

# Avaliação de um mecanismo de previsão adaptativa de perdas de pacotes com aplicação à transmissão de Voz sobre IP\*

Fabício Murai, Hugo H. Costa Sato,  
Edmundo de Souza e Silva, Daniel R. Figueiredo

<sup>1</sup>Programa de Engenharia de Sistemas e Computação – COPPE/UFRJ  
Caixa Postal: 68.511 – 21941-972 – Rio de Janeiro, RJ – Brasil

{fabricio,hugosato,edmundo,daniel}@land.ufrj.br

***Abstract.** Packet loss directly affects the quality of service provided by multimedia applications in computer networks. These negative effects are even more harmful when the loss process shows great variability. These losses can be masked by using redundancy, however, the quality and efficiency of this mechanism depends on loss process. Thus, an accurate forecast for the loss process in near future can help to improve the performance of redundancy-based mechanisms. and efficiency of this mechanism. In this article, we evaluate different prediction models, such as hidden Markov models (HMM) and autoregressive models (AR), with respect to their capability of forecasting the loss process. We also evaluate how this prediction affects the recovery of loss packets, in terms of recovery rate and overhead. Our results show that recovery performance is directly related to forecast performance.*

***Resumo.** A perda de pacotes é um fator determinante na qualidade de serviço de aplicativos multimídia em redes. Os efeitos negativos são ainda mais graves quando o processo de perda demonstra grande variabilidade. O uso de redundância pode mascarar essas perdas, entretanto, a qualidade e eficiência deste tipo de mecanismo depende do comportamento do processo de perda. Assim, uma boa previsão do processo de perda no futuro próximo pode auxiliar a melhorar o desempenho de mecanismos baseados em redundância. Neste artigo, avaliamos diferentes modelos preditivos, tais como modelos de Markov ocultos (HMM) e modelos auto-regressivos (AR), com relação à capacidade de previsão do processo de perda. Mais importante, avaliamos também o impacto que essa previsão tem na recuperação dos pacotes perdidos, em termos da taxa de recuperação e da quantidade de redundância utilizada. Nossos resultados indicam que a qualidade da recuperação está diretamente ligada à qualidade da previsão.*

## 1. Introdução

Uma das principais causas para a degradação da qualidade de serviço oferecida pelas aplicações multimídia em redes é perda de pacotes, em particular, a variabilidade do processo de perdas. Para oferecer uma melhor qualidade de serviço, as aplicações implementam algum tipo de mecanismo para mascarar os efeitos da perda de pacotes. Um mecanismo frequentemente adotado é o uso de redundância, que consiste em transmitir informação

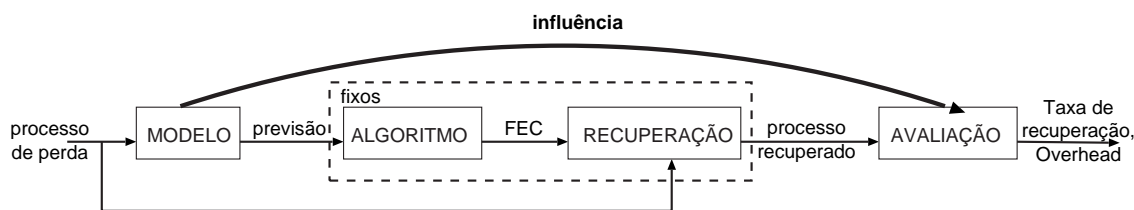
---

\*Este trabalho teve suporte da FAPERJ e do CNPq.

redundante de forma a recuperar pacotes que tenham sido descartados pela rede. Um dos problemas fundamentais deste tipo de mecanismo é decidir a quantidade de informação redundante que deve ser transmitida. Por um lado, a transmissão de uma maior quantidade de informação redundante oferece, geralmente, uma maior capacidade de recuperação no caso de perdas. Por outro lado, se não há perdas, a informação redundante passa a ser um desperdício dos recursos da rede, podendo ter efeitos negativos até mesmo para a aplicação que gera esta redundância.

Idealmente, a quantidade de redundância transmitida pela aplicação deveria ser apenas o suficiente para recuperar as perdas que irão ocorrer. Se o aplicativo soubesse como a rede irá se comportar, então ele poderia se prevenir e utilizar exatamente a quantidade de redundância necessária para realizar a recuperação. Mas como saber como o processo de perdas irá se comportar? Uma possível estratégia é tentar *prever* o comportamento do processo de perdas dentro de um intervalo de tempo futuro. Desta forma, a aplicação poderia se adaptar de maneira pró-ativa, utilizando a redundância adequada para o que ainda vai acontecer.

A Figura 1 ilustra como um modelo de previsão de perdas de pacotes pode ser utilizado por uma aplicação para determinar a quantidade de redundância que deve ser utilizada. A partir das observações do processo de perdas real, o modelo gera previsões acerca das perdas que ocorrerão no futuro próximo. Essas previsões servem de entrada para um algoritmo, que irá selecionar o esquema de redundância que deve ser utilizado. Tal escolha terá um impacto direto sobre os pacotes que serão efetivamente recuperados. Espera-se que quanto melhor a previsão, mais apropriada será a escolha da quantidade de redundância. Desta forma, um bom modelo de previsão do processo de perdas pode garantir uma melhor qualidade de serviço ao usuário e ao mesmo tempo utilizar melhor os recursos da rede (reduzindo a quantidade de redundância a ser transmitida). Na prática, aplicações multimídia tendem a utilizar mecanismos de redundância



**Figura 1. Utilizando um modelo de previsão de perdas de pacotes para determinar a quantidade de redundância a ser transmitida.**

chamados de (*Forward Error Correction*) (**FEC**) e a implementar apenas alguns esquemas pré-estabelecidos. Diferentes esquemas de **FEC** podem possuir a mesma quantidade de redundância (ou não), mas cuja eficácia irá depender do processo de perdas. O problema passa a ser então, utilizar a previsão para escolher um dos esquemas de **FEC** oferecidos pela aplicação [Filho et al. 2006, Duarte et al. 2003, Bolot et al. 1999].

Neste trabalho, abordaremos exatamente o problema de previsão do processo de perda de pacotes e como esta previsão influencia na recuperação dos pacotes perdidos. Basearemos nossos estudos em modelos preditivos, como modelos de Markov ocultos (HMM) e modelos auto-regressivos (AR), e avaliaremos a qualidade do esquema de FEC escolhido em termos da capacidade de recuperação e da quantidade de redundância uti-

lizada. Esta avaliação será feita utilizando *traces* de pacotes reais, obtidos na Internet, e *traces* sintéticos, gerados a partir de modelos de tráfego.

Em particular, utilizaremos o algoritmo adaptativo para parametrizar o modelo de previsão proposto no trabalho de [Filho and de Souza e Silva 2006]. Esta parametrização do modelo preditivo é feita dinamicamente de acordo com as variações do processo de perda real ao longo do tempo. O modelo preditivo é então utilizado para prever as estatísticas de interesse do processo de perdas no futuro próximo, condicionado às observações das perdas passadas.

Os autores de [Filho and de Souza e Silva 2006] utilizam um modelo de Markov oculto (HMM) hierárquico para descrever as características do processo de perdas. A exatidão deste modelo preditivo foi comparada com a de outros modelos de Markov ocultos em trabalhos passados. Neste artigo estendemos o trabalho anterior estabelecendo uma comparação entre este modelo de Markov oculto e outros modelos, como o modelo auto-regressivo, que é freqüentemente utilizado em análise preditiva.

Além da análise de sensibilidade do modelo preditivo, e da comparação com modelos auto-regressivos, a mais importante contribuição deste trabalho é o estudo da influência da qualidade das previsões do processo de perdas no resultado da recuperação (4a. caixa da Figura 1).

Em suma, nosso objetivo é estender a avaliação do uso de um modelo preditivo para o do processo de perdas de pacotes, iniciado em [Filho and de Souza e Silva 2006] de forma a: **1**– avaliar mais detalhadamente a capacidade de previsão do algoritmo adaptativo utilizando os modelos de Markov ocultos, modelos auto-regressivos, e alguns preditores simples. Para isto, iremos considerar algumas medidas de desempenho; **2**– verificar a eficácia do algoritmo de previsão em relação a seus parâmetros com os modelos e cenários aqui considerados; e **3**– comparar o desempenho do modelo HMM e dos modelos auto-regressivos em relação à incerteza da previsão e à taxa de recuperação dos pacotes.

Este trabalho está organizado nas seguintes seções: na segunda seção, faremos uma revisão dos trabalhos relacionados. Na terceira seção, abordaremos resumidamente os principais conceitos teóricos. Na quarta seção, exibiremos os experimentos referentes ao mecanismo de previsão adaptativa. Na quinta seção, estudaremos a aplicação do mecanismo de previsão em uma simulação aplicação de voz sobre IP. Enfim, concluiremos este trabalho na sexta seção.

