

UNIVERSIDADE FEDERAL DO PARANÁ

EDGARD MASSAHIRO MUNETIKO

*ANÁLISE DE UM NOVO MODELO DE TRÁFEGO VOIP*

Curitiba

2010

EDGARD MASSAHIRO MUNETIKO

*ANÁLISE DE UM NOVO MODELO DE TRÁFEGO VOIP*

Dissertação apresentada ao Curso de Pós-Graduação em Engenharia Elétrica, Área de Concentração em Telecomunicações, Departamento de Engenharia Elétrica UFPR, como parte das exigências para a obtenção do título de MESTRE em Engenharia Elétrica.

**Orientador: Prof. Dr. Eduardo Parente Ribeiro**

**Co-orientador: Prof. Dr. Carlos Marcelo Pedroso**

Curitiba

2010

# TERMO DE APROVAÇÃO

EDGARD MASSAHIRO MUNETIKO

## *ANÁLISE DE UM NOVO MODELO DE TRÁFEGO VOIP*

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre no Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Paraná

Aprovado em 30 de Julho de 2010

### **BANCA EXAMINADORA**

---

Prof. Dr. Eduardo Parente Ribeiro

---

Prof. Dr. Carlos Marcelo Pedroso  
Doutor em Engenharia Elétrica - UFPR

---

Prof. Dr. Evelio Martín García Fernández  
Doutor em Engenharia Elétrica - UFPR

---

Prof. Dr. Ricardo Nabhen  
Doutor em Engenharia da Computação - PUCPR

*À minha família, pelo apoio.*

*Aos amigos, pelo apoio e companheirismo.*

## Agradecimentos

À toda a minha família, pelo encorajamento, apoio e compreensão nos momentos de ausência.

Aos professores Eduardo Parente Ribeiro e Carlos Marcelo Pedroso pelos ensinamentos, orientação, e principalmente, pela paciência, sem a qual este trabalho não se realizaria.

Aos Professores Solivan Arantes Valente e Marcia Olandoski pelos valiosos conhecimentos transmitidos durante o período.

Aos Professores Pavanello e Tonidandel pela oportunidade de complementar as disciplinas.

Aos engenheiros João Marcos Menon, Hermes Assis e Mauro Antonio da Silveira, pelo incentivo e confiança.

Aos colegas pelo tempo e experiência compartilhados.

## Resumo

O crescimento da migração do tráfego de telefonia tradicional para os serviços VoIP (Voice over IP) tem atraído a atenção das operadoras de telefonia (fixa e celular), prestadoras de serviço VoIP, fornecedores de equipamentos e outros, para o problema da eficiência do dimensionamento das redes de backbone, de forma a se ter a maior economia e confiabilidade possível.

Os modelos existentes até há pouco tempo atrás (geralmente modelos Poissonianos ou exponenciais) se mostraram ineficientes neste aspecto quando aplicado a redes VoIP. Este trabalho vem propor a utilização de um modelo originalmente proposto para tráfego web (SURGE) adaptado para a modelagem de tráfego VoIP. Este modelo faz uma análise de parâmetros obtidos facilmente através de monitoramento feito diretamente na rede em estudo: tempo de duração da sessão, intervalo entre sessões, intervalo entre pacotes e tempo de envio dos pacotes.

Para se verificar a qualidade do modelo proposto, será gerado um tráfego simulado com as distribuições e parâmetros obtidos e os resultados serão comparados com os valores empíricos.

Palavras-chaves: VoIP, Modelo de trafego, rede de voz.

# Abstract

The growth in migration from traditional voice traffic to VoIP has attracted the attention of telecom operators (land and mobile), VoIP service providers, equipment suppliers and others to the efficiency problem of scaling the backbone networks, in order to have the largest possible economy and reliability.

The existing models until recently (usually Poissonian or exponential models) are inefficient in this respect when applied to VoIP networks. This paper proposes the use of a model originally proposed for web traffic (SURGE) adapted for the modeling of VoIP traffic. This model analyzes the parameters obtained easily by directly monitoring the network under study: duration of the session, interval time between sessions, interval time between packets and time sending packets.

In order to assess model behavior, simulated traffic was generated with the parameters and distributions obtained from real traffic and results were compared with empirical values.

Keywords: VoIP, Traffic model, voice network.

# Sumário

<b>Lista de Figuras</b>	<b>vii</b>
<b>Lista de Tabelas</b>	<b>ix</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Objetivos . . . . .	1
1.2 Contexto . . . . .	1
1.3 Estrutura . . . . .	3
<b>2 Revisão da Literatura</b>	<b>4</b>
2.1 Modelo SURGE . . . . .	4
2.2 Outros trabalhos relacionados ao assunto . . . . .	6
2.3 Modelo ON-OFF . . . . .	9
2.4 Modelos de tráfego modernos . . . . .	10
2.4.1 Distribuições de cauda pesada . . . . .	10
2.4.2 Auto-similaridade . . . . .	11
2.5 Protocolos VoIP . . . . .	14
2.5.1 BICC (Bearer Independent Call Control) . . . . .	15
2.5.2 RTP . . . . .	17
<b>3 Descrição do Modelo Proposto</b>	<b>20</b>
3.1 Coleta dos dados . . . . .	21

3.2	Topologia da rede em estudo . . . . .	23
<b>4</b>	<b>Resultados</b>	<b>24</b>
4.1	Intervalo entre sessões . . . . .	24
4.2	Tempo de duração da sessão . . . . .	26
4.3	Análise do tráfego de pacotes . . . . .	28
4.3.1	Tempo de transmissão dos pacotes . . . . .	28
4.3.2	Intervalo entre pacotes . . . . .	29
4.4	Simulação . . . . .	30
<b>5</b>	<b>Considerações Finais</b>	<b>36</b>
5.1	Análise dos resultados e discussão . . . . .	36
5.2	Conclusão e estudos futuros . . . . .	37
	<b>Referências Bibliográficas</b>	<b>38</b>

## Lista de Figuras

2.1	ON/OFF aplicado no SURGE . . . . .	5
2.2	ACF para codec G.723 . . . . .	7
2.3	CDF e ACF para codecs de taxa variável (AMR e iSAC) . . . . .	8
2.4	Comparação entre cauda leve e cauda pesada . . . . .	10
2.5	Processo Auto-similar do conjunto de Cantor . . . . .	12
2.6	Auto-similaridade no tráfego de pacotes . . . . .	13
2.7	BICC e RTP . . . . .	16
2.8	Troca de sinalização BICC . . . . .	17
2.9	Pilha RTP . . . . .	18
3.1	Modelo Proposto . . . . .	20
3.2	Topologia da rede em estudo . . . . .	23
4.1	Distribuição acumulada do intervalo entre sessões . . . . .	25
4.2	Comparação do QQPlot do intervalo entre sessões com outras distribuições . . . . .	26
4.3	Distribuição acumulada do tempo de duração da sessão . . . . .	27
4.4	Comparação do QQPlot do tempo de duração da sessão com outras distribuições . . . . .	27
4.5	Distribuição complementar em escala log do tempo de duração da sessão . . . . .	29
4.6	Distribuição acumulada do tempo de intervalo entre pacotes . . . . .	30
4.7	Comparação do QQPlot do intervalo entre pacotes com outras distribuições . . . . .	31
4.8	Topologia da simulação . . . . .	32

