Modelos de Fluido para Tráfego Multimídia e Aplicações

por

Sidney Cunha de Lucena

UFRJ

Tese submetida para a obtenção do título de

Doutor em Ciências em Engenharia de Sistemas e Computação

ao Programa de Pós-Graduação de Engenharia de Sistemas e Computação

da COPPE/UFRJ

por

Sidney Cunha de Lucena

Março 2004

MODELOS DE FLUIDO PARA TRÁFEGO MULTIMÍDIA E APLICAÇÕES

Sidney Cunha de Lucena

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Aprovada por:

Prof. Edmundo Albuquerque de Souza e Silva, Ph.D.

Prof. Rosa Maria Meri Leão, Dr.

Prof. Rosângela Fernandes Coelho, Docteur ENST

Prof. Nelson Luis Saldanha da Fonseca, Ph.D.

Prof. José Ferreira de Rezende, Dr.

Prof. Jorge Moreira de Souza, Dr.

RIO DE JANEIRO, RJ - BRASIL

MARÇO DE 2004

LUCENA, SIDNEY CUNHA DE

Modelos de Fluido para Tráfego Multimídia e Aplicações [Rio de Janeiro] 2004

XX, 183 p. 29,7 cm (COPPE/UFRJ, D.Sc., Engenharia de Sistemas e Computação, 2004)

Tese - Universidade Federal do Rio de Janeiro, COPPE

1. Modelos de Tráfego
2. Descritores de Tráfego
3. Modelos Markovianos

I. COPPE/UFRJ II. Título (Série)

Dedico aos meus pais Silvério e Marne, pelo amor na minha criação, pelo sacrifício para a minha formação e por sempre acreditarem em mim.

Ao meu irmão Sérgio, sua esposa Valéria e minha encantadora sobrinha Giovana, para os quais desejo toda a felicidade de uma vida.

À minha namorada Mônica, que sempre esteve ao meu lado nesses anos, me amparando nos momentos de tristeza e se superando nos momentos de alegria.

Aos meus parentes e amigos de todas as horas, por jamais deixarem de me dar força.

Em especial à memória de meu tio Armindo, que me ensinou a nunca deixar de fazer o que gosto - um beijo, meu padrinho!

Agradecimentos

Agradeço o apoio e o empenho de meus orientadores nestes longos anos de doutorado.

Agradeço à todos no **LAND** pela amizade e o apoio de sempre, presente nas várias gerações de alunos desde o início do laboratório. Em especial, sem desmerecimento dos demais colegas, ao Flávio Pimentel Duarte, camarada que por mais tempo esteve comigo no **LAND**, e à Carolina Maria de Vilar Moreira Vieira, que me socorreu inúmeras vezes sempre com um sorriso acolhedor.

Agradeço o apoio financeiro da **CAPES** durante os anos de bolsa.

Agradeço ao Programa de Engenharia de Sistemas e Computação da **COPPE** por toda a infra-estrutura acadêmica que me permitiu a realização do trabalho de tese.

Agradeço à **RNP** e meus colegas de trabalho pelo apoio incessante que me foi dado para a conclusão do doutorado - em especial, ao Alexandre Grojsgold, Ari Frazão e Raniery Pontes, além dos demais colegas do Centro de Engenharia e Operações: Cybelle Oyama, Fábio Okamura, Marcel Faria, Alex Soares e Guilherme Domingues.

E agradeço a Deus por eu ser brasileiro!

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

MODELOS DE FLUIDO PARA TRÁFEGO MULTIMÍDIA E APLICAÇÕES

Sidney Cunha de Lucena

Março/2004

Orientadores: Edmundo de Souza e Silva
Rosa Maria Meri Leão

Programa: Engenharia de Sistemas e Computação

Uma das principais medidas de desempenho que afetam a qualidade de uma aplicação de rede é a taxa de perda de pacotes. A previsão das perdas auxilia o dimensionamento de recursos da rede e permite a avaliação de mecanismos e protocolos que coexistem numa rede de pacotes. Portanto, predizer medidas relacionadas com a perda de pacotes em uma rede multimídia é um dos problemas que tem recebido muita atenção na literatura nos últimos anos. A criação de modelos que produzam resultados precisos é ainda objeto de pesquisa, dado as características peculiares do tráfego gerado pelas aplicações.

Vários são os modelos propostos na literatura para capturar as características do tráfego que mais influenciam nas medidas relacionadas a perda de pacotes. Cada modelo em geral é talhado para um determinado tipo de tráfego e ainda pode ou não capturar dependências temporais das taxas, dependendo do uso a que ele se destina.

Nesta tese é proposto um modelo markoviano com recompensas para representar o tráfego gerado por aplicações de vídeo e tráfego agregado. O modelo incorpora um parâmetro de desempenho, além de novos descritores do tráfego sendo modelado. Estes descritores são obtidos a partir da duração de sobrecargas de tráfego em relação a uma dada taxa. No trabalho são feitos vários estudos usando tráfegos de vídeo codificados em MPEG4 e tráfegos WAN da rede *Abilene* e suas limitações são exploradas. Os resultados mostram que o modelo prevê taxas de perdas com maior precisão do que outros, para uma faixa de parâmetros relevante.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

FLOW MODELS FOR MULTIMEDIA TRAFFIC AND APPLICATIONS

Sidney Cunha de Lucena

March/2004

Advisors: Edmundo de Souza e Silva
Rosa Maria Meri Leão

Department: Computer and System Engineering

One of the most important performance measurements that affects the quality of a network application is the packet loss rate. Predicting packet loss rates is useful for the dimensioning of the networks resources and permits the evaluation of control mechanisms and protocols in a packet switched network. Therefore, predicting packet losses related measurements in a multimedia network is one of the issues that has received considerable attention in the literature over the last many years. Despite of the large existing literature on the subject, the development of models that produce accurate results is still an active research area, due to the wide variety of characteristics of the traffic generated by multimedia applications, and the difficulty in capturing these characteristics in a single model.

A large set of models have been proposed in literature in order to capture traffic characteristics that are most influential on the packet loss related measurements. Each model, in general, is tailored to a certain traffic type and may or may not capture temporal dependencies among the traffic rates, depending on the use one envision for the model.

In the present thesis it is proposed a Markovian model with rewards to represent the traffic generated by video applications and aggregated traffic. The model incorporates a performance parameter, in addition to new descriptors proposed for the traffic being modeled. These descriptors are obtained from the random variable that is equal to the duration of traffic overloads periods with respect to a given packet rate. In this work, studies are performed using different MPEG4 coded video traffic and WAN traffic from *Abilene* network. The limitations of the model are explored. The results show that the proposed model predicts loss rates with higher precision than other models for a relevant parameter range.

Sumário

Lista de Figuras

Capítulo 1

Introdução

1.1 Considerações Iniciais

Durante a última década, muitos foram os resultados publicados visando resolver o problema de caracterização e modelagem dos novos perfis de tráfego que surgiam nas redes de pacotes [1]. O suporte para estes resultados foi uma série de medidas estatísticas que mostravam que os modelos até então utilizados, oriundos do tráfego em redes de telefonia, não mais se aplicava às novas redes multimídia.

O objetivo da modelagem de tráfego é obter um ferramental matemático capaz de tratar probabilisticamente o impacto, na qualidade final das transmissões, dos recursos de rede. Dentre as medidas estatísticas que caracterizam um tráfego, temos descritores de primeira ordem, como a média e o *burstiness*, e os de segunda ordem, como a autocorrelação. A autocorrelação, por exemplo, pode ser aplicada às taxas de transmissão ou aos intervalos de tempo entre chegadas de pacotes. Esta medida estatística foi muito usada nos novos modelos de tráfego por capturar dependências no tempo entre (por exemplo) as taxas de transmissão. Essas dependências têm um impacto não desprezível nas medidas de desempenho estudadas, como a distribuição do tamanho das filas nos canais de saída dos roteadores.

Nos últimos quinze anos, foram propostos vários descritores de tráfego e estudados a sua eficácia em capturar as estatísticas mais relevantes deste tráfego. Basicamente, as pesquisas e propostas que surgiram desenvolveram-se visando casar as distribuições de taxa e funções de autocorrelação dos modelos com as dos tráfegos reais. Uma das poucas exceções que podem ser consideradas é o *parâmetro de Hurst*, embora esta seja uma medida intimamente relacionada com a função de autocorrelação, sendo inclusive matematicamente definida através desta.

Uma das características mais investigada durante o processo de modelagem é a dependência de longa duração entre as taxas de um determinado tráfego. Esta característica pode ser avaliada a partir do comportamento da função de autocorrelação das taxas de um tráfego. Um decaimento lento desta função sugere que o valor do processo num tempo t afeta seu comportamento futuro durante um longo período de tempo.

Inúmeros estudos indicaram que a lenta queda da função de autocorrelação com o "*lag*" do tempo é uma característica presente em diferentes tipos de tráfego. Portanto, modelos recentes passaram a incorporar essa característica em oposição a modelos mais tradicionais que não levavam tal comportamento em consideração [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14].

Em 1994, Leland et al publicaram o artigo seminal na área de modelos de tráfego. Em [2], os autores mostram que o tráfego agregado em redes locais possui características auto-similares. Modelos de tráfego que não eram capazes de reproduzir tais características passaram a ser alvo de críticas. No entanto, a natureza assintótica dos critérios para avaliar o grau de auto-similaridade dos tráfegos, e outras possibilidades não exploradas nos modelos ditos "não auto-similares", direcionaram novas pesquisas em busca de alternativas para os modelos auto-similares.

Por exemplo, outros autores mostraram que para prever certas medidas de desempenho, como a distribuição do tamanho de uma fila com espaço de armazenamento finito, é necessário apenas capturar a autocorrelação para um tempo finito, chamado em [15] de "horizonte de correlação". Isto ocorre porque o processo de ocupação da fila se renova sempre que a fila esvazia. Esta renovação implica em uma "perda de memória", tornando o processo de ocupação da fila independente de seus estados anteriores. Ou seja, para tempos acima dos períodos máximos de ocupação da fila (*busy-periods*), não há necessidade de se capturar o comportamento da autocorrelação. Como conseqüência disto, modelos markovianos continuam sendo atrativos para capturar as características de tráfego que tenham impacto relevante nas medidas de desempenho [16].

A motivação inicial deste trabalho foi o da pesquisa por novos descritores de tráfego que sirvam de base para parametrizar um modelo de forma que a construção do modelo e sua parametrização seja mais simples do que para outros modelos existentes e ainda possam predizer com precisão medidas de desempenho tais como a fração de perda de pacotes em uma fila.

Em [17], Taralp, Devetsikiotis e Lambadaris mostram que modelos possuindo mesma distribuição de taxa e mesma função de autocorrelação podem exibir distribuições absolutamente diferentes para a ocupação de uma fila infinita. Este resultado é aqui confirmado ao se refazer os testes para o caso de perdas numa fila finita (seção 3.1). Neste mesmo artigo, mostrou-se que as diferenças estatísticas existentes entre os modelos podem ser observadas através dos gráficos no $\mathcal{R}^3$ das respectivas funções geradoras de momento logarítmico (explicadas na seção 2.2.6), recurso largamente usado para se computar a banda efetiva de uma fila finita onde se deseja limitar a probabilidade máxima de "transbordo" (*overflow*).

Neste trabalho de tese, para se diferenciar estatisticamente modelos que possuam mesma distribuição de taxa e mesma função de autocorrelação, considerou-se o comportamento dos níveis de taxa e ainda a duração dos períodos nos quais o tráfego originado permanece acima destes níveis. A razão para esta opção origina-se na idéia de caracterizar volumes de tráfego associados a períodos de tempo, visto que tal associação é o que costuma diferenciar dois tráfegos com a mesma função de autocorrelação quando estes têm seus gráficos de taxa ao longo do tempo comparados. Além disso, dependendo da taxa de serviço da fila, os níveis de taxa acima deste valor são mais importantes de serem caracterizados uma vez que as perdas ocorrem durante estes períodos.

Através de gráficos no $\mathcal{R}^3$ mostrando a distribuição da duração dos períodos onde o tráfego encontra-se acima de um certo nível de taxa, para diversos níveis de taxa (ver seção 3.3), é possível diferenciar modelos que contenham a mesma distribuição de taxa e função de autocorrelação. Da mesma forma, é possível mostrar que estes modelos causam impacto diferenciado na distribuição da perda de pacotes dada uma taxa de serviço e diversos tamanhos de fila.

Em outros estudos realizados nesta tese, comparou-se dois diferentes modelos caracterizando uma mesma seqüência. As funções de autocorrelação para estes modelos apresentam diferenças entre si: uma possuindo decaimento rápido, aproximando-se da autocorrelação extraída para seqüência real apenas para tempos curtos, a outra possuindo decaimento lento, aproximando-se da seqüência real para tempos longos. Também as frações de perda, extraídas para várias taxas de serviço e um mesmo tamanho de fila, apresentam diferenças na aproximação da fração de perda da seqüência real. Nestes exemplos, também foi possível relacionar as probabilidades de perda com as respectivas distribuições para a duração dos períodos de tempo onde a taxa de chegada de pacotes de cada modelo esteve acima de cada taxa de serviço.

Diante dos exemplos apresentados, ficou claro que esta nova abordagem é útil para descrever tráfegos e é uma alternativa para a construção de modelos de tráfegos. Mais ainda, a interpretação desta informação é intuitiva e permite a construção de modelos direcionados para a obtenção das medidas de interesse desejadas, dependentes de um parâmetro apenas a ser considerado na medida de desempenho: a taxa de serviço da fila sujeita ao tráfego. Este tipo de abordagem foi recentemente considerado em outros trabalhos [18].

Como contribuição inicial desta tese, mostrou-se a influência sobre medidas de desempenho da distribuição do período de taxas em que o tráfego está acima de determinados níveis. Em seguida, elabora-se uma proposta de modelagem que considera os resultados obtidos. Estes resultados indicam que aproximando-se adequadamente a duração dos períodos de tempo onde o volume de tráfego fica acima da taxa de serviço da fila possibilita uma boa predição da fração de perda de pacotes para diversos tamanhos de fila.

Assim sendo, procurou-se caracterizar a distribuição das durações de *sobrecarga* e das durações de *folga*, onde *sobrecarga* e *folga* dizem respeito, respectivamente, a volumes de tráfego acima e abaixo da taxa de serviço da fila. Para tanto, foram usadas distribuições hiper-exponenciais pela sua propriedade de capturar um lento decaimento de $P[D > t]$, onde D é a variável aleatória representando a duração que se deseja aproximar.

O modelo de tráfego proposto é markoviano. Os estados da cadeia de Markov aproximam os tempos de sobrecarga ou folga. A cada estado associam-se valores de taxa de acordo com aquelas praticadas durante os respectivos períodos. O método usado para aproximar uma distribuição de duração, para uma seqüência real, por uma distribuição hiper-exponencial, na forma markoviana, também é contribuição desta tese assim como todo o modelo proposto. Por fim, o modelo proposto possui apenas seis estados, o que o torna bastante atrativo em termos de custos computacionais para métodos analíticos de solução e simulações.

A eficiência da proposta pode ser mostrada pelos testes executados. Nestes, seqüências MPEG4 são modeladas para avaliação do modelo resultante. É notório que tais seqüências caracterizam-se pela dificuldade de se capturar estatísticamente seu comportamento e predizer medidas de perda. Também foram usadas seqüências de tráfego WAN para os testes. Os resultados aqui obtidos demonstram que a estratégia de usar as definições de sobrecarga e folga simplificam o problema e possibilitam prever com boa aproximação a fração de perda ao longo de uma extensa faixa para tamanhos máximos de fila.

Não somente a eficiência do modelo mas também suas limitações são exploradas no trabalho. No caso, é mostrado que a eficiência do modelo tende a degradar quanto menor é a taxa de serviço em relação à taxa média do tráfego-fonte, ou seja, quanto maior a carga do sistema. Os motivos para isto são discutidos assim como são sugeridas maneiras de se melhorar o modelo proposto de forma a minimizar o problema.

É importante enfatizar que o modelo aqui proposto não é necessariamente um modelo de vídeo. Trata-se de um modelo genérico para diferentes fontes, incluindo tráfego agregado.

1.2 Estrutura da Tese

A tese está organizada em seis capítulos a partir deste de introdução. O capítulo 2 faz uma revisão bibliográfica da teoria que serve de base para o trabalho. Neste capítulo são tratados temas sobre medidas de desempenho e são mostrados alguns dos descritores de tráfego mais conhecidos, assim como modelos de tráfego mais usuais e que sejam de algum interesse para o trabalho. Uma vez que o modelo aqui proposto pertence à classe de modelos markovianos com recompensa de taxa, expressões analíticas de algumas medidas são também mostradas para esta classe de modelos.

O capítulo 3 apresenta resultados que mostram como diferenciar tráfegos com mesma distribuição de taxa e função de autocorrelação. A partir de medidas de períodos de tempo associados com volume de tráfego, estes modelos são estatisticamente distinguíveis e novos descritores são propostos. Conforme anteriormente mencionado, o mesmo experimento foi realizado para dois modelos distintos, inclusive quanto às respectivas funções de autocorrelação, modelando uma seqüência de vídeo em MPEG4.

O capítulo 4 apresenta de maneira detalhada a proposta de modelagem que exemplifica o uso dos descritores apresentados no capítulo 3. Inicialmente é mostrada a idéia básica e, passo a passo, faz-se a evolução para a proposta final. Vários resultados envolvendo a modelagem da mesma seqüência MPEG4 do capítulo 3 são mostrados para exemplificar a proposta. Neste capítulo são feitas também considerações sobre a sensibilidade do modelo proposto em função de seus parâmetros e são discutidas possíveis modificações que possam tornar a proposta mais genérica com relação ao conhecimento da taxa de serviço da fila.

O capítulo 5 explora as possibilidades do modelo proposto elaborando um "teste de esforço" para este. Primeiramente, três seqüências MPEG4 são modeladas para duas cargas distintas e os resultados de perda são comparados com os fornecidos por outros três modelos, dois deles bastante conhecidos na literatura e com características distintas. A comparação entre os modelos é feita através de seus descritores, o que inclui as distribuições para duração de sobrecarga e duração de folga. Após os testes, o modelo proposto é avaliado segundo suas possibilidades para caracterizar tráfegos na *Internet*. Para este caso, foi usado como exemplo seqüências de tráfego extraídas de um dos enlaces da rede **Abilene**, a *Internet* acadêmica dos EUA, também conhecida como *Internet 2*. No início do capítulo, algumas características relacionadas a tráfegos MPEG e tráfegos de *Internet* são mostradas.

O capítulo 6 conclui o trabalho, resumindo as contribuições científicas, destacando as observações mais importantes e mostrando direcionamentos para futuros trabalhos. Por fim, três apêndices são anexados: o primeiro para descrever a ferramenta TANGRAM II, usada em todos os experimentos relacionados neste trabalho, o segundo para mostrar a implementação dos modelos na mesma e o terceiro contendo os valores numéricos para os parâmetros dos modelos usados nos testes do capítulo 5.

Capítulo 2

Fundamentação Teórica e Revisão Bibliográfica

Neste capítulo encontra-se resumida a revisão bibliográfica para o trabalho desenvolvido. As seções abaixo relacionadas fazem um apanhado da literatura sobre questões referentes a medidas de desempenho, descritores de tráfego e modelos de fonte. Ao final do capítulo, são mostradas algumas expressões analíticas para a obtenção de descritores de tráfego em modelos markovianos com recompensa de taxa.

2.1 Medidas de Desempenho

A tarefa de dimensionar tamanhos de fila, escolher a capacidade de um canal para satisfazer um dado requisito de desempenho e parametrizar políticas de serviço em equipamentos de comutação não é tarefa simples. Dependendo dos valores adotados, vários fatores relacionados à qualidade do serviço desejado são modificados. Sendo assim, torna-se necessário avaliar de forma objetiva o grau de degradação que uma transmissão sofre e relacionar este resultado com os parâmetros do sistema, principalmente aqueles que podem ser ajustados.

Medidas de desempenho são definidas com este propósito. Através delas, o comportamento de um sistema genérico consegue ser dimensionado e avaliado segundo sua qualidade. Estas medidas também se prestam para avaliar modelos de tráfego. Como todo modelo é uma aproximação, avaliar certas medidas de desempenho para um determinado modelo de tráfego, frente a diferentes e possíveis cenários, e relacioná-las com medidas extraídas para as situações reais que se deseja modelar, mostra a eficiência deste modelo e sua aplicabilidade.

O objetivo de um modelo de tráfego é representar uma seqüência real de forma que este modelo, em combinação com o modelo do sistema no qual esta seqüência se insere, possa razoavelmente prever o desempenho do "serviço". Por exemplo, caso se deseje prever o atraso sofrido pelos pacotes de uma determinada aplicação de vídeo sendo transmitida via rede, então deseja-se um modelo de fonte de vídeo que, quando combinado com o modelo da rede, possa prever esta medida com uma precisão satisfatória.

As primeiras aplicações envolvendo comunicação de dados tinham como preocupação principal a integridade dos dados transmitidos. Esta integridade, na maioria das vezes, é comprometida por perdas na rede devido a congestionamentos. Sendo assim, as medidas de desempenho mais tradicionais e mais relevantes para os sistemas de comunicação estão relacionadas com medidas de perda por sobrecargas em filas. Dentre as mais comuns, podemos citar a fração de perda, NP/NT, onde NP é a quantidade de dados perdidos e NT é a quantidade de dados transmitidos.

A forma mais usual de avaliar o desempenho de um modelo de tráfego, associado a um modelo de rede (em geral representada por filas), é através dos gráficos da *fração de dados perdidos versus tamanho da fila* e da *fração de dados perdidos versus carga (ou utilização)*. A carga costuma ser expressa pela razão da *taxa média da fonte* pela *taxa média de serviço da fila*. Um gráfico tridimensional relacionando fração de perda, capacidade de armazenamento da fila e carga fornece uma boa visão do desempenho do sistema.

Há também outras medidas relacionadas com as durações das perdas. Assim, também são usadas como medida de desempenho a *fração de tempo onde ocorrem perdas* e a *distribuição da duração dos períodos com perdas*. Estas medidas são bastante apropriadas para tráfegos multimídia onde é mais importante determinar quanto tempo uma aplicação fica sem receber o áudio ou o vídeo sendo transmitido.

Como a perda está associada à ocupação da fila, a *distribuição do tamanho da fila*, $P[Q > q]$ (Q é o número de ocupações na fila em um instante qualquer), também é uma medida muito usada. Esta medida é sempre dependente da taxa de serviço, portanto o gráfico 3D envolvendo a distribuição do tamanho da fila, a taxa de serviço e a capacidade de armazenamento fornece uma boa indicação sobre o comportamento da fila. Outras medidas decorrentes desta são o *tamanho médio da fila* e o *desvio-padrão do tamanho da fila*, ambas em função da taxa de serviço [19].

Estas medidas aqui apresentadas podem ser calculadas tanto para o caso estacionário como para o caso transiente. A análise transiente consiste em limitar as medidas ao tempo de observação desejado. Ao contrário da medida em estado estacionário, a medida em estado transiente é função do comportamento inicial dos processos que envolvem o sistema.

Outras medidas menos tradicionais relacionadas com a perda podem ser verificadas na literatura. Por exemplo, em [20] é calculada a *proporção de tempo, em um intervalo* $(0, t)$*, onde ocorre "alta-perda"*, ou seja, onde a fila está cheia e a fonte gera dados à taxa de pico. Já em [21] é calculada a *média e o coeficiente de variação dos períodos onde a fila se encontra com um tamanho maior e menor que um determinado nível.*

2.2 Descritores de Tráfego

Todo modelo de tráfego procura aproximar determinadas medidas estatísticas coletadas para o referido tráfego. Tais medidas procuram descrever as características mais relevantes do tráfego em questão e por isso são chamadas descritores. Os descritores visam identificar padrões que sejam capazes de prever o comportamento destes tráfegos e, por conseguinte, guiar a alocação de recursos necessários para a transmissão dos mesmos. Portanto, um modelo é avaliado observando-se os descritores extraídos deste modelo e comparando com aqueles extraídos do tráfego real.

A seguir, serão mostrados alguns dos descritores mais tradicionais utilizados para caracterizar tráfego.

2.2.1 Média, Pico, variância e distribuição de taxa

A taxa média é a razão entre o volume total de tráfego enviado e o período no qual este tráfego foi enviado. A taxa de pico informa o volume máximo transmitido num determinado intervalo de tempo. Em determinados cenários, este descritor é mais importante que a taxa média para a alocação de recursos, como numa situação onde não é permitida qualquer perda de dados.

Alguns modelos são parametrizados de forma que a variância ou a distribuição de taxa seja igual à do tráfego real. Obviamente, quando esta distribuição é próxima da distribuição do tráfego real, a taxa de pico e a taxa média também são iguais. Vale ressaltar que distribuições de taxa diferentes podem ter mesma taxa média e mesma variância e taxa de pico. Modelos que se propõem a modelar a distribuição de taxa tendem a ser mais complexos devido ao grau de flexibilidade que devem possuir. Em geral, muitos modelos procuram ajustar apenas a taxa média e a variância, e por vezes a taxa de pico, mas não a distribuição completa.

2.2.2 Autocorrelação

Trata-se de uma função no tempo, adimensional, mas que contém a mesma informação estatística da função de autocovariância. Esta, por sua vez, é uma extensão do conceito de covariância.

A covariância entre duas variáveis aleatórias X e Y é dada por:

$$cov(X,Y) = E[XY] - E[X]E[Y] \ .$$

Esta medida indica o grau de relação linear existente entre as variáveis aleatórias X e Y, ou seja, a tendência delas variarem de forma conjunta [22]. Se a covariância for positiva, quando X crescer, Y tende também a crescer; se for negativa, quando X crescer, Y tende a decrescer; e se for zero não haverá relação conjunta entre as variáveis. Variáveis aleatórias independentes necessariamente têm covariância zero.

Esta medida pode ser apresentada de forma adimensional e compreendida no intervalo $[-1, 1]$ se dividirmos a covariância pelo produto dos desvios-padrões das duas variáveis aleatórias em questão. Tem-se, então, o coeficiente de correlação:

$$\Psi(X, Y) = \frac{E[XY] - E[X]E[Y]}{\sigma_X \sigma_Y} .$$

Neste caso, coeficientes de correlação 1 ou -1 indicam relação totalmente linear entre as variáveis (Ex.: $Y = 2X$ ou $Y = 1 - 2X$), e coeficiente zero indica que não há relação linear.

As funções de autocovariância e de autocorrelação são as respectivas medidas acima mostradas onde substitui-se as duas variáveis aleatórias em questão por variáveis aleatórias correspondentes a um mesmo processo estocástico defasado no tempo. Assim sendo, tem-se:

$$cov(\tau) = E[X(t)X(t + \tau)] - \mu^2 \quad e \quad \Psi(\tau) = \frac{E[X(t)X(t + \tau)] - \mu^2}{\sigma^2} ,$$

onde μ e σ^2 são, respectivamente, a média e a variância do processo estocástico, supostamente estacionário.

Através destes descritores pode-se identificar se um tráfego possui a tendência de ter súbitos aumentos de taxa que se mantêm durante um certo tempo, que são chamados de rajadas. Autocorrelações positivas sugerem rajadas no tráfego, de acordo com o grau destas correlações [23]. Nos últimos anos, a autocorrelação tem sido considerada um descritor fundamental, tendo em vista que um decaimento lento desta, em função do tempo, indica que o processo possui dependência de longa duração [2, 5, 24]. Isto significa que uma variável aleatória no tempo t está positivamente correlacionada com a variável aleatória deste mesmo processo no tempo $t - \tau$ para um τ relativamente grande. Processos com tais características podem implicar que mais recursos precisam ser alocados para satisfazer, por exemplo, uma dada fração de perda, em comparação com um processo cujo decaimento é mais rápido [25]. É o caso dos tráfegos de vídeo, onde um aumento na capacidade de armazenamento de uma fila submetida a este tráfego não implica em redução proporcional na fração de pacotes perdidos - seria necessário um aumento muito maior desta capacidade.

2.2.3 Espectro de potência

O espectro de potência é a transformada de Fourier da autocovariância [19, 7]. Portanto, traz consigo a mesma informação que a autocovariância, porém mapeada para o domínio da freqüência. Alguns autores aplicam esta transformada à autocorrelação ou ao termo $E[X(t)X(t+\tau)]$. Dependendo do que for usado, outros nomes podem ser atribuídos para este descritor, como densidade espectral, por exemplo [26, 3]. Supondo que o processo de chegada seja descrito por uma cadeia de Markov, o espectro de potência estará relacionado com os autovalores de sua matriz de transição Q. Q pode ter autovalores complexos e, neste caso, eles aparecem em pares conjugados. De uma maneira informal, pode-se dizer que para cada autovalor de Q tem-se uma forma de "sino" no espectro de potência, onde o centro deste sino é a parte imaginária do autovalor e sua largura média é dada pela componente real do autovalor. No domínio do tempo, pode-se dizer que as componentes reais dos autovalores representam um termo exponencial na função de autocorrelação e que as componentes imaginárias representam um termo periódico nesta mesma função [19]. Processos auto-similares [2, 15, 27] possuem valor infinito no espectro de potência para a freqüência zero, o que está em concordância com o decaimento hiperbólico da autocovariância, uma vez que o espectro de potência no ponto zero é igual à integral da autocovariância.

Em [28], Hajek e He concluem que a média e o espectro de potência não são suficientes para predizer o comportamento de um processo com relação à ocupação numa fila, principalmente quando o espectro de potência está concentrado em freqüências baixas. Seria necessário também modelar outro descritor, como a distribuição de probabilidade marginal do fluxo de chegada. Li e Hwang, em [7], propõem modelar a distribuição do fluxo de chegada e o espectro de potência para baixas freqüências com o intuito de predizer o comportamento numa fila. Em [19], ambos afirmam que o espectro de potência em baixas freqüências é o que tem maior influência sobre o comportamento numa fila, quando submetida ao tráfego em questão.

2.2.4 Índice de dispersão para contagem

O índice de dispersão para contagem, IDC, para um determinado tempo t, é definido como [29, 30]:

$$IDC(t) = \frac{Var[N_t]}{E[N_t]} \ ,$$

onde N_t é o número de chegadas de pacotes num tempo t.

Como a variância na chegada de pacotes está relacionada a rajadas, pode-se dizer que um IDC crescente no tempo significaria um aumento de rajadas no tráfego ao longo do tempo. É o que ocorre em processos com dependência de longa duração [2, 31]; eles possuem IDC que cresce monotonicamente com o tempo. Em processos **não** auto-similares, como os markovianos, o IDC tende para um valor finito. Esta característica do IDC, que separa processos com dependência de longa duração (por exemplo, os auto-similares) daqueles com dependência de curta duração, torna este descritor atrativo.

2.2.5 Parâmetro de Hurst

O parâmetro de Hurst (H) é um dos indicadores do grau de auto-similaridade que um processo possui [2, 31]. Auto-similaridade, em processos estocásticos, pode ser entendida como comportamentos estatísticos que se mantêm ao longo de múltiplas escalas de tempo. Este aspecto auto-similar pode ser, algumas vezes, percebido visualmente através de gráficos amostrais, mas é melhor caracterizado pelo comportamento da função de autocorrelação do tráfego quando submetido a agrupamentos cada vez maiores de amostras ao longo do tempo [2].

Em outras palavras, analisa-se inicialmente a função de autocorrelação da seqüência de tráfego original. Novamente a partir da seqüência original, obtém-se a média de cada grupo de m amostras consecutivas e analisa-se a função de autocorrelação para esta seqüência de médias, repetindo-se o procedimento para m cada vez maior. Num tráfego com características auto-similares, ter-se-ão funções de autocorrelação com aspecto igual para cada m diferente, ou quando m for muito grande.

Nos modelos mais convencionais, como processos de Poisson e cadeias de Markov com poucos estados, a agregação de tráfego tende ao ruído branco. Já para modelos que possuem a propriedade de auto-similaridade, esta agregação manteria o padrão originalmente observado.

Suponha que determinado processo estocástico X, estacionário, com média μ e variância σ^2, tenha função de autocovariância $cov(k)$, onde k é o k-ésimo instante de tempo discreto, tal que [2]:

$$\lim_{k\to\infty} cov(k) = k^{-\beta}C \ ,$$

$0 < \beta < 1$ e C constante.

Seja $cov^{(m)}(k)$ a função de autocovariância para agrupamentos de m amostras em seqüência, onde k é a média do k-ésimo agrupamento, um processo é dito assintoticamente auto-similar de 2ª ordem se:

$$cov^{(m)}(k) = r(k)$$

quando $m \to \infty$ e $k \geq 0$; e é dito exatamente auto-similar de 2ª ordem se:

$$cov^{(m)}(k) = r(k)$$

para $k \geq 0$. O grau de auto-similaridade é dado por H (chamado parâmetro de auto-similaridade ou de *"Hurst"*) e corresponde a:

$$H = 1 - \frac{\beta}{2} \ .$$

H igual a 0.5 indica que o tráfego não possui auto-similaridade. A medida que H se aproxima de 1, maior a intensidade da auto-similaridade no tráfego.

Outras propriedades pertinentes aos processos com auto-similaridade, e que também servem para caracterizar os processos como tal, seguem abaixo [23, 2]:

1. Seja $Var(X^{(m)})$ a variância da média de m amostras geradas pelo processo X. $Var(X^{(m)}) = \frac{\sigma^2}{m^{-\beta}}$ quando $m \to \infty$, ou seja, a variância da média amostral decai mais vagarosamente que $\frac{\sigma^2}{m}$, que é o caso dos modelos mais convencionais;

2. A função de autocovariância não é somável, ou seja, $\sum_k cov(k) = \infty$, devido à função de autocovariância decair hiperbolicamente ($cov(k) \sim k^{-\beta}C$, $k \to \infty$ e C constante);

3. A função de densidade espectral destes processos tem distribuição em torno da origem na forma
$$f(\lambda) \sim a\lambda^{-\gamma}\ ,\ \lambda \to 0,$$
sendo λ a variável aleatória no domínio da freqüência e a uma constante qualquer. Aqui, $\gamma = 1 - \beta$ e $f(\lambda) = \sum_k cov(k)e^{ik\lambda}$.

Nos modelos que não apresentam características auto-similares, $\lim_{k\to\infty} cov(k) = \rho^k$, $0 < \rho < 1$, ($cov(k)$ decai exponencialmente). Assim sendo, $cov^{(m)}(k) \to 0$ quando $m \to \infty$ e $\sum_k cov(k) < \infty$.

Um grande problema que existe é a estimação do parâmetro de Hurst para seqüências reais. Isso se deve ao fato deste parâmetro denotar o comportamento de um processo cujo tempo de vida tende para infinito. Vários são os métodos propostos que variam muito em complexidade e grau de acurácia [31]. Vários deles são capazes de dar respostas diferentes para uma mesma seqüência analisada, dependendo do tamanho da seqüência. Quanto maior a seqüência, mais eficazes se tornam os métodos.

Dentre os métodos mais usados, podemos citar o que analisa o decaimento da variância [2, 31] e outro que analisa o crescimento do IDC [31]. Ambos são equivalentes em termos de acurácia e se valem da taxa de decaimento (variância), ou crescimento (IDC), em escala *log x log* para estimar o H. Além destes, outros métodos também usados são a análise do R/S (*"Rescaled Adjusted Range Statistics"*) [2, 31], o método baseado em análise do periodograma (espectro de potência), que é medido no domínio da freqüência [31], e o método baseado em *"wavelets"* [9].

Vale lembrar que processos que possuem dependência de longa duração, também chamados de processos LRD ("Long Range Dependence") não necessariamente são processos auto-similares. Para ser LRD, basta que o processo possua autocorrelação com decaimento vagaroso ao longo do tempo, podendo este ser finito. Ainda assim, para processos LRD, o parâmetro de Hurst é usado para indicar o grau de dependência de longa duração que o processo possui. Todavia, em [32] é mostrado que o parâmetro H, por si só, não é suficiente para predizer o comportamento de uma fila quando submetida a um processo LRD. No entanto, é de notório conhecimento que, para processos exibindo dependência de longa duração, a perda decai muito lentamente a medida que o tamanho máximo da fila aumenta. Quanto maior o grau de dependência, mais lento é este decaimento. Outro estudo sobre a relevância do parâmetro de Hurst pode ser encontrado em [33].

2.2.6 Função Geradora de Momento Logarítmico

É definido em [34, 35] como sendo

$$\mathcal{B}(\theta, t) = \frac{1}{\theta t} \log E[e^{\theta \mathcal{X}[0,t]}] \quad ,$$

onde $\mathcal{X}[0,t]$ é o tráfego total produzido por uma fonte no intervalo $[0,t)$. Assume-se que $\mathcal{X}[0,t]$ tem incrementos estacionários.

O parâmetro t define uma escala de tempo onde o tráfego $\mathcal{X}[0,t]$ é acumulado e para a qual é extraída a média indicada na expressão. O parâmetro θ serve para ponderar os valores obtidos entre o valor médio e o valor máximo de $\mathcal{X}[0,t]$. Pode-se verificar que, quando $\theta \to 0$, a expressão converge para o valor médio ($\frac{E[\mathcal{X}[0,t]]}{t}$) e que, quando $\theta \to \infty$, a expressão converge para o valor de pico ($\frac{E[sup\ \mathcal{X}[0,t]]}{t}$, onde $sup\ x$ é o supremo de x) [35].

Seja Q o número de ocupações numa fila infinita e b um valor de referência qualquer. Se b for suficientemente grande e $P[Q > b]$ bastante pequena, tem-se que, para uma fila sendo alimentada por $\mathcal{X}$,

$$P[Q > b] \approx e^{-\theta b} \quad ,$$

onde θ é escolhido de acordo com a capacidade de serviço da fila [17]. De uma outra maneira, sendo $\gamma = P[Q > b]$, pode-se dizer que

$$\theta = \frac{\partial \gamma}{\partial b} \ \ e \ \ \theta t = \frac{\partial \gamma}{\partial C} \quad ,$$

onde C é a capacidade de serviço da fila [35].

Atribuindo-se valores para b e $P[Q > b]$, a capacidade C resultante é conhecida como "banda efetiva" ou "capacidade efetiva". Graças à sua propriedade aditiva, a banda efetiva é uma medida muito usada em controles de admissão de conexão

(CAC) para prever a capacidade exigida por uma fonte, ou um agregado delas, de forma que a perda não ultrapasse um determinado limite. Todavia, dado as características assintóticas das aproximações envolvidas, esses algoritmos costumam desperdiçar muita banda. Esta técnica é analítica por natureza e a interpretação dos parâmetros θ e t não é facilmente conseguida [17]. Além disso, em [36], é dito que o decaimento de $P[Q > b]$ com b, para processos de chegada com dependência de longa duração, não é exponencial. Isso invalida alguns métodos de controle que usam banda efetiva. Ainda assim, outros métodos podem ser desenvolvidos para suportar processos com dependência de longa duração.

2.3 Modelos de Fonte

Nesta seção são apresentados, de maneira compacta, alguns modelos tradicionalmente usados na literatura. São eles os processos de Poisson modulados por Markov (MMPP), os modelos markovianos com recompensa de taxa, os modelos auto-regressivos, alguns processos de renovação, modelos exibindo auto-similaridade e os que são aqui chamados de modelos híbridos, aqueles provenientes da junção de dois ou mais tipos de modelo diferentes. Cada "classe" de modelo tem suas características e limitações quanto ao casamento de determinados descritores, como será mostrado adiante.

2.3.1 MMPPs e derivados

O MMPP (*"Markov-Modulated Poisson Process"*) é um modelo markoviano onde cada estado representa a taxa média de um processo de Poisson [37, 23, 38, 39, 40], ou seja, tem-se uma cadeia de Markov modulando a taxa média de um processo de Poisson.

Este tipo de modelo é muito usado para caracterizar tráfegos de voz e dados [37, 39]. No caso de agregados de tráfegos de voz, é comum usar um modelo *"on-off"* para cada fonte de voz, onde o estado *"off"* representa silêncio e tem taxa de transmissão zero. A cadeia de Markov resultante da multiplexação destas fontes de voz é semelhante ao modelo de nascimento e morte descrito em 2.3.2. Um procedimento similar foi também proposto para o caso de tráfegos de vídeo com mudança de cena vagarosa [39]. Este método para a multiplexação de fontes *"on-off"* se apóia no fato da soma de processos de Poisson resultar num novo processo de Poisson cuja taxa média corresponde à soma das taxas médias dos processos originais.

O modelo MMPP também é usado para se modelar vários outros tipos de tráfego. Em [41], é proposto um MMPP com oito estados, baseado em histograma linear, para representar uma fonte de vídeo com um dado algoritmo de compressão. Em [42], os processos de Poisson que representam as taxas existentes no tráfego real modelado são calculados de forma que haja uma justaposição dos intervalos $[\lambda_i - \delta_i, \lambda_i + \delta_i]$ para cada i, onde $lambda_i$ é a taxa média do processo de Poisson no estado i e δ_i o desvio em relação à $lambda_i$, segundo uma determinada probabilidade, para as taxas geradas neste estado. $\lambda_n + delta_n$, onde n representa o estado com $lambda_i$ máximo, deve coincidir com a taxa máxima do tráfego real. Estes modelos têm a desvantagem de somente capturar dependências de curta duração. Em [19], Li *et al* definem um MMPP com diagrama de estados em círculo chamado CMPP (*"Circulant Modulated Poisson Process"*), capaz de se ajustar às distribuições de taxa e espectros de potência genéricos. A parametrização deste modelo é muito complicada, assim como também é a obtenção do espectro de potência de uma seqüência real. A modelagem por CMPP é feita através de uma ferramenta chamada *SMAQ Tool* onde são efetuadas: a captura da distribuição de taxa e do espectro de potência; a parametrização do modelo envolvendo o número de estados, as taxas de transição e as taxas médias dos processos de Poisson associados; e a análise do CMPP resultante para filas com serviço exponencial. As medidas obtidas para um CMPP aplicado a uma fila com serviço exponencial podem ser tanto em estado estacionário como transiente.

Em [8, 43], Andersen e Nielsen propõem um MMPP para aproximar dependência de longa duração. Naquele modelo, vários MMPPs de 2 estados são superpostos, cada um com suas respectivas taxas de transição e taxas médias para os processos de Poisson associados. Essas taxas são escolhidas de forma tal que os MMPPs possuam, cada um, uma autocorrelação com decaimento exponencial a taxas cada vez menores. Ao se somar as autocorrelações destes MMPPs para se obter a autocorrelação do processo superposto, o decaimento desta será aproximadamente hiperbólico para um tempo finito. Seja $\gamma(k)$ a autocorrelação do processo superposto e $\gamma_i(k)$ a autocorrelação do i-ésimo MMPP:

$$\gamma(k) = \sum_{i=1}^{d} \gamma_i(k) \approx \sigma^2 k^{-\beta} \ , \ \ 1 \leq k \leq 10^n$$

onde d é o número de MMPPs de 2 estados, σ^2 é a variância, n é a escala de tempo e β é tal que $H = 1 - (\beta/2)$. Os parâmetros deste modelo são a taxa média, a autocorrelação para intervalo de observação igual a 1, o parâmetro de *Hurst*, o número de MMPPs de 2 estados e o número de escalas de tempo a serem modeladas. O uso do parâmetro de Hurst na parametrização do modelo torna este sujeito aos problemas de estimação deste parâmetro em tráfegos reais.

Existem várias expressões analíticas capazes de fornecer medidas de interesse para fontes MMPP alimentando filas com taxa de serviço exponencial. Vale ressaltar que, se a taxa média do processo de Poisson associado a qualquer dos estados da cadeia resulta em intervalos médios de transmissão que são muito menores que o tempo médio de permanência no respectivo estado, então podemos aproximar este MMPP por um modelo markoviano de fluido (ver seção 2.3.2).

2.3.2 Modelos markovianos com recompensa de taxa

Trata-se de uma cadeia de Markov onde cada estado possui uma taxa determinística associada. Seja uma cadeia de Markov de tempo contínuo $\mathcal{M} = \{\mathcal{X}(t) : t = 0, \ldots \infty\}$ e espaço de estados discreto. Considere um vetor de recompensas (taxas) $\lambda =< \lambda_1, \lambda_2, \ldots, \lambda_N >$ associado à cadeia moduladora $\mathcal{M}$. O modelo é tal que o sistema, enquanto no estado i, transmite a uma taxa constante λ_i.

Note que num modelo MMPP, o estado i representa uma transmissão de dados a intervalos exponencialmente distribuídos com média $\frac{1}{\lambda_i}$. Quando esta média é muito menor que o tempo médio de permanência no estado i, diz-se também que trata-se de modelo markoviano de fluido.

A definição dos parâmetros para esta classe de modelos, assim como para os

MMPPs, pode ser feita através do casamento de descritores de tráfego do modelo com descritores obtidos para o tráfego real.

Um exemplo disso é o conhecido modelo de nascimento e morte usado para tráfegos de vídeo que não possuam mudanças bruscas de cena (e.g., videotelefonia) [44, 45, 37]. Como dito anteriormente, neste modelo diversas fontes *"on-off"* idênticas representando mini-fontes individuais são multiplexadas resultando num modelo de nascimento e morte. Cada estado deste modelo representa o número de fontes ativas. Uma vez determinado o número máximo de mini-fontes, e por conseguinte o número de estados da cadeia resultante, bastam as taxas de transição de estado da fonte *"on-off"* que representa cada mini-fonte e a taxa de transmissão de dados no estado *"on"* desta mini-fonte para se obter o modelo completo. Ou seja, independente do número de estados deste modelo, bastam apenas 4 parâmetros para determiná-lo completamente. Isto mostra que modelos markovianos com um grande número de estados não necessariamente demandam um número de parâmetros igualmente grande.

Outro exemplo de construção de modelos markovianos com recompensa são os modelos baseados em histograma linear e em histograma logarítmico. Como o próprio nome diz, eles são construídos de forma a casar as distribuições de probabilidade (linear ou logarítmica) da taxa de transmissão. Para tal, divide-se a seqüência de tráfego que se deseja modelar em diferentes níveis de taxa. Estes níveis podem ter suas fronteiras igualmente espaçadas (histograma linear) ou exponencialmente espaçadas (histograma logarítmico) [6]. Para cada nível de taxa, associa-se a taxa média das amostras compreendidas neste nível como recompensa de estado. O número de estados do modelo será igual ao número de níveis. As probabilidades de transição entre os estados são calculadas de acordo com as freqüências de transição entre os níveis.

O exemplo do modelo de nascimento e morte é destinado a casos particulares de seqüências de vídeo que possuam dependência de curta duração - seqüências de vídeo, em geral, apresentam dependência de longa duração. É justamente neste aspecto que têm se baseado as principais críticas aos modelos markovianos: eles não são capazes de capturar dependências de longa duração pois o parâmetro H

é sempre 0.5, qualquer que seja a cadeia de Markov. Esta crítica foi e ainda é rebatida com o argumento de que é possível construir modelos markovianos que exibam dependência para um tempo finito mas longo - trata-se de uma questão de adequar as transições do modelo com as escalas de tempo encontradas no tráfego a ser modelado. Neste caso, o modelo markoviano torna-se capaz de gerar seqüências finitas que, a partir de estimadores de parâmetro de Hurst, apresentem H maior que 0.5 para este tempo finito. Dependendo do valor atribuído a este tempo limite, o modelo markoviano pode tornar-se suficientemente preciso para prever medidas de desempenho mais tradicionais, como a perda numa fila finita [3, 4, 46, 47, 8, 43].

Este é o caso do modelo pseudo auto-similar, proposto em [3]. A Figura 2.1 mostra o diagrama de estados deste modelo para uma cadeia com 5 estados. As probabilidades de transição são determinadas em função de 2 parâmetros, a e b.

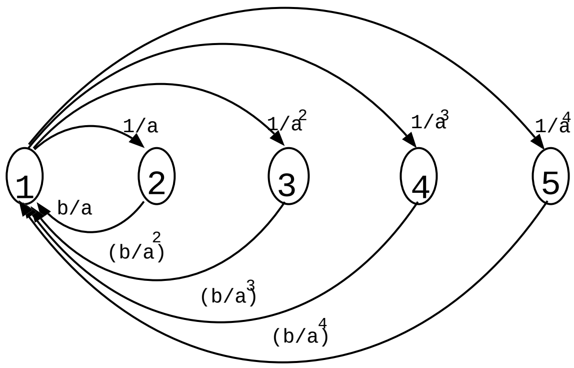


Figura 2.1: Modelo pseudo auto-similar

Este modelo foi concebido para caracterizar tráfegos em canais ATM. Trata-se de uma cadeia de Markov de tempo discreto onde, a cada *slot* de tempo, pode haver a transmissão de apenas uma célula ou nenhuma. Portanto, podemos considerar este modelo como uma fonte *on-off* onde o estado 1 é o estado *on* com recompensa igual a 1 célula/slot. Os demais estados têm recompensa zero e representam estados de silêncio (*off*). A cada transição de *on* para *off*, tem-se um tempo em silêncio cuja duração é exponencialmente distribuída e possui média que varia de acordo com o estado *off* de destino. Cada estado *off* representa uma escala de tempo de silêncio diferente. Por conseguinte, o intervalo de tempo entre transmissões tem distribuição de cauda longa (ver seções 2.3.5 e 4.1). O parâmetro b determina a taxa média, o parâmetro a indica o parâmetro de *Hurst* e o número de estados da cadeia indica por quanto tempo têm-se uma dependência significativa entre amostras (sendo um tempo longo, tem-se uma dependência de longa duração). Em [46] e [47], não são mostradas expressões para o cálculo de H ou alguma forma de estimar um tempo limite para a dependência. É proposto em [47] que a seja achado de forma recorrente, variando-se o valor deste e medindo o parâmetro de Hurst resultante da simulação do modelo. Na seção 5.4 encontram-se a expressão para a taxa média do modelo e também uma expressão, proposta aqui neste trabalho de tese, para encontrar a.

Os estudos da literatura têm mostrado que os modelos markovianos são vantajosos pela sua flexibilidade e pela quantidade de expressões analíticas conhecidas para a obtenção de diversos descritores de tráfego e medidas de interesse [16, 48, 49]. Dependendo de como são definidas as transições da cadeia, é possível obter modelos que capturem, por exemplo, a distribuição marginal da taxa de envio de dados assim como o seu espectro de potência e, conseqüentemente, a autocorrelação [7].

2.3.3 Modelos auto-regressivos

Nos modelos auto-regressivos [23], a variável aleatória de uma seqüência é uma função explícita das variáveis anteriores dado uma janela de tempo. A expressão para esta variável aleatória, portanto, é uma fórmula recursiva. Alguns destes modelos foram propostos para caracterizar seqüências de vídeo com pouca variação em sua taxa de transmissão, ou seja, que não possuam mudanças bruscas de cena.

Existem dois tipos mais conhecidos de modelos auto-regressivos:

- Modelo auto-regressivo linear.
- Modelo TES (Transform Expand Sample).

Modelo auto-regressivo linear

A seqüência de variáveis aleatórias $\{X_n\}$ é obtida da seguinte maneira para o caso de modelos auto-regressivos lineares:

$$X_n = a_0 + \sum_{r=1}^{P} a_r X_{n-r} + \varepsilon_r \ , \ n > 0 \ ,$$

onde ε_r é i.i.d. e possui média η, normalmente próxima de zero. P é a ordem da auto-regressão (tamanho da janela).

Este costuma ser um modelo interessante para tráfegos autocorrelacionados dentro de um certo limite de tempo, pois consegue-se reproduzir a função de autocorrelação empírica até um intervalo de tempo igual a P. Isto ocorre porque o modelo se vale das P variáveis anteriores para calcular a próxima variável. Todavia, este modelo não consegue reproduzir uma distribuição marginal de taxa qualquer [23]. No modelo auto-regressivo linear de ordem 1 (janela 1), a função de distribuição de probabilidade de X tem forma de sino centrada na média e sua função de autocovariância tem decaimento exponencial.

Modelo TES

O modelo TES [23, 50, 51] objetiva reproduzir a função de autocorrelação e a distribuição marginal de taxa, ambas empíricas e calculadas a partir de uma determinada seqüência. Para tal é usada uma função $\{U_n\}_{n=0}^{\infty}$ dita de *"background"* onde, aplicando-se a inversa da função de distribuição de probabilidade de taxa encontrada ($\hat{H}$), obtém-se uma função (D_h) que possui a mesma densidade de distribuição de probabilidade que $\hat{H}$, ou seja, $\hat{h}$. Assim, ter-se-á uma seqüência de variáveis aleatórias com a mesma distribuição de $\hat{H}$.

Para que a teoria da inversa da função se aplique, é necessário que U_n tenha distribuição uniforme:

$$U_n^+ = \begin{cases} U_0 & , n = 0 \\ \langle U_{n-1}^+ + V_n \rangle & , n > 0 \end{cases} \qquad U_n^- = \begin{cases} U_n^+ & \text{, n par} \\ 1 - U_n^+ & \text{, n ímpar} \end{cases}$$

sendo U_0 uniforme entre $[0,1]$ e V_n definido abaixo. Os indicadores $+$ e $-$ designam seqüências com autocorrelações positivas e negativas, respectivamente. U_n^- tem autocorrelação negativa porque, quanto maior U_{n-1}^-, menor será U_n^-, e vice-versa.

Os colchetes angulares indicam que apenas é computada a parte fracionária da operação. A seqüência estacionária $\{V_n\}_{n=0}^{\infty}$, chamada função de inovação, possui variáveis i.i.d. com distribuição qualquer e é independente de $U_0, ..., U_{n-1}$. U_n tem distribuição uniforme em $[0,1)$ [50].

D_h é dita função de *"foreground"*, ou distorção, e corresponderá ao processo que se deseja modelar com distribuição de probabilidade marginal $\hat{H}$. A expressão para $\hat{H}$ é encontrada a partir de inspeção nos dados experimentais [50]. Tem-se que

$$D_h = \hat{H}^{-1}(U) \ .$$

O anacronismo TES significa *"Transform Expand Sample"*. A etapa de transformação corresponde a, no caso de U_n^-, usar os pontos de U_n^+ para gerar um novo ponto (U_n^+ não tem etapa de transformação); a etapa de expansão verifica-se ao se somar a variável aleatória V a U_{n-1}^+, o que corresponde a expandir o campo de possibilidades de U_n; e, a partir deste campo, a escolha de U_n, segundo a distribuição de V, corresponde à etapa de amostragem [50]. As características da função de *"background"*, em especial da distribuição de V, é que determinarão a autocorrelação desejada.

Devido ao fato do campo de possibilidades de U_n muitas vezes se encontrar na fronteira de 0 com 1 (podemos imaginar um círculo com circunferência unitária), acontecerá de se ter um gráfico de amostras resultantes que ficará temporariamente saltando entre valores muito distantes (perto de 0 e perto de 1). Para se diminuir este aspecto "dente de serra" do gráfico de amostras, o que é indesejável, é aplicada uma função de suavização (*"Stiching Transformation"*) à U_n. Esta função tem por finalidade efetuar uma transformação que aproxime os valores de U que se encontram próximas ao limite em 0 e ao limite em 1:

$$S_\xi(U) = \begin{cases} \frac{U}{\xi} & , 0 \leq U < \xi \\ \frac{1-U}{1-\xi} & , \xi \leq U < 1 \end{cases}$$

$S_\xi(U)$ também possui distribuição uniforme. A função de *"foreground"* será, portanto [50]:

$$D_y = H^{-1}(S_\xi(U)).$$

A escolha de ξ, assim como da distribuição de V e da distorção D_h, determinarão a função de autocorrelação do modelo. Vale notar que $S_\xi(0)$ é igual a $S_\xi(1)$, desde que $0 < \xi < 1$ e, se $\xi = 0.5$, $S(U_n^+) = S(U_n^-)$. A medida que ξ se afasta de 0.5 aumenta-se a discrepância entre $S(U_n^+)$ e $S(U_n^-)$ [50].

A função de autocorrelação da seqüência em análise é obtida via dados experimentais e, posteriormente, ajusta-se ξ e V no modelo TES para que este forneça uma função de autocorrelação que se aproxime satisfatoriamente da função empírica encontrada.

O processo TES está incluído numa família de processos chamada ARM, *"Autoregressive Modular Processes"* [17], que também inclui o E-TES (*"Extended TES"*). O E-TES é uma generalização do TES que usa uma seqüência de pré-inovação, $\{Z_n\}$, que alimenta um filtro que produzirá a seqüência de inovação $\{V_n\}$ [17]:

$$V_n = \sum_{i \geq 0} \alpha_i Z_{n-i} \quad .$$

Apesar da flexibilidade deste modelo para capturar a distribuição de taxa e a função de autocorrelação, a parametrização do mesmo é complexa e necessita de ferramentas de *software* sofisticadas, como o *TES tool* [51].

2.3.4 Processo SRP

O *"Spatial Renewal Process"* (SRP) [17] utiliza 2 processos de *"background"*, $\{X_n\}$ e $\{T_n\}$, onde cada um modela, de forma independente, a distribuição marginal e a autocorrelação, respectivamente. Ambas seqüências são i.i.d.. O processo SRP é composto por uma cadeia de períodos de renovação onde o n-ésimo período tem tamanho T_n e o caminho amostral ao longo deste período tem valor X_n [17]. A distribuição marginal do processo será a mesma de $\{X_n\}$. A autocorrelação do processo está relacionada com a função de distribuição de $\{T_n\}$, $F_T(t)$:

$$1 - \rho(t) = \mu^{-1} \int_0^t (1 - F_T(u))du \ \ , t \geq 0$$

$$-\frac{d}{dt}\rho(t) = \mu^{-1}(1 - F_T(t)) \ \ , \rho(0) = 1 \ , t \geq 0 \quad ,$$

onde $\mu = \int_0^\infty (1 - F_T(u))du$. Para garantir que $F_T(t)$ é uma distribuição válida, $\rho'(t) < 0$ e $\rho''(t) < 0$ para todo $t > 0$.

Este modelo não é muito comum na literatura e foi aqui comentado pelo fato dele ser utilizado em alguns dos testes descritos na seção 3.1 pois, assim como o modelo

TES, este modelo é capaz de aproximar a distribuição de taxa e a função de autocorrelação. No entanto, também possui a desvantagem de ter uma parametrização complexa.