## **2. Trabalhos relacionados**

A variabilidade espacial e temporal das características da rede [Paxson 1997], tais como a taxa de perda de pacotes e o tempo gasto nas filas dos roteadores intermediários, motivou a crescente aplicação de medições em protocolos adaptativos. No âmbito de controle de congestionamento, diversas propostas de protocolos de transporte realizam medições para ajustar a taxa de envio de dados, [Brakmo and Peterson 1995, Mascolo et al. 2001] são alguns exemplos. Em transmissão de dados multimídia, é freqüente a utilização de medições da taxa de perda para calibrar a quantidade de informação redundante enviada junto aos dados originais, e.g. [Bolot et al. 1999].

Neste contexto de perdas de pacotes, o trabalho de [Yajnik et al. 1999] mostra que os eventos de perdas de pacotes apresentam correlação em largas escalas de tempo. Com

isto, é de se esperar que esquemas de recuperação que façam uso de previsões futuras da taxa de perda futura condicionada ao estado recente apresentem bons resultados. Por exemplo, os autores de [Su et al. 2004] propõem um algoritmo para prever as perdas futuras condicionado às observações modelando o processo de perdas por cadeia de Markov de Gilbert-Elliot simplificado. Entretanto, cadeias de Markov dificilmente representam autocorrelações de larga escala de tempo com baixa complexidade computacional para aplicações em tempo real.

Já em [Salamatian and Vaton 2001], os autores propõem um modelo de Markov oculto onde a série de observações  $\{X_t\}$  é um símbolo binário ( $M = 2$ ) representando a perda do pacote individualmente. Com isto, o número de parâmetros do modelo é proporcional apenas ao número de estados  $N$ . Neste mesmo trabalho, os autores mostram que com poucos estados, em geral 4, o modelo é capaz de representar as autocorrelações significativas da série  $\{X_t\}$  observadas em [Yajnik et al. 1999], mantendo a complexidade baixa. Porém, devido à estrutura específica deste modelo, as suas previsões de estatísticas transientes convergem rapidamente para as respectivas estatísticas estacionárias ( $t \rightarrow \infty$ ).

Em [Duarte 2003], cada observação do modelo de Markov oculto é o número agregado de perdas em um bloco de  $S$  pacotes, ou seja, o número de símbolos possivelmente emitidos é  $M = S + 1$ . Além disto, neste trabalho, o autor considera uma matriz de transição particular motivado por medições realizadas à época. Estudos com este modelo foram realizados em [Filho 2006].

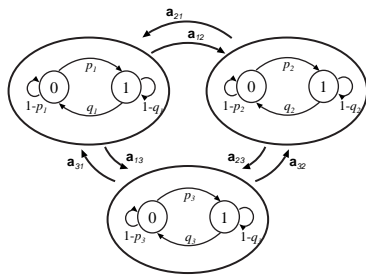
No trabalho de [Filho 2006, Filho and de Souza e Silva 2006], temos um HMM onde cada observação é um vetor de  $S$  valores binários. É fácil notar que o número de possíveis símbolos emitidos é  $2^S$ . Para evitar uma explosão no espaço paramétrico do modelo, o autor considera que as  $S$  observações são geradas a partir de processo Gilbert-Elliot simplificado, ilustrado na Figura 2(a). Assim, o número de parâmetros por estado oculto cai de  $2^S$  para 3. A motivação deste modelo é que, em janelas de tempo pequenas do processo original, as características podem ser capturadas por um processo simples enquanto, em escalas de tempo maiores do processo original, as dinâmicas podem ser capturadas pelas transições dos estados ocultos. A Figura 2(b) mostra a evolução no tempo dos estados ocultos do modelo bem como a emissão de símbolos. Os nossos experimentos utilizarão este modelo por apresentar desempenho superior aos anteriores [Filho 2006].

### 3. Conceitos teóricos

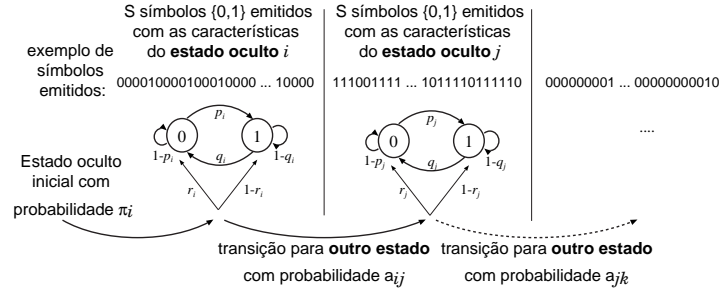
Para analisarmos a estrutura contida no conjunto de medidas e realizar previsões utilizaremos a noção de processo estocásticos. Desta forma, ao observamos a seqüência de medidas  $\mathbf{x} = x_1, \dots, x_n$ , teremos uma realização de um processo  $\{X_t\}$  qualquer. Duas classes de processos estocásticos serão consideradas neste trabalho, os modelos de Markov ocultos [Rabiner 1989] e os modelos auto-regressivos [Brockwell and Davis 2002]. A seguir, descrevemos estes modelos e seus processos de estimação de parâmetros.

#### 3.1. Modelos de Markov ocultos

Modelos Markovianos ocultos, ou simplesmente HMM (**Hidden Markovian Models**), têm sido utilizados extensivamente em análise de séries temporais [Elliot et al. 1995], e.g. reconhecimento de padrões, processamento de sinais, etc.



(a) Exemplo da estrutura do modelo HMM considerado com 3 estados



(b) Análise transiente do modelo HMM considerado

**Figura 2. Modelo de Markov oculto considerado**

Considere um processo estocástico Markoviano  $\{Y_t\}$  em tempo discreto com  $N$  estados e o processo estocástico de observações  $\{X_t\}$  com  $M$  valores. Um modelo de Markov oculto é definido como uma extensão do processo  $\{Y_t\}$  onde o valor observado  $x_t$  é uma **função probabilística** do estado  $y_t$ . Ou seja, os processos estocásticos  $\{X_t\}$  e  $\{Y_t\}$  são dependentes entre si. Neste modelo resultante,  $\{Y_t\}$  deixa de ser observado diretamente, daí o termo oculto. Todo HMM é definido pelos parâmetros:  $N$ ,  $M$ , vetor  $\pi = [\pi_i]_{n \times 1}$  de distribuição do estado inicial, matriz de probabilidade de transição entre estados  $\mathbf{A} = [a_{ij}]_{N \times N}$  e matriz de distribuição probabilística das observações  $\mathbf{B} = [b_{ij}]_{N \times M}$ . Sucintamente, denotamos o HMM pela tripla  $\lambda = (\mathbf{A}, \mathbf{B}, \pi)$ . Assim, para o conjunto de observações  $\mathbf{x} = x_1 \cdots x_n$ , estimaremos  $\lambda = (\mathbf{A}, \mathbf{B}, \pi)$  de forma a obter a “melhor” parametrização possível. O conceito de melhor é dado pela maximização da função de verossimilhança,  $P[\mathbf{x}|\lambda]$ .

Dada a complexidade da função objetivo, não existe uma expressão analítica para este problema. A solução é obtida pelo método iterativo de **Baum-Welch** que não garante a convergência ao ótimo global. Todavia, a cada iteração a função objetivo é garantidamente incrementada, convergindo a um máximo local [Bilmes 1997] para qualquer solução inicial. Na prática, a parametrização inicial é escolhida de forma aleatória.

Primeiramente, calculamos as variáveis auxiliares  $\alpha_t(i) = P(x_1, x_2 \cdots x_t, y_t = i|\lambda)$  e  $\beta_t(i) = P(x_{t+1}, x_{t+2} \cdots x_n | y_t = i, \lambda)$ , para  $1 \leq t \leq n$  e  $1 \leq i \leq N$ . Essas variáveis podem ser calculadas de forma eficiente usando as recursões **forward-backward** (ver detalhes em [Rabiner 1989]). Com esses valores pode-se avançar para o próximo passo, que é o cálculo do número esperado de visitas ao estado  $i$  no tempo  $t$ ,  $\gamma_t(i)$ , e o número esperado de transições de  $i$  para  $j$ ,  $\xi_t(i, j)$ .

$$\gamma_t(i) = P(Y_t = i|\mathbf{X}, \lambda) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^N \alpha_t(j)\beta_t(j)}, \quad (1a)$$

$$\xi_t(i, j) = P(Y_t = i, Y_{t+1} = j|\mathbf{X}, \lambda) = \frac{\alpha_t(i)a_{ij}b_{jx_{t+1}}\beta_{t+1}(j)}{\sum_{j=1}^N \alpha_t(j)\beta_t(j)}. \quad (1b)$$

Para a próxima iteração, reestimamos  $\lambda$  através das seguintes equações:

$$\bar{\pi}_i = \gamma_1(t) \quad \bar{a}_{ij} = \frac{\sum_{t=1}^{n-1} \xi_t(i, j)}{\sum_{t=1}^{n-1} \gamma_t(i)} \quad \bar{b}_{jk} = \frac{\sum_{t=1: x_t=k}^n \gamma_t(j)}{\sum_{t=1}^n \gamma_t(j)} \quad (2)$$

Com os novos parâmetros, a função de verossimilhança dado o modelo reestimado é calculada. Se o ganho relativo à iteração anterior for pequeno pode-se parar. Também é possível parar o algoritmo pelo número de iterações, principalmente se este algoritmo estiver sendo usado em um aplicativo de tempo real.