4.9	Distribuição acumulada do tempo de intervalo entre pacotes do tráfego agregado para os dados da simulação . . . . .	32
4.10	Comparação do QQPlot para os dados da simulação do tempo de intervalo entre pacotes do tráfego agregado . . . . .	33
4.11	Comparação da função de auto-correlação do tráfego agregado da simulação em diferentes escalas . . . . .	34
4.12	Comparação da ACF do tráfego agregado da simulação para diferentes valores de alfa . . . . .	35

## Lista de Tabelas

2.1	Mensagens BICC . . . . .	16
3.1	Período de coleta dos dados . . . . .	22
4.1	Média e desvio padrão do intervalo entre sessões . . . . .	25
4.2	Tempo de transmissão dos pacotes . . . . .	29
4.3	Resultados da modelagem . . . . .	32
4.4	Estatísticas da quantidade de bytes enviados dos dados simulados . . . . .	33
4.5	Parâmetro de Hurst para as simulações . . . . .	35

# 1 Introdução

## 1.1 Objetivos

O objetivo deste trabalho é propor um modelo de tráfego de chamadas de uma rede de telecomunicações comercial que utiliza o VoIP para transporte de chamadas originadas da rede celular com destino a rede fixa e outras operadoras celulares.

Este trabalho irá adaptar o modelo SURGE [1] (*Scalable URL Reference Generator*) (descrito no capítulo 2.1), criado para ser aplicado em tráfego web, para o problema específico de um tráfego VoIP, onde será estudado o comportamento do fluxo de pacotes RTP [2] (descrito no capítulo 2.5.2) dentro do intervalo de duração de uma sessão, gerando o modelo composto proposto (descrito no capítulo 3). A superposição de diversas sessões fornecerá o tráfego agregado total, e espera-se que tenha uma boa aderência com o tráfego agregado total observado na rede. Desta forma, será obtido um modelo simples que permite a realização de estudos de desempenho de maneira relativamente fácil e com poucos parâmetros que podem ser estimados também facilmente.

## 1.2 Contexto

O rápido crescimento do número de usuários de internet está acelerando a migração da rede de telefonia tradicional, baseada em circuitos, para o VoIP, baseado em rede de pacotes. Serviços oferecidos por operadoras de telefonia (como o Vono da GVT e o NetFone da Net) ou o Skype, por exemplo, estão cada vez mais em destaque no leque de produtos oferecidos. Esta situação tem atraído a atenção de operadoras, fornecedores de equipamentos de telecomunicações e consumidores, que exigem cada vez mais qualidade e preço menor.

Este crescimento da implementação da voz sobre pacotes causa um incremento de tráfego nos backbones, fazendo com que os provedores de rede sejam obrigados a investir em

sua ampliação para manter a qualidade do serviço. Por outro lado, para reduzir ao mínimo o investimento nestas ampliações, é necessário se fazer um dimensionamento adequado de forma a não se ter desperdício com recursos ou, ao contrário, um congestionamento na rede que pode reduzir a qualidade do serviço. Empresas que prestam serviços baseados em VoIP necessitam oferecer redes de alta performance com baixo *jitter* e atraso nos pacotes e largura de banda suficiente para o tráfego.

Uma forma de se otimizar a ocupação da largura de banda das redes de comunicações é reduzir ao mínimo possível o tráfego de pacotes da voz e uma das formas de se fazer isto é utilizando codecs (codificadores do sinal analógico de voz para um fluxo de pacotes) eficientes. No caso do GSM (*Global System for Mobile Communications*, ou Sistema Global para Comunicações Móveis), que foi utilizada na topologia analisada neste trabalho, é utilizado o codec AMR (*Adaptive Multi-Rate*) [3][4]. O AMR é um codec com taxa de compressão variável, que depende da qualidade do enlace por onde o tráfego é transmitido.

Com uma modelagem adequada do tráfego VoIP pode-se também otimizar a utilização dos recursos de rede, pois, pode-se com isto, desenvolver uma ferramenta computacional de simulação com o objetivo de garantir a qualidade do serviço, simulando e prevendo o que acontece quando se altera alguma característica do tráfego como, por exemplo, o perfil da duração da chamada ou a quantidade de usuários da rede.

Os modelos utilizados até há pouco tempo atrás [5][6][7] eram baseados em modelos markovianos, geralmente Poissonianos ou Gaussianos, não atendendo a necessidade das redes de comunicações modernas que se utilizam do VoIP, já que este tipo de tráfego não apresenta uma regularidade estatística que mostra algum tipo de convergência para um processo Gaussiano comum (ruído branco Gaussiano). Esta tecnologia possui características próprias, diferentes da telefonia convencional, e esta nova característica do tráfego VoIP fez com que fosse necessário uma nova visão de análise, que levasse em conta também as características do tráfego dos pacotes IP. Um bom modelo deve levar em consideração aspectos relevantes para caracterização do tráfego e estudos de desempenho. A base do modelo utilizado neste trabalho é o modelo SURGE e entre os aspectos de desempenho analisados encontram-se a duração das sessões, intervalo entre as sessões sucessivas, o tempo de transmissão dos pacotes

e o intervalo de tempo entre os pacotes.

Existem diversos trabalhos publicados, alguns abordados no Capítulo 2, onde são estudados os modelos SURGE, ON-OFF ou MMPP em diversos cenários [1] [5] [6] [7] [8] [9] [10]. Este trabalho irá abordar uma análise baseada no modelo SURGE com algumas variações, gerando um modelo composto para um tráfego VoIP, o que ajudará em simulações, previsões e no dimensionamento adequado da rede. Este modelo composto deverá considerar além da sessão, as características do fluxo dos pacotes RTP. A aplicação de uma variação do SURGE para tráfego VoIP é inédita e apresenta vantagens sobre a abordagem tradicional, pois deve gerar um modelo simples com parâmetros facilmente obtidos através de monitoramento dos pacotes de dados e da sinalização das chamadas na própria rede em estudo.

