



Universidade Federal de Campina Grande  
Centro de Engenharia Elétrica e Informática  
Departamento de Engenharia Elétrica  
Programa de Pós Graduação em Engenharia Elétrica

Tese de Doutorado

**Modelo para a Distribuição de Probabilidade  
do Comprimento dos Pacotes em Redes de  
Computadores**

Ewerton Rômulo Silva Castro

Campina Grande – PB  
Março de 2011

Universidade Federal de Campina Grande  
Centro de Engenharia Elétrica e Informática  
Departamento de Engenharia Elétrica  
Programa de Pós Graduação em Engenharia Elétrica

## **Modelo para a Distribuição de Probabilidade do Comprimento dos Pacotes em Redes de Computadores**

Ewerton Rômulo Silva Castro

Tese de Doutorado submetida Coordenação do Programa de Pós Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica da Universidade Federal de Campina Grande como requisito necessário para obtenção do grau de Doutor em Ciências no Domínio da Engenharia Elétrica.

Área de Concentração: Comunicações.

Marcelo Sampaio de Alencar, Ph.D. UFCG  
Orientador

Iguatemi Eduardo Fonseca, D.Sc., UFERSA  
Orientador

Campina Grande – PB, Março de 2011

©Ewerton Rômulo Silva Castro – ewerton.castro@ee.ufcg.edu.br

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

C355m

Castro, Ewerton Rômulo Silva

Modelo para a Distribuição de Probabilidade de Comprimento dos Pacotes em Redes de Computadores / Ewerton Rômulo Silva Castro. — Campina Grande, 2011.

112 f.: il.

Tese (Doutorado em Engenharia Elétrica) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática.

Orientador: Prof<sup>o</sup> Ph.D. Marcelo Sampaio de Alencar, Prof<sup>o</sup>. Dr. Iguatemi Eduardo Fonseca.

Referências.

1. Comprimento de Pacotes. 2. Densidade de Probabilidade. 3. Distribuição Cumulativa. 4. Rede de Computadores. 5. Tráfego de Internet.  
I. Título.

CDU 004.7(043)

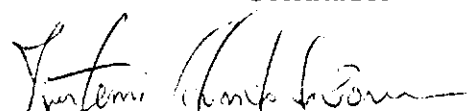
UFCG - BIBLIOTECA - CAMPUS I	
2574	05-05-011

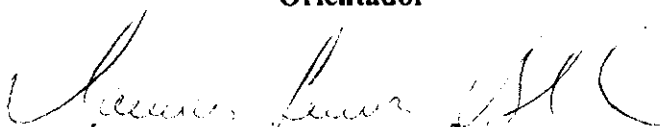
**MODELO PARA A DISTRIBUIÇÃO DE PROBABILIDADE DO COMPRIMENTO  
DOS PACOTES EM REDES DE COMPUTADORES**

**EWERTON RÔMULO SILVA CASTRO**

Tese Aprovada em 25.02.2011

  
**MARCELO SAMPAIO DE ALENCAR, Ph.D., UFCG**  
Orientador

  
**IGUATEMI EDUARDO DA FONSECA, Dr., UFERSA**  
Orientador

  
**MAURÍCIO FERREIRA MAGALHÃES, Dr., UNICAMP**  
Componente da Banca

  
**MARCELO LUIS FRANCISCO ABBADE, Dr., PUC-Campinas-SP**  
Componente da Banca

  
**JOSÉ ANTÃO BELTRÃO MOURA, Ph.D., UFCG**  
Componente da Banca

  
**MARCOS ANTONIO GONÇALVES BRASILEIRO, D.Sc.**  
Componente da Banca

  
**JOSÉ EWERTON POMBO DE FARIAS, Dr., UFCG**  
Componente da Banca

**CAMPINA GRANDE – PB**  
**FEVEREIRO - 2011**

*Dedico esta tese a meus pais, Romualdo e Jadilene, aos meus irmãos, Raudek e João Neto, à minha avó Palmira e minha noiva Elizabete, que sempre me apoiaram e incentivaram para concluir esta etapa de minha vida.*

---

---

# Agradecimentos

---

A Deus.

Aos meus pais, Romualdo e Jadilene, aos meus irmãos, Raudek Wolney e João Neto, à minha avó Palmira e minha noiva Elizabete, pelo apoio para concluir este trabalho;

Aos professores Marcelo Sampaio de Alencar e Iguatemi E. Fonseca, pela sábia orientação e amizade durante todo caminho percorrido neste trabalho;

Ao professor José Ewerton Pombo de Farias, pelas valiosas contribuições ao longo deste trabalho e pela amizade;

Aos professores Marcos A. Barbosa e Rômulo Valle pela amizade;

Aos alunos e ex-alunos de pós-graduação Évio, Vínicius, Edvaldo, Danilo, Jerônimo, Paulo, Portela, Rafael, Raíssa, Fausy, Erik, Késia, Fabrício, Luiz Felipe, Luiz Gonzaga, Manoel Alves, Ajey Kumar, Jacinto, Tomás, Thiago, Rex, Alfranke, pela amizade e apoio durante o trabalho no laboratório e na vida pessoal;

Aos funcionários Ângela, Pedro, Galba e Ronaldo, do DEE, pela presteza e amizade;

À CAPES pelo aporte financeiro.

Ao Iecom, que forneceu toda a infraestrutura necessária para a realização deste trabalho.

*Mais glorioso não é quem vence em batalhas milhares de homens, mas quem a si mesmo vence, este é o maior entre os conquistadores.*

— (Buda)

---

---

# Resumo

---

Os modelos tradicionais de tráfego telefônico não se adaptam ao tráfego de pacotes em rede de computadores. Portanto, as pesquisas para encontrar novos modelos enfatizaram o estudo de diversas características associadas à rede, como a duração do fluxo, o tempo entre chegadas de pacotes, o comprimento dos pacotes, além de suas propriedades estatísticas. Buscou-se equações mais adequadas para o modelamento. Esta tese apresenta um modelo que leva a uma equação para a função densidade de probabilidade, e sua função de distribuição cumulativa, do comprimento do pacote para rede de computadores com tráfego bimodal. Esse tipo de tráfego é importante por aparecer na literatura em diversas medições reais de comprimento dos pacotes. O modelo pode ser usado para geração de tráfego sintético, comparação, simulação e estimação do tráfego em redes de computadores. O modelo proposto é validado por comparações com dados reais de tráfego medidos pelo autor, resultados experimentais da literatura e comparações com outras distribuições de probabilidade.

**Palavras-chave:** Comprimento de Pacotes, Densidade de Probabilidade, Distribuição Cumulativa, Rede de computadores, Tráfego de Internet.



---

---

# Abstract

---

The traditional models of telephone traffic are not suitable for packet traffic in computer networks. Therefore, research to find new models focused on the study of several features associated with duration of flow, time between packet arrivals, packet lengths and their statistical properties. It was a search for more suitable equations for its modeling. This thesis presents a model that leads to an equation for the probability density function, and its cumulative distribution function, of the packet length for computer network with bimodal traffic. This type of traffic is important because it appears in the literature in several actual measurements of packet length. The model can be used for synthetic traffic generation, comparison, simulation and estimation of traffic on computer networks. The proposed model is validated by comparisons with actual data traffic measured by the author, experimental results of the literature and comparisons with other probability distributions.

**Keywords:** Packet Length, Probability Density, Cumulative Distribution, Computer Network, Internet Traffic.

---

---

# Sumário

---

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Modelamento Matemático</b>	<b>5</b>
2.1	Introdução . . . . .	5
2.2	Modelo para o Comprimento do Pacote . . . . .	5
2.3	Modelo da função densidade de probabilidade . . . . .	10
2.3.1	O denominador $\left \frac{d\ell}{dx}\right $ . . . . .	11
2.3.2	O numerador $p_X(x)$ . . . . .	12
2.3.3	Função Densidade de Probabilidade . . . . .	13
2.4	Cálculo da Função Cumulativa de Probabilidade . . . . .	15
2.4.1	Função cumulativa de probabilidade . . . . .	16
2.5	Primeiro Momento, Segundo Momento, Variância e Desvio Padrão . . . . .	17
2.5.1	Primeiro momento . . . . .	17
2.5.2	Segundo momento . . . . .	21
2.5.3	Variância . . . . .	24
2.5.4	Desvio padrão . . . . .	28
2.6	Conclusão . . . . .	30
<b>3</b>	<b>Medidas de Tráfego em Redes de Computadores</b>	<b>31</b>
3.1	Introdução . . . . .	31
3.2	Medidas das Frequências Relativas do Comprimento dos Pacotes . . . . .	31
3.2.1	Frequências Relativas . . . . .	32
3.3	Medidas cumulativas dos comprimentos dos pacotes . . . . .	58
3.3.1	Medições de dados cumulativos . . . . .	58
3.3.2	Dados obtidos da literatura . . . . .	67
3.4	Conclusão . . . . .	89
<b>4</b>	<b>Análise dos resultados</b>	<b>90</b>
4.1	Conclusão . . . . .	95

---

<b>5 Conclusão</b>	<b>96</b>
5.1 Conclusões . . . . .	96
5.2 Contribuições da Tese . . . . .	98
5.3 Sugestões para Trabalhos Futuros . . . . .	98
<b>A Publicações</b>	<b>100</b>
<b>B Função Gama</b>	<b>101</b>
<b>C Distribuição Exponencial</b>	<b>103</b>
<b>D Distribuição de Weibull</b>	<b>104</b>
<b>E Distribuição de Log-normal</b>	<b>105</b>
<b>F Distribuição de Pareto</b>	<b>106</b>

---

---

# Lista de Figuras

---

2.1	Fluxo de dados de uma rede com acesso à Internet. . . . .	6
2.2	Transformação não linear de fdp uniforme (área equivalente). (a) Densidade de probabilidade uniforme, (b) Modelo do comprimento do pacote e (c) Função densidade de probabilidade. . . . .	7
2.3	Função densidade de probabilidade $p(x)$ e função cumulativa de probabilidade $P(x)$ . . . . .	8
2.4	Transformação não linear de fdp não uniforme (área equivalente). (a) Densidade de probabilidade não uniforme, (b) Modelo do comprimento do pacote e (c) Função densidade de probabilidade. . . . .	9
2.5	Comprimento do pacote $\ell$ em função da variável aleatória $x$ . . . . .	10
2.6	Variável aleatória $x$ em função do comprimento do pacote $\ell$ . . . . .	10
2.7	Densidade de probabilidade da distribuição Beta para diferentes valores de $\alpha$ e $\beta$ . . . . .	13
2.8	Superfície de $E[L]$ em função de $\alpha$ e $\beta$ . . . . .	19
2.9	Gráfico de $E[L]$ em função de $\alpha$ . . . . .	20
2.10	Gráfico de $E[L]$ em função de $\beta$ . . . . .	20
2.11	Superfície de $\text{Var}[L]$ em função de $\alpha$ e $\beta$ . . . . .	26
2.12	Gráfico de $\text{Var}[L]$ em função de $\alpha$ . . . . .	27
2.13	Gráfico de $\text{Var}[L]$ em função de $\beta$ . . . . .	27
2.14	Superfície de $\sigma$ em função de $\alpha$ e $\beta$ . . . . .	28
2.15	Gráfico de $\sigma$ em função de $\alpha$ . . . . .	29
2.16	Gráfico de $\sigma$ em função de $\beta$ . . . . .	29
3.1	Tabela gerada pelo programa IPTRAF – Distribuição de comprimento dos pacotes. . . . .	32
3.2	Tabela com o arquivo <i>log</i> em formato texto gerado pelo programa IPTRAF – Distribuição de comprimento dos pacotes, obtido com o uso de um computador do laboratório durante um dia. . . . .	33
3.3	O modelo de referência ISO-OSI e o modelo TCP/IP. . . . .	34
3.4	Comprimento médio dos pacotes versus frequência de ocorrência dos pacotes. . . . .	40
3.5	Comprimento médio dos pacotes versus frequência de ocorrência dos pacotes, valores normalizados. . . . .	41

---

3.6	Modelo de densidade de probabilidade versus valores medidos (Diversos). . . . .	42
3.7	Modelo de densidade de probabilidade (dois intervalos) versus valores medidos (Diversos). . . . .	43
3.8	Tabela com o arquivo <i>log</i> em formato texto gerado pelo programa IPTRAF- <i>YouTube</i> . . . . .	44
3.9	Modelo de densidade de probabilidade versus valores medidos ( <i>YouTube</i> ). . . . .	45
3.10	Modelo de densidade de probabilidade (dois intervalos) versus valores medidos ( <i>YouTube</i> ). . . . .	45
3.11	Tabela com o arquivo <i>log</i> em formato texto gerado pelo programa IPTRAF- <i>Torrent</i> . . . . .	46
3.12	Tabela com o arquivo <i>log</i> em formato texto gerado pelo programa IPTRAF- <i>Torrent2</i> . . . . .	47
3.13	Modelo de densidade de probabilidade versus valores medidos ( <i>Torrent</i> ). . . . .	48
3.14	Modelo de densidade de probabilidade (dois intervalos) versus valores medidos ( <i>Torrent</i> ). . . . .	49
3.15	Modelo de densidade de probabilidade versus valores medidos ( <i>Torrent2</i> ). . . . .	49
3.16	Modelo de densidade de probabilidade (dois intervalos) versus valores medidos ( <i>Torrent2</i> ). . . . .	50
3.17	Tabela com o arquivo <i>log</i> em formato texto gerado pelo programa IPTRAF- <i>FTP</i> . . . . .	51
3.18	Modelo de densidade de probabilidade versus valores medidos ( <i>FTP</i> ). . . . .	52
3.19	Modelo de densidade de probabilidade (dois intervalos) versus valores medidos ( <i>FTP</i> ). . . . .	52
3.20	Tabela com o arquivo <i>log</i> em formato texto gerado pelo programa IPTRAF- <i>Lab Comp</i> . . . . .	53
3.21	Modelo de densidade de probabilidade versus valores medidos ( <i>Lab Comp</i> ). . . . .	54
3.22	Modelo de densidade de probabilidade (dois intervalos) versus valores medidos ( <i>Lab Comp</i> ). . . . .	55
3.23	Tabela com o arquivo <i>log</i> em formato texto gerado pelo programa IPTRAF- <i>Indústria</i> . . . . .	56
3.24	Modelo de densidade de probabilidade versus valores medidos ( <i>Indústria</i> ). . . . .	57
3.25	Modelo de densidade de probabilidade (dois intervalos) versus valores medidos ( <i>Indústria</i> ). . . . .	57
3.26	Modelo de distribuição cumulativa versus valores medidos ( <i>Diversos</i> ). . . . .	60
3.27	Modelo de distribuição cumulativa versus valores medidos ( <i>YouTube</i> ). . . . .	61
3.28	Modelo de distribuição cumulativa versus valores medidos ( <i>Torrent</i> ). . . . .	62
3.29	Modelo de distribuição cumulativa versus valores medidos ( <i>Torrent2</i> ). . . . .	63
3.30	Modelo de distribuição cumulativa versus valores medidos ( <i>FTP</i> ). . . . .	64
3.31	Modelo de distribuição cumulativa versus valores medidos ( <i>Lab Comp</i> ). . . . .	65
3.32	Modelo de distribuição cumulativa versus valores medidos ( <i>Indústria</i> ). . . . .	66

---

3.33	Modelo de distribuição cumulativa versus valores medidos ( <i>IP</i> ). . . . .	68
3.34	Modelo de distribuição cumulativa versus valores medidos ( <i>TCP</i> ). . . . .	69
3.35	Modelo de distribuição cumulativa versus valores medidos ( <i>HTTP</i> ). . . . .	70
3.36	Modelo de distribuição cumulativa versus valores medidos ( <i>UDP</i> ). . . . .	71
3.37	Modelo de distribuição cumulativa versus valores medidos ( <i>DNS</i> ). . . . .	72
3.38	Modelo de distribuição cumulativa versus valores medidos ( <i>SMTP</i> ). . . . .	73
3.39	Modelo de distribuição cumulativa versus valores medidos ( <i>SSH</i> ). . . . .	74
3.40	Modelo de distribuição cumulativa versus valores medidos ( <i>Tafvelin</i> ) e outras distribuições. . . . .	75
3.41	Modelo de distribuição cumulativa versus valores medidos ( <i>Rastin1</i> ) e outras distribuições. . . . .	76
3.42	Modelo de distribuição cumulativa versus valores medidos ( <i>Rastin2</i> ) e outras distribuições. . . . .	77
3.43	Modelo de distribuição cumulativa versus valores medidos ( <i>Rastin3</i> ) e outras distribuições. . . . .	78
3.44	Modelo de distribuição cumulativa versus valores medidos ( <i>Kos</i> ) e outras distribuições. . . . .	79
3.45	Modelo de distribuição cumulativa versus valores medidos ( <i>Rishi1</i> ) e outras distribuições. . . . .	80
3.46	Modelo de distribuição cumulativa versus valores medidos ( <i>Rishi2</i> ) e outras distribuições. . . . .	81
3.47	Modelo de distribuição cumulativa versus valores medidos ( <i>Rishi3</i> ) e outras distribuições. . . . .	82
3.48	Modelo de distribuição cumulativa versus valores medidos ( <i>Rishi4</i> ) e outras distribuições. . . . .	83
3.49	Modelo de distribuição cumulativa versus valores medidos ( <i>Rishi5</i> ) e outras distribuições. . . . .	84
3.50	Modelo de distribuição cumulativa versus valores medidos ( <i>Cheng</i> ) e outras distribuições. . . . .	85
3.51	Modelo de distribuição cumulativa versus valores medidos ( <i>Sprint1</i> ) e outras distribuições. . . . .	86
3.52	Modelo de distribuição cumulativa versus valores medidos ( <i>Sprint2</i> ) e outras distribuições. . . . .	87
3.53	Modelo de distribuição cumulativa versus valores medidos ( <i>CAIDA</i> ) e outras distribuições. . . . .	88
B.1	Função Gama $\Gamma(z)$ . . . . .	102

---

---

# Lista de Tabelas

---

3.1	Unidade Máxima de Transmissão (MTU) para diferentes tipos de rede. . . . .	34
3.2	Comprimentos mínimos e máxios dos pacotes para diferentes tipos de protocolos. . . . .	36
3.3	Tamanho, tamanho médio e frequência dos pacotes. . . . .	37
3.4	Tamanho médio ( $TM$ ) por diversos métodos. . . . .	38
3.5	Desvio padrão, variância e momentos. . . . .	38
3.6	Valores da Tabela 3.3 normalizados. . . . .	39
3.7	Comparação entre as medidas de Tafvelin, $P_L$ e outras distribuições. . . . .	75
3.8	Comparação entre as medidas de Pries (Rastin1), $P_L$ e outras distribuições. . . . .	77
3.9	Comparação entre as medidas de Pries (Rastin2), $P_L$ e outras distribuições. . . . .	77
3.10	Comparação entre as medidas de Pries (Rastin3), $P_L$ e outras distribuições. . . . .	78
3.11	Comparação entre as medidas de Kos, $P_L$ e outras distribuições. . . . .	79
3.12	Comparação entre as medidas de Rishi1, $P_L$ e outras distribuições. . . . .	81
3.13	Comparação entre as medidas de Rishi2, $P_L$ e outras distribuições. . . . .	81
3.14	Comparação entre as medidas de Rishi3, $P_L$ e outras distribuições. . . . .	82
3.15	Comparação entre as medidas de Rishi4, $P_L$ e outras distribuições. . . . .	83
3.16	Comparação entre as medidas de Rishi5, $P_L$ e outras distribuições. . . . .	84
3.17	Comparação entre as medidas de Cheng, $P_L$ e outras distribuições. . . . .	85
3.18	Comparação entre as medidas de SPRINT1, $P_L$ e outras distribuições. . . . .	86
3.19	Comparação entre as medidas de SPRINT2, $P_L$ e outras distribuições. . . . .	87
3.20	Comparação entre as medidas de CAIDA, $P_L$ e outras distribuições. . . . .	88
4.1	Resumo das comparações entre $p_L(\ell)$ e os valores experimentais. . . . .	90
4.2	Resumo das comparações entre $p_L(\ell)$ e os valores experimentais para dois intervalos. . . . .	91
4.3	Resumo das comparações entre $P_L(\ell)$ e os valores experimentais pelo autor. . . . .	92
4.4	Resumo das comparações entre $P_L(\ell)$ e os valores experimentais obtidos por Mattila. . . . .	93
4.5	Resumo das comparações entre $P_L(\ell)$ e os valores experimentais de outros autores. . . . .	94

---

---

# Lista de Siglas

---

<b>3G</b>	<i>Third Generation of Mobile Communications</i>	Terceira geração de comunicações móveis
<b>ADSL</b>	<i>Asymmetric Digital Subscriber Line</i>	Linha digital assimétrica para assinante
<b>DSL</b>	<i>Digital Subscriber Line</i>	Linha digital para assinante
<b>FTP</b>	<i>File Transfer Protocol</i>	Protocolo de transferência de arquivos
<b>HMM</b>	<i>Horário de Maior Movimento</i>	
<b>HTTP</b>	<i>Hypertext Transfer Protocol</i>	Protocolo de transferência de hipertexto
<b>ICMP</b>	<i>Internet Control Message Protocol</i>	Protocolo de controle de mensagens na Internet
<b>IETF</b>	<i>Internet Engineering Task Force</i>	Força tarefa de engenharia na Internet
<b>IP</b>	<i>Internet Protocol</i>	Protocolo da Internet
<b>IPTRAF</b>	<i>IP Network Monitoring Software</i>	Programa de monitoramento de rede IP
<b>ISDN</b>	<i>Integrated Services Digital Network</i>	Redes digitais de serviços integrados
<b>ITU</b>	<i>International Telecommunication Union</i>	União Internacional de Telecomunicações
<b>LAN</b>	<i>Local Area Network</i>	Redes locais
<b>MAC</b>	<i>Medium Access Control</i>	Controle de acesso ao meio
<b>MTU</b>	<i>Maximum Transmission Unit</i>	Unidade máxima de transmissão
<b>p2p</b>	<i>Peer-to-Peer</i>	Par-a-par, entre pares
<b>QoS</b>	<i>Quality of service</i>	Qualidade de serviço
<b>RFC</b>	<i>Request for Comments</i>	Pedido de comentários
<b>RMSE</b>	<i>Root Mean Square Error</i>	Erro da Raiz Média Quadrática
<b>SSE</b>	<i>Sum of Squares due to Error</i>	Soma dos Quadrados devido ao Erro
<b>SMTP</b>	<i>Simple Mail Transfer Protocol</i>	Protocolo simples de transferência de correspondência



<b>SNMP</b>	<i>Simple Network Management Protocol</i>	Protocolo simples de gerenciamento de redes
<b>SSH</b>	<i>Secure Shell</i>	Interpretador de comandos seguro
<b>TCP</b>	<i>Transmission Control Protocol</i>	Protocolo de controle de transmissão
<b>UDP</b>	<i>User Datagram Protocol</i>	Protocolo do datagrama de usuário
<b>VPN</b>	<i>Virtual Private Network</i>	Rede privada virtual
<b>WAN</b>	<i>Wide Area Network</i>	Redes de larga área

---

---

# Lista de Símbolos

---

$x$	Variável aleatória
$L_m$	Comprimento mínimo do pacote de bits ou bytes normalizados
$L_M$	Comprimento máximo do pacote de bits ou bytes normalizados
$l$	Variável aleatória que representa o comprimento do pacote normalizados
$B(\cdot)$	Função Beta
$I_x(\cdot)$	Função Beta incompleta normalizada
$\alpha$ e $\beta$	Parâmetros da distribuição relacionados ao tipo de tráfego
$\Gamma(\cdot)$	Função Gama
$p_L(\cdot)$	Função densidade de probabilidade do comprimento do pacote
$P_L(\cdot)$	Função de distribuição cumulativa do comprimento do pacote
$E[\cdot]$	Valor esperado
$\text{Var}[\cdot]$	Variância
$\sigma$	Desvio Padrão
${}_1F_1(\cdot)$	Função hypergeometrica degenerada ou Função hypergeometrica confluyente
$L_n^k(\cdot)$	Função de Laguerre

## CAPÍTULO 1

---

---

# Introdução

---

A Internet tem sido cada vez mais usada como ferramenta de pesquisa no mundo moderno. No Brasil, o envio e recebimento de informações tais como compras virtuais, transações comerciais e bancárias tem crescido, como demonstram dados de 2010, divulgados em agosto do mesmo ano, em que o número de transações comerciais *on line* se elevou em 40% em relação ao ano anterior. Em 2009, o faturamento do setor foi de R\$ 4,8 bilhões no primeiro semestre e em 2010, R\$ 6,7 bilhões no mesmo período. A projeção é que o faturamento seja de R\$ 14,3 bilhões para o ano de 2010, os valores reais serão divulgados em Abril de 2011. Se confirmado, será um crescimento de 35% em relação a 2009, quando o setor faturou cerca de R\$ 10,6 bilhões (EBIT, 2010).

A Pesquisa Nacional por Amostra de Domicílios (PNAD) 2009, do Instituto Brasileiro de Geografia e Estatística (IBGE), mostra ainda que o Brasil apresenta um dos maiores crescimentos de uso da rede. Em 2009, 27,4% (16 milhões) dos domicílios investigados em todo o país tinham acesso à Internet, contra 23,8% em 2008. Também em 2009, 67,9 milhões de pessoas com 10 ou mais anos de idade declararam ter usado a Internet, o que representa um aumento de 12 milhões (21,5%) sobre 2008. Em 2005, a Internet tinha 31,9 milhões de usuários, o aumento no período foi de 112,9% e observado em todas as regiões (PNAD-2009, 2010).

Esse crescimento rápido na demanda do número de acessos é acompanhado por uma maior demanda na taxa de transmissão de dados nas redes de computadores e nas redes de acesso à Internet. No entanto, muitas vezes, o crescimento da demanda na taxa de transmissão não é acompanhado por uma expansão da infra-estrutura para atendê-la. Por isso, há necessidade de se elaborar mais estudos teóricos para manter os níveis de qualidade de serviço (QoS) (ALTMAN *et al.*, 1997), utilizando a mesma infra-estrutura de rede instalada, principalmente, em Horários de Maior Movimento (HMM). A otimização da rede de acesso disponível pode ser conseguida com um tratamento estatístico do tráfego que flui na rede. Assim, é possível uma otimização sem comprometer o desempenho. Isso se torna ainda mais relevante com a crescente tendência de convergência de diversos tipos de redes e serviços sobre as redes baseadas no protocolo IP (IP – Internet Protocol).

Estudos mostram que os modelos estatísticos usados no tráfego telefônico não são realistas na descrição do tráfego em redes de computadores ou em redes convergentes, (LELAND *et al.*, 1994), (PAXSON; FLOYD, 1995), (CROVELLA; BESTAVROS, 1995), (WILLINGER *et al.*, 1998), (PARK; WILLINGER, 2000), (QUAN; CHUNG, 2002), (TANENBAUM, 2003). Assim, a busca por novos modelos estatísticos para o tráfego em redes e que supram as deficiências encontradas nos modelos clássicos tornou-se ênfase de diversas pesquisas (IACOVAZZI; BAIOCCHI, 2010).

Nessa busca por novos modelos, um dos caminhos encontrados é a caracterização do tráfego em redes de computadores usando o comprimento do pacote da rede. Esta caracterização tem sido estudada e alguns resultados estão disponíveis na literatura, (BEVERLY; CLAFFY, 2003), (KOS *et al.*, 2003), (SPRINT, 2005), (ZHAO BIAO *et al.*, 2006), (SINHA *et al.*, 2007), (YU *et al.*, 2007) e (CAIDA, 2008). Essas pesquisas, usualmente, envolvem as medições dos comprimentos dos pacotes, tendo em vista a possibilidade de usar estas informações no projeto e estimativa da infra-estrutura de redes e suas aplicações.

Em seguida, usando essa caracterização, busca-se encontrar um modelamento da distribuição de probabilidade do comprimento do pacote. No trabalho de Li Bo é sugerido que uma distribuição do comprimento do pacote pode ser usada para identificação de diferentes tipos de aplicações TCP (Protocolo de Controle de Transmissão) (BO *et al.*, 2006). Em outro trabalho, Alfonso Iacovazzi destaca que a classificação estatística do tráfego é possível com base em algumas das características do fluxo de dados IP. Ele menciona que o comprimento do pacote é uma característica chave na classificação da camada de aplicação do fluxo de pacotes e que esta classificação pode ser útil nas políticas de segurança, filtro de tráfego e no suporte aos mecanismos de qualidade de serviço (IACOVAZZI; BAIOCCHI, 2010).

Uma característica importante que essa distribuição de probabilidade teria, segundo Tafvelin, é um comportamento bimodal <sup>1</sup>, em que 40% dos pacotes tem tamanho menor que 44 *bytes* (primeiro pico) e que outros 40% dos pacotes estão entre 1400 *bytes* e 1500 *bytes* (segundo pico) (JOHN; TAFVELIN, 2007). Resultados similares foram encontrados por Rastin Pries (PRIES *et al.*, 2009) e Castro (CASTRO *et al.*, 2010). Os trabalhos de McCreary (MCCREARY; CLAFFY, ) e Karagiannis (KARAGIANNIS *et al.*, 2003) fazem referências ao comprimento dos pacotes com uma distribuição de tráfego trimodal. Enquanto que Cheng Yu vai mais longe e classifica o tráfego de dados sobre o ponto de vista do comprimento dos pacotes com distribuição multi-modal (YU *et al.*, 2007). No entanto, Rishi Sinha (SINHA *et al.*, 2007) e Tafvelin (JOHN; TAFVELIN, 2007) verificaram que houve uma mudança do senso comum do comportamento trimodal com picos próximos a 40, 576 e 1500 *bytes*, para um comportamento bimodal com pico próximos a 40 e 1500 *bytes*. Essa mudança, segundo esses autores, se deve à evolução dos sistemas operacionais

---

<sup>1</sup> A distribuição é chamada bimodal porque em estatística descritiva, a moda é o valor que detém o maior número de observações, ou seja, o valor ou valores mais frequentes. A moda não é necessariamente única, ao contrário da média ou da mediana. É especialmente útil quando os valores ou observações não são numéricos, uma vez que a média e a mediana podem não ser bem definidas. Um exemplo simples é a série {1, 1, 3, 5, 6, 6} que apresenta duas modas (bimodal): 1 e 6.

e do aumento do uso do padrão *Ethernet* com 1500 bytes de Unidade Máxima de Transmissão (MTU—*Maximum Transmission Unit*) (SINHA *et al.*, 2007).

Então, o objetivo desta tese é apresentar um modelo matemático para a função densidade de probabilidade de comprimento de pacotes com esta característica bimodal na distribuição do tráfego. E como mencionado por Mushtaq, a função de densidade de probabilidade ou a função de distribuição cumulativa, podem auxiliar no projeto, controle, gerência, interpolação e extrapolação de redes (MUSHTAQ; RIZVI, 2005).

Em outros dois trabalhos, Torabzadeh e Jung utilizam o comprimento do pacote para auxiliar no modelamento do comportamento desejado. No artigo apresentado por Torabzadeh (TORABZADEH; AJIB, 2010), é destacado que tradicionalmente nos agendamentos de pacotes por equidade proporcional (*Proportional Fairness Scheduler*) não são considerados as características do tráfego (Processo de chegada e comprimento dos pacotes) quando ocorre a atualização da taxa de transferência média de usuários na sua função de utilidade (ROUGHAN; VEITCH, 1998), (JALALI *et al.*, 2000), (ANDREWS, 2004), (SANG *et al.*, 2006), (ANIBA; AISSA, 2007). Assim, Torabzadeh apresenta um algoritmo de agendamento de pacotes que considera o comprimento do pacote, o comprimento da fila do usuário, a taxa de transmissão do usuário e a garantia de serviço. Jung destaca em seu artigo que para aumentar a transferência de dados, uma série de pacotes que compartilham um mesmo cabeçalho podem ser agregadas em um quadro (JUNG; SOHRABY, 2010). Esse processo de agregação foi proposto acerca de uma década (GOPALA-KRISHNA, 2003), (SHAFFER *et al.*, 1999). Em seu trabalho, é apresentado um modelo matemático de um sistema de agregação de pacotes assumindo uma distribuição geral para o comprimento do pacote (JUNG; SOHRABY, 2010), (HONG *et al.*, 2006), (HONG; SOHRABY, 2007).

Normalmente, as aplicações de Internet são identificadas por seus números de portas. No entanto, aplicações recentes e de tempo real nem sempre podem ser detectadas por esse método simples e, portanto, outras técnicas, tais como classificação de pacotes ou análise profunda de pacotes foram desenvolvidas (LEGEDZA *et al.*, 1998), (GUPTA; MCKEOWN, 2000), (GUPTA; MCKEOWN, 2001), (MACIAN; FINTHAMMER, 2001), (MCKEOWN; VARGHESE, 2001). Isso por causa do crescimento das aplicações de Internet que não usam um número de porta registrado (NIRKHE; BAUGHER, 1995), (DECASPER *et al.*, 1998). Parish considera uma abordagem alternativa para a detecção de aplicações em tempo real e obteve uma estatística da impressão digital do fluxo de tráfego gerado por esse aplicativos usando a distribuição do comprimento dos pacotes (PARISH *et al.*, 2003).

A maioria das abordagens que visam identificar o tráfego da camada de aplicação do IP ou do TCP, usando algum tipo de medida de fluxo de tráfego (KARAGIANNIS *et al.*, 2005). Crotti usou o comprimento e o tempo entre chegadas dos primeiros  $n$  pacotes para descrever estatisticamente (impressão digital) a camada de aplicação (CROTTI *et al.*, 2007). Essa impressão digital é usada para medir as similaridades entre um determinado fluxo e o seu protocolo correspondente. Os trabalhos de Sun (SUN *et al.*, 2002), McGregor (MCGREGOR *et al.*, 2004), Moore (MOORE; ZUEV, 2005), Zander (ZANDER *et al.*, 2005), Liberatore (LIBERATORE *et al.*, 2006),

(LIBERATORE; LEVINE, 2006), Wright (WRIGHT *et al.*, 2006), (WRIGHT *et al.*, 2007), (WRIGHT *et al.*, 2009), Alshammari (ALSHAMMARI; ZINCIR-HEYWOOD, 2007) e Dusi (DUSI *et al.*, 2009) apresentam algoritmos que utilizam algumas das características do tráfego: duração do fluxo, tempo entre chegadas de pacotes, comprimento dos pacotes e suas estatísticas (média, variância, desvio padrão, etc) para realizar a identificação do tráfego na aplicação.

Neste contexto, esta tese apresenta um modelo para a função de densidade de probabilidade (fdp) do comprimento dos pacotes em redes de computadores com distribuição de tráfego bimodal. As equações obtidas apresentam uma boa aproximação quando comparados com valores reais da densidade de probabilidade do tráfego dados. Essa aproximação é ainda melhor quando utiliza-se a função de distribuição cumulativa. Os resultados também são melhores que outras distribuições, Exponencial, Log-normal, Weibull e Pareto.

Para obter o modelo da função densidade de probabilidade, algumas condições matemáticas importantes são utilizadas e em seguida, a função de distribuição cumulativa é obtida. Também são obtidos a média, o segundo momento, a variância e o desvio padrão da distribuição, apresentados no Capítulo 2. Em seguida, no Capítulo 3, os resultados são comparados com as medidas de comprimento de pacote. Também são comparados a fórmula da distribuição cumulativa com os valores experimentais apresentados por vários autores da literatura. No Capítulo 4 são analisados essas comparações e os resultados obtidos. Finalmente, o Capítulo 5 apresenta as conclusões do trabalho.

## CAPÍTULO 2

---

---

# Modelamento Matemático

---

Neste capítulo é apresentado o modelo matemático da função densidade de probabilidade (fdp), e a função cumulativa de probabilidade (FCP), do comprimento dos pacotes em redes de computadores com tráfego bimodal. O objetivo do capítulo é apresentar o desenvolvimento matemático para se obter esse modelo de tráfego de dados e no capítulo posterior compara-se os resultados obtidos por diferentes autores com os resultados do modelo teórico abordado neste capítulo. Também são apresentadas as equações da média, segundo momento, variância e desvio padrão.

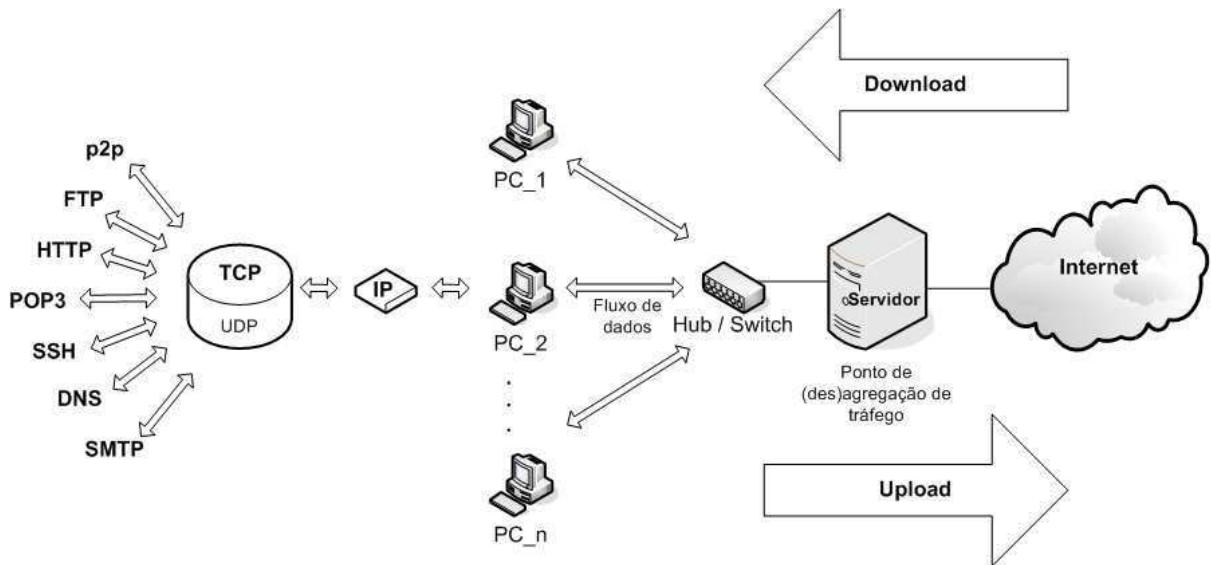
## 2.1 Introdução

Nas páginas a seguir são apresentadas as demonstrações matemáticas das fórmulas propostas para a fdp e a FCP do comprimento dos pacotes, também chamado de tamanho dos pacotes por alguns autores, em redes de computadores com tráfego bimodal. Inicialmente, apresentam-se um modelo de comprimento do pacote e algumas definições matemáticas utilizadas. Em seguida, é utilizado o conceito de mapeamento biunívoco (área equivalente) para se chegar à função densidade de probabilidade. Fazendo sua integração, obtém-se a função de distribuição cumulativa. Resolvendo integrais com termos envolvendo o comprimento do pacote e a fdp, encontram-se as equações da média e do segundo momento. A partir destas duas últimas equações, a equação da variância é obtida e, finalmente, com ela, o desvio padrão. No capítulo posterior, os modelos propostos são comparados com diversas medições realizadas, com valores de medidas obtidas da literatura e com outras distribuições de probabilidade.

## 2.2 Modelo para o Comprimento do Pacote

Nesta seção são apresentados um modelo de comprimento do pacote e algumas definições matemáticas utilizadas no mesmo. Em seguida, é usado no mapeamento biunívoco para se chegar à função densidade de probabilidade.

Considera-se o modelo ilustrado na Figura 2.1, uma configuração de rede com acesso à Internet, bastante comum em conexões residenciais e de muitas empresas. A principal diferença entre ambos é a substituição do servidor no ponto de (des)agregação do tráfego por um modem ou roteador de linha digital assimétrica para assinante (ADSL).



**Figura 2.1** Fluxo de dados de uma rede com acesso à Internet.

No passo seguinte, considera-se que o tráfego gerado por cada usuário seja uma distribuição uniforme, como ilustrado na Figura 2.2a, em que  $x$ ,  $L_m$  e  $L_M$  são definidos por

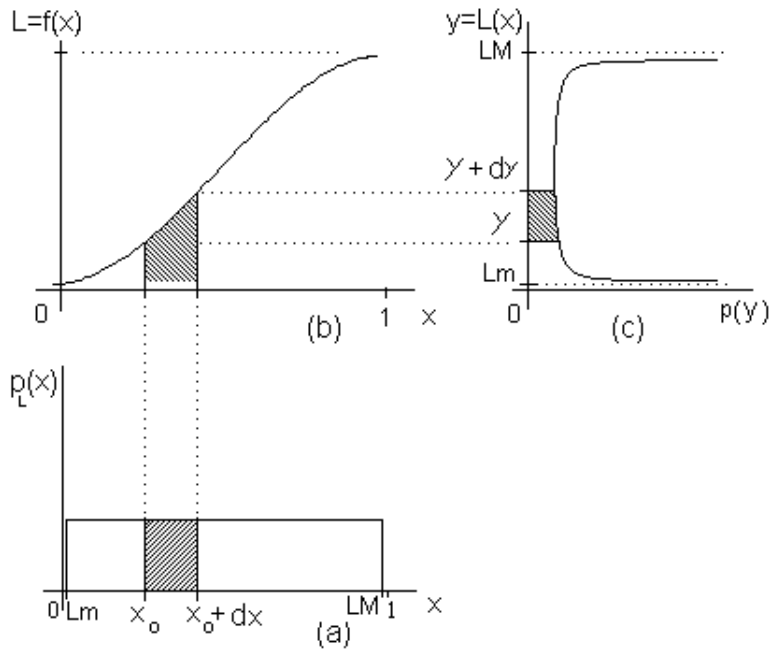
**Definição 1:** Considere que  $x$  seja uma variável aleatória e que  $0 \leq x \leq 1$ .

**Definição 2:** Suponha que  $L_m$ ,  $0 \leq L_m < 1$ , seja uma variável que representa o comprimento mínimo do pacote em número de *bits* ou *bytes* normalizados. Então,  $L_m$  é a razão entre o comprimento mínimo do pacote que pode ser enviado pela interface de rede e  $N_{max}$ . O termo  $N_{max}$  é o número máximo de *bits* ou *bytes* possível de ser enviado por uma interface de rede qualquer em um intervalo de tempo  $t_0$ .

