo	Periodo ligado (de "on")	Número de células geradas	Distribuição do intervalo entre chegadas
rajada	N/λ_p	N	determinístico igual a $1/\lambda_p$
Poisson	$1/f$	Poisson com média N	exponencial com média igual a $1/(f \cdot N)$
uniforme	$1/f$	N	$\approx$ exponencial com média igual a $1/(f \cdot N)$
determinístico	$1/f$	N	determinístico com média igual a $1/(f \cdot N)$

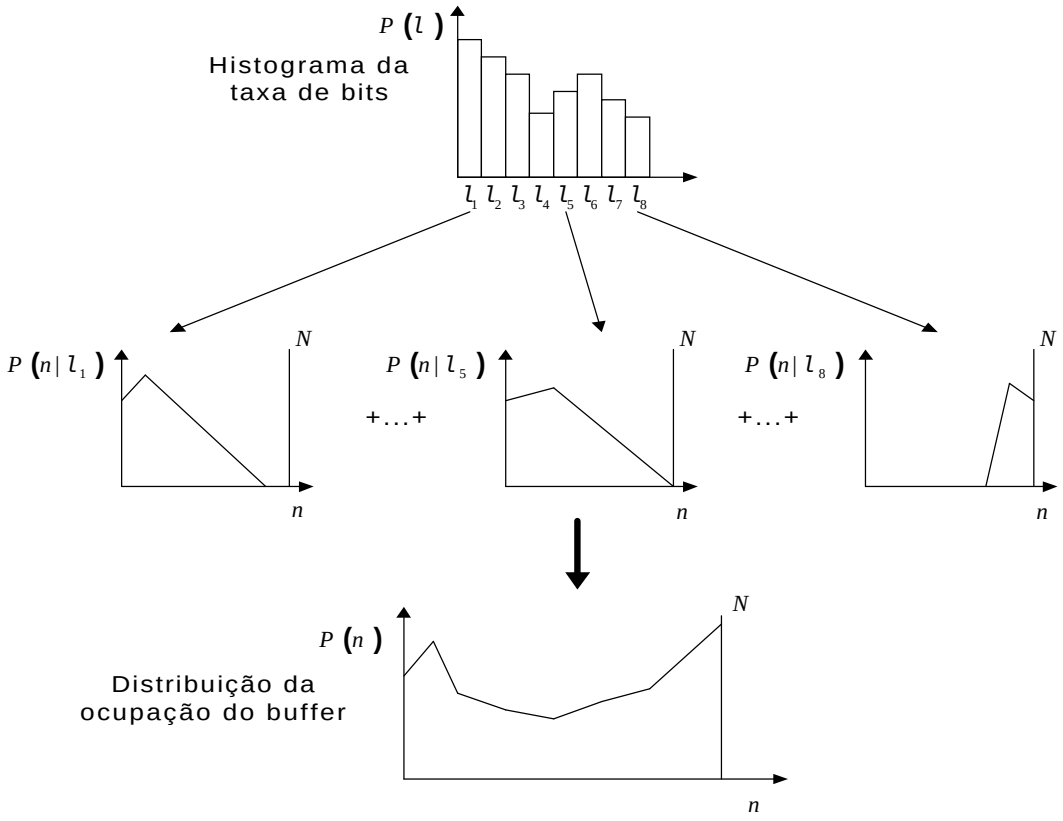
Intuitivamente podemos estimar que, para uma única fonte alimentando um *buffer* de tamanho finito, o modo em rajada seja o pior caso pois deve requerer um tamanho de buffer maior do que aquele requerido pelos modelos mais suaves para absorver a rajada de dados do frame corrente antes que estes dados sejam transmitidos através da rede. Para situações em que temos várias fontes multiplexadas a situação pode ser pior caso haja uma correlação inerente às fontes de vídeo. Para o caso de rajada caso tenhamos todas as fontes sincronizadas e com os mesmos tamanhos de

frame, a sobrecarga no multiplex irá ocorrer sempre, ao início de cada frame. Para questões de modelagem e simplicidade consideremos as fontes não sincronizadas, ocasionando a suavização do tráfego fazendo com que seja menos freqüente a sobrecarga no buffer.

Cosideremos o modo uniforme. Notemos que este tipo de suavização aleatória não pode ser considerada como realista entretanto podemos ter uma visão qualitativa do que ocorre na realidade. Sabemos que se as células de vídeo são distribuidas aleatoriamente dentro do frame com uma distribuição uniforme elas terão um intervalo de chegadas exponencialmente distribuidos aproximadamente. Desde que as células ATM tem tamanho fixo consideremos que elas terão serviço determinístico. Logo podemos considerar o sistema como uma fila *M/D/1/K* frame a frame , com *K* sendo o tamanho do buffer. Olhando para um frame de uma determinada fonte o processo considerado pode ser de Poisson com taxa λ , porém, considerando-se uma seqüência, λ terá uma distribuição $f_\lambda(x)$. Desde que as taxas de chegada e saída das células são muito maiores que a taxa de frames, o sistema chega rapidamente ao estado estacionário, em relação ao tamanho do frame. Através deste sistema podemos observar as características de termo-longo na ocupação do buffer. Por exemplo a ocupação do buffer (dada uma fonte particular alimentando o multiplex) pode ser dada por :

$$P(n) = \sum_{I=1}^{N} \Pr(n|\lambda = \lambda_I) \cdot \Pr(\lambda = \lambda_I), \qquad (23)$$

onde $\Pr(\lambda = \lambda_I)$ é a aproximação do histograma de $f_\lambda(x)$ para a fonte de vídeo, $\Pr(n|\lambda = \lambda_I)$ é a ocupação do buffer dado que a taxa de chegada é λ_I e *N* é o número de intervalos que dividem o histograma da taxa de bits. Isto pode ser visto na figura a seguir.



A modelagem baseada no histograma parece desprezar a correlação intraframe da seqüência. Em [20] são mostradas três seqüências (bits por frame X número do frame na seqüência) diferentes e a distribuição do buffer para os três. A primeira seqüência é a original, a segunda e a terceira são versões reordenadas da primeira. A distribuição do tamanho do buffer são quase as mesmas, não havendo surpresa, uma vez que estamos trabalhando com o mesmo histograma. Entretanto o caso em que uma maior diferença entre a seqüência original e a seqüência modificada é onde é apresentada uma variação rápida de frame para frame. Isso maximiza o transiente em cada novo frame. Conclui-se então que quando assumimos que o sistema atinge o estado estacionário rapidamente no início de cada frame, a taxa de bits de frames adjacentes seja correlacionada. Entretanto a forma como ocorre a correlação entre os frames não é importante.