### 3.2. Modelos auto-regressivos

Uma subclasse de processos lineares já bem utilizada em problemas de análise de séries temporais são os modelos auto-regressivos ou simplesmente **AR**. A definição formal de um modelo AR por [Brockwell and Davis 2002] é

**Definição 3.1.** *Uma série temporal  $\{X_t\}$  é um modelo AR( $p$ ) se é estacionária e satisfaz a seguinte equação para todo  $t$*

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t \quad (3)$$

onde a série  $\{Z_t\}$  é independente e identicamente distribuída  $N(0, \sigma^2)$ .

Seja a função de autocovariância de um processo estocástico  $X_t$  de lag  $h$  dada por  $\gamma_X(h) = Cov(X_h, X_0) = E[(X_h - E[X_h])(X_0 - E[X_0])]$ . Para um conjunto de observações  $\mathbf{x} = x_1 \dots x_n$ , estimaremos  $\phi = (\phi_1, \dots, \phi_p)$  e  $\sigma^2$  de forma que a função de autocovariância do modelo  $\gamma_X(\cdot)$ , ou seja, as suas características de segunda ordem, se aproximem à das observações  $\hat{\gamma}_X(\cdot)$ .

Para resolver este problema, apresentaremos o método de estimação de Yule-Walker. A idéia deste método é utilizar a autocovariância amostral  $\hat{\gamma}(h)$  para determinar os parâmetros  $\phi$  e  $\sigma^2$ . Para isto, manipulando algebricamente a Definição 3.1, teremos as seguintes fórmulas de estimação:

$$\begin{aligned} \phi &= \mathbf{\Gamma}_p^{-1} \boldsymbol{\gamma}_p, & \boldsymbol{\gamma}_p &= (\gamma_X(1), \dots, \gamma_X(p))', \\ \sigma^2 &= \gamma(0) - \phi \boldsymbol{\gamma}_p. & \mathbf{\Gamma}_p &= [\gamma_X(i-j)]_{p \times p}. \end{aligned} \quad (4)$$

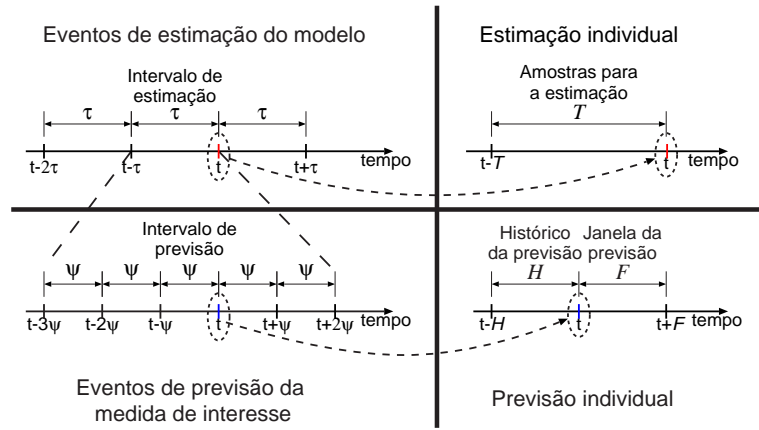
Desta forma, basta utilizar  $\hat{\gamma}(h)$  nas Equações de 4 para obter os coeficientes do modelo auto-regressivo resolvendo o sistema de equações lineares. Este conjunto de equações é freqüentemente chamado de estimadores de Yule-Walker.

### 3.3. Mecanismo de previsão adaptativa

Neste trabalho, supomos que o estado da rede é observado indiretamente através dos eventos de perda na transmissão de cada pacote individualmente. Isto é, cada elemento do processo  $\{X_t\}$  assumirá valor 0 (1) caso o pacote transmitido no tempo  $t$  seja transmitido (perdido). Consideremos que, em uma janela com  $F$  observações, o estado da rede é caracterizado pela sua taxa de perda média  $R_t^F = \frac{1}{F} \sum_{f=0}^{F-1} X_{t+f}$  e pelo seu tamanho médio da rajada de perdas  $B_t^F = \frac{1}{n} \sum_{i=0}^{n-1} Z_i$  onde  $\{Z_t\}$  é a série que representa o tamanho das  $n$  rajadas de perdas determinada por  $\{X_t\}$  que ocorrem no intervalo discreto  $\{t, t +$

$F - 1$ }. Nos restringimos a estas estatísticas do processo de perda porque as aplicaremos no mecanismo de recuperação de perdas de pacotes (**FEC**).

Em [Filho and de Souza e Silva 2006], os autores propuseram um mecanismo de previsão adaptativa utilizando HMM. Neste trabalho, consideramos o uso do mecanismo também para modelos auto-regressivos. *Note que, para ambos os modelos, a previsão é dita adaptativa porque, em intervalos de  $\tau$  segundos, os parâmetros do modelo são reestimados de forma ótima, utilizando as  $\tau$  últimas observações de  $\{X_t\}$ .* Essa reestimação é fundamental, pois a seqüência de perdas, em geral, não é estacionária. Enquanto isso, a previsão para uma janela de tempo futuro de  $F$  segundos é feita em intervalos de  $\psi$  segundos, condicionada ao passado traduzido pelas últimas  $H$  observações de  $\{X_t\}$ . A Figura 3 ilustra como estes eventos ocorrem ao longo do tempo e os respectivos parâmetros.



**Figura 3. Mecanismo de previsão adaptativa em função dos parâmetros de estimação ( $T, \tau$ ) e de previsão ( $\psi, H, F$ )**

### Previsão com HMM e AR

Com um único modelo HMM, em cada evento de previsão teremos uma estimativa da taxa de perda média e do tamanho médio da rajada de perdas futuras a partir da análise transiente do HMM condicionada às  $H$  últimas amostras. Para obter a previsão da taxa de perdas  $R_t^F$ , faremos uma análise transiente do modelo HMM com estado inicial condicionado aos eventos de perdas  $X_{t-H}, \dots, X_{t-1}$ . Assim teremos uma estimativa da distribuição probabilística de  $R_t^F$ . Isto é, para calcular a distribuição de  $R_t^F$  condicionada às últimas observações, temos a seguinte equação:

$$P(R_t^F = j | X_{t-H}, \dots, X_{t-1}) = \sum_{\forall y_{t-1}} P(R_t^F = j | Y_{t-1}) P(Y_{t-1} | X_{t-H}, \dots, X_{t-1}). \quad (5)$$

Em [Filho 2006], o autor apresenta um algoritmo recursivo para calcular esta distribuição da segunda etapa utilizando apenas operações matriciais. De posse da estimativa de  $P(R_t^F = j | X_{t-H}, \dots, X_{t-1})$ , utilizaremos como previsão  $\hat{R}_t^F = E[R_t^F | X_{t-H}, \dots, X_{t-1}]$ .

Para estimar o tamanho médio da rajada, iremos utilizar a seguinte aproximação:

$$\hat{B}_t^F = \sum_{i=1}^N P(Y_t = i | X_{t-H}, \dots, X_{t-1}) E[R_t^F | Y_t = i] \quad (6)$$

Já usando modelos auto-regressivos, é necessário manter dois modelos em paralelo, um para descrever a taxa de perda ao longo do tempo e a outro para o tamanho médio da rajada de perdas. Denotando a série da taxa de perda e tamanho médio da rajada de perda por  $\{R_t\}$  e  $\{B_t\}$ , temos a estimativa da previsão dada pelas equações:

$$\hat{R}_{t+1} = \mu_R + \phi_1^R(R_t - \mu_R) + \dots + \phi_p^R(R_{t-p+1} - \mu_R) \quad (7a)$$

$$\hat{B}_{t+1} = \mu_B + \phi_1^B(B_t - \mu_B) + \dots + \phi_p^B(B_{t-p+1} - \mu_B) \quad (7b)$$

Neste, o histórico de previsão  $H$  é representado pela ordem de auto-regressão  $p$ .

#### 4. Avaliação da capacidade de previsão

Neste conjunto de experimentos, avaliaremos a capacidade de previsão do mecanismo de previsão adaptativa em um conjunto de medições reais obtidas na *Internet*.

##### 4.1. Metodologia

Para avaliar o mecanismo de previsão adaptativo, utilizaremos um conjunto de medições de perdas de pacotes fim-a-fim. Estas medições foram obtidas com auxílio do módulo *Traffic Generator* da ferramenta Tangram-II [de Souza e Silva et al. 2006] e utilizadas em [Duarte 2003, Duarte et al. 2003] e depois em [Filho and de Souza e Silva 2006, Filho et al. 2006].

A cada intervalo de tempo de 20 milissegundos, um pacote com 324 *bytes* de dados é enviado pelo gerador de tráfego para um destino arbitrado – somando os cabeçalhos do par UDP-IP (8+20 *bytes*) teremos uma taxa constante de 140.80 kbps, desprezível se comparada com a capacidade dos canais envolvidos. No destinatário, são feitos os registros (*trace*) dos pacotes perdidos (1) e recebidos (0) compondo nosso processo de perdas  $\{X_t\}$ . A motivação era emular uma aplicação de voz sobre IP (VoIP) simplificada com codificação **PCM** na *Internet*.