**Definição 3:** Defina  $L_M$  como uma variável que representa o comprimento máximo do pacote em número de *bits* ou *bytes* normalizados. Logo,  $L_M$  é a razão entre o comprimento máximo do pacote que pode ser enviado pela interface de rede e  $N_{max}$ . O valor de  $L_M$  está dentro do intervalo  $0 < L_M \leq 1$  e  $L_m < L_M$ .

Nessas condições foram realizadas algumas medições de comprimento de pacote, no ponto de agregação do tráfego, Figura 2.1. Essas medidas indicaram uma distribuição bimodal, Figura 2.2c, com picos próximos a 40 e 1500 bytes (JOHN; TAFVELIN, 2007), (PRIES *et al.*, 2009). Se não houvesse qualquer influência do ponto de agregação no fluxo de dados que o atravessa, este ponto da rede também teria uma distribuição uniforme para o comprimento dos pacotes. Como isso não ocorre, há indícios da ocorrência de alguma transformação não linear de fdp no ponto de agregação, Figura 2.2b. A Transformação não linear de fdp pode ser obtida utilizando o conceito de mapeamento biunívoco (área equivalente), ilustrado na Figura 2.2 (ALENCAR, 2009), que leva à fórmula para transformação de variável aleatória dada por





**Figura 2.2** Transformação não linear de fdp uniforme (área equivalente). (a) Densidade de probabilidade uniforme, (b) Modelo do comprimento do pacote e (c) Função densidade de probabilidade.

$$p_X(x)dx = p_L(\ell)d\ell, \tag{2.1}$$

$$p_L(\ell) = \frac{p_X(x)}{\left| \frac{d\ell}{dx} \right|}, \quad x = f^{-1}(\ell). \tag{2.2}$$

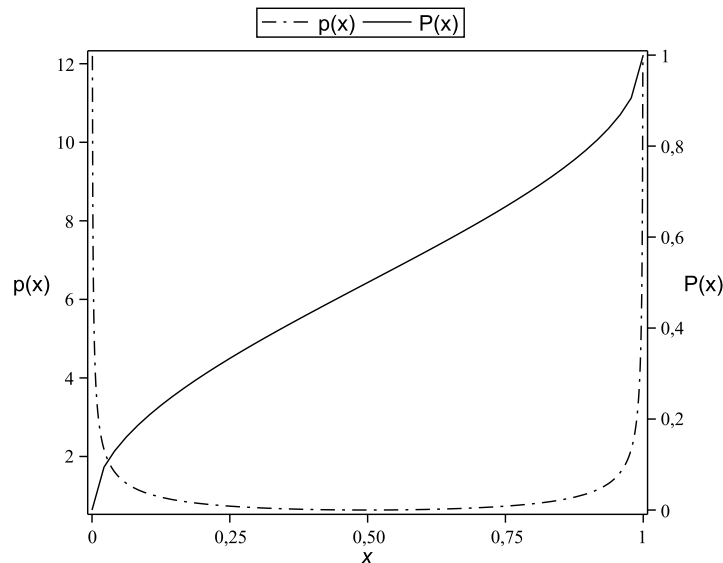
O objetivo é encontrar uma função adequada para ser utilizada na transformação não linear de fdp, Figura 2.2b. Para isso, primeiro, observou-se que a função densidade de probabilidade (fdp) de um tom senoidal (ALENCAR, 2009) é dada por

$$p(x) = \frac{b}{\pi\sqrt{V^2 - (x-a)^2}}, \quad |x-a| < V, \tag{2.3}$$

em que  $V$  é a amplitude máxima da senoide,  $a \in R$  e  $b \in R$ . Essa função densidade de probabilidade e sua função cumulativa estão representadas na Figura 2.3.

Ao se observar a Equação 2.3 e a Figura 2.3, verifica-se que:

- A curva da fdp tem um formato côncavo (‘U’);
- O termo  $V$  ajusta o tamanho da abertura do gráfico. Então, se o valor de  $V$  cresce o formato côncavo (‘U’) aumenta. Da mesma forma, se o valor de  $V$  diminui, o formato côncavo (‘U’) diminui;



**Figura 2.3** Função densidade de probabilidade  $p(x)$  e função cumulativa de probabilidade  $P(x)$ .

- O termo  $b$  ajusta a escala do gráfico no eixo  $p(x)$ . Ou seja, se o valor de  $b$  aumenta, a escala do gráfico de  $p(x)$  aumenta. Da mesma forma, se o valor de  $b$  diminui, a escala do gráfico de  $p(x)$  diminui;
- O parâmetro  $a$  ajusta o deslocamento do gráfico no eixo  $x$ . Se o valor de  $a$  for maior que zero o gráfico de  $p(x)$  se desloca para direita. Da mesma forma, se o valor de  $a$  for menor que zero o gráfico se desloca para esquerda.

O comportamento bimodal, descrito por Tafvelin (JOHN; TAFVELIN, 2007) e Rastin Pries (PRIES *et al.*, 2009), é semelhante ao comportamento da fdp,  $p(x)$ , apresentado na Figura 2.3. As medidas cumulativas do comprimento dos pacotes apresentada por Rastin Pries (PRIES *et al.*, 2009), têm o formato similar à FCP,  $P(x)$ , da Figura 2.3. Além disso, observa-se que

$$\frac{d}{dx} \left[ \arcsin \frac{u}{a} \right] = \frac{1}{\sqrt{a^2 - u^2}} \frac{du}{dx} \quad (2.4)$$

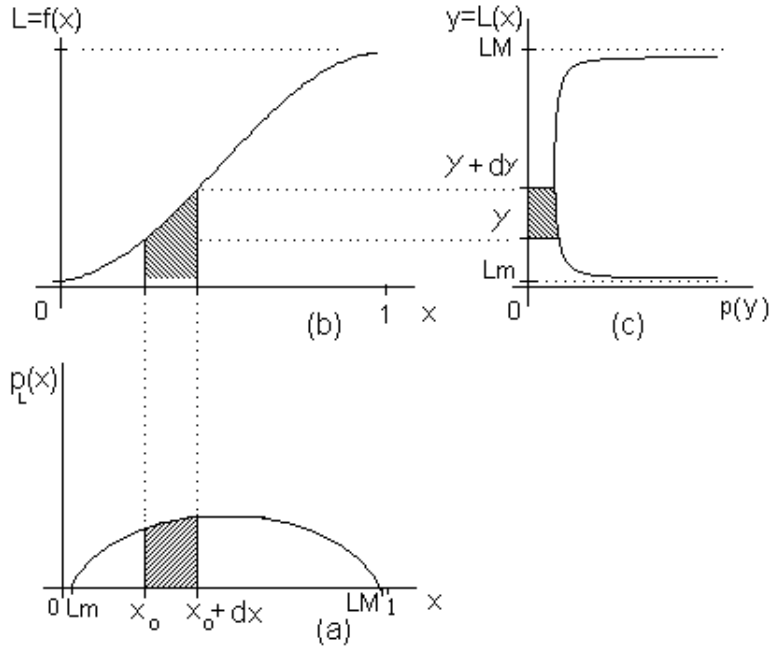
e

$$\frac{d}{dx} \left[ \arccos \frac{u}{a} \right] = \frac{-1}{\sqrt{a^2 - u^2}} \frac{du}{dx}. \quad (2.5)$$

Em outras palavras, significa que é possível encontrar uma equação similar à Equação (2.3), ou seja, usar uma equação senoidal (2.4) ou (2.5) na transformação não linear de fdp no ponto de agregação, ilustrado na Figura 2.2, para modelar a função densidade de probabilidade do comprimento dos pacotes.

A Figura 2.2 ilustra a transformação não linear de fdp uniforme e uma primeira equação de densidade de probabilidade foi obtida, com bons resultados de aproximação entre os valores medidos e o modelo da fdp (CASTRO *et al.*, 2010). No entanto, observou-se que melhores

resultados poderiam ser obtidos se fosse considerado o comportamento do usuário com uma distribuição não uniforme, Figura 2.4a. A seção seguinte descreve testes com diversas distribuições utilizando o Matlab (MATLAB, 2010) e o Maple (MAPLE, 2010), que levaram à distribuição Beta, a que melhor se adaptou ao modelo não uniforme considerado nesta tese, como ilustra a Figura 2.4.



**Figura 2.4** Transformação não linear de fdp não uniforme (área equivalente). (a) Densidade de probabilidade não uniforme, (b) Modelo do comprimento do pacote e (c) Função densidade de probabilidade.

Após sucessivas tentativas de ajustes, encontrou-se uma equação para a transformação não linear de fdp no ponto de agregação, ilustrado na Figura 2.4b, dada pela definição a seguir:

**Definição 4:** Considere que  $\ell(x)$ , ou apenas  $\ell$ , seja uma variável aleatória que represente o comprimento do pacote em número de *bits* ou *bytes* normalizados enviado por uma interface de rede qualquer em um intervalo de tempo  $t$  e que seja possível expressar  $\ell$  por

$$\ell = L_M - \left( \frac{L_M - L_m}{2} \right) \left[ \cos\left( \frac{\pi x}{n} \right) + 1 \right], \quad 0 \leq x \leq 1, \quad n \in N^*, \quad (2.6)$$

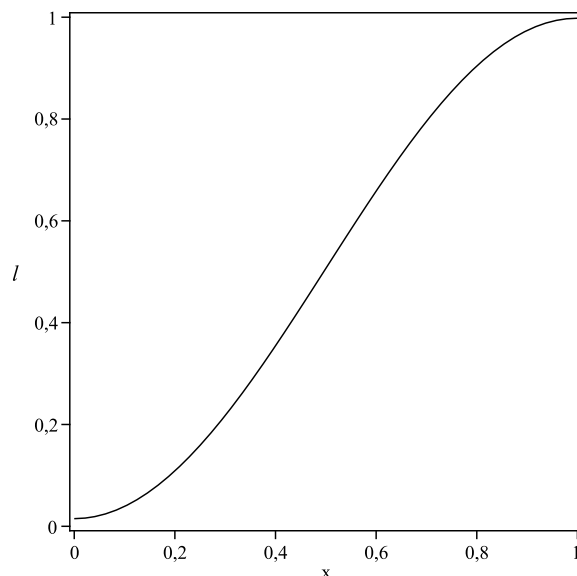
a Figura 2.5 apresenta o comprimento do pacote  $\ell$  em função da variável aleatória  $x$ , para  $n = 1$ ,  $L_m = 0.1$  e  $L_M = 0.9$ .

A partir da Equação (2.6), obtém-se a variável aleatória  $x$  em função do comprimento do pacote  $\ell$ , ou seja,

$$x(\ell) = \frac{n}{\pi} \arccos \left[ 2 \left( \frac{L_M - \ell}{L_M - L_m} \right) - 1 \right], \quad L_m \leq \ell \leq L_M, \quad (2.7)$$

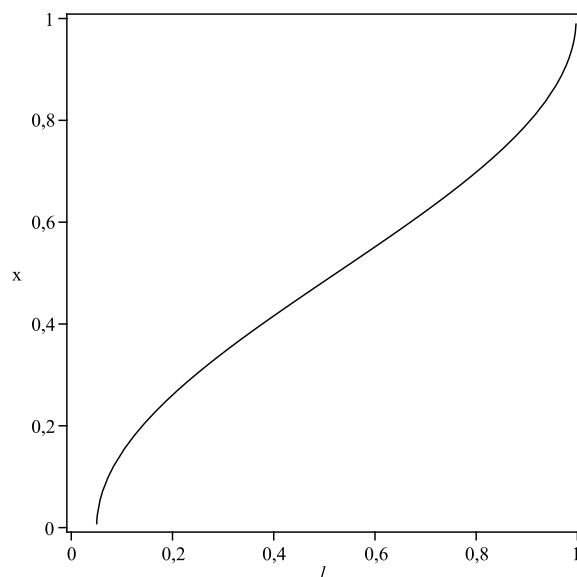
ou

$$x(\ell) = \frac{n}{\pi} \arccos \left[ \frac{L_M + L_m - 2\ell}{L_M - L_m} \right], \quad (2.8)$$



**Figura 2.5** Comprimento do pacote  $\ell$  em função da variável aleatória  $x$ .

a Figura 2.6 apresenta a variável aleatória  $x$  em função do comprimento do pacote  $\ell$ , para  $n = 1$ ,  $L_m = 0.1$  e  $L_M = 0.9$ .



**Figura 2.6** Variável aleatória  $x$  em função do comprimento do pacote  $\ell$ .

## 2.3 Modelo da função densidade de probabilidade

Nesta seção é descrito o procedimento para obtenção de uma fórmula para a função densidade de probabilidade. Para isso se faz uso do modelo de comprimento do pacote e do mapeamento biunívoco, ambos apresentados na seção anterior. Os dois termos apresentados na Equação (2.2), numerador e denominador, são obtidos nas duas subseções a seguir.

### 2.3.1 O denominador $\left|\frac{d\ell}{dx}\right|$

O primeiro termo obtido é o denominador da Equação (2.2),  $\left|\frac{d\ell}{dx}\right|$ . Para isso, inicia-se derivando a Equação (2.6) em relação a  $x$ , encontrando

$$\frac{d\ell}{dx} = -\left(\frac{L_M - \ell}{2}\right) \left[-\frac{\pi}{n} \operatorname{sen}\left(\frac{\pi x}{n}\right)\right], \quad (2.9)$$

$$\frac{d\ell}{dx} = \frac{\pi}{n} \left(\frac{L_M - \ell}{2}\right) \left[\operatorname{sen}\left(\frac{\pi x}{n}\right)\right]. \quad (2.10)$$

Mas

$$\begin{aligned} \operatorname{sen}^2 a + \cos^2 a &= 1, \\ \operatorname{sen}^2 a &= 1 - \cos^2 a, \\ \operatorname{sen}(a) &= \pm \sqrt{1 - \cos^2 a}. \end{aligned} \quad (2.11)$$

Logo,

$$\frac{d\ell}{dx} = \pm \frac{\pi}{n} \left(\frac{L_M - \ell}{2}\right) \sqrt{1 - \cos^2\left(\frac{\pi x}{n}\right)}. \quad (2.12)$$

Sabendo-se que a Equação (2.6), pode ser reescrita como

$$\left[2\left(\frac{L_M - \ell}{L_M - L_m}\right) - 1\right] = \cos\left(\frac{\pi x}{n}\right), \quad (2.13)$$

e substituindo esta equação em (2.12), leva a

$$\frac{d\ell}{dx} = \pm \frac{\pi}{n} \left(\frac{L_M - \ell}{2}\right) \sqrt{1 - \left[2\left(\frac{L_M - \ell}{L_M - L_m}\right) - 1\right]^2}, \quad (2.14)$$

ou

$$\frac{d\ell}{dx} = \pm \frac{\pi}{n} \sqrt{\left(\frac{L_M - L_m}{2}\right)^2 - \left[\ell - \left(\frac{L_M + L_m}{2}\right)\right]^2}, \quad (2.15)$$

E finalmente,

$$\left|\frac{d\ell}{dx}\right| = \left|\pm \frac{\pi}{n} \sqrt{\left(\frac{L_M - L_m}{2}\right)^2 - \left[\ell - \left(\frac{L_M + L_m}{2}\right)\right]^2}\right|, \quad (2.16)$$

ou

$$\frac{1}{\left|\frac{d\ell}{dx}\right|} = \frac{n}{\left|\pm \pi \sqrt{\left(\frac{L_M - L_m}{2}\right)^2 - \left[\ell - \left(\frac{L_M + L_m}{2}\right)\right]^2}\right|}, \quad (2.17)$$

como o denominador da Equação (2.17) é sempre maior que zero por causa do modulo e real para  $L_m < \ell < L_M$ , então

$$\left| \frac{d\ell}{dx} \right| = \frac{n}{\pi \sqrt{\left(\frac{L_M - L_m}{2}\right)^2 - \left[\ell - \left(\frac{L_M + L_m}{2}\right)\right]^2}}. \quad (2.18)$$

Comparando (2.18) com (2.3), verifica-se que o  $V = \frac{L_M - L_m}{2}$ ,  $b = n$ ,  $x = \ell$  e  $a = \frac{L_M + L_m}{2}$ .

### 2.3.2 O numerador $p_X(x)$

Para o termo  $p_X(x)$ , numerador da Equação (2.2), foram feitas diversas tentativas de ajuste utilizando diferentes distribuições de probabilidade. Entre as funções avaliadas, a distribuição que proporcionou o melhor ajuste da curva com os dados medidos, como ilustrado no mapeamento apresentado na Figura 2.4a, foi a distribuição Beta. Essa distribuição é uma família de distribuições de probabilidade definidas no intervalo  $(0, 1)$  parametrizado por dois termos de forma,  $\alpha$  e  $\beta$ .

A distribuição Beta é dada por

$$f(x, \alpha, \beta) = \left[ \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right] x^{\alpha-1} (1-x)^{\beta-1}, \quad (2.19)$$

ou

$$f(x, \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad (2.20)$$

em que,  $\Gamma(\cdot)$  é a função Gama (B) e  $B(\cdot)$  é a função Beta dada por

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx, \quad (2.21)$$

ou

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}. \quad (2.22)$$

A Figura 2.7 ilustra a distribuição Beta em função de  $x$  para diferentes valores de  $\alpha$  e  $\beta$ .

A função cumulativa é dada por

$$F(x, \alpha, \beta) = \frac{I_x(\alpha, \beta)}{B(\alpha, \beta)}, \quad (2.23)$$

em que

$$I_x(\alpha, \beta) = \int_0^c x^{\alpha-1} (1-x)^{\beta-1} dx, \quad (2.24)$$

com  $0 \leq c \leq 1$ ,  $I_x(\alpha, \beta)$  é chamada de função Beta incompleta normalizada,  $I_0(\alpha, \beta) = 0$  e  $I_1(\alpha, \beta) = 1$ .

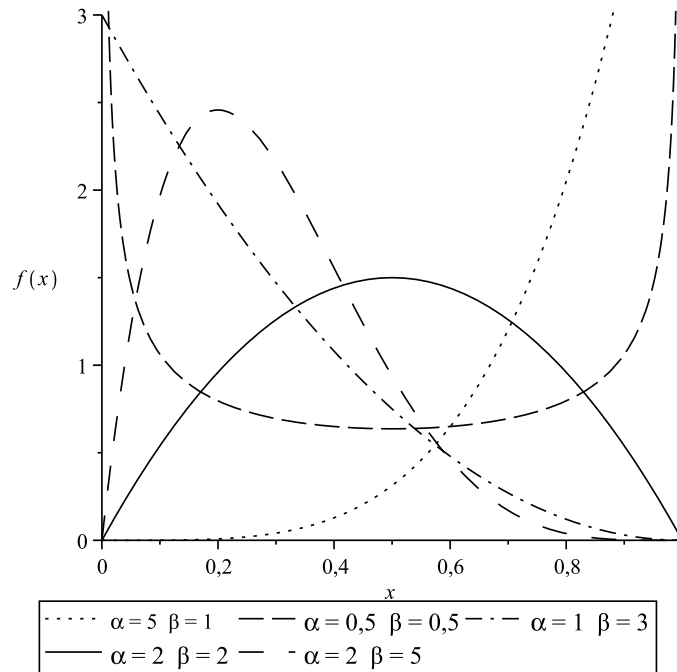


Figura 2.7 Densidade de probabilidade da distribuição Beta para diferentes valores de  $\alpha$  e  $\beta$ .

### 2.3.3 Função Densidade de Probabilidade

A função densidade de probabilidade é obtida substituindo as Equações (2.19) e (2.17) na Equação (2.2),

$$p_L(\ell) = \frac{\left[ \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right] x^{\alpha-1} (1-x)^{\beta-1}}{\frac{\pi}{n} \sqrt{\left(\frac{L_M - L_m}{2}\right)^2 - \left[\ell - \left(\frac{L_M + L_m}{2}\right)\right]^2}}, \quad (2.25)$$

substituindo a Equação (2.7) em (2.25), chega-se à função densidade de probabilidade do comprimento do pacote

$$p_L(\ell) = \frac{n}{\pi \sqrt{\left(\frac{L_M - L_m}{2}\right)^2 - \left[\ell - \left(\frac{L_M + L_m}{2}\right)\right]^2}} \cdot \left[ \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right] \cdot \left\{ \frac{n}{\pi} \arccos \left[ 2 \left( \frac{L_M - \ell}{L_M - L_m} \right) - 1 \right] \right\}^{\alpha-1} \cdot \left\{ 1 - \frac{n}{\pi} \arccos \left[ 2 \left( \frac{L_M - \ell}{L_M - L_m} \right) - 1 \right] \right\}^{\beta-1} \quad (2.26)$$

em que,  $n \in \mathbb{N}^*$ ,  $\Gamma(\cdot)$  é a função Gama,  $\ell$  é o comprimento do pacote,  $\alpha$  e  $\beta$  são parâmetros da distribuição relacionados ao tipo de tráfego.

Analisando a Equação (2.26) apresentada, tem-se:

- Para  $L_m < \ell < L_M$ , os comprimentos dos pacotes são maiores que o comprimento mínimo ( $L_m$ ) e menores que o comprimento máximo ( $L_M$ ). Este intervalo é uma restrição para os valores do comprimento dos pacotes, resultado coerente com a prática e dentro da validade da equação.
- Para  $\ell = L_m$ , a raiz do denominador igual a zero e no numerador aparece  $0^{\alpha-1}$ . Esse valor tende para  $\infty$  quando  $0 < \alpha < 1$  e este é o primeiro ponto de indeterminação da Equação (2.26).
- Para  $\ell = L_M$ , a raiz do denominador igual a zero e no numerador tem-se  $0^{\beta-1}$ . Esse valor tende para  $\infty$  quando  $0 < \beta < 1$  e este é o segundo ponto de indeterminação da Equação (2.26).
- Para  $\ell > L_M$  ou  $\ell < L_m$ ,  $p_L(\ell)$ , apresenta valores complexos, o que esta fora da situação estudada aqui.
- Para que a função dada Equação (2.26), seja considerada uma densidade de probabilidade, é preciso que

$$\int_{-\infty}^{\infty} p(\ell) d\ell = 1, \quad (2.27)$$

porém,  $p_L(\ell) \neq 0$  apenas para o comprimento normalizado  $0 \leq \ell \leq 1$ . Logo,

$$\int_0^1 p_L(\ell) d\ell = 1. \quad (2.28)$$

No entanto, pelos comentários feitos nos itens anteriores sobre o intervalo de validade da Equação (2.26),  $p_L(\ell)$  é definida no intervalo  $L_m < \ell < L_M$ . Então, os limites da integral são de  $\ell = L_m + \varepsilon$  a  $\ell = L_M - \varepsilon$ , com  $\varepsilon \rightarrow 0$ . Logo,

$$\int_{L_m+\varepsilon}^{L_M-\varepsilon} p_L(\ell) d\ell = 1. \quad (2.29)$$

Para esses limites de integração, tem-se que  $\ell \rightarrow L_m$  e  $\ell \rightarrow L_M$ , quando  $\varepsilon \rightarrow 0$  na Equação (2.6) e que  $x \rightarrow 0$  e  $x \rightarrow 1$  na Equação (2.7), também para  $\varepsilon \rightarrow 0$ .

Substituindo a Equação (2.26) em (2.29),

$$\int_{L_m+\varepsilon}^{L_M-\varepsilon} \frac{n}{\pi \sqrt{\left(\frac{L_M-L_m}{2}\right)^2 - \left[\ell - \left(\frac{L_M+L_m}{2}\right)\right]^2}} \cdot \left[ \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right] \cdot \left\{ \frac{n}{\pi} \arccos \left[ 2 \left( \frac{L_M - \ell}{L_M - L_m} \right) - 1 \right] \right\}^{\alpha-1} \cdot \left\{ 1 - \frac{n}{\pi} \arccos \left[ 2 \left( \frac{L_M - \ell}{L_M - L_m} \right) - 1 \right] \right\}^{\beta-1} d\ell = 1. \quad (2.30)$$



Sabendo que  $x$  é dado pela Equação (2.7), reescrita a seguir,

$$x(\ell) = \frac{n}{\pi} \arccos \left[ 2 \left( \frac{L_M - \ell}{L_M - L_m} \right) - 1 \right] \quad (2.31)$$

e que

$$\frac{d}{dx} \arccos \left( \frac{u}{a} \right) = - \frac{1}{\sqrt{a^2 - u^2}} \frac{du}{dx}, \quad (2.32)$$

então

$$dx = \frac{n}{\pi \sqrt{\left( \frac{L_M - L_m}{2} \right)^2 - \left[ \ell - \left( \frac{L_M + L_m}{2} \right) \right]^2}} d\ell. \quad (2.33)$$

Logo, substituindo (2.31), (2.33) na integral da Equação (2.30) e considerando o limite para  $\varepsilon \rightarrow 0$  que leva a  $x \rightarrow 0$  e  $x \rightarrow 1$  nos limites da integração, então

$$\int_0^1 \left[ \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right] x^{\alpha-1} (1-x)^{\beta-1} dx = 1, \quad (2.34)$$

e retirando os termos da função gama de dentro do integrando

$$\left[ \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right] \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx = 1. \quad (2.35)$$

Substituindo a Equação (2.21) na Equação (2.35), encontra-se

$$\left[ \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right] B(\alpha, \beta) = 1. \quad (2.36)$$

Usando, agora, a Equação (2.22) na Equação (2.36),

$$\left[ \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right] \left[ \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \right] = 1, \quad (2.37)$$

como desejado. Assim, a Equação (2.26) é uma distribuição de probabilidade válida.

## 2.4 Cálculo da Função Cumulativa de Probabilidade

Nesta seção é calculada a função de distribuição cumulativa. Para isso, faz-se uso da função densidade de probabilidade obtida na seção anterior.

### 2.4.1 Função cumulativa de probabilidade

A função cumulativa é dada por

$$P_L(\ell) = \int_{-\infty}^{\ell} p(l)dl, \quad (2.38)$$

com  $x$  dado pela Equação (2.7), reescrita a seguir,

$$x(\ell) = \frac{n}{\pi} \arccos \left[ 2 \left( \frac{L_M - \ell}{L_M - L_m} \right) - 1 \right]. \quad (2.39)$$

Para o cálculo da Equação (2.38), repete-se o procedimento da seção anterior até a Equação (2.34). Em seguida, inicia-se com a Equação (2.35), mas para um intervalo de integração de 0 a  $\ell$  com  $0 \leq \ell \leq 1$ , logo

$$P_L(\ell) = \left[ \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right] \int_0^{\ell} x^{\alpha-1} (1-x)^{\beta-1} dx. \quad (2.40)$$

Mas sabendo que a função Beta incompleta normalizada é

$$I_x(\alpha, \beta) = \frac{1}{B(\alpha, \beta)} \int_0^{\ell} x^{\alpha-1} (1-x)^{\beta-1} dx, \quad (2.41)$$

ou ainda

$$I_x(\alpha, \beta)B(\alpha, \beta) = \int_0^{\ell} x^{\alpha-1} (1-x)^{\beta-1} dx. \quad (2.42)$$

Assim, a Equação (2.40) pode ser reescrita como

$$P_L(\ell) = \left[ \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right] I_x(\alpha, \beta)B(\alpha, \beta). \quad (2.43)$$

Usando a Equação (2.22) em (2.43)

$$P_L(\ell) = \left[ \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right] I_x(\alpha, \beta) \left[ \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \right], \quad (2.44)$$

ou

$$P_L(\ell) = I_x(\alpha, \beta). \quad (2.45)$$

Ou seja, para se encontrar a função de distribuição cumulativa do comprimento dos pacotes de uma rede de computadores com tráfego bimodal, basta calcular a função Beta incompleta normalizada dada por

$$I_x(\alpha, \beta) = \frac{1}{B(\alpha, \beta)} \int_0^\ell x^{\alpha-1} (1-x)^{\beta-1} dx, \quad (2.46)$$

em que  $x = x(\ell)$  é dado pela Equação (2.47) e  $I_x(\alpha, \beta)$  possui as propriedades de que  $I_0(\alpha, \beta) = 0$  e  $I_1(\alpha, \beta) = 1$ . E  $\alpha$  e  $\beta$  são parâmetros reguladores da equação.

$$x(\ell) = \frac{n}{\pi} \arccos \left[ 2 \left( \frac{L_M - \ell}{L_M - L_m} \right) - 1 \right]. \quad (2.47)$$

## 2.5 Primeiro Momento, Segundo Momento, Variância e Desvio Padrão

Nesta seção são apresentadas as equações do primeiro momento, segundo momento, variância e desvio padrão, a partir dos resultados obtidos na seção anterior.

### 2.5.1 Primeiro momento

O primeiro momento pode ser obtido usando a equação a seguir

$$E[L] = \int_{-\infty}^{\infty} \ell p_L(\ell) d\ell, \quad (2.48)$$

mas  $p_L(\ell) \neq 0$  apenas para  $0 \leq \ell \leq 1$ . Logo,

$$E[L] = \int_0^1 \ell p_L(\ell) d\ell. \quad (2.49)$$

substituindo (2.6) em (2.49) e respeitando os limites de integração, semelhante ao procedimento usado para se obter (2.29) e (2.34), encontra-se

$$E[L] = \lim_{\varepsilon \rightarrow 0} \int_{L_m+\varepsilon}^{L_M-\varepsilon} \left\{ L_M - \left( \frac{L_M - L_m}{2} \right) \left[ \cos \left( \frac{\pi x}{n} \right) + 1 \right] \right\} p_L(\ell) d\ell \quad (2.50)$$

ou ainda,

$$E[L] = L_M - \left( \frac{L_M - L_m}{2} \right) - \left( \frac{L_M - L_m}{2} \right) \lim_{\varepsilon \rightarrow 0} \int_{L_m+\varepsilon}^{L_M-\varepsilon} \left[ \cos \left( \frac{\pi x}{n} \right) \right] p_L(\ell) d\ell. \quad (2.51)$$

No entanto,

$$\lim_{\varepsilon \rightarrow 0} \int_{L_m+\varepsilon}^{L_M-\varepsilon} \left[ \cos \left( \frac{\pi x}{n} \right) \right] p_L(\ell) d\ell \simeq \left[ \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right] \int_0^1 \cos(\pi x) x^{\alpha-1} (1-x)^\beta dx, \quad (2.52)$$

para  $n = 1$ .

A tabela de integrais (GRADSHTEYN; RYZHIK, 2007) revela que

$$\int_0^1 \cos(\pi x) x^{\alpha-1} (1-x)^{\beta-1} dx = \frac{1}{2} B(\alpha, \beta) \left[ {}_1F_1(\alpha; \alpha + \beta; j\pi) + {}_1F_1(\alpha; \alpha + \beta; -j\pi) \right] \quad (2.53)$$

em que,  ${}_1F_1(a; a + b; c)$  com  $a = \alpha$ ,  $b = \beta$  e  $c = \pm j\pi$ , é a função hypergeometrica degenerada ou função hypergeometrica confluyente, que pode ser calculada por

$${}_1F_1(a; a + b; c) = \sum_{n=0}^{\infty} \frac{(a)_n c^n}{(b)_n n!}, \quad (2.54)$$

com  $(a)_n = a(a+1)(a+2)\cdots(a+n-1)$  e  $(b)_n = b(b+1)(b+2)\cdots(b+n-1)$ , para  $a > 0$  e  $b > 0$ . Assim, o primeiro momento pode ser calculado por

$$E[L] = L_M - \left( \frac{L_M - L_m}{2} \right) - \left( \frac{L_M - L_m}{2} \right) \left( \frac{1}{2} \right) B(\alpha, \beta) \left[ \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right] \cdot \left[ {}_1F_1(\alpha; \alpha + \beta; j\pi) + {}_1F_1(\alpha; \alpha + \beta; -j\pi) \right], \quad (2.55)$$

mas, usando a Equação (2.22) em (2.55), chega-se a

$$E[L] = \left( \frac{1}{2} \right) (L_M + L_m) - \left( \frac{1}{4} \right) (L_M - L_m) \left[ {}_1F_1(\alpha; \alpha + \beta; j\pi) + {}_1F_1(\alpha; \alpha + \beta; -j\pi) \right]. \quad (2.56)$$

Consultando a tabela no sítio de Wolfram (WOLFRAM, 2001), encontra-se

$${}_1F_1(t; u; v) = \frac{\Gamma(1-t)\Gamma(u)}{\Gamma(u-t)} L_{-t}^{u-1}(v), \quad (2.57)$$

em que  $L_n^k(y)$  é a função de Laguerre, que pode ser calculada por

$$L_n^k(y) = \sum_{m=0}^n (-1)^m \frac{(n+k)!}{(n-m)!(k+m)!m!} y^m. \quad (2.58)$$

Então, outra forma de representar (2.56) é

$$E[L] = \left( \frac{1}{2} \right) (L_M + L_m) - \left( \frac{1}{4} \right) (L_M - L_m) \left[ \frac{\Gamma(1-\alpha)\Gamma(\alpha + \beta)}{\Gamma(\beta)} \right] \cdot \left[ L_{-\alpha}^{\alpha+\beta-1}(j\pi) + L_{-\alpha}^{\alpha+\beta-1}(-j\pi) \right], \quad (2.59)$$

que é uma função mais simples para o tratamento computacional que (2.56). Para representar (2.59) em uma forma mais compacta, basta considerar

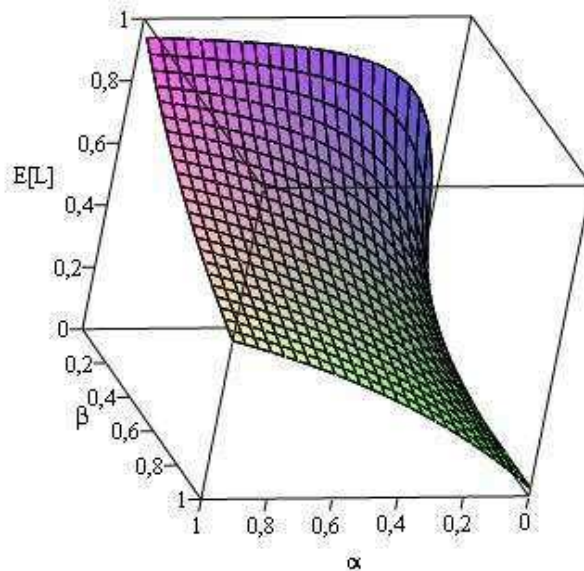
$$A_0 = \left[ \frac{\Gamma(1-\alpha)\Gamma(\alpha+\beta)}{\Gamma(\beta)} \right], \tag{2.60}$$

$$A_1 = L_{-\alpha}^{\alpha+\beta-1}(j\pi), \tag{2.61}$$

e

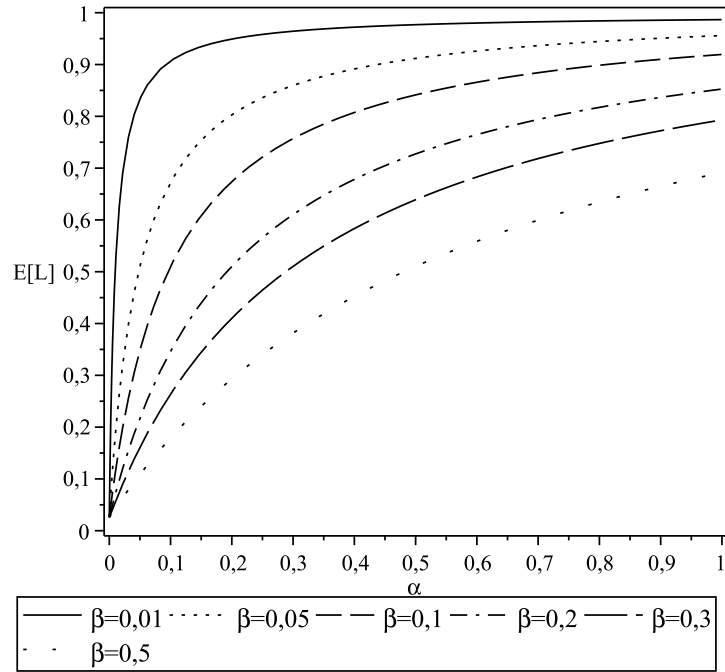
$$A_2 = L_{-\alpha}^{\alpha+\beta-1}(-j\pi), \tag{2.62}$$

$$E[L] = \left(\frac{1}{2}\right)(L_M + L_m) - \left(\frac{1}{4}\right)(L_M - L_m)A_0[A_1 + A_2]. \tag{2.63}$$

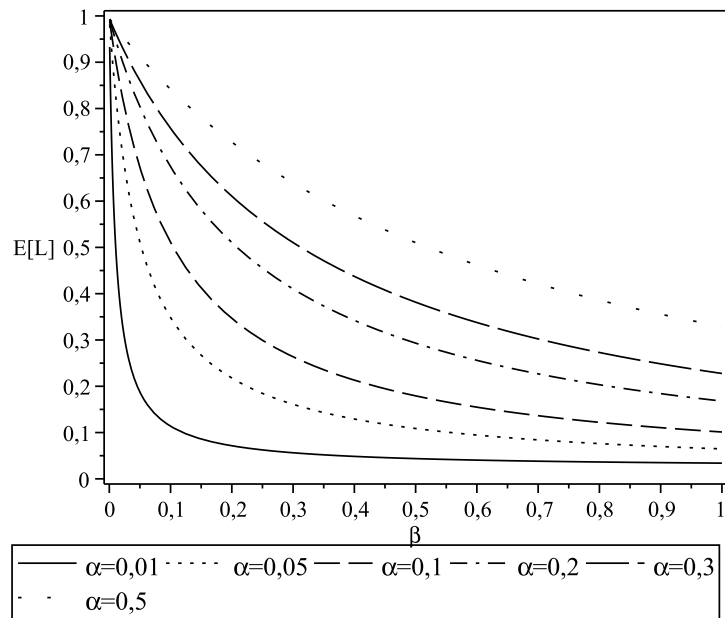


**Figura 2.8** Superfície de  $E[L]$  em função de  $\alpha$  e  $\beta$ .

As Figuras 2.8, 2.9 e 2.10, ilustram a variação de  $E[L]$ , Equação (2.59), em função de  $\alpha$  e  $\beta$ . A primeira figura é uma superfície que ilustra o comportamento da média em função de  $\alpha$  e  $\beta$ . Verifica-se que a média aumenta quando ocorre um aumento do valor do parâmetro  $\alpha$ . Esse comportamento é observado no gráfico da variação de  $E[L]$  em função de  $\alpha$ , Figura 2.9, para diferentes valores de  $\beta$ . Fenômeno inverso corre com o valor do parâmetro  $\beta$ , a média diminui quando o valor de  $\beta$  cresce e aumenta quando o valor de  $\beta$  diminui. A Figura 2.10 ilustra esta variação de  $E[L]$  em função de  $\beta$  para alguns valores de  $\alpha$ .