## 1.3 Estrutura

Este trabalho está dividido em cinco capítulos, além das referências bibliográficas. Além deste capítulo introdutório, esta dissertação faz, no segundo capítulo, uma introdução ao leitor do estado da arte e dos conhecimentos básicos necessários para o entendimento deste trabalho, como a definição de protocolos, distribuições de probabilidade utilizadas para o desenvolvimento, os modelos de tráfego relacionados e os demais trabalhos relacionados ao tema. O terceiro capítulo apresenta a descrição do modelo proposto e suas qualidades. O quarto capítulo mostra e explica os resultados das análises realizadas, além de apresentar uma simulação computacional que gera um tráfego sintético que possibilita testar a qualidade do modelo proposto. Finalmente, o quinto capítulo apresenta a análise dos resultados, a conclusão e as sugestões de trabalhos futuros.

## 2 Revisão da Literatura

O objetivo deste capítulo é relacionar trabalhos representativos sobre modelagem de tráfegos, onde os principais modelos para caracterização de tráfego VoIP são apresentados em [5][11][12][13]. Este trabalho foi inspirado no modelo SURGE, que é uma ferramenta desenvolvida para modelar e simular tráfego web. Muitos trabalhos publicados sobre o assunto abordam o modelo ON-OFF, que tem algumas características semelhantes a este trabalho, como por exemplo o tempo de duração da sessão. Mas este trabalho, apesar de ter o mesmo objetivo de se obter um modelo de um tráfego VoIP, possui algumas diferenças, que o tornam um modelo eficiente, e com parâmetros de fácil obtenção, já que para isto basta se fazer um monitoramento dos pacotes e da sinalização em tempo real ou mesmo através de dados já coletados anteriormente.

### 2.1 Modelo SURGE

O SURGE (Scalable URL Reference Generator) [1] é uma ferramenta desenvolvida para gerar tráfego web, baseado em seis parâmetros que os autores utilizaram para a modelagem. São eles:

- Tamanho de arquivos: Tamanho do arquivo no servidor.
- Tamanho das requisições: Tamanho do arquivo enviado (nem sempre é o mesmo salvo no servidor).
- Popularidade: Número de requisições feitas para um arquivo no servidor.
- Referências incorporadas: Refere-se a estrutura do objeto web. Este parâmetro pode influenciar no tamanho do tempo de OFF. O tempo de OFF entre referências incorporadas é menor do que o tempo de OFF entre objetos web.

- Localização temporal: probabilidade de um objeto web ser requisitado novamente em um futuro próximo.
- OFF Times: Dividido em active OFF e inactive OFF. O primeiro é o intervalo entre as requisições de referências incorporadas e o segundo é o intervalo entre uma nova requisição de página web e o fim do envio da requisição anterior (para um usuário).

A ferramenta analisa os efeitos da alta variabilidade baseada no conceito de *User Equivalent* (UE), que pode ser definida como um processo em um loop infinito que alterna entre períodos de requisições ao servidor e períodos de inatividade. Cada UE é, portanto, um processo ON/OFF. A Figura 2.1, ilustra os períodos de ON e OFF no modelo SURGE.

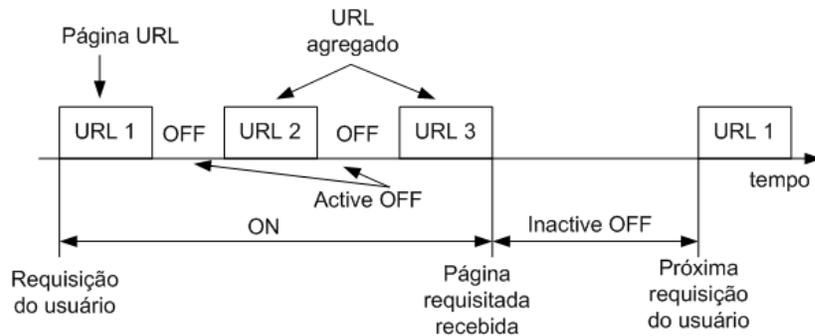


Figura 2.1: ON/OFF aplicado no SURGE (fonte: [1])

Alguns parâmetros do modelo SURGE podem ser comparados com os deste trabalho. A diferença com o modelo apresentado aqui é que o tempo de intervalo entre duas sessões é a diferença entre o início de duas sessões (para usuários diferentes) e o tempo de duração da sessão é o tempo total de cada chamada. O modelo faz um tratamento semelhante a este trabalho, levantando as características dos parâmetros, comparando as funções de distribuição complementar e gerando um tráfego simulado, comparando o resultado com um modelo já existente.

## 2.2 Outros trabalhos relacionados ao assunto

Chen et al. [11] mostram que as características de tráfego VoIP são diferentes das características de tráfego de chamadas tradicionais. Os autores propõem uma nova aproximação para o tempo de duração das chamadas, utilizando uma distribuição que se aproxima de uma união de duas distribuições log-normal. O artigo mostra que o tempo de duração de uma chamada celular VoIP não é exponencial, como era esperado.

Adas [5] publicou um abrangente artigo sobre vários modelos de tráfego, como o modelos ON-OFF, MMPP, baseados em fluidos e modelos regressivos. Descreve também modelos de tráfego estacionário de dependência de longa duração, de curta duração, auto-similaridade e modelo browniano fracionário. Grande parte dos tipos de modelos utilizados em estudos de tráfego estão descritos neste artigo. O mesmo defende que o modelo utilizado deve ser adequado ao tipo de tráfego a ser estudado. Caso contrário, corre-se o risco de se sobre-estimar, ou sub-estimar a performance da rede.

O trabalho de Wenyu et al. [12] também faz uma análise experimental dos padrões ON-OFF, para dois detectores de silêncio (ITU G.729 Annex B *Voice Activity Detector* (VAD) e NeVoT *Silence Detector* (SD)). Os autores fazem uma análise gráfica da função de distribuição complementar (CDF) comparando os dois detectores de silêncio. Os resultados indicam que as distribuições de ON e OFF são sensíveis aos níveis de volume e tipo de detectores de silêncio, mas ambos mostraram que as suposições tradicionais de distribuição exponencial também não se ajustaram às sessões de áudio gravadas. As duas distribuições (ON e OFF) se aproximaram mais de uma distribuição de cauda pesada do que uma curva exponencial.