A análise do tráfego agregado para fontes modeladas pelo histograma de suas taxas não é complicada. O histograma do tráfego agregado é obtido fazendo-se a convolução do histograma das diversas fontes.

3.5 Outros métodos de modelagem de fontes de vídeo :

3.5.1 TES (Transform-expand-sample) [23] [24]:

A principal deste método é que ele pode gerar uma distribuição arbitrária para o número de bits dentro de um frame bem como modelar a estrutura de correlação do frame. Para um conjunto de parâmetros dado, a função autocorrelação do modelo TES é obtido de uma forma fechada. A seguir, através de uma procura sistemática no espaço de parâmetros e computação numérica a função correlação resultante do modelo é possível aproximar a função autocorrelação de uma dada seqüência de vídeo VBR [25] .

3.5.2 Modelagem Auto-Similar [18] :

Será visto ao final do presente trabalho.

4. Modelos para fontes de voz :

4.1 Introdução :

Nos sistemas típicos de voz por pacotes o sinal é digitalizado, codificado e posteriormente os pacotes são formados para em seguida serem transmitidos. O método modulação mais comum é o PCM, onde um sinal analógico é convertido ao formato digital codificado, que representa a amplitude quantizada do sinal analógico original. As técnicas mais tradicionais para sistemas de transmissão de voz recaem em codificação CBR porque tais sistemas não permitem a variação da banda passante alocada para tais aplicações. A qualidade do som depende da taxa de amostra do sinal analógico (número de amostras por unidade de tempo) bem como da resolução da amostra (bits por amostra). A capacidade necessária seria então o produto destas duas quantidades [1].

Uma fonte de voz apresenta períodos ativos e períodos de silêncio. A voz CBR transmite os períodos de silêncio (que não tem nenhum tipo de informação) bem como os períodos ativos, fazendo, consequentemente, o uso ineficiente dos recursos providos pelo sistema. Para uma utilização eficiente do sistema a detecção dos períodos ativos é necessária para que os pacotes sejam gerados somente quando as fontes estão ativas. Ainda para melhorar a eficiência da transmissão novas técnicas de modulação são utilizadas como o DPCM (só é transmitida a diferença entre amostras consecutivas, com esta diferença sendo codificada por um número constante de bits) e o ADPCM (só é transmitida a diferença entre amostras consecutivas, com esta

diferença sendo codificada por um número adaptativo de bits, conforme a diferença entre as amostras).

Desde que as fontes de voz VBR geram pacotes "periodicamente", as propriedades estatísticas do processo de chegadas dos pacotes precisam ser conhecidas e o processo precisa ser modelado para que o projeto de tais sistemas estejam de acordo com os requisitos de retardo e perda de pacotes para a reconstrução da voz no receptor.

Para os modelos de análise de filas em um multiplex estatístico, alimentados por fontes de voz utilizaremos o modelo representado na figura 8 abaixo.

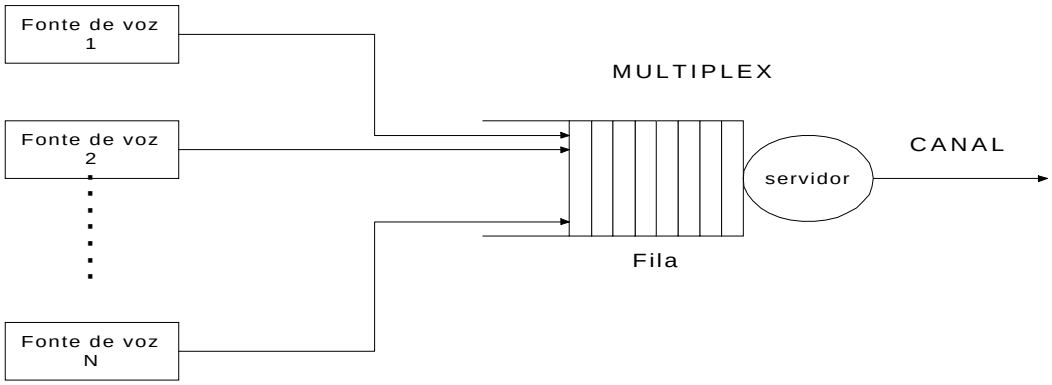


figura 8

O processo de chegada de pacotes ao multiplex é um tanto complexo e pode possuir correlação no número de chegadas em intervalos de tempo adjacentes o que afetará de forma significante o desempenho do multiplex. Até mesmo se o processo de geração de pacotes é aproximado por um processo de renovação, com pacotes deterministicamente espaçados durante o período ativo da fonte seguido por um período de silêncio exponencialmente distribuído, o processo resultante da superposição não é um processo de renovação e a sua análise exata seria intratável, especialmente se o sistema contém buffer finito e mecanismos de controle de sobrecarga. Para a análise do multiplex então são feitas aproximações para modelar a taxa agregada por processos mais simples que serão vistos a seguir.

4.2 Modelos Semi-Markoviano, Markoviano em tempo contínuo e por fluxo de fluido [7]:

Considemos o modelo da fonte como sendo VBR, ou seja, só há geração de pacotes durante o período ativo. Consideraremos N diferentes fintes de voz

independentes, cada uma gerando um pacote a cada *1/V* s quando ativa. Ao ser gerado o pacote entra na fila e o tempo de transmissão por pacote é de *1/VC* s. Logo *C* fontes ativas são necessárias para saturar o sistema. A distribuição dos períodos ativos e de silêncio serão considerados exponenciais, com parâmetros α e β respectivamente. Logo o número de fontes ativas pode ser modelado por uma cadeia de markov de tempo contínuo (phase process) onde o número de fontes ativas é dado por γ. Esta cadeia é mostrada na figura 9. As considerações feitas acima serão válidas para 4.2.1, 4.2.3 e 4.2.3.