**Figura 2.9** Gráfico de  $E[L]$  em função de  $\alpha$ .



**Figura 2.10** Gráfico de  $E[L]$  em função de  $\beta$ .

## 2.5.2 Segundo momento

O segundo momento é obtido usando a equação

$$E[L^2] = \int_{-\infty}^{\infty} \ell^2 p_L(\ell) d\ell, \quad (2.64)$$

ou

$$E[L^2] = \int_0^1 \ell^2 p_L(\ell) d\ell, \quad (2.65)$$

substituindo (2.6) com  $n = 1$  em (2.64) e respeitando os limites de integração,

$$E[L^2] = \lim_{\varepsilon \rightarrow 0} \int_{L_m + \varepsilon}^{L_M - \varepsilon} \left\{ L_M - \left( \frac{L_M - L_m}{2} \right) [\cos(\pi x) + 1] \right\}^2 p_L(\ell) d\ell. \quad (2.66)$$

Chamando o integrando de  $I_1$ , então

$$I_1 = \left\{ L_M - \left( \frac{L_M - L_m}{2} \right) [\cos(\pi x) + 1] \right\}^2 = L_M^2 - 2L_M \left( \frac{L_M - L_m}{2} \right) [\cos(\pi x) + 1] + \left( \frac{L_M - L_m}{2} \right)^2 [\cos(\pi x) + 1]^2, \quad (2.67)$$

$$I_1 = L_M^2 - 2L_M \left( \frac{L_M - L_m}{2} \right) \cos(\pi x) - 2L_M \left( \frac{L_M - L_m}{2} \right) + \left( \frac{L_M - L_m}{2} \right)^2 [\cos^2(\pi x) + 2\cos(\pi x) + 1], \quad (2.68)$$

$$I_1 = L_M^2 - 2L_M \left( \frac{L_M - L_m}{2} \right) \cos(\pi x) - 2L_M \left( \frac{L_M - L_m}{2} \right) + \left( \frac{L_M - L_m}{2} \right)^2 \cos^2(\pi x) + 2 \left( \frac{L_M - L_m}{2} \right)^2 \cos(\pi x) + \left( \frac{L_M - L_m}{2} \right)^2, \quad (2.69)$$

relembrando que

$$\begin{aligned} \cos^2 a &= \frac{1}{2} [\cos(2a) + 1], \\ \cos^2 a &= \frac{1}{2} \cos(2a) + \frac{1}{2}, \end{aligned} \quad (2.70)$$

então

$$I_1 = L_M^2 - 2L_M \left( \frac{L_M - L_m}{2} \right) \cos(\pi x) - 2L_M \left( \frac{L_M - L_m}{2} \right) + \left( \frac{L_M - L_m}{2} \right)^2 \left[ \left( \frac{1}{2} \right) \cos(2\pi x) + \left( \frac{1}{2} \right) \right] + 2 \left( \frac{L_M - L_m}{2} \right)^2 \cos(\pi x) + \left( \frac{L_M - L_m}{2} \right)^2, \quad (2.71)$$

$$I_1 = L_M^2 - 2L_M \left( \frac{L_M - L_m}{2} \right) \cos(\pi x) - 2L_M \left( \frac{L_M - L_m}{2} \right) + \left( \frac{L_M - L_m}{2} \right)^2 \left( \frac{1}{2} \right) \cos(2\pi x) + \left( \frac{L_M - L_m}{2} \right)^2 \left( \frac{1}{2} \right) + 2 \left( \frac{L_M - L_m}{2} \right)^2 \cos(\pi x) + \left( \frac{L_M - L_m}{2} \right)^2, \quad (2.72)$$

$$I_1 = L_M^2 + \left( \frac{3}{2} \right) \left( \frac{L_M - L_m}{2} \right)^2 - 2L_M \left( \frac{L_M - L_m}{2} \right) + \left[ 2 \left( \frac{L_M - L_m}{2} \right)^2 - 2L_M \left( \frac{L_M - L_m}{2} \right) \right] \cos(\pi x) + \left( \frac{L_M - L_m}{2} \right)^2 \left( \frac{1}{2} \right) \cos(2\pi x), \quad (2.73)$$

$$I_1 = \left( \frac{3}{8} \right) \left( L_M^2 + \frac{2}{3} L_M L_m + L_m^2 \right) - \left( \frac{1}{2} \right) \left( L_M^2 - L_m^2 \right) \cos(\pi x) + \left( \frac{1}{8} \right) \left( L_M - L_m \right)^2 \cos(2\pi x). \quad (2.74)$$

Substituindo (2.74) em (2.50),

$$E[L^2] = \left( \frac{3}{8} \right) \left( L_M^2 + \frac{2}{3} L_M L_m + L_m^2 \right) \lim_{\varepsilon \rightarrow 0} \int_{L_m + \varepsilon}^{L_M - \varepsilon} p_L(\ell) d\ell - \left( \frac{1}{2} \right) \left( L_M^2 - L_m^2 \right) \lim_{\varepsilon \rightarrow 0} \int_{L_m + \varepsilon}^{L_M - \varepsilon} \cos(\pi x) p_L(\ell) d\ell + \left( \frac{1}{8} \right) \left( L_M - L_m \right)^2 \lim_{\varepsilon \rightarrow 0} \int_{L_m + \varepsilon}^{L_M - \varepsilon} \cos(2\pi x) p_L(\ell) d\ell. \quad (2.75)$$

A primeira integral da Equação (2.75) é aproximadamente igual a 1 para  $\varepsilon \rightarrow 0$ , como visto anteriormente. Usando a Equação (2.53), pode-se calcular as outras duas integrais. Assim, a Equação (2.75), fica



$$\begin{aligned}
 E[L^2] &= \left(\frac{3}{8}\right) \left(L_M^2 + \frac{2}{3}L_M L_m + L_m^2\right) - \\
 &\quad - \left(\frac{1}{4}\right) \left(L_M^2 - L_m^2\right) \left[ {}_1F_1(\alpha; \alpha + \beta; j\pi) + {}_1F_1(\alpha; \alpha + \beta; -j\pi) \right] + \\
 &\quad + \left(\frac{1}{16}\right) \left(L_M - L_m\right)^2 \left[ {}_1F_1(\alpha; \alpha + \beta; j2\pi) + {}_1F_1(\alpha; \alpha + \beta; -j2\pi) \right], \tag{2.76}
 \end{aligned}$$

ou usando a função de Laguerre (WOLFRAM, 2001),

$$\begin{aligned}
 E[L^2] &= \left(\frac{3}{8}\right) \left(L_M^2 + \frac{2}{3}L_M L_m + L_m^2\right) - \\
 &\quad - \left(\frac{1}{4}\right) \left(L_M^2 - L_m^2\right) \left[ \frac{\Gamma(1 - \alpha)\Gamma(\alpha + \beta)}{\Gamma(\beta)} \right] \left[ L_{-\alpha}^{\alpha+\beta-1}(j\pi) + L_{-\alpha}^{\alpha+\beta-1}(-j\pi) \right] + \\
 &\quad + \left(\frac{1}{16}\right) \left(L_M - L_m\right)^2 \left[ \frac{\Gamma(1 - \alpha)\Gamma(\alpha + \beta)}{\Gamma(\beta)} \right] \left[ L_{-\alpha}^{\alpha+\beta-1}(j2\pi) + L_{-\alpha}^{\alpha+\beta-1}(-j2\pi) \right]. \tag{2.77}
 \end{aligned}$$

Semelhante ao caso do primeiro momento, (2.77) pode ser representada em uma forma mais compacta se for considerado

$$A_0 = \left[ \frac{\Gamma(1 - \alpha)\Gamma(\alpha + \beta)}{\Gamma(\beta)} \right], \tag{2.78}$$

$$A_1 = L_{-\alpha}^{\alpha+\beta-1}(j\pi), \tag{2.79}$$

$$A_2 = L_{-\alpha}^{\alpha+\beta-1}(-j\pi), \tag{2.80}$$

$$A_3 = L_{-\alpha}^{\alpha+\beta-1}(j2\pi), \tag{2.81}$$

e

$$A_4 = L_{-\alpha}^{\alpha+\beta-1}(-j2\pi), \tag{2.82}$$

que leva a

$$\begin{aligned}
 E[L^2] &= \left(\frac{3}{8}\right) \left(L_M^2 + \frac{2}{3}L_M L_m + L_m^2\right) - \left(\frac{1}{4}\right) \left(L_M^2 - L_m^2\right) A_0(A_1 + A_2) + \\
 &\quad + \left(\frac{1}{16}\right) \left(L_M - L_m\right)^2 A_0(A_3 + A_4). \tag{2.83}
 \end{aligned}$$

### 2.5.3 Variância

A variância é obtida em função das equações do primeiro e segundo momentos, usando a equação,

$$\text{Var}[L] = E[L^2] - (E[L])^2. \quad (2.84)$$

Então, recorrendo às Equações (2.56) e (2.76). E considerando

$C_0 = {}_1F_1(\alpha; \alpha + \beta; j\pi)$ ,  $C_1 = {}_1F_1(\alpha; \alpha + \beta; -j\pi)$ ,  $D_0 = {}_1F_1(\alpha; \alpha + \beta; j2\pi)$  e  $D_1 = {}_1F_1(\alpha; \alpha + \beta; -j2\pi)$ , tem-se

$$E[L^2] = \left(\frac{3}{8}\right) \left(L_M^2 + \frac{2}{3}L_M L_m + L_m^2\right) - \left(\frac{1}{4}\right) (L_M^2 - L_m^2) [C_0 + C_1] + \left(\frac{1}{16}\right) (L_M - L_m)^2 [D_0 + D_1]. \quad (2.85)$$

e

$$(E[L])^2 = \left\{ \left(\frac{1}{2}\right) (L_M + L_m) - \left(\frac{1}{4}\right) (L_M - L_m) [C_0 + C_1] \right\}^2, \quad (2.86)$$

$$(E[L])^2 = \left(\frac{1}{2}\right)^2 (L_M + L_m)^2 - \left(\frac{1}{4}\right) (L_M + L_m) (L_M - L_m) [C_0 + C_1] + \left(\frac{1}{4}\right)^2 (L_M - L_m)^2 [C_0 + C_1]^2, \quad (2.87)$$

$$(E[L])^2 = \left(\frac{1}{2}\right)^2 (L_M + L_m)^2 - \left(\frac{1}{4}\right) (L_M^2 - L_m^2) [C_0 + C_1] + \left(\frac{1}{4}\right)^2 (L_M - L_m)^2 [C_0 + C_1]^2. \quad (2.88)$$

Fazendo a substituição de (2.85) e (2.88) em (2.84), chega-se a

$$\begin{aligned} \text{Var}[L] = E[L^2] - (E[L])^2 = & \left(\frac{3}{8}\right) \left(L_M^2 + \frac{2}{3}L_M L_m + L_m^2\right) + \left(\frac{1}{4}\right) (L_M^2 - L_m^2) [C_0 + C_1] \\ & + \left(\frac{1}{16}\right) (L_M - L_m)^2 [D_0 + D_1] - \left(\frac{1}{2}\right)^2 (L_M + L_m)^2 + \left(\frac{1}{4}\right) (L_M^2 - L_m^2) [C_0 + C_1] \\ & - \left(\frac{1}{4}\right)^2 (L_M - L_m)^2 [C_0 + C_1]^2, \quad (2.89) \end{aligned}$$

ou

$$\begin{aligned} \text{Var}[L] = & \left(\frac{1}{8}\right) (L_M - L_m)^2 + \left(\frac{1}{2}\right) (L_M^2 - L_m^2) [C_0 + C_1] \\ & + \left(\frac{1}{16}\right) (L_M - L_m)^2 \left\{ [D_0 + D_1] + [C_0 + C_1]^2 \right\}^2, \end{aligned} \quad (2.90)$$

e de forma mais simplificada

$$\text{Var}[L] = \left(\frac{1}{2}\right) (L_M^2 - L_m^2) [C_0 + C_1] + \left(\frac{1}{16}\right) (L_M - L_m)^2 \left\{ 2 + [D_0 + D_1] + [C_0 + C_1]^2 \right\}. \quad (2.91)$$

E finalmente, substituindo os valores de  $C_0$ ,  $C_1$ ,  $D_0$  e  $D_1$ ,

$$\begin{aligned} \text{Var}[L] = & \left(\frac{1}{2}\right) (L_M^2 - L_m^2) [{}_1F_1(\alpha; \alpha + \beta; j\pi) + {}_1F_1(\alpha; \alpha + \beta; -j\pi)] + \\ & + \left(\frac{1}{16}\right) (L_M - L_m)^2 \left\{ 2 + [{}_1F_1(\alpha; \alpha + \beta; j2\pi) + \right. \\ & \left. + {}_1F_1(\alpha; \alpha + \beta; -j2\pi)] + [{}_1F_1(\alpha; \alpha + \beta; j\pi) + {}_1F_1(\alpha; \alpha + \beta; -j\pi)]^2 \right\}, \end{aligned} \quad (2.92)$$

ou, usando a função de Laguerre (WOLFRAM, 2001),

$$\begin{aligned} \text{Var}[L] = & \left(\frac{1}{2}\right) (L_M^2 - L_m^2) \left[ \frac{\Gamma(1 - \alpha)\Gamma(\alpha + \beta)}{\Gamma(\beta)} \right] \left[ L_{-\alpha}^{\alpha + \beta - 1}(j\pi) + L_{-\alpha}^{\alpha + \beta - 1}(-j\pi) \right] + \\ & + \left(\frac{1}{16}\right) (L_M - L_m)^2 \left\{ 2 + \left[ \frac{\Gamma(1 - \alpha)\Gamma(\alpha + \beta)}{\Gamma(\beta)} \right] \left[ L_{-\alpha}^{\alpha + \beta - 1}(j2\pi) + L_{-\alpha}^{\alpha + \beta - 1}(-j2\pi) \right] + \right. \\ & \left. + \left[ \frac{\Gamma(1 - \alpha)\Gamma(\alpha + \beta)}{\Gamma(\beta)} \right]^2 \left[ L_{-\alpha}^{\alpha + \beta - 1}(j\pi) + L_{-\alpha}^{\alpha + \beta - 1}(-j\pi) \right]^2 \right\}. \end{aligned} \quad (2.93)$$

Representando (2.93) de forma mais compacta, basta considerar

$$A_0 = \left[ \frac{\Gamma(1 - \alpha)\Gamma(\alpha + \beta)}{\Gamma(\beta)} \right], \quad (2.94)$$

$$A_1 = L_{-\alpha}^{\alpha + \beta - 1}(j\pi), \quad (2.95)$$

$$A_2 = L_{-\alpha}^{\alpha + \beta - 1}(-j\pi), \quad (2.96)$$

$$A_3 = L_{-\alpha}^{\alpha + \beta - 1}(j2\pi), \quad (2.97)$$

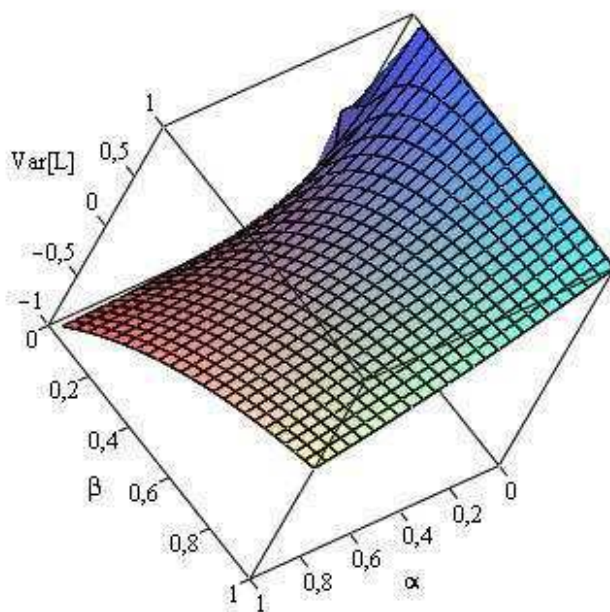
e

$$A_4 = L_{-\alpha}^{\alpha+\beta-1}(-j2\pi), \quad (2.98)$$

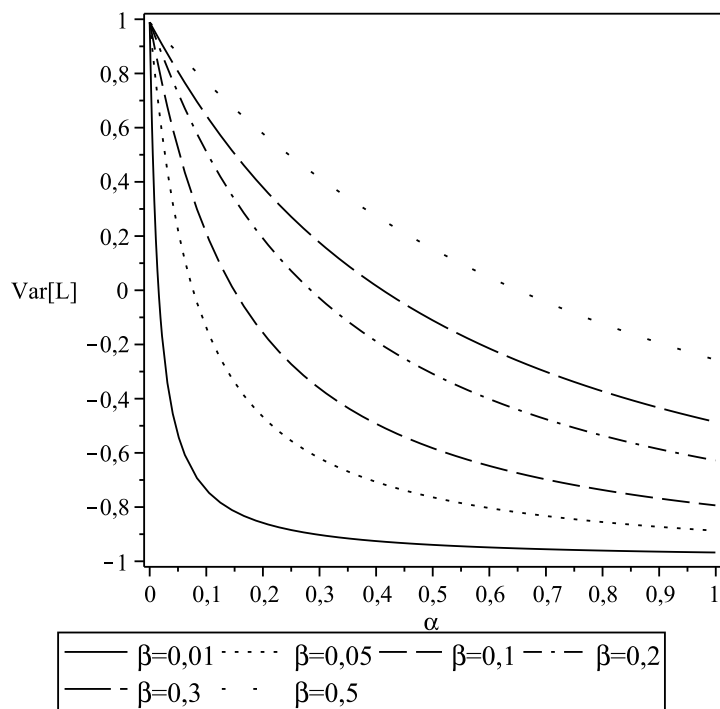
logo,

$$\begin{aligned} \text{Var}[L] = & \left(\frac{1}{2}\right) \left(L_M^2 - L_m^2\right) A_0(A_1 + A_2) + \\ & + \left(\frac{1}{16}\right) (L_M - L_m)^2 \left[2 + A_0(A_3 + A_4) + (A_0)^2(A_1 + A_2)^2\right]. \end{aligned} \quad (2.99)$$

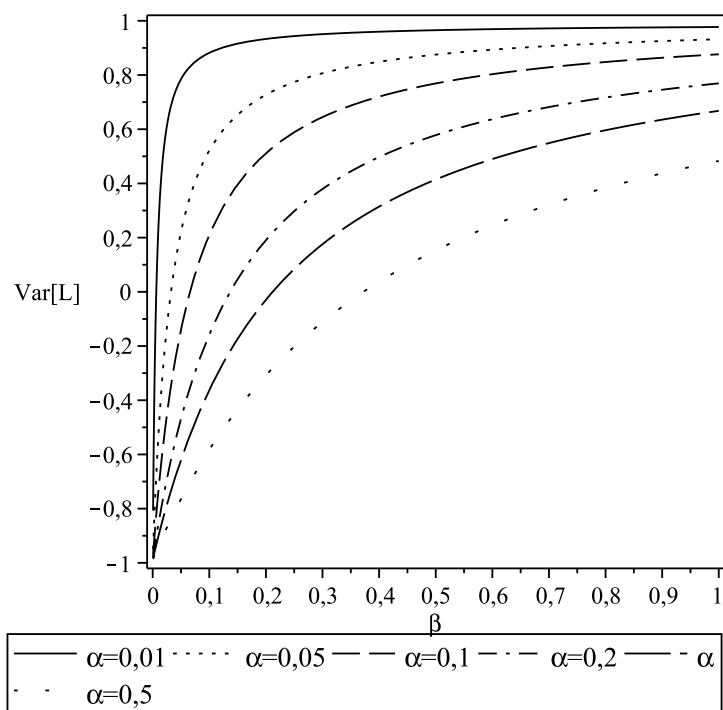
As Figuras 2.11, 2.12 e 2.13, ilustram a variação de  $\text{Var}[L]$ , Equação (2.93), em função de  $\alpha$  e  $\beta$ . A primeira figura é uma superfície que ilustra o comportamento da variância em função de  $\alpha$  e  $\beta$ . Verifica-se que a variância aumenta quando ocorre um aumento do valor do parâmetro  $\beta$ . Este comportamento é observado no gráfico da variação de  $\text{Var}[L]$  em função de  $\beta$ , Figura 2.13, para diferentes valores de  $\alpha$ . Fenômeno inverso corre com o valor do parâmetro  $\alpha$ , a variância diminui quando o valor de  $\alpha$  cresce e a variância aumenta quando o valor de  $\alpha$  diminui. A Figura 2.12 ilustra esta variação de  $\text{Var}[L]$  em função de  $\alpha$  para alguns valores de  $\beta$ .



**Figura 2.11** Superfície de  $\text{Var}[L]$  em função de  $\alpha$  e  $\beta$ .



**Figura 2.12** Gráfico de Var[L] em função de  $\alpha$ .



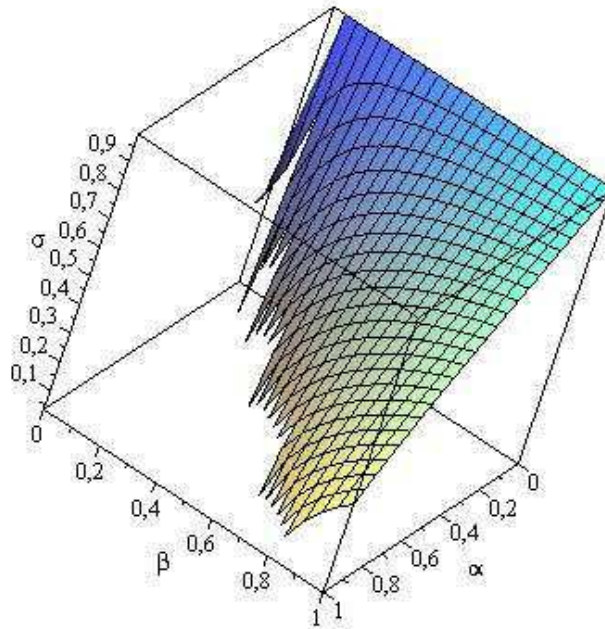
**Figura 2.13** Gráfico de Var[L] em função de  $\beta$ .

### 2.5.4 Desvio padrão

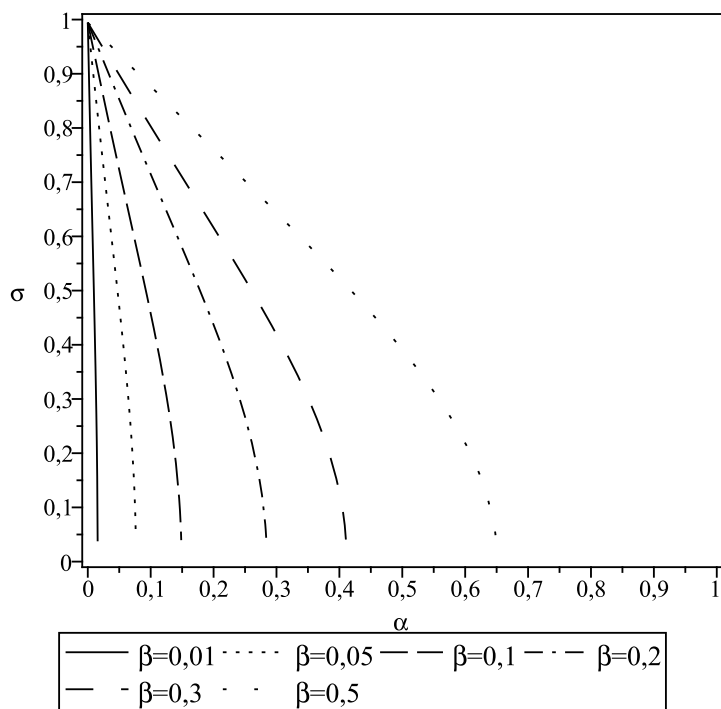
Sabendo que  $\text{Var}[L] = \sigma^2$ , então para encontrar o desvio padrão basta tomar a raiz quadrada da Equação (2.99), logo

$$\sigma = \left\{ \left( \frac{1}{2} \right) (L_M^2 - L_m^2) A_0 (A_1 + A_2) + \left( \frac{1}{16} \right) (L_M - L_m)^2 \left[ 2 + A_0 (A_3 + A_4) + (A_0)^2 (A_1 + A_2)^2 \right] \right\}^{\frac{1}{2}}. \quad (2.100)$$

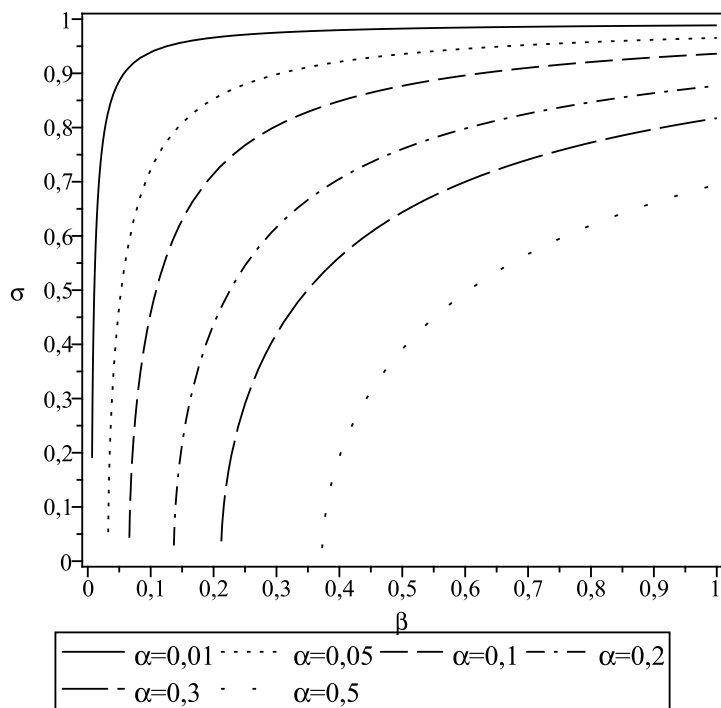
As Figuras 2.14, 2.15 e 2.16, ilustram a variação do desvio padrão,  $\sigma$ , Equação (2.100), em função de  $\alpha$  e  $\beta$ . A variação do desvio padrão é semelhante à da variância, ou seja, o desvio padrão aumenta quando ocorre um aumento do valor do parâmetro  $\beta$ . E o valor desvio padrão diminui quando o valor de  $\alpha$  cresce e aumenta quando o valor de  $\alpha$  diminui.



**Figura 2.14** Superfície de  $\sigma$  em função de  $\alpha$  e  $\beta$ .



**Figura 2.15** Gráfico de  $\sigma$  em função de  $\alpha$ .



**Figura 2.16** Gráfico de  $\sigma$  em função de  $\beta$ .

## 2.6 Conclusão

Neste capítulo foram apresentados os desenvolvimentos matemáticos para se obter as fórmulas matemáticas da função densidade de probabilidade (fdp) e da função cumulativa de probabilidade (FCP) do comprimento dos pacotes em redes de computadores com tráfego bimodal. Além disso, encontrou-se também as equações da média, segundo momento, variância e desvio padrão.

No próximo capítulo, o modelo proposto é comparado com diversas medições realizadas pelo autor e com valores de medidas obtidas da literatura.



## CAPÍTULO 3

---

---

# Medidas de Tráfego em Redes de Computadores

---

### 3.1 Introdução

Neste capítulo são apresentadas as comparações entre os valores teóricos e práticos dos comprimentos dos pacotes em rede de computadores. Os valores reais ou práticos utilizados são algumas medições de dados, obtidas pelo autor, em várias situações, e medidas de tráfego de dados encontrados na literatura obtidas por diferentes autores. Os valores teóricos são obtidos das fórmulas matemáticas desenvolvidas no capítulo anterior.

Esses dados podem ser divididos em dois subconjuntos. No primeiro, apresentam-se os resultados da densidade de probabilidade do comprimento dos pacotes, também chamado de tamanho dos pacotes, por alguns autores. No segundo subconjunto são ilustrados os resultados da distribuição cumulativa do comprimento dos pacotes.

### 3.2 Medidas das Frequências Relativas do Comprimento dos Pacotes

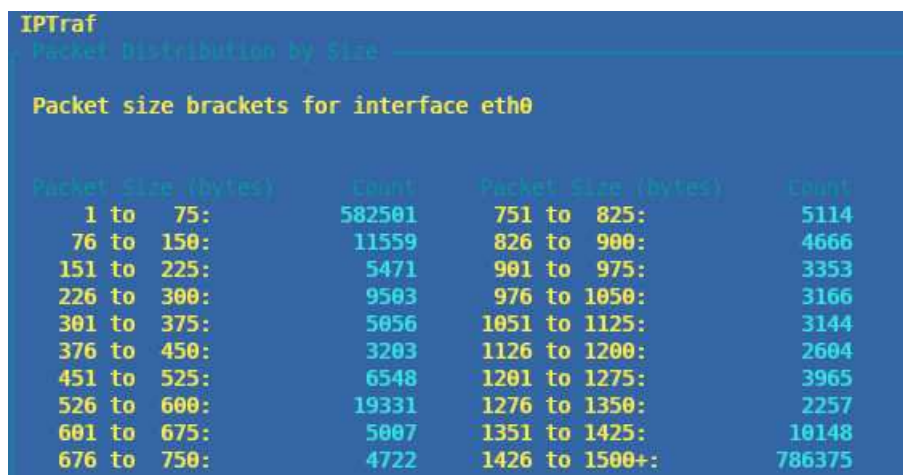
Nesta seção são apresentados alguns conjuntos de dados de tráfego em redes de computadores coletados pelo autor, com foco nas medidas de comprimento dos pacotes em redes de computadores. Primeiro, é descrito, passo-a-passo, o procedimento para se obter os histogramas da distribuição do comprimento dos pacotes. Em seguida, esses dados serão comparados com o modelo matemático da densidade de probabilidade do comprimento dos pacotes e as diferenças entre ambos são calculadas.

### 3.2.1 Frequências Relativas

A Figura 2.1, ilustra uma configuração bastante comum para redes de computadores com acesso a Internet e é nessa configuração em que os dados foram medidos.

O primeiro conjunto de dados disponibilizados descreve a distribuição do comprimento dos pacotes presentes no tráfego de uma conexão de Internet que usa um servidor de rede local (*gateway*), Figura 2.1.

Para o conjunto de medições realizadas pelo autor foi utilizado o programa IPTRAF (IPTRAF, 2010) em um computador com o sistema operacional Linux. Esse programa captura as dimensões dos pacotes IP em bytes de entrada e saída da placa de rede com acesso à Internet durante um intervalo de tempo determinado (Figura 3.1). Em seguida, essas informações são salvas em um arquivo *log* do tipo texto (Figura 3.2) com a gravação das informações feita de forma periódica.



```

IPtraf
Packet Distribution by Size

Packet size brackets for interface eth0

Packet Size (bytes)      Count      Packet Size (bytes)      Count
  1 to 75:                582501    751 to 825:              5114
 76 to 150:              11559     826 to 900:              4666
151 to 225:               5471     901 to 975:              3353
226 to 300:               9503     976 to 1050:             3166
301 to 375:               5056     1051 to 1125:            3144
376 to 450:               3203     1126 to 1200:            2604
451 to 525:               6548     1201 to 1275:            3965
526 to 600:              19331    1276 to 1350:            2257
601 to 675:               5007     1351 to 1425:            10148
676 to 750:               4722     1426 to 1500+:           786375
  
```

**Figura 3.1** Tabela gerada pelo programa IPTRAF – Distribuição de comprimento dos pacotes.

Algumas observações sobre as Figuras 3.1 e 3.2 se fazem necessárias. A primeira, observando a primeira linha da Figura 3.1 e a primeira da Figura 3.2, verifica-se que a quantidade de pacotes no intervalo 1 – 75 é de 582.501 na primeira e de 582.510 na segunda figura. Essa diferença de valores ocorre devido à diferença de tempo entre a captura da tela do IPTRAF e a gravação do arquivo de *log*. A segunda observação é que na estatística de coleta do comprimento dos pacotes da rede foi utilizado um intervalo de 75 bytes. Esse intervalo é padrão do IPTRAF e não pode ser alterado. A escolha do IPTRAF foi pela facilidade de uso e por ser um programa padrão já instalado nos servidores Linux. Outra observação importante é sobre o comprimento mínimo e máximo dos pacotes nas redes. Uma descrição detalhada sobre essa observação é apresentada na seção a seguir.

```
Running time: 86460 seconds
*** Packet Size Distribution, generated Wed Oct 7 10:15:41 2009

Interface: eth0 MTU: 1500

Packet Size (bytes)    Count
1 to 75:                582510
76 to 150:              11559
151 to 225:             5471
226 to 300:             9506
301 to 375:             5056
376 to 450:             3203
451 to 525:             6548
526 to 600:            19331
601 to 675:             5007
676 to 750:            4722
751 to 825:            5114
826 to 900:            4666
901 to 975:            3353
976 to 1050:           3166
1051 to 1125:          3144
1126 to 1200:          2604
1201 to 1275:          3965
1276 to 1350:          2257
1351 to 1425:          10148
1426 to 1500:          786375

Running time: 86514 seconds
Wed Oct 7 10:15:41 2009; ***** Packet size distribution facility stopped *****
```

**Figura 3.2** Tabela com o arquivo *log* em formato texto gerado pelo programa IPTRAF – Distribuição de comprimento dos pacotes, obtido com o uso de um computador do laboratório durante um dia.

### Comprimento mínimo e máximo dos pacotes

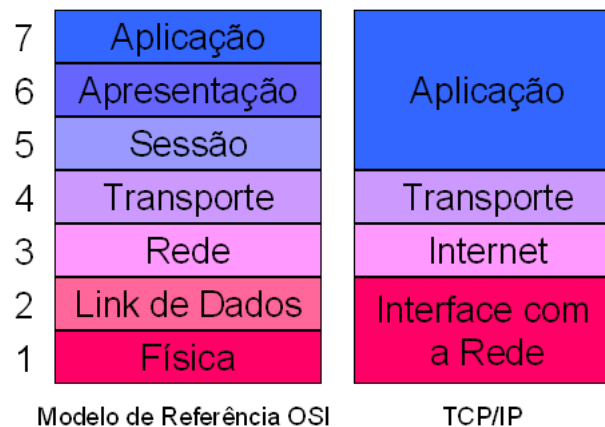
Em redes de computadores, o MTU (*Maximum Transmission Unit* – Unidade Máxima de Transmissão) refere-se ao comprimento do maior datagrama que uma camada de um protocolo de comunicação pode transmitir. Na Tabela 3.1 são apresentados alguns valores padrões de MTU e suas respectivas tecnologias de rede. Os valores, em algumas situações, podem ser alterados pelo administrador da rede.

O protocolo IP permite a fragmentação de pacotes, possibilitando que um datagrama seja dividido em pedaços, cada um pequeno o suficiente para poder ser transmitido por uma conexão com o MTU menor que o datagrama original. Essa fragmentação acontece na camada IP, camada 3 do modelo OSI, Figura 3.3, e usa o parâmetro MTU da interface de rede que envia o pacote pela conexão. O processo de fragmentação marca os fragmentos do pacote original para que a camada IP do destinatário possa montar os pacotes recebidos, reconstituindo o datagrama original.

O protocolo da Internet define o MTU do caminho de uma transmissão Internet como o menor valor MTU de qualquer um dos saltos (*hops* - cada roteador no meio do caminho é também conhecido como salto) do IP do caminho desde o endereço de origem até o endereço

Rede	MTU padrão (bytes)
PPP	296
X.25	576
IEEE 802.3	1.492
Ethernet	1.500
FDDI	4.352
4Mb Token Ring	4.464
16Mb Token Ring	17.914
Hyperchannel	65.535

**Tabela 3.1** Unidade Máxima de Transmissão (MTU) para diferentes tipos de rede.



**Figura 3.3** O modelo de referência ISO-OSI e o modelo TCP/IP.

de destino. Visto de outro modo, o MTU do caminho define o maior valor de MTU que pode passar pelo caminho sem que os seus pacotes sofram posterior fragmentação.

O RFC 1191 descreve a descoberta do MTU do caminho (*Path MTU discovery*) (RFC-1191, 1990), uma técnica para determinar o MTU do caminho entre dois hospedeiros IP de forma a evitar fragmentação IP. Essa técnica utiliza a resposta (*feedback*) ao endereço de origem, quando este envia datagramas de tamanhos progressivamente maiores, com o *bit* DF (Não fragmentar – *Don't Fragment*) ligado (*on*) – qualquer dispositivo que ao longo do caminho precise de fragmentação, o pacote será descartado e uma resposta ICMP (*Internet Control Message Protocol*–Protocolo de Controle de Mensagens Internet) “datagrama demasiadamente grande” é enviado para o endereço de origem. Com esse teste, o hospedeiro de origem “aprende” qual o valor máximo de MTU que atravessa a rede sem que haja necessidade de fragmentação de pacote. Infelizmente, um número crescente de redes bloqueia todo o tráfego ICMP, por exemplo, para evitar ataques de negação de serviço (*denial of service*), e isso impede o funcionamento da técnica explicada.

Paralelamente, em uma rede IP, o caminho desde o endereço de origem até ao endereço de destino pode ser alterado dinamicamente, em resposta a variados eventos, balanceamento de carga, congestionamento, queda do enlace, etc. Isto pode resultar em que o MTU do caminho

se modifique, repetidamente, durante uma transmissão, o que pode introduzir ainda mais perda de pacotes antes que o *host* descubra o novo valor seguro de MTU.

Assim, nas redes *Ethernet* usa-se, por padrão, um MTU de 1500 bytes ou 1492 bytes no IEEE 802.3 (RFC-1042, 1988). Isso porque, dos 1500 bytes do quadro utilizados no datagrama IP, 8 bytes são utilizadas na camada de enlace (LLC-Controle do Link Lógico e SNAP), o que limita o tamanho máximo do datagrama.

Cada datagrama IP pode ter um tamanho máximo de 65.536 bytes (64 *kbytes*), incluindo seu cabeçalho, que pode usar 20 ou 24 bytes, dependendo se um campo “opções” for usado ou não. Ou seja, o comprimento mínimo é de 20 bytes, se apenas o cabeçalho sem nenhum dado no datagrama for transmitido. Dessa forma, os datagramas IP podem transportar até 65.516 bytes de dados. No entanto, nas redes *Ethernet*, se o pacote de dados recebidos da camada de Transporte for maior do que o MTU do caminho, o protocolo IP fragmenta os pacotes em quantos datagramas forem necessários, de acordo com este valor.

Acima da camada IP, geralmente, utilizam-se os protocolos UDP ou TCP. Nota-se que durante a transferência de dados no protocolo TCP, ele implementa o mecanismo de confirmação de recebimento. Portanto, apesar de o protocolo IP não verificar se o datagrama chegou ao destino, o protocolo TCP faz esta verificação. A conexão é confiável, apesar do IP sozinho ser um protocolo não confiável.

Retornando à análise dos intervalos do comprimento dos pacotes, considera-se que:

- O primeiro intervalo, Figuras 3.1 e 3.2, é de 1 a 75 bytes, para pacotes do protocolo IP. No entanto, pelo que é apresentado nos dois parágrafos anteriores, na prática, o comprimento mínimo é 28 bytes, ou seja, o intervalo será de 28 a 75 bytes. Isso porque são utilizados 8 bytes no LLC mais 20 bytes do cabeçalho do Protocolo de Internet (IP v4).
- Os pacotes do tipo IP/TCP têm comprimento mínimo de 48 bytes, ou seja, o intervalo é de 48 a 75 bytes, dos quais, 8 bytes são do LLC, 20 bytes do cabeçalho do Protocolo de Internet (IP v4) e 20 bytes do protocolo TCP.
- Os pacotes do tipo IP/UDP têm comprimento mínimo de 36 bytes, ou seja, o intervalo é de 36 a 75 bytes. Isso porque são utilizados 8 bytes na camada de enlace, 20 bytes do cabeçalho do Protocolo de Internet (IP v4) e 8 bytes do protocolo UDP.
- No caso do ICMP, o comprimento mínimo é de 32 bytes. Ou seja, 8 bytes são do LLC, 4 bytes do ICMP e 20 bytes do cabeçalho do Protocolo de Internet (IP v4). Outros comprimentos mínimos de pacotes são possíveis dependendo dos protocolos de rede e transporte utilizados, mas não são abordados nesta tese.
- Para o comprimento máximo, o pacote é de 1492 bytes, como descrito. Assim, o último intervalo, na prática, apresentado nas Figuras 3.1 e 3.2 é de 1426 a 1492 bytes.

Tipo	Comp. mínimo (bytes) Tipo + LLC	Comp. máximo (bytes)	Comp. máximo na rede <i>Ethernet</i> (bytes) (MTU - LLC)
IP	28	65.536	1492
TCP/IP	48	65.536	1492
UDP/IP	36	65.536	1492
ICMP/IP	32	65.536	1492

**Tabela 3.2** Comprimentos mínimos e máxios dos pacotes para diferentes tipos de protocolos.

### Coleta de dados de comprimento de pacotes

A primeira coleta de dados foi realizada em uma situação de utilização de um computador no laboratório do Iecom (IECOM, 2010). O principal objetivo foi obter a frequência de ocorrência dos comprimentos dos pacotes para um computador com acesso à Internet e obter as frequências de ocorrência do comprimento dos pacotes, Figura 3.1, Figura 3.2 e Tabela 3.3. O conteúdo obtido durante o período de coleta foi o mais diversificado possível, incluindo sítios de notícias (Globo, Folha, etc), sítio de *blogs*, portais (Uol, Terra, Ig, Yahoo, etc), sítio de vídeo (YouTube), *webmail* (Yahoo, Gmail, Hotmail), *download* de arquivos de vídeos, programas e imagens de CD's. Não foram acessados conteúdos hospedados em FTP e *Torrent* (p2p), bloqueados pelo *firewall* da UFCG. As Tabelas 3.1 e 3.2 ilustram os valores que foram coletados no experimento. A duração de tempo foi de 86.514 segundos, aproximadamente 24 horas (do dia 6 ao dia 7 de outubro de 2009). É para esse conjunto de dados que é apresentado o procedimento completo de obtenção do gráfico de distribuição de comprimentos de pacotes em forma de histograma. Para os demais conjuntos de dados, apenas as frequências de ocorrência do comprimento dos pacotes medidos e os histogramas são apresentados. O primeiro conjunto de dados recebe a nome de “*Diversos*”.

### Tratamento estatísticos dos dados

Os valores da Tabela 3.3 são os mesmos do arquivo texto de *log* do IPTRAF ilustrado na Figura 3.2. Na primeira linha tem-se 582.510 ocorrências de pacotes com comprimentos que variam de 1 a 75 bytes. Na prática para pacotes IP, esse primeiro intervalo é de 28 a 75 bytes como explicado anteriormente, o último é de 1426 a 1492 bytes e os demais intervalos não sofrem alterações. Com o programa IPTRAF não é possível saber qual a distribuição de tamanhos de pacotes dentro de cada um desses intervalos. Assim, usando o conceito de intervalos, limites de classe e ponto médio de uma classe da teoria estatística (SPIEGEL, 2006), obtém-se, a terceira coluna da Tabela 3.3, o Tamanho Médio dos Pacotes ( $TMP_i$ ) do intervalo  $i$ , a partir de

$$TMP_i = \frac{(V_{m_i} + V_{M_i})}{2} \quad (3.1)$$

em que  $V_{m_i}$  é o menor valor do  $i$ -ésimo intervalo e  $V_{M_i}$  é o maior valor do mesmo intervalo, ambos apresentados na coluna Tamanho do pacote ( $T$ ) da Tabela 3.3.  $N_i$  é número do  $i$ -ésimo intervalo,  $T$  é o tamanho do pacote em (bytes) e  $FP$  é a frequência de ocorrência dos pacotes.

N do intervalo ( $i$ )	Tamanho do pacote (bytes) ( $T$ )	Tamanho médio dos pacotes ( $TMP$ )	Frequência dos pacotes ( $FP$ )
1	1 a 75 (28 a 75)	38 ( $\sim 52$ )	582510
2	76 a 150	113	11559
3	151 a 225	188	5471
4	226 a 300	263	9506
5	301 a 375	338	5056
6	376 a 450	413	3203
7	451 a 525	488	6548
8	526 a 600	563	19331
9	601 a 675	638	5007
10	676 a 750	713	4722
11	751 a 825	788	5114
12	826 a 900	863	4666
13	901 a 975	938	3353
14	976 a 1050	1013	3166
15	1051 a 1125	1088	3144
16	1126 a 1200	1163	2604
17	1201 a 1275	1238	3965
18	1276 a 1350	1313	2257
19	1351 a 1425	1388	10148
20	1426 a 1500+ (1426 a 1429)	1463 (1459)	786375
Total		15.010 (15.020)	1.477.705

**Tabela 3.3** Tamanho, tamanho médio e frequência dos pacotes.

A partir dos dados da Tabela 3.3, encontra-se o tamanho médio dos pacotes de todo tráfego, usando diversos tipos de médias, definidos na Tabela 3.4. Nos cálculos das Tabelas 3.4 e 3.5 considerou-se que todos os pacotes do intervalo  $i$  fossem de tamanho  $TMP_i$ .

A Tabela 3.5 ilustra os valores de desvio médio, desvio padrão, variância, primeiro e segundo momentos centrados na média.

O valor  $TMP$  normalizado ( $TMP_n$ ) apresentado na coluna 3 da Tabela 3.6 é obtido pela divisão do  $TMP$  pelo valor do MTU de 1500, padrão para redes Ethernet. Na última coluna da mesma Tabela, os valores de  $FP$  relativos ou normalizados ( $FP_n$ ) são obtidos pela divisão do número de ocorrência do tamanho do pacote pelo número total de pacotes.