Menth et. al [13] publicaram um artigo onde é usado um conjunto representativo de amostras de voz especificamente feitos para pesquisa, codificados com diversos codecs, analisado suas saídas e parametrizados os diferentes modelos. O modelo se ajustou satisfatoriamente tanto à função de distribuição acumulativa complementar dos períodos ON-OFF quanto ao tamanho dos pacotes. Além disso, o modelo reproduz também a função de autocorrelação (ACF) do tamanho dos pacotes quando muitas rajadas de voz são enviadas para

o servidor. Os detectores de silêncio (SD e VAD) utilizados junto com os codecs, podem acrescentar um atraso no início ou no fim do trem de pacotes. Foram analisados os seguintes codecs: G.711 e G.729 (codecs com taxa constante-CBR), G.723 e iLBC (codecs com detector de silêncio) e finalmente, os codecs AMR e iSAC (codecs com taxa variável - VBR). Por causa da dificuldade da análise dos codecs com detector de silêncio, foram discutidos três aproximações para se tratar os períodos de OFF a nível de pacotes durante um período de fala. A análise destas aproximações mostraram que estes codecs, para o período de comprimento de fala e intervalo de silêncio, foi ajustada uma distribuição geométrica e uma distribuição binomial negativa. A análise da ACF mostra que períodos de ON-OFF curtos possuem um rápido decaimento da função ACF, e ao contrário, períodos de ON-OFF mais longos possuem um lento decaimento da função de ACF. É feita uma análise da ACF para vários codecs utilizados, conforme gráfico ilustrado na Figura 2.2. Nesta figura, verifica-se que o modelo que mais se aproxima dos dados empíricos é uma distribuição geométrica e negativa binomial para uma das aproximações (em destaque).

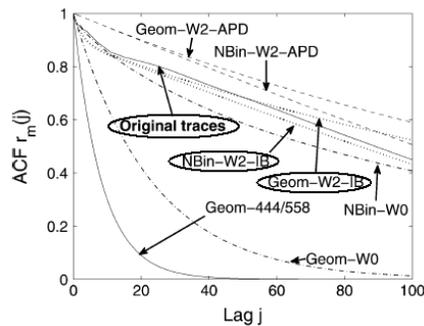


Figura 2.2: Curva ACF para o codec G.723 (fonte: [13])

Foi feita a análise também para os codecs com taxa variável (AMR e iSAC). Para ambos os codecs, obteve-se uma boa aproximação na função de distribuição complementar e na função de auto-correlação, conforme pode-se ver na Figura 2.3. O modelo de Menth et al. possui uma característica que pode dificultar a análise pois, conforme o tipo de codec e detector de silêncio utilizado, é necessário fazer aproximações devido à estrutura irregular do fluxo de pacotes.

Pedroso et al. [9] trazem uma análise da evolução das características do tráfego

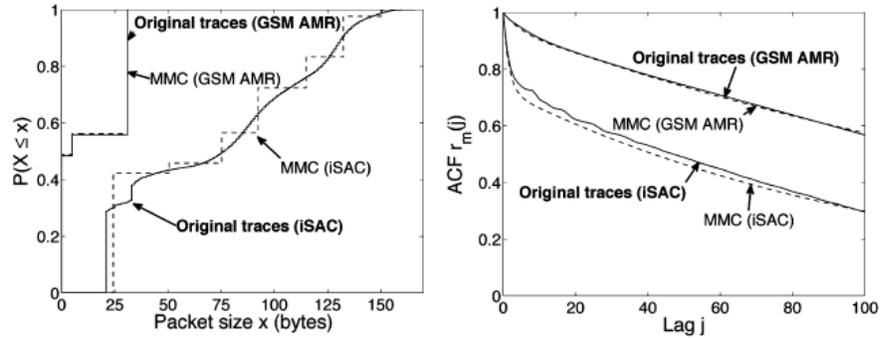


Figura 2.3: CDF (à esquerda) e ACF (à direita) para os codecs de taxa variável (AMR e iSAC)(fonte: [13])

VoIP de uma grande operadora de telecomunicações brasileira que oferece o serviço de VoIP, comparando as características do tráfego dos anos de 2006 e 2007. O trabalho analisou as seguintes características: protocolo, codec, intervalo entre pacotes RTP, tamanho dos pacotes RTP e tráfego agregado total. Em relação aos protocolos utilizados, verificou-se que mais de 90% dos pacotes trafegavam RTP. O restante tratavam a sinalização SIP e outros protocolos. Os autores detectaram que houve um aumento da utilização do codec G729, em detrimento do codec G711, e também uma alteração nas características do intervalo entre pacotes, suavizando sua característica auto-similar, que pode ter sido causado pelo aumento do número de usuários. A análise do tamanho dos pacotes mostrou uma distribuição praticamente discreta nos dois períodos de coleta (2006 e 2007) e mostrou que houve uma perda de sua característica auto-similar verificando-se um rápido decaimento da função de auto-correlação. O estudo do tráfego agregado total mostrou um lento decaimento da função de auto-correlação, indicando que este tráfego é auto-similar. Estas análises foram realizadas através das funções de distribuição acumulada e da função de auto-correlação.

Em outro artigo, Pedroso et al. [10] mostram um modelo que pode descrever o comportamento das sessões de conversação de duas grandes operadoras de telecomunicações do Brasil. Os parâmetros analisados são o tempo de duração da sessão e o intervalo entre chegadas de sessões. Esta análise foi feita com dados empíricos coletados na rede de cada operadora. O estudo da distribuição acumulada de probabilidade do intervalo entre chegadas

mostrou uma boa aderência deste parâmetro a uma distribuição exponencial para ambas as operadoras. O teste da função de auto-correlação também mostrou que o intervalo de tempo entre chegadas é independente em relação ao deslocamento, mostrando ser um processo Markoviano. A análise do tempo de duração da sessão mostrou em ambas as operadoras uma característica de cauda pesada, indicando a ocorrência de chamadas de longa duração. Para comprovar a aderência dos dados ao modelo foi utilizado o gráfico quantile-quantile (QQPlot), que mostra que os dados dos dois conjuntos (empíricos e o modelo) possuem a mesma distribuição de probabilidade.

Como foi visto, existem diversas formas de modelagem. Este trabalho propõe, fundamentalmente, uma abordagem caracterizando o comportamento do usuário ao invés do tráfego agregado, ou seja, a análise dos parâmetros relacionados às sessões.

## 2.3 Modelo ON-OFF

O ON-OFF é um padrão amplamente utilizado para caracterizar a fala humana. Foi estudado por Willinger [14] para caracterizar tráfego ethernet, onde estudos realizados mostraram que os modelos tradicionais utilizados na época mostravam-se inadequados. Este padrão caracteriza-se pela alternância de períodos de atividade (ON) e períodos de inatividade (OFF). O período de ON é o intervalo de tempo onde a fala é transmitida e o período de OFF é o intervalo de tempo onde não é transmitido nenhum sinal.

A princípio, os períodos de ON e OFF podem assumir qualquer tipo de distribuição, mas nos modelos tradicionais de tráfego [15] [11] os períodos de ON e OFF são modelados por uma distribuição exponencial. Estudos mais recentes [12] mostraram que o tempo de duração de sessão numa rede VoIP pode ser modelado por uma distribuição de cauda pesada (Capítulo 2.4.1).