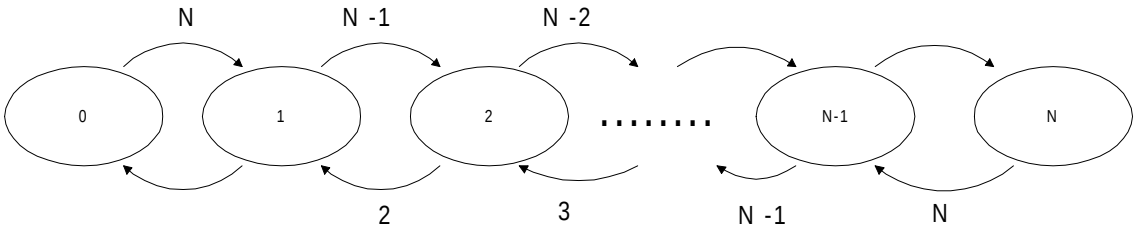


figura 9

4.2.1 Modelagem utilizando um processo Semi-Markoviano :

Neste modelo modelaremos a fila de modo que a taxa de mudança do tamanho da fila é proporcional à diferença entre a capacidade do canal (*C*) e o número de fontes ativas (*j*) . Durante o período em que *j* fontes de voz estão ativas o tamanho da fila altera-se (cresce ou diminui de acordo com a relação entre *j* e *C*) por um número geométrico de pacotes onde o parâmetro desta distribuição geométrica é baseada na diferença entre *j* e *C.* Sob estas condições se o estado do sistema é dado pelo número de fontes ativas e pelo número de pacotes na fila teremos um processo Semi-Markoviano (tempo de permanência no estado não tem uma distribuição de probabilidade exponencial como em um processo de Markov, considerando espaço de estados contínuo).

Para este processo temos que se $\gamma = j = C$ o tamanho da fila não muda. Se temos $\gamma = j < C$ então, se afila não está ocupada por pacotes, ela esvazia-se de um pacote a cada $1 / \left[V\left(j - C \right) \right]$ segundos começando do ponto em que o processo entrou no estado *j* .E por último, se temos $\gamma = j > C$ o tamanho da fila aumenta de um pacote a cada $1 / \left[V\left(j - C \right) \right]$ segundos começando do ponto em que o processo entrou no estado *j* . As aproximações acima têm algumas implicações. Por exemplo, em um sistema real se qualquer fonte gera um pacote durante a transmissão de outro o tamanho da fila aumentará independente da relação entre *j* e *C.* Ainda podemos citar

que após a transmissão de um pacote o tamanho da fila diminuirá, independendo da relação entre *j* e *C* (o modelo não considera estas possibilidades). O modelo assume que um aumento no tamanho da fila ocorrerá somente quando $\gamma > C$ (à direitra da linha em negrito que indica a capacidade do sistema na figura 10) ou a diminuição somente quando $\gamma < C$ (à esquerda da linha em negrito), o que não é verdadeiro . Este modelo claramente irá subestimar a probabilidade da fila estar vazia.

Representaremos o processo como na figura 10 a seguir.

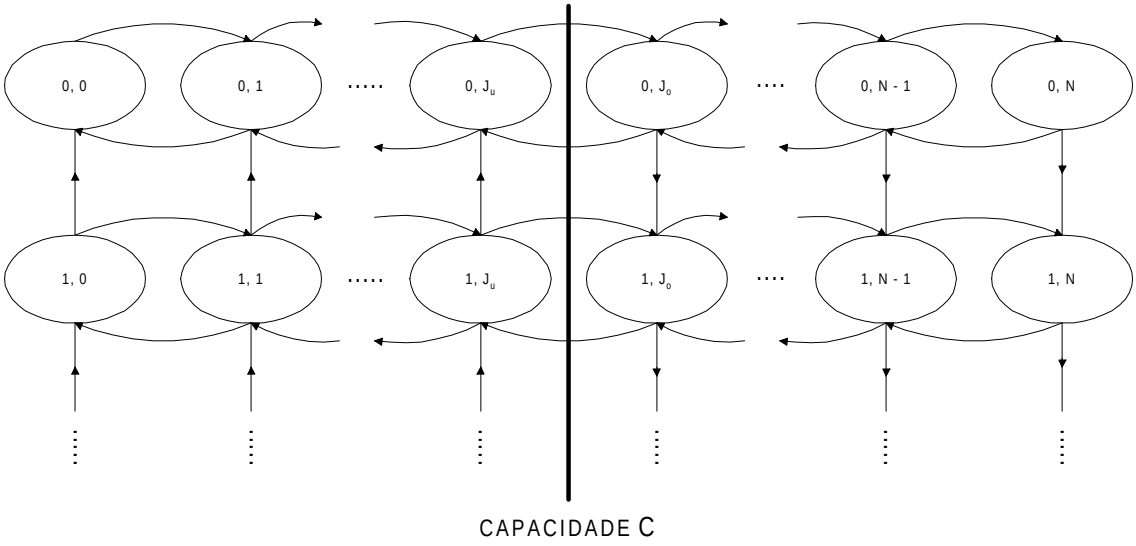


figura 10

Seja $p_{i,j} = \lim_{t\to\infty} \Pr\{l(t) = i,\ \gamma(t) = j\}$, onde ($l(t)$, $\gamma(t)$) representa o estado do diagrama acima (número de pacotes no tamanho da fila no instante *t* e número de fontes ativas no instante *t* respectivamente) . Seja $q_{i\,j}$ a proporção das transições que levam o processo para o estado *(i, j)* e $m_{i\ j}$ o tempo de permanencia do processo no estado *(i, j)* do processo. De [4, pag. 325-326] temos :

$$p_{i,j} = \frac{q_{i,j} \cdot m_{i,j}}{\sum_{i=0}^{\infty}\sum_{l=0}^{N} q_{k,l} \cdot m_{k,l}} \qquad (24)$$

Em [7] temos maiores detalhes da resolução do sistema, a qual consiste da obtenção de π_i que representa a probabilidade de que o tamanho da fila seja *i*, ou seja :

$$\pi_i = \Pr\{l = i\} = \sum_{j=0}^{N} p_{i,j} \qquad (25)\,.$$

4.2.2 Modelo utilizando Cadeia de Markov de Parâmetro Contínuo :

Agora consideraremos que os pacotes, durante os períodos ativos, são gerados de acordo com um processo de Poisson de parâmetro β ao invés da taxa constante apresentada no modelo anterior. A distribuição dos tamanhos dos períodos de silêncio e ativo continuam sendo exponenciais. Como a superposição de Processos de Poisson consiste de um novo Processo de Poisson, quando observarmos j fontes ativas, teremos uma taxa agregada de $j\beta$ para o processo de chegada de pacotes. Similarmente assumiremos que os tempos de serviço são exponencialmente distribuídos com parâmetro ν. Como anteriormente $C = \nu/\beta$ será a capacidade do canal .