A Figura 3.4 apresenta o comprimento médio dos pacotes ( $TMP$ ) versus frequência de ocorrência dos pacotes ( $FP$ ). Os valores são apresentados sem a normalização na Tabela 3.6, segunda e quarta colunas respectivamente. A Figura 3.5 ilustra os mesmos valores da Figura

Tamanho médio ( $TM$ )	Equação	Valor numérico
Média aritmética com pesos	$\frac{\sum_{i=1}^{20} TMP_i \cdot FP_i}{\sum_{i=1}^{20} FP_i}$	845,185
Média aritmética sem pesos	$\frac{\sum_{i=1}^{20} TMP_i}{20}$	750,500
Raiz média quadrática sem pesos	$\sqrt{\frac{\sum_{i=1}^{20} TMP_i^2}{20}}$	865,887
Raiz média quadrática com pesos	$\sqrt{\frac{\sum_{i=1}^{20} TMP_i^2 \cdot FP_i}{20}}$	1084,93
Média geométrica sem pesos	$\left(\prod_{i=1}^{20} TMP_i\right)^{\frac{1}{20}}$	571,14
Média harmônica sem pesos	$\frac{20}{\sum_{i=1}^{20} \frac{1}{TMP_i}}$	341,226

**Tabela 3.4** Tamanho médio ( $TM$ ) por diversos métodos.

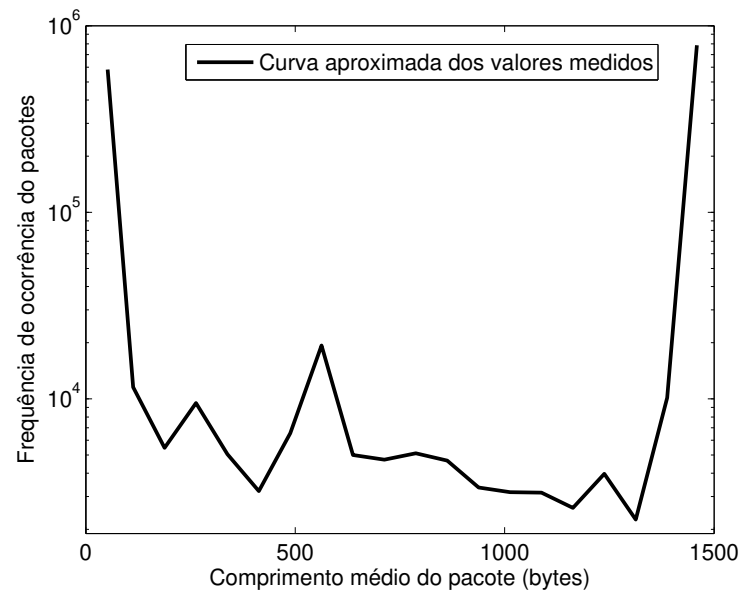
Tipo	Equação	Valor Numérico
Desvio médio	$\frac{\sum_{i=1}^{20}  TMP_i - TM  FP_i}{\sum_{i=1}^{20} FP_i}$	667,69
Desvio padrão	$\sqrt{\frac{1}{(\sum_{i=1}^{20} FP_i) - 1} (\sum_{i=1}^{20} (TMP_i - TM)^2 FP_i)}$	680,25
Variância	(Desvio Padrão) <sup>2</sup>	462.742,478
1º Momento	$\frac{\sum_{i=1}^{20} (TMP_i - TM) FP_i}{\sum_{i=1}^{20} FP_i}$	0
2º Momento	$\frac{\sum_{i=1}^{20} (TMP_i - TM)^2 FP_i}{\sum_{i=1}^{20} FP_i}$	462.742,164
Raiz 2º momento	$\sqrt{2^\circ Momento}$	680,25

**Tabela 3.5** Desvio padrão, variância e momentos.



i	Tamanho médio dos pacotes (TMP)	TMP normalizado	Frequência dos pacotes (FP)	FP normalizado
1	52	0,0347	582510	0,3942
2	113	0,0753	11559	0,0078
3	188	0,1253	5471	0,0037
4	263	0,1753	9506	0,0064
5	338	0,2253	5056	0,0034
6	413	0,2753	3203	0,0022
7	488	0,3253	6548	0,0044
8	563	0,3753	19331	0,0131
9	638	0,4253	5007	0,0034
10	713	0,4753	4722	0,0032
11	788	0,5253	5114	0,0035
12	863	0,5753	4666	0,0032
13	938	0,6253	3353	0,0023
14	1013	0,6753	3166	0,0021
15	1088	0,7253	3144	0,0021
16	1163	0,7753	2604	0,0018
17	1238	0,8253	3965	0,0027
18	1313	0,8753	2257	0,0015
19	1388	0,9253	10148	0,0069
20	1459	0,9727	786375	0,5322
Total	15.020		1.477.705	

**Tabela 3.6** Valores da Tabela 3.3 normalizados.



**Figura 3.4** Comprimento médio dos pacotes versus frequência de ocorrência dos pacotes.

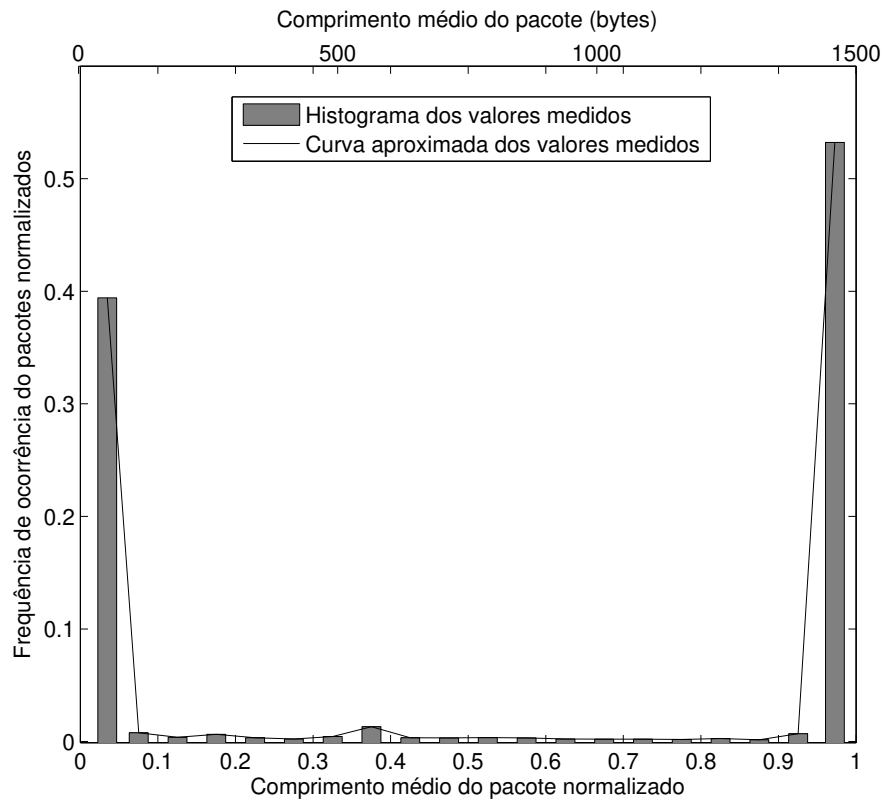
3.4, só que normalizados, colunas três e cinco da Tabela 3.6. Na Figura 3.5, o histograma em forma de barra representa os dados coletados e normalizados em cada intervalo, última coluna da Tabela 3.6. O gráfico em linha contínua é a curva aproximada dos valores medidos. Nesta mesma figura, é possível observar o comportamento bimodal, com o primeiro pico próximo a 44 *bytes* (primeiro pico) e outro entre 1400 *bytes* e 1500 *bytes* (JOHN; TAFVELIN, 2007). Destaca-se ainda o pico de baixa intensidade em 576 *bytes*, resquícios de um comportamento trimodal, como destacado por Rishi Sinha (SINHA *et al.*, 2007), McCreary (MCCREARY; CLAFFY, ), Karagiannis (KARAGIANNIS *et al.*, 2003), Cheng Yu (YU *et al.*, 2007) e Rastin Pries (PRIES *et al.*, 2009).

O processo de captura dos dados, geração de tabelas, normalização dos valores e apresentação em forma de histograma foram realizados para cada um dos conjuntos de dados presentes nesta tese. No entanto, apenas os dados coletados e os resultados em forma de histogramas normalizados são apresentados nas demais medições.

As comparações entre a função densidade de probabilidade  $p_L(\ell)$  dada pela Equação (2.26) e o conjunto de dados “*Diversos*” são apresentados nas Figuras 3.6 e 3.7.

Na Figura 3.6 há três gráficos. O primeiro em barra, mostra o histograma dos valores medidos. O segundo, em linha tracejada, ilustra a curva aproximada dos valores medidos. Por fim, o terceiro, em linha escura contínua, a função densidade de probabilidade  $p_L(\ell)$  proposta nesta tese e descrita pela Equação (2.26).

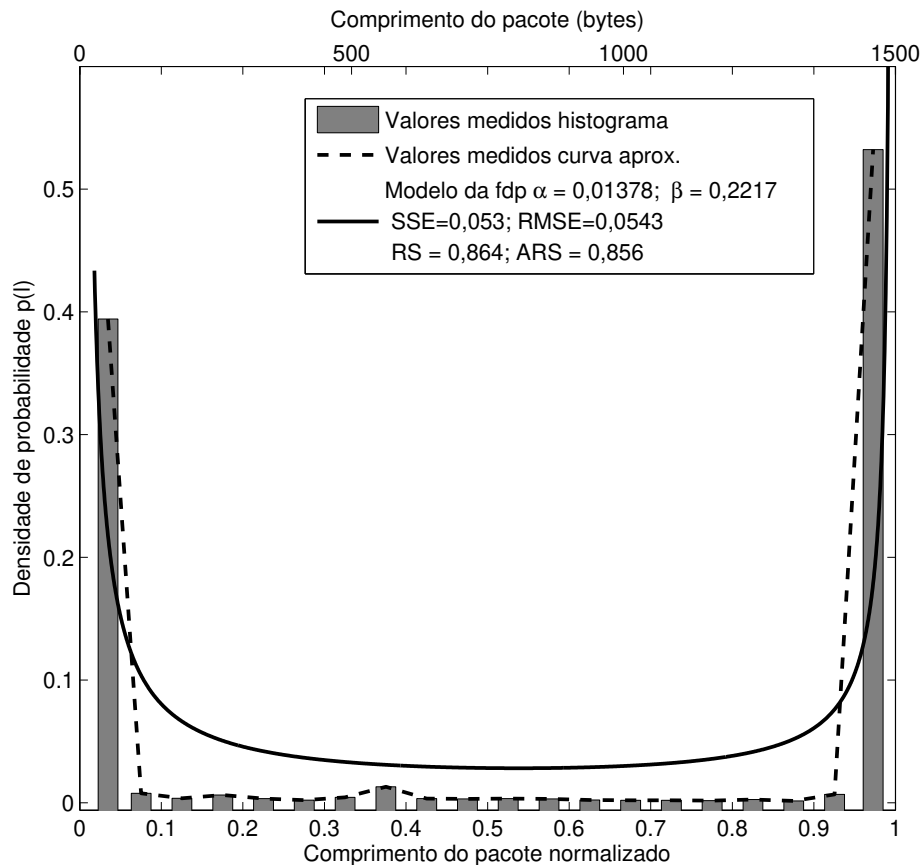
Os valores teóricos obtidos com a fórmula de  $p_L(\ell)$  foram ajustados pelo método dos mínimos quadrados para encontrar os melhores valores dos parâmetros  $\alpha$  e  $\beta$ , considerando  $\alpha, \beta > 0$ . As métricas Soma dos Quadrados dos Erros (SSE–*Sum of Squares due to Error*), Erro da raiz média quadrática (RMSE–*Root Mean Square Error*), R-quadrado (RS–*R-square*)



**Figura 3.5** Comprimento médio dos pacotes versus frequência de ocorrência dos pacotes, valores normalizados.

e R-quadrado ajustado (*ARS-Adjusted R-square*) foram utilizadas no cálculo das diferenças entre os valores analíticos e os valores experimentais. Os valores de SSE e RMSE próximos de zero indicam que o modelo possui um pequeno componente de erro aleatório, e que o ajuste obtido é bom na predição dos valores. O valor de RS varia entre zero e um e é o quadrado da correlação entre os valores experimentais e os valores analíticos. Os valores de RS próximos a um indicam que o modelo proposto é adequado para representar o conjunto de dados. O ARS é o RS ajustado. Seu valor está entre zero e um, com os valores mais próximos a um, indicando o melhor ajuste entre o modelo e o experimento (MATLAB, 2010).

Os valores obtidos nesta primeira comparação são  $\alpha = 0,01378$ ,  $\beta = 0,2217$ ,  $SSE = 0,05297$  (5,297%) e  $RMSE = 0,05425$  (5,426%). Como mencionado no parágrafo anterior, quanto mais próximos a zero forem os valores de SSE e RMSE, melhor será o ajuste da curva aos dados. Os valores obtidos indicam um bom ajuste. Isto significa uma boa aproximação entre o valor medido e o valor teórico. O valor obtido para o RS foi de 0,8638, ou seja, 86,38% do total de variação dos dados estão representados pelo modelo. Para o ARS este valor foi 0,8563, ou 85,63%. Graficamente, o modelo capturou bem o comportamento bimodal do tráfego como sugerido por Tafvelin (JOHN; TAFVELIN, 2007) e Pries (PRIES *et al.*, 2009), além de aderir bem aos resultados. Apesar de que, na parte central do gráfico, a curva de  $p_L(\ell)$  ficou um pouco acima dos valores medidos. Isto se refletiu nos valores dos erros calculados. Numericamente, as

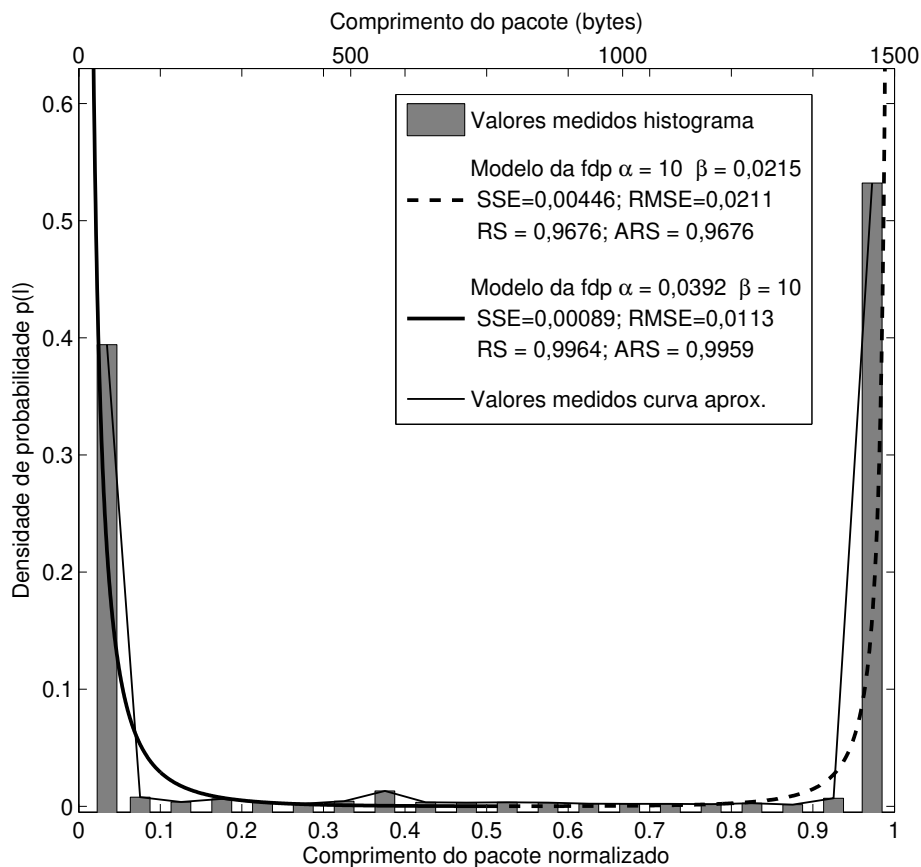


**Figura 3.6** Modelo de densidade de probabilidade versus valores medidos (Diversos).

quatro métricas utilizadas para verificação da qualidade do modelo, para este primeiro conjunto de dados medidos, apresentaram bons resultados.

Os resultados apresentados no parágrafo anterior podem ser otimizados, diminuição da diferença entre os valores medidos e a curva teórica, se forem considerados dois intervalos de dados. Nesse caso, se utilizam os mesmos dados, a mesma fórmula para  $p_L(\ell)$ , mas se divide o intervalo de comprimento dos pacotes  $l$  em dois segmentos. O primeiro de zero a 0,5 (750 bytes) e o segundo de 0,5 a 1 (1500 bytes). A Figura 3.7 ilustra esta nova situação.

A Figura 3.7 apresenta quatro gráficos. O primeiro em barra, é o histograma dos valores medidos. O segundo, em linha fina e contínua, é a curva aproximada dos valores medidos. O terceiro, em linha grossa e contínua, a  $p_L(\ell)$ , avaliada no primeiro intervalo de dados (0 a 0,5). E finalmente, em linha grossa e tracejada,  $p_L(\ell)$ , avaliada no segundo intervalo de dados (0,5 a 1). Na nova configuração, os valores obtidos são de  $\alpha = 0,03919$ ,  $\beta = 10$ ,  $SSE = 0,004464$ ,  $RMSE = 0,02113$ ,  $RS = 0,9676$  e  $ARS = 0,9676$  no primeiro intervalo. E para o segundo intervalo,  $\alpha = 10$ ,  $\beta = 0,02153$ ,  $SSE = 0,0008932$  (0,089%),  $RMSE = 0,0113$  (1,13%),  $RS = 0,9964$  (99,64%) e  $ARS = 0,9959$  (99,59%). Estes valores reforçam os excelentes resultados apresentados na Figura 3.6 e diminuem de forma considerável, os erros no intervalo (0,1–0,9) localizado entre os dois picos, aproximando os valores teóricos e práticos.



**Figura 3.7** Modelo de densidade de probabilidade (dois intervalos) versus valores medidos (Diversos).

Devido aos bons resultados obtidos com a divisão com dois intervalos, não houve necessidade de divisão do intervalo em um número maior.

Nos parágrafos a seguir, são repetidas as comparações descritas, mas em situações distintas de tráfego de dados.

### Comprimento de pacotes–*YouTube*

O segundo conjunto de dados foi obtido em uma situação de utilização de um computador com acesso a um sítio com conteúdo de vídeo. Foram abertos diversos vídeos de tamanhos e tempos variados no sítio do *YouTube*. Com esse conjunto de dados, tem-se uma pequena amostra do tráfego de *stream* de vídeo e de como seria o comportamento da distribuição dos comprimentos dos pacotes neste caso específico. Para esse conjunto de dados deu-se o nome de “*YouTube*”. As Figuras 3.8, 3.9 e 3.10 ilustram os resultados obtidos.

```
Thu Nov 19 16:08:31 2009; ***** Packet size distribution facility started *****  
*** Packet Size Distribution, generated Thu Nov 19 16:51:18 2009  
  
Interface: eth0  MTU: 1500  
  
Packet Size (bytes)  Count  
1 to 75:             83928  
76 to 150:           381  
151 to 225:          267  
226 to 300:          685  
301 to 375:          293  
376 to 450:          1410  
451 to 525:          562  
526 to 600:          322  
601 to 675:          132  
676 to 750:          131  
751 to 825:          225  
826 to 900:          172  
901 to 975:          141  
976 to 1050:         147  
1051 to 1125:        151  
1126 to 1200:        119  
1201 to 1275:        184  
1276 to 1350:        152  
1351 to 1425:        109  
1426 to 1500:        114253  
  
Running time: 2567 seconds  
Thu Nov 19 16:51:18 2009; ***** Packet size distribution facility stopped *****
```

**Figura 3.8** Tabela com o arquivo *log* em formato texto gerado pelo programa IPTRAF-*YouTube*.

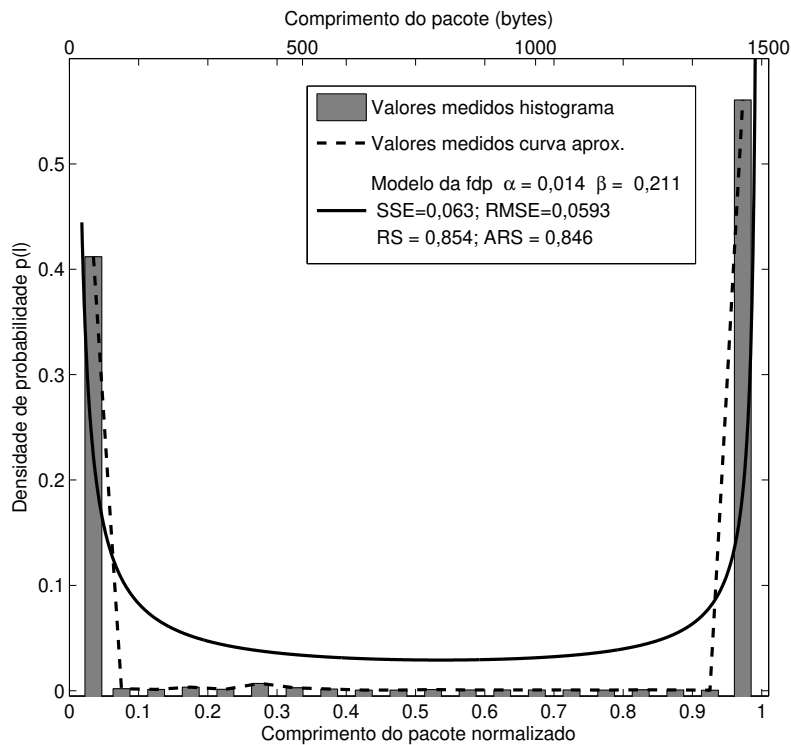


Figura 3.9 Modelo de densidade de probabilidade versus valores medidos (YouTube).

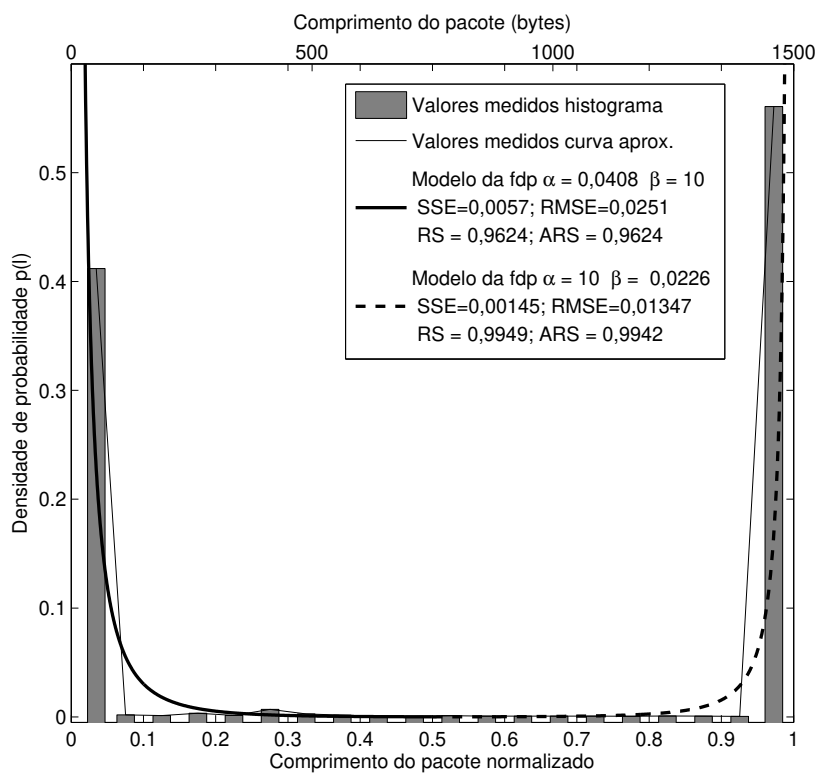


Figura 3.10 Modelo de densidade de probabilidade (dois intervalos) versus valores medidos (YouTube).

A Figura 3.8 apresenta os dados coletados pelo IPTRAF. A Figura 3.9 ilustra a comparação entre os dados medidos e o modelo proposto. Os resultados são  $\alpha = 0,0142$ ,  $\beta = 0,2109$ ,  $SSE = 0,06332$  (6,332%),  $RMSE = 0,05931$  (5,931%),  $RS = 0,8542$  (85,42%) e  $ARS = 0,8461$  (84,61%).

Repetindo os valores do primeiro conjunto de dados, numericamente, as quatro métricas utilizadas apresentaram bons resultados. Graficamente, o modelo apresentou o comportamento bimodal previsto. Mas como no caso anterior, a parte central do gráfico,  $p_L(\ell)$ , ficou um pouco afastada dos valores medidos, refletindo nos valores dos erros. Esses resultados podem ser otimizados se forem considerados dois intervalos de dados, Figura 3.10. Neste caso, os valores do primeiro intervalo são  $\alpha = 0,04078$ ,  $\beta = 10$ ,  $SSE = 0,005672$  (0,5672%),  $RMSE = 0,02511$  (2,511%),  $RS = 0,9624$  e  $ARS = 0,9624$ . Para o segundo intervalo,  $\alpha = 10$ ,  $\beta = 0,02265$ ,  $SSE = 0,001452$  (0,014%),  $RMSE = 0,01347$  (1,34%),  $RS = 0,9949$  (99,49%) e  $ARS = 0,9942$  (99,42%). No caso, repetem-se os bons resultados do caso anterior e novamente a redução da diferença entre o modelo teórico e os valores práticos.

### Comprimento de pacotes–*Torrent*

```
Running time: 21600 seconds
*** Packet Size Distribution, generated Tue Oct 27 06:07:16 2009

Interface: eth1  MTU: 1500

Packet Size (bytes)      Count
1 to 75:                 91211
76 to 150:               11532
151 to 225:              653
226 to 300:              403
301 to 375:              896
376 to 450:              201
451 to 525:              62
526 to 600:              349
601 to 675:              121
676 to 750:              45
751 to 825:              25
826 to 900:              638
901 to 975:              19
976 to 1050:             65
1051 to 1125:            74
1126 to 1200:            60
1201 to 1275:            13
1276 to 1350:            78
1351 to 1425:            451
1426 to 1500:           107210

Running time: 23791 seconds
Tue Oct 27 06:07:16 2009; ***** Packet size distribution facility stopped *****
```

**Figura 3.11** Tabela com o arquivo *log* em formato texto gerado pelo programa IPTRAF–*Torrent*.



O terceiro conjunto de dados foi subdividido em dois subconjuntos. Em ambos, os dados foram obtidos em situações em que um computador baixava conteúdos da Internet usando um programa do tipo *Torrent* (p2p). No Primeiro, foram efetuados diversos *downloads* de arquivos de tamanhos variados (5MB, 10MB, 12MB, 15MB e 17MB) (Figura 3.11). Este conjunto é nomeado “*Torrent*”. Em seguida, foi feito o *download* da imagem de um DVD de aproximadamente 2,1 *Gbytes* (Figura 3.12) e nomeado “*Torrent2*”. Semelhante aos casos anteriores, a tentativa foi obter uma amostra de como ficaria a distribuição do comprimento de pacotes quando a rede possui um tráfego do tipo *Torrent*.

```
Running time: 3600 seconds
*** Packet Size Distribution, generated Thu Dec 31 17:04:55 2009

Interface: eth1  MTU: 1500

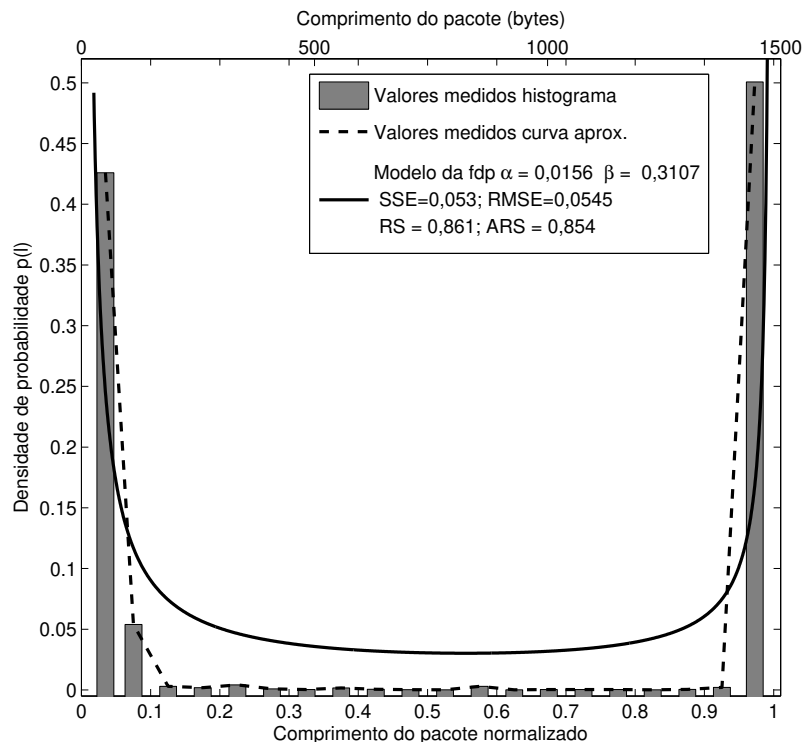
Packet Size (bytes)      Count
1 to 75:                 1050053
76 to 150:               34547
151 to 225:              3758
226 to 300:              8856
301 to 375:              2388
376 to 450:              2052
451 to 525:              1849
526 to 600:              1704
601 to 675:              1453
676 to 750:              1134
751 to 825:              1179
826 to 900:              873
901 to 975:              1368
976 to 1050:             511
1051 to 1125:            1248
1126 to 1200:            663
1201 to 1275:            416
1276 to 1350:            662
1351 to 1425:            942
1426 to 1500:            1502556

Running time: 3926 seconds
Thu Dec 31 17:04:55 2009; ***** Packet size distribution facility stopped
*****
```

**Figura 3.12** Tabela com o arquivo *log* em formato texto gerado pelo programa *IPTRAF-Torrent2*.

A Figura 3.13 ilustra a comparação entre os dados do experimento e o modelo teórico. Os resultados obtidos são  $\alpha = 0,0156$ ,  $\beta = 0,3107$ ,  $SSE = 0,0534$  (5,34%),  $RMSE = 0,05447$  (5,447%),  $RS = 0,8614$  (86,14%) e  $ARS = 0,8537$  (85,37%).

Numericamente, as quatro métricas apresentam-se dentro da faixa dos resultados anteriores. Graficamente, o comportamento bimodal similar ao anterior e a parte central do gráfico,  $p_L(\ell)$ , afastada dos valores experimentais, com esta diferença reflete nos valores dos erros. Para melhorar os resultados, considera-se novamente, os dois intervalos de dados, e os resultados do primeiro intervalo são  $\alpha = 0,04311$ ,  $\beta = 10$ ,  $SSE = 0,001025$  (0,1025%),  $RMSE = 0,0121$  (1,21%),  $RS = 0,9936$  e  $ARS = 0,9918$ . Para o segundo intervalo,  $\alpha = 10$ ,  $\beta = 0,02025$ ,  $SSE = 0,001067$  ou (0,1067%),  $RMSE = 0,01155$  ou (1,155%),  $RS = 0,9953$  ou (99,53%)



**Figura 3.13** Modelo de densidade de probabilidade versus valores medidos (*Torrent*).

e  $ARS = 0,9947$  ou (99,47%). A Figura 3.14 ilustra os bons resultados obtidos com redução da diferença entre o modelo teórico e os valores práticos.

A Figura 3.15 ilustra a comparação entre o experimento e o modelo de  $p_L(\ell)$ . Os valores observados são  $\alpha = 0,01392$ ,  $\beta = 0,1889$ ,  $SSE = 0,05997$  (5,997%),  $RMSE = 0,05772$  (5,772%),  $RS = 0,8638$  (86,38%) e  $ARS = 0,8563$  (85,63%).

Os bons resultados numéricos e gráficos repetem-se neste caso. A parte central da curva de  $p_L(\ell)$ , fica afastada dos valores experimentais. Assim, para otimizar os resultados e minimizar os erros consideram-se dois intervalos. Os valores do primeiro intervalo são  $\alpha = 0,03993$ ,  $\beta = 10$ ,  $SSE = 0,003911$  (0,3911%),  $RMSE = 0,02211$  (2,211%),  $RS = 0,9727$  e  $ARS = 0,9693$ . E para o segundo intervalo,  $\alpha = 10$ ,  $\beta = 0,02318$ ,  $SSE = 0,001545$  ou (0,1545%),  $RMSE = 0,0139$  ou (1,39%),  $RS = 0,9948$  ou (99,48%) e  $ARS = 0,9941$  ou (99,41%). A Figura 3.16, ilustra a redução da diferença entre o modelo teórico e os valores experimentais.

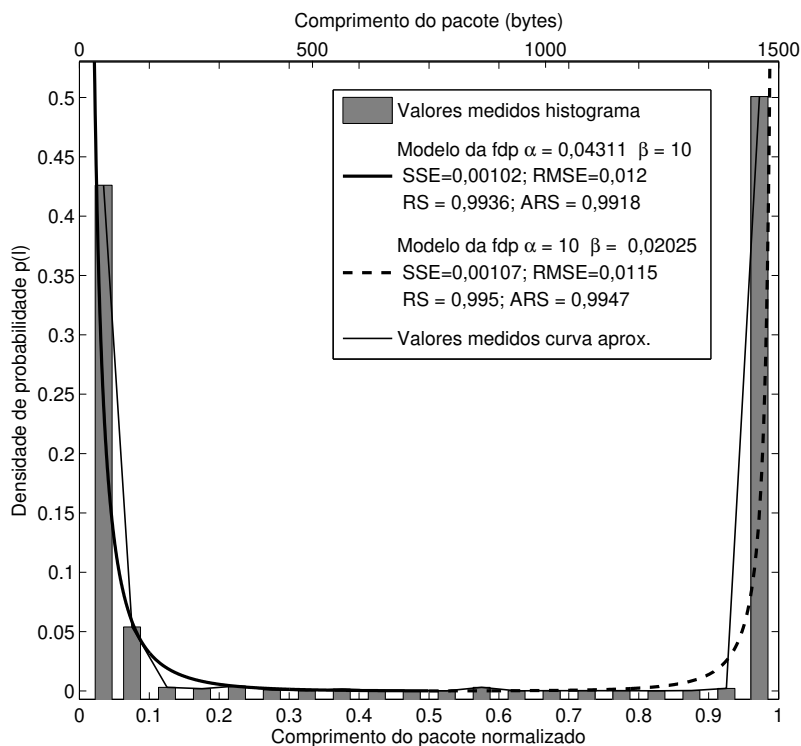


Figura 3.14 Modelo de densidade de probabilidade (dois intervalos) versus valores medidos (*Torrent*).

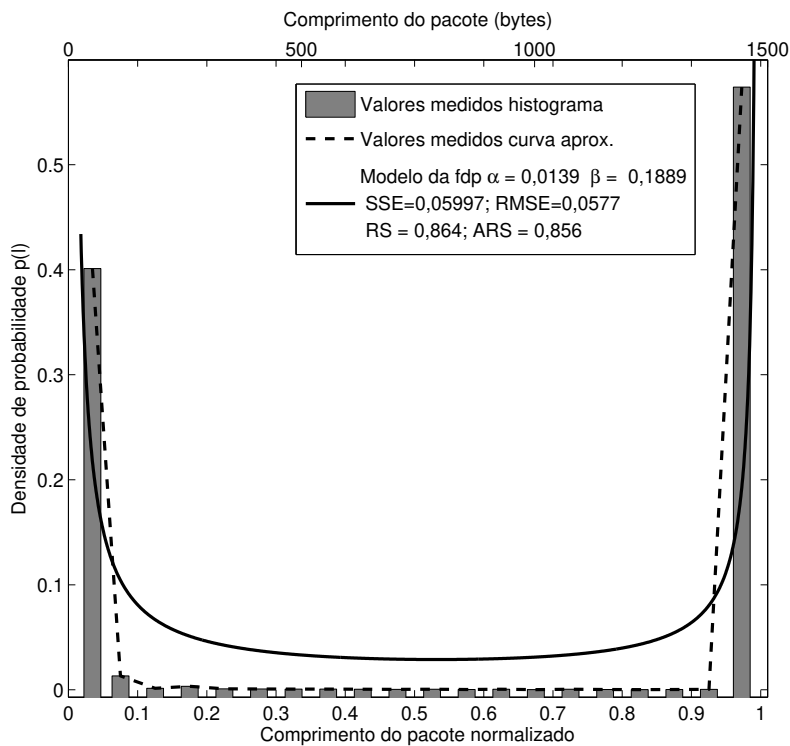
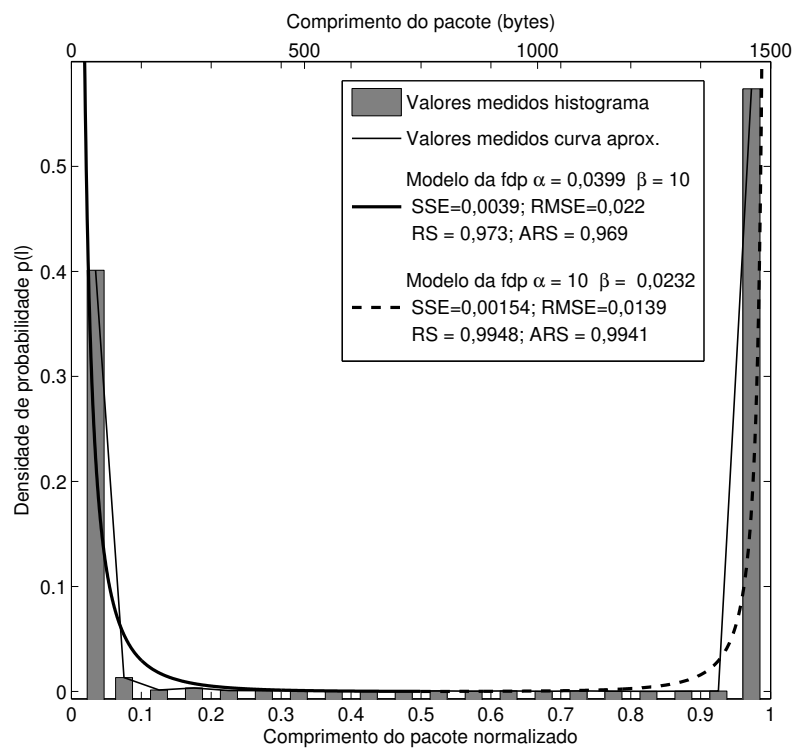


Figura 3.15 Modelo de densidade de probabilidade versus valores medidos (*Torrent2*).



**Figura 3.16** Modelo de densidade de probabilidade (dois intervalos) versus valores medidos (*Torrent2*).

**Comprimento de pacotes–FTP**

```

Running time: 10800 seconds
*** Packet Size Distribution, generated Thu Dec 31 15:57:06 2009

Interface: eth1  MTU: 1500

Packet Size (bytes)  Count
1 to 75:             1821207
76 to 150:           294555
151 to 225:          59732
226 to 300:           28155
301 to 375:          21637
376 to 450:          18691
451 to 525:          17301
526 to 600:          16340
601 to 675:          15652
676 to 750:          15262
751 to 825:          14522
826 to 900:          13492
901 to 975:          12906
976 to 1050:         11976
1051 to 1125:        11346
1126 to 1200:        10538
1201 to 1275:        10209
1276 to 1350:        9901
1351 to 1425:        10487
1426 to 1500:       1192452

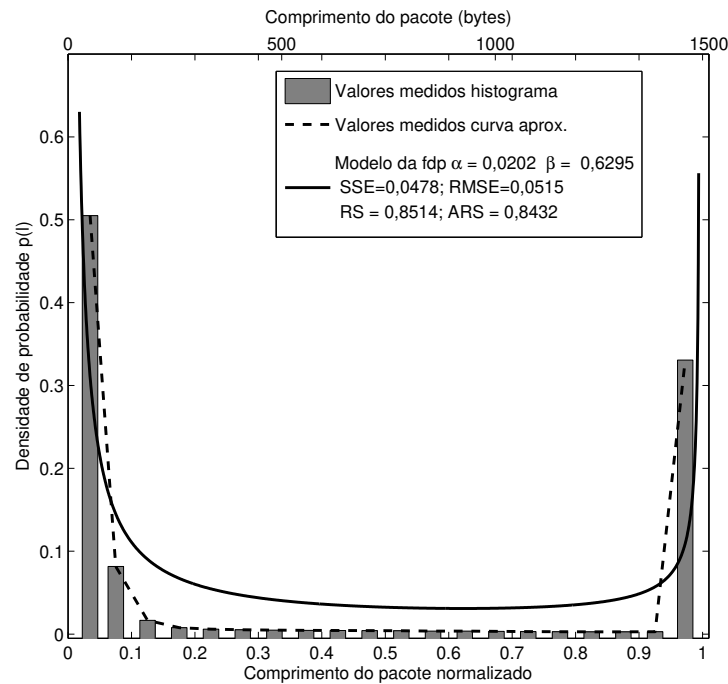
Running time: 13464 seconds
Thu Dec 31 15:57:06 2009; ***** Packet size distribution facility stopped
*****

```

**Figura 3.17** Tabela com o arquivo *log* em formato texto gerado pelo programa *IPTRAF–FTP*.

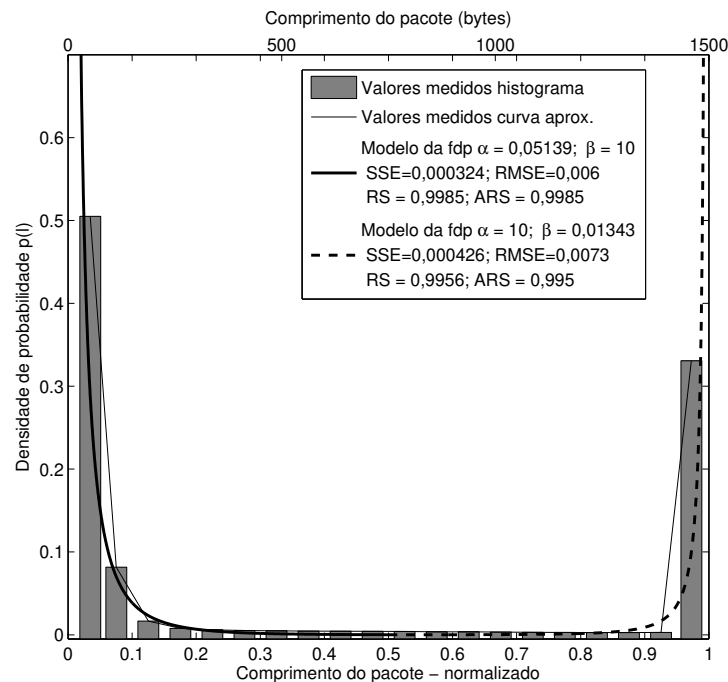
A quarta coleta de dados foi obtida do tráfego de dados de utilização de um computador que transferia conteúdo da Internet usando o protocolo de transferência de arquivos *FTP* (Figura 3.17). Com o mesmo objetivo das medições anteriores de verificar o comportamento do comprimento dos pacotes para uma determinada situação específica, foi feito o *download* de uma imagem de um DVD com 1,8 *Gbytes* utilizando *FTP*. O nome desse conjunto de dados é *FTP* e a conexão com a Internet utilizada foi do tipo ADSL. Essa conexão, também foi usada nas medições anteriores, *Torrent* e *Torrent2*.