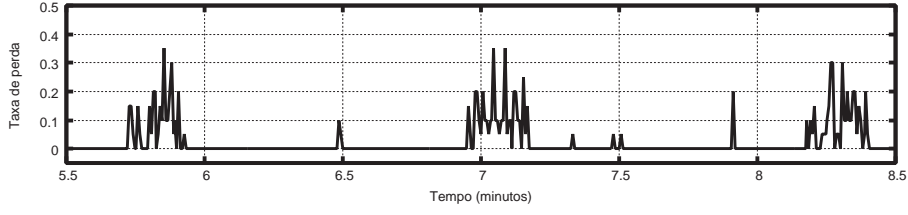
Dentre todo o conjunto de amostras, analisaremos três *traces* que apresentavam algumas características representativas na taxa de perda. Em geral, observamos variações periódicas na taxa de perda (Fig. 4(a)), variabilidade intensa (Fig. 4(b)) e taxas de perdas constantes com altas variações de curto tempo (Fig. 4(c)). Os *traces* selecionados possuem autocorrelação significativa (Fig. 5), justificando o uso de modelos. Para verificar a capacidade de previsão da taxa de perda, definiremos um conjunto de medidas quantitativas. Antes disto, notemos que o desempenho da aplicação aumenta com a sua capacidade de prever variações da taxa de perda. Entretanto, se olharmos o caso onde a taxa de perda  $R_t^F$  é nula durante um longo período de tempo com elevações bruscas de curta duração (picos), um preditor  $\hat{R}_t^F \equiv 0$  seria muito bom de acordo com as medidas acima mas seria inútil para a aplicação. Evitaremos estas situações particionando o tempo de observação em instantes que apresentam ou não uma variabilidade arbitrada. A taxa de perda em instante de tempo  $t$  será dita **variante** quando

$$R_t^F \notin \bigcup_{i=1..k} [R_{t-i}^F - \delta, R_{t-i}^F + \delta]$$

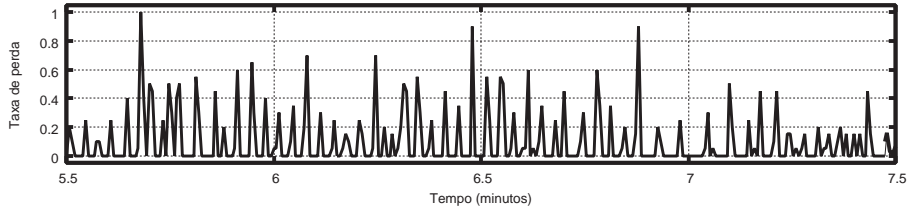
ou **invariante** caso contrário.

Considerando apenas a previsão e a taxa de perda real nos instantes variantes, definimos as seguintes medidas de desempenho:

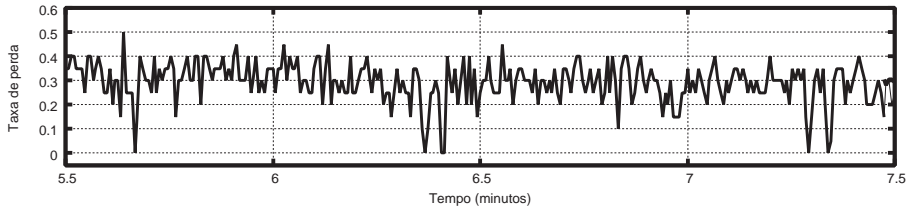




(a) Taxa de perda com rajadas periódicas em uma medição entre UMd e UFRJ



(b) Taxa de perda com alta variabilidade em uma medição entre UMass e UFRJ



(c) Taxa de perda elevada em uma medição entre UFMG e UFRJ

**Figura 4. Taxa de perda de três *traces* selecionados para uma janela de previsão  $F = 400$  milissegundos**

1. A primeira medida utilizada é o erro médio quadrático  $\mathbf{MSE} = E[(\hat{R}_t^F - R_t^F)^2]$ .
2. Como segunda medida de desempenho, temos a taxa de acerto relativa. Ela será muito útil por ter uma interpretação simples. Além disto, para determinadas aplicações pode ser tolerável uma incerteza relativamente pequena.

Formalmente, seja a variável aleatória que indica o acerto relativo

$$I_t^\alpha = \begin{cases} 1 & \hat{R}_{t+1} \in [R_{t+1}(1 - \alpha), R_{t+1}(1 + \alpha)], \\ 0 & \text{caso contrário.} \end{cases}$$

A taxa de acerto relativa é definida como  $E[I_t^\alpha] = \mathbf{I}(\alpha)$ .

3. A última medida de desempenho é a correlação cruzada (**COR**), medindo a correlação entre a série de previsões e a série de observações. Com ela, temos a noção se a variabilidade de  $\hat{R}_t^F$  está associada com a variabilidade de  $R_t^F$ .

## 4.2. Resultados experimentais

Para podermos comparar os modelos HMM e AR, iremos utilizar os mesmos valores para intervalo de treinamento ( $\tau_{HMM} = \tau_{AR}$ ) e intervalo de previsão ( $\psi_{HMM} = \psi_{AR}$ ).

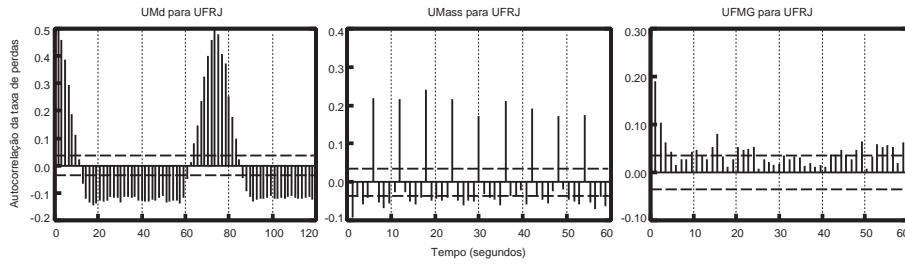


Figura 5. Autocorrelação dos processos de perdas da Figura 4 .

Tomemos ainda a janela de previsão igual ao intervalo de previsão. As transições entre os estados ocultos do HMM acontecem a cada  $S$  observações. As amostras que servem de entrada para o modelo AR foram obtidas agrupando-se  $S$  observações. Note que ambos os modelos irão computar  $f' = \psi/S$  previsões. Vamos denotar por  $R_{t+i}$   $1 \leq i \leq f'$  a  $i$ -ésima previsão calculada no instante  $t$ .

Além dos modelos HMM e AR, vamos utilizar os seguintes modelos de referência:

- **Replicador:** A cada intervalo de previsão  $\psi$ , as últimas estatísticas são usadas como previsão para o próximo intervalo:

$$R_{t+i}^* = \frac{1}{f'} \sum_{j=0}^{f'-1} R_{t-j} \quad B_{t+i}^* = \frac{1}{f'} \sum_{j=0}^{f'-1} B_{t-j} \quad 1 \leq i \leq f'$$

- **Preditor Média:** Semelhante ao Replicador, mas o intervalo de previsão é  $\tau$ .

Tabela 1. Comparação de medidas de desempenho nos *traces* da Figura 4

Modelo	UMd - UFRJ			UMass - UFRJ			UFMG - UFRJ		
	MSE	COR	$I(0.4)$	MSE	COR	$I(0.4)$	MSE	COR	$I(0.4)$
<b>HMM</b>	0.01244	0.23599	10%	0.04599	0.43518	25%	0.00782	0.26941	16%
<b>AR</b>	0.01330	0.22376	9%	0.05625	0.41024	16%	0.01039	0.16756	14%
Replicador	0.01823	0.16268	0%	0.11260	-0.16332	0%	0.01373	0.12242	0%
Média	0.02189	0.10525	1%	0.08989	0.00520	0%	0.01511	0.15834	10%

Para o 1o. trace da Tabela 1, vemos que **AR** tem desempenho equivalente ao **HMM** no mecanismo de previsão adaptativa. Isto se deve ao fato que o processo (Figura 4(a)) tem uma estrutura fortemente auto-regressiva: quando (não) ocorre uma perda, ela tende a (não) ocorrer novamente ( $x_t \sim x_{t-1}$ ).

Para o 2o. trace analisado, o modelo **HMM** e o modelo **AR** têm um desempenho geral semelhante no segundo *trace*(Figura 4(b)). Como neste caso existe uma estrutura auto-regressiva vista na periodicidade da taxa de perda, o modelo **AR** terá um desempenho geral comparável ao modelo **HMM**.

Por fim, para os segmentos do último trace (Figura 4(c)) com variabilidade o **HMM** é notoriamente melhor do que o modelo auto-regressivo.