O comportamento do modelo sobre um dado período de tempo durante o qual o processo está no estado $\gamma = j$, $j = 0, 1, ..., N$ é idêntico ao comportamento transiente do sistema *M/M/1* com taxa de chegadas de $j\beta$, taxa de serviço de ν e ocupação inicial igula àquela do início do período. Este sistema experimenta maiores flutuações na sua ocupação devido a sua natureza aleatória. Ou seja, se este modelo fosse utilizado para modelar um sistema real, por exemplo uma probabilidade de overflow estimada seria maior que aquela acontecendo no sistema real. O diagrama de estados do sistema pode ser visto na figura 11 abaixo.

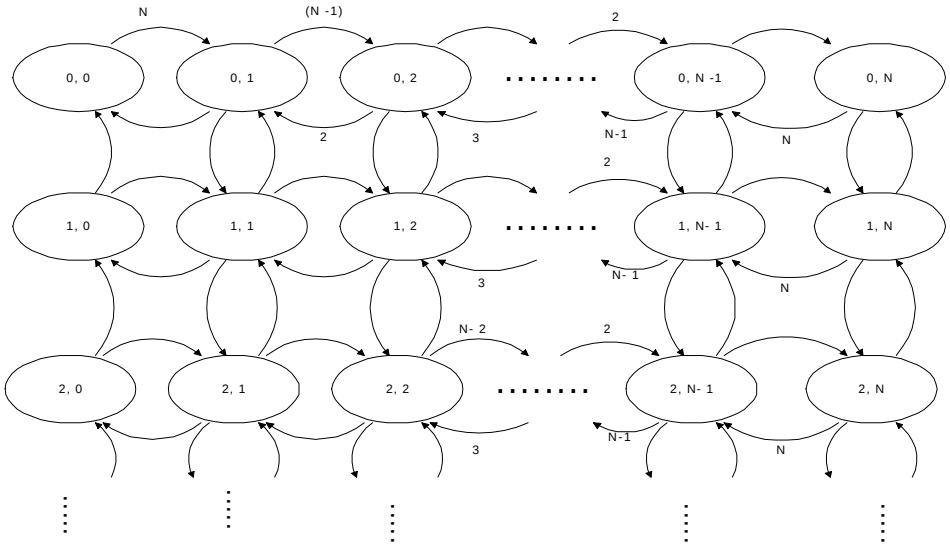


figura 11

O estado desta Cadeia de Markov é denotado por *(s, γ)* onde *s* é o número de pacotes no sistema e γ o número de fontes ativas. Mais uma vez a solução do sistema é dada pelo vetor π_i , ou seja :

$$\pi_i = \Pr\ \{s = i+1\} = \sum_{j=0}^{N} p_{i+1,\,j} \qquad (26)$$

Para maiores detalhes sobre a obtenção de π_i consultar [7].

4.2.3 Modelo por Fluxo de Fluido :

Consideraremos que cada fonte ativa gera informação à taxa uniforme de uma unidade de informação por unidade de tempo e o servidor remove a informação a uma taxa uniforme que não excede a capacidade do canal *C*.

Em um sistema real a unidade de informação não entra no buffer de transmissão (e consequentemente não pode ser transmitida) até que uma fonte particular complete a geração do pacote. Neste modelo, no entanto, é possível que a transmissão esteja sendo feita enquanto a geração ainda não foi concluída. O desempenho deste modelo tende a ser menos preciso a medida que o conteúdo do buffer é menor e que o número de fontes ativas é menor que a capacidade do sistema.

Consideremos que a duração média de um período ativo, 1/α, seja tomada como unidade de tempo. A unidade de informação é tida como aquela que pode ser gerada durante a duração média de um período ativo, ou seja, seria equivalente a V/ α pacotes. Seja *B(t)* e *γ(t)* o tamanho do conteúdo do buffer e o número de fontes ativas em *t*. Definimos $P_i(t,\ b) = \Pr\{\gamma(t) = i,\ B(t) \le b\}$ para $0 \le i \le N,\ t \ge 0,\ b \ge 0$. A solução do sistema é dado pela seguinte equação diferencial , considerando $F_i(b) = \lim_{t\to\infty} P_i(t,\ b)$ e que $F_k(b) = 0$ para $k < 0$ e $k > N$:

$$(i - C)\frac{dF_i(b)}{db} = (N - i + 1)\frac{\lambda}{\alpha}F_{i-1}(b) - \left\{(N - i)\frac{\lambda}{\alpha} + i\right\}\cdot F_i + (i+1)\cdot F_{i+1}(b) \qquad (27).$$

A equação acima só permite transição entre estados adjacentes.Para maiores detalhes sobre a solução da equação 21 consultar [7].

4.2.4 Conclusões e alguns resultados [7] :

Para os três útimos modelos a distribuição do número de pacotes na fila, para um conjunto de parâmetros, foi analisada em [7]. Os parâmetros escolhidos foram $1/\lambda$, a média do tempo inativo da fonte de voz ; $1/\alpha$, a média do tempo ativo; V (pacotes/seg), a taxa de geração de pacotes no período ativo ; N, o número de fontes de voz e C a capacidade do link.

Em todos os testes realizados teremos V= 62.5 pacotes/seg. São escolhidos três conjuntos de parâmetros : PS0 ($1/\lambda$ = 1.65s, $1/\alpha$ = 1.35s), PS1 ($1/\lambda$ = 0.825s $1/\alpha$ = 0.625s) e PS2 ($1/\lambda$ = 3.3s, $1/\alpha$ = 2.7s). São utilizados três conjuntos para observar-se a sensibilidade dos modelos aos tamanhos médios de duração dos períodos ativos e inativos. Assim o processo de geração de pacotes em PS1 será menos "rajada" que em PS0 que por sua vez será menos "rajada" que em PS2.Os parâmetros N e C serão utilizados para variação da utilização do servidor. Os resultados apresentados referem-se ao comportamento da função de distribuição complementar cumulativa, ou seja, a probabilidade do buffer exceder um certo valor ($F(x) = \Pr\{q(t) > x\}$, onde *q(t)* é o tamanho da fila e *x* o número de pacotes).

