

**UNIVERSIDADE DE BRASÍLIA  
FACULDADE DE TECNOLOGIA  
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**UMA METODOLOGIA DE ENGENHARIA DE TRÁFEGO  
BASEADA NA ABORDAGEM AUTO-SIMILAR PARA A  
CARACTERIZAÇÃO DE PARÂMETROS E A  
OTIMIZAÇÃO DE REDES MULTIMÍDIA**

**PRISCILA AMÉRICA SOLÍS MENDEZ BARRETO**

**ORIENTADOR: PAULO HENRIQUE PORTELA DE CARVALHO**

**TESE DE DOUTORADO EM ENGENHARIA ELÉTRICA**

**PUBLICAÇÃO: PPGENE.TD-019/07  
BRASÍLIA-DF: MARÇO, 2007**

**UNIVERSIDADE DE BRASÍLIA  
FACULDADE DE TECNOLOGIA  
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**UMA METODOLOGIA DE ENGENHARIA DE TRÁFEGO  
BASEADA NA ABORDAGEM AUTO-SIMILAR PARA A  
CARACTERIZAÇÃO DE PARÂMETROS E OTIMIZAÇÃO  
DE REDES MULTIMÍDIA**

**PRISCILA AMÉRICA SOLÍS MENDEZ BARRETO**

**TESE DE DOUTORADO SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA ELÉTRICA DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM ENGENHARIA ELÉTRICA.**

**APROVADA POR:**

---

**Prof. Dr. PAULO HENRIQUE PORTELA DE CARVALHO, ENE/UNB  
(ORIENTADOR)**

---

**Prof. Dr. ANTONIO JOSÉ MARTINS SOARES, ENE/UNB  
(EXAMINADOR INTERNO)**

---

**Prof. Dr. RICARDO STACIARINI PUTTINI, ENE/UNB  
(EXAMINADOR INTERNO)**

---

**Prof. Dr. RODRIGO PINTO LEMOS, EEE/UFG  
(EXAMINADOR EXTERNO)**

---

**Prof. Dr. MARIO ANTONIO RIBEIRO DANTAS, INE/UFSC  
(EXAMINADOR EXTERNO)**

## FICHA CATALOGRÁFICA

SOLIS BARRETO, PRISCILA

Uma Metodologia de Engenharia de Tráfego Baseada na Abordagem Auto-Similar para a Caracterização de Parâmetros e a Otimização de Redes Multimídia [Distrito Federal] 2007.

xvi, 164p., 210x297mm (ENE/FT/UnB, Doutor, Tese de Doutorado – Universidade de Brasília. Faculdade de Tecnologia).

Departamento de Engenharia Elétrica

- |                              |                          |
|------------------------------|--------------------------|
| 1. Tráfego auto-similar      | 2. Engenharia de Tráfego |
| 3. Caracterização de tráfego | 4. Otimização            |
| 5. Planejamento de Rede      | 6. Métricas de QoS       |
| I. ENE/FT/UnB                | II. Título (série)       |

## REFERÊNCIA BIBLIOGRÁFICA

**SOLIS BARRETO, P. (2007).** Uma Metodologia de Engenharia de Tráfego Baseada na Abordagem Auto-Similar para a Caracterização de Parâmetros e a Otimização de Redes Multimídia. Tese de Doutorado em Engenharia Elétrica, Publicação PPGEN.TD-019/07, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 164p.

## CESSÃO DE DIREITOS

**AUTOR:** Priscila América Solís Mendez Barreto

**TÍTULO:** Uma Metodologia de Engenharia de Tráfego Baseada na Abordagem Auto-Similar para a Caracterização de Parâmetros e a Otimização de Redes Multimídia.

GRAU: Doutor

Ano: 2007

É concedida à Universidade de Brasília permissão para reproduzir cópias desta dissertação de mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta tese de doutorado pode ser reproduzida sem autorização por escrito do autor.

---

Priscila América Solís Mendez Barreto  
SHIS QL 14 Conjunto 6 Casa 13, CEP 71640-065 Brasília, DF, Brasil

## **AGRADECIMENTOS**

Durante o desenvolvimento das diferentes atividades do Doutorado na Universidade de Brasília, tive a sorte de trabalhar com excelentes pessoas com as quais aprendi muito.

Agradeço ao meu orientador, o Prof. Paulo Henrique Portela de Carvalho, pela ajuda constante, paciência, confiança e suporte ao longo do desenvolvimento deste trabalho. Agradeço também pela sua exigência e bom senso, o que resultou em um período de bastante aprendizado, acadêmico, profissional e humano.

Todo trabalho de pesquisa não se faz de forma isolada, o suporte de um grupo é fundamental. Agradeço aos funcionários e professores do Departamento de Engenharia Elétrica, principalmente aos professores Humberto Abdalla Júnior e Antônio Martins Soares, pelo seu interesse e dedicação em ajudar aos alunos a desenvolver suas potencialidades profissionais e acadêmicas. Devo dizer que aprendi muito como aluna e aumentei minha admiração à profissão de professor através deles.

Agradeço também aos meus colegas, entre alunos de graduação e pós-graduação do Labcom e Lemom, principalmente a Roque Lambert e Georges Amvame-Nze pelo entusiasmo e desafios que compartilhamos na atividade de pesquisa. Assim também gostaria de agradecer o suporte e interesse dado pelos meus colegas do Departamento de Ciência de Computação da UnB.

Aos professores do Departamento de Computação da Universidade Católica de Goiás, em especial ao Prof. Luiz Carlos de Souza, que foi um grande incentivador e facilitador deste projeto. Agradeço à minha amiga, Maria José pelo seu apoio e bons conselhos diante dos dilemas científicos e familiares.

Finalmente, e não por menos, agradeço aos meus pais pelo exemplo de força e perseverança, pela dedicação incondicional e apoio moral. Aos meus filhos, Ernesto, Lorenzo e a pequena Sofia, pelos momentos de dispersão e alegre distração que permitiram o desenvolvimento deste trabalho. A Marcelo pela companhia e incentivo.

Dedico este trabalho à minha família

## **RESUMO**

### **UMA METODOLOGIA DE ENGENHARIA DE TRÁFEGO BASEADA NA ABORDAGEM AUTO-SIMILAR PARA A CARACTERIZAÇÃO DE PARÂMETROS E A OTIMIZAÇÃO DE REDES MULTIMÍDIA**

**Autor: Priscila América Solís Mendez Barreto**

**Orientador: Paulo Henrique Portela de Carvalho**

**Programa de Pós-Graduação em Engenharia Elétrica**

**Brasília, março de 2007.**

Este trabalho propõe uma metodologia de engenharia de tráfego que tem como base a caracterização do tráfego em uma rede multimídia. A metodologia tem como objetivos a otimização das capacidades dos enlaces da rede e do tamanho das filas assim como a preservação da estabilidade operativa da rede ao otimizar o mapeamento de novos fluxos em caminhos virtuais. É proposto um modelo de tráfego híbrido que considera a topologia de uma rede do tipo *backbone* com tecnologia de comutação de etiquetas e tráfego agregados auto-similar e de Poisson. O modelo de tráfego proposto permite, então, o cálculo de métricas de QoS que estabelecem limites superiores para o tráfego agregado da rede. Com o cálculo dessas métricas é possível implementar o processo de otimização de planejamento da rede e avaliar o impacto da introdução de novas aplicações. Foi proposto e avaliado um algoritmo que integra o processo de otimização de planejamento e mapeamento do tráfego em rotas. Com diversos tipos de séries de tráfego foram feitas avaliações do algoritmo proposto e os resultados foram comparados com simulações. A comparação confirmou que o método proposto consegue, na maioria dos casos, estabelecer um limite superior aceitável para os valores das métricas de QoS na rede, o que permite a realização do processo de otimização. Os resultados obtidos validam a metodologia proposta neste trabalho como de utilidade prática em redes operativas de grande porte devido ao baixo custo computacional que um modelo analítico tem em relação à simulação destes ambientes.

## **ABSTRACT**

### **A TRAFFIC ENGINEERING METHODOLOGY BASED IN A SELF-SIMILAR APPROACH FOR PARAMETERS CHARACTERIZATION AND THE OPTIMIZATION OF A MULTIMEDIA NETWORK**

**Author: Priscila América Solís Mendez Barreto**

**Supervisor: Paulo Henrique Portela de Carvalho**

**Programa de Pós-Graduação em Engenharia Elétrica**

**Brasília, March, 2007.**

This work proposes a traffic engineering methodology based in multimedia traffic characterization. The methodology goals are the optimization of capacities of network links and buffers as well as the preservation of the network operation stability when optimizing the task of mapping flows into network virtual paths.

A hybrid traffic model is proposed, which considers the network topology of a backbone network with label switching technology and aggregated self-similar and Poisson traffic flows. The proposed traffic model permits the calculus of more accurate QoS metrics, which allows the establishment of upper QoS limits for aggregated traffic flows in the network. With the metrics calculus, it is possible to implement the optimization procedure of network planning and evaluate the impact of introducing new applications. Also, an algorithm which integrates the optimization of network planning and flows mapping is proposed. With different traffic series, an evaluation of the proposed algorithm is made and the results were compared with simulations. The comparison showed that the proposed method achieves, in most of the cases, the establishment of an acceptable upper limit for the QoS metric values in the network, which allows the execution of the optimization process. The results validate the proposed methodology of this work as a practical method for wide operative networks due the litte computational effort that an analytical model has in comparison to the simulation of these environments.

# SUMÁRIO

<b>1 -</b>	<b>INTRODUÇÃO .....</b>	<b>1</b>
1.1.	O CONTEÚDO DESTA TESE .....	5
1.2.	OBJETIVOS DESTA TESE.....	6
1.3.	PUBLICAÇÕES REALIZADAS PELA AUTORA .....	7
1.3.1.	PUBLICAÇÕES REALIZADAS PELA AUTORA RELACIONADAS A ESTA TESE.....	7
1.3.2.	OUTRAS PUBLICAÇÕES .....	8
<b>2 -</b>	<b>PROJETO DE REDES E ENGENHARIA DE TRÁFEGO .....</b>	<b>9</b>
2.1.	PLANEJAMENTO DE REDES DE PACOTES .....	10
2.1.1.	O PROJETO INICIAL DE REDE.....	11
2.1.2.	GERÊNCIA DA REDE.....	22
2.2.	ENGENHARIA DE TRÁFEGO EM REDES MULTIMÍDIA .....	23
2.2.1.	TECNOLOGIA MPLS .....	25
2.2.2.	PROBLEMAS DA ENGENHARIA DE TRÁFEGO (TE).....	29
2.2.3.	MÉTRICAS DE QOS.....	31
2.2.4.	COMPLEXIDADE COMPUTACIONAL PARA A ENGENHARIA DE TRÁFEGO.....	32
2.2.5.	O PROBLEMA DE OTIMIZAÇÃO COM BASE NA CARACTERIZAÇÃO DO TRÁFEGO .....	33
2.3.	RESUMO DO CAPÍTULO .....	37
<b>3 -</b>	<b>MODELOS DE TRÁFEGO PARA REDES MULTIMÍDIA.....</b>	<b>38</b>
3.1.	TRÁFEGO DE POISSON E TEORIA DE FILAS.....	39
3.2.	TRÁFEGO AUTO-SIMILAR.....	49
3.2.1.	INVARIÂNCIA À ESCALA E DEPENDÊNCIA DE LONGA DURAÇÃO .....	51
3.2.2.	MOVIMENTO BROWNIANO FRACIONÁRIO (FBM) .....	53
3.2.3.	RUÍDO GAUSSIANO FRACIONÁRIO (FGN).....	53
3.2.4.	MODELO DE TRÁFEGO PROPOSTO POR NORROS E PROCESSOS ENVOLTÓRIA FBM 54	
3.2.4.1.	ESCALA DE TEMPO MÁXIMA E TAMANHO DA FILA.....	57
3.2.4.2.	PERDA DE PACOTES .....	60



3.2.5.	CARACTERIZAÇÃO DO TRÁFEGO AUTO-SIMILAR E A AGREGAÇÃO DE FLUXOS...	63
3.2.5.1.	ESTIMATIVA DO PARÂMETRO DE HURST .....	63
3.2.5.2.	MODELO PARA OBTENÇÃO DOS PARÂMETROS DO TRÁFEGO AGREGADO.....	66
3.2.5.3.	INFLUÊNCIA DO PARÂMETRO $H$ E DO COEFICIENTE DE VARIÂNCIA.....	72
<b>3.3.</b>	<b>RESUMO DO CAPÍTULO .....</b>	<b>74</b>
<b>4 -</b>	<b>UMA FERRAMENTA PARA GERAÇÃO E ANÁLISE DE TRÁFEGO.....</b>	<b>76</b>
<b>4.1.</b>	<b>INFERÊNCIA.....</b>	<b>77</b>
4.1.1.	ESPECIFICAÇÃO DE DISTRIBUIÇÕES TEÓRICAS COM BASE EM DADOS OBSERVADOS .....	79
4.1.2.	TESTE DE HIPÓTESE .....	80
<b>4.2.</b>	<b>GERAÇÃO DE TRÁFEGO .....</b>	<b>82</b>
4.2.1.	GERAÇÃO POR INVERSÃO .....	83
4.2.2.	GERAÇÃO DE PROCESSO FGN .....	83
<b>4.3.</b>	<b>VALIDAÇÃO DA FERRAMENTA .....</b>	<b>85</b>
4.3.1.	VALIDAÇÃO DA GERAÇÃO DE TRÁFEGO.....	87
4.3.2.	ESTIMAÇÃO DE TRÁFEGOS REAIS .....	89
<b>4.4.</b>	<b>RESUMO DO CAPÍTULO .....</b>	<b>97</b>
<b>5 -</b>	<b>UM MODELO DE TRÁFEGO HÍBRIDO PARA CARACTERIZAÇÃO DE TRÁFEGO MULTIMEDIA.....</b>	<b>98</b>
<b>5.1.</b>	<b>CENÁRIOS DE SIMULAÇÃO E CONSIDERAÇÕES INICIAIS PARA A DEFINIÇÃO DE UM MODELO DE TRÁFEGO .....</b>	<b>98</b>
5.1.1.	O MODELO M/M/1 E SUAS LIMITAÇÕES .....	100
5.1.2.	O MODELO DE ENVOLTÓRIA FBM PARA CARACTERIZAÇÃO DO TRÁFEGO.....	103
5.1.3.	UM MODELO DE TRÁFEGO HÍBRIDO PARA CÁLCULO DO ATRASO .....	106
<b>5.2.</b>	<b>DEFINIÇÃO DE UM MODELO DE TRÁFEGO PARA CÁLCULO DE ATRASO E PERDAS DE PACOTES EM UMA REDE MULTIMEDIA .....</b>	<b>108</b>
5.2.1.	O MODELO DE TRÁFEGO HÍBRIDO .....	118
5.2.2.	ANÁLISE DE RESULTADOS DO MODELO DE TRÁFEGO HÍBRIDO.....	119
<b>5.3.</b>	<b>RESUMO DO CAPÍTULO .....</b>	<b>130</b>

<b>6 - PLANEJAMENTO DE REDE: OTIMIZAÇÃO DE PARÂMETROS DE UMA REDE MULTIMÍDIA.....</b>	<b>132</b>
<b>6.1. O PROBLEMA DA ALOCAÇÃO DAS CAPACIDADES.....</b>	<b>132</b>
<b>6.2. OS PROBLEMAS DA DEFINIÇÃO DO ROTEAMENTO.....</b>	<b>133</b>
<b>6.3. ALGORITMO GENÉTICO COMO FERRAMENTA DE OTIMIZAÇÃO.....</b>	<b>134</b>
<b>6.4. OTIMIZAÇÃO DE UMA REDE PARA MÉTRICAS DE PERDAS E ATRASO A UM CUSTO MINIMO.....</b>	<b>137</b>
<b>6.5. EXTENSÕES DO MODELO PARA ENGENHARIA DE TRÁFEGO.....</b>	<b>145</b>
<b>6.6. RESUMO DO CAPÍTULO.....</b>	<b>151</b>
<b>7 - CONCLUSÕES E TRABALHOS FUTUROS.....</b>	<b>153</b>
<b>7.1. CONCLUSÕES.....</b>	<b>153</b>
<b>7.2. TRABALHOS FUTUROS.....</b>	<b>155</b>
<b>BIBLIOGRAFIA.....</b>	<b>158</b>

## LISTA DE TABELAS

TABELA 2.1. FORMULAÇÃO GERAL PARA O PROBLEMA DE DIMENSIONAMENTO DE REDE (GERLA, M. E KLEINROCK L., 1977).....	11
TABELA 2.2. FORMULAÇÃO GERAL PARA O PROBLEMA OTIMIZAÇÃO DE ROTEAMENTO .....	19
TABELA 3.1. CÁLCULO COMPARATIVO DO ATRASO NO MODELO SIMULADO E ANALÍTICO .....	45
TABELA 3.2. PROPRIEDADES DO MODELO DE TRÁFEGO DE NÓRROS.....	54
TABELA 3.3. AVALIAÇÃO DOS MÉTODOS DE ESTIMATIVA DO PARÂMETRO DE HURST H. ....	66
TABELA 3.4. PARÂMETROS DOS TRAÇOS GERADOS. ....	70
TABELA 3.5. AVALIAÇÃO DOS MÉTODOS DE ESTIMATIVA DO TRÁFEGO AGREGADO DE DOIS FLUXOS. ....	71
TABELA 3.6. AVALIAÇÃO DOS MÉTODOS DE ESTIMATIVA DO TRAFEGO AGREGADO DE DOIS FLUXOS. ....	71
TABELA 3.7. RESULTADO DA SIMULAÇÃO. ....	73
TABELA 3.8. RESULTADO DA SIMULAÇÃO. ....	74
TABELA 4.1. VALORES ESPERADOS E OBSERVADOS E VALORES QUI-QUADRADO (30 GRAUS DE LIBERDADE) ...	88
TABELA 4.2. INFERÊNCIA ESTATÍSTICA NO TRÁFEGO AUTO-SIMILAR GERADO PELO GTAR.....	89
TABELA 4.3. PARÂMETRO DE HURST PARA AMOSTRAS DE 5 MINUTOS DE TRÁFEGO .....	94
TABELA 4.4. PARÂMETRO DE HURST PARA HTTP PARA AMOSTRAS DE 5 MINUTOS DE TRÁFEGO .....	95
TABELA 4.5. PARÂMETRO DE HURST PARA P2P PARA AMOSTRAS DE 5 MINUTOS DE TRÁFEGO.....	96
TABELA 4.6. PARÂMETRO DE HURST PARA AMOSTRAS NA ESCALA DE 1MS .....	96
TABELA 5.1. DESCRIÇÃO DOS FLUXOS PARA O CENÁRIO (A) .....	99
TABELA 5.2. FLUXOS DEFINIDOS PARA O CENÁRIO (B) .....	99
TABELA 5.3. TABELA DE ROTAS PARA O CENÁRIO (B) .....	100
TABELA 5.4. COMPARAÇÃO DOS ATRASOS PARA O CENÁRIO (A).....	100
TABELA 5.5. COMPARAÇÃO DOS ATRASOS PARA O CENÁRIO (A) NO MODELO SIMULADO .....	101
TABELA 5.6. FLUXOS DEFINIDOS PARA O CENÁRIO (A).....	104
TABELA 5.7. FLUXOS DEFINIDOS PARA O CENÁRIO (B).....	104
TABELA 5.8. PARÂMETROS DOS FLUXOS AUTO-SIMILARES.....	109
TABELA 5.9. RESULTADOS PARA A SÉRIE 9VBR01 .....	120
TABELA 5.10. RESULTADOS PARA A SÉRIE 9VBR02 .....	122
TABELA 5.11. RESULTADOS PARA A SÉRIE 9VBR03 .....	124
TABELA 5.12. RESULTADOS PARA A SÉRIE 9VBR04.....	125
TABELA 5.13. RESULTADOS PARA A SÉRIE 8VBR01 .....	126
TABELA 5.14. RESULTADOS PARA A SÉRIE 8VBR02 .....	127
TABELA 5.15. RESULTADOS PARA A SÉRIE 8VBR03 .....	128
TABELA 5.16. RESULTADOS PARA A SÉRIE 8VBR04 .....	129
TABELA 6.1. FORMULAÇÃO DO PROBLEMA DA ALOCAÇÃO DAS CAPACIDADES EM UMA REDE .....	133
TABELA 6.2. FORMULAÇÃO DO PROBLEMA DA ALOCAÇÃO DOS FLUXOS .....	133
TABELA 6.3. FORMULAÇÃO DO PROBLEMA DA ALOCAÇÃO DAS CAPACIDADES E DO ROTEAMENTO .....	134
TABELA 6.4. VALORES MÁXIMOS DE ATRASOS E PERDAS PARA OS FLUXOS.....	138
TABELA 6.5. CAPACIDADES OTIMIZADAS DA REDE PARA SÉRIE 9VBR01 .....	139

TABELA 6.6. ATRASOS E PERCENTUAL DE PERDAS DE PACOTES PARA REDE OTIMIZADA – SÉRIE 9VBR01 .....	139
TABELA 6.7. CAPACIDADES DA REDE COM M/M/1/K SÉRIE 9VBR01 .....	141
TABELA 6.8. MÉTRICAS DA REDE OTIMIZADA PARA SÉRIE 9VBR02 .....	142
TABELA 6.9. CAPACIDADES OTIMIZADAS DA REDE PARA SÉRIE 9VBR02 .....	142
TABELA 6.10. MÉTRICAS PARA A REDE OTIMIZADA COM FILAS MENORES – SÉRIE 9VBR01 .....	143
TABELA 6.11. ATRASOS E PERCENTUAL DE PERDAS DE PACOTES PARA REDE OTIMIZADA – SÉRIE 8VBR01 .....	144
TABELA 6.12. CAPACIDADES OTIMIZADAS DA REDE PARA SÉRIE 8VBR01 .....	145
TABELA 6.13. REQUERIMENTOS DE QoS DAS CLASSES .....	147
TABELA 6.14. FORMULAÇÃO DO PROCESSO DE TE COM CARACTERIZAÇÃO DE TRÁFEGO .....	147
TABELA 6.15. MÉTRICAS APÓS AGREGAÇÃO DO FLUXO 8VBR01 .....	150
TABELA 6.16. MÉTRICAS PARA VÁRIOS MAPEAMENTOS DE 8VBR01 .....	150

## LISTA DE FIGURAS

FIGURA 2.1- PROCESSO DE PLANEJAMENTO E OTIMIZAÇÃO DE REDE (RIEDL A., 2003).....	11
FIGURA 2.2. PROCESSO DE DESENHO DE UM PROJETO DE REDE (RIEDL A., 2003, COM MODIFICAÇÕES).....	12
FIGURA 2.3. OS ROTEADORES DE BORDA PARA A MATRIZ DE TRÁFEGO.....	13
FIGURA 2.4. K-CONECTIVIDADE PARA OS NÓS 1 E 3: (A) 2-CONECTADO (B) 2-CONECTADO (C) 3-CONECTADO .	15
FIGURA 2.5. EXEMPLO DE EXECUÇÃO DO ALGORITMO DE KLEITMAN PARA VERIFICAÇÃO DA CONECTIVIDADE DE UM GRAFO.....	16
FIGURA 2.6. ÁRVORE DE AMPLITUDE MÍNIMA EM UM GRAFO CONECTADO.....	16
FIGURA 2.7. COMPONENTES LER, LSR E LSP.....	27
FIGURA 2.8. ESTRUTURA DE UM NÓ MPLS.....	28
FIGURA 3.1. FILAS DE TRANSMISSÃO ENCADEADAS.....	41
FIGURA 3.2. EXEMPLO DE FILA EM REDE DE DADOS.....	42
FIGURA 3.3. REDE PARA O EXEMPLO DE UTILIZAÇÃO DO MODELO M/M/1.....	43
FIGURA 3.4. COMPARAÇÃO DE TRÁFEGO DE POISSON E TRÁFEGO INTERNET.....	47
FIGURA 3.5. SÉRIE DE TRÁFEGO AUTO-SIMILAR.....	56
FIGURA 3.6. SÉRIE DE TRÁFEGO AGREGADO E PROCESSO ENVOLTÓRIA FBM.....	57
FIGURA 3.7. LIMITE SUPERIOR DO TAMANHO DA FILA.....	59
FIGURA 3.8. CENÁRIO DE SIMULAÇÃO PARA AVALIAÇÃO DA INFLUÊNCIA DOS PARÂMETROS NO DESEMPENHO DA REDE.....	73
FIGURA 4.1. TELA PRINCIPAL PARA GERAÇÃO DE TRÁFEGO DO GTAR.....	86
FIGURA 4.2. DISTRIBUIÇÃO DE TEMPO ENTRE PACOTES PARA O TRÁFEGO POISSON GERADO PELO GTAR.....	88
FIGURA 4.3. DISTRIBUIÇÃO DE TEMPO ENTRE PACOTES PARA O TRÁFEGO POISSON GERADO PELO MGEN.....	88
FIGURA 4.4. ESQUEMA DO PROCESSO DE T <sub>E</sub> COM CARACTERIZAÇÃO DE TRÁFEGO.....	91
FIGURA 4.5. PERÍODO DE UTILIZAÇÃO PARA TRÁFEGO DOWNSTREAM – 31 DIAS.....	92
FIGURA 4.6. COMPORTAMENTO DO TRÁFEGO POR SERVIÇO PARA SETE DIAS.....	92
FIGURA 4.7. COMPORTAMENTO DO TRÁFEGO POR SERVIÇO - 24 HORAS.....	93
FIGURA 4.8. PERCENTUAIS DE TRÁFEGO POR SERVIÇO.....	93
FIGURA 5.1. REDE DO CENÁRIO (A) E DEFINIÇÃO DAS ROTAS POR FLUXO.....	99
FIGURA 5.2. REDE DO CENÁRIO (B).....	99
FIGURA 5.3. REDE COMPARAÇÃO DOS ATRASOS POR FLUXO.....	102
FIGURA 5.4. COMPARATIVO DOS ATRASOS PARA O FLUXO F4 E H=0.8.....	105
FIGURA 5.5. COMPARATIVO DOS ATRASOS PARA O FLUXO F4 E H=0.8 E VARIÂNCIA MAIOR.....	106
FIGURA 5.6. REDE COMPARAÇÃO DOS ATRASOS POR FLUXO – CENÁRIO (A).....	107
FIGURA 5.7. REDE COMPARAÇÃO DOS ATRASOS POR FLUXO – CENÁRIO (B).....	108
FIGURA 5.8. SÉRIE DE TRÁFEGO 9VBR01.....	110
FIGURA 5.9. ENVOLTÓRIA FBM PARA A SÉRIE 9VBR01.....	110
FIGURA 5.10. SÉRIE DE TRÁFEGO 9VBR02.....	111
FIGURA 5.11. ENVOLTÓRIA FBM PARA A SÉRIE 9VBR02.....	111
FIGURA 5.12. SÉRIE DE TRÁFEGO 9VBR03.....	112

FIGURA 5.13. ENVOLTÓRIA FBM PARA A SÉRIE 9VBR03 .....	112
FIGURA 5.14. SÉRIE DE TRÁFEGO 9VBR04 .....	113
FIGURA 5.15. ENVOLTÓRIA FBM PARA A SÉRIE 9VBR04 .....	113
FIGURA 5.16. SÉRIE DE TRÁFEGO 8VBR01 .....	114
FIGURA 5.17. ENVOLTÓRIA FBM PARA A SÉRIE 8VBR01 .....	114
FIGURA 5.18. SÉRIE DE TRÁFEGO 8VBR02 .....	115
FIGURA 5.19. ENVOLTÓRIA FBM PARA A SÉRIE 8VBR02 .....	115
FIGURA 5.20. ENVOLTÓRIA FBM PARA A SÉRIE 8VBR03 .....	116
FIGURA 5.21. ENVOLTÓRIA FBM PARA A SÉRIE 8VBR03 .....	116
FIGURA 5.22. ENVOLTÓRIA FBM PARA A SÉRIE 8VBR04 .....	117
FIGURA 5.23. ENVOLTÓRIA FBM PARA A SÉRIE 8VBR04 .....	117
FIGURA 6.1. DIAGRAMA DE FLUXO DO FUNCIONAMENTO DE UM AG TÍPICO .....	135
FIGURA 6.2. ESQUEMA DO PROCESSO DE TÊ COM CARACTERIZAÇÃO DE TRÁFEGO .....	146

## **LISTA SÍMBOLOS, NOMENCLATURA E ABREVIACÕES.**

ADSL	Assymmetric Digital Subscriber Line
ATM	Assynchronous Transfer Mode
BBT	Bidirectional Traffic Trunk
BER	Bit Error Rate
bps	bits por segundo
BRAS	Broadband Router Access Server
CBR	Constraint Based Routing - Constant Bit Rate
Diffserv	Differentiated Services
DWDM	Dense Wavelength Division Multiplexing
fBm	Fractional Brownian Motion
FEC	Forwarding Equivalence Class
fGn	Fractional Gaussian Noise
FIB	Forwarding Information Base
HTTP	Hyper Text Transfer Protocol
i.i.d.	independentes e identicamente distribuidos
IETF	Internet Engineering Tasking Force
IP	Internet Protocol
IS-IS	Intermediate System to Intermediate System
ITU	International Telecommunications Union
kbps	kbits por segundo
LAN	Local Área Network
LDP	Label Distribution Protocol
LER	Label Edge Router
LFIB	Label Forwarding Information Base
LPM	Least Prefix Match
LSP	Label Switched Path
LSR	Label Switch Router
MC	Multicommodity flow
MCP	Minimum Critical Path
MPLS	Multiprotocol Label Switching
MST	Minimum Spanning Tree

NP	Non Polynomiam
NS	Network Simulator
OSPF	Open Shortest Path First
p.d.f	probability density function
POP	Point of Presence
PPP	Point to Point Protocol
pps	pacotes por segundo
QoS	Quality of Service
RCS	Rate Controlled Service
RTFC	Rede de Telefonia fixa comutada
RTT	Round Trip Time
SDH	Synchronous Digital Hierarchy
TCP	Transmission Control Protocol
TE	Traffic Engineering
UDP	User Datagram Protocol
VBR	Variable Bit Rate
VoIP	Voice over IP
VPN	Virtual Private Network
VTP	Variance Time Plot
WDM	Wavelength Division Multiplexing



# 1 - INTRODUÇÃO

No cenário atual da tecnologia, as redes de telefonia fixa e móvel e as redes de pacotes têm-se integrado para dar lugar ao conceito de rede integrada ou multimídia. A integração dessas redes ainda se encontra em processo de padronização, pesquisa e desenvolvimento (Lambert R., 2005). Um ponto essencial nessa área é a implementação de qualidade de serviço para aplicações multimídia em que a caracterização de tráfego, a otimização de roteamento e o planejamento da rede são pontos essenciais do projeto.

Com o objetivo de se oferecer suporte de níveis de serviço para diferentes aplicações, a capacidade física dos núcleos de rede tem sido constantemente aumentada junto com a implementação de tecnologias tais como o MPLS (*Multi-protocol Label Switching*) com serviços diferenciados e reserva de recursos para a prática da engenharia de tráfego em uma rede multimídia (Solis Barreto P., de Carvalho, P. H. P., Soares, A. M., Abdalla Jr H., Lambert, R., Amvame-Nze, G., 2004)

Uma rede orientada a oferecer qualidade de serviço é fundamentada no agrupamento de diferentes tipos de tráfego em um conjunto de classes de serviços com prioridades específicas que são traduzidas em métricas de QoS a serem cumpridas. O MPLS é a tecnologia que permite implementar a engenharia de tráfego em uma rede global com suporte para aplicações multimídia (Awduche D. et al., 1999). Na prática, se traduz em redes com maiores capacidades e com técnicas mais eficientes de encaminhamento de pacotes e diferenciação de serviços para reserva de banda. Todavia com a persistência de problemas tais como a dificuldade no estabelecimento de garantias de QoS entre dois nós, devido, em grande parte, ao comportamento das fontes de tráfego e ao tratamento dos fluxos de tráfego em cada nó da rede.

O problema de QoS tem uma das suas origens na caracterização do tráfego na rede. Este processo é de fundamental importância para o processo de engenharia de tráfego. A obtenção de métricas confiáveis de QoS deriva-se de uma correta caracterização do tráfego. Historicamente, como um legado do sistema telefônico, na área de redes tem-se utilizado modelos de Markov (Bertsekas e Gallager, 1992) e suas extensões para redes de pacotes desenvolvidas por Kleinrock (Gerla, M. e Kleinrock L., 1977, Stallings W., 2002) para o planejamento da rede. Nesse cenário se parte da premissa que a fonte de tráfego tem comportamento de processos incrementais conforme processos de Poisson.

Com o trabalho pioneiro de Leland, Wilson e outros em 1994 (Leland W. E., Taqqu M. S., Willinger W., Wilson D. V., 1994) verifica-se a natureza auto-similar do tráfego em redes de dados e a sua característica de longa dependência em claro confronto com o bem aceito modelo de Poisson, que apresenta curta dependência entre as amostras.

Várias simulações e estudos analíticos (Mayor G. e Silvester J., 1997, Norros I, 1994, Bodamer S. e Charzinski J., 2002) indicaram que a característica de longa dependência presente no tráfego auto-similar tem efeito perturbador no tamanho das filas em uma rede. Nesse cenário, a distribuição do comprimento da fila decai muito lentamente em comparação ao comportamento observado com o tráfego Poisson, em que a fila exibe um comportamento exponencial de decrementos. Este comportamento é observado em diferentes escalas de tempo, pois a natureza auto-similar do processo preserva as propriedades estatísticas do tráfego ao longo das escalas.

Conclui-se, então, que o tráfego auto-similar produz um efeito perturbador no desempenho da rede, naturalmente inesperado quando a mesma foi projetada com um modelo de Markov. Essas conclusões vieram a confirmar as observações e percepções nas redes de grande porte nos últimos anos, principalmente após o aumento de tráfego vindo de aplicações com origem nos protocolos da Internet.

No processo atual de convergência de redes, com o chamado tráfego multimídia e onde a multiplexação estatística é o elemento dominante nos backbones das operadoras, medidas e avaliações do tráfego, confirmam altos graus de auto-similaridade (Perlingueiro F, Ling L, 2005, Rodrigues L, e Guardieiro P.R, 2004).

Esta observação se verifica para diversas aplicações e se mantém com a agregação sucessiva do tráfego nos núcleos de rede. As séries de tráfego que apresentam a longa dependência aparecem como as dominantes. Surge, então, a dificuldade de fornecer garantias de qualidade de serviço, definir tamanho de enlaces e filas e estabelecer políticas de roteamento, pois não existe um modelo de consenso que, diante da natureza auto-similar do tráfego, estabeleça a metodologia para a aplicação de técnicas de engenharia de tráfego na concepção atual de rede multimídia.

Na prática, o modelo de Markov é a base dos modelos utilizados como uma aproximação do desempenho que a rede deveria ter. Os enlaces são então aumentados em percentuais entre 50% e 80% para garantir banda para a desconhecida periodicidade das rajadas de

tráfego, característica intrínseca do tráfego auto-similar, e que venham a influenciar a taxa média calculada para um conjunto de séries agregadas de tráfego. Naturalmente, este método teórico-empírico tem um forte reflexo na relação custo-benefício do processo de planejamento de rede e na sua gerência.

Diversos trabalhos de pesquisa (Willinger W. e Paxson V, 1998, Sahinoglu Z. e Tekinay S, 1999, Riedl Anton, 2003) têm-se preocupado em desenvolver modelos que possam capturar o comportamento real do tráfego na rede, para assim poder implementar técnicas de engenharia de tráfego que permitam a obtenção de métricas de QoS próximas das observadas. Na última década, têm-se realizado diversos estudos que verificam a natureza do tráfego em diferentes tipos de redes (Willinger W. e Paxson V, 1998, Leland W. E., Taqqu M. S., Willinger W., Wilson D. V., 1994, Jerkins J. e Wang J.L, 1999, Sahinoglu Z. e Tekinay S, 1999, Gomes D., Agoulmine N., de Souza J.N, 2004). Muitos destes trabalhos utilizam as idéias formuladas por Norros (Norros I., 1994) para a modelagem do tráfego auto-similar como um processo fBm (fractional Brownian motion).

Os modelos propostos têm sido difíceis de serem levados à prática para operar em linha, principalmente pela sua alta complexidade computacional, pois a caracterização de um fluxo de tráfego como auto-similar exige o cálculo de parâmetros que podem precisar de longos tempos de observação para apresentarem precisão. O cálculo de métricas de QoS, a definição de rotas e o agrupamento de fluxos em função da banda efetiva e de QoS, assim como a otimização dos parâmetros de dimensionamento da rede são processos que dificilmente podem ser realizados em linha em redes em produção (Solis Barreto P., de Carvalho, P. H. P., Soares, A. M., Abdalla Jr H., 2005),.

Na prática, o projeto inicial de planejamento da rede com uma caracterização do tráfego que considere a fonte como auto-similar, com métricas de QoS mínimas a serem cumpridas pelos fluxos mapeados em rotas definidas previamente, pode ser o primeiro passo para a entrada em operação da rede, com a expectativa de se ter um desempenho mais aproximado do real e um maior tempo de estabilidade operativa.

A manutenção do desempenho pode-se realizar a partir da monitoração das métricas de QoS e da previsão das mesmas diante da entrada de novos fluxos em que o critério de preservação das métricas seria o que definiria o posicionamento em uma rota ou outra de um novo fluxo na rede.

Quando, após o processo de manutenção do desempenho, não é possível cumprir as métricas de QoS estipuladas para as aplicações, é necessário modificar a topologia da rede. Essa modificação, em contraste a uma prática teórico-empírica, pode ser realizada com modelos analíticos ou de simulação. O desenvolvimento de um modelo analítico pode, em certos casos, fornecer maior flexibilidade e facilidade de uso, principalmente quando grandes volumes de fluxos de tráfego podem ser caracterizados a partir de poucos parâmetros e permitir ainda a aplicação de técnicas de otimização para mudar a topologia da rede e prever seu comportamento, atividades menos flexíveis nos modelos de simulação e que para redes de grande porte tem apresentado fortes limitações (Fujimoto R., Perrumalla K., Park A., Wu H., Ammar M., Riley G, 2003). Existe então o potencial de aplicar técnicas de otimização para uma relação estabelecida de custo-benefício que permitam definir caminhos virtuais assim como o mapeamento do tráfego e dimensionamento de enlaces e filas, com um custo computacional mais baixo que aquele exigido pela simulação de redes com grandes volumes de tráfego.

Existem diversos trabalhos que têm focado na otimização do roteamento e do planejamento para redes multimídia (Elwalid, C. Jin, S. Low, e I. Widjaja, 2001, Riedl A., 2002, Banerjee G. e Sidu D, 2002). Um problema recorrente observado é o método para obter métricas de QoS confiáveis a serem utilizadas, seja para definição de políticas de roteamento e agrupamento de tráfego para o cálculo de capacidades dos enlaces e filas. A complexidade computacional da otimização aumenta quando se consideram múltiplas variáveis a serem otimizadas conjuntamente. A particularização de situações para reduzir esta complexidade tem sido utilizada em diversos casos com resultados aceitáveis onde se considera que a solução obtida é sub-ótima e que pode, dentro de certo intervalo de confiança, descrever o comportamento real da rede.

Existe então, no cenário tecnológico atual, a necessidade de um modelo analítico que permita que, em uma rede multimídia, seja realizado um cálculo de métricas de QoS para um conjunto de fluxos agregados com características de auto-similaridade. Este modelo em conjunto com um método de otimização que permita calcular as capacidades de enlaces e filas e definir rotas e agrupamento de fluxos em uma rede multimídia pode ser de grande utilidade no processo de planejamento, operação e gerência de redes.

## 1.1. O CONTEÚDO DESTA TESE

Esta tese tem como objetivo geral propor uma metodologia que tenha como base a caracterização do tráfego em redes multimídia para otimizar o planejamento, a operação e promover a estabilidade.

A metodologia se fundamenta em um modelo analítico para a avaliação de métricas de QoS que funcionem como limites superiores para a garantia operativa da rede com diversas configurações de enlaces, filas e rotas. O modelo analítico considera a modelagem de uma rede do tipo backbone, com tecnologia MPLS e conjuntos de fluxos agregados com diversos parâmetros. Assim também, o modelo permite avaliar o impacto da introdução de novos fluxos na rede e otimizar a escolha do LSP (*Label Switched Path*) em função de duas métricas de QoS. O trabalho tem como base dois elementos principais:

(a) um modelo de tráfego híbrido que, a partir da caracterização do tráfego e do seu comportamento nos diferentes enlaces da rede, define o método a ser aplicado para a obtenção de métricas máximas de QoS;

(b) uma heurística para a implementação da otimização dos elementos de rede que ajude a garantir as métricas de QoS dentro de valores máximos estipulados e com um custo mínimo.

Os itens (a) e (b) requerem um conjunto de conceitos básicos que inicialmente são descritos no capítulo 2, onde se apresenta uma visão geral sobre o projeto de rede e a engenharia de tráfego. Na primeira parte desse capítulo, é detalhada a produção dos elementos necessários para a implantação de uma rede. São discutidos elementos tais como confiabilidade, a formulação do problema da otimização em função das capacidades e do roteamento, a gerência de rede e seus elementos, onde se dá destaque ao processo de engenharia de tráfego. Posteriormente, ainda nesse capítulo, descreve-se brevemente a tecnologia MPLS e os problemas da engenharia de tráfego, identificando-se então o foco do modelo proposto nesta tese.

No capítulo 3, são apresentadas as bases para os modelos de caracterização de tráfego para redes multimídia. Inicia-se com uma breve ilustração dos modelos M/M/1 e M/M/1/K e da sua aplicabilidade no cenário atual. Posteriormente, discutem-se os parâmetros que caracterizam o tráfego auto-similar e as agregações com outros tráfegos. Discutem-se os

diversos enfoques para o cálculo de métricas de QoS em uma rede com circuitos virtuais ao usar esta caracterização. Descreve-se o movimento Browniano fracionário e o processo de envoltória com base nesse movimento para a descrição do tráfego agregado de uma fonte auto-similar.

O capítulo 4 descreve a ferramenta utilizada neste trabalho para a geração de tráfego sintético assim como para a caracterização de séries de tráfego real. A ferramenta, chamada de GTAR é descrita em seus fundamentos conceituais e é efetuada uma validação dos módulos de geração de tráfego e inferência estatística com o uso de séries reais de tráfego obtidas em um provedor Internet de grande porte.

O capítulo 5 apresenta um modelo de tráfego híbrido com base nos modelos M/M/1k e fBm, apresentado no capítulo 3. Esse modelo de tráfego híbrido tem como propósito a produção de métricas de QoS, atraso e perdas de pacotes, que sejam mais precisas para fluxos de tráfego agregado. O modelo de tráfego híbrido é avaliado com uma série de séries de tráfego experimentais e é definido o seu domínio de aplicação.

O capítulo 6 formula os problemas tradicionais do projeto de rede de pacotes, tais como identificados nos trabalhos pioneiros de Gerla e Kleinrock em 1977 e os estende para a configuração de uma rede multimídia com circuitos virtuais. É mostrada a viabilidade da utilização de uma ferramenta de otimização evolucionária junto com o modelo de tráfego híbrido para o processo de dimensionamento de uma rede. Os resultados são confrontados com diversas simulações com o propósito de validar a metodologia proposta. É apresentado um algoritmo que permite avaliar o melhor mapeamento de rotas para novos fluxos na rede promovendo-se assim a estabilidade operativa da mesma.

O capítulo 7 apresenta as conclusões obtidas deste trabalho de tese e as possíveis vertentes para trabalhos futuros.

## **1.2. OBJETIVOS DESTA TESE**

- Analisar e formular um modelo de tráfego híbrido com base no modelo M/M/1k e processos envoltória fBm. Validar a aplicabilidade do modelo para diferentes relações de média, variância, explosividade e níveis de carga no enlace;
- Definir e analisar a implementação de um modelo analítico e seus métodos para que, com base no modelo de tráfego híbrido, em um ambiente de rede com circuitos

virtuais conforme a tecnologia MPLS, permita a obtenção de métricas de QoS tais como perdas e atraso para troncos de tráfego multimídia mapeados em diferentes LSPs ;

- Implementar uma heurística de otimização de baixo custo computacional para o dimensionamento de enlaces e filas em uma rede com circuitos virtuais para diferentes volumes de tráfego e diferentes níveis de carga nos enlaces em função de dois requerimentos mínimos de QoS;
- Definir e implementar um método de promoção da estabilidade da rede a partir da caracterização de novos fluxos para a otimização do agrupamento em troncos de tráfego já existentes;
- Analisar e discutir sobre a utilidade do modelo de tráfego híbrido, do modelo analítico e da heurística de otimização para o mapeamento de fluxos e otimização adaptativa do roteamento em uma rede multimídia com tecnologia MPLS;
- Caracterizar e analisar diferentes séries de tráfego sintéticos e reais e analisar a relação dos parâmetros das diferentes fontes de tráfego que os geram;
- Implementar uma ferramenta para uso acadêmico de geração de tráfego sintético de múltiplos tipos em escalas de tempo de até  $10^{-3}$  segundos e para o caso do tráfego auto-similar, para diferentes relações de média e variância, o que é de grande utilidade na pesquisa de extensões futuras ao modelo de tráfego híbrido para o caso do tráfego multifractal (Vianna Neto, César A., 2005).

### **1.3. PUBLICAÇÕES REALIZADAS PELA AUTORA**

#### **1.3.1. Publicações Realizadas pela Autora Relacionadas a esta Tese**

- Solis Barreto P., de Carvalho, P. H. P., Soares, A. M., Abdalla Jr H. (2005). “The Influence of Self-similar Traffic in Real Time Applications.” In: Yu.I. Shokin, O.I. Potaturkin (eds.) *IASTED International Conference on Communications Systems*, ACTA Press, Novosibirsk, Russia.
- Solis Barreto P., de Carvalho, P. H. P., Soares, A. M., Abdalla Jr H., Lambert, R. (2005). “Hybrid Model of Traffic Characterization for Optimization in GMPLS

Optical Networks.” In: *IMOC 2005, International Microwave and Optoelectronics Conference*, Brasilia, Brasil.

- Solis Barreto P., de Carvalho, P. H. P., Soares, A. M., Abdalla Jr H. (2005), “A Traffic Model for Multimedia Applications in Converged Networks.” In: *13th IEEE Modelling, Analysis and Simulation in Computer and Telecommunications*, IEEE Computer Society, 163-170, Atlanta, E.U.A.
- Solis Barreto P., de Carvalho, P. H. P., Soares, A. M., Abdalla Jr H., Lambert, R., Amvame-Nze, G., Bizerra R.S. (2006). “Open Source Software for Evaluation of Applications in an Experimental Testbed for Converged Networks.” In: *2nd. International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities*, IEEE Computer Society, Barcelona, Espanha.
- Solis Barreto P., de Carvalho, P. H. P., Soares, A. M., Abdalla Jr H., Lambert R., Amvame-Nze, G., Bizerra R.S. (2005). “Implementação de um Ambiente de Teste e Medição em Redes Convergentes.” In: *XXII Simpósio Brasileiro de Telecomunicações*, Campinas, São Paulo, Brasil.
- Solis Barreto P., de Carvalho, P. H. P., Soares, A. M., Abdalla Jr H., Bizerra R. S. (2005). “Uma Ferramenta de Código Aberto para Análise de Desempenho em Redes Convergentes.” In: *XXII Simpósio Brasileiro de Telecomunicações*, Campinas, São Paulo, Brasil.
- Solis Barreto P., de Carvalho, P. H. P., Soares, A. M., Abdalla Jr H., Lambert, R., Amvame-Nze, G. (2004). “Performance Evaluation of Shortest Path Computation for IP and MPLS Multi-service Networks over Open Source Implementation.” In: *Lecture Notes in Computer Science*, Springer-Verlag, Volume 3126, 267-278.

### **1.3.2. Outras Publicações**

- Solis Barreto P. ; de Carvalho, P. H. P., Soares, A. M., Abdalla Jr H., Puttini R., Lambert, R., Amvame-Nze, G. (2005). “A Study of Billing Schemes in an Experimental Next Generation Network.” In: *Lecture Notes in Computer Science*, Springer-Verlag, Volume 3420, 61-67.



## 2 - PROJETO DE REDES E ENGENHARIA DE TRÁFEGO

Tradicionalmente, o desenho de uma rede de comutação de pacotes tem sido mais complexo que uma rede de comutação de circuitos. Inicialmente, as complexidades aparecem ao considerar fatores tais como a presença de mais de uma rota entre origem e destino, a análise entre o atraso de pacotes, a utilização do enlace e filas, entre outros.

Com o crescimento massivo da Internet e a sua organização em sistemas autônomos, diferentes operadores de serviço gerenciam porções independentes da rede. O backbone da Internet é formado por redes de trânsito especialmente projetadas para transmissão de dados em longas distâncias. Conectados a estas redes estão os provedores regionais que provêm o acesso a organizações, empresas, universidades e clientes de natureza privada.

Um dos principais interesses dos provedores de serviço comerciais é otimizar sua operação para minimizar os custos. Este processo implica a realização de gerenciamento de rede que disponha de medidas que possam ser traduzidas em variáveis monetárias relacionadas à operação da rede. Este processo pode se tornar complexo ao considerar um ambiente de rede multiserviço, com diferentes fluxos e requerimentos de QoS (*Quality of Service*). Neste texto o termo QoS é utilizado no mesmo entendimento do IETF (*Internet Engineering Tasking Force*), que se diferencia da interpretação do ITU (*International Telecommunications Union*).

O QoS é a capacidade da rede de fornecer garantia para um determinado tipo de tráfego que seus requerimentos serão satisfeitos. Ainda que o QoS seja usado como o objetivo de um problema de otimização de rede, são necessárias medidas quantitativas que o expressem como resultado de uma configuração de rede.

Tradicionalmente, o planejamento de redes de pacotes tem se concentrado em otimizar custo de desempenho em uma topologia de rede, assim como a capacidade dos enlaces e as estratégias de roteamento. Nos anos recentes, com a utilização massiva de diferentes serviços, a introdução de novos conceitos e tecnologias e um melhor conhecimento das características estatísticas do tráfego multimídia, o processo de planejamento tem apresentado diferentes alternativas, complexidades e desafios (Jerkins J. e Wang J.L , 1999)

Neste capítulo se apresenta uma discussão sobre planejamento de redes de pacotes e o processo de projeto inicial de rede, assim como os fundamentos da teoria para a caracterização de tráfego em redes multimídia.

## **2.1. PLANEJAMENTO DE REDES DE PACOTES**

Um processo de planejamento de redes tem como objetivo a otimização de uma ou mais variáveis relacionadas à operação da rede. A operação da rede é o objetivo central do processo de planejamento. Sob esta consideração, definem-se processos que promovem a otimização de um conjunto de variáveis pré-definidas, entre as quais geralmente se incluem o custo de operação e parâmetros de desempenho traduzidos em métricas de QoS.

Os diferentes processos e suas tarefas associadas são mostrados na figura 2.1. O projeto inicial de rede consiste em um ponto de partida de operação. As etapas do processo de projeto inicial de rede serão discutidas mais detalhadamente em seções posteriores.

Quando a rede entra em operação, deve-se monitorar um conjunto de parâmetros pré-definidos a partir de um processo de Engenharia de Tráfego que permita realizar ajustes e otimizações. Entre os controles de engenharia de tráfego, inclui-se o controle de admissão, que é uma medida que permite avaliar fluxos de tráfego e o cumprimento dos requerimentos de QoS. Este processo também identifica o momento de carga de tráfego que, com a infra-estrutura implantada, não permite novos ajustes, o que exige um novo dimensionamento e uma nova topologia.

O processo de adaptação de rede recebe as novas configurações que permitirão à rede voltar a operar corretamente e cumprir os requerimentos de desempenho estabelecidos. Esse processo geralmente envolve novos investimentos em infra-estrutura e equipamentos.

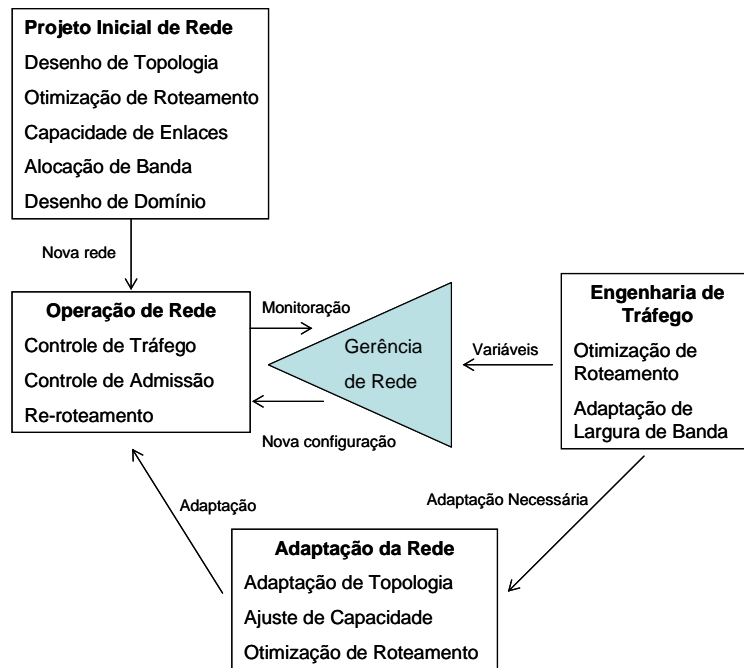


Figura 2.1- Processo de planejamento e otimização de rede (Riedl A., 2003).