A Figura 3.18 apresenta a comparação entre os valores obtidos e os resultados obtidos com a fórmula de  $p_L(\ell)$ . Os valores dos parâmetros após o procedimento de ajuste de curva são  $\alpha = 0,02021$ ,  $\beta = 0,6295$  com  $SSE = 0,04779$  (4,779%),  $RMSE = 0,05152$  (5,152%),  $RS = 0,8514$  (85,14%) e  $ARS = 0,8432$  (84,32%). Considerando os dois intervalos para otimizar os resultados e minimizar os erros, os valores do primeiro intervalo são  $\alpha = 0,03993$ ,  $\beta = 10$ ,  $SSE = 0,003911$  (0,3911%),  $RMSE = 0,02211$  (2,211%),  $RS = 0,9727$  e  $ARS = 0,9693$ . E para o segundo intervalo,  $\alpha = 10$ ,  $\beta = 0,02318$ ,  $SSE = 0,001545$  (0,1545%),  $RMSE = 0,0139$



**Figura 3.18** Modelo de densidade de probabilidade versus valores medidos (FTP).

(1,39%), RS=0,9948 (99,48%) e ARS=0,9941 (99,41%). A Figura 3.19, ilustra os resultados obtidos.



**Figura 3.19** Modelo de densidade de probabilidade (dois intervalos) versus valores medidos (FTP).

### Comprimento de pacotes–*Lab Comp*

O quinto conjunto de dados foi coletado no servidor de um laboratório do Departamento de Ciência da Computação da UFCG (Figura 3.20). Esse servidor provê conexão de acesso a Internet a 56 computadores divididos em três salas. A primeira possui 10 *desktops* com acesso via LAN e 16 *notebooks* com acesso via WLAN. Na segunda sala são 10 *desktops* e na terceira, 20 *desktops*. Esse conjunto é chamado de “Lab Comp” e foi coletado com o mesmo propósito das medições anteriores. O período de coleta foi de aproximadamente 15 horas, nas outras 9 horas restantes do dia, o laboratório permanece vazio e com seus computadores desligados.

```
Running time: 50400 seconds
*** Packet Size Distribution, generated Tue Nov 10 22:06:56 2009

Interface: eth0  MTU: 1500

Packet Size (bytes)    Count
1 to 75:               9548569
76 to 150:             2572441
151 to 225:           9858194
226 to 300:           1526444
301 to 375:           290618
376 to 450:           33011
451 to 525:           17358
526 to 600:           15225
601 to 675:           29164
676 to 750:           80706
751 to 825:           32233
826 to 900:           12842
901 to 975:           32419
976 to 1050:          77453
1051 to 1125:         23852
1126 to 1200:         364242
1201 to 1275:        461900
1276 to 1350:        244812
1351 to 1425:        1486286
1426 to 1500:        14196059

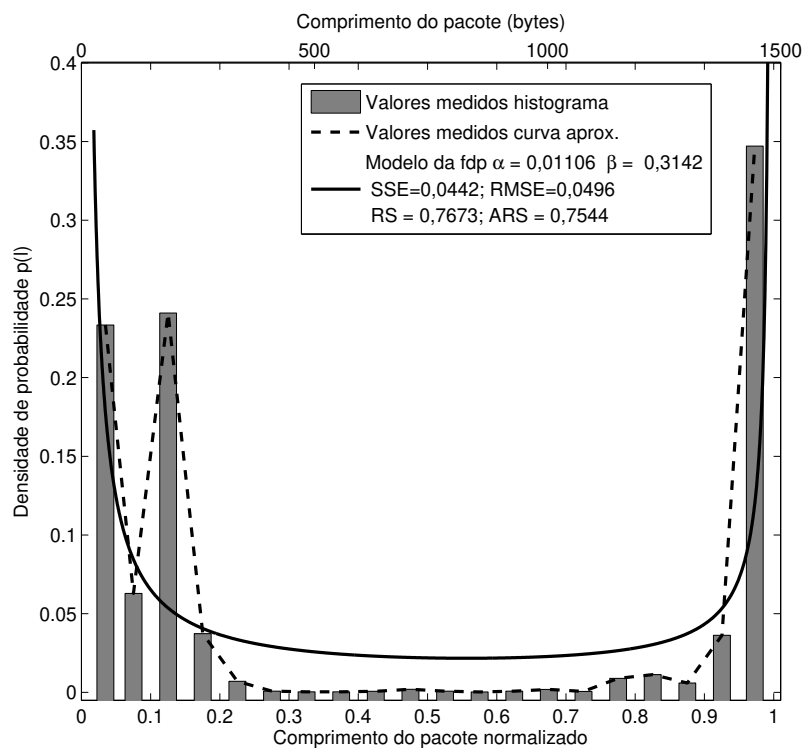
Running time: 52938 seconds
Tue Nov 10 22:06:56 2009; ***** Packet size distribution facility stopped *****
```

**Figura 3.20** Tabela com o arquivo *log* em formato texto gerado pelo programa *IPTRAF–Lab Comp*.

A Figura 3.21 apresenta a comparação entre os valores obtidos e o modelo de  $p_L(\ell)$ . Os valores dos parâmetros após o procedimento de ajuste de curva são  $\alpha = 0,01106$ ,  $\beta = 0,3142$  com  $SSE = 0,04421$  (4,421%),  $RMSE = 0,04956$  (4,956%),  $RS = 0,7673$  (76,73%) e  $ARS = 0,7544$  (75,44%). Observa-se que estes dois últimos valores,  $RS$  e  $ARS$ , estão abaixo dos resultados das medições anteriores. Os resultados são observados na Figura 3.21, por causa do terceiro pico de tráfego de dados na distribuição medida. O surgimento do pico intermediário provavelmente ocorre devido a algum tipo de aplicação específica presente na rede (ataque de negação de serviço, VoIP, Skype, serviços de e-mail, etc). Assim, seria necessário uma análise mais profunda para sua completa caracterização. Outro ponto importante a ser destacado

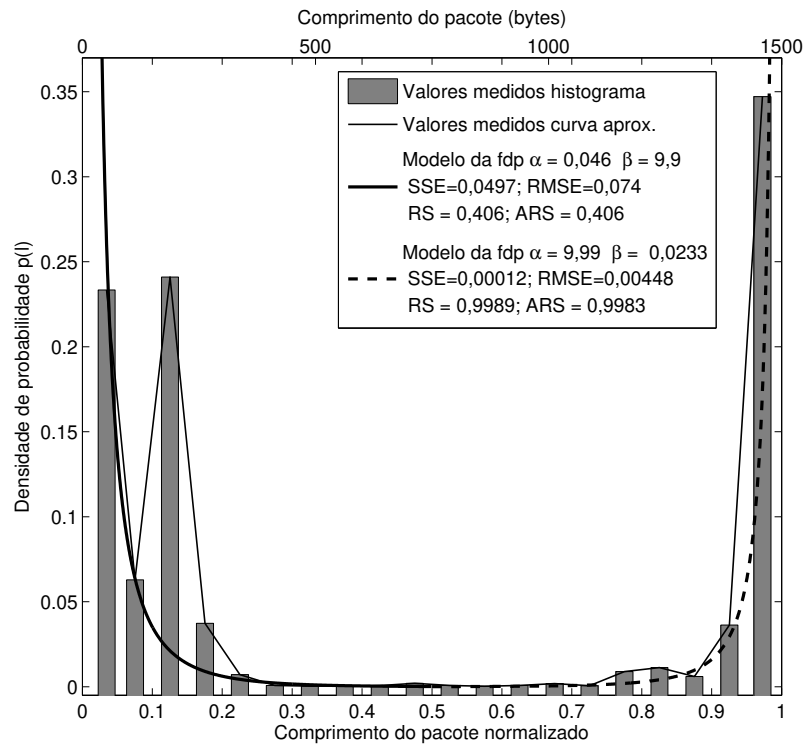
é que dependendo do tipo de aplicação que esteja gerando este tipo do comportamento, talvez seja possível utilizar o modelo de  $p_L(\ell)$  na identificação deste tipo de aplicação na rede. Como já mencionado, para esse tipo de comportamento de tráfego é dado o nome de distribuição trimodal (MCCREARY; CLAFFY, ), (KARAGIANNIS *et al.*, 2003), (YU *et al.*, 2007). O modelo de distribuição de comprimento de pacotes proposto nesta tese é para distribuições de tráfego bimodais justificando assim, a menor precisão do modelo para o caso de tráfego trimodal.

Considerando os dois intervalos, com o objetivo de otimizar os resultados e minimizar os erros, obtém-se os seguintes valores,  $\alpha = 0,046$ ,  $\beta = 9,9$ ,  $SSE = 0,04968$  (4,968%),  $RMSE = 0,0743$  (7,43%),  $RS = 0,4059$  e  $ARS = 0,4059$ , no primeiro intervalo e de  $\alpha = 9,99$ ,  $\beta = 0,0233$ ,  $SSE = 0,0001205$  (0,01205%),  $RMSE = 0,004481$  (0,4481%),  $RS = 0,9989$  (99,89%) e  $ARS = 0,9983$  (99,83%), no segundo intervalo. A Figura 3.22, ilustra os resultados obtidos. Verifica-se que o pico intermediário de dados está dentro do primeiro intervalo. O modelo proposto nesta tese não abrange este tipo de comportamento e a diferença entre os valores medidos e o modelo proposto é bastante acentuada. No entanto, no segundo intervalo, o mesmo não ocorre e a diferença diminui para menos de 0,5% nas métricas utilizadas.



**Figura 3.21** Modelo de densidade de probabilidade versus valores medidos (*Lab Comp*).





**Figura 3.22** Modelo de densidade de probabilidade (dois intervalos) versus valores medidos (*Lab Comp*).

### Comprimento de pacotes–*Indústria*

Para a sexta coleta de dados foram utilizados os dados de tráfego de um servidor em uma indústria (Figura 3.23). Esse servidor tem uma conexão via modem ADSL de 1 Mbit/s com a Internet e é um *gateway* de acesso para 80 computadores distribuídos em 5 salas da empresa. Nomeado “Indústria”, o conjunto de dados busca verificar o comportamento do comprimento de pacotes para um fluxo de dados maior devido ao número de computadores da rede.

```
Running time: 90960 seconds
*** Packet Size Distribution, generated Wed Oct 28 15:42:50 2009

Interface: eth0  MTU: 1500

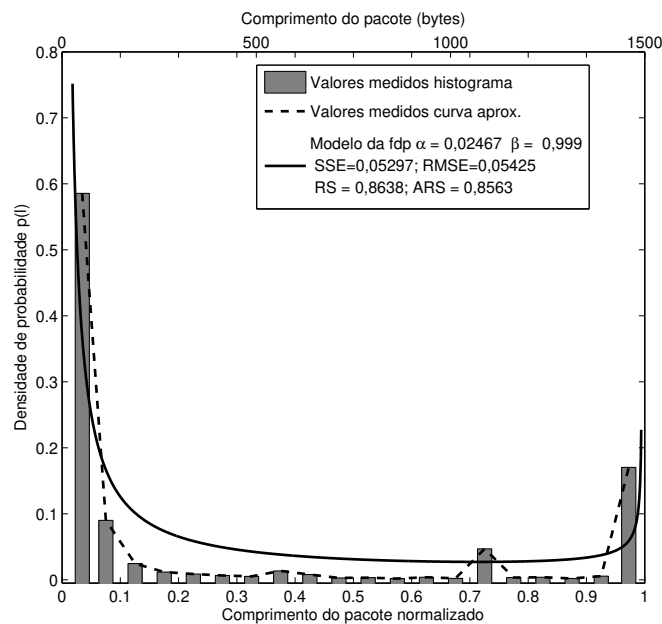
Packet Size (bytes)    Count
1 to 75:              5942389
76 to 150:            913553
151 to 225:           249872
226 to 300:           120022
301 to 375:           89056
376 to 450:           65876
451 to 525:           52794
526 to 600:           136097
601 to 675:           79902
676 to 750:           29466
751 to 825:           32219
826 to 900:           19521
901 to 975:           39866
976 to 1050:          22256
1051 to 1125:         476203
1126 to 1200:         33550
1201 to 1275:         39897
1276 to 1350:         20431
1351 to 1425:         57321
1426 to 1500:        1729663

Running time: 91005 seconds
Wed Oct 28 15:42:50 2009; ***** Packet size distribution facility stopped *****
```

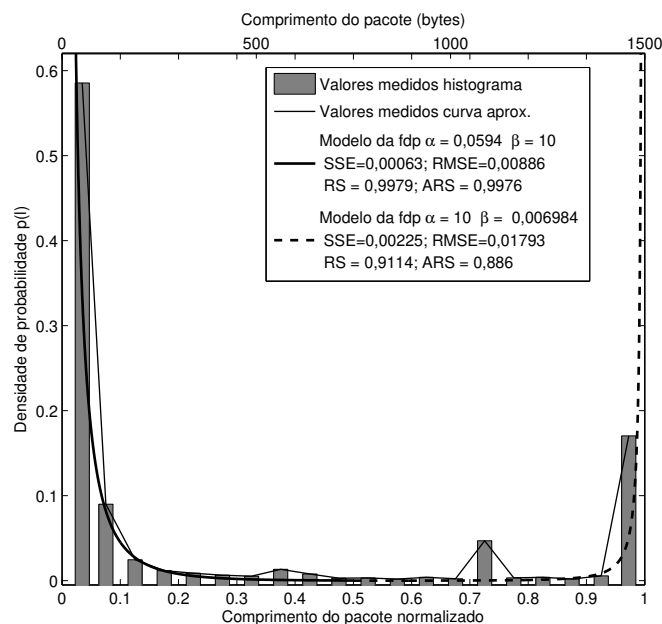
**Figura 3.23** Tabela com o arquivo *log* em formato texto gerado pelo programa IPTRAF–*Indústria*.

O conjunto de dados apresenta uma distribuição trimodal, mas com um terceiro pico de dados de baixa intensidade. Isso favorece a aproximação da distribuição por um modelo de tráfego bimodal. A Figura 3.24 apresenta a comparação entre os valores obtidos e o modelo de  $p_L(\ell)$ . Os valores dos parâmetros após o procedimento de ajuste de curva são  $\alpha = 0,02467$ ,  $\beta = 0,999$  com SSE= 0,04521 (4,521%), RMSE= 0,04878 (4,878%), RS= 0,8644 (86,44%) e ARS= 0,8644 (86,44%). Considerando os dois intervalos para otimizar os resultados e minimizar os erros, os valores do primeiro intervalo são  $\alpha = 0,0594$ ,  $\beta = 10$ , SSE= 0,000628 (0,0628%), RMSE= 0,00886 (0,886%), RS= 0,9979 e ARS= 0,9976. E para o segundo intervalo,  $\alpha = 10$ ,  $\beta = 0,006984$ , SSE= 0,00225 (0,225%), RMSE= 0,01793 (1,793%), RS= 0,9114 (91,14%) e ARS= 0,8861 (88,61%). A Figura 3.25, ilustra os resultados obtidos. O surgimento do pico intermediário, provavelmente ocorre devido alguma aplicação específica

presente na rede da empresa. No caso, existem 6 servidores (e-mail, web, ftp, gateway, banco de dados e arquivos de usuários e de impressão) e 5 deles com acesso remoto (Internet => LAN). Assim, seria necessário uma análise mais profunda para caracterizar o pico de dados próximo a 1100 bytes. Verifica-se que mesmo para uma distribuição trimodal, se houver um pico intermediário com valor máximo próximo a curva descrita pelo modelo de  $p_L(\ell)$ , ou seja, de baixa intensidade, o modelo bimodal proposto apresenta uma boa aproximação.



**Figura 3.24** Modelo de densidade de probabilidade versus valores medidos (*Indústria*).



**Figura 3.25** Modelo de densidade de probabilidade (dois intervalos) versus valores medidos (*Indústria*).

### 3.3 Medidas cumulativas dos comprimentos dos pacotes

Nesta seção são apresentados os histogramas cumulativos dos dados apresentados na seção anterior. Também estão inclusos vários resultados da distribuição cumulativa do comprimento dos pacotes obtidos por outros autores.

#### 3.3.1 Medições de dados cumulativos

As comparações entre o modelo proposto para a função cumulativa de probabilidade  $P_L(\ell)$  dada pela Equação (2.45) e os dados da Seção 3.2 são apresentados nesta subseção. Os resultados são ilustrados nas Figuras de 3.27 a 3.32. Nessas figuras, os resultados são representados por três gráficos. O primeiro em barra, mostra o histograma dos valores cumulativos medidos da seção 3.2. O segundo, em linha tracejada, ilustra a curva aproximada dos valores cumulativos destes mesmos valores. Por fim, o terceiro, em linha escura contínua, o resultado da função de distribuição cumulativa  $P_L(\ell)$  proposto nesta tese, Equação (2.45).

#### Dados Cumulativos–*Diversos*

Para o primeiro conjunto, *Diversos*, os dados cumulativos são comparados com os resultados obtidos de  $P_L(\ell)$  dado pela Equação (2.45). A Figura 3.26 apresenta os resultados obtidos em cinco gráficos. O primeiro, em barra, é o histograma dos valores cumulativos das medidas. Em linha fina contínua, a curva aproximada para os valores cumulativos. O terceiro, em linha contínua grossa, o resultado obtido de  $P_L(\ell)$  ajustado pelo método dos mínimos quadrados para encontrar os melhores valores dos parâmetros  $\alpha$  e  $\beta$ , considerando  $\alpha, \beta > 0$ . Em linha tracejada grossa, o resultado obtido de  $P_L(\ell)$  usando os valores de  $\alpha$  e  $\beta$  obtidos com a Equação (2.26),  $p_L(\ell)$ , em um intervalo (Figura 3.6). E finalmente, linha tracejada fina,  $P_L(\ell)$  usando os valores de  $\alpha$  e  $\beta$  obtidos com a equação de  $p_L(\ell)$  em dois intervalos (Figura 3.7).

Gráficamente, o valor de  $P_L(\ell)$  usando os valores de  $\alpha$  e  $\beta$  obtidos na Figura 3.6 apresenta valores bem acima do valores experimentais. Estes resultados são comprovados pelos valores numéricos das métricas  $SSE= 4,31$ ,  $RMSE= 0,5035$ <sup>1</sup>. Existem duas justificativas para esta diferença. A primeira é que na parte central do gráfico (0,1–0,9) da figura 3.6, o  $p_L(\ell)$  está acima dos valores reais. Isso se reflete no gráfico cumulativo com valores acima dos valores reais. A segunda justificativa é que  $p_L(\ell)$  possui duas indeterminações em  $\ell = L_m$  e  $\ell = L_M$ , por isso usa-se  $\ell = L_m + \varepsilon$  e  $\ell = L_M - \varepsilon$ , com  $\varepsilon \rightarrow 0$ . Essa indeterminação dificulta a representação, de forma mais aproximada, dos dois picos da curva, justificando o salto maior na função cumulativa.

No gráfico de  $P_L(\ell)$  usando os valores de  $\alpha$  e  $\beta$  obtidos na Figura 3.7, dois intervalos, apresenta valores abaixo do valores experimentais. Isso ocorre por causa da diferença numérica entre o primeiro pico, próximo a 40 bytes, e a curva do modelo  $p_L(\ell)$ . Não são utilizados os

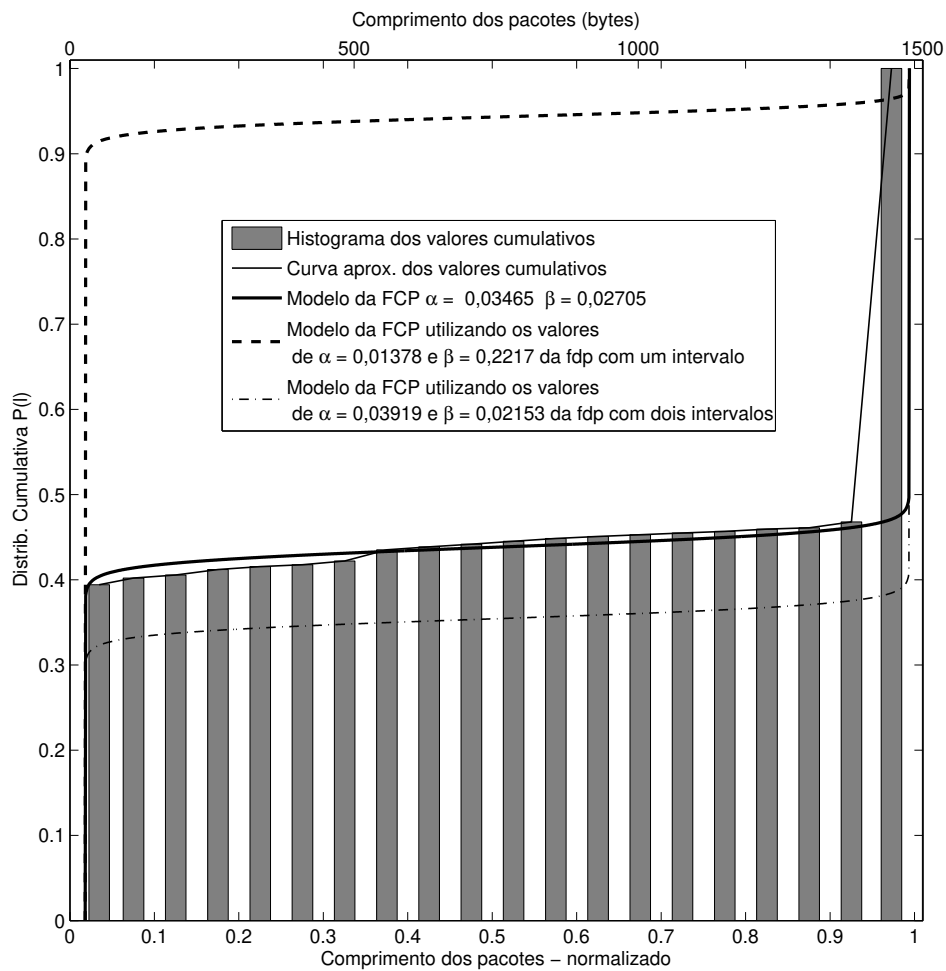
<sup>1</sup>Não foi possível obter os valores das métricas RS e ARS

valores  $\alpha = 10$  e  $\beta = 10$  na função cumulativa e os resultados das métricas são  $SSE = 0,1199$ ,  $RMSE = 0,08398$ <sup>2</sup>. Observa-se que o crescimento do valor de  $\alpha$  faz com que o primeiro salto da função cumulativa diminua, ou seja, inversamente proporcional à amplitude do primeiro salto. Para o valor de  $\beta$ , a diminuição de seu valor faz com que o segundo salto da função cumulativa aumente, ou seja, inversamente proporcional à amplitude do segundo salto.

E finalmente, o gráfico obtido de  $P_L(\ell)$  ajustado pelo método dos mínimos quadrados para encontrar os melhores valores dos parâmetros  $\alpha$  e  $\beta$  é ilustrado pela linha contínua grossa. Verifica-se que as igualdades  $I_0(\alpha, \beta) = 0$  e  $I_1(\alpha, \beta) = 1$  são duas propriedades importantes porque  $x = 0$  e  $x = 1$  podem ser obtidos quando se faz  $\ell = L_m + \varepsilon$  e  $\ell = L_M - \varepsilon$ , com  $\varepsilon \rightarrow 0$ , ou seja,  $x = 0$  para  $\ell = L_m$  e  $x = 1$  para  $\ell = L_M$ . Na prática isto significa que no processo de aproximação dos dados pelo modelo de  $P_L(\ell)$ , esses dois valores e outros que estejam muito próximos a eles, são excluídos. Em seguida, com os valores restantes, utilizam-se o método dos mínimos quadrados para encontrar os melhores valores dos parâmetros  $\alpha$  e  $\beta$ , com  $\alpha, \beta > 0$ . As métricas utilizadas para comparar os valores experimentais e os teóricos foram as mesmas da seção anterior, SSE, RMSE, RS e ARS (MATLAB, 2010). Os resultados obtidos são  $\alpha = 0,03465$ ,  $\beta = 0,02705$ ,  $SSE = 0,0003979$  (0,03979%),  $RMSE = 0,004838$  (0,04838%),  $RS = 0,9564$  (95,64%) e  $ARS = 0,9538$  (95,38%). Os valores obtidos nas métricas SSE e RMSE estão próximos a zero, indicando uma excelente aproximação entre os valores medidos e os valores teóricos obtidos com o modelo proposto. Este resultado é reforçado pelos valores de RS e ARS, acima dos 95%, ou seja, próximos a um e sem a necessidade de divisão do intervalo em dois subintervalos, como realizado para o modelo de  $p_L(\ell)$  (Figura 3.7). Graficamente, Figura 3.26, o modelo cumulativo capturou bem o comportamento bimodal do tráfego e isto se refletiu nos menores valores dos erros calculados. Outro ponto a ser destacado é a maior aderência da curva aos resultados teóricos, principalmente na região central do gráfico.

---

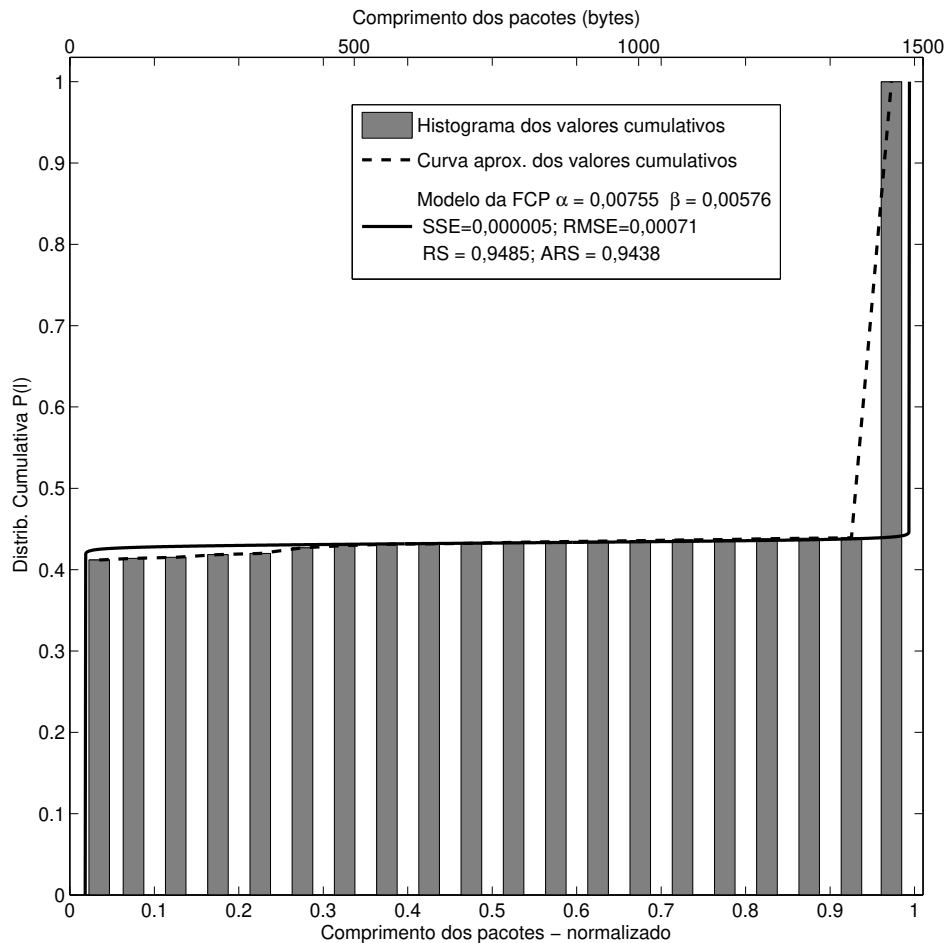
<sup>2</sup>Não foi possível obter os valores das métricas RS e ARS



**Figura 3.26** Modelo de distribuição cumulativa versus valores medidos (*Diversos*).

### Dados Cumulativos–YouTube

O segundo conjunto de dados cumulativos é o *YouTube* e os resultados obtidos após o ajuste dos parâmetros são:  $\alpha = 0,007552$ ,  $\beta = 0,005761$ ,  $SSE= 0,000005535$ ,  $RMSE= 0,0007093$ ,  $RS= 0,9485$  (94,85%) e  $ARS= 0,9438$  (94,38%). Novamente, os valores de SSE e RMSE estão muito próximos a zero e com RS e ARS, acima dos 94%. Graficamente na Figura 3.27, o modelo cumulativo reproduz bem o comportamento bimodal do tráfego.

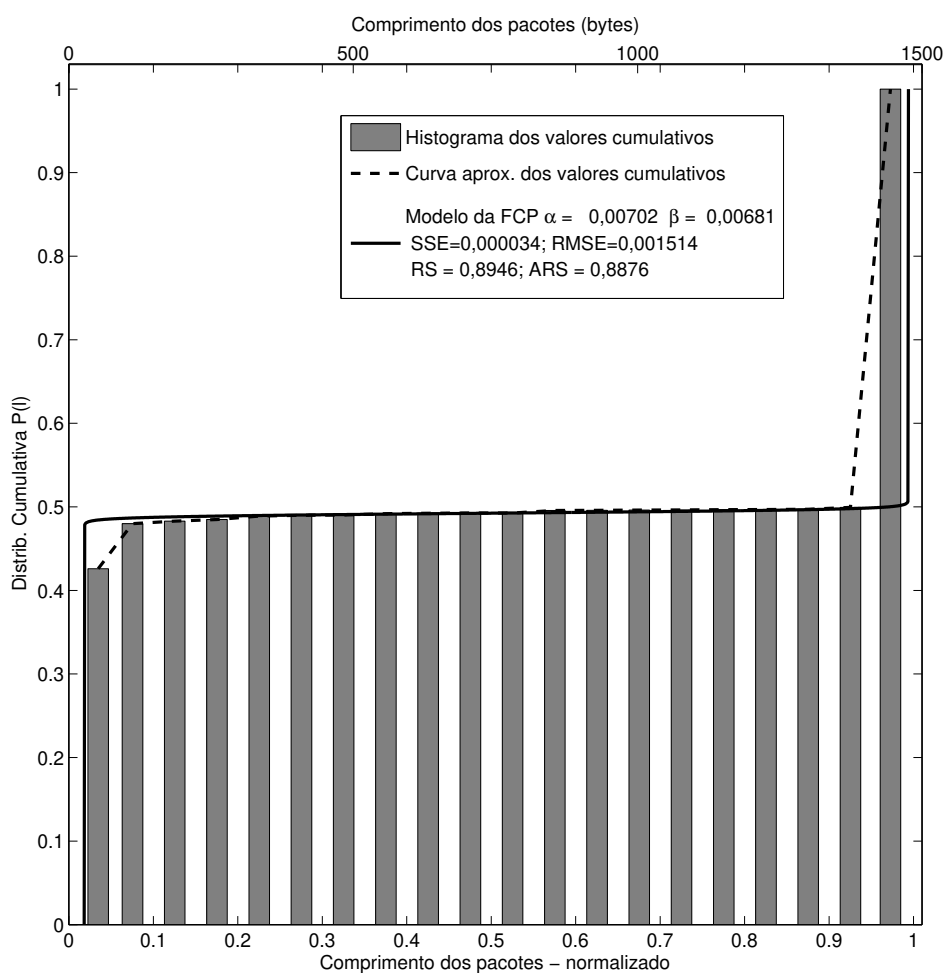


**Figura 3.27** Modelo de distribuição cumulativa versus valores medidos (*YouTube*).

### Dados Cumulativos–*Torrent*

O terceiro conjunto de dados cumulativos são os dados dos subconjuntos *Torrent* e *Torrent2*. Para o primeiro subconjunto, *Torrent*, os resultados obtidos com o ajuste dos parâmetros são:  $\alpha = 0,007016$ ,  $\beta = 0,006808$ ,  $SSE = 0,0000344$ ,  $RMSE = 0,001514$ ,  $RS = 0,8946$  (89,46%) e  $ARS = 0,8876$  (88,76%). Os valores de SSE e RMSE continuaram muito próximos a zero, com RS e ARS, acima dos 88%.

Para o segundo subconjunto, *Torrent2*, os resultados são:  $\alpha = 0,005133$ ,  $\beta = 0,003758$ ,  $SSE = 0,00001536$ ,  $RMSE = 0,001012$ ,  $RS = 0,8826$  e  $ARS = 0,8748$ . Graficamente os resultados são ilustrados nas Figuras 3.28 e 3.28. Apesar da pequena queda nos valores de RS e ARS, a função cumulativa reproduziu bem o comportamento bimodal do tráfego em ambos os casos.



**Figura 3.28** Modelo de distribuição cumulativa versus valores medidos (*Torrent*).



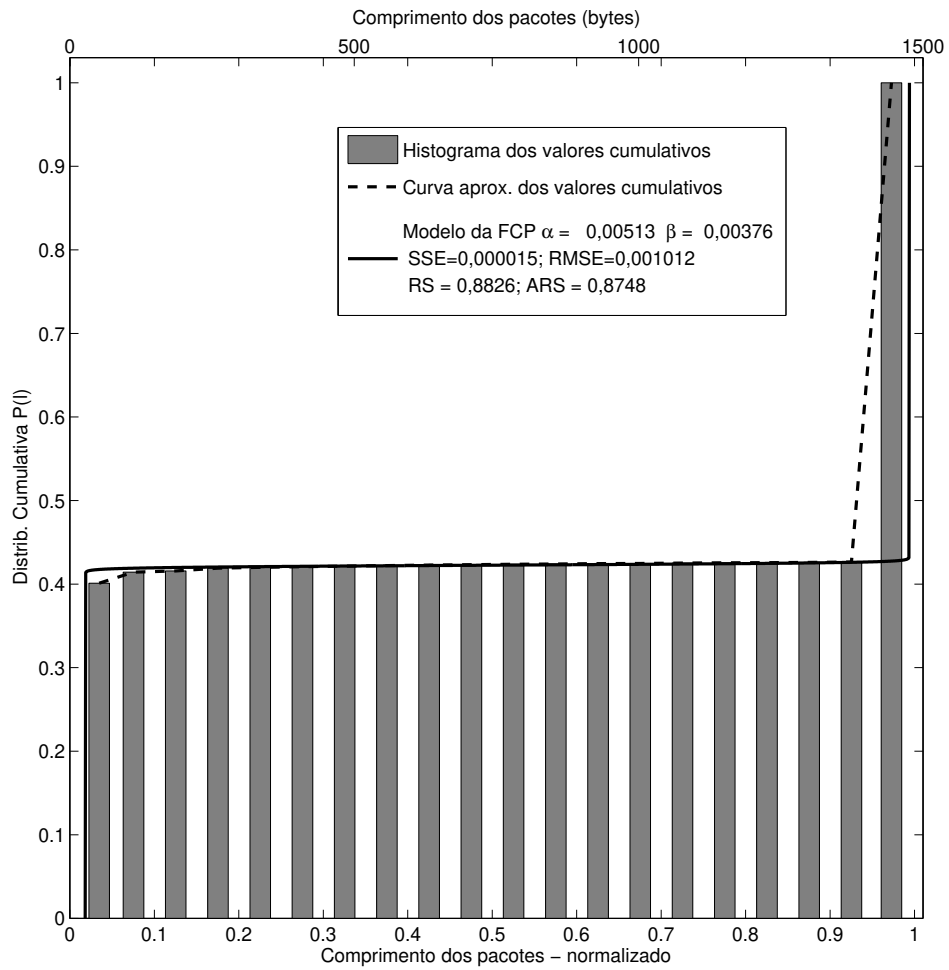
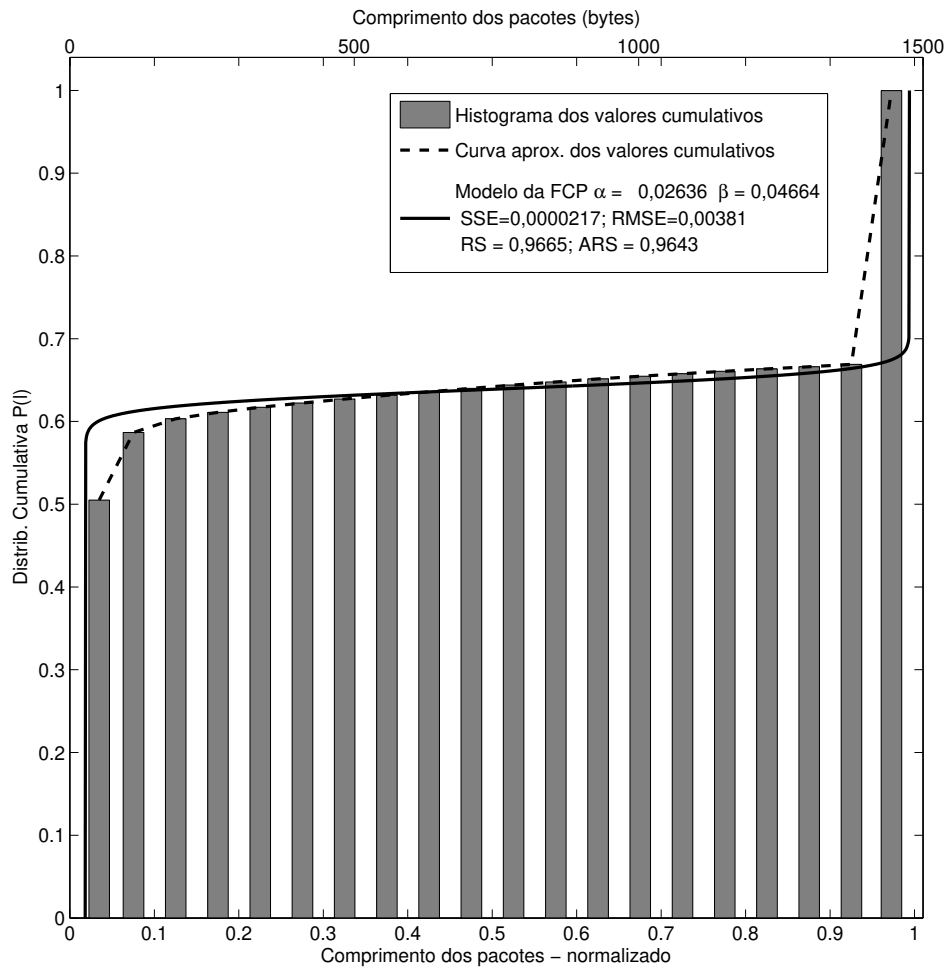
Dados Cumulativos–*Torrent2*

Figura 3.29 Modelo de distribuição cumulativa versus valores medidos (*Torrent2*).

### Dados Cumulativos–FTP

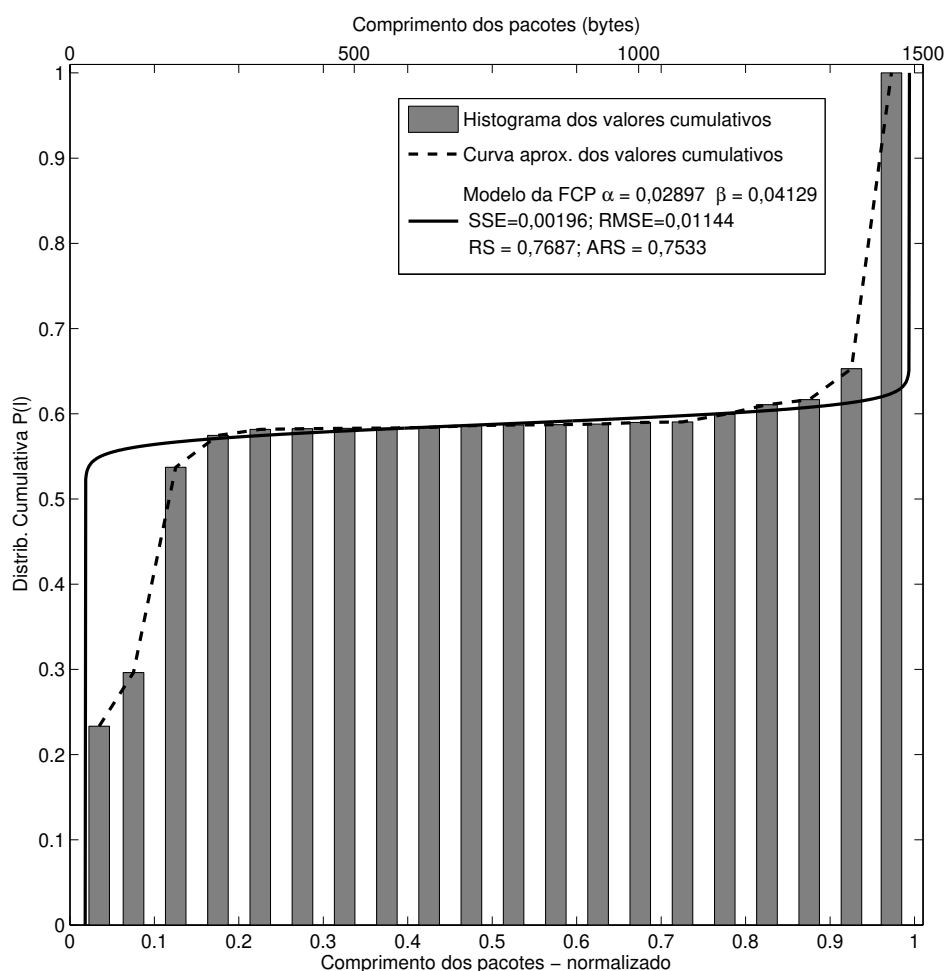
O quarto conjunto de dados cumulativos é o *FTP* com os resultados:  $\alpha = 0,02636$ ,  $\beta = 0,04664$ ,  $SSE = 0,0002173$ ,  $RMSE = 0,003806$ ,  $RS = 0,9665$  (96,65%) e  $ARS = 0,9643$  (96,43%). Novamente, os valores de  $SSE$  e  $RMSE$  estão muito próximos a zero, com  $RS$  e  $ARS$ , acima dos 96%. Na Figura 3.30, o modelo  $P_L(\ell)$  reproduz de forma aproximada o comportamento bimodal do tráfego.