2.3.5 Modelos exibindo auto-similaridade

Os modelos com auto-similaridade [23, 2, 10] são aqueles que possuem uma função de autocorrelação cujo decaimento ao longo do tempo está de acordo com a definição de auto-similaridade. Seguem abaixo alguns modelos com auto-similaridade mais conhecidos. De uma maneira geral, este modelos conseguem capturar o parâmetro de Hurst e a taxa média. Em alguns casos, a variância e a autocorrelação para durações curtas também são capturados, como será visto a seguir.

Fontes *On-Off* com distribuição de cauda longa

A modelagem do fenômeno de auto-similaridade pode ser conseguida agregando-se vários processos de renovação que possuam tempos entre renovações independentes e identicamente distribuídos com distribuição do tipo [2]:

$$P\{U \geq u\} \sim u^{-\alpha} h(u) \ ,$$

$\alpha = 3 - 2H$, $1 < \alpha < 2$ e $h(u) \sim C$ quando $u \to \infty$. H é o parâmetro de Hurst e C uma constante.

Esta característica da distribuição acima é chamada de cauda longa, ou *"heavy-tail"*. Este termo vem do fato de que o decaimento hiperbólico é bem mais suave do que o decaimento exponencial, o que provoca uma "cauda" mais longa no formato deste tipo de função.

Esta maneira de se modelar tráfegos auto-similares é também usada para explicar porque o tráfego gerado na *Internet* possui características auto-similares. Como qualquer tráfego pode ser modelado por uma fonte *"ON-OFF"* genérica, se a distribuição dos tempos de permanência em cada estado possuir cauda longa, a multiplexação destas fontes resultará num tráfego auto-similar.

"Fractional Brownian Motion" (FBM) e *Fractional Gaussian Noise* (FGN)

Um processo FBM (*"Fractional Brownian Motion"*), Z_t, é um processo exatamente auto-similar de segunda ordem com distribuição Gaussiana de média zero [11, 26, 2]. Trata-se de um processo de incrementos estacionários cuja variância é igual a $|t|^{2H}$ para todo t, onde H é o parâmetro de Hurst. A derivada de Z_t é chamada de FGN (*"Fractional Gaussian Noise"*). A função de autocorrelação de um processo FGN é [2]:

$$r(k) = 1/2(|k+1|^{2H} - |k|^{2H} + |k-1|^{2H}) \ , \ k > 0 \ .$$

Os parâmetros do FGN são a média, a variância e o H. Assim como o FBM, o FGN também será exatamente auto-similar de segunda ordem com distribuição Gaussiana, mas não será crescente e sim estacionário.

Fractional ARIMA (*Autoregressive Integrated Moving Average*)

O F-ARIMA é um processo assintoticamente auto-similar de segunda ordem. É capaz de modelar dependências de curta e longa duração, enquanto que o FBM, por ter uma estrutura mais rígida de autocorrelação, só captura dependências de longa duração. O F-ARIMA tem distribuição Gaussiana e seus parâmetros costumam ser a média, a variância e o parâmetro de Hurst. Na verdade, estes 3 parâmetros correspondem a uma forma reduzida de se usar o F-ARIMA, que é bem mais genérico. Além da média e da variância, o F-ARIMA possui mais 3 parâmetros: p, q e d. p indica a ordem da auto-regressão, q indica o tamanho da janela para a média móvel e d equivale a $H + 1/2$, onde H é o parâmetro de Hurst. Como os parâmetros p e q são difíceis de se estipular, costuma-se atribuir valores fixos a cada um deles (0 ou 1, em geral [2]), restando apenas d como parâmetro. p e q são responsáveis pela captura da dependência de curta duração [26].

2.3.6 Modelos híbridos

Há também muitas propostas de modelagem que procuram juntar características de modelos distintos num mesmo processo. Tais modelos são aqui classificados como modelos híbridos.

Os casos mais comuns de modelos híbridos são os que usam cadeias de Markov para modular os parâmetros de outro processo estocástico, assim como ocorre com o MMPP.

Um exemplo bastante conhecido é o *Markov Renewal Modulated* TES (MRMT) [13], onde um processo TES é modulado por uma cadeia de Markov com o intuito de modelar tráfegos de vídeo VBR. Para tal, uma seqüência de vídeo é dividida em cenas e estas cenas são agrupadas em classes. Para cada classe de cena tem-se um processo TES associado modelando a taxa de bits e a duração da cena. A transição entre cenas é governada pela cadeia de Markov.

Um outro exemplo pode ser encontrado em [14], onde os autores usam distribuições Gamma [22] e cadeia de Markov para modelar tráfegos de vídeo codificados em MPEG. Neste modelo, a seqüência de vídeo é dividida e classificada por GOP (*Group of Pictures*, ver seção 5.1.1). Cada classe de GOP tem uma distribuição Gamma associada modelando o tamanho dos quadros **I**, **P** e **B**. A transição entre classes de GOP é regida por uma cadeia de Markov.

2.4 Expressões Analíticas para Modelos Markovianos com Recompensas de Taxa

Como o modelo a ser proposto neste trabalho de tese, assim como boa parte dos exemplos a serem utilizados, são modelos markovianos com recompensas de taxa, seguem abaixo as expressões analíticas para alguns dos descritores definidos na seção 2.2. Estas expressões também podem ser encontradas em [48, 49].

O cálculo dos descritores transientes são baseados na técnica de uniformização [52]. Neste método, uma cadeia de Markov de tempo contínuo é transformada numa cadeia análoga de tempo discreto, para a qual medidas transientes podem ser calculadas. Trabalhar com o problema em tempo discreto é mais vantajoso dado uma maior simplicidade em sua análise.

Considere uma cadeia de Markov de tempo contínuo com gerador infinitesimal $\mathbf{Q}$ (matriz de taxa). Seja Λ um valor maior ou igual ao máximo valor absoluto dos elementos da diagonal de $\mathbf{Q}$, define-se que

$$\mathbf{P} = \mathbf{I} + \mathbf{Q}/\Lambda ,$$

onde $\mathbf{I}$ é a matriz identidade.

É possível mostrar que $\mathbf{P}$ é uma matriz estocástica e que podemos transformar a cadeia de Markov original numa cadeia discreta com matriz de transição $\mathbf{P}$, onde o tempo de permanência em qualquer estado antes de uma transição (incluindo transições para o mesmo estado) tem distribuição exponencial com média $1/\Lambda$.

A partir desta transformação, várias medidas transientes podem ser encontradas. Maiores detalhes encontram-se em [52]. Por exemplo, o vetor de probabilidades de estado no tempo t ($\mathbf{p(t)}$) pode ser facilmente calculado por:

$$p(t) = p(0) \sum_{n=0}^{\infty} e^{-\Lambda t} \frac{(\Lambda t)^n}{n!} P^n , \tag{2.1}$$

onde $\mathbf{p(0)}$ é a distribuição de probabilidade inicial dos estados. O parâmetro Λ é chamado taxa de uniformização [53, 54, 52].

Esta maneira de representar a cadeia determina o número de eventos que ocorrem num intervalo de tempo t como sendo um processo de Poisson, o que facilita os cálculos e a interpretação probabilística do sistema. A taxa média do processo de Poisson é, justamente, Λ e os eventos ditados por este processo são as transições da cadeia discreta. Assim sendo, as medidas transientes são calculadas condicionando-se o tempo de observação ao número de transições da cadeia e, posteriormente, descondicionando-se.

Vários dos cálculos mostrados a seguir fazem uso da distribuição de probabilidade estacionária e da distribuição de probabilidade transiente dos estados da cadeia de Markov. Tais medidas podem ser obtidas por vários métodos, alguns conhecidos na literatura [54] e outros recentemente desenvolvidos [55]. Apenas as expressões necessárias para a formulação analítica dos descritores são apresentadas.

Taxa Média e variância para modelos markovianos

A taxa média do modelo é calculada em função da probabilidade estacionária π_i de se estar no estado i da cadeia, que possui taxa de recompensa λ_i. Desta forma, para uma cadeia com N estados:

$$E[\lambda] = \sum_{i=1}^{N} \pi_i \lambda_i \ .$$

O cálculo da variância da taxa é dado por:

$$Var[\lambda] = \sum_{i=1}^{N} \pi_i (\lambda_i - E[\lambda])^2 \ .$$

Autocovariância para modelos markovianos

A função de autocovariância, como anteriormente mostrada, corresponde a:

$$Cov(\tau) = E\{X(t)X(t+\tau)\} - \mu^2 \ ,$$

onde τ é o tempo e μ é a taxa média, anteriormente calculada.

Usando o conceito de uniformização, a medida $E\{X(t)X(t+\tau)\}$ será, para o caso de um modelo markoviano com N estados:

$$E[X(t)X(t+\tau)] = \sum_{m=0}^{\infty} \sum_{n \geq m} E[X(m)X(n)](e^{-\Lambda t} \frac{(\Lambda t)^m}{m!})(e^{-\Lambda \tau} \frac{(\Lambda \tau)^{(n-m)}}{(n-m)!}) \ , \qquad (2.2)$$

onde

$$E[X(m)X(n)] = \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j P[X(m) = i, X(n) = j] \ .$$

Considerando que a cadeia é homogênea e estacionária, tem-se que:

$$\begin{aligned} E[X(m)X(n)] &= \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j P[X(n) = j | X(m) = i] P[X(m) = i] \qquad (2.3) \\ &= \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j p_{ij}(n-m) \pi_i \qquad (2.4) \end{aligned}$$

$$= \sum_{i=1}^{N} \lambda_i \pi_i \sum_{j=1}^{N} \lambda_j p_{ij}(n-m) \tag{2.5}$$

$$= E[X(0)X(k)], \text{ para } k = m - n \tag{2.6}$$

$$= \sum_{i=1}^{N} \lambda_i \pi_i \sum_{j=1}^{N} \lambda_j p_{ij}(k) \tag{2.7}$$

onde $p_{ij}(k)$ é a probabilidade do modelo transicionar do estado i para j em k passos.

Definindo-se que

$$\gamma(n) = \lambda(\mathbf{P}(n))^T \ , \tag{2.8}$$

onde $\mathbf{A}^T$ é a transposta da matriz $\mathbf{A}$ e $\mathbf{P}(n)$ é a matriz cujo i,j-ésimo elemento é igual a $p_{ij}(n)$, tem-se que:

$$E[X(m)X(n)] = \sum_{i=1}^{N} \lambda_i \pi_i \gamma_i(n-m) \ , \tag{2.9}$$

onde $\gamma_i(n-m)$ é o i-ésimo elemento do vetor $\gamma(n-m)$.

Sabendo-se que a cadeia é homogênea e estacionária, e logo $E[X(t)X(t+\tau)]$ depende somente da diferença de tempo τ, a partir de (2.9) e (2.2) obtém-se que:

$$E[X(t)X(t+\tau)] = \sum_{i=1}^{N} \lambda_i \pi_i \sum_{n=0}^{\infty} \gamma_i(n) e^{-\Lambda\tau} \frac{(\Lambda\tau)^n}{n!} \tag{2.10}$$

Sejam dois vetores não negativos $\mathbf{x} = \langle x_1, \ldots, x_N \rangle$ e $\mathbf{y} = \langle y_1, \ldots, y_N \rangle$, o seu produto interno é dado por $\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{N} x_i y_i$. Desta maneira, a equação (2.10) pode ser reescrita da forma

$$E[X(t)X(t+\tau)] = \lambda \cdot \pi \cdot \delta(\tau) \tag{2.11}$$

onde

$$\delta(\tau) = \sum_{n=0}^{\infty} \gamma(n) e^{-\Lambda\tau} \frac{(\Lambda\tau)^n}{n!} \ . \tag{2.12}$$

Note que, na equação (2.12), o vetor $\gamma(n)$ pode ser facilmente calculado a partir da recursão $\gamma(n) = \gamma(n-1)\mathbf{P}^T$ com $\gamma(0) = \lambda$.

Nota-se também que, como a cadeia de Markov é estacionária, a medida mencionada independe do tempo inicial t, mas apenas de τ. Além disso, a soma infinita em 2.12 pode ser truncada de maneira a se obter os resultados com uma dada tolerância.

Índice de dispersão para modelos markovianos

Recordando,

$$IDC(t) = \frac{Var[N(t)]}{E[N(t)]} \ ,$$

onde $Var[N(t)] = E[N^2(t)] - E[N(t)]^2$ e $N(t)$ é o número de chegadas no intervalo t.

Seja $\lambda(t)$ a taxa de chegada de pacotes no instante t, note que $N(t) = \int_0^t \lambda(s)ds$ e portanto

$$E[N(t)] = \int_0^t E[\lambda(s)]ds \ .$$

$E[\lambda(s)]$ pode ser calculado uniformizando a cadeia e usando resultados para cadeias de Markov com recompensas (ver [52]). Então,

$$E[N(t)] = \int_0^t \sum_{n=0}^{\infty} e^{-\Lambda s} \frac{(\Lambda s)^n}{n!} \lambda \cdot \mathbf{v}(n) ds \ ,$$

onde $\mathbf{v}(n)$ é o vetor de probabilidade de estados da cadeia após n passos.

Resolvendo a integral [52] tem-se que:

$$E[N(t)] = t \sum_{n=0}^{\infty} e^{-\Lambda t} \frac{(\Lambda t)^n}{n!} \left[\frac{\sum_{j=0}^{n} \lambda \cdot \mathbf{v}(j)}{n+1} \right] \ . \qquad (2.13)$$

O termo entre colchetes pode ser escrito recursivamente da forma

$$f(n) = \frac{n}{n+1} f(n-1) + \frac{1}{n+1} [\lambda \cdot \mathbf{v}(n)]$$

$$f(0) = \lambda \cdot \mathbf{v}(0) \ ,$$

onde $\mathbf{v}(0)$ é o vetor de probabilidades iniciais.

Em [56], é desenvolvido um método para calcular o segundo momento de uma variável aleatória que é uma função do estado de uma cadeia de Markov. Para o cálculo do IDC, temos que obter o segundo momento do número de pacotes que chegam num período de $[0, T]$. Usando os resultados de [56] tem-se que:

$$N(T)^2 = [\int_0^T \lambda(t)dt]^2 = 2 \int \int_{0<s<t<T} \lambda(s)\lambda(t) ds dt \ ,$$

portanto,

$$E[N(T)^2] = 2 \int \int_{0 \leq s \leq t \leq T} E[\lambda(s)\lambda(t)] ds dt \ .$$

$E[\lambda(s)\lambda(t)]$ é dado por (2.2), logo:

$$E[N(T)^2] = 2\sum_{0\le i}\sum_{j\ge i} E[\lambda_i\lambda_j]\int\int_{0<s<t<T} e^{-\Lambda s}\frac{(\Lambda s)^i}{i!}e^{-\Lambda(t-s)}\frac{(\Lambda(t-s))^{(j-i)}}{(j-i)!}dsdt\ .$$

Calculando a integral acima obtém-se que

$$E[N(T)^2] = \sum_{0\le i}\sum_{j\ge i} E[\lambda_i\lambda_j]\left\{\frac{2}{(\Lambda)^2}E_{j+2,\Lambda}(T)\right\}$$

onde $E_{j+2,\Lambda}(T)$ é a distribuição de Erlang com $(j+2)$ estágios.

Nota-se agora que $E[\lambda_i\lambda_j]$ foi obtido em (2.4). Substituindo (2.4) na equação acima e trocando a ordem dos somatórios, obtém-se que

$$E[N(T)^2] = \frac{2}{\Lambda^2}\sum_{j=0}^{\infty} E_{j+2,\Lambda}(T)\sum_{k=1}^{N}\lambda_k D_k^j\ , \tag{2.14}$$

onde

$$D_k^j = \sum_{i=0}^{j}\sum_{l=1}^{N}\lambda_l p_{lk}(j-i)v_k(j).$$

e $v_k(j)$ é a probabilidade da cadeia estar no estado k no passo j. D_k^j pode ser obtido recursivamente por:

$$D_k^j = \sum_{s=1}^{N} D_s^{j-1}p_{sk} + \lambda_k v_k(j)\ ,$$

com $D_k^0 = \lambda_k v_k(0)$.

Medidas de tempo operacional

Algumas medidas utilizadas no neste trabalho para caracterizar tráfegos são baseadas em medidas de tempo operacional, muito usadas em sistemas de confiabilidade. As expressões para estas medidas foram obtidas de [53] e o detalhamento destes cálculos pode ser encontrado em [54, 53].

No capítulo 3 são apresentadas medidas relativas a períodos de rajada, que são estatísticas relacionadas com períodos nos quais o tráfego permanece acima de um determinado nível de taxa. O cálculo destes descritores relativos a períodos de rajada, para uma fonte modelada por uma cadeia de Markov com recompensas de taxa, pressupõe a definição de um subconjunto $\mathcal{R}$ de estados do modelo onde a taxa da fonte esteja acima do nível em questão.

Seja r um estado do modelo pertencente a $\mathcal{R}$. A variável aleatória *duração de um período de rajada em um dado intervalo iniciado no estado* r é igual ao tempo total em $\mathcal{R}$, iniciado a partir de r, antes do modelo sair deste subconjunto.

Seja DR a duração de um período de rajada, $\overline{DR(t)}$ a duração média de um período de rajada num intervalo t e $\overline{DR}$ a duração média estacionária de um período de rajada, onde $\overline{DR}$ é $\lim_{t\to\infty} \overline{DR(t)}$. $P[DR > t]$, $\overline{DR(t)}$ e $\overline{DR}$ podem ser calculados da mesma maneira que medidas de confiabilidade, obtidas avaliando-se tempos operacionais em modelos de disponibilidade.

Portanto, de [54, 53] obtém-se imediatamente as seguintes expressões:

$$P[DR > t] = \sum_{n=0}^{\infty} e^{-\Lambda t}\frac{\Lambda t}{n!}\Omega[n, n+1] \ , \tag{2.15}$$

$$\overline{DR(t)} = \frac{1}{\Lambda}\sum_{n=0}^{\infty} E_{n+1,\Lambda}(t)\Omega[n, n+1] \ , \tag{2.16}$$

$$\overline{DR} = \frac{1}{\Lambda}\sum_{n=0}^{\infty} \Omega[n, n+1] \ . \tag{2.17}$$

$E_{n+1,\Lambda}(t)$ é a distribuição de Erlang com $(n+1)$ estágios.

$\Omega[n, k]$ pode ser calculado recursivamente por

$$\Omega[n, k] = \sum_{i=1}^{N} \Omega[n, k, a_i]$$

e

$$\Omega[n, k, a_i] = \sum_{j=1}^{N} \Omega[n-1, k-1, a_j]p_{ji}, \quad \text{se } a_i \in \mathcal{R} \text{ e}$$

$$\Omega[n, k, a_i] = \sum_{j=1}^{N} \Omega[n-1, k, a_j]p_{ji}, \quad \text{caso contrário.}$$

$$\Omega[0, 1, a_i] = \begin{cases} \pi_i(0) & \text{se } a_i \in \mathcal{R} \text{ e} \\ 0 & \text{se } a_i \notin \mathcal{R}. \end{cases}$$

$$\Omega[0, 0, a_i] = \begin{cases} 0 & \text{se } a_i \in \mathcal{R} \text{ e} \\ \pi_i(0) & \text{se } a_i \notin \mathcal{R}. \end{cases}$$

a_i é o último estado visitado e $\pi(0)$ é o vetor de probabilidades iniciais.

As expressões para estas medidas foram calculadas a partir do instante inicial de uma rajada. Sendo assim, o vetor de probabilidades iniciais, $\pi(0)$, representa a probabilidade de entrada no subconjunto $\mathcal{R}$. Esta probabilidade é obtida supondo-se a cadeia em estado estacionário e observando-se as probabilidades condicionais de transição dos estados não pertencentes a $\mathcal{R}$ para os estados em $\mathcal{R}$ [54].

2.5 Resumo e Conclusões do Capítulo

Este capítulo procurou discutir o uso de medidas de desempenho para avaliar qualidade de serviço e mostrou as estatísticas mais comumente usadas para caracterizar tráfego, estatísticas estas chamadas de descritores de tráfego. Também foi aqui relacionado um conjunto de modelos de tráfego representando as correntes usualmente mais seguidas com exceção do processo SRP, aqui mostrado devido à sua relevância para o capítulo 3. Por fim, como as propostas resultantes deste trabalho de tese são baseadas em modelos markovianos com recompensa de taxa, expressões analíticas para o cálculo de alguns descritores de tráfego apresentados foram desenvolvidas para tais modelos.

Todos os modelos aqui apresentados conseguem aproximar melhor um ou mais descritores. Em geral, os descritores mais usados são a distribuição de taxa e a função de autocorrelação. No entanto, como será mostrado no capítulo 3, tais descritores não são necessariamente suficientes para determinar com precisão a fração de perda de pacotes qualquer que seja a fila. Por este motivo, os resultados do capítulo 3 serão usados na formulação da proposta de modelagem contida no capítulo 4.

Capítulo 3

Caracterização de Tráfego usando a Duração dos Períodos de Tempo com Taxas acima de Determinado Nível

Conforme apresentado na seção 2.3 e ao longo do capítulo 2, um modelo de tráfego, que reproduz um subconjunto dos descritores apresentados, não necessariamente é capaz de reproduzir o comportamento numa fila qualquer que seja a sua capacidade de armazenamento e taxa de serviço. Neste capítulo, este argumento é discutido e novas formas de caracterização de tráfego são apresentadas. Todos os intervalos de confiança mostrados foram extraídos usando 95% de probabilidade.

3.1 Desempenho de Modelos com mesma Distribuição de Taxa e mesma Autocorrelação

Dentre os descritores apresentados no capítulo 2.2, a distribuição de taxa (também chamada *distribuição marginal de taxa*) e a função de autocorrelação são aqueles mais comumente utilizados para parametrizar um modelo. Em outras palavras, um modelo é dito adequado quando um desses descritores obtidos do modelo é aproximadamente igual àquele obtido da seqüências real. A pergunta para a qual se deseja resposta é: modelos possuindo mesma distribuição de taxa e mesma função de autocorrelação resultam em mesmo comportamento quando aplicados a uma fila genérica? Ou, olhando-se por um outro prisma: capturar a distribuição de taxa e a função de autocorrelação para uma determinada seqüência de tráfego é suficiente para se prever seu comportamento numa fila genérica?

A resposta para estas perguntas é negativa e ela foi mostrada por Taralp, Devtsikiotis e Lambadaris em [17] e [57]. Nestes artigos de 1998, os autores usam 3 diferentes modelos, todos com mesma distribuição de taxa e mesma função de autocorrelação, para mostrar que a distribuição da ocupação numa fila infinita, $P[Q > x]$, onde Q é o tamanho da fila, é diferente para os três casos dado uma mesma taxa de serviço.

Os modelos usados em [17] e [57] são o *Spatial Renewal Process*, o *Extended TES*, e o *Distorted Gaussian*, os dois primeiros já comentados na seção 2.3. No caso do *Distorted Gaussian*, trata-se de um modelo Gaussiano, com uma dada função de autocorrelação, sobre o qual é realizada uma operação de distorção para se obter a distribuição marginal desejada [17]. Graças aos geradores de seqüência para estes três modelos, gentilmente cedidos pelos autores, o mesmo teste é aqui apresentado, mas desta vez focando o comportamento da perda numa fila finita.

Os descritores usados em [17] e [57] foram aqui recalculados e apresentados. A Figura 3.1 traz a distribuição de taxa e a Figura 3.2 traz a função de autocorrelação para os três modelos. Pode-se notar que as curvas para cada descritor são muito próximas e possuem um formato normalmente encontrado em seqüências amostrais de tráfego agregado. A Figura 3.3 mostra as funções de autocovariância para um tempo maior que os da Figura 3.2 e acrescentamos também, na Figura 3.4, as medidas de índice de dispersão para os mesmos modelos, estas também com características próximas.

Na Figura 3.5, podemos visualizar a fração de perda para estes modelos dado uma fila finita com carga de 33%. A fração de perda é mostrada em função do tamanho máximo da fila num gráfico em escala *log x log*. É possível notar diferenças significativas na fração de perdas para os modelos embora os valores dos descritores mostrados nas Figuras 3.1, 3.2 e 3.4 sejam os mesmos. É possível concluir, então, que outras formas de se caracterizar um tráfego devam ser exploradas.

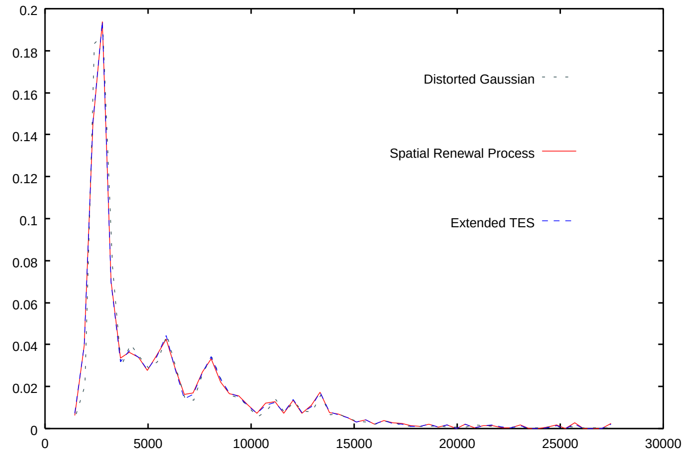


Figura 3.1: Distribuição de taxa para modelos-exemplo

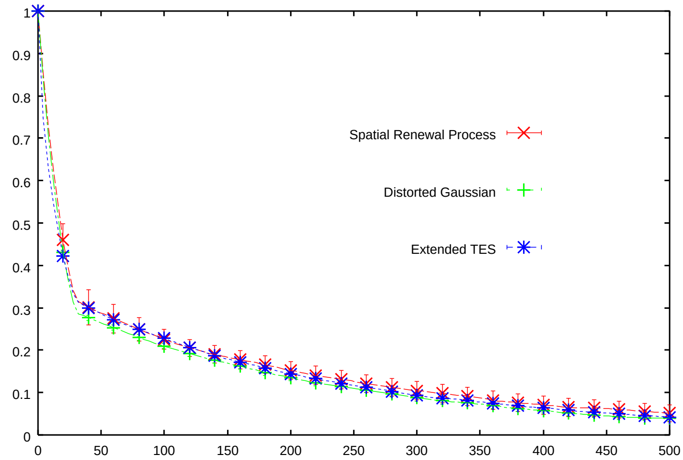


Figura 3.2: Função de autocorrelação para modelos-exemplo

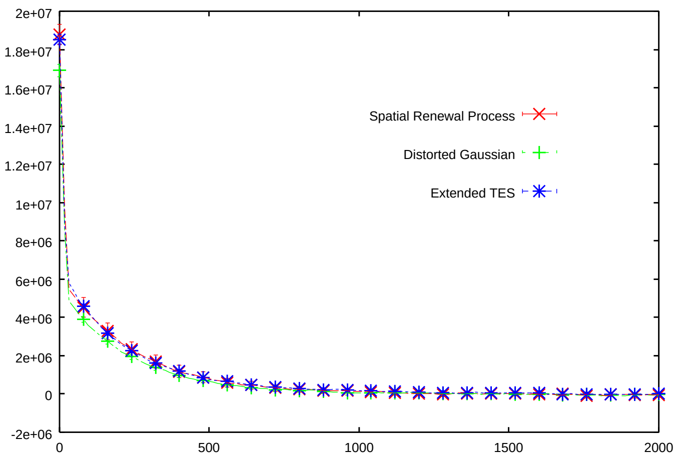


Figura 3.3: Função de autocovariância para modelos-exemplo

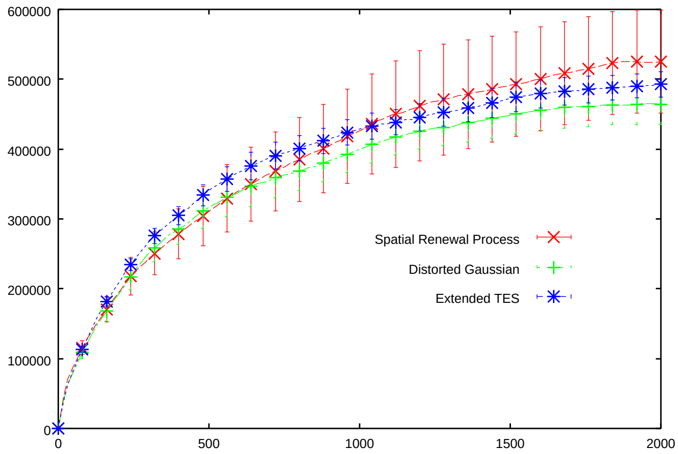


Figura 3.4: Índice de dispersão para cada modelos-exemplo

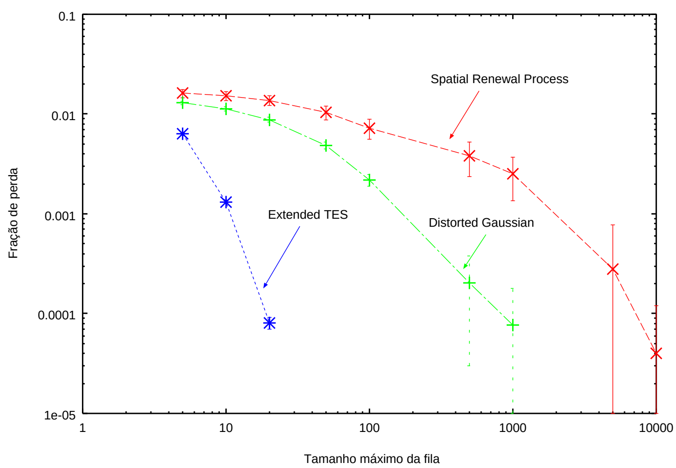


Figura 3.5: Perdas numa fila finita para modelos-exemplo (carga de 33%)

3.2 Períodos de Tempo em Taxas "Altas"

Qualquer nova maneira de se caracterizar um tráfego deve levar em consideração características não capturadas, ou parcialmente capturadas, pela distribuição de taxa e pela função de autocorrelação, que são os descritores tradicionalmente mais usados. Além disso, é desejável que estas características tenham uma associação lógica com a quantidade de perda numa fila finita uma vez que esta medida de desempenho é uma das que melhor caracteriza a qualidade de uma transmissão multimídia.

Pela definição de ambos os descritores, fica claro que a distribuição de taxa não fornece informação de ordem temporal e que a autocorrelação apenas caracteriza à média entre dois instantes de tempo qualquer. No entanto, uma informação simples e de extrema importância para o comportamento das perdas numa fila finita não é diretamente capturada nestes descritores: **"quanto tempo o tráfego permanece fluindo acima de uma determinada taxa"**.

Vejamos o exemplo construído na Figura 3.6. Nesta figura temos duas seqüências periódicas com mesma distribuição de taxa (50% de amostras com taxa 3 e 50% de amostras com taxa 1), portanto mesma média.

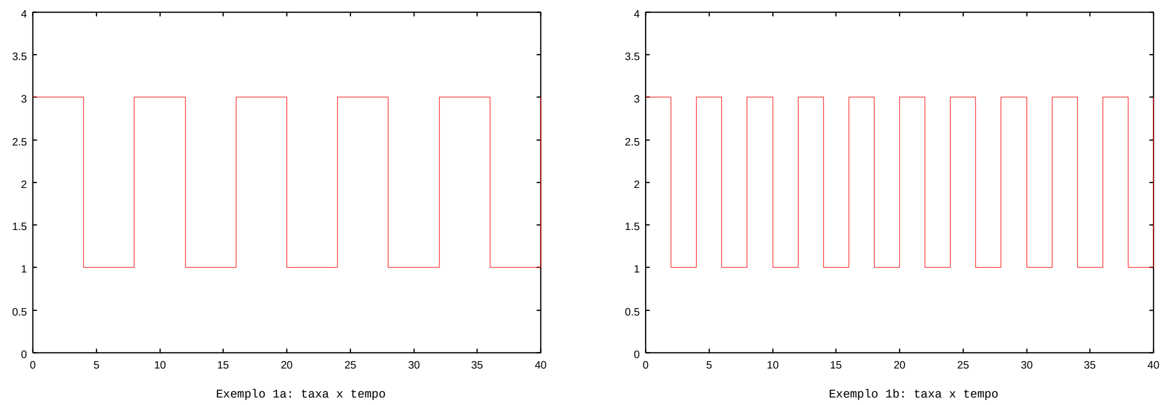


Figura 3.6: Exemplo 1, seqüências periódicas com mesma distribuição de taxa

Imaginemos que ambas as seqüências são aplicadas a uma fila com taxa de serviço igual a 1.7 e que taxas acima deste valor são consideradas taxas altas. Fica claro, neste exemplo, que a seqüência da esquerda, pelo fato da duração dos períodos em taxa alta ser maior que os da seqüência da direita, resulta em uma fração perda maior que a da segunda, caso a fila em questão não seja suficientemente grande para absorver a quantidade transmitida por qualquer das seqüências durante seus períodos de alta taxa. Assim sendo, podemos dizer que a duração dos períodos em taxa alta pode diferenciar estas seqüências quanto às suas perdas se soubermos a taxa de serviço da fila.

3.3 Reavaliando Modelos com mesma Distribuição de Taxa e mesma Autocorrelação

Em [17] e [57], os autores usaram a *função geradora de momento logarítmico* [34, 35] para diferenciar estatísticamente as seqüências geradas pelos três modelos. Dado o que foi apresentado na seção anterior, utilizaremos como variável aleatória a *duração dos períodos com taxas acimas de um determinado nível* r, $D(r)$, para diferenciarmos estatísticamente as mesmas seqüências. Por simplificação, chamaremos este nível r de *nível de rajada*. Nas figuras 3.7, 3.8 e 3.9 podemos ver a distribuição complementar da *duração dos períodos de rajada*, $P[D(r) > t]$, para os três modelos dado diferentes níveis de rajada. Comparando-se as três figuras, fica claro que a função $P[D(r) > t]$ é bem diferente para cada um dos modelos.

Para melhor visualizar esta diferença, usaremos a média da *duração dos períodos de rajada*, $\overline{D}(r)$, em função dos *níveis de rajada* de forma a termos um gráfico bidimensional que facilite esta comparação. A Figura 3.10 mostra tal medida para os três modelos em escala logarítmica. Pela Figura 3.10, podemos inferir o comportamento da fração de perdas mostrado na Figura 3.5, ou seja, o modelo que possui rajadas de maior duração é o *Spatial Renewal Process* e é este modelo que possui a maior fração de perda. Como a perda foi extraída para uma carga de 33%, iremos visualizar a média da *duração dos períodos de rajada* para os *níveis de rajada* que correspondem a taxas maiores que a taxa de serviço da fila para esta carga, no caso, em torno de 17500. A Figura 3.11 mostra a distribuição complementar da *duração dos períodos de rajada*, em escala logarítmica, para este nível de rajada. Podemos verificar, por estas figuras, que maiores frações de perda correspondem a *durações de períodos de rajada* maiores.

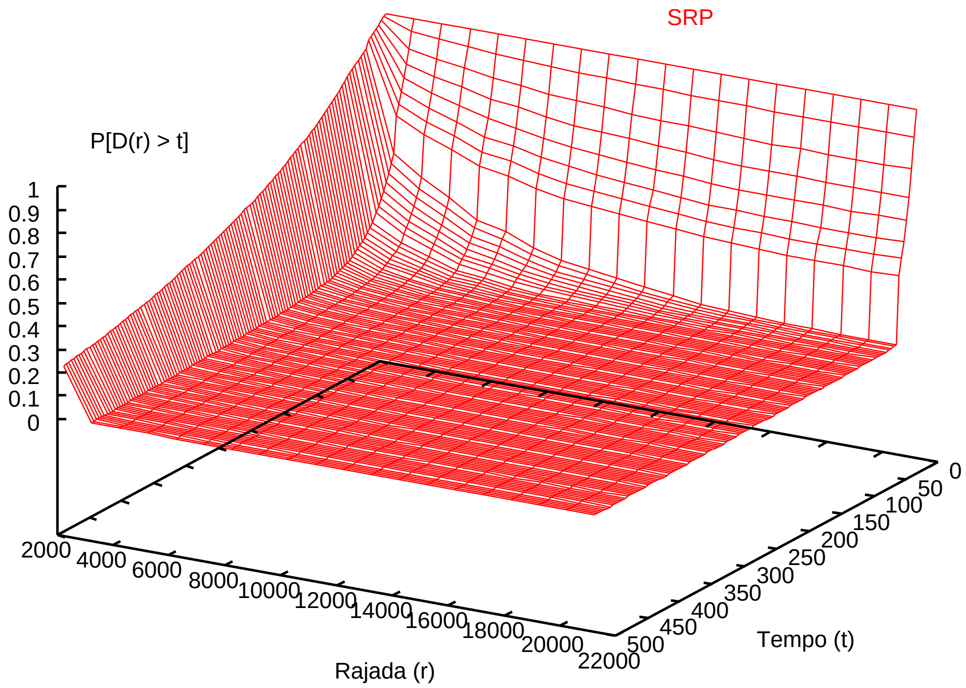


Figura 3.7: Distribuição da duração dos períodos de rajada para o *Spatial Renewal Process*

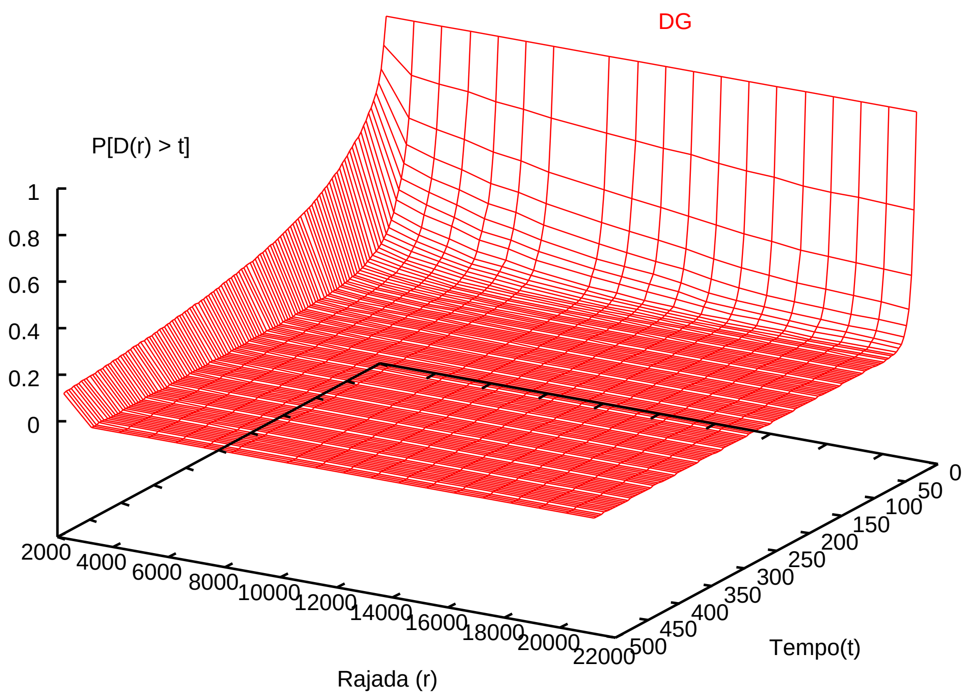


Figura 3.8: Distribuição da duração dos períodos de rajada para o *Distorted Gaussian*

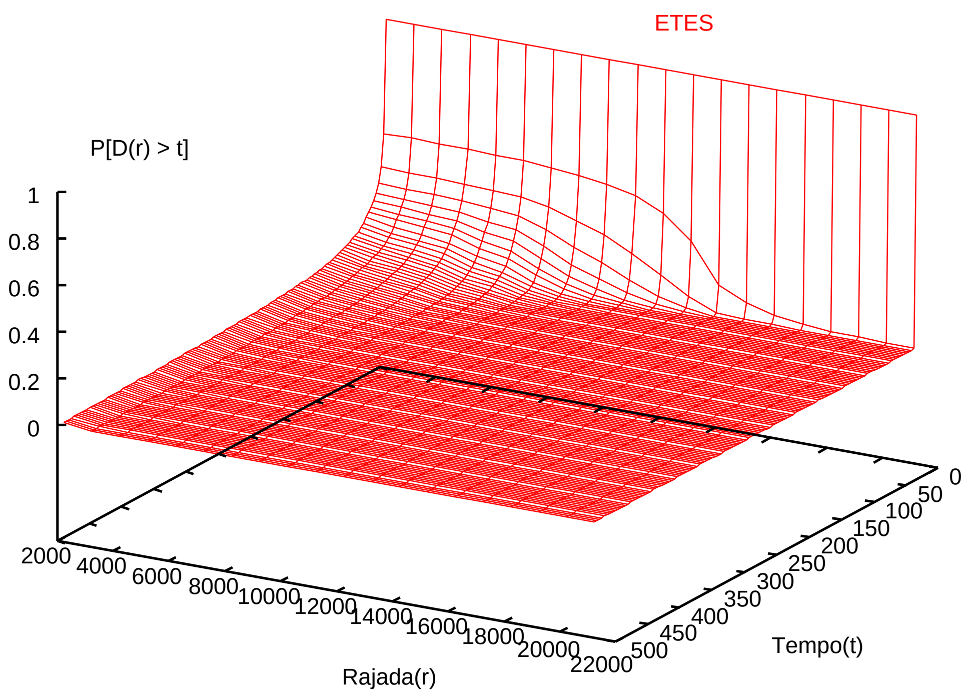


Figura 3.9: Distribuição da duração dos períodos de rajada para o *Extended* TES

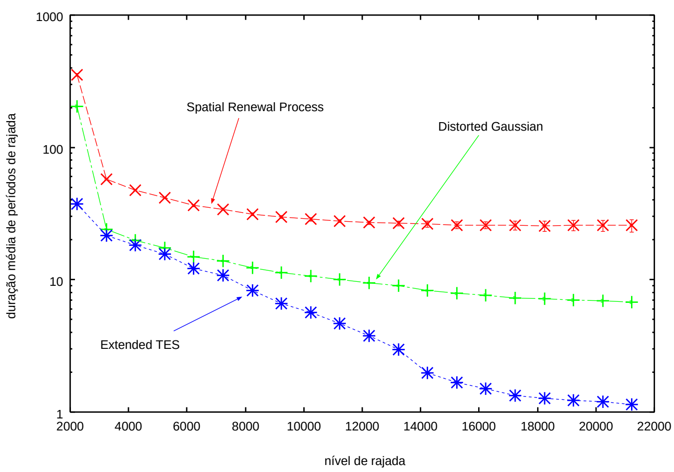


Figura 3.10: Duração média da rajada em função do nível de rajada

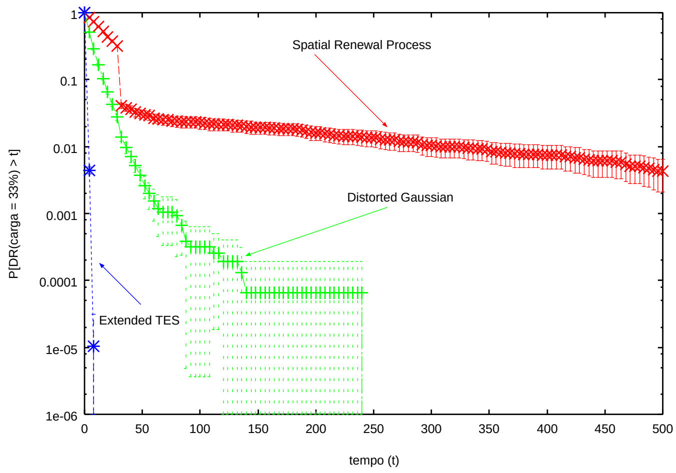


Figura 3.11: Distribuição da duração dos períodos de rajada para um nível de rajada correspondente a uma carga de 33%

3.4 Comparando Modelos Distintos para mesma seqüência de Vídeo

De maneira a complementar os testes apresentados na seção 3.3, usaremos um modelo markoviano de histograma e um modelo F-ARIMA para modelar uma seqüência de vídeo codificada em MPEG4. Esta seqüência contém o número de pacotes de 1500 *bytes* transmitidos a cada quadro. No caso, assume-se que os pacotes são igualmente espaçados no tempo a cada intervalo entre quadros. Tal suposição se apóia nas seguintes razões:

- Primeiramente, deseja-se que os resultados não sejam por demais focados em um cenário específico quanto ao ambiente de rede. Ao se determinar um canal de transmissão com taxa conhecida, sendo este dedicado, os pacotes de cada quadro serão transmitidos em seqüência na taxa do canal, que ficará ocioso até o próximo quadro. No entanto, na chegada destes pacotes numa fila de saída de uma interface de algum equipamento comutador qualquer, esta relação não é mais garantida uma vez que outros fluxos estarão disputando este mesmo recurso. Isto significa dizer que, em termos práticos, nada garante o mesmo espaçamento inicial entre pacotes na chegada destes numa fila de saída.

- Tal impossibilidade na garantia deste espaçamento torna-se ainda maior se o caminho até a fila de saída em estudo, que pode ser um *buffer* de recepção do aplicativo que vai tocar o filme, passar por diversos canais, cada qual com uma taxa de transmissão diferente.

- Em ambientes com meio compartilhado, o embaralhamento de pacotes de fontes distintas é bastante comum.

A análise da perda será feita para uma fila de 100 pacotes, onde a taxa de serviço será variada. O objetivo é comparar os modelos de acordo com o comportamento da *duração dos períodos de rajada* de cada um deles em função dos diversos *níveis de rajada*, principalmente daqueles que correspondem a taxas maiores que as taxas de serviço sendo usadas na fila.

A seqüência de vídeo escolhida, codificada em MPEG4, corresponde a 60 minutos do filme *Silence of the Lambs*, obtida em [58]. A taxa de apresentação dos quadros do filme foi codificada em 25 quadros/segundo, sendo que o tamanho médio do quadro corresponde a 1,954 *pacotes* de 1500 bytes, com variância igual a 2,584. O modelo markoviano de histograma para esta seqüência possui 8 estados, cada um com uma taxa de acordo com o seguinte vetor de recompensas:

$$\Lambda = < 0.8487 \;\; 2.3353 \;\; 4.3398 \;\; 6.3552 \;\; 8.3426 \;\; 10.5082 \;\; 12.2767 \;\; 14.0000 > \;\; .$$

A matriz de probabilidade de transição do modelo é

$$\mathbf{P} = \begin{bmatrix} 0.844230 & 0.144623 & 0.010572 & 0.000487 & 0.000088 & 0 & 0 & 0 \\ 0.195305 & 0.725893 & 0.066722 & 0.010588 & 0.001372 & 0.000089 & 0 & 0.000030 \\ 0.062267 & 0.288563 & 0.571337 & 0.063811 & 0.010421 & 0.001287 & 0.002316 & 0 \\ 0.004783 & 0.156957 & 0.220000 & 0.555217 & 0.056087 & 0.005652 & 0.001304 & 0 \\ 0.001639 & 0.073770 & 0.140984 & 0.216393 & 0.488525 & 0.065574 & 0.013115 & 0 \\ 0 & 0.006557 & 0.032787 & 0.029508 & 0.147541 & 0.672131 & 0.098361 & 0.013115 \\ 0 & 0 & 0.094340 & 0.050314 & 0.044025 & 0.194969 & 0.603774 & 0.012579 \\ 0 & 0 & 0 & 0.100000 & 0 & 0.300000 & 0.300000 & 0.300000 \end{bmatrix} .$$

O modelo F-ARIMA usa a média e a variância da taxa de pacotes mostradas mais o parâmetro de Hurst, que foi estimado como sendo igual a 0,987. Este valor foi obtido a partir da média dos parâmetros de Hurst fornecidos pelos estimadores R/S, Higuchi e pelo método da variância. As implementações utilizadas estão relacionadas com o trabalho desenvolvido em [59].

As Figuras 3.12, 3.13 e 3.15 trazem, respectivamente, a função de probabilidade de massa das taxas (histograma), a autocorrelação e o índice de dispersão para os modelos e a seqüência MPEG4. No caso da função de probabilidade de massa, estas foram computadas usando 8 níveis uma vez que o modelo markoviano de histograma possui apenas 8 níveis. Todos os descritores aqui apresentados foram calculados analiticamente para o modelo de histograma segundo as expressões mostradas em 2.4. Já para o modelo F-ARIMA, os descritores foram calculados a partir da média de 8 caminhos amostrais gerados por simulação e contendo o mesmo número de amostras contido na seqüência MPEG4 real - para as finalidades deste estudo comparativo, este número de caminhos amostrais foi verificado como sendo suficiente para mostrar o perfil dos descritores para o modelo F-ARIMA. É possível verificar, pelas curvas de autocorrelação e de índice de dispersão, que o modelo de histograma captura apenas a dependência de curtíssima duração e que o modelo F-ARIMA captura a dependência de longa duração, como era de se esperar. Isto também pode ser visto na Figura 3.14, que contém as curvas de autocorrelação para um tempo mais longo.

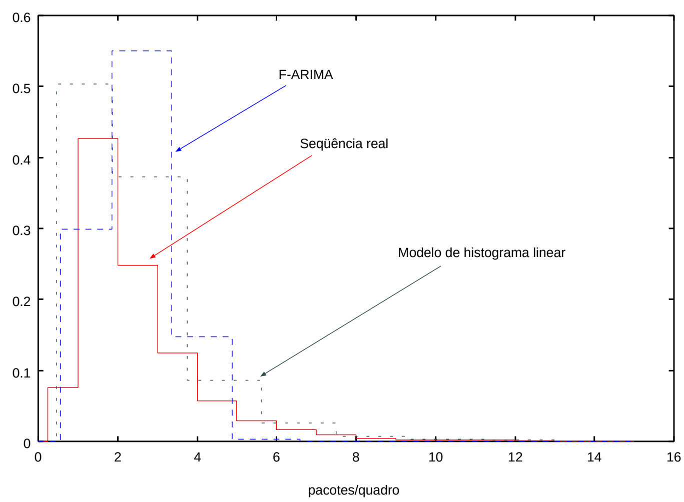


Figura 3.12: Histogramas da taxa de pacotes por quadro

As Figuras 3.16 e 3.17 trazem, respectivamente, a *duração média dos períodos de rajada*, em função dos *níveis de rajada*, e a *fração de perda*, em função da taxa de serviço, para ambos os modelos e a seqüência de vídeo. Assim como nos demais descritores, a *duração média dos períodos de rajada* foi calculada analiticamente para o modelo de histograma segundo o procedimento descrito em 2.4.

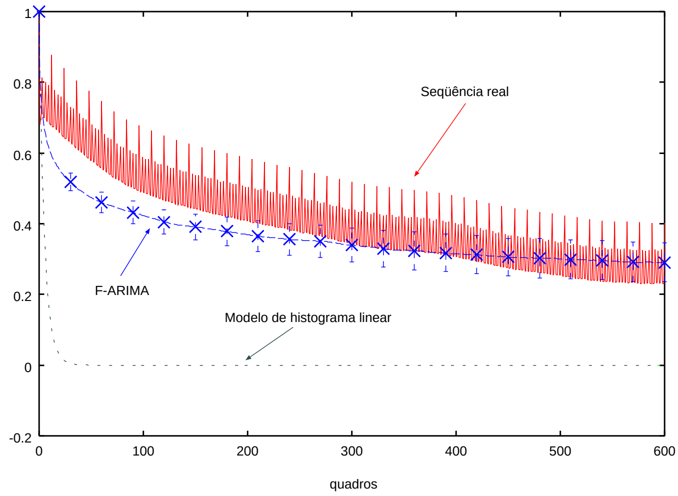


Figura 3.13: Função de autocorrelação da taxa de pacotes por quadro

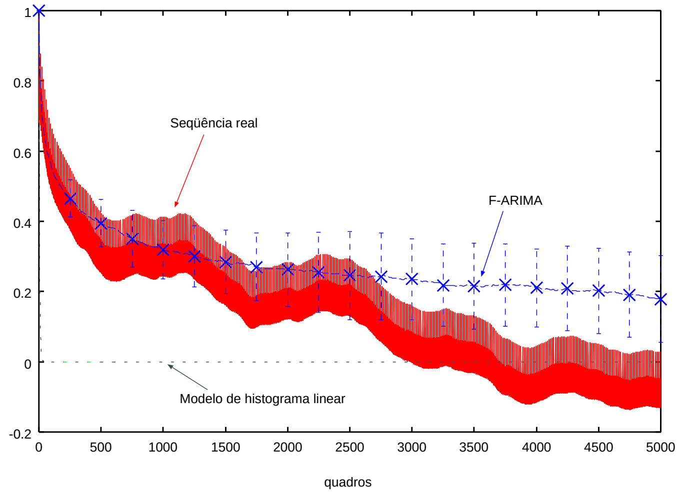


Figura 3.14: Função de autocorrelação da taxa de pacotes por quadro para intervalos de até 5000 quadros

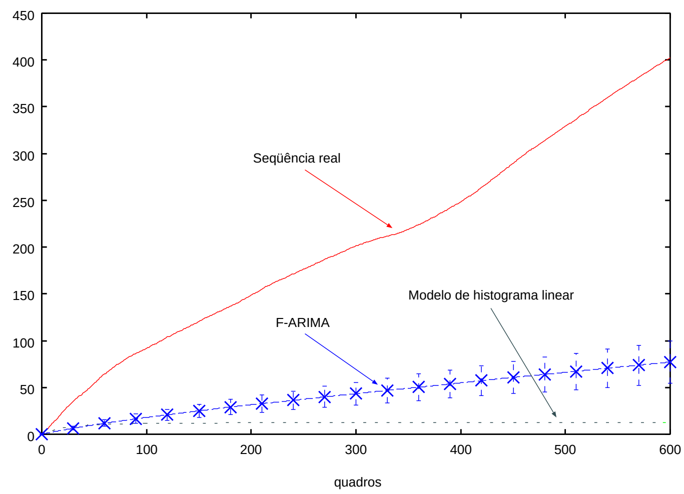


Figura 3.15: Índice de dispersão da taxa de pacotes gerados

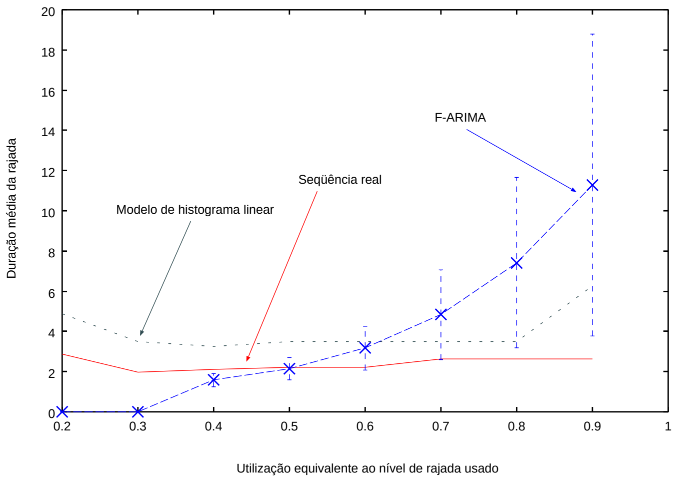


Figura 3.16: Duração média das rajadas

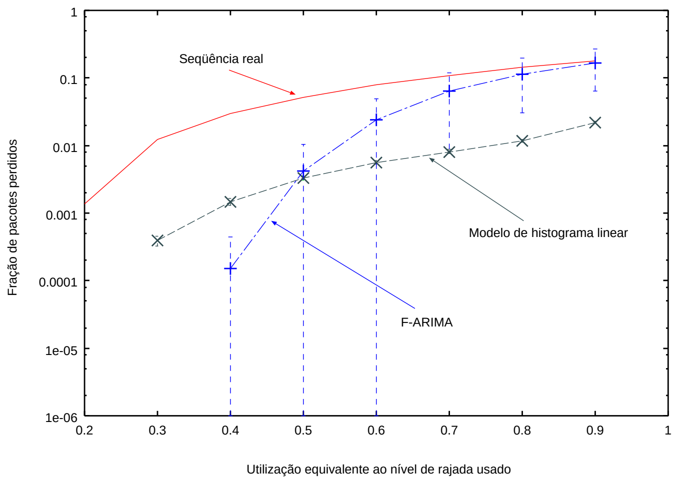


Figura 3.17: Perdas de pacotes para modelos e seqüência real

Observando-se as Figuras 3.16 e 3.17, é possível verificar que não existe uma "equivalência" entre as curvas no que diz respeito às durações médias e perdas da seqüência real e as mesmas curvas para cada modelo. Ou seja, nem para todos os níveis de rajada onde um modelo tem duração média superior ao outro, a fração de perda também se mostra superior. Isso fica mais evidente ao compararmos as durações médias para a seqüência real e o modelo de histograma. Enquanto que o modelo de histograma tem durações médias para os períodos acima de r maiores que os da seqüência real, esta última tem frações de perda maiores que as do modelo de histograma para qualquer r. A explicação para tal é que a média das durações das rajadas não traz informação suficiente para uma uma análise mais precisa dos modelos com relação às perdas. Assim sendo, as Figuras 3.18, 3.19, 3.20, 3.21, 3.22, 3.23 e 3.24 trazem as distribuições das durações das rajadas para os níveis de rajada equivalentes às cargas de 0.2 a 0.9.