### 2.1.1. O Projeto Inicial de Rede

O processo para definir o projeto inicial de uma rede é complexo, podendo ser subdividido em diferentes tarefas a serem tratadas individualmente. Uma formulação geral para o projeto inicial é mostrada na Tabela 2.1. O parâmetro a minimizar é o custo total da linha, o que não impede que outros parâmetros possam ser utilizados, tais como o atraso médio sujeito a um custo específico, etc.

Tabela 2.1. Formulação geral para o problema de dimensionamento de rede (Gerla, M. e Kleinrock L., 1977)

dados de entrada	localização dos nós horários de pico de tráfego entre os nós
minimizar	custo total da linha ou da rede
considerando as variáveis	topologia capacidade dos canais políticas de roteamento
sujeito a	restrições de capacidade dos canais restrições de atraso de pacotes restrições de confiabilidade

A formulação da Tabela 2.1 considera uma rede com um único tipo de tráfego, representado por pacotes onde os requerimentos de tráfego entre os nós acontecem em intervalos aleatórios, e o comportamento do sistema é como o de uma rede estocástica de filas. Nesse caso, focaliza-se o problema do projeto inicial em uma rede de pacotes com

uma única classe de tráfego. Os modelos matemáticos associados a esse enfoque serão discutidos em seções posteriores deste texto.

Nos ambientes de redes multiserviço que existem hoje, onde é necessária a diferenciação dos serviços para requerimentos de QoS, o processo de projeto inicial mostrado na Tabela 2.1 pode ser estendido para considerar essa nova realidade. O conjunto de passos individuais para este processo é mostrado na figura 2.2.

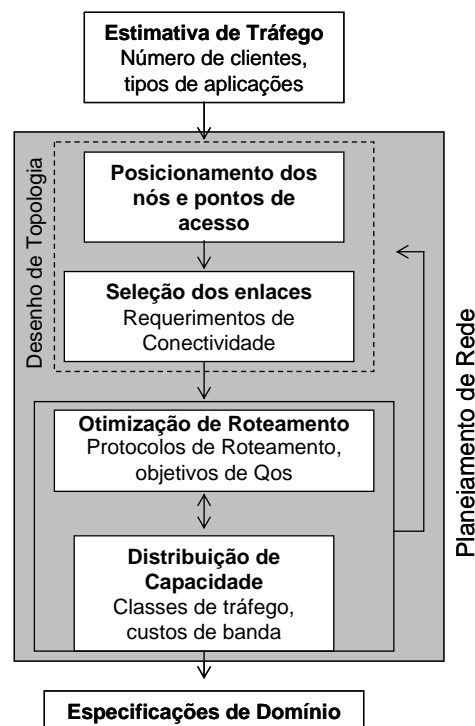


Figura 2.2. Processo de Desenho de um projeto de rede (Riedl A., 2003, com modificações).

### Estimativa do Tráfego

Nesse novo cenário, o processo se inicia com uma estimativa do tráfego na rede. Nos primeiros anos da utilização da rede de pacotes, existia a visão de uma única classe de tráfego, o que em certa forma simplificava a análise. Hoje em dia, nas redes multiserviço, cada tipo de aplicação deve ser classificada em um tipo de serviço, e deve-se estimar o tráfego associado a essa aplicação, utilizando uma matriz de tráfego, para estabelecer as relações de tráfego entre os diferentes nós. Isto é necessário, pois cada um dos serviços tem diferentes necessidades de utilização da rede e diferentes perfis de usuários.

Uma forma de classificar as aplicações é utilizar as métricas de QoS. Em uma rede multiserviço, definem-se classes de tráfego que comportam estas métricas. A especificação da aplicação, em função das métricas que deve cumprir, identifica a classe de tráfego em que ela deve ser mapeada. A especificação das classes de tráfego e do seu comportamento é uma característica tecnológica da rede.

Uma matriz de tráfego é uma estrutura que, em suas linhas e colunas, possui os nós de fronteira da rede e o fluxo estimado entre estes nós. Ou seja, esta matriz mostra o tráfego entrante na rede que vem dos diferentes usuários e seu agrupamento para a entrada na rede do provedor de serviço.

Como mostrado na figura 2.3, os roteadores de borda são os localizados nos POPs (Pontos de Presença) do provedor de serviço, que representam o acesso à rede. As áreas atendidas em cada POP têm diferentes características demográficas, sócio-econômicas, comportamentais e tecnológicas que influenciam a utilização de serviços e seu conseqüente tráfego. Nos roteadores de borda então, agregam-se os diferentes fluxos de tráfego que vêm de cada uma destas áreas. Note-se que a agregação dos fluxos deve preservar as especificações individuais de cada classe de tráfego relativa ao serviço.

Na fase inicial, a matriz de tráfego representa estimativas gerais do tráfego. Mas durante a operação da rede, os dados da matriz de tráfego devem ser atualizados periodicamente para acompanhar, ainda que de forma aproximada, o dinamismo comportamental dos usuários e de suas aplicações (Elwalid, C. Jin, S. Low, e I. Widjaja, 2001, Iovanna P., Sabella R., Settembre M., 2003). Essa periodicidade da atualização depende também da operação da rede e das suas políticas de mercado.

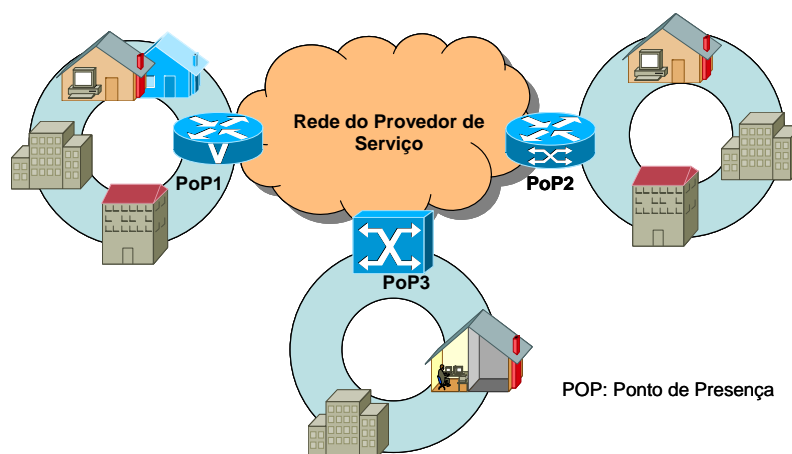


Figura 2.3. Os roteadores de borda para a matriz de tráfego.

## Desenho de Topologia

O desenho da topologia contém duas tarefas: posicionamento dos nós e seleção dos enlaces.

A localização dos nós se dá após conhecer a distribuição de clientes na área de serviço e as demandas de tráfego. Este processo agrupa a escolha dos tipos de equipamentos e as tecnologias a serem utilizadas. As especificações de posicionamento, equipamentos e tecnologias devem influenciar nos padrões de tráfego dentro da rede.

A escolha dos enlaces surge da necessidade de interconectar os diferentes roteadores de borda. Neste ponto, o tipo de topologia é escolhido e se dá preferência na prática às topologias híbridas. Um fator fundamental na escolha dos enlaces é a confiabilidade, ou seja, a capacidade de a rede se manter conectada na presença de falha em um enlace, o que naturalmente afeta a sua topologia.

O problema de escolher os enlaces deve considerar a conectividade da rede em relação aos custos. Normalmente, trabalha-se sobre um grafo de enlaces candidatos que devem ser escolhidos. Para o cálculo de custos, podem-se fazer suposições de roteamento e infraestrutura dos enlaces, para um conjunto de demandas de tráfego conhecidas.

O requerimento de conectividade é um elemento importante do projeto de rede e muitas vezes, quando otimizada a topologia, podem-se obter graus baixos de conectividade. É necessária a utilização da teoria de grafos para a análise da conectividade, que pode ser feita em conjunto com o processo de otimização.

Conceitualmente, dois nós  $i, j$  de um grafo não direcionado são  $k$ -conectados se existe um caminho que conecta  $i$  e  $j$  em cada subgrafo obtido ao eliminar  $(k-1)$  nós exceto  $i$  e  $j$  junto com suas arestas adjacentes do grafo (Bertsekas D., Gallager R., 1992). Um exemplo deste conceito pode ser visualizado na figura 2.4.

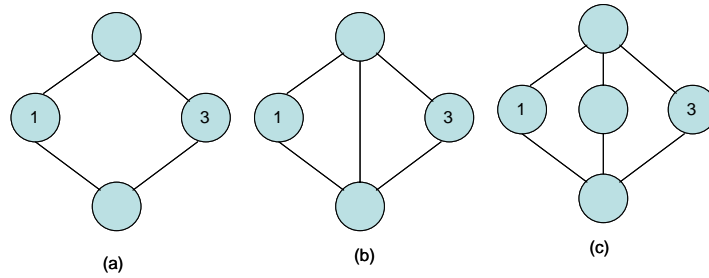


Figura 2.4.  $k$ -conectividade para os nós 1 e 3: (a) 2-conectado (b) 2-conectado (c) 3-conectado

Um grafo é  $k$ -conectado se cada par de nós é  $k$ -conectado. A verificação da conectividade pode ser realizada ao fazer uso do algoritmo de Kleitman (Bertsekas D., Gallager R., 1992), descrito nos seguintes passos:

1.  $i \leftarrow 0$ ,  $k \leftarrow$  grau de conectividade a ser verificado
2. Repetir
  - Selecionar um nó arbitrário  $n_i$
  - Conectado  $\leftarrow$   $(k-i)$ -conectividade entre  $n_i$  e os outros nós
  - Eliminar o  $n_i$  e suas arestas adjacentes do grafo
  - $i \leftarrow i + 1$
  - até  $(i > k-1)$  ou  $(\text{conectado} = 0)$
3. Se conectado então grafo é  $k$ -conectado

Um exemplo da aplicabilidade do algoritmo é mostrado na figura 2.5. Em 2.5(a) se verifica a 3-conectividade entre  $n_0$  e os outros nós. Depois em (b),  $n_0$  é eliminado e se escolhe um nó arbitrário  $n_1$  para se verificar a 2-conectividade e, em (c), obtém-se um grafo não-conectado e então se conclui que o grafo original de (a) não é 3-conectado. Em 2.5(d), inicia-se uma nova verificação para 3-conectividade. Em (e) se elimina  $n_0$  e se verifica a conectividade entre  $n_1$  e os outros nós. Em (f), verifica-se que o grafo é conectado e se conclui a 3-conectividade do grafo original.

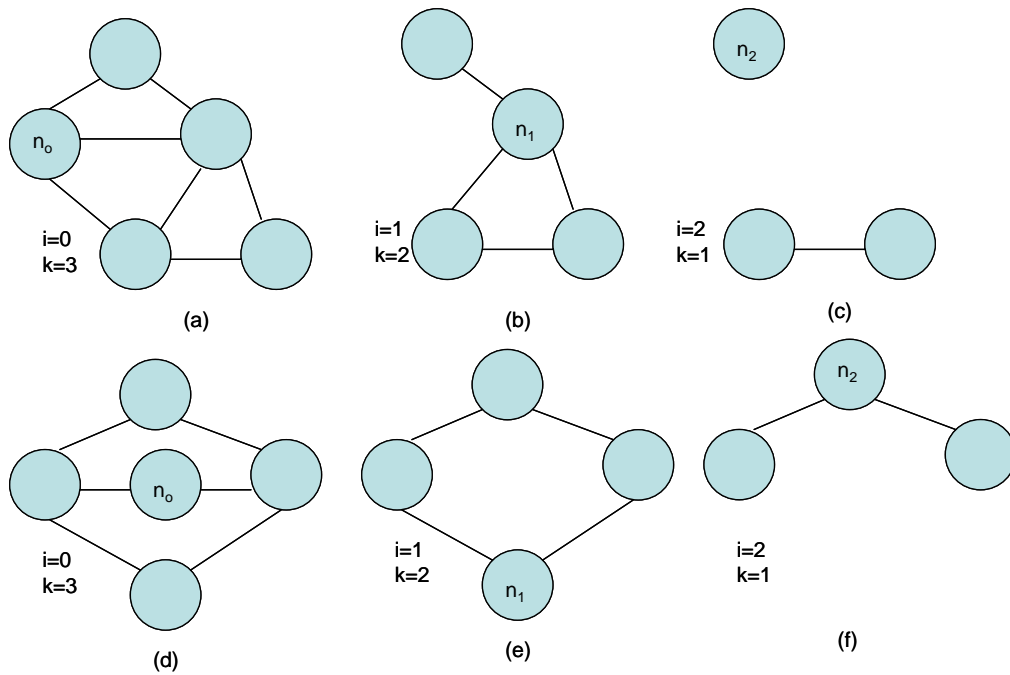


Figura 2.5. Exemplo de execução do algoritmo de Kleitman para verificação da conectividade de um grafo.

Para as redes em que conectividade não é importante, pode-se utilizar uma topologia do tipo árvore de amplitude (spanning tree), de tal forma que a capacidade é concentrada em alguns poucos enlaces. Uma árvore de amplitude é um subgrafo conectado de uma rede que inclui todos os nós e não possui ciclos. Uma árvore de amplitude mínima (MST-*Minimum Spanning Tree*) é uma árvore de amplitude com uma soma mínima de pesos em suas arestas, com  $G=(V,A)$ , em que  $V$  representa o conjunto de nós,  $A$  o conjunto de arestas e  $w_{ij}$  o peso para cada aresta  $(i,j) \in A$ . A figura 2.6 mostra um exemplo de uma árvore de amplitude mínima, onde os enlaces da árvore de amplitude mínima calculada estão em linhas duplas e os pesos  $w_{ij}$  são mostrados em cada aresta.

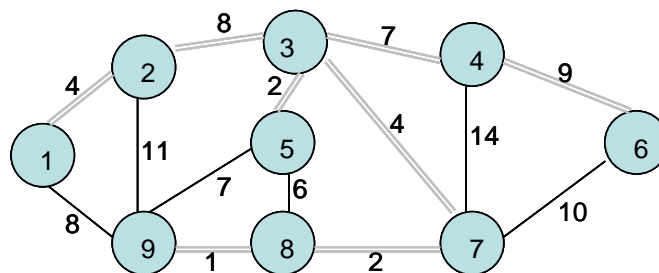


Figura 2.6. Árvore de amplitude mínima em um grafo conectado.



Entre os vários algoritmos para o cálculo da árvore de amplitude mínima em um grafo conectado, encontram-se os de Kruskal e de Prim, ambos com base em uma estratégia de abordagem gulosa (Cormen T., et al, 2002). O algoritmo de Prim opera de modo muito semelhante ao algoritmo de Dijkstra para localizar caminhos mais curtos em um grafo em conjunto com uma estratégia gulosa que aumenta a árvore de amplitude mínima uma aresta de cada vez. Algoritmos de roteamento tais como o OSPF (*Open Shortest Path First*), utilizam a estratégia da árvore de amplitude para a descoberta da topologia.

Um problema potencial é a restrição de quantidade de tráfego que pode ser transmitida em um enlace. Nesse caso, um algoritmo que considere a restrição é necessário e pode ser resolvido ao modificar o algoritmo de Prim de tal forma que cada vez que uma aresta deve ser adicionada, verifica-se se cumpre as restrições de tráfego e, caso não as cumpra não é adicionada.

Um caso especial do MST com restrições é o algoritmo de Essau Williams (Cormen T., et al, 2002), que trabalha com um nó central  $N$  por onde todo o tráfego deve passar, algoritmo muito utilizado no projeto de redes de acesso, com o nó central como o concentrador de tráfego.

### **Otimização de Roteamento**

A otimização de roteamento consiste em distribuir o tráfego na rede de forma ótima. O termo roteamento nas redes de pacotes descreve duas atividades: cálculo de rotas e encaminhamento de pacotes. As tecnologias de roteamento então devem realizar as seguintes tarefas:

- descobrir e divulgar estados da rede e mudanças nos enlaces (falhas, carga etc.);
- cálculo de rotas na rede;
- encaminhamento de pacotes da origem ao destino.

A otimização do roteamento descreve o processo de melhorar uma solução de rede ao encontrar os caminhos mais vantajosos para uma relação específica, sem alterar a estrutura da rede. Tradicionalmente, o projeto de redes tem-se focado em otimizar o custo de desempenho ao ajustar a topologia de rede, a capacidade dos enlaces e as estratégias de roteamento. Nessas estratégias, incluem-se as tecnologias de protocolos de roteamento.

Ao conhecer os padrões de tráfego entre os nós da rede, uma alternativa para a otimização parte da medição de congestionamento no enlace em termos do tráfego médio que carrega. Considere que  $F_{ij}$  seja o fluxo do enlace  $(i,j)$ . Este fluxo pode ser expresso em medidas convenientes de dados transmitidos por unidade de tempo, tais como pacotes/segundo ou bits/segundo. O processo de otimização precisa definir uma função a ser otimizada. Essa função pode ter a forma da Eq. 2.1, onde  $D_{ij}$  é definida como uma função monótona crescente que representa o custo (Bertsekas D., Gallager R., 1992).

$$\sum_{(i,j)} D_{ij}(F_{ij}) \quad (2.1)$$

A forma da função da Eq. 2.1 pode relacionar o fluxo do enlace à sua capacidade com o cálculo de um parâmetro de QoS, por exemplo o atraso. Em uma rede com diferentes nós, o problema do roteamento ótimo pode ser formulado da forma como aparece na Tabela 2.2.

Pode-se observar que existe uma estreita relação entre a otimização de roteamento e a capacidade dos enlaces e que a otimização não pode ser avaliada sem ter passado pelo dimensionamento da rede. Assim também, as restrições impostas pelos algoritmos de roteamento, a serem estudados no próximo capítulo, e sua boa compreensão são fundamentais para este processo.

Tabela 2.2. Formulação geral para o problema otimização de roteamento

definições	<p><math>A =</math> conjunto de todos os nós origem-destino, onde <math>a=(i,j)</math> é uma dupla de <math>A</math></p> <p><math>P_a =</math> conjunto de todos os caminhos que conectam um par <math>a</math></p> <p><math>x_p =</math> fluxo do caminho <math>p</math></p> <p><math>r_a =</math> fluxo que entra na rede no nó <math>i</math> e que vai para o nó destino de <math>j</math> (em unidades de dados/segundo).</p> <p><math>F_{ij} = \sum_{\substack{\text{todos os caminhos } p \\ \text{que contém}(i,j)}} x_p</math></p>
restrições	<p><math>r_a = \sum_{p \in P_a} x_p \quad \forall a \in A,</math></p> <p><math>x_p \geq 0 \quad \forall p \in P_a, a \in A</math></p>
minimizar	<p><math>\sum_{(i,j)} D_{ij}(F_{ij}) =</math> função de custo</p>

Ao otimizar, assume-se que  $D_{ij}$  é uma função diferenciável de  $F_{ij}$ , definida no intervalo  $[0, C_{ij})$ , com  $C_{ij} > 0$ . Seja  $\mathbf{x}$  vetor de fluxos  $x_p$ . A Eq. 2.2 a seguir define a função de custo do problema de otimização e a Eq. 2.3, a derivada parcial de  $D$  em relação aos fluxos do caminho  $p$ , denotados por  $x_p$ . O termo  $D'_{ij}$  é avaliado nos fluxos totais de  $x$ .

$$D(x) = \sum_{(i,j)} D_{ij} \left( \sum_{\substack{\text{todos os } p \\ \text{que contém}(i,j)}} x_p \right) \quad (2.2)$$

$$\frac{\partial D(x)}{\partial x_p} = \sum_{\substack{\text{todos os enlaces}(i,j) \\ \text{de } p}} D'_{ij} \quad (2.3)$$

A derivada parcial representa o comprimento do caminho  $p$  quando o comprimento de cada enlace  $(i,j)$  é tido como a primeira derivada  $D'_{ij}$  avaliada em  $x$ . Esta derivada parcial é chamada de primeira derivativa de comprimento do caminho  $p$ . Se  $x^* = \{x_p^*\}$  é um vetor ótimo de caminho de fluxo e  $x_p^* > 0$  para algum caminho  $p$  de um par  $a$ , deve ser possível deslocar uma quantidade de tráfego deste caminho  $p$  para outro caminho  $p'$  do mesmo par  $a$  sem melhorar o custo, pois em caso contrário  $x^*$  não seria ótimo. Dessa forma, se  $y$  é essa quantidade de tráfego, a Eq. 2.4 expressa esta relação. Ou seja, a mudança no custo deve ser positiva, de tal forma que se conclui a relação expressa na equação 2.5, para  $x_p^* > 0$ .

$$y \frac{\partial D(x^*)}{\partial x_p} - y \frac{\partial D(x^*)}{\partial x_p} \leq 0 \quad (2.4)$$

$$\frac{\partial D(x^*)}{\partial x_p} \geq \frac{\partial D(x^*)}{\partial x_p} \quad (2.5)$$

A conclusão principal das equações 2.4 e 2.5 é que o roteamento ótimo direciona o tráfego exclusivamente nos caminhos mais curtos em relação aos pesos dos enlaces. Na teoria de grafos, o comprimento do enlace é equivalente ao peso e pode ser expresso como  $G=(V,A)$  em que  $V$  é o conjunto de nós,  $A$  é o mesmo da Tabela 2.2, e  $w_{ij}$ , o peso para cada aresta  $(i,j) \in A$ . Ao otimizar o roteamento,  $w_{ij}$  deve ser expresso como um valor que relaciona o fluxo de carga no enlace em relação à sua capacidade. Em uma rede, esse conhecimento produz mecanismos de roteamento adaptativo.

A otimização de roteamento focada na engenharia de tráfego procura os caminhos que permitam alcançar as métricas de QoS pré-estabelecidas para uma matriz de tráfego determinada ou um estado de tráfego conhecido, tentando diminuir as capacidades dos enlaces.

As métricas dependem da topologia da rede e da natureza do tráfego. Então, a boa caracterização do tráfego é fundamental nesse processo.

### **Distribuição de Capacidade**

Com o conhecimento das demandas do tráfego, a topologia da rede e um padrão fixo de roteamento, as capacidades dos enlaces podem ser estabelecidas. Pode-se observar que a otimização do roteamento e a distribuição da capacidade são atividades muito relacionadas, pois a capacidade do enlace é uma das variáveis utilizadas para a otimização do roteamento. Espera-se que a capacidade e fluxo em cada enlace cumpram certas restrições de desempenho assim como de confiabilidade enquanto os custos são minimizados, o que resulta na maioria dos casos em um problema conjunto complexo. Em redes com vários tipos de serviço, utilizam-se tecnologias em que com técnicas de escalonamento a capacidade do enlace é dividida em subgrupos nos quais cada classe de tráfego é mapeada com um método específico. Os elementos de desempenho frente ao tráfego agregado e a sua interação com várias classes ainda não são totalmente entendidos. Alguns trabalhos (Wu K. e Reeves S, 2003) estudam a possibilidade de planificação da capacidade para duas classes de tráfego em um ambiente de serviços diferenciados. É observado que a correta

caracterização do tráfego para estes processos é de extrema importância e alguns trabalhos focam a metodologia de dimensionamento em classificações tais como tráfego elástico e tráfego de *stream* (Riedl A., 2003).

O problema de roteamento ótimo se dá quando, dado um enlace  $(i,j)$ , o custo da rede deve ser otimizado em função das capacidades  $C_{ij}$  e dos fluxos  $F_{ij}$ . O problema pode ser resolvido ao minimizar o custo com respeito a  $F_{ij}$  e depois obter as capacidades ótimas em uma equação que expresse  $C_{ij}$  em função de  $F_{ij}$ . O resultado deste processo pode ser uma topologia ótima em custo, mas muitas vezes fraca em conectividade, ao concentrar o fluxo em poucos enlaces. Na prática, é um enfoque pouco utilizado e uma opção alternativa é o uso de métodos heurísticos.

Um método heurístico utiliza uma topologia de rede específica e de forma sucessiva, perturba esta topologia ao mudar parâmetros em busca do atendimento de uma função objetivo. Os passos de uma heurística consideram que:

- (a) os nós da rede e as demandas de tráfego para cada par de nós são conhecidos;
- (b) um modelo de roteamento é utilizado de tal forma que defina os fluxos  $F_{ij}$  nos enlaces  $(i,j)$ , dado um conjunto de capacidade  $C_{ij}$ .

A definição de fluxos normalmente se faz ao otimizar o roteamento em função de uma métrica de QoS, onde comumente é mais utilizado o atraso médio por pacote. Existem então as restrições de QoS, as de confiabilidade que dependem do grau de conectividade da rede, mencionadas anteriormente, e as de custo.

Considere uma rede representada por  $G=(V,A)$ , em que  $V$  é o conjunto de nós,  $A$  o conjunto de arestas e  $c_{ij}$  a capacidade de cada enlace  $(i,j) \in A$ . Inicialmente,  $G$  representa a topologia a ser testada. A variável  $G\_Melhor$  representa a topologia atual, que na primeira execução da heurística é a primeira topologia estabelecida, após o processo de dimensionamento. Os passos para a heurística são enumerados a seguir (Bertsekas D., Gallager R., 1992, com adaptações):

1. Calcular os fluxos  $F_{ij}$  em  $G$  com o método de roteamento (como definido anteriormente em (b))
2. Se  $Métrica\_QoS(G, F) > Metricas\_Requeridas$  ir ao passo 1.
3.  $Conf \leftarrow$  Verificar confiabilidade para  $G$
4. Se  $Conf < Conf\_Requerido$ , ir ao passo 8
5.  $Custo \leftarrow$  Verificar o custo para  $G$
6.  $Custo\_G\_Melhor \leftarrow$  Verificar o custo para  $G\_Melhor$
7. Se  $Custo < Custo\_G\_Melhor$   
então  $G\_Melhor \leftarrow G$
8.  $G \leftarrow$  Gerar uma nova topologia ao modificar a capacidade de um ou vários enlaces
9. Ir ao passo 1.

Uma heurística, como a mostrada acima, não garante uma solução ótima, e esse procedimento é, na área de redes um tópico aberto a melhoras e novos procedimentos (Bertsekas e Gallager, 1992, Gerla M. e Kleinrock L, 1977, Youserfí'zadeh H., Jonckheere E., Silvester, J., 2003). Pode-se observar a relação entre os diferentes passos e pontualmente, no caso dos passos 1,2 e 8, o grau de complexidade de cada procedimento. Particularmente, no passo 2, é fundamental a existência de um modelo adequado de tratamento de tráfego para obter métricas de QoS que correspondam à realidade. Por outro lado, o passo 8 deve trabalhar de tal forma que as novas topologias geradas tendam a oferecer melhores soluções que as anteriores, para evitar longos tempos de processamento. O passo 1 deve ter um critério de roteamento, que considere os fluxos, suas características e as capacidades dos enlaces. É nesses tópicos então, que se foca parte da proposta deste trabalho de pesquisa.

### **Especificações de Domínio**

As especificações de domínio são orientadas a arquiteturas de serviços, onde se procura identificar entidades centrais tais como gerenciadores de recursos ou servidores de políticas. Os procedimentos relativos não são objetos de estudo deste trabalho e não serão abordados no texto.

#### **2.1.2. Gerência da Rede**

No momento em que a rede se encontra operativa, o processo de gerência descreve o acompanhamento e a influência na operação para manter uma qualidade de serviço

específica. Isso implica o estabelecimento de atividades e procedimentos de monitoração que resultem em decisões de gerência para manter a capacidade operativa da rede.

O tópico de gerência de redes é hoje um dos assuntos mais abrangentes e importantes na área. No contexto deste trabalho, as atividades de gerência de interesse são as relacionadas à otimização de roteamento e a utilização de banda, que, pela sua importância hoje, representam o processo de engenharia de tráfego e, no processo evolutivo para redes convergentes multimídia, têm incorporado ainda mais relevância.

A engenharia de tráfego permite que um provedor possa ajustar sua rede para permitir níveis de serviço que não impliquem necessariamente investimentos adicionais em infraestrutura.

Em redes multimídia, os pacotes que pertencem a fluxos de tráfego diferentes com endereços de origem e destino iguais podem atravessar a rede por caminhos diferentes. Esses caminhos são escolhidos em função da sua capacidade de prover os requerimentos de QoS das aplicações associadas a esses fluxos. Constantemente, esses caminhos podem ser modificados, ao agregar diferentes fluxos para manter o desempenho dentro dos limites requeridos. A alocação de banda também é um processo fundamental no processo de engenharia de tráfego, pois complementa o processo de alocação de caminhos, ao movimentar largura de banda entre diferentes fluxos e influenciar diretamente nas métricas de QoS e custos operacionais da rede.

O equilíbrio entre estas variáveis tem motivado um estudo mais específico da modelagem do tráfego como um dos fatores fundamentais do processo de engenharia de tráfego.

## **2.2. ENGENHARIA DE TRÁFEGO EM REDES MULTIMÍDIA**

No modelo de referência TCP/IP, a camada 3 ou camada Internet tem como função principal o roteamento de datagramas. Os nós precisam regularmente de entregas indiretas de datagramas e, para esse efeito, são utilizados os roteadores.

No datagrama IP, existe o campo Endereço Destino que contém o número IP do nó para o qual o datagrama deve ser enviado. Quando o datagrama entra na rede, o roteador deve decidir que caminho será percorrido para chegar ao seu destino. Podem existir vários caminhos possíveis, o que motiva a necessidade de mecanismos na rede que decidam qual caminho deve ser tomado.

Estes mecanismos são definidos por protocolos de roteamento, que podem ser divididos em externos e internos. Os protocolos de roteamento interno encaminham datagramas no mesmo domínio, enquanto os protocolos de roteamento externo fazem o roteamento entre domínios diferentes. Os roteadores internos ou de trânsito encaminham pacotes entre roteadores e efetuam o processo de encaminhamento IP.

Nos roteadores reside a tabela de encaminhamento IP, também chamada de FIB (*Forwarding Information Base*). Quando um roteador concentra vários enlaces e várias rotas, a tabela de roteamento cresce e para cada datagrama deve-se realizar uma busca, o que implica em tempos de recuperação altos e, conseqüentemente, maiores atrasos (Comer D., 2000).

Uma alternativa é utilizar, em vez de um processo de roteamento, um processo de comutação. No caso da comutação, podem ser utilizadas operações de indexação para a pesquisa em tabelas, o que implica em uma menor complexidade computacional.

No conceito de comutação por rótulos, é introduzido no pacote um rótulo que serve de índice para a busca na tabela de encaminhamento. Essa busca é realizada por meio de um acesso direto à tabela de comutação, ou seja, possui uma complexidade mais baixa que a do algoritmo LPM (*Longest Prefix Match*), utilizado nas redes IP. Uma vez encontrada a entrada correspondente ao rótulo, o rótulo do pacote é substituído por aquele correspondente da tabela e envia-se o pacote. Este processo é bem mais simples e mais rápido que o procedimento de encaminhamento nas redes IP.

Até 1996, existiam várias tecnologias que procuravam utilizar a comutação de rótulos para encaminhamento de pacotes. Entre estas temos a IP Switching (Ipsilon), CSR – Cell Switched Router (Toshiba), Tab Switching (Cisco Systems), ARIS –Aggregate Route-based IP Switching (IBM) e SITA – Switching IP Through ATM (Telecom Finland). Diante disso, a força tarefa de Engenharia de Internet (IETF – Internet Engineering Tasking Force) criou um grupo de trabalho com o objetivo de padronizar estas tecnologias e surgiu dessa forma o padrão MPLS (*Multi-protocol Label Switching*).

O MPLS é um modelo híbrido que surgiu como uma forma de tirar proveito das melhores propriedades do roteamento de pacotes e da comutação de circuitos e formou-se a partir da padronização de diversos protocolos que deram origem ao conceito de comutação por rótulos.



A comutação por rótulos busca aproveitar, ao mesmo tempo, as vantagens da camada 3, como flexibilidade e alcançabilidade, e a velocidade de encaminhamento oferecida pela camada 2. Pode, por exemplo, fazer a seleção de caminhos em função do destino final e considerar parâmetros de QoS de camada 2, diferente da dinâmica do encaminhamento IP, no qual é importante somente saber o destino imediato (paradigma *hop by hop*). Outra motivação para o uso de rótulos é identificar diferentes tipos de fluxos e tratá-los de forma diferenciada, ou seja, suporte à qualidade de serviço e engenharia de tráfego. Assim também, esta tecnologia visa a implementação de novos serviços, tais como VPN (*Virtual Private Network*), gerenciamento e provisionamento mais simples para provedores e assinantes de rede.

### **2.2.1. Tecnologia MPLS**

A primeira estrutura do MPLS é a de controle. No plano de controle, estão localizadas as funções de controle, tais como sinalização, roteamento, conversão de endereços, policiamento de tráfego, entre outras. Neste plano, a representação da rede é a mesma dos protocolos de roteamento (Alwayn V., 2002).

A segunda estrutura do MPLS é a de encaminhamento, cuja operação é ditada pelo plano de controle e agrega funções relativas à propagação de datagramas IP, como por exemplo, encapsulamento, segmentação, remontagem e rotulação de datagrama.

A superposição de um rótulo ao datagrama tem a propriedade de imprimir à comunicação uma característica de orientação à conexão.

Um rótulo é um identificador de tamanho fixo e significado local. Quando um pacote entra em um domínio MPLS, a ele é atribuído um rótulo (*label*). Um domínio MPLS é formado por um conjunto de roteadores de núcleo e um conjunto de roteadores de borda, que interligam subdomínios. O rótulo permite o desacoplamento entre o roteamento e o encaminhamento. Assim, os roteadores só analisam os rótulos para poder encaminhar o pacote.

Em redes que já trabalham com circuitos virtuais, a estrutura do pacote permite a introdução do rótulo sem necessidade de adicionar novos campos.

A figura 2.7 mostra os principais componentes de uma arquitetura MPLS, como definidas em Rosen E, 2001.

A FEC (*Forwarding Equivalence Class*) é uma representação de um grupo de pacotes que compartilham das mesmas exigências para o transporte. Todos os pacotes pertencentes a este grupo recebem o mesmo tratamento por todo encaminhamento (origem-destino).

Os LER e os LSR são os roteadores de comutação de rótulos. O LER (*Label Edge Router*) está situado na entrada do domínio MPLS e tem a função de inserir rótulos aos pacotes, agrupá-los em uma FEC e encaminhá-los por meio de um LSP (*Label Switched Path*). Quando o pacote chega no LER, este deve fazer a primeira colocação do rótulo com base em uma busca do IP destino baseado no prefixo da rede. O rótulo é colocado no pacote e este é encaminhado ao roteador adjacente seguinte. Quando o LER está situado na saída do domínio MPLS, é responsável pela retirada dos rótulos e pela entrega dos pacotes ao destino.

Os LSR (*Label Switch Router*) do núcleo têm a função de encaminhar os pacotes com base apenas no rótulo. Ao receber um pacote, cada LSR troca o rótulo existente por outro, passa o pacote para o próximo LSR e assim por diante até chegar no LER de saída. O LER somente pode fazer encaminhamento. Ele utiliza o valor do rótulo para obter a interface de saída. É importante notar que o pacote é encaminhado sem a necessidade de analisar o cabeçalho da camada de rede. Essa é uma diferença fundamental para o processo de encaminhamento IP mencionado anteriormente. Quando o pacote alcança o roteador de saída da rede, o rótulo é retirado e o pacote é entregue ao destinatário.

O *Label Switch Path* (LSP) é um caminho percorrido pelos pacotes na rede. Como a criação do LSP ocorre somente na entrada de uma rede MPLS, os demais roteadores, ou seja, os LSR do núcleo, irão somente chavear os rótulos ao encaminhar o pacote de acordo com a LSP predeterminada, sem a necessidade de fazer o roteamento dos pacotes. Os rótulos são distribuídos no momento do estabelecimento das LSP.

Um LSP é unidirecional, portanto é preciso ter dois LSP para uma comunicação entre duas entidades.

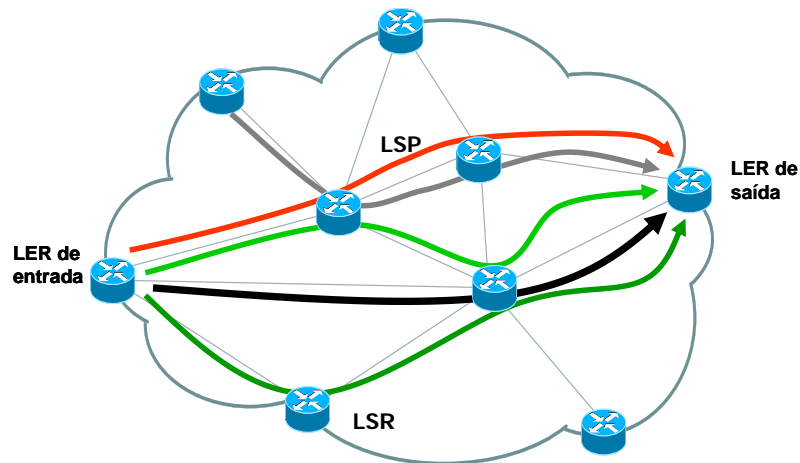


Figura 2.7. Componentes LER, LSR e LSP.

Cada FEC é representada na entrada do domínio por um rótulo e cada LSP é associado a uma FEC. Ao receber um pacote, o LER da rede MPLS verifica a qual FEC ele pertence e o encaminha através da LSP correspondente. Portanto, há uma associação pacote-rótulo-FEC-LSP.

Quando o pacote entra numa rede MPLS, a ele é associado a uma FEC e então é criada uma LSP para esta FEC. Um fato importante é que para cada FEC existe um LSP, mas um LSP pode servir várias FECs. A associação pacote-FEC acontece somente uma vez, quando os pacotes entram em um domínio MPLS.

A FEC pode ser determinada por um ou mais parâmetros. Alguns desses parâmetros são: endereço IP da fonte ou destino ou endereço IP da rede; número da porta da fonte ou destino; ID do protocolo IP e métricas de QoS.

O MPLS apresenta dois passos para solucionar o problema de encaminhamento de pacotes na rede. No primeiro passo, são analisados todos os pacotes ingressantes e classificados em FECs. No segundo passo, é localizada a interface associada à FEC para envio aos LSR intermediários. Esses procedimentos utilizam um único rótulo para identificar determinada FEC, que funcionalmente é um identificador de tamanho fixo de 20 bits, afixado no cabeçalho de cada pacote, e que se torna um indexador para todas as decisões tomadas com relação ao encaminhamento do pacote. O processo de análise do rótulo, chamado também de *Label Swapping* (Rosen E, 2001) tem como essência de que em cada salto, o LSR descarta o rótulo existente e aplica um novo rótulo que indica ao próximo nó como enviar o pacote. Esses valores não são constantes e podem mudar de roteador para roteador.

A análise de cabeçalho do nível de enlace, além de determinar o próximo salto serve também para aplicar políticas de prioridades, pois no MPLS os protocolos não determinam a prioridade ou classe de serviços com base no cabeçalho e sim na verificação a que classe de serviços ou precedência as FECs dos pacotes pertencem, indicação dada pelo rótulo. Dessa forma, se observa que um rótulo possui as funções de representar, ao mesmo tempo, rota e classes de serviço de um pacote.

O MPLS é formado por dois planos: o plano de controle e o plano de encaminhamento. O plano de encaminhamento utiliza os rótulos contidos nos pacotes e na tabela de encaminhamento LFIB (*Label Forwarding Information Base*) de cada LSR para encaminhar os pacotes. O plano de controle é responsável por manter essas tabelas entre um determinado número de LSR. A figura 2.8 mostra a estrutura de um nó MPLS e o funcionamento dos dois planos.

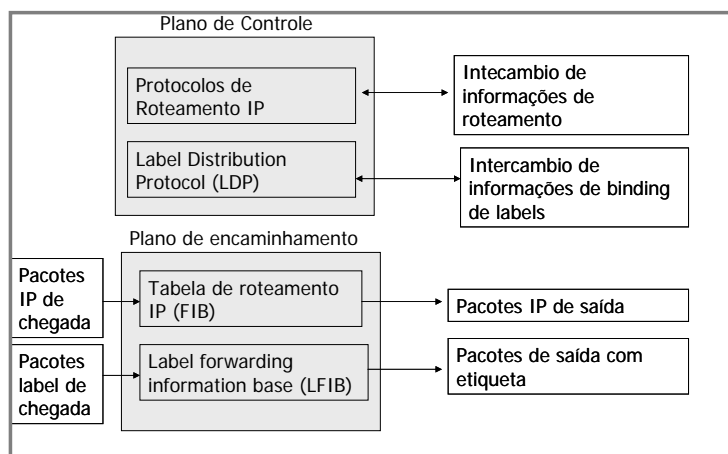


Figura 2.8. Estrutura de um nó MPLS

Na borda de uma rede MPLS, os LER fazem a classificação e decisões de encaminhamento ao examinar o cabeçalho dos pacotes IP.

O rótulo serve como uma representação curta para o cabeçalho IP e reduz a complexidade de processamento nos próximos nós do caminho. O rótulo gerado durante o processamento do cabeçalho no LER é usado nos próximos nós da rede para suas respectivas decisões de encaminhamento. Quando o pacote emerge do interior de uma rede MPLS, os LER descobrem que devem encaminhar pacotes a uma interface não-rotulada e simplesmente removem qualquer encapsulamento de pacote antes de os encaminhar. (Davie B., e Rekhter J., 2000).

Quando um LSR recebe um pacote com rótulo, este é extraído e usado como índice até ser encontrado na tabela de encaminhamento LFIB residente no LSR. A seguir, é verificado qual deve ser a interface de saída e o rótulo de saída é adicionado ao pacote, que é então enviado à interface de saída.

O mais importante é que apenas um único algoritmo de encaminhamento é necessário para todos os tipos de comutação. Isso pode ser implementado em *hardware* para maior velocidade.

Essa simplicidade no processo de encaminhamento possibilita a separação do plano de encaminhamento do plano de controle. Portanto, novas funções de roteamento (controle) podem ser desenvolvidas sem que haja a necessidade de modificar o processo de encaminhamento.

A separação das funcionalidades permite a introdução de novos procedimentos para roteamento, independente de modificar os outros planos do protocolo.

Em um LSP, o fluxo segue na direção do LER de ingresso para o LER de egresso. Para um determinado enlace, o LSR que transmite o fluxo é denominado *upstream* LSR e o que recebe é o *downstream* LSR.

Portanto, quando um *downstream* LSR recebe um pacote com o rótulo, é necessário que seja tomada uma decisão. A responsabilidade dessa tarefa pertence ao plano de controle, que utiliza tabelas de encaminhamento LFIB como base para essa decisão. Cada roteador deve possuir a sua tabela.

### **2.2.2. Problemas da Engenharia de Tráfego (TE)**

A engenharia de tráfego tem como enfoque a otimização do desempenho da rede, em relação à aplicação de tecnologias e princípios científicos para a medição, modelagem, caracterização e controle de tráfego. A otimização do desempenho pode ser orientada ao tráfego ou aos recursos. É hoje a principal aplicação do MPLS devido a possibilidade de controlar o fluxo de tráfego na rede, com o objetivo de reduzir problemas de congestionamento e conseguir uma utilização homogênea dos recursos disponíveis.

Dentro do contexto operacional de uma rede, a TE é um processo que funciona dentro da gerência da rede. É hoje um processo de fundamental importância, pois permite a evolução

da rede em função de variações de carga de tráfego, comportamento e número de usuários e custos de operação.

O desempenho orientado ao tráfego inclui elementos que tem como objetivo cumprir requerimentos de QoS para fluxos de tráfego. Como exemplo, na Internet, com o modelo de serviço do melhor esforço que utiliza uma única classe de fluxo, os objetivos de QoS incluem: minimização de perda de pacotes, minimização de atrasos, maximização de processamento e execução de acordos de nível de serviço (*Service Level Agreements – SLA*). Em contrapartida, o desempenho orientado aos recursos se concentra na otimização da utilização de recursos de rede. No contexto deste trabalho, o interesse se concentra no desempenho orientado a tráfego.

O congestionamento, que produz perda de pacotes e atrasos, se manifesta em dois cenários: quando os recursos da rede são insuficientes ou inadequados para acomodar a carga oferecida e quando os fluxos de tráfego são mapeados de forma ineficiente sobre os recursos disponíveis, o que cria um problema de balanço na rede, onde alguns recursos são sobre-utilizados enquanto outros são inutilizados.

Um tronco de tráfego é um agregado de fluxos de tráfego que pertencem à mesma classe. Nos modelos de serviço de classe única, como o de melhor esforço da Internet, um tronco de tráfego encapsula todo o tráfego entre um LSR de ingresso e um LSR de egresso. Operacionalmente, um tronco de tráfego é um objeto roteável e é distinto de um LSP pois pode ser movido de um caminho para outro.

Na prática, um tronco de tráfego pode ser caracterizado pelos LSRs de ingresso e egresso. Uma FEC é mapeada sobre ele assim com um conjunto de atributos que determinam suas características comportamentais. Conceitualmente, os troncos de tráfego são unidirecionais, mas pode-se definir o BTT (*Bidirectional Traffic Trunk*) que consiste em dois troncos de tráfego que são acoplados logicamente, com um tronco que transmite tráfego de um nó de origem para um nó destino e um tronco que leva tráfego do nó de destino para o nó de origem. Ambos os troncos podem ser inicializados por ações atômicas de um LSR e nenhum dos troncos de tráfego pode existir sem o outro. Este tipo de troncos, mencionados na RFC 2702, são apenas umas conveniências administrativas e na prática, a maioria das funções da engenharia de tráfego consideram os troncos do tipo unidirecional.

Os parâmetros de tráfego podem ser usados para capturar as características dos fluxos de tráfego a serem transportados pelo tronco. Tais características surgem da correta caracterização do tráfego que permita indicar as exigências de recursos que o tronco precisa e prever o comportamento dessa agregação de fluxos na topologia operacional da rede.

As operações básicas em um tronco de tráfego se resumem a seguir:

- a) Estabelecimento ou criação de um tronco de tráfego.
- b) Ativação do tronco, que consiste em que comece a passar tráfego.
- c) Desativação, processo inverso de (b).
- d) Modificação de atributos, que consiste em alterar parâmetros que influenciam as características comportamentais do tronco.
- e) Re-encaminhamento, que consiste em mudar a rota pela qual passa o tronco.
- f) Destruir, que consiste em remover o tronco de tráfego e recuperar todos os recursos alocados a ele.
- g) Policiamento e condicionamento do tráfego, que determina as ações que deveriam ser tomadas quando um tronco de tráfego excede seu contrato especificado nos parâmetros de tráfego.

A distribuição da carga por múltiplos troncos de tráfego paralelos entre dois nós permite que no caso de que o tráfego agregado ultrapasse a capacidade do enlace para um conjunto de métricas de QoS, essa carga poderia ser dividida adequadamente em sub-fluxos e dirigi-los através de múltiplos caminhos entre dois nós (Mittra D. e Ramakrishnan K.G, 2001).

### **2.2.3. Métricas de QoS**

Para que uma rede suporte vários requerimentos de QoS, é necessário que o roteamento considere várias métricas, tais como banda, atraso, probabilidade de perda de pacotes, e que seja capaz de encontrar caminhos que satisfaçam esses requerimentos. O conjunto de métricas influenciam a complexidade do processo de identificação de rotas ótimas.

Considere-se uma rede  $G=(V,A)$ , com  $V$  como o número de nós e  $A$  o número de enlaces. Seja  $m(i,j)$  uma métrica que representa certa propriedade do enlace  $(i,j)$  e um caminho  $p=(i,j,\dots,n,o)$  que assume que a mesma propriedade  $m(p)$  é uma função de  $m(i,j),m(j,k)\dots m(n,o)$  e que as restrições em  $m(p)$  definem a complexidade do cômputo do

caminho. Neste conjunto de definições, se definem as métricas classificadas a seguir (Wang Z, 1999):

- a) Métrica aditiva se  $m(p)=m(i,j)+m(j,k)+\dots+m(n,o)$
- b) Métrica multiplicativa se  $m(p)=m(i,j)\times m(j,k)\times\dots\times m(n,o)$
- c) Métrica côncava se  $m(p)=\min[m(i,j),m(j,k),\dots,m(n,o)]$
- d) Métrica convexa se  $m(p)=\max(m(i,j),m(j,k),\dots,m(n,o))$

Ao considerar os itens de (a) até (d) pode-se identificar que atraso, jitter e custo são métricas aditivas. A perda de pacotes segue uma regra de composição  $m(p)=1-(1-m(i,j))(1-m(j,k))\dots(1-m(n,o))$  e é considerada uma métrica multiplicativa. A largura de banda é uma métrica côncava.

#### **2.2.4. Complexidade computacional para a engenharia de tráfego**

O roteamento com base em restrições (*Constraint Based Routing – CBR*) é usado para selecionar caminhos para troncos de tráfego sujeito a restrições impostas pelos itens que dizem respeito ao desempenho. Os atributos associados aos troncos de tráfego e recursos, assim como os parâmetros associados ao roteamento, representam de forma coletiva as variáveis de controle que podem ser modificadas por ação administrativa ou por agentes automatizados.

Os algoritmos do caminho mais curto, onde são utilizados protocolos tais como OSPF (*Open Shortest Path First*) e IS-IS (*Integrated Intermediate System-to-Intermediate System*), visam encontrar na rede, um caminho para alcançar o destino que represente o menor custo, em função de métricas associadas a cada enlace entre os nós. Este processo é com base no algoritmo de Dijkstra (Cormen T., et al, 2002) tem como resultado uma concentração do roteamento nestes enlaces, sem se considerar outros elementos da rede, tais como sobrecarga de recursos e requerimentos de atraso e latência para certa transmissão. Na engenharia de tráfego, o cômputo de um caminho modelado que satisfaz um conjunto arbitrário de restrições pode diferir do menor caminho e geralmente a métrica associada ao caminho corresponde à largura de banda do enlace.

Os modelos de roteamento podem ser classificados como *online* e *offline*. No primeiro, os pedidos chegam um atrás do outro e os pedidos futuros são desconhecidos. No segundo, todas as demandas são conhecidas com antecedência. Um modelo intermediário necessita ter um conhecimento estatístico do tráfego. Naturalmente, os algoritmos para roteamento



*online* precisam de uma maior rapidez mas, em função do processo dinâmico e operativo da rede, em muitas situações são mais desejáveis que os processos *offline*.

O tempo de complexidade de um algoritmo de roteamento é determinado pelo número de métricas consideradas (Banerjee G. e Sidhu D., 2002). Os problemas que trabalham com uma ou mais métricas côncavas e uma ou mais métricas aditivas podem ser resolvidos em tempo polinomial. Mas o problema de computar rotas ótimas sujeitas a restrições de uma ou mais métricas aditivas ou multiplicativas é conhecido como MCP (*Multiple Constrained Path Selection*) e constitui um problema NP completo. As únicas combinações que não produzem problemas NP completos se dão ao considerar a banda e um dos outros quatro parâmetros seguintes: atraso, jitter, custo e probabilidade de perdas (Wang Z., 1999).

### **2.2.5. O problema de otimização com base na caracterização do tráfego**

A engenharia de tráfego no MPLS apresenta basicamente três problemas (Awduche D. et al, 1999): (a) como mapear pacotes em FECs, (b) como mapear FECs em troncos de tráfego e (c) como mapear troncos de tráfego sobre uma topologia de uma rede física através de LSPs.

O problema em (a) tem sido muito pouco abordado e para sua solução é necessária uma correta caracterização do tráfego, que permita estabelecer um processo de decisão com base nas características estatísticas do tráfego, seus requerimentos de QoS e seu impacto na agregação de um tronco já existente. Assim também, a conformação de um tronco de tráfego que agrupe várias FECs com características similares pode ser utilizado como um processo que promova o balanceamento de carga na rede e a sua decisão sem intervenção humana permitiria uma maior escalabilidade da rede e insumo para o seu planejamento, além de promover a estabilidade operativa.

Alguns autores consideram o processo de mapeamento de fluxos de tráfego em uma topologia física como um dos itens de maior importância no processo de TE (Girish M., Zhou B., Hu J., 2000, Banerjee G. e Sidhu D., 2002, Awduche D., et al, 1999, Alves Resende R, 2001, Gomes D., Agoulmine N., de Souza J.N, 2004).

Uma vez que um caminho satisfatório é calculado, um protocolo de sinalização é usado para estabelecer um LSP ao longo daquele caminho, e o tráfego que satisfaz uma

determinada FEC é enviado pelo LSP. Na perspectiva da TE, o cômputo desse caminho para um LSP visa satisfazer um conjunto de exigências associadas aos troncos de tráfego a serem levados pelo LSP. Na formulação do roteamento com base em restrições, são definidos a topologia da rede e um conjunto de atributos relativos aos recursos e suas restrições. As demandas dos fluxos de tráfego ou os LSPs que devem ser roteados na rede são também descritos e então, o problema consiste em selecionar o posicionamento ótimo dos LSPs através da rede ao se aderir a estas restrições.

O problema do re-roteamento ocorre quando em função de um número de variáveis, tais como carga excessiva, falhas ou mudança de topologia, os caminhos LSPs devem ser alterados ou re-computados. Este problema consiste em um problema de admissão que consiste em determinar se um LSP pode ser admitido ou não, e no caso de que seja, qual é o caminho para esse LSP.