Em resumo, em todos os casos considerados, os modelos **HMM** e **AR** foram muito melhores do que os preditores replicador e média em termos de erro médio quadrático, correlação cruzada e taxa de acerto. Além disto, o modelo **HMM** foi tão ou mais bem sucedido em acertar a taxa de perda durante os segmentos com certa variabilidade do que o modelo **AR**. Na próxima seção avaliaremos a influência da capacidade de previsão na qualidade da recuperação de pacotes pelo mecanismo de **FEC**.

## 5. Aplicação da previsão da taxa de perda

Nesta seção descreveremos brevemente o funcionamento do algoritmo de seleção de FEC desenvolvido em [Filho and de Souza e Silva 2006]. Em seguida, propomos uma modificação deste algoritmo que nos permitirá relacionar a capacidade de previsão de diferentes modelos (Seção 4) com o desempenho da recuperação.

Baseado na suposição de que as condições do canal a curto prazo são bem representadas por um modelo de Gilbert-Elliot simplificado, o algoritmo de seleção de FEC em [Filho and de Souza e Silva 2006] determina um esquema de FEC dentre os disponíveis segundo a heurística  $\mathbb{C}$ :

*Se a taxa de perda média prevista for menor que  $\theta$ , não use redundância.*

*Senão, utilize o esquema de FEC que mantém a taxa de perda média pós-recuperação abaixo de  $\theta$ . Se houver mais de um esquema que satisfaça essa condição os critérios de desempate são, nesta ordem: A) menor overhead; B) menor atraso de reconstrução. Por outro lado, se não houver nenhum esquema, escolha aquele cuja taxa de perda média pós-recuperação é mais próxima de  $\theta$ .*

Iremos restringir os esquemas disponíveis àqueles que tem janela de tamanho  $\leq 6$  a fim de limitar o atraso de reconstrução. O *overhead* atribuído a cada um desses esquemas de FEC pode variar de 16.7% a 100%.

No modelo HMM, cada estado oculto corresponde a um Gilbert. Durante a previsão, a distribuição de probabilidade de cada estado nos próximos instantes de tempo é determinada pela análise transiente da cadeia oculta. O algoritmo utiliza essa distribuição para selecionar o esquema de FEC a ser empregado da seguinte forma:

*Sejam  $i$  e  $j$  os estados ocultos mais prováveis no instante  $t$  (sendo  $\pi_{t|H}(i) \geq \pi_{t|H}(j)$ ) e  $\omega$  um grau de garantia arbitrário.*

- Se  $\pi_{t|H}(i) \geq \omega$ , utilize o FEC dado pela heurística  $\mathbb{C}$  para o modelo de Gilbert embutido no estado  $i$ .
- Senão, combine o esquema de FEC dado pela heurística para o modelo de Gilbert do estado  $i$  com o esquema dado para o modelo de Gilbert do estado  $j$ .

Note que a seleção de FEC do algoritmo é baseada no estado oculto mais provável e, em alguns casos, nos 2 estados mais prováveis. Isso significa desconsiderar grande parte dos estados e sua probabilidade transiente. Sendo a previsão da taxa de perda do HMM dada em função de todos os estados, seria impossível estabelecer uma relação entre a previsão e o resultado da utilização do algoritmo.

Outros modelos para o processo de perdas (e.g. AR, Replicador) não têm suas previsões dadas em função de modelos de Gilbert assim como o HMM. Por isso, também não poderíamos aplicar diretamente a heurística  $\mathbb{C}$  para obter os resultados da recuperação dos outros modelos, caracterizando um óbice à comparação dos resultados do HMM.

Para permitir o estudo da relação entre a previsão do HMM e o resultado da recuperação e também a comparação dos resultados da recuperação feita a partir da previsão de diferentes modelos, propomos uma modificação no algoritmo de seleção de FEC:

**procedure SelectFEC**(  $R_{t+1}^*$ ,  $B_{t+1}^*$  ).

**início**

$$q := \frac{1}{B_{t+1}^*};$$

$$p := \frac{q \times R_{t+1}^*}{1 - R_{t+1}^*};$$

Use o esquema de FEC dado pela heurística  $\mathbb{C}$  para o modelo de Gilbert(p,q);

**fim**

onde  $R_{t+1}^*$  e  $B_{t+1}^*$  denotam as previsões da taxa média de perda e do tamanho médio da rajada de perda respectivamente,  $p$  denota a probabilidade de transição do estado 0 para o estado 1 em um modelo de Gilbert e  $q$  denota a probabilidade de transição de 1 para 0. Note que o Gilbert determinado por  $p$  e  $q$  possui taxa média de perda  $R_{t+1}^*$  e tamanho médio de rajada  $B_{t+1}^*$ .

Após a introdução do algoritmo modificado de seleção de FEC, precisamos mostrar que *quanto menor o grau de incerteza da previsão, melhor o resultado após a recuperação*, caso contrário não haveria sentido em se utilizar modelos para o processo de perdas. De fato, os resultados obtidos comprovaram a eficácia do algoritmo.

## 5.1. Metodologia

Para mostrar que o algoritmo de seleção de FEC modificado pode proporcionar bons resultados, vamos supor a existência do seguinte preditor:

**Preditor Ótimo:** A taxa de perda e o tamanho médio da rajada a cada  $S$  observações são conhecidos pelo preditor, ou seja,  $R_{t+i}^* = R_{t+i}$  e  $B_{t+i}^* = B_{t+i}$   $1 \leq i \leq f'$ .

É importante notar que usar o preditor ótimo para escolha de FEC não significa necessariamente escolher FEC de maneira ótima, ou seja, pode existir uma escolha de FEC que recupere mais pacotes com menor *overhead* para um dado intervalo de tempo. Isso pode acontecer porque: 1– O modelo de Gilbert determinado por  $R_{t+i}$  e  $B_{t+i}$  pode não caracterizar bem a distribuição das perdas naquele intervalo; ou 2– Ainda que o modelo caracterize bem a distribuição das perdas, a heurística do algoritmo é baseada no valor esperado de pacotes recuperados e não leva em consideração a variância. Por esse motivo, usaremos também como referência a **Heurística Ótima**:

**Heurística Ótima:** a escolha do FEC é feita de forma idêntica a heurística  $\mathbb{C}$ , mas no lugar da previsão taxa de perda pós-recuperação é utilizado o valor real da taxa de perda após a reconstrução dos pacotes perdidos.

Note que a Heurística Ótima não recebe um modelo de Gilbert como entrada. A taxa de perda em cada intervalo de tempo é calculada executando-se o mecanismo de recuperação uma vez para cada esquema de FEC disponível. A escolha do esquema  $FEC_{t+i}$  a ser usado é feita sem levar em consideração  $FEC_{t+i-1}$  e  $FEC_{t+i+1}$ . Na prática, a escolha de  $FEC_{s_{t+i-1}}$  e  $FEC_{s_{t+i+1}}$  pode melhorar a recuperação em  $t+i$ , mas considerar essas escolhas aumentaria muito o custo computacional do experimento.

**Parametrização.** Os parâmetros dos modelos foram especialmente arbitrados para este estudo de caso. Idealmente um preditor deveria ser capaz de antever o resultado da transmissão de cada pacote (sucesso ou falha), ou seja, deveríamos tomar

$S = 1$  observação. No entanto, iremos adotar  $S = 25$  por uma série de motivos:

- A cada intervalo  $\psi$  o par (taxa de perda média, tamanho médio da rajada) pode ser modificado  $\psi/S$  vezes. Conseqüentemente, o esquema de FEC selecionado pelo algoritmo também pode mudar  $\psi/S$  vezes. A troca muito freqüente de esquema de redundância se traduz em maior *overhead* devido às condições de contorno da implementação e maior custo computacional.
- Dado que  $\psi = f' \times S, f' \in \mathbb{N}^*$ , valores pequenos de  $S$  implicam menores intervalos de previsão. A cada previsão do HMM são necessárias as últimas  $H$  observações. Portanto, quanto menor o valor de  $S$ , maior é a periodicidade com que as observações precisam ser recebidas. Note que a quantidade de informação necessária à realização das previsões é independente de  $S$ . Ela é proporcional apenas à freqüência de transmissão dos dados.
- Durante a análise transiente do HMM, o vetor  $\pi_{t|H}$  é multiplicado pela matriz de transição dos estados ocultos  $\psi/S - 1$  vezes. Fixando  $\psi$ , o aumento no valor de  $S$  reflete em menor número de multiplicações matriciais.

Fixamos  $\tau = T = 4$  minutos. O restante dos parâmetros dos modelos HMM e AR não foram fixados *a priori*. Nos experimentos, fizemos com que tais parâmetros variassem por um grande conjunto de valores e selecionamos para comparação aqueles com os melhores resultados de acordo com as métricas descritas a seguir. No primeiro caso, tais parâmetros são: estrutura da cadeia (proposta em [Filho 2006] ou proposta em [Duarte 2003]), no. de estados ocultos ( $N \in \{5, 10, 20\}$ ), intervalo de previsão ( $\psi \in \{50, 100\}$ ) e memória de previsão ( $H \in \{50, 500, 1000\}$ ). Já no caso do modelo auto-regressivo os parâmetros não-fixados são: grau de regressão ( $p \in \{2, 20, 40\}^*$ ) e intervalo de previsão ( $\psi \in \{50, 100\}$ ).