## 2.4 Modelos de tráfego modernos

### 2.4.1 Distribuições de cauda pesada

Em estatística, distribuições de cauda pesada são utilizadas para explicar variados fenômenos naturais, como a distribuição geográfica da população mundial, a distribuição da riqueza, etc [16]. A característica desta distribuição é uma cauda longa, significando que a probabilidade de existir valores extremos no conjunto é significativa.

Uma variável aleatória  $X$ , possui uma distribuição de cauda pesada se obedecer a equação:

$$Pr\{X > x\} \sim cx^{-\alpha}, x \rightarrow \infty \quad (2.1)$$

onde  $\alpha$  é chamado de parâmetro de forma e  $c$  é uma constante positiva.

Isto significa que a cauda da distribuição decai hiperbolicamente, de maneira diferente de uma distribuição de cauda leve, como a exponencial ou gaussiana, cujo decaimento da cauda é exponencial. A Figura 2.4 ilustra a diferença de uma distribuição de cauda leve (linha contínua) e cauda pesada (linha tracejada).

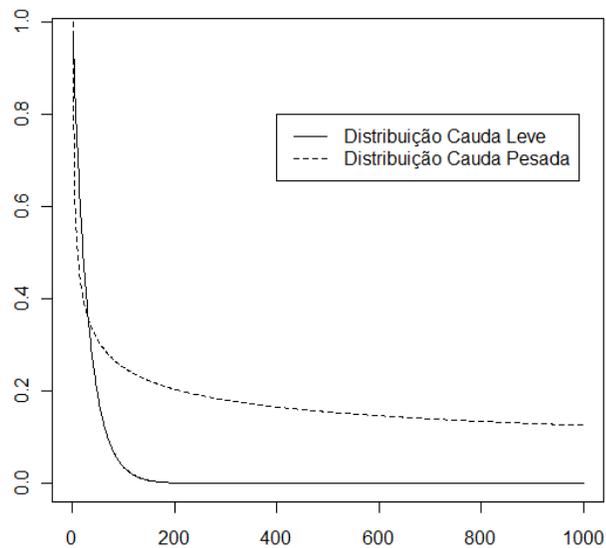


Figura 2.4: Comparação entre cauda leve e cauda pesada

## Distribuição de Pareto

A distribuição de Pareto é uma distribuição de cauda pesada, utilizada originalmente por Vilfredo Pareto para descrever a distribuição de riqueza na população, onde se descobriu que 80% da riqueza estava distribuída em apenas 20% da população [16]. Esta relação está presente em outros fenômenos na internet como o tamanho dos arquivos enviados por servidores web [1][17] ou o tempo de duração das sessões VoIP [18].

Uma variável aleatória  $X$ , possui a probabilidade de ser maior do que um valor  $x$ , com distribuição de Pareto, se obedecer a equação (2.2)

$$Pr\{X > x\} = \left(\frac{\beta}{x}\right)^\alpha \quad (2.2)$$

onde  $X$  é uma variável aleatória com uma Distribuição de Pareto,  $\alpha$  é o parâmetro de forma positivo e  $\beta$  é chamado de escala ou parâmetro de localização.

O valor esperado de  $X$ , para  $\alpha > 1$  é dado pela equação (2.3).

$$E(X) = \left(\frac{\alpha\beta}{\alpha - 1}\right) \quad (2.3)$$

e a variância de  $X$ , para  $\alpha > 2$  é definido pela equação (2.4).

$$var(X) = \left(\frac{\beta}{\alpha - 1}\right)^2 \left(\frac{\alpha}{\alpha - 2}\right) \quad (2.4)$$

Neste trabalho foi utilizada uma variação da distribuição de Pareto (Pareto tipo II) [19] definida como:

$$Pr\{X > x\} = \beta\left(\frac{1}{x^\alpha}\right) - 1 \quad (2.5)$$

### 2.4.2 Auto-similaridade

O fenômeno da auto-similaridade foi introduzido por Mandelbrot [20], em um trabalho em que mostra que as características do fenômeno são preservadas mesmo se forem alteradas as escalas de tempo ou espaço dos dados, de modo a satisfazer a relação dada por [21]:

$$X_{mt} \cong_d m^H X_t, m > 0 \quad (2.6)$$

onde  $X_{mt}$  é um processo estocástico auto-similar de um processo  $X_t$ , reescalonado multiplicado por um fator  $m^H$ .

Como exemplo, pode-se citar o conjunto de Cantor [22]. Este processo é ilustrado na Figura 2.5, onde inicia-se com um bloco de lado unitário. Divide-se cada lado em 3 partes iguais e obtém-se 4 cópias do bloco original reescalonado em 1/3. Repetindo-se este processo infinitas vezes, os blocos mantêm sua característica inicial redimensionada.

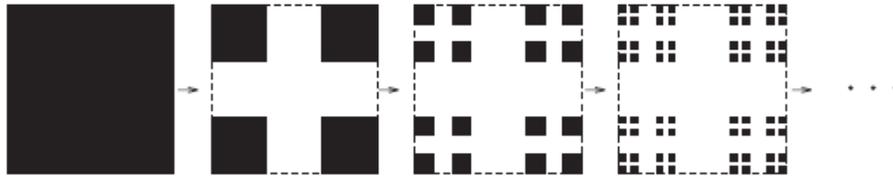


Figura 2.5: Processo Auto-similar do conjunto de Cantor (fonte: [22])

A auto-similaridade em tráfego de rede é uma propriedade que pode ser visualizada graficamente. A Figura 2.6 ilustra o tráfego de uma rede, em bytes, em quatro escalas de tempo diferentes. Percebe-se a semelhança visual entre os gráficos nas diferentes escalas. Esta característica é adquirida quando várias sessões com propriedades com distribuição de cauda pesada são sobrepostas [14].