As capacidades dos enlaces não podem ser excedidas com a adição de novos LSPs, o fluxo agregado não pode ultrapassar a capacidade disponível ao longo de todo o caminho do LSP e o nó de entrada e saída devem ter o LSP com origem e terminação neles respectivamente. Então, para o re-roteamento, o problema do controle de admissão se aplica a um conjunto de LSPs que precisam ser re-allocados (Mitra D. e Ramakrishnan K.G, 2001).

Existe uma preocupação em tentar minimizar o re-roteamento de muitos LSPs, por exemplo no caso de uma falha, com o objetivo de prevenir oscilações. Isto não é desejável, pois, quando um nó ou enlace entra em operação novamente, é necessário processar para todos os LSPs as rotas ótimas que pode resultar em mudanças substanciais nas rotas existentes e paralelamente existe um maior consumo de tempo no processo de otimização e degradação de performance em redes operativas.

Muitos fabricantes de equipamentos MPLS (Alwayn V., 2002) têm desenvolvido soluções de engenharia de tráfego que trabalham com extensões dos protocolos OSPF e IS-IS, de tal forma que as informações sobre as características de QoS de nós e enlaces sejam disseminadas na rede. Dessa forma, os algoritmos padrões de Dijkstra são aplicados a topologias de rede podadas com base na informação de QoS. Este método, embora funcional, não soluciona a natureza NP do roteamento e proporciona soluções sub-otimas.

Outro dos problemas na área de TE é desenhar uma topologia de rede que supra um conjunto de demandas solicitadas pelos LSPs (Iovanna P., Sabella R., Settembre M, 2003).

Existem diversos trabalhos científicos que têm procurado dar soluções ao problema da otimização do roteamento em redes MPLS. Em Pereira T. e Ling L.L (2002), é mostrada uma estratégia OSPF adaptativa, com base na estimativa efetiva de banda. A estimativa de banda é realizada com base em um método que foca sua atenção na aproximação para tráfego auto-similar, que considera uma função de otimização com base no tamanho do buffer e no parâmetro de Hurst. O algoritmo OSPF atualiza os custos dos enlaces com base na estimativa de banda efetiva de cada interface dos roteadores. A função de otimização somente considera um critério de perda de pacotes, sem considerar o atraso e percentual de perda de pacotes.

Em Rasiah P. e Chung J. (2002) é investigado o uso de RCS (*Rate Controlled Service*) em uma rede de melhor esforço para estabelecer limites de atraso e perda de pacotes e em função disso otimizar o estabelecimento dos LSPs. É utilizada uma modificação do algoritmo de Dijkstra com restrições de banda para computar os LSPs e um modelo de tráfego conhecido como modelo *leaky bucket*, que toma como parâmetro a rajada máxima de tráfego a ser enviada na rede. Porém, em outros trabalhos (Fonseca N.L.S, M. Gilberto e Neto, 2000, Mayor G. e Silvester J, 1997), esse modelo de tráfego mostrou limites de atraso muito altos em relação aos observados em simulações e aos que caracterizam o tráfego como auto-similar, o que pode resultar em um sobre-dimensionamento das capacidades da rede e, em conseqüência, em custos maiores de operação.

Em (Elwalid et al, 2001), é utilizado um modelo com base no trabalho de Bertsekas D. e Gallager R. (1992), formulação mostrada na Tabela 2.2. Neste trabalho é proposto um conjunto de algoritmos chamados de MATE (*MPLS Adaptive Traffic Engineering*) que tem como objetivo mapear o tráfego do tipo melhor esforço que não precisa reserva de banda. Com a utilização de medidas de desempenho da rede obtidas a partir de ‘pacotes teste’ que estabelecem estatísticas de atraso de pacotes e perda, são realizados os processos de monitoração da rede e balanceamento de carga. Os algoritmos mostram estabilidade e com o processo de balanceamento são reduzidas as taxas de perda. São mostrados modelos de tráfego que exibem curta dependência (Poisson) e processos discretos autoregressivos de ordem  $p$  (onde  $p$  determina a escala de tempo que exhibe a dependência e autocorrelação

do tráfego). No trabalho não se aprofunda o desempenho do processo de TE em redes reais com tráfego agregado e com diferentes graus de auto-similaridade.

Em Buriol L.S. et al. (2003), é mostrado um algoritmo genético híbrido para resolver o problema de estabelecimento dos pesos com roteamento OSPF. Com a utilização de uma função de custo definida em trabalhos anteriores pelos autores e a utilização de um algoritmo genético com busca local, é realizada a otimização de roteamento. Os autores argumentam que a dinâmica dos algoritmos do caminho mais curto é adequada e eficiente e que o problema se concentra em um estabelecimento mais assertivo dos pesos nos enlaces, ou seja, pesos que promovam uma melhor divisão do tráfego nos caminhos da rede, em vez de realizar extensões aos algoritmos que possam incrementar a sua complexidade.

Conforme as vertentes de pesquisa anteriormente mencionadas se reconhece a importância de um modelo de caracterização do tráfego para a obtenção de medidas de desempenho real em uma rede multimídia para o processo de engenharia de tráfego que abrange principalmente:

- a) O problema de caracterização dos fluxos, seu agrupamento em troncos de tráfego e a produção de métricas de QoS;
- b) Um processo de planejamento otimizado de rede que beneficie o desempenho da rede e evite oscilações e cálculos frequentes do processo de otimização de roteamento na introdução de novos troncos de tráfego, o que promove a estabilidade da rede;
- c) Que o processo de otimização se beneficie da técnica de otimização aplicada por algoritmos genéticos que permitam estabelecer para uma rede com um conjunto de restrições de QoS e custo máximo, um conjunto solução de topologias que considerem capacidades e posicionamento de LSPs, para fornecer insumos para o processo de expansão da rede.

Como foi mencionado anteriormente, o estudo dos modelos de tráfego para a obtenção de métricas reais de QoS é fundamental, e nesse sentido, o capítulo seguinte dissertará sobre esse assunto.

### **2.3. RESUMO DO CAPÍTULO**

Neste capítulo foram apresentados os conceitos básicos sobre o projeto de rede, os seus componentes e tarefas e a necessidade de métodos que permitam a sua aplicação nas redes operativas com tráfego multimídia.

Foram apresentados também os conceitos básicos da tecnologia MPLS e seus componentes, onde foi dado um destaque ao processo de engenharia de tráfego, onde foram enumerados os principais problemas e possíveis metodologias para implementá-la.

Com o intuito de limitar o escopo de aplicação e contribuição desta tese par ao leitor, dentro dos conceitos apresentados neste capítulo, foi apresentado um breve resumo do estado da arte e do interesse de pesquisas na área para a implementação dos processos de engenharia de tráfego, onde aparecem como elementos recorrentes, a necessidade de caracterização de tráfego em redes multimídia e a implementação de processos e métodos que permitam otimizar os processos de planejamento e roteamento de fluxos na rede.

### **3 - MODELOS DE TRÁFEGO PARA REDES MULTIMÍDIA**

O processo de Poisson estacionário, com base em incrementos independentes, tem sido amplamente utilizado para modelar o tráfego das redes de dados, modelagem essa que é herança do modelo proposto por Erlang para o tráfego telefônico (Willinger W. e Paxson V, 1998).

Um processo cujo comportamento em regime se caracteriza por picos de tráfegos que se alteram com tráfegos pequenos, em torno de uma média, como é o caso do processo de Poisson tem, como resultado, a compensação do uso da fila e o rápido escoamento do tráfego em certos nós da rede.

Quando o processo que modela o tráfego é um processo que apresenta dependência de longa duração, ou seja, é um processo estocástico lento caracterizado pelo fenômeno da persistência temporal, os picos de tráfego podem perdurar por um tempo suficientemente grande. Redes nas quais os tráfegos apresentam dependência de longa duração podem apresentar desempenho muito aquém do esperado.

Esses tipos de tráfego, caracterizados como processos auto-similares (Willinger W. e Paxson V, 1998, Leland W. E., Taqqu M. S., Willinger W., Wilson D. V., 1994), apresentam rajadas em diferentes escalas de tempo, ou seja, ao se observar um processo auto-similar em diferentes escalas de tempo, constata-se uma semelhança no padrão das amostras do processo. O parâmetro de Hurst, denotado por  $H$ , é o parâmetro que caracteriza o grau de auto-similaridade desses processos.

Quando vários tráfegos do tipo Poisson são agregados, o tráfego agregado resultante tende a suavizar-se, e facilita assim sua análise do tráfego. No agrupamento de tráfegos auto-similares individuais não ocorre suavização do tráfego agregado resultante, o que torna mais complicado o estudo analítico das redes sujeitas a esse tipo de tráfego.

Como foi mencionado anteriormente, parte da proposta deste trabalho é a utilização de medidas reais de QoS para o processo de otimização do planejamento da rede e seu roteamento. Nesse contexto, nas seguintes seções, descrevem-se o modelo de Markov e o modelo de tráfego auto-similar.

### 3.1. TRÁFEGO DE POISSON E TEORIA DE FILAS

Nas redes de voz, o modelo de tráfego de Poisson tem sido usado nos últimos cinquenta anos. Este modelo considera que o tráfego é formado por chamadas com duração que seguem uma distribuição exponencial. Outras considerações relativas a este tráfego é que as taxas de crescimento são previsíveis, o que permite ajuste e planificação de capacidade no longo prazo. Dessa forma, os modelos de tráfego nessas redes sempre foram tratáveis matematicamente e foram utilizados com sucesso no dimensionamento e na predição do desempenho.

A confiança derivada dos modelos de redes de voz influenciou a sua utilização nas incipientes redes de pacotes surgidas no século XX. O estudo analítico do desempenho das redes de dados é tradicionalmente fundamentado em modelos desenvolvidos a partir da teoria de filas para redes de voz (Gerla M. e Kleinrock L, 1977). O uso dessa teoria permite a definição e a utilização de um conjunto de equações das quais se pode extrair a estimativa de parâmetros importantes para o projeto de rede.

Nas redes de pacotes, os pacotes são transmitidos com o uso de técnicas de armazenamento e encaminhamento, de tal forma que um pacote que teve origem no nó  $s$  e deve ser entregue ao nó  $d$  é recebido e armazenado em uma fila em qualquer nó intermediário  $k$  enquanto espera a transmissão, e é encaminhado ao nó  $p$ , o nó seguinte na rota entre  $s$  e  $d$ , quando o canal  $(k, p)$  o permite. Ainda quando o canal está vazio, o pacote deve ser primeiro recebido no nó  $k$  antes que a transmissão para o nó  $p$  seja feita. A decisão do nó  $p$  é feita sobre uma política de roteamento pré-definida.

As necessidades de tráfego entre os nós ocorrem em intervalos variáveis e as necessidades dos requerimentos também são variáveis. Dessa forma, nos enlaces, são construídas filas de pacotes e o sistema se comporta como uma rede estocástica de filas.

No encaminhamento de pacotes, somente é necessário conhecer o destino do pacote. Então, os primeiros métodos analíticos para redes de pacotes consideraram a modelagem de uma rede de filas com  $n$  tipos de clientes, onde  $n$  é o número de destinos diferentes.

Com esse enfoque, o modelo analítico considera a medida de atraso de pacotes  $T$  como expresso na Eq. 3.1.

$$T = \sum_{\substack{j,k \\ j \neq k}} \frac{\gamma_{jk}}{\gamma} Z_{jk} \quad (3.1)$$

em que:

$\gamma_{jk}$  é a taxa média de pacotes do nó origem  $j$  ao nó destino  $k$ ;

$\gamma$  é a taxa média de pacotes da rede;

$Z_{jk}$  é o atraso médio de pacotes (enfileiramento e transmissão) de  $j$  e  $k$ .

Ao aplicar o teorema de Little (Bertsekas e Gallager, 1992) a esse modelo, pode-se chegar a uma expressão para  $T$  da forma mostrada na Eq. 3.2 (Gerla e Kleinrock, 1977).

$$T = \sum_{i=1}^b \frac{\lambda_i T_i}{\gamma} \quad (3.2)$$

em que  $b$  é o número de enlaces,  $\lambda_i$  é a taxa média de tráfego e  $T_i$  é o tempo médio de enfileiramento mais o atraso de transmissão no enlace  $i$ .

O conhecimento dos valores  $\lambda_i$  e  $T_i$  se fundamenta nas seguintes suposições:

- a) os processos externos de chegada seguem uma distribuição de Poisson;
- b) o tamanho de pacotes segue uma distribuição exponencial;
- c) os nós têm capacidade infinita de armazenamento;
- d) usa-se roteamento fixo;
- e) os canais têm BER (*bit error rate*) igual a zero;
- f) na independência entre os tempos de chegada de pacotes e os tempos de transmissão em cada canal.

Com as características anteriores, o modelo da rede pode ser analisado como um conjunto de filas M/M/1, conforme a notação de Kendall (Bertsekas e Gallager, 1992), em que primeiro  $M$  indica a natureza sem memória do processo de chegada (*memoryless*), o que caracteriza um processo de Poisson, o segundo  $M$  indica a natureza da distribuição de probabilidade dos tempos de serviço e o último número indica o número de servidores, que neste caso é 1. Analiticamente, sob essa consideração, poder-se-ia tornar mais complicada a modelagem do processo de chegada e de serviço de cada fila. A dificuldade surge pela



alta correlação que existe entre o tamanho do pacote a ser servido e o tempo de serviço deste pacote.

Como exemplo, considere-se a linha de transmissão da figura 3.1. Se os pacotes têm todos o mesmo tamanho e que o processo de chegada é do tipo Poisson, pode-se modelar a primeira fila como do tipo  $M/D/1$ , onde  $D$  corresponde a tempos determinísticos (constantes) de chegada entre pacotes. A máxima taxa de chegada do pacote na segunda fila é  $1/\mu$ , onde  $\mu$  é a taxa de serviço, que é a taxa máxima de serviço da primeira fila, pois todos os pacotes têm o mesmo tamanho, de tal forma que não haverá pacotes em espera na segunda fila, e conseqüentemente poder-se-ia concluir que o modelo de Poisson é inapropriado para este caso.

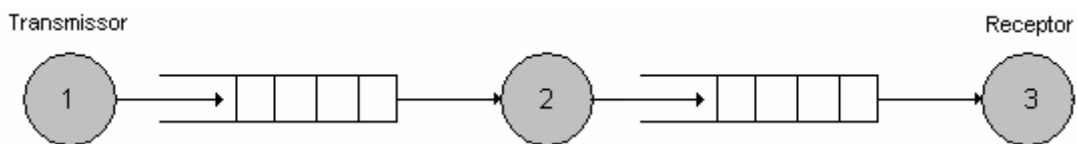


Figura 3.1. Filas de transmissão encadeadas.

Considere agora um segundo exemplo, em que a variável aleatória que representa o tamanho dos pacotes que chegam à primeira fila tem uma função de densidade de probabilidade exponencial. Pode-se modelar a primeira fila como  $M/M/1$ . O intervalo entre a chegada de dois pacotes consecutivos na segunda fila é maior ou igual ao intervalo do tempo de serviço do segundo pacote. Assim, existe uma forte correlação entre o tamanho do pacote e o intervalo de tempo entre pacotes que chegam à segunda fila, e seria inadequada uma modelagem  $M/M/1$  para o segundo nó.

O teorema de independência de Kleinrock (Bertsekas D., Gallager R., 1992) sugere um método para aproximar o número médio de pacotes no sistema e o atraso médio sofrido por estes pacotes.

Considerem-se vários fluxos na rede, em que cada um segue um único caminho, que consiste em uma seqüência de enlaces que conectam origem e destino (circuito virtual). Nesse caso, seja  $x_s$  a taxa de chegada do fluxo  $S$  no sistema. A taxa total de chegada no enlace  $(i, j)$  é a soma da taxa de chegada de todos os fluxos que passam por esse enlace. Para rede de pacotes, é necessário utilizar um modelo mais geral, pois nesse tipo de rede,

um mesmo fluxo pode seguir caminhos distintos, mas a idéia básica e do modelo de circuitos virtuais se preserva.

Como exemplo, considere o enlace  $(i,j)$  da figura 3.2. Nesta figura, mostra-se uma rede com 4 nós, interligados por cinco enlaces. No caso do nó  $i$ , a taxa de chegada de pacotes é igual a  $\lambda_{i,j}=x_1 + x_3$ .

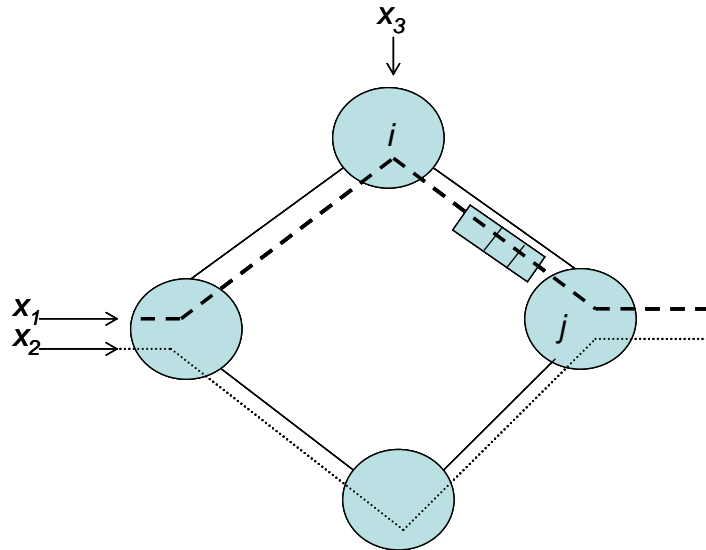


Figura 3.2. Exemplo de fila em rede de dados.

O teorema de Kleinrock sugere que, ao se mesclarem vários fluxos de pacotes em um único enlace de transmissão, produz-se o efeito de restauração da independência entre os tempos de chegada e o tamanho dos pacotes. Este teorema conclui que a utilização de M/M/1 como modelo em cada enlace é adequada. Dessa forma, é válida a aplicação das equações 3.1 e 3.2, assim como a Eq. 3.3 para o cálculo de  $T_i$  e as equações 3.4 e 3.5, para os números médios de pacotes no enlace e no sistema, respectivamente.

$$T_i = \sum_{\substack{(j,k) \\ j \neq k}} \left( \frac{\lambda_{jk}}{\mu_{jk}(\mu_{jk} - \lambda_{jk})} + \frac{1}{\mu_{jk}} + d_{jk} \right) \quad (3.3)$$

$$N_{ij} = \frac{\lambda_{ij}}{\mu_{ij} - \lambda_{ij}} \quad (3.4)$$

$$N = \sum_{(i,j)} N_{ij} = \sum_{(i,j)} \frac{\lambda_{ij}}{\mu_{ij} - \lambda_{ij}} \quad (3.5)$$

Nas Eqs. 3.3, 3.4 e 3.5 as variáveis são as seguintes:

$T_i =$  atraso no nó  $i$

$\lambda_{jk} =$  taxa de chegada do nó  $j$  ao nó  $k$

$\mu_{jk} =$  taxa de serviço entre o nó  $j$  e o nó  $k$

$N_{ij} =$  número de elementos no sistema entre o nó  $i$  e o nó  $j$

Como exemplo de aplicação do modelo M/M/1, suponha uma topologia simples, como a mostrada na figura 3.3.

Os tráfegos  $X_1$ ,  $X_2$  e  $X_3$ , representados na figura 3.3, seguem uma distribuição de Poisson com pacotes de tamanho médio igual a 400 bits,  $X_4$  é um tráfego CBR (*constant bit rate*) com pacotes de tamanho fixo igual a 512 bits e taxa constante de 1 kbps.

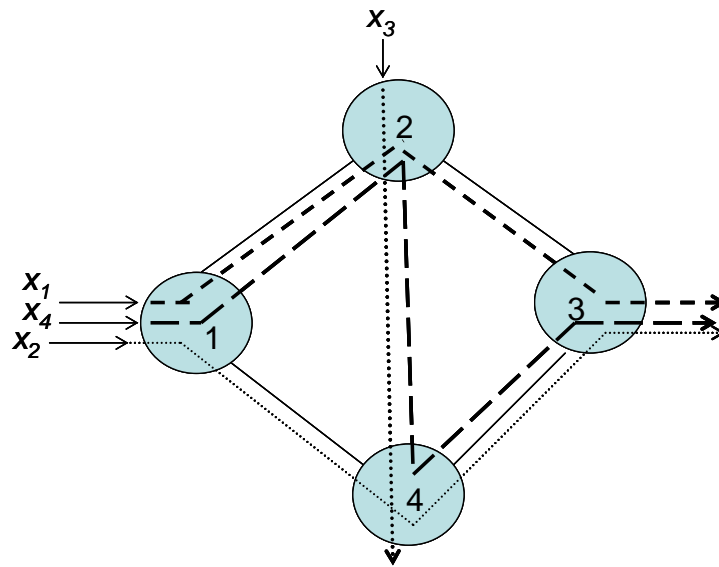


Figura 3.3. Rede para o exemplo de utilização do modelo M/M/1

No exemplo, todos os sub-enlaces têm capacidade de 16 kbps e os tráfegos  $X_1$ ,  $X_2$  e  $X_3$  têm, cada um, 8 kbps de taxa média. O tamanho médio dos pacotes do tráfego agregado é igual à média ponderada do tamanho dos pacotes dos tráfegos individuais como mostrado na Eq. 3.6. A taxa média de chegada de pacotes e a taxa média de serviço de cada subenlace são calculadas com as Eqs. 3.7 e 3.8, em que  $B$  é o tamanho médio dos pacotes do tráfego agregado,  $B_i$  e  $R_i$  são, respectivamente, o tamanho médio dos pacotes e a taxa média dos fluxos individuais e  $C$  é a capacidade do subenlace.

$$B = \frac{\sum_i B_i R_i}{\sum_i R_i} \quad (3.6)$$

$$\lambda = \frac{\sum_i R_i}{B} \quad (3.7)$$

$$\mu = \frac{C}{B} \quad (3.8)$$

Neste exemplo o objetivo é calcular o atraso médio dos pacotes do fluxo CBR. Esse tráfego está agrupado com o tráfego  $X_1$  no enlace (1,2), com o  $X_2$  no enlace (1,4) e com o tráfego  $X_3$  no enlace (2,4). Para o subenlace (1,2), o tamanho médio dos pacotes é  $B \cong 412,5$  bits/pacotes, a taxa média de chegada de pacotes é  $\lambda_{12} \cong 21,8$  pacotes/segundo e a taxa média de atendimento é  $\mu_{12} \cong 38,8$  pacotes/segundo. Pela Eq. 3.3, o atraso no caminho com origem no nó 1 e com destino no nó 3 pode ser calculado como mostrado na Eq. 3.9.

$$T_{13} = \left( \frac{\lambda_{12}}{\mu_{12}(\mu_{12} - \lambda_{12})} + \frac{1}{\mu_{12}} \right) + \left( \frac{\lambda_{24}}{\mu_{24}(\mu_{24} - \lambda_{24})} + \frac{1}{\mu_{24}} \right) + \left( \frac{\lambda_{43}}{\mu_{43}(\mu_{43} - \lambda_{43})} + \frac{1}{\mu_{43}} \right) \quad (3.9)$$

Logo, o atraso médio sofrido pelos pacotes do fluxo CBR é igual a, aproximadamente, 0,177 segundos. Esta primeira topologia, chamada de cenário 1, pode ser modificada quando o enlace (2,4) da figura 3.3 é substituído por um enlace de 30 kbps e com um fluxo  $X_3$  de 16 kbps, o que resulta no cenário 2. No cenário 2, o novo valor do atraso médio dos pacotes do tráfego CBR é de 0,149 segundos. Finalmente, para o cenário 3, alteram-se todos os enlaces da rede da figura 3.3 para um valor 25 kbps e todos os tráfegos de Poisson para 16 kbps, o que resulta em um novo valor do atraso do caminho (1,3) igual a 0,152 (cenário 3).

As três variações da rede da figura 3.3 foram implementadas no simulador de redes de código aberto NS -*Network Simulator* (NSimulator, 2006). Uma comparação dos resultados é mostrada na Tabela 3.1, onde em cada fila é mostrado o atraso em segundos para cada um dos cenários descritos anteriormente. O atraso no modelo analítico com base no modelo M/M/1 é muito próximo do atraso obtido na simulação.

Tabela 3.1. Cálculo comparativo do atraso no modelo simulado e analítico

	<b>Atraso Simulação (segundos)</b>	<b>Atraso Calculado (segundos)</b>
<b>Cenário 1</b>	0,160	0,177
<b>Cenário 2</b>	0,141	0,149
<b>Cenário 3</b>	0,144	0,152

A elegância matemática do modelo com tráfego de Poisson e os resultados, como os observados na Tabela 3.1, exemplificam a consistência das suposições do modelo M/M/1, também conhecido como modelo de Markov, diante de medições reais de tráfego nas redes de pacotes. A equação 3.9 permite o cálculo do atraso, mas outra métrica de extrema importância é a probabilidade de perda de pacotes entre dois nós. O atraso é uma métrica aditiva enquanto a probabilidade de perda de pacotes é uma métrica multiplicativa, conforme foi mencionado na seção 2.2.

A probabilidade de perdas se obtém a partir da consideração de que não existe mais uma fila de tamanho infinito, mas limitada por um valor constante  $K$ , que representa a quantidade de bits/bytes ou pacotes que a fila é capaz de armazenar diante da impossibilidade do canal dar vazão ao fluxo. Então o modelo M/M/1 é estendido para um modelo M/M/1/ $K$ , onde  $K$  especifica o tamanho máximo da fila. Nesse caso, a taxa de chegada  $\lambda_n$ , vai depender do estado  $n$  do sistema, como mostrado na Eq. 3.10

$$\lambda_i = \begin{cases} \lambda, n = 0,1,2,\dots,K \\ 0 \end{cases} \quad (3.10)$$

A taxa média de entrada no sistema  $\bar{\lambda}$  é definida pela Eq. 3.11, onde  $P_K$  é a probabilidade de perda por falta de capacidade do sistema para atender clientes quando está cheio.

$$\bar{\lambda} = \sum_{n=0}^{K-1} \lambda P_n = \lambda \sum_{n=0}^{K-1} P_n = \lambda(1 - P_K) \quad (3.11)$$

Dessa forma, a taxa de ocupação do sistema é definida por  $\bar{\lambda}/\mu$  com no máximo  $K$  elementos no sistema e  $\rho = \lambda/\mu$  define a taxa de pressão no sistema. O número de elementos no sistema é definido pela Eq. 3.12. O sistema pode estar em equilíbrio para valores de  $\rho$  maiores que 1 e, nesse caso, haverá um número elevado de usuários que não chegam a ser servidos.

$$N = \frac{\rho}{1-\rho} - \frac{(K+1)\rho^{K+1}}{1-\rho^{K+1}} \quad (3.12)$$

A probabilidade de perdas é então definida pela Eq. 3.13.

$$P_K = \frac{(1-\rho)\rho^K}{1-\rho^{K+1}} \quad (3.13)$$

A evolução das aplicações e a utilização da multiplexação estatística confirmaram com observações posteriores a percepção de que o tráfego nas redes de pacotes possuía características de rajadas ou *bursty*.

O fato de que o tráfego seja em rajadas, ou seja, que não acompanhe uma taxa de transmissão estável, como é a suposição no caso de que este siga uma distribuição de Poisson, pode ter implicações profundas no modelo matemático mostrado nos exemplos anteriores.

O modelo analítico mostrado anteriormente foi utilizado nas últimas décadas para os projetos de redes de dados, seja para seu dimensionamento ou para a predição de seu desempenho. Em muitos casos, o resultado da predição pela análise difere substancialmente do resultado observado na prática e em simulações. Principalmente, isso acontece em função de que o comportamento do tráfego como um processo de Poisson não é geralmente o observado no tráfego na Internet, principalmente pela presença de rajadas em intervalos aleatórios.

Um enfoque natural para observar o fenômeno de rajadas é em termos de escalas de tempo, sobre as quais ocorrem as rajadas. No caso do tráfego de Poisson, a escala de tempo está relacionada à taxa  $\lambda$  do processo.

As observações do tráfego em redes de pacotes mostram que as rajadas de tráfego acontecem em muitas escalas de tempo (segundos, milissegundos etc.) e que tal característica não corresponde na prática aos modelos analíticos com base no tráfego de Poisson. Uma apreciação visual desta discrepância é mostrada na figura 3.4. Os gráficos são séries de tráfego, disponíveis em <http://www.acm.org/sigcomm/ITA> e coletados por J.

Mogul em 1995 entre a Internet e a DEC (*Digital Equipment Corporation*), com duração de uma hora, que reportam conexões TCP (*Transmission Control Protocol*).

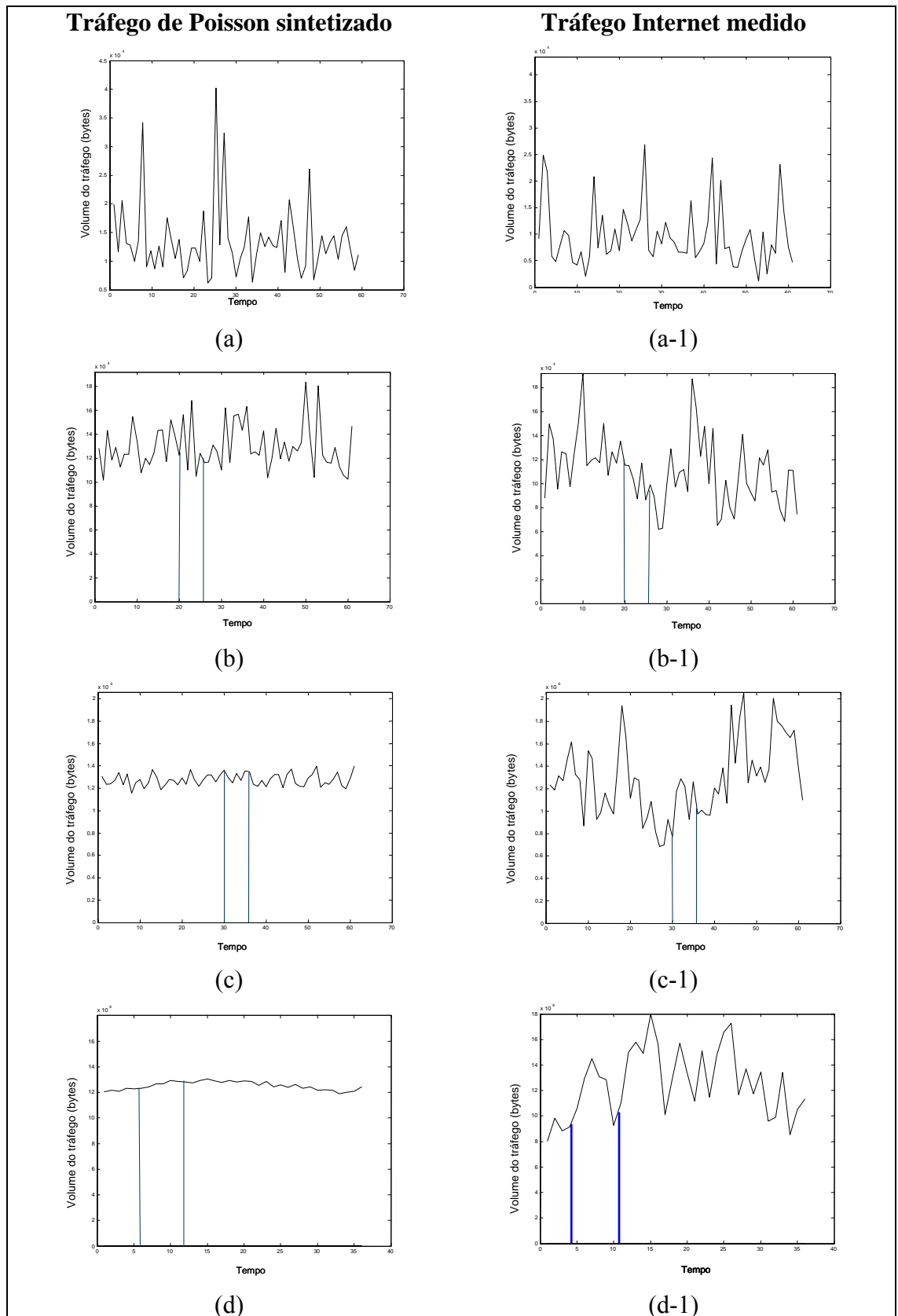


Figura 3.4. Comparação de tráfego de Poisson e tráfego Internet.

Desta série, foram extraídos os parâmetros de interesse e foi usada uma distribuição de Poisson para a geração da série, que toma como base a média e a variância, que foram calculadas na escala de 0,1 segundos, que é a escala da coleta da série de tráfego real.

Na Fig. 3.4, na coluna da esquerda (a,b,c,d) estão as diferentes escalas do tráfego da série sintetizada gerada por um processo de Poisson. A coluna da direita (a-1,b-1,c-1,d-1) mostra as escalas do tráfego medido. As escalas são de 0,1, 1, 10,0 e 100,0 segundos para as linhas (a) e (a-1), (b) e (b-1), (c) e (c-1), (d) e (d-1) respectivamente.

Na linha (a) e (a-1) mostra-se um intervalo de 6 segundos de cada uma das séries, na escala de 0,1 segundos. Pode-se observar que, nesta escala, as rajadas são facilmente visíveis em cada uma das séries. Nas linhas (b) e (b-1), (c) e (c-1), (d) e (d-1), são enquadrados entre duas linhas verticais os intervalos mostrados na escala menor subsequente. Para a linha (b) e (b-1), são mostrados 60 segundos de tráfego na escala de 1 segundo. Pode-se observar que as rajadas são menos acentuadas nesta escala que na anterior.

Na linha (c) e (c-1), mostra-se o tráfego correspondente a 600 segundos em uma escala de 10 segundos. Visualmente, as rajadas da coluna da esquerda são bem menos aparentes que as da direita.

Finalmente, na linha (d) e (d-1) mostra-se a série total do tráfego em uma escala de 100 segundos. O tráfego de Poisson não captura as rajadas das medições reais, diferenças que aparecem mais claramente conforme as escalas de tempo aumentam. Do ponto de vista da engenharia de tráfego, esta diferença é crucial, pois, um tráfego que se comporte como o da coluna esquerda permite uma melhor modelagem, pois a partir de certas escalas não acontecem mudanças bruscas de comportamento.

Conclui-se esta análise ao afirmar que o modelo de tráfego mostrado na seção anterior dimensiona uma rede que considera um comportamento do tráfego como o da coluna esquerda enquanto que, na realidade, o que acontece é o mostrado na coluna direita. Isto resulta em uma rede que apresenta comportamento aquém do esperado, com atrasos e perdas de pacotes em maiores taxas, com o resultado de um dimensionamento ineficiente e que pode ter um alto custo operacional, o que é posteriormente mostrado nesta tese através de exemplos no capítulo 4.



Em função da falta de um modelo apropriado, uma solução para o comportamento do tráfego da coluna da direita da Fig. 3.4 tem sido superestimar o tráfego. O experimento mostrado na Fig. 3.4 reproduz um resultado já observado e reportado (Willinger W. e Paxson V, 1998) nesta mesma série de tráfego.

O comportamento do tráfego em redes de pacotes, tais como a Internet, mostra as características de alta ou extrema variabilidade e, estatisticamente, esta característica é capturada por uma dependência de longo prazo, ou seja, funções de autocorrelação que mostram um decremento lento e com distribuições *heavy-tails* (Morin P. R., Neilson J., 1995), que possuem variância infinita. Este comportamento é chamado de comportamento monofractal ou auto-similar.

### **3.2. TRÁFEGO AUTO-SIMILAR**

Duas características do tráfego auto-similar são de extrema importância na sua caracterização: a propriedade da invariância à escala (fractal probabilístico), o que significa que existe uma equivalência estatística do processo em uma escala de tempo pequena é em uma escala de tempo maior; e a propriedade da dependência de longa duração, caracterizada pela persistência temporal, ou seja, picos de tráfego podem perdurar por um tempo suficientemente grande.

O trabalho de Leland W. E., Taqqu M. S., Willinger W. e Wilson D. V. (1994), em que a partir de rigorosas análises estatísticas, se demonstra a auto-similaridade do tráfego Ethernet, e constitui a base para muitas publicações que se seguiram e comprovaram esta realidade com diferentes tipos de tráfego. De forma resumida, os resultados mais importantes obtidos neste trabalho foram:

- tráfego LAN Ethernet é estatisticamente auto-similar, independente do período de coleta dos dados e de onde os dados foram coletados;
- o grau de auto-similaridade medido em termos do parâmetro de Hurst, chamado também de parâmetro  $H$ , pode ser utilizado como medida de elasticidade do tráfego;

- os modelos correntes de tráfego de pacotes (com base em modelos de Poisson), considerados na literatura, não conseguem capturar a propriedade de auto-similaridade;
- medidas comumente utilizadas para caracterizar tráfego em rajadas, tais como o índice de dispersão, a taxa média de pico ou o coeficiente de variação (tempo entre as chegadas) não têm significado para tráfego auto-similar, e podem ser substituídas pela medição do parâmetro  $H$ .

Em 1995, Beran, J. Sherman, R., Taqqu, M. e Wilinger, W. concluíram que o tráfego de vídeo VBR (*variable bit rate*) apresenta longa dependência. Foram analisados vinte grandes conjuntos de dados de vídeo VBR gerados por uma variedade de aplicações. No mesmo ano, foram feitos estudos das características do tráfego das WANs (Paxson V. e Floyd S., 1995) e foram encontradas também características auto-similares para esse tipo de tráfego. Em Sahinoglu Z. e Tekinay S., (1999), apresenta-se um estudo da auto-similaridade de tráfego multimídia em redes convergentes e as suas implicações em desempenho e métricas de QoS e ainda se conclui sobre a limitação da validade dos modelos convencionais de alocação de recursos, tais como a banda, na presença de tráfego auto-similar.

A natureza do congestionamento produzido por tráfegos auto-similares difere drasticamente das previsões com base em modelos formais padrão e mostram um quadro muito mais complicado do que tipicamente era assumido no passado.

Uma definição comum de processos estocásticos auto-similares é feita na escala de tempo contínuo. Seja o processo estocástico  $Y(t)$  para  $t \in \mathbf{R}$ . Diz-se que  $Y(t)$  é um processo auto-similar com parâmetro de Hurst  $\frac{1}{2} \leq H \leq 1$ , denotado  $H$ -ss, se  $Y(t) \equiv a^{-H}Y(at)$ , para  $a > 0$ ,  $t \geq 0$ , em que a igualdade tem sentido probabilístico, ou seja, as características expressas nas Eqs. 3.14, 3.15 e 3.16 para média, variância e autocorrelação, respectivamente, devem ser satisfeitas.

$$E[Y(t)] = \frac{E[Y(at)]}{a^H} \quad (3.14)$$

$$\text{var}[Y(t)] = \frac{\text{var}[Y(at)]}{a^{2H}} \quad (3.15)$$

$$R_y(t_1, t_2) = \frac{R_y(at_1, at_2)}{a^{2H}} \quad (3.16)$$

$Y(t)$  pode ser interpretada como uma versão escalada de  $Y(at)$ , após a normalização por  $a^{-H}$ . O parâmetro  $H$  de auto-similaridade é a chave para a identificação da auto-similaridade de um processo. Valores de  $H$  inferiores a 0,5 indicam a inexistência de auto-similaridade e definem o processo como um processo de Poisson, e valores mais próximos da unidade indicam um grande grau de persistência ou dependência de longa duração.

Na continuação, serão mostrados os conceitos matemáticos do tráfego auto-similar que fornecem a base para o entendimento do modelo de caracterização de tráfego que será utilizado neste trabalho.

### 3.2.1. Invariância à escala e dependência de longa duração

Considere o processo estocástico de tempo discreto  $X(t)$  que representa o volume de tráfego seja em pacotes, *bytes* ou *bits* no instante  $t \in \mathbf{N}$ . Nas modelagens propostas, é necessário que o processo  $X(t)$  seja estacionário no sentido amplo. Define-se um novo processo a partir de  $X(t)$  para a formulação da propriedade da invariância à escala, expresso na Eq. 3.17.

$$X^m(t) = \frac{1}{m} (X(t)_{im-m+1} + X(t)_{im-m+2} + \dots + X(t)_{im}) \quad (3.17)$$

para  $m = 1, 2, 3, \dots$  em que  $X^m(t)$  é definido como um processo agregado, obtido pela divisão da série temporal original em  $m$  blocos não sobrepostos. A nova série é uma réplica da série original, mas com suavização, reduzida em  $m$  vezes e, como a série original, também é um processo estacionário no sentido amplo.

Como exemplo, considere o processo  $X(t) = (1, 3, 4, 2, 9, 1, 7, 5, 2, 6, 5, 7, \dots)$ . Para  $m = 2$ , temos que  $X^2(t) = (2=(1+3)/2, 3=(3+2)/2, 5=(9+1)/2, 6=(7+5)/2, 4, 6, \dots)$  e, para  $m = 3$ ,  $X^3(t) = (8/3=(1+3+4)/3, 4=(2+9+1)/3, 14/3=(7+5+2)/3, 6, \dots)$ . Se o processo original  $X(t)$  for auto-similar,  $X^m(t)$  possuirá a mesmo coeficiente de autocorrelação de  $X(t)$ , com a relação expressa na Eq. 3.18.

$$R^{(m)}(\tau) = R(\tau) \quad (3.18)$$

Para descrever a propriedade da dependência de longa duração, considere que o processo  $X(t)$  tenha parâmetro de Hurst,  $\frac{1}{2} \leq H \leq 1$ . Esse processo será exatamente auto-similar de 2.<sup>a</sup> ordem se o coeficiente de autocorrelação do processo for dado pela equação 3.19.

$$R(k) = R^{(m)}(k) = \frac{1}{2} \left[ |k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H} \right], m \geq 1 \quad (3.19)$$

De maneira análoga, diz-se que o processo  $X(t)$  é assintoticamente auto-similar de 2.<sup>a</sup> ordem se o coeficiente de autocorrelação do processo for dado pela equação 3.20.

$$\lim_{m \rightarrow \infty} R^{(m)}(k) = R(k) = \frac{1}{2} \left[ |k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H} \right] \quad (3.20)$$

Em geral, nos processos com dependência de curta duração, a função de autocovariância decai rapidamente de acordo com uma função exponencial, como mostrado nas Eqs. 3.21 e 3.22.

$$C_x(k) \approx a^{|k|}, |k| \rightarrow \infty, 0 < a < 1 \quad (3.21)$$

$$\sum_{k=1}^{\infty} C_x(k) = \frac{1}{1-a} \quad (3.22)$$

O modelo M/M/1/K para o tráfego de redes de dados considera que os processos de chegada têm dependência de curta duração. Em contraste com esses processos, existem os processos com dependência de longa duração que têm uma função de autocorrelação que decai lentamente de acordo com uma função hiperbólica, como é o caso dos processos auto-similares com incrementos estacionários.

Os processos auto-similares de incrementos estacionários têm a função de autocovariância descrita pelas Eqs. 3.23 e 3.24, em que esses processos realmente apresentam dependência de longa duração.

$$C_x(k) \approx |k|^{-\beta}, |k| \rightarrow \infty, 0 < \beta < 1 \quad (3.23)$$

$$C_x(k) = \infty \quad (3.24)$$

### 3.2.2. Movimento Browniano Fracionário (fBm)

O Movimento Browniano Fracionário (fBm) é um exemplo de processo estocástico auto-similar, proposto por Mandelbrot e Van Ness, definido como  $A(t)=\beta t^H$ , em que  $\beta$  é uma variável aleatória com distribuição normal, média zero e variância 1 e  $H$  é o parâmetro de Hurst. Logo,  $A(t)$  é uma variável aleatória com distribuição normal e média zero. A variância de  $A(t)$  e a variância do processo de incrementos  $[A(t)-A(s)]$ , são dadas respectivamente pelas Eqs. 3.25 e 3.26 respectivamente.

$$\text{var}[A(t)] = \text{var}[\beta.t^H] = \text{var}[\beta]t^{2H} = t^{2H} \quad (3.25)$$

$$\text{var}[A(t) - A(s)] = E[(A(t) - A(s))^2] = |t - s|^{2H} \quad (3.26)$$

Um importante parâmetro de interesse nos processos auto-similares é a função de autocorrelação e, no fBm, essa função é dada pela Eq. 3.27.

$$R_{B_H}(s, t) = E[A(t)A(s)] = \frac{1}{2}(t^{2H} + s^{2H} - |t - s|^{2H}) \quad (3.27)$$

Pode-se observar as três condições para que um processo seja auto-similar, que são apresentadas nas Eqs. 3.28, 3.29 e 3.30.

$$E[A(at)] = a^H E[A(t)] = 0 \quad (3.28)$$

$$\text{var}[A(at)] = a^{2H} \text{var}[A(t)] \quad (3.29)$$

$$R_{B_H}(at, as) = \frac{1}{2}[(at)^{2H} + (as)^{2H} - |at - as|^{2H}] = a^{2H} R_{B_H}(t, s) \quad (3.30)$$

O fBm de fato é um processo auto-similar e, em conjunto com o ruído gaussiano fracionário, que será apresentado na próxima seção, é um dos os processos mais utilizados na modelagem do tráfego auto-similar.

### 3.2.3. Ruído Gaussiano Fracionário (fGn)

O Movimento Browniano Fracionário (fBm) pode ser trabalhado como um processo de incrementos, onde a amostra  $X(t)$  do fBm representa o incremento do processo em um

intervalo de tempo  $t-s$ . Esse processo, conhecido como Ruído Gaussiano Fracionário (fGn) e definido na Eq. 3.31, com  $t,s \in \mathbb{Z}$ .

$$X(t) = A(t) - A(s) \quad (3.31)$$

Esse processo tem duas propriedades fundamentais: dependência de longo prazo e variância com decaimento lento, ou alta variabilidade, quando  $\frac{1}{2} < H < 1$ . Sua função de autocorrelação é dada pela Eq. 3.32.

$$R(k) = \frac{1}{2} \left( |k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H} \right) \quad (3.32)$$

### 3.2.4. Modelo de tráfego proposto por Norros e Processos Envoltória fBm

Seja  $Z(t)$  um fBm normalizado, com parâmetro  $H \in [1/2, 1)$ . O trabalho de Norros (Norros I., 1995), que apresenta um modelo para o tráfego auto-similar com base no fBm e no fGn, formula o processo de chegada cumulativo  $A(t)$  que representa o número de *bits* que chegam no intervalo  $(0, t]$ , em que  $A(t) = at + \sigma Z(t)$ . O processo cumulativo  $A(t)$  é um processo auto-similar de segunda ordem e apresenta as propriedades enumeradas na Tabela 3.2.

Tabela 3.2. Propriedades do modelo de tráfego de Norros

média	$E\{A(t)\} = a$	(3.33)
variância	$\text{var}\{A(t)\} = t^{2H}$	(3.34)
covariância	$C_A(t, s) = \frac{1}{2} [s^{2H} + t^{2H} - (s-t)^{2H}]$ , $t > s > 0$	(3.35)

O fBm  $A(t) = at + \sigma Z(t)$  tem as seguintes propriedades:

- $Z(t)$  tem incrementos estacionários;
- $Z(0) = 0$  e  $E[Z(t)] = 0$ , para todo  $t$ ;
- $\text{Var}\{Z(t)\} = t^{2H}$ ;
- $Z(t)$  é um processo gaussiano, e dessa forma, todas as distribuições marginais de dimensão finita são gaussianas.

Do ponto de vista da modelagem de tráfego de pacotes, o modelo fBm é uma representação razoável de tráfego agregado, resultado da multiplexação de um conjunto de fontes independentes de tráfego. Ainda quando o tráfego real não seja exatamente auto-similar, é assintoticamente auto-similar.

Os três parâmetros no modelo fBm capturam as características do tráfego real:  $a$  como a taxa média, mede o volume ou quantidade do tráfego,  $\sigma$  e  $H$  capturam a explosividade (*burstiness*, em inglês) do tráfego. A variância e sua relação com a média é uma medida da magnitude das flutuações de tráfego ao redor da taxa média. O parâmetro  $H$  é um indicador da taxa de decaimento das correlações no tráfego, o que, de certa forma, descreve o grau de explosividade do tráfego.

Para um fBm  $A(t)$ , com média  $a$  e variância  $\sigma^2$ , o processo envoltória  $\hat{A}(t)$  é definido como mostrado na Eq. 3.36.

$$\hat{A}(t) = at + k\sqrt{\sigma^2 t} = at + k\sigma \frac{1}{2} \quad (3.36)$$

em que parâmetro  $k$  determina a probabilidade de que  $A(t)$  ultrapasse  $\hat{A}(t)$  no instante  $t$ . Se  $A(t)$  é um processo fBm, a relação da Eq. 3.37 é verdadeira e  $\phi(y)$  é uma função de distribuição residual de uma distribuição Gaussiana padrão. Aproximar  $\phi(y)$  pela Eq. 3.38, tem-se que  $\phi(k) < \epsilon$ , de tal forma que a Eq. 3.39 define o valor de  $k$ . Observe-se que o parâmetro  $k$  determina a probabilidade de que  $A(t)$  exceda  $\hat{A}(t)$  no tempo  $t$ , ou seja  $P(A(t) > \hat{A}(t)) \approx \epsilon$ .

$$P\left(\frac{A(t) - at}{\sigma t^{1/2}} > k\right) = \phi(k) \quad (3.37)$$

$$\phi(y) \approx (2\pi)^{-1/2} (1 + y)^{-1} \exp(-y^2 / 2) \approx \exp(-y^2 / 2) \quad (3.38)$$

$$k = \sqrt{-2 \ln \epsilon} \quad (3.39)$$

O interesse do desenvolvimento do processo envoltória se foca em derivar uma expressão para o cálculo de métricas de QoS para tráfego auto-similar em uma rede de pacotes. Com esse objetivo, uma extensão da definição da Eq. 3.36, que derivará em uma equação para o cálculo de atraso em um nó considerando tráfego auto-similar, é feita em Mayor G. e Silvester J, (1997). Define-se  $A_H(t)$  como um processo fBm com média  $a$ . A lei de Hurst

estabelece que a variância de um incremento deste processo é dada por  $\text{Var}[A_H(t=s)-A_H(t)]=\sigma^2 s^{2H}$  com  $H \in [1/2,1]$ . O envoltória fBm é definida pela equação 3.40.

$$\hat{A}_H(t) = at + k\sigma t^H \quad (3.40)$$

O processo envoltória fBm é um caso especial quando  $H = 1/2$ . Como o processo possui longa dependência, se  $A_H(t)$  ultrapassa  $\hat{A}_H(t)$ , é possível que permaneça nesse estado por um longo período.

Para investigar a precisão do cálculo do processo envoltória fBm, foi realizado o cálculo para a série mostrada na figura 3.5, com  $a=1,389389 \times 10^3$ ,  $\sigma^2=1,472771 \times 10^5$  e  $H=0,85$ , na escala de 0,1 segundo. Foi calculado o  $\hat{A}_H(t)$ , como é mostrado na figura 3.6, para um valor de  $\varepsilon = 10^{-1}$ . Pode-se observar que  $\hat{A}_H(t)$  consegue aproximar o comportamento da série real de tráfego agregado, representada pelo fBm  $A_H(t)$ . Este comportamento, em face do conhecimento dos parâmetros do tráfego, permite a derivação de expressões para o cálculo do atraso e da probabilidade de perda de pacotes em função de uma máxima escala de tempo, como será mostrado a continuação.