**Métricas.** As métricas utilizadas na avaliação serão: a *taxa de recuperação*  $r$  e *overhead*  $o$  (dados pelas equações 8). Em uma aplicação de voz sobre IP, a recuperação de pacotes se reflete como ganho em termos de QoS. De maneira geral, quanto mais redundância adicionamos ao fluxo, maior é a chance de recuperarmos pacotes perdidos. Em contrapartida, do ponto de vista da rede, é importante minimizar o *overhead* no canal de transmissão. Por isso, esta é a outra métrica que iremos considerar.

$$r = \frac{\text{no. de pacotes recuperados}}{\text{no. de pacotes perdidos}} \quad o = \frac{\text{no. de pacotes adicionados ao fluxo}}{\text{no. de pacotes do fluxo original}} \quad (8)$$

Devido ao problema em se atribuir pesos a  $r$  e  $o$  para obter uma avaliação unidimensional dos resultados, vamos nos restringir a compará-los segundo a Definição 5.1.

**Definição 5.1.** *Sejam  $o_i$  e  $r_i$  o overhead e a taxa de recuperação do modelo  $i$ , respectivamente. O resultado  $R_A = (o_A, r_A)$  de um modelo  $A$  é dito **melhor** que o resultado  $R_B = (o_B, r_B)$  de um modelo  $B$  se uma das seguintes condições forem satisfeitas:*

- $r_A > r_B$  e  $o_A < o_B$ ; ou
- $|r_A - r_{\text{Preditor Ótimo}}| < |r_B - r_{\text{Preditor Ótimo}}|$  e  $|o_A - o_{\text{Preditor Ótimo}}| < |o_B - o_{\text{Preditor Ótimo}}|$

Os modelos foram implementados na ferramenta de transmissão de voz sobre IP conhecida por Vivavoz [Azevedo et al. 2006]. No entanto, os resultados a seguir foram obtidos *offline* para reduzir o tempo de simulação. Arbitramos  $\theta = 0.03$  conforme sugerido pelo trabalho em que foi desenvolvido o algoritmo original.

---

\*Note que este conjunto de valores é justo, visto que, para  $S = 25$ , tomar valores de  $p$  iguais a 2,20 e 40 equivale a considerar as últimas 50, 100 e 1000 observações, respectivamente.

## 5.2. Estudo de Caso

A seguir iremos avaliar o desempenho dos modelos anteriormente citados quanto à taxa de recuperação e ao *overhead*, para cada um dos traces descritos na seção de experimentos. As previsões serão usadas pelo algoritmo de seleção de FEC a partir de  $t = 4$  minutos. Assim como na seção 4, iremos avaliar os traces somente no intervalo entre 4 e 8 minutos.

Inicialmente iremos mostrar os resultados obtidos através da variação no espaço paramétrico do HMM. Serão omitidos: 1) os resultados para o modelo HMM proposto em [Duarte 2003]; 2) os resultados advindos das parametrizações do modelo AR por concisão. Em seguida, comparamos os melhores resultados (obtidos conforme a Definição 5.1) de cada modelo.

A interpretação das Figuras 6(a-d) deve ser feita levando em conta que cada resultado do HMM possui: 1) uma localização que indica a taxa de recuperação e o *overhead*; e 2) uma coloração que indica a incerteza (MSE) a ele associado.

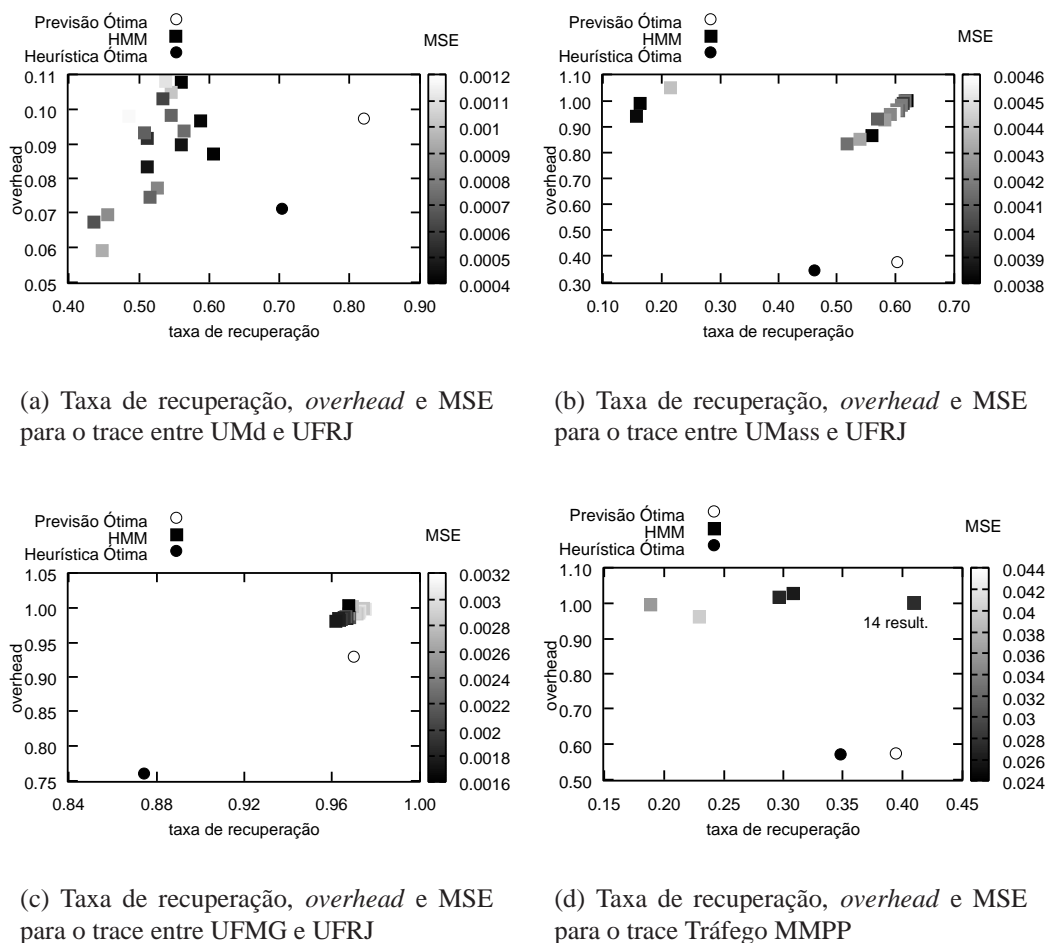


Figura 6. Experimentos usando o Algoritmo de Seleção de FEC Modificado

**Tabela 2. Algumas Parâmetrizações do HMM e Resultados das Figuras 7(a,b)**

Parametrização			UMd - UFRJ			UMass - UFRJ		
N	$\psi$	$H$	$r$	$o$	MSE	$r$	$o$	MSE
5	50	50	<b>0.56</b>	<b>0.11</b>	<b>4.08e-04</b>	0.60	0.95	4.58e-03
5	50	500	0.56	0.09	4.57e-04	<b>0.62</b>	<b>1.00</b>	<b>3.80e-03</b>
5	50	1000	0.61	0.09	4.09e-04	<b>0.16</b>	<b>0.94</b>	<b>3.83e-03</b>
10	50	50	<b>0.59</b>	<b>0.10</b>	<b>4.19e-04</b>	0.61	0.98	4.19e-03
10	50	1000	0.54	0.10	6.12e-04	<b>0.16</b>	<b>0.99</b>	<b>3.80e-03</b>
10	100	500	<b>0.54</b>	<b>0.11</b>	<b>1.11e-03</b>	0.57	0.93	4.10e-03
20	50	500	0.44	0.07	6.60e-04	<b>0.56</b>	<b>0.87</b>	<b>3.80e-03</b>
20	100	500	<b>0.55</b>	<b>0.10</b>	<b>1.01e-03</b>	0.52	0.83	4.15e-03
20	100	1000	<b>0.49</b>	<b>0.10</b>	<b>1.17e-03</b>	0.59	0.95	4.22e-03

### Trace entre UMd e UFRJ

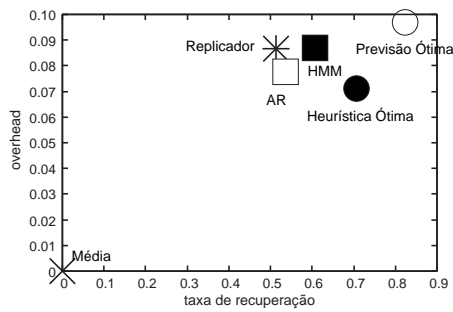
Este trace (Figura 4(a)) possui uma baixa taxa de perda média ao longo do trecho avaliado (2,2%). Além disso, as variações na taxa de perda ocorrem de maneira periódica e esparsa. Nesse caso, a qualidade do resultado está intimamente ligada à qualidade da previsão (ver Figura 6(a)). Podemos observar que os resultados com menor **MSE** são aqueles que estão mais próximos da Previsão e Heurística Ótimas. Note ainda que a dispersão dos resultados para este trace é pequena, indicando a robustez da recuperação quanto a variação dos parâmetros do modelo. Na Tabela 2 são listados os valores encontrados de alguns experimentos de 2 dos traces, sendo destacados os mais relevantes.