O trabalho publicado por Leland et al. [21] em 1993 colocou em destaque a auto-similaridade no tráfego de rede. Este trabalho analisou o paradigma dos modelos existentes, onde os dados de tráfego eram caracterizados como modelos markovianos que não apresentavam característica auto-similar ou fractal, como, por exemplo, o modelo Poissoniano. O trabalho de Leland faz uma demonstração de que o tráfego de uma rede LAN é estatisticamente auto-similar, o que não é corretamente representado nos modelos normalmente utilizados na época que não mostraram este comportamento fractal, ou seja, um comportamento que mantém suas características mesmo alterando a escala de visualização. Este comportamento tem sérias implicações no projeto, controle e análise da rede e a agregação de pacotes deste tráfego tipicamente intensifica a característica da auto-similaridade, ao invés de amortecê-lo. A característica crítica do tráfego auto-similar é que não há um comprimento padrão de rajada: em observações realizadas em todas as escalas de tempo, de alguns

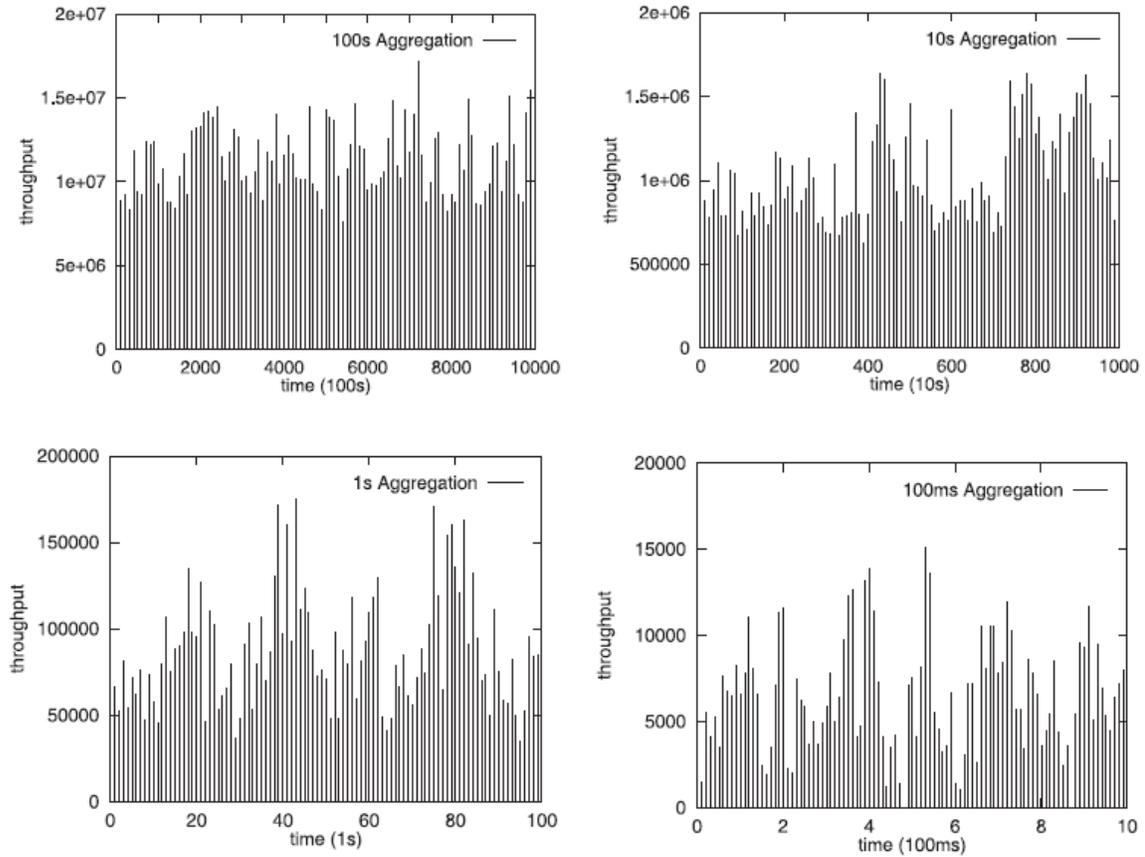


Figura 2.6: Auto-similaridade no tráfego de pacotes(fonte: [22])

milisegundos até minutos ou horas, a rajada de tráfego auto-similar é evidente.

O comportamento de um processo auto-similar tem importantes implicações no estudo do dimensionamento da disponibilidade da largura de banda, quando o estudo do tráfego é realizado em várias escalas de tempo. Um dimensionamento feito sem levar em conta as características auto-similares da rede pode levar a perda de pacotes, prejudicando a qualidade do serviço.

Outro termo utilizado neste trabalho é "dependência de longa duração" ("LRD" do termo em inglês *Long Range Dependence*) e refere-se ao fenômeno que ocorre quando uma análise de uma série de dados temporais mostra alguma correlação no tempo. Este termo se refere ao comportamento da cauda da função de auto-correlação de séries temporais estacionárias que apresenta um lento decaimento.

A LRD é caracterizada pelo parâmetro H, o parâmetro de Hurst<sup>1</sup> [23]. Hurst observou que seus dados apresentavam um comportamento com características de LRD e este parâmetro oscilava no intervalo entre 0,5 e 1. Há várias formas de se calcular este parâmetro [24][25] e neste trabalho o parâmetro será calculado através da transformada de wavelet.

## 2.5 Protocolos VoIP

Os protocolos VoIP podem ser divididos em dois grupos: protocolos de sinalização e protocolos de transporte.

Entre os protocolos de sinalização destacam-se o SIP [26], o H.323 [27] e o BICC [28] [29]. O Protocolo de Iniciação de Sessão (SIP) é um protocolo de aplicação, que utiliza o modelo "requisição-resposta", similar ao HTTP, para iniciar sessões de comunicação interativa entre usuários. É um padrão da IETF (*Internet Engineering Task Force*) padronizado na RFC 3261 de 2002 [26]. O H.323 é parte da família de recomendações ITU-T (*International Telecommunication Union*) H.32x, que pertence a série H da ITU-T. O BICC (descrito no capítulo 2.5.1), utilizado neste trabalho, é um protocolo, também do ITU-T, utilizado em redes de telecomunicações para controle de chamadas baseado em sinalização de banda estreita.

Os protocolos de transporte são protocolos utilizados pelas *Media Gateway* (MGW) para envio dos pacotes do fluxo de áudio. Os mais utilizados são o RTP e o RTCP. O RTP (descrito no Capítulo 2.5.2) é um protocolo de rede utilizado para aplicações em tempo real, que define a forma de fragmentação dos pacotes, adicionando o número de sequência e o tempo. O RTCP é um protocolo que trabalha em conjunto com o RTP e tem a finalidade de troca de informações entre os participantes da sessão. Ambos são definidos pela RFC 3550 do IETF (*Internet Engineering Task Force*).

Neste trabalho serão estudados os protocolos BICC e RTP, que são os protocolos

---

<sup>1</sup>Hurst foi um hidrólogo que se dedicou a estudar, em 1951, o nível do rio Nilo, com o objetivo de projetar um reservatório para regular o fluxo deste rio

utilizados na rede analisada.

### 2.5.1 BICC (Bearer Independent Call Control)

O *Bearer Independent Call Control* (BICC) é um protocolo de sinalização de controle de chamadas baseado nas especificações existentes no protocolo equivalente de banda estreita. O BICC foi desenvolvido para permitir que se forneça os mesmos serviços sobre um backbone de banda larga que a telefonia convencional provê em redes de banda estreita.