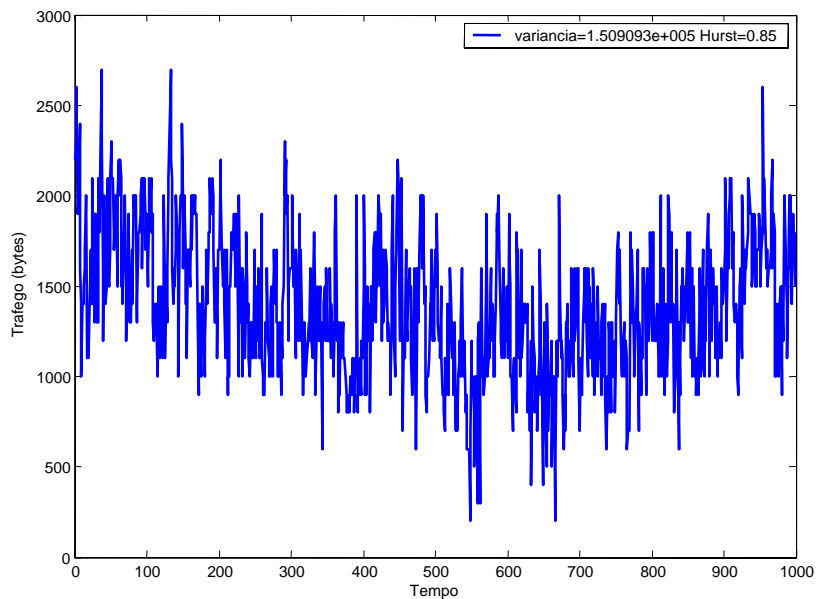


Figura 3.5. Série de tráfego auto-similar



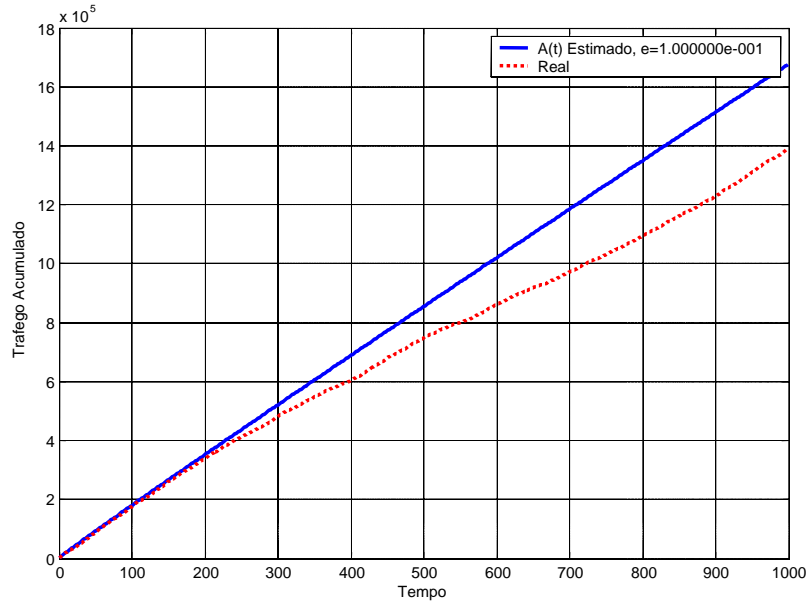


Figura 3.6. Série de tráfego agregado e processo envoltória fBm

### 3.2.4.1. Escala de Tempo Máxima e Tamanho da Fila

O atraso máximo que um pacote experimenta em uma fila é limitado pelo período máximo de ocupação (Ng K.J., Song S., Tang B., Zhao W, 2002). Com a definição da Eq.3.40, considera-se uma fila tipo FIFO com taxa constante de serviço  $c$ . Se  $A_H(t)$  é o processo envoltória, o comprimento de cada período ocupado é limitado por uma constante  $d$ , como mostrado na Eq. 3.41. Esta definição pode ser encontrada em (Mayor G. e Silvester J, 1997) e (Fonseca N.L.S, M. Gilberto e Neto C.A.V, 2000).

$$d = \inf\{t \geq 1 : A_H(t) - ct \leq 0\} \quad (3.41)$$

Considere-se que durante um período ocupado, que começa em 0 e vai até  $t$ , o número de elementos em um nó é dado por  $q(t)$ . Ou seja,  $q(t)$  pode ser expresso como aquilo que chega no nó menos aquilo que consegue ser servido nesse período, Eq. 3.42.

$$q(t) = A(t) - ct \geq 0 \quad (3.42)$$

Ao considerar a Eq. 3.40, em que  $\hat{A}_H(t)$  é o envoltória probabilístico de  $A_H(t)$ , pode-se definir a Eq. 3.43, de tal forma que  $P(q(t) > \hat{q}(t)) = P(A_H(t) > \hat{A}_H(t)) \leq \varepsilon$ .

$$\hat{q}(t) = \hat{A}(t) - ct \geq 0 \quad (3.43)$$

O atraso máximo de um elemento no nó é dado pelo tamanho máximo de  $q(t)$  em um período ocupado, ou seja  $\max(q(t))$ . Pode-se definir então, ao considerar a Eq. 3.43, que o tamanho máximo da fila é dado pelo processo envoltória, dentro de uma probabilidade específica, conforme a Eq. 3.44. Ao limitar o comportamento do processo de chegada, é possível transformar a obtenção de um limite probabilístico de um sistema estocástico definido por  $q(t)=A_H(t)-ct \geq 0$  em um problema mais simples que se restringe em encontrar o máximo de um sistema determinístico dado pela Eq. 3.43.

$$q_{\max} = \max(\hat{q}(t)) = \max(\hat{A}(t) - ct) \quad (3.44)$$

Da Eq. 3.44 se percebe que é necessário o cálculo de um  $t_{\max}$ , onde  $q_{\max}$  alcança seu maior valor. Substitui-se 3.40 em 3.44 e se deriva a expressão resultante de tal forma que

$$\frac{d\hat{q}(t_{\max})}{dt} = 0, \text{ o que resulta na Eq. 3.45.}$$

$$t_{\max} = \left( \frac{k\sigma H}{c-a} \right)^{\frac{1}{1-H}} \quad (3.45)$$

A expressão 3.45 expressa então um resultado de extrema importância, pois  $t_{\max}$  é a máxima escala de tempo que é o intervalo de 0 até  $t_{\max}$ , no qual a fila alcança a mais alta ocupação em um sentido probabilístico. Percebe-se então que o valor de  $t_{\max}$  permitiria o cálculo do tempo máximo que, probabilisticamente, um elemento teria que esperar a ser servido em um nó. Os estudos que fundamentam estas definições se originaram em redes ATM, mas podem ser também considerados para tecnologias de comutação de pacotes orientadas a conexão, em que uma conexão representa um caminho virtual por onde passa um fluxo de pacotes enviados de uma fonte para um destino.

Ao considerar a máxima escala de tempo conforme a Eq. 3.45, ao se substituir na Eq. 3.42, pode-se derivar o tamanho máximo da fila ao considerar a máxima escala de tempo, e dessa forma se chega à expressão 3.46, cuja derivação formal pode ser conferida em (Fonseca N.L.S, M. Gilberto e Neto C.A.V, 2000).

$$q_{\max} = (c-a)^{\frac{H}{H-1}} (k\sigma)^{\frac{1}{1-H}} H^{\frac{H}{1-H}} (1-H) \quad (3.46)$$

Para visualizar melhor a relação do  $q_{\max}$  com a utilização do canal, a figura 3.7 ilustra um exemplo dessa relação com a mesma serie da figura 3.4. No eixo horizontal aparecem as

capacidades normalizadas pela média e no eixo vertical a capacidade máxima da fila. Foi utilizado para o cálculo um  $\epsilon = 0.001$ . Pode-se observar que para uma utilização do canal entre 33% ( $c/a = 3$ ) e 40% ( $c/a = 2.5$ ) o aumento da fila é maior que 200%.

Observe-se que da Eq.3.46 podem-se obter também os valores para a capacidade de um enlace em função do tamanho da fila.

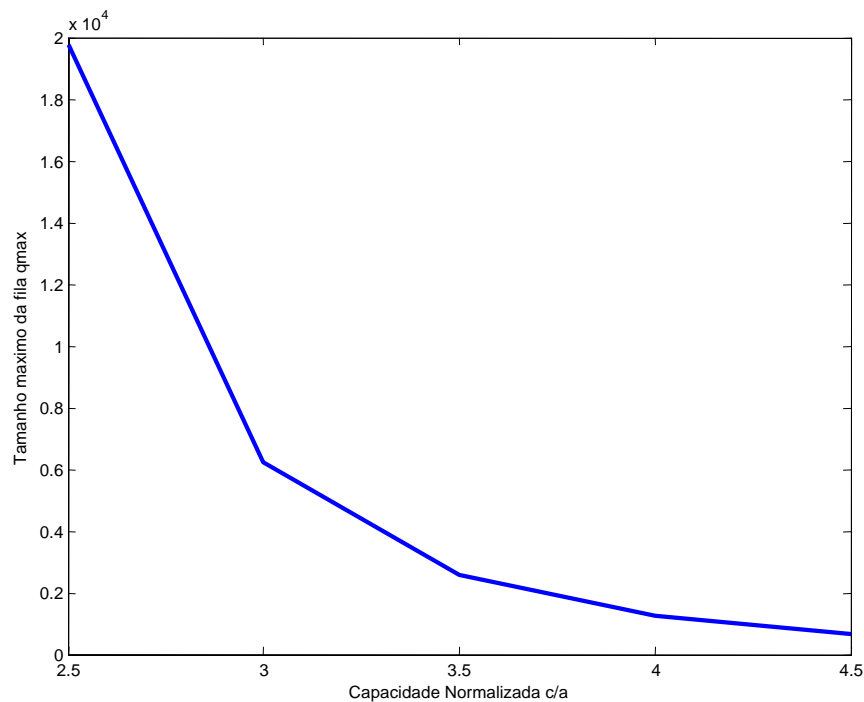


Figura 3.7. Limite Superior do Tamanho da Fila

Em Ng K.J., Song S., Tang B., Zhao W. (2002), com base no trabalho de Mayor G. e Silvester J (1997), se deriva um método para o cálculo do atraso estatístico fim-a-fim para tráfego auto-similar e um conjunto de servidores FIFO (*First In-First Out*). Este método se generaliza para uma série orientada a conexão e é apresentado a continuação.

Seja  $S$  um servidor com  $m$  conexões, onde  $A_H^i(t)$  é o processo fBm para a conexão  $M_i$ , expresso como  $A_H^i(t) = at + \sigma Z(t)$ , com  $Z(t)$  normalizado como um processo fBm com parâmetro  $H$ . O processo fBm envoltória pode-se expressar como na Eq. 3.40. Assume-se que a taxa do serviço do servidor é suficientemente grande de tal forma que o tempo de transmissão de um pacote de tamanho máximo é menor que a unidade de tempo. Se o tamanho do intervalo de tempo ocupado  $B$  satisfaz  $P\{B < b\} \leq p$ , e o processo envoltória de  $A(t)$  com erro  $\epsilon$  e tamanho  $b$  é  $\hat{A}_H^i(t)$ , a probabilidade de que o pior atraso de pacote  $M_i$  não

exceda  $d$  é maior que  $(1-p)(1-\sum_{j=1}^m \varepsilon_j)$ , de tal forma que  $d$  é dado pela Eq. 3.47, em que  $c$  é a taxa de serviço constante do servidor.

$$d = \max_t \left\{ \frac{1}{c} \sum_{j=1}^m \hat{A}_j(t) - t \right\} \quad (3.47)$$

A Eq. 3.47 não é mais que o tamanho máximo de fila na máxima escala de tempo dividido pela taxa de serviço. Assim, a Eq. 3.47 é equivalente à Eq. 3.48, onde se particulariza para o caso de um único nó.

$$d = \frac{q_{\max}}{c} \quad (3.48)$$

Em uma rede de pacotes, em que cada elemento de comutação é um servidor e cada caminho é servido por vários servidores, o atraso entre dois nós é a soma do atraso em cada um dos servidores.

Quando um elemento de comutação recebe como entrada um processo auto-similar, que é em realidade o tráfego, a agregação deste tráfego com outros preserva as características da auto-similaridade. Sob esta consideração, pode-se concluir que o processo fBm de envoltórias e a Eq.3.48 podem ser considerados para o cálculo do atraso máximo entre dois nós em uma rede de pacotes com tráfego auto-similar, dentro de um certo limite probabilístico. Ou seja, se um caminho percorre  $I$  servidores, e o servidor  $S_i(i=1,\dots,I)$  tem um valor máximo de atraso  $d_i$  com probabilidade  $1-\delta_i$ , então o limite de atraso é  $\sum_{i=1}^I d_i$  com uma probabilidade de  $1-\sum_{j=1}^m \delta_j$ .

### 3.2.4.2. Perda de Pacotes

Os roteadores usam as filas para o armazenamento de pacotes durante momentos de congestionamento nos enlaces. Quando o tamanho da fila é ultrapassado pelos pacotes que estão à espera de serem servidos pelo enlace, ocorrem perdas de pacotes. O tamanho das filas têm, então, uma forte influência na probabilidade de perda de pacotes e na variação do atraso. Espera-se, então, que, diante da importância da função que estes elementos de rede

têm, existam métodos bem fundamentados para definir seu tamanho diante da natureza auto-similar do tráfego em redes multimídia. Não é assim.

Na prática, o tamanho das filas é definido utilizando-se uma regra empírica, fundamentado no trabalho de Villamizar C. e Song C. (1994). Com base em medidas experimentais com oito fluxos do tipo TCP em enlaces de 40Mbps, estabeleceu-se a conclusão que um nó precisa um tamanho de fila equivalente ao tempo *round-trip* (*RTT*, em inglês) de um fluxo típico que passa pelo nó multiplicado pela capacidade das interfaces de rede do roteador. Esta regra é conhecida como  $B=RTT \times C$ , em que  $B$  é o tamanho da fila,  $RTT$  é o *round-trip-time* e  $C$  é a capacidade do canal. Como exemplo, um operador de rede que trabalhe com um  $RTT$  de 250ms, em um nó roteador com uma interface de 10 Gbps, precisa de  $250\text{ms} \times 10\text{Gbps} = 2,5 \text{ Gbps}$  como tamanho da fila, é esse tamanho cresce linearmente com o aumento da capacidade do canal.

Com filas maiores, aumentam os atrasos em uma rede. Para fluxos pequenos e aplicações interativas, deseja-se trabalhar com atrasos pequenos. Tamanhos exagerados de filas podem fazer que muitas destas aplicações não sejam utilizáveis em casos de congestionamento na rede. Mas se as filas são menores, a probabilidade de perda de pacotes tende a aumentar com o congestionamento da rede. Encontrar então um ponto de equilíbrio entre essas duas métricas para definir o tamanho das filas, relacionando este processo com o tamanho do enlace e considerar as características do tráfego multimídia, não é uma tarefa trivial. No caso do tráfego auto-similar, o comportamento das probabilidades de perdas de pacotes observadas mostram que pode-se combater a perda de pacotes de melhor forma ao incrementar a capacidade do enlace em vez do tamanho da fila (Tsybakov B., Georganas N, 2000).

Vários trabalhos têm sido realizados para prover modelos matemáticos de tráfego de longa dependência com uma visão que permita facilitar a análise de desempenho de uma rede no sentido de teoria de filas. Estes trabalhos têm sido importantes no sentido de que estabelecem fronteiras básicas de desempenho ao pesquisar o comportamento das filas com fontes auto-similares que mostram características de desempenho fundamentalmente diferentes daquelas correspondentes aos sistemas com fontes de Poisson. Particularmente, o tamanho da distribuição de uma fila nos sistemas de filas infinitas tem um comportamento de decrescimento mais lento que exponencial, em claro contraste com o tráfego de curta dependência em que este decrescimento é exponencial.

Os trabalhos de pesquisa realizados têm mostrado que em função da distribuição do tamanho da fila, que pode apresentar comportamentos conforme uma distribuição de Weibull (Norros I, 1994), não existe uma vantagem em aumentar consideravelmente o tamanho das filas pois se incorre em uma penalidade no atraso em relação ao ganho obtido na redução de perdas de pacotes.

A maior fraqueza dos resultados obtidos no estudo de filas com fontes auto-similares está na sua natureza assintótica ao estabelecer limites superiores ou inferiores. Por exemplo, no caso da fila infinita, os limites superiores e inferiores são derivados conforme o tamanho da fila se aproxima do infinito.

Dentro do contexto deste trabalho, ao considerar o tráfego em uma rede multimídia modelado a partir de um processo fBm, considera-se uma fila do tipo FIFO com tamanho máximo  $B$  e taxa de serviço constante  $C$ . O problema está em encontrar uma expressão que relacione o processo de chegada com a taxa de serviço e o tamanho da fila para calcular a probabilidade de perda de pacotes. É também possível, a partir desta expressão, calcular o tamanho do enlace e da fila para que a probabilidade de perda de pacotes se limite a um valor máximo.

Ao considerar as expressões de máxima escala de tempo e tamanho máximo da fila apresentadas na seção anterior, pode-se inferir que, ao limitar o tamanho da fila para um valor constante, pode-se encontrar uma expressão para o cálculo da probabilidade de perda de pacotes.

Seja  $A(t)$  o tráfego que chega a um sistema de filas que opera com taxa de serviço constante  $C$ . O trabalho não realizado neste sistema é dado pela Eq. 3.42.

Como foi visto na seção anterior, o processo  $\hat{A}_H(t)$  é um limitante superior para o volume de tráfego  $A_H(t)$ . A Eq. 3.43, que expressa a máxima escala de tempo, define o momento em que a fila alcança seu tamanho máximo, ou seja, o momento em que o processo de chegada de tráfego alcança uma diferença máxima em relação à capacidade de serviço. Então, um limitante superior para o tamanho da fila pode ser expresso pela Eq. 3.43.

Seja  $B_f$  o tamanho máximo da fila. A máxima probabilidade de perdas está então dada pela relação entre a ocupação virtual da fila menos o tamanho máximo da fila e o limitante superior para o volume de tráfego  $A(t_{max})$ , onde  $t_{max}$  é a máxima escala de tempo, conforme

a Eq. 3.45. Dessa forma, se define a máxima probabilidade de perda de pacotes conforme a Eq. 3.51.

$$P_{\max} = \frac{\hat{A}_H(t_{\max}) - Ct_{\max} - B_f}{\hat{A}_H(t_{\max})} \quad (3.49)$$

A Eq. 3.51 relaciona a ocupação máxima da fila menos o tamanho máximo, o que efetivamente representa aqueles elementos do nó que não encontraram espaço na fila para serem armazenados. Como a perda é uma métrica multiplicativa, o cálculo da probabilidade de perdas de cada nó em um caminho virtual deve seguir a expressão para as perdas apresentada na seção 2.2.3.

### 3.2.5. Caracterização do Tráfego auto-similar e a agregação de fluxos

#### 3.2.5.1. Estimativa do Parâmetro de Hurst

O primeiro parâmetro para caracterização do tráfego como auto-similar é o parâmetro de Hurst. Existem muitos métodos de estimativa do parâmetro de Hurst, entre eles os mais utilizados são:

- *Variance-Time-Plot* (VTP) que utiliza a propriedade de variância de decaimento lento (Morin P., 1995, Pereira Dantas M.J., 2000);
- método de Higuchi, que considera a dimensão fractal de uma série temporal na estimativa de  $H$  (Morin P., 1995, Pereira Dantas M.J., 2000);
- método Abry-Veitch, com base na decomposição de amostras em coeficientes de aproximação (filtro passa-baixos) e coeficientes de detalhes (filtro passa-altos), com o uso da transformada de Wavelet (Pereira Dantas M.J., 2000);
- Método de neurocomputação, que utiliza redes neurais na estimativa do parâmetro de Hurst (Gomes D., Fonseca N., Agoulmine N., Souza, J. N, 2000);
- Método Kettani-Gubner (Kettani H., Gubner J.A., 2002), que utiliza a função de autocorrelação do tráfego. Este método ainda não é muito utilizado, mas é um método promissor e de baixa complexidade computacional.

Neste trabalho foram comparados os métodos VTP e de Kettani-Gubner, e o último foi escolhido por parecer mais eficiente em termos computacionais e bastante simples de implementar. Este método de estimativa do parâmetro de Hurst é fundamentado no coeficiente de correlação do tráfego auto-similar que considera um processo como exatamente auto-similar de segunda ordem se seu coeficiente de autocorrelação é descrito pela equação  $R(k)=1/2(|k+1|^{2H}-2|k|^{2H}+|k-1|^{2H})$ . Para  $k=1$  tem-se  $R(1)=2^{2H-1}-1$ . Ao isolar  $H$ , tem-se  $\hat{H}=1/2[1+\log(1+R(1))]$ .

Sob a suposição de que o processo seja ergódico, pode-se calcular o parâmetro de Hurst a partir da Eq. 3.50.

$$\hat{H}_n = \frac{1}{2} \left[ 1 + \log(1 + \hat{R}_n(1)) \right] \quad (3.50)$$

Se  $X(t)$  é um processo gaussiano exatamente auto-similar de segunda ordem, ou seja, ruído gaussiano fracionário, para  $0 < H < 3/4$ , e para uma função amostra relativamente grande,  $\hat{R}_n(1)$  é aproximadamente uma variável aleatória com função densidade de probabilidade normal  $N(\mu_n, \sigma_n^2)$ , em que a média e a variância são respectivamente dadas pelas Eqs. 3.51 e 3.52.

$$\mu_n = R(1) - (1 - R(1))n^{2H-2} \quad (3.51)$$

$$\sigma_n^2 = \frac{\log n}{n} [2H(2H-1)(1+R(1))]^2, H = 0,75 \quad (3.52)$$

Com as equações 3.51 e 3.52, pode-se calcular o intervalo de confiança de  $\hat{R}_n(1)$  e, conseqüentemente, o intervalo de confiança de  $\hat{H}_n$ . Com os valores de uma tabela de uma função de densidade de probabilidade normal (gaussiana) pode-se chegar à Eq. 3.53.

$$P \left( \left| \frac{\hat{R}_n(1) - \mu_n}{\sigma_n} \right| \leq 1.96 \right) = 0.95 \quad (3.53)$$

A equação 3.55 expressa que a probabilidade da equação 3.54 ser válida é 95%.

$$\mu_n - 1.96\sigma_n \leq \hat{R}_n(1) \leq \mu_n + 1.96\sigma_n \quad (3.54)$$

onde o limite inferior  $h_-$  e o limite superior  $h_+$  do intervalo de confiança são  $h_- \leq \hat{H}_n \leq h_+$ , e  $h_-$  e  $h_+$  são dados pela Eq. 3.55.



$$h_{\pm} = \mu_n \pm 1.96\sigma_n$$

$$h_{\pm} = \frac{1}{2} \left\{ 1 + \log_2 \left[ 1 + \rho(1) - (1 - \rho(1))n^{2H-2} \pm 1.96\sigma_n \right] \right\} \quad (3.55)$$

Para  $H > \frac{3}{4}$ , a distribuição de probabilidade existe e não é normal, a proposta é assumir a mesma média e a mesma variância do caso que  $H < \frac{3}{4}$ , seguido de um pequeno ajuste. Considera-se  $w_n$  como a diferença entre os limites do intervalo de confiança, obtém-se e Eq. 3.56 e 3.57, e, após algumas simplificações e aproximações, chega-se a Eq. 3.58 para o cálculo do intervalo de confiança.

$$\mu_n - 1.96\sigma_n \leq \hat{\rho}_n(1) \leq \mu_n + 1.96\sigma_n \quad (3.56)$$

$$w_n = h_+ - h_- = 2(1.96\sigma_n) \quad (3.57)$$

$$w_n \equiv \frac{5}{\sqrt{n}} \quad (3.58)$$

Para ilustrar a aplicabilidade do método de Kettani e Gubner e compará-lo com métodos amplamente aceitos, como é o caso do VTP, foi realizada uma comparação entre os valores de  $H$  obtidos em cada método utilizando-se séries temporais que representam fluxos de tráfego. A avaliação foi feita com um conjunto de 20 séries de tráfego sintéticas, geradas para cada valor de  $H$ , com  $H$  no intervalo de 0,2 até 0,9, com variações de 0,1.

Em cada série temporal, o valor da estimativa do parâmetro  $H$  com um mesmo método pode-se considerar uma variável aleatória. Então, com essa consideração, a média dessa variável para esta comparação, é considerada a estimativa do parâmetro  $H$ . Assim também, o erro do valor estimado e do valor real do  $H$  na série é o desvio padrão. O resultado das estimativas está na Tabela 3.3.

Tabela 3.3. Avaliação dos métodos de estimativa do parâmetro de Hurst  $H$ .

$H$ (real)	VTP		Kettani-Gubner	
	$H$	Erro	$H$	Erro
0,200	0,435	0,016	0,250	0,015
0,300	0,419	0,014	0,304	0,011
0,400	0,432	0,012	0,405	0,008
0,500	0,512	0,013	0,504	0,008
0,600	0,597	0,014	0,593	0,011
0,700	0,692	0,023	0,695	0,012
0,800	0,794	0,028	0,797	0,014
0,900	0,904	0,016	0,924	0,010

Os resultados da Tabela 3.3 mostram que para  $H$  real igual a 0,2, ambos os métodos aparecem ser ineficazes e, com o método VTP, o erro é ainda maior visto que o valor estimado é duas vezes superior ao valor real, enquanto que, para o método Kettani-Gubner, o erro é de 25%. Para  $H$  real igual a 0,3 e 0,4, o método VTP tem erros muito grandes, enquanto que o método Kettani-Gubner tem erros inferiores a 2%.

Para valores de  $H$  entre 0,5 a 0,9, os dois métodos são igualmente eficazes na estimativa do parâmetro  $H$ , uma vez que o maior erro com relação ao valor real de  $H$  foi com o método Kettani-Gubner para  $H$  igual 0,9, onde o erro foi de 0,024, que é menor que 3%.

O método VTP faz uso da propriedade de invariância à escala para a estimativa do parâmetro  $H$ , através da variância do processo agregado. O método de Kettani-Gubner faz uso da propriedade da longa dependência através do coeficiente de autocorrelação.

### 3.2.5.2. Modelo para obtenção dos parâmetros do tráfego agregado

O tráfego das redes convergentes ou multimídia é a superposição de vários fluxos de fontes independentes com características distintas. Conhecidas as características individuais de cada fluxo, gostar-se-ia que fosse possível prever as características do tráfego agregado, elemento necessário para o processo de otimização e dimensionamento da rede.

Sabe-se que o processo estocástico  $A(t)$ , auto-similar  $H$ -ss de incrementos estacionários, é razoavelmente aproximado por uma fonte de tráfego fBm. Este tipo de tráfego é descrito por apenas 3 parâmetros: taxa média do processo de incrementos  $a$ , a variância  $\sigma^2$  e parâmetro de Hurst  $H$ .

Define-se a função  $V(t)=\text{var}(A(t))$  como a variância do processo  $A(t)$  no instante  $t$ . Se supõe que  $A(t)$  possa ser aproximado por um fBm, pode-se aproximar a função  $V(t)$  por uma fórmula relativamente simples, expressa na Eq. 3.59, em que  $H$  é o parâmetro de Hurst e  $\sigma^2$  é a variância do processo de incrementos  $A(t)-A(s)$ .

$$V_{fbm}(t; H, \sigma^2) = \sigma^2 t^{2H} \quad (3.59)$$

Para que a aproximação de  $A(t)$  por uma fonte de tráfego fBm seja satisfatória, é necessário que  $V(t)$  e  $V_{fbm}(t; H, \sigma^2)$  sejam suficientemente próximos.

Considere o tráfego agregado  $A_g(t)$ , formado pela superposição de processos de chegada independentes  $A_g(t) = \sum_s A_s(t)$  e que cada fonte individual possa ser aproximada por uma fonte de tráfego fBm. Naturalmente, a taxa média  $a_g$  do tráfego agregado é a soma da taxa média de cada fluxo individual  $a_s$ , e expressões para  $H_g$  e  $v_g$  devem ser encontradas. Como as fontes são independentes e cada fonte individual é aproximadamente um processo fBm, pode-se expressar a relação da Eq. 3.60.

$$V_g(t) = \text{var}(A_g(t)) = \text{var}\left(\sum_s A_s(t)\right) = \sum_s \text{var}(A_s(t)) \cong \sum_s V_{fbm}(t; H_s, \sigma_s^2) \quad (3.60)$$

Se o tráfego agregado pode ser aproximado por um processo fBm, então a relação da Eq. 3.61 é válida.

$$V_{fbm}(t; H_g, \sigma_g^2) \cong \sum_s V_{fbm}(t; H_s, \sigma_s^2) \quad (3.61)$$

Considere uma nova função decorrente da função de variância do processo fBm, expressa na Eq. 3.62, e pode-se observar que existe uma correspondência linear entre o logaritmo da escala de tempo de observação  $\tau = \ln(t)$  e o logaritmo da variância. Ao aplicar a função da Eq.3.61 para o tráfego agregado, obtém-se a Eq. 3.63.

$$W_{fbm}(\tau; H, \sigma^2) = \ln(V_{fbm}(e^\tau; H, \sigma^2)) = \ln(\sigma^2 \cdot e^{2H\tau}) = 2H\tau + \ln(\sigma^2) \quad (3.62)$$

$$W_g(\tau) = \ln(V_g(e^\tau)) = \ln\left(\sum_s V_s(e^\tau)\right) \cong \ln\left(\sum_s \sigma_s^2 e^{2H_s\tau}\right) = W_{g0}(\tau)$$

$$W_g(\tau) = \ln(V_g(e^\tau))$$

$$W_{g0}(\tau) = \ln\left(\sum_s \sigma_s^2 e^{2H_s \tau}\right) \quad (3.63)$$

Para o caso em que os fluxos individuais não tenham o mesmo parâmetro  $H$ , a função  $W_g(\tau)$  é não linear como na equação 3.63. Se assume que  $W_{g0}(\tau)$  seja uma boa representação de  $W_g(\tau)$ , e ao fazer  $\tau \rightarrow \infty$ , obtém-se a expressão 3.64.

$$W_{g0}(\tau) = \ln\left(\sigma_1^2 e^{2H_1 \tau} + \sigma_2^2 e^{2H_2 \tau} + \dots + \sigma_n^2 e^{2H_n \tau}\right)$$

$$(\tau \rightarrow \infty, i \neq j) \rightarrow \sigma_i^2 e^{2\max(H_i)\tau} \gg \sigma_j^2 e^{2H_j \tau} \quad (3.64)$$

$$W_{g0}(\tau) \cong \ln\left(\sigma_i^2 e^{2\max(H_i)\tau}\right), \tau \rightarrow \infty$$

$$W_{g0}(\tau) \cong 2 \max(H_i) \tau$$

Em Orenstein P., Kim H., e Lau C.L. (2001), é utilizada a aproximação da Eq. 3.64 para encontrar os parâmetros de interesse  $(H_g, \sigma_g^2)$ , ao supor que estes valores são os mesmos encontrados em  $W_{g0}(\tau)$  quando  $\tau \rightarrow \infty$ . As equações 3.65 e 3.66 mostram os valores estimados do parâmetro  $H$  e da variância, respectivamente.

$$H_g = \max(H_s) \quad (3.65)$$

$$\sigma_g^2 = \sum_s \left\{ \sigma_s^2 \mid H_s = \max(H_s) \right\} \quad (3.66)$$

Um segundo método, também proposto em Orenstein P., Kim H., Lau C.L. (2001), consiste em fixar um tempo arbitrário de escala logarítmica  $x_i$  e encontrar  $H_g = H(x_i)$  e  $\sigma_g^2 = \sigma^2(x_i)$ . A reta  $W_{fBm}(\tau; H(x_i), \sigma^2(x_i))$  e tangente a  $W_{g0}(\tau)$  no ponto  $\tau = x_i$ , logo, obtém-se a expressão 3.67.

$$2H(x_i) = \frac{\partial W_{g0}(\tau)}{\partial \tau} = \frac{\sum_s 2H_s \sigma_s^2 e^{2H_s \tau}}{\sum_s \sigma_s^2 e^{2H_s \tau}} \quad (3.67)$$

No ponto  $\tau = x_i$ ,  $W_{fBm}(\tau; H(x_i), \sigma(x_i))$  e  $W_{g0}(\tau)$  têm o mesmo valor, logo se obtém a expressão 3.68.

$$\ln(\sigma^2(x_i)) + 2H_g x_i = \ln\left(\sum_s \sigma_s^2 e^{2H_s x_i}\right)$$

$$\frac{\sigma^2(x_i)}{\sum_s \sigma_s^2 e^{2H_s x_i}} = e^{-2H_g x_i} \quad (3.68)$$

$$\sigma^2(x_i) = \sum_s \sigma_s^2 e^{2(H_s - H_g)x_i}$$

Considera-se que  $x_i=0$  o que implica que  $e^{x_i}=1$ , e podem-se encontrar valores simplificados para  $H_g=H(x_i)$  e  $\sigma_g=\sigma(x_i)$ , como mostrados nas Eqs. 3.69 e 3.70.

$$H_g = H(0) = \frac{\sum_s H_s \sigma_s^2}{\sum_s \sigma_s^2} \quad (3.69)$$

$$\sigma_g^2 = \sigma^2(0) = \sum_s \sigma_s^2 \quad (3.70)$$

Para avaliar a eficiência dos métodos de estimativa analítica dos parâmetros do tráfego agregado, foram gerados traços de tráfegos auto-similares com características distintas, como mostrado na Tabela 3.4. Para cada classe, foram gerados 20 traços. Todos os traços foram gerados com a mesma média. Foram considerados 3 coeficientes de variâncias distintos e 4 valores de  $H$ , para um total de 12 possíveis combinações. Na primeira coluna da Tabela 3.4 são referenciados os 12 tipos de tráfego, na segunda e na terceira colunas estão os coeficientes de variância e parâmetro  $H$  correspondentes. Os tráfegos foram agregados dois a dois, de maneira a obter os casos de interesse. O método da Kettani-Gubner foi utilizado na estimativa do parâmetro de Hurst. A estimativa dos parâmetros de interesse com os dois métodos e o resultado da simulação são apresentados na Tabela 3.5.

Tabela 3.4. Parâmetros dos traços gerados.

Identificador	Coefficiente de Variância ( $v=\sigma^2/a$ )	Hurst (H)
A	0,5	0,55
B	1,0	0,55
C	1,5	0,55
D	0,5	0,65
E	1,0	0,65
F	1,5	0,65
G	0,5	0,75
H	1,0	0,75
I	1,5	0,75
J	0,5	0,85
K	1,0	0,85
L	1,5	0,85

Para efeitos de análise, a estimativa dos parâmetros como mostrada nas Eqs. 3.65 e 3.66 será chamado de método 1, e de método 2, a estimativa com base nas Eqs. 3.69 e 3.70.

Na Tabela 3.5 pode-se observar que os dois métodos são bastante acurados na estimativa dos parâmetros de interesse para os casos particulares, em que o parâmetro  $H$  é o mesmo para os dois fluxos individuais (entradas 1 a 4 da Tabela 3.11). Para os casos mais gerais o método 1 mostrou-se menos preciso na estimativa dos parâmetros, como se pode observar no 16.<sup>a</sup> entrada da tabela. O método 2 mostrou-se bastante acurado para todos os casos, pois o erro na estimativa do parâmetro é sempre menor que 2%.

Ao verificar a validade do método 2 na estimativa do parâmetro  $H$  e do coeficiente de variância do tráfego agregado, quando considerada a agregação de 2 fluxos, parte-se para uma nova etapa que é a validação do método 2 para a agregação de 4 fluxos. Os resultados seguem na Tabela 3.6.

Tabela 3.5. Avaliação dos métodos de estimativa do tráfego agregado de dois fluxos.

Nº.	Fluxo 1		Fluxo 2		Método 1		Método 2		Simulação	
	$H_1$	$v_1$	$H_2$	$V_2$	$H_g$	$v_g$	$H_g$	$v_g$	$H_g$	$v_g$
1	0,55	0,50	0,55	1,50	0,55	1,00	0,55	1,00	0,55±0,00	1,00±0,01
2	0,65	0,50	0,65	1,50	0,65	1,00	0,65	1,00	0,65±0,01	1,04±0,19
3	0,75	0,50	0,75	1,50	0,75	1,00	0,75	1,00	0,75±0,01	1,09±0,27
4	0,85	0,50	0,85	1,50	0,85	1,00	0,85	1,00	0,85±0,01	1,04±0,20
5	0,55	1,00	0,65	1,00	0,65	0,50	0,60	1,00	0,60±0,01	0,94±0,11
6	0,55	0,50	0,75	0,50	0,75	0,25	0,65	0,50	0,66±0,01	0,50±0,00
7	0,55	1,50	0,85	1,50	0,85	0,75	0,70	1,50	0,72±0,00	1,50±0,02
8	0,65	1,50	0,75	1,50	0,75	0,75	0,70	1,50	0,70±0,01	1,50±0,02
9	0,65	0,50	0,85	0,50	0,85	0,25	0,75	0,50	0,76±0,01	0,50±0,01
10	0,75	1,00	0,85	1,00	0,85	0,50	0,80	1,00	0,80±0,01	1,00±0,03
11	0,55	0,50	0,65	1,50	0,65	0,75	0,63	1,00	0,63±0,00	1,01±0,01
12	0,55	1,50	0,65	0,50	0,65	0,25	0,58	1,00	0,63±0,01	1,00±0,01
13	0,55	0,50	0,75	1,50	0,75	0,75	0,70	1,00	0,71±0,00	1,01±0,01
14	0,55	1,50	0,75	0,50	0,75	0,25	0,60	1,00	0,60±0,01	1,00±0,01
15	0,55	0,50	0,85	1,50	0,85	0,75	0,78	1,00	0,79±0,00	1,01±0,01
16	0,55	1,50	0,85	0,50	0,85	0,25	0,63	1,00	0,64±0,01	1,01±0,01
17	0,65	0,50	0,75	1,50	0,75	0,75	0,73	1,00	0,73±0,01	1,01±0,01
18	0,65	1,50	0,75	0,50	0,75	0,25	0,68	1,00	0,68±0,00	1,01±0,01
19	0,65	0,50	0,85	1,50	0,85	0,75	0,80	1,00	0,80±0,01	1,04±0,17
20	0,65	1,50	0,85	0,50	0,85	0,25	0,70	1,00	0,71±0,01	1,00±0,01
21	0,75	0,50	0,85	1,50	0,85	0,75	0,83	1,00	0,83±0,01	1,01±0,02
22	0,75	1,50	0,85	0,50	0,85	0,25	0,78	1,00	0,78±0,01	1,01±0,03

Pode-se observar que um aumento no número de fluxos agregados faz que o erro na predição do parâmetro  $H$  para o tráfego agregado aumente ligeiramente, como é observado nos casos 3 e 4. Mesmo com um aumento do erro, o método 2 mostrou-se bastante preciso na estimativa dos parâmetros de interesse, uma vez que os erros continuam muito pequenos.

Tabela 3.6. Avaliação dos métodos de estimativa do tráfego agregado de dois fluxos.

Nº.	Fluxo 1		Fluxo 2		Fluxo 3		Fluxo 4		Método 2		Simulação	
	$H_1$	$v_1$	$H_2$	$v_2$	$H_3$	$v_3$	$H_4$	$v_4$	$H_5$	$v_g$	$H_g$	$v_g$
<b>1</b>	0,55	0,50	0,65	0,50	0,75	1,00	0,85	1,50	0,75	0,88	0,76±0,01	0,90±0,09
<b>2</b>	0,55	1,50	0,65	1,50	0,75	1,50	0,85	0,50	0,67	1,25	0,68±0,01	1,26±0,02
<b>3</b>	0,55	1,00	0,65	1,00	0,75	0,50	0,85	1,00	0,69	0,88	0,71±0,01	0,85±0,06
<b>4</b>	0,55	0,50	0,55	1,00	0,55	1,50	0,85	1,50	0,65	1,13	0,67±0,01	1,14±0,02

### 3.2.5.3. Influência do parâmetro $H$ e do coeficiente de variância

Para avaliar a influência do parâmetro  $H$  e do coeficiente de variância no desempenho da rede, foram feitos dois experimentos que avaliam, em uma rede simulada no NS-2 com as extensões MPLS, as variações de atraso e de perda de pacotes em função destes parâmetros. Foram gerados tráfegos com oito parâmetros  $H$  distintos, para  $H$  entre 0,5 e 0,85, todos com um mesmo coeficiente de variância.

Para cada valor de  $H$ , foi feita uma simulação da qual foram extraídos os parâmetros de desempenho de interesse. Para cada simulação, foram gerados quatro fluxos: F1 (n0 para n5), F2 (n0 para n6), F3 (n8 para n9) e F4 (n0 para n7) que seguem os circuitos virtuais representados na figura 3.8 (a), (b), (c) e (d), respectivamente. Os fluxos F1, F2 e F3 com taxa média de 8 kbps e pacotes de tamanho médio de 400 *bits*, têm o mesmo  $H$  em cada simulação. O fluxo F4, CBR, tem taxa fixa de 1 kbps e tamanho de pacote fixo de 512 *bits*.

As capacidades dos enlaces foram ajustadas de modo a obter o fator de utilização de interesse. As medições dos parâmetros de desempenho foram feitas em cima do tráfego F4. Foram feitas simulações para três valores de fator de utilização distintos. O resultado da simulação segue na Tabela 3.7.

Pode-se observar na Tabela 3.7 que quando se tem um fator de utilização  $\rho$  de 0,6 e 0,7, o tráfego auto-similar só tem expressiva influência no atraso médio dos pacotes da rede para  $H \geq 0,65$ . Para  $\rho = 0,8$ , tráfegos com  $H \geq 0,55$  já surtem efeito negativo sobre o atraso médio sofrido pelos pacotes. Isto nos leva a pensar que para fatores de utilização altos da rede e altos graus de auto-similaridade, podem-se esperar cenários de baixo desempenho em uma rede deste tipo. Também é observado que com  $\rho = 0,6$ , começa a ocorrer perdas de pacotes para  $H \geq 0,80$ , com  $\rho = 0,7$  as perdas começam para  $H \geq 0,75$  e, com  $\rho = 0,8$  para  $H \geq 0,65$ . Conclui-se que o atraso é o parâmetro de desempenho mais sensível ao tráfego auto-similar, ou seja, valores de  $H$  relativamente baixo causam uma degradação significativa no atraso médio sofrido pelos pacotes. Observa-se também que, quanto maior o fator de utilização do sistema, maior o impacto do tráfego auto-similar no desempenho da rede.



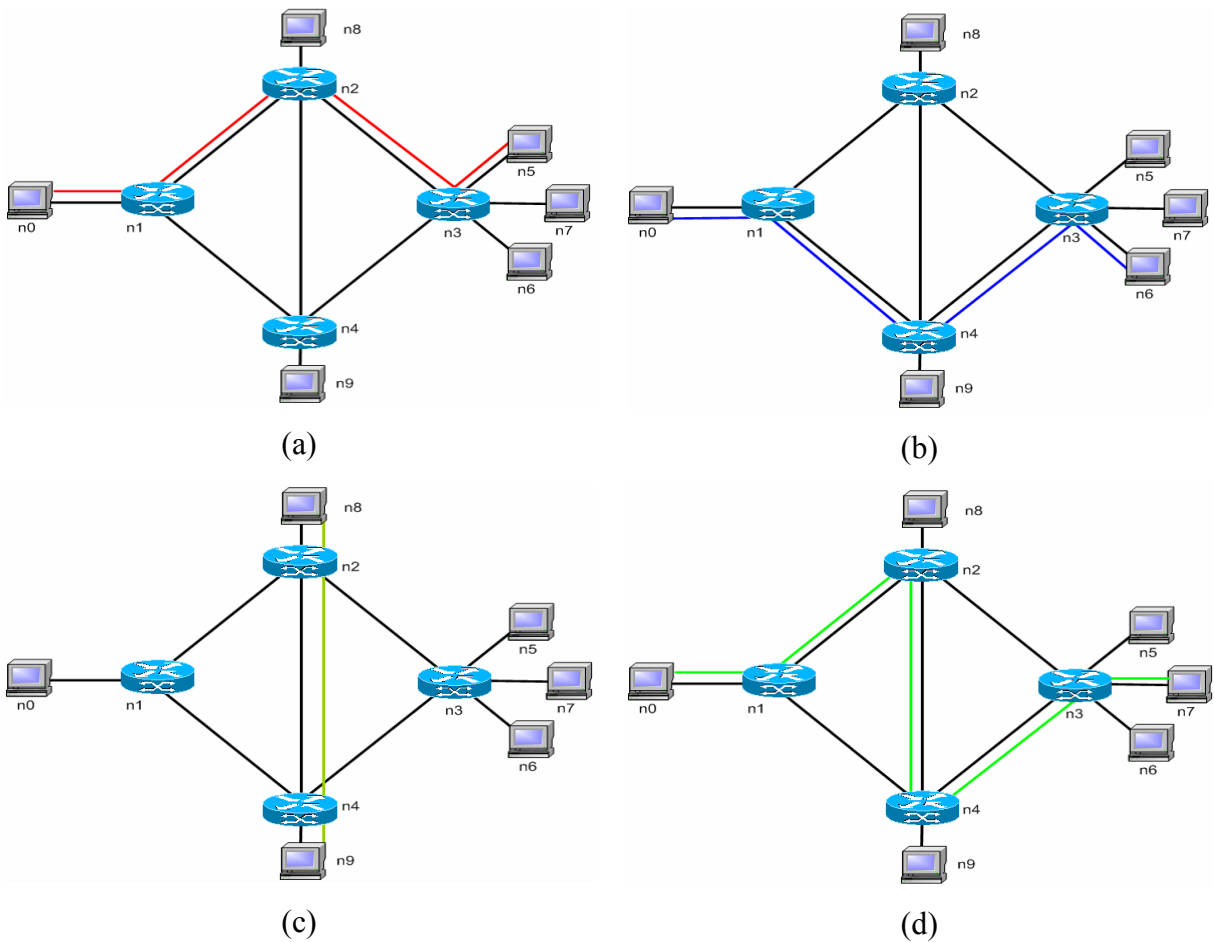


Figura 3.8. Cenário de simulação para avaliação da influência dos parâmetros no desempenho da rede

A análise anterior induz então, a necessidade de considerar a carga do enlace como um elemento no processo de dimensionamento da rede, assim também como para um modelo de cálculo de métricas de QoS.

Tabela 3.7. Resultado da simulação.

H	$\rho = 0,6$			$\rho = 0,7$			$\rho = 0,8$		
	Atraso (seg)	Jitter (seg)	Perda (%)	Atraso (seg)	Jitter (seg)	Perda (%)	Atraso (seg)	Jitter (seg)	Perda (%)
0,50	0,180	0,072	0	0,244	0,091	0	0,418	0,120	0
0,55	0,181	0,072	0	0,248	0,097	0	0,450	0,135	0
0,60	0,188	0,072	0	0,269	0,099	0	0,590	0,132	0
0,65	0,224	0,079	0	0,343	0,104	0,13	0,741	0,163	0,64
0,70	0,253	0,087	0	0,411	0,121	0	0,938	0,194	1,28
0,75	0,287	0,085	0	0,499	0,133	0,64	1,149	0,263	3,33
0,80	0,367	0,095	0,13	0,791	0,151	1,53	1,550	0,375	7,04
0,85	0,668	0,127	1,15	1,117	0,247	5,13	1,752	0,504	12,93

Para avaliar a influência do coeficiente de variância foram gerados tráfegos com seis coeficientes de variância distintos, para  $\nu$  entre 50 e 300 todos com um mesmo parâmetro  $H = 0,7$  e com um mesmo fator de utilização  $\rho$ . Para cada valor de  $\nu$  foi feita uma simulação no NS na mesma rede mostrada na figura 3.9, e foram extraídos os parâmetros de desempenho de interesse.

Pode-se observar na Tabela 3.8 que, um aumento do coeficiente de variância influencia nos parâmetros de atraso, *jitter* e perda de pacotes.

Tabela 3.8. Resultado da simulação.

Coeficiente de variância ( $\nu$ )	Atraso médio	Jitter médio	Pacotes perdidos
050	0,370	0,048	2
100	0,626	0,069	14
150	0,842	0,089	28
200	1,226	0,129	89
250	1,418	0,144	80
300	1,518	0,169	115

Das análises anteriores, conclui-se que, a caracterização do tráfego em redes convergentes com um único parâmetro como é o caso do realizado no modelo M/M/1 e M/M/1/k, é pouco descritivo e não permite avaliar a real influência do tráfego no desempenho da rede. No caso do tráfego auto-similar, a quantidade de tráfego é dada pela média  $a$ , mas a qualidade do tráfego é descrita pelos parâmetros  $H$  e  $\nu$ , e se comprova a sua influência no cálculo das métricas de QoS.

### 3.3. RESUMO DO CAPÍTULO

Neste capítulo foram discutidos os conceitos básicos dos modelos markovianos, M/M/1 e M/M/1/K dentro do contexto de planejamento de redes. Foram apresentados os conceitos básicos dos processos auto-similares e foi feita uma análise comparativa das diferenças fundamentais entre fontes de tráfego de cada tipo.

Foi apresentado o uso do modelo de envoltória fBm que permite a partir da caracterização do tráfego agregado, a derivação de expressões para o cálculo do atraso máximo e probabilidade de perdas de pacotes em uma rede com tráfego multimídia. Foram definidos os conceitos e apresentaram-se expressões para o cálculo da máxima escala de tempo, atraso máximo e probabilidade de perda de pacotes de uma fonte auto-similar para um nó. Estas expressões foram estendidas para uma rede de circuitos virtuais.

Foi feita uma análise do processo de agregação de fluxos auto-similares e foram apresentados os métodos para o cálculo dos parâmetros de Hurst e variância em um tronco de tráfego. Foram feitas análises experimentais da relação entre os parâmetros que definem uma fonte auto-similar, o nível de carga no enlace e as métricas de QoS. Foi descrito o método utilizado para a geração das séries de tráfego auto-similares experimentais deste trabalho.

## 4 - UMA FERRAMENTA PARA GERAÇÃO E ANÁLISE DE TRÁFEGO

Como foi apresentado no capítulo 3, uma das principais vantagens do modelo fBm para tráfego agregado, está na sua simplicidade de caracterização. No entanto, a definição de auto-similaridade apresentada no capítulo 3 dificilmente é atendida por processos reais. Pela definição, a auto-similaridade existe para qualquer escala e a relação entre os fatores de escala se mantém a mesma para todas as escalas. Em tráfegos reais, no entanto, a auto-similaridade só existe para um número finito de escalas e a relação entre os fatores de escala se mantém apenas aproximadamente igual para diferentes escalas. De fato, os tráfegos reais são, tipicamente, assintoticamente auto-similares, ou seja, apresentam características auto-similares apenas para escalas muito grandes.

A primeira generalização de um processo exatamente auto-similar pode ser feita ao considerar que o processo será exatamente auto-similar de segunda ordem quando a relação  $Y(t) \equiv a^{-H}Y(at)$ , para  $a > 0$ ,  $t \geq 0$ , é válida para as estatísticas de até segunda ordem. Essa definição ainda é muito estrita para ser usada como modelo de processos reais, pois generaliza a auto-similaridade apenas no que diz respeito aos momentos de ordem  $n$ . Por esse modelo, todas as estatísticas de até segunda ordem (como variância, função de autocorrelação e coeficiente de correlação) ainda são completamente determinadas pelo parâmetro  $H$  em todas as escalas.

Outra possível generalização é considerar o processo com dependência de longa duração definido pela Eq. 4.1. Essa definição generaliza a função de autocorrelação de um processo auto-similar, em que este processo exhibe, assintoticamente, exatamente o mesmo comportamento da função de autocorrelação de um processo de longa duração. Ou seja, a relação não precisa ser válida para todas as escalas, basta que ela seja válida assintoticamente para escalas grandes.

$$\sum_k r_x(k) = \infty \quad (4.1)$$

Outros modelos que generalizam o modelo auto-similar no que diz respeito à escala são o processo forte assintoticamente auto-similar de segunda ordem e o processo assintoticamente auto-similar de segunda ordem. O processo exatamente auto-similar de

segunda ordem é mais restrito e o assintoticamente auto-similar de segunda ordem é mais geral.

Conforme explicado em (M. S. Taqqu, W. Willinger and B. Sherman, 1997) é possível gerar tráfego auto-similar ao agregar várias fontes de tráfegos VBR (*variable bit rate*) que apresentem uma distribuição do tipo *heavy-tail* para o período *on*. Mas não é possível para uma única fonte gerar o tráfego desta forma devido à necessidade de que o número de fontes agregadas deve ser alto, o que demandaria um alto custo computacional.

Neste capítulo é apresentada uma ferramenta que utiliza os conceitos definidos no capítulo 3 assim como o conceito de tráfego assintoticamente auto-similar discutido anteriormente para a geração de séries sintéticas de tráfego auto-similar em diferentes escalas. Neste capítulo, inicialmente é discutido o processo de inferência estatística, utilizado na ferramenta para a caracterização de tráfego. Posteriormente é descrito o processo de geração de tráfego utilizado pela ferramenta. Na última seção do capítulo a ferramenta é validada com um conjunto de comparações e testes realizados em dados sintéticos e dados reais.

#### **4.1. INFERÊNCIA**

A maior parte dos sistemas reais possui uma ou mais fontes de aleatoriedade. Tipicamente, essas fontes são representadas pela sua esperança ou pela distribuição de probabilidade. Muitas vezes, a representação da aleatoriedade em um sistema por sua esperança pode ser enganosa e normalmente é usada apenas quando não é possível extrair informações suficientes da fonte de aleatoriedade. Todavia, caso não se escolha a distribuição correta para modelagem das fontes de aleatoriedade, a precisão do modelo também pode ser afetada.

Caso seja possível coletar dados do processo estocástico que se deseja modelar, estes podem ser utilizados para simular o processo das seguintes formas:

- a) utilizar os próprios valores dos dados coletados na simulação;
- b) utilizar os próprios valores para definir uma função de distribuição empírica e gerar os valores da simulação de acordo com essa simulação;

- c) utilizar técnicas de inferência estatística para estimar uma forma de distribuição teórica. Os valores da simulação podem ser gerados a partir dessa distribuição teórica.

Se for utilizado o método (a), a simulação só poderá reproduzir o que já aconteceu historicamente no sistema. Quando se utiliza o método (b), pode-se por exemplo utilizar o método de geração por inversão para gerar valores i.i.d (independentes e identicamente distribuídos) com a distribuição empírica dos dados obtidos. Dessa forma, quaisquer valores entre o mínimo valor observado e o máximo valor observado pode ser obtido, evitando-se assim o problema do método (a).