**Figura 3.30** Modelo de distribuição cumulativa versus valores medidos (*FTP*).

### Dados Cumulativos–*Lab Comp*

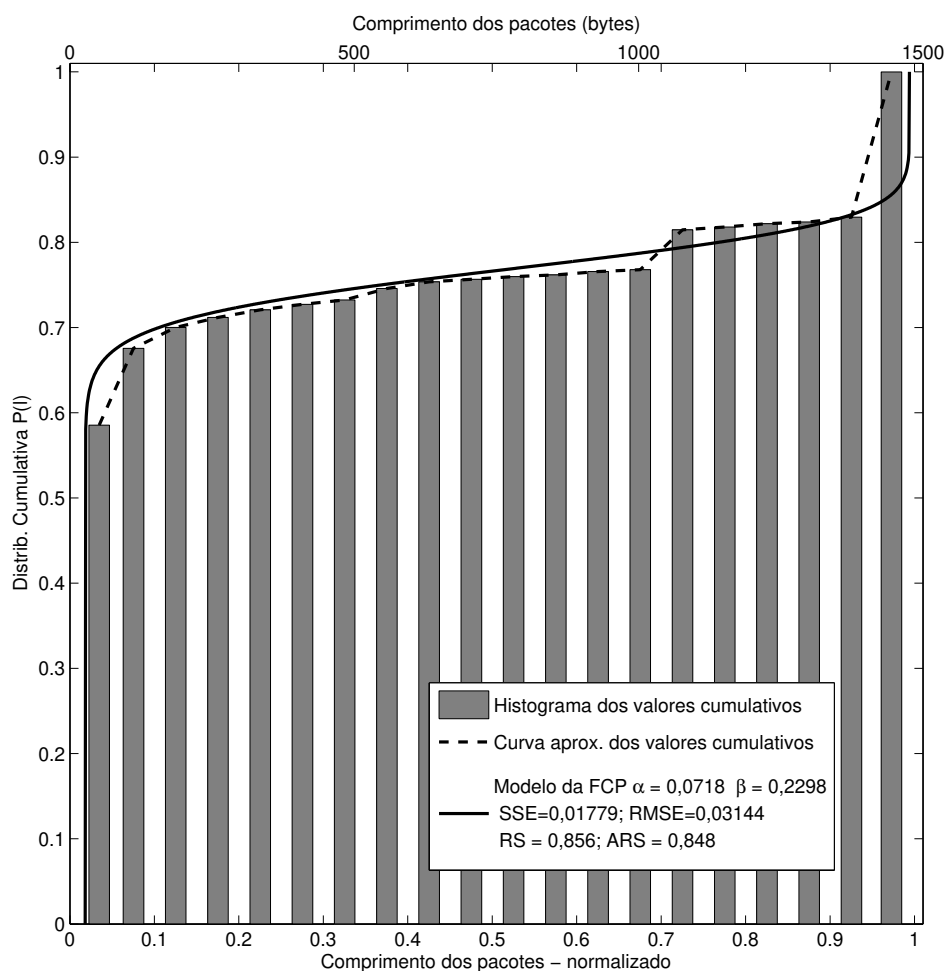
Para o quinto conjunto de dados cumulativos, *Lab Comp*, os resultados são:  $\alpha = 0,02897$ ,  $\beta = 0,04129$ ,  $SSE = 0,001962$ ,  $RMSE = 0,01144$ ,  $RS = 0,7687$  (76,87%) e  $ARS = 0,7533$  (75,33%). Apesar dos valores de SSE e RMSE próximos de zero, os valores de RS e ARS, estão próximos a 76%. Esses resultados estão na mesma faixa de valores apresentados quando comparados aos dados de densidade com o modelo de  $p_L(\ell)$ . A maior diferença dos valores se deve ao aparecimento do pico intermediário, distribuição trimodal, como apresentado anteriormente e ilustrado nas Figuras 3.21 e 3.22. Na Figura 3.31, o valor próximo a 0,1 do eixo horizontal ilustra bem o impacto do pico intermediário na distribuição cumulativa e a diferença para o modelo bimodal. A curva  $P_L(\ell)$  tenta se ajustar a maior parte do dados, gerando algumas diferenças, entre os valores medidos e a curva teórica.



**Figura 3.31** Modelo de distribuição cumulativa versus valores medidos (*Lab Comp*).

**Dados Cumulativos–Indústria**

O sexto e último conjunto de dados cumulativos desta sequência, é chamado de *Indústria* e seus resultados são:  $\alpha = 0,07183$ ,  $\beta = 0,2298$ ,  $SSE= 0,01779$ ,  $RMSE= 0,03144$ ,  $RS= 0,8561$  (85,61%),  $ARS= 0,8481$  (84,81%) e apresentados na Figura 3.32. Semelhante aos resultados das Figuras 3.24 e 3.25, quando se tem uma distribuição de tráfego trimodal, mas com o terceiro pico de baixa intensidade, o modelo bimodal da distribuição cumulativa consegue representar com boa aproximação os valores medidos.



**Figura 3.32** Modelo de distribuição cumulativa versus valores medidos (*Indústria*).

### 3.3.2 Dados obtidos da literatura

Nesta seção tem início as comparações entre a distribuição cumulativa  $P_L(\ell)$  dada, pela Equação (2.45), e as medições obtidas da literatura.

O primeiro conjunto de dados pode ser encontrado no sítio de Ville Mattila (MATTILA, 2010). Ele descreve a distribuição do comprimento dos pacotes presentes no tráfego de uma conexão de Internet utilizando linha digital para assinante (DSL) em uma configuração semelhante a da Figura 2.1. A principal diferença, do caso descrito, é a substituição ou troca do servidor no ponto de (des)agregação do tráfego por um modem/roteador de linha digital assimétrica para assinante (ADSL).

Na estatística de coleta do comprimento dos pacotes da rede foi utilizada uma resolução de 16 *bytes*. Essas estatísticas de tráfego foram coletadas em um período de 14 dias entre 11 e 24 de setembro de 2003.

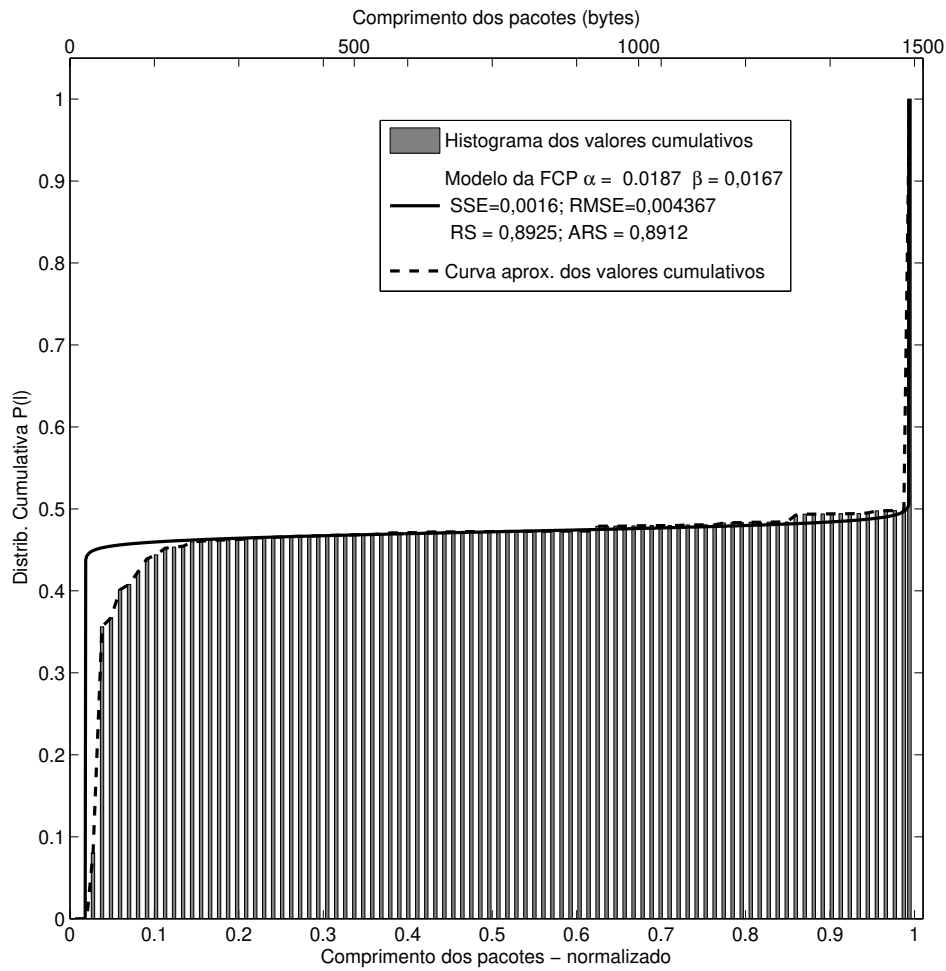
Algumas das características dos dados coletados (MATTILA, 2010):

- Entre cinco e dez conexões SSH abertas por dia,
- Algumas centenas de megabytes transferidos utilizando um FTP seguro durante o período de amostragem,
- Os dados de SMTP são de caixas de *e-mails* de cinco usuários,
- Os serviços de DNS são utilizados apenas pelos serviços de SMTP e HTTP,
- O tráfego do HTTP foi coletado de dois servidores com baixo tráfego em seus sítios.

As Figuras 3.33 a 3.39 ilustram os valores obtidos (MATTILA, 2010) e apresentados na forma de histogramas cumulativos. Ao longo do texto, esses são referidos como “Internet” seguido do tipo de dados utilizados. Por exemplo, “Internet TCP” significa o conjunto coletado por Ville Mattila e disponibilizados na Internet para o tráfego do protocolo TCP. Nessas figuras, as espessuras das barras dos histogramas são proporcionais à 16 *bytes* utilizados nas medições de cada intervalo de comprimento dos pacotes. A distribuição cumulativa do comprimento dos pacotes é indicada no eixo vertical. Esses e outros procedimentos estão descritos passo-a-passo na seção anterior.

#### Dados Cumulativos–Internet IP

O primeiro conjunto de dados desta sequência, é o chamado de *Internet IP* (MATTILA, 2010). Os resultados obtidos na comparação com  $P_L(\ell)$  são:  $\alpha = 0,0187$ ,  $\beta = 0,01673$ ,  $SSE = 0,001602$  (0,1602%),  $RMSE = 0,004367$  (0,4367%),  $RS = 0,8925$  (89,25%),  $ARS = 0,8912$  (89,12%) e apresentados na Figura 3.33. Nas duas primeiras métricas usadas, os resultados estão abaixo de 0,5% e nas duas últimas, os resultados estão acima de 89%. Semelhante aos

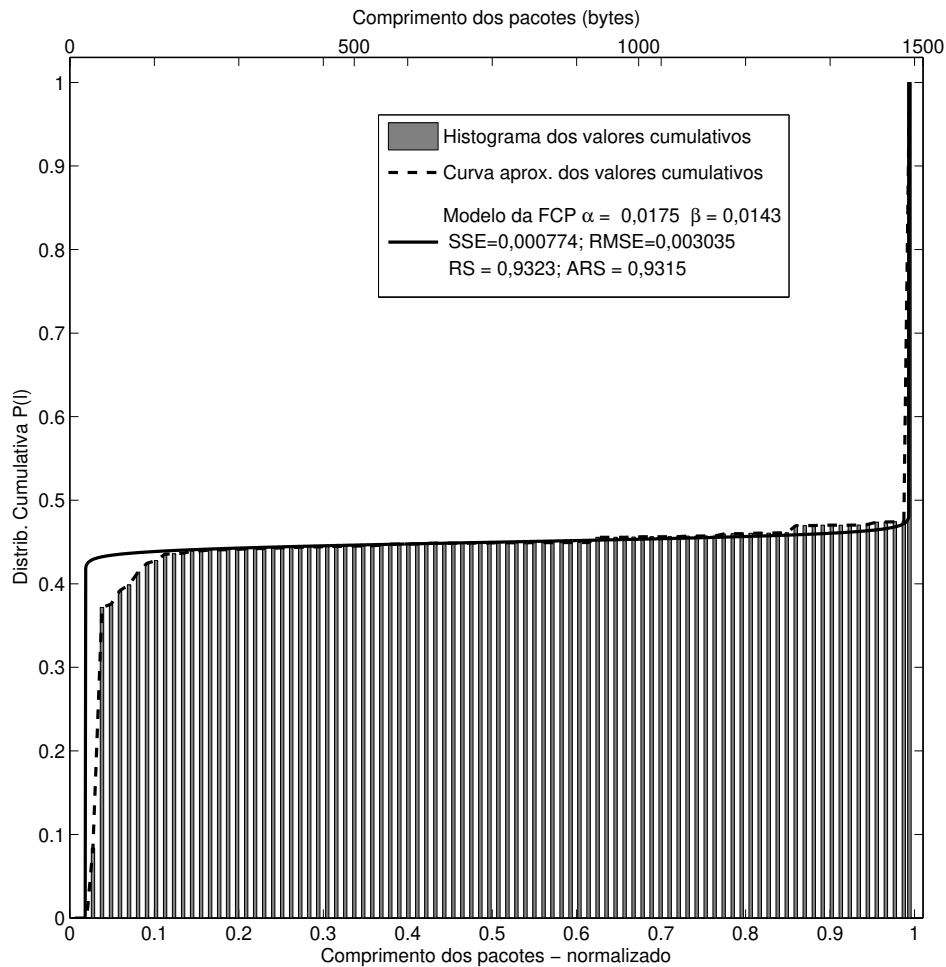


**Figura 3.33** Modelo de distribuição cumulativa versus valores medidos (*IP*).

resultados obtidos anteriormente, o modelo proposto adere bem aos valores experimentais comprovado pelos bons resultados gráficos e calculados.

**Dados Cumulativos–Internet TCP**

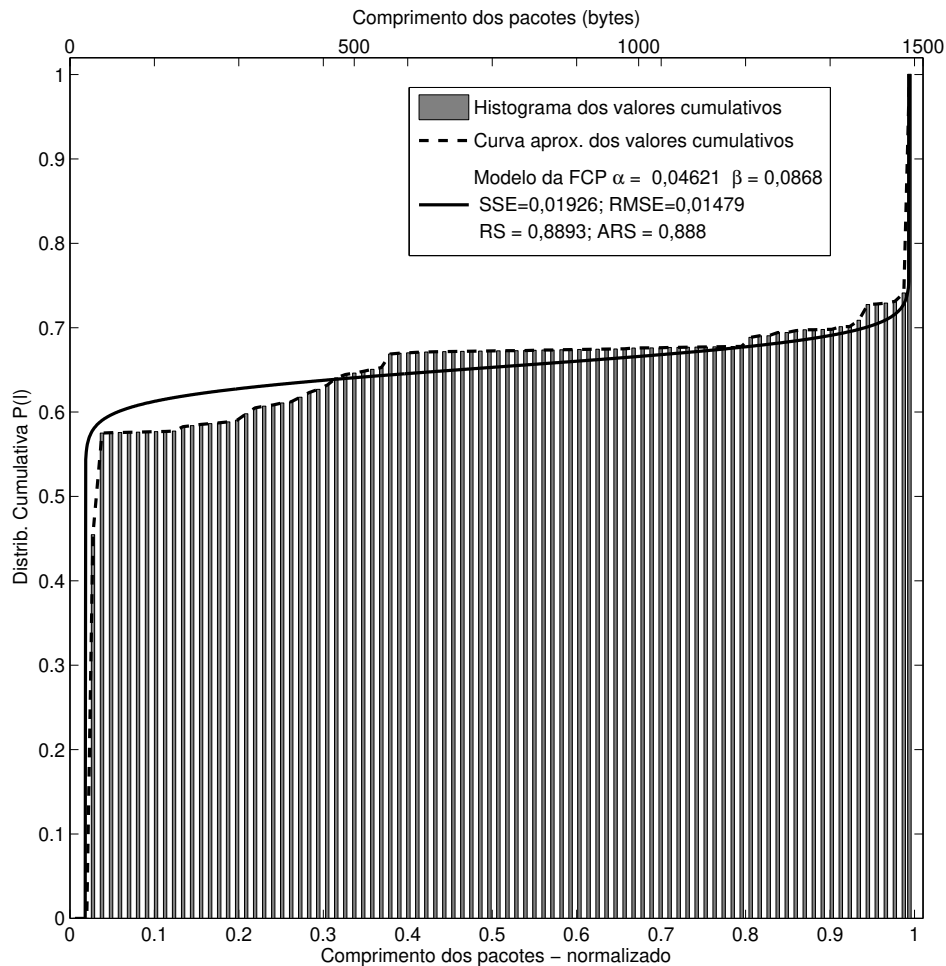
No segundo conjunto de dados, chamado de *Internet TCP* (MATTILA, 2010). A comparação com  $P_L(\ell)$  resulta em  $\alpha = 0,0175$ ,  $\beta = 0,01431$ ,  $SSE= 0,0007737$ ,  $RMSE= 0,003035$ ,  $RS= 0,9323$ ,  $ARS= 0,9315$  e apresentados na Figura 3.34. As duas primeiras métricas usadas, estão muito próximas a zero e nas duas últimas, os resultados estão acima de 93%. Estes resultados reforçam a boa aproximação entre o modelo cumulativo e os valores experimentais.



**Figura 3.34** Modelo de distribuição cumulativa versus valores medidos (*TCP*).

**Dados Cumulativos–Internet HTTP**

Para o terceiro conjunto de dados, chamado de *Internet HTTP* (MATTILA, 2010), os resultados da comparação com  $P_L(\ell)$  são  $\alpha = 0,04621$ ,  $\beta = 0,0868$ ,  $SSE= 0,01926$ ,  $RMSE= 0,01479$ ,  $RS= 0,8893$  e  $ARS= 0,888$ . E a Figura 3.35 ilustra estes resultados. Esta é mais uma comparação entre valores experimentais e valores teóricos com bons resultados de aproximação.

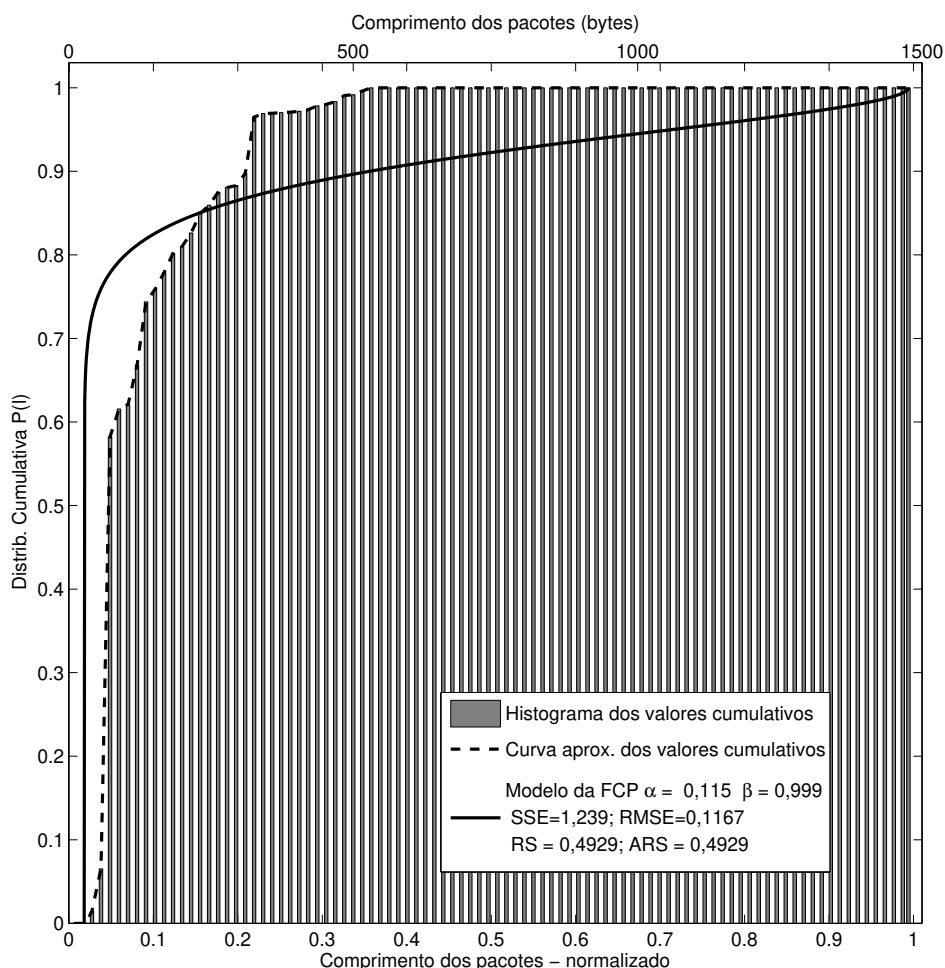


**Figura 3.35** Modelo de distribuição cumulativa versus valores medidos (*HTTP*).



### Dados Cumulativos–Internet UDP

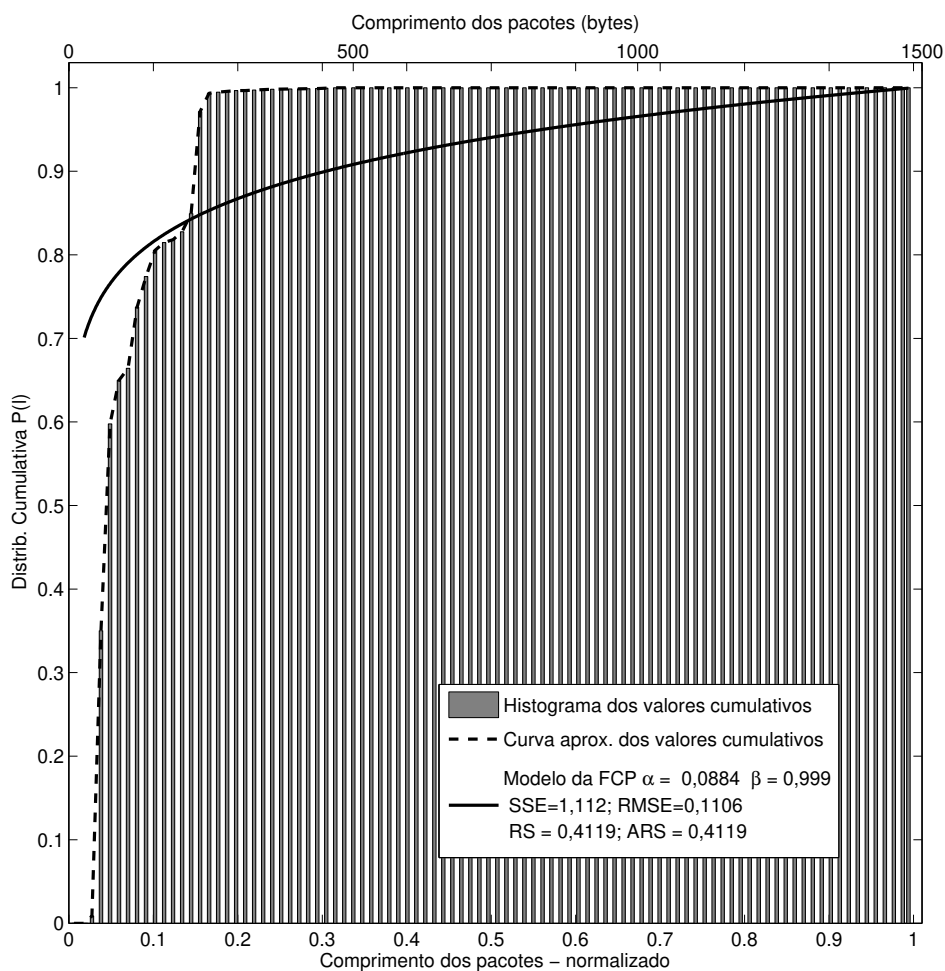
No quarto conjunto de dados, chamado de *Internet UDP* (MATTILA, 2010), os resultados comparativos com  $P_L(\ell)$  são  $\alpha = 0,115$ ,  $\beta = 0,999$ ,  $SSE= 1,239$ ,  $RMSE= 0,1167$ ,  $RS= 0,4929$  e  $ARS= 0,4929$ . É ilustrado na Figura 3.36. Neste caso o modelo proposto não se adaptou bem aos resultados. Os possíveis motivos são o comportamento monomodal e a inexistência de pacotes de comprimento acima de 576 bytes. O motivo dessa limitação de comprimento de pacotes necessitaria de uma análise mais profunda do desse protocolo e na realização de mais medições de comprimento de pacotes, focados na sua caracterização. Essa situação, monomodal, é bem diferente da condição essencial de validade em que o modelo desta tese está embasado, a distribuição bimodal do comprimento dos pacotes. O comportamento monomodal é melhor modelado pelas distribuições Exponencial, Log-normal, Weibull e Pareto.



**Figura 3.36** Modelo de distribuição cumulativa versus valores medidos (*UDP*).

**Dados Cumulativos–Internet DNS**

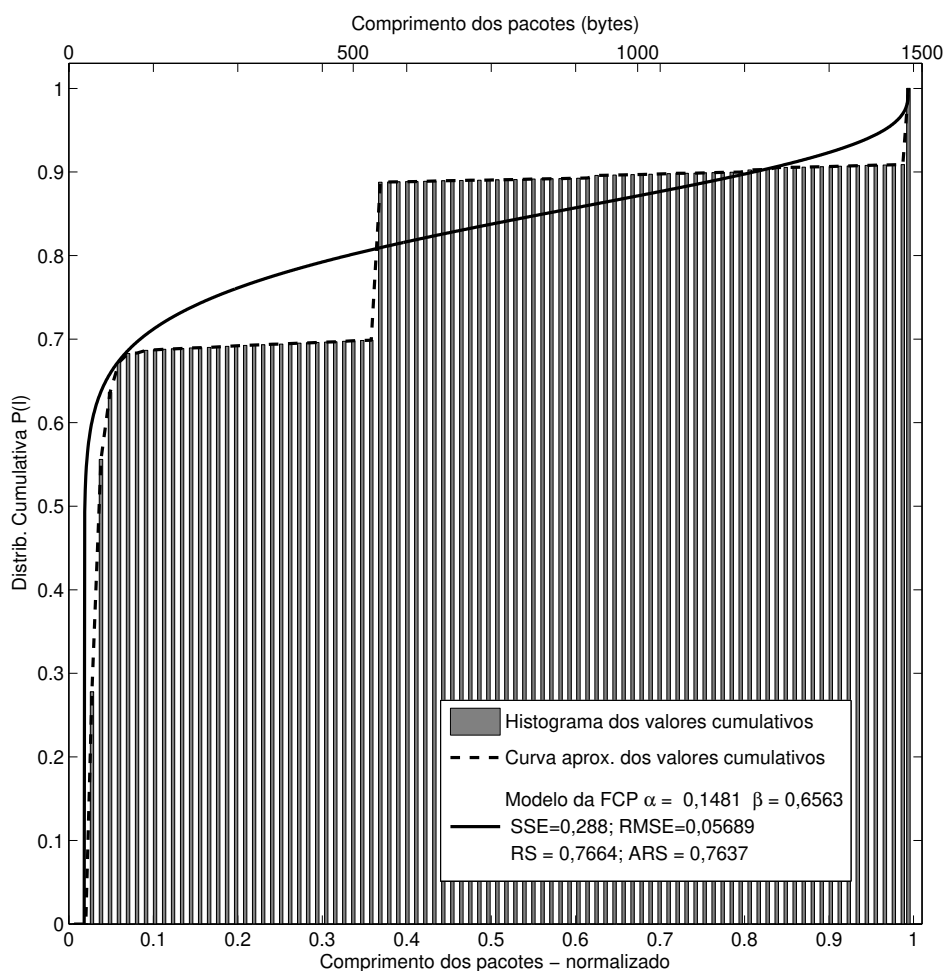
Como quinto conjunto de dados, *Internet DNS* (MATTILA, 2010), os resultados da comparação com  $P_L(\ell)$  são:  $\alpha = 0,0884$ ,  $\beta = 0,999$ ,  $SSE= 1,112$ ,  $RMSE= 0,1106$ ,  $RS= 0,4119$  e  $ARS= 0,4119$ . A Figura 3.37 ilustra os resultados comparativos. Situação semelhante ao conjunto anterior com distribuição monomodal, pacotes de comprimento limitados a 300 bytes e possibilidade de ser melhor modelado pelas distribuições Exponencial, Log-normal, Weibull e Pareto.



**Figura 3.37** Modelo de distribuição cumulativa versus valores medidos (*DNS*).

**Dados Cumulativos–Internet SMTP**

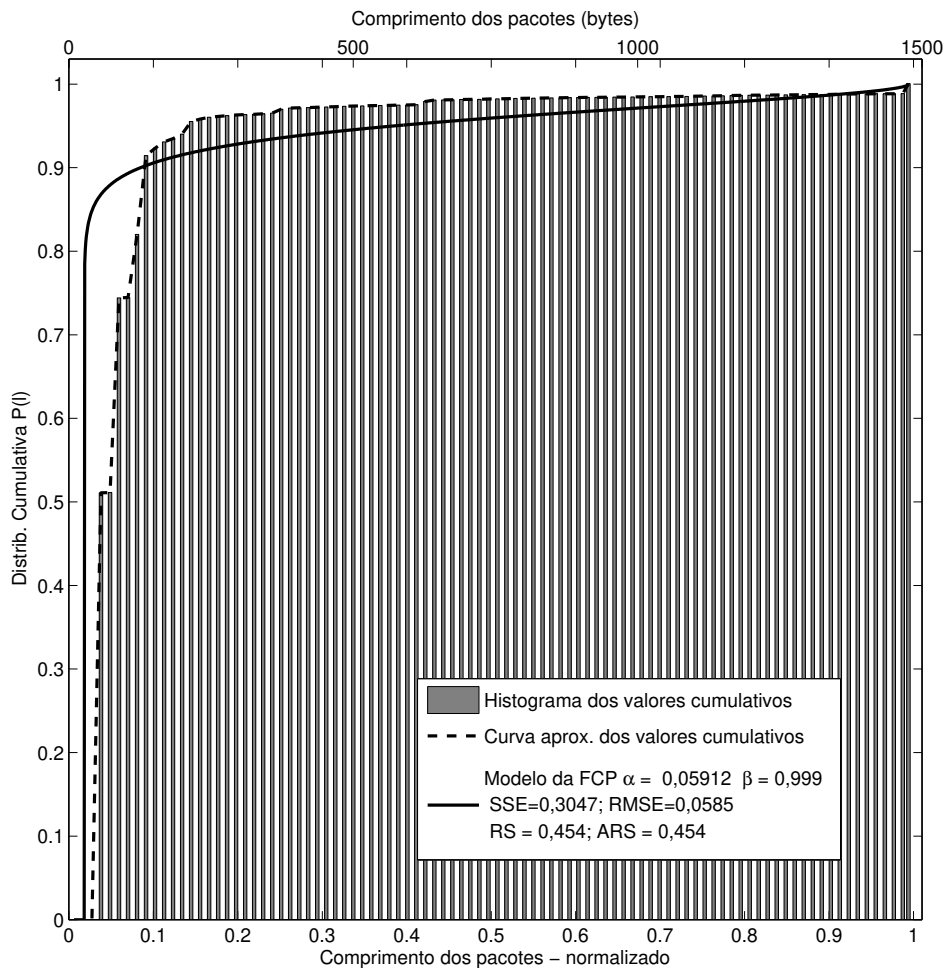
No sexto conjunto de dados, *Internet SMTP* (MATTILA, 2010), a sua comparação com a curva de  $P_L(\ell)$  tem os seguintes resultados  $\alpha = 0,1481$ ,  $\beta = 0,6563$ ,  $SSE= 0,288$ ,  $RMSE= 0,05689$ ,  $RS= 0,7664$  e  $ARS= 0,7637$ . A Figura 3.38 ilustra a comparação dos resultados. Situação em que ocorre a distribuição trimodal com pico intermediário de alta intensidade, salto grande de valores. É semelhante aos resultados obtidos da comparação com o modelo de  $p_L(\ell)$  com o modelo proposto não sendo adequado na sua caracterização.



**Figura 3.38** Modelo de distribuição cumulativa versus valores medidos (*SMTP*).

### Dados Cumulativos–Internet SSH

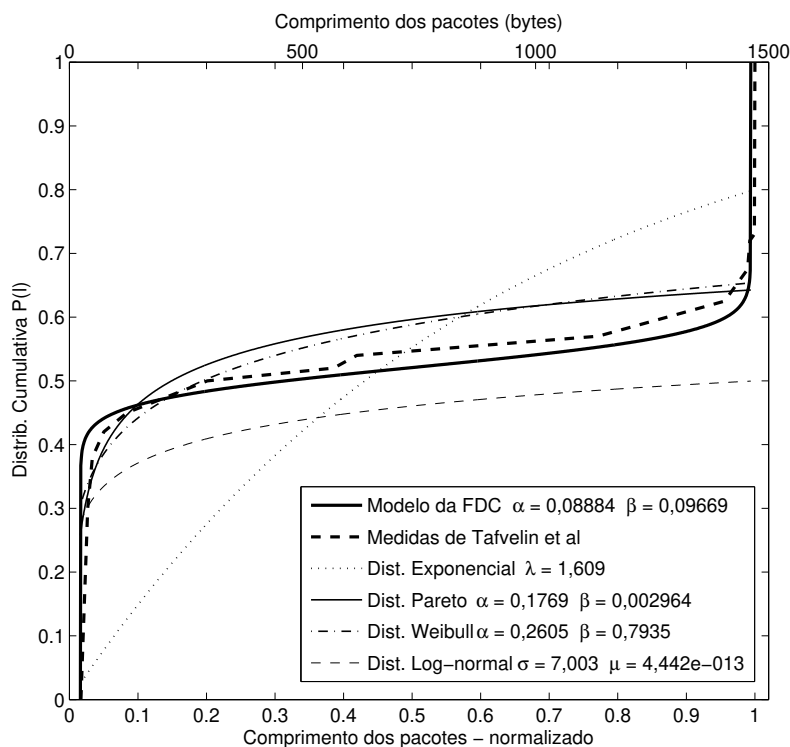
Finalmente, o sétimo conjunto de dados, *Internet SSH* (MATTILA, 2010), tem os parâmetros  $\alpha = 0,05912$ ,  $\beta = 0,999$  e as medidas,  $SSE=0,3047$ ,  $RMSE=0,05851$ ,  $RS=0,454$  e  $ARS=0,454$ . Os resultados da comparação estão ilustrados na Figura 3.39. Esta é mais uma situação em que existe um grande número de pacotes concentrados nos valores baixos e alguns poucos pacotes nos valores altos, ou seja, poucos pacotes de comprimento acima de 300 bytes. Esta distribuição é tipicamente monomodal com possibilidade de ser melhor modelado pelas distribuições Exponencial, Log-normal, Weibull e Pareto.



**Figura 3.39** Modelo de distribuição cumulativa versus valores medidos (SSH).

### Dados Cumulativos–Tafvelin

O conjunto de dados obtidos do trabalho publicado por Tafvelin, foi coletado durante 20 dias consecutivos em abril de 2006 do tráfego bidirecional de um *backbone* OC-192. Nesse enlace, foram usados *splitters* ópticos ligados a duas placas Endace DAG6.2SE (JOHN; TAFVELIN, 2007). A Figura 3.40 e a Tabela 3.7 ilustram as comparações entre os valores medidos por Tafvelin, o modelo proposto da função de distribuição cumulativa e as distribuições de probabilidade Exponencial, Log-normal, Pareto e Weibull. Os resultados apresentados ilustram que o modelo de  $P_L(\ell)$  é o que melhor representa os valores experimentais, principalmente, na representação do segundo pico da distribuição bimodal. Isso se reflete nos valores numéricos, com  $P_L(\ell)$  apresentando os menores erros e as maiores correlações.



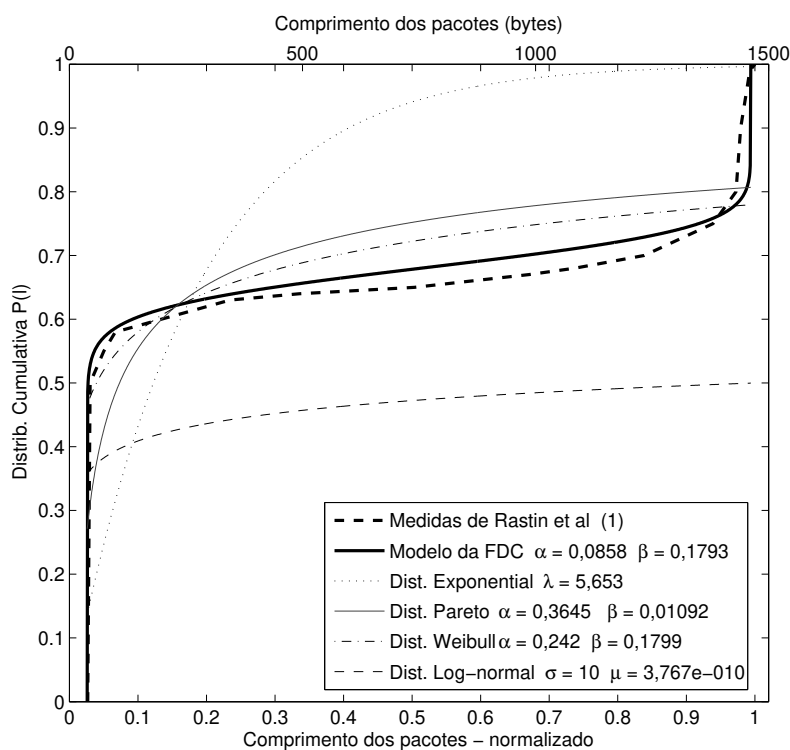
**Figura 3.40** Modelo de distribuição cumulativa versus valores medidos (Tafvelin) e outras distribuições.

Dist.	par.1	par.2	SSE	RMSE	RS	ARS
Exponencial	$\lambda = 1,609$	–	0,587	0,2048	0,1721	0,1721
Log-normal	$\mu = 4,442e - 13$	$\sigma = 7,003$	0,5368	0,1958	0,2429	0,2429
Pareto	$\alpha = 0,1769$	$\beta = 0,002964$	0,02724	0,0497	0,8678	0,8558
Weibull	$\alpha = 0,2605$	$\beta = 0,7935$	0,02317	0,0459	0,8876	0,8774
$P_L$	$\alpha = 0,0888$	$\beta = 0,09669$	0,01219	0,03188	0,9697	0,9671

**Tabela 3.7** Comparação entre as medidas de Tafvelin,  $P_L$  e outras distribuições.

### Dados Cumulativos–*Rastin*

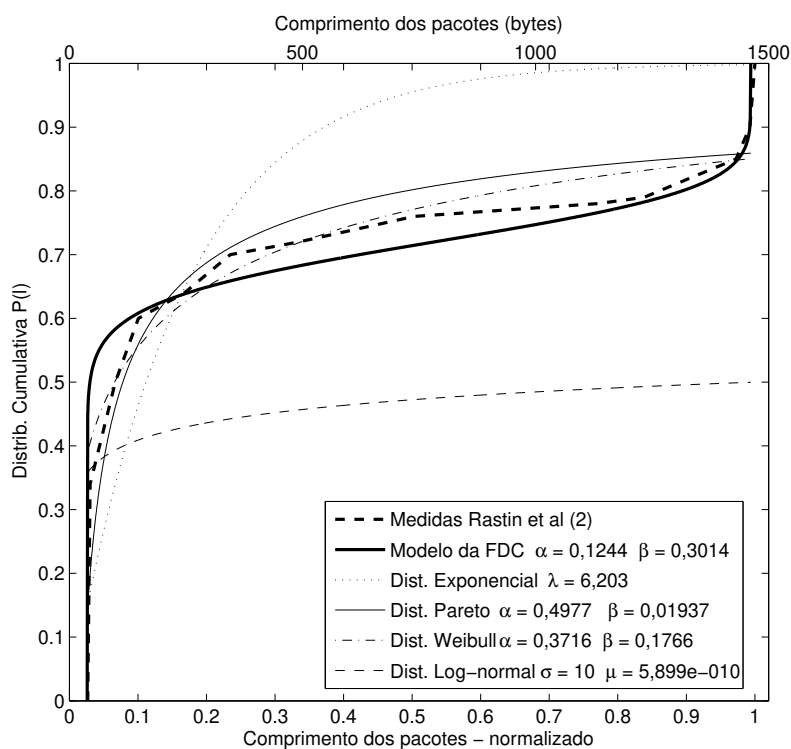
As medidas de dados de Rastin Pries foram obtidas de um centro de comutação de um provedor de serviços de Internet que provê acesso a 250 residências. Os usuários têm a conexão disponibilizada por meio de diversos pontos de acessos de redes sem fio. O tráfego é multiplexado usando um enlace de IEEE 802.11a (PRIES *et al.*, 2009). Das medidas apresentadas por Rastin Pries, utilizaram-se três. A primeira é chamada por Rastin Pries de *all packets*, aqui nomeada “Rastin1”, Figura 3.41. A segunda classificada como *BitTorrent*, foi nomeada “Rastin2”, Figura 3.42. E a terceira intitulada de “Rastin3”, Figura 3.43, tem as medidas chamadas *eDonkey*. As Figuras 3.41, 3.42, 3.43 e as Tabelas 3.8, 3.9, 3.10 ilustram as comparações entre os valores medidos por *Rastin Pries*, o modelo proposto da função de distribuição cumulativa e as distribuições de probabilidade Exponencial, Log-normal, Pareto e Weibull. Novamente, os resultados apresentados ilustram que o modelo de  $P_L(\ell)$  é o que melhor representa os valores experimentais, principalmente, na representação do segundo pico da distribuição bimodal. Isso se reflete nos valores numéricos, com  $P_L(\ell)$  apresentando os menores erros e as maiores correlações.