O primeiro conjunto de características BICC foi especificado pela ITU-T nas recomendações Q.1901 em 2000. Provê controle de chamada em uma rede ATM, e foi evoluído na segunda versão na recomendação Q.1902 em 2001. Esta versão adicionou o BCP (*Bearer Control Protocol*), que é a habilidade para se controlar o transporte de voz de entidades remotas e a capacidade de oferecer negociação de codec e modificações. Após isto, o BICC não sofreu grandes alterações.

Este protocolo separa as camadas de controle de chamadas e controle de conexão de fluxo do tráfego de pacotes de voz (Figura 2.7), transportando a sinalização independente da sinalização de estabelecimento da portadora de voz. Os servidores de rede designados como *Server 1* e *Server 2* possuem a função de controle das chamadas, trocando sinalização com os outras centrais de comutação e enviando solicitações para os elementos de rede designados como *Media Gateways* (MGW1 e MGW2), cuja função é fazer a codificação entre mídias de voz (por exemplo, converter TDM em VoIP) e estabelecer as conexões entre os canais de voz. O estabelecimento da portadora do tráfego de voz pode ser iniciado por qualquer um dos lados do par de comunicação. O BICC estabelece o enlace entre os dois nós e negocia o codec a ser utilizado na comunicação entre o par, através de mensagens específicas.

O BICC herda as mensagens e parâmetros da sinalização da telefonia convencional e pode, portanto, suportar os mesmos serviços. Isto permite também um interfuncionamento natural entre os protocolos, sem perdas de funcionalidades.

Alguns serviços básicos de comunicação são suportados pelo BICC, como voz, fax, dados 64Kbps e identificador de chamadas.

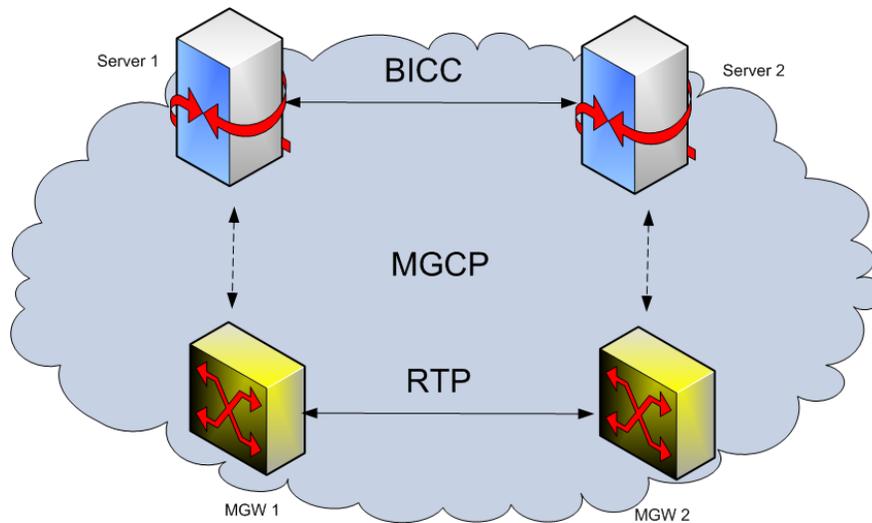


Figura 2.7: BICC e RTP

As principais mensagens utilizadas neste trabalho estão listadas na Tabela 2.1 e explicadas na sequência:

Tabela 2.1: Mensagens BICC

Mensagem	Significado	Descrição da Mensagem
IAM	Initial Address Message	Indica do início da chamada
APM	AP Message	Faz a negociação de Codec e endereço IP
ACM	Acknowledge Message	Confirmação de endereço
ANM	Answer Message	Atendimento.
REL	Release Message	Desconexão
RLC	Release Complete Message	Confirmação de desconexão

- IAM é a mensagem que inicia a troca de sinalização entre dois equipamentos que fazem conexão BICC, enviada pela central de onde se originou a chamada.
- APM é a mensagem que faz a negociação de codec e define os endereços IPs das duas Media Gateway que irão fazer o transporte da voz.
- ACM é a mensagem que indica que a central de destino recebeu o número discado e que pode tratar o mesmo.

- ANM é a mensagem que indica que o assinante de destino atendeu a chamada e irá começar a conversação.
- REL é a mensagem indicando que uma das partes desconectou a chamada.
- RLC é a mensagem que indica que a outra parte entendeu a mensagem de desconexão e irá proceder à liberação do canal.

A sequência básica da troca de mensagens do protocolo BICC entre duas entidades está ilustrada na Figura 2.8.

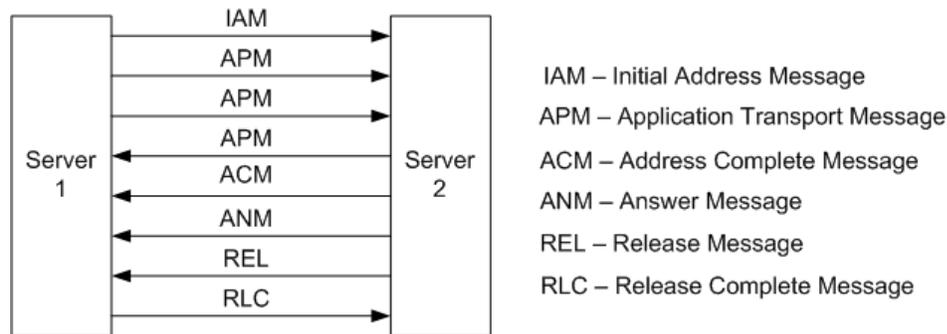


Figura 2.8: Troca de sinalização BICC

## 2.5.2 RTP

O RTP (*Real-Time Transport Protocol*) [2] é o protocolo utilizado que oferece funções para transporte de rede de sinais de áudio ou vídeo digitalizado, ou outro tipo qualquer de dados em tempo real. É um protocolo que não contém mecanismos que garantem reserva de recursos ou qualidade de serviço, mas atribui esses requisitos para serviços como o IntServ e DiffServ. Ao invés disso, o RTP provê duas funcionalidades: número de sequência em cada pacote transmitido e o tempo, o primeiro para detectar pacotes fora de ordem e o segundo para controle de tempo. Como o RTP foi desenvolvido para transportar uma grande variedade de tipos de dados de tempo real, não provê uma interpretação uniforme de semânticas. Ao invés disto, cada pacote começa com um cabeçalho fixo; campos no cabeçalho especificam como interpretar estes campos e como interpretar o conteúdo útil do pacote (*payload*). A Figura 2.9 ilustra o formato do cabeçalho fixo RTP.