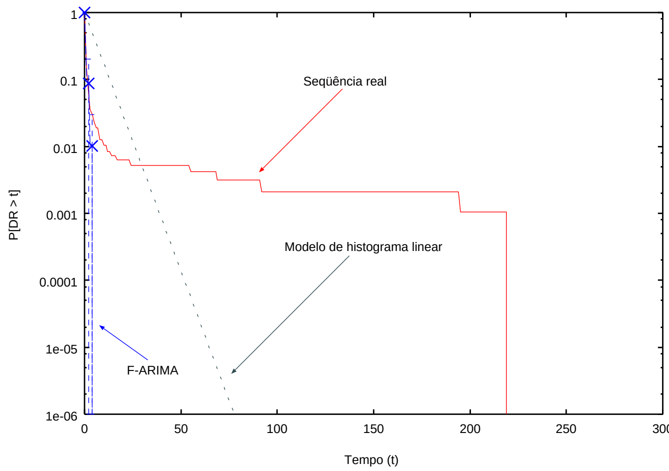


Figura 3.18: Distribuição da duração das rajadas para r equivalente à carga de 0.3

Podemos verificar que, para níveis de rajada inferiores a uma carga de 0.5, a cauda das distribuições de duração de rajada para o modelo F-ARIMA não é muito extensa e, portanto, este modelo possui fração de perda inferior às do modelo de histograma e da seqüência real. Já para r relativo às cargas de 0.5 a 0.7, a cauda desta distribuição para o modelo F-ARIMA passa a ser maior que a do modelo de histograma porém menor que a da seqüência real, conseqüentemente as perdas também são subestimadas para estes valores de carga (comportamento este que se reflete nas curvas de perda da Figura 3.17). Finalmente, para r relativo às cargas de 0.8 e 0.9, a cauda de $P[DR > t]$ para o modelo F-ARIMA passa a ser maior que

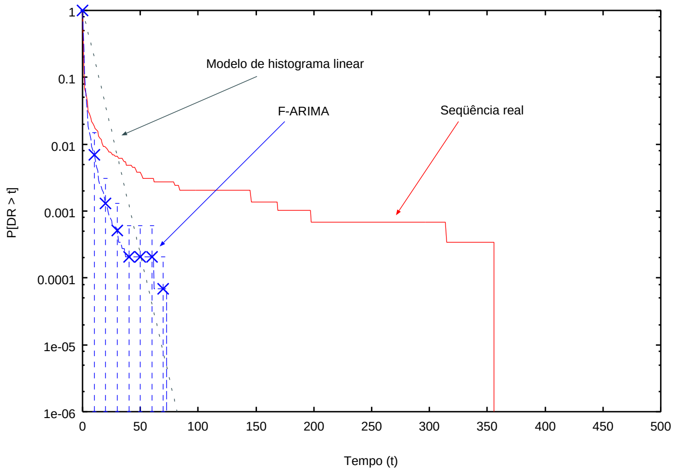


Figura 3.19: Distribuição da duração das rajadas para r equivalente à carga de 0.4

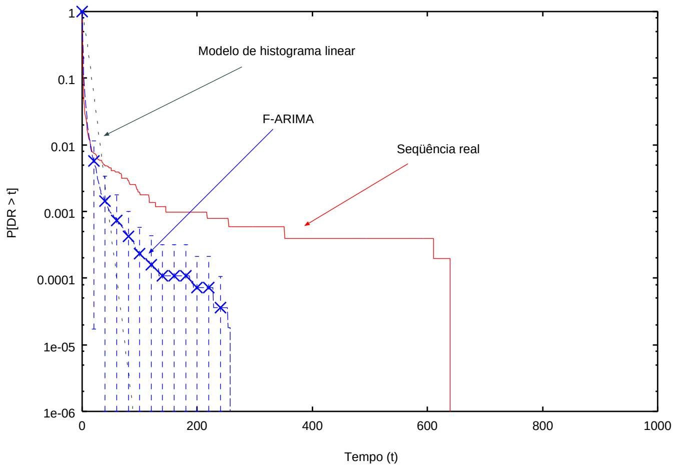


Figura 3.20: Distribuição da duração das rajadas para r equivalente à carga de 0.5

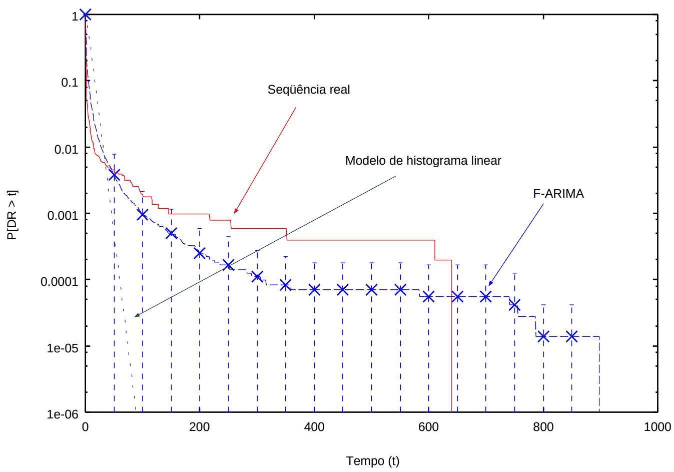


Figura 3.21: Distribuição da duração das rajadas para r equivalente à carga de 0.6

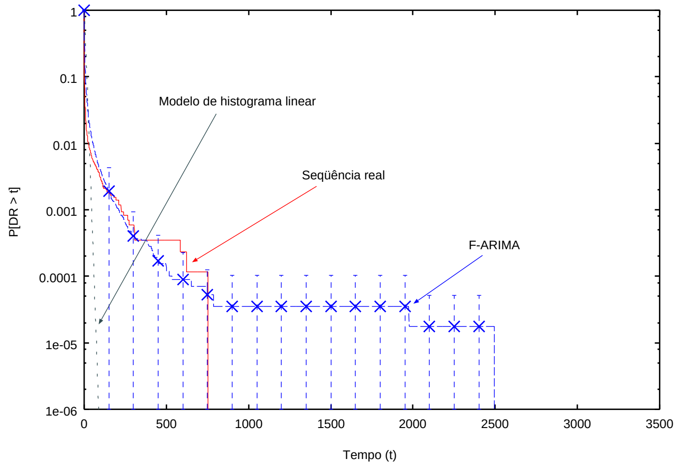


Figura 3.22: Distribuição da duração das rajadas para r equivalente à carga de 0.7

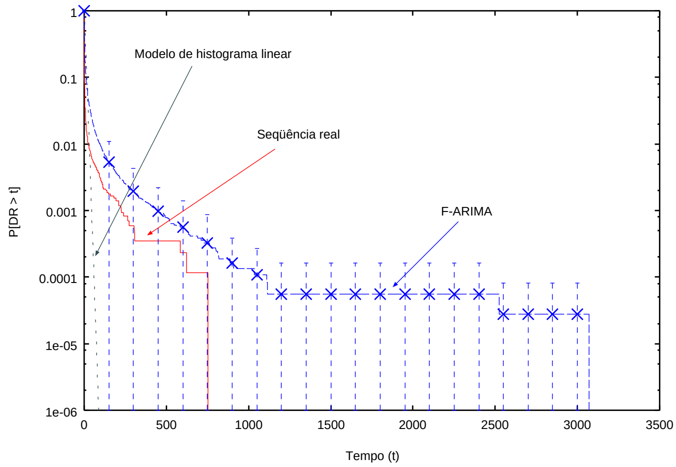


Figura 3.23: Distribuição da duração das rajadas para r equivalente à carga de 0.8

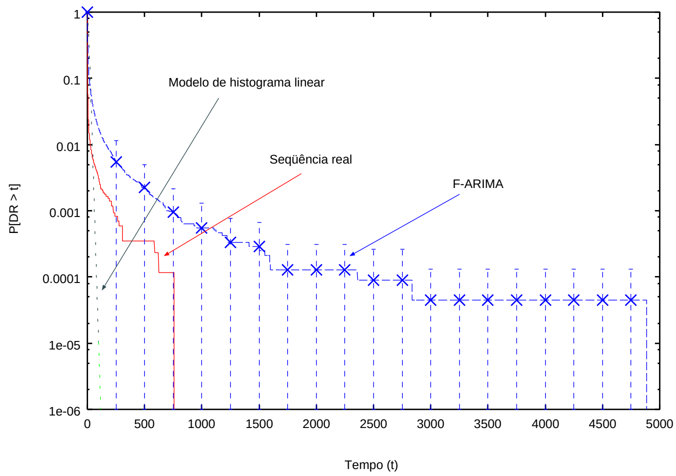


Figura 3.24: Distribuição da duração das rajadas para r equivalente à carga de 0.9

a do modelo de histograma e da seqüência real. Através da figura 3.17 podemos observar que a fração de perda estimada pelo modelo F-ARIMA para cargas altas fica bem próxima da fração de perda calculada para a seqüência real.

Esta correlação entre as frações de perda e as caudas das distribuições de rajada é um comportamento esperado, uma vez que uma cauda mais longa significa longos períodos nos quais a fila está enchendo, pois a taxa de chegada de pacotes é superior à taxa de serviço da fila. Estes períodos, embora raros, quando ocorrem são suficientes para aumentar bastante a fração de pacotes perdidos.

Assim como ocorreu para os demais descritores, a distribuição da duração das rajadas foi obtida analiticamente para o modelo markoviano de histograma e por inspeção da seqüência real e das seqüências geradas pela simulação do modelo F-ARIMA. No caso deste último, a distribuição resultante é uma média das distribuições das seqüências geradas em cada rodada de simulação. Para o modelo markoviano, da mesma forma que na medida de *duração média*, fez-se uso das expressões descritas na seção 2.4. Outra possível medida, que não foi explorada neste trabalho, é a *duração média transiente*, $\overline{DR}(t)$, ou seja, a duração média dado um tempo t de observação. A seção 2.4 traz a expressão analítica genérica para a obtenção desta medida em modelos markovianos com recompensa de taxa.

3.5 Resumo e Conclusões do Capítulo

Neste capítulo foi mostrado que modelos apresentando mesma distribuição de taxa e mesma função de autocorrelação, e que apresentam diferentes perdas numa fila finita, podem ser diferenciados através de novos descritores baseados na duração de períodos nos quais a taxa se encontra acima de determinados valores, aqui chamados de *níveis de rajada* (seção 3.3). Estes novos descritores são a distribuição da duração das rajadas e a duração média das rajadas. Foi possível obter o modelo que melhor caracteriza a fração de perda na fila através da observação destes descritores. Ou seja, aquele modelo onde a distribuição da duração das rajadas se aproxima da distribuição da seqüência real é o que melhor caracteriza a fração de perda na fila. Nos exemplos mostrados pudemos observar para quais taxas de serviço a fração de perda numa fila é melhor capturada por um certo modelo que por outro, tenham eles mesma função de autocorrelação ou não.

Capítulo 4

Proposta de Modelagem

Neste capítulo, será descrita uma proposta de modelagem baseada nos resultados mostrados no capítulo anterior. Esta proposta objetiva desenvolver um modelo simples e capaz de prever o comportamento das perdas numa fila finita.

Frente aos resultados da literatura esta tarefa é complexa. Freqüentemente, modelos adaptados para uma determinada classe de tráfego não são capazes de ter a mesma eficiência quando aplicados a outras classes. Um exemplo disso é o modelo *pseudo auto-similar*, proposto em [3], [46] e [47]. Nestes trabalhos, os autores conseguem uma razoável aproximação quando modelam, a partir de uma cadeia de Markov por eles definida, o tráfego agregado de Bellcore, apresentado em [2]. A medida de desempenho predita pelo modelo é a perda de células ATM numa fila finita com capacidade de 200 células [3], variando-se sua carga. No entanto, alguns trabalhos encontrados na literatura, como em [60], relatam situações onde este mesmo modelo apresenta problemas para prever o comportamento das perdas considerando outros perfis de tráfego.

Assim sendo, é razoável assumir as limitações das diversas classes de modelos de tráfego e procurar incorporar, ao processo de modelagem, as informações do sistema que mais influenciam a medida de interesse desejada. Desta forma, tem-se não simplesmente um modelo de fonte genérico com a pretensão de prever o comportamento de qualquer medida de desempenho para qualquer sistema de fila, mas um modelo que leva em consideração parâmetros que influenciam determinada medida de desempenho extraída para uma fila com características conhecidas. Este procedimento resulta em modelos mais simplificados e mais eficientes.

Um exemplo de modelo que utiliza parâmetros do sistema na sua descrição pode ser encontrado em [18], onde Schwefel e Lipsky incorporam o *Maximum Burst Size* (MBS), usado em mecanismos de policiamento de tráfego em redes, no modelo de tráfego por eles proposto. Em um sistema de fila finita com um servidor e política do tipo *FIFO*, a taxa de serviço é um parâmetro candidato a ser incorporado no modelo e foi o escolhido neste trabalho de acordo com o estudo feito no capítulo anterior. Doravante, chamaremos este parâmetro de *nível de sobrecarga* e o usaremos para extrair informações para prever a perda numa fila.

O cenário de aplicação do modelo de tráfego envolve, portanto, situações onde a taxa de serviço para uma fila seja conhecida. Exemplos mais comuns são: reserva de banda em interfaces de saída de roteadores para um tráfego específico, a capacidade efetiva estimada por controles de admissão mediante uma especificação do tráfego, a capacidade de um enlace de rede sendo alimentado por um tráfego agregado.

A seqüência do filme *Silêncio dos Inocentes* codificado em MPEG4 será usada para exemplificar o modelo e verificar seu desempenho em aproximar a fração de perda. Vale lembrar que seqüências codificadas em MPEG representam uma classe de tráfego de difícil caracterização [61, 4, 5, 16], como será abordado no capítulo 5. Outros exemplos envolvendo outras seqüências de vídeo em MPEG4 e seqüências de tráfego agregado também serão mostrados no capítulo 5.

4.1 Usando Durações de Sobrecarga e Folga para Caracterização do Tráfego

Como mostrado na seção 3.2, os períodos de tempo com taxas acima de um determinado nível de sobrecarga (períodos de sobrecarga) trazem informações relevantes para o processo de ocupação numa fila que não costumam ser capturadas por outros descritores. Da mesma maneira, os períodos de tempo com taxas abaixo do nível de sobrecarga (períodos de folga) são importantes para caracterizar a inatividade no processo de ocupação da fila em foco.

Outros trabalhos na literatura também se valem da descrição dos períodos de sobrecarga e folga. Em [21], os autores modelam um sistema de fonte e fila por um processo *Quasi Birth-Death* (QBD). Trata-se de um modelo markoviano para o qual é extraída uma expressão analítica para a distribuição dos períodos de sobrecarga. Esta informação é então usada para dimensionar sistemas de descarte preemptivo. Em [62], expressões para a distribuição dos períodos de sobrecarga e folga são obtidas para fontes *ON-OFF* multiplexadas. As distribuições são então usadas para caracterizar o processo de saída do multiplexador. E ainda em [18], é dito que é o expoente da distribuição dos períodos de sobrecarga num agregado de fontes *ON-OFF* que determina o desempenho numa fila.

Assim sendo, durações de sobrecarga (taxas acima do nível de sobrecarga) e folga (taxas abaixo do nível de sobrecarga) serão usadas neste trabalho para descrever o tráfego aplicado a uma determinada fila. No caso, serão capturadas as distribuições dessas durações de uma seqüência real, uma vez que as médias das mesmas, por elas só, podem não trazer informação suficiente para a predição da perda. Relembrando o discurso feito no capítulo anterior, seja o caso de um tráfego possuindo uma duração de sobrecarga muito longa porém de rara ocorrência. Se este tráfego estiver aplicado a uma fila muito longa, mas não o suficiente para evitar perdas, durante a sobrecarga as raras ocorrências destas sobrecargas ocasionarão perdas que não podem ser previstas apenas com a informação de média de período de sobrecarga.

Distribuições complementares ($P[X > x]$) são úteis para avaliar se uma distribuição possui cauda longa ou não. Para tal, é necessário verificar a seguinte condição [63]:

$$\lim_{x\to\infty} P[X > x] \sim x^{-\alpha},\ 0 < \alpha < 2\ .$$

Um exemplo de distribuição de cauda longa é a distribuição de Pareto [6].

Como a cauda das durações é um dado importante para o processo de ocupação de uma fila, dois descritores são necessários para o modelo a ser proposto neste capítulo:

$$P[Ds > t] \text{ e } P[Df > t]\ ,$$

onde Ds e Df são, respectivamente, as variáveis aleatórias indicando *duração de sobrecarga* e *duração da folga.*

Quanto maior a cauda de $P[Ds > t]$, menor a tendência de decaimento da perda em relação ao tamanho máximo da fila. Isto se justifica pelo fato das maiores perdas serem provocadas pelas maiores durações de sobrecarga. Apenas quando o tamanho máximo da fila for suficientemente grande para absorver estes longos períodos de sobrecarga, é que a perda se reduzirá mais acentuadamente. Já no caso de $P[Df > t]$, uma cauda longa indica longos períodos de inatividade na ocupação da fila. Como perdas só ocorrem durante períodos de sobrecarga, a cauda de $P[Df > t]$ pode não influenciar tanto na fração de perda. O capítulo 4.4 traz uma discussão maior sobre a sensibilidade da perda com relação a estes descritores.

4.2 Usando Hiper-exponenciais para Modelar a Distribuição das Durações de Sobrecarga e Folga

As distribuições das durações de sobrecarga e folga devem ser parametrizadas e modeladas. Optou-se pelo uso de distribuições *hiper-exponenciais* para este fim. Distribuições hiper-exponenciais correspondem a uma soma ponderada de distribuições exponenciais, o que permite representá-las facilmente através de cadeias de Markov. Além disso, elas são capazes de aproximar distribuições com coeficiente de variação maior que um. Nossos estudos mostraram que é comum as distribuições de Ds e Df, tanto para as seqüências MPEG4 como para o tráfego agregado, possuírem coeficiente de variação maior que um.

Como exemplo, as Figuras 4.1 e 4.2 mostram as distribuições das durações de sobrecarga e folga para a seqüência em MPEG4 do filme *Silence of the Lambs*, considerando que a seqüência alimenta uma fila com carga de 33%. Ambas distribuições possuem cauda maior que a de uma distribuição exponencial com mesma média. Pode-se observar que a hiper-exponencial, dependendo dos parâmetros, pode apresentar uma cauda próxima à da seqüência real. Tais características também foram observadas para outros níveis de sobrecarga e também em outras seqüências MPEG4, como será mostrado no capítulo 5. Também para seqüências de tráfego WAN foi possível verificar esta característica (ver capítulo 5).

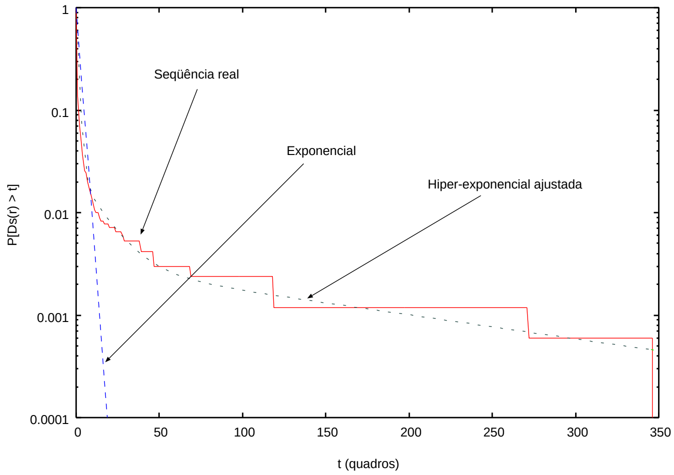


Figura 4.1: Distribuição das durações de sobrecarga para r equivalente à 33% de carga

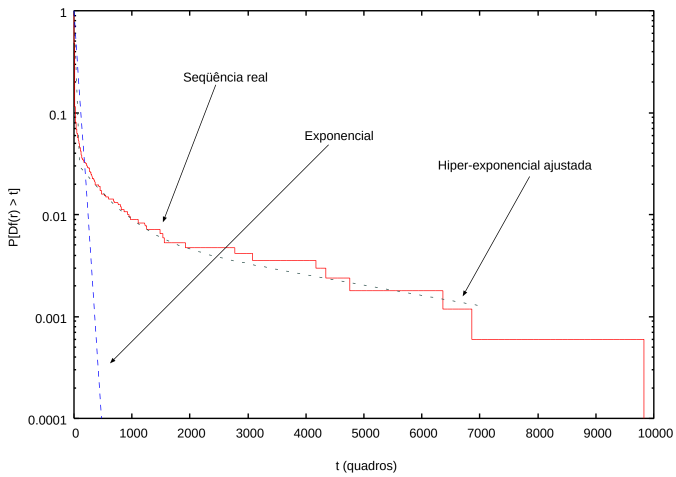


Figura 4.2: Distribuição das durações de folga para r equivalente à 33% de carga

A Figura 4.3 mostra o diagrama ilustrando os parâmetros de uma distribuição hiper-exponencial. No diagrama, p_i é a probabilidade de ocorrência do estado i e $1/\alpha_i$ é a duração média do estado i. A duração de cada estado é uma variável aleatória exponencial.

A média ($\overline{x}$) e o segundo momento ($\overline{x^2}$) da distribuição hiper-exponencial são:

$$\overline{x} = \sum_{i}^{N} p_i \frac{1}{\alpha_i} \text{ e}$$

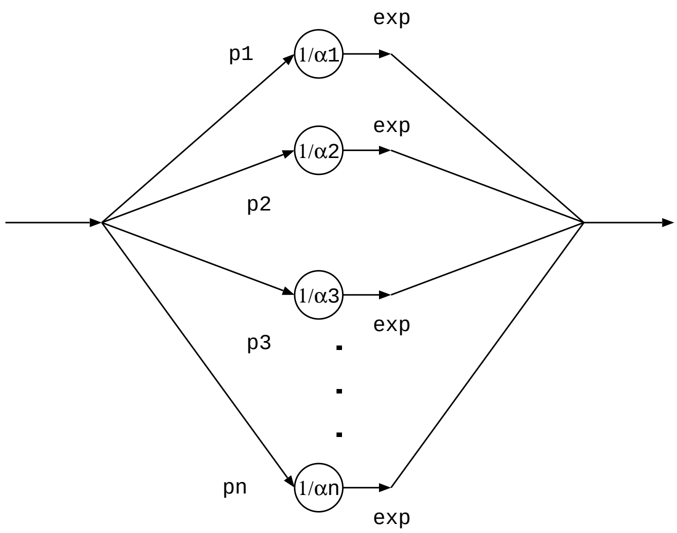


Figura 4.3: Diagrama com os parâmetros de uma distribuição hiper-exponencial

$$\overline{x^2} = \sum_{i}^{N} p_i \frac{2}{{\alpha_i}^2} \ .$$

O próximo passo, após a escolha da distribuição que será usada para representar as durações de sobrecarga e folga, é a sua parametrização. A escolha do número de estados para a hiper-exponencial envolve um compromisso entre obter uma aproximação satisfatória e proporcionar um número reduzido de estados para o modelo de tráfego a ser apresentado na seção 4.3.1.

A partir de vários exemplos como os das Figuras 4.1 e 4.2, onde são mostrados os logaritmos das distribuições, foi possível observar que a cauda dessas distribuições conseguem ser razoavelmente aproximadas por distribuições exponenciais ponderadas por uma constante $C < 1$ e com durações médias muito acima das médias de cada distribuição real. Assim sendo, definiu-se a hiper-exponencial a ser usada como tendo três estados: um deles representando a cauda da distribuição e os outros dois representando sua parte inicial.

A Figura 4.4 mostra um exemplo da aproximação desejada onde é usado o ln das distribuições. Supondo-se que as distribuições dos períodos de sobrecarga ou folga podem apresentar cauda longa, tem-se que

$$\ln x^{-\alpha} = -\alpha \ln x \ , \ 0 \ < \ \alpha \ < \ 2 \ ,$$

é uma função convexa como a que está apresentada na figura. As taxas das três exponenciais ponderadas que conjuntamente aproximam esta distribuição são representadas pelas retas indicadas na figura visto que:

$$\ln e^{-\lambda_i t} = -\lambda_i t \ , \ i = 1, 2, 3 \ ,$$

onde λ_3 é a taxa da exponencial que representa a cauda da distribuição (região 3), λ_1 é a taxa que representa a região inicial (região 1) e λ_2 é a taxa que representa a região intermediária da distribuição (região 2). Portanto, o problema de parametrização da hiper-exponencial escolhida envolve achar as inclinações (λ_i) e os fatores de ponderação (p_i) de forma que as três retas indicadas na figura aproximem da melhor maneira possível a distribuição.

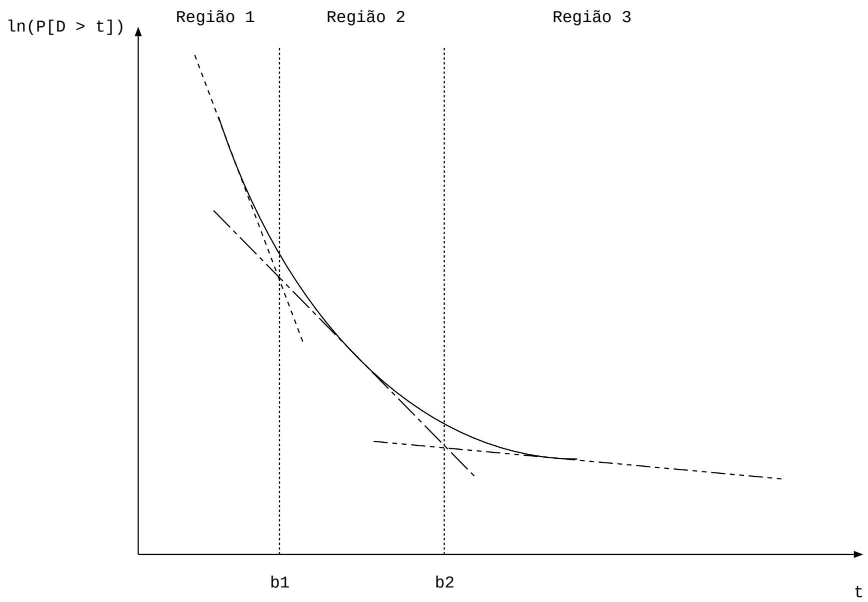


Figura 4.4: Exemplo de aproximação de uma distribuição de cauda longa por uma hiper-exponencial de 3 estados

É importante verificar que as retas apresentadas na Figura 4.4 para cada região representam a soma ponderada das três exponenciais naquela região. A abordagem usada parte do pressuposto que, dadas as respectivas ponderações:

- na região 3 predomina unicamente a exponencial de taxa λ_3;
- na região 2 predomina a exponencial de taxa λ_2 acrescida da exponencial de taxa λ_3;
- na região 1 predomina a exponencial de taxa λ_1 acrescida das outras duas.

Existem vários métodos para a obtenção de parâmetros de uma distribuição, a partir de um conjunto de amostras, que podem ser usados para resolver este problema. Em [64] é apresentado um algoritmo bastante genérico e eficiente para obter parâmetros de uma distribuição hiper-exponencial. Este algoritmo divide o histograma das amostras em regiões, de acordo com um coeficiente de variação máximo, e calcula uma hiper-exponencial para cada região a ser ponderada num cálculo final. Claramente, este procedimento não se aplica à hiper-exponencial aqui usada, que tem apenas três estados. Outro algoritmo existente é o *Expectation-Maximization* (EM) [65], que procura aproximar uma distribuição qualquer por uma distribuição *phase-type*. A hiper-exponencial é um caso particular de uma *phase-type* e, portanto, este método se aplica ao problema aqui descrito. No entanto, trata-se de um método muito mais genérico e iterativo que depende de soluções iniciais que podem não convergir. Além disso, em [64] é dito que este método pode não capturar satisfatoriamente a cauda das distribuições. Em [66] encontra-se a ferramenta **EMpht** para ajustes de distribuição usando o EM. Foram realizados testes com esta ferramenta para algumas seqüências MPEG4 e problemas de convergência foram encontrados.

Além dos métodos citados, há também a possibilidade de transformar o problema de ajuste das três retas num problema de otimização para ser resolvido usando métodos genéricos para tal. No entanto, as restrições do problema, assim como sua função objetivo, devem ser muito bem definidas para garantir um ajuste adequado da cauda da distribuição, o que pode resultar em um problema de otimização complexo. No capítulo 4.4 será mostrado que a cauda das distribuições dos períodos de sobrecarga tem forte influência no resultado das perdas.

Tendo em vista o que foi dito, foi elaborado um novo método para ajustar a distribuição de Ds e Df pela hiper-exponencial escolhida. A estratégia usada consiste em encontrar os pontos b_1 e b_2 (ver Figura 4.4) que delimitam, aproximadamente, as regiões. De posse destes pontos, basta investigar o histograma das durações para encontrar os valores de λ_i e p_i. Vale ressaltar que esta solução para parametrização da distribuição hiper-exponencial faz parte do conjunto de contribuições deste trabalho.

Inicialmente busca-se b_2, depois b_1. A busca de b_2 envolve 2 etapas:

- encontrar λ_3 através de $\frac{d}{dt}(\ln P[D > t]_{r_3})$, onde r_i indica que t situa-se na região i:

$$P[D > t]_{r_3} \sim p_3 e^{-\lambda_3 t}$$

$$\ln P[D > t]_{r_3} \sim \ln(p_3 e^{-\lambda_3 t})$$

$$\ln P[D > t]_{r_3} \sim \ln p_3 + \ln e^{-\lambda_3 t}$$

$$\ln P[D > t]_{r_3} \sim \ln p_3 - \lambda_3 t$$

$$\frac{d}{dt}(\ln P[D > t]_{r_3}) \sim -\lambda_3 \ ;$$

- encontrar b_2 tal que a média das durações na região 3 seja $\Delta_3 = \frac{1}{\lambda_3}$.

De forma semelhante, a busca de b_1 envolve as mesmas duas etapas:

- encontrar λ_2 através de:

$$P[D > t]_{r2} - p_3 e^{-\lambda_3 t} \sim p_2 e^{-\lambda_2 t}$$

$$\ln(P[D > t]_{r2} - p_3 e^{-\lambda_3 t}) \sim \ln p_2 + \ln e^{-\lambda_2 t}$$

$$\ln(P[D > t]_{r2} - p_3 e^{-\lambda_3 t}) \sim \ln p_2 - \lambda_2 t$$

$$\frac{d}{dt}\ln(P[D > t]_{r_2} - p_3 e^{-\lambda_3 t}) \sim -\lambda_2 \ .$$

- encontrar b_1 tal que a média das durações na região 2 seja $\Delta_2 = \frac{1}{\lambda_2}$.

Os parâmetros p_1, p_2 e p_3 são encontrados através do histograma de durações. Ou seja, cada um dos três estados da hiper-exponencial representa uma região diferente do histograma de durações, como mostrado na Figura 4.5 para o exemplo da seção 3.4, com r equivalente à 90% de carga. A Figura 4.6 mostra um exemplo de como cada exponencial influencia a hiper-exponencial.

O problema para encontrar λ_3 equivale a aproximar um grupo de pontos através de uma reta. No entanto, como b_2 é inicialmente desconhecido, não é possível saber o tamanho deste grupo. Sendo assim, o método usado para aproximar a reta que representa a taxa média na região 3 envolve um processo iterativo que calcula várias retas até que um critério de parada seja atingido.

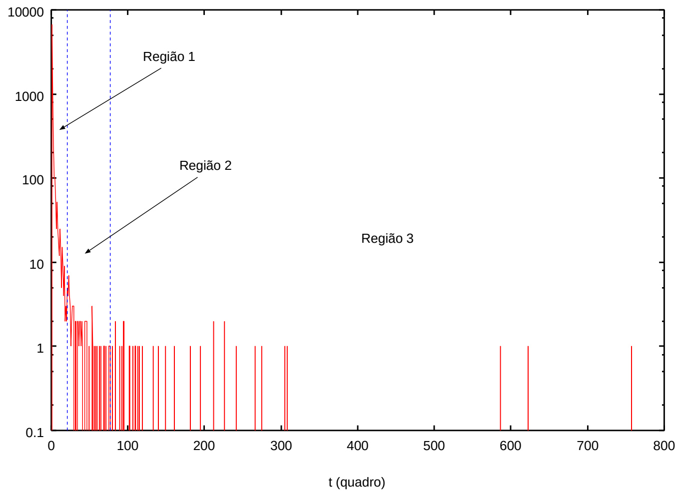


Figura 4.5: Histograma da duração das sobrecargas para uma carga de 90% e sua divisão nas 3 regiões a serem representadas pela hiper-exponencial

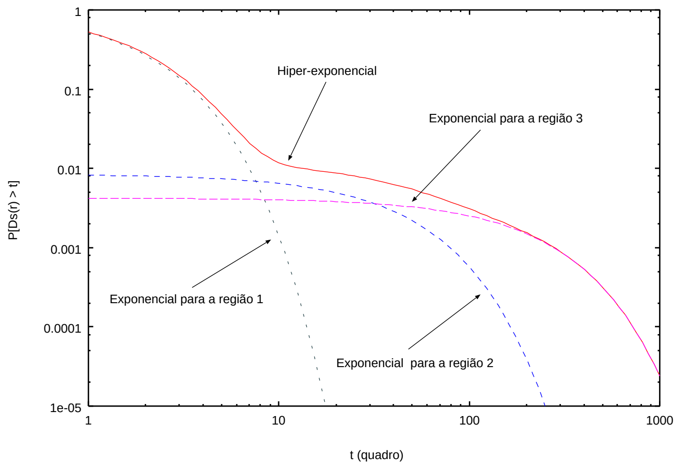


Figura 4.6: Distribuição hiper-exponencial escolhida e a parcela de cada exponencial ponderada

O processo usado para buscar λ_3 calcula a inclinação de sucessivas retas ligando o último ponto da distribuição $[t_n, P_n]$ aos pontos à esquerda em ordem decrescente. Quando a diferença percentual entre a inclinação atual e a anterior for menor que uma dada tolerância ϵ, tem-se λ_3. ϵ é aqui chamado "tolerância da taxa". A Figura 4.7 ilustra a idéia do algoritmo. Assume-se que uma diferença mínima entre duas inclinações sucessivas indica que os pontos fazem parte da reta correspondente à cauda da distribuição. O final da distribuição, que aparece na área destacada na figura, foi propositalmente exagerado para mostrar possíveis decaimentos decorrentes do truncamento existente na distribuição real.

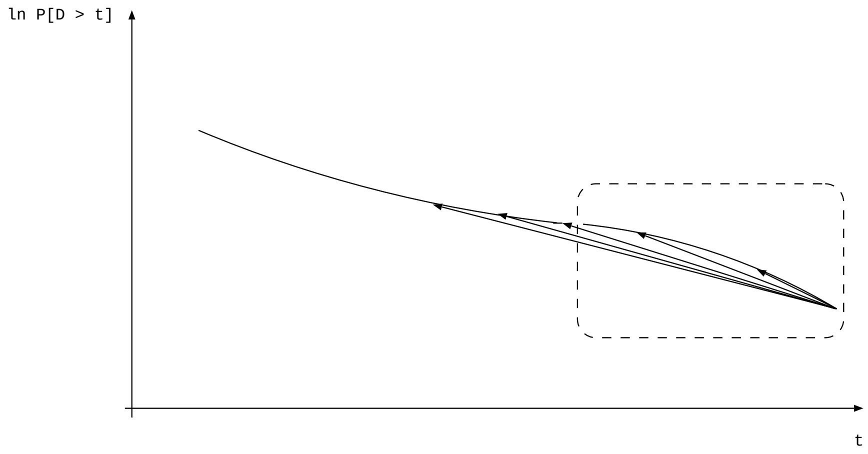


Figura 4.7: Exemplo ilustrando método de busca para λ_3

Seja uma seqüência real de n pontos $[t_k, P_k]$ representando uma distribuição de D, onde D pode ser Ds ou Df, P_k é o valor correspondente a $P[D > t]$ para $t = t_k$, e $[t_n, P_n]$ o índice ponto mais a direita da distribuição cujo $P_n \neq 0$. Temos o seguinte passo de iteração:

$$\lambda_3(i) = -\frac{\ln P_{i-1} - \ln P_n}{t_{i-1} - t_n} = -\frac{\ln(\frac{P_{i-1}}{P_n})}{t_{i-1} - t_n} \ ,$$

onde i assume o valor inicial de n e é decrementado unitariamente a cada passo da iteração.

λ_3 corresponderá a $\lambda_3(i-1)$ para o menor i tal que

$$\frac{\lambda_3(i) - \lambda_3(i-1)}{\lambda_3(i)} \leq \epsilon \ .$$

Tem-se, então, que

$$\Delta_3 = \frac{1}{\lambda_3} \ ,$$

onde Δ_3 é a duração média atribuída à região 3.

Note que $P[D > t]$ é uma curva que decresce monotonicamente, logo $P_{i-1} \geq P_i$ para qualquer $i > 0$.

Uma vez determinado λ_3, torna-se necessário encontrar b_2 e, conseqüentemente, p_3. Para tal, investiga-se o histograma da seqüência estudada (Figura 4.5). Seja γ_i, o número de durações D com valor t_i. O algoritmo proposto encontra o menor valor de t para o qual a média da região 3 é aproximadamente Δ_3. Ou seja, procura-se s igual ao maior valor de v que satisfaz

$$\frac{\sum_{j=0}^{v} \gamma_{n-j} t_{n-j}}{\sum_{l=0}^{v} \gamma_{n-l}} = \Delta_3 \pm \psi \ ,$$

onde ψ é aqui chamado "tolerância da média". O valor t_{n-s} limita a região 3 e corresponde a b_2. Desta forma,

$$p_3 = \frac{\sum_{j=0}^{s} \gamma_{n-j}}{\sum_{u=0}^{n} \gamma_u} \ .$$

Embora Δ_3 seja a média sugerida pela inclinação da reta na região 3, para mantermos a coerência com o histograma das durações a duração média para a região 3 será também extraída do histograma, e este será o parâmetro usado na distribuição hiper-exponencial:

$$\Delta_3' = \frac{\sum_{j=0}^{s} \gamma_{n-j} t_{n-j}}{\sum_{u=0}^{n} \gamma_u} \ .$$

O mesmo procedimento é adotado para determinar λ_2 e b_1. Porém, neste caso, a inclinação de reta buscada a partir de $P[D > t]$ deve considerar a exponencial encontrada para a região 3. Assim, temos que:

$$\frac{\ln\left(\frac{P_{n-i-2} - p_3 e^{-\lambda_3 t_{n-i-2}}}{P_{n-s-1} - p_3 e^{-\lambda_3 t_{n-s-1}}}\right)}{t_{n-i-2} - t_{n-s-1}} = -\lambda_2(i) \ ,$$

onde o valor inicial de i é s e é decrementado unitariamente a cada iteração. λ_2 corresponderá a $\lambda_2(i-1)$ para o menor i tal que

$$\frac{\lambda_2(i) - \lambda_2(i-1)}{\lambda_2(i)} \leq \epsilon \ .$$

Desta forma,

$$\Delta_2 = \frac{1}{\lambda_2} .$$

onde Δ_2 é a duração média atribuída à região 2.

Da mesma forma que anteriormente, para se obter p_2 procura-se o maior v tal que:

$$\frac{\sum_{j=1}^{v} \gamma_{n-s-j} t_{n-s-j}}{\sum_{l=1}^{v} \gamma_{n-s-l}} = \Delta_2 \pm \psi .$$

O primeiro tempo da região 2, b_1, corresponderá a t_{n-s-w}, onde w é o maior v que satisfaz a relação acima. Desta forma,

$$p_2 = \frac{\sum_{j=0}^{w} \gamma_{n-s-j}}{\sum_{u=0}^{n} \gamma_u} .$$

Do mesmo modo que para a região 3, a duração média para a região 2 será também extraída a partir do histograma:

$$\Delta_2' = \frac{\sum_{j=0}^{w} \gamma_{n-s-j} t_{n-s-j}}{\sum_{u=0}^{n} \gamma_u} .$$

Finalmente, tem-se a última região (região 1), de onde se extrai Δ_1' e p_1 facilmente por intermédio do histograma γ. Seja $g = n - s - w$,

$$\Delta_1' = \frac{\sum_{z=0}^{g-1} \gamma_z t_z}{\sum_{z=0}^{g-1} \gamma_z} \qquad \text{e}$$

$$p_1 = 1 - p_2 - p_3 = \frac{\sum_{z=0}^{g-1} \gamma_z}{\sum_{u=0}^{n} \gamma_u} .$$

Os parâmetros da hiper-exponencial que representa as distribuições de Ds ou Df consistem nas exponenciais com taxas $\frac{1}{\Delta_1'}$, $\frac{1}{\Delta_2'}$ e $\frac{1}{\Delta_3'}$ respectivamente ponderadas por p_1, p_2 e p_3.

A escolha de ϵ e ψ de ser cuidadosa pois pode não haver solução. Para determinar seus valores foi implementado o seguinte algoritmo:

1. escolhe-se um valor inicial de ϵ bem pequeno (foi usado 0.0001);

2. caso ϵ não seja suficiente para encontrar λ_3, aumenta-se progressivamente ϵ de um certo valor (e.g., 0.0001) até que λ_3 seja encontrado;

3. escolhe-se um valor inicial de ψ bem pequeno (foi usado 0.001);

4. caso ψ não seja suficiente para encontrar p_3, aumenta-se progressivamente ψ de um certo valor (e.g., 0.001) até que p_3 seja encontrado;

5. se os valores encontrados para ϵ e ψ não permitirem encontrar λ_2 e π_2, aumenta-se ϵ de um certo valor, encontra-se um novo λ_3 e retorna-se ao passo 3.

A Figura 4.8 mostra uma aproximação para a distribuição da duração das sobrecargas para a seqüência em MPEG4 do filme *Silence of the Lambs*, carga de 90%, usando o algoritmo proposto. No caso, foram usadas uma tolerância de taxa igual a 0.15% e uma tolerância de média igual a 0.55%. A Figura 4.9 mostra a robustez deste algoritmo frente a diferentes valores de tolerância. Estes valores devem ser tais que possibilitem a convergência da solução, ou seja, que possibilitem ao algoritmo encontrar três exponenciais ponderadas. Como podemos observar pela figura, todas as distribuições resultantes são muito próximas entre si, indicando que o algoritmo é bem tolerante quanto a seus parâmetros.

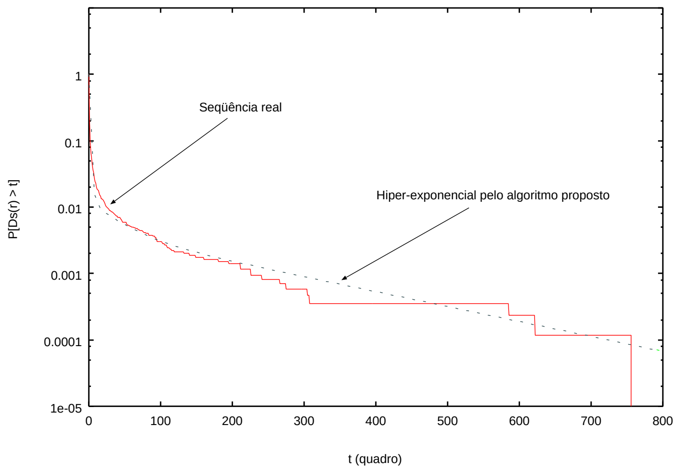


Figura 4.8: Distribuição para a seqüência real sendo aproximada pela hiper-exponencial extraída segundo o algoritmo proposto (para 90% de carga)

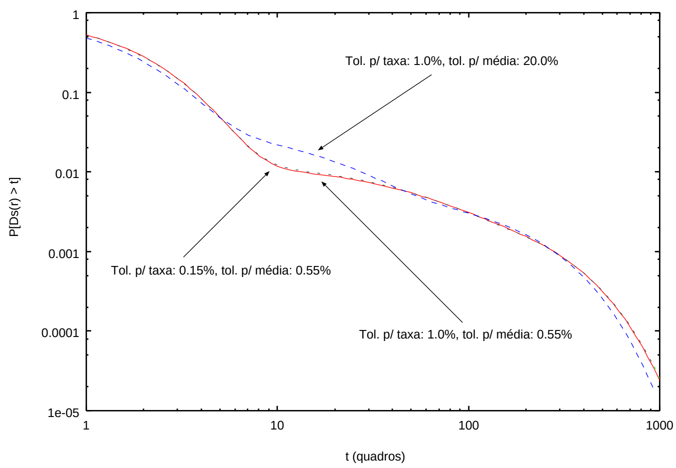


Figura 4.9: Distribuições de hiper-exponenciais geradas pelo algoritmo proposto a partir de diferentes valores de "tolerância de taxa" e "tolerância da média" (para 90% de carga)

4.3 Modelo de Tráfego usando Sobrecarga e Folga

O modelo de tráfego aqui apresentado tem a finalidade de representar as características de folga e sobrecarga para um determinado tráfego. Como já foi dito, um dos objetivos deste modelo é prever a perda numa fila FIFO, cuja taxa de serviço indica o *nível de sobrecarga* usado para dividir o modelo em folga e sobrecarga.

4.3.1 Cadeia de Markov do modelo proposto

O modelo proposto está ilustrado na Figura 4.10. Nele temos dois subconjuntos de estados, um representando os estados de sobrecarga (s_1, s_2 e s_3) e outro os estados de folga (f_1, f_2 e f_3). As taxas de transição entre os estados são calculadas a partir dos parâmetros das distribuições hiper-exponenciais das durações de sobrecarga e de folga. Estes parâmetros são obtidos segundo o algoritmo da seção 4.2. Por exemplo, a taxa de transição do estado s_1 para o estado f_2 é igual a $\alpha_1 q_2$, que são parâmetros das hiper-exponenciais.

A cada estado da cadeia está associada uma recompensa que indica o volume de tráfego gerado enquanto a cadeia estiver naquele estado. A parametrização das recompensas será descrita na seção 4.3.2.

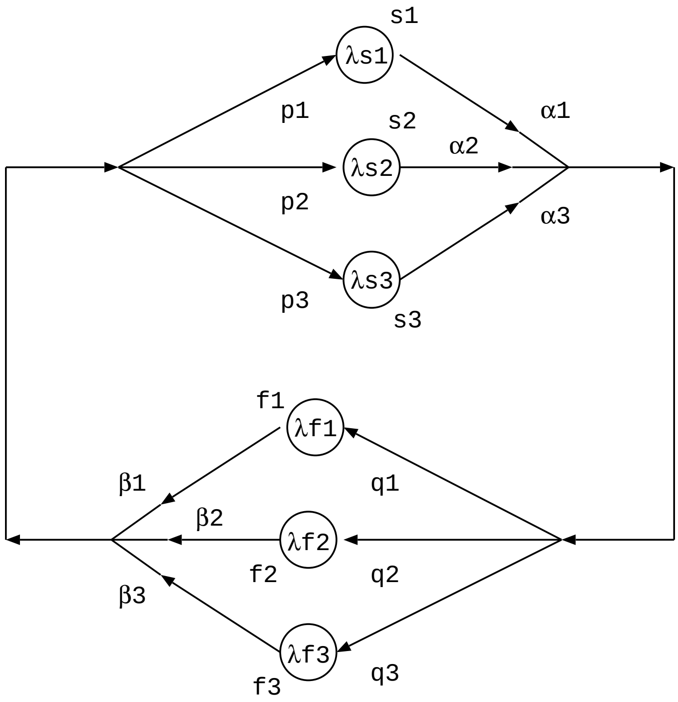


Figura 4.10: Cadeia de Markov para modelo de sobrecarga e folga

As recompensas associadas aos estados devem obedecer a seguinte relação:

$$\overline{\lambda_s} = \frac{\sum_{i=1}^{3} \pi_{s_i} \lambda_{s_i}}{\sum_{i=1}^{3} \pi_{s_i}} = \lambda_{s_{medido}} \ , \ \overline{\lambda_f} = \frac{\sum_{i=1}^{3} \pi_{f_i} \lambda_{f_i}}{\sum_{i=1}^{3} \pi_{f_i}} = \lambda_{f_{medido}} \ , \tag{4.1}$$

onde $\overline{\lambda_s}$ é a taxa média em sobrecarga do modelo, $\overline{\lambda_f}$ é a taxa média em folga do modelo, π_{s_i} é a probabilidade do estado s_i, π_{f_i} é a probabilidade do estado f_i, λ_{s_i} é a recompensa associada à s_i, λ_{f_i} é a recompensa associada à f_i, $\lambda_{s_{medido}}$ é a taxa média em sobrecarga para a seqüência real e $\lambda_{f_{medido}}$ a taxa média em folga.

Ou seja, a soma das recompensas dos estados de sobrecarga (ou folga) ponderada pela probabilidade estacionária do estado condicionada ao subconjunto de sobrecarga (ou folga) deve ser igual à taxa média em sobrecarga (ou folga) da seqüência real.

A taxa média do modelo é obtida por:

$$\overline{\lambda} = \pi_s \overline{\lambda_s} + \pi_f \overline{\lambda_f} \ .$$

Uma vez que podemos aglutinar os estados da cadeia em dois subconjuntos, um de sobrecarga e um de folga, cujas taxas de transição de sobrecarga para folga e de folga para sobrecarga são, respectivamente, $\frac{1}{\overline{Ds}}$ e $\frac{1}{\overline{Df}}$ [54], temos que as probabilidades em estado estacionário do modelo são calculadas a partir de:

$$\pi_s = \frac{\overline{Ds}}{\overline{Ds} + \overline{Df}} \ ,$$

$$\pi_f = \frac{\overline{Df}}{\overline{Ds} + \overline{Df}} \ ,$$

onde π_s é a probabilidade do subconjunto de sobrecargas e π_f é a probabilidade do subconjunto de folgas.

As expressões para $\overline{Ds}$ e $\overline{Df}$ são análogas às mostradas na subseção 4.2 para a média de uma variável aleatória hiper-exponencial.

4.3.2 Associando recompensas de taxa aos estados

Existem várias maneiras de associarmos recompensas de taxa aos estados. No entanto, a distribuição destas recompensas de taxa pode impactar as medidas de desempenho na fila. Uma maneira de perceber isto é observando a influência destas recompensas nos descritores mais tradicionais como a distribuição de taxas e a função de autocorrelação da taxa.

A influência da escolha das recompensas na distribuição de taxas é fácil de ser avaliada. Tomemos a maneira mais simples de associação de recompensas: todos os estados de folga possuem a mesma recompensa (λ_f) e todos os estados de sobrecarga a mesma recompensa (λ_s). Estas recompensas equivalem às taxas médias em sobrecarga e em folga da seqüência real conforme equação 4.1. Certamente que a distribuição de taxas para este caso, onde só existem duas recompensas será diferente da distribuição para o caso de termos recompensas diferentes associadas aos estados.

No caso da função de autocorrelação, podemos também verificar a influência das recompensas de taxa escolhidas para o modelo. Recordando a expressão para a autocorrelação, temos que:

$$r(\tau) = \frac{E[\lambda(t)\lambda(t+\tau)] - \overline{\lambda(t)}^2}{\sigma^2} \ ,$$

onde σ^2 é a variância da taxa representada pelo processo $\lambda(t)$.

Dos parâmetros que compõe a expressão, o termo $E[\lambda(t)\lambda(t+\tau)]$ pode variar bastante dependendo da escolha das recompensas, assim como a variância da taxa também irá se alterar.

A Figura 4.11 mostra um exemplo numérico da forma mais simples para associar recompensas de taxa ao modelo. A seqüência MPEG4 modelada neste exemplo é, outra vez, o filme *Silence of the Lambs*, alimentando uma fila com 33% de carga. Os valores de α_i, p_i, β_i e q_i, $i = 1,\ 2,\ 3$, foram calculados a partir do algoritmo da seção 4.2 com $\epsilon = 0.08$ e $\psi = 0.19$, para as durações de sobrecarga, e $\epsilon = 0.008$ e $\psi = 0.07$, para as durações de folga.

Neste método, basta atribuir como recompensa a taxa média em sobrecarga, para os três estados de sobrecarga, e a taxa média em folga, para os três estados de folga. No entanto, esta parametrização possui a desvantagem de apresentar pouca variabilidade na taxa de geração de dados do modelo. Alguns testes foram realizados com esta parametrização e alguns dos resultados obtidos para a fração de perda na fila ficaram bem diferentes dos obtidos para a seqüência real.

Uma forma de prover maior variabilidade às taxas do modelo é associar uma recompensa de taxa diferente para cada um dos estados. Observando a Figura 4.5, podemos notar que o histograma da duração das sobrecargas foi dividido em três regiões, segundo o algoritmo da seção 4.2. O mesmo foi feito para o histograma da duração das folgas. Cada uma dessas seis regiões é representada através de um estado no modelo.

As recompensas de taxa para cada um dos estados serão definidas baseadas nas taxas da seqüência real obtidas para cada uma das regiões:

$$\lambda_{s_i} = \sum_{j=1}^{n} \frac{\lambda_{s_j}^{(i)}}{n} ,$$

onde λ_{s_i} é a recompensa associada ao estado i, $\lambda_{s_j}^{(i)}$ é a taxa da seqüência real na região i e n o número de amostras da região i.

A Figura 4.12 mostra um exemplo numérico desta parametrização com múltiplas recompensas de taxa. O modelo representa a mesma seqüência usada no modelo da Figura 4.11. A Figura 4.13 mostra os histogramas de taxa da seqüência real para

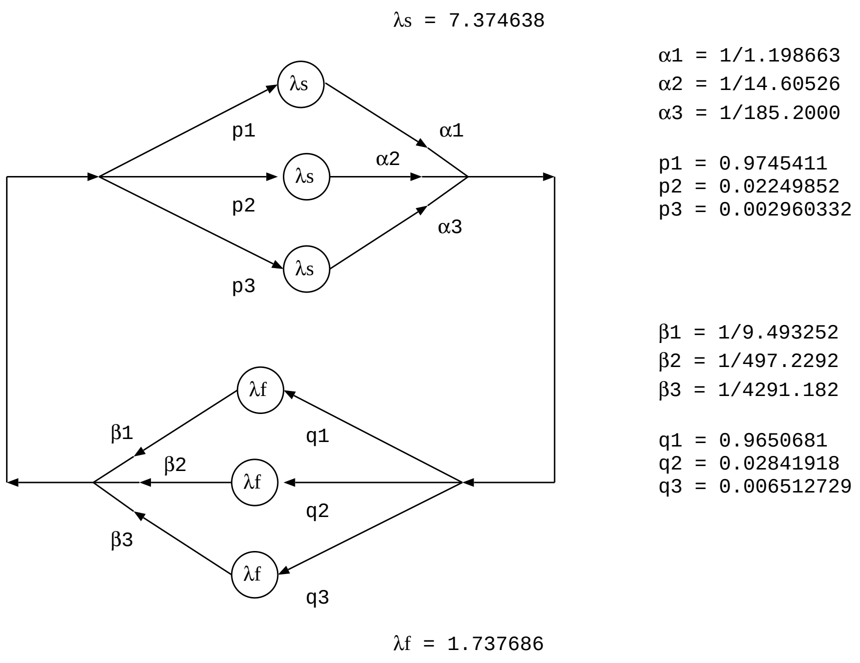


Figura 4.11: Parametrização do modelo proposto, para 33% de carga, usando duas recompensas de taxa, uma para sobrecarga e outra para folga

cada uma das três regiões do histograma de durações de sobrecarga e a Figura 4.14 mostra estes mesmos histogramas em escala logarítmica. Podemos observar que o histograma para cada uma das regiões é bem diferente.

Uma outra possibilidade é associar um modelo de histograma [41] a cada estado de acordo com seu nível, sobrecarga ou folga. Esta variante traz o benefício de ser possível casar melhor a distribuição de taxa. Todavia, aumenta bastante o número de estados do modelo e prejudica sua simplicidade. Parametrizando-se a cadeia desta forma tem-se que, uma vez entrando num estado de sobrecarga, as taxas são geradas segundo o modelo de histograma associado às sobrecargas e, passando-se para um estado de folga, estas passam a ser geradas segundo o modelo de histograma associado às folgas. Cada estado da cadeia, de sobrecarga ou folga, tem associado a ele os seguintes parâmetros (que são os parâmetros do modelo de histograma): vetor de recompensas, matriz de probabilidade de transição e vetor com a distribuição de probabilidade inicial.