Durante a fase de treinamento a taxa de perda média é baixa e, por isso, o algoritmo não emprega nenhuma redundância quando suas entradas são alimentadas pelo Preditor Média. As perdas estão concentradas principalmente em 3 períodos ao longo do intervalo avaliado, assumindo (a taxa de perda) valores em torno de 12%. Na Figura 7(a) podemos observar que o modelo que mais se aproxima dos resultados obtidos pela Heurística Ótima e da Previsão Ótima é do HMM. Pela natureza estável do processo de perdas, a qualidade da previsão das perdas e o resultado do Replicador são bons. Ainda assim, o HMM recupera mais pacotes com um pouco menos de *overhead*.

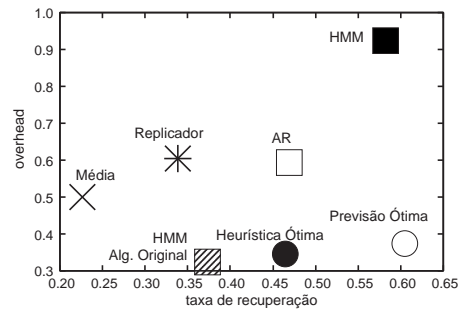
### Trace entre UMass e UFRJ

A taxa média de perda no trecho avaliado deste trace é 9.5%(Figura 4(b)). O processo de perdas se caracteriza por picos periódicos cuja amplitude varia de 12 a 68%. Os resultados obtidos usando o modelo HMM possuem um *overhead* consideravelmente maior do que aqueles dados pelo Preditor Ótimo e pela Heurística Ótima. A grande maioria dos resultados se situa na mesma faixa de recuperação que a Previsão Ótima, porém com uso mais intensivo de redundância. Novamente isso indica que, em geral, a recuperação não é afetada pela parametrização do modelo.

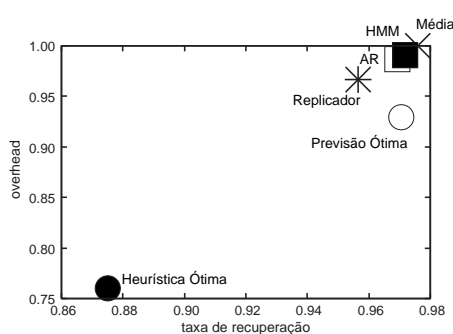
Das 4 melhores previsões em termos de **MSE**, 2 estão localizadas numa área de muito *overhead* e pouca recuperação, ao passo que previsões com maior **MSE** obtêm melhores resultados(ver Figura 6(b) e Tabela 2). Isso é de fato contra-intuitivo, mas pode ser explicado pela situação hipotética da Figura 8.



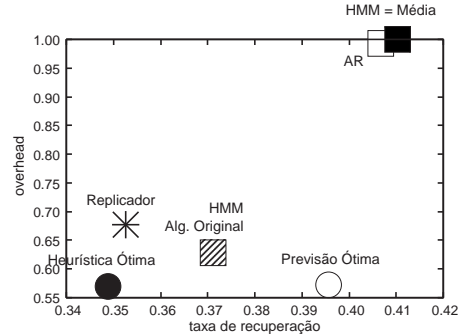
(a) Taxa de recuperação e *overhead* para o trace entre UMd e UFRJ



(b) Taxa de recuperação e *overhead* para o trace entre UMass e UFRJ



(c) Taxa de recuperação e *overhead* para o trace entre UFMG e UFRJ

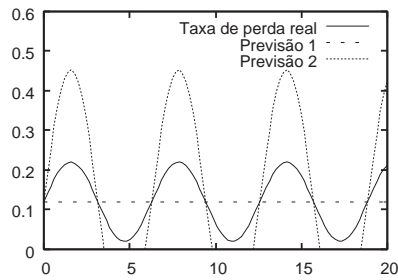


(d) Taxa de recuperação e *overhead* para o trace Tráfego MMPP

**Figura 7. Resultados dos experimentos usando diversos modelos**

A Previsão 1 dada pela reta horizontal representa uma previsão estacionária ( $R_t = 0.12$ ) que, quando fornecida ao algoritmo de seleção de FEC implicará no uso de um mesmo esquema de redundância ao longo de todo o intervalo considerado. A redundância será adicionada ao fluxo mesmo nos trechos em que a taxa de perda cai, contribuindo para o *overhead* total sem trazer benefícios substanciais. Por outro lado, nos trechos em que a taxa de perda sobe, o algoritmo alimentado pela Previsão 1 pode não selecionar uma FEC tão robusta quanto a necessária para manter a taxa de perda a níveis aceitáveis. A Previsão 2, por possuir uma amplitude muito maior que a da taxa de perda real, tem um grau de incerteza elevado. No entanto, ela não irá fazer com que o algoritmo adicione redundância nos trechos em que a perda é baixa e, possivelmente, irá fazer com que empregue um esquema de redundância satisfatório aos requisitos da aplicação. Nesta situação, a previsão com maior incerteza é aquela que dá o melhor resultado. O fenômeno descrito está associado a escolhas de  $S$  grande. Devido aos motivos que apresentamos para escolha de  $S = 25$ , em vez de propormos a redução deste parâmetro, sugerimos o uso do algoritmo de seleção de FEC Original. Por se basear nos estados ocultos mais prováveis, a seleção feita daquela maneira exprime uma variabilidade maior na escolha do FEC. Na figura 7(b) mostramos, além dos métodos de referência, o melhor resultado da recuperação quando o algoritmo original é utilizado.





**Figura 8. Situação em que uma previsão com incerteza maior pode recuperar melhor.**

Os preditores média e replicador não geram bons resultados devido à grande variabilidade do processo de perdas. Se considerarmos o algoritmo de seleção de FEC original, o *overhead* cai de 92% para 32%. Em compensação, a recuperação diminui de 58% para 38%. É importante notar que o uso do algoritmo original nem sempre melhora o resultado. De fato, na maior parte dos casos, fixada a parametrização, melhores resultados eram obtidos quando suas previsões eram utilizadas pelo algoritmo modificado. Por concisão, omitimos os resultados dos experimentos que empregaram o algoritmo original.

A diferença entre os resultados da Previsão Ótima e da Heurística Ótima mostra que com um pouco mais de *overhead* (1%) algumas vezes é possível obter ganhos substanciais (12%) na taxa de recuperação.

### Trace entre UFMG e UFRJ

Este trace (Figura 4(c)) possui alta taxa de perda (29%) no trecho avaliado. As perdas em geral são espaçadas, o que facilita a recuperação. Os resultados mostrados na figura 6(c) são bastante semelhantes e todos se encontram próximos ao da Previsão Ótima. Note que os eixos de variação estão bem reduzidos e que quanto melhor a previsão, mais próximo da Heurística Ótima o resultado se situa. Na Figura 7(c) os resultados também são semelhantes. Para este trace, a relação entre *overhead* e recuperação é praticamente linear.

### Simulação com Tráfego MMPP

A motivação deste experimento é simular um único gargalo multiplexado pela agregação de várias aplicações multimídias. A principal característica deste tipo de tráfego agregado é observada na auto-similaridade da sua intensidade [Sahinoglu and Tekinay 1999].

Os resultados foram obtidos usando um processo MMPP (*Markovian Modulated Poisson Process*) como tráfego de fundo. O processo MMPP é simplesmente um processo Poisson com intensidade  $\lambda_t$ , sendo  $\{\lambda_t\}$  um processo markoviano. Pela proposta de [Robert and Le Boudec 1997] temos uma cadeia de Markov específica que nos permite gerar tráfego pseudo auto-similar com baixo custo computacional.

A variabilidade na intensidade do tráfego apresentada em escalas de tempo elevadas ocasiona perdas por congestionamento com uma correlação temporal maior, como vemos na Figura 9, do que em outros modelos de tráfego como o Poisson por exemplo.

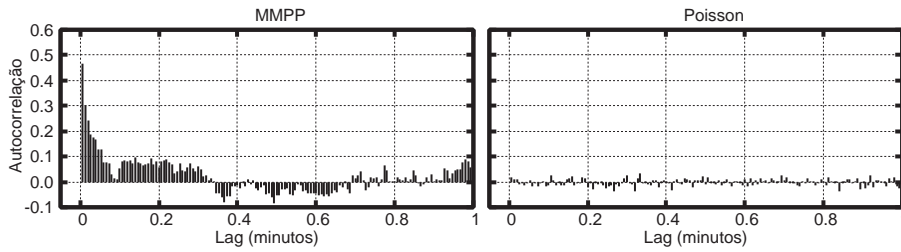
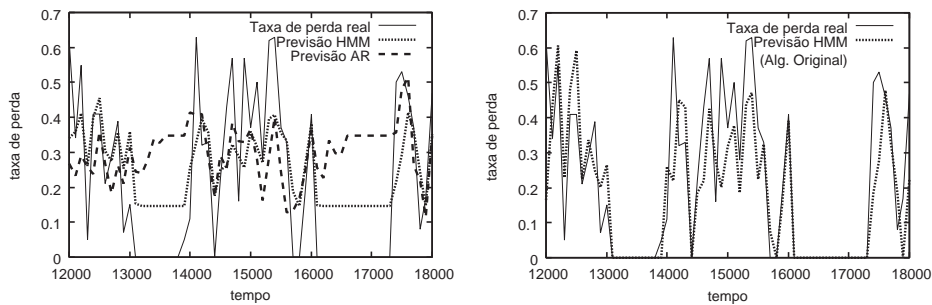


Figura 9. Autocorrelação da taxa de perdas



(a) Os modelos HMM e AR mostram dificuldades em representar a taxa de perda nula.