Caso seja possível encontrar uma distribuição teórica que modele os dados observados, o método (c) é preferível aos outros métodos pelas seguintes razões:

- a distribuição teórica obtida a partir dos dados observados pode ter algumas irregularidades, principalmente se a quantidade de dados for pequena. Uma distribuição teórica ameniza essas irregularidades e pode fornecer mais informações sobre o processo estocástico, a partir de parâmetros dessa distribuição. Dessa forma, pode-se gerar um conjunto de valores a partir dessa distribuição ao modificar seus parâmetros de forma a estimar o comportamento do sistema em diferentes situações. Por exemplo, caso se determine que os tempos entre chegadas em uma fila se adaptam a uma distribuição exponencial com determinada média, pode-se estimar como o processo se comportaria caso o tempo médio entre chegadas fosse menor. Assim, seria possível modelar o processo para o caso de uma taxa de utilização maior.
- Com o método (b) para a geração de valores, não é possível gerar valores fora do intervalo dos dados observados. No entanto, algumas vezes o desempenho do sistema depende da probabilidade de ocorrência de eventos extremos, que podem estar fora do intervalo de dados observados.
- Quando a estimação é feita a partir de  $x$  valores, é necessário armazenar  $2x$  valores para representação da distribuição empírica, enquanto que para uma distribuição teórica basta sua probabilidade de distribuição (p.d.f, em inglês) definida por poucos parâmetros.

Além da distribuição, podem ser simuladas outras características do processo que influenciam no seu desempenho, como a auto-similaridade e a dependência entre as amostras. Em alguns casos, a modelagem do processo com base somente em sua distribuição, ao desconsiderar qualquer dependência entre os valores de suas funções amostrais, pode ser enganosa e é necessário inferir sobre outras características estatísticas do processo para sua correta modelagem.

#### 4.1.1. Especificação de Distribuições Teóricas com Base em Dados Observados

A especificação da distribuição teórica que melhor se aproxima a um determinado processo estocástico envolve basicamente os seguintes passos: (1) definição de famílias de distribuições; (2) estimação dos parâmetros das distribuições; (3) determinação do quão representativo é a distribuição escolhida.

O primeiro passo pode-se ajudar do conhecimento teórico sobre o processo estocástico que se deseja modelar. No entanto, dificilmente se tem o conhecimento suficiente sobre o processo físico que permita selecionar apenas uma distribuição. As medidas de média e variância podem dar uma idéia das possíveis distribuições. Outro método de se verificar possíveis famílias de distribuição é ao verificar a forma do histograma, que é basicamente uma estimativa gráfica da função densidade de probabilidade.

Depois de definir um ou mais candidatos para as distribuições, deve-se especificar os valores de seus parâmetros para identificar as possíveis distribuições a serem usadas. Dois métodos clássicos para estimação dos parâmetros são o método de máxima semelhança (MLE) e o método dos momentos (MOM). O método dos momentos estima o parâmetro ao igualar os momentos teóricos aos momentos amostrais (caso o parâmetro possa ser dado em função dos momentos da distribuição) e por meio da resolução da equação da distribuição em função do parâmetro. Por exemplo, no caso de uma distribuição exponencial, o parâmetro  $\lambda$  é igual a  $1/a$ , em que  $a$  é a média do processo. Pelo método dos momentos, o estimador de  $\lambda$  é dado pela Eq. 4.2, onde  $X_i$  são os valores observados e  $n$  é o número de observações.

$$\hat{\lambda} = \frac{1}{\hat{a}} = \frac{n}{\sum_{i=1}^n X_i} \quad (4.2)$$

O estimador da máxima semelhança parte da definição de uma função de semelhança que depende do parâmetro  $\theta$  desejado, como sendo a função de densidade de probabilidade conjunta de se obter os  $n$  valores observados para aquela distribuição. Se os valores  $X_1, X_2, \dots, X_n$  são i.i.d., a função de máxima semelhança é dada pela Eq. 4.3. Dessa forma, o estimador de máxima semelhança é definido como o valor de  $\theta$  que maximiza a função de máxima semelhança. Para o caso das distribuições discretas, a função de máxima semelhança dá a probabilidade de se obter os  $n$  valores observados para a distribuição ao escolher um certo parâmetro  $\theta$ . O parâmetro que maximize a Eq. 4.3 é então o parâmetro que melhor explica os dados observados para determinada distribuição.

$$L(\theta) = p_\theta(X_1)p_\theta(X_2)\dots p_\theta(X_n) \quad (4.3)$$

Como exemplo, para o caso de uma distribuição de Poisson, o método da máxima semelhança parte da Eq. 4.4.

$$L(\lambda) = \prod_{i=1}^n p_\lambda(X_i) = \prod_{i=1}^n \lambda e^{-\lambda X_i} = \lambda^n e^{-\lambda \sum_{i=1}^n X_i} \quad (4.4)$$

Ao resolver a Eq.4.4, ao maximizar  $L(\lambda)$ , o estimador pelo método da máxima semelhança de  $\lambda$  é dado pela Eq. 4.5, o que mostra uma equivalência entre este método e o método dos momentos.

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n X_i} \quad (4.5)$$

Após a determinação das distribuições que podem corresponder aos dados observados, deve-se então verificar o quão bem elas representam a distribuição do processo observado e determinar qual distribuição seria a melhor escolha. Com esses objetivos, são utilizados os métodos de *goodness-of-fit*, que utilizam hipóteses estatísticas para definir se as observações  $X_i$  são amostras independentes de uma determinada distribuição.

#### 4.1.2. Teste de hipótese

Um dos mais usados e mais antigos teste de hipótese é o teste do qui-quadrado. Este teste é uma comparação formal de um histograma com a função de densidade de probabilidade teórica.



Para calcular o valor do qui-quadrado, primeiro deve-se dividir a distribuição em  $k$  intervalos adjacentes. Em seguida verifica-se o número  $N_j$  de observações em cada intervalo. Depois, calcula-se a proporção esperada  $p_j$  de ocorrências em cada intervalo para a distribuição desejada. Para o caso contínuo, considere-se que  $N_j$  seja o número de observações no  $j$ -ésimo intervalo  $[a_{j-1}, a_j]$ , e  $p_j$  definida como a proporção de ocorrência em cada intervalo para a distribuição desejada, definida pela Eq. 4.6, em que  $f(x)$  é a função de densidade probabilidade da distribuição.

$$p_j = \int_{a_{j-1}}^{a_j} f(x) dx \quad (4.6)$$

O valor do qui-quadrado é dado pela Eq. 4.7. Como  $np_j$  é a quantidade esperada de observações que cairiam no  $j$ -ésimo intervalo, se a hipótese fosse verdadeira, o valor de  $\chi^2$  deveria ser pequeno. Logo, a hipótese seria rejeitada caso o valor de  $\chi^2$  fosse muito grande.

$$\chi^2 = \sum_{j=1}^k \frac{(N_j - np_j)^2}{np_j} \quad (4.7)$$

Para a definição dos valores de aceitação da hipótese, parte-se do raciocínio de que, se todos os parâmetros da distribuição fossem conhecidos de antemão e que se os valores observados realmente correspondessem a variáveis aleatórias independentes com a distribuição teórica estimada, então o valor obtido pelo  $\chi^2$  converge, na medida que  $n \rightarrow \infty$ , para uma distribuição de  $\chi^2$ , com  $k-1$  graus de liberdade, em que  $k$  é o número de intervalos. Dessa forma, pode-se definir um valor limite  $\chi^2_{k-1, 1-\alpha}$  para o qual a hipótese seria aceita. O valor limite corresponde ao ponto em que, para uma variável aleatória  $X$  que possua uma distribuição de qui-quadrado com  $k-1$  graus de liberdade, a probabilidade de  $X$  ser maior que esse valor seja  $P(X > \chi^2_{k-1, 1-\alpha}) = \alpha$ .

Uma das maiores dificuldades de se implementar um teste de qui-quadrado consiste em escolher o tamanho e o número de intervalos. Existem basicamente dois modos de se escolherem os intervalos: intervalos iguais ou probabilidades iguais. Em qualquer um dos casos, o número de intervalos deve ser escolhido de forma que cada um tenha um número significativo de amostras. Geralmente, é aconselhado que o método escolhido seja o método de probabilidades iguais de forma que  $np_j \geq 5$  para todo  $j$ .

## 4.2. GERAÇÃO DE TRÁFEGO

Para simular um tráfego, deve-se determinar os tempos de saída (ou chegada) de cada pacote. Como uma série de tráfego real é um processo estocástico, seus valores são aleatórios. Outra possível fonte de aleatoriedade de um tráfego é o tamanho dos seus pacotes. Dessa forma, há a necessidade de gerar variáveis aleatórias que possam representar os tempos entre dois pacotes sucessivos e os tamanhos de cada pacote.

Uma ferramenta básica para a geração de variáveis aleatórias de acordo com qualquer distribuição é uma fonte de variáveis aleatórias independentes e identicamente distribuídas com distribuição uniforme entre 0 e 1. A maior parte das linguagens de programação possuem geradores de números pseudo-aleatórios. Eles são “pseudo” por serem totalmente reproduzíveis com algoritmos matemáticos, mas cumprem os requerimentos estatísticos de serem valores independentes com distribuição uniforme.

Grande parte dos geradores presentes nas linguagens de programação utiliza o método cônico-linear, que é um algoritmo recursivo definido pela Eq. 4.8, em que  $t$ ,  $z$  e  $w$  são números inteiros positivos, com  $t < w$  e  $z < w$ , em que  $r$  é chamado de valor semente.

$$r_{n+1} = (r_n t + z) \bmod w \quad (4.8)$$

Um valor inicial de  $r_0$  deve ser escolhido para ser o primeiro elemento da seqüência de números aleatórios, de tal forma que, com essa relação, pode-se gerar uma seqüência periódica com período máximo de  $w-1$ . Os valores escolhidos para  $z$ ,  $w$  e  $t$  determinarão se a seqüência gerada terá características satisfatórias, ou seja, se seus valores não podem ser distinguidos de variáveis aleatórias i.i.d. com distribuição uniforme. Valores de  $w$  muito usados em softwares de simulação são  $w=2^{31}$  e  $w=2^{48}$  (Law A.M., Kelton, W.D, 1991).

Para a geração dos diferentes tipos de tráfego são utilizados vários métodos. Para o caso do tráfego de Poisson, é utilizado o método da inversão (A.M Law and W. D. Kelton, 1991), em que variáveis aleatórias distribuídas exponencialmente são utilizadas para representar o tempo de chegada entre pacotes.

Os tipos de tráfego VBR (*variable bit rate*) gerados a partir de aplicações tais como vídeo compactado e transferência de arquivos são caracterizados por períodos de atividade (*on*) e períodos sem atividade (*off*). Para modelar esses tráfegos, deve ser configurada a distribuição dos períodos *on* e *off* e o tipo de tráfego enviado durante o período *on*. A

seguir serão detalhados os métodos principais utilizados para a geração de tráfego da ferramenta apresentada neste capítulo.

#### 4.2.1. Geração por Inversão

O método por inversão pode ser usado para gerar variáveis aleatórias i.i.d. que possuam uma distribuição conhecida ou mesmo uma distribuição empírica. O algoritmo é muito simples e consiste basicamente em gerar uma variável aleatória  $U$  com distribuição uniforme entre 0 e 1. A variável aleatória desejada é dada por  $X=F^{-1}(U)$ , em que  $F^{-1}(U)$  é o inverso da p.d.f desejada.

Como  $F^{-1}(U)=X$ , tem-se que  $P(X\leq x)=P(F^{-1}(U)\leq x)=P(U\leq F(x))$ . Como  $U$  é uniformemente distribuída entre 0 e 1, então  $P(U\leq a)=a$ . Logo  $P(X\leq x)=F(x)$  e a variável aleatória  $X$  possui a p.d.f. desejada.

#### 4.2.2. Geração de Processo fGn

Um processo fGn não pode ser gerado pelo método da geração por inversão, pois, nesse caso, não basta uma variável aleatória i.i.d. com distribuição gaussiana. Por ser um processo auto-similar, deve-se modelar também o efeito do parâmetro  $H$  que influencia diretamente a função de autocorrelação do processo. Ou seja, os valores gerados devem ser correlacionados de forma que a função amostra possua o parâmetro de Hurst desejado.

O método utilizado para a geração de processos fGn escolhido para a ferramenta de geração de tráfego apresentada neste capítulo se baseia no espectro de potência dos processos fGn e foi apresentado por Vern Paxson (Paxson V., 1997). Esse método gera um processo auto-similar em tempo discreto  $G(t)$  definido conforme a Eq. 4.9, em que  $G(t)$  representa o número de ocorrências no  $t$ -ésimo intervalo de tamanho  $\delta$ . No caso da geração de tráfego, cada ocorrência representa um pacotes enviado ou recebido, ou seja,  $G(t)$  representa o número de pacotes enviados no  $t$ -ésimo intervalo.

$$G(t) = Y(t\delta) = X((t+1)\delta) - X(t\delta) \quad (4.9)$$

A geração de uma função amostra fGn consiste em construir inicialmente uma seqüência de valores  $\{f_1, f_2, \dots, f_{n/2}\}$  correspondentes ao espectro de potência de um processo fGn para as frequências de  $2\pi/n$  a  $\pi$ . Cada  $f_i$  deve ser multiplicada por variáveis aleatórias exponenciais i.i.d. com média 1, o que resulta na seqüência  $\{f'_i\}$ . Caso  $X$  seja um processo

auto-similar, as ordenadas do periodograma de sua função amostral são assintoticamente independentes e exponencialmente distribuídas com média  $f(\psi, H)$ , em que  $f(\psi, H)$  é o valor do espectro de potência de um processo fGn com parâmetro  $H$  na frequência  $\psi$ .

Após a geração da seqüência  $\{f'_i\}$  deve-se gerar uma seqüência de valores complexos  $\{z_1, z_2, \dots, z_{n/2}\}$ , de tal forma que  $|z_i| = \sqrt{f'_i}$  e a fase de  $z_i$  seja uniformemente distribuída entre  $0$  e  $2\pi$ . Essa fase aleatória preserva o espectro de potência do processo, já que este depende unicamente do módulo de  $z_i$ , com a garantia de que diferentes funções amostras geradas por esse método sejam diferentes. Todavia, essa fase aleatória faz com que as distribuições marginais do resultado final sejam gaussianas, um requerimento para o processo fGn. A seqüência  $\{z_1, z_2, \dots, z_{n/2}\}$  representa a transformada de Fourier discreta do sinal para as frequências positivas.

Após a geração da seqüência  $\{z_1, z_2, \dots, z_{n/2}\}$ , deve-se expandir esta seqüência para obter a seqüência  $\{z'_i\}$ , que representa a transformada de Fourier discreta para frequências positivas e negativas. Como o objetivo é geração de tráfego, o sinal deve ser real, e dessa forma, deve-se expandir  $\{z_1, z_2, \dots, z_{n/2}\}$  de forma que seu módulo seja uma função par e sua fase seja uma função ímpar. Finalmente, é calculada a transformada de Fourier inversa de  $\{z'_i\}$  para gerar a função amostra fGn  $\{x_i\}$ .

Como a amostra fGn  $\{x_i\}$  é gerada a partir do espectro de potencia de um processo fGn, ela possui propriedades de autocorrelação de um processo fGn, já que a autocorrelação de um processo estocástico e seu espectro de potencia formam um par de transformada de Fourier. A seqüência gerada possui média  $0$ , no entanto, como em um processo fGn a média, a variância e o parâmetro  $H$  são independentes, pode-se obter qualquer média e variância por médio de transformadas lineares, sem alterar o parâmetro de Hurst.

No caso da geração de tráfego, cada valor da seqüência fGn  $\{x_i\}$  representa o número de pacotes em um intervalo de tamanho  $\delta$ . Com esse critério, a geração de tráfego deve-se utilizar o tempo de saída ou chegada de cada pacote, e não a quantidade de pacotes por intervalo de tempo. Dessa forma, deve-se definir o tamanho de intervalo  $\delta$  e distribuir os pacotes dentro desse intervalo. Como em um intervalo somente pode-se gerar um número inteiro de pacotes, devem-se converter os valores reais da seqüência obtida para valores inteiros.

O problema maior está em distribuir os valores dentro do intervalo de forma a garantir a auto-similaridade para escalas menores que  $\delta$ . No entanto, como foi discutido no início deste capítulo, os tráfegos reais são apenas assintoticamente auto-similares, ou seja, não são auto-similares para escalas pequenas. Com esse critério, podem-se distribuir os pacotes de maneira uniforme dentro do intervalo  $\delta$  pequeno, o que garante a auto-similaridade para escalas maiores que essa.

### 4.3. VALIDAÇÃO DA FERRAMENTA

Existem hoje ferramentas que geram séries de tráfego (Netspec 2006, Netperf 2006, D-ITG 2006, MGEN 2006) com diferentes características. A ferramenta desenvolvida para gerar as séries de tráfego auto-similar utilizadas neste trabalho, a GTAR (Gerador de Tráfego e Analisador de QoS na Rede), é um *software* de código aberto multiplataforma que oferece a funcionalidade de gerar séries de tráfego para múltiplas fontes e destinos.

A ferramenta está dividida em quatro módulos básicos:

1. Módulo de sincronismo
2. Módulo de geração de tráfego
3. Módulo de inferência
4. Análise de desempenho

O módulo 1 é utilizado para sincronizar os elementos de rede, tarefa prévia à geração de tráfego entre eles, com o objetivo de obter métricas de desempenho em redes reais, tarefa realizada pelo módulo 4. Como o foco da utilização do GTAR neste trabalho de tese está nos módulos 2 e 3, os módulos 1 e 4 não serão detalhados neste texto. Maiores detalhes sobre estes módulos podem ser encontrados em (Solis Barreto P., de Carvalho, P. H. P., Soares, A. M., Abdalla Jr H., Bizerra R. S, 2005, Solis Barreto P., de Carvalho, P. H. P., Soares, A. M., Abdalla Jr H., Lambert, R., Amvame-Nze, G., Bizerra R.S 2006).

A interface principal para a geração de séries de tráfego sintéticas do GTAR é apresentada na Fig. 4.1, onde aparecem as telas de especificação dos parâmetros tais como média, variância, parâmetro  $H$  e tamanho do intervalo para um processo de incrementos gaussiano auto-similar.

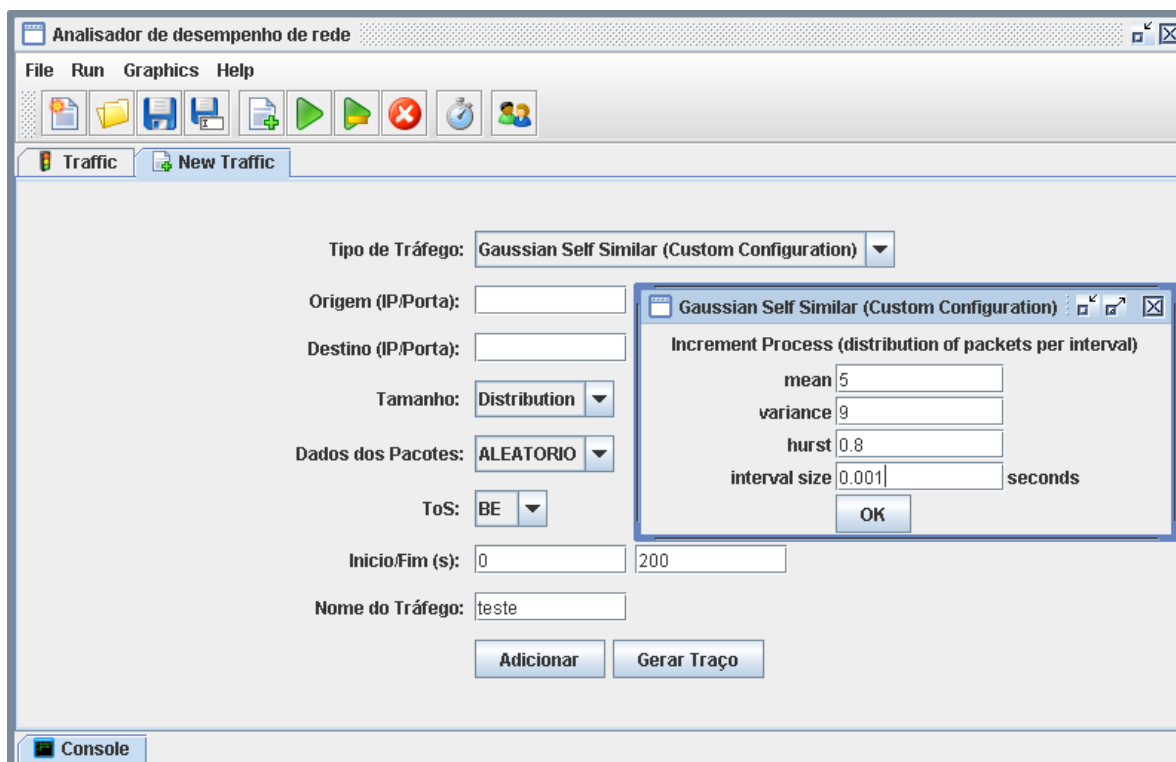


Figura 4.1. Tela principal para geração de tráfego do GTAR

O módulo de geração de tráfego permite que vários tipos de tráfego sejam gerados. Por exemplo, um dos casos mais simples, a geração de tráfego CBR (*Constant Bit Rate*) é feita ao enviar um pacote a cada  $1/T_x$  segundos, onde  $T_x$  é a taxa de transmissão de pacotes por segundo definido pelo usuário. A geração dos outros tipos de tráfego, por exemplo o tráfego de Poisson, utiliza o método da inversão. A geração do tráfego auto-similar é feita com o método detalhado na seção 4.2.2.

A carga útil dos pacotes enviados pelo GTAR podem ser dados aleatórios ou dados de um arquivo especificado pelo usuário. O tamanho do pacote também deve ser especificado pelo usuário. A carga útil é então encapsulada em um pacote UDP. O cabeçalho de aplicação do pacote tem dois campos: Timestamp (8 bytes) e número de seqüência (4 bytes). O campo de timestamp é usado para calcular as métricas de QoS da série de tráfego, principalmente para avaliar desempenho de séries de tráfego em séries operativas, elemento útil para análise de desempenho. O número de seqüência serve para, no nó receptor, identificar os pacotes que são perdidos na rede.

Para cada fluxo de tráfego gerado pelo GTAR, um arquivo de log é criado. Este arquivo contém o número de seqüência, tamanho e tempo de partida para cada pacote. Esse tipo de funcionalidade é especialmente interessante para a geração de séries de tráfego sintéticas

que depois podem ser utilizadas como arquivos de entrada em ambientes de simulação, como o NS-2.

O módulo 3 do GTAR, um módulo de inferência permite a avaliação das propriedades estatísticas, tais como média, variância e parâmetro  $H$ , o que permite a análise de séries de tráfego coletadas de ambientes operativos reais.

#### **4.3.1. Validação da Geração de Tráfego**

Para avaliar o GTAR foram feitos dois tipos de geração de tráfego: o primeiro, para testar a eficácia da geração do tráfego baseado em variáveis independentes com uma distribuição conhecida e o segundo para analisar se o tráfego auto-similar gerado possui as características de auto-similaridade esperada. Em ambos casos, é feita uma comparação com outra ferramenta bastante utilizada no ambiente acadêmico para geração de tráfego, o MGEN (MGEN, 2006).

Para o primeiro tipo, foram geradas series de tráfego Poisson com altas taxas. Com os arquivos de log das series geradas foram inferidos os parâmetros estatísticos. A Tabela 4.1 mostra o resultado da media, variância (esperada e observada) e o valor qui-quadrado, com 30 graus de liberdade para cada fluxo.

Os dados da Tabela 4.1 mostram que o tráfego gerado pela ferramenta mantém a media e variancia próximas daquela esperada para taxas de 50 e 100 pacotes por segundo. Entretanto, para altas taxas, por exemplo, 500 pps, a variância do MGEN é consideravelmente diferente da esperada. Os valores  $\chi^2$  mostram que a distribuição de tempo entre pacotes do tráfego gerado pelo GTAR é mais próxima de uma distribuição exponencial que a do tráfego gerado pelo MGEN.

Tabela 4.1. Valores esperados e observados e valores qui-quadrado (30 graus de liberdade)

	Esperado			Observado		$\chi^2$
	Taxa (pps)	Media (s)	Variância (s <sup>2</sup> )	Média (s)	Variância (s <sup>2</sup> )	
GTAR	50	0,02	0,0004	0,0200034	0,000403112	34,15
	100	0,01	0,0001	0,0100829	0,00010173	24,37
	500	0,002	0,000004	0,00199886	0,0000039757	167,32
MGEN	50	0,02	0,0004	0,01968649	0,0003972880	33842
	100	0,01	0,0001	0,01001523	0,000116622	107624
	500	0,002	0,000004	0,00202174	0,0000164278	494132

As figuras 4.2 e 4.3 mostram as distribuições de tempo entre pacotes para os tráfegos gerados pelo GTAR e MGEN, respectivamente, para uma taxa de 100 pps. Pode-se conferir que a distribuição do tráfego gerado pelo GTAR é mais próxima de uma distribuição exponencial.

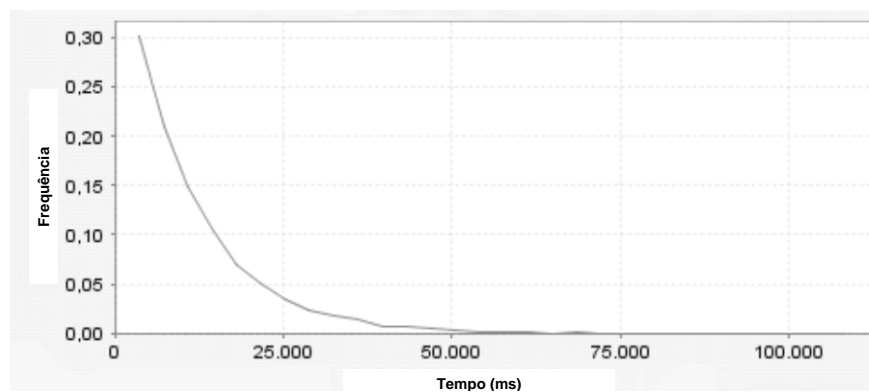


Figura 4.2. Distribuição de tempo entre pacotes para o tráfego Poisson gerado pelo GTAR.

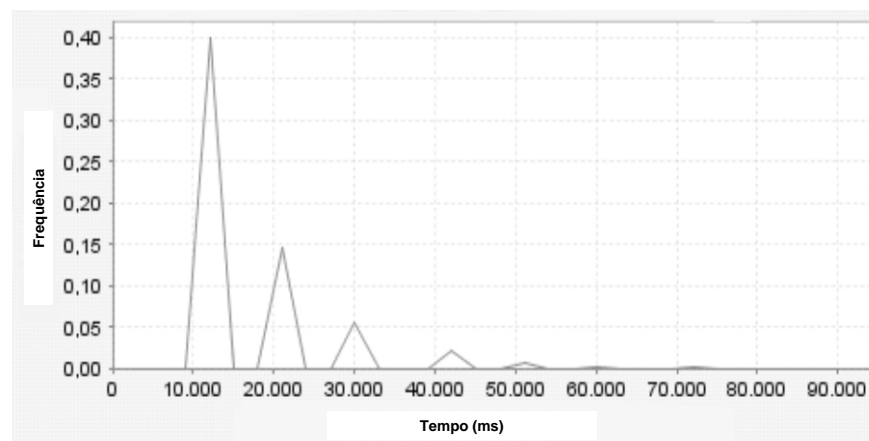


Figura 4.3. Distribuição de tempo entre pacotes para o tráfego Poisson gerado pelo MGEN.



Para avaliar o processo de geração de tráfego auto-similar foram gerados seis tipos de tráfego com diferentes valores de  $H$ , com uma taxa de 200 pps. O tempo de duração de cada avaliação foi de uma hora. O processo de incrementos para este tráfego deve ser um processo fGn com  $H$  igual ao especificado pelo usuário. O número de pacotes enviados por unidade de tempo deve ter uma distribuição fGn.

A Tabela 4.2 mostra a média e o parâmetro de Hurst para o processo de incrementos do tráfego. A escala de tempo escolhida foi de 0,1 segundos. O parâmetro  $H$  das séries foi calculado usando quatro métodos diferentes: o método Whittle (V. Paxson, 1997, T. Karagiannis, M. Faloutsos e R.H. Riedi, 2002, Richard Clegg, 2005), o VTP (V. Paxson, 1997), o método do periodograma (T. Karagiannis, M. Faloutsos e R.H. Riedi, 2002, Richard Clegg, 2005) e o método Kettani-Gubner (Kettani H., Gubner J. A., 2002). Os resultados obtidos mostram que os parâmetros estimados estão próximos dos esperados. Este resultado demonstra que a ferramenta consegue gerar o tráfego auto-similar de acordo com os parâmetros especificados.

Tabela 4.2. Inferência estatística no tráfego auto-similar gerado pelo GTAR

Esperado		Observado				
Média	H	Média	(Whittle)	(VTP)	Periodograma	(K-G)
20	0.65	19.995	0.654	0.653	0.658	0.654
20	0.70	20.120	0.699	0.700	0.712	0.698
20	0.75	19.953	0.750	0.751	0.763	0.749
20	0.80	20.202	0.791	0.787	0.808	0.788
20	0.85	20.351	0.848	0.830	0.848	0.837
20	0.90	20.743	0.894	0.880	0.912	0.882

#### 4.3.2. Estimação de tráfegos reais

Ao estimar o comportamento de um processo real, deve-se realizar um pré-processamento dos dados para remover possíveis *trends* e periodicidade, que podem afetar a estimação. Um processo que não possua dependência de longa duração e apresenta componentes com certa periodicidade pode enganar os estimadores, podendo revelar, falsamente, valores altos para o parâmetro  $H$ .

Uma forma muito utilizada para a remoção dessas componentes indesejadas é passar o sinal por um filtro passa-altas. No entanto, o interesse desta validação se foca no parâmetro

$H$ , principalmente em sua influência em grandes escalas, ou seja, na dependência de longa duração do processo. Por isso, deve-se avaliar o comportamento do sinal em grandes escalas, o que significa avaliá-lo em baixas frequências, impossibilitando o uso de um filtro passa-baixas para remoção de *trends* e periodicidade. Em (Clegg R., 2005) são propostos três outros métodos para realizar o pré-processamento dos dados:

- calcular o logaritmo nos dados originais (apropriado somente quando os dados são positivos);
- remoção do trend linear (retirar a reta que melhor se ajusta aos dados);
- remoção do polinômio de décima ordem que melhor se ajusta aos dados.

O método utilizado no conjunto de dados avaliados nesta seção é o método da remoção do polinômio de décima ordem. Apesar de esses métodos serem capazes de remover o *trend* do sinal, muitas vezes eles não são capazes de remover a periodicidade.

Os dados utilizados nesta seção correspondem ao tráfego médio do *backbone* de uma grande operadora de Internet. As medidas de tráfego reais são classificadas por aplicação, coletadas em diferentes períodos de tempo. O tráfego é caracterizado utilizando-se 4 métodos diferentes do parâmetro  $H$ . É feita uma comparação do parâmetro de Hurst para diferentes aplicações em diferentes escalas de tempo.

A rede avaliada é um provedor de acesso de grande porte nacional, com mais de dez milhões de assinantes RTFC (Rede de Telefonia fixa comutada) e mais de um milhão de assinantes ADSL (*Assymetric Digital Subscriber Line*). A rede IP é formada pelos níveis mostrados na Fig. 4.4, com níveis de acesso, distribuição, núcleo e Internet. Cada camada de acesso tem um roteador PPPoX (*Point to Point Protocol*) chamada BRAS (*Broadband Router Access Server*), e é conectada a dois roteadores de distribuição pelo menos. O roteador de distribuição tem capacidade de 200 Gbps para agregação de tráfego e 40 Mpps de encaminhamento. Todas as interfaces IP são Gigabit ou 10 Gigabit Ethernet. Em alguns casos, a interface destas redes é a rede de transporte SDH sobre DWDM.

O protocolo MPLS é utilizado no *backbone* desde o nível de acesso até o nível de núcleo. A rede usa Diffserv e Intserv.

O ponto de coleta de dados foi uma interface Gigabit Ethernet com mais de 50.000 assinantes de banda larga ADSL. As escalas para o tráfego coletado foram: 5 minutos, um minuto, um milissegundo, classificando-se o tráfego por aplicação ao usar o Cisco Netflow (Netflow, 2006) e o aplicação de código aberto Ethereal (Ethereal, 2006).

As aplicações da rede podem ser divididas em duas categorias principais: residencial e corporativo. A principal diferença entre ambas é o período de tempo de maior utilização. No caso da categoria corporativa este período é de segunda-feira a sexta-feira das 08:30 às 19:00 com picos das 11:00 às 15:00 como pode ser conferido na Fig. 4.5 para o tráfego downstream. As series de tráfego nas unidades de Mbps foram normalizadas em um período de amostragem de 31 dias. Na Fig. 4.6 é mostrado um maior detalhe do comportamento das amostras de tráfego em um período de sete dias.

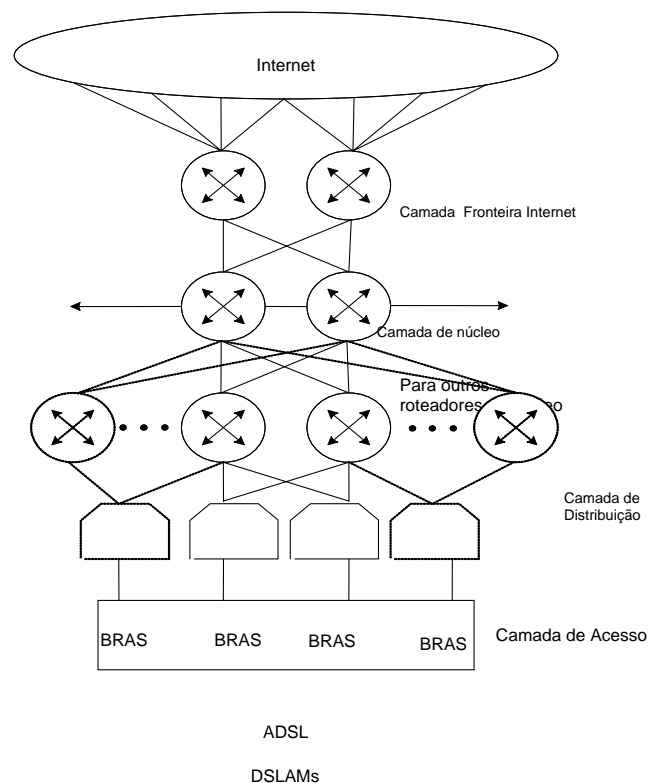


Figura 4.4. Esquema do processo de TE com caracterização de tráfego

Na Fig. 4.7, é mostrada uma análise mais detalhada das aplicações em um período de 24 horas. Na Fig. 4.8 é mostrada a classificação por aplicação para tráfego *downstream* durante o dia. Como pode ser observado, a fonte mais importante de tráfego é o HTTP (*Hyper Text Transfer Protocol*) seguido pelas aplicações *peer-to-peer*. Também as aplicações VoIP (*Voice over IP*) não são representativas em termos de volume de tráfego.

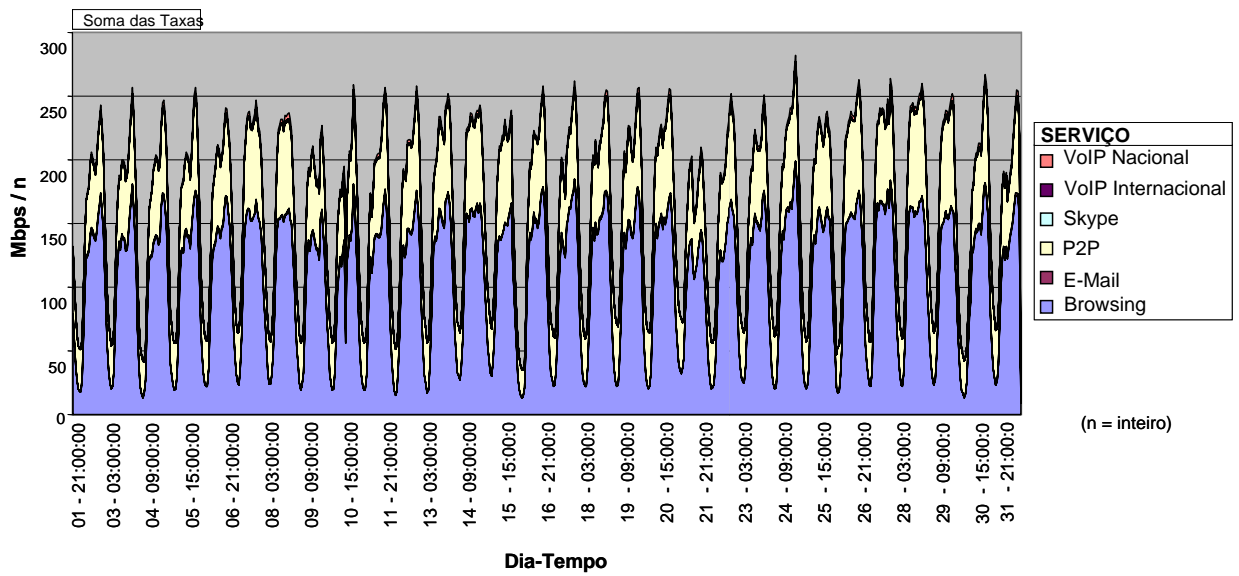


Figura 4.5. Período de utilização para tráfego downstream – 31 dias

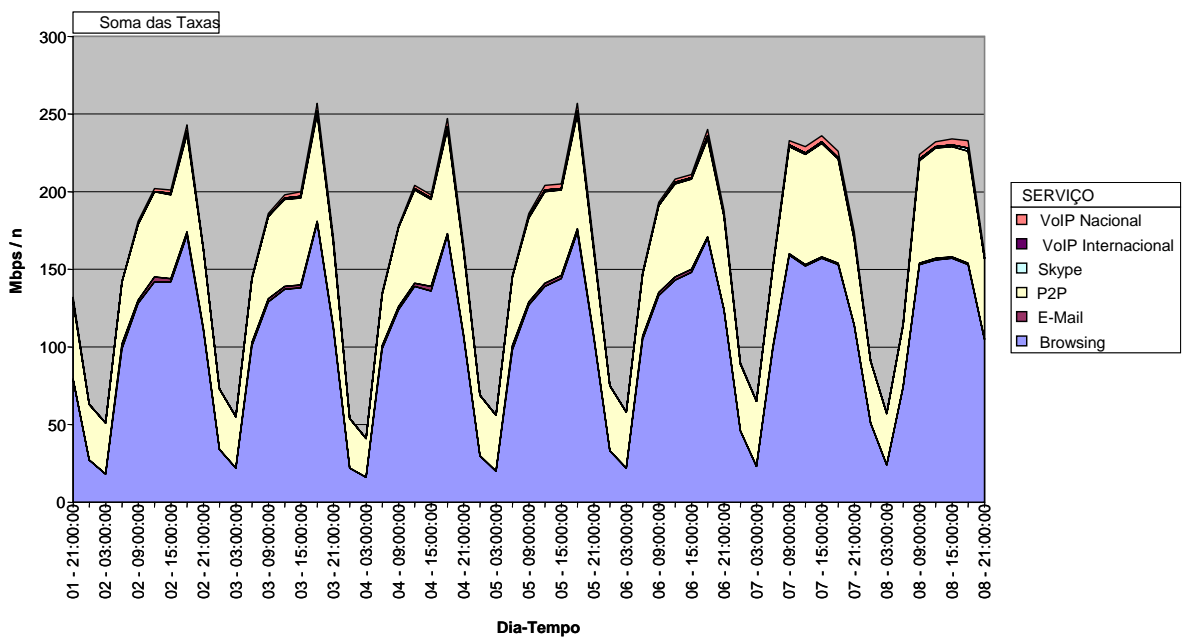


Figura 4.6. Comportamento do tráfego por serviço para sete dias

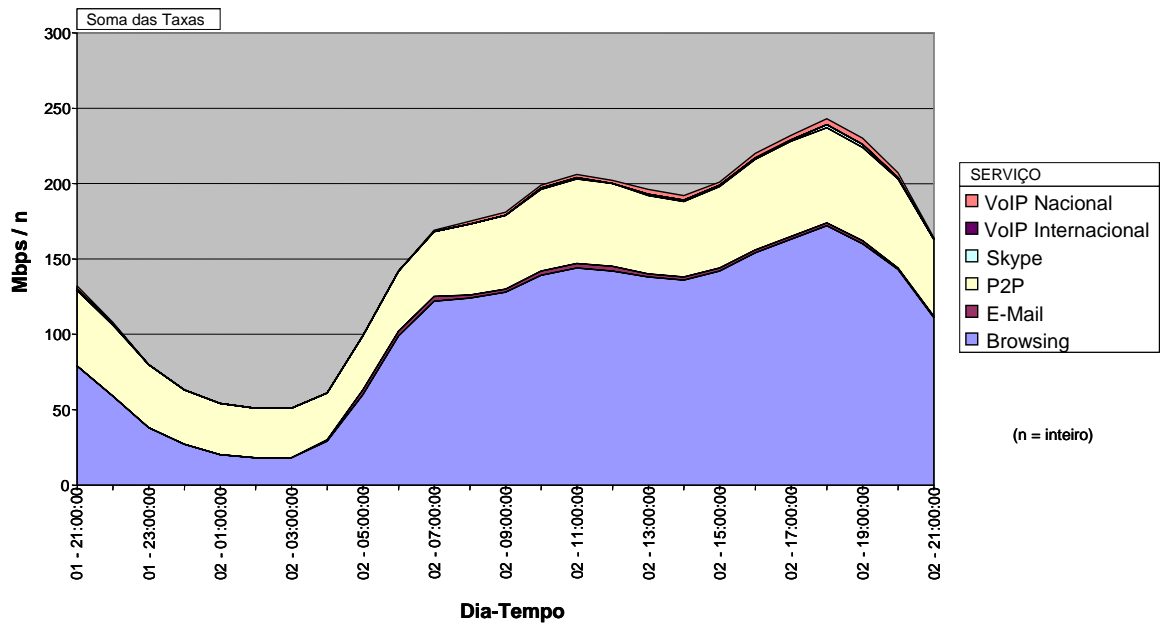


Figura 4.7. Comportamento do tráfego por serviço - 24 horas

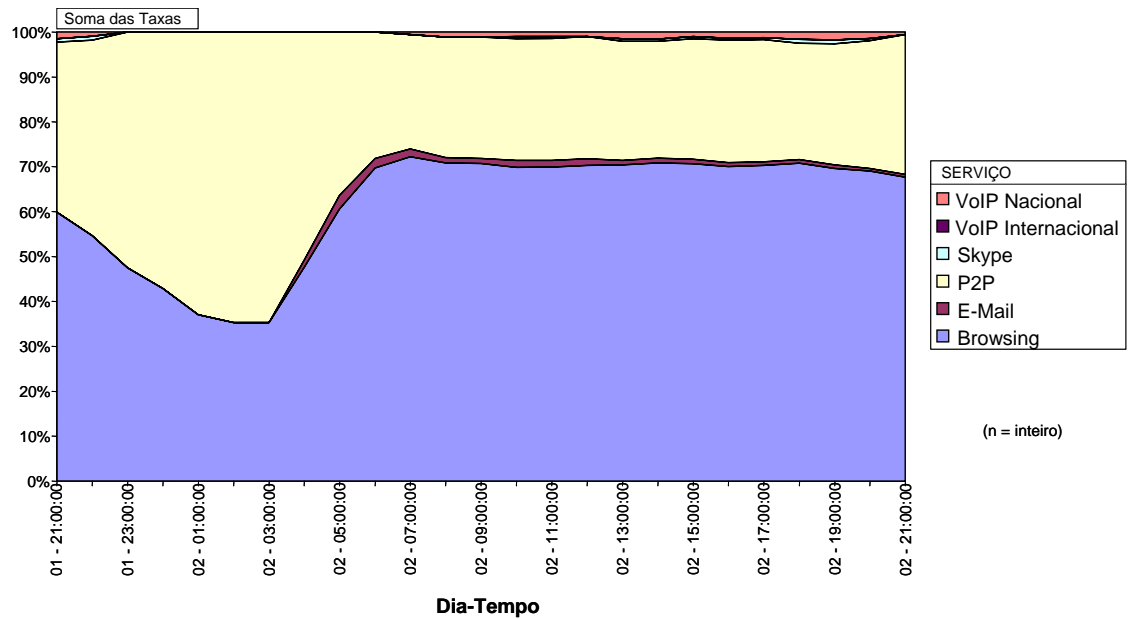


Figura 4.8. Percentuais de tráfego por serviço

A análise considera uma separação do tráfego por aplicações. Para as diferentes séries de tráfego foi calculado o parâmetro  $H$  com dois métodos diferentes: o método VTP e o método Kettani-Gubner. Foi realizada uma análise qui-quadrado como um teste de significância, devido ao fato da necessidade de verificar a similaridade da distribuição. Os enlaces onde foram feitas as coletas mostram uma carga de no máximo 70%.

Na Tabela 4.3, é mostrado o parâmetro  $H$  para medidas de tráfego na escala de 5 minutos. Como pode ser verificado, o parâmetro tende a valores maiores que 0,9, indicando um alto grau de longa dependência nas amostras. A terceira coluna que mostra a variável qui-quadrado para a distribuição Gaussiana, infere que estes valores não foram aceitáveis para outras distribuições. Os valores mostram que na escala avaliada, o tráfego no *backbone* realmente possui características auto-similares e segue uma distribuição gaussiana, o que indica que o tráfego deve ser, pelo menos, assintoticamente auto-similar.

Tabela 4.3. Parâmetro de Hurst para amostras de 5 minutos de tráfego

Dia	VTP	Kettani-Gubner	$\chi^2$ Distribuição Gaussiana
1	0,928	0,965	47,618
2	0,951	0,976	70,653
3	0,944	0,967	51,924
4	0,916	0,943	37,069
5	0,910	0,943	16,403
6	0,911	0,941	26,306
7	0,955	0,974	42,451
8	0,945	0,973	31,903
9	0,955	0,979	46,972
10	0,939	0,964	48,694
11	0,887	0,944	21,354
12	0,916	0,947	52,785
13	0,954	0,978	35,132
14	0,947	0,970	51,924
15	0,948	0,973	45,250
16	0,951	0,975	23,722
17	0,938	0,966	38,146

Na Tabela 4.4, é mostrada a estimação do parâmetro  $H$  para aplicações HTTP (*Hyper Text Transfer Protocol*). Os valores resultantes estão entre 0,67 e 0,93. Observar que o valor de 0,63 aparece somente no dia 12.

Na Tabela 4.5, são mostrados os valores de  $H$  para aplicações P2P (*Point to Point*). Os valores calculados estão entre 0,86 e 0,96.

Tabela 4.4. Parâmetro de Hurst para *http* para amostras de 5 minutos de tráfego

Dia	VTP	Kettani-Gubner	$\chi^2$ Distribuição Gaussiana
1	0,843	0,895	31,042
2	0,812	0,878	71,299
3	0,901	0,926	38,146
4	0,900	0,934	52,569
5	0,815	0,879	32,549
6	0,816	0,904	62,042
7	0,865	0,906	17,910
8	0,870	0,916	39,653
9	0,907	0,935	28,028
10	0,867	0,919	21,785
11	0,869	0,906	27,167
12	0,671	0,861	35,778
13	0,878	0,909	36,208
14	0,839	0,894	44,604
15	0,874	0,907	30,611
16	0,753	0,850	23,292
17	0,851	0,914	40,299

A Tabela 4.4 mostra os resultados das series de tráfego na escala de 1 ms. Como pode ser observado, as aplicações VoIP apresentam um parâmetro  $H$  entre 0,79 e 0,83.

Tabela 4.5. Parâmetro de Hurst para P2P para amostras de 5 minutos de tráfego

Dia	VTP	Kettani-Gubner	$\chi^2$ Distribuição Gaussiana
1	0,915	0,948	63,549
2	0,942	0,962	49,771
3	0,937	0,963	45,250
4	0,902	0,935	28,243
5	0,902	0,928	20,708
6	0,901	0,942	30,181
7	0,939	0,964	52,354
8	0,932	9,964	52,354
9	0,937	0,968	39,007
10	0,922	0,948	38,576
11	0,860	0,926	37,500
12	0,904	0,942	33,840
13	0,937	0,965	55,799
14	0,922	0,958	46,972
15	0,935	0,963	46,757
16	0,935	0,967	49,986
17	0,933	0,964	37,285

Tabela 4.6. Parâmetro de Hurst para amostras na escala de 1ms

Aplicação	VTP	Kettani-Gubner
Tráfego Total	0,927	0,945
http	0,916	0,924
P2P	0,890	0,871
VoIP	0,839	0,799

O número de assinantes xDSL em todos os casos é de 50.000 pelo menos, com um média de 40% de conexões simultâneas.

Os resultados mostrados confirmam o alto grau de auto-similaridade para o tráfego agregado na maioria das aplicações.



#### 4.4. RESUMO DO CAPÍTULO

Neste capítulo foram apresentados os conceitos para a implementação do GTAR (Geração de Tráfego e Análise de Rede). A ferramenta tem quatro módulos principais: o sincronismo, a geração de tráfego, a inferência e a análise de rede. A utilidade básica da ferramenta neste trabalho de tese está na geração de séries de tráfego sintéticas com características de Poisson e auto-similares conforme os parâmetros  $H$ , variância e média. O módulo de inferência estatística permite avaliar séries de tráfego originárias de ambientes operativos reais.

O processo de geração de tráfego da ferramenta foi avaliado para os tráfegos Poisson e auto-similar, a partir de uma comparação com o tráfego gerado por uma outra ferramenta de geração de tráfego amplamente conhecida e de uso acadêmico, o MGEN. O resultado da comparação mostra que o GTAR gera séries de tráfego sintéticas conforme os parâmetros especificados e com uma maior precisão de que o MGEN.

O módulo de inferência do GTAR foi utilizado para caracterizar o tráfego real coletado em um período de 17 dias, em uma escala de 5 minutos de um provedor de grande porte. A análise dos resultados, para o tráfego total resultante das agregações de várias aplicações e para as aplicações *http* e P2P mostraram um grau de longa dependência nas séries de tráfego real, com o parâmetro  $H$  calculado por três métodos diferentes. O teste de hipótese, com a utilização do qui-quadrado com 30 graus de liberdade, comprovou que uma distribuição gaussiana é a que se ajusta melhor às séries de tráfego, o que permite concluir que o tráfego é assintoticamente auto-similar para este tipo de aplicações.

## **5 - UM MODELO DE TRÁFEGO HÍBRIDO PARA CARACTERIZAÇÃO DE TRÁFEGO MULTIMÉDIA**

Neste capítulo, é apresentada a proposta de um modelo de tráfego híbrido para o cálculo de métricas de QoS em uma rede multimídia. As métricas de interesse são a probabilidade de perda de pacotes e o atraso máximo.

São apresentados dois cenários de simulação para validar o modelo de tráfego híbrido ao comparar resultados analíticos com resultados simulados em um ambiente de simulação de redes com extensões para MPLS. Nestes cenários é feito um conjunto de análises experimentais realizadas com o objetivo de definir um modelo de tráfego que seja capaz de produzir métricas de QoS mais precisas para uma rede com tráfego multimídia. As análises são realizadas com o objetivo de entender as características essenciais do comportamento do tráfego e de suas métricas de QoS com o objetivo de definir o domínio de validade do modelo.

### **5.1. CENÁRIOS DE SIMULAÇÃO E CONSIDERAÇÕES INICIAIS PARA A DEFINIÇÃO DE UM MODELO DE TRÁFEGO**

Nesta seção, o objetivo é fazer uma análise comparativa entre os atrasos obtidos em dois cenários de rede em que é feito o cálculo dos atrasos ao considerar que as fontes de tráfego são do tipo Poisson e este mesmo cálculo quando as fontes de tráfego são auto-similares. São estabelecidos dois cenários: a) uma rede pequena, com um núcleo de roteamento com 4 nós e 4 fluxos de tráfego, ilustrada na figura 5.1, onde à direita da rede aparecem as rotas fixas de cada um dos fluxos; b) um núcleo de roteamento com 12 nós e 10 fluxos de tráfego, ilustrado na figura 5.2, com os fluxos roteados como mostrado na Tabela 5.3.

Nas seções seguintes, para os dois cenários anteriores serão avaliados os modelos de tráfego M/M/1 e o modelo fBm, com diferentes tipos de agregações de tráfego com o objetivo de obter métricas de atraso e perda de pacotes.

Para o cenário (a), as características de cada fluxo são detalhadas na Tabela 5.1 assim como o atraso máximo tolerável de cada um, informação que o modelo de tráfego deve considerar para o cálculo de capacidades que permitam obter essas métricas. Então é esperado que o processo de dimensionamento seja capaz de fornecer um conjunto de capacidades  $C = \{C_{ij}\}$  que obedeça à restrição dos atrasos para cada fluxo.

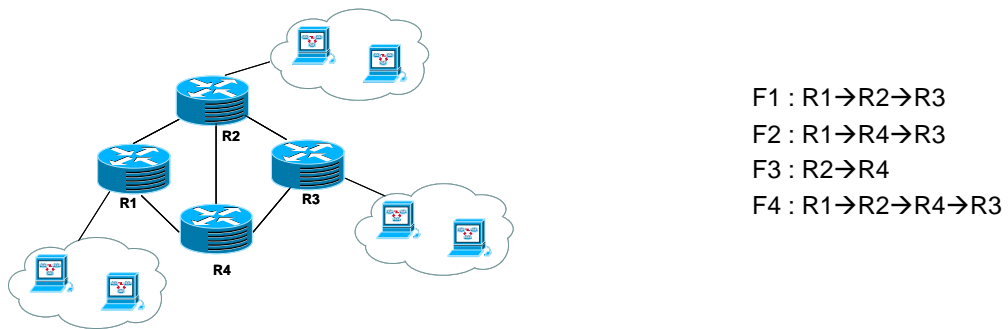


Figura 5.1. Rede do cenário (a) e definição das rotas por fluxo.