Resultados em [7] revelam que os modelos apresentados em 4.2.1, 4.2.2 e 4.2.3, usando os parâmetros PS0, prevêem filas maiores que a simulação com o modelo em 4.2.2 afastando-se um pouco mais devido a sua maior variabilidade como discutido anteriormente.

Os modelos apresentados em 4.2.1 e 4.2.3 superestimam a probabilidade da fila estar vazia uma vez que nenhum dos dois modelos prevê um aumento do tamanho das filas quando o número de fontes ativas é menor que a capacidade do sistema. Em um sistema real, para este caso, o tamanho da fila variaria em torno de zero.

Os resultados obtidos para os parâmetros PS1 e PS2 tendem a ser bem próximos aos gráficos apresentados em [7] quando utiliza-se os parâmetros PS0. Também observa-se que *F(x)* estimada pelos modelos analíticos afastam-se da simulação a medida que o número de fontes aumenta.

Simulações também mostram que a medida que o número de fontes aumenta o modelo apresentado em 4.2.1 aproxima-se do comportamento da fila *M/D/1*, assim como o modelo de 4.2.2 aproxima-se da fila *M/M/1*, neste caso com convergência mais lenta. As filas *M/D/1* e *M/M/1* prevêem probabilidades muito menores para um mesmo de tamanho de fila porque não leva em consideração a correlação existente no processo de geração dos pacotes. Esta convergência também é consequência de que a

superposição de um grande número de processos de renovação (os tamanhos dos períodos ativos de cada fonte são exponencialmente distribuídos, formando um processo de renovação) converge para um processo em que a distribuição dos intervalos entre eventos converge para uma distribuição exponencial [4].

Em [29] é utilizada a modelagem por fluxo de fluido porém é considerado o caso em que temos fila de tamanho limitado.

4.3 Modelagem de um processo de chegadas em um Multiplex Estatístico de Voz e Dados por MMPP [8] [9] :

Realizaremos um estudo da modelagem de um multiplex estatístico cuja entrada consiste da superposição de processo de chegadas de pacotes de voz e dados. O processo de chegada agregado em [8] e [9] é aproximado por MMPP (Markov Modulated Poisson Process) de dois estados.

Começaremos modelando uma única fonte através de um processo de renovação. Este processo de renovação será modelado da seguinte forma : quando a fonte está no período ativo o intervalo entre geração dos pacotes é determinístico. Os períodos ativos são seguidos de períodos de silêncio exponencialmente distribuídos. Utilizaremos resultados provenientes de teoria de renovação para obtermos a média, variância, razão variância-média e terceiro momento do número de chegadas em um intervalo de tempo para a superposição de fontes de pacotes de voz. Consideraremos, para fonte única, que o processo de chegadas seja um processo de renovação com distribuição dos intervalos entre chegadas dada por :

$$F(t) = \left[(1-\alpha T) + \alpha T\left(1 - e^{-\beta(t-T)}\right)\right] \cdot U(t-T) \qquad (28)$$

$$\tilde{f}(s) = \int_0^\infty e^{-st} dF(t) = \left[1 - \alpha T + \alpha T\beta / (s+\beta)\right] \cdot e^{-st} \qquad (29)$$

Com *U(t)* sendo o degrau unitário. Logo a taxa média de chegada de pacotes de uma fonte única é dada por :

$$\lambda = \frac{-1}{\tilde{f}'(0)} = \frac{1}{(T + \alpha T / \beta)} \qquad (30)$$

Isto corresponde ao número de pacotes de voz geometricamente distribuídos (com média $1/\alpha T$) durante um período ativo com duração aproximadamente exponencialmente distribuído com média α^{-1} seguido por um período de silêncio exponencialmente distribuído com média β^{-1}. Seja *N(0,t)* o número de pacotes que

chegam considerando-se o processo de renovação estacionário no intervalo *(0,t)*. Definimos então $M_r(t) = E\left[N^r(0,t)\right]$ e seja $\tilde{M}(s) = L\left[M_r(t)\right]$ a transformada de Laplace correspondente.Pode ser mostrado que :

$$\tilde{M}_1(s) = \lambda / s^2 \qquad (31)$$

$$\tilde{M}_2(s) = \lambda / s^2 \left(\frac{1 + \tilde{f}(s)}{1 - \tilde{f}(s)} \right) \qquad (32)$$

$$\tilde{M}_3(s) = \lambda / s^2 \left(\frac{1 + 4 \cdot \tilde{f}(s) + \tilde{f}(s)^2}{\left(1 - \tilde{f}(s)\right)^2} \right) \qquad (33)$$

onde $\tilde{f}(s)$ é a transformada de Laplace da distribuição do intervalo de tempo entre chegadas e λ^{-1} é a media do intervalo entre chegadas. Podemos verificar que $M_1(t) = \lambda \cdot t$. Podemos ainda verificar que o índice de dispersão para contagens satisfaz :

$$\lim_{t\to\infty} I(t) = \lim_{t\to\infty} \frac{\mathrm{var}\left(N(0,t)\right)}{M_1(t)} = \frac{1 - (1 - \alpha T)^2}{\left(\alpha T + \beta T\right)^2} \qquad (34)$$

Os segundo e terceiro momentos do n'umero de chegadas em um intervalo de tempo finito podem ser obtidos pela inversão numérica de $\tilde{M}_2(s)$ e $\tilde{M}_3(s)$ no intervalo de tempo apropriado.

Considerando a superposição dos processos de *n* fontes de voz independentes e denotamos $N_i(0,t)$ o número de pacotes que chegam no intervalo *(0,t)* .O Número de pacotes que chegam na superposição é dado por: $N^S(0,t) = \sum_{i=1}^{n} N_i(0,t)$ e conseqüentemente temos $M_1^S(t) = E\left[N^S(0,t)\right] = n \cdot M_1(t)$ .O indice de dispersão de contagem satisfaz à seguinte expressão :

$$\frac{\mathrm{var}\left[N^S(0,t)\right]}{E\left[N^S(0,t)\right]} = \frac{\mathrm{var}\left[N(0,t)\right]}{E\left[N(0,t)\right]} \qquad (35)$$

onde usamos *S* para denotar a variável de superposição. O terceiro momento central para o processo de superposição é dado por :

$$\mu_3^S(0,t) = E\left\{\left[N^S(0,t) - E\left(N^S(0,t)\right)\right]^3\right\} = n\cdot\left[M_3(t) - 3M_2(t)M_1(t) + 2M_1^3(t)\right] \qquad (36)$$

ou seja as estatísticas consideradas acima para o processo de superposição podem ser obtidas através do processo de fonte única. Estes resultados nos permite calcular a m'edia, a razão variância-média e o terceiro momento central para o processo de superposição considerando-se um intervalo de tempo finito e para um intervalo de tempo infinito a razão variância-média.