(b) Comparação entre a previsão do HMM e a taxa estacionária do estado oculto mais provável.

Figura 10. Trecho do trace Tráfego MMPP.

Assim como o trace anterior, este possui alta taxa de perda média (26%) e as perdas também ocorrem em rajadas. A taxa de perda após a recuperação usando os diferentes modelos se situou próxima aos 16%. Os resultados da Figura 6(d) estão localizados na região superior do gráfico, indicando o uso excessivo de redundância quando comparado com o *overhead* imposto pelo Preditor e Heurística Ótimos (entre 55% e 60%).

Um mesmo esquema de redundância foi selecionado por 14 parametrizações diferentes e mantido ao longo de todo o intervalo observado. Esses resultados estão sobrepostos no ponto (0.41, 1.00) e representam um caso extremo de robustez em relação à escolha dos parâmetros. Observe também que os 2 resultados cujo MSE é maior estão à esquerda do gráfico, indicando que a taxa de recuperação foi menor. Como a taxa de perda é alta ao longo da fase de treinamento, os modelos HMM e AR apresentam dificuldades para prever corretamente períodos em que a taxa de perda é nula (ver Figura 10(a)). Note na Figura 7(d) que o HMM e o Preditor Média geram os mesmos resultados. O resultado do AR também está muito próximo dos destes.

Se considerarmos o uso do algoritmo de seleção de FEC original o resultado do HMM se aproxima da Heurística Ótima consideravelmente. O *overhead* cai de 100% para 63% sem que a taxa de recuperação descreça muito (4% de queda). Isso acontece porque nesse caso a previsão dada pela taxa de perda estacionária do estado oculto mais provável captura melhor os trechos de taxa de perda nula (Figura 10(b)).

O Replicador nesse caso possui menor *overhead* porque representa corretamente os trechos em que a taxa de perda se mantém nula por um tempo  $t \geq \psi$ .

## 6. Conclusões

Neste trabalho, avaliamos a capacidade de previsão da taxa de perda dos modelos HMM e AR quanto a métricas pré-definidas para *traces* representativos de um grande número cenários de perda. Poderíamos considerar vários outros modelos ou ainda diferentes estruturas de HMM para comparação, mas deixamos esse estudo para um futuro trabalho. Tomando-se o melhor resultado para cada modelo estudado, observou-se que o desempenho dos modelos em geral é semelhante, mas que em segmentos com maior variabilidade o **HMM** pode obter uma previsão superior a do **AR**.

Para estudar a influência da previsão quando empregada na seleção de FEC, foi proposto um algoritmo baseado naquele proposto em [Filho and de Souza e Silva 2006]. Este algoritmo recebe como entrada as previsões da taxa de perda média e do tamanho médio da rajada para determinar qual esquema de redundância adotar. Mostramos que este algoritmo opera perto do resultado da Heurística Ótima quando alimentado pela Previsão Ótima. Além disso, para os processos de perdas considerados, o algoritmo satisfaz a condição desejável a um algoritmo de seleção de FEC alimentado por previsões: *a qualidade da recuperação deve ser monotonicamente decrescente na incerteza da previsão*. Combinando o uso do modelo HMM com o algoritmo proposto em, por exemplo, uma ferramenta de Voz sobre IP, são obtidos ganhos significativos em termos de recuperação, os quais refletem diretamente na qualidade do áudio recebido.

A partir do algoritmo de seleção de FEC modificado, concluímos que não só a previsão, mas também os resultados da recuperação são robustos em relação à variação no espaço paramétrico do modelo HMM.

Não foram encontrados, dentre os experimentos realizados, resultados superiores ao do Preditor Ótimo. Isso indica que, a princípio, seria possível fazer previsões mais precisas e obter resultados de recuperação ainda melhores.

## Referências

- [Azevedo et al. 2006] Azevedo, J. A., Netto, B. C. M., de Souza e Silva, E., and Leão, R. M. M. (2006). Freemeeting: um ambiente para trabalho cooperativo e ensino a distância. In *Anais do 7 Fórum Internacional de Software Livre*, pages 319–323.
- [Bilmes 1997] Bilmes, J. (1997). A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. Technical Report, University of Berkeley, ICSI-TR-97-021.
- [Bolot et al. 1999] Bolot, J. C., Parisi, S. F., and Towsley, D. F. (1999). Adaptive FEC-based error control for Internet telephony. In *Proceedings of the IEEE INFOCOM*, pages 1453–1460.
- [Brakmo and Peterson 1995] Brakmo, L. S. and Peterson, L. L. (1995). TCP Vegas: End to end congestion avoidance on a global Internet. *IEEE Journal on Selected Areas in Communications*, 13(8):1465–1480.
- [Brockwell and Davis 2002] Brockwell, P. J. and Davis, R. A. (2002). *Introduction to Time Series and Forecasting*. Springer.
- [de Souza e Silva et al. 2006] de Souza e Silva, E., da Silva, A. P., de A. Rocha, A. A., Leão, R. M. M., Duarte, F. P., Filho, F. J. S., Jaime, G. D. G., and Muntz, R. R. (2006).

Modeling, analysis, measurement and experimentation with the Tangram-II integrated environment. In *VALUETOOLS*, page 7.

- [Duarte 2003] Duarte, F. P. (2003). Algoritmo adaptativo para previsão e recuperação de perda de pacotes em aplicações multimídias usando cadeias de Markov ocultas. Master's thesis, Universidade Federal do Rio de Janeiro - COPPE - Programa de Engenharia de Sistemas e Computação.
- [Duarte et al. 2003] Duarte, F. P., de Souza e Silva, E., and Towsley, D. (2003). An adaptive FEC algorithm using hidden Markov chains. *SIGMETRICS Performance Evaluation Review*, 31(2):11–13.
- [Elliot et al. 1995] Elliot, R. J., Aggoun, L., and Moore, J. B. (1995). *Hidden Markov Models: Estimation and Control*. Springer-Verlag.
- [Filho 2006] Filho, F. J. S. (2006). Previsão de estatísticas de perdas de pacotes usando modelos de markov ocultos. Master's thesis, Universidade Federal do Rio de Janeiro - COPPE - Programa de Engenharia de Sistemas e Computação.
- [Filho and de Souza e Silva 2006] Filho, F. J. S. and de Souza e Silva, E. A. (2006). Modelling the short-term dynamics of packet losses. In *Performance Evaluation Review*.
- [Filho et al. 2006] Filho, F. J. S., Watanabe, E. H., and de Souza e Silva, E. A. (2006). Adaptive forward error correction for interactive streaming over the Internet. In *Proceedings of the IEEE Globecom*.
- [Mascolo et al. 2001] Mascolo, S., Casetti, C., Gerla, M., Sanadidi, M. Y., and Wang, R. (2001). TCP Westwood: Bandwidth estimation for enhanced transport over wireless links. In *Mobile Computing and Networking*, pages 287–297.
- [Paxson 1997] Paxson, V. (1997). End-to-end Internet packet dynamics. In *SIGCOMM '97: Conference proceedings on Applications, technologies, architectures, and protocols for computer communication*, pages 139–152.
- [Rabiner 1989] Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–285.
- [Robert and Le Boudec 1997] Robert, S. and Le Boudec, J. Y. (1997). New Models for Pseudo Self-Similar Traffic. *Performance Evaluation*, 30(1-2):57–68.
- [Sahinoglu and Tekinay 1999] Sahinoglu, Z. and Tekinay, S. (1999). On multimedia networks: self-similar traffic and network performance. *Communications Magazine, IEEE*, 37(1):48–52.
- [Salamatian and Vaton 2001] Salamatian, K. and Vaton, S. (2001). Hidden Markov modeling for network communication channels. In *Proceedings of the ACM SIGMETRICS*, pages 92–101.
- [Su et al. 2004] Su, Y. C., Yang, C. S., and Lee, C. W. (2004). The analysis of packet loss prediction for Gilbert-model with loss rate uplink. *Information Processing Letters*, 90:155–159.
- [Yajnik et al. 1999] Yajnik, M., Moon, S. B., Kurose, J. F., and Towsley, D. F. (1999). Measurement and modeling of the temporal dependence in packet loss. In *Proceedings of the IEEE INFOCOM*, pages 345–352.