Tabela 5.1. Descrição dos fluxos para o cenário (a)

	F1	F2	F3	F4
tipo	VBR	VBR	VBR	CBR
taxa (bps)	8.000	8.000	8.000	1.000
tamanho médio do pacote (bits)	400	400	400	512
atraso máximo (segundos)	0,11	0,11	0,6	0,18

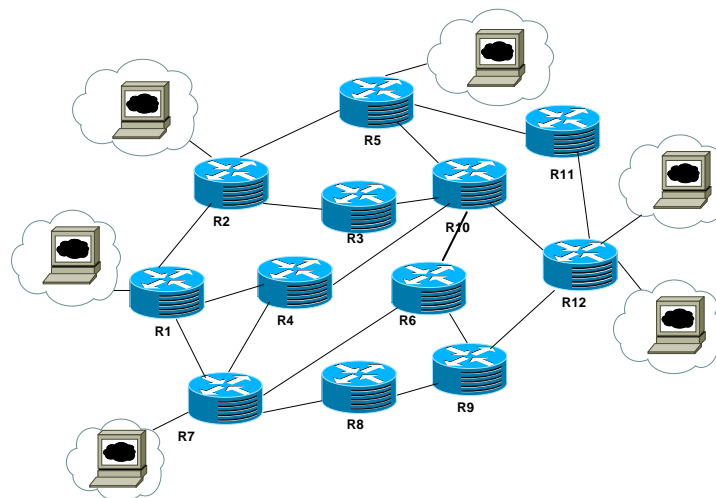


Figura 5.2. Rede do cenário (b).

Tabela 5.2. Fluxos definidos para o cenário (b)

Fluxo	F1	F1	F3	F4	F5	F6	F7	F8	F9	F10
Tipo	VBR	VBR	CBR	VBR	CBR	VBR	CBR	VBR	CBR	VBR
Tamanho do pacote (bits)	400	400	512	400	512	400	512	400	512	400
Taxa média (bps)	8.000	8.000	1.000	8.000	1.000	8.000	1.000	8.000	1.000	8.000

Tabela 5.3. Tabela de rotas para o cenário (b)

F1,F3	R1→R4→R10→R12
F2,F4	R1→R4→R7→R6→R9→R12
F5	R1→R2→R3→R10→R12
F6	R5→R10→R12
F7	R1→R4→R10→R6→R9→R12
F8	R7→R8→R9→R12
F9	R1→R2→R5→R11→R12
F10	R2→R3→R10→R12

### 5.1.1. O Modelo M/M/1 e suas limitações

Para avaliar o comportamento de cada um dos cenários quando dimensionados com diferentes modelos de tráfego, considera-se que o valor do atraso para cada fluxo, conforme aparece nas Tabelas 5.1 representa um conjunto de restrições, ou seja, é o valor máximo admissível do atraso para cada fluxo. Para o cálculo do atraso, considera-se que o processo de chegada de tráfego é um processo de Poisson e cada um dos roteadores se comporta como uma fila M/M/1, de tal forma que o atraso individual por fluxo é calculado pela Eq. 3.9.

Para cada cenário, foi realizada uma simulação no NS-2 para validar os resultados do modelo analítico, neste caso o M/M/1. Na Tabela 5.4 é mostrada, para o cenário (a), uma comparação dos resultados obtidos no modelo analítico com os resultados da topologia de rede simulada no NS.

Tabela 5.4. Comparação dos atrasos para o cenário (a)

Modelo	F1	F2	F3	F4
Modelo analítico	0,08	0,10	0,03	0,14
Modelo simulado	0,06	0,08	0,03	0,14

Os resultados da Tabela 5.4 consideram que os fluxos têm um processo de chegadas de Poisson e modelam cada roteador como uma fila M/M/1. Os fluxos F1, F2 e F3 têm um parâmetro  $H = 0,5$ , que tipifica um tráfego de Poisson. Os valores de pacotes perdidos para todos os fluxos na simulação foram iguais a zero pois neste caso objetiva-se somente analisar o atraso e dessa forma o tamanho máximo das filas considera-se ilimitado.

O resultado anterior fornece uma topologia para um roteamento fixo e fluxos de tráfego descritos por um processo de Poisson. Se os fluxos apresentam a mesma taxa média, mas não seguem uma distribuição de Poisson para o processo de chegadas, espera-se que os resultados de atraso na rede não correspondam aos calculados pelo modelo M/M/1. Para comprovar este argumento, os fluxos foram substituídos por séries de tráfego em que o processo de chegadas apresenta longa dependência, mas com a mesma taxa média utilizada nos fluxos originais. Foram feitas simulações no NS-2 e comprovou-se que o atraso sofre variações consideráveis em relação àquelas mostradas na Tabela 5.4. O resultado é mostrado na Tabela 5.5.

Tabela 5.5. Comparação dos atrasos para o cenário (a) no modelo simulado

Modelo	F1	F2	F3	F4
Tráfego Poisson	0,06	0,08	0,03	0,14
Tráfego auto-similar $H=0,8$	0,17	0,20	0,07	0,29

Como era esperado, o resultado da Tabela 5.5 comprova que a função de atraso, que considera cada roteador como uma fila M/M/1, não consegue fornecer resultados que permitam obter um desempenho mínimo para fluxos auto-similares, pois subestima o valor do atraso para todos os fluxos, com diferenças percentuais de até 100%. Esse resultado corresponde às observações práticas em redes operativas e experimentais (Figueiredo Daniel R., Liu B., Feldmann A., Misra V., Towsley D., Willinger W., 2005) e tem suas motivações nos aspectos discutidos no Capítulo 3.

A rede do cenário (a) é uma rede pequena e a aplicação do modelo analítico é simples. No caso de redes maiores, calcular um conjunto de capacidades que definem uma topologia inicial de operação para um conjunto de métricas de desempenho para vários fluxos, torna-se um processo mais trabalhoso.

Para ilustrar uma comparação para tipos diferentes de tráfego, foram alocadas as capacidades de rede no cenário (b). Esta rede trabalha com 10 fluxos, detalhados na Tabela 5.2, e utiliza uma tabela de roteamento detalhada na Tabela 5.3.

Na figura 5.3, ilustra-se para cada fluxo o atraso calculado com o modelo analítico M/M/1, com a simulação com fluxos do tipo Poisson e com a simulação com fluxos com  $H=0,8$ , em todos os casos, para a mesma topologia.

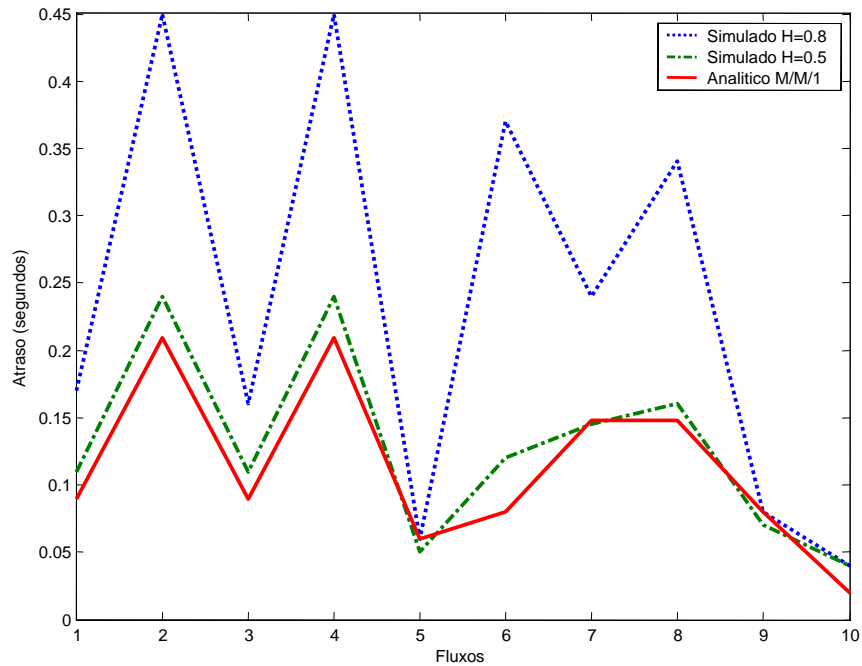


Figura 5.3. Rede Comparação dos atrasos por fluxo

Como pode-se observar, se os fluxos são do tipo Poisson, o modelo M/M/1 é bastante adequado. Mas, como na prática, nas redes de pacotes, os fluxos geralmente apresentam diversos graus de auto-similaridade, com valores de  $H > 0,5$  e altas variâncias, comportamento simulado na rede do cenário (b) com os fluxos da Tabela 5.2, se comprova que as capacidades calculadas com o uso do modelo M/M/1 não são adequadas para esse tipo de tráfego, conforme ilustrado na figura 5.3 em que a linha tracejada mostra o atraso calculado em simulações no NS-2.

Como foi visto anteriormente na Tabela 5.5, podem existir diferenças de até de 100% entre os resultados do modelo analítico o da simulação. No caso do cenário (b), e com a informação da Tabela 5.3 pode-se observar que existe uma maior agregação de fluxos nos enlaces e que essa agregação, em função de cada parâmetro das fontes de tráfego, tende a preservar a auto-similaridade, observação também comprovada na análise da seção 3.2.5.2.

Da análise anterior, pode-se comprovar que, ao considerar o fluxo como uma fonte de Poisson, quando de fato a fonte é auto-similar, obtêm-se maiores atrasos que os esperados, após o processo de planejamento de rede. É dessa forma que se ressalta a importância da correta modelagem do tráfego auto-similar para a obtenção de métricas mais precisas de QoS.

### 5.1.2. O modelo de envoltória fBm para caracterização do tráfego

Como foi apresentado na seção 3.2 os modelos fBm e fGn podem descrever as características do tráfego auto-similar. A envoltória fBm é um processo que descreve, dentro de uma margem probabilística, o comportamento de tráfego agregado de uma fonte auto-similar. Assim também na seção 3.2 esses conceitos foram estendidos para uma rede com circuitos virtuais. As considerações iniciais para caracterizar o tráfego em uma rede multimídia são as seguintes:

- a) os fluxos são caracterizados pelas variáveis  $(H, a, \sigma^2)$ , onde  $H$  é o parâmetro de Hurst,  $a$  é a média e  $\sigma^2$  é a variância;
- b) a agregação dos fluxos produz tráfego com características de auto-similaridade;
- c) A rede tem roteamento fixo e é orientada a conexão, como é o caso da tecnologia MPLS.

Com as considerações anteriores, observa-se que o ponto de partida para a caracterização de tráfego é o cálculo do parâmetro  $H$ , assim como a variância e a média. Assim, também, pelo processo de agregação dos fluxos nos diferentes enlaces da rede com diversos parâmetros  $H$ ,  $a$  e  $\sigma^2$  os valores resultantes destes parâmetros após a agregação em um enlace devem ser calculados e espera-se que a auto-similaridade seja uma propriedade a ser preservada no tráfego agregado (Solis Barreto P., de Carvalho, P. H. P., Soares, A. M., Abdalla Jr H., 2005).

Em uma rede operativa, o cálculo destes parâmetros deve ser realizado de forma contínua, conforme os conceitos apresentados no capítulo 3. Então, é natural que se deseje que o custo computacional desse cálculo seja baixo, mas ao mesmo tempo preciso.

O ponto de interesse nesta seção é avaliar se os conceitos apresentados na seção 3.2 para a modelagem do tráfego auto-similar permitem obter métricas mais precisas de QoS. Para esse efeito inicialmente será analisado o cálculo do atraso, com os mesmos cenários das figuras 5.1 e 5.2.

Para o cenário (a) e o cenário (b) respectivamente foram calculados os parâmetros de interesse nos fluxos agregados nos diferentes enlaces e para a caracterização dos fluxos foi utilizado o método de Kettani e Gubner, mostrado na seção 3.2.5 para o cálculo do

parâmetro de Hurst e o método 2 da seção 3.2.5 para o cálculo dos parâmetros do tráfego agregado para cada um dos cenários de avaliação. Os fluxos utilizados no cenário (a) e no cenário (b), com os parâmetros de caracterização, estão mostrados nas Tabelas 5.6 e 5.7, respectivamente.

Tabela 5.6. Fluxos definidos para o cenário (a)

	Fluxo 1	Fluxo 2	Fluxo 3	Fluxo 4
Tipo	VBR	VBR	VBR	CBR
Taxa média (bps)	8.000	8.000	8.000	1.000
Tamanho médio do pacote (bits)	400	400	400	512
Variância (bps)	1.250	1.250	1.250	50
Parâmetro de Hurst	0,8	0,8	0,8	0,5

Tabela 5.7. Fluxos definidos para o cenário (b)

Fluxo	F1	F1	F3	F4	F5	F6	F7	F8	F9	F10
Tipo	VBR	VBR	CBR	VBR	CBR	VBR	CBR	VBR	CBR	VBR
Tamanho do pacote (bits)	400	400	512	400	512	400	512	400	512	400
Taxa média (bps)	8.000	8.00	1.000	8.000	1.000	8.000	1.000	8.000	1.000	8.000
Variância (bps)	1.250	1.250	50	1.250	50	1.250	50	1.250	50	1.250
Parâmetro de Hurst	0,8	0,8	0,5	0,8	0,5	0,8	0,5	0,8	0,5	0,8

Na figura 5.4 é apresentada uma comparação entre os atrasos para o fluxo F4 da Tabela 5.7, para capacidades de rede previamente calculadas com o modelo M/M/1 e com o modelo de envoltória fBm. Pode-se observar que, para cargas no enlace inferiores a 0,3, o modelo M/M/1 é bastante aproximado do valor simulado, com uma tendência a subestimar o valor real. No caso do método da envoltória fBm, os valores estimados dos atrasos, para cargas acima de 0,3 e 0,6, estabelecem limites máximos do atraso. Esses resultados são equivalentes aos observados em outros trabalhos (Mayor G. e Silvester J, 1998) para o caso de redes ATM.



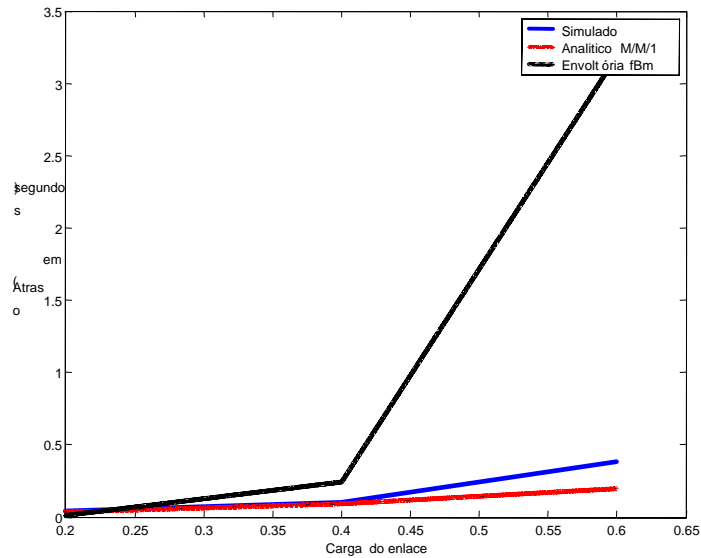


Figura 5.4. Comparativo dos atrasos para o fluxo F4 e H=0.8

Na figura 5.5 é apresentada uma outra comparação entre os atrasos para o fluxo F4 da Tabela 5.10. Para este caso, o fluxo F4 foi modificado para uma relação de coeficiente de variância ( $a/\sigma^2$ ) 150 vezes maior que o apresentado na Tabela 5.7 e foram calculadas as capacidades da rede com o modelo M/M/1 e com o modelo de envoltória fBm. Pode-se observar que, neste caso, a sobre-estimação do atraso do modelo de envoltória fBm diminuiu em relação ao observado na Figura 5.4. Novamente, para cargas no enlace inferiores a 0,3, o modelo M/M/1 é bastante aproximado do valor simulado, com uma tendência a subestimar o valor real. No caso do método da envoltória fBm, os valores estimados dos atrasos, para cargas acima de 0,3 e 0,6, estabelecem limites máximos do atraso.

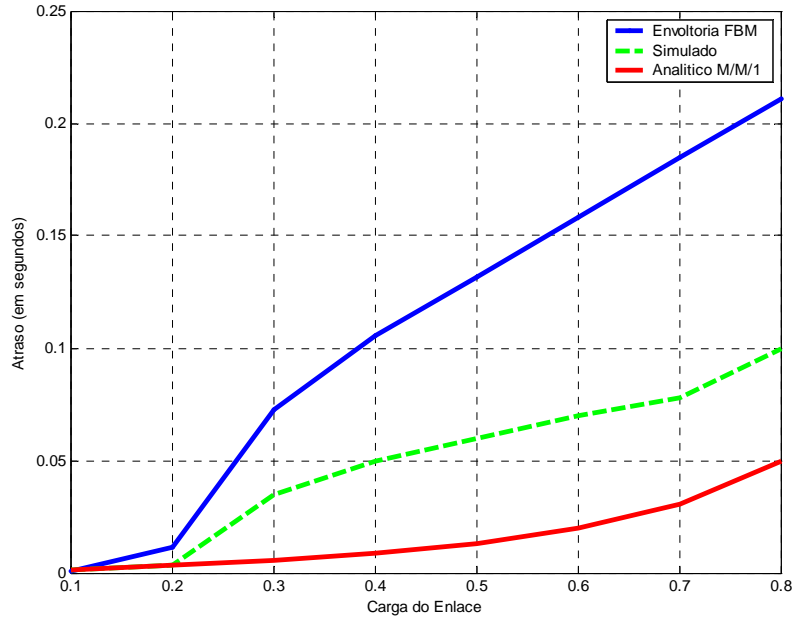


Figura 5.5. Comparativo dos atrasos para o fluxo F4 e  $H=0.8$  e variância maior

Na avaliação dos cálculos mostrados na figura 5.5 pode-se observar que para cargas de até 0,20 ou 0,30, o modelo M/M/1 para o cálculo do atraso, ainda quando a fonte é auto-similar, é bastante próximo do simulado, mas para cargas acima deste valor existe uma tendência a sub-estimar.

O modelo de envoltória fBm com o uso da Eq. 3.7 tende a sobre-estimar o valor de atraso, pois como foi explicado na seção 3.2, estabelece valores limites para esse valor. Na comparação mostrada nas Figuras 5.4 e 5.5 considerará-se que o tamanho máximo da fila é infinito, o que resulta em valores ainda maiores para a sobre-estimação do atraso com o modelo da envoltória fBm.

### 5.1.3. Um modelo de tráfego híbrido para cálculo do atraso

Nesta seção o objetivo é avaliar o comportamento da combinação dos dois modelos, o modelo M/M/1 e o modelo de envoltória fBm para o cálculo do atraso nos dois cenários de simulação. O uso do modelo analítico do cálculo de atraso discrimina para cada fluxo em função dos parâmetros  $H$ ,  $a$  e  $\sigma^2$  e da carga do enlace, a obtenção do atraso com o modelo M/M/1 e com o modelo de envoltória fBm. Os critérios básicos definem um **modelo de tráfego híbrido** e são os seguintes:

1. A caracterização da agregação dos fluxos em cada um dos enlaces se realiza com o método de Kettani e Gubner, mostrado na seção 3.2.5 para o cálculo do parâmetro de Hurst agregado e o método 2 da seção 3.2.5 para o cálculo da variância agregada;
2. Para cargas abaixo de 30%, independentemente do valor de  $H$  e  $\sigma^2$  ou quando o parâmetro  $H$  é menor ou igual a 0,5, o cálculo das métricas de QoS se realiza com o modelo M/M/1;
3. Para cargas entre 30% e 100%, o cálculo das métricas se realiza ao calcular a máxima escala de tempo do fluxo agregado. O atraso se calcula conforme a Eq. 3.47.

Para o cenário (a), Figura 5.1, foram estabelecidos atrasos máximos associados a cada fluxo. o processo de dimensionamento forneceu um conjunto de capacidades dos enlaces que produziram valores de atraso para cada fluxo mostrados na figura 5.6. Para o cenário (b), os resultados dos atrasos de cada fluxo são mostrados na figura 5.7. Neste caso, pode-se observar, que para os fluxos F1, F2, F6 e F9, o modelo analítico em relação à simulação, subestimou os valores do atraso, mas com uma margem muito pequena. No caso do F1, a subestimação foi de 0,009, para F2, foi de 0,002, para F6 foi de 0,013 e para F9 foi de 0,008.

Pode-se observar que o modelo analítico forneceu um conjunto de capacidades com valores maiores daqueles obtidos na simulação.

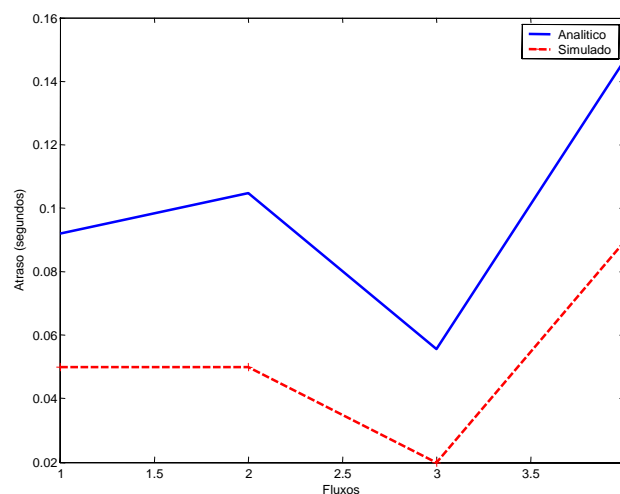


Figura 5.6. Rede Comparação dos atrasos por fluxo – cenário (a)

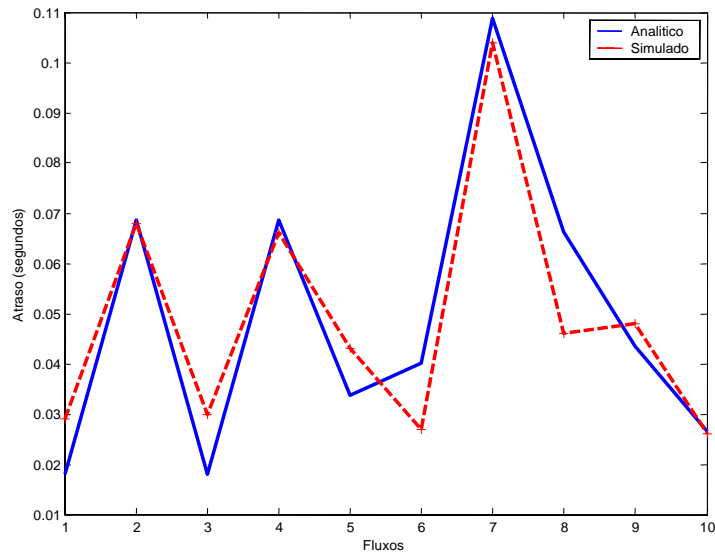


Figura 5.7. Rede Comparação dos atrasos por fluxo – cenário (b)

As tabelas de fluxos 5.6 e 5.7 mostram uma combinação de parâmetros de tráfego bastante simplificada. É interessante poder modificar relações de coeficiente de variância ( $a/\sigma^2$ ) e valores de  $H$  e avaliar o impacto nas métricas de QoS sob as considerações anteriormente feitas. Tem-se observado que, para valores altos do coeficiente de variância, existe uma maior dificuldade de descrever o tráfego (Jerkins J. e Wang J.L, 1999, Solis Barreto P., de Carvalho, P. H. P., Soares, A. M., Abdalla Jr H., 2005) e conseqüentemente, suas métricas. O mesmo se observa para valores altos do parâmetro  $H$ . Outro elemento de interesse é observar o comportamento das métricas de QoS diante de diferentes cargas no enlace. Na seguinte seção pretende-se avaliar o modelo híbrido de tráfego proposto para diferentes variações do parâmetro  $H$  e  $\sigma^2$  e estendê-lo para o cálculo do percentual de perdas de pacotes.

## 5.2. DEFINIÇÃO DE UM MODELO DE TRÁFEGO PARA CÁLCULO DE ATRASO E PERDAS DE PACOTES EM UMA REDE MULTIMÍDIA

Conforme foi mencionado no capítulo 3, os modelos para o tráfego auto-similar provêm aproximações assintóticas já que permitem estabelecer fronteiras máximas e mínimas das métricas de QoS para um certo fluxo. Existe também, conforme foi mostrado na seção 3.2.5.4, uma relação entre o parâmetro  $H$ , a variância e a carga do enlace para o cálculo das métricas de desempenho. O ponto de interesse nesta seção é, então, avaliar o

comportamento do modelo de tráfego híbrido em uma rede com circuitos virtuais para diferentes ternas de valores  $(H, a, \sigma^2)$  e a sua relação com diferentes cargas no enlace.

Para esse efeito, com o uso de um gerador de tráfego, o GTAR, já descrito no Capítulo 4, foram geradas várias séries sintéticas de tráfego auto-similar com diferentes valores de  $H$  e  $\sigma^2$ . Foram geradas oito séries de tráfego auto-similar sintético na escala de 0,01 segundos conforme a tabela 5.8.

Assim também foram geradas 4 séries de tráfego Poisson, com  $H=0,5$ , taxa média de 39 kbps e variância de  $4,998377 \times 10^4$  na escala de 0,1 segundos. Dessa forma, os fluxos F1, F2, F4, F6, F8 e F10 são auto-similares e os fluxos F3, F5, F7 e F9 são markovianos.

Na figura 5.8 são mostrados 100 segundos da série 9vbr01 na escala de 0,1 segundos. Para esse agrupamento, foi feito o cálculo da variância utilizando-se o método VTP (Morin P. R., Neilson J., 1995, Pereira Dantas M.J., 2000). A figura 5.9 mostra a envoltória fBm calculada para a série 9vbr01 com uma probabilidade  $\epsilon=0,1$ .

Tabela 5.8. Parâmetros dos fluxos auto-similares

No.	Identificador	Média	Variância (escala 0.1s)	H
1	9vbr01	121 kbps	$1,600439 \times 10^6$	0,84
2	9vbr02	124 kbps	$5,565116 \times 10^5$	0,87
3	9vbr03	122 kbps	$2,421986 \times 10^5$	0,87
4	9vvr04	122 kbps	$1,509093 \times 10^5$	0,85
5	8vbr01	125 kbps	$9,793116 \times 10^5$	0,78
6	8vbr02	124 kbps	$3,304075 \times 10^5$	0,74
7	8vbr03	123 kbps	$1,903485 \times 10^5$	0,79
8	8vbr04	123 kbps	$1,191379 \times 10^5$	0,78

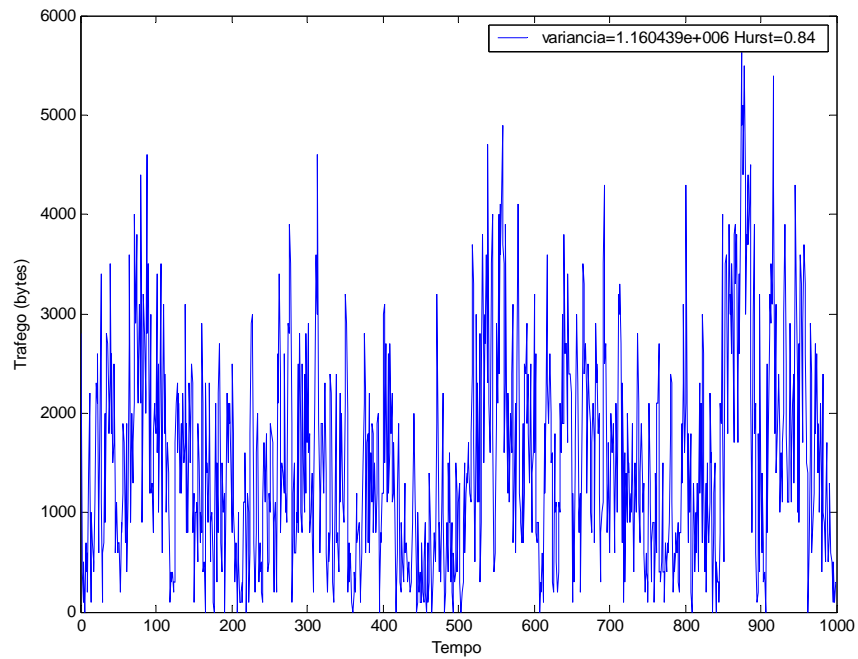


Figura 5.8. Série de tráfego 9vbr01

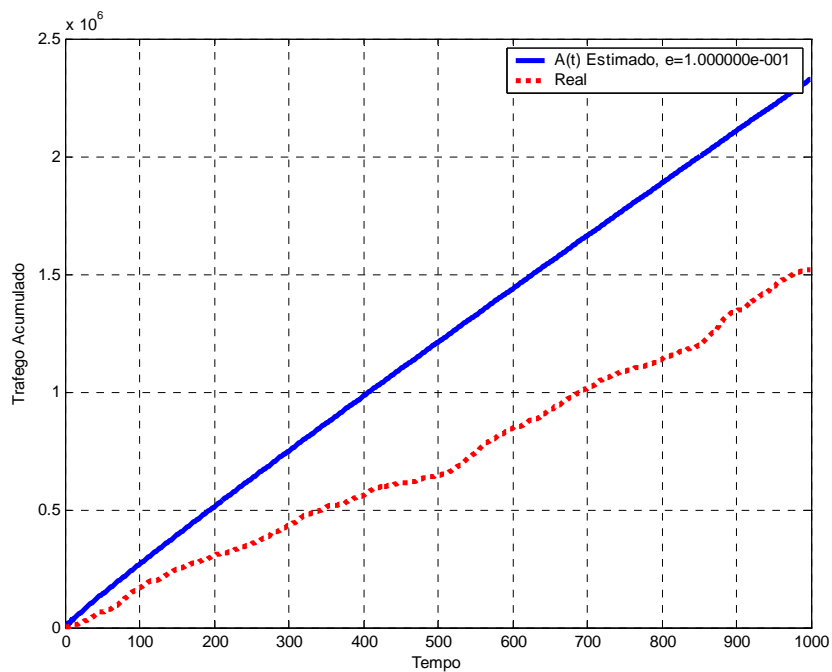


Figura 5.9. Envoltória fBm para a série 9vbr01

As figuras 5.10 a 5.15 mostram as séries e as envoltórias fBm para 9vbr02, 9vbr03 e 9vbr04, respectivamente. A análise dos gráficos das figuras correspondentes às envoltórias fBm, mostram uma maior tendência à sobre-estimação quando a variância da série é maior, ao considerar que o  $H$  é praticamente o mesmo para todas as séries.

Pode-se observar na Tabela 5.8 que os aumentos da variância nas séries com prefixo 9vbr são da ordem de 2:1 e esta variação aparece claramente como uma influência para uma melhor estimação da envoltória fBm.

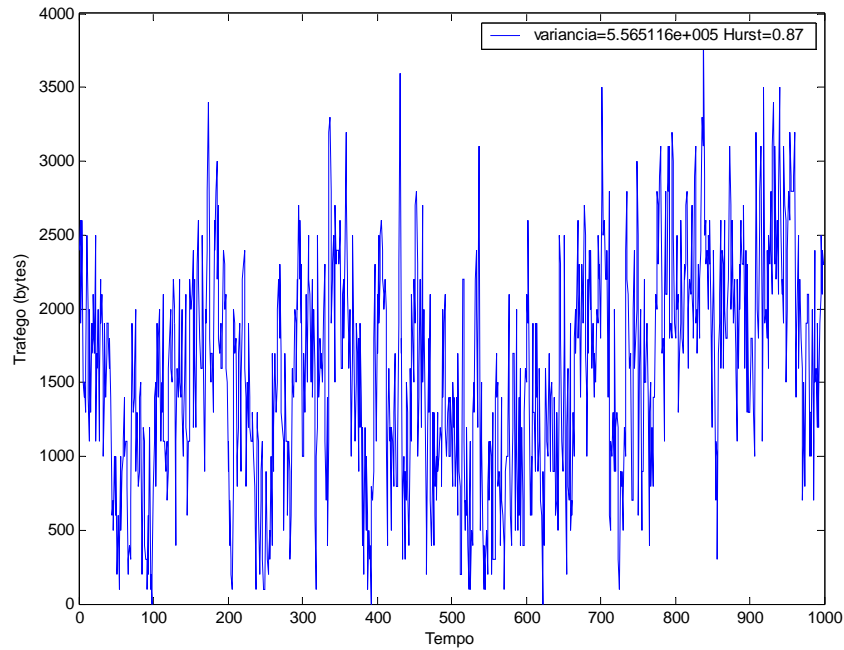


Figura 5.10. Série de tráfego 9vbr02

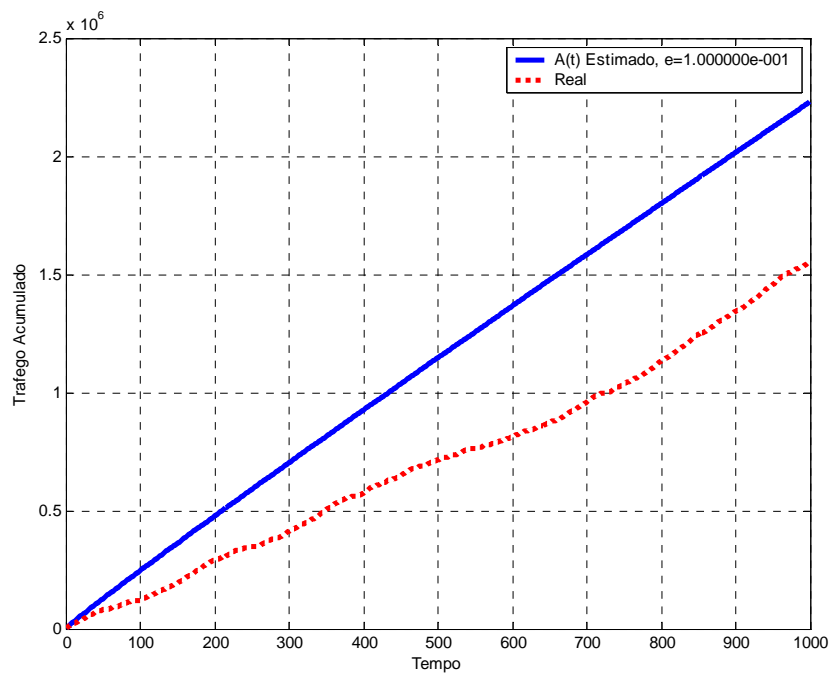


Figura 5.11. Envoltória fBm para a série 9vbr02

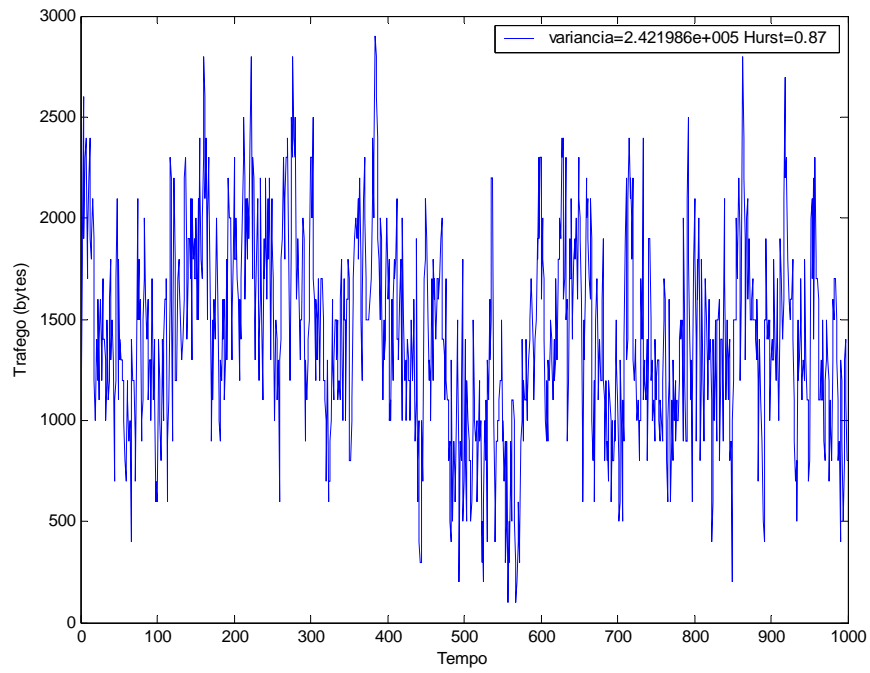


Figura 5.12. Série de tráfego 9vbr03

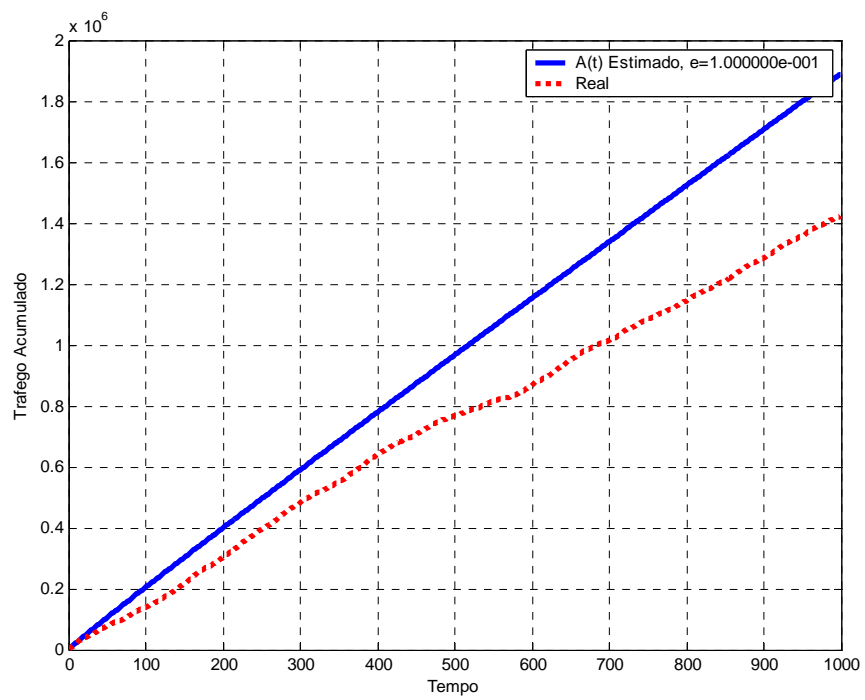


Figura 5.13. Envoltória fBm para a série 9vbr03



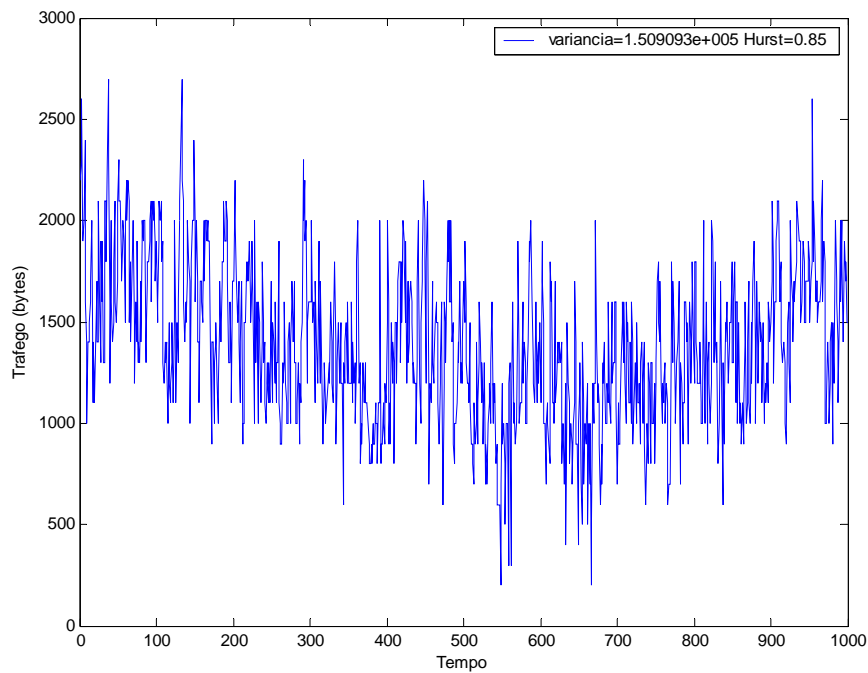


Figura 5.14. Série de tráfego 9vbr04

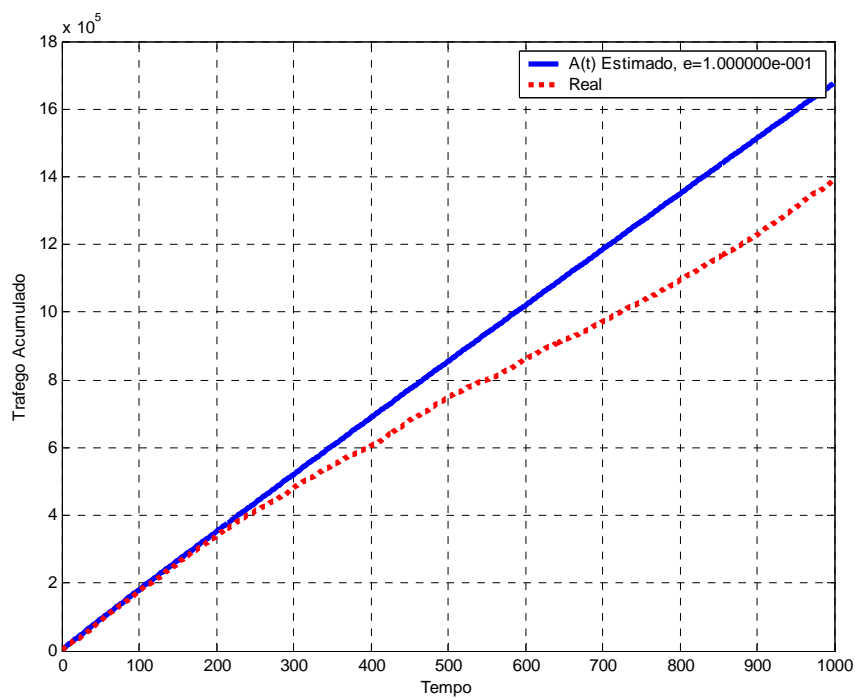


Figura 5.15. Envoltória fBm para a série 9vbr04

Para as séries da Tabela 5.8 com prefixo 8vbr, as figuras 5.16, 5.17, 5.18 e 5.19 mostram as séries e as envoltórias fBm para 8vbr01 e 8vbr02 respectivamente.

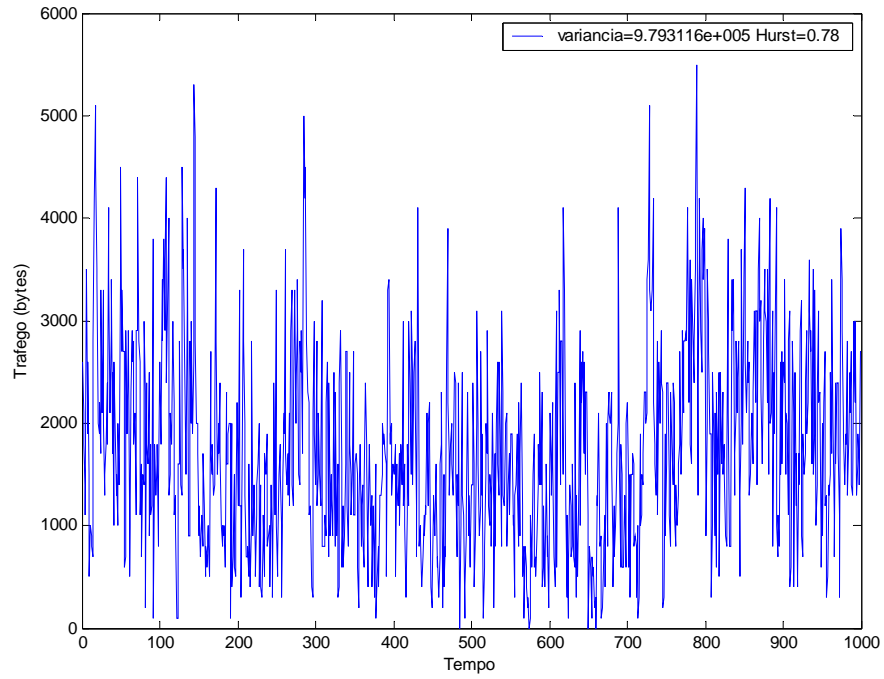


Figura 5.16. Série de tráfego 8vbr01

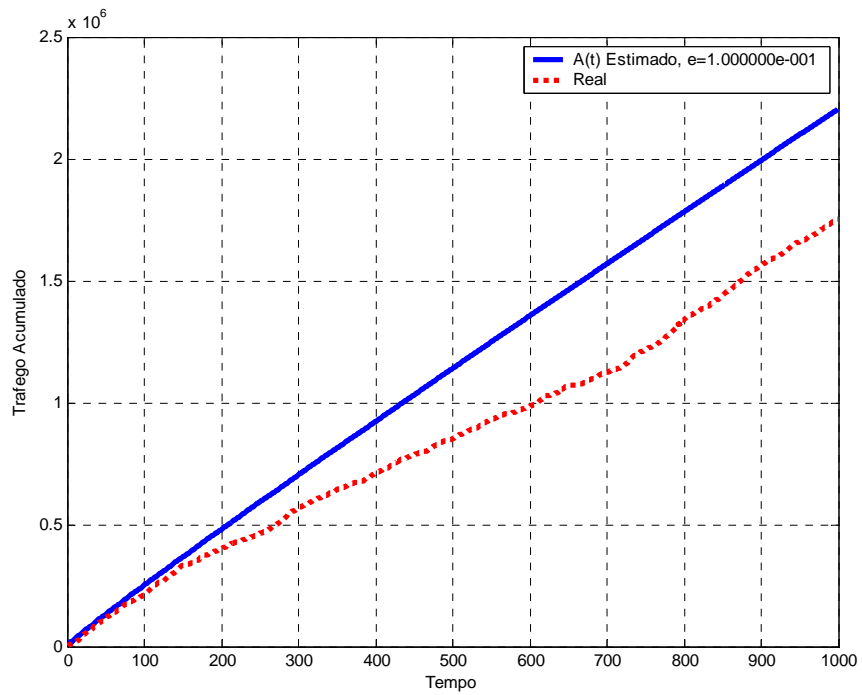


Figura 5.17. Envoltória fBm para a série 8vbr01

Pode-se observar que a envoltória fBm para ambas séries apresenta uma menor sobre-estimação que para as séries do prefixo 9vbr. Isto mostra a tendência de que, para valores menores de  $H$  e  $\sigma^2$ , o impacto da sobre-estimação diminui.

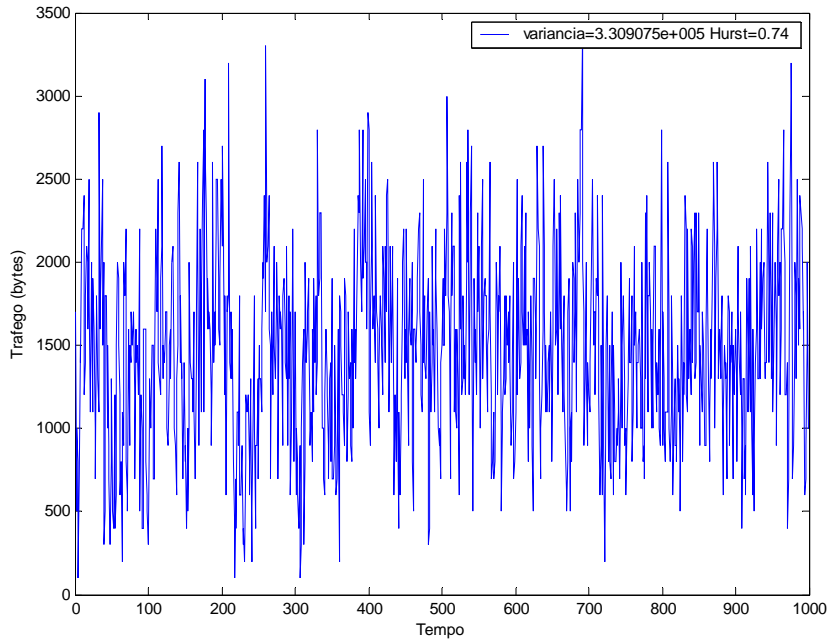


Figura 5.18. Série de tráfego 8vbr02

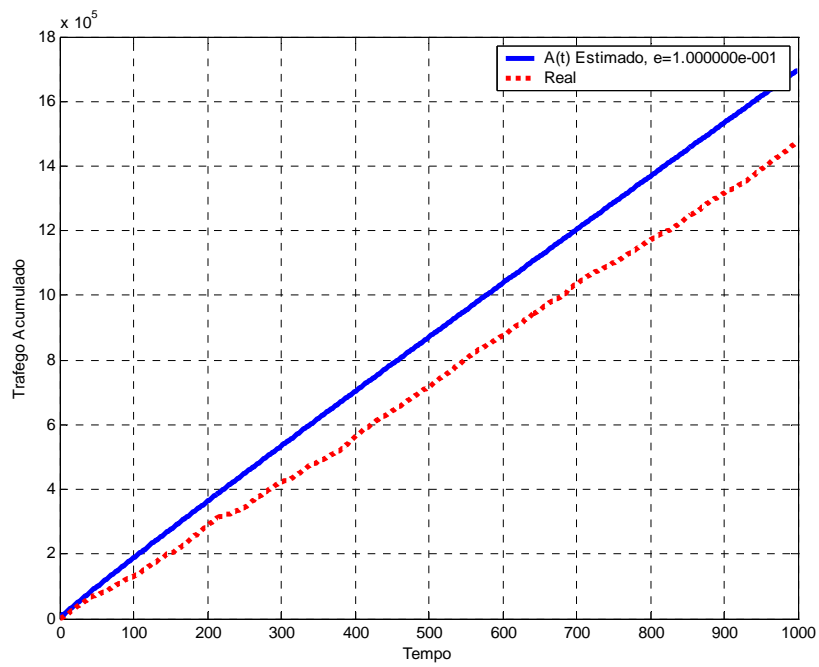


Figura 5.19. Envoltória fBm para a série 8vbr02

As séries 8vbr03 e 8vbr04 são apresentadas nas figuras 5.20 a 5.23. Observa-se claramente nestas figuras que o grau de explosividade do tráfego é menor e que a envoltória fBm para o tráfego agregado é um descritor bastante preciso.