Tais modelos de histograma são construídos inspecionando-se toda a seqüência real, ou seja, juntando-se todas sobrecargas e folgas em seqüências separadas para posterior inspeção de cada uma, conforme o procedimento mostrado em [41]. No

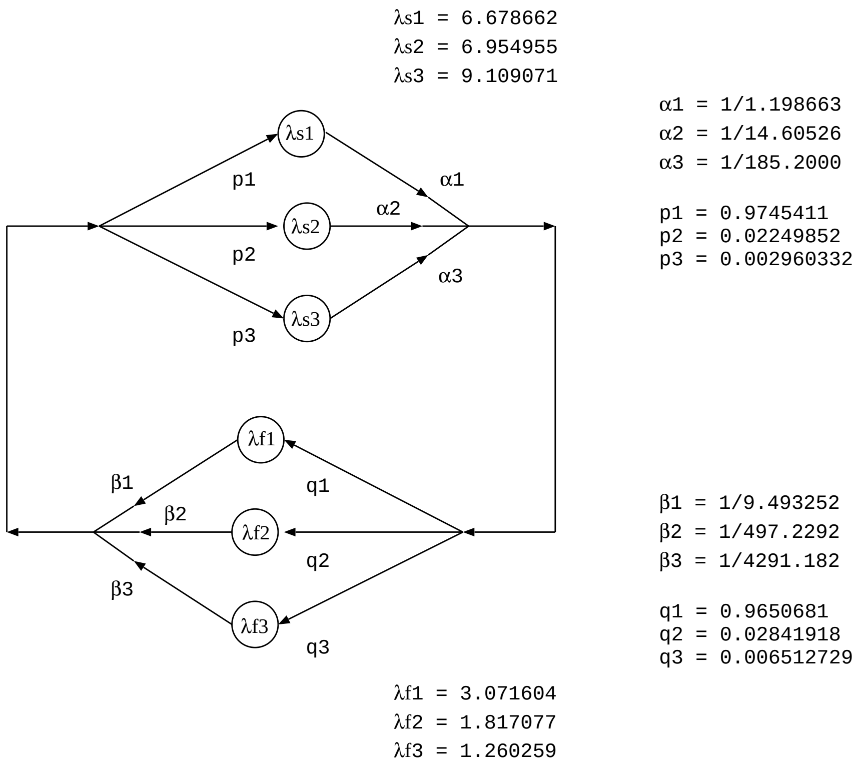


Figura 4.12: Parametrização do modelo proposto, para 33% de carga, usando múltiplas recompensas de taxa

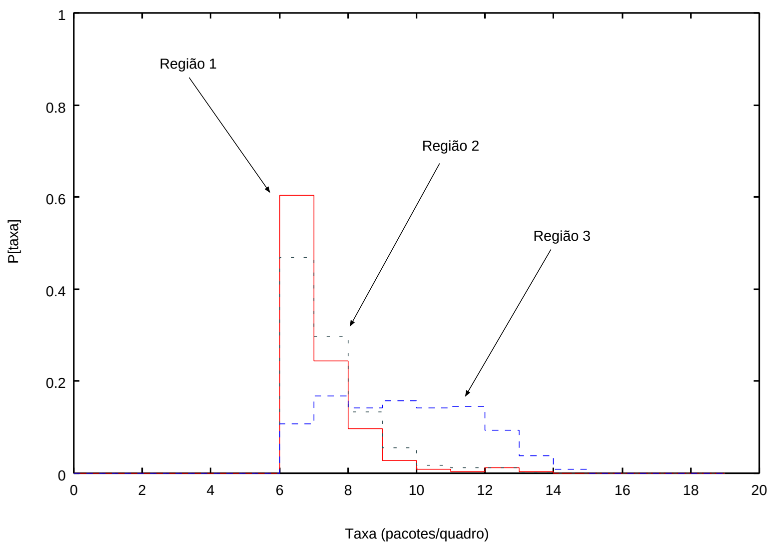


Figura 4.13: Histograma das taxas da seqüência real para cada região de duração de sobrecarga, dado 33% de carga

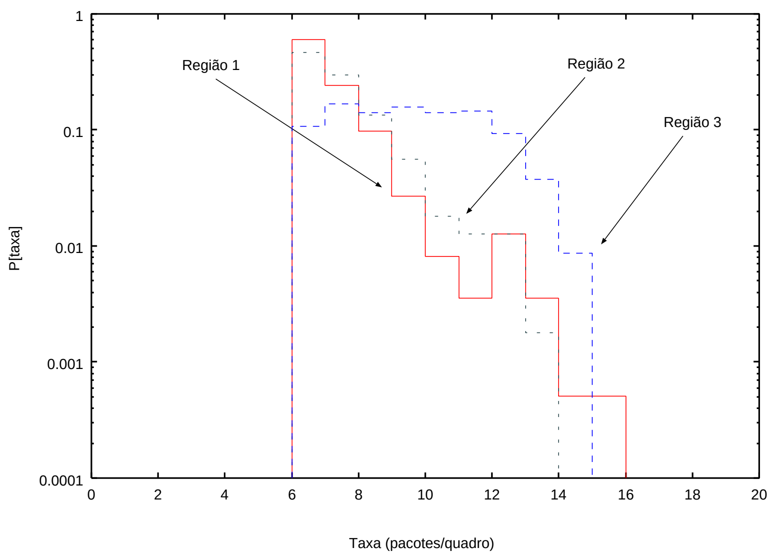


Figura 4.14: Histograma em escala logarítmica das taxas da seqüência real para cada região de duração de sobrecarga, dado 33% de carga

caso, as recompensas de taxa são obtidas dividindo-se o intervalo entre a taxa mínima e a máxima em n faixas, cada uma correspondendo a um estado do respectivo modelo de histograma. Portanto, o modelo de histograma conterá um número de estados igual ao número de faixas em que as taxas foram divididas. Além de mais complexa, esta maneira de representar a variação das taxas quando em sobrecarga e quando em folga revelou-se inadequada, como explicado adiante na seção 4.3.3, e portanto não foi escolhida.

4.3.3 Resultados experimentais usando seqüências reais

Dadas as três propostas para associação de recompensas de taxa aos estados, verificaremos qual delas caracteriza melhor o tráfego em relação às perdas. Para tal, vamos utilizar o exemplo da seção 3.4, o filme *Silence of the Lambs*, e caracterizá-lo em função do nível de sobrecarga referente à carga de 33%. As probabilidades e taxas de transição das distribuições hiper-exponenciais são as mesmas obtidas no exemplo da seção anterior para este mesmo modelo.

Como visto na seção 4.3.2, a mudança nas recompensas do modelo vai alterar o seu histograma de taxa e a sua função de autocorrelação. Portanto, iremos analisar o comportamento destes descritores para as diferentes associações de recompensa e relacioná-los com a fração de perda obtida. Primeiramente serão mostradas as frações de perda para modelos e seqüência, em seguida estas medidas serão avaliadas comparando-se os descritores acima.

As Figuras 4.11 e 4.12, mostradas na seção anterior, contêm, respectivamente, os modelos usando duas e múltiplas recompensas para sobrecarga e folga. O modelo que usa o modelo de histograma para representar as recompensas é o mesmo da Figura 4.11, sendo que no lugar de λ_s e λ_f tem-se um vetor de recompensas com sua respectiva matriz de probabilidades de transição. Abaixo, seguem os vetores e matrizes de probabilidades para os estados de sobrecarga e folga. Também aparecem listados os vetores com as probabilidades iniciais usadas para se entrar em um estado de sobrecarga ou em um estado de folga. Os histogramas para as sobrecargas têm 6 estados e os histogramas para as folgas têm 3 estados. Estes números de estados foram escolhidos observando-se o gráfico da Figura 4.28, que contém as taxas da seqüência real.

Parâmetros do modelo de histograma para os estados de sobrecarga para 33% de carga:

$$\Lambda = < 6.331087 \;\; 8.000000 \;\; 9.356234 \;\; 11.000000 \;\; 12.261111 \;\; 14.375008 > \;\; ,$$

$$\Pi_{inicial} = < 0.854944 \;\; 0.099467 \;\; 0.028419 \;\; 0.001776 \;\; 0.014802 \;\; 0.000592 > \;\; ,$$

$$\mathbf{P} = \begin{bmatrix} 0.885377 & 0.065318 & 0.037927 & 0.002528 & 0.008850 & 0 \\ 0.401554 & 0.494819 & 0.077720 & 0.005181 & 0.018137 & 0.002591 \\ 0.226463 & 0.086514 & 0.569975 & 0.081425 & 0.035623 & 0 \\ 0.041379 & 0.013793 & 0.220690 & 0.531034 & 0.186207 & 0.006897 \\ 0.122222 & 0.022222 & 0.094444 & 0.144444 & 0.6 & 0.016666 \\ 0.125 & 0 & 0 & 0.25 & 0.25 & 0.375 \end{bmatrix} .$$

Parâmetros do modelo de histograma para os estados de folga para 33% de carga:

$$\Lambda = < 0.851682 \;\; 2.333443 \;\; 4.347132 > \;\; ,$$

$$\Pi_{inicial} = < 0.017170 \;\; 0.402605 \;\; 0.580225 > \;\; ,$$

$$\mathbf{P} = \begin{bmatrix} 0.845576 & 0.144054 & 0.010370 \\ 0.195372 & 0.735833 & 0.068796 \\ 0.060815 & 0.292066 & 0.647119 \end{bmatrix} .$$

A Figura 4.15 mostra as frações de perdas dada a carga de 33%. Estas perdas foram extraídas por simulação em função do tamanho máximo da fila. Os intervalos de confiança representam 95% de probabilidade. Pode-se claramente verificar que o modelo com múltiplas recompensas apresenta a melhor caracterização da fração de perda. É possível notar também a ineficiência do modelo que usa o modelo de histogramas para representar as recompensas nos estados de sobrecarga e folga, uma vez que o número de estados aumenta consideravelmente sem a correspondente melhora na predição da perda.

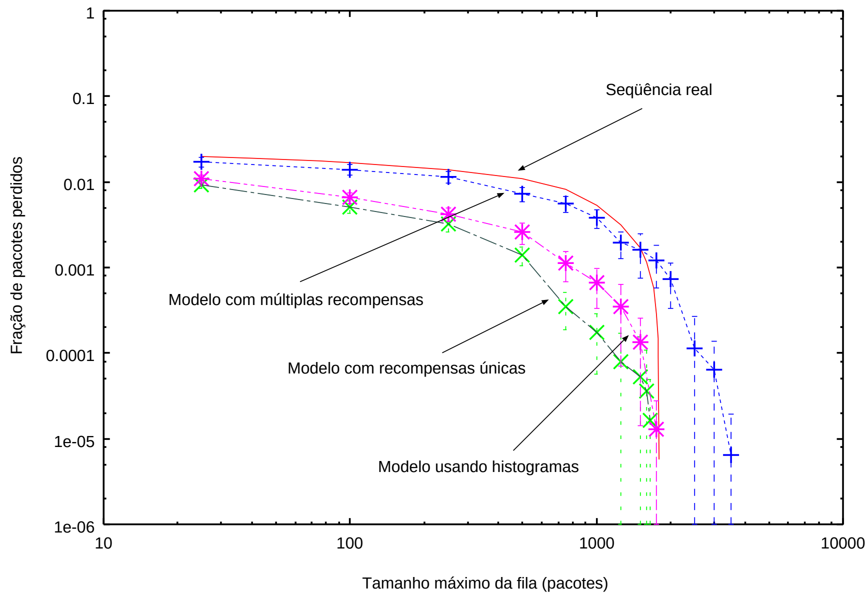


Figura 4.15: Fração de perda de pacotes para carga de 33%

As Figuras 4.16 e 4.18 mostram o histograma de taxa e a função de autocorrelação, respectivamente. O nível de sobrecarga r aparece indicado no histograma através de uma linha vertical. Para melhor visualizar as probabilidades referentes às taxas mais altas, a Figura 4.17 traz o histograma mostrado com escala logarítmica. Ambos os descritores foram extraídos analiticamente para os modelos. No caso dos histogramas, estes foram gerados a partir das probabilidades estacionárias dos estados.

A avaliação que pode ser feita através dos histogramas, para estes exemplos, é pouco conclusiva. O modelo que usa o modelo de histograma para representar as recompensas de taxa dos estados possui a melhor aproximação para este descritor mas apresenta pior casamento com a curva que indica a fração de perda. Por outro lado, a autocorrelação, para este exemplo, indica qual modelo aproxima melhor a fração de perda da seqüência real.

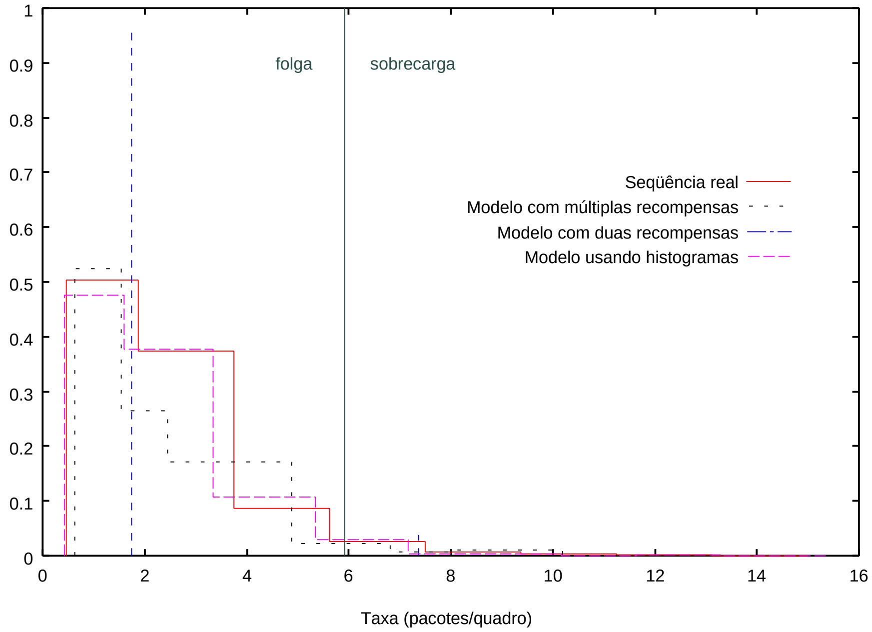


Figura 4.16: Histogramas de taxa para modelos ajustados para 33% de carga e seqüência MPEG4

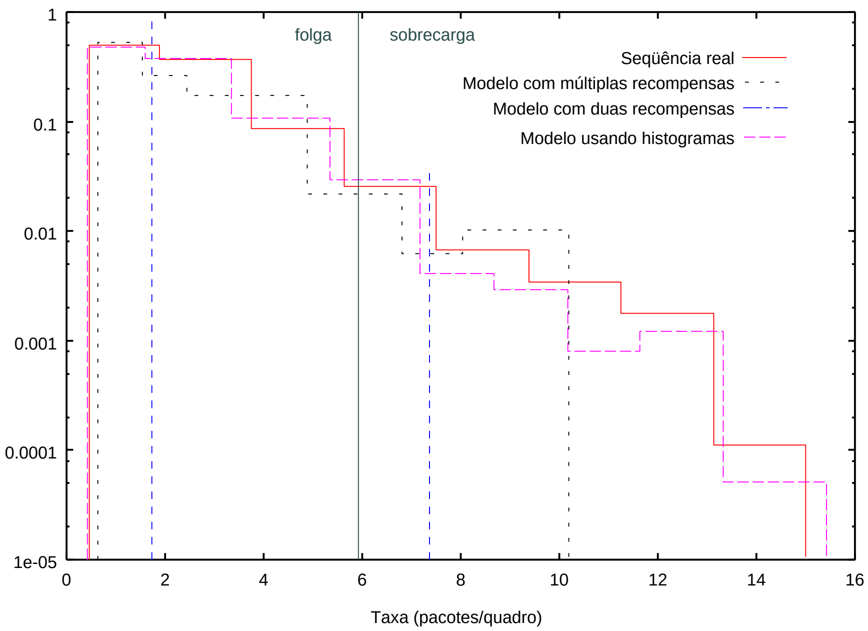


Figura 4.17: Histogramas de taxa para modelos ajustados para 33% de carga e seqüência MPEG4, escala logarítmica

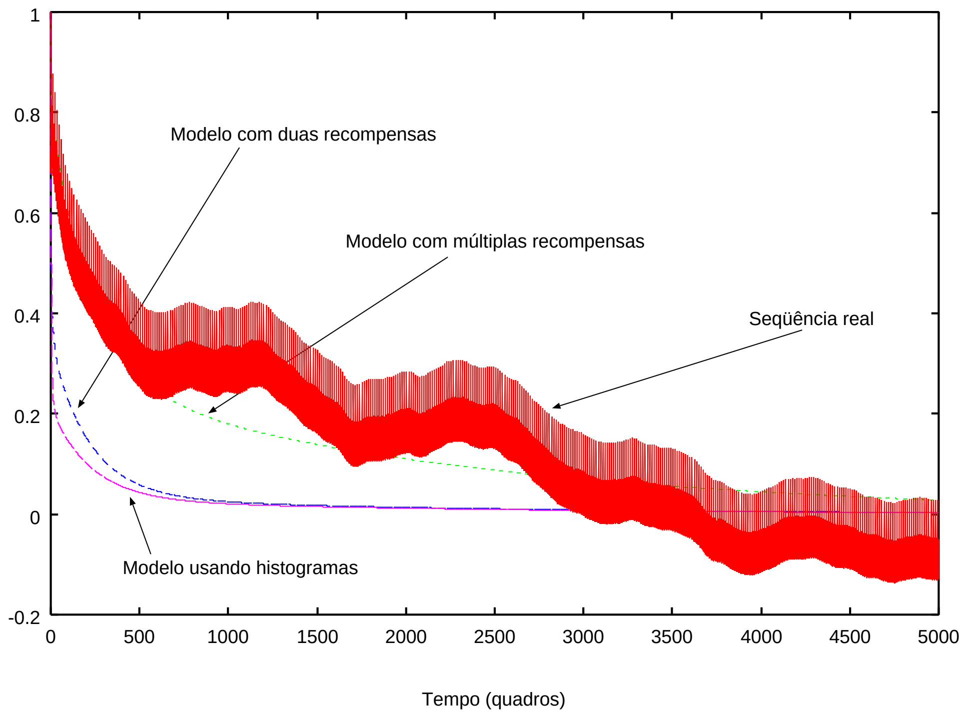


Figura 4.18: Autocorrelação da taxa para modelos ajustados para 33% de carga e seqüência MPEG4

No caso do modelo que usa modelos de histograma representando taxas em sobrecarga e folga, a taxa média em sobrecarga ou folga será dada pela evolução destes modelos de histograma durante os períodos de sobrecarga ou folga partindo sempre de uma mesma distribuição inicial. Isto significa que esta configuração não necessariamente é capaz de fornecer a mesma taxa média em sobrecarga ou folga da seqüência real uma vez que, para tal, é necessário que o período em sobrecarga ou folga seja suficientemente grande para que os modelos de histograma entrem em regime estacionário. Conseqüentemente, a taxa média do modelo não necessariamente será a mesma da seqüência. Abaixo são mostradas as taxas médias extraídas para a seqüência e o modelo usando modelo de histograma. Pode-se verificar que este modelo não conseguiu capturar a taxa média do tráfego em função de não modelar corretamente as taxas médias em sobrecarga e folga.

	seqüência real	Modelo c/ histog. p/ 33%
Taxas médias	1.954032	2.025086

Comparando-se as curvas de autocorrelação para os modelos, é possível notar que tanto a curva de autocorrelação quanto a curva de perda, para o modelo com múltiplas recompensas, se aproximam melhor da seqüência real. Neste exemplo podemos dizer que o modelo mais adequado é aquele com a função de autocorrelação mais próxima da autocorrelação da seqüência real, o que é coerente com os estudos apresentados na literatura [15, 27, 25]. No entanto, como este descritor não é usado para parametrizar o modelo proposto, cabe aqui uma análise de um dos descritores usados para parametrizar o modelo, a taxa média em sobrecarga para as três escalas de tempo consideradas.

A Figura 4.19 mostra as diferenças nos histogramas de taxa de sobrecarga, dado 33% de carga, para as três escalas de tempo indicando o percentual de quadros em cada escala. O mesmo histograma é apresentado na Figura 4.20, porém com escala logarítmica. Pode-se verificar nas figuras que a taxa média em sobrecarga para a região 3 (região com as maiores durações) é maior que nas demais. Trata-se de uma informação importante que não é capturada caso se associe uma única recompensa de taxa aos três estados. Isto significa dizer que, no caso de múltiplas recompensas, a distribuição das taxas passa a ser melhor capturada.

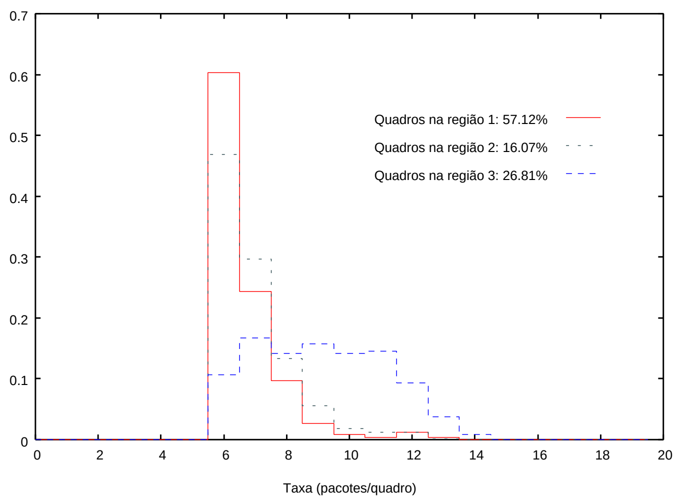


Figura 4.19: Histogramas de taxa em sobrecarga dada a escala de duração associada, 33% de carga

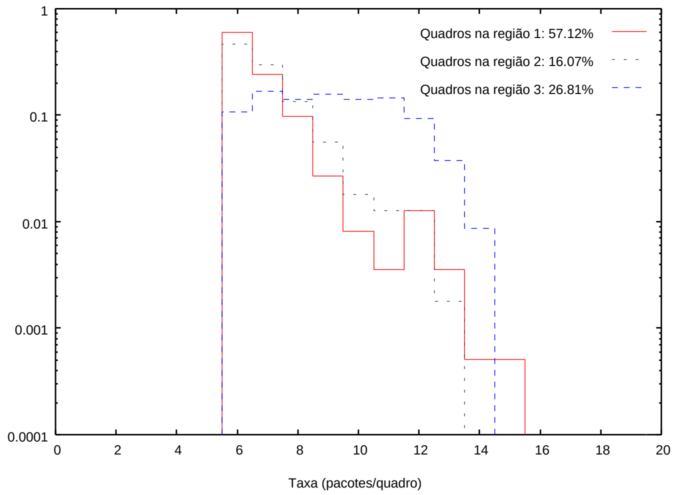


Figura 4.20: Histogramas de taxa em sobrecarga em escala logarítmica dada a escala de duração associada, 33% de carga

4.4 Considerações sobre a Sensibilidade da Fração de Perda aos Parâmetros do Modelo Proposto

Na elaboração do modelo nas seções anteriores, buscou-se obter um modelo simples, de fácil caracterização, mas que pudesse prever com precisão as medidas de interesse. Obviamente é possível simplificar ainda mais o modelo, ou pela diminuição do número de estados ou pelo uso de um menor número de recompensas , conforme verificado em um dos modelos testados na seção anterior. Naquele exemplo usou-se uma mesma recompensa de taxa para os estados de sobrecarga e outra para os estados de folga. Uma outra possibilidade de simplificação seria manter recompensas distintas nos estados de sobrecarga, mas usar apenas a taxa média em folga como recompensa única para todos os estados de folga; ou ainda, usar apenas um estado para a folga. O objetivo desta seção é estudar a sensibilidade do modelo proposto frente aos parâmetros usados em sua construção. A seqüência utilizada nos exemplo é o filme *Silence of the Lambs* codificado em MPEG4. Todos os intervalos de confiança foram calculados com 95% de probabilidade. Todos os modelos desta seção foram extraídos para uma carga de 33%.

4.4.1 Sensibilidade em relação aos estados de folga

As folgas são estados nos quais não ocorrem perdas. Sua influência na ocupação da fila está relacionada somente com o esvaziamento da mesma. Para que a taxa de transmissão durante a folga e a duração da folga tenham pouca influência na perda observou-se que, na maioria dos casos estudados, é necessário que as durações das folgas sejam suficientes para esvaziar a fila. Em geral, o evento de esvaziamento durante a folga é mais freqüente quanto menor é a carga e, conseqüentemente, maior a taxa de serviço alocada. No entanto, se a avaliação de desempenho estiver baseada em medidas transientes de perda para um tempo relativamente pequeno, este tipo de simplificação pode prejudicar a precisão do modelo em prever a distribuição das perdas.

A Figura 4.21 compara a fração de perda obtida para o modelo com recompensas múltiplas, doravante chamado *modelo de referência*, com aquela obtida para o modelo onde as recompensas de taxa nos estados de folga são iguais à taxa média em folga.

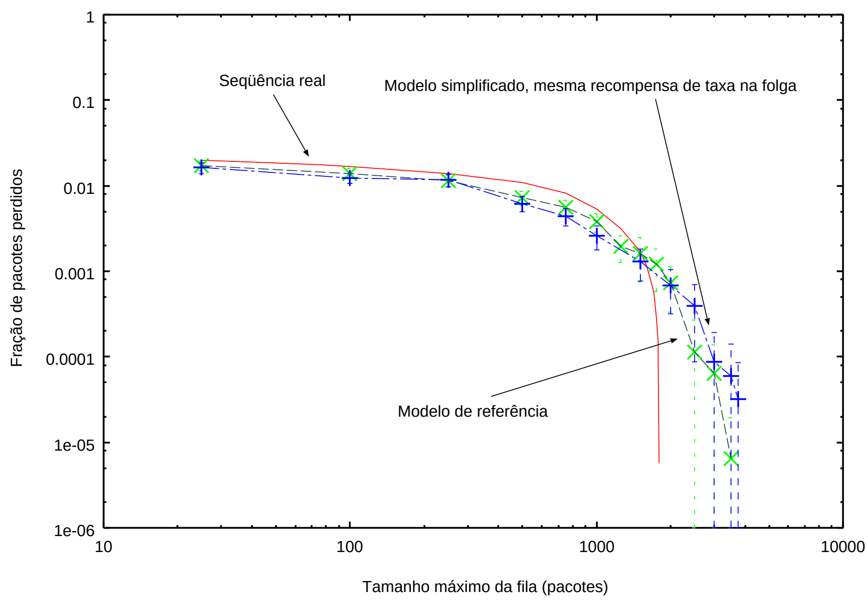


Figura 4.21: Comparação da fração de perda para modelo com apenas um valor de recompensa de taxa nos estados de folga, carga de 33%

Pode-se ver claramente que esta simplificação teve pouquíssimo impacto na fração de perda dada a comparação com a mesma para o modelo de referência, o que está de acordo com o comentário acima. Portanto é baixa a sensibilidade da fração de perda em relação à distribuição da recompensa nos períodos de folga.

Seja o caso onde a folga passa a ser representada apenas por um único estado. Esta simplificação significa que a duração da folga é aproximada por uma distribuição exponencial. A Figura 4.22 compara os resultados para este modelo.

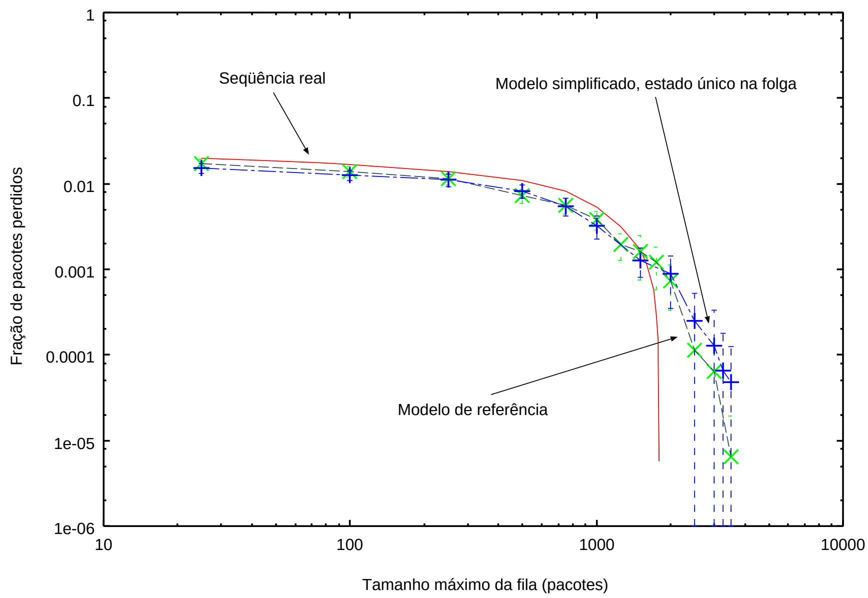


Figura 4.22: Comparação da fração de perda para modelo com apenas um estado de folga, carga de 33%

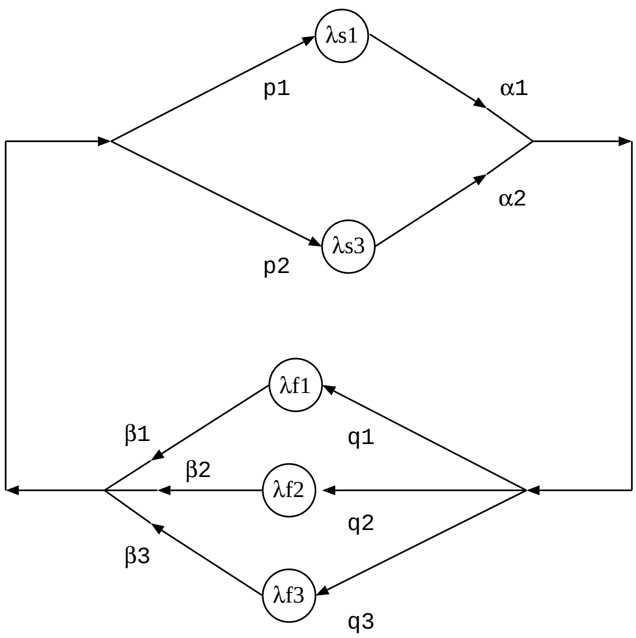
Como é possível notar, a simplificação feita ainda é aceitável mesmo para uma faixa maior de tamanho máximo de fila. Assim sendo, para este exemplo, a sensibilidade da perda aos estados de folga, de uma maneira geral, é bastante baixa.

4.4.2 Sensibilidade em relação aos estados de sobrecarga

Intuitivamente que a sensibilidade da perda aos parâmetros que descrevem o comportamento da sobrecarga deve ser maior que no caso das folgas, já que são nos períodos de sobrecarga que as perdas ocorrem. Como a subseção 4.3.3 mostra a sensibilidade do modelo com relação à distribuição das recompensas de taxa nos estados de sobrecarga, nesta seção será mostrado um exemplo do comportamento da distribuição das perdas frente a diferentes caracterizações das durações de sobrecargas.

Reduzindo o número de estados

Primeiramente, veremos como as perdas se comportam perante uma redução de três para dois no número de estados usados para descrever a hiper-exponencial de sobrecarga. Estes dois estados foram obtidos juntando-se as duas regiões de menor escala de tempo do modelo de referência e associando-se, ao estado resultante, uma recompensa de taxa que mantenha a mesma taxa média em sobrecarga, o que está de acordo com o procedimento descrito na seção 4.3. Desta forma, preserva-se a cauda da distribuição, uma vez que o estado que descreve a escala de tempo maior é o mesmo. Como já mencionado em 4.1, a cauda desta distribuição tem grande influência na perda, principalmente em filas grandes. O modelo com dois estados de sobrecarga pode ser visualizado na Figura 4.23. A Figura 4.24 mostra o gráficos com a distribuição complementar das sobrecargas para os modelos e a seqüência real.



33% :

λs1 = 6.739320
λs2 = 9.109071

α1 = 1/1.501188
α2 = 1/185.2000

p1 = 0.9970397
p2 = 0.0029603

λf1 = 3.071604
λf2 = 1.817077
λf3 = 1.260259

β1 = 1/9.493252
β2 = 1/497.2292
β3 = 1/4291.182

q1 = 0.9650681
q2 = 0.02841918
q3 = 0.006512729

Figura 4.23: Modelo reduzido com 2 estados em sobrecarga

A Figura 4.25 compara a fração de perdas para os modelos citados e a fração de perdas para a seqüência real. Pode-se observar que a redução de estados feita não

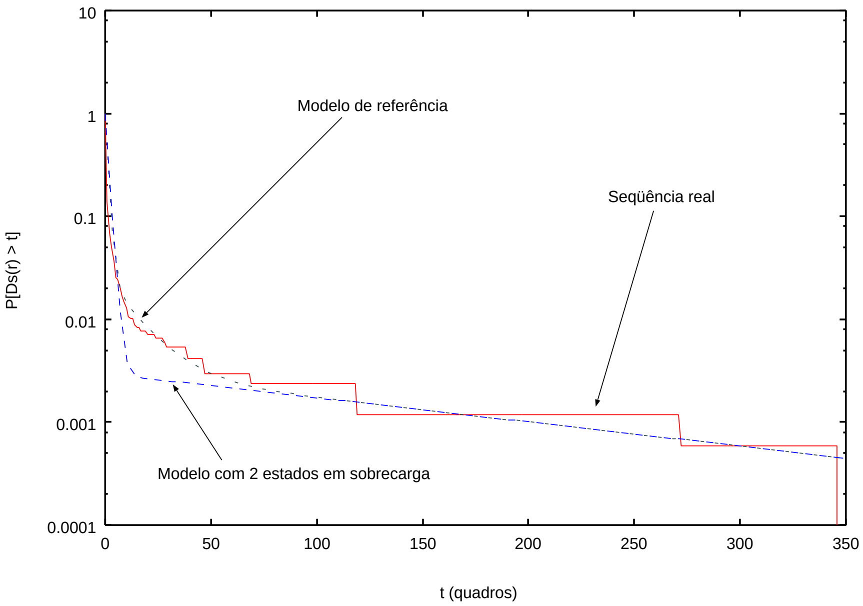


Figura 4.24: Distribuição das durações de sobrecarga para modelo com 2 estados de sobrecarga e r referente a 33%

ocasionou nenhum impacto na predição da fração de perdas para esta utilização. Esta insensibilidade está relacionada com a manutenção no modelo da cauda da distribuição da duração de sobrecarga, como poderá ser comprovado a seguir.

Ajuste empírico da distribuição hiper-exponencial

O próximo teste procura verificar como a taxa de perdas se comporta variando os parâmetros que descrevem a hiper-exponencial de sobrecarga. Neste caso, ao invés de se usar o algoritmo proposto na seção 4.2, duas hiper-exponenciais foram obtidas através de um ajuste visual a partir do gráfico de distribuição das durações de sobrecarga para a seqüência real. Este ajuste é feito variando-se os parâmetros da distribuição até que esta possua a aproximação desejada. O objetivo deste ajuste é obter dois modelos, um aproximando a distribuição para valores relativos a tempos "pequenos" (isto é, $P[D > t]$ para t pequeno) e outro para tempos maiores. A Figura 4.26 mostra as distribuições resultantes das aproximações "visuais" (empíricas) para uma carga de 33%. Vale lembrar que, em decorrência deste novo ajuste, novas recompensas de taxa em sobrecarga são obtidas para o novo modelo, uma vez que as *escalas de duração* das sobrecargas foram alteradas.

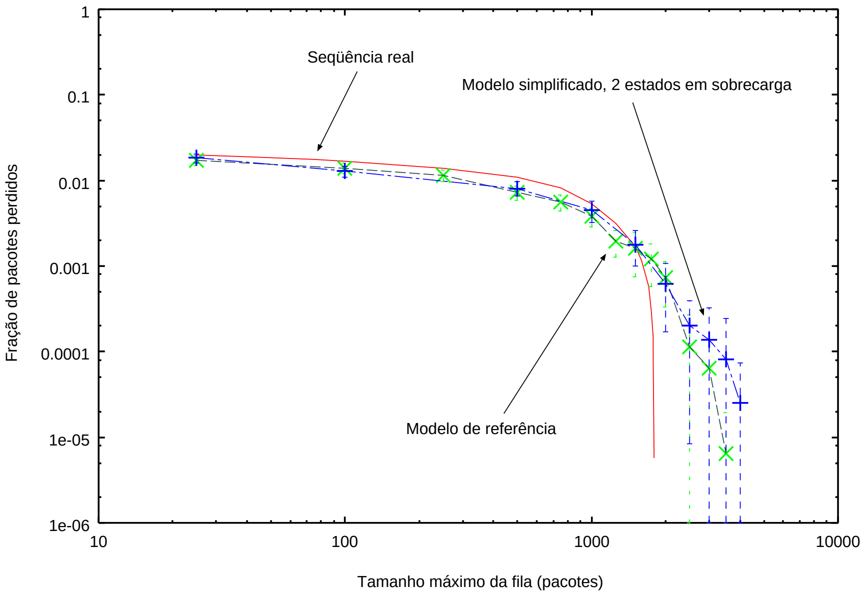


Figura 4.25: Comparação da fração de perda para modelo com 2 estados de sobrecarga, carga de 33%

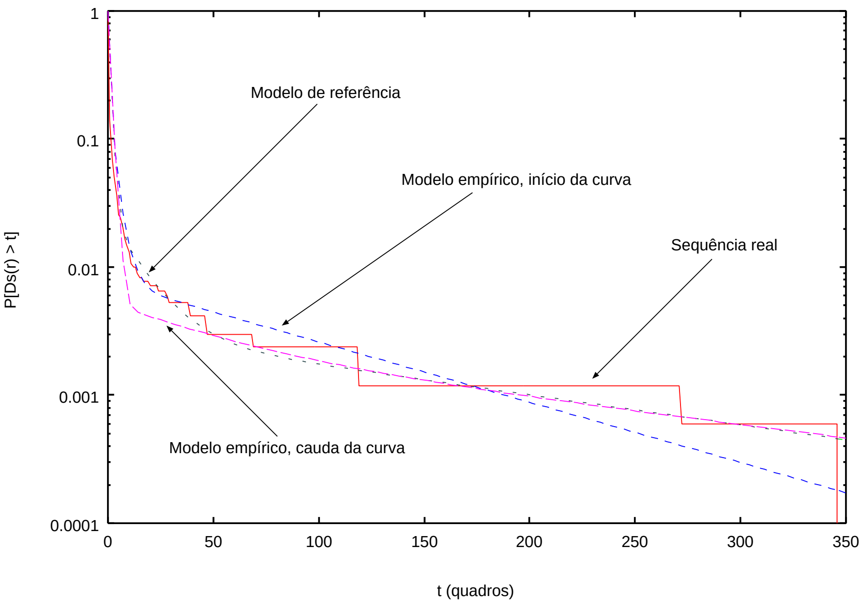


Figura 4.26: Distribuição das durações de sobrecarga para modelos com ajuste empírico da distribuição de sobrecarga e r referente a 33%

A Figura 4.27 mostra a fração de perdas para os modelos contendo as distribuições empíricas em comparação com o modelo de referência e a seqüência real. Como se pode notar, os modelos que não aproximaram a cauda da distribuição sofreram degradação na predição das perdas. Já os modelos que apenas aproximaram melhor a cauda das distribuições resultaram em perdas bem mais próximas que as da seqüência real.

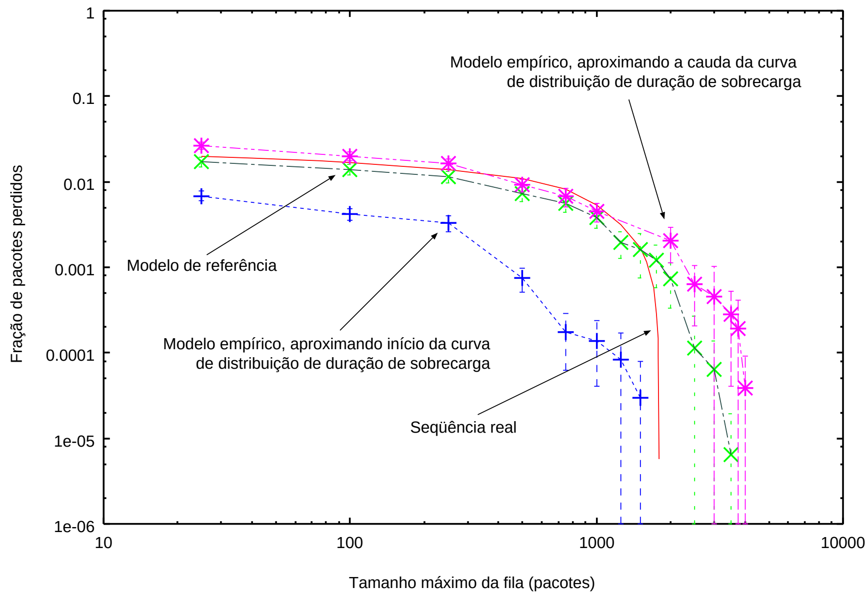


Figura 4.27: Comparação da fração de perda para modelos com ajuste empírico da distribuição de sobrecarga e r referente a 33%

4.5 Eficiência do Modelo em Função do Nível de Sobrecarga

Como visto neste capítulo, as medidas de perda são muito mais sensíveis à distribuição das sobrecargas do que a das folgas. A distribuição das sobrecargas está relacionada com dois descritores usados para parametrizar o modelo: a distribuição dos períodos de sobrecarga e as taxas médias associadas a estes períodos.

O modelo proposto não se preocupa em caracterizar diretamente a distribuição de taxas em sobrecarga - as recompensas que representam as taxas são escolhidas em função dos tempos de duração de sobrecargas. A precisão do modelo é relacionada ao casamento da distribuição das taxas em sobrecarga do modelo com a seqüência real, em particular com a cauda da distribuição.

No modelo proposto neste trabalho, três recompensas de taxa, com suas respectivas probabilidades de ocorrência durante as sobrecargas, modelam a distribuição de taxas em sobrecarga para a seqüência real. Quanto maior a diferença entre as taxas em sobrecargas, mais difícil é aproximar sua distribuição usando apenas três recompensas de taxa. Isto é, quanto maior a diferença entre o nível de sobrecarga e a taxa de pico, mais difícil tende ser a descrição desta distribuição.

A Figura 4.28 mostra a seqüência de taxas para 60 minutos do filme *Silence of the Lambs* codificado em MPEG4. As marcas no gráfico indicam os níveis de sobrecarga para uma carga de 33% e 90%. Podemos ver a aproximação da distribuição de taxas em sobrecarga para cada carga nas Figuras 4.29 e 4.30, onde aparecem os histogramas da seqüência real e os valores (em forma de impulsos) para os modelos de referência. Pode-se notar pela distância entre a taxa de pico da seqüência real e a dos modelos que a aproximação para a carga de 33% deverá ser melhor que para a carga de 90%.

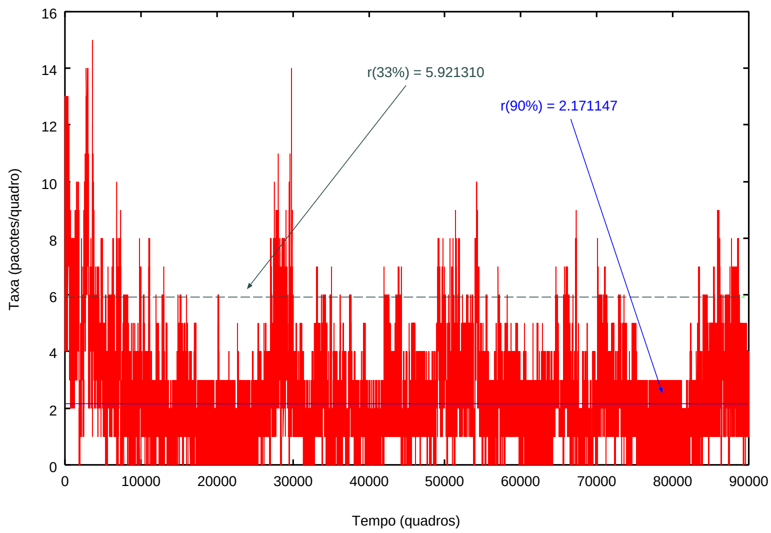


Figura 4.28: seqüência de pacotes para o filme *Silence of the Lambs* em MPEG4

As Figuras 4.31 e 4.32 mostram a aproximação da perda para os dois níveis de sobrecarga, suportando o comentário feito acima.

Uma forma de verificar melhor a influência da distribuição das taxas em sobre-

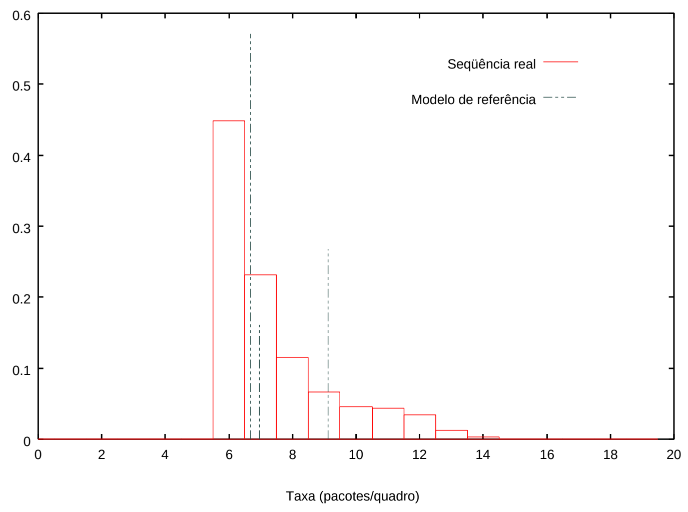


Figura 4.29: Histogramas das taxas em sobrecarga para modelo de referência e seqüência real, a 33%

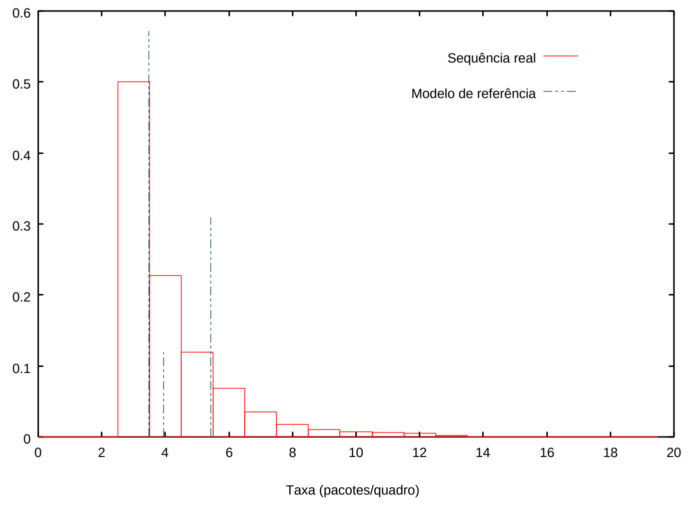


Figura 4.30: Histogramas das taxas em sobrecarga para modelo de referência e seqüência real, a 90%

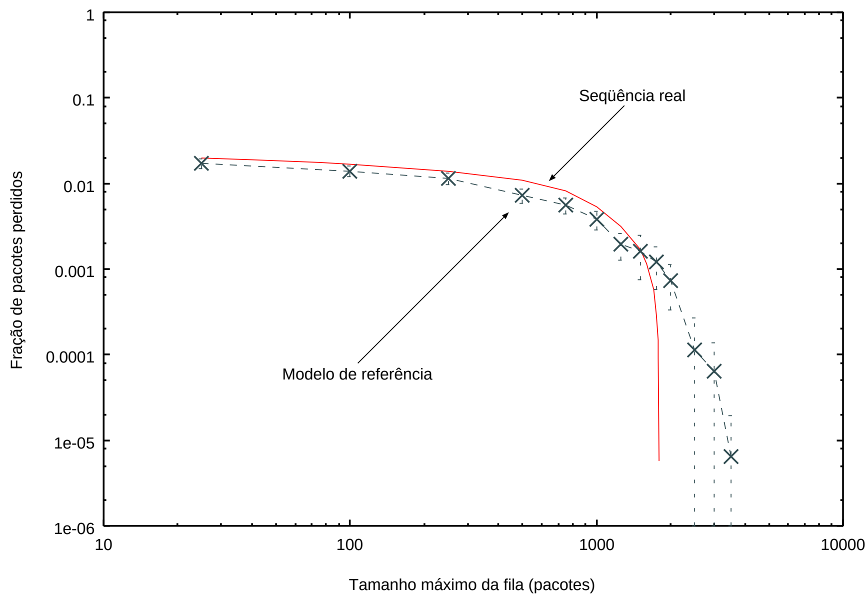


Figura 4.31: Aproximação da fração de perda para carga igual a 33%

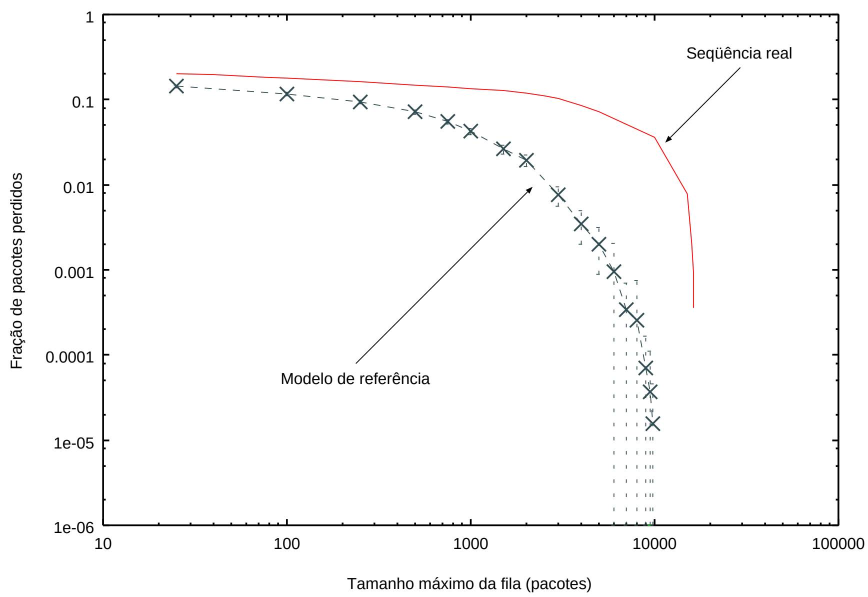


Figura 4.32: Aproximação da fração de perda para carga igual a 90%

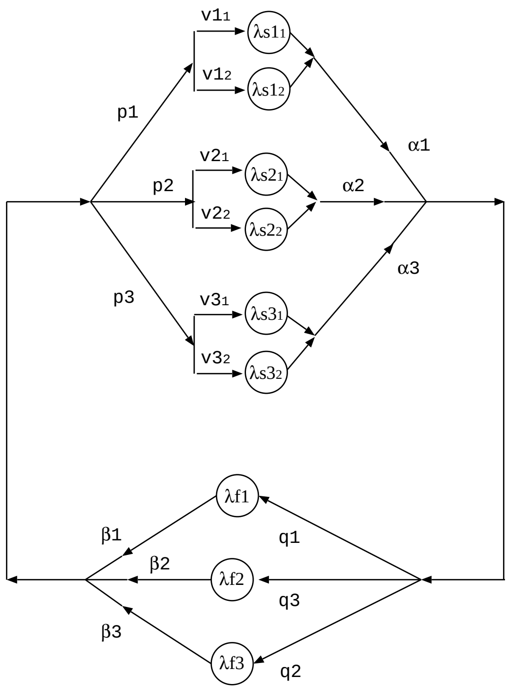
carga, dado uma determinada escolha para os intervalos de duração, é aumentar o número de recompensas associadas a cada intervalo de tempo. Nos exemplos a seguir, cada estado de sobrecarga do modelo referente à carga de 90% foi convertido em dois outros estados. Estes dois estados têm, cada qual, uma recompensa diferente. Estas recompensas, ao serem ponderadas pelas probabilidades de entrada nos seus respectivos estados, equivalem à recompensa original. A Figura 4.33 mostra um diagrama referente à modificação proposta e os valores dos respectivos parâmetros para dois diferentes exemplos. As novas recompensas, assim como as probabilidades de entrada para os respectivos estados, foram obtidas investigando-se o histograma de taxas condicionado a cada intervalo de tempo. No exemplo 1, os histogramas para cada região (intervalo de tempo) foram divididos em duas sub-regiões onde a sub-região contendo as taxas maiores representa, aproximadamente, 10% do total de taxas para a região. No exemplo 2, fez-se o mesmo usando 25% do total de taxas para a região.

A Figura 4.34 mostra as perdas para os dois exemplos citados. Como se pode observar, em ambos exemplos houve uma melhora na aproximação da perda para tamanhos de fila maiores - melhora esta decorrente da melhora na descrição da distribuição de taxas em sobrecarga. Mesmo assim, tal melhora na aproximação é pequena considerando-se os resultados para a seqüência real.

4.6 Construção de um Modelo com Múltiplos Níveis de Sobrecarga

Pelos experimentos deste capítulo e da proposta de um modelo voltado para uma medida de desempenho levando em conta a taxa de serviço da fila, é natural indagarmos se não seria factível um modelo que fosse capaz de capturar estas mesmas características para múltiplos níveis. Mais uma vez, vale lembrar que a proposta da tese é a de uma modelagem que leve em consideração o sistema estudado e uma medida de interesse que caracterize a eficiência de seu funcionamento.

É possível dividir o tráfego em quantos níveis forem especificados, medir a distri-



Ex. 1:	Ex. 2:
$\lambda s1_1$ = 3.215433	$\lambda s1_1$ = 3.000000
$\lambda s1_2$ = 5.496241	$\lambda s1_2$ = 4.558846
$\lambda s2_1$ = 3.652500	$\lambda s2_1$ = 3.418582
$\lambda s2_2$ = 6.450549	$\lambda s2_2$ = 5.630573
$\lambda s3_1$ = 4.873681	$\lambda s3_1$ = 4.426143
$\lambda s3_2$ = 10.46724	$\lambda s3_2$ = 8.661853
$v1_1$ = 0.8861833	$v1_1$ = 0.6952700
$v1_2$ = 0.1138167	$v1_2$ = 0.3047300
$v2_1$ = 0.8978676	$v2_1$ = 0.7650580
$v2_2$ = 0.1021324	$v2_2$ = 0.2349420
$v3_1$ = 0.8990799	$v3_1$ = 0.7610696
$v3_2$ = 0.1009201	$v3_2$ = 0.2389304

$\alpha 1$ = 1/1.525697
$\alpha 2$ = 1/37.12500
$\alpha 3$ = 1/193.2222

p1 = 0.9873433
p2 = 0.00843783
p3 = 0.004218915

$\lambda f1$ = 1.461650
$\lambda f2$ = 0.985885
$\lambda f3$ = 0.762348

$\beta 1$ = 1/4.784158
$\beta 2$ = 1/38.60791
$\beta 3$ = 1/223.8333

q1 = 0.9306223
q2 = 0.06515880
q3 = 0.004218915

Figura 4.33: Modelo de sobrecarga-folga modificado, dois estados por escala de duração de sobrecarga ponderados de acordo com o respectivo histograma de taxa

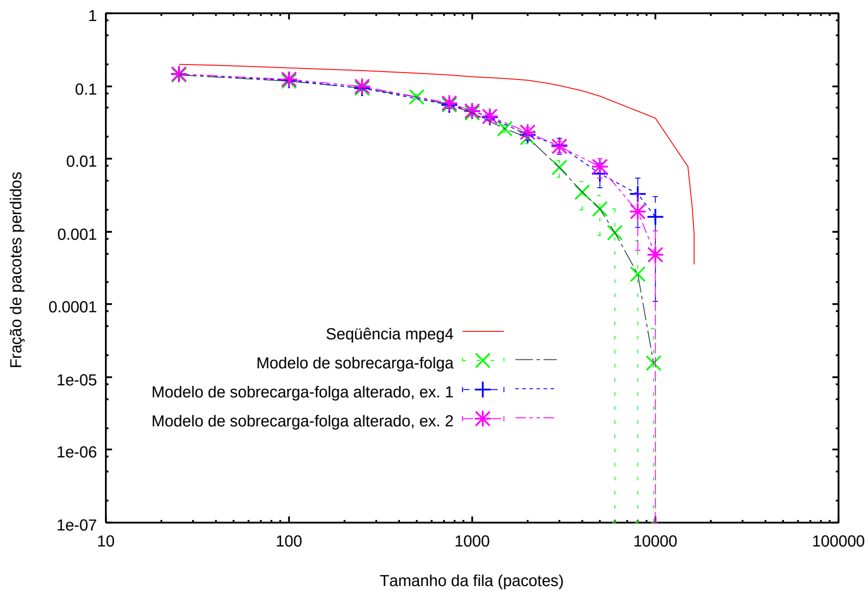


Figura 4.34: Fração de perda para carga de 90%, comparando modelos de sobrecarga-folga usando 2 estados para cada escala de duração de sobrecarga

buição da duração dos períodos em cada nível e modelar estas distribuições usando as hiper-exponenciais descritas na seção 4.2. Podemos também medir as probabilidades de transição entre os níveis, tal qual é feito para um modelo de histograma [41], e construir uma cadeia correspondente.

A Figura 4.36 mostra um exemplo usando dois níveis, 33% e 90%, e portanto três faixas de taxas. A Figura 4.35 mostra as distribuições de duração para cada faixa. Vale reparar que o nível com as menores taxas corresponde à folga para r relativo a 90% de carga e que o nível com as maiores taxas corresponde à sobrecarga para r referente a 33% de utilizaçao.

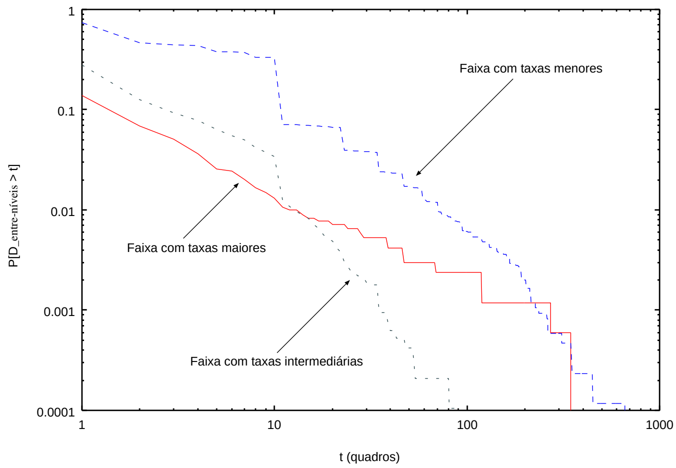


Figura 4.35: Distribuição das durações para cada faixa de taxas

Seja **A** o modelo representado na Figura 4.36 e sejam os modelos **B** e **C** aqueles representados na Figura 4.37 para as cargas de 90% e 33%, respectivamente. Deseja-se que o modelo **A** seja capaz de apresentar, de forma conjunta, as mesmas características que os modelos **B** (carga de 90%) e **C** (carga de 33%).

As Figuras 4.38 e 4.39 fazem uma comparação entre as taxas de perdas previstas pelo modelo **A** com as dos demais modelos e a seqüência real para suas respectivas cargas. Podemos verificar que, no caso de carga igual a 33% e 90%, os resultados não apresentam diferença significativa.