**Figura 3.41** Modelo de distribuição cumulativa versus valores medidos (*Rastin1*) e outras distribuições.

Dist.	par.1	par.2	SSE	RMSE	RS	ARS
Exponencial	$\lambda = 5,653$	–	0,8515	0,2175	–	–
Log-normal	$\mu = 3,767e - 10$	$\sigma = 10$	1,362	0,2751	–	–
Pareto	$\alpha = 0,3645$	$\beta = 0,01092$	0,2729	0,1267	0,6611	0,6412
Weibull	$\alpha = 0,242$	$\beta = 0,1799$	0,08779	0,0765	0,6502	0,6269
$P_L$	$\alpha = 0,0858$	$\beta = 0,1793$	0,03385	0,046	0,9036	0,8976

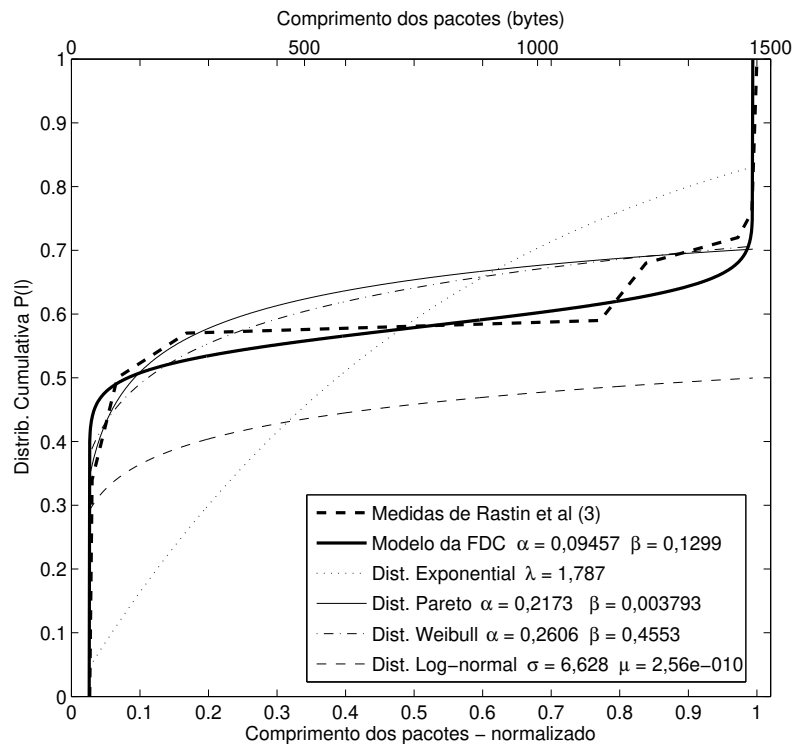
**Tabela 3.8** Comparação entre as medidas de Pries (Rastin1),  $P_L$  e outras distribuições.



**Figura 3.42** Modelo de distribuição cumulativa versus valores medidos (*Rastin2*) e outras distribuições.

Dist.	par.1	par.2	SSE	RMSE	RS	ARS
Exponencial	$\lambda = 6,203$	–	0,3151	0,15	0,6291	0,6291
Log-normal	$\mu = 5,899e - 10$	$\sigma = 10$	1,291	0,3036	–	–
Pareto	$\alpha = 0,4977$	$\beta = 0,01937$	0,07874	0,07783	0,9073	0,9002
Weibull	$\alpha = 0,3716$	$\beta = 0,1766$	0,01446	0,03626	0,9492	0,9445
$P_L$	$\alpha = 0,1244$	$\beta = 0,3014$	0,0389	0,05693	0,8953	0,8866

**Tabela 3.9** Comparação entre as medidas de Pries (Rastin2),  $P_L$  e outras distribuições.



**Figura 3.43** Modelo de distribuição cumulativa versus valores medidos (*Rastin3*) e outras distribuições.

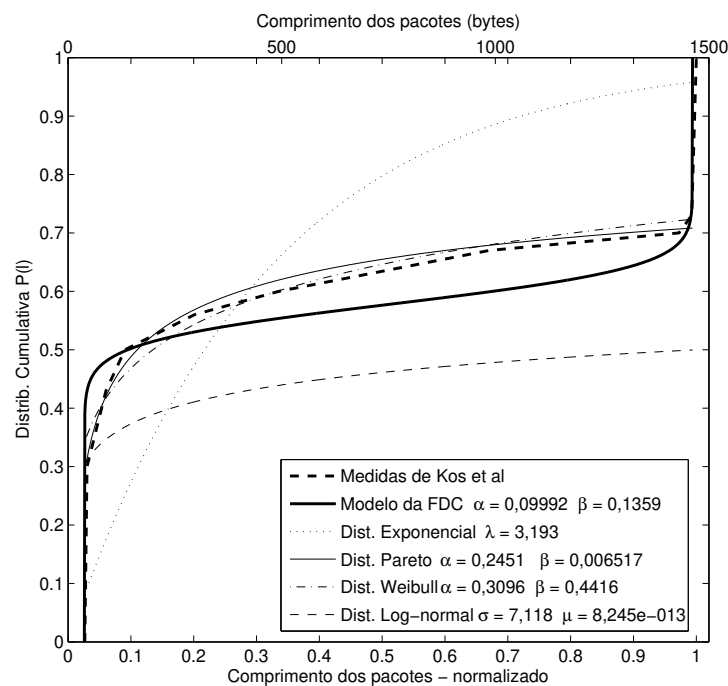
Dist.	par.1	par.2	SSE	RMSE	RS	ARS
Exponencial	$\lambda = 1,787$	–	0,4199	0,216	0,3571	0,3571
Log-normal	$\mu = 2,56e - 10$	$\sigma = 6,628$	0,6003	0,2583	0,0808	0,0808
Pareto	$\alpha = 0,2173$	$\beta = 0,003793$	0,01479	0,04966	0,89	0,8717
Weibull	$\alpha = 0,2606$	$\beta = 0,4553$	0,0173	0,0537	0,8714	0,85
$P_L$	$\alpha = 0,09457$	$\beta = 0,1299$	0,01649	0,04854	0,9392	0,9306

**Tabela 3.10** Comparação entre as medidas de Pries (*Rastin3*),  $P_L$  e outras distribuições.



### Dados Cumulativos–Kos

Os valores apresentados por Kos são medidas de tráfego real do *backbone* acadêmico da universidade de Ljubljana que conecta todos os membros de seu corpo docente (KOS *et al.*, 2003). A Figura 3.44 e a Tabela 3.11, apresentam as comparações entre os valores medidos por Kos, o modelo proposto da função de distribuição cumulativa e as distribuições de probabilidade Exponencial, Log-normal, Pareto e Weibull. O melhor resultado foi obtido pela distribuição de Pareto seguido pela distribuição de Weibull. O modelo proposto apresentou o terceiro melhor resultado. No entanto, nenhum das duas primeiras distribuições apresentou o comportamento bimodal. Isso é importante porque 40% dos comprimentos dos pacotes concentram-se no segundo pico, próximo a 1500 bytes (SINHA *et al.*, 2007), (JOHN; TAFVELIN, 2007) e (PRIES *et al.*, 2009). Os melhores resultados numéricos obtidos pelas distribuições de Pareto e Weibull foram obtidos desconsiderando os pontos do segundo pico e aproximando as curvas aos demais valores medidos.



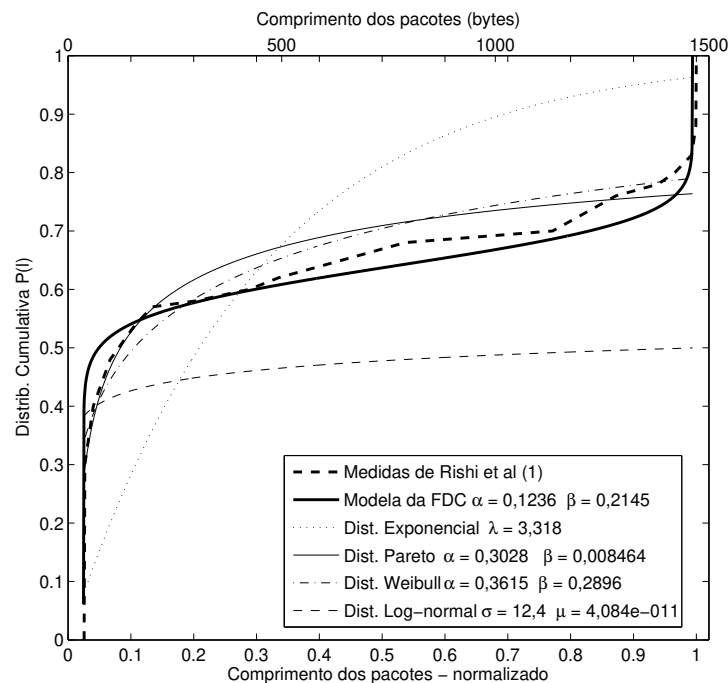
**Figura 3.44** Modelo de distribuição cumulativa versus valores medidos (*Kos*) e outras distribuições.

Dist.	par.1	par.2	SSE	RMSE	RS	ARS
Exponencial	$\lambda = 3,193$	–	0,3515	0,1976	0,4644	0,4644
Log-normal	$\mu = 8,245e - 13$	$\sigma = 7,118$	0,5468	0,2465	0,1668	0,1668
Pareto	$\alpha = 0,2451$	$\beta = 0,006517$	0,001918	0,01788	0,9872	0,9851
Weibull	$\alpha = 0,3096$	$\beta = 0,4416$	0,005624	0,03062	0,9626	0,9563
$P_L$	$\alpha = 0,09992$	$\beta = 0,1359$	0,02723	0,06237	0,9153	0,9032

**Tabela 3.11** Comparação entre as medidas de *Kos*,  $P_L$  e outras distribuições.

### Dados Cumulativos–*Rishi*

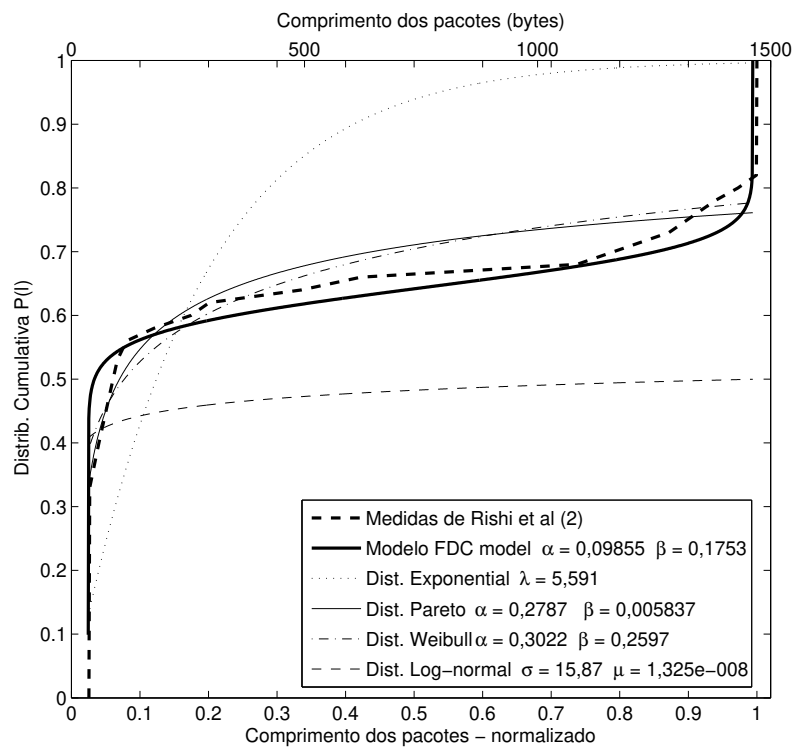
Dos dados coletados e apresentados por Rishi Sinha (SINHA *et al.*, 2007), foram usados cinco. O primeiro chamado de “Los Nettos” que foi obtido de um provedor de serviços de Internet com tráfego de dados acadêmicos e comerciais. Esse conjunto recebeu o título de *Rishi1*, e os resultados são apresentados na Figura 3.45 e na Tabela 3.12. O segundo conjunto é chamado de “USC Internet2”, proveniente de usuários usando VPN em conexões de rede sem fio. O nome usado para esses dados é *Rishi2* e resultados apresentados na Figura 3.46 e na Tabela 3.13. Os outros três conjuntos utilizados foram coletados pelo NLANR (*National Laboratory for Applied Network Research*) no GigaPOP, na Universidade de Memphis e no Centro de Supercomputação de Pittsburgh. A comparação com os dados do GigaPOP, chamado de *Rishi3*, são apresentados na Figura 3.47 e na Tabela 3.14. Os dados de Memphis, *Rishi4*, são ilustrados na Figura 3.48 e na Tabela 3.15. Por fim, *Rishi5*, Figura 3.49 e Tabela 3.16, representa os dados do Centro de Supercomputação de Pittsburgh. O modelo proposto apresentou os melhores resultados numéricos quando comparados com os dados *Rishi1* e *Rishi3*. Para os dados *Rishi2* e *Rishi4* as distribuições de Pareto e Weibull obtiveram melhores resultados. E em *Rishi5*, o melhor resultado foi obtido pela distribuição de Pareto. No entanto, nenhum das distribuições (Pareto e Weibull) apresentaram o comportamento bimodal. E como já destacado, 40% dos comprimentos dos pacotes concentram-se no segundo pico, próximo a 1500 bytes (SINHA *et al.*, 2007), (JOHN; TAFVELIN, 2007) e (PRIES *et al.*, 2009). Os melhores resultados numéricos obtidos pelas distribuições de Pareto e Weibull foram obtidos desconsiderando os pontos do segundo pico e aproximando as curvas aos demais valores medidos.



**Figura 3.45** Modelo de distribuição cumulativa versus valores medidos (*Rishi1*) e outras distribuições.

Dist.	par.1	par.2	SSE	RMSE	RS	ARS
Exponencial	$\lambda = 3,318$	–	0,519	0,1747	0,4296	0,4296
Log-normal	$\mu = 4,084e - 11$	$\sigma = 12,4$	1,106	0,2551	–	–
Pareto	$\alpha = 0,3028$	$\beta = 0,008464$	0,03347	0,0489	0,9113	0,905
Weibull	$\alpha = 0,3615$	$\beta = 0,2896$	0,02419	0,04157	0,9359	0,9313
$P_L$	$\alpha = 0,1236$	$\beta = 0,2145$	0,02801	0,04321	0,9443	0,9406

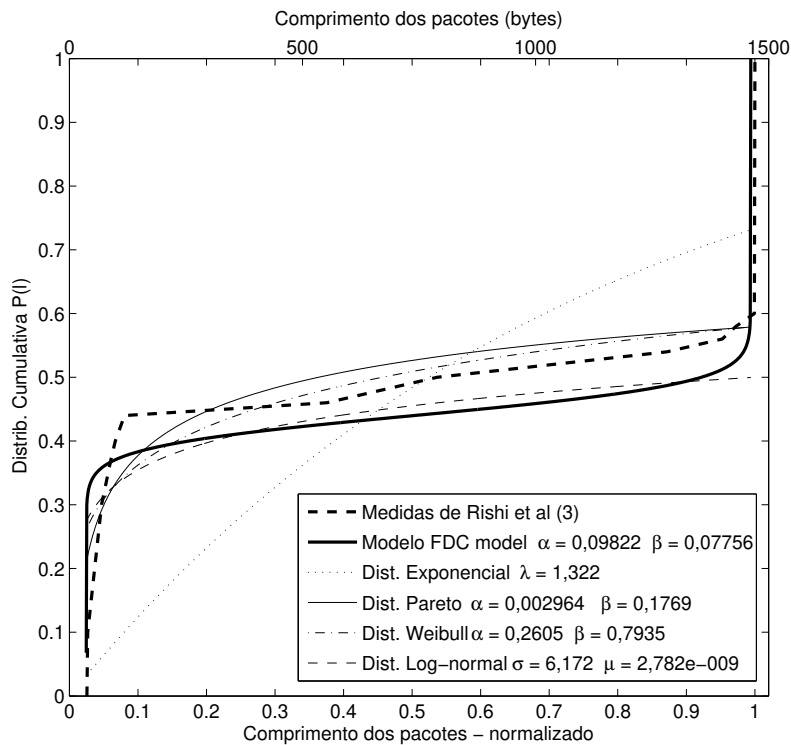
**Tabela 3.12** Comparação entre as medidas de *Rishi1*,  $P_L$  e outras distribuições.



**Figura 3.46** Modelo de distribuição cumulativa versus valores medidos (*Rishi2*) e outras distribuições.

Dist.	par.1	par.2	SSE	RMSE	RS	ARS
Exponencial	$\lambda = 5,591$	–	0,571	0,1951	0,2758	0,2758
Log-normal	$\mu = 1,325e - 8$	$\sigma = 15,87$	0,9353	0,2497	–	–
Pareto	$\alpha = 0,2787$	$\beta = 0,005837$	0,01626	0,03681	0,9379	0,9327
Weibull	$\alpha = 0,3022$	$\beta = 0,2597$	0,02114	0,04197	0,9192	0,9125
$P_L$	$\alpha = 0,09855$	$\beta = 0,1753$	0,04058	0,05587	0,897	0,8891

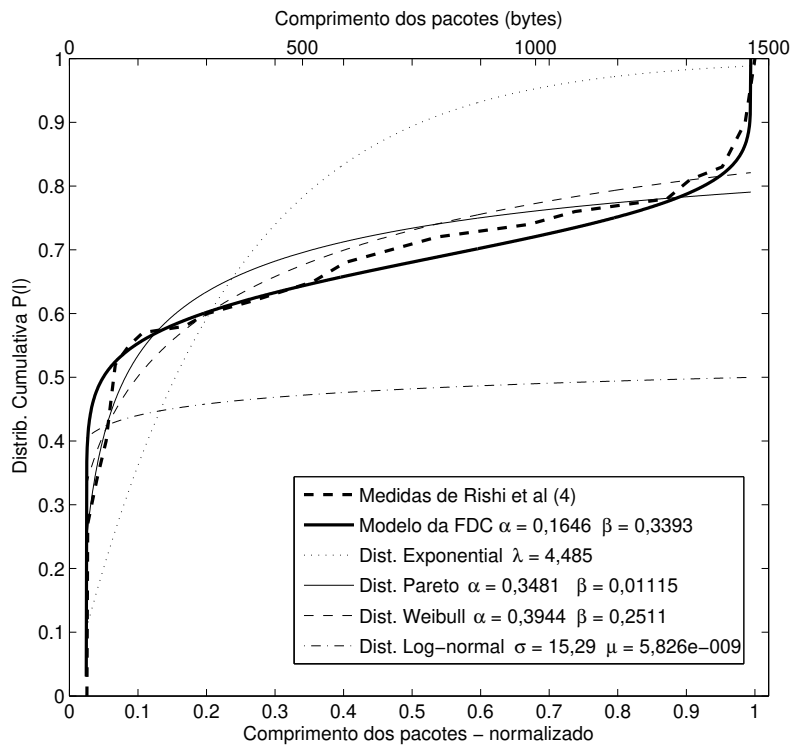
**Tabela 3.13** Comparação entre as medidas de *Rishi2*,  $P_L$  e outras distribuições.



**Figura 3.47** Modelo de distribuição cumulativa versus valores medidos (*Rishi3*) e outras distribuições.

Dist.	par.1	par.2	SSE	RMSE	RS	ARS
Exponencial	$\lambda = 1,322$	–	0,5225	0,2087	0,2755	0,2755
Log-normal	$\mu = 2,782e-9$	$\sigma = 6,172$	0,07386	0,08594	0,6569	0,6569
Pareto	$\alpha = 0,1703$	$\beta = 0,006206$	0,03333	0,06085	0,8452	0,828
Weibull	$\alpha = 0,2847$	$\beta = 1,654$	0,04567	0,07123	0,7879	0,7643
$P_L$	$\alpha = 0,09822$	$\beta = 0,07756$	0,07003	0,08368	0,8604	0,8465

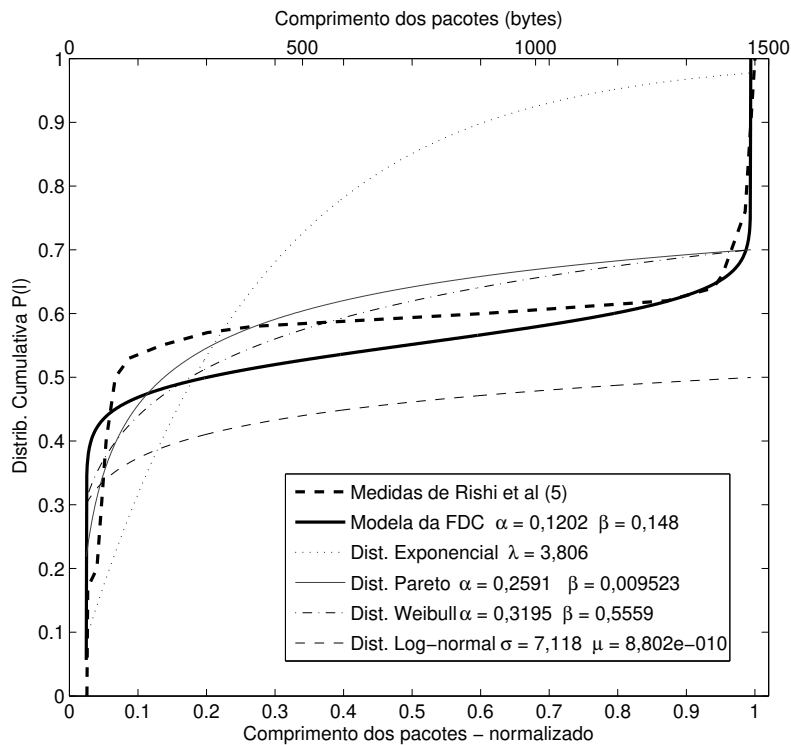
**Tabela 3.14** Comparação entre as medidas de *Rishi3*,  $P_L$  e outras distribuições.



**Figura 3.48** Modelo de distribuição cumulativa versus valores medidos (*Rishi4*) e outras distribuições.

Dist.	par.1	par.2	SSE	RMSE	RS	ARS
Exponencial	$\lambda = 4,485$	—	0,5344	0,1677	0,4811	0,4811
Log-normal	$\mu = 5,826e-9$	$\sigma = 15,29$	1,248	0,2563	—	—
Pareto	$\alpha = 0,3481$	$\beta = 0,01115$	0,02982	0,04317	0,9404	0,9366
Weibull	$\alpha = 0,3944$	$\beta = 0,2511$	0,03147	0,04435	0,9371	0,9331
$P_L$	$\alpha = 0,1646$	$\beta = 0,3393$	0,04721	0,0527	0,9251	0,9207

**Tabela 3.15** Comparação entre as medidas de *Rishi4*,  $P_L$  e outras distribuições.



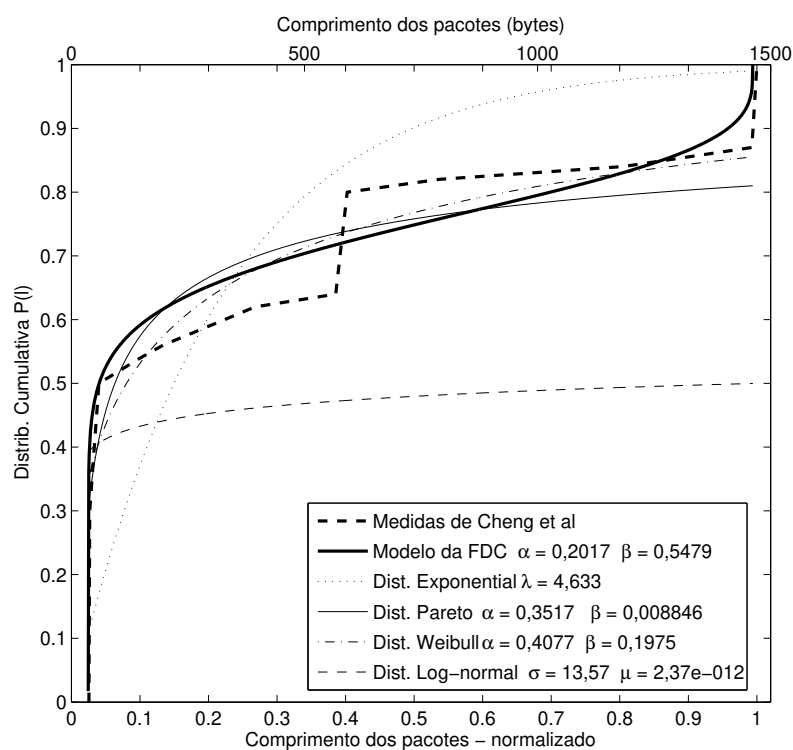
**Figura 3.49** Modelo de distribuição cumulativa versus valores medidos (*Rishi5*) e outras distribuições.

Dist.	par.1	par.2	SSE	RMSE	RS	ARS
Exponencial	$\lambda = 3,806$	–	0,6913	0,2222	0,192	0,192
Log-normal	$\mu = 8,802e - 10$	$\sigma = 7,118$	0,6511	0,2156	0,239	0,239
Pareto	$\alpha = 0,2591$	$\beta = 0,009523$	0,05748	0,07229	0,838	0,8232
Weibull	$\alpha = 0,3195$	$\beta = 0,5559$	0,09015	0,09053	0,7459	0,7228
$P_L$	$\alpha = 0,1202$	$\beta = 0,148$	0,1248	0,102	0,78	0,7616

**Tabela 3.16** Comparação entre as medidas de *Rishi5*,  $P_L$  e outras distribuições.

### Dados Cumulativos–Cheng

Os resultados apresentados no trabalho de Cheng Yu utilizou medidas coletadas pela Associação Cooperativa para Análise de Dados da Internet (CAIDA–*Cooperative Association for Internet Data Analysis*) do tráfego da *NASA Ames Internet Exchange (AIX)* entre 21 e 27 de fevereiro de 2000 (YU *et al.*, 2007). Os resultados da comparação são apresentados na Figura 3.50 e na Tabela 3.17. Cheng destaca que o comportamento trimodal, Figura 3.50, ocorreria devido ao comportamento do protocolo TCP, com os picos de pacotes em 40 *bytes* ocorrendo devido aos pacotes de *acknowledgements* do TCP, em 576 *bytes* por causa da transmissão de pacotes sem o uso do *path discovery* e em 1500 *bytes* com o *path discovery* implementado. A curva de  $P_L(\ell)$  adaptou-se bem aos valores experimentais apesar do comportamento trimodal com terceiro pico com intensidade moderada.



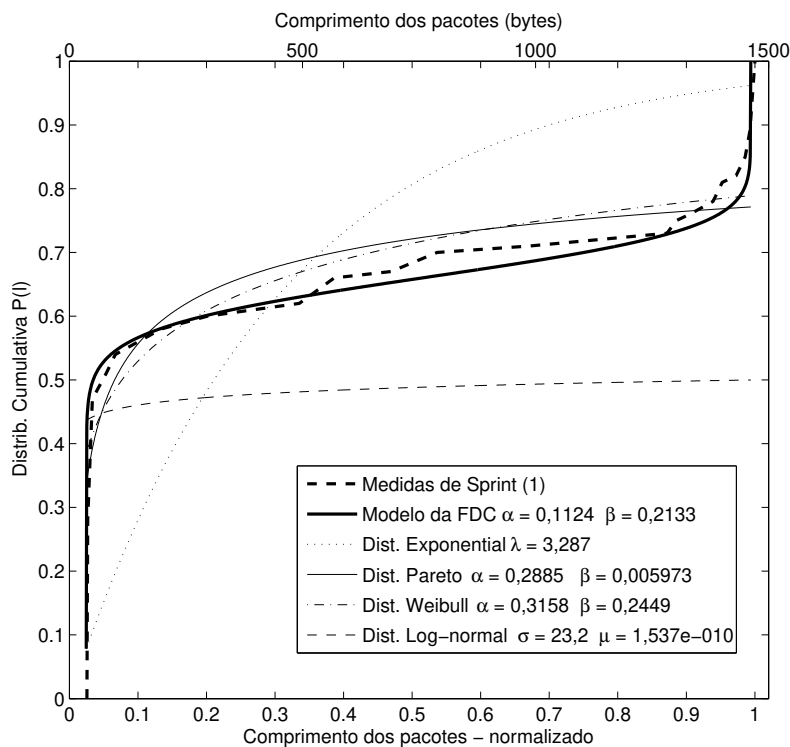
**Figura 3.50** Modelo de distribuição cumulativa versus valores medidos (*Cheng*) e outras distribuições.

Dist.	par.1	par.2	SSE	RMSE	RS	ARS
Exponencial	$\lambda = 4,633$	–	0,282	0,1601	0,6639	0,6639
Log-normal	$\mu = 2,37e-12$	$\sigma = 13,57$	0,9798	0,2984	–	–
Pareto	$\alpha = 0,3517$	$\beta = 0,008846$	0,0401	0,0708	0,8672	0,8506
Weibull	$\alpha = 0,4077$	$\beta = 0,1975$	0,02998	0,06121	0,9008	0,8884
$P_L$	$\alpha = 0,2017$	$\beta = 0,5479$	0,04294	0,06907	0,8963	0,8848

**Tabela 3.17** Comparação entre as medidas de *Cheng*,  $P_L$  e outras distribuições.

### Dados Cumulativos–*Sprint*

Os dados cumulativos apresentados a seguir, foram obtidos dos sítios do *Sprint Academic Research Group* (SPRINT, 2005), o primeiro chamado de *Sprint1*, foi coletado em 6 de fevereiro de 2004 em um *backbone* IP de San Jose com 84 *Mbps* de taxa de utilização com duração de coleta de três horas. Os resultados para este conjunto são apresentados na Figura 3.51 e na Tabela 3.18. O segundo, *Sprint2*, coletado em 10 de janeiro de 2005 em um *backbone* IP de San Jose com 14 *Mbps* de taxa de utilização com duração de coleta de três horas e os resultados ilustrados na Figura 3.52 e na Tabela 3.19. Outro conjunto de dados em que o modelo cumulativo se adaptou bem aos valores experimentais, comprovado pelos bons resultados numéricos e gráficos.



**Figura 3.51** Modelo de distribuição cumulativa versus valores medidos (*Sprint1*) e outras distribuições.

Dist.	par.1	par.2	SSE	RMSE	RS	ARS
Exponencial	$\lambda = 3,287$	–	0,8358	0,1949	0,1366	0,1366
Log-normal	$\mu = 1,537e - 10$	$\sigma = 23,2$	1,449	0,2567	–	–
Pareto	$\alpha = 0,2885$	$\beta = 0,005973$	0,05514	0,05387	0,8687	0,8618
Weibull	$\alpha = 0,3158$	$\beta = 0,2449$	0,04855	0,05055	0,8844	0,8784
$P_L$	$\alpha = 0,1124$	$\beta = 0,2133$	0,03972	0,04457	0,9249	0,9212

**Tabela 3.18** Comparação entre as medidas de *SPRINT1*,  $P_L$  e outras distribuições.



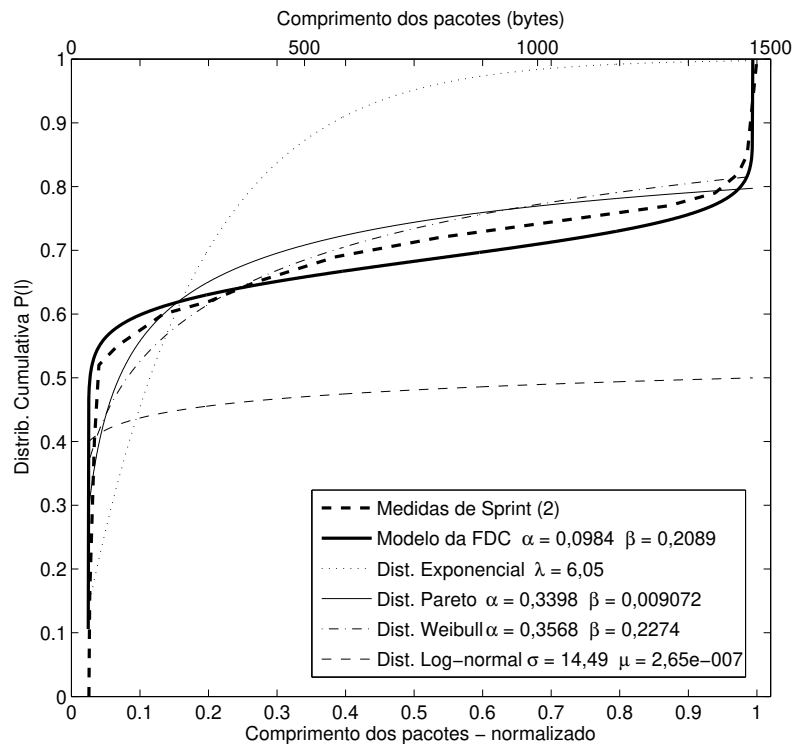


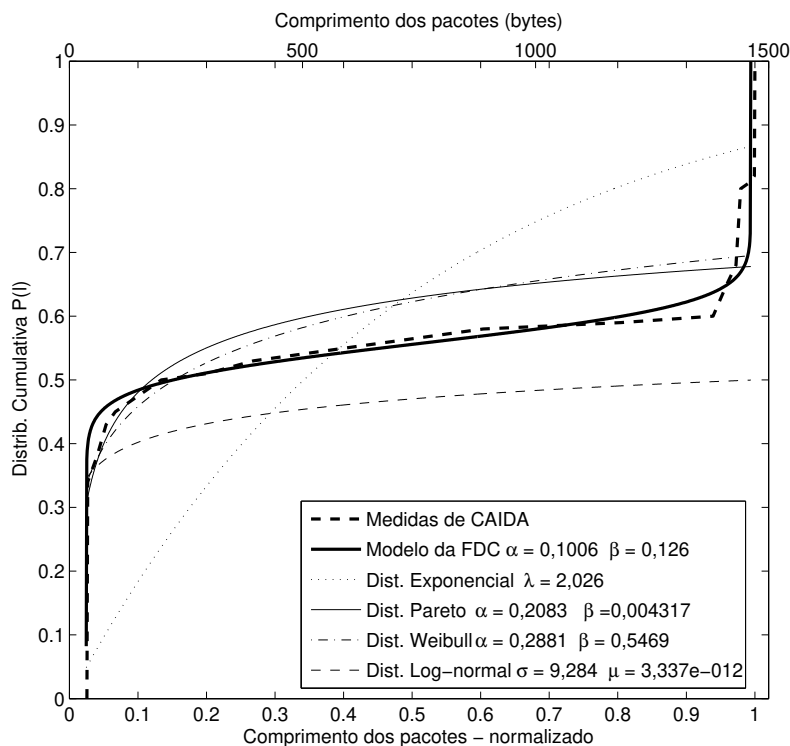
Figura 3.52 Modelo de distribuição cumulativa versus valores medidos (*Sprint2*) e outras distribuições.

Dist.	par.1	par.2	SSE	RMSE	RS	ARS
Exponencial	$\lambda = 6,05$	–	0,5563	0,1865	0,4337	0,4337
Log-normal	$\mu = 2,65e-7$	$\sigma = 14,49$	1,156	0,2777	–	–
Pareto	$\alpha = 0,3398$	$\beta = 0,009072$	0,05045	0,0623	0,8884	0,8798
Weibull	$\alpha = 0,3568$	$\beta = 0,2274$	0,0691	0,07291	0,8471	0,8353
$P_L$	$\alpha = 0,0984$	$\beta = 0,2089$	0,01529	0,0357	0,9285	0,9226

Tabela 3.19 Comparação entre as medidas de *SPRINT2*,  $P_L$  e outras distribuições.

### Dados Cumulativos–CAIDA

O último conjunto de dados foi obtido do sítio da Associação Cooperativa para Análise de Dados da Internet (CAIDA–*Cooperative Association for Internet Data Analysis*) (CAIDA, 2008). Os dados foram coletados em março de 2008 em um OC-192 *backbone* IP chamado de “Equinix-Chicago” em um enlace que integra Chicago e Seattle, nos EUA. Os resultados obtidos na comparação são ilustrados na Figura 3.53 e na Tabela 3.20. Os resultados numéricos e gráficos obtidos, mantêm a tendência dos resultados anteriores de boa aderência aos valores experimentais que estão dentro das condições de validade do modelo proposto para a função de distribuição cumulativa.



**Figura 3.53** Modelo de distribuição cumulativa versus valores medidos (CAIDA) e outras distribuições.

Dist.	par.1	par.2	SSE	RMSE	RS	ARS
Exponencial	$\lambda = 2,026$	–	0,8158	0,2072	0,03531	0,03531
Log-normal	$\mu = 3,337e - 12$	$\sigma = 9,284$	0,7946	0,2045	0,06039	0,06039
Pareto	$\alpha = 0,2083$	$\beta = 0,004317$	0,08037	0,07087	0,7617	0,7468
Weibull	$\alpha = 0,2881$	$\beta = 0,5469$	0,068	0,06519	0,7984	0,7858
$P_L$	$\alpha = 0,1006$	$\beta = 0,126$	0,02669	0,03962	0,9481	0,945

**Tabela 3.20** Comparação entre as medidas de CAIDA,  $P_L$  e outras distribuições.

## 3.4 Conclusão

Neste capítulo foram apresentadas as comparações entre os valores teóricos e medidos dos comprimentos dos pacotes em rede de computadores. Os valores práticos utilizados são medições de dados obtidas pelo autor e medidas de tráfego de dados encontrados na literatura. Os valores teóricos são obtidos dos modelos matemáticos da função densidade de probabilidade  $p_L(\ell)$ , e função cumulativa de probabilidade  $P_L(\ell)$ , abordados no capítulo anterior.

As medidas obtidas pelo autor foram comparadas com a função densidade de probabilidade e com a função de distribuição cumulativa. Enquanto as medidas obtidas na literatura foram comparadas com a função de distribuição cumulativa. Análises e comentários sobre os resultados são apresentados no próximo capítulo.

## CAPÍTULO 4

# Análise dos resultados

Este capítulo ilustra as análises e comentários dos resultados apresentados nos dois capítulos anteriores.

A partir da definição de  $\ell$ , Equação (2.6), que representa o comprimento do pacote enviado por um interface de rede, o modelo matemático da função densidade de probabilidade do comprimento dos pacotes em redes de computadores  $p_L(\ell)$ , Equação (2.26), é obtida. No capítulo anterior, compara-se  $p_L(\ell)$  com valores experimentais e um resumo dos resultados obtidos é apresentado na Tabela 4.1.

Tipo	$\alpha$	$\beta$	SSE	RMSE	RS	ARS
Diversos	0,01378	0,2217	0,05297	0,05425	0,8638	0,8563
YouTube	0,0142	0,2109	0,06332	0,05931	0,8542	0,8461
Torrent	0,0156	0,3107	0,0534	0,05447	0,8614	0,8537
Torrent2	0,01392	0,1889	0,05997	0,05772	0,8638	0,8563
FTP	0,02021	0,6295	0,04779	0,05152	0,8514	0,8432
Lab Comp	0,01106	0,3142	0,04421	0,04956	0,7673	0,7544
Indústria	0,02467	0,999	0,04521	0,04878	0,8644	0,8644

**Tabela 4.1** Resumo das comparações entre  $p_L(\ell)$  e os valores experimentais.

A Tabela 4.1 resume os resultados obtidos na comparação entre o modelo proposto de  $p_L(\ell)$  e os valores experimentais realizados por este autor. Os experimentos Diversos, YouTube, Torrent, Torrent2 e FTP apresentam um comportamento bimodal na distribuição dos comprimentos dos pacotes. A comparação com  $p_L(\ell)$  resultando nos valores das quatro métricas utilizadas nas medições dos erros. SSE e RMSE têm valores próximos a 5% nos cinco conjuntos citados e de aproximadamente 85% para os valores de RS e ARS. Estes indicam uma boa aderência aos resultados experimentais. Os dois conjuntos restantes da Tabela 4.1, Lab Comp e Indústria, apresentam tráfego com distribuição trimodal. Os dados do Lab Comp apresentam uma intensidade alta <sup>1</sup> no pico intermediário dos dados que se reflete na queda da aderência do

<sup>1</sup>A amplitude do pico é maior que o valor teórico previsto pela curva gerada pela equação de  $p_L(\ell)$ .

modelo (de  $\sim 85\%$  para  $\sim 76\%$ ), pela métrica RS e ARS. Os dados do conjunto Indústria, não sofreram da mesma redução devido à baixa intensidade<sup>2</sup> do pico intermediário de dados. Isso faz com que o surgimento do novo pico, abaixo da curva gerada pela equação de  $p_L(\ell)$ , diminua os valores dos erros. Assim, os dados do conjunto Indústria apresentam os melhores resultados das comparações realizadas.

Os resultados apresentados na Tabela 4.1, foram otimizados quando considerados dois intervalos de dados. Neste caso, utilizaram-se os mesmos dados, a mesma  $p_L(\ell)$ , mas dividiu-se o intervalo de comprimento dos pacotes  $\ell$  em dois segmentos. O primeiro de zero a 0,5 (750 bytes) e o segundo de 0,5 a 1 (1500 bytes). A Tabela 4.2 apresenta um resumo desta nova situação. Os valores obtidos reforçam a aderência do modelo aos dados de tráfego com comportamento bimodal. Nas medidas Diversos, YouTube, Torrent, Torrent2 e FTP, o modelo proposto para a densidade de probabilidade apresenta excelentes resultados. Isso é confirmado pelas quatro métricas utilizadas nas medições dos erros. Nos cinco conjuntos citados, SSE e RMSE têm valores entre 0% e 2%, RS e ARS entre 96% e 99%. Para os dados do Lab Comp, a intensidade alta do terceiro pico de dados se reflete na baixa precisão calculada no primeiro intervalo, próximo a 40% para RS e ARS. No entanto, para o segundo intervalo a aderência é de aproximadamente 99%. A comparação com os dados do conjunto Indústria, foram excelentes. Confirmando a previsão de que se o tráfego de dados for trimodal, mas com o pico intermediário de baixa intensidade, o modelo bimodal proposto apresenta bons resultados de aproximação. Outro detalhe importante é que nos ajustes de  $\alpha$  e  $\beta$  com os valores experimentais com dois intervalos, limitou-se o valor máximo destes parâmetros em 10, pois com valores maiores não se obteve grandes melhoras nos resultados.

Tipo	$\alpha$	$\beta$	SSE	RMSE	RS	ARS
Diversos	0,03919	10	0,004464	0,02113	0,9676	0,9676
	10	0,02153	0,000893	0,0113	0,9964	0,9959
YouTube	0,04078	10	0,005672	0,02511	0,9624	0,9624
	10	0,02265	0,001452	0,01347	0,9949	0,9942
Torrent	0,04311	10	0,001025	0,0121	0,9936	0,9918
	10	0,02025	0,001067	0,01155	0,9953	0,9947
Torrent2	0,03993	10	0,003911	0,02211	0,9727	0,9693
	10	0,02318	0,001545	0,0139	0,9948	0,9941
FTP	0,05139	10	0,0003238	0,005998	0,9985	0,9985
	10	0,01343	0,0004259	0,007297	0,9956	0,995
Lab Comp	0,046	9,9	0,04968	0,0743	0,4059	0,4059
	9,99	0,0233	0,0001205	0,004481	0,9989	0,9983
Indústria	0,0594	10	0,000628	0,00886	0,9979	0,9976
	10	0,006984	0,00225	0,01793	0,9114	0,8861

**Tabela 4.2** Resumo das comparações entre  $p_L(\ell)$  e os valores experimentais para dois intervalos.

<sup>2</sup>A amplitude do pico é menor que o valor teórico previsto pela curva gerada pela equação de  $p_L(\ell)$ .