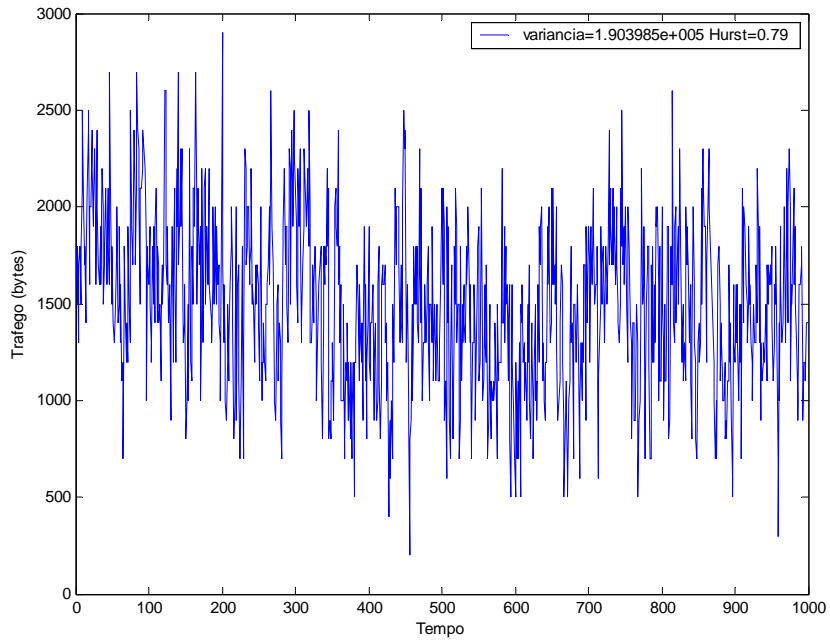


Figura 5.20. Envoltória fBm para a série 8vbr03

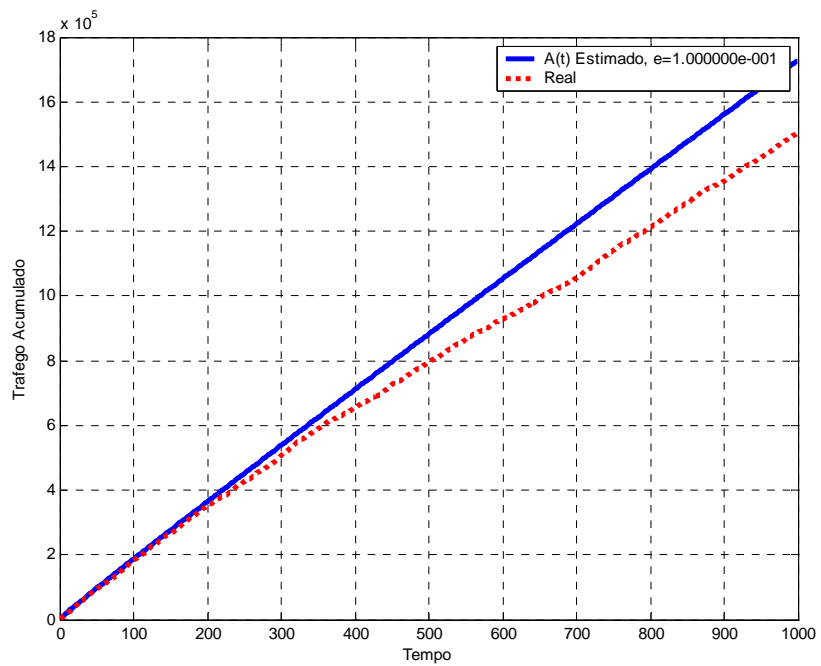


Figura 5.21. Envoltória fBm para a série 8vbr03

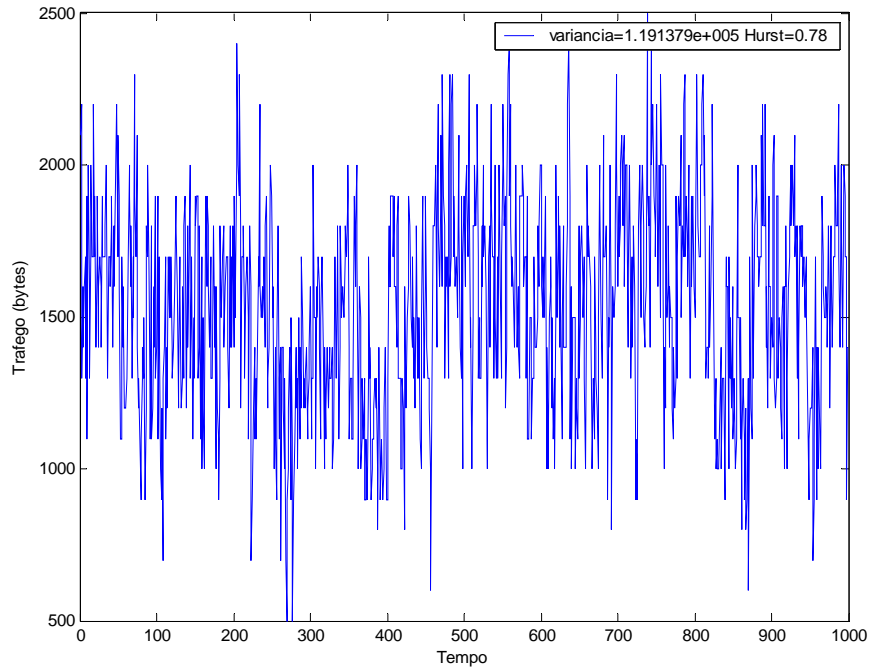


Figura 5.22. Envoltória fBm para a série 8vbr04

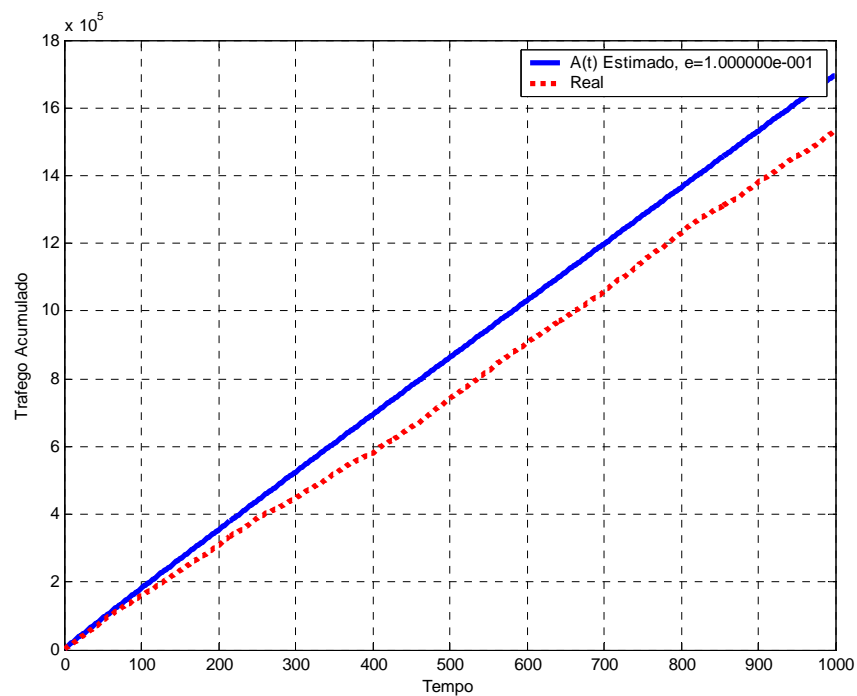


Figura 5.23. Envoltória fBm para a série 8vbr04

Para estender o modelo de tráfego híbrido proposto na seção 5.1.3 para o cálculo do percentual de perdas de pacotes é interessante entender como se relaciona o processo de estimação da envoltória fBm com a carga do enlace e a sua relação com o tamanho da fila em cada nó e a implicação no grau de sobre-estimação sobre os valores das métricas de

QoS. Um alto grau de sobre-estimação prejudicaria a otimização de um processo de planejamento de rede. Por outro lado, ao observar a Eq. 3.46, a relação entre a média da série e a capacidade do enlace pode propiciar valores da máxima escala de tempo particulares a cada configuração resultante para diferentes valores de capacidades de enlaces e tamanhos de filas em cada nó.

Então, o foco da melhoria do modelo estaria em diminuir os percentuais de sobre-estimação das métricas de QoS para tráfegos com variâncias maiores e maior longa dependência, conforme analisado nas séries da Tabela 5.8. É a partir desta análise que se pode pensar em um modelo de tráfego híbrido que, em função dos parâmetros do fluxo agregado no enlace e da carga nesse enlace, permita o cálculo das métricas de QoS diminuindo a sobre-estimação que a envoltória fBm apresenta para valores altos da variância e da longa dependência em relação à capacidade do enlace. Cabe mencionar que valores altos do parâmetro  $H$  e da variância são comuns nas redes operativas. Detalhes destas caracterizações sobre séries reais de tráfego em uma operadora de telecomunicações de grande porte podem ser encontradas no Capítulo 4.

### 5.2.1. O modelo de tráfego híbrido

Para a definição do modelo de tráfego híbrido estendido, foram estabelecidos diferentes níveis de carga nos enlaces da rede hipotética da Fig. 5.3 e foram calculados os valores de atraso e percentual de perda de pacotes. A definição do **modelo de tráfego híbrido estendido** se fundamenta nas seguintes preposições:

1. a caracterização da agregação dos fluxos em cada um dos enlaces se realiza com o método de Kettani e Gubner, mostrado na seção 3.2.5 para o cálculo do parâmetro de Hurst agregado e o método 2 da seção 3.2.5 para o cálculo da variância agregada;
2. para cargas abaixo de 30%, independentemente do valor de  $H$  e  $\sigma^2$  ou quando o parâmetro  $H$  é menor ou igual a 0,5, o cálculo das métricas de QoS se realiza com o modelo M/M/1/K;
3. para cargas entre 30% e 100%, o cálculo das métricas se realiza com base na máxima escala de tempo do fluxo agregado. O atraso se calcula conforme a Eq. 3.47 e o percentual de perdas conforme a Eq. 3.51;

4. o percentual de perdas de pacotes no nó anterior influencia os parâmetros do fluxo entrante no nó seguinte. O valor de  $H$  do fluxo agregado depende da variância agregada. O coeficiente de variância é uma medida da qualidade do tráfego. Os novos valores da variância e de  $H$  são então calculados ao preservar a relação do coeficiente de variância original de fluxo.

Foram avaliados três níveis de carga nos enlaces: 40%, 60% e 80% para cada uma das séries da Tabela 5.8, com filas de tamanho fixo máximo de 20 pacotes para todos os nós. Esses resultados foram confrontados com simulações realizadas para cada configuração de rede no NS-2.

Os resultados das métricas de atraso e percentual de perdas de pacotes para cada uma das séries da Tabela 5.8 serão discutidos na seguinte seção.

### **5.2.2. Análise de Resultados do modelo de tráfego híbrido**

Os resultados para a série 9vbr01, com a maior variância e com  $H > 0,8$ , são mostrados na tabela 5.9. A variável  $\rho$  é a carga em cada um dos enlaces por onde passa o fluxo. Na tabela 5.9, pode-se observar que, para a maior variância nas diferentes cargas do enlace avaliadas, as estimações de atraso e percentual de perdas, na maioria dos casos apresentam um limite superior. Em todos os casos, os valores do atraso representam limites superiores do atraso simulado.

Para o percentual de perdas de pacotes, o limite superior é bastante amplo, como é o dos fluxos F3, F5 e F7 na carga de 40%. Esses fluxos são todos do tipo Poisson. O fluxo F9, também do tipo Poisson, apresenta valores bastantes mais precisos. Ao observar a tabela de roteamento definido para os fluxos, Tabela 5.3, pode-se observar que o fluxo F9 somente é agrupado com o fluxo F5 no enlace 1-2. O agrupamento deste fluxo acarreta, como resultado, um tráfego de Poisson.

Tabela 5.9. Resultados para a série 9vbr01

	$\rho=0.4$			
	Modelo Analítico		Simulação NS	
	Atraso	% perdas	Atraso	% perdas
F1	0,06	0,25	0,04	0,17
F2	0,12	0,41	0,04	0,25
F3	0,06	0,25	0,04	0,0
F4	0,12	0,41	0,05	0,17
F5	0,10	0,25	0,08	0,002
F6	0,07	0,18	0,06	0,17
F7	0,10	0,34	0,11	0,0
F8	0,12	0,26	0,06	0,17
F9	0,05	0,0	0,05	0,0
F10	0,09	0,26	0,05	0,17
	$\rho=0.6$			
	Modelo Analítico		Simulação NS	
	Atraso	% perdas	Atraso	% perdas
F1	0,09	0,22	0,06	0,17
F2	0,16	0,34	0,05	0,32
F3	0,09	0,22	0,06	0,005
F4	0,16	0,34	0,07	0,28
F5	0,16	0,22	0,13	0,008
F6	0,10	0,14	0,09	0,17
F7	0,15	0,29	0,17	0,0
F8	0,18	0,21	0,09	0,17
F9	0,11	0,0001	0,19	0,0
F10	0,14	0,22	0,07	0,17
	$\rho=0.8$			
	Modelo Analítico		Simulação NS	
	Atraso	% perdas	Atraso	% perdas
F1	0,12	0,13	0,10	0,17
F2	0,21	0,22	0,07	0,40
F3	0,12	0,13	0,09	0,03
F4	0,21	0,22	0,08	0,41
F5	0,23	0,13	0,18	0,02
F6	0,13	0,08	0,12	0,18
F7	0,23	0,18	0,37	0,004
F8	0,24	0,12	0,12	0,18
F9	0,30	0,009	0,38	0,0
F10	0,19	0,13	0,10	0,18

No caso dos fluxos F3, F5 e F7, são agrupados em diferentes enlaces com tráfegos auto-similares, como é o caso do F3 nos enlaces 1-4-10-12 e o fluxo F5 no enlace 10-12. Assim também, o fluxo F7 é agrupado com um fluxo auto-similar nos enlaces 1-4 e 1-10.

O agrupamento destes fluxos, com fluxos auto-similares, conforme o método 2 da seção 3.2.5, resulta em variâncias maiores que as individuais de cada fluxo. No caso do parâmetro  $H$ , o valor resultante da agregação de fluxos Poisson e auto-similares tende a dar



valores de  $H$  muito próximos ao maior  $H$  dos fluxos individuais, conforme foi mostrado na Tabela 3.5. Pode-se observar, então, que o impacto da sobre-estimação do percentual de perdas é maior para aqueles fluxos que, ao serem caracterizados individualmente, não apresentam a característica de longa dependência.

Para uma carga no enlace de 0,6, pode-se observar que o mesmo padrão se mantém mas com valores de percentual de perdas e atraso menos sobre-estimados para os fluxos auto-similares. Para uma carga no enlace de 0,8, os valores do atraso para todos os fluxos mostram-se bastante aproximados dos simulados.

O percentual de perdas para o fluxo F7 continua apresentando uma sobre-estimação considerável e para os fluxos F2 e F4 apresenta sub-estimação. Ambos os fluxos, F2 e F4, são auto-similares e são agrupados nos enlaces 1-4 com mais três fluxos: dois auto-similares e um de Poisson.

No enlace 4-10, o fluxo F1 é agrupado com dois fluxos de Poisson. No enlace 9-12, o fluxo F4 é agrupado com um fluxo de Poisson. O agrupamento dos fluxos auto-similares com fluxos de Poisson tende a diminuir o coeficiente de variância do fluxo original.

No modelo analítico, o maior percentual de perdas acontece para os fluxos F2 e F4 no enlace 1-4, diminuindo então o impacto das perdas nos enlaces posteriores, pois a correção da variância dos fluxos auto-similares se faz em função de se preservar o coeficiente de variância.

Na simulação, o percentual de perdas no enlace 1-4 é menor que no modelo analítico, produzindo-se então maiores perdas nos enlaces posteriores, o que induziu à sub-estimação para este caso. Foi observado que pequenas variações do tamanho da fila nos nós no ambiente de simulação produzia forte impacto nos valores do percentual de perdas. Considera-se, então, que a variação do tamanho da fila é um forte elemento de variação das métricas no ambiente de simulação, situação que não necessariamente pode-se generalizar.

A tabela 5.10 apresenta os resultados para a série 9vbr02.

Tabela 5.10. Resultados para a série 9vbr02

	$\rho=0.4$			
	Modelo Analítico		Simulação NS	
	Atraso	% perdas	Atraso	% perdas
F1	0,05	0,06	0,04	0,08
F2	0,09	0,13	0,04	0,13
F3	0,05	0,06	0,02	0,0
F4	0,09	0,13	0,04	0,08
F5	0,08	0,15	0,02	0,0
F6	0,06	0,08	0,06	0,08
F7	0,09	0,09	0,04	0,0
F8	0,10	0,15	0,06	0,08
F9	0,05	0,0	0,05	0,0
F10	0,08	0,15	0,05	0,08
	$\rho=0.6$			
	Modelo Analítico		Simulação NS	
	Atraso	% perdas	Atraso	% perdas
F1	0,09	0,16	0,05	0,08
F2	0,16	0,23	0,05	0,26
F3	0,09	0,16	0,03	0,006
F4	0,16	0,23	0,07	0,21
F5	0,15	0,16	0,04	0,007
F6	0,10	0,11	0,08	0,08
F7	0,15	0,21	0,05	0,0
F8	0,18	0,16	0,08	0,08
F9	0,11	0,0001	0,08	0,0
F10	0,14	0,16	0,07	0,08
	$\rho=0.8$			
	Modelo Analítico		Simulação NS	
	Atraso	% perdas	Atraso	% perdas
F1	0,12	0,12	0,09	0,08
F2	0,21	0,18	0,07	0,34
F3	0,12	0,12	0,04	0,03
F4	0,21	0,18	0,08	0,38
F5	0,22	0,12	0,06	0,02
F6	0,13	0,09	0,11	0,06
F7	0,22	0,16	0,10	0,0
F8	0,24	0,09	0,12	0,09
F9	0,27	0,001	0,13	0,0
F10	0,18	0,11	0,10	0,09

A série analisada na Tabela 5.10 apresenta uma variância menor que a série 9vbr01. Com uma carga nos enlaces de 0,4, pode-se observar que o modelo analítico estabeleceu valores máximos para todos os atrasos que foram confirmados na simulação.

Para o percentual de perdas, os fluxos F3, F5 e F7 apresentam novamente uma maior sobre-estimação mas em um percentual menor que o observado para a mesma carga com a série 9vbr01.

Para uma carga de 0,6, todos os atrasos mostram-se como limites superiores para cada um dos fluxos. No caso do percentual de perdas, não existe nenhum caso de subestimação, à exceção do F2, mas vale observar que com um percentual muito pequeno.

Para uma carga de 0,8, os atrasos novamente estabelecem limites superiores para os simulados. Para o percentual de perdas de pacotes os fluxos F2 e F4 apresentam subestimação. O grau de sobre-estimação se mantém aproximadamente no mesmo padrão que para as outras cargas analisadas anteriormente.

Ao comparar os resultados da tabela 5.9 e 5.10, pode-se observar que se manteve o mesmo padrão de cálculo das métricas de QoS para todos os fluxos. Para cada série os fluxos que apresentaram uma sub-estimação foram os mesmos e se induz que existe uma tendência maior de sobre-estimação para aqueles fluxos Poisson ao serem agrupados com os auto-similares, conforme foi analisado anteriormente.

Uma limitação evidente nesta análise é que o tamanho das filas se manteve constante para todos os nós, o que não é necessariamente o que ocorre em um ambiente de operação real.

A Tabela 5.11 mostra os resultados para a série 9vbr03.

Pode-se observar que para uma carga de 0,4 existe uma maior tendência a sub-estimar os valores do percentual de perdas de pacotes. Os atrasos, por sua vez, apresentam um grau de sobre-estimação bem menor que as séries anteriores.

Para uma carga de 0,6, os valores do atraso e do percentual de perdas mostram-se mais próximos. Pode-se comparar com as tabelas anteriores que, da mesma forma, para uma carga de 0,6, os valores obtidos pelo modelo analítico encontram-se mais próximos daqueles obtidos pela simulação. Para uma carga de 0,8, novamente os fluxos F2 e F4 apresentam uma sub-estimação para o percentual de perdas. Para os outros fluxos o grau de sobre-estimação do percentual de perdas é bastante aceitável. No caso dos atrasos pode-se observar uma maior tendência a sobre-estimar em relação à carga de 0,6.

Tabela 5.11. Resultados para a série 9vbr03

	$\rho=0.4$			
	Modelo Analítico		Simulação NS	
	Atraso	% perdas	Atraso	% perdas
F1	0,02	0,0	0,04	0,02
F2	0,06	0,0	0,04	0,06
F3	0,02	0,0	0,01	0,0
F4	0,06	0,0	0,02	0,03
F5	0,08	0,0	0,03	0,0
F6	0,06	0,06	0,05	0,02
F7	0,04	0,0	0,04	0,0
F8	0,11	0,12	0,06	0,02
F9	0,05	0,0	0,05	0,0
F10	0,07	0,0	0,04	0,02
	$\rho=0.6$			
	Modelo Analítico		Simulação NS	
	Atraso	% perdas	Atraso	% perdas
F1	0,08	0,04	0,06	0,02
F2	0,15	0,10	0,06	0,15
F3	0,08	0,04	0,03	0,003
F4	0,15	0,10	0,07	0,13
F5	0,14	0,11	0,04	0,004
F6	0,10	0,05	0,09	0,02
F7	0,16	0,08	0,05	0,0
F8	0,18	0,10	0,09	0,02
F9	0,11	0,0	0,08	0,0
F10	0,13	0,12	0,07	0,02
	$\rho=0.8$			
	Modelo Analítico		Simulação NS	
	Atraso	% perdas	Atraso	% perdas
F1	0,13	0,12	0,11	0,02
F2	0,22	0,16	0,07	0,27
F3	0,13	0,12	0,05	0,03
F4	0,22	0,16	0,08	0,36
F5	0,24	0,12	0,07	0,08
F6	0,14	0,07	0,12	0,04
F7	0,24	0,16	0,10	0,004
F8	0,26	0,09	0,13	0,04
F9	0,27	0,009	0,13	0,0
F10	0,20	0,12	0,10	0,03

A tabela 5.12 mostra os resultados para a série 9vbr04.

Tabela 5.12. Resultados para a série 9vbr04

	$\rho=0.4$			
	Modelo Analítico		Simulação NS	
	Atraso	% perdas	Atraso	% perdas
F1	0,01	0,0	0,03	0,007
F2	0,01	0,0	0,04	0,03
F3	0,01	0,0	0,02	0,0
F4	0,01	0,0	0,05	0,007
F5	0,03	0,0	0,03	0,0
F6	0,04	0,0	0,06	0,006
F7	0,03	0,0	0,04	0,0
F8	0,01	0,0	0,06	0,006
F9	0,05	0,0	0,06	0,0
F10	0,02	0,0	0,05	0,006
	$\rho=0.6$			
	Modelo Analítico		Simulação NS	
	Atraso	% perdas	Atraso	% perdas
F1	0,03	0,0	0,06	0,007
F2	0,11	0,0	0,06	0,12
F3	0,03	0,0	0,06	0,004
F4	0,11	0,0	0,07	0,14
F5	0,13	0,0	0,13	0,002
F6	0,09	0,05	0,09	0,006
F7	0,09	0,0	0,16	0,0
F8	0,18	0,09	0,10	0,007
F9	0,11	0,0	0,08	0,0
F10	0,03	0,0	0,07	0,006
	$\rho=0.8$			
	Modelo Analítico		Simulação NS	
	Atraso	% perdas	Atraso	% perdas
F1	0,12	0,05	0,10	0,007
F2	0,21	0,09	0,07	0,20
F3	0,12	0,05	0,05	0,03
F4	0,21	0,09	0,09	0,35
F5	0,22	0,09	0,07	0,003
F6	0,14	0,03	0,12	0,02
F7	0,24	0,08	0,11	0,0
F8	0,26	0,07	0,13	0,02
F9	0,27	0,008	0,14	0,0
F10	0,19	0,09	0,10	0,01

Pode-se observar que para a menor carga nos enlaces, os atrasos são levemente subestimados. No caso do percentual de perdas também existe essa tendência.

A mesma tendência se observa para as demais cargas, com exceção dos atrasos para a carga de 0,8 em que todos os valores dos atrasos calculados com o modelo analítico se apresentam como limites superiores dos atrasos simulados. A tendência de subestimação do percentual de perdas também diminui ao aumentar a carga do enlace.

A Tabela 5.13 mostra os resultados para a série 8vbr01.

Tabela 5.13. Resultados para a série 8vbr01

	$\rho=0.4$			
	Modelo Analítico		Simulação NS	
	Atraso	Perdas	Atraso	Perdas
F1	0,05	0,15	0,04	0,17
F2	0,09	0,26	0,03	0,22
F3	0,05	0,15	0,01	0,0
F4	0,09	0,26	0,04	0,17
F5	0,09	0,22	0,02	0,0
F6	0,06	0,14	0,05	0,17
F7	0,09	0,21	0,04	0,0
F8	0,14	0,35	0,05	0,17
F9	0,05	0,0	0,06	0,0
F10	0,08	0,22	0,04	0,16
	$\rho=0.6$			
	Modelo Analítico		Simulação NS	
	Atraso	Perdas	Atraso	Perdas
F1	0,08	0,23	0,05	0,17
F2	0,14	0,32	0,05	0,32
F3	0,08	0,23	0,03	0,002
F4	0,14	0,32	0,06	0,24
F5	0,14	0,24	0,04	0,008
F6	0,09	0,18	0,07	0,17
F7	0,14	0,29	0,06	0,0
F8	0,16	0,26	0,08	0,17
F9	0,11	0,0001	0,08	0,0
F10	0,13	0,24	0,06	0,17
	$\rho=0.8$			
	Modelo Analítico		Simulação NS	
	Atraso	Perdas	Atraso	Perdas
F1	0,11	0,19	0,08	0,17
F2	0,18	0,30	0,06	0,41
F3	0,11	0,19	0,04	0,02
F4	0,18	0,30	0,07	0,40
F5	0,21	0,18	0,06	0,02
F6	0,12	0,12	0,10	0,17
F7	0,21	0,25	0,10	0,0
F8	0,21	0,18	0,10	0,17
F9	0,27	0,008	0,14	0,0
F10	0,17	0,19	0,08	0,17

Pode-se observar que, para a carga de 0,4, os atrasos calculados para todos os fluxos correspondem a limites superiores para os atrasos simulados. O percentual de perdas de pacotes também parece ser, na maioria dos fluxos, um limite superior.

Novamente, como aconteceu com a série 9vbr01, a sobre-estimação para os fluxos do tipo Poisson é maior que para os fluxos auto-similares.

Ao examinar na Tabela 5.13, as cargas de 0,6 e 0,8, pode-se observar que existe um comportamento similar ao observado para a série 9vbr01, mas com uma maior sobre-estimação do modelo analítico para o percentual de perdas de pacotes.

Tabela 5.14. Resultados para a série 8vbr02

	$\rho=0.4$			
	Modelo Analítico		Simulação NS	
	Atraso	Perdas	Atraso	Perdas
F1	0,03	0,0	0,02	0,04
F2	0,09	0,01	0,04	0,0
F3	0,03	0,0	0,01	0,0
F4	0,09	0,01	0,04	0,04
F5	0,10	0,01	0,03	0,0
F6	0,06	0,09	0,05	0,04
F7	0,06	0,0	0,04	0,0
F8	0,11	0,13	0,06	0,04
F9	0,05	0,0	0,05	0,0
F10	0,09	0,01	0,04	0,04
	$\rho=0.6$			
	Modelo Analítico		Simulação NS	
	Atraso	Perdas	Atraso	Perdas
F1	0,09	0,06	0,06	0,04
F2	0,15	0,14	0,06	0,20
F3	0,09	0,06	0,03	0,01
F4	0,15	0,14	0,07	0,15
F5	0,15	0,13	0,04	0,005
F6	0,10	0,10	0,09	0,04
F7	0,16	0,09	0,06	0,0
F8	0,18	0,18	0,09	0,04
F9	0,11	0,0001	0,08	0,0
F10	0,14	0,13	0,07	0,04
	$\rho=0.8$			
	Modelo Analítico		Simulação NS	
	Atraso	Perdas	Atraso	Perdas
F1	0,12	0,14	0,09	0,04
F2	0,21	0,20	0,07	0,32
F3	0,12	0,14	0,05	0,03
F4	0,22	0,20	0,08	0,34
F5	0,23	0,14	0,07	0,01
F6	0,14	0,11	0,12	0,05
F7	0,23	0,17	0,10	0,0
F8	0,25	0,15	0,12	0,05
F9	0,27	0,0	0,14	0,0
F10	0,19	0,14	0,10	0,05

Na tabela 5.14, pode-se observar o resultado para a série 8vbr02. Novamente, apresenta-se o mesmo padrão de comportamento que para a série 9vbr02, como por exemplo, a sub-estimação das perdas para os fluxos F2 e F4, no caso da maior carga no enlace.

Pode-se observar, também, que no caso da carga 0,6, a sobre-estimação do percentual de perda de pacotes para todos os fluxos apresenta um menor grau que para a série 9vbr02.

As Tabelas 5.15 e 5.16 mostram os resultados para as séries 8vbr03 e 8vbr04.

Tabela 5.15. Resultados para a série 8vbr03

	$\rho=0.4$			
	Modelo Analítico		Simulação NS	
	Atraso	Perdas	Atraso	Perdas
F1	0,01	0,0	0,04	0,01
F2	0,02	0,0	0,04	0,07
F3	0,01	0,0	0,01	0,0
F4	0,02	0,0	0,04	0,02
F5	0,04	0,0	0,03	0,0
F6	0,03	0,0	0,04	0,0
F7	0,02	0,0	0,04	0,0
F8	0,07	0,0	0,06	0,02
F9	0,05	0,0	0,05	0,0
F10	0,03	0,0	0,02	0,02
	$\rho=0.6$			
	Modelo Analítico		Simulação NS	
	Atraso	Perdas	Atraso	Perdas
F1	0,03	0,0	0,05	0,02
F2	0,12	0,0	0,05	0,18
F3	0,03	0,0	0,03	0,005
F4	0,12	0,0	0,07	0,15
F5	0,14	0,0	0,04	0,003
F6	0,09	0,06	0,09	0,02
F7	0,10	0,0	0,06	0,0
F8	0,17	0,09	0,09	0,02
F9	0,11	0,0001	0,08	0,0
F10	0,13	0,0	0,07	0,002
	$\rho=0.8$			
	Modelo Analítico		Simulação NS	
	Atraso	Perdas	Atraso	Perdas
F1	0,12	0,06	0,09	0,02
F2	0,20	0,11	0,05	0,18
F3	0,12	0,06	0,03	0,009
F4	0,20	0,11	0,07	0,14
F5	0,22	0,09	0,04	0,004
F6	0,14	0,04	0,09	0,006
F7	0,23	0,08	0,06	0,0
F8	0,23	0,09	0,09	0,006
F9	0,27	0,008	0,08	0,0
F10	0,18	0,09	0,07	0,007



Pode-se observar que para as cargas de 0,4 e 0,6 existe uma forte tendência a sub-estimar os valores do atraso e do percentual de perdas de pacotes.

Para a série 8vbr03, com uma carga de 0,8, pode-se considerar que o cálculo das métricas não apresenta um alto grau de sobre-estimação. Os fluxos F2 e F4 tiveram um percentual de perda de pacotes sub-estimado em um grau bastante reduzido em comparação com os cálculos para as séries com maior variância e maior longa dependência.

Tabela 5.16. Resultados para a série 8vbr04

	$\rho=0.4$			
	Modelo Analítico		Simulação NS	
	Atraso	Perdas	Atraso	Perdas
F1	0,01	0,0	0,02	0,07
F2	0,005	0,0	0,04	0,06
F3	0,01	0,0	0,02	0,0
F4	0,003	0,0	0,04	0,007
F5	0,02	0,0	0,03	0,0
F6	0,002	0,0	0,06	0,007
F7	0,03	0,0	0,04	0,0
F8	0,005	0,0	0,06	0,007
F9	0,05	0,0	0,05	0,0
F10	0,02	0,0	0,05	0,007
	$\rho=0.6$			
	Modelo Analítico		Simulação NS	
	Atraso	Perdas	Atraso	Perdas
F1	0,03	0,0	0,06	0,007
F2	0,04	0,0	0,05	0,18
F3	0,03	0,0	0,03	0,0
F4	0,04	0,0	0,07	0,15
F5	0,07	0,0	0,04	0,002
F6	0,06	0,0	0,09	0,007
F7	0,07	0,0	0,06	0,0
F8	0,11	0,0	0,09	0,007
F9	0,22	0,0	0,08	0,0
F10	0,06	0,0	0,07	0,007
	$\rho=0.8$			
	Modelo Analítico		Simulação NS	
	Atraso	Perdas	Atraso	Perdas
F1	0,11	0,0	0,06	0,007
F2	0,17	0,04	0,05	0,19
F3	0,11	0,0	0,03	0,006
F4	0,17	0,04	0,07	0,15
F5	0,21	0,03	0,04	0,007
F6	0,12	0,04	0,09	0,007
F7	0,23	0,01	0,06	0,0
F8	0,22	0,07	0,09	0,007
F9	0,27	0,01	0,08	0,0
F10	0,17	0,03	0,07	0,007

Para a série 8vbr04 pode-se observar que o cálculo das métricas para a carga de 0,8 apresenta valores aceitáveis para o atraso mas sub-estima a perda de pacotes em quase todos os casos.

Os resultados mostrados em todas as tabelas anteriores mostram que existe uma maior dificuldade em aproximar o percentual de perdas de pacotes.

Observa-se também que existe uma tendência a sobre-estimar o percentual de perda de pacotes para fluxos de Poisson que são agregados com fluxos auto-similares. Assim, também, a associação dos fluxos auto-similares com os de Poisson, ao reduzir o coeficiente de variância do fluxo agregado, pode induzir a uma sub-estimação do percentual de perdas de pacotes para cargas no enlace acima de 0,6.

O modelo apresenta imprecisão maior para variâncias menores e parâmetro  $H$  menor que 0,8. Os melhores resultados do modelo se observam para cargas no enlace de 0,6 e altas variâncias.

A comparação dos resultados analíticos com simulações no NS-2 é um ponto inicial de validação do modelo, sujeito em certos casos a limitações na simulação, como é o caso da variação do tamanho da fila e o seu impacto no cálculo do percentual de perdas de pacotes.

Um outro ponto de interesse é a influência do tamanho da fila na variação das métricas. Essa avaliação pode ser realizada ao variar o tamanho das filas e as capacidades nos enlaces e observar o comportamento das métricas para cada nova topologia.

### **5.3. RESUMO DO CAPÍTULO**

Neste capítulo foi proposto um modelo de tráfego híbrido para a caracterização do tráfego seja Poisson ou auto-similar, que considera a carga em cada um dos enlaces parciais da rede para estabelecer um conjunto de métricas de QoS. Esses valores das métricas representam valores máximos que não podem ser excedidos pelos fluxos de tráfego.

Para efetuar a validação do modelo de tráfego proposto, foram apresentados exemplos, primeiro utilizando uma única métrica, o atraso, em dois cenários de rede que foram comparados com simulações no NS-2. Os resultados mostraram que o modelo  $M/M/1/k$  e o modelo de envoltória  $fBm$  podem-se complementar para formular um modelo de tráfego que permita a obtenção de métricas de QoS mais precisas.

Posteriormente, foi proposto o modelo de tráfego híbrido com base nos modelos M/M/1/k e envoltória fBm e foi avaliado para vários tipos de séries de tráfego sintéticas, diferenciadas pelos valores do parâmetro  $H$  e variância. Para um tamanho de fila fixo, foram calculados os valores de atraso e perda para o mesmo cenário, uma rede com 12 nós e 10 fluxos de tráfego.

A análise realizada confirmou que, o modelo de tráfego híbrido efetua um cálculo dos parâmetros de QoS mais aproximado daqueles observados com os outros modelos analisados, o fBm e o M/M/1/k. Observou-se que o cálculo das métricas de QoS apresenta uma tendência a sub-estimar valores de perdas de pacotes para altas variâncias, resultantes da agregação de vários fluxos auto-similares. Essa sobre-estimação apresenta um maior grau quando a carga do enlace é maior que 60% e a variância é alta. Para cargas do enlace entre 40% e 60% e variâncias moderadas, como as observadas em séries de tráfego real, o modelo de tráfego híbrido obteve na maioria dos casos, valores de métricas de QoS que representam cotas superiores, para as métricas de QoS reais observadas nas simulações. Assim também, existe um menor grau de sobre-estimação para o cálculo do atraso que para a probabilidade de perdas.

A análise do modelo foi realizada considerando-se uma fila de tamanho fixo. No capítulo seguinte considera-se a variação do tamanho da fila e da capacidade do enlace como parte de um processo de otimização de elementos da rede para avaliar a utilidade do modelo de tráfego híbrido em um processo de otimização de rede.

## **6 - PLANEJAMENTO DE REDE: OTIMIZAÇÃO DE PARÂMETROS DE UMA REDE MULTIMÍDIA**

Neste capítulo é apresentada uma metodologia que com o uso do modelo de tráfego híbrido apresentado no Capítulo 5 e de um algoritmo genético como ferramenta de otimização, permite a planificação de uma rede para cumprir métricas específicas de desempenho a um custo mínimo.

Assim também é apresentado um modelo de engenharia de tráfego com base em um processo de caracterização de tráfego. É apresentado um algoritmo para realizar este processo. O algoritmo é avaliado com o objetivo de se verificar a possibilidade de que, dado um novo fluxo entrante em uma rede operacional, possa-se calcular o melhor caminho para mapear este fluxo na rede, com o objetivo de preservar as métricas de QoS, o que promove a estabilidade operativa da rede e minimiza a modificação constante dos seus elementos.

### **6.1. O PROBLEMA DA ALOCAÇÃO DAS CAPACIDADES**

Na Tabela 6.1, é apresentada a formulação do problema de alocação das capacidades em uma rede.

Como pode ser observada na Tabela 6.1, a função de custo a ser otimizada pode ser uma função linear ou uma função côncava. Essa função de custo, por exemplo, pode representar o valor de implantação da rede em função da capacidade dos enlaces. Se o tráfego é de Poisson e se a função é linear, a solução ótima pode ser obtida a partir da utilização do método de multiplicadores de Lagrange (Bertsekas D., Gallager R., 1992).

Tabela 6.1. Formulação do problema da alocação das capacidades em uma rede

dado	<ul style="list-style-type: none"> <li>▪ uma topologia</li> <li>▪ uma matriz de tráfego R</li> <li>▪ as políticas de roteamento (ou seja, os caminhos dos fluxos <math>f_1, f_2, \dots, f_m</math> na rede)</li> </ul>
minimizar	<ul style="list-style-type: none"> <li>▪ uma função de custo</li> </ul>
ao considerar as variáveis	<ul style="list-style-type: none"> <li>▪ <math>C=(C_1, C_2, \dots, C_b)</math>, em que <math>b</math>=número de enlaces</li> </ul>
sujeito a	<ul style="list-style-type: none"> <li>▪ <math>f \leq C</math>: os fluxos são menores que as capacidades</li> <li>▪ uma ou mais restrições de desempenho</li> </ul>

## 6.2. OS PROBLEMAS DA DEFINIÇÃO DO ROTEAMENTO

O problema da definição do roteamento, no enfoque mais simples, consiste em encontrar uma política de roteamento que defina rotas a serem seguidas pelos fluxos de tal forma que se minimize uma variável de desempenho. A Tabela 6.2 mostra a formulação do problema.

Tabela 6.2. Formulação do problema da alocação dos fluxos

dados	<ul style="list-style-type: none"> <li>▪ uma topologia</li> <li>▪ uma matriz de tráfego R</li> <li>▪ as capacidades dos enlaces <math>C_i</math></li> </ul>
minimizar	<ul style="list-style-type: none"> <li>▪ uma função de desempenho</li> </ul>
ao considerar as variáveis	<ul style="list-style-type: none"> <li>▪ <math>f=(f_1, f_2, \dots, f_m)</math></li> </ul>
sujeito a	<ul style="list-style-type: none"> <li>▪ <math>f</math> é um MC (<i>multicommodity flow</i>) que satisfaz a matriz de tráfego R</li> <li>▪ <math>f \leq C</math></li> </ul>

A aplicação direta de técnicas de otimização comuns na literatura para este problema de roteamento resulta em redes com baixa conectividade (poucos enlaces com alta carga) e pode violar restrições de confiabilidade (Bertsekas e Gallager, 1992, Mitra D., e Ramakrishnan K.G, 2001). Com técnicas heurísticas que produzem resultados sub-ótimos, melhores desempenhos podem ser obtidos (Gerla, M. e Kleinrock L, 1977).

O problema de estabelecimento das capacidades e dos fluxos é formulado na Tabela 6.3.

Tabela 6.3. Formulação do problema da alocação das capacidades e do roteamento

dados	<ul style="list-style-type: none"> <li>▪ uma topologia</li> <li>▪ uma matriz de tráfego R</li> <li>▪ as funções de custo da capacidade <math>d_i(C_i)</math></li> </ul>
minimizar	<ul style="list-style-type: none"> <li>▪ uma função de custo</li> </ul>
ao considerar as variáveis	<ul style="list-style-type: none"> <li>▪ <math>f=(f_1, f_2, \dots, f_b)</math> e <math>C=(C_1, C_2, \dots, C_b)</math></li> </ul>
sujeito a	<ul style="list-style-type: none"> <li>▪ <math>f</math> é um MC que satisfaz a matriz de tráfego R</li> <li>▪ <math>f \leq C</math></li> <li>▪ Um ou mais requerimentos de desempenho (atraso, perda de pacotes, <i>jitter</i>).</li> </ul>

O problema apresentado na Tabela 6.3 exige uma otimização simultânea das capacidades e dos fluxos. A existência de um grande número de mínimos locais faz com que uma solução exata seja difícil de se obter.

### 6.3. ALGORITMO GENÉTICO COMO FERRAMENTA DE OTIMIZAÇÃO

Os algoritmos genéticos (AG), vem sendo principalmente utilizados como métodos de pesquisa e otimização (Buriol L.S, Resende M.G.C, Ribeiro C.C, Thorup M. 2002, Anton Riedl, 1998). Dado um conjunto grande de possíveis soluções, deseja-se escolher o ponto que otimiza uma função objetivo devendo-se cumprir um conjunto de restrições. No caso do problema mostrado na Tabela 6.1, pode-se identificar a função de custo como a função objetivo e o atraso máximo como o conjunto de restrições.

Os AG se fundamentam na idéia da seleção natural. Considera-se que a tendência para que um indivíduo sobreviva em um ambiente depende da sua estrutura genética e que, em várias gerações de indivíduos, somente os bons genes prevaleçam e os maus genes sejam descartados. Assim, também, ao juntar indivíduos com boas características e reproduzi-los, os filhos destes indivíduos podem ser ainda mais aptos a sobreviver, pois concentram a combinação de bons genes.

Os AG aplicam este princípio a problemas de otimização ao representar possíveis alternativas de solução através de conjuntos adequados de genes e de operações de seleção natural destes genes. Inicialmente, de forma aleatória, é formada uma geração de partida ou inicial.

Logo em seguida, são aplicados três processos sobre essa geração inicial: reprodução, cruzamento e mutação, em que são formadas novas gerações até alcançar algum critério específico de interrupção do processo. Em cada iteração, sobre uma geração, os indivíduos são transformados e sua adequação é avaliada. A figura 6.1 ilustra o procedimento geral do AG.

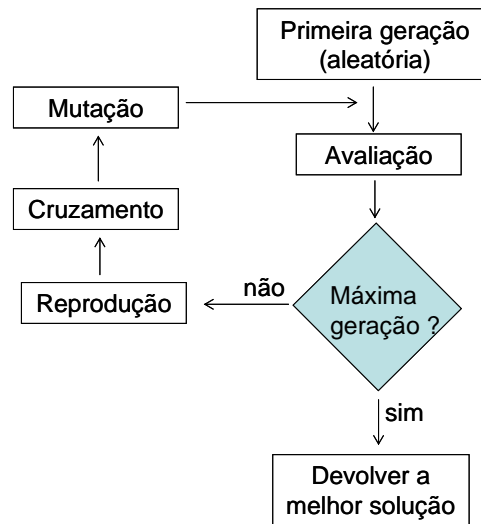


Figura 6.1. Diagrama de fluxo do funcionamento de um AG típico

No processo de reprodução, é criada uma nova geração. A partir de uma geração existente, seqüências de genes são reproduzidas com probabilidade proporcional à qualidade da solução correspondente. As seqüências que representam soluções com boas características têm uma maior probabilidade de sobrevivência que aquelas seqüências com características não desejáveis.

A operação de cruzamento seleciona pares de seqüências, quebra sua seqüência genética em posições aleatórias e promove a troca de informação genética. Finalmente, o processo de mutação introduz novo material genético ao selecionar aleatoriamente a mudança em genes individuais.

A mutação é considerada um processo importante para, parcialmente, deslocar o processo de busca a novos sub-espacos dentro do espaco-solução, pois no caso contrário, o processo de busca convergiria a um ponto local mínimo sem considerar outros sub-espacos.

Na área de redes de telecomunicações, existem alguns trabalhos que propõem o uso do AG como ferramenta para implementar heurísticas que ajudem no processo de planejamento de

redes e sua otimização (Riedl A., 1998),(Riedl A., 2002), (Buriol L.S, Resende M.G.C., Ribeiro C.C., Thorup M , 2002).

Em (Riedl A., 1998), é utilizado AG para planejamento de redes ópticas passivas na área de acesso. Neste mesmo trabalho, o AG é utilizado para planejamento integrado de redes de pacotes, que consiste em considerar os aspectos de topologia e otimização de roteamento, de tal forma que o atraso fim-a-fim da rede não exceda um valor específico. Nesse trabalho, o AG considera uma topologia pré-estabelecida, em uma rede com 10 nós e caminhos definidos a partir de protocolos de roteamento do caminho mais curto, as capacidades ótimas e o custo mínimo são calculados, sob a consideração de um processo de chegada de tráfego de Poisson e roteadores modelados como um sistema de filas M/M/1.

Em (Riedl, A., 2002), um AG é utilizado para otimização de roteamento no contexto da Engenharia de Tráfego, com o uso de protocolos de roteamento de caminho mais curto e suas métricas. O trabalho é focado em tecnologias de roteamento que utilizam vários tipos de métricas, tais como atraso e largura de banda, com o objetivo de derivar caminhos mais curtos para cada destino na rede, mas não são mostrados resultados experimentais com diferentes tipos de tráfego e suas medições representadas em métricas de QoS.

Em (Buriol L.S., Resende M.G.C., Ribeiro C.C., Thorup M., 2002), um AG é utilizado para um melhoramento local do problema de alocação de pesos nos enlaces no protocolo OSPF (*Open Shortest Path First*). Os resultados do AG são comparados com outros algoritmos, em que o primeiro se mostra mais eficiente para o processo, assim como mais flexível diante de várias topologias de rede.

Então, pode-se concluir que o uso do AG no problema de otimização de redes é uma alternativa interessante. Entretanto, os seus resultados dependem de funções objetivo adequadas ao problema. E, no caso deste trabalho, a caracterização do tráfego tem a função de fornecer ao AG o conjunto de restrições dentro das quais se deve encontrar o conjunto solução.

Para ilustrar a importância do processo de caracterização do tráfego, nas seguintes seções são mostradas aplicações do AG ao problema de dimensionamento de redes como mostrado na Tabela 6.1.



#### 6.4. OTIMIZAÇÃO DE UMA REDE PARA MÉTRICAS DE PERDAS E ATRASO A UM CUSTO MINIMO

O objetivo desta seção é observar a aplicabilidade do modelo de tráfego híbrido proposto e avaliado no capítulo 5 em conjunto com um algoritmo genético para a otimização das capacidades dos enlaces e tamanhos das filas dos nós em uma rede hipotética.

Para a avaliação, é utilizada a rede de 10 nós mostrada na Figura 5.2 com as rotas definidas conforme a Tabela 5.3. Os fluxos utilizados para a avaliação são definidos da seguinte forma: os fluxos F3, F5, F7 e F9 são do tipo Poisson, gerados com  $H=0,5$ , taxa média de 39 kbps e variância de  $4,998377 \times 10^4$  na escala de 0,1 segundos. Os fluxos F1, F2, F4, F6, F8 e F10 são auto-similares, conforme os parâmetros da série 9vbr01, descrita na Tabela 5.8.

Para o processo de otimização, são definidos para cada um dos fluxos valores máximos admissíveis de atraso e perdas de pacotes. Dessa forma, o processo de otimização implementado por meio de um algoritmo genético utilizará um conjunto de indivíduos representados pelas capacidades e tamanhos das filas e, conforme o processo mostrado na Fig. 6.1, avaliará diferentes topologias com a utilização do modelo de tráfego híbrido, com o objetivo de encontrar aquela topologia que, com um menor custo, produza métricas de atraso e percentual de perda de pacotes menores que aqueles estabelecidos como limites superiores para cada fluxo. Esta proposta é uma extensão à proposta apresentada na Tabela 6.1, pois para este caso, também serão obtidos valores limites para os tamanhos das filas.

O elemento inovador da proposta do modelo de tráfego híbrido em conjunto com a otimização, está no cálculo de dois elementos da topologia da rede, capacidades e tamanhos máximos das filas, em função de duas métricas simultâneas, atraso e percentual de perda de pacotes.

O custo da rede, função a ser otimizada para se obter as métricas de desempenho estabelecidas, pode ser uma função linear, conforme proposto em (Bertsekas D., Gallager R., 1992) e mostrada na Equação 6.1. em que  $p_{ij}$  é o valor da unidade da capacidade do enlace.

$$Custo = \sum_{i,j} p_{ij} C_{ij} \quad (6.1)$$

Com o objetivo de considerar um custo conjunto, que expresse as capacidades dos enlaces e tamanho das filas, a Equação 6.1 pode-se estender para a Equação 6.2, onde a variável  $q_i$  representa o preço por unidade de dimensionamento da fila e  $B_i$  representa o tamanho da fila no nó  $i$ .

$$Custo = \sum_{i,j} p_{ij} C_{ij} + \sum_i^n q_i B_i \quad (6.2)$$

Para cada um dos fluxos, foram estabelecidos valores máximos de atraso e o percentual de perdas de pacotes, conforme mostrados na Tabela 6.4.

Tabela 6.4. Valores máximos de atrasos e perdas para os fluxos

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
Atraso Máximo	0,10	0,15	0,10	0,15	0,10	0,15	0,10	0,15	0,10	0,15
% Máximo de Perdas	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05	0,05

Conforme a análise realizada na seção 5.2.2, foram estabelecidos os seguintes limites:

- a) o tamanho da fila tem como valor mínimo 20 e valor máximo 200 pacotes. Foi observado que, no ambiente de simulação do NS-2, para valores das filas abaixo de 20 pacotes, os percentuais de perda de pacotes se mantinham constantes para diferentes valores de carga nos enlaces. O valor de limite superior 200 foi escolhido em função de dar para o processo de escolha dos melhores indivíduos um maior grau de liberdade para o algoritmo genético.
- b) A carga dos enlaces tem como valor mínimo 0,2 e como valor máximo 0,9.

Os resultados do processo de otimização podem ser observados na Tabela 6.5, em que são apresentadas as capacidades dos enlaces e as capacidades das filas.

Tabela 6.5. Capacidades otimizadas da rede para série 9vbr01

Capacidade dos enlaces (kbps)		Capacidade das filas (pacotes)	
Nós	Capacidade	Nós	Capacidade
1 2	153	1 2	64
1 4	1450	1 4	70
2 3	488	2 3	21
2 5	87	2 5	106
3 10	524	3 10	165
4 7	712	4 7	72
4 10	651	4 10	144
5 10	291	5 10	36
5 11	74	5 11	25
6 9	825	6 9	183
7 6	533	7 6	29
7 8	145	7 8	31
8 9	311	8 9	78
9 12	1313	9 12	118
10 6	73	10 6	110
10 12	1435	10 12	162
11 12	108	11 12	49

A rede da Fig.5.2 teve as suas capacidades e tamanho de filas definidas conforme a Tabela 6.5. Foram obtidos os valores de atraso e percentual de perda de pacotes com o modelo de tráfego híbrido e com a simulação no NS-2. Os resultados são mostrados na Tabela 6.6.

Tabela 6.6. Atrasos e percentual de perdas de pacotes para rede otimizada – série 9vbr01

F	Híbrido Analítico (1)		M/M/1/k (2)		Híbrido Simulado (3)		Markov Simulado (4)	
	Atraso	Perdas	Atraso	Perdas	Atraso	Perdas	Atraso	Perdas
F1	0,05	0,04	0,03	0,0	0,03	0,05	0,13	0,02
F2	0,06	0,09	0,05	0,008	0,03	0,003	0,17	0,04
F3	0,06	0,04	0,03	0,0	0,03	0,04	0,12	0,005
F4	0,06	0,09	0,05	0,008	0,04	0,04	0,18	0,07
F5	0,06	0,04	0,05	0,0	0,04	0,02	0,18	0,0
F6	0,04	0,04	0,03	0,0	0,04	0,02	0,11	0,07
F7	0,08	0,04	0,06	0,008	0,07	0,0	0,14	0,04
F8	0,03	0,04	0,03	0,0	0,03	0,008	0,15	0,04
F9	0,09	0,0	0,08	0,0	0,08	0,0	0,08	0,0
F10	0,05	0,04	0,05	0,0	0,04	0,04	0,23	0,03

Como pode-se observar na Tabela 6.6, na coluna (1) são mostrados os valores dos atrasos e percentual de perdas para cada fluxo. No caso dos atrasos, ao compará-los com os da Tabela 6.5, observa-se que para todos os fluxos, o processo de otimização encontrou

capacidades dos enlaces e capacidades das filas que permitem que cada fluxo tenha um valor menor que o máximo estabelecido.

No caso do percentual de perdas, para o fluxo F2 e o fluxo F4 o processo de otimização, depois de repetidas iterações, conseguiu alcançar valores mínimos percentuais acima dos máximos permitidos. Novamente, conforme mostrado na seção 5.2.2, os fluxos F2 e F4 apresentam uma maior dificuldade para terem o cálculo de percentual de perdas de pacotes corretamente aproximado. Para todos os outros fluxos, os valores do percentual de perdas permaneceram menores que os máximos estabelecidos.

No caso do percentual de perdas, para o fluxo F2 e o fluxo F4 o processo de otimização, depois de repetidas iterações, conseguiu alcançar valores mínimos percentuais acima dos máximos permitidos. Novamente, conforme mostrado na seção 5.2.2, os fluxos F2 e F4 apresentam uma maior dificuldade para terem o cálculo de percentual de perdas de pacotes corretamente aproximado. Para todos os outros fluxos, os valores do percentual de perdas permaneceram menores que os máximos estabelecidos.

Os resultados da rede simulada no NS-2 aparecem na coluna (3) da Tabela 6.6. Pode-se observar que para o caso dos atrasos, em todos os fluxos, o modelo analítico estabeleceu um conjunto de capacidades que produziram métricas de atrasos superiores que foram confirmados na simulação. Para o caso do percentual de perdas de pacotes em todos os casos, a simulação mostra valores menores que aqueles calculados pelo modelo. Confirma-se também um menor grau de sobre-estimação daquele observado na análise da seção 5.2.2. Para os fluxos F2 e F8 os percentuais de sobre-estimação foram bastante mais acentuados que para o resto dos fluxos. Para o F8, a carga média no enlace é de 0,30. Para o fluxo F2, a carga média no enlace é de 0,45, mas no enlace 9-12 é de 0,80. Novamente, como foi analisado na seção 5.2.2, a alta carga no enlace 9-12 tende a sobre-estimar o valor do percentual de perdas de pacotes. Para o fluxo F8, a carga média nos enlaces é 0,45, mas este fluxo também passa pelo enlace 9-12.