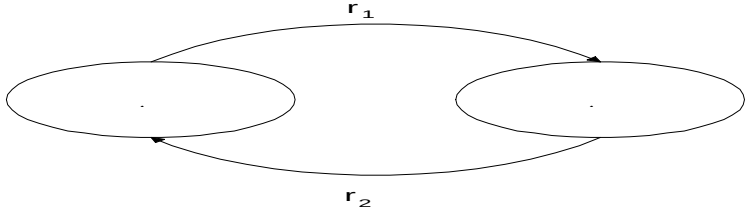
A estrutura de correlação do processo de superposicao pode ser totalmente pela curva variancia-tempo $V^2(t) = \text{var}\left[N^S(0,t)\right]$. Isto é, se :

$$C(t) = \text{cov}\left[N^S(0,t), N^S(t,2\cdot t)\right] \text{ entao}$$

$$C(t) = \frac{V^S(2\cdot t)}{2} - V^S(t) \qquad (37)$$

O que faríamos a seguir seria modelar a fonte agregada pelo MMPP fazendo com que estas estatíticas sejam igualadas. Os parâmetros do MMPP serão obtidos tal que as express~oes da taxa média de chegadas do processo total, do *idc*, do *idc* quando o tempo tende a infinito e a do terceiro momento central para o processo de superposição sejam igualadas às express~oes correspondentes obtidas do MMPP. Esta decis~ao recai sobre o MMPP porque este 'e um processo de tratamento anal'itico relativamente simples e pode ser considerado vers'atil uma vez que seus par^ametros podem ser facilmente modificados.

O MMPP é totalmente caracterizados pelos valores $r_1, r_2, \lambda_1, \lambda_2$ mostrados na figura a seguir:



Para o MMPP podemos escrever as seguintes expressões [8]:

$$E[N_t] = \frac{\lambda_1 r_2 + \lambda_2 r_1}{r_1 + r_2} \cdot t \qquad (38)$$

$$\frac{\mathrm{var}(N_t)}{\bar{N}_t} = 1 + \frac{2(\lambda_1 + \lambda_2)^2 r_1 r_2}{(r_1 + r_2)^2 (\lambda_1 r_2 + \lambda_2 r_1)} - \frac{2(\lambda_1 + \lambda_2)^2 r_1 r_2}{(r_1 + r_2)^3 (\lambda_1 r_2 + \lambda_2 r_1) t} \cdot \left(1 - e^{-(r_1 - r_2)t}\right) \qquad (39)$$

$$\lim_{t \to \infty} \frac{\mathrm{var}(N_t)}{\bar{N}_t} = 1 + \frac{2(\lambda_1 + \lambda_2)^2 r_1 r_2}{(r_1 + r_2)^2 (\lambda_1 r_2 + \lambda_2 r_1)} = b_\infty - 1 \qquad (40)$$

Os parâmetros do MMPP são obtidos através de métodos numéricos apresentados em [8] a partir dos parâmetros das fontes individuais. Obtidos os parâmetros do MMPP podemos estudar por exemplo a distribuição de probabilidade de retardo experimentado pelos pacotes, como feito em [8].

5. Modelos de Tráfego Auto-Similares :

5.1 Introdução [2]:

Estudos recentes de medidas de tráfego com alta-qualidade e alta resolução têm revelado um novo fenômeno com ramificações em potencial para modelagem, projeto e controle da redes de banda larga . Isto inclui análise de alguns milhões de pacotes observados em uma LAN Ethernet em um ambiente de pesquisa e desenvolvimento e análise e observação de alguns milhões de frames de dados provenientes de serviços de vídeo VBR. Nestes estudos o tráfego de pacotes parece ser estatisticamente auto-similar. O fenômeno auto-similar (ou fractal) faz com que o tráfego medido exiba uma estrutura similar sobre várias escalas de tempo. Em tráfego de pacotes, a auto-similaridade manifesta-se independentemente do tamanho das rajadas de dados :em todas as escalas de tempo consideradas (de alguns milisegundos até minutos ou horas), a natureza de rajada do tráfego mostra-se similar. Modelos estocásticos Auto-similares incluem ainda *fractional Gaussian noise* [26] e *fractional ARIMA processes* [27].

A Auto-similaridade manifesta-se de diferentes formas : uma função autocorrelação cujo somatório não tende a um valor finito, a variância da média de amostragem que decresce (como uma função do tamanho da amostra *n*) mais rapidamente que *1/n*, etc... .O parâmetro chave para caracterizar o fenômeno da auto-similaridade é o parâmetro de Hurst *H*, que é utilizado para capturar o grau de auto-similaridade de uma dada seqüência.

Do ponto de vista da matemática, o modelo de tráfego auto-similar difere dos outros modelos de algumas formas : seja *s* uma unidade de tempo representativa de uma escala de tempo, tal que $s = 10^m$ segundos $(m = 0, \pm 1, \pm 2, \ldots)$.Para qualquer escala de tempo *s* seja $X^{(s)} = \{X_n^{(s)}\}$ a representação para as séries de tempo computadas como número de unidades (pacotes, bytes, células, etc...) por unidade de tempo no fluxo de dados. Para modelos de tráfego tradicionais observamos que à medida que *s* aumenta a "agregação" do tráfego tende para uma seqüência de variáveis aleatórias independentes e indenticamente distribuidas similares ao ruído puro (branco). Para modelos auto-similares, repetindo-se esta mesma operação, as seqüências resultantes não se distinguem entre si ("exatamente auto-similar") mas distinguem-se do ruído branco ou convergem para séries de tempo com estruturas de auto-correlação não degenerativa ("assintoticamente auto-similar"). Os modelos tradicionais rapidamente convergem para o ruído branco após o aumento de normalmente duas ou três ordens de grandeza nas escalas de tempo.

Implicações em potencial do tráfego auto-similar em questões relacionadas com projeto, controle e desempenho de redes de alta velocidade vem sendo estudadas. Por exemplo, pode ser mostrado que algumas das medidas de rajada que vem sendo utilizadas não caracterizam a natureza auto-similar do tráfego.