A partir de  $p_L(\ell)$ , encontra-se o modelo matemático para a função de distribuição cumulativa do comprimento dos pacotes em redes de computadores  $P_L(\ell)$ , Equação (2.45). Esse modelo usa a função Beta incompleta normalizada.

Em seguida, comparou-se  $P_L(\ell)$  com valores experimentais Diversos. O valor de  $P_L(\ell)$  usando os valores de  $\alpha$  e  $\beta$  obtidos na Figura 3.6, Tabela 4.1, apresenta valores bem acima do valores experimentais (Figura 3.26). As justificativas para esta diferença são: Primeiro, que na parte central do gráfico (0,1–0,9) da figura 3.6 está acima dos valores reais. Isso se reflete no gráfico cumulativo com valores acima dos valores reais. A segunda justificativa é que  $p_L(\ell)$  possui duas indeterminações em  $\ell = L_m$  e  $\ell = L_M$ , por isso usa-se  $\ell = L_m + \varepsilon$  e  $\ell = L_M - \varepsilon$ , com  $\varepsilon \rightarrow 0$ . Essa indeterminação dificulta a representação, de forma mais aproximada, dos dois picos da curva, justificando o salto maior na função cumulativa.

No gráfico de  $P_L(\ell)$  usando os valores de  $\alpha$  e  $\beta$  obtidos na Figura 3.7, dois intervalos, dados Diversos da Tabela 4.2, apresentou valores abaixo do valores experimentais (Figura 3.26). Isso ocorre por causa da diferença numérica entre o primeiro pico, próximo a 40 bytes, e a curva do modelo  $p_L(\ell)$ .

O passo seguinte foi comparar  $P_L(\ell)$  com valores experimentais e realizar o ajuste de curva pelo método dos mínimos quadrados para encontrar os melhores valores dos parâmetros  $\alpha$  e  $\beta$ , considerando  $\alpha, \beta > 0$ . Um resumo dos resultados obtidos é apresentado na Tabela 4.3.

Tipo	$\alpha$	$\beta$	SSE	RMSE	RS	ARS
Diversos	0,03465	0,02705	0,0003979	0,004838	0,9564	0,9538
YouTube	0,007552	0,005761	0,0000055	0,000709	0,9485	0,9438
Torrent	0,007016	0,00681	0,0000344	0,001514	0,8946	0,8876
Torrent2	0,005133	0,003758	0,0000154	0,001012	0,8826	0,8748
FTP	0,02636	0,04664	0,0002173	0,003806	0,9665	0,9643
Lab Comp	0,02897	0,04129	0,001962	0,01144	0,7687	0,7533
Indústria	0,07183	0,2298	0,01779	0,03144	0,8561	0,8481

**Tabela 4.3** Resumo das comparações entre  $P_L(\ell)$  e os valores experimentais pelo autor.

O modelo proposto para a distribuição cumulativa  $P_L(\ell)$  apresentou resultados melhores que os obtidos com o modelo de  $p_L(\ell)$ . Para os conjuntos de dados Diversos, YouTube, Torrent, Torrent2 e FTP, os valores de SSE e RMSE ficaram abaixo de 1%. Os valores de RS e ARS ficaram entre 87% e 96%. A comparação com os dados Lab Comp, trimodal de alta intensidade, revela que o nível de aderência permaneceu no mesmo patamar do obtido para  $p_L(\ell)$ . Repetindo as observações anteriores, verifica-se que o modelo proposto, apresenta um bom nível de aderência para a distribuição trimodal com pico intermediário de baixa intensidade.

Dando continuidade às comparações entre os dados experimentais e o modelo proposto, a Tabela 4.4 ilustra o resumo dos resultados obtidos das comparações com a série de conjuntos de dados de Ville Mattila (MATTILA, 2010). Verifica-se que para os conjuntos de dados “Internet IP”, “Internet TCP” e “Internet HTTP”, o modelo  $P_L(\ell)$  apresenta resultados no mesmo nível

Tipo	$\alpha$	$\beta$	SSE	RMSE	RS	ARS
Internet IP	0,0187	0,01673	0,001602	0,004367	0,8925	0,8912
Internet TCP	0,0175	0,01431	0,0007737	0,003035	0,9323	0,9315
Internet HTTP	0,04621	0,0868	0,01926	0,01479	0,8893	0,888
Internet UDP	0,115	0,999	1,239	0,1167	0,4929	0,4929
Internet DNS	0,0884	0,999	1,112	0,1106	0,4119	0,4119
Internet SMTP	0,1481	0,6563	0,288	0,05689	0,7664	0,7637
Internet SSH	0,05912	0,999	0,3047	0,05851	0,454	0,454

**Tabela 4.4** Resumo das comparações entre  $P_L(\ell)$  e os valores experimentais obtidos por Mattila.

de precisão dos resultados anteriores, SSE e RMSE abaixo de 2%, RS e ARS acima de 88%. Para os demais conjuntos, UDP, DNS, SMTP e SSH, a comparação com  $P_L(\ell)$  revelou que este modelo não é adequado para modelar estas aplicações. Isso se justifica pela característica dessas, com baixa quantidade de pacotes, tráfego monomodal e distribuição do comprimento dos pacotes concentrados em valores baixos. Essa situação, monomodal, é bem diferente da condição essencial de validade em que o modelo desta tese está embasado, a distribuição bimodal do comprimento dos pacotes. O comportamento monomodal é melhor modelado pelas distribuições Exponencial, Log-normal, Weibull e Pareto.

A última série de comparações é resumida na Tabela 4.5. Os resultados apresentados ilustram como o modelo  $P_L(\ell)$  consegue reproduzir bem os valores experimentais de tráfego com distribuição bimodal. Para o caso de distribuição trimodal, o modelo cumulativo adapta-se de forma razoável bem quando o pico intermediário tem uma intensidade de baixa para média. Desta série de comparações com dados coletados em diferentes ambientes de rede, os dados comparativos *Rishi3*, apresentam resultados regulares enquanto que os dados *Rishi5* apresentam os maiores erros da Tabela 4.4.

A partir dos resultados comparativos apresentados, verifica-se como o modelo de distribuição cumulativa  $P_L(\ell)$  proposto nesta tese, depende do tipo de aplicação que gera o tráfego, uma característica interessante que pode ser usada na classificação de tráfego.

Além do modelo proposto para a função densidade de probabilidade  $p_L(\ell)$  e sua função de distribuição cumulativa  $P_L(\ell)$ , ainda foram apresentados as equações do primeiro momento ( $E[L]$ ), segundo momento ( $E[L^2]$ ), variância ( $\text{Var}[L]$ ), e desvio padrão, ( $\sigma$ ).

As Figuras 2.8, 2.9 e 2.10, ilustram a variação de  $E[L]$ , Equação (2.59), em função de  $\alpha$  e  $\beta$ . A primeira figura é uma superfície que ilustra o comportamento da média em função de  $\alpha$  e  $\beta$ . Verifica-se que a média aumenta quando ocorre um aumento do valor do parâmetro  $\alpha$ . Esse comportamento é observado no gráfico da variação de  $E[L]$  em função de  $\alpha$ , Figura 2.9, para diferentes valores de  $\beta$ . Fenômeno inverso ocorre com o valor do parâmetro  $\beta$ , a média diminui quando o valor de  $\beta$  cresce e aumenta quando o valor de  $\beta$  diminui. A Figura 2.10 ilustra esta variação de  $E[L]$  em função de  $\beta$  para alguns valores de  $\alpha$ .

A  $\text{Var}[L]$ , Equação (2.93), é função de  $\alpha$  e  $\beta$ . Verificou-se que a variância aumenta quando ocorre um aumento do valor do parâmetro  $\beta$  e que fenômeno inverso corre com o valor do parâmetro  $\alpha$ , a variância diminui quando o valor de  $\alpha$  cresce e a variância aumenta quando o valor de  $\alpha$  diminui.

As Figuras 3.40 a 3.53 ilustram os resultados das comparações entre diferentes modelos teóricos e os dados cumulativos obtidos por diferentes autores da literatura. Dentre os modelos teóricos foram escolhidos as distribuições Exponencial (Apêndice C), por ser muito utilizada em alguns modelos de tráfego de dados (HONG *et al.*, 2006), (HONG; SOHRABY, 2007), (JUNG; SOHRABY, 2010); As distribuições Weibull (Apêndice D), Log-normal (Apêndice E) e Pareto (Apêndice F) bastante utilizadas no modelamento do comprimento do pacotes (MUSHTAQ; RIZVI, 2005), (TORABZADEH; AJIB, 2010); e a distribuição cumulativa  $P_L(\ell)$  proposta nesta tese.

Dos resultados obtidos, Tabelas 3.7 a 3.20, o modelo  $P_L(\ell)$  proposto é o que apresenta melhores resultados. As distribuições Weibull e Pareto apresentam resultados razoáveis. Enquanto que as distribuições Exponencial e Log-normal são as menos aderentes. Estes resultados se refletem nos gráficos das Figuras 3.40 a 3.53. A curva da distribuição Exponencial é bem discrepante dos dados experimentais. A curva da log-normal fica bem abaixo dos valores reais. Enquanto, as distribuições Weibull e Pareto estão próximas dos valores medidos, mas não reproduzem o segundo pico da distribuição do comprimento dos pacotes. Esse segundo pico representa 40% dos pacotes (JOHN; TAFVELIN, 2007), (PRIES *et al.*, 2009). E finalmente, a função cumulativa  $P_L(\ell)$  que mais se aproxima dos valores experimentais e do comportamento da curva dos dados. A Tabela 4.5 resume as comparações entre  $P_L(\ell)$  e os valores experimentais de outros autores.

Tipo	$\alpha$	$\beta$	SSE	RMSE	RS	ARS
Tafvelin	0,08884	0,09669	0,01219	0,03188	0,9697	0,9671
Rastin1	0,0858	0,1793	0,03385	0,046	0,9036	0,8976
Rastin2	0,1244	0,3014	0,0389	0,05693	0,8953	0,8866
Rastin3	0,09457	0,1299	0,01649	0,04854	0,9392	0,9306
Kos	0,09992	0,1359	0,02723	0,06237	0,9153	0,9032
Rishi1	0,1236	0,2145	0,02801	0,04321	0,9443	0,9406
Rishi2	0,09855	0,1753	0,04058	0,05587	0,897	0,8891
Rishi3	0,09822	0,07756	0,07003	0,08368	0,8604	0,8465
Rishi4	0,1646	0,3393	0,04721	0,0527	0,9251	0,9207
Rishi5	0,1202	0,148	0,1248	0,102	0,78	0,7616
Cheng	0,2017	0,5479	0,04294	0,06907	0,8963	0,8848
Sprint1	0,1124	0,2133	0,03972	0,04457	0,9249	0,9212
Sprint2	0,0984	0,2089	0,01529	0,0357	0,9285	0,9226
CAIDA	0,1006	0,126	0,02669	0,03962	0,9481	0,945

**Tabela 4.5** Resumo das comparações entre  $P_L(\ell)$  e os valores experimentais de outros autores.



## 4.1 Conclusão

Neste capítulo foram apresentadas as análises dos resultados obtidos nesta tese. Verificou-se que o modelo proposto  $p_L(\ell)$  obteve bons resultados quando comparado ao tráfego de dados com distribuição bimodal e que os resultados são otimizados quando se divide o intervalo avaliado em dois. Observou-se ainda que a função cumulativa  $P_L(\ell)$  apresenta uma aderência maior que  $p_L(\ell)$  aos dados experimentais. Ainda, foi observado que a fórmula  $P_L(\ell)$  depende do tipo de aplicação que gera o tráfego da rede e se este é unimodal, bimodal ou trimodal. Comparando-se com os resultados de outras distribuições a função cumulativa proposta apresenta valores mais próximos dos valores reais, principalmente na região do segundo pico em que as demais distribuições não conseguem representá-lo.

## CAPÍTULO 5

---

---

# Conclusão

---

### 5.1 Conclusões

Esta tese apresentou um modelo para a função de densidade de probabilidade  $p_L(\ell)$  e a partir desta, obteve-se a função cumulativa de probabilidade  $P_L(\ell)$  do comprimento dos pacotes em redes de computadores. A partir dessas duas funções foram obtidas as equações para o cálculo do primeiro momento, segundo momento, variância e desvio padrão.

Em seguida, foram realizadas comparações entre os modelos teóricos propostos e os valores reais de tráfego. A partir dessas comparações, verificou-se que o modelo de densidade de probabilidade do comprimento de pacotes se adaptou bem ao tráfego de dados com distribuição bimodal do comprimento dos pacotes, com bons resultados numéricos. Esses resultados são otimizados quando se avalia os dados em dois intervalos.

O comportamento do tráfego bimodal descrito por Tafvelin (JOHN; TAFVELIN, 2007), Pries (PRIES *et al.*, 2009) e outros autores, foi formulado analiticamente nesta tese. Com o modelo obtido, pode-se estimar com uma boa precisão a função densidade de probabilidade do comprimento dos pacotes  $p_L(\ell)$ . Outro resultado importante é que, se no tráfego de dados com distribuição trimodal, o pico de dados intermediário for de baixa intensidade o modelo de  $p_L(\ell)$  ainda apresenta uma boa aproximação dos valores reais.

Para o modelo da função de distribuição cumulativa do comprimento dos pacotes  $P_L(\ell)$ , foi demonstrado que seu valor é aproximadamente igual à função Beta incompleta normalizada parametrizada pela equação de  $x$ . Esse modelo foi comparado com várias medidas encontradas na literatura e os resultados comparativos mostraram que ele adere bem aos valores experimentais e sem a necessidade de divisão do intervalo, em dois, como realizado para otimizar os resultados de  $p_L(\ell)$ . A comparação entre os dados experimentais de Tafvelin (JOHN; TAFVELIN, 2007) e as distribuições, Exponencial, Log-normal, Weibull, Pareto e  $P_L(\ell)$ , ilustraram que o modelo proposto apresenta os melhores resultados.

O modelo proposto nesta tese,  $p_L(\ell)$  e sua função cumulativa  $P_L(\ell)$ , dependem mais do tipo de aplicação geradora do tráfego que do tipo de rede na qual este tráfego é transportado.

Esse resultado reforça a importância do modelo por envolver as medições do comprimento dos pacotes, que por sua vez, possibilita o uso dessas informações no projeto e estimativa da infraestrutura de redes e suas aplicações, como destacado por Li Bo (BO *et al.*, 2006) e Iacovazzi (IACOVAZZI; BAIOCCHI, 2010).

Foram feitas análises da influência dos parâmetros  $\alpha$  e  $\beta$  na distribuição do comprimento dos pacotes. Verificou-se que a média é diretamente proporcional ao parâmetro  $\alpha$  e inversamente proporcional ao parâmetro  $\beta$ . No caso da variância e desvio padrão, seus valores são diretamente proporcionais ao parâmetro  $\beta$  e inversamente proporcionais ao parâmetro  $\alpha$ .

Para  $p_L(\ell)$  e  $P_L(\ell)$ , se  $\alpha$  for muito próximo a zero e  $\beta$  é for grande, maior que 1, a distribuição de probabilidade e distribuição cumulativa concentram-se na região próxima a 40 bytes. Isso sugere que  $\alpha$  regula a frequência de pacotes pequenos, os chamados pacotes de reconhecimento do TCP (TCP *acknowledgements* (YU *et al.*, 2007), (SINHA *et al.*, 2007)). No caso contrário, se  $\alpha$  for grande, maior que 1 e se  $\beta$  é for muito próximo a zero, a distribuição de probabilidade e distribuição cumulativa concentram-se na região próxima a 1500 bytes. Isso sugere que  $\beta$  regula a frequência de pacotes de carga máxima do padrão *Ethernet* (1492 bytes) (YU *et al.*, 2007), (SINHA *et al.*, 2007).

Na Figura 3.26 observa-se que o crescimento do valor de  $\alpha$  faz com que o primeiro salto da função cumulativa diminua, ou seja, inversamente proporcional à amplitude do primeiro salto. Para o valor de  $\beta$ , a diminuição de seu valor faz com que o segundo salto da função cumulativa aumente, ou seja, inversamente proporcional à amplitude do segundo salto.

No caso de se utilizar  $P_L(\ell)$  na classificação de tráfego (SUN *et al.*, 2002), (MCGREGOR *et al.*, 2004), (MOORE; ZUEV, 2005), (ZANDER *et al.*, 2005), (LIBERATORE *et al.*, 2006), (LIBERATORE; LEVINE, 2006), (WRIGHT *et al.*, 2006), (WRIGHT *et al.*, 2007), (WRIGHT *et al.*, 2009), (ALSHAMMARI; ZINCIR-HEYWOOD, 2007), (DUSI *et al.*, 2009) seria possível realizar várias comparações entre  $P_L(\ell)$  e as medidas de um mesma aplicação (http, p2p, VoIp, etc.) para identificar as faixas de valores de  $\alpha$  e  $\beta$  de cada uma das aplicações. Com isso, teria-se uma metodologia de classificação de tráfego alternativa as apresentadas por Callado (CALLADO *et al.*, 2009).

No trabalho publicado por Cucej (CUCEJ; FRAS, 2009) é apresentado uma metodologia para estimação de parâmetros de uma distribuição de probabilidade a partir do dados de tráfego. O autor utiliza este procedimento com as distribuições de Pareto e Exponencial. No entanto, se for utilizado a distribuição apresentada nesta tese, resultados mais próximos do real podem ser obtidos.

Para projetos de dimensionamento de enlaces em pontos de agregação, Figura 2.1, a seguinte situação poderia ser usada: Para uma determinado tráfego médio ( $TM_E$ ) e desvio padrão ( $\sigma_E$ ) do enlace, determina-se os valores de  $\alpha_E$  e  $\beta_E$  utilizando as equações 2.59 e 2.100 e um método numérico de iteração. Determinado os valores de  $\alpha_E$  e  $\beta_E$ , encontra-se a distribuição cumulativa do enlace ( $P_{L_E}(\ell)$ ) usando a equação 2.45. O passo seguinte é determinar o tráfego médio ( $TM_U$ ) dos  $N$  usuários conectados no enlace usando  $TM_U = \frac{TM_E}{N}$ . Para um determinado desvio padrão ( $\sigma_U$ ) do usuários, determina-se os valores de  $\alpha_U$  e  $\beta_U$  utilizando as equações 2.59

e 2.100 e um método numérico de iteração. Determinado os valores de  $\alpha_U$  e  $\beta_U$ , encontra-se a distribuição cumulativa dos usuários ( $P_{L_U}(\ell)$ ) usando a equação 2.45. Em seguida, busca-se encontrar qual a taxa de transmissão de cada usuário ( $TU$ ) e quantos pacotes ( $n$ ) são necessários para que com a distribuição  $P_{L_U}(\ell)$ , se tenha o valor médio  $TM_U$  e desvio padrão  $\sigma_U$ . Encontrado o valor de  $n$  de cada usuário, multiplica-se este valor pelo número de usuário  $N$ . E finalmente determina-se qual a taxa de transmissão do enlace ( $TU_E$ ) capaz de atender a  $n.N$  pacotes com o tráfego médio ( $TM_E$ ) e desvio padrão ( $\sigma_E$ ).

Observa-se que a função densidade de probabilidade e a função distribuição cumulativa propostas nesta tese, podem auxiliar nos projetos, controle, gerência, interpolação e extrapolação das redes (MUSHTAQ; RIZVI, 2005). Sendo possível ainda, melhorar o desenvolvimento de simuladores e um desenvolvimento mais rápido e eficiente de algoritmos para otimização das redes de comunicações e suas aplicações. O modelo ainda pode ser usado para geração de tráfego sintético (CUCEJ; FRAS, 2009), comparação e simulação do tráfego em redes de computadores.

## 5.2 Contribuições da Tese

A principal contribuição desta tese foi o desenvolvimento da fórmula que modela a função densidade de probabilidade do comprimento dos pacotes em redes de computadores com tráfego bimodal. Também foram obtidas as fórmulas da função de distribuição cumulativa, primeiro momento, segundo momento, variância e desvio padrão.

Verificou-se que as fórmulas encontradas apresentaram uma boa aproximação quando comparados com valores medidos de tráfego. E ainda, que quando a distribuição é trimodal com baixa intensidade do pico intermediário, o modelo também apresenta bons resultados.

Na comparação da função de distribuição cumulativa proposta nesta tese e as distribuições Exponencial, Weibull, Log-normal e Pareto, o modelo proposto obteve resultados melhores que as demais distribuições.

E finalmente, que os tipos de aplicações presentes na rede têm uma influência maior no formato da função densidade de probabilidade dos comprimentos dos pacotes que o tipo de rede na qual o tráfego é transportado.

## 5.3 Sugestões para Trabalhos Futuros

- Obter as equações de  $p_L(\ell)$ ,  $P_L(\ell)$ ,  $E(L)$  e  $Var(L)$  para o caso discreto, mais adequado ao tráfego de dados em redes de computadores.
- Realizar várias comparações entre  $P_L(\ell)$  e as medidas de um mesma aplicação (http, p2p, VoIp, etc.) para tentar identificar as faixas de valores de  $\alpha$  e  $\beta$  de cada uma das aplicações.
- Comparar as equações propostas com tráfego em redes que usam IPv6.

- 
- Realizar comparações com outras bases de dados e em situações de tráfego não abordadas na tese.
  - Modelar o tráfego de dados para distribuições trimodais, equações contínuas e discretas.
  - Modelar o tráfego de dados para aplicações que usam pacotes pequenos e em pequenas quantidades.
  - Geração de tráfego sintético a partir das equações apresentadas e comparação com dados reais.
  - Tentar identificar os diferentes tipos de aplicações TCP/IP usando o modelo de distribuição cumulativa ou a densidade de probabilidade.
  - Buscar a existência de alguma relação das equações apresentadas e a autosimilaridade.

## APÊNDICE A

---

---

# Publicações

---

- “A Packet Distribution Traffic Model for Computer Networks”, ITS 2010 – The International Telecommunications Symposium (ITS), 6 a 9 de setembro de 2010 - Manaus - AM.
- “Comparison Results of a Mathematical Model and Experimental Measurements for the Distribution Function of the Packet Length in Computer Networks”, IWT 2011 – International Workshop on Telecommunications, 3 a 6 de Maio de 2011 - Rio de Janeiro - RJ.

## APÊNDICE B

---

---

# Função Gama

---

A função Gama é definida como

$$\Gamma(z) = \int_0^{\infty} t^{z-1} \exp^{-t} dt \quad (\text{B.1})$$

e verifica-se que (para  $n$  natural)

$$\Gamma(n+1) = n! \quad (\text{B.2})$$

Assim,

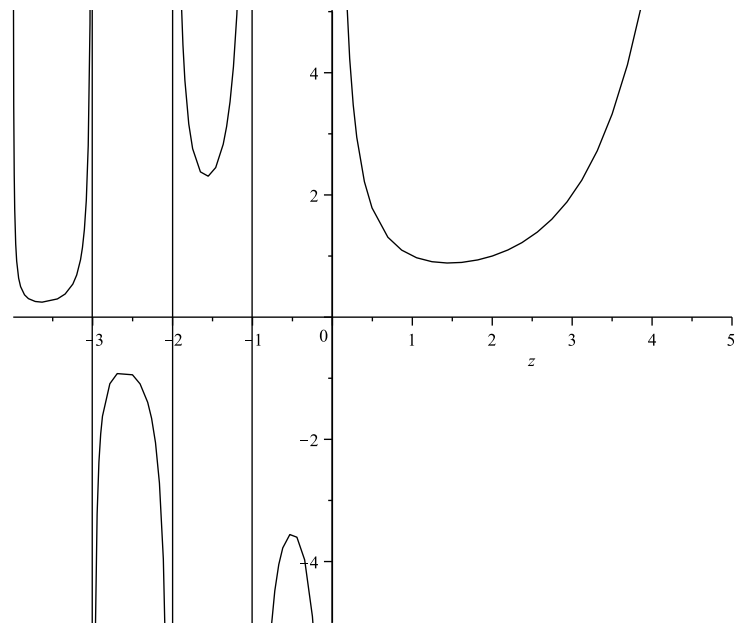
$$\begin{aligned} \Gamma(1) &= 1, \\ \Gamma(2) &= 1. \end{aligned} \quad (\text{B.3})$$

De modo geral

$$\Gamma(n+1) = n \cdot \Gamma(n), \quad \forall n \in \mathbb{N} \quad (\text{B.4})$$

e em particular

$$\Gamma\left(\frac{1}{2}\right) = \pi. \quad (\text{B.5})$$



**Figura B.1** Função Gama  $\Gamma(z)$ .



## APÊNDICE C

---

---

# Distribuição Exponencial

---

A distribuição Exponencial é definida como

$$p(x) = \lambda e^{-x\lambda}, \quad \lambda > 0. \quad (\text{C.1})$$

A função de distribuição cumulativa é

$$P(x) = 1 - e^{-x\lambda}. \quad (\text{C.2})$$

O primeiro momento ou valor esperado é dado por

$$m_1 = E[x] = \frac{1}{\lambda}. \quad (\text{C.3})$$

E a variância é calculada usando

$$\text{Var}[x] = \sigma^2 = \frac{1}{\lambda^2}. \quad (\text{C.4})$$

## APÊNDICE D

---

---

# Distribuição de Weibull

---

A distribuição Weibull é definida como

$$p(x) = \alpha x^{\alpha-1} \beta^{-\alpha} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}, \quad \alpha, \beta > 0. \quad (\text{D.1})$$

A função de distribuição cumulativa é

$$P(x) = 1 - e^{-\left(\frac{x}{\beta}\right)^{\alpha}}. \quad (\text{D.2})$$

O primeiro momento ou valor esperado é dado por

$$m_1 = E[x] = \beta \Gamma\left(1 + \frac{1}{\alpha}\right) \quad (\text{D.3})$$

E a variância é calculada usando

$$\text{Var}[x] = \sigma^2 = \beta^2 \left( \Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma\left(1 + \frac{1}{\alpha}\right)^2 \right) \quad (\text{D.4})$$

## APÊNDICE E

---

---

# Distribuição de Log-normal

---

A distribuição Log-normal é definida como

$$p(x) = \frac{e^{-\frac{(\log(x)-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}x\sigma} \quad \sigma, \mu > 0. \quad (\text{E.1})$$

A função de distribuição cumulativa é

$$P(x) = \frac{1}{2} \left[ \operatorname{erf} \left( \frac{(\log(x) - \mu)}{\sqrt{2}\sigma} \right) + 1 \right]. \quad (\text{E.2})$$

O primeiro momento ou valor esperado é dado por

$$m_1 = E[x] = e^{\mu + \frac{\sigma^2}{2}} \quad (\text{E.3})$$

E a variância é calculada usando

$$\operatorname{Var}[x] = \sigma^2 = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2} \quad (\text{E.4})$$

## APÊNDICE F

---

---

# Distribuição de Pareto

---

A distribuição Pareto é definida como

$$p(x) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}}, \quad \alpha, \beta > 0. \quad (\text{F.1})$$

A função de distribuição cumulativa é

$$P(x) = 1 - \left(\frac{\beta}{x}\right)^\alpha, \quad x > \beta. \quad (\text{F.2})$$

O primeiro momento ou valor esperado é dado por

$$m_1 = E[x] = \frac{\alpha\beta}{(\alpha - 1)}, \quad \alpha > 1. \quad (\text{F.3})$$

E a variância é calculada usando

$$\text{Var}[x] = \sigma^2 = \frac{\alpha\beta^2}{(\alpha - 1)^2(\alpha - 2)}. \quad (\text{F.4})$$

---

---

# Referências Bibliográficas

---

ALENCAR, M. S. *Probabilidade e Processos Estocásticos*. 1. ed. São Paulo, Brasil.: Editora Érica Ltda, 2009. ISBN 978-85-365-0216-8.

ALSHAMMARI, R.; ZINCIR-HEYWOOD, A. N. A flow based approach for ssh traffic detection. *IEEE International Conference on Systems, Man and Cybernetics, ISIC 2007*, Montreal, Canada, p. 296–301, October 2007.

ALTMAN, E.; ARTIGES, D.; TRAORE, K. *On the Integration of Best-Effort and Guaranteed Performance Services*. France, July 1997.

ANDREWS, M. Instability of the proportional fair scheduling algorithm for hdr. *Wireless Communications, IEEE Transactions on*, v. 3, n. 5, p. 1422–1426, sept. 2004.

ANIBA, G.; AISSA, S. Adaptive scheduling for mimo wireless networks: cross-layer approach and application to hsdpa. *Wireless Communications, IEEE Transactions on*, v. 6, n. 1, p. 259–268, jan. 2007.

BEVERLY, R.; CLAFFY, K. C. Wide-area IP multicast traffic characterization. *IEEE Network*, v. 17, n. 1, p. 8–15, 2003.

BO, L.; PARISH, D. J.; SANDFORD, J. M. Using tcp packet size distributions for application detection. *The 7th Annual PostGraduate Symposium on The Convergence of Telecommunications, Networking and Broadcasting*, 2006. Loughborough University.

CAIDA. *The Cooperative Association for Internet Data Analysis – Packet size distribution comparison between Internet links in 1998 and 2008*. 2008. [www.caida.org/research/traffic-analysis/pkt\\_size\\_distribution/graphs.xml](http://www.caida.org/research/traffic-analysis/pkt_size_distribution/graphs.xml). Access in October 2010.

CALLADO, A. *et al.* A survey on internet traffic identification. *Survey and Tutorial, IEEE Communications*, v. 11, n. 3, p. 37–53, Oct. 2009.

CASTRO, E. *et al.* A packet distribution traffic model for computer networks. *The International Telecommunications Symposium, ITS 2010, Manaus, Brazil*, p. 1–5, September 2010.

- CROTTI, M. *et al.* Traffic classification through simple statistical fingerprinting. *SIGCOMM Comput. Commun. Rev.*, ACM, New York, NY, USA, v. 37, p. 5–16, January 2007.
- CROVELLA, M. E.; BESTAVROS, A. *Explaining World Wide Web Traffic Self Similarity*. Boston University, 1995.
- CUCEJ, Z.; FRAS, M. Data source statistics modeling based on measured packet traffic: A case study of protocol algorithm and analytical transformation approach. In: *Telecommunication in Modern Satellite, Cable, and Broadcasting Services, 2009. TELSISKS '09. 9th International Conference on*. [S.l.: s.n.], 2009. p. 55–64.
- DECASPER, D. *et al.* Router plugins: a software architecture for next generation routers. *SIGCOMM Comput. Commun. Rev.*, ACM, New York, NY, USA, v. 28, p. 229–240, October 1998.
- DUSI, M. *et al.* Using gmm and svm-based techniques for the classification of ssh-encrypted traffic. *44th IEEE international conference on Communications 2009, ICC'09*, Dresden, Germany, p. 702–707, June 2009.
- EBIT. *22a edição do relatório WebShoppers, elaborado pela e-bit, com o apoio da Câmara Brasileira de Comércio Eletrônico*. 2010. [www.webshoppers.com.br/webshoppers/WebShoppers22.pdf](http://www.webshoppers.com.br/webshoppers/WebShoppers22.pdf). Acessado em Novembro de 2010.
- GOPALAKRISHNA, R. A. Network packet aggregation. *U. S. Patent US 6614808 B1, Filed Sep. 2, 1999*, 2003.
- GRADSHTEYN, I. S.; RYZHIK, I. M. *Table of Integrals, Series and Products*. 7th edition. ed. [S.l.]: Academic Press – ELSEVIER, 2007. ISBN 0-1237-3637-4.
- GUPTA, P.; MCKEOWN, N. Classifying packets using hierarchial intelligent cuttings. *IEEE Micro*, v. 20, n. 1, p. 34–41, 2000.
- GUPTA, P.; MCKEOWN, N. Algorithms for packets classification. *IEEE Networking*, v. 15, n. 2, p. 24–32, 2001.
- HONG, J. H. *et al.* Performance analysis of packet encapsulation and aggregation. In: *Modeling, Analysis, and Simulation of Computer and Telecommunication Systems, 2006. MASCOTS 2006. 14th IEEE International Symposium on*. [S.l.: s.n.], 2006.
- HONG, J. H.; SOHRABY, K. On the asymptotic analysis of packet aggregation systems. In: *Modeling, Analysis, and Simulation of Computer and Telecommunication Systems, 2007. MASCOTS '07. 15th International Symposium on*. [S.l.: s.n.], 2007.
- IACOVAZZI, A.; BAIOCCHI, A. Optimum packet length masking. *22nd International Teletraffic Congress (ITC), 2010, Amsterdam, The Netherlands, 7-9, September 2010*.

- 
- IECOM. *IECOM – Instituto de Estudos Avançados em Comunicações*. 2010. <http://www.iecom.org.br/>. Accessed in February de 2010.
- IPTRAF. *IPTRAF – IP Network Monitoring Software*. 2010. <http://iptraf.seul.org/>. Accessed in February 2010.
- JALALI, A.; PADOVANI, R.; PANKAJ, R. Data throughput of cdma-hdr a high efficiency-high data rate personal communication wireless system. In: *Vehicular Technology Conference Proceedings, 2000. VTC 2000-Spring Tokyo. 2000 IEEE 51st*. [S.l.: s.n.], 2000. v. 3, p. 1854–1858 vol.3.
- JOHN, W.; TAFVELIN, S. Analysis of Internet backbone traffic and header anomalies observed. *IMC '07: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, New York, NY, USA*, p. 111–116, 2007.
- JUNG, H. H.; SOHRABY, K. On modeling, analysis, and optimization of packet aggregation systems. *IEEE Transactions on Communications*, v. 58, n. 2, p. 660–668, 2010.
- KARAGIANNIS, T. *et al.* *File-sharing in the Internet: A characterization of P2P traffic in the backbone*. November 2003. University of California, Riverside, USA, CA 92521, Tech. Rep.
- KARAGIANNIS, T.; PAPAGIANNAKI, K.; FALOUTSOS, M. *BLINC: Multilevel Traffic Classification in the Dark*. August 2005. SIGCOMM'05, Philadelphia, Pennsylvania, USA.
- KOS, A.; PUSTISEK, M.; BESTER, J. Characteristics of real packet traffic captured at different network locations. *Computer as a Tool. The IEEE Region 8 EUROCON*, v. 1, p. 164–168, 2003.
- LEGEDZA, U.; WETHERALL, D.; GUTTAG, J. Improving the performance of distributed applications using active networks. In: *INFOCOM '98. Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*. [S.l.: s.n.], 1998. v. 2, p. 590–599.
- LELAND, W. E. *et al.* On the self-similar nature of ethernet traffic. *IEEE ACM Trans. on Networking*, v. 2, n. 1, p. 1–15, February 1994.
- LIBERATORE, M.; LEVINE, B. N. Inferring the source of encrypted http connections. *Proceedings of the 13th ACM conference on Computer and communications security, CCS '06*, Alexandria, Virginia, USA, October 2006.
- LIBERATORE, M.; TEIXEIRA, R.; SALAMATIAN, K. Early application identification. *CoNEXT, 2006*, Lisboa, Portugal, December 2006.
- MACIAN, C.; FINTHAMMER, R. An evaluation of the key design criteria to achieve high update rates in packet classifiers. *Network, IEEE*, v. 15, n. 6, p. 24–29, nov 2001.

- 
- MAPLE. *Maple – The Essential Tool for Mathematics and Modeling*. 2010. <http://www.maplesoft.com/products/Maple/index.aspx>. Accessed October 2010.
- MATLAB. *Matlab – The Language of Technical Computing*. 2010. <http://www.mathworks.com/products/matlab/>. Accessed October 2010.
- MATTILA, V. *Traffic Analysis – A review of Internet traffic packet size distributions*. 2010. [http://poliisi.iki.fi/~ville/sekalaiset/Internet/traffic\\_analysis/packet\\_size\\_distribution](http://poliisi.iki.fi/~ville/sekalaiset/Internet/traffic_analysis/packet_size_distribution). Accessed November 2010.
- MCCREARY, S.; CLAFFY, K. C. *Trends in Wide Area IP Traffic Patterns - A View from Ames Internet Exchange*. Proceedings of the 13th ITC Specialist Seminar on Internet Traffic Measurement and Modeling. Monterey, CA, 2000.
- MCGREGOR, A. *et al.* Flow clustering using machine learning techniques. *Passive and Active Network Measurement, PAM 2004*, Antibes Juan-les-Pins, France, p. 205–214, April 2004.
- MCKEOWN, M.; VARGHESE, G. Fast ip packet forwarding and classification for next generation internet. *IEEE Networking*, v. 15, p. 6–7, 2001.
- MOORE, A. W.; ZUEV, D. Internet traffic classification using bayesian analysis techniques. *SIGMETRICS Perform. Eval. Rev.*, ACM, New York, NY, USA, v. 33, p. 50–60, June 2005.
- MUSHTAQ, S. A.; RIZVI, A. A. Statistical analysis and mathematical modeling of network (segment) traffic. *Proceedings of the IEEE Symposium on Emerging Technologies*, p. 246–251, September 2005.
- NIRKHE, V.; BAUGHER, M. Quality of service support for networked media players. In: *Proceedings of the 40th IEEE Computer Society International Conference, COMPCON '95*. Washington, DC, USA: IEEE Computer Society, 1995. ISBN 0-8186-7029-0.
- PARISH, D. *et al.* Using packet size distributions to identify real-time networked applications. *IEEE Proceedings Communication*, v. 4, n. 150, p. 221–227, August 2003.
- PARK, K.; WILLINGER, W. *Self-similar network traffic and performance evaluation*. 1. ed. United States of America: Wiley-Interscience, 2000. (1, 1). ISBN 0-4713-1974-0.
- PAXSON, V.; FLOYD, S. Wide-area traffic: The failure of poisson modeling. *IEEE/ACM Transactions on Networking*, v. 3, n. 3, p. 226–244, Jun 1995.
- PNAD-2009. *Pesquisa Nacional por Amostra de Domicílios (PNAD) 2009 – Instituto Brasileiro de Geografia e Estatística (IBGE)*. 2010. <http://www.ibge.gov.br>. Acessado em Novembro de 2010.



- 
- PRIES, R. *et al.* Traffic measurement and analysis of a broadband wireless Internet access. *IEEE 69th Vehicular Technology Conference. VTC Spring*, p. 1–5, April 2009.
- QUAN, Z.; CHUNG, J.-M. A novel analysis of queue length in differentiated services networks with self-similar arrival processes. *Circuits and Systems*, v. 3, August 2002. The 2002 45th Midwest Symposium, MWSCAS-2002.
- RFC-1042. *A Standard for the Transmission of IP Datagrams over IEEE 802 Networks*. February 1988. <http://www.faqs.org/rfcs/rfc1042.html>. Access in October 2010.
- RFC-1191. *Path MTU Discovery*. November 1990. <http://www.faqs.org/rfcs/rfc1191.html>. Access in October 2010.
- ROUGHAN, M.; VEITCH, D. On-line estimation of the parameters of long-range dependence. *IEEE GLOBECOM'98, Sydney, Australia*, v. 6, n. 1, p. 3716–3721, November 1998.
- SANG, A. *et al.* A flexible downlink scheduling scheme in cellular packet data systems. *Wireless Communications, IEEE Transactions on*, v. 5, n. 3, p. 568 – 577, March 2006.
- SHAFFER, S.; WEISS, D.; CASUBA, J. Method for constructing adaptive packet lengths in a congested network. *U. S. Patent US 6003089 B1, Filed Mar. 31, 1997*, 1999.
- SINHA, R.; PAPADOPOULOS, C.; HEIDEMANN, J. *Internet Packet Size Distributions: Some Observations*. [S.l.], 2007.
- SPIEGEL, M. R. *Estatística*. 3st edition. ed. São Paulo, Brasil: Makron Books, Coleção Schaum, 2006. ISBN 8-5346-0120-8.
- SPRINT. *The Academic Research group at Sprint*. 2005. <https://research.sprintlabs.com/packstat/packetoverview.php>. Access in October 2010.
- SUN, Q. *et al.* Statistical identification of encrypted web browsing traffic. In: *Proceedings of the 2002 IEEE Symposium on Security and Privacy*. Washington, DC, USA: IEEE Computer Society, 2002. p. 19–. ISBN 0-7695-1543-6.
- TANENBAUM, A. S. *Redes de Computadores*. 4. ed. Rio de Janeiro: Editora Campus, 2003. ISBN 8-5352-1185-3.
- TORABZADEH, M.; AJIB, W. Packet scheduling and fairness for multiuser mimo systems. *Vehicular Technology, IEEE Transactions on*, v. 59, n. 3, p. 1330 –1340, mar 2010.
- WILLINGER, W.; PAXSON, V.; TAQQU, M. S. *Self-similarity and Heavy-Tails: Structural Modeling of Network Traffic*. Boston: Birkh auser, 1998. *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. ISBN 0-8176-3951-9.

---

WOLFRAM. *The Wolfram Functions Site*. October 2001. <http://functions.wolfram.com/07.20.27.0001.01>. Access in October 2010.

WRIGHT, C. V. *et al.* Language identification of encrypted voip traffic: Alejandra y roberto or alice and bob? In: *Proceedings of 16th USENIX Security Symposium on USENIX Security Symposium*. Berkeley, CA, USA: USENIX Association, 2007.

WRIGHT, C. V. *et al.* Traffic morphing: An efficient defense against statistical traffic analysis. *16th Annual Network and Distributed System Security Symposium (NDSS)*, San Diego, CA, February 2009.

WRIGHT, C. V.; MONROSE, F.; MASSON, G. M. On inferring application protocol behaviors in encrypted network traffic. *J. Mach. Learn. Res.*, v. 7, p. 2745–2769, December 2006.

YU, C.; RAVINDRAN, V.; LEON-GARCIA, A. Internet traffic characterization using packet-pair probing. *26th IEEE International Conference on Computer Communications. INFOCOM 2007.*, p. 1766–1774, 2007.

ZANDER, S.; NGUYEN, T.; ARMITAGE, G. Automated traffic classification and application identification using machine learning. *IEEE Conference on Local Computer Networks, LCN 2005*, Sydney, Australia, November 2005.

ZHAOBIAO, L. *et al.* Performance analysis of burst assembly under self-similar traffic with measured wan packet size distribution. *First International Conference on Communications and Networking in China. ChinaCom '06.*, p. 1–5, 2006.