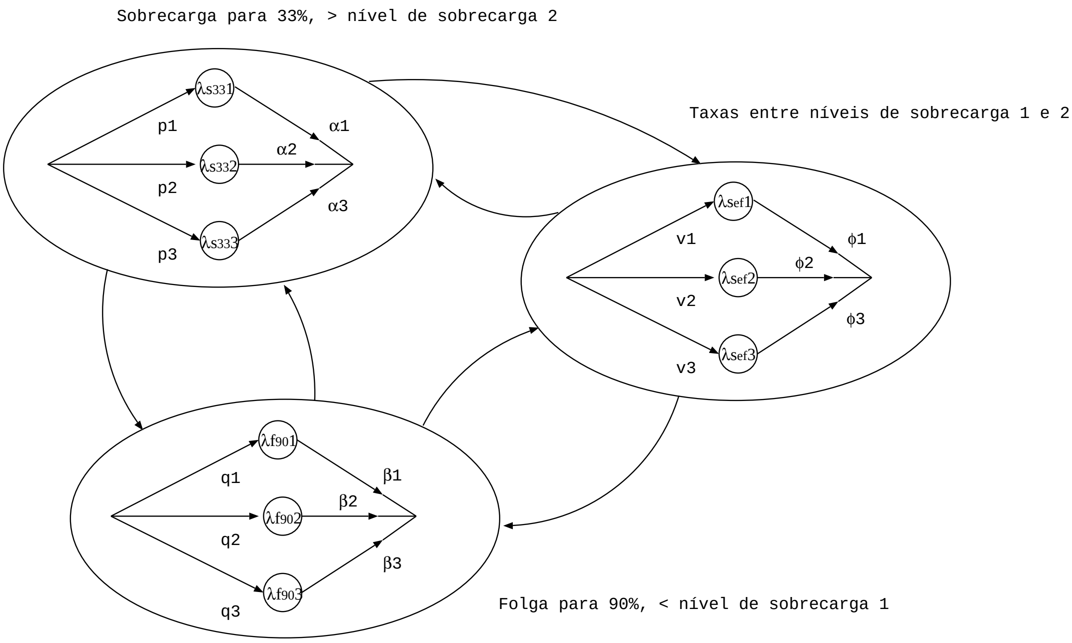
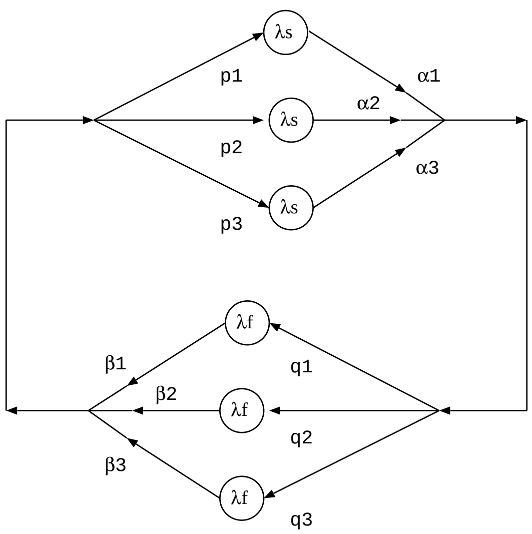


Figura 4.36: Modelo incorporando 2 níveis de sobrecarga



33% :	90% :
λs = 7.374638	λs = 4.137482
α1 = 1/1.198663	α1 = 1/1.525697
α2 = 1/14.60526	α2 = 1/37.12500
α3 = 1/185.2000	α3 = 1/193.2222
p1 = 0.9745411	p1 = 0.9873433
p2 = 0.02249852	p2 = 0.00843783
p3 = 0.002960332	p3 = 0.004218915
λf = 1.737686	λf = 1.226927
β1 = 1/9.493252	β1 = 1/4.784158
β2 = 1/497.2292	β2 = 1/38.60791
β3 = 1/4291.182	β3 = 1/223.8333
q1 = 0.9650681	q1 = 0.9306223
q2 = 0.02841918	q2 = 0.06515880
q3 = 0.006512729	q3 = 0.004218915

Figura 4.37: Parametrização dos modelos para 33% e 90% de carga

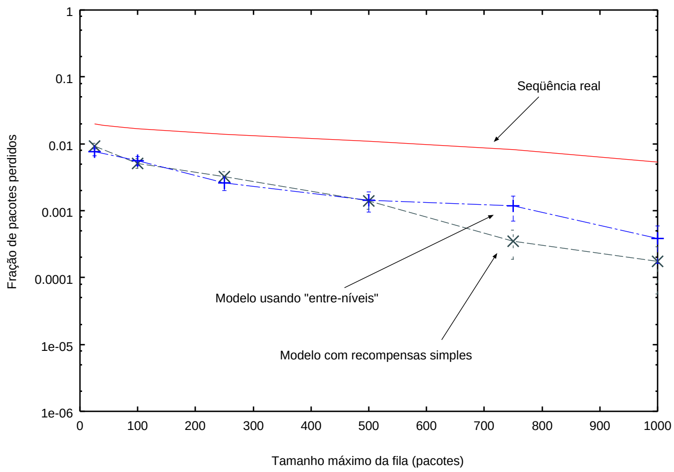


Figura 4.38: Comparação da fração de perdas para modelo usando 2 níveis de sobrecarga aplicado a fila com 33% de carga

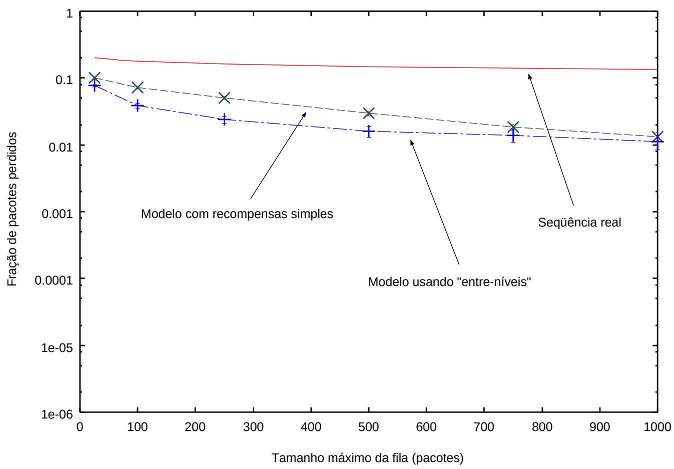


Figura 4.39: Comparação da fração de perdas para modelo usando 2 níveis de sobrecarga aplicado a fila com 90% de carga

4.7 Resumo e Conclusões do Capítulo

Este capítulo descreveu uma proposta de modelagem baseada em um modelo markoviano com recompensas. O modelo foi parametrizado usando os seguintes descritores: a distribuição da duração das sobrecargas, a distribuição da duração das folgas e o histograma de taxas da seqüência real. O modelo possui dois subconjuntos de estados: um onde é gerado tráfego acima da taxa de serviço da fila (estados de sobrecarga) e o outro onde é gerado tráfego abaixo da taxa de serviço da fila (estados de folga). Os parâmetros dos estados de cada um desses subconjuntos são obtidos através da aproximação da distribuição da duração das sobrecargas e das folgas por distribuições hiper-exponenciais. Foi proposto um novo algoritmo para a obtenção dos parâmetros da distribuição hiper-exponencial a partir da distribuição real das durações.

Foram apresentadas três formas para definir as recompensas associadas aos estados. A que obteve melhor desempenho e menor custo computacional foi aquela em que uma recompensa diferente é associada a cada estado do modelo (modelo com múltiplas recompensas).

Pelos testes efetuados, foi possível verificar a robustez do algoritmo usado para parametrizar as hiper-exponenciais e, conseqüentemente, o modelo. Os resultados obtidos para a fração de perda são muito próximos daqueles obtidos para a seqüência real usada como exemplo para uma determinada carga, o filme *Silence of the Lambs* codificado em MPEG4.

Este capítulo também fez algumas considerações a respeito da sensibilidade do modelo proposto frente aos parâmetros usados em sua construção. Foram mostrados resultados para diferentes possibilidades de parametrização tanto na distribuição das durações quanto na atribuição de recompensas de taxa, seja para folga ou sobrecarga. Além disso, é apresentado um estudo sobre a possibilidade de um modelo incorporar mais de um nível seguindo a mesma filosofia de modelagem adotada. O resultado deste estudo mostra que o uso de mais de um nível não traz nenhum ganho e aumenta muito a complexidade do modelo.

Fica claro que, dentre os possíveis parâmetros ligados à proposta de modelagem deste trabalho, a distribuição de taxas em sobrecarga e a cauda da distribuição de duração das sobrecargas são aqueles que mais influenciam a fração de perda resultante.

Capítulo 5

Resultados e Análises Comparativas

Este capítulo mostra os resultados da aplicação do modelo proposto, que chamaremos de modelo de *sobrecarga-folga*, a outras seqüências reais além da mostrada como exemplo no capítulo 4. Tais resultados são avaliados comparando-se a fração de perda e os descritores deste modelo com a fração de perda e os descritores extraídos para as seqüências reais e gerados por outros modelos mais conhecidos na literatura, cada qual com suas características particulares. Todos os intervalos de confiança mostrados foram calculados para 95% de probabilidade.

5.1 Tráfegos de Fontes Individuais e de Agregados

Quando se trata de caracterizar e modelar seqüências reais de tráfego, pode-se pensar em duas categorias: tráfegos de fontes individuais e tráfegos de fontes agregadas. Tais categorias têm suas peculiaridades quanto a ambientes de aplicação e requisitos de qualidade. Por exemplo, tráfegos agregados são mais apropriados em estudos de dimensionamento para roteadores de grande porte de uma rede WAN. Já tráfegos de fontes individuais são mais apropriados para estudos de dimensionamento e de qualidade de serviço das aplicações e dos protocolos fim-a-fim. Quanto aos requisitos de qualidade, pode-se dizer que o grau de tolerância pode ser diferente em ambos os casos: 1% de perda num agregado de tráfego *http* pode ter efeito desprezível, mas 1% de perda numa conversação usando voz sobre IP, dependendo do codificador usado, pode tornar a qualidade da aplicação muito ruim [67, 68, 69].

Outra diferença existente entre estas duas categorias de tráfego vem da dificuldade para se capturar tráfegos agregados na *Internet*. Leituras SNMPs feitas em interfaces de saída ou entrada de roteadores requerem intervalos de tempo de amostragem da ordem de dezenas de segundos para se garantir respostas confiáveis destes roteadores, mesmo aqueles de maior capacidade. Tais intervalos costumam ser demasiadamente longos para capturar variações de tráfego relevantes para os tamanhos de fila de tais equipamentos. Por conta disso, a maneira mais confiável para obter amostras relevantes é interceptando fisicamente o tráfego, o que pode ter um custo razoavelmente alto dependendo da infra-estrutura local do equipamento.

Assim sendo, a proposta para modelos de fonte deste trabalho será avaliada para seqüências que representem fontes individuais e seqüências que representem tráfego agregado. No caso de fontes individuais, seqüências de vídeo MPEG4 serão usadas para avaliação do modelo proposto, como feito nos capítulos anteriores. Tais seqüências foram escolhidas por conta da popularidade do formato MPEG4 e por sua conhecida característica de periodicidade no tamanho dos quadros enviados, o que pode dificultar a tarefa de modelagem (tal característica será explicada a seguir). No caso de fontes agregadas, seqüências de entrada e saída coletadas em um dos enlaces da rede acadêmica Abilene, nos EUA, foram usadas como exemplo.

5.1.1 Características da codificação MPEG4

O formato MPEG4 de codificação de vídeo é popularmente usado para compressão de áudio e imagem. Ele é capaz de fornecer altas taxas de compressão de acordo com a qualidade exigida na recuperação do processo. Como os dados resultantes da codificação de imagem são muito maiores que os dados resultantes da codificação de áudio, a estratégia usada para codificar as imagens é que determina o comportamento do tráfego originado pela transmissão de um vídeo em MPEG4.

O processo de codificação tem basicamente por objetivo a compressão dos dados. Para tal, são utilizados algoritmos que aproveitem a redundância espacial das imagens assim como sua dependência temporal. Entende-se por esta dependência o fato dos quadros de vídeo constantemente se repetirem em seqüência, com poucas variações entre as imagens. Este comportamento corresponde a movimentos lentos de cena, com poucas mudanças.

Sendo assim, o formato MPEG4 incorpora um algoritmo diferencial que se vale de predições para traduzir as mudanças de cena a cada quadro. São gerados três tipos de quadro: quadros com codificação absoluta e chamados *Intra* (quadros **I**), quadros decorrentes de predição e chamados *Predictive* (quadros **P**), e quadros oriundos da interpolação destes (**I** e **P**) e chamados *Bidirectional* (quadros **B**).

A partir de uma imagem absolutamente codificada segundo suas redundâncias espaciais, como num formato JPEG (quadros **I**), os quadros futuros passam a ser constituídos pela diferença entre o quadro real e aquele predito pelo algoritmo (quadros **P**). Além disto, também é usado um recurso de interpolação para predizer quadros situados entre os quadros codificados com as técnicas anteriores (quadros **I** e **P**) onde, da mesma forma, apenas a diferença entre esta predição e o quadro real é transmitida (quadros **B**) [70]. Desta forma, uma seqüência de quadros em MPEG4 tem o seguinte padrão:

I B B B P B B B P B B B I B B B P B ...

Este padrão corresponde à codificação de cada quadro. No entanto, na transmissão destes, os quadros **P** vêm na frente dos quadros **B** para facilitar a decodificação:

I P B B B P B B B P B B B I P B B B ...

O número de quadros entre os quadros **I** é fixo e define o tamanho do *GOP*, "Group of Pictures". Geralmente seu valor costuma ser igual a 12 para filmes codificados numa taxa de 25 quadros/segundo.

Tais técnicas representam uma redução significativa no número de bits enviados a cada quadro. Do lado do receptor, este usará as mesmas técnicas de previsão de quadros e se valerá da informação transmitida para recompor a imagem. Porém, como os quadros **I**, em média, são significativamente maiores que os quadros **P**, principalmente se não ocorre mudança brusca de cena, e estes são maiores que os quadros **B**, é fácil observar um comportamento periódico numa seqüência MPEG4. Tal comportamento pode ser verificado através do formato "dente-de-serra´´ de sua função de autocorrelação, como mostrado na Figura 5.1 para o filme *Silence of the Lambs*. A transformada de Fourier desta função resulta no espectro de potência das freqüências associadas à seqüência.

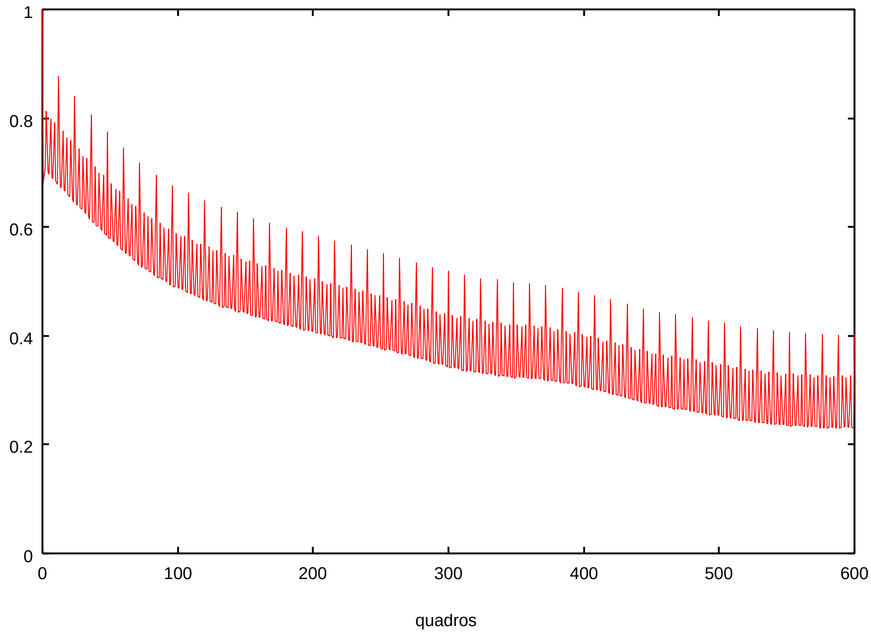


Figura 5.1: Periodicidade de uma seqüência MPEG4 verificada pela sua função de autocorrelação

Certamente que este comportamento periódico torna a tarefa de caracterização mais complexa, além da característica de dependência de longa duração que este tipo de tráfego possui [61, 4, 71]. Além disso, a sensibilidade à perda das aplicações para esta codificação de vídeo é razoavelmente grande, principalmente quando ela acontece num pacote pertencente a um quadro **I**. Maiores detalhes e referências sobre MPEG4 podem ser encontrados em [70].

5.1.2 Características de tráfego nas grandes redes

O perfil de tráfego nas grandes redes possui grande variabilidade ao longo do dia. Sujeito ao comportamento humano dos usuários das redes, o volume de tráfego varia conforme horários comerciais e dias da semana. Dependendo da forma de conexão, o volume de tráfego pode ser maior durante a madrugada, hora em que as tarifas telefônicas são mais baixas, e menor no restante do dia. Em geral, isso é mais comum quando a maioria dos usuários tem acesso doméstico, já que estes costumam ser realizados fora do horário comercial. Já em casos onde o uso é tipicamente realizado no horário de trabalho, como em provedores com um vasto número de clientes comerciais ou em redes acadêmicas, este perfil se inverte.

O dimensionamento de recursos em provedores, em geral, se baseia nos horários de pico, aqueles nos quais o tráfego apresenta maior volume. Porém, existe ainda a necessidade de se distinguir o que é chamado de "tráfego legítimo", ou seja, aquele que **não** está relacionado com ataques ou mau uso da rede, dos "tráfegos ilegítimos", que oneram o uso da banda e devem, até certo ponto, ser desconsiderados no dimensionamento da rede. Esta distinção entre o que é legítimo e ilegítimo por vezes é muito difícil de se alcançar. Casos mais simples, como ataques de DOS ("denial of service"), conseguem ser verificados graças ao súbito aumento no número de pacotes pequenos destinados a uma determinada sub-rede. No entanto, o uso massivo de aplicativos P2P ("peer to peer") para transferências de vídeos e músicas, em geral codificados em MPEG4 e MP3, respectivamente, é difícil de ser verificado. Tal uso pode mudar bastante as características do tráfego além de representar um custo maior para aqueles que financiam o provimento do acesso. Isto é mais crítico quando este financiamento objetiva outros fins, como é o caso das redes acadêmicas, subsidiadas para promover ensino e pesquisa.

Paralelo ao debate sobre a legitimidade do que passa nas redes, outros pontos costumam ser foco de grande discussão no que concerne ao perfil deste tipo de tráfego. O principal deles diz respeito sobre a existência ou não de características auto-similares. Em [2], os autores afirmam que o tráfego oriundo de um agregado de redes locais é auto-similar. Já em [72], os autores dizem que é possível usar modelos markovianos para modelar o mesmo tipo de tráfego. Esta controvérsia é esclarecida em [73], onde os autores mostram que, dependendo da utilização do canal e do intervalo de amostragem, ambos comportamentos podem ser encontrados.

Basicamente, se um canal está saturado, a taxa de envio de dados tem pouca variação uma vez que esta se encontra muito próxima de seu limitante superior. Em decorrência disto, tráfegos nesta condição tendem a não apresentar auto-similaridade.

Da mesma forma, se um canal tem utilização muito baixa, também haverá pouca variabilidade e, conseqüentemente, tráfegos nesta condição também tendem a não apresentar auto-similaridade. Segundo a análise em [73], quando a utilização do canal é de 50%, o tráfego que passa pelo canal tem melhores condições para apresentar auto-similaridade. Além disso, intervalos de amostragem muito grandes podem "suavizar" demais as variações do tráfego, interferindo na avaliação do grau de auto-similaridade. Maiores detalhes podem ser encontrados em [73] onde os autores se valem de *wavelets* para formalizar a análise.

5.2 Seqüências Usadas nos Testes

Para evitar possíveis resultados tendenciosos decorrentes de características particulares a uma determinada seqüência, três exemplos de tráfego MPEG4 e dois de tráfego WAN foram usados nos testes.

5.2.1 Seqüências MPEG4

As seqüências MPEG4 escolhidas representam 3 filmes de ação: *Jurassic Park*, *Star Trek - First Contact* e o filme já usado como exemplo nos capítulos anteriores, *Silence of the Lambs*. Estas seqüências podem ser obtidas em [58] e um estudo estatístico destas e de outras seqüências, tanto na codificação MPEG4 quanto na H.263, pode ser visto em [74].

A razão da escolha por filmes de ação é que estes costumam possuir um número maior de mudanças bruscas de cena, o que resulta em quadros com maior variabilidade estatística de tamanho. Todas as seqüências MPEG4 usadas nos exemplos foram codificadas a 25 quadros/segundo e o número de amostras de cada uma é igual a 90000. As Figuras 5.2, 5.3 e 5.4 mostram a taxa de pacotes por quadro ao longo do tempo para as seqüências escolhidas.

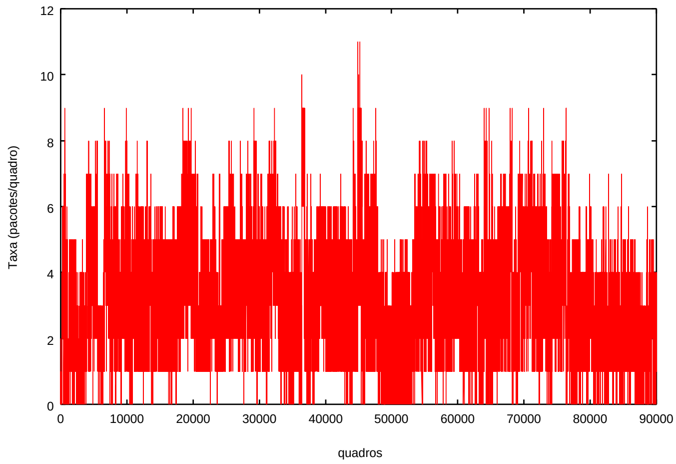


Figura 5.2: Taxa de pacotes/quadro para o filme *Jurassic Park* codificado em MPEG4

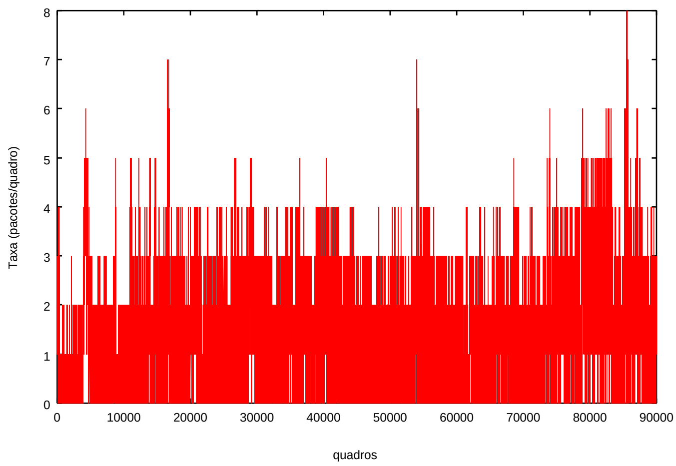


Figura 5.3: Taxa de pacotes/quadro para o filme *Star Trek - First Contact* codificado em MPEG4

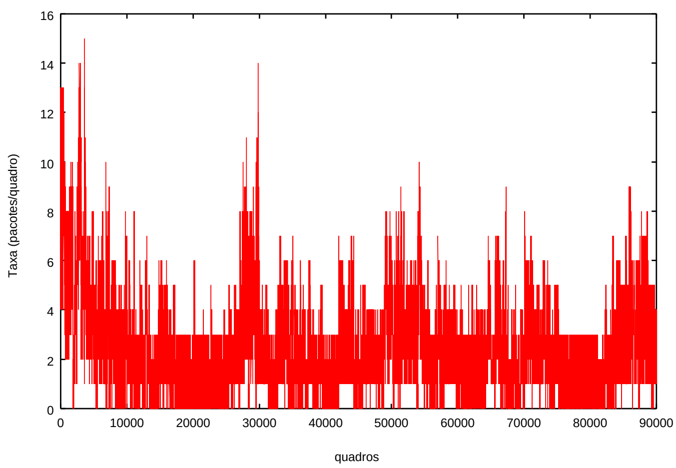


Figura 5.4: Taxa de pacotes/quadro para o filme *Silence of the Lambs* codificado em MPEG4

5.2.2 Seqüências de tráfego agregado na Internet

Para representar tráfegos em redes WAN, foram usadas 2 seqüências obtidas da rede Abilene [75]. Estas seqüências representam, cada uma, dez minutos de amostragem do tráfego de entrada e saída coletados em uma das interfaces POS OC-48 do roteador que representa o nó da rede Abilene em Indianapolis. Esta medição foi realizada em 14 de agosto de 2002, das 10:30hs às 10:40hs e, na época, o enlace amostrado ligava Indianapolis a Cleveland sem haver saturação em nenhum dos sentidos. Os tráfegos de entrada e saída estão separados nos arquivos IPLS-CLEV-20020814-103000-0 e IPLS-CLEV-20020814-103000-1, respectivamente, e podem ser obtidos em [76].

O aparato usado para coletar tais seqüências envolvia "separadores" ópticos e circuitos eletrônicos capazes de registrar o instante de tempo de recepção e transmissão de cada pacote. O resultado destes registros foi processado para fornecer novas seqüências contendo, cada qual, o número de pacotes transmitidos a cada 100 milisegundos. A conversão do formato original para um contendo o número de pacotes a um intervalo fixo de tempo é necessário para se extrair os descritores de tráfego. No caso, o valor escolhido para este intervalo é considerado em [77] como sendo suficiente para uma análise detalhada do tráfego num *backbone*. Mais informações sobre o processo de aquisição destas amostras, assim como o formato original resultante, podem ser obtidas em [76].

As Figuras 5.5 e 5.6 mostram, respectivamente, a taxa de pacotes dos tráfegos de entrada e saída relatados. Ambas seqüências possuem 6000 amostras, o equivalente a 10 minutos de tráfego com granularidade de 100 milisegundos, como dito anteriormente.

5.3 Modelos Usados nos Testes

Para avaliar a eficiência do modelo proposto, que objetiva obter a fração de perda para uma dada taxa de serviço na fila, outros modelos aplicados às mesmas condições foram usados como referência. No caso, o resultado desejado é a predição

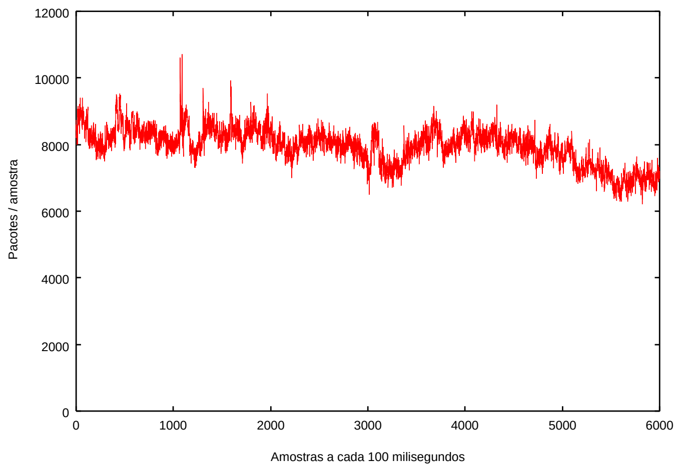


Figura 5.5: Número de pacotes a cada 100msec para tráfego de entrada na interface POS OC-48 do nó de Indianapolis da Abilene ligando Cleveland

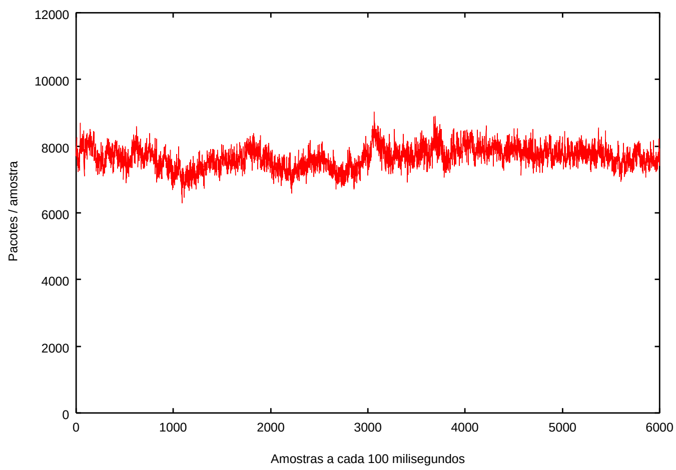


Figura 5.6: Número de pacotes a cada 100msec para tráfego de saída na interface POS OC-48 do nó de Indianapolis da Abilene ligando Cleveland

da fração de perda e, portanto, será esta a medida de desempenho extraída para cada modelo.

Os modelos escolhidos possuem características particulares quanto ao casamento dos descritores de tráfego *distribuição de taxas* e *função de autocorrelação*. Cada um deles faz parte de uma das categorias apresentadas na seção 2.3. São eles:

- Modelo markoviano de histograma linear
- Modelo F-ARIMA
- Modelo pseudo auto-similar modificado

O modelo de histograma linear é o de construção mais intuitiva. De acordo com o número de estados representando os níveis de taxa da fonte, consegue capturar a distribuição da taxa - objetivo para o qual ele foi concebido. Todavia, da maneira como é parametrizado, não é possível capturar dependências de longa duração. Para os exemplos deste capítulo, este modelo contém 8 estados e as transições da cadeia representam a variação da taxa de pacotes a cada quadro. Este modelo representa a captura da distribuição de taxa e da dependência de curtíssima duração.

O modelo F-ARIMA não parametriza detalhadamente a distribuição da taxa mas casa apenas a média e a variância. Porém, este modelo captura a dependência de longa duração através do casamento do parâmetro de Hurst, o que pode ser visto pela comparação da função de autocorrelação. Para os exemplos deste capítulo, este modelo representa a captura da dependência de longa duração.

O modelo pseudo auto-similar também objetiva unicamente o casamento da função de autocorrelação para o longo termo, restringindo-se apenas à média no que diz respeito à distribuição de taxa. O grande atrativo deste modelo é que ele captura uma dependência de duração longa usando cadeia de Markov. Este modelo é de interesse especial para o trabalho de tese aqui apresentado pois trata-se de um caso particular do modelo de sobrecarga-folga, como será esclarecido a seguir.

O apêndice B traz as implementações destes modelos, assim como do modelo de sobrecarga-folga, na ferramenta TANGRAM II [49].

5.4 Modelo pseudo auto-similar como caso particular do modelo de sobrecarga-folga

Como descrito na seção 2.3.2, o modelo pseudo auto-similar foi introduzido em [46] visando modelar o tráfego de células ATM passando por um enlace com capacidade conhecida e sendo servido por uma fila. Trata-se de uma cadeia de Markov de tempo discreto cujo tempo entre eventos corresponde ao *time slot* do enlace, ou seja, o espaço de tempo no qual uma célula ATM pode ou não ser enviada para o enlace e, por conseguinte, chegar ou não na fila. A fila, por sua vez, possui um tempo de serviço constante medido em número de *time slots*. Só existem duas possíveis taxas na fonte: 1 célula por *slot* ou 0 células por *slot*. A taxa média da fonte é, portanto, a probabilidade de haver (ou chegar) uma célula no *slot* k, onde k é inteiro.

No modelo pseudo auto-similar proposto em [46], tem-se um estado representando a chegada de 1 célula por *slot* e quatro estados representando a ausência de células por *slot*. Segundo a concepção dos autores, após cada chegada de célula, existe uma probabilidade de nova chegada ou de *slot* vazio, sendo que o número de *slots* vazios entre cada chegada de célula possui distribuição de cauda longa. Assim sendo, as probabilidades de transição seguem o padrão mostrado na Figura 5.7.

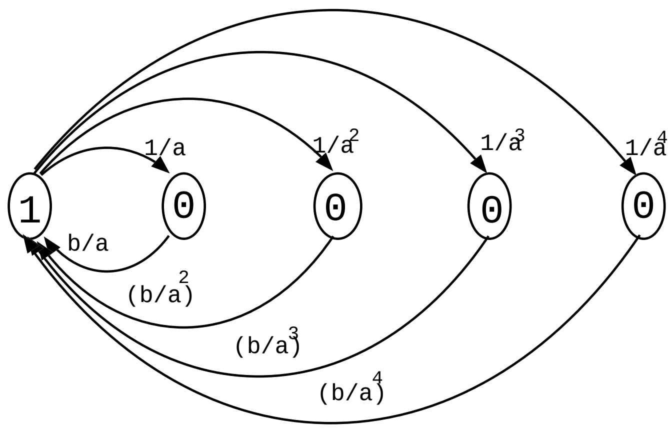


Figura 5.7: Modelo pseudo auto-similar

Como só existem duas taxas possíveis neste modelo, e como a taxa de serviço para uma fila onde possa ocorrer perda obrigatoriamente tem valor intermediário a estas, fica claro que trata-se de um modelo de sobrecarga-folga particular. No caso, a duração das sobrecargas, representadas pelo estado de taxa 1, tem distribuição exponencial (ou geométrica) com taxa média igual a

$$1 - \sum_{i=1}^{4} \frac{1}{a^i}$$

e a duração das folgas, representada pelos estados com taxa 0, tem distribuição hiper-exponencial com taxa média igual a

$$\sum_{i=1}^{4} \frac{1}{a^i} (\frac{b}{a})^i .$$

Certamente que a escolha deste formato tem razões matemáticas objetivando um número mínimo de parâmetros e um comportamento que aproxime esta cadeia de um modelo auto-similar quando o número de estados tende a infinito [3].

A analogia com o modelo de sobrecarga-folga aqui proposto facilita o trabalho de parametrização do modelo pseudo auto-similar. Segundo o que é proposto em [47], a escolha do parâmetro b é função da média e a escolha do parâmetro a depende do parâmetro de Hurst (H). O artigo propõe um procedimento iterativo para a obtenção de a até que H seja alcançado satisfatoriamente. No entanto, é trivial perceber que a poderia ser facilmente **calculado** usando-se a expressão analítica para a duração média de sobrecarga do modelo:

$$\overline{Ds} = \frac{1}{\sum_{i=1}^{4} \frac{1}{a^i}} .$$

Para os exemplos aqui mostrados, o modelo pseudo auto-similar tem sua semântica alterada de forma a não considerar somente casos onde as únicas possíveis taxas são 1 e 0. Para que assim o fosse, seria necessário definir uma capacidade de enlace para se obter um valor de *time slot*. No entanto, a proposta de análise aqui apresentada independe desta grandeza. Assim sendo, as duas taxas associadas aos estados do modelo serão a taxa média em sobrecarga (λs) e a taxa média em folga (λf). Ou seja, o modelo pseudo auto-similar será explicitamente alterado para ser um modelo de sobrecarga-folga nos moldes propostos, porém mantendo suas distribuições características para durações de sobrecarga e folga. A Figura 5.8 mostra o modelo pseudo auto-similar modificado.

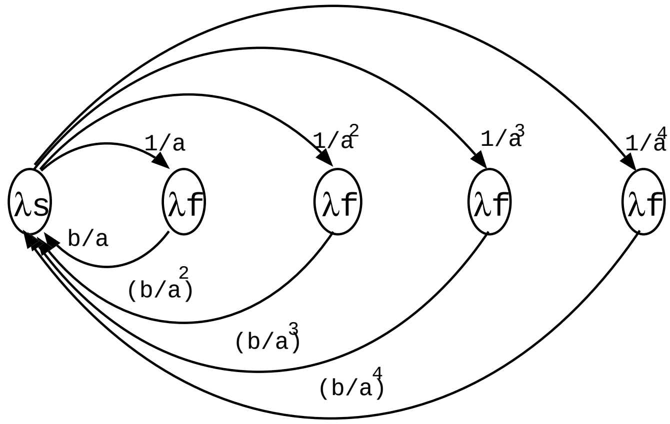


Figura 5.8: Modelo pseudo auto-similar modificado

5.5 Análise dos Modelos para Seqüências MPEG4

Esta seção traz os resultados das perdas para as seqüências MPEG4 escolhidas e os respectivos modelos. A análise destas perdas, em cuja aproximação com as da seqüência real determina a eficiência do modelo, é feita frente aos descritores mais tradicionais, que são o histograma de taxa e a função de autocorrelação, e frente aos descritores aqui estudados e usados na parametrização do modelo de sobrecarga-folga proposto: a distribuição das durações de sobrecarga e a distribuição das durações de folga. As distribuições de taxas em sobrecarga e em folga foram omitidas mas podem ser inferidas nos gráficos contendo os histogramas de taxa.

Os modelos de sobrecarga-folga destinados para cada seqüência foram construídos segundo a metodologia descrita nas seções 4.2 e 4.3.2 para a captura das distribuições hiper-exponenciais e associação de múltiplas recompensas de taxa aos estados. Como já mencionado, a construção destes modelos depende dos níveis de sobrecarga usados nos testes. Estes níveis foram escolhidos de forma a se ter cargas de 50% e 90%. O motivo destes valores é representar o que pode ser considerado como carga média e carga alta.

Os resultados das perdas para as seqüências MPEG4 e seus respectivos modelos são extraídas para cada exemplo e mostradas nos gráficos. Primeiramente, são apresentadas as medidas de perda para cada nível de sobrecarga em função do tamanho da fila. Em seguida, é mostrado um gráfico com os histogramas de taxa e outro com as funções de autocorrelação. No caso destes dois últimos, são mostradas as curvas para cada modelo de sobrecarga-folga referente a cada nível de sobrecarga. Por fim, as distribuições das durações de sobrecarga e as distribuições das durações de folga são mostradas para cada nível de sobrecarga.

As medidas de perda foram extraídas a partir de simulação. Para os modelos markovianos com recompensa de taxa, existem métodos analíticos que resolvem a distribuição da ocupação da fila em estado estacionário, distribuição esta que é usada para extrair a medida de perda. Todavia, este método ainda não estava implementado no **TANGRAM** na época dos experimentos. No caso dos modelos de histograma, pseudo auto-similar modificado e sobrecarga-folga, foram realizadas 100 rodadas de 90000 quadros cada. No caso do modelo F-ARIMA, 8 seqüências de 90000 quadros foram geradas e as medidas de perda foram extraídas a partir da simulação delas. Como as médias de pacotes por quadro das seqüências MPEG4 são próximas de zero, a geração das seqüências para o modelo F-ARIMA apresentou algumas amostras com valores negativos que foram truncados para zero. No caso das seqüências reais, simulou-se a mesma em apenas uma rodada, com 90000 quadros.

Os descritores foram calculados analiticamente para os modelos de histograma, pseudo auto-similar modificado e sobrecarga-folga. As expressões analíticas para os descritores dos modelos markovianos encontram-se na seção 2.4. Para obtenção dos histogramas de taxa destes modelos, calculou-se as probabilidades estacionárias dos estados e associou-se a elas as respectivas recompensas. Para o modelo F-ARIMA, os descritores foram extraídos da mesma forma que para a seqüência real, usando-se as mesmas 8 seqüências de 90000 quadros usadas para extrair a fração de perda.

As tolerâncias de taxa (ϵ) e tolerâncias de média (ψ) usadas na captura das distribuições de duração para a construção dos modelos de sobrecarga-folga seguem indicadas:

ϵ/ψ	50%		90%	
	Sobrecarga	Folga	Sobrecarga	Folga
Jurassic Park	0.003/0.53	0.001/0.03	0.01/0.06	0.01/0.1
Star Trek - First Contact	0.005/0.05	0.0007/0.003	0.003/0.03	0.005/0.005
Silence of the Lambs	0.002/0.03	0.0005/0.16	0.015/0.055	0.026/0.65

Os parâmetros dos modelos usados em cada exemplo podem ser verificados no apêndice C.

5.5.1 Seqüência do filme *Jurassic Park*

Trata-se de um filme de ação e ficção com taxa média igual a 2.554 *pacotes/quadro* e variância igual a 2.434 $(pacotes/quadro)^2$, codificado a 25 *quadros/segundo*. O parâmetro de Hurst, usado pelo modelo F-ARIMA e estimado pelo mesmo método adotado na seção 3.4, é igual a 0.861. A duração média em sobrecarga, usada para construir o modelo pseudo auto-similar modificado, é igual a 1.335 quadros para carga de 50% e 2.894 para carga de 90%. As Figuras 5.9 e 5.10 trazem as frações de perda dado o tamanho da fila para cada carga, com este tamanho variando em escala logarítmica.

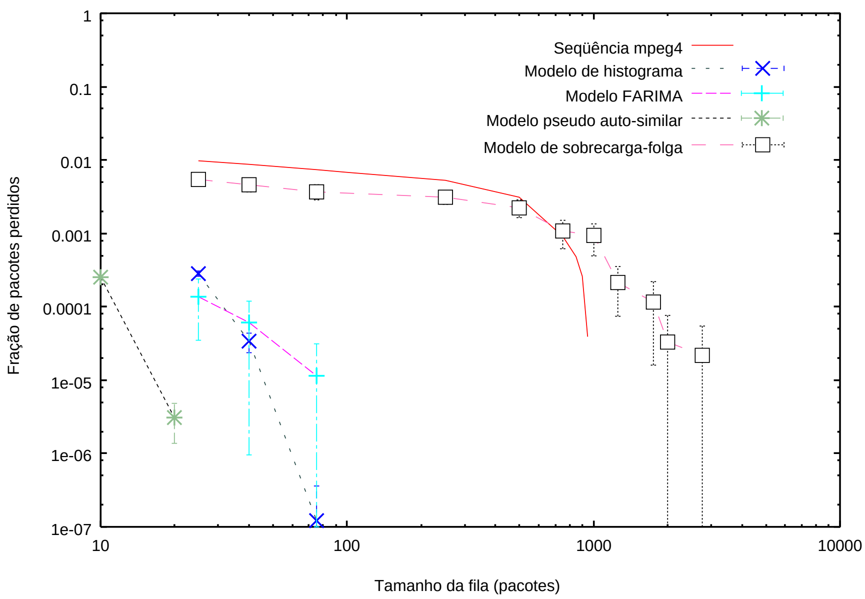


Figura 5.9: Fração de pacotes perdidos, fila em escala logarítmica, filme *Jurassic Park* em MPEG4, r a 50% de carga

Pode-se notar que o modelo de sobrecarga-folga foi superior aos demais para a carga de 50%. No entanto, o mesmo não ocorreu para a carga de 90%. Como mencionado na seção 4.5, a diminuição do nível de sobrecarga (aumento da carga) prejudica a caracterização das taxas em sobrecarga para este modelo, o que influencia diretamente a fração de perdas. As Figuras 5.11 e 5.12 mostram o histograma de taxas dos modelos e seqüência real, onde são ilustrados os níveis de sobrecarga

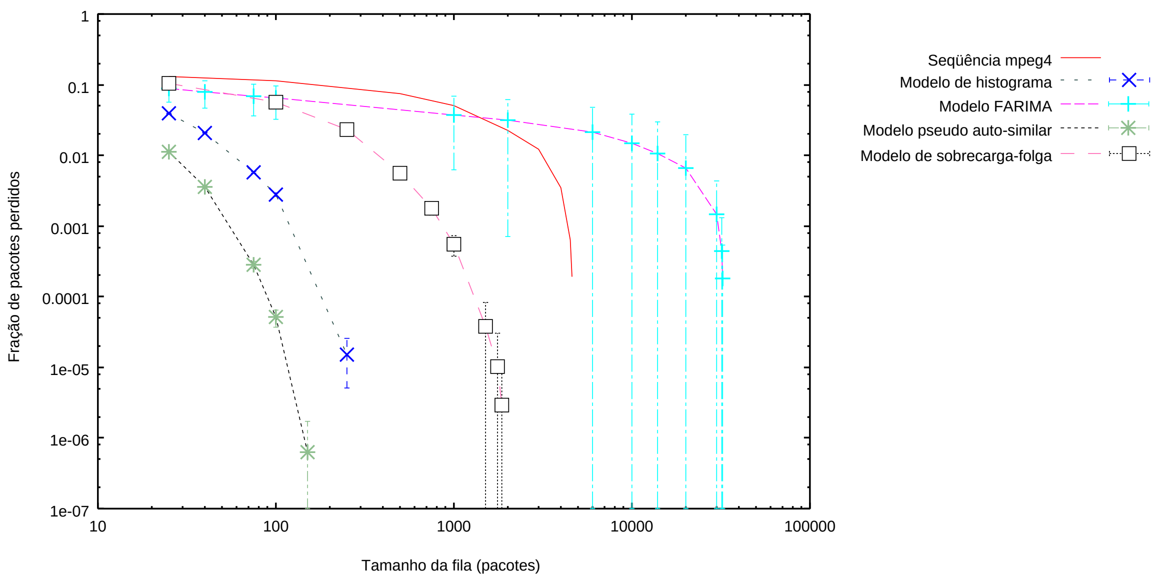


Figura 5.10: Fração de pacotes perdidos, fila em escala logarítmica, filme *Jurassic Park* em MPEG4, r a 90% de carga

usados nos exemplos para cargas de 50 e 90%. Como pode ser verificado, a maior taxa em sobrecarga para o modelo de sobrecarga-folga relativo à carga de 90% fica bem abaixo da taxa de pico da seqüência real.

A Figura 5.13 mostra a função de autocorrelação para os modelos e a seqüência real. Como já era esperado, o modelo F-ARIMA foi quem melhor aproximou este descritor, enquanto que o modelo de histograma teve a pior aproximação. Mas é interessante notar que o modelo de sobrecarga-folga para 50% de carga aproximou melhor esta função do que o modelo de sobrecarga-folga para 90% de carga, o que também reflete o comportamento das perdas para estes modelos.

As Figuras de 5.14 a 5.17 mostram as distribuições de duração de sobrecarga para cargas de 50 e 90%. A duração aparece em escala linear e logarítmica. Pode-se notar que as distribuições hiper-exponenciais dos modelos de sobrecarga-folga copiam de forma bastante aproximada a distribuição real. No entanto, para a carga de 90%, a hiper-exponencial escolhida não captura corretamente a cauda da distribuição, o que contribui para sua ineficiência. Interessante notar que, apesar do modelo F-ARIMA não apresentar uma boa aproximação para esta distribuição, para a carga de 90% e tamanho máximo de fila até 3000 pacotes a fração de perda deste modelo foi a que melhor caracterizou a seqüência. A seção 5.7 faz algumas considerações

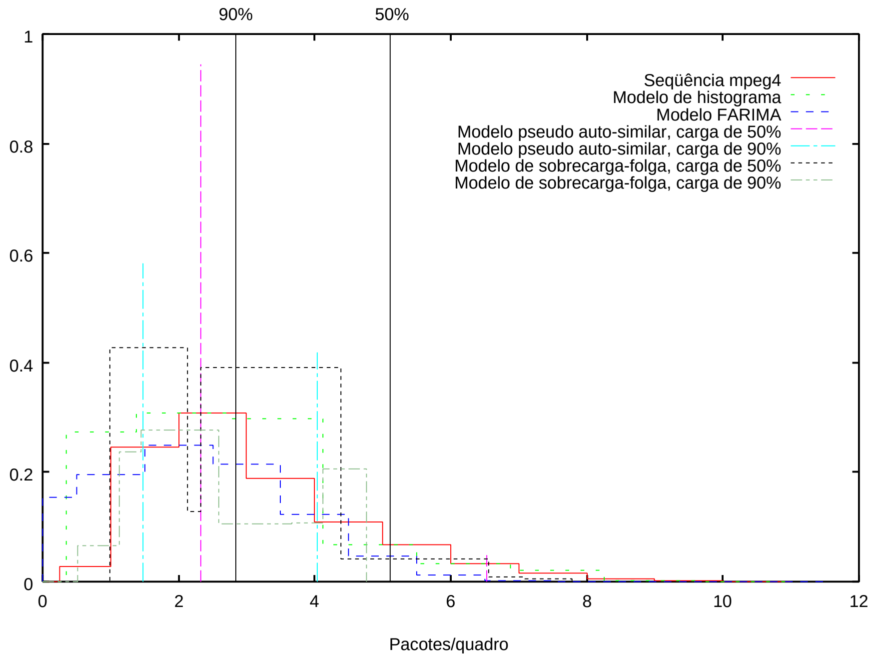


Figura 5.11: Histogramas de taxa, filme *Jurassic Park* em MPEG4

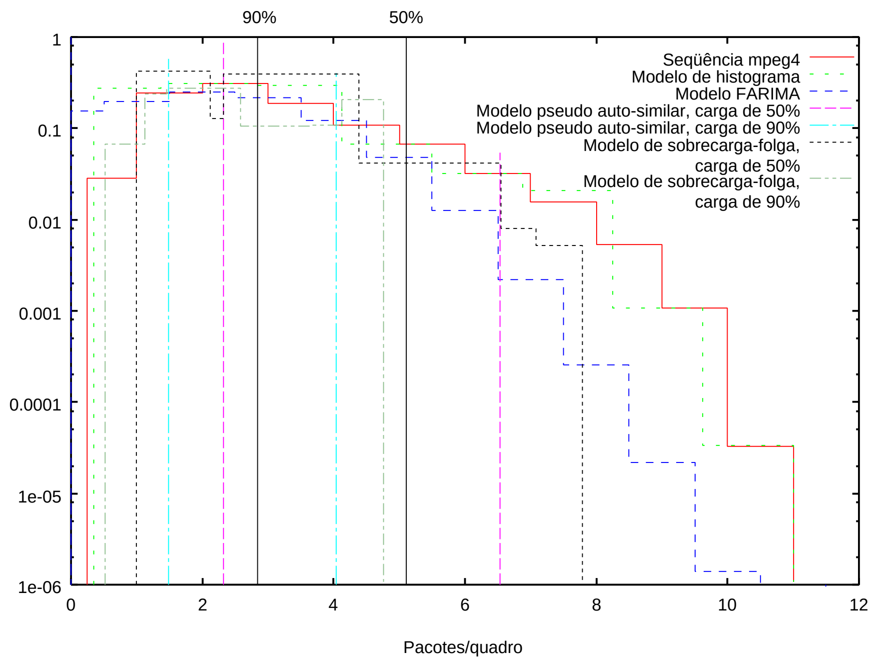


Figura 5.12: Histogramas de taxa, escala logarítmica, filme *Jurassic Park* em MPEG4

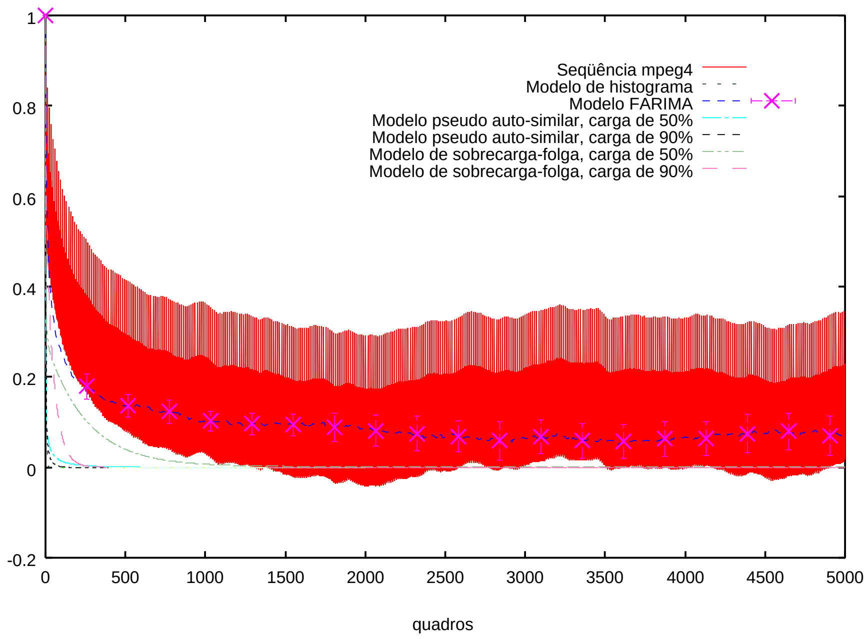


Figura 5.13: Funções de autocorrelação, filme *Jurassic Park* em MPEG4

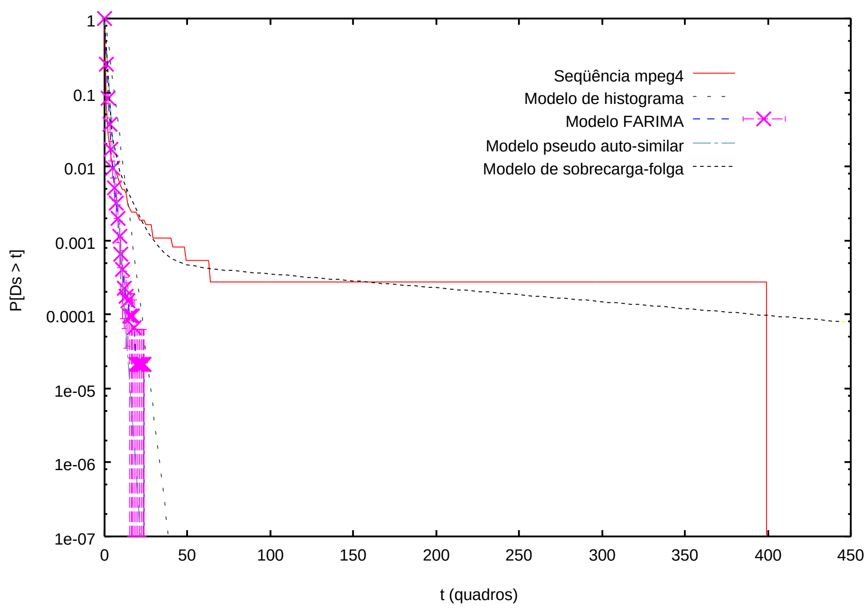


Figura 5.14: Distribuição de duração de sobrecarga, filme *Jurassic Park* em MPEG4, r a 50% de carga

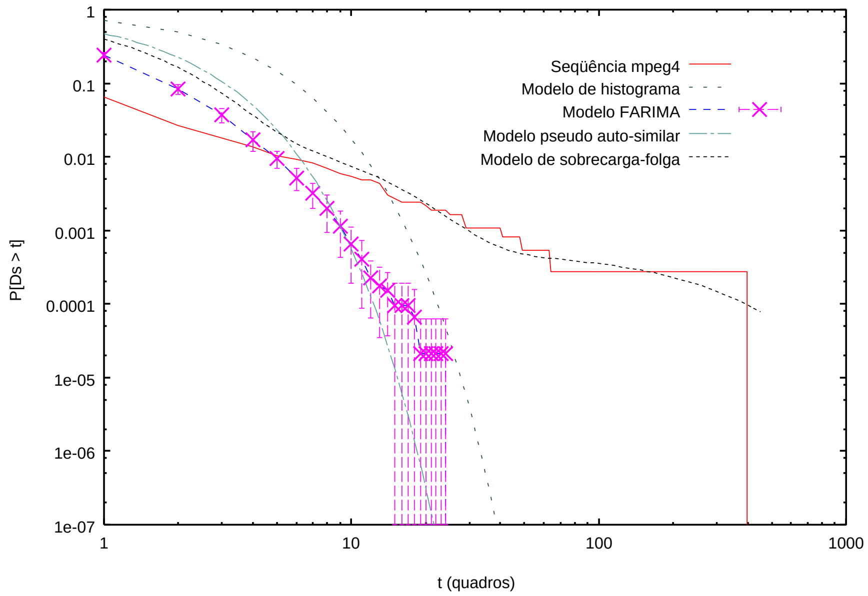


Figura 5.15: Distribuição de duração de sobrecarga, escala logarítmica, filme *Jurassic Park* em MPEG4, r a 50% de carga

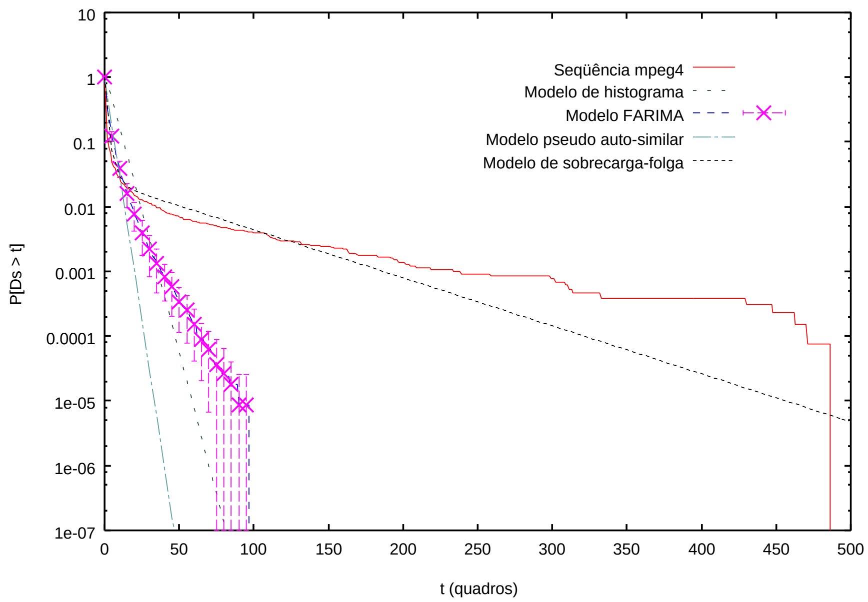


Figura 5.16: Distribuição de duração de sobrecarga, filme *Jurassic Park* em MPEG4, r a 90% de carga

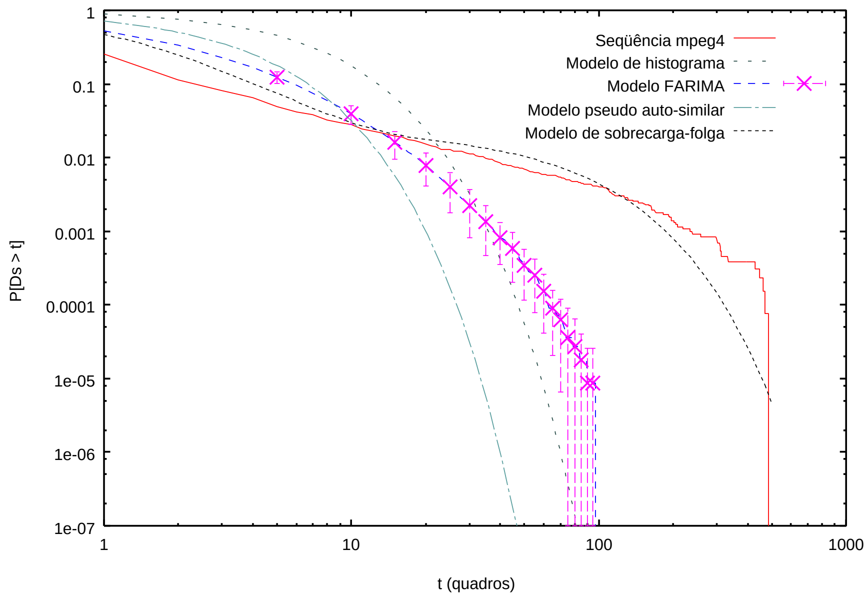


Figura 5.17: Distribuição de duração de sobrecarga, escala logarítmica, filme *Jurassic Park* em MPEG4, r a 90% de carga

a respeito. É importante lembrar que o modelo F-ARIMA não usa a carga como forma de parametrização e é idealizado para capturar a perda qualquer que seja a carga, o que não se verifica para a carga de 50%.

As Figuras de 5.18 a 5.21 trazem as distribuições de folga de forma análoga às de sobrecarga. Mais uma vez, os modelos de sobrecarga-folga têm melhor eficiência nesta captura, no entanto, para a carga de 90%, todos os modelos capturam satisfatoriamente esta distribuição.