A coluna (2) mostra os valores de atraso e perdas ao utilizar o modelo M/M/1/k com o processo de otimização e a coluna (4) mostra a simulação da rede com as capacidades sugeridas pela otimização usando esse modelo.

Para comparar os resultados do modelo de tráfego híbrido com o modelo M/M/1/k, a rede foi otimizado com este modelo. Os resultados das capacidades dos enlaces e capacidades das filas são mostrados na Tabela 6.7.

Tabela 6.7. Capacidades da rede com M/M/1/k série 9vbr01

Capacidade dos enlaces (kbps)		Capacidade das filas (pacotes)	
Nós	Capacidade	Nós	Capacidade
1 2	120	1 2	152
1 4	549	1 4	89
2 3	201	2 3	171
2 5	78	2 5	96
3 10	264	3 10	146
4 7	303	4 7	158
4 10	265	4 10	185
5 10	166	5 10	51
5 11	82	5 11	47
6 9	324	6 9	20
7 6	343	7 6	174
7 8	175	7 8	175
8 9	211	8 9	148
9 12	470	9 12	60
10 6	109	10 6	53
10 12	541	10 12	118
11 12	63	11 12	43

Pode-se observar que em relação à Tabela 6.5, que as capacidades dos enlaces são consideravelmente menores. A capacidade das filas apresenta uma diferença menos marcante. A média da coluna 2 para a Tabela 6.7 é 110 e a da Tabela 6.5 é 86. A média das capacidades dos enlaces da Tabela 6.5 tende a ser consideravelmente menor que a da Tabela 6.7. Este resultado é interessante pois mostra que o modelo de tráfego híbrido induz o aumento das capacidades dos enlaces mais que as capacidades das filas, e isso é coerente com as recomendações observadas em outros trabalhos (Appenzeller Guido, 2005).

A comparação dos atrasos e das perdas quando utilizado o modelo de Markov para fazer a otimização, conforme mostrado na Tabela 6.6 colunas 2 e 4, confirma a tendência deste modelo a sub-estimar estas métricas ao tratar a fonte como um processo de chegadas de Poisson. Pode-se observar que os atrasos calculados pelo modelo analítico M/M/1/k são consideravelmente menores aos da simulação, coluna 4. Os valores das perdas calculado pelo modelo analítico M/M/1/k são sub-estimados para todos os fluxos.

As tabelas 6.8 e 6.9 mostram as métricas da rede otimizada para a série 9vbr02 e as capacidades dos enlaces e filas calculados para essa série, respectivamente.

Tabela 6.8. Métricas da rede otimizada para série 9vbr02

F	Híbrido Analítico		Híbrido Simulado	
	Atraso	Perdas	Atraso	Perdas
F1	0,07	0,01	0,07	0,0
F2	0,13	0,06	0,07	0,0
F3	0,07	0,01	0,03	0,0
F4	0,13	0,06	0,11	0,0
F5	0,10	0,03	0,03	0,0
F6	0,10	0,05	0,10	0,0
F7	0,09	0,03	0,04	0,0
F8	0,14	0,07	0,10	0,02
F9	0,07	0,0	0,07	0,0
F10	0,09	0,03	0,06	0,05

Pode-se observar que nestas tabelas que as métricas de atraso e perdas calculadas pelo modelo analítico apresentam um grau de sobre-estimação menor e que pode ser considerado aceitável na maioria dos casos. Pode-se observar, na Tabela 6.9, que o tamanho das filas está na maioria dos casos 50% abaixo do valor máximo admissível.

Tabela 6.9. Capacidades otimizadas da rede para série 9vbr02

Capacidade dos enlaces (kbps)		Capacidade das filas (pacotes)	
Nós	Capacidade	Nós	Capacidade
1 2	217	1 2	35
1 4	1074	1 4	113
2 3	484	2 3	20
2 5	121	2 5	82
3 10	517	3 10	72
4 7	529	4 7	33
4 10	464	4 10	49
5 10	397	5 10	36
5 11	77	5 11	30
6 9	831	6 9	86
7 6	775	7 6	155
7 8	251	7 8	26
8 9	412	8 9	175
9 12	626	9 12	42
10 6	113	10 6	182
10 12	696	10 12	44
11 12	57	11 12	60

Na análise mostrada nas tabelas anteriores, induze-se uma certa relevância sobre o intervalo do tamanho possível para a capacidade máxima das filas. Com o intuito de avaliar este elemento e a sua influência na precisão das métricas calculadas pelo modelo de tráfego híbrido, foi introduzida uma variação nesse intervalo. Assim, foi feito um cálculo considerando que a fila teria como tamanho mínimo 20 e como tamanho máximo 60. Os resultados dos atrasos e do percentual de atrasos são mostrados na tabela 6.10.

Tabela 6.10. Métricas para a rede otimizada com filas menores – série 9vbr01

F	Híbrido Analítico		Híbrido Simulado	
	Atraso	Perdas	Atraso	Perdas
F1	0,09	0,14	0,09	0,002
F2	0,23	0,16	0,10	0,05
F3	0,09	0,14	0,03	0,0
F4	0,23	0,16	0,10	0,22
F5	0,30	0,18	0,03	0,0
F6	0,09	0,14	0,08	0,14
F7	0,22	0,15	0,07	0,0
F8	0,27	0,18	0,23	0,05
F9	0,08	0,0	0,06	0,0
F10	0,28	0,18	0,15	0,01

Como pode-ser observado na Tabela 6.10, o tamanho máximo das filas têm uma influência no grau de sobre-estimação das métricas para o processo de otimização da rede. Pode-se observar que os atrasos e as perdas têm um maior grau de sobre-estimação que aquele observado na Tabela 6.6 e em geral, a precisão do modelo para este caso não é aceitável. O tamanho do intervalo na escolha do tamanho das filas é então um elemento de melhora no processo de otimização.

Para ilustrar a influência do tamanho do intervalo para o tamanho da fila, foi feita uma otimização para a série 8vbr01, com dois intervalos. Os resultados são mostrados na Tabela 6.11.

Tabela 6.11. Atrasos e percentual de perdas de pacotes para rede otimizada – série 8vbr01

F	20 < Filas < 200				20 < Filas < 300			
	Híbrido Analítico (1)		Híbrido Simulado (2)		Híbrido Analítico (3)		Híbrido Simulado (4)	
	Atraso	Perdas	Atraso	Perdas	Atraso	Perdas	Atraso	Perdas
F1	0,04	0,03	0,01	0,0	0,06	0,03	0,05	0,001
F2	0,04	0,22	0,004	0,0	0,14	0,10	0,11	0,06
F3	0,02	0,0	0,01	0,0	0,06	0,03	0,03	0,0
F4	0,05	0,15	0,004	0,0	0,14	0,10	0,12	0,08
F5	0,03	0,0	0,04	0,05	0,10	0,01	0,04	0,0
F6	0,14	0,13	0,11	0,09	0,11	0,09	0,10	0,0
F7	0,04	0,0	0,01	0,0	0,09	0,02	0,04	0,0
F8	0,14	0,003	0,07	0,10	0,14	0,07	0,13	0,08
F9	0,08	0,0	0,06	0,0	0,08	0,0	0,08	0,0
F10	0,05	0,02	0,04	0,05	0,09	0,01	0,08	0,0

As colunas (1) e (2) correspondem ao modelo híbrido analítico e ao híbrido simulado, respectivamente, para um intervalo do tamanho da fila de 20 como mínimo e 200 como máximo. Como se pode observar nestas colunas, os valores de atraso do modelo analítico estão na maioria dos casos, muito próximos dos valores simulados. O mesmo não acontece com o percentual de perdas em que, em alguns fluxos são sub-estimados, enquanto em outros são sobre-estimados. Nas colunas (3) e (4), são mostradas as métricas para o tamanho da fila no intervalo (20,300). Pode-se observar que os atrasos do modelo analítico para todos os casos representam valores máximos em relação à simulação.

Para o percentual de perda de pacotes, pode-se observar que em todos os casos, os valores representam também valores máximos, com uma sobre-estimação menos acentuada que a observada nas colunas (1) e (2). Conclui-se então que, neste modelo de tráfego híbrido, a escolha do intervalo de valores possíveis da fila para um conjunto de fluxos agregados em uma certa topologia tem uma influência relevante no cálculo das métricas.

A Tabela 6.12 mostra um comparativo das capacidades dos enlaces e do tamanho das filas para cada um dos casos mostrados na Tabela 6.11. A média das capacidades para o tamanho das filas entre 20 e 200 é 752 kbps e o tamanho médio das filas é 39 pacotes. Para o tamanho das filas entre 20 e 300, a média das capacidades é 551 kbps e o tamanho médio das filas é 101 pacotes.



Tabela 6.12. Capacidades otimizadas da rede para série 8vbr01

20 < Tamanho Máximo < 200				20 < Tamanho Máximo < 300			
Capacidade dos enlaces (kbps)		Capacidade das filas (pacotes)		Capacidade dos enlaces (kbps)		Capacidade das filas (pacotes)	
Nós	Capacidade	Nós	Capacidade	Nós	Capacidade	Nós	Capacidade
1 2	335	1 2	64	1 2	112	1 2	29
1 4	1848	1 4	24	1 4	774	1 4	42
2 3	546	2 3	26	2 3	580	2 3	259
2 5	85	2 5	24	2 5	112	2 5	22
3 10	853	3 10	32	3 10	590	3 10	28
4 7	1157	4 7	85	4 7	823	4 7	85
4 10	592	4 10	40	4 10	709	4 10	239
5 10	223	5 10	30	5 10	434	5 10	217
5 11	146	5 11	20	5 11	64	5 11	48
6 9	1395	6 9	48	6 9	1133	6 9	199
7 6	995	7 6	20	7 6	397	7 6	21
7 8	313	7 8	27	7 8	188	7 8	29
8 9	593	8 9	54	8 9	452	8 9	133
9 12	1701	9 12	49	9 12	989	9 12	165
10 6	161	10 6	37	10 6	117	10 6	29
10 12	1786	10 12	39	10 12	1833	10 12	24
11 12	57	11 12	59	11 12	62	11 12	149

Então, dos resultados analisados anteriormente, observa-se que o aumento nas filas teve uma influência em uma melhor aproximação dos valores de percentual de perdas de pacotes e atraso. Pode-se concluir que o processo de otimização é então capaz de encontrar aqueles valores das capacidades e tamanho das filas que cumpram, com um melhor custo, os requerimentos de QoS estabelecidos para cada um dos fluxos.

## 6.5. EXTENSÕES DO MODELO PARA ENGENHARIA DE TRÁFEGO

Na Fig. 6.2 é mostrado o esquema de um modelo de engenharia de tráfego que utiliza a caracterização de tráfego para a tomada de decisões de mapeamento de fluxos.

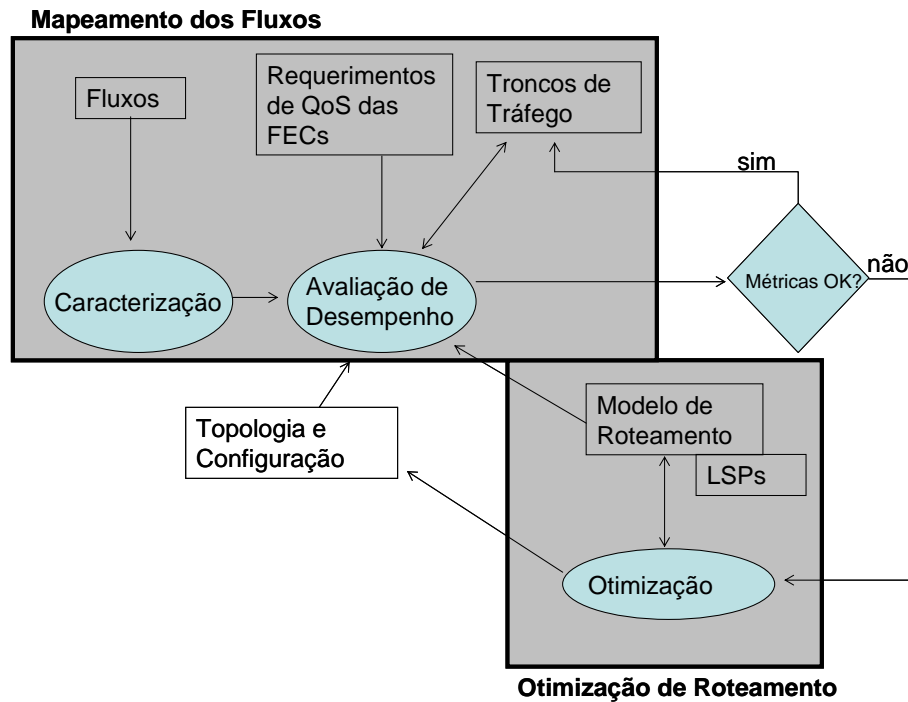


Figura 6.2. Esquema do processo de TE com caracterização de tráfego

Considere a matriz  $R_H$  em que são detalhados os fluxos relativos aos roteadores de borda de um domínio. Essa matriz é resultado de uma caracterização dos volumes de tráfego e representa um perfil dos fluxos esperados em cada roteador de borda.

Em um primeiro momento, os fluxos são caracterizados e, conforme um conjunto de requerimentos de QoS associado a classes de fluxo, é avaliado o desempenho do fluxo nos troncos de tráfego, para que, a partir dessa informação, seja decidido, conforme o modelo de roteamento, o tronco de tráfego onde será mapeado esse fluxo.

Quando o fluxo encontra um tronco de tráfego onde são cumpridos os requisitos das métricas para ele e para os outros fluxos, então esse fluxo é mapeado nesse tronco. Se as métricas não são possíveis de serem alcançadas para esse fluxo e para os outros troncos em que o fluxo é mapeado, então se inicia um processo de otimização do roteamento, que consistiria em otimizar as rotas de cada um dos LSPs da rede para alcançar o conjunto de métricas requeridas pelos troncos de tráfego existentes. Assim, também, nesse processo de otimização, podem-se alterar os troncos de tráfego, agrupando-se os fluxos de formas diferentes, conforme os resultados da avaliação de desempenho para os fluxos.

Pode-se observar então que a parte central do processo é a capacidade de poder avaliar o desempenho dos fluxos, e é nesse ponto que o modelo de tráfego híbrido proposto neste trabalho tem a sua relevância.

Para um conjunto de aplicações específicas, uma definição hipotética dos requerimentos de QoS para essas classes é mostrada na Tabela 6.13.

Tabela 6.13. Requerimentos de QoS das classes

Identificador da Classe	Taxa de perdas máxima	Atraso máximo
EF	$P_{EF}$	$A_{EF}$
BE	$P_{BE}$	$A_{BE}$
AF <sub>11</sub>	$P_{af11}$	$A_{af11}$
AF <sub>12</sub>	$P_{af12}$	$A_{af12}$
AF <sub>13</sub>	$P_{af13}$	$A_{af13}$

Tabela 6.14. Formulação do processo de TE com caracterização de tráfego

dados	<ul style="list-style-type: none"> <li>▪ Uma topologia</li> <li>▪ Uma matriz de tráfego <math>R_H</math> (tráfego caracterizado)</li> <li>▪ As funções de custo da capacidade <math>d_i(C_i)</math></li> <li>▪ Um conjunto de troncos de tráfego agrupados por FECs=<math>F</math></li> </ul>
minimizar	<ul style="list-style-type: none"> <li>▪ Uma função de custo</li> </ul>
ao considerar as variáveis	<ul style="list-style-type: none"> <li>▪ <math>LSP_R = (L_1, L_2, \dots, L_b)</math></li> <li>▪ <math>C = (C_1, C_2, \dots, C_n)</math></li> <li>▪ <math>F_L = \{(f, l) \in M: F \rightarrow LSP_R\}</math></li> </ul>
restrições	<ul style="list-style-type: none"> <li>▪ <math>\sum_i (f_i, L) &lt; \rho(L)</math>  <math>\forall f_i \in (f_i, L), QoSM(f_i, L) \leq QoSR(f_i)</math></li> <li>▪ onde <math>L \in LSP_R</math> e existe um <math>f_i \in F</math> onde <math>f_i = f_i</math></li> </ul>

A Tabela 6.14 apresenta a formulação do processo de TE com caracterização de tráfego. A função  $QoSR$  se refere aos requerimentos de QoS de uma classe de serviço específica e a função  $QoSM$  se refere ao cálculo de parâmetros de QoS específicos dessa classe quando mapeada em um LSP. Os parâmetros de QoS utilizados para este processo são o atraso e a

probabilidade de perda de pacotes. A função  $\rho(L)$  se refere ao cálculo da carga máxima admissível no LSP para a alocação da classe em função dos parâmetros de QoS e da capacidade física deste. O conjunto de troncos de tráfego  $F$  são os fluxos agrupados em FECs.

Cada fluxo  $f_n$  é caracterizado por uma terna  $(S,A,P)$  onde  $S=(H,a,\sigma^2)$ , em que  $H$  é o parâmetro de Hurst,  $a$  é a média e  $\sigma^2$  a variância. Os parâmetros  $A$  e  $P$  representam o atraso máximo que o fluxo tolera e a taxa máxima de perdas de pacotes admissível pelo fluxo, respectivamente.

De tal forma que, para que um fluxo possa ser associado a uma classe de equivalência de encaminhamento (FEC), deve-se cumprir:  $\forall f_i \in T_m \in FEC_s, AgH(H_i, T_m) < H_{FEC}, QoS\_Atraso(f_i, T_m) < A_{FEC}, QoS\_Perdas(f_i, T_m) < P_{FEC}$  em que  $H_{FEC}$  é o  $H$  máximo associado à classe,  $A_{FEC}$  é o atraso máximo associado à classe e  $P_{FEC}$  é a taxa máxima de perdas associada à classe. A função  $AgH$  faz a estimação do  $H$  resultante ao agregar o fluxo  $f_i$  no tronco de tráfego  $T_m$ .

Conforme a Fig.6.2 e a Tabela 6.14, podem-se sintetizar esses processos em um algoritmo que após a classificação e caracterização de um novo fluxo de tráfego entrante na rede, defina, com um conhecimento prévio da topologia da rede e da política de roteamento, o agrupamento de fluxos, ou tronco de tráfego, onde esse novo fluxo deve entrar para cumprir as métricas de QoS requeridas por aquele tronco de tráfego.

Esse algoritmo, formado por seis passos, é apresentado a continuação:

```

Mapear_Fluxos( $f_i(H,A,P), G(V,E,L), T_D$ )
1  Tipo_de_Classe ← Classifica_fluxo( $f_i(H,A,P)$ )
2  B ← Calcular_Mapeamento( $f_i(H,A,P), G(V,E,L), Tipo_de_Classe$ )
3  Se B = 1
4  então Atualizar_Tabela( $T_D, f_i$ )
5         Aceitar_Fluxo( $f_i, G(V,E,L)$ )
6  senão Otimizar_Rotas( $G(V,E,L), T_D$ )

```

O algoritmo recebe como parâmetros um fluxo, a topologia da rede com os LSPs e uma tabela  $T_D$ , que contém as definições como estão mapeados os fluxos em classes e a associação dessas classes em LSPs.

Na linha 1 do algoritmo, é realizada a identificação do tipo de classe que deve ser associada ao fluxo de tráfego. Em função das especificações desse fluxo, seleciona-se através de uma busca o tipo de classe que esse fluxo deve ter. Na linha 2, realiza-se o processo de mapeamento do fluxo, em função da sua classe, em um dos LSPs. Este processo precisa avaliar os parâmetros de QoS e a carga nos enlaces para identificar em que classe associada a um LSP esse fluxo deve-se agrupar. O processo de avaliação dos parâmetros de desempenho deve ser realizado para o fluxo entrante e para os fluxos restantes associados ao mesmo tronco de tráfego.

O passo 2, então, é um processo de otimização que deve escolher, em função da classe de tráfego do fluxo, dos seus parâmetros de desempenho e dos mapeamentos já existentes de classes em LSPs, o melhor posicionamento ou agrupamento desse fluxo entrante em um tronco de tráfego já existente.

No algoritmo, a variável  $B$  funciona como um gatilho para o processo de otimização. Se o valor retornado na variável  $B$  é igual a 1, este é um indicador de que o mapeamento do fluxo foi possível na topologia de rede existente. Caso contrário, indica que a rede está com uma necessidade de re-dimensionamento e re-adaptação e, na linha 6, este processo é realizado.

Para ilustrar o funcionamento dos passos 1-5 do algoritmo, considere-se o cenário de rede conforme a Fig. 5.2, a tabela de rotas conforme a Tabela 5.3 e os parâmetros dos fluxos F1, F2, F4, F6, F8 e F10 correspondem aos da série 9vbr01 da Tabela 5.8 e os tráfegos do tipo Poisson, F3, F5, F7 e F9, gerados com  $H=0.5$ , taxa média de 39 kbps e variância de  $4.998377 \times 10^4$ . Todos os fluxos foram gerados na escala de 0,1 segundos.

O atraso máximo admissível para todos os fluxos é de 0,15 segundos. O percentual de perdas máximo admissível para todos os fluxos é de 15%. A rede foi otimizada para cumprir essas métricas para cada um dos fluxos, conforme mostrado na seção 6.4. O fluxo entrante é um fluxo auto-similar com os parâmetros correspondentes ao fluxo 8vbr01 da Tabela 5.8. Os resultado do processo de busca da melhor rota para cumprir as métricas resultou na rota  $R1 \rightarrow R4 \rightarrow R10 \rightarrow R12$ , correspondente aos fluxos F1 e F3 da Tabela 5.3.

O resultado das métricas de QoS, com o fluxo 8vbr01 agregado aos fluxos F1 e F3 é mostrado na Tabela 6.15. O fluxo F11 corresponde ao novo fluxo agregado na rede. Pode-se observar que o atraso calculado para cada um dos fluxos representa, como foi observado

nos casos anteriores, valores máximos para os atraso reais, conforme o resultado da simulação da rede. Para os valores das perdas, o modelo analítico teve leves sub-estimações para os fluxos F1, F4, F6 e F11. Observe-se também que os valores das perdas na simulação se mantiveram abaixo dos valores máximos estipulados.

Tabela 6.15. Métricas após agregação do fluxo 8vbr01

F	Híbrido Analítico		Híbrido Simulado	
	Atraso	Perdas	Atraso	Perdas
F1	0,13	0,08	0,04	0,10
F2	0,11	0,12	0,04	0,12
F3	0,13	0,08	0,02	0,0
F4	0,11	0,12	0,06	0,10
F5	0,14	0,10	0,03	0,0
F6	0,13	0,11	0,06	0,10
F7	0,19	0,12	0,06	0,0
F8	0,13	0,07	0,10	0,0
F9	0,10	0,0	0,09	0,0
F10	0,13	0,10	0,07	0,07
F11	0,13	0,08	0,04	0,10

Para verificar que foi feito o mapeamento do fluxo no caminho que representa as melhores métricas em relação às outras possibilidades, na Tabela 6.16 são mostrados os valores das métricas para o mapeamento do fluxo na rota correspondente ao fluxo F2 e ao fluxo F10, nas colunas enumeradas como (1) e (2) respectivamente.

Tabela 6.16. Métricas para vários mapeamentos de 8vbr01

F	Rota F2 (1)		Rota F10 (2)	
	Atraso	Perdas	Atraso	Perdas
F1	0,09	0,08	0,09	0,09
F2	0,26	0,11	0,11	0,13
F3	0,09	0,08	0,09	0,09
F4	0,26	0,11	0,11	0,13
F5	0,14	0,10	0,53	0,11
F6	0,12	0,11	0,12	0,11
F7	0,19	0,11	0,15	0,13
F8	0,09	0,05	0,09	0,07
F9	0,10	0,0	0,10	0,0
F10	0,13	0,10	0,51	0,11
F11	0,26	0,11	0,51	0,11

Pode-se observar que para o caso da rota do fluxo F2, os valores do atraso dos fluxos F2 e F11 são maiores que os atrasos máximos permitidos, que são de 0,15s para cada fluxo. No

caso do percentual de perdas, todos os valores se mantêm abaixo dos valores máximos permitidos. Para o caso do mapeamento do fluxo novo na rota do fluxo F10, os valores dos atrasos do fluxo F5, F10 e F11 resultam estar bastante acima do valor permitido. Os valores percentuais de perdas de pacotes se mantêm abaixo dos valores permitidos.

As tentativas de mapeamento do fluxo nas rotas dos fluxos F5 e F9 não foram consideradas pelo algoritmo em função de apresentarem cargas nos enlaces maiores que 100%, o que é um critério de descarte para esses caminhos.

Se não tivesse sido encontrada uma rota para o fluxo entrante que permitisse a manutenção dos valores de perda e atraso abaixo dos valores limite, o algoritmo deveria executar a linha 6 com a função  $Otimizar\_Rotas(G(V, E, L), T_D)$ . Com esta função deve-se, para o conjunto de troncos de tráfego, de forma conjunta otimizar o agrupamento dos fluxos e dimensionar a rede em função dos requerimentos de qualidade de serviço, processo parcialmente ilustrado na seção 6.4, onde, foi mostrado o processo de otimização do dimensionamento para um conjunto de métricas de QoS considerando-se um roteamento fixo.

A função  $Otimizar\_Rotas$  é então um processo de otimização que permitiria posicionar os LSPs em função das capacidades e dos requerimentos de QoS. Esse procedimento não foi desenvolvido neste trabalho de tese e é considerado um trabalho futuro, derivado dos conceitos apresentados e exemplificados neste capítulo.

## **6.6. RESUMO DO CAPÍTULO**

Neste capítulo foram apresentados os problemas de dimensionamento de rede e da definição de roteamento. Um dos elementos para a solução destes problemas está em dispor, como uma ferramenta de uso, um modelo de tráfego que permita que em uma rede com tráfego multimídia possam-se obter métricas de QoS que sejam melhores daquelas obtidas com os modelos tradicionais, principalmente com base no modelo M/M/1/k.

A ferramenta de otimização utilizada foi um algoritmo genético que minimiza uma função de custo linear para cumprir as métricas estabelecidas. Verificou-se que a topologia de rede sugerida pelo processo de otimização apresentou resultados satisfatórios que foram confirmados com simulações no NS-2.

Com o modelo de tráfego híbrido, apresentado no Capítulo 5, para o cálculo das métricas de QoS, foi realizado o processo de otimização do dimensionamento de dois elementos da rede: as capacidades dos enlaces e o tamanho das filas. Observou-se que o intervalo de variação do tamanho da fila no processo de otimização é um elemento primordial para uma melhora na aproximação das métricas de QoS.

Nos exemplos mostrados, verificou-se uma aproximação aceitável das métricas QoS para a topologia de rede sugerida pelo processo de otimização. Esses resultados foram confirmados com simulações no NS-2, que validaram a aplicabilidade do modelo de tráfego híbrido para otimizar o dimensionamento de dois importantes elementos de rede: capacidades e tamanho das filas.

A última parte do capítulo apresenta uma extensão para o processo de engenharia de tráfego em uma rede multimídia. É apresentado e discutido um esquema para a realização deste processo com base no modelo de tráfego proposto.

É apresentado e detalhado um algoritmo de 6 passos que permite a realização da engenharia de tráfego e que procura preservar a estabilidade da rede ao mapear fluxos novos nas rotas mais adequadas, ou seja, aquelas que permitiriam manter as métricas de QoS estabelecidas para os fluxos.

Exemplificou-se o mapeamento de um fluxo entrante, usando novamente o cenário de uma rede com 12 nós e 11 fluxos para este caso. Foi efetuada a validação deste processo de mapeamento com simulações no NS-2 e comparações com outros potenciais mapeamentos. Discutiu-se, finalmente, a extensão deste trabalho para o processo de otimização de roteamento, que é o passo 6 do algoritmo apresentado na última seção deste capítulo.



## 7 - CONCLUSÕES E TRABALHOS FUTUROS

### 7.1. CONCLUSÕES

O foco desta tese foi a elaboração de uma metodologia para otimizar o planejamento e promover a estabilidade operativa de uma rede com tráfego multimídia. Como componente principal da proposta foi apresentado um modelo de tráfego híbrido com base nos processos envoltória fBm e modelos de filas markovianas. O modelo de tráfego híbrido considera a carga no enlace e os parâmetros do tráfego agregado para produzir simultaneamente métricas de atraso e perdas. O processo de otimização do planejamento da rede minimiza uma função de custo da rede ao estabelecer os valores de capacidades dos enlaces e tamanhos das filas necessários para cumprir um conjunto de métricas de QoS. A otimização é realizada com uma ferramenta evolucionária, um algoritmo genético. O modelo de tráfego híbrido também foi utilizado no processo de mapeamento de novos fluxos na rede, proposta que promove a estabilidade operativa da rede.

No capítulo 2, com o intuito de limitar o escopo do problema estudado nesta tese, foi apresentado um resumo do estado da arte e foram apresentados os fundamentos e conceitos teóricos sobre planejamento de rede. Foram discutidos os problemas de engenharia de tráfego para redes multimídia com tecnologia MPLS. Como resultado deste estudo, identificou-se a necessidade da caracterização de tráfego em redes multimídia e a implementação de processos e métodos que permitam otimizar os processos de planejamento e roteamento de fluxos na rede.

No capítulo 3, foram introduzidos os conceitos dos modelos M/M/1 e M/M/1/K dentro do contexto de planejamento de redes. Foi apresentado o uso do modelo de envoltória fBm e foram derivadas as expressões para o cálculo do atraso máximo e da probabilidade de perdas de pacotes em uma rede com tráfego multimídia. Estas expressões foram estendidas para uma rede de multiplexação estatística com circuitos virtuais. Foi feita uma análise experimental do processo de agregação de fluxos auto-similares através da comparação de vários métodos. Foi estabelecida uma relação entre as características de uma fonte auto-similar, o nível de carga no enlace e as métricas de QoS. Os resultados permitiram o estabelecimento da relação entre variâncias e carga no enlace para o aumento dos valores de atraso e perda de pacotes.

O capítulo 4 foi apresentada a ferramenta de geração de tráfego utilizada neste trabalho, o GTAR. O processo de geração de tráfego da ferramenta foi avaliado para os tráfegos

Poisson e auto-similar, a partir de uma comparação com o tráfego gerado por uma outra ferramenta de uso acadêmico. O resultado da comparação mostra que o GTAR gera séries de tráfego sintéticas conforme os parâmetros especificados e com uma maior precisão de que a outra ferramenta. Foi avaliado também o processo de inferência estatístico da ferramenta para a caracterização de tráfego auto-similar. Para esse efeito foram coletadas séries de tráfego reais de um grande provedor de Internet brasileiro. Os resultados mostraram a precisão da ferramenta para a caracterização do tráfego.

O capítulo 5 foram estabelecidos dois cenários hipotéticos de rede com roteamento fixo para o cálculo de métricas de QoS onde se comprovou que, quando a fonte de tráfego é auto-similar, as métricas calculadas pelos métodos markovianos apresentam uma forte tendência a sub-estimar os valores na rede, o que na prática resulta em altas taxas de congestionamento nos nós. Foi proposto um modelo de tráfego híbrido com base nos modelos M/M/1k e fBm. Com um conjunto de séries sintéticas de tráfego, foi avaliado o modelo de tráfego híbrido para valores decrescentes de variância, em diferentes intervalos de carga nos enlaces e com um tamanho fixo máximo de fila. Essa análise mostrou, por meio de simulações, que o modelo de tráfego híbrido calcula métricas de QoS que representam valores limites. As métricas obtidas por simulação, na maioria dos casos, sempre estiveram abaixo dos valores calculados pelo modelo analítico. Nos casos em que os valores das métricas de QoS produzidas pela simulação estiveram acima desses valores limites, verificou-se que existia uma forte agregação de fluxos auto-similares, o que resulta em altas variâncias para a série de tráfego agregado. Observou-se também que o grau de sub-estimação do modelo analítico para estes casos era pequeno, com uma maior aproximação que a obtida com os modelos markovianos. Também foi observado que existia uma maior sobre-estimação do modelo analítico para os fluxos de tráfego do tipo Poisson e que apresentavam pouca agregação com fluxos auto-similares.

O capítulo 6 apresentou uma metodologia para otimização de parâmetros em uma rede multimídia. Avaliou-se a utilização de um algoritmo genético para a otimização das capacidades das redes hipotéticas em função de valores máximos do atraso e perdas de pacotes calculados pelo modelo de tráfego híbrido apresentado no capítulo 5. Neste capítulo também foram definidos os problemas de estabelecimento das capacidades em uma rede com roteamento fixo. Foi realizado o processo de otimização simultâneo das capacidades dos enlaces e das filas para uma rede hipotética com dez fluxos de tráfego.

O processo de otimização minimizou uma função de custo ao estabelecer esses parâmetros da rede ao produzir métricas de QoS abaixo de valores limites estabelecidos para cada fluxo. Os resultados do processo de otimização mostraram uma maior precisão em relação às simulações no NS-2 quando o intervalo de variação do tamanho máximo das filas era maior. Os resultados das métricas para cada um dos fluxos apresentaram resultados bastante aceitáveis, o que permitiu concluir a validação do modelo de tráfego híbrido para a otimização de dois importantes elementos de rede: as capacidades dos enlaces e o tamanho das filas. Este resultado, a um custo computacional baixo, permite que a partir de coletas de tráfego, se defina o ponto de partida de operação de uma rede. A utilização de um modelo analítico para este efeito é relevante dada a reconhecida dificuldade que existe na configuração de modelos de simulação de redes que trabalham com grandes volumes de tráfego.

Foi apresentada uma metodologia de engenharia de tráfego que apresenta três processos principais: avaliação do desempenho, caracterização do tráfego e otimização. Esta metodologia parte de uma rede otimizada, onde, a partir do mapeamento em função de métricas de QoS dos novos fluxos entrantes, permite preservar a estabilidade operativa da rede evitando-se assim o re-dimensionamento constante das capacidades dos enlaces e tamanho das filas.

Foi apresentado e avaliado o algoritmo para implementar a metodologia de tráfego proposta. Conclui-se que o modelo de tráfego híbrido é fundamental no processo de otimização do planejamento da rede e no mapeamento dos fluxos.

## **7.2. TRABALHOS FUTUROS**

Nesta tese, o foco concentrou-se na implementação de métodos de engenharia de tráfego para redes multimídia, que com o uso de modelos de tráfego e processos de otimização, apresentam resultados que têm um uso potencial em redes operativas. A metodologia de tráfego proposta tem vários elementos que não foram abordados na sua totalidade e que apresentam campos férteis de futuras pesquisas.

Um dos elementos de maior importância no processo de otimização de uma rede é o roteamento. Em Banerjee S. e Sidhu D. (2002), considera-se o problema de cálculo de caminhos para TE com os requisitos de banda e atraso. São propostos dois algoritmos que procuram incrementar a receita da rede, reduzir o custo e distribuir a carga. Ambos algoritmos são comparados com outros que utilizam as mesmas restrições e conseguem

melhores resultados em relação ao desempenho. Os autores não fazem nenhuma menção aos procedimentos para o cálculo das métricas em função das características do tráfego.

Em Riedl A. (2003), são utilizados algoritmos genéticos heurísticos para otimizar o roteamento e estabelecer as capacidades dos enlaces de uma rede IP multimídia. O autor caracteriza o tráfego em dois grupos: elástico e *stream* e são mostrados procedimentos diferentes para a otimização do roteamento em função dessa caracterização e de várias métricas de QoS e custo. O desempenho dos algoritmos genéticos mostrou-se favorável para a solução do problema, mas o modelo não permite que sejam agregados fluxos de diferentes grupos, que em algumas redes operacionais pode não ser viável. Em Fortz, B., Rexford J., e Thorup M. (2002), argumenta-se a utilização dos protocolos IP tradicionais como OSPF e IS-IS para realizar TE. No trabalho é explicado como adaptar a configuração dos pesos sob uma visão geral da rede que considera o tráfego e a topologia dentro de um domínio. Os autores mostram resultados em que com a utilização destes algoritmos de roteamento e uma política adaptativa para a colocação dos pesos, obtém-se resultados quase ótimos em relação às implementações dos algoritmos com extensões, a um menor custo computacional. O trabalho mostra redes que partem de uma topologia com os pesos com base na largura de banda e depois adapta esses pesos como funções da carga de tráfego e da topologia e se concentra em modificar poucos pesos, apresentando resultados satisfatórios de desempenho e baixa complexidade. O trabalho enfoca a capacidade do enlace como um parâmetro que indica a carga desejável de tráfego e no caso de tráfego elástico esta carga se estabelece até 60% da capacidade do enlace, mas não se faz referência ao método utilizado para o cálculo desse percentual e a sua implicação em métricas de QoS tais como atraso e perda de pacotes.

Todos os trabalhos anteriores apresentam a necessidade de obter métricas mais precisas que permitiriam dar maiores insumos ao processo de ajuste dos parâmetros dos algoritmos de roteamento para a otimização do estabelecimento dos LSPs na rede. O modelo de tráfego, proposto e o processo de otimização implementado com um algoritmo genético são ferramentas úteis para este efeito. A otimização de roteamento completaria o passo 6 do algoritmo proposto no capítulo 4 desta tese.

A partir da identificação do tráfego em redes multimídia com características de longa dependência tem-se realizado diversos estudos com diversas séries de tráfego real para avaliar a precisão desta caracterização (Vianna Neto, C A., 2005, Ji Y., 2006). Resultados experimentais mostram que o tráfego multimídia pode apresentar variações ao longo do tempo que o parâmetro  $H$  não consegue capturar o que resulta em um cascadeamento não

uniforme e é introduzida para esse efeito a função de Holder. Um dos resultados mais importantes mostra que quando o tráfego é observado em pequenas escalas apresenta uma dinâmica que os modelos de tráfego baseados em processos auto-similares sobre-estimam. Modelos de tráfego com base em processos multifractais têm sido derivados para caracterizar o tráfego multimídia em pequenas escalas de tempo. Esses modelos permitem a definição de um processo envoltória mBm (*multifractal Brownian Motion*). Como foi observado nos resultados do capítulo 4 desta tese, o modelo de tráfego proposto estabelece valores máximos para as métricas de QoS. Uma direção de pesquisa para a melhoria do modelo de tráfego híbrido seria integrar os processos envoltória mBm. Isto derivaria no estudo da relação para o estabelecimento da dinâmica de cálculo das métricas de QoS a partir de um processo de decisão resultante da caracterização do tráfego. Um resultado esperado deste processo seria que as métricas calculadas apresentem um menor grau de superestimação o que teria um resultado positivo no processo de otimização para os passos do algoritmo de engenharia de tráfego proposto nesta tese.

## BIBLIOGRAFIA

- A.M Law, D. Kelton. (1991). *Simulation Modeling and Analysis*, 2a. Edição, McGraw-Hill New York, E.U.A.
- Alves Resende R. (2001). *Roteamento de Tráfego Adaptativo Baseado em Caminho Mínimo em Redes MPLS*, Tese de Mestrado, Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Brasil.
- Alwayn V. (2002). *Advanced MPLS Design and Implementation*, Cisco Press, Indianapolis, E.U.A.
- Appenzeler Guido. (2005). *Sizing Router Buffers*, Tese de Doutorado, Departamento de Ciencia da Computação, Universidade de Stanford, E.U.A.
- Awduche D., et al. (1999). *Requirements for Traffic Engineering over MPLS*, Network Working Group, RFC 2702.
- Banerjee G. e Sidu D. (2002). “Comparative Analysis of Path Computation Techniques for MPLS Traffic Engineering.” In: *The International Journal of Computer and Telecommunications Networking*, Volume 4-1, 149-165.
- Beran, J. Sherman, R., Taqqu, M., and Wilinger, W. (1995). “Long-range Dependence in Variable-Bit-Rate Video Traffic.” In: *IEEE Transactions on Communications*, Volume 43, 1566-1579.
- Bertsekas D. e Gallager R. (1992). *Data Networks*, 2ª. Edição, Prentice-Hall, New Jersey, E.U.A.
- Bodamer S. e Charzinski J. (2002). “Evaluation of Effective Bandwidth Schemes for Self Similar Traffic.” In: *Proceedings 13<sup>th</sup> ITC Specialist Seminar on IP Measurement, Modeling and Management*, Vol. 21, 1-10, Monterrey, E.U.A.
- Buriol L.S, Resende M.G.C, Ribeiro C.C, Thorup M. (2003). “A Hybrid Genetic Algorithm for the Weight Setting Problem in OSPF/IS-IS Routing.” In: *Networks*, 46, 36-56
- Clegg Richard. (2005). “A Practical Guide to Measuring the Hurst Parameter.” In: *Proceedings of 21st UK Performance Engineering Workshop*, School of Computing Science, Inglaterra.
- Comer D. (2000). *Internetworking with TCP/IP-Principles, Protocols and Architectures*, Prentice-Hall, New Jersey, E.U.A.
- Cormen T. et al. (2002). *Algoritmos – Teoría e Prática*, 2ª. Edição, Editora Campus Ltda, São Paulo, Brasil.

- Davie B., e Rekhter J. (2000). *MPLS Technology and Applications*, Morgan Kaufmann Publishers, San Diego, E.U.A.
- D-ITG [Online]. Disponível: <http://www.grid.unina.it/software/ITG/index.php>, (2006, abril 8)
- Elwalid A., C. Jin, Low S., e Widjaja I. (2001). "MATE: MPLS Adaptive Traffic Engineering." In: *Proceedings INFOCOM - Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies*, Volume 3, 1300-1309 Anchorage, E.U. A.
- Ethereal application, [On-line] Disponível: <http://www.ethereal.com/>, (2006, 17 Out)
- Figueiredo Daniel R., Liu B., Feldmann A., Misra V., Towsley D., Willinger W.. (2005). "On TCP and self-similar traffic." In: *Performance Evaluation*, Elsevier Science Publishers, Volume 61(2-3), 129-141, Amsterdam, Países Baixos.
- Fonseca N.L.S, M. Gilberto e Neto C.A.V. (2000). "On the Equivalent Bandwidth of Self-Similar Sources." In: *ACM Transactions on Modeling and Computer Simulation*, Volume 10-2, 104-124, Nova Iorque, E.U.A.
- Fortz, B., Rexford J., e Thorup M. (2002). "Traffic Engineering with Traditional IP Routing Protocols." In: *IEEE Communications Magazine*, Volume 4-10, 118-124.
- Fujimoto R., Perrumalla K., Park A., Wu H., Ammar M., Riley G. (2003). "Large Scale Network Simulation: How Big? How Fast?," In: *11th IEEE Modelling, Analysis and Simulation in Computer and Telecommunications*, 116-123, IEEE Computer Society, Florida, E.U.A.
- Gerla, M. e Kleinrock L. (1977), "On the Topological Design of Distributed Computer Networks." In: *IEEE Transactions on Communications*, Vol 25(1), 48-60.
- Girish M, Zhou B., e Hu J. (2000). "Formulation of the Traffic Engineering Problems in MPLS Based IP Network." In: *Proceedings of the Fifth IEEE Symposium on Computers and Communications*, 214-219, Washington, E.U.A.
- Gomes D., Agoulmine N., de Souza J.N. (2004). "Bandwidth Allocation Management based on Neural Networks Prediction for VoD System Providers." In: *Lecture Notes in Computer Science*, Volume 3126 , 279 – 290.
- Gomes D., Fonseca N., Agoulmine N., Souza, J. N. (2000). "Neurocomputação na Estimativa do Parâmetro de Hurst." In: *19º Simpósio Brasileiro de Redes de Computadores*, Brasil.

- Iovanna P., Sabella R., Settembre M. (2003). “A Traffic Engineering System for the Multilayer Networks Based on the GMPLS Paradigm.” In: *IEEE Network*, Volume 17-2, 28-37.
- Jerkins J. e Wang J.L. (1999). “From Network measurement collection to traffic performance modeling: challenges and lessons learned.” In: *Journal of the Brazilian Computer Society*, Volume 5-3.
- Ji Y. (2006). “Multi-Scale Internet Traffic Analysis Using Piecewise Self-Similar Processes.” In: *IEICE Trans. Commun*, Volume E89-B, 2125-2133.
- Kettanni H., Gubner J. A. (2002). “A novel approach to the estimation of the Hurst parameter in self-similar traffic.” In: *Proceedings of the 27<sup>th</sup> IEEE International Conference on Local Computers Networks*, Tampa, Florida, E.U.A.
- Lambert R. (2004). *Evolução das tecnologias e protocolos para comunicação multimídia*, Dissertação de Mestrado, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasil.
- Leland W. E., Taquq M. S., Willinger W., Wilson D. V. (1994). “On the Self-Similar Nature of Ethernet Traffic (extended version).” In: *IEEE/ACM Transactions on Networking (TON)*, Volume 2-1, 1 – 15,
- Mayor G. e Silvester J. (1997). “Time Scale Analysis of an ATM Queueing System with Long-Range Dependent Traffic.” In: *Proceedings of the INFOCOM '97, 16th Annual Joint Conference of the IEEE Computer and Communications Societies*, Volume 1, 205.
- MGEN [Online]. Disponível: <http://mgen.pf.itd.nrl.navy.mil> software (2005, maio 3)
- Mitra D. e Ramakrishnan K.G. (2001). “Techniques for Traffic Engineering of Multiservice, Multipriority Networks.” In: *Bell Labs Technical Journal*, Volume 1, 139-151.
- Morin P. R., Neilson J. (1995). The impact of self-similar on network Performance Analysis, Relatório Técnico, Departamento de Ciências da Computação, Carleton University, E.U.A.
- Netflow, [On-line] Disponível : [www.cisco.com/en/US/tech/tk812](http://www.cisco.com/en/US/tech/tk812), (2006, outubro 17)
- Netperf [Online]. Disponível: <http://www.netperf.org/netperf/NetperfPage.html>, (2006, abril 8)
- Netspec [Online]. Disponível: <http://www.ittc.ku.edu/netspec>, (2006, abril 8)



- Ng K.J., Song S., Tang B., Zhao W. (2002). The Statistical End-to-end Delay Guarantee for Networks with Self-similar Traffic, Technical Report, Department of Computer Science, Baptist University, Hong Kong, China.
- Norros I. (1994). “On the use of fractional Brownian motion in the theory of connectionless networks.” In: *IEEE Journal on Selected Areas in Communications*, Volume 13, 953–962.
- NSimulator, [On-line] <http://www.isi.edu/nsnam/ns>, (2006, agosto 8)
- Orenstein P., Kim H., Lau C.L. (2001). “Bandwidth Allocation for Self Similar Traffic consisting of multiple traffic classes with distinct characteristics.” In: *IEEE Proceedings of Globecom 2001*, Vol. 4.
- Park K. (1997). “On the Effect and Control of Self-Similar Network Traffic: A Simulation Perspective.” In: *Proceedings of the 1997 Winter Simulation Conference*, IEEE Computer Society, 989-996, , Atlanta, E.U.A.
- Paxson V. (1997). “Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic.” In: *Computer Communication Review*, Volume 27, 5-18.
- Paxson V. e Floyd S. (1995). “Wide Area Traffic: The Failure of Poisson Modeling.” In: *IEEE/ACM Transactions on Networking*, Volume 3-3, 226-244.
- Pereira B.T e Ling L.L. (2002). “Network Performance Analysis of an Adaptive OSPF Routing Strategy – Effective Bandwidth Estimation.” In: *International Telecommunications Symposium*, Natal, Brasil.
- Pereira Dantas M.J. (2000). Predição Multiescala de Tráfego em Redes ATM utilizando modelagem Fractal e Análise de Wavelets, Dissertação de Mestrado, Universidade Federal de Goiás, Goiânia, Brasil.
- Perlingeiro Firmiano, Ling Lee. (2005). “Estudo de Estimação de Banda Efetiva para Tráfego Auto-Similar com Variância Infinita.” In: *XII Simpósio Brasileiro de Telecomunicações*, Campinas, São Paulo, Brasil.
- Rasiah P., e Chung J. (2002). “Traffic Engineering Optimal Routing for LSP Setup in MPLS.” In: *The 2002 Symposium 45th on Circuits and Systems*, Volume III, 272-275.
- Riedl Anton. (1998). “A Versatile Genetic Algorithm for Network Planning.” In: *Proceedings of EUNICE'98*, Munich, Alemanha.

- Riedl Anton. (2002). "A Hybrid Genetic Algorithm for Routing Optimization in IP Networks Utilizing Bandwidth and Delay Metrics." In: *IEEE Workshop on IP Operations*, 166-170.
- Riedl Anton. (2003). Routing Optimization and Capacity Assignment in Multi-Service IP Networks, Tese de Doutorado, Lehrstuhl für Kommunikationsnetze, Universidade Técnica de Munich, Alemanha.
- Rodrigues L, e Guardieiro P.R. (2004). "A Spatial and Temporal Analysis of Internet Aggregate Traffic at the Flow Level." In: *Global Telecommunications Conference, Globecom 2004*, Volume 2, 685-691.
- Rosen E. (2001). Multiprotocol Label Switching Architecture, Network Working Group, RFC-3031.
- Sahinoglu Z. e Tekinay S. (1999). "On Multimedia Networks: Self-Similar Traffic and Network Performance." In: *IEEE Communications Magazine* Volume1, 48-52.
- Samorodnitsky G. (2006). "Long memory and self-similar processes." In : *Annales de la Faculté des Sciences de Toulouse-Mathématiques*, Tomo XV, Número 1, 107-123.
- Solis Barreto P., de Carvalho, P. H. P., Soares, A. M., Abdalla Jr H. (2005). "The Influence of Self-similar Traffic in Real Time Applications." In: *Yu.I. Shokin, O.I. Potaturkin (eds.) IASTED International Conference on Communications Systems*, ACTA Press, Novosibirsk, Russia.
- Solis Barreto P., de Carvalho, P. H. P., Soares, A. M., Abdalla Jr H., Lambert, R. (2005). "Hybrid Model of Traffic Characterization for Optimization in GMPLS Optical Networks." In: *IMOC 2005, International Microwave and Optoelectronics Conference*, Brasilia, Brasil.
- Solis Barreto P., de Carvalho, P. H. P., Soares, A. M., Abdalla Jr H. (2005), "A Traffic Model for Multimedia Applications in Converged Networks." In: *13th IEEE Modelling, Analysis and Simulation in Computer and Telecommunications*, IEEE Computer Society, 163-170, Atlanta, E.U.A.
- Solis Barreto P., de Carvalho, P. H. P., Soares, A. M., Abdalla Jr H., Lambert, R., Amvame-Nze, G., Bizerra R.S. (2006). "Open Source Software for Evaluation of Applications in an Experimental Testbed for Converged Networks." In: *2nd. International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities*, IEEE Computer Society, Barcelona, Espanha.

- Solis Barreto P., de Carvalho, P. H. P., Soares, A. M., Abdalla Jr H., Lambert R., Amvame-Nze, G., Bizerra R.S. (2005). "Implementação de um Ambiente de Teste e Medição em Redes Convergentes." In: *XXII Simpósio Brasileiro de Telecomunicações*, Campinas, São Paulo, Brasil.
- Solis Barreto P., de Carvalho, P. H. P., Soares, A. M., Abdalla Jr H., Bizerra R. S. (2005). "Uma Ferramenta de Código Aberto para Análise de Desempenho em Redes Convergentes." In: *XXII Simpósio Brasileiro de Telecomunicações*, Campinas, São Paulo, Brasil.
- Solis Barreto P., de Carvalho, P. H. P., Soares, A. M., Abdalla Jr H., Lambert, R., Amvame-Nze, G. (2004). "Performance Evaluation of Shortest Path Computation for IP and MPLS Multi-service Networks over Open Source Implementation." In: *Lecture Notes in Computer Science*, Springer-Verlag, Volume 3126, 267-278.
- Stallings W. (2002). *High Speed Networks and Internets – Performance and Quality of Service*, 2a. edição, Prentice-Hall, New Jersey, E.U.A.
- T. Karagiannis, M. Faloutsos and Riedi R.H. (2002). "Long-Range Dependence: Now You See It, Now You Don't!" In: *Proc. IEEE Global Telecommunications Conf. Global Internet Symposium*, Volume 3, 2165- 2169.
- Taqq M. S., Willinger W. and Sherman B. (1997). "Proof of a fundamental result in self-similar traffic modeling." In: *Computer Communication Review*, Volume 27, 5-23.
- Tsybakov B., Georganas N. (2000). "Overflow and losses in a network queue with a self-similar input." In: *Queueing Systems*, Volume 35, 201-235.
- Vianna Neto, C. (2005). *Modelagem e Computação de Banda Passante Equivalente de Fluxos Multifractais*, Tese de Doutorado, Unicamp, Campinas, Brasil.
- Villamizar C. e Song C. (1994). "High Performance TCP in ANSNET." In: *ACM Computer Communications Review*, Volume 24-5, 45-60.
- Wang Z. (1999). "On the complexity of quality of service routing." In: *Information Processing Letters*, Volume 69, 111-114.
- Willinger W. e Paxson V. (1998). "Where Mathematics Meets the Internet." In: *Notices of the ACM*, Volume 45(8), 961-970.
- Xiao X., Hannan A., Bailey B., Ni L.M. (2000). "Traffic Engineering with MPLS in the Internet." In: *IEEE Network Magazine*, Volume 14(2), 28-33.
- Yousefi'zadeh H., Jonckheere E.A., Silvester J.A. (2003). "Utilizing Neural Networks to Reduce Packet Loss in Self-Similar Teletraffic Patterns." In: *Proceedings of IEEE*

*ICC Communication QoS, Reliability, and Performance Modeling, Volume 3,*  
1942-1946.