5.2 Processos Auto-similares [17,18]:

Seja $X = (X_t : t = 0,1,2,\ldots)$ um processo estocástico com média μ, variância σ^2 e função autocorrelação $r(k), k \geq 0$. Em particular assumiremos que X tem uma função autocorrelação da seguinte forma :

$$r(k) = \frac{E\{(X_i - \mu)(X_{i+k} - \mu)\}}{E\{(X_i - \mu)^2\}} \sim k^{-\beta} L_1(t), \text{ a medida que } k \to \infty, \tag{41}$$

onde 0 <β<1 e a função $L_1(.)$ é tal que $\lim_{t\to\infty} L_1(tx)/L_1(x) = 1$, para todo $x > 0$. Para cada $m = 1,2,3,\ldots$, seja $X^{(m)} = \left(X_k^{(m)} : k = 1,2,3,\ldots\right)$ um novo processo, com sua correspondente função autocorrelação , obtida pela média de elementos da série original X tomados em blocos de tamanho m que não se sobrepoem. Isto é, para cada m=1,2,3,..., $X^{(m)}$ é dado por $X_k^{(m)} = \frac{1}{m}\left(X_{km-m+1} + \ldots + X_{km}\right), k \geq 1$. Ou seja :

$$X_1^{(1)} = X_1; X_2^{(1)} = X_2; \ldots$$
$$X_1^{(2)} = \frac{1}{2}\left(X_1 + X_2\right); X_2^{(2)} = \frac{1}{2}(X_3 + X_4); \ldots$$
$$X_1^{(3)} = \frac{1}{3}(X_1 + X_2 + X_3); X_2^{(3)} = \frac{1}{3}(X_4 + X_5 + X_6); \ldots$$
$$X_1^{(4)} = \frac{1}{4}(X_1 + X_2 + X_3 + X_4); X_2^{(4)} = \frac{1}{4}(X_5 + X_6 + X_7 + X_8); \ldots$$

O processo X é chamado auto-similar exato de segunda ordem com parâmetro auto-similar $H = 1 - \beta/2$ se para todo $m = 1,2,\ldots$, $\text{var}\left(X^{(m)}\right) = \sigma^2 m^{-\beta}$ e $r^{(m)}(k) = r(k), k \geq 0.$

O processo X é chamado auto-similar assintótico de segunda ordem com parâmentro de auto-similaridade $H = 1 - \beta/2$ se para k grande o bastante,

$$r^{(m)}(k) \to r(k), \text{ a medida que } m \to \infty$$

Com $r(k)$ dado por (1). Em outras palavras, X é exatamente ou assintóticamente auto-similar se o processo agregado correspondente $X^{(m)}$ é o mesmo que X ou assintóticamente é indistingüivel de X (pelo menos com relação à sua função de autocorrelação.

Intuitivamente a característica mais interessante de um processo auto-similar (exato ou assintótico) é que o seu processo agregado possui uma estrutura de correlação não degenerativa, à medida que $m \to \infty$. A estrutura de correlação degenerativa está presente na grande maioria dos modelos de tráfego de pacotes tradicionais, ou seja os seus processos agregados $X^{(m)}$ tendem a um ruído puro, isto é, para todo $k \geq 1$ temos :

$$r^{(m)}(k) \to 0, \text{ a medida que } m \to \infty.$$

Os processos que exibem a propriedade descrita por (1) são ditos possuirem dependência de termo longo *(Long-range dependence-LRD)*, ou seja uma função de autocorrelação que decai hiperbolicamente. Pode ser provado que para processos que exibem esta característica temos $\sum_k r(k) = \infty$. Essa não "somabilidade" das correlações capturam a intuição por trás da *LRD*; enquanto as correlações para grandes valores de *k* são pequenas, seu efeito cumulativo não pode ser desprezado e dá origem a características que são drasticamente diferentes das caracteríticas dos modelos usados tradicionalmente *(*modelos *Short-range dependence-SRD)*. Estes últimos são caracterizados por um decaimento exponencial da função de auto-correlação, isto é :

$$r(k) \sim \rho^k \text{ com } (0 < \rho < 1) \text{ resultando } \sum_k r(k) < \infty$$

5.3 Outras metodologias para descrição de processos auto-similares **[17,18]:**

Para o mesmo processo $X^{(m)} = \left(X_k^{(m)} : k = 1,2,3,\ldots\right)$ descrito no item anterior, pode ser mostrado que a seqüencia $\left(\text{var}\left(X^{(m)}\right) : m \geq 1\right)$ também pode ser usada para a verificação de processos auto-similares. Pode ser mostrado que para processos auto-similares temos :

$$\text{var}\left(X^{(m)}\right) \sim c \cdot m^{-\beta}, \text{ a medida que } m \to \infty$$

com 0 <β<1. Por outro lado, para modelos que não apresentam *LRD* (por exemplo modelos Markovianos), o processo agregado $X^{(m)}$ tende a um processo similar ao ruído branco, ou seja :

$$\text{var}\left(X^{(m)}\right) \sim c \cdot m^{-1}, \text{ a medida que } m \to \infty$$

O efeito de Hurst também pode ser utilizado para a descrição de seqüências auto-similares. Dadas as observações $\left(X_k : k = 1,2,3,\ldots,n\right)$ com média $\bar{X}(n)$ e variância $S^2(n)$ definimos a estatística R/S como :

$$R(n)/S(n) = 1/S(n)\left[\max\left(0, W_1, W_2, W_3, \ldots, W_n\right) - \min\left(0, W_1, W_2, W_3, \ldots, W_n\right)\right]$$

onde $W_k = \left(X_1 + X_2 + X_3 + ... + X_k\right) - k \cdot \bar{X}(n), 1 \le k \le n$. Para seqüências LRD observa-se que :

$$E\left[R(n)/S(n)\right] \sim c \cdot n^H, \text{ a medida que } n \to \infty$$

com o parâmentro de Hurst tipicamente próximo a 0.7. Por outro lado se a seqüência de X_k`s é obtida de um ruído puro Gaussiano, ou uma seqüência *SRD* teremos :

$$E\left[R(n)/S(n)\right] \sim c \cdot n^{0.5}, \text{ a medida que } n \to \infty$$

A discrepância entre as duas últimas expressões é chamada de efeito de Hurst.