5.5.2 Seqüência do filme *Star Trek - First Contact*

Este é outro filme de ação e ficção científica codificado a 25 $quadros/segundo$. Sua taxa média é igual a 1.097 $pacotes/quadro$ e sua variância é igual a 0.821 $(pacotes/quadro)^2$. O parâmetro de Hurst, usado pelo modelo F-ARIMA e também estimado como na seção 3.4, é igual a 0.975. A duração média em sobrecarga, usada para construir o modelo pseudo auto-similar modificado, é igual a 1.376 quadros para carga de 50% e 1.713 para carga de 90%. As Figuras 5.22 e 5.23 trazem as perdas para as cargas de 50 e 90%, com tamanho de fila variando em escala logarítimica.

Como é possível notar, para a carga de 50% o modelo de sobrecarga-folga prediz

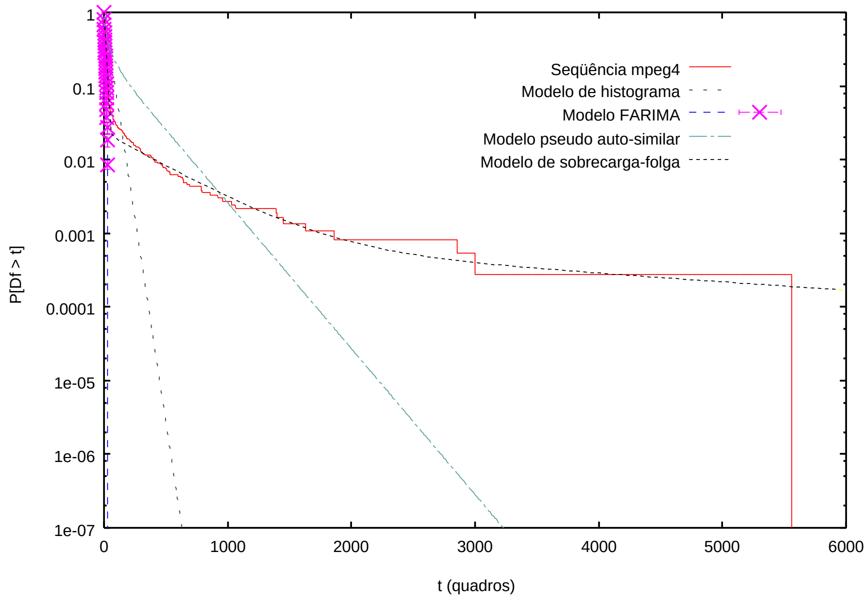


Figura 5.18: Distribuição de duração de folga, filme *Jurassic Park* em MPEG4, r a 50% de carga

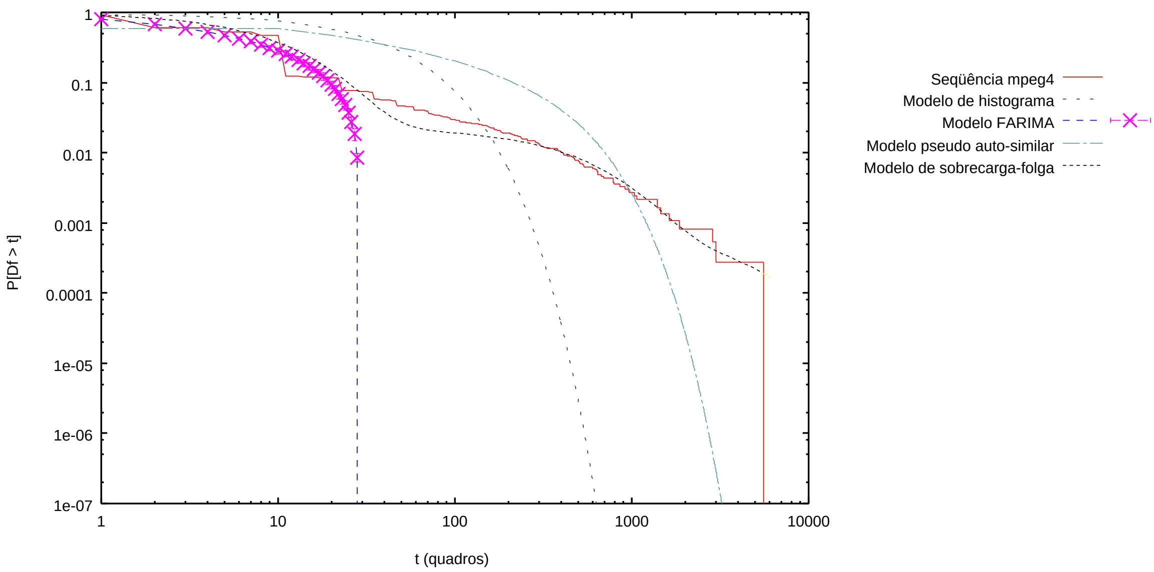


Figura 5.19: Distribuição de duração de folga, escala logarítmica, filme *Jurassic Park* em MPEG4, r a 50% de carga

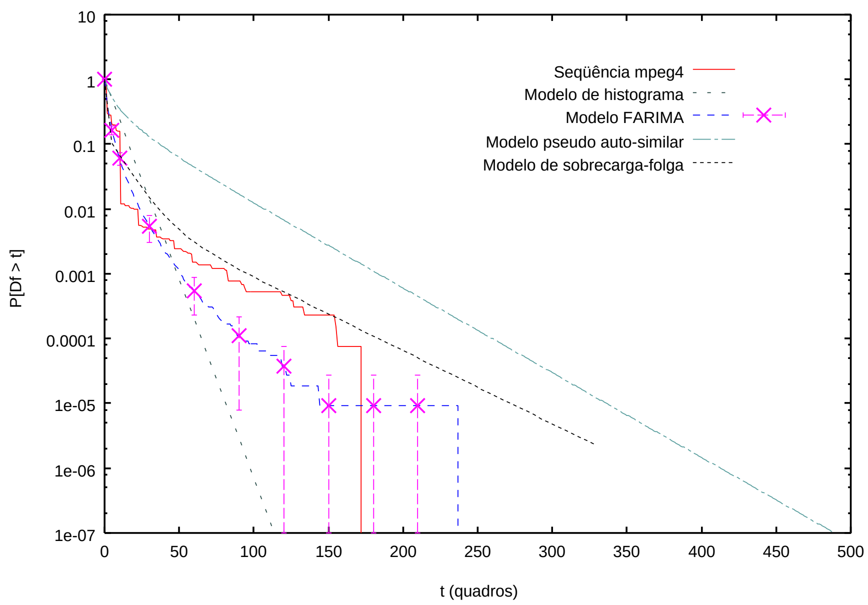


Figura 5.20: Distribuição de duração de folga, filme *Jurassic Park* em MPEG4, r a 90% de carga

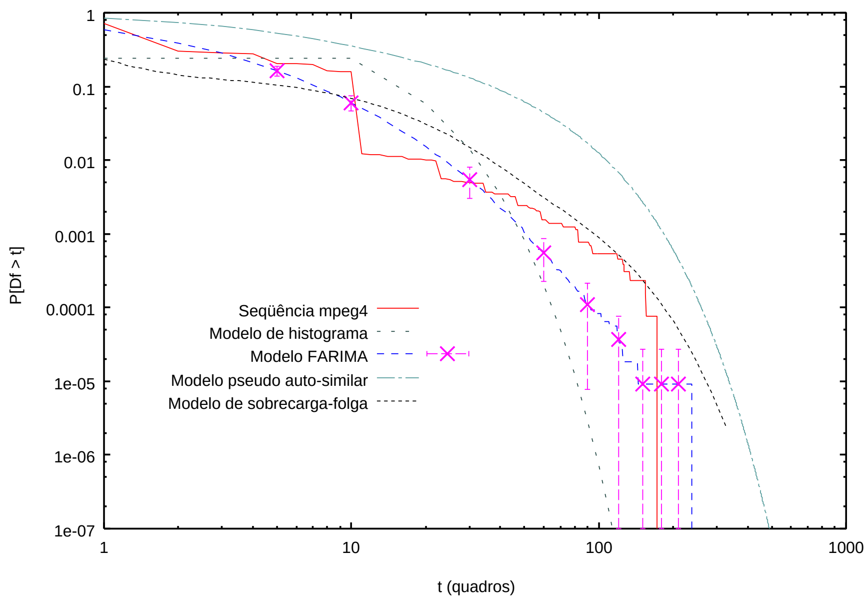


Figura 5.21: Distribuição de duração de folga, escala logarítmica, filme *Jurassic Park* em MPEG4, r a 90% de carga

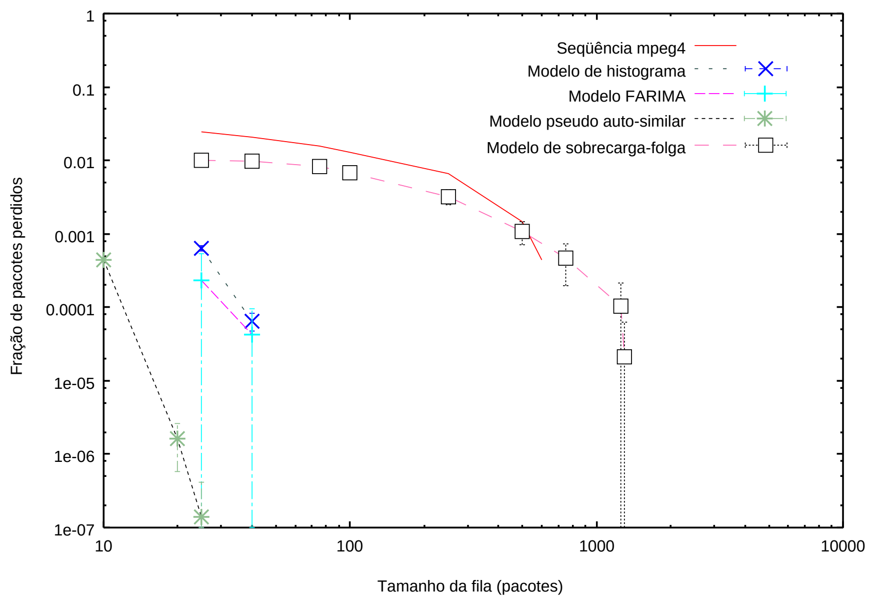


Figura 5.22: Fração de pacotes perdidos, escala logarítmica, filme *Star Trek - First Contact* em MPEG4, r a 50% de carga

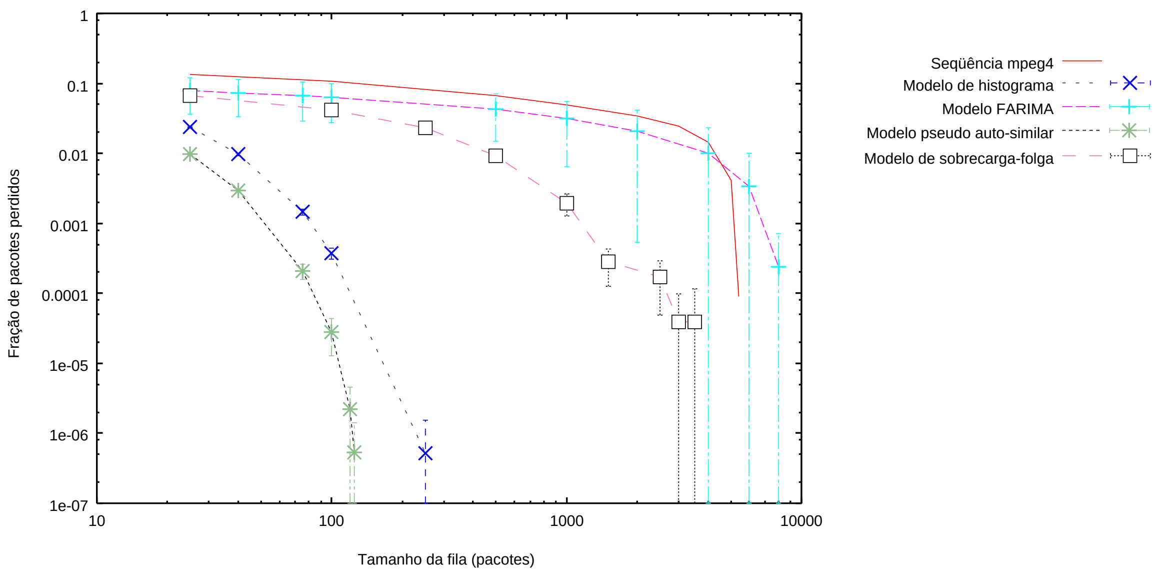


Figura 5.23: Fração de pacotes perdidos, escala logarítmica, filme *Star Trek - First Contact* em MPEG4, r a 90% de carga

a perda de pacotes real com muito melhor aproximação que os demais modelos. Para carga de 90%, uma vez mais o modelo F-ARIMA tem melhor aproximação para tamanho máximo de fila até aproximadamente 5000 pacotes. O modelo de sobrecarga-folga consegue aproximar a perda apenas para um tamanho de fila inferior a 100 pacotes.

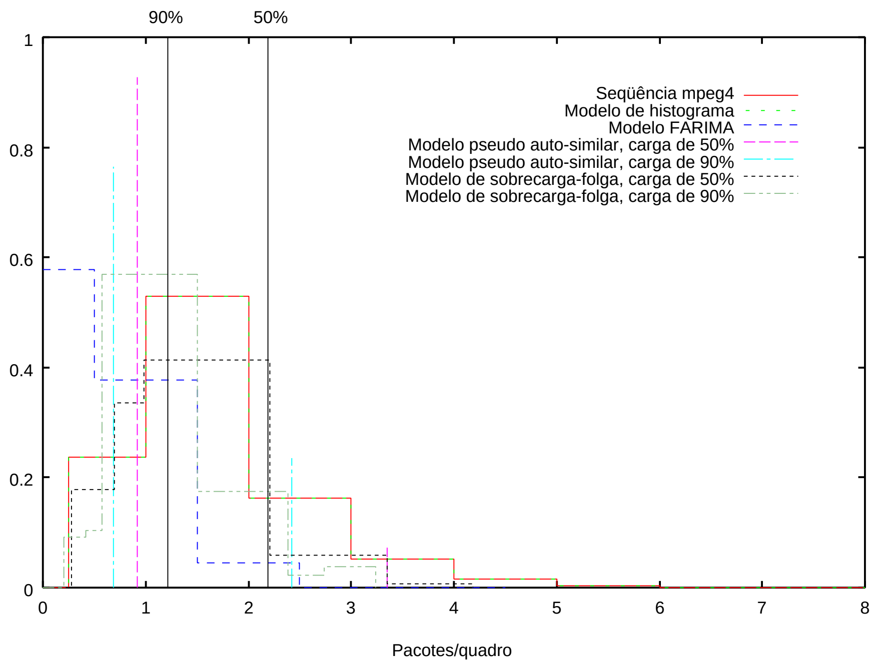


Figura 5.24: Histogramas de taxa, filme *Star Trek - First Contact* em MPEG4

As Figuras 5.24 e 5.25 mostram os histogramas de taxa para a seqüência e os modelos, tanto em escala linear quanto logarítmica. Novamente, é possível notar que o histograma de taxas para o modelo de sobrecarga-folga com r referente à 90% de carga não aproxima a cauda do histograma para a seqüência real. Isto também ocorre com o modelo F-ARIMA, contudo sua função de autocorrelação é mais próxima da real, como mostrado na Figura 5.26. Ainda assim, é possível verificar que a autocorrelação para os modelos de sobrecarga-folga têm razoável aproximação com a função real, sendo que há um decaimento mais rápido para a função do modelo com carga de 90%, o que é coerente com seu comportamento de perda já que esta decai mais que a real quando o tamanho da fila aumenta.

As Figuras de 5.27 a 5.30 trazem as distribuições de duração de sobrecarga para seqüência e modelos. Mais uma vez, as distribuições hiper-exponenciais dos modelos de sobrecarga-folga são as que melhor aproximam as distribuições da seqüência para

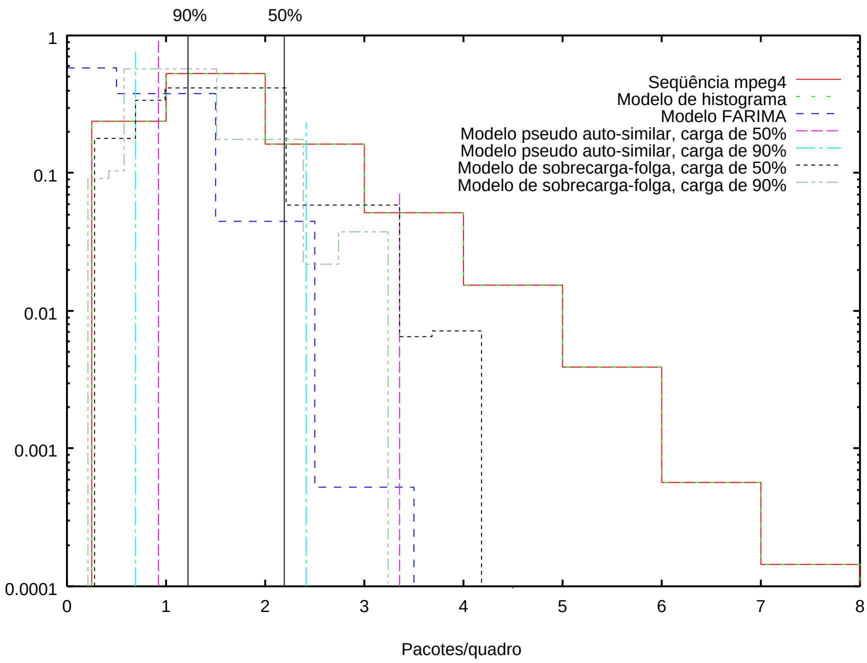


Figura 5.25: Histogramas de taxa, escala logarítmica, filme *Star Trek - First Contact* em MPEG4

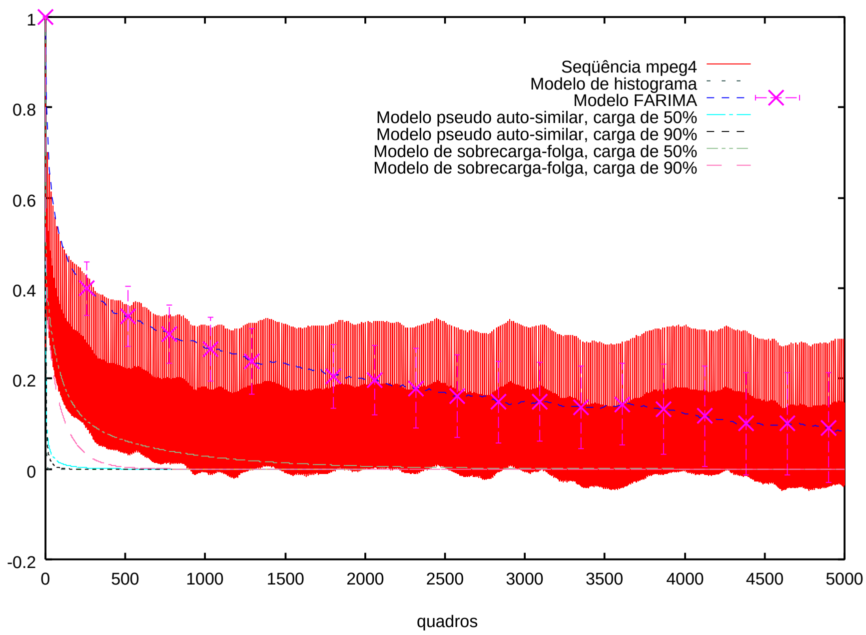


Figura 5.26: Funções de autocorrelação, filme *Star Trek - First Contact* em MPEG4

ambas as cargas. No entanto, é importante notar que no modelo F-ARIMA este descritor casa satisfatoriamente com o da seqüência real para a carga de 90%. Esta característica, aliada à distribuição de taxa e autocorrelação mais próximas das reais, o torna o melhor candidato a prever nestas condições de alta carga.

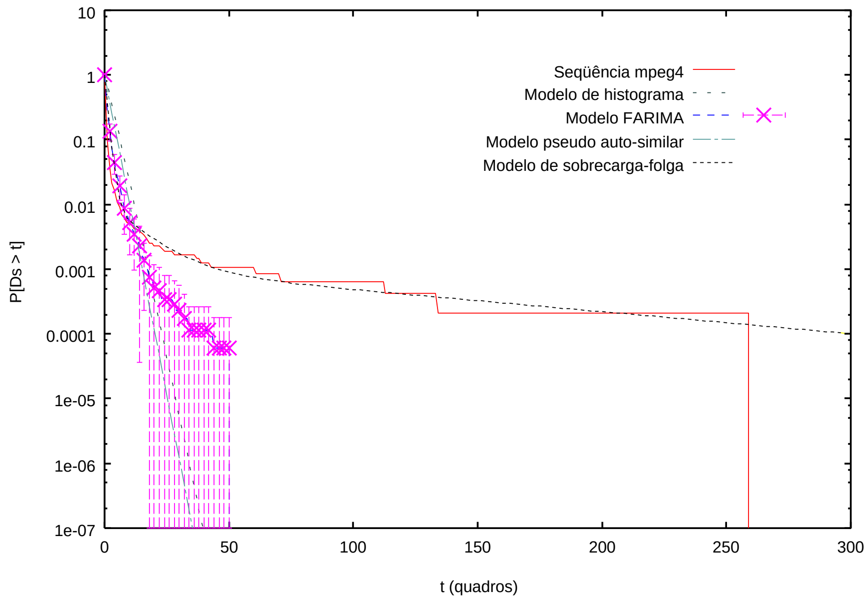


Figura 5.27: Distribuição de duração de sobrecarga, filme *Star Trek - First Contact* em MPEG4, r a 50% de carga

As Figuras de 5.31 a 5.34 mostram as distribuições de duração de folga para seqüência e modelos. As distribuições dos modelos de sobrecarga-folga se adequam melhor para ambas as cargas, embora a cauda deste modelo para a carga de 50% fique razoavelmente abaixo do caso real. O modelo F-ARIMA foi o que pior modelou esta distribuição, mantendo sua cauda bem superior à cauda real. No entanto, como mostrado na seção 4.4.1, a descrição da folga pouco influi na medida de perda.

5.5.3 Seqüência do filme *Silence of the Lambs*

Mais uma vez será usado este filme de ação policial como exemplo, cuja taxa média é de 1.954 $pacotes/quadro$ e variância igual a 2.584 $(pacotes/quadro)^2$, sendo em 25 $quadros/segundo$ a taxa de transmissão dos quadros. Nos testes do capítulo 4, este filme foi descrito para a carga de 33%. Desta vez os testes para carga de 90% envolvem a comparação com outros modelos. Como já mostrado no capítulo

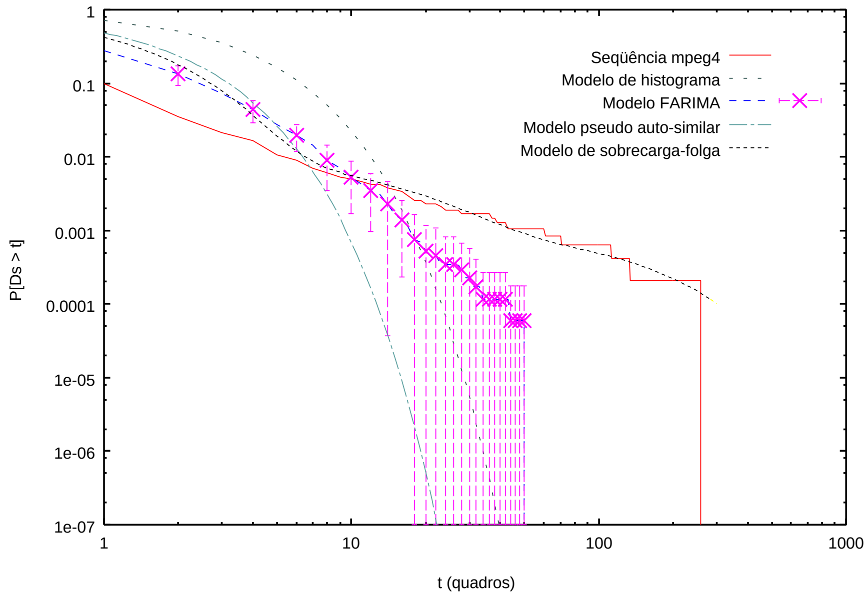


Figura 5.28: Distribuição de duração de sobrecarga, escala logarítmica, filme *Star Trek - First Contact* em MPEG4, r a 50% de carga

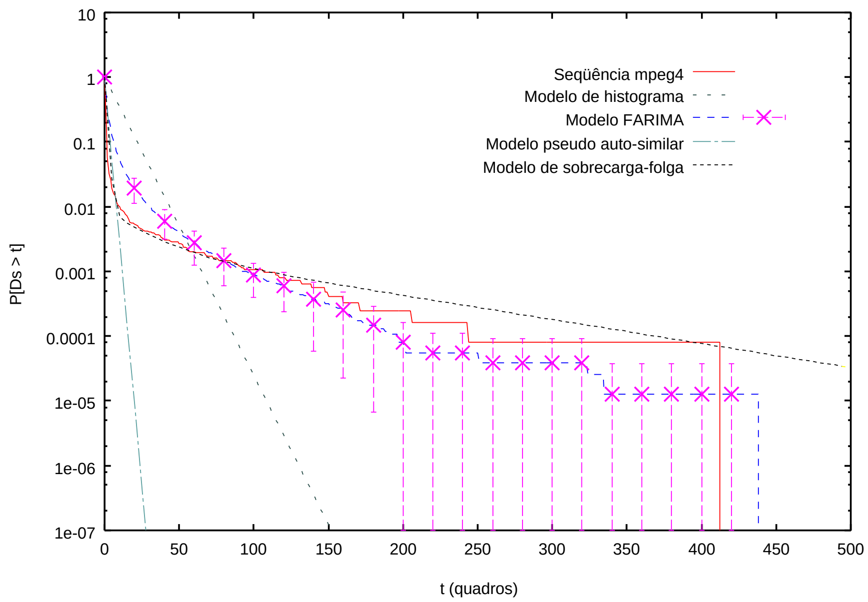


Figura 5.29: Distribuição de duração de sobrecarga, filme *Star Trek - First Contact* em MPEG4, r a 90% de carga

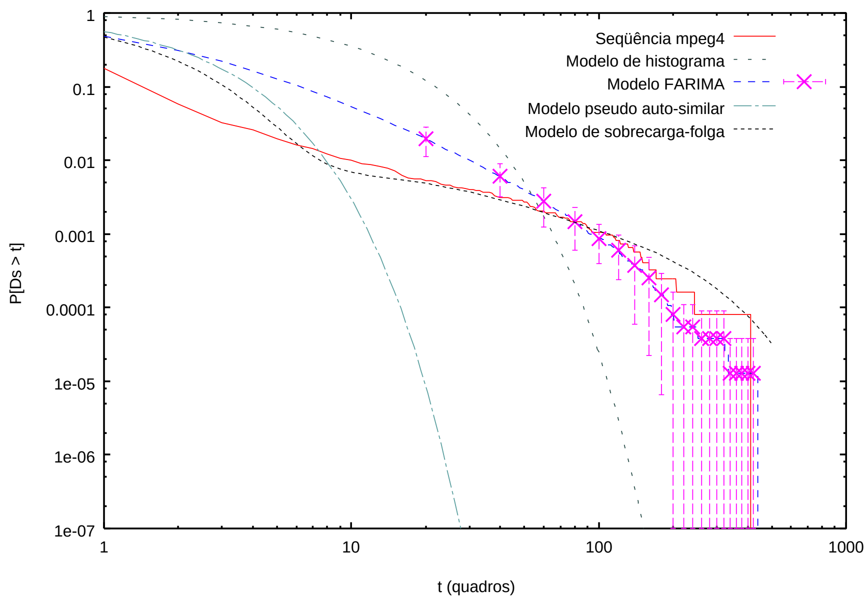


Figura 5.30: Distribuição de duração de sobrecarga, escala logarítmica, filme *Star Trek - First Contact* em MPEG4, r a 90% de carga

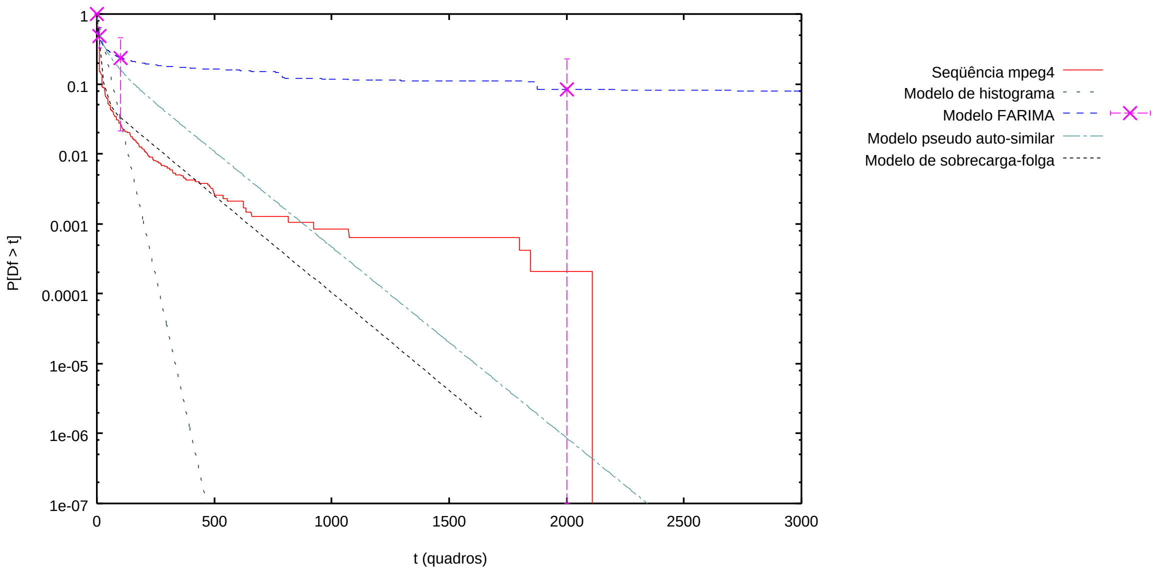


Figura 5.31: Distribuição de duração de folga, filme *Star Trek - First Contact* em MPEG4, r a 50% de carga

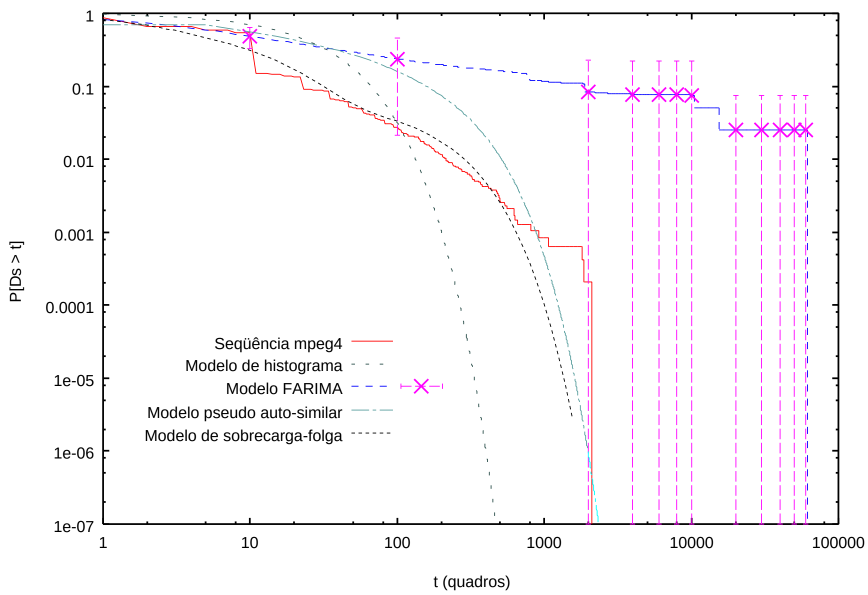


Figura 5.32: Distribuição de duração de folga, escala logarítmica, filme *Star Trek - First Contact* em MPEG4, r a 50% de carga

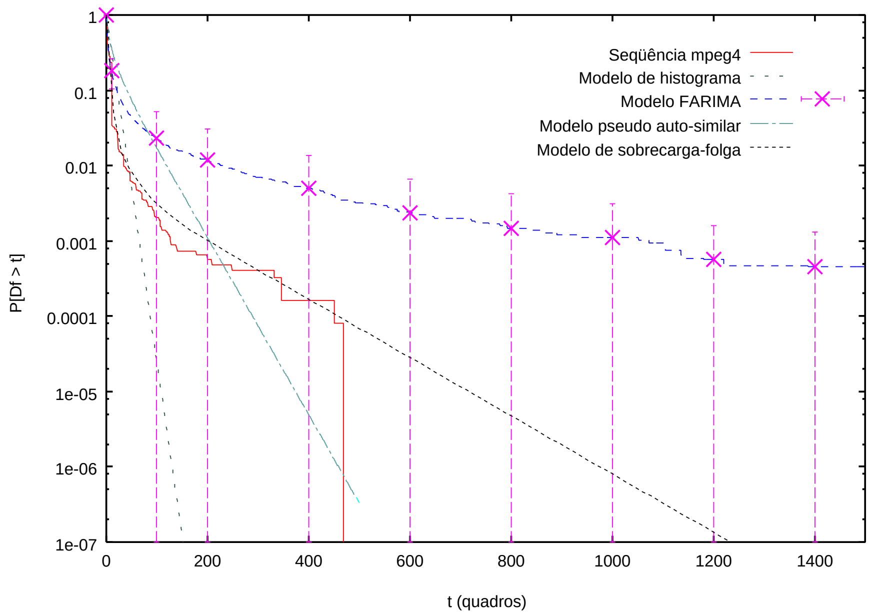


Figura 5.33: Distribuição de duração de folga, filme *Star Trek - First Contact* em MPEG4, r a 90% de carga

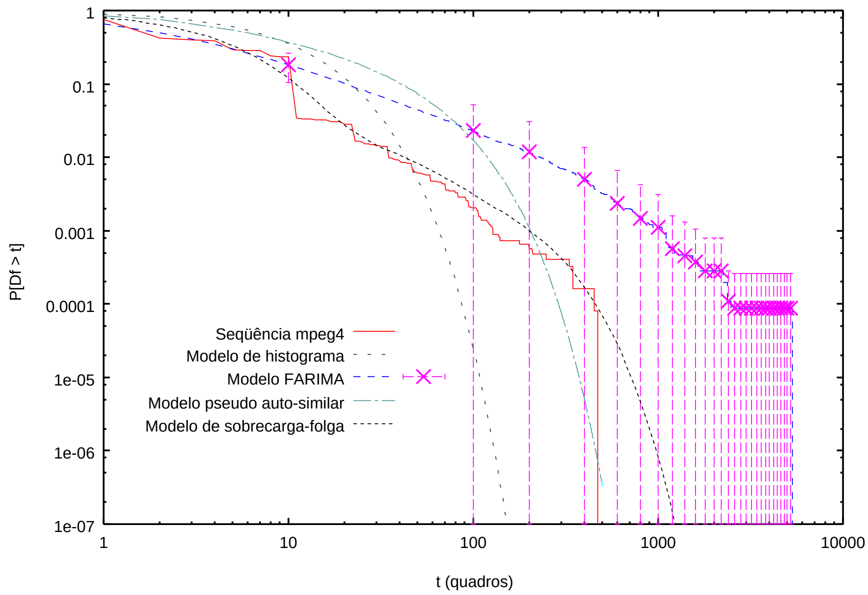


Figura 5.34: Distribuição de duração de folga, escala logarítmica, filme *Star Trek - First Contact* em MPEG4, r a 90% de carga

3, para esta seqüência o parâmetro de Hurst é igual a 0.987. A duração média em sobrecarga, usada para construir o modelo pseudo auto-similar modificado, é igual a 2.198 quadros para carga de 50% e 2.635 quadros para carga de 90%.

As Figuras 5.35 e 5.36 mostram as frações de perda para os modelos e a seqüência em ambas as cargas. Para a carga de 50%, a melhor aproximação é a do modelo de sobrecarga-folga, todavia a previsão piora à medida que o tamanho máximo da fila aumenta para valores muito grandes, acima dos comumente praticados. O modelo F-ARIMA subestimou a perda, sendo este o mais aproximado somente para um tamanho de fila superior a 3500 pacotes. Já para a carga de 90%, o modelo de sobrecarga-folga teve um comportamento semelhante ao observado para as outras seqüências: a fração de perda foi subestimada para filas com tamanhos máximos grandes. O modelo F-ARIMA, que costuma ser mais eficiente para cargas maiores, superestimou bastante a perda para a carga de 90% e tamanho máximo de fila superior a 2000 pacotes. Demais modelos tiveram desempenho muito inferior para ambas as cargas.

As Figuras 5.37 e 5.38 mostram os histogramas para modelos e seqüência em escala linear e logarítmica. As mesmas observações para os exemplos anteriores aqui se aplicam, sendo que o modelo F-ARIMA desta vez não aproximou bem a

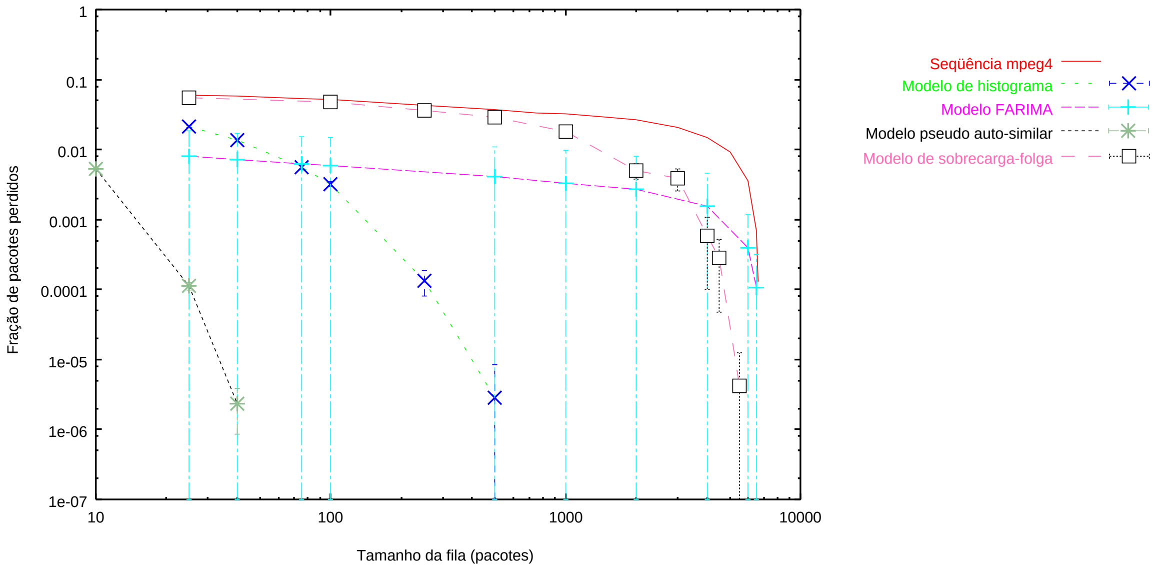


Figura 5.35: Fração de pacotes perdidos, escala logarítmica, filme *Silence of the Lambs* em MPEG4, r a 50% de carga

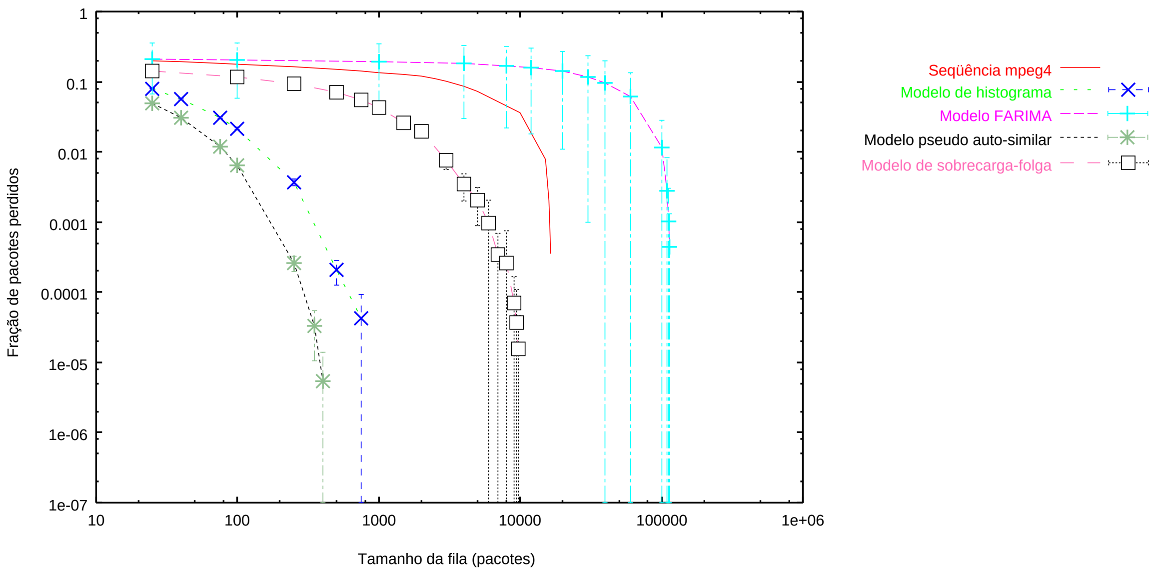


Figura 5.36: Fração de pacotes perdidos, escala logarítmica, filme *Silence of the Lambs* em MPEG4, r a 90% de carga

cauda do histograma.

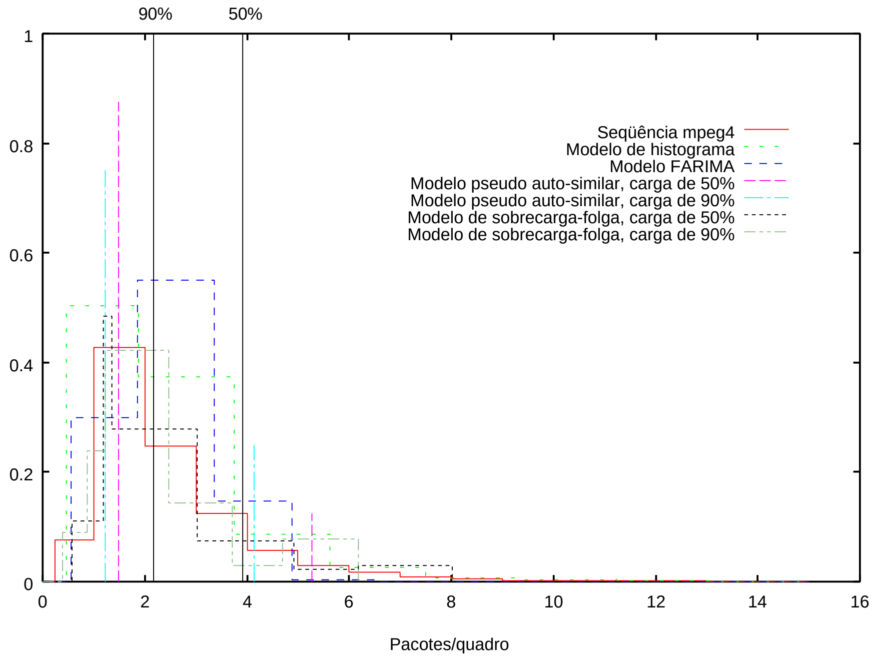


Figura 5.37: Histogramas de taxa, filme *Silence of the Lambs* em MPEG4

A Figura 5.39 traz as funções de autocorrelação para modelos e seqüência. Desta vez o modelo de sobrecarga-folga orientado para carga de 50% teve este descritor um pouco mais distanciado do real que nos demais exemplos, o que é coerente com o afastamento das curvas de perda para este modelo e a seqüência real quando o tamanho da fila possui valores maiores que 3500 pacotes.

As Figuras de 5.40 a 5.43 mostram as distribuições de duração de sobrecarga para modelos e seqüência. Novamente, os modelos de sobrecarga-folga capturam com precisão esta distribuição. Interessante notar o caso do modelo F-ARIMA para a carga de 90%, cuja curva para esta distribuição aparece acima da curva real. Isto está de acordo com a maneira como este modelo super-estimou a perda para esta carga.

As Figuras de 5.44 a 5.47 trazem as distribuições de duração de folga para as cargas de 50 e 90%. Em ambos os casos, o modelo de sobrecarga-folga teve a melhor captura deste descritor, sendo que para a carga de 90% quase todos os modelos, excetuando o modelo de histograma, tiveram distribuição razoavelmente próxima da real. Para a carga de 50%, tanto o modelo de histograma quanto o F-ARIMA tiveram as curvas para este descritor bem diferentes da curva da seqüência real.

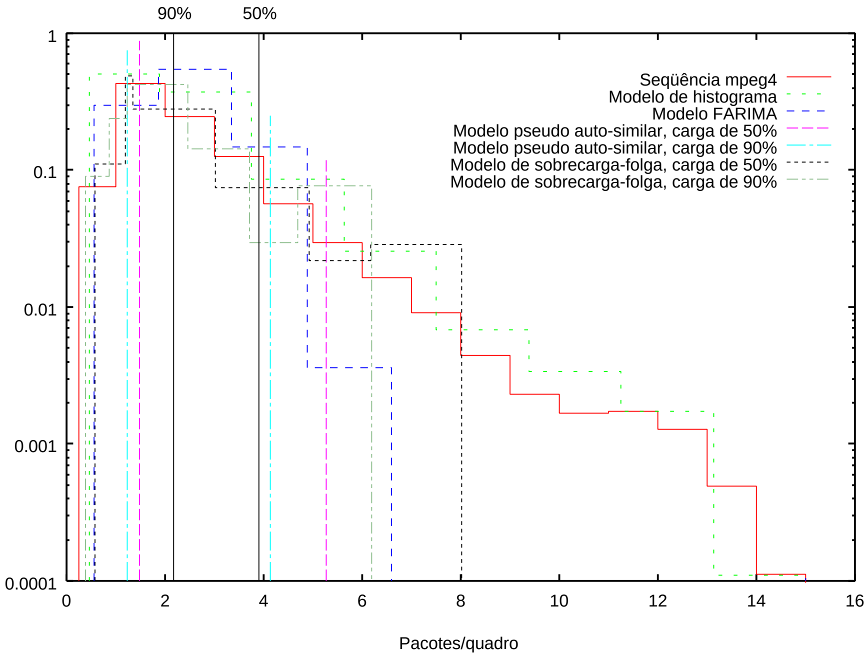


Figura 5.38: Histogramas de taxa, escala logarítmica, filme *Silence of the Lambs* em MPEG4

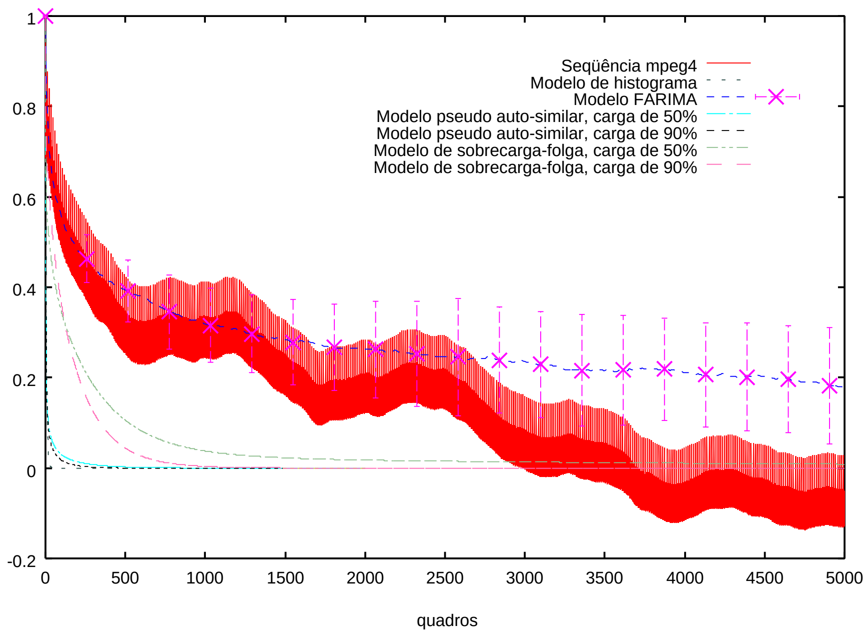


Figura 5.39: Funções de autocorrelação, filme *Silence of the Lambs* em MPEG4

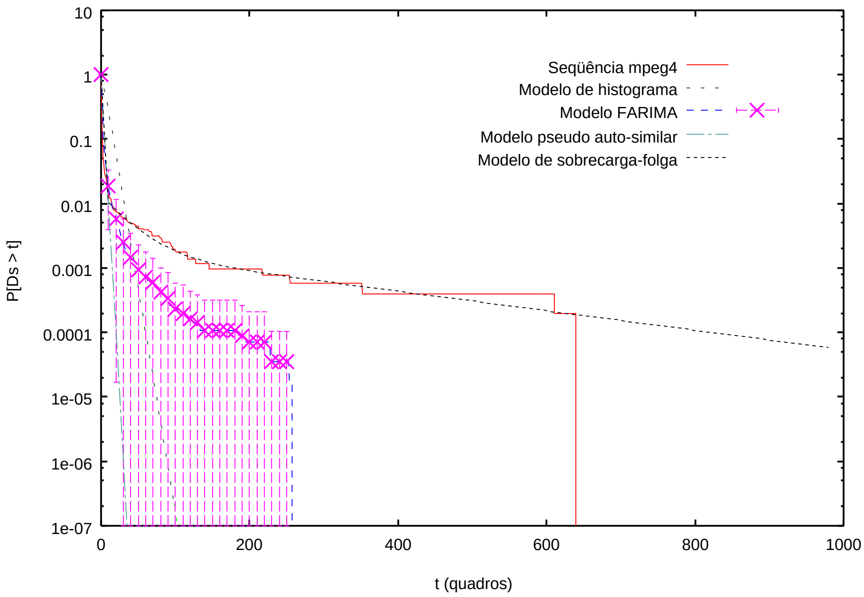


Figura 5.40: Distribuição de duração de sobrecarga, filme *Silence of the Lambs* em MPEG4, r a 50% de carga

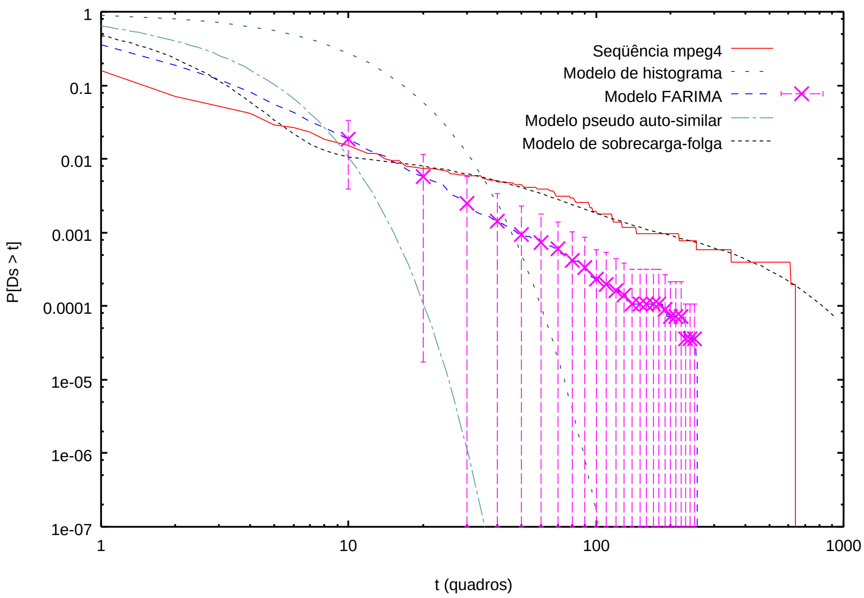


Figura 5.41: Distribuição de duração de sobrecarga, escala logarítmica, filme *Silence of the Lambs* em MPEG4, r a 50% de carga

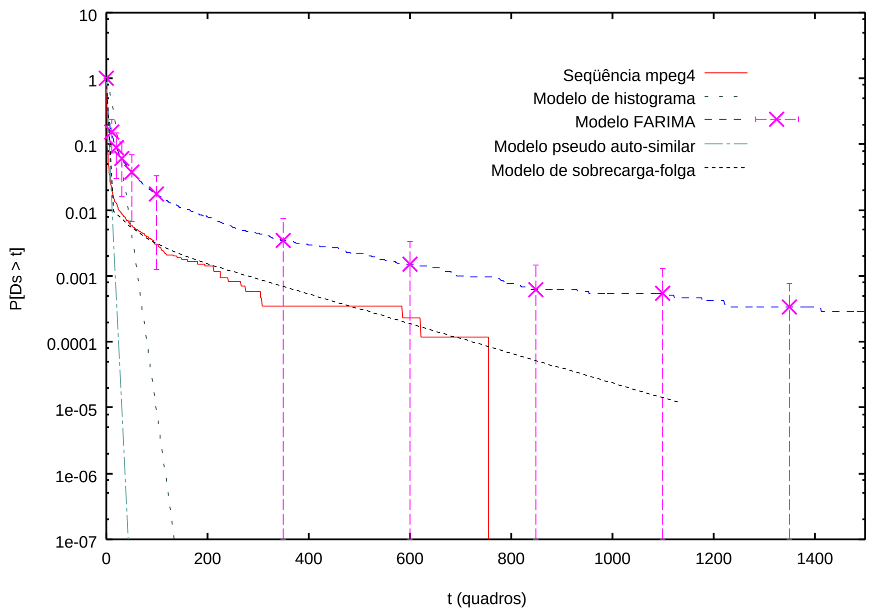


Figura 5.42: Distribuição de duração de sobrecarga, filme *Silence of the Lambs* em MPEG4, r a 90% de carga

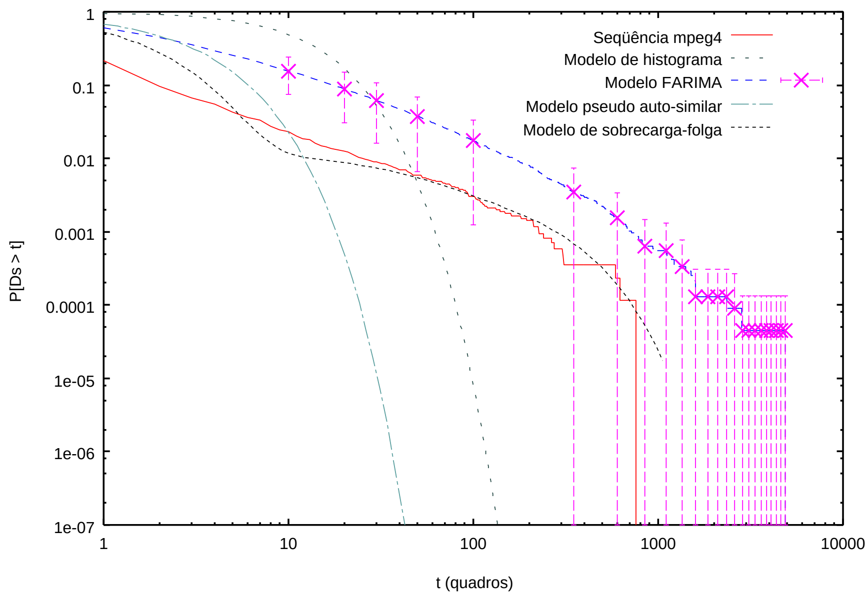


Figura 5.43: Distribuição de duração de sobrecarga, escala logarítmica, filme *Silence of the Lambs* em MPEG4, r a 90% de carga

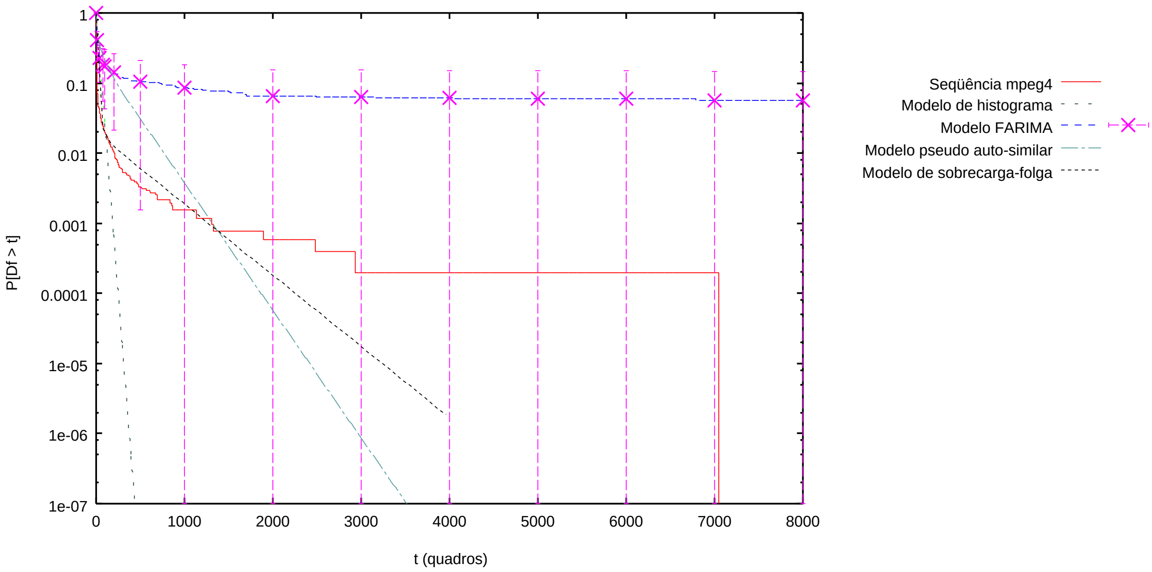


Figura 5.44: Distribuição de duração de folga, filme *Silence of the Lambs* em MPEG4, r a 50% de carga

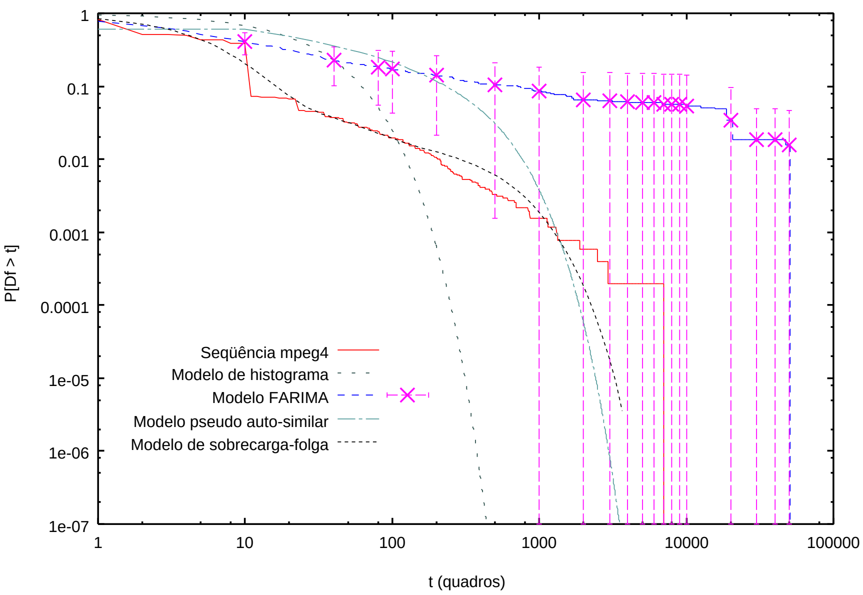


Figura 5.45: Distribuição de duração de folga, escala logarítmica, filme *Silence of the Lambs* em MPEG4, r a 50% de carga

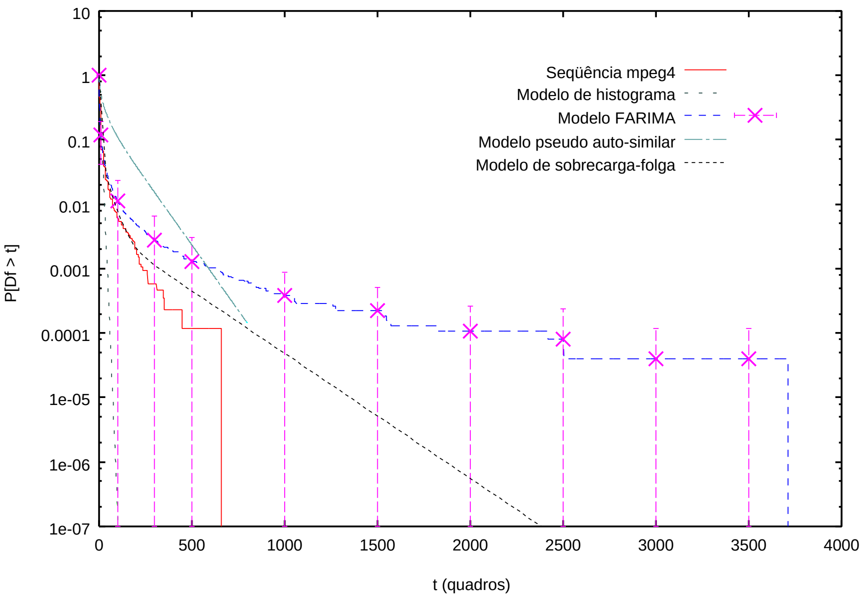


Figura 5.46: Distribuição de duração de folga, filme *Silence of the Lambs* em MPEG4, r a 90% de carga

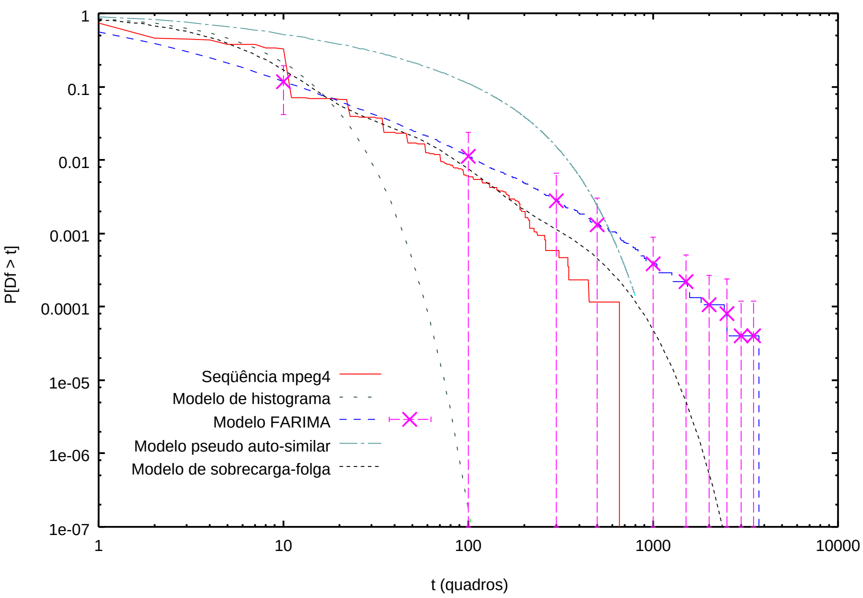


Figura 5.47: Distribuição de duração de folga, escala logarítmica, filme *Silence of the Lambs* em MPEG4, r a 90% de carga

5.6 Análise dos Modelos para Seqüências de Tráfego WAN da Rede Abilene

Nesta seção, apenas os resultados referentes ao modelo de sobrecarga-folga são analisados. As conclusões extraídas das comparações entre modelos na seção anterior mostraram a influência dos diferentes descritores na aproximação da perda uma vez que os modelos testados têm características distintas. Em prosseguimento a este estudo, deseja-se analisar o comportamento do modelo de sobrecarga-folga quando usado para caracterizar tráfegos de redes WAN. Este tipo de tráfego tem padrão bastante distinto dos tráfegos MPEG4 já que a agregação do fluxo de pacotes de diferentes aplicações tende a suavizar, até certo ponto, o perfil de tráfego.

Para a construção dos modelos de sobrecarga-folga que serão vistos adiante foram usadas as seguintes tolerâncias de taxa (ϵ) e média (ψ) na captura das distribuições de duração:

	ϵ/ψ	
Seqüência, carga	Sobrecarga	Folga
Entrada, 90%	0.045/0.40	0.001/0.14
Saída, 90%	-	0.001/0.40
Saída, 95%	0.070/0.30	0.004/0.03

5.6.1 Seqüência de tráfego WAN de entrada

A seqüência contendo o tráfego de entrada na interface do roteador de Indianapolis voltada para Cleveland possui uma média de 7954.958 pacotes a cada 100 milisegundos. Sua variância é igual a 304165.166 $(\text{pacotes})^2$ e o pico é igual a 10712 pacotes, ambas medidas também extraídas para a escala de tempo de 100 milisegundos. Como a carga mínima para haver perda deve ser superior a 74.26%, o modelo de sobrecarga-folga calculado para esta seqüência foi avaliado para uma carga igual a 90%, a mesma utilizada na seção anterior para mostrar o comportamento do modelo proposto frente a cargas altas.

A Figura 5.48 mostra a aproximação da perda para esta carga. As Figuras 5.49 e

5.50 mostram, respectivamente, as aproximações para as distribuições de duração de sobrecarga e folga. Elas foram obtidas através do algoritmo proposto na seção 4.2. A Figura 5.51 traz as curvas de autocorrelação para modelo e seqüência. Podemos observar que o modelo possui esses descritores bem próximos aos obtidos para a seqüência real.

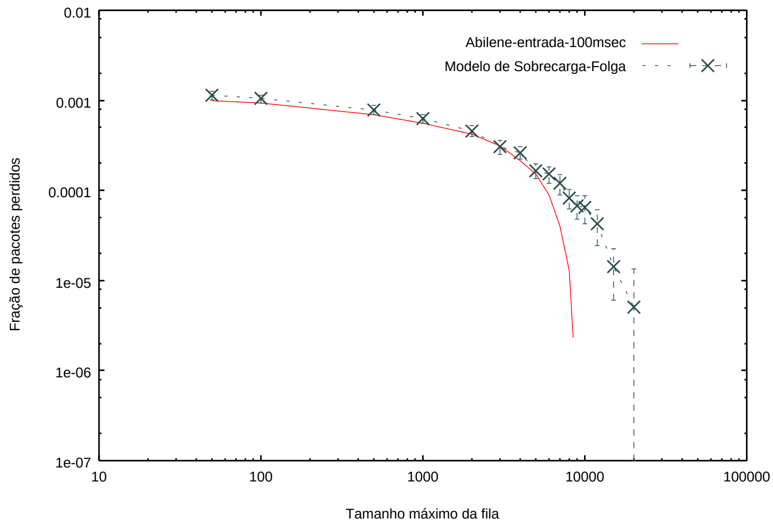


Figura 5.48: Fração de perda para tráfego de entrada e modelo de sobrecarga-folga, carga de 90%

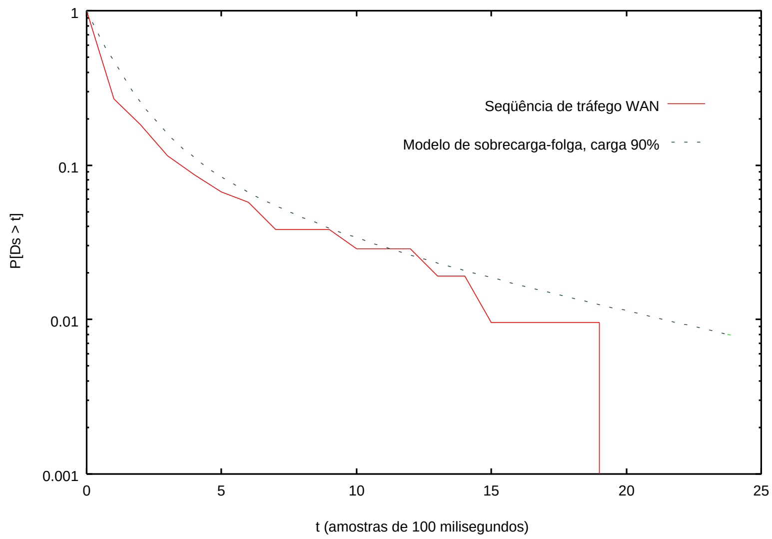


Figura 5.49: Distribuição das durações de sobrecarga para tráfego de entrada, carga de 90%

Pode-se verificar que a aproximação da perda através do modelo de sobrecarga-folga é bastante satisfatória para esta seqüência do tráfego de entrada na interface

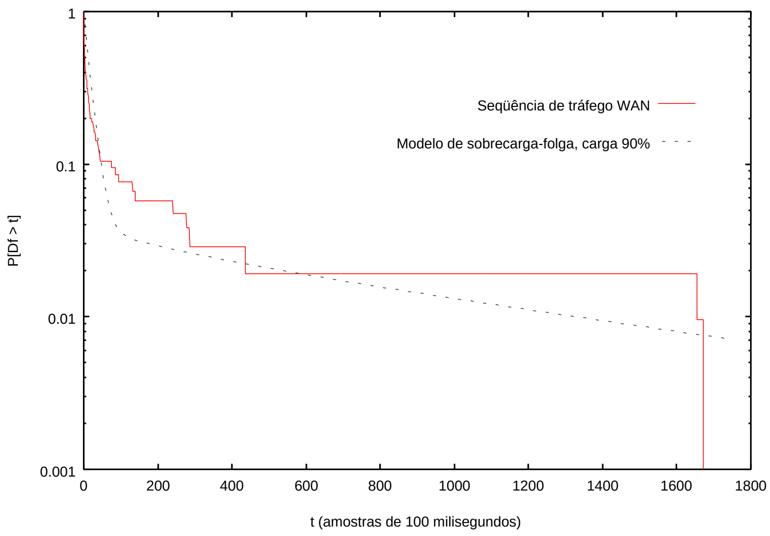


Figura 5.50: Distribuição das durações de folga para tráfego de entrada, carga de 90%

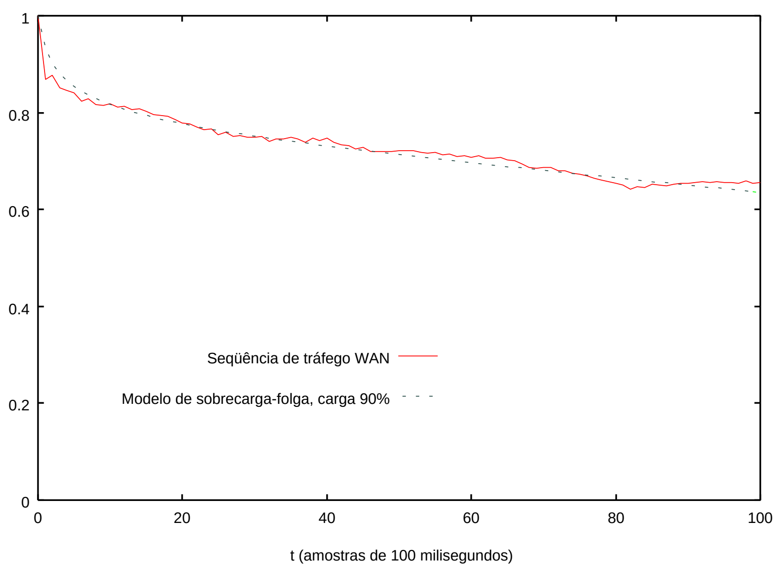


Figura 5.51: Função de autocorrelação para tráfego de entrada e modelo de sobrecarga-folga com carga de 90%

mencionada. Embora trate-se de uma carga alta, o modelo de sobrecarga-folga consegue se adaptar bem já que este tipo de tráfego não possui as mesmas variações de taxa encontradas nas seqüências MPEG4.

5.6.2 Seqüência de tráfego WAN de saída

No caso da seqüência do tráfego de saída para esta mesma interface, a média de pacotes a cada 100 milisegundos é igual a 7660.626, a variância é igual a 113111.228 $(\text{pacotes})^2$ e o pico igual a 9028 pacotes. Para estes valores, a carga mínima para perda é ainda maior - 84.85%. Em decorrência destas características, neste exemplo foram usadas cargas de 90% e 95%.

O exemplo com carga de 90% para esta seqüência tem um aspecto bastante interessante. Como o nível de sobrecarga fica muito próximo da taxa de pico, apenas duas durações distintas de sobrecarga são encontradas - 100 e 200 milisegundos. A conseqüência disto foi uma variação do modelo de sobrecarga-folga até então usado: construiu-se um modelo de sobrecarga-folga onde as durações de sobrecargas foram modeladas por apenas dois estados. As probabilidades de entrada nestes estados, após o término de uma folga, foram obtidas diretamente do histograma das durações de sobrecarga: 0.914286 para durações de 100 milisegundos e 0.085714 para durações de 200 milisegundos. O processo de obtenção das taxas em sobrecarga foi o mesmo definido na seção 4.3.2, adaptado para dois estados. O processo para a modelagem das folgas não sofreu alteração, as tolerâncias de taxa (ϵ) e média (ψ) usadas foram respectivamente 0.001 e 0.40.

A aproximação da fração de perda se encontra na Figura 5.52. A Figura 5.53 mostra a aproximação para a distribuição das durações de folga. As curvas de autocorrelação para modelo e seqüência podem ser vistas na Figura 5.54. Como o número de durações de sobrecarga que ocorrem para este caso é igual a dois, suas probabilidades são diretamente parametrizadas no modelo e, portanto, não é apresentada figura mostrando esta distribuição. Os valores das probabilidades estão indicados no parágrafo anterior.

Para o exemplo com carga de 95%, as distribuições de duração de sobrecarga

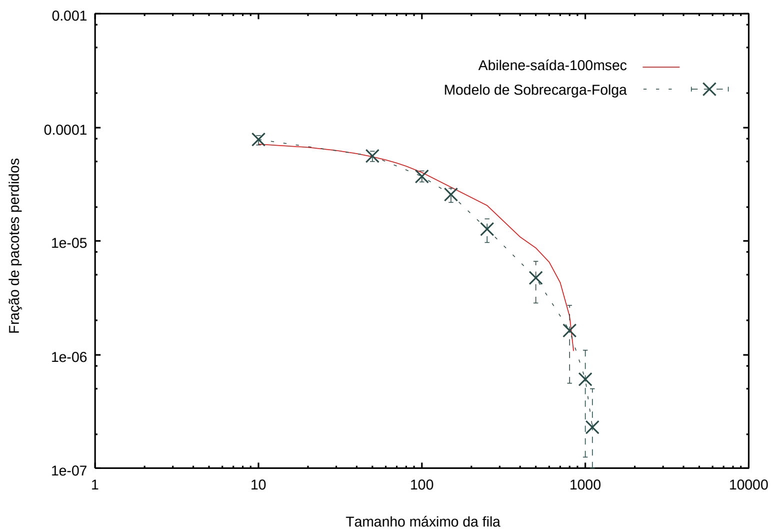


Figura 5.52: Fração de perda para tráfego de saída e modelo de sobrecarga-folga, carga de 90%

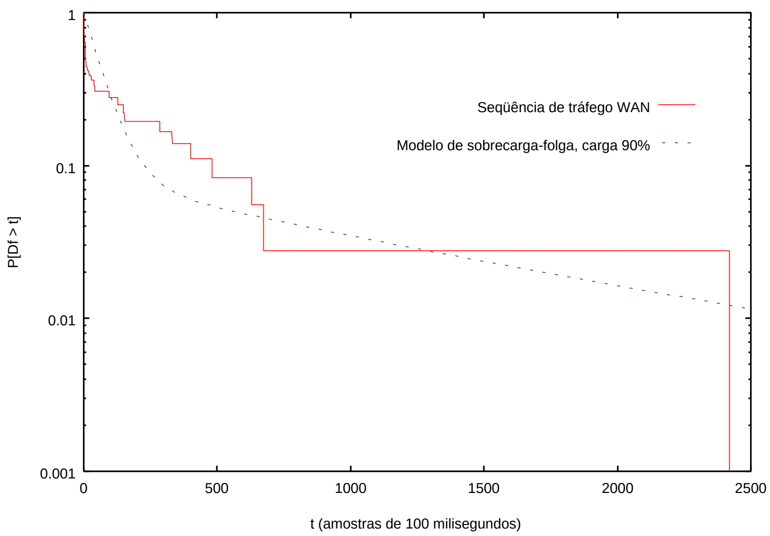


Figura 5.53: Distribuição das durações de folga para tráfego de saída, carga de 90%

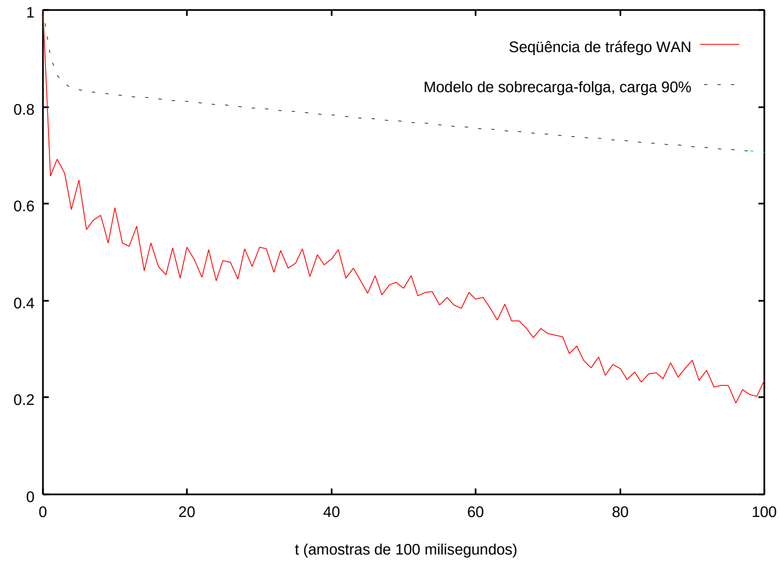


Figura 5.54: Função de autocorrelação para tráfego de saída e modelo de sobrecarga-folga com carga de 90%

e de folga são mostradas nas Figuras 5.56 e 5.57. As tolerâncias usadas foram 0.07 (ϵ) e 0.30 (ψ), para as durações de sobrecarga, e 0.004 (ϵ) e 0.03 (ψ), para as durações de folga. As curvas de perda podem ser vistas na Figura 5.55. As curvas de autocorrelação podem ser vistas na Figura 5.58.

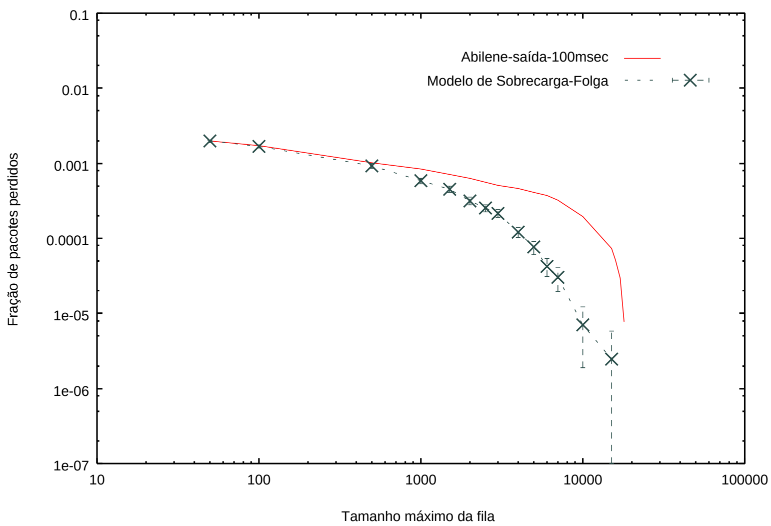


Figura 5.55: Fração de perda para tráfego de saída e modelo de sobrecarga-folga, carga de 95%

Como também é possível verificar pelos exemplos usando o tráfego de saída da interface de rede citada, o modelo de sobrecarga-folga é bastante efetivo para se modelar tráfegos de redes WAN mesmo para o caso de cargas altas. Apenas no

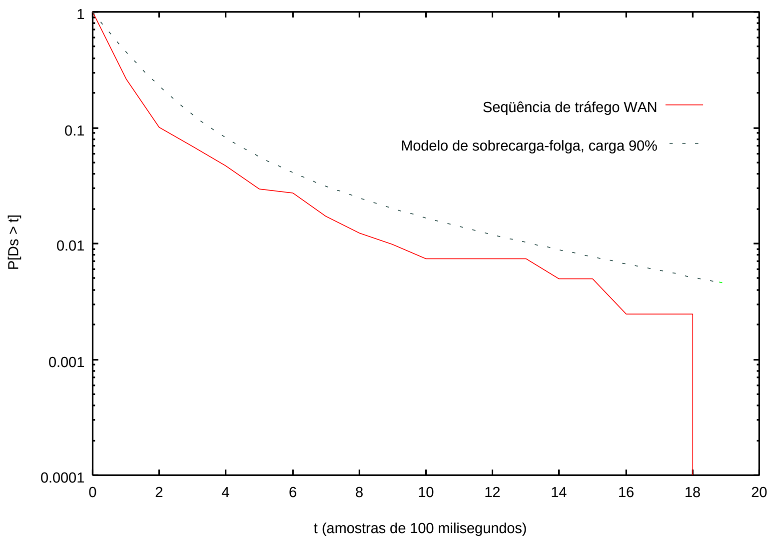


Figura 5.56: Distribuição das durações de sobrecarga para tráfego de saída, carga de 95%

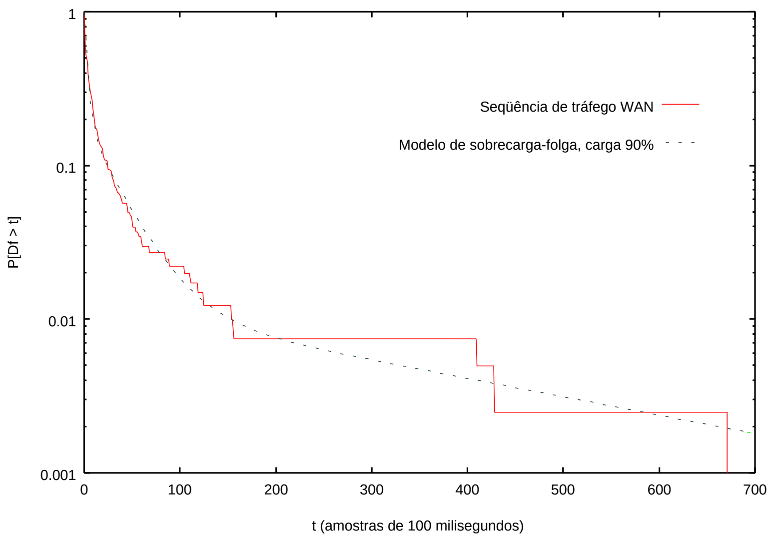


Figura 5.57: Distribuição das durações de folga para tráfego de saída, carga de 95%

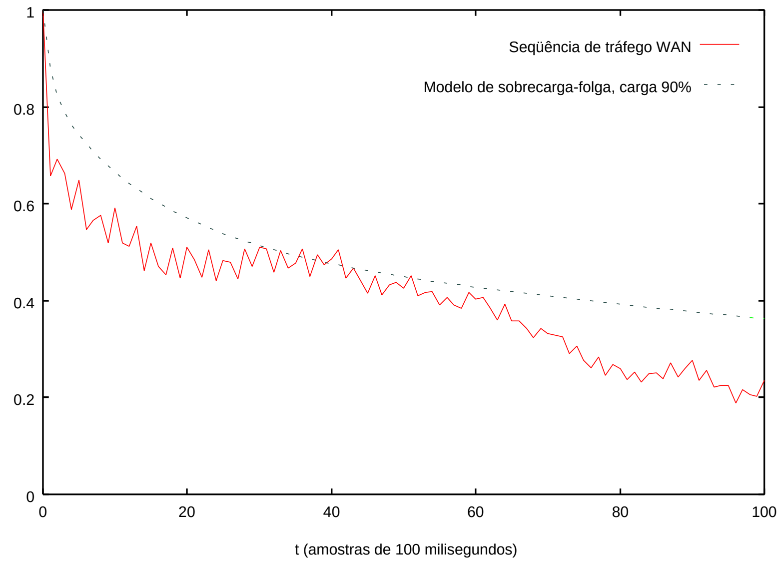


Figura 5.58: Função de autocorrelação para tráfego de saída e modelo de sobrecarga-folga com carga de 95%

exemplo usando carga de 95% é que pode-se notar alguma deficiência na aproximação da perda quando o tamanho máximo da fila ultrapassa 1000 pacotes.

Interessante notar que, como visto no capítulo 3, o casamento da autocorrelação para modelo e seqüência real não significa que a fração de perda terá casamento semelhante. Na Figura 5.54 verifica-se que as funções de autocorrelação são bem distintas embora as frações de perda para modelo e seqüência sejam bem casadas (ver Figura 5.52).

Os parâmetros dos modelos usados para os tráfegos WAN podem ser verificados no apêndice C.

5.7 Resumo e Conclusões do Capítulo

Este capítulo mostrou algumas características peculiares às seqüências de vídeo codificadas em MPEG4. Tais características conferem a este tipo de tráfego um grau de dificuldade relativamente alto no que diz respeito a um processo de modelagem visando a predição da perda. Maior ainda é a dificuldade quando se deseja fazê-lo para diversos tamanhos máximos de fila, desde algumas dezenas de pacotes até a ordem de milhares - o mesmo para diversos valores de carga. Também neste capítulo, fez-se um comentário geral sobre o perfil de tráfego em redes WAN relatando algumas considerações envolvendo a captura e modelagem de suas características.

Uma vez que objetiva-se aproximar satisfatoriamente a fração de perda, analisou-se aqui, de forma comparativa, o modelo de sobrecarga-folga aplicado a tráfegos MPEG4. Demais modelos usados para comparação foram escolhidos por serem eficientes na captura de um descritor específico, um para cada tipo de modelo: o modelo de histograma captura bem a distribuição de taxa, o modelo F-ARIMA captura bem a autocorrelação e o modelo pseudo auto-similar modificado captura bem a duração média em sobrecarga.

De uma maneira geral, os resultados para as seqüências MPEG4 aqui estudadas ratificam o que foi apresentado nos capítulos anteriores. Para cargas médias e baixas, os modelos de sobrecarga-folga têm bom desempenho. No entanto, este desempenho se degrada conforme as taxas em sobrecarga passam a ter uma variabilidade muito grande, o que ocorre a medida que a carga cresce.

O modelo pseudo auto-similar modificado possui distribuição exponencial para a duração de sobrecarga qualquer que seja a carga. Apesar dele apresentar distribuição hiper-exponencial para a duração de folga, esta característica deste modelo para as sobrecargas limita por demais sua eficácia na predição da perda, o que pode ser comprovado em todos os exemplos desta seção.

O modelo de histograma não leva em conta as distribuições de duração de folga e sobrecarga, que são importantes para o cálculo das medidas de interesse. Os resultados mostram sua ineficácia para prever a fração de perda em uma fila.

O modelo F-ARIMA mostrou-se mais eficiente em carga alta que em carga média, chegando a ser a melhor aproximação para as perdas em carga alta nos três exemplos para seqüências MPEG4. É importante notar que, para o caso de cargas altas, este modelo superestimou as frações de perda em diversas ordens de grandeza para tamanhos máximos de fila acima de um determinado valor. Para tamanhos máximos de fila altos, uma boa aproximação para a função de autocorrelação torna-se mais relevante onde os períodos de fila ocupada são maiores, conforme já observado no estudo realizado em [16].

Através dos exemplos realizados para as seqüências MPEG4, pode-se concluir que os descritores que melhor indicam a fração de perda na fila são a autocorrelação e a distribuição da duração de sobrecarga.

Além da análise comparativa para tráfegos de vídeo MPEG4, analisou-se também a eficiência do modelo de sobrecarga-folga para aproximar a perda envolvendo seqüências de tráfego WAN. Para tais casos, o modelo de sobrecarga-folga aproximou bem a perda mesmo para cargas altas. Vale ressaltar que para o tráfego agregado o descritor que melhor indicou a fração de perda na fila foi a distribuição da duração de sobrecarga.

Importante também dizer que, nos testes de perda, os tamanhos usados para as filas de pacotes foram estendidos até valores próximos daqueles onde não ocorrem mais perdas para a seqüência simulada. Tais valores são muito superiores aos que costumam ser praticados em situações reais. Geralmente, em interfaces seriais, ethernets, fast-ethernets ou ATM, as filas de saída não passam da ordem de 1000 pacotes, o que já é um valor alto se considerarmos uma interface com carga alta. Em casos como este, a fila permanece quase sempre cheia e um tamanho grande de fila implica em latências altas que degradam aplicações que utilizam voz e vídeo interativo, muito comuns atualmente. Também nos casos das aplicações e protocolos fim-a-fim para a transmissão interativa de vídeo e áudio, as filas nos receptores costumam ser limitadas de acordo com um retardo máximo que não comprometa a qualidade da conversação. Este valor é o mesmo usado para voz sobre IP e costuma ser igual a 200 milisegundos [69]. Numa transmissão de vídeo a 25 quadros/segundo, se considerarmos os valores de pico em pacotes/quadro das seqüências MPEG4 usadas nos testes, o limite máximo para estas filas não ultrapassaria 100 pacotes.

Capítulo 6

Conclusões e Direcionamentos Futuros

Este trabalho teve como foco principal propor um modelo markoviano com recompensas com o objetivo de representar o tráfego gerado tanto por uma aplicação de vídeo quanto o tráfego agregado fluindo sobre um canal de comunicação de dados.

O capítulo 1 motiva o trabalho desenvolvido. No capítulo 2 são apresentadas as estatísticas mais comumente usadas para caracterizar um tráfego, particularmente um tráfego de vídeo e um agregado, assim como um conjunto de modelos dentre os mais usualmente encontrados na literatura. Os modelos apresentados neste capítulo, em geral, tentam aproximar um ou dois descritores do tráfego modelado, por exemplo a função de autocorrelação e o histograma de taxas.

No capítulo 3 foram apresentados alguns modelos que, embora possuam descritores idênticos como a função de autocorrelação e o histograma de taxas, eles obtêm resultados bem distintos para a fração de perda em uma fila. Esta foi a motivação para o estudo de novos descritores capazes de capturar o comportamento desta medida de desempenho. Foram então definidos novos descritores baseados na distribuição da variável aleatória igual à duração dos períodos em que a fonte se encontra acima de um determinado nível de taxa. Como o tempo que a fonte permanece com a taxa acima de um determinado nível é uma variável que influencia fortemente a perda de pacotes em uma fila, investigamos dois descritores relacionados com este tempo. São eles: a distribuição da duração dos períodos em que a fonte se encontra acima de um certo nível (distribuição da duração da rajada) e a duração média destes períodos (duração média da rajada). Foram mostrados exemplos onde modelos com a mesma função de autocorrelação e o mesmo histograma de taxa podem gerar valores bem distintos para esses descritores. Foi possível relacionar os descritores com o comportamento da fila do modelo de desempenho correspondente (modelo fonte mais fila), ou seja, o modelo que possui probabilidade mais alta de gerar rajadas de maior duração é aquele que apresenta a maior fração de perda na fila.

Com base nos resultados obtidos no capítulo 3, foi proposto um modelo markoviano com recompensas no capítulo 4. Como o objetivo do modelo é o de representar as características dos descritores definidos no capítulo 3, e prever acuradamente a fração de perda em uma fila, o nível associado à fonte para definição dos períodos de rajada foi escolhido como sendo igual à taxa de serviço desta fila. O modelo foi então parametrizado usando três descritores obtidos da seqüência real: a distribuição da duração das sobrecargas, a distribuição da duração das folgas e a distribuição das taxas em cada um desses períodos. O modelo proposto possui dois subconjuntos de estados: um onde é gerado tráfego acima da taxa de serviço da fila associada (estados de sobrecarga) e o outro onde é gerado tráfego abaixo da taxa de serviço da fila (estados de folga). Os parâmetros dos estados de cada um desses subconjuntos são obtidos através da aproximação da distribuição da duração das sobrecargas e das folgas por distribuições hiper-exponenciais. Foi proposto um algoritmo para a obtenção dos parâmetros da distribuição hiper-exponencial a partir da distribuição dessas durações obtidas da seqüência sendo modelada. Dentre as formas apresentadas para associação de recompensas aos estados, a que obteve melhor desempenho e menor custo computacional foi aquela em que uma recompensa diferente é associada a cada estado do modelo (modelo com múltiplas recompensas). No capítulo 4 é realizada ainda uma breve análise da sensibilidade do modelo proposto frente aos parâmetros usados em sua construção. Desta análise fica evidente que, dentre os possíveis parâmetros relacionados à proposta de modelagem, a distribuição de taxas em sobrecarga e a cauda da distribuição da duração das sobrecargas são aqueles que mais influenciam a fração de perda resultante.

O capítulo 5 foi destinado à apresentação dos resultados dos diversos experimentos realizados. Foram usados como exemplo tráfegos de vídeo e tráfegos de redes WAN. A escolha de tráfegos de vídeo como forma de demonstrar a eficiência do modelo proposto deve-se à reconhecida dificuldade que existe para se obter bons resultados com estas seqüências. Isto se justifica uma vez que estas se caracterizam por apresentarem uma variabilidade muito grande da taxa e por suas funções de autocorrelação apresentarem dependência de longa duração. Por outro lado, a utilização de tráfegos de redes WAN mostra que o modelo é útil como ferramenta de dimensionamento, além do fato desta classe de tráfego ser de grande interesse para pesquisadores e engenheiros de redes.

A partir dos resultados obtidos para as seqüências MPEG4, pôde-se concluir que, para cargas médias e baixas, os modelos de sobrecarga-folga têm bom desempenho. No entanto, este desempenho se degrada conforme as taxas em sobrecarga passam a ter uma variabilidade muito grande, o que ocorre a medida que a carga cresce. Os modelos pseudo auto-similar e de histograma se mostraram ineficazes para prever a fração de perda na fila. Uma das razões para este comportamento é o fato destes modelos não capturarem a distribuição da duração das sobrecargas. O modelo F-ARIMA mostrou-se mais eficiente em carga alta do que em carga média. É importante notar que, para o caso de cargas altas, este modelo superestimou as frações de perda em diversas ordens de grandeza para tamanhos máximos de fila acima de um determinado valor. Através dos exemplos realizados para as seqüências MPEG4, pôde-se concluir que os descritores mais importantes a serem levados em consideração para prever a fração de perda na fila são a autocorrelação e a distribuição da duração de sobrecarga. Nos testes realizados com as seqüências de tráfego WAN o modelo de sobrecarga-folga aproximou bem a perda mesmo para cargas altas. Vale ressaltar que para o tráfego agregado o descritor mais importante para prever a fração de perda na fila foi a distribuição da duração de sobrecarga.

6.1 Resumo das Contribuições da Tese

As contribuições decorrentes do trabalho de tese aqui apresentado podem ser resumidas nos ítens a seguir. A relação segue as etapas percorridas durante o desenvolvimento da tese:

- Foi mostrado que tráfegos, ou modelos de tráfego, com mesma distribuição de taxa e mesma função de autocorrelação podem ser estatísticamente diferenciados por descritores que associam as taxas de transmissão com as durações dos períodos em que estas se encontram acima de um determinado nível;
- Foram definidos dois descritores relacionados com a duração dos períodos em que a taxa da fonte está acima ou abaixo de um determinado nível. O nível considerado foi a taxa de serviço da fila. São eles a distribuição da duração das sobrecargas e a distribuição da duração das folgas;
- Elaborou-se um algoritmo para aproximar uma distribuição (obtida de uma seqüência real), com coeficiente de variação maior do que um, por uma distribuição hiper-exponencial de três estados e aplicou-se este método para capturar as distribuições de duração de sobrecarga e folga;
- Foi proposto um modelo markoviano de 6 estados, aqui chamado "modelo de *sobrecarga-folga*", capaz de capturar as distribuições das durações de sobrecarga e de folga e ainda associar recompensas de taxa de acordo com a distribuição das taxas em cada um desses períodos;
- Mostrou-se que o modelo proposto, uma cadeia de Markov de 6 estados com recompensas de taxa, é capaz de predizer a fração de perda para seqüências MPEG4 e de redes WAN para valores de carga na fila bastante usuais e uma vasta abrangência de tamanhos máximos de fila. Mostrou-se também que as distribuições de duração de sobrecarga têm grande relevância na predição da fração de perda.

6.2 Trabalhos Futuros

As propostas para trabalhos futuros podem ser agrupadas em duas direções: evolução do modelo proposto e aprimoramento do algoritmo definido para parametrizar a distribuição hiper-exponencial.

Outro trabalho subseqüente é o de realizar um maior número de testes com tráfegos agregados, e usar o modelo como base do dimensionamento do *backbone* de uma rede. Em outras palavras, determinar as capacidades dos enlaces de forma a satisfazer critérios de desempenho. Como o modelo proposto é um modelo markoviano simples, soluções analíticas eficientes são possíveis para a tarefa de dimensionamento.

Apêndice A

A Ferramenta de Trabalho TANGRAM II

A ferramenta computacional utilizada para a obtenção dos resultados mostrados ao longo deste trabalho de tese foi o **Tangram II** [49]. Esta ferramenta é desenvolvida e mantida pelo Laboratório de Modelagem, Análise e Desenvolvimento de Redes e Sistemas de Computação, o LAND, localizado na UFRJ [78] e vinculado ao Programa de Engenharia de Sistemas e Computação (PESC) do Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia (COPPE) [79].

O **Tangram II** é um ambiente que permite o desenvolvimento de modelos probabilísticos voltados para tráfegos de redes ou sistemas de computação. No entanto, as características genéricas dos métodos envolvidos com os processos de modelagem associados à ferramenta permite a construção de modelos com propósitos muito mais amplos que os citados.

O ambiente de modelagem da ferramenta pode ser dividido nos seguintes módulos:

- TGIF (Tangram Graphic Interface Facility): usado como interface gráfica para a construção de modelos, o que inclui uma linguagem de programação orientada a eventos, a interação entre objetos e a associação de desenhos aos objetos construídos;

- Simulador: usado para simular o modelo construído, o que inclui um simulador de eventos raros que se vale da técnica de *splitting/RESTART* e, dependendo de como for construído o modelo, um simulador de fluido;
- Gerador de cadeia de Markov: usado para gerar, a partir da descrição no TGIF, uma cadeia de Markov com recompensas de taxa e/ou impulso que seja representativa do modelo, quando este assim permitir;
- Métodos analíticos de solução: destinado a extrair soluções estacionárias e transientes para a cadeia de Markov gerada;
- Medidas de interesse: usado para obter medidas de interesse para a cadeia de Markov a partir dos resultados obtidos com os métodos analíticos de solução;
- Descritores de tráfego: caso o modelo construído seja um modelo de tráfego, este módulo é usado para extrair descritores de tráfego a partir da cadeia de Markov gerada e também para seqüências reais que contenham amostras a intervalos fixos de tempo.

Dentre os métodos analíticos disponíveis para soluções envolvendo cadeias de Markov tem-se:

- métodos exatos e iterativos para obtenção das probabilidades estacionárias dos estados;
- métodos para obtenção das probabilidades transientes dos estados;
- métodos para a obtenção de distribuições relacionadas com as recompensas de taxa e de impulso;
- métodos envolvendo a obtenção de distribuições para medidas de tempo operacional.

Paralelo a este ambiente, o **Tangram II** disponibiliza uma ferramenta para conversação usando voz sobre IP, chamada *Vivavoz*, e uma versão do TGIF, chamada TGWB, para ser usada como "quadro-branco". Ambas ferramentas também foram desenvolvidas no LAND e cada qual tem características particulares: o *Vivavoz* implementa um algoritmo de FEC desenvolvido no LAND [80] e o TGWB possui suporte a multicast coerente [81]. Com isto, pode-se dizer que o **Tangram II** foi concebido para ser uma ferramenta com suporte à trabalho colaborativo.

Também em paralelo, há um ambiente para geração de tráfego, unicast ou multicast, com suporte a protocolos IP e ATM.

O **Tangram II** é gratuitamente distribuído segundo a licença GPL, *Gnu Public License*, e pode ser obtido em http://www.land.ufrj.br. Manuais do **Tangram II** também podem ser lá obtidos.

A.1 Telas do Tangram II

A figura A.1 mostra a tela principal do **Tangram II**, onde é possível ver os módulos citados: o ambiente de modelagem, o *vivavoz*, o "quadro-branco" e o gerador de tráfego.

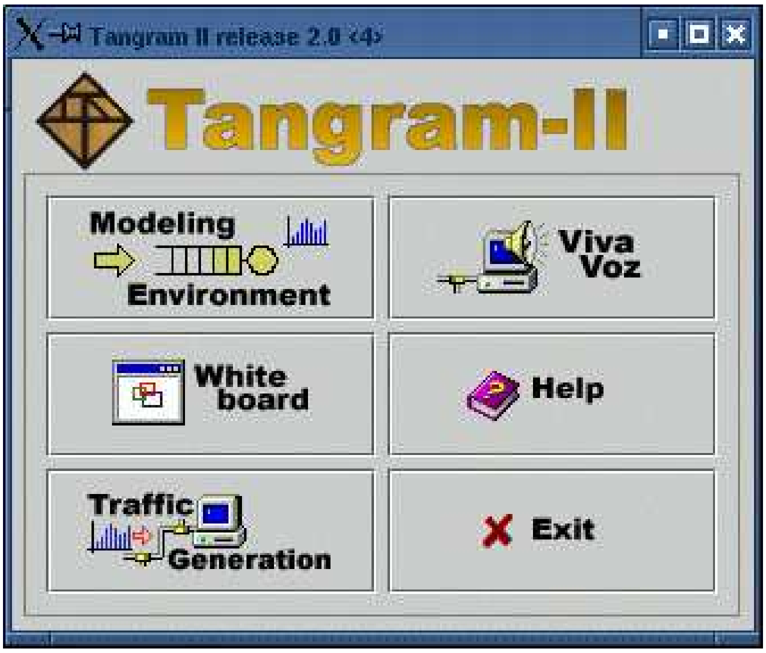


Figura A.1: Menu principal do grupo de ferramentas associadas ao **Tangram II**

A figura A.2 mostra o ambiente de modelagem. No canto inferior direito vê-se a descrição do arquivo carregado e no inferior esquerdo o diretório deste. A coluna à esquerda mostra os botões para os módulos citados. De cima para baixo, tem-se o

TGIF, o gerador de cadeias de Markov, o módulo contendo os métodos para solução analítica de cadeias de Markov, o módulo para cálculo de medidas de interesse para as soluções a partir de cadeias de Markov, o módulo para descritores de tráfego e o simulador.

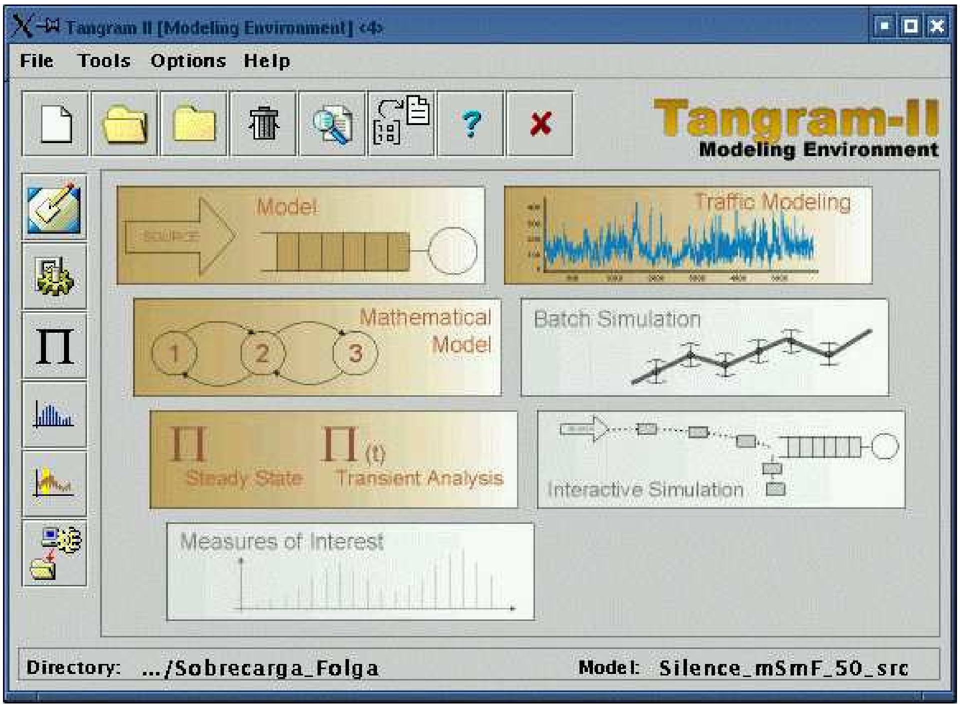


Figura A.2: Janela principal para o ambiente de modelagem do **Tangram II**

As figuras de A.3 a A.8 mostram exemplos de telas para estes respectivos módulos. Maiores detalhes podem ser obtidos em [78].

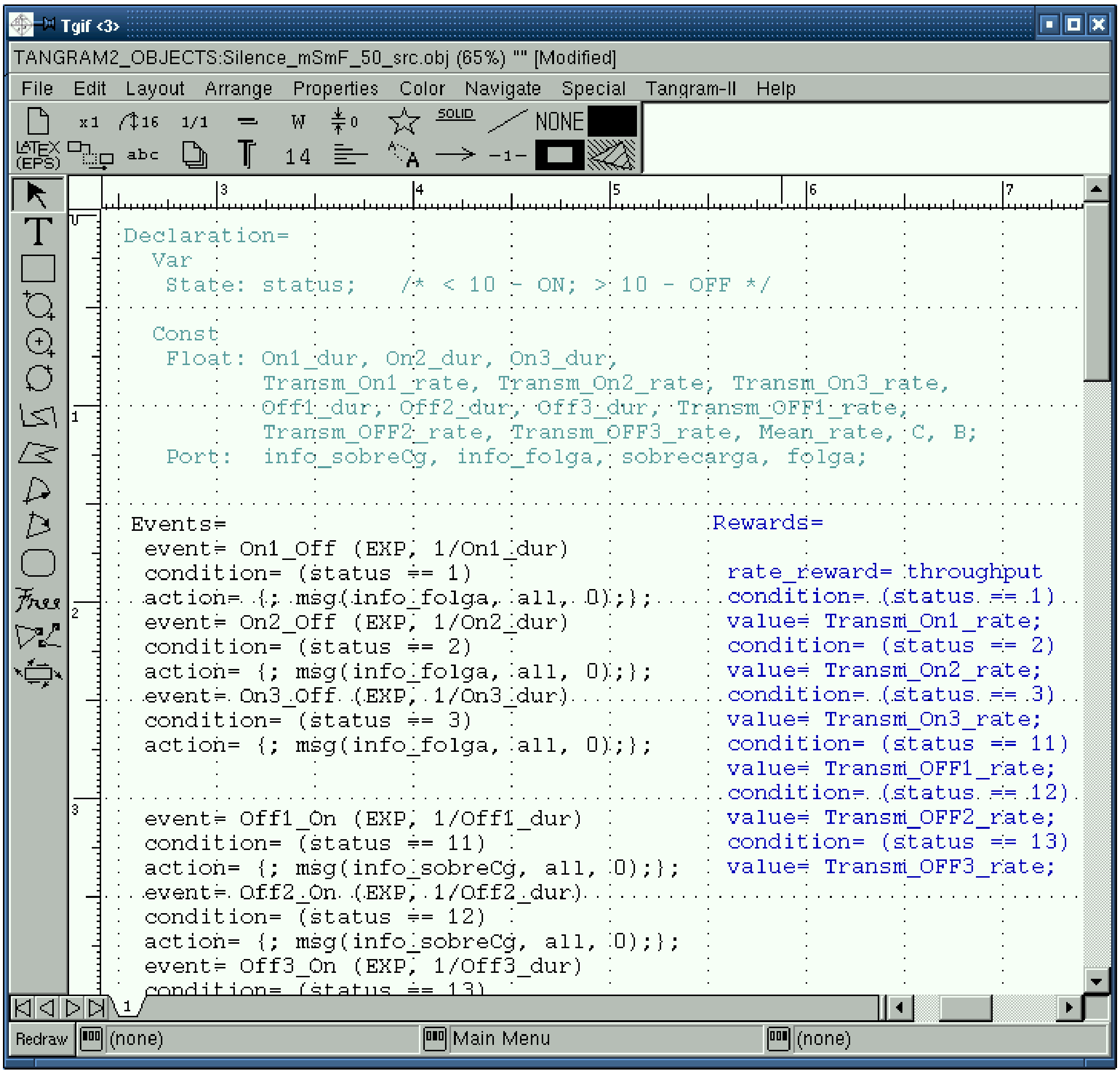


Figura A.3: Janela do TGIF

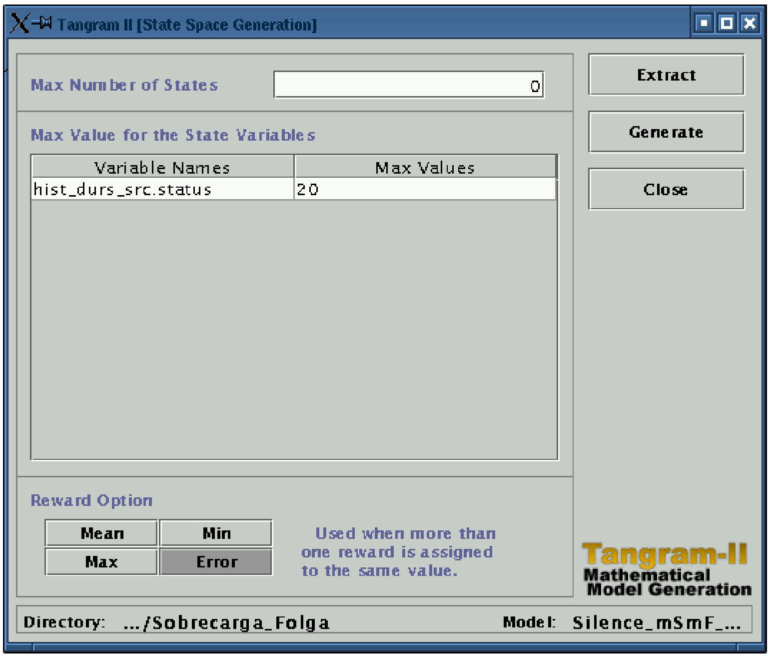


Figura A.4: Janela do gerador de cadeia de Markov

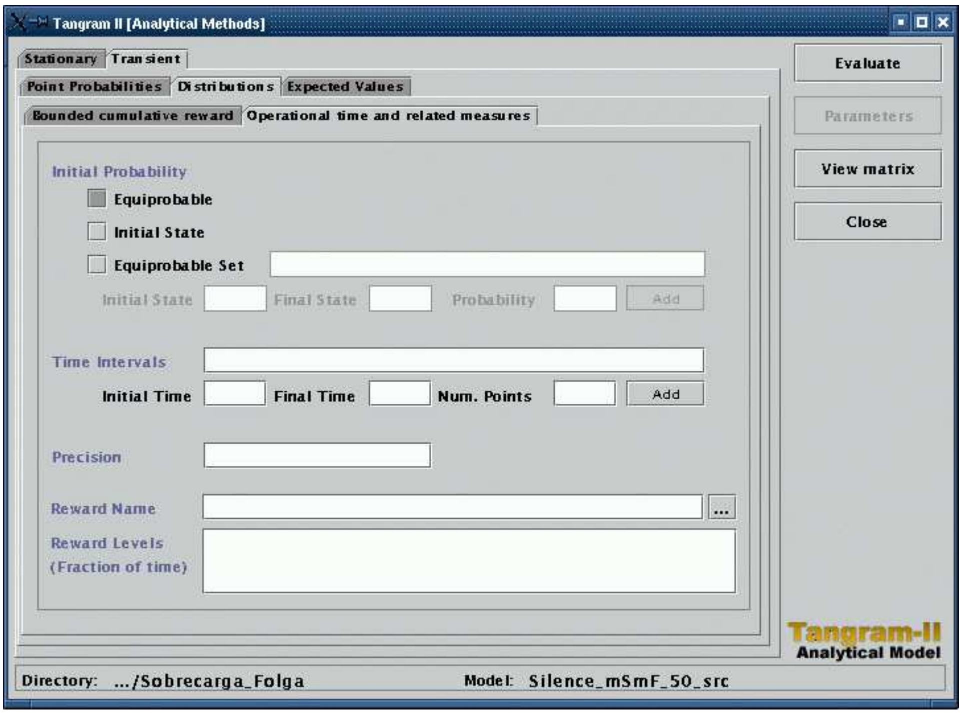


Figura A.5: Janela do ambiente de soluções analíticas para cadeias de Markov

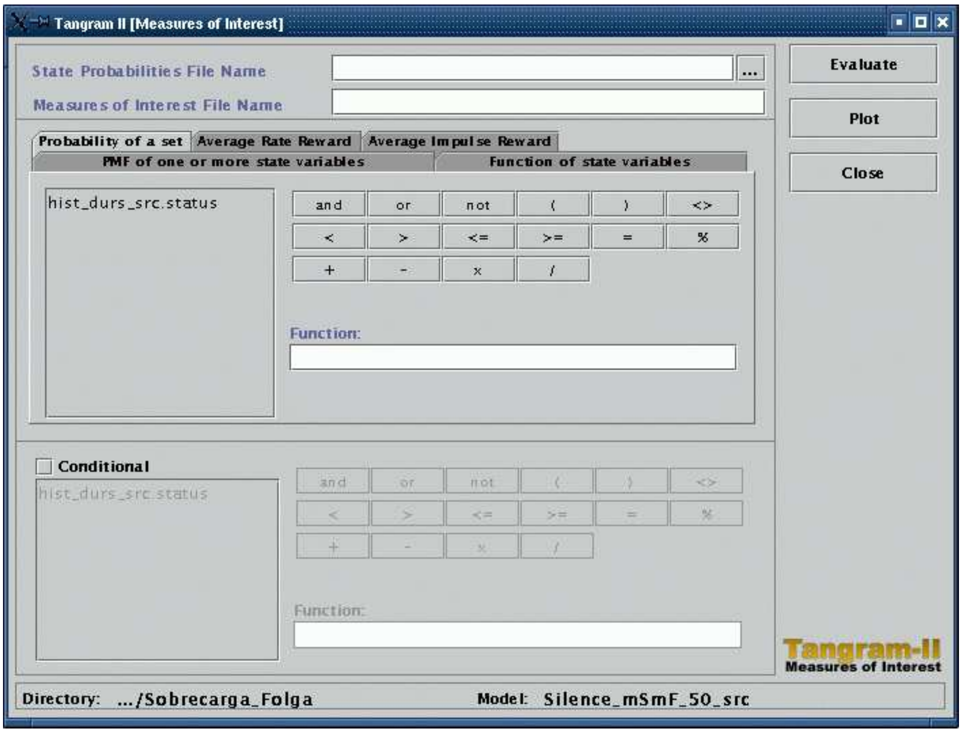


Figura A.6: Janela do módulo para cálculo de medidas de interesse

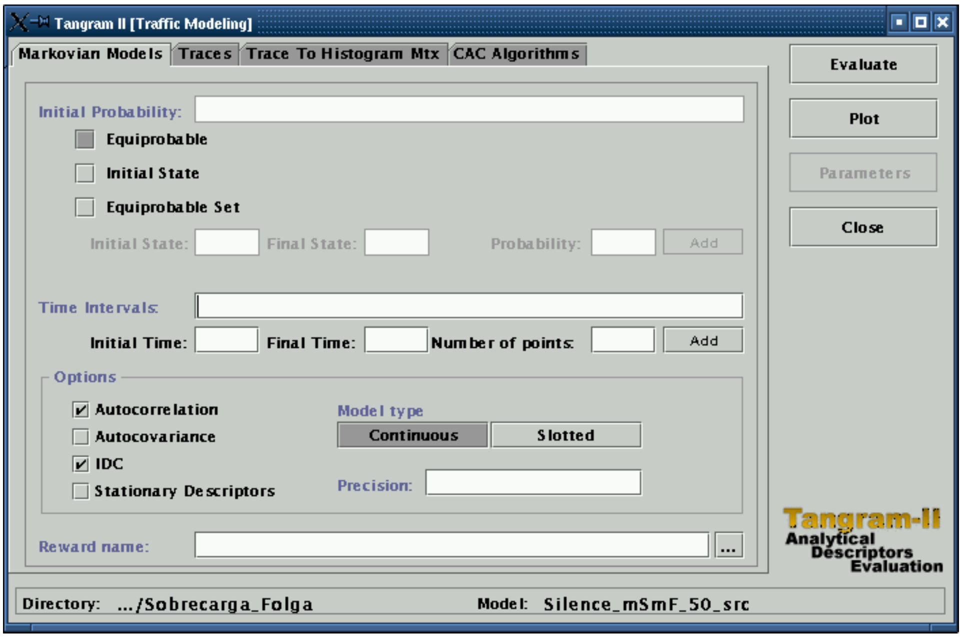


Figura A.7: Janela para o ambiente de cálculo de descritores de tráfego

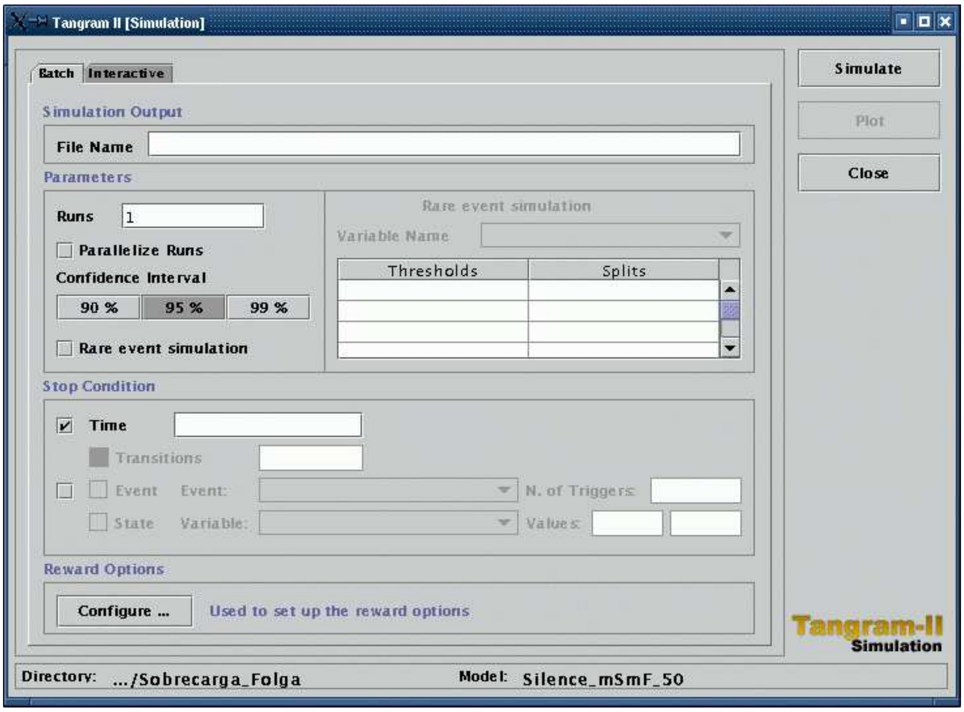


Figura A.8: Janela do simulador

Apêndice B

Apêndice B - Implementação em TANGRAM II de Modelos usados nos Testes

Neste apêndice são mostradas as implementações em Tangram II para os modelos de histograma, F-ARIMA, pseudo auto-similar modificado e sobrecarga-folga. No caso, a título de exemplo, estarão apenas sendo usados os modelos parametrizados para a seqüência do filme *Silence of the Lambs* codificada em MPEG4.