6. Conclusão :

A modelagem de fontes de tráfego é um tópico de pesquisa de importância significante para o projeto de redes que transportam serviços de diferentes tipos. No presente trabalho tentou-se sumarizar as principais idéias relacionadas ao assunto bem como a apresentação de algumas idéias propostas em literaturas especializadas.

7. Bibliografia :

[1] Onvural, Raif O. , " Asynchronous Transfer Mode Networks : Performence Issues ", Artech House, inc ,1994.

[2] V.S. Frost e B. Melamed ," Traffic Modeling For Telecommunications Networks", IEEE Communication Magazine, março de 1994, pags 70-81.

[3] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson and J. Robbins , "Performance models of statistical multiplexing in packet vídeo communications ", IEEE Trans. on Commun., vol 36 pags 834-844, july 1988.

[4] Ross, Sheldon M. , " Introduction to Probability Models ", Fourth Edition ,Academic Press, Inc 1989.

[5] L. Kleinrock, Queuering Systems, vol 1 , New York : Willey 1975.

[6] B. Maglaris, D. Anastassiou, P. Sen, N. Rikli, " Models for Packet Switching of Variable-Bit-Rate Video Sourses ", IEEE Journal on Selected Areas in Communications, vol. 7, junho 1989, pags 865-869.

[7] J. N. Diagle e J. D. Langford , " Models for Analysis of Packet Voice Communication Systems ", IEEE Journal on Selected Areas in Communications, vol 6 , setembro 1986, pags 847-855.

[8] H. Heffes e D. M. Lucantoni, "A Markov modulated characterization of packet voice and data traffic and related statistical multiplexer performance ", IEEE Journal on Selected Areas in Communications, vol 4 pags 856-868, setembro de 1986.

[9] R. Nagarajan, J. F. Kurose, Don Towsley, "Approximation Techniques for Computing Packet Loss in Finite-Buffered Voice Multiplexers", IEEE Journal on Selected Areas in Communications, vol 9, pags 368-377, abril de 1991.

[10] Commission of European Communities, "COST 224 Performance evaluation and design of multiservice networks", outubro de 1991.

[11] J. A. Suruagy Monteiro, "Rede Digital de Serviços Integrados de Faixa Larga (RDSI-FL), IX Escola de Computação, Recife, 24 a 31 de julho de 1994.

[12] Gusella, R. , "Characterizing the Variability of Arrival Processes with Indexes of Dispersion", IEEE Journal on Selected Areas in Communications, vol 9, pags 203-211, fevereiro de 1991.

[13] R. Guerin, H. Ahmadi e M. Naghshineh, "Equivalent capacity and its application to bandwidth application in high-speed networks", IEEE Journal on Selected Areas in Communications, vol 9, pags 968-981, setembro de 1991.

[14] Anwar I. Elwalid and Debasis Mitra, "Effective Bandwidth of General Markovian Traffic Sources and Admission Control of High Speed Networks", IEEE/ACM Trasactions on Networking, pags 329-343, junho de 1993.

[15] Gagan L. Choudury, David M. Lucantoni, and Ward With, "Squeezing the Most Out ATM" , IEEE Trans. on Commun., vol. 44,pags 203-217, fevereiro de 1996.

[16] Aurel A. Lazar, Predrag R. Jelenkovic, "On Dependence of Queue Tail Distribution on Multiple Times Scales of ATM Multiplexers"

[17] W. E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson, "On the Self-Similar Nature of Ethernet Traffic (Extended Version)", IEEE/ACM Trasactions on Networking , vol 2, pags 1-16, fevereiro de 1994.

[18] J. Beran, R. Sherman, M. S. Taqqu and W. Willinger, "Long-Range Dependence in Variable-Bit-Rate Video Traffic", IEEE Trans. on Commun., vol. 43, pags 1566-1579 fevereiro/março/abril de 1995.

[19] H. J. Fowler, Will E. Leland, "Local Area Network Traffic Characteristics, with Implications for Broadband Network Congestion Management",IEEE Journal on Selected Areas in Communications, vol. 9, setembro de 1991.

[20] P. Skelly, M. Schwartz e S. Dixit, " A histogram-based model for video traffic behavior in an ATM multiplexer", IEEE/ACM Transactions on Networking, vol 1, pags 446-459, agosto de 1993.

[21] P. Skelly, M. Schwartz e S. Dixit, " A Histogram-Based Model for Video Traffic Behavior in an ATM Network Node with an Application to Congestion Control " , IEEE INFOCOM 1992 pags 95-104, maio de 1992.

[22] R. Gruenefelder, J. P. Cosmas, S. Manthrope e A. Odinma-Okafor , "Characterization of video codecs as autoregressive moving average processes and related queuring system performance", IEEE Journal on Selected Areas in Communications, vol. 9, abril de 1991, pags 284-293.

[23] B. Melamed, D. Raychaudhuri, B. Sengupta e J. Zdepski , "TES-Based Traffic Modeling For Performance Evaluation Of Integraded Networks", IEEE INFOCOM 1992 pags 75-84, maio de 1992.

[24] B. Melamed, D. Raychaudhuri, B. Sengupta e J. Zdepski , "TES-Based Traffic Modeling For Performance Evaluation Of Integraded Networks", Proceedings of the IEEE Global Communications Conference, (Phoenix, Arizona),dezembro de 1993.

[25] Aurel A. Lazar, Giovanni Pacifini e Dimitrios E. Pendarakis, "Modelling Video Sources for Real-Time Scheduling", Department of Electrical Engineering and Center for Telecommunications Research, Columbia University, New York.

[26] B. B. Mandelbrot e J. W. Van Ness, "Fractional Brownian Motions, Fractional Noises and Apllications", SIAM Review, vol. 10, 1968, pags 422-437.

[27] C. W. Grange e R. Joyeux, "An Introduction to Long-Memory Time Series Models and Fractional Differing", Times Series Anal.,vol 1,1980, pags 15-29.

[28] G. D. Stamoulis, M. E. Anagnostou e A. D. Georgantas, "Traffic source models for ATM networks : a survey ", Computer Communications, vol. 17 , número 6, junho de 1994, pags 428-438.

[29] R. C. F. Tucker, "Acurate Method for Analysis of a Packet_Speech Multiplexer with Limited Delay", IEEE Trans. on Commun.,vol 36, número 4, abril de 1988, pags 479-483.