A figura B.1 mostra como a seqüência real é simulada para obtenção da fração de perdas. Ao final da simulação contabilizam-se os pacotes perdidos e os enviados para se efetuar o cálculo. Neste exemplo, o arquivo chamado "Silence_pcktrace" contém, em sintaxe definida para o **Tangram II**, os intervalos de tempo para cada chegada de pacote da seqüência real. Para este tipo de simulação, a fila é explicitamente modelada (objeto "Server").

A figura B.2 mostra o modelo F-ARIMA aplicado a uma fila determinística. Este modelo "fonte + fila" é orientado a simulação - a cada rodada um caminho amostral representativo do modelo F-ARIMA é gerado e usado para alimentar a fila indicada. É também possível gerar os vários caminhos amostrais em separado, tantos quantas forem as rodadas, e depois usar a forma indicada na figura B.1 para simular o processo.

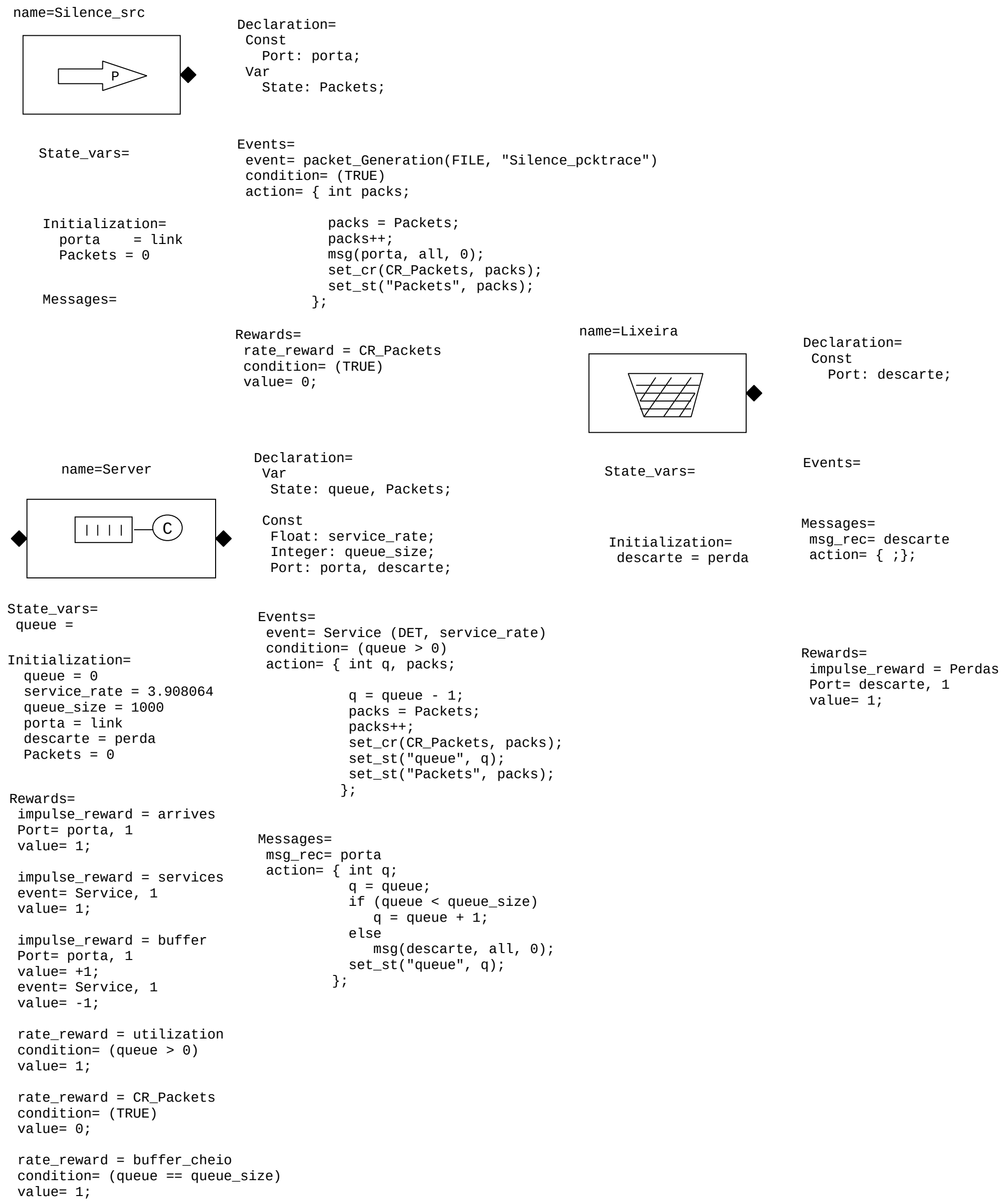


Figura B.1: Modelo de fonte + fila para seqüência real

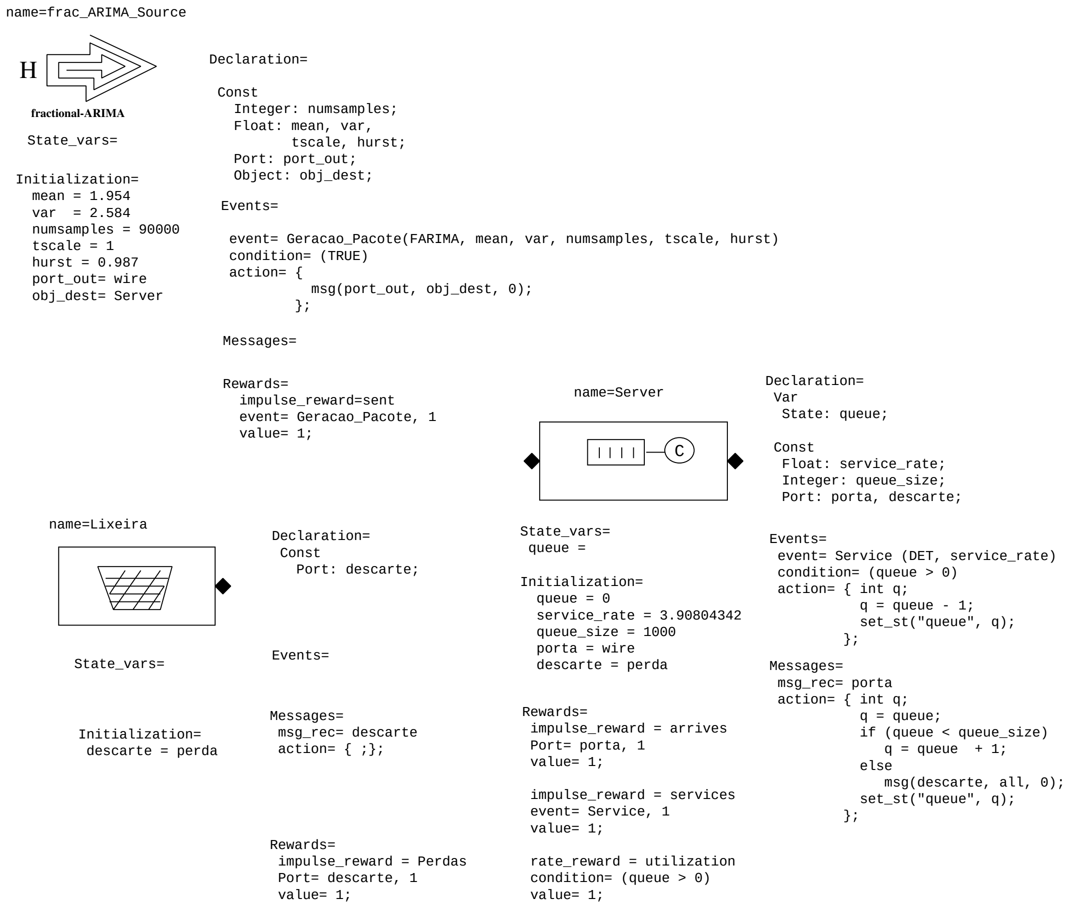


Figura B.2: Modelo de fonte + fila para fonte F-ARIMA

A figura B.3 mostra o modelo pseudo auto-similar modificado com a representação da fila através da recompensa "queue". Ao final da simulação, tem-se a fração de tempo na qual esta recompensa esteve maior ou igual ao valor "B", que indica o tamanho máximo da fila. De posse deste valor, para se obter a fração de pacotes perdidos (FPP), basta usar a expressão

$$FPP = \frac{(\text{Transm\_ON\_rate} - C)FT}{\overline{\lambda}} ,$$

onde "Transm_ON_rate" representa a taxa média em sobrecarga, C é a taxa de serviço da fila, $\overline{\lambda}$ é a taxa média da fonte e FT é a fração de tempo onde "queue" é maior ou igual a B.

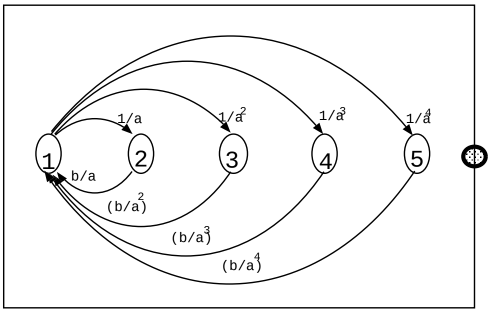
A figura B.4 mostra o modelo de sobrecarga-folga, com múltiplas recompensas de taxa para sobrecarga e folga, mais a representação da fila. No caso desta representação, as operações com as recompensas permitem obter o valor da fração de perda ao final da simulação, bastando observar o valor médio no tempo para o total acumulado pela recompensa "Perda". As recompensas "Perda_1", "Perda_2" e "Perda_3" descrevem as perdas para cada estado de sobrecarga e servem apenas para fins didáticos e de depuração.

Este tipo de descrição para modelo de "fonte + fila" se aplica à simulação por fluido. Pode-se notar que, a cada evento, as taxas que descrevem a recompensa de fluido são modificadas. Dois recursos em especial foram usados nesta modelagem:

- o evento do tipo "REWARD_REACHED", que ocorre sempre que o valor acumulado de uma recompensa ("queue", no caso) cresce até atingir um determinado valor ("B", no caso);
- o objeto "reflector", usado apenas para que o objeto modelando a "fonte + fila" possa mandar mensagens para si mesmo.

Outras implementações para este modelo são também possíveis.

A figura B.5 mostra o modelo de histograma para a fonte de tráfego, sem incluir as recompensas que descrevem a fila e instruções associadas. Pode-se reparar que, para cada estado, são descritas as probabilidades de transição para os demais name=PseudoSS_src



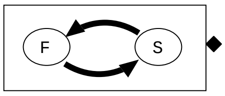

```
Declaration=
 Var
  State: stage;
 Const
  Float: a, b,
         Transm_ON_rate, Transm_OFF_rate,
         B, C;
```

```
State_vars=
  stage=

Initialization=
stage = 1
a = 3.17665
b = 0.80828948412
Transm_ON_rate  = 5.274457
Transm_OFF_rate = 1.479957

B = 1000
C = 3.908064

Messages=


Rewards=
rate_reward= pacotes
condition= (stage == 1)
value= Transm_ON_rate;
condition= (stage != 1)
value= Transm_OFF_rate;

rate_reward= queue
bounds= 0, B
condition= (stage == 1)
value= Transm_ON_rate - C;
condition= (stage != 1)
value= Transm_OFF_rate - C;

rate_reward= not_OpTime
condition= (stage > 1)
value= 1;
```

```
Events=
 event= up_trans_1_2 (EXP, (1/a))
 condition = (stage == 1)
 action= { ; set_st("stage", 2); };

 event= up_trans_1_3 (EXP, (1/a)*(1/a))
 condition = (stage == 1)
 action= { ; set_st("stage", 3); };

 event= up_trans_1_4 (EXP, (1/a)*(1/a)*(1/a))
 condition = (stage == 1)
 action= { ; set_st("stage", 4); };

 event= up_trans_1_5 (EXP, (1/a)*(1/a)*(1/a)*(1/a))
 condition = (stage == 1)
 action= { ; set_st("stage", 5); };

 event= down_trans_2_1 (EXP, (b/a))
 condition = (stage == 2)
 action= { ; set_st("stage", 1); };

 event= down_trans_3_1 (EXP, (b/a)*(b/a))
 condition = (stage == 3)
 action= { ; set_st("stage", 1); };

 event= down_trans_4_1 (EXP, (b/a)*(b/a)*(b/a))
 condition = (stage == 4)
 action= { ; set_st("stage", 1); };

 event= down_trans_5_1 (EXP, (b/a)*(b/a)*(b/a)*(b/a))
 condition = (stage == 5)
 action= { ; set_st("stage", 1); };
```

Figura B.3: Modelo de fonte + fila para fonte pseudo auto-similar modificada

```
name=hiper_durs_src
```




```
Declaration=
  Var
   State: status;   /* < 10 - ON; > 10 - OFF */

  Const
   Float: On1_dur, On2_dur, On3_dur,
          Transm_On1_rate, Transm_On2_rate, Transm_On3_rate,
          Off1_dur, Off2_dur, Off3_dur, Transm_OFF1_rate,
          Transm_OFF2_rate, Transm_OFF3_rate, Mean_rate, C, B;
   Port:  info_sobreCg, info_folga, sobrecarga, folga;
```

```
State_vars=
  status =

Initialization=
status = 1

On1_dur = 1.324089
On2_dur = 34.08621
On3_dur = 287

Transm_On1_rate = 4.574057
Transm_On2_rate = 5.257967
Transm_On3_rate = 7.099497

Off1_dur = 8.618437
Off2_dur = 534.2979
Off3_dur = 4991

Mean_rate = 1.9540211170

Transm_OFF1_rate = 1.229061
Transm_OFF2_rate = 1.479592
Transm_OFF3_rate = 1.164706

info_sobreCg = aviso_sobreCg
info_folga   = aviso_folga
sobrecarga   = seta_sobreCg
folga        = seta_folga

C = 3.908064
B = 1000
```

```
Events=
 event= On1_Off (EXP, 1/On1_dur)
 condition= (status == 1)
 action= {; msg(info_folga, all, 0);};
 event= On2_Off (EXP, 1/On2_dur)
 condition= (status == 2)
 action= {; msg(info_folga, all, 0);};
 event= On3_Off (EXP, 1/On3_dur)
 condition= (status == 3)
 action= {; msg(info_folga, all, 0);};


 event= Off1_On (EXP, 1/Off1_dur)
 condition= (status == 11)
 action= {; msg(info_sobreCg, all, 0);};
 event= Off2_On (EXP, 1/Off2_dur)
 condition= (status == 12)
 action= {; msg(info_sobreCg, all, 0);};
 event= Off3_On (EXP, 1/Off3_dur)
 condition= (status == 13)
 action= {; msg(info_sobreCg, all, 0);};


 event= Buffer_Full(REWARD_REACHED)
 condition= (get_cr(queue) /\ B)
 action= {  float ir;
            switch (status)
            {
               case 1: ir = (Transm_On1_rate - C)/Mean_rate;
                       set_ir(Perda, ir);
                       set_ir(Perda_1, ir);
                       break;
               case 2: ir = (Transm_On2_rate - C)/Mean_rate;
                       set_ir(Perda, ir);
                       set_ir(Perda_2, ir);
                       break;
               case 3: ir = (Transm_On3_rate - C)/Mean_rate;
                       set_ir(Perda, ir);
                       set_ir(Perda_3, ir);
                       break;
            }
        };
```

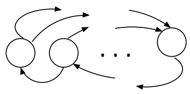
```
Rewards=

 rate_reward= pacotes
 condition= (TRUE)
 value= Transm_On1_rate;

 rate_reward= queue
 bounds= 0, B
 condition= (FALSE)
 value= 0;

 rate_reward= Perda
 condition= (FALSE)
 value= 0;

 rate_reward= Perda_1
 condition= (FALSE)
 value= 0;

 rate_reward= Perda_2
 condition= (FALSE)
 value= 0;

 rate_reward= Perda_3
 condition= (FALSE)
 value= 0;
```

```
Messages=
msg_rec = sobrecarga
action = {  float ir;
            set_ir(pacotes, Transm_On1_rate);
            ir = Transm_On1_rate - C;
            set_ir(queue, ir);
            set_st("status", 1);
         }
         : prob = 0.9869013;
         {  float ir;
            set_ir(pacotes, Transm_On2_rate);
            ir = Transm_On2_rate - C;
            set_ir(queue, ir);
            set_st("status", 2);
         }
         : prob = 0.0113392;
         {  float ir;
            set_ir(pacotes, Transm_On3_rate);
            ir = Transm_On3_rate - C;
            set_ir(queue, ir);
            set_st("status", 3);
         }
         : prob = 0.001759531;

msg_rec = folga
action = {  float ir;
            set_ir(pacotes, Transm_OFF1_rate);
            ir = Transm_OFF1_rate - C;
            set_ir(queue, ir);
            set_ir(Perda, 0);
            set_ir(Perda_1, 0);
            set_ir(Perda_2, 0);
            set_ir(Perda_3, 0);
            set_st("status", 11);
         }
         : prob = 0.9904203;
         {  float ir;
            set_ir(pacotes, Transm_OFF2_rate);
            ir = Transm_OFF2_rate - C;
            set_ir(queue, ir);
            set_ir(Perda, 0);
            set_ir(Perda_1, 0);
            set_ir(Perda_2, 0);
            set_ir(Perda_3, 0);
            set_st("status", 12);
         }
         : prob = 0.009188661;
         {  float ir;
            set_ir(pacotes, Transm_OFF3_rate);
            ir = Transm_OFF3_rate - C;
            set_ir(queue, ir);
            set_ir(Perda, 0);
            set_ir(Perda_1, 0);
            set_ir(Perda_2, 0);
            set_ir(Perda_3, 0);
            set_st("status", 13);
         }
         : prob = 0.0003910068;
```

```
name=reflector

 State_vars=


 Events=


 Rewards=


 Declaration=
 Const
  Port: port_sobrecarga, port_folga,
        sign_sobrecarga, sign_folga;

Messages=
msg_rec = port_sobrecarga
action = {; msg(sign_sobrecarga, all, 0);};

msg_rec = port_folga
action = {; msg(sign_folga, all, 0);};

Initialization=

port_sobrecarga = aviso_sobreCg
port_folga = aviso_folga
sign_sobrecarga = seta_sobreCg
sign_folga = seta_folga
```

Figura B.4: Modelo de fonte + fila para fonte de sobrecarga-folga estados. De posse deste modelo, gera-se a cadeia de Markov associada e obtém-se os descritores de tráfego. A variável "Op_1st" serve para o cálculo das medidas de tempo operacional, usadas para se calcular a distribuição das durações de sobrecarga (recompensa "op_time") e de folga (recompensa "not_op_time"). A implementação da fila se dá nos mesmos moldes usados para o modelo de sobrecarga-folga. Sua descrição foi omitida por questões de espaço.

```
name=histogram_src
```



```
Declaration=
Var
 State: state;

Const
 Float: Tx_1, Tx_2, Tx_3, Tx_4, Tx_5, Tx_6, Tx_7, Tx_8;
 Integer: Op_1st;
```

```
 State_vars=

 Initialization=
  state    = 1

  Tx_1 =  0.848656
  Tx_2 =  2.335312
  Tx_3 =  4.339766
  Tx_4 =  6.355217
  Tx_5 =  8.342623
  Tx_6 = 10.508197
  Tx_7 = 12.276730
  Tx_8 = 14.000000

  Op_1st = 2


Messages=


Rewards=

rate_reward= throughput
 condition= (state == 1)
 value= Tx_1;
 condition= (state == 2)
 value= Tx_2;
 condition= (state == 3)
 value= Tx_3;
 condition= (state == 4)
 value= Tx_4;
 condition= (state == 5)
 value= Tx_5;
 condition= (state == 6)
 value= Tx_6;
 condition= (state == 7)
 value= Tx_7;
 condition= (state == 8)
 value= Tx_8;

rate_reward= op_time
 condition= (state >= Op_1st)
 value= 1;

rate_reward= not_op_time
 condition= (state < Op_1st)
 value= 1;
```

```
Events=
 event= transition_1 (EXP, 1.000000)
 condition= (state == 1)
 action=
   {; set_st("state", 1);} : prob= 0.8442296633;
   {; set_st("state", 2);} : prob= 0.1446233468;
   {; set_st("state", 3);} : prob= 0.0105719467;
   {; set_st("state", 4);} : prob= 0.0004865750;
   {; set_st("state", 5);} : prob= 0.0000884682;

 event= transition_2 (EXP, 1.000000)
 condition= (state == 2)
 action=
   {; set_st("state", 1);} : prob= 0.1953052763;
   {; set_st("state", 2);} : prob= 0.7258925643;
   {; set_st("state", 3);} : prob= 0.0667223432;
   {; set_st("state", 4);} : prob= 0.0105884809;
   {; set_st("state", 5);} : prob= 0.0013720285;
   {; set_st("state", 6);} : prob= 0.0000894801;
   {; set_st("state", 8);} : prob= 0.0000298267;

 event= transition_3 (EXP, 1.000000)
 condition= (state == 3)
 action=
   {; set_st("state", 1);} : prob= 0.0622668210;
   {; set_st("state", 2);} : prob= 0.2885629744;
   {; set_st("state", 3);} : prob= 0.5713366783;
   {; set_st("state", 4);} : prob= 0.0638106265;
   {; set_st("state", 5);} : prob= 0.0104206870;
   {; set_st("state", 6);} : prob= 0.0012865046;
   {; set_st("state", 7);} : prob= 0.0023157082;

 event= transition_4 (EXP, 1.000000)
 condition= (state == 4)
 action=
   {; set_st("state", 1);} : prob= 0.0047826087;
   {; set_st("state", 2);} : prob= 0.1569565217;
   {; set_st("state", 3);} : prob= 0.2200000000;
   {; set_st("state", 4);} : prob= 0.5552173914;
   {; set_st("state", 5);} : prob= 0.0560869565;
   {; set_st("state", 6);} : prob= 0.0056521739;
   {; set_st("state", 7);} : prob= 0.0013043478;

 event= transition_5 (EXP, 1.000000)
 condition= (state == 5)
 action=
   {; set_st("state", 1);} : prob= 0.0016393443;
   {; set_st("state", 2);} : prob= 0.0737704918;
   {; set_st("state", 3);} : prob= 0.1409836066;
   {; set_st("state", 4);} : prob= 0.2163934426;
   {; set_st("state", 5);} : prob= 0.4885245902;
   {; set_st("state", 6);} : prob= 0.0655737705;
   {; set_st("state", 7);} : prob= 0.0131147541;

 event= transition_6 (EXP, 1.000000)
 condition= (state == 6)
 action=
   {; set_st("state", 2);} : prob= 0.0065573770;
   {; set_st("state", 3);} : prob= 0.0327868852;
   {; set_st("state", 4);} : prob= 0.0295081967;
   {; set_st("state", 5);} : prob= 0.1475409836;
   {; set_st("state", 6);} : prob= 0.6721311476;
   {; set_st("state", 7);} : prob= 0.0983606557;
   {; set_st("state", 8);} : prob= 0.0131147541;

 event= transition_7 (EXP, 1.000000)
 condition= (state == 7)
 action=
   {; set_st("state", 3);} : prob= 0.0943396226;
   {; set_st("state", 4);} : prob= 0.0503144654;
   {; set_st("state", 5);} : prob= 0.0440251572;
   {; set_st("state", 6);} : prob= 0.1949685535;
   {; set_st("state", 7);} : prob= 0.6037735849;
   {; set_st("state", 8);} : prob= 0.0125786164;

 event= transition_8 (EXP, 1.000000)
 condition= (state == 8)
 action=
   {; set_st("state", 4);} : prob= 0.1000000000;
   {; set_st("state", 6);} : prob= 0.3000000000;
   {; set_st("state", 7);} : prob= 0.3000000000;
   {; set_st("state", 8);} : prob= 0.3000000000;
```

Figura B.5: Modelo de fonte usando histograma

Apêndice C

Apêndice C - Parâmetros dos Modelos usados nos Exemplos do Capítulo 5

Seguem abaixo os valores dos parâmetros de cada modelo citado no capítulo 5 de acordo com a seqüência modelada. Os parâmetros do modelo de sobrecarga-folga seguem a nomenclatura usada na figura 4.12. Os parâmetros do modelo pseudo auto-similar modificado seguem a nomenclatura usada na figura 5.8. As taxas de transição entre estados para o modelo de sobrecarga-folga foram escritas explicitando as durações médias de cada escala de tempo usada: taxa na escala i igual a $1/D_i$, onde D_i é a duração média na escala de tempo i.

C.1 Modelos para as Seqüências de Vídeo codificadas em MPEG4

Para cada seqüência foram utilizados os modelos de histograma, F-ARIMA, pseudo auto-similar modificado e de sobrecarga-folga. A unidade de tempo usada nos modelos é *quadros*, ou seja, o tempo de transmissão de um quadro. No caso das seqüências MPEG4 modeladas, este tempo é 1/25 segundos.

C.1.1 Seqüência do filme *Jurassic Park*

Modelo de histograma

$$\Lambda =< 0.8961 \;\; 2.0000 \;\; 3.3670 \;\; 5.0000 \;\; 6.0000 \;\; 7.2583 \;\; 9.0000 \;\; 10.3333 >$$

$$\mathbf{P} = \begin{bmatrix} 0.728275 & 0.182343 & 0.069375 & 0.013298 & 0.005490 & 0.001220 & 0 & 0 \\ 0.161728 & 0.632353 & 0.168176 & 0.021878 & 0.011084 & 0.004745 & 0.000036 & 0 \\ 0.065492 & 0.170056 & 0.663059 & 0.059447 & 0.027876 & 0.013845 & 0.000224 & 0 \\ 0.053216 & 0.107593 & 0.255968 & 0.488064 & 0.061340 & 0.032659 & 0.001160 & 0 \\ 0.041767 & 0.111495 & 0.260614 & 0.121850 & 0.416983 & 0.042458 & 0.004833 & 0 \\ 0.009605 & 0.073106 & 0.197439 & 0.107791 & 0.064034 & 0.530416 & 0.015009 & 0.001601 \\ 0.010417 & 0.010417 & 0.114583 & 0.093750 & 0.093750 & 0.281250 & 0.395833 & 0 \\ 0 & 0 & 0 & 0 & 0.666667 & 0.333333 & 0 & 0 \end{bmatrix}$$

Modelo F-ARIMA

$$\begin{aligned} \text{Média} &= 2.554 \\ \text{Variância} &= 2.434 \\ \text{Parâmetro de Hurst} &= 0.861 \end{aligned}$$

Modelo pseudo auto-similar modificado

- Carga de 50%

$$\begin{aligned} \text{a} &= 2.285710 \\ \text{b} &= 0.594073 \\ \lambda\text{s} &= 6.531950 \\ \lambda\text{f} &= 2.324581 \end{aligned}$$

- Carga de 90%

$$\begin{aligned} \text{a} &= 3.880770 \\ \text{b} &= 1.616721 \\ \lambda\text{s} &= 4.041643 \\ \lambda\text{f} &= 1.479751 \end{aligned}$$

Modelo de sobrecarga-folga

- Carga de 50%

$$\lambda s1 = 6.382306 \quad \lambda f1 = 2.384277$$
$$\lambda s2 = 6.726389 \quad \lambda f2 = 1.985940$$
$$\lambda s3 = 7.433190 \quad \lambda f3 = 2.262745$$

$$\alpha 1 = 1/1.040446 \quad \beta 1 = 1/9.764951$$
$$\alpha 2 = 1/7.578947 \quad \beta 2 = 1/455.9643$$
$$\alpha 3 = 1/232 \quad \beta 3 = 1/3805$$

$$p1 = 0.9736556 \quad q1 = 0.9763715$$
$$p2 = 0.0258012 \quad q2 = 0.0228137$$
$$p3 = 0.0005432 \quad q3 = 0.0008148$$

- Carga de 90%

$$\lambda s1 = 3.813488 \quad \lambda f1 = 1.673652$$
$$\lambda s2 = 3.495531 \quad \lambda f2 = 1.233059$$
$$\lambda s3 = 4.440636 \quad \lambda f3 = 1.029982$$

$$\alpha 1 = 1/1 \quad \beta 1 = 1/2.263571$$
$$\alpha 2 = 1/3.122456 \quad \beta 2 = 1/11.00155$$
$$\alpha 3 = 1/58.53312 \quad \beta 3 = 1/38.43226$$

$$p1 = 0.7422198 \quad q1 = 0.8401809$$
$$p2 = 0.2334815 \quad q2 = 0.1479381$$
$$p3 = 0.0242986 \quad q3 = 0.0118810$$

C.1.2 Seqüência do filme *Star Trek - First Contact*

Modelo de histograma

$$\Lambda =< 0.0000 \;\; 1.0000 \;\; 2.0000 \;\; 3.0000 \;\; 4.0000 \;\; 5.0000 \;\; 6.0000 \;\; 7.0769 >$$

$$\mathbf{P} = \begin{bmatrix} 0.558858 & 0.373136 & 0.049052 & 0.015238 & 0.003339 & 0.000376 & 0 & 0 \\ 0.167649 & 0.706962 & 0.095315 & 0.023803 & 0.005261 & 0.000947 & 0.000042 & 0.000021 \\ 0.072673 & 0.306058 & 0.545933 & 0.054504 & 0.017144 & 0.003005 & 0.000478 & 0.000205 \\ 0.062032 & 0.251123 & 0.169198 & 0.466524 & 0.039358 & 0.010695 & 0.000423 & 0.000642 \\ 0.040404 & 0.209957 & 0.159452 & 0.137806 & 0.419913 & 0.025253 & 0.006494 & 0.000722 \\ 0.011331 & 0.118980 & 0.155807 & 0.121813 & 0.099150 & 0.461756 & 0.031161 & 0 \\ 0 & 0.039216 & 0.137255 & 0.058824 & 0.215686 & 0.156863 & 0.392157 & 0 \\ 0 & 0.076923 & 0.153846 & 0.230769 & 0.153846 & 0 & 0 & 0.384615 \end{bmatrix}$$

Modelo F-ARIMA

$$\begin{aligned} &\text{Média} && = 1.097 \\ &\text{Variância} && = 0.821 \\ &\text{Parâmetro de Hurst} && = 0.975 \end{aligned}$$

Modelo pseudo auto-similar modificado

- Carga de 50%

$$\begin{aligned} &\text{a} && = 2.329560 \\ &\text{b} && = 0.655955 \\ &\lambda\text{s} && = 3.353103 \\ &\lambda\text{f} && = 0.920143 \end{aligned}$$

- Carga de 90%

$$\begin{aligned} &\text{a} && = 2.679450 \\ &\text{b} && = 1.086952 \\ &\lambda\text{s} && = 2.418060 \\ &\lambda\text{f} && = 0.691343 \end{aligned}$$

Modelo de sobrecarga-folga

- Carga de 50%

$$\lambda s1 = 3.273911 \quad \lambda f1 = 1.149131$$
$$\lambda s2 = 3.435897 \quad \lambda f2 = 0.825781$$
$$\lambda s3 = 3.934272 \quad \lambda f3 = 0.565541$$

$$\alpha 1 = 1/1.130277 \quad \beta 1 = 1/8.379248$$
$$\alpha 2 = 1/12.71739 \quad \beta 2 = 1/108.8309$$
$$\alpha 3 = 1/127.8 \quad \beta 3 = 1/799.5$$

$$p1 = 0.9892428 \quad q1 = 0.9371441$$
$$p2 = 0.0097026 \quad q2 = 0.0586374$$
$$p3 = 0.0010546 \quad q3 = 0.0042185$$

- Carga de 90%

$$\lambda s1 = 2.286801 \quad \lambda f1 = 0.727920$$
$$\lambda s2 = 2.489253 \quad \lambda f2 = 0.427966$$
$$\lambda s3 = 2.987614 \quad \lambda f3 = 0.415391$$

$$\alpha 1 = 1/1.290729 \quad \beta 1 = 1/4.305651$$
$$\alpha 2 = 1/24.12346 \quad \beta 2 = 1/26.94842$$
$$\alpha 3 = 1/116.9310 \quad \beta 3 = 1/112.3014$$

$$p1 = 0.9910794 \quad q1 = 0.9657773$$
$$p2 = 0.0065688 \quad q2 = 0.0283027$$
$$p3 = 0.0023518 \quad q3 = 0.0059200$$

C.1.3 Seqüência do filme *Silence of the Lambs*

Modelo de histograma

$$\Lambda =< 0.8487 \;\; 2.3353 \;\; 4.3398 \;\; 6.3552 \;\; 8.3426 \;\; 10.5082 \;\; 12.2767 \;\; 14.0000 >$$

$$\mathbf{P} = \begin{bmatrix}
0.844230 & 0.144623 & 0.010572 & 0.000487 & 0.000088 & 0 & 0 & 0 \\
0.195305 & 0.725893 & 0.066722 & 0.010588 & 0.001372 & 0.000089 & 0 & 0.000030 \\
0.062267 & 0.288563 & 0.571337 & 0.063811 & 0.010421 & 0.001287 & 0.002316 & 0 \\
0.004783 & 0.156957 & 0.220000 & 0.555217 & 0.056087 & 0.005652 & 0.001304 & 0 \\
0.001639 & 0.073770 & 0.140984 & 0.216393 & 0.488525 & 0.065574 & 0.013115 & 0 \\
0 & 0.006557 & 0.032787 & 0.029508 & 0.147541 & 0.672131 & 0.098361 & 0.013115 \\
0 & 0 & 0.094340 & 0.050314 & 0.044025 & 0.194969 & 0.603774 & 0.012579 \\
0 & 0 & 0 & 0.100000 & 0 & 0.300000 & 0.300000 & 0.300000
\end{bmatrix}$$

Modelo F-ARIMA

$$\begin{aligned}
\text{Média} &= 1.954 \\
\text{Variância} &= 2.584 \\
\text{Parâmetro de Hurst} &= 0.987
\end{aligned}$$

Modelo pseudo auto-similar modificado

- Carga de 50%

$$\begin{aligned}
a &= 3.176650 \\
b &= 0.808289 \\
\lambda s &= 5.274457 \\
\lambda f &= 1.479957
\end{aligned}$$

- Carga de 90%

$$\begin{aligned}
a &= 3.619480 \\
b &= 1.125424 \\
\lambda s &= 4.137482 \\
\lambda f &= 1.226927
\end{aligned}$$

Modelo de sobrecarga-folga

- Carga de 50%

λs1 = 4.574057 λf1 = 1.229061

λs2 = 5.257967 λf2 = 1.479592

λs3 = 7.099497 λf3 = 1.164706

$\alpha 1 = 1/1.324089$ $\beta 1 = 1/4.574057$

$\alpha 2 = 1/34.08621$ $\beta 2 = 1/5.257967$

$\alpha 3 = 1/287$ $\beta 3 = 1/7.099497$

p1 = 0.9869013 q1 = 0.9904203

p2 = 0.0113392 q2 = 0.0091887

p3 = 0.0017595 q3 = 0.0003910

- Carga de 90%

λs1 = 3.475027 λf1 = 1.461650

λs2 = 3.938272 λf2 = 0.985885

λs3 = 5.438183 λf3 = 0.762348

$\alpha 1 = 1/1.525697$ $\beta 1 = 1/4.784158$

$\alpha 2 = 1/37.12500$ $\beta 2 = 1/38.60791$

$\alpha 3 = 1/193.2222$ $\beta 3 = 1/223.8333$

p1 = 0.9873433 q1 = 0.9306223

p2 = 0.0084378 q2 = 0.0651588

p3 = 0.0042189 q3 = 0.0042189

C.2 Modelos para as Seqüências de Tráfego WAN em Enlace da Rede Abilene

Para estas seqüências, apenas modelos de sobrecarga-folga foram utilizados. A unidade de tempo usada é igual a 100 milisegundos. Isto significa dizer que um tempo de valor 1 para os modelos corresponde a 100 milisegundos do tráfego real.

C.2.1 Tráfego de entrada com carga de 90%

$$
\begin{array}{ll}
\lambda\mathrm{s}1 = 8975.657895 & \lambda\mathrm{f}1 = 8276.930219 \\
\lambda\mathrm{s}2 = 9062.918033 & \lambda\mathrm{f}2 = 8122.272727 \\
\lambda\mathrm{s}3 = 9197.589744 & \lambda\mathrm{f}3 = 7729.663479
\end{array}
$$

$$
\begin{array}{ll}
\alpha 1 = 1/1 & \beta 1 = 1/17.16832 \\
\alpha 2 = 1/2.904762 & \beta 2 = 1/286 \\
\alpha 3 = 1/11.14286 & \beta 3 = 1/1255
\end{array}
$$

$$
\begin{array}{ll}
\mathrm{p}1 = 0.7307692 & \mathrm{q}1 = 0.9619048 \\
\mathrm{p}2 = 0.2019231 & \mathrm{q}2 = 0.0095238 \\
\mathrm{p}3 = 0.0673077 & \mathrm{q}3 = 0.0285714
\end{array}
$$

C.2.2 Tráfego de saída com carga de 90%

$$
\begin{array}{ll}
\lambda\mathrm{s}1 = 8593.093750 & \lambda\mathrm{f}1 = 7831.896829 \\
\lambda\mathrm{s}2 = 8665 & \lambda\mathrm{f}2 = 7620.620032 \\
 & \lambda\mathrm{f}3 = 7533.225598
\end{array}
$$

$$
\begin{array}{ll}
\alpha 1 = 1/1 & \beta 1 = 1/67.84848 \\
\alpha 2 = 1/2 & \beta 2 = 1/629 \\
 & \beta 3 = 1/1547
\end{array}
$$

$$
\begin{array}{ll}
\mathrm{p}1 = 0.91428572 & \mathrm{q}1 = 0.9166667 \\
\mathrm{p}2 = 0.08571428 & \mathrm{q}2 = 0.0277778 \\
 & \mathrm{q}3 = 0.0555556
\end{array}
$$

C.2.3 Tráfego de saída com carga de 95%

$$
\begin{array}{ll}
\lambda\mathrm{s}1 = 8189.197987 & \lambda\mathrm{f}1 = 7824.056030 \\
\lambda\mathrm{s}2 = 8230.126829 & \lambda\mathrm{f}2 = 7654.282186 \\
\lambda\mathrm{s}3 = 8302.805195 & \lambda\mathrm{f}3 = 7369.647802
\end{array}
$$

$$\alpha 1 = 1/1 \qquad \beta 1 = 1/3.230061$$
$$\alpha 2 = 1/2.356322 \qquad \beta 2 = 1/33.37838$$
$$\alpha 3 = 1/8.052632 \qquad \beta 3 = 1/364$$

$$\mathrm{p1} = 0.7376238 \qquad \mathrm{q1} = 0.8049383$$
$$\mathrm{p2} = 0.2153465 \qquad \mathrm{q2} = 0.1827160$$
$$\mathrm{p3} = 0.0470297 \qquad \mathrm{q3} = 0.0123457$$

Referências Bibliográficas

[1] Vern Paxson and Sally Floyd. Wide Area Traffic: The Failure of Poisson Modeling. *IEEE/ACM Transactions on Networking*, 3(3):226–244, 1995.

[2] W. Leland, W. Willinger, M. Taqqu, and D. Wilson. On the self-similar nature of ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, 2(1):1–15, February 1994.

[3] Stephan Robert. *Modélisation Markovienne du Trafic dans les Réseaux de Communication*. PhD thesis, Ecole Polytechnique Fédérale de Lausanne, Switzerland, 3 1996.

[4] Oliver Rose. Simple and efficient models for variable bit rate MPEG video traffic. *Performance Evaluation*, 30(1-2):69–85, 1997.

[5] K. R. Krishnan and Gopalakrishnan Meempat. Long-range dependence in VBR video streams and ATM traffic engineering. *Performance Evaluation*, 30(1-2):45–56, 1997.

[6] Norbert Vicari. Models of WWW-Traffic: a Comparison of Pareto and Logarithmic Histogram Models. Technical report, Institute of Computer Science, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany, 3 1998.

[7] San qi Li and Chia-Lin Hwang. On the Convergence of Traffic Measurement and Queueing Analysis: A Statistical-Matching and Queueing (SMAQ) Tool. *IEEE/ACM Transactions on Networking*, 5(1):95–110, 1997.

[8] Allan T. Andersen and Bo Friis Nielsen. A markovian aproach for modeling packet traffic with long-range dependence. *IEEE Journal on Select Areas in Communication*, 16(5):719–732, 1998.

[9] Vinay J. Ribeiro, Rudolf H. Riedi, Matthew S. Crouse, and Richard G. Baraniuk. Simulation of nonGaussian Long-Range-Dependent Traffic using Wavelets. In *ACM SIGMETRICS*, 1999.

[10] Anestis Karasaridis and Dimitrios Hatzinakos. A Non-Gaussian Self-Similar Process for Broadband Heavy Traffic Modeling. In *Globecom*, pages 1513–1518, 1998.

[11] Ilkka Norros. On the Use of Fractional Brownian Motion in the Theory of Connectionless Networks. *IEEE Journal on Select Areas in Communications*, 13(6):953–962, 1995.

[12] Paulo Salvador, Rui Valadas, and António Pacheco. Multiscale Fitting Procedure using Markov Modulated Poisson Proceses. Technical report, Universidade de Aveiro/Instituto de Telecomunicações, 1 2003.

[13] Benjamin Melamed and Dimitrios E. Pendarakis. Modeling Full-Length VBR Video Using Markov-Renewal-Modulated TES Models. *IEEE Journal on Select Areas in Communications*, 16(5):600–611, 1998.

[14] Uttam K. Sarkar, Subramanian Ramakrishnan, and Dilip Sarkar. Modeling full-length video using Markov-modulated Gamma-based framework. *IEEE/ACM Transactions on Networking*, 11(4):638–649, 2003.

[15] Matthias Grossglauser and Jean-Chrysostome Bolot. On the relevance of long-range dependence in network traffic. Technical report, INRIA, 06902 Sophia Antipolis Cedex - France, 1996.

[16] Daniel P. Heyman and T. V. Lakshman. What Are the Implications of Long-Range Dependence for VBR-Video Traffic Engineering? *IEEE/ACM Transactions on Networking*, 4(3):301–317, 1996.

[17] Tarkan Taralp, Michael Devetsikiotis, and Ioannis Lambadaris. In Search of Better Statistics for Traffic Characterization. In *CAMAD-98*, pages 103–110, 1998.

[18] Hans-Peter Schwefel and Lester Lipsky. Impact of Aggregated, Self-Similar ON/OFF Traffic on Delay in Stationary Queueing Models (extended version). *Performance Evaluation*, (43):203–221, 2001.

[19] San qi Li and Chia-Lin Hwang. Queue response to input correlation functions: Continuous spectral analysis. *IEEE/ACM Transactions on Networking*, 1(6):678–692, 1993.

[20] Lalita A. Kulkarni and San qi Li. Transient behaviour of queueing systems with correlated traffic. *Performance Evaluation*, 27&28:117–146, 1996.

[21] Wing cheong Lau and San qi Li. Sojourn-time Analysis on Nodal Congestion in Broadband Networks and its Impact on QoS Specifications. In *IEEE INFOCOM*, pages 1327–1337, 1996.

[22] Kishor Shridharbhai Trivedi. *Probability and Statistics with Reliability, Queuing, and Computer Science Applications.* John Wiley & Sons, Inc., second edition, 2002.

[23] V.S. Frost and B. Melamed. Traffic Modeling for Telecommunications Networks. *IEEE Communications Magazine*, 32(3):70–81, 1994.

[24] Walter Willinger, Siddharta Devadhar, Andrew T. Heybey, Robert Sherman, Mark Sullivan, and John R. Vollaro. Measuring ATM traffic cell-by-cell: Experiences and preliminary findings from BAGNet. In *International Conference on the Performance and Management of Complex Communication Networks - PMCCN'97*, pages 91–110, 1997.

[25] Ashok Erramilli, Onuttom Narayan, and Walter Willinger. Experimental queueing analysis with long-range dependent packet traffic. *IEEE/ACM Transactions on Networking*, 4(2):209–223, 1996.

[26] Heman Michiel and Koen Laevens. Teletraffic Engineering in a Broad-Band Era. *Proceedings of the IEEE*, 85(12):2007–2033, 1997.

[27] Jan Beran, Robert Sherman, Murad S. Taqqu, and Walter Willinger. Long-Range Dependence in Variable-Bit-Rate Video Traffic. *IEEE Transactions on Communications*, 43(2/3/4):1566–1579, 1995.

[28] Bruce Hajek and Linhai He. On variations of queue response for inputs with the same mean and autocorrelation function. *IEEE/ACM Transactions on Networking*, 6(5):588–598, 1998.

[29] Ricardo Gusella. Characterizing the Variability of Arrival Processes with Indexes of Dispersion. *IEEE Journal on Select Areas in Communication*, 9(2):203–211, 1991.

[30] Kotikalapudi Sriram and Ward Whitt. Characterizing Superposition Arrival Processes in Packet Multiplexers for Voice and Data. *IEEE Journal on Select Areas in Communication*, 4(6):833–846, 1986.

[31] Sándor Molnár, Attila Vidács, and Arne A. Nilsson. Bottlenecks on the way towards fractal characterization of network traffic: Estimation and interpretation of the hurst parameter. In *International Conference on the Performance and Management of Complex Communication Networks - PMCCN'97*, pages 125–144, 1997.

[32] Ronn Ritke, Xiaoyan Hong, and Mario Gerla. Contradictory Relationship between Hurst Parameter and Queueing Performance. Technical report, UCLA - Computer Science Department, 405 Hilgard Ave., Los Angeles, CA 90024, 1998.

[33] Arnold L. Neidhardt and Jonathan L. Wang. The Concept of Relevant Time Scales and Its Application to Queuing Analysis of Self-Similar Traffic (or Is Hurst Naughty or Nice?). In *Sigmetrics 98*, pages 222–232, 1998.

[34] Frank Kelly. *Notes on Effective Bandwidth*, volume 4 of *Royal Statistical Society Lecture Notes*, pages 141–168. Oxford University Press, 1996. Stochastic Networks: Theory and Applications.

[35] Costas Courcoubetis, Vasilios A. Siris, and George D. Stamoulis. Application of the many sources asymptotic and effective bandwidths to traffic engineering. In *ACM SIGMETRICS*, 1998.

[36] Anestis Karasaridis and Dimitrios Hatzinakos. Resource Allocation Issues For Long-Tailed LRD Internet Wan Traffic. In *Globecom*, pages 1495–1499, 1999.

[37] Jaime Jungok Bae and Tatsuya Suda. Survey of Traffic Control Schemes and Protocols in ATM Networks. *Proceedings of the IEEE*, 79(2):170–189, 1991.

[38] Harry Heffes and David M. Lucantoni. A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance. *IEEE Journal on Select Areas in Communication*, 4(6):856–868, 1986.

[39] John N. Daigle and Joseph D. Langford. Models for Analysis of Packet Voice Communications Systems. *IEEE Journal on Select Areas in Communication*, 4(6):847–855, 1986.

[40] Wolfgang Fischer and Kathleen Meier-Hellstern. The Markov-Modulated Poisson Process (MMPP) Cookbook. *Performance Evaluation*, 18:149–171, 1992.

[41] Paul Skelly, Mischa Schwartz, and Sudhir Dixit. A Histogram-Based Model for Video Traffic Behavior in an ATM Multiplexer. *IEEE/ACM Transactions on Networking*, 1(4):445–459, 1993.

[42] Daniel Heyman and David Lucantoni. Modeling Multiple IP Traffic Streams with Rate Limits. In *Proceedings of ITC 17 - Teletraffic Engineering at the Internet Era*.

[43] Allan T. Andersen and Bo Friis Nielsen. An application of two state markovian sources to the modeling of self-similar behaviour. In *IEEE INFOCOM*, pages 196–204, 1997.

[44] Prodip Sen, Basil Maglaris, Nasser-Eddine Rikli, and Dimitris Anastassiou. Models for Packet Switching of Variable-Bit-Rate Video Sources. *IEEE Journal on Select Areas in Communication*, 7(5):865–869, June 1989.

[45] Basil Maglaris, Dimitris Anastassiou, Prodip Sen, Gunnar Karlsson, and John D. Robbins. Performance Models of Statistical Multiplexing in Packet Video Communications. *IEEE Transactions on Communications*, 36(7):834–844, 1988.

[46] Stephan Robert and Jean-Yves Le Boudec. On a Markov Modulated Chain Exhibiting Self-similarities over Finite Timescale. *Performance Evaluation*, 27:159–173, 1996.

[47] Stephan Robert and Jean-Yves Le Boudec. New models for pseudo self-similar traffic. *Performance Evaluation*, 30(1-2):57–68, 1997.

[48] Sidney Cunha de Lucena. Uma ferramenta para o cálculo de descritores de tráfego para modelos markovianos de fontes multimídia. Master's thesis, Coppe Sistemas - UFRJ, 8 1997.

[49] Rosa M. M. Leão and Edmundo de Souza e Silva and Sidney C. de Lucena. A Set of Tools for Traffic Modeling, Analysis and Experimentation. In *Computer Performance Evaluation / TOOLS*.

[50] David L. Jagerman and Benjamin Melamed. The Transition and Autocorrelation Structure of TES Processes, Part I: General Theory. *Communications in Statistics - Stochastic Models*, 8(2):193–219, 1992.

[51] David L. Jagerman and Benjamin Melamed. The Transition and Autocorrelation Structure of TES Processes, Part II: Special Cases. *Communications in Statistics - Stochastic Models*, 8(3):499–527, 1992.

[52] E. de Souza e Silva and H. R. Gail. The Uniformization Method in Performability Analysis. In G. Rubino B. R. Haverkort, R. Marie and K. S. Trivedi, editors, *Performability Modelling: Techniques and Tools*, chapter 3, pages 31–58. Wiley, 2001.

[53] Edmundo de Souza e Silva and H. Richard Gail. Calculating Cumulative Operational Time Distributions of Repairable Computer Systems. *IEEE Transactions on Computers*, c-35(4):322–332, 1986.

[54] Edmundo A. de Souza e Silva and Richard R. Muntz. *Métodos Computacionais de Solução de Cadeias de Markov: Aplicações a Sistemas de Computação e Comunicação.* Escola de Computação. Instituto de Informática da UFRGS, 1992.

[55] Ana Paula Couto da Silva. Métodos de Solução para Modelos Markovianos com Recompensa. Master's thesis, COPPE / UFRJ, 2001.

[56] Winfried Grassmann. Means and variances of time averages in Markovian environments. *European Journal of Operational Research*, (31):132–139, 1987.

[57] Tarkan Taralp, Michael Devetsikiotis, and Ioannis Lambadaris. Traffic Characterization for QoS Provisioning in High-Speed Networks. In *Proceedings of the 31st Annual Hawaii International Conference on System Sciences*, pages 485–492, 1998.

[58] http://www-tkn.ee.tu-berlin.de/ fitzek/trace/trace.html.

[59] Adenílson Raniery Sarges Pontes. Controle de admissão de conexões de vídeo com dependência temporal. Master's thesis, Pontifícia Universidade Católica do Rio de Janeiro, 10 2000.

[60] Rachid El Abdouni Khayari, Ramin Sadre, and Boudewijn Haverkort. The pseudo self-similar traffic model: Application and validation. Technical report, Laboratory for Performance Evaluation and Distributed Systems, RWTH Aachen, Ericsson Eurolab, Germany, 2002.

[61] Oliver Rose. Statistical Properties of MPEG Video Traffic and Their Impact on Traffic Modeling in ATM Systems. In *Proceedings of the 20th Annual Conference on Local Computer Networks*, pages 397–406, 1995.

[62] A. Lombardo, S. Palazzo, and G. Schembra. Characterization of Intermedia Synchronization and its Effects on the Performance of an ATM Multiplexer Loaded by Multimedia Traffic. Technical report, Istituto di Informatica e Telecomunicazioni, University of Catania, V.le A. Doria, 6 - 95125 Catania - Italia, 1999.

[63] Mark E. Crovella, Murad S. Taqqu, and Azer Bestavros. Heavy-tailed probability distributions in the world wide web. 1996.

[64] Alma Riska, Vesselin Diev, and Evgenia Smirni. Efficient fitting of long-tailed data sets into hyperexponential distributions. Technical report, Department of

Computer Science, College of William and Mary, Williamsburg, VA 23187-8795, USA, 2003.

[65] Marwan M. Krunz and Armand M. Makowski. Modeling Video Traffic Using $M/G/\infty$ Input Processes: A Compromise Between markovian and LRD Models. *IEEE Journal on Select Areas in Communications*, 16(5):733–748, 1998.

[66] M. Olson. The EMpht-programme.

[67] D. Towsley. Providing quality of service in packet switched networks. *Performance Evaluation of Computer and Communication Systems*, pages 560–586, 1993.

[68] A.Markopoulou, F.Tobagi, and M.Karam. Assessment of VoIP quality over Internet backbones. In *Infocom 2002*, June 2002.

[69] A. Markopoulou, F. Tobagi, and M. Karam. Assessing the quality of Voice Communications over Internet Backbones. *IEEE Transactions on Networking*, October 2003.

[70] Jie Liang. New Trends in Multimedia Standards: MPEG4 and JPEG2000.

[71] S. Asmussen, O. Nerman, and M. Oslon. Fitting phase type distributions via the EM algorithm. *Scandinavian Journal of Statistics*, 23:419–441, 1996.

[72] Robert Morris and Dong Lin. Variance of Aggregated Web Traffic. In *Infocom-2000*, pages 360–366, 2000.

[73] Xusheng Tian, Jie Wu, and Chuanyi Ji. A Unified Framework for Understanding Network Traffic Using Independent Wavelet Models. In *Infocom-2002*, 2002.

[74] F. Fitzek and M. Reisslein. MPEG–4 and H.263 Video Traces for Network Performance Evaluation. *IEEE Network*, 15(6):40–54, November/December 2001. video traces available at `http://www-tkn.ee.tu-berlin.de/research/trace/trace.html` or `http://www.eas.asu.edu/trace`.

[75] http://abilene.internet2.edu/.

[76] http://pma.nlanr.net/traces/long/ipls1.html.

[77] Chadi Barakat, Patrick Thiran, Gianluca Iannaccone, Christophe Diot, and Philippe Owezarski. Modeling Internet backbone traffic at the flow level. *IEEE Transactions on Signal Processing - Special Issue on Networking*, 51(8), 2003.

[78] http://www.land.ufrj.br.

[79] http://www.cos.ufrj.br.

[80] D.R. Figueiredo and E. de Souza e Silva. Efficient Mechanisms for Recovering Voice Packets in the Internet. In *Proceedings of IEEE/Globecom'99, Global Internet: Application and Technology Symposium.*

[81] Jorge Allyson Azevedo. Transmissão Multicast Confiável: Aplicação na Ferramenta Tangram Whiteboard e Experimentos na Internet. Master's thesis, NCE / UFRJ, 2002.