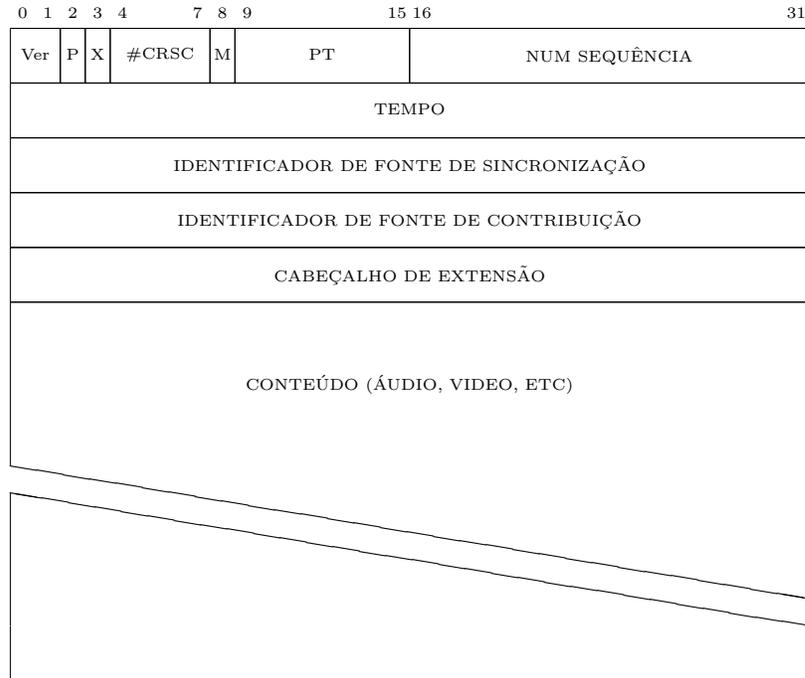


Figura 2.9: Pilha RTP

Como mostra a figura, cada pacote começa com um campo designando a versão do RTP, com 2 bits. O campo do décimo-sexto bit SEQUENCE NUM contém um número de sequência para o pacote. O primeiro número é escolhido aleatoriamente.

Algumas aplicações solicitam um cabeçalho opcional de extensão, inserido entre o cabeçalho fixo e o conteúdo útil. Neste caso, o campo X é utilizado como um indicador para especificar se a extensão está presente ou não no cabeçalho.

A interpretação dos campos remanescentes dependem do valor do campo PT (PTYPE), que especifica o tipo de conteúdo útil do pacote. O campo P especifica se existem "zeros" preenchendo o campo do conteúdo do pacote; este valor é utilizado em casos de encriptação que exigem que os dados sejam alocados em blocos de tamanho fixo.

O campo M (*Marker*) depende da aplicação. É utilizado em aplicações que necessitam indicar pontos no fluxo de dados (por exemplo, o início de cada quadro que está enviando vídeo).

O tipo de conteúdo útil do pacote também afeta a interpretação do campo do tempo (*timestamp*). Este campo é um valor de 32 bits que fornece o tempo que o primeiro octeto digitalizado foi amostrado, com um tempo inicial para uma sessão escolhido ao acaso. A norma especifica que o tempo inserido no cabeçalho é incrementado continuamente, enquanto existir dados para serem enviados.

### 3 Descrição do Modelo Proposto

Varios trabalhos publicados sobre o assunto até o momento tratavam de modelos do trafego de dados/voz a nivel do pacote ou entao de modelo classicos ON-OFF, com estudos analisando a duracao da sessao e o intervalo entre elas como distribuicoes exponenciais.

Este trabalho vem propor uma adaptação do modelo SURGE, caracterizando o fluxo da rajada de pacotes transmitidos durante o período de duração da sessão, tratando de forma mais detalhada os parâmetros de tempo dos pacotes. Dentro do período de duração da sessão, serão estudados o intervalo de tempo entre os pacotes e o período de duração de transmissão do pacote (Figura 3.1).

Com este modelo será possível reproduzir o tráfego VoIP com tratamento à nível de pacotes. Como existem vários codecs com diferentes taxas, fixas ou variáveis, que podem ser utilizados na transmissão dos pacotes, espera-se modelar o tráfego dos pacotes com melhor detalhamento, principalmente com codecs com taxa variável.

Para se fazer a modelagem do tráfego em análise, serão estudados os parâmetros de tempo de duração da sessão, intervalo entre sessões, intervalo entre pacotes e o tempo de envio dos pacotes (Figura 3.1).

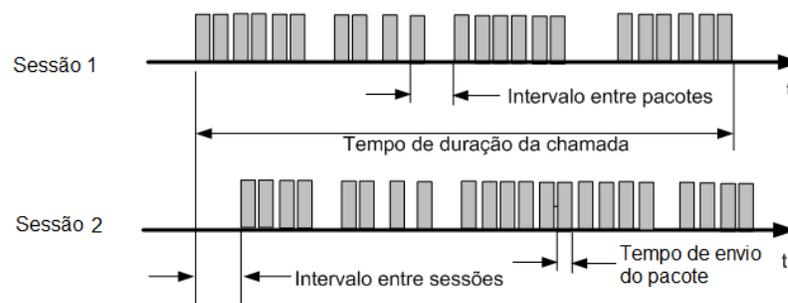


Figura 3.1: Modelo Proposto

- Tempo de duração da sessão

O tempo de duração da sessão é o tempo de duração de uma chamada, e foi obtido

através da diferença de tempo entre as mensagens BICC RLC e IAM de uma mesma chamada, que são, respectivamente, a mensagem final e inicial da sinalização da chamada.

- Intervalo entre sessões

O intervalo entre sessões é o tempo entre duas sessões subseqüentes, e foi obtido fazendo-se a diferença do tempo de chegada entre de duas mensagens IAM subseqüentes.

- Tempo de envio dos pacotes

Para se obter o período de envio dos pacotes, foi analisado o tamanho dos pacotes. Sabendo-se o tamanho dos pacotes, obtém-se o tempo de transmissão do pacote através da razão entre o tamanho do pacote (em bits) e a velocidade do enlace (em bits/seg).

- Intervalo entre pacotes

Para se caracterizar o intervalo entre pacotes deste modelo, analisou-se o tempo entre os pacotes RTP e obteve-se a diferença entre o tempo de chegada de dois pacotes consecutivos .

Após a caracterização de todos os parâmetros que compõem o modelo proposto será possível desenvolver um simulador para se gerar um tráfego agregado que tenha uma boa aderência com o tráfego agregado total observado na rede.

Dentre os principais diferenciais deste modelo, podem ser destacados a modelagem de sessões individuais, ou seja, é analisado o comportamento individual de cada sessão, a partir de poucos parâmetros facilmente estimados, coletados diretamente da rede.

### **3.1 Coleta dos dados**

Para se obter os dados referentes às características da sessão (tempo de duração da sessão e tempo de intervalo entre sessões) foram coletadas informações do protocolo BICC de uma rede celular comercialmente ativa no período descrito na Tabela 3.1 com uma aplicação

proprietária do fornecedor das centrais de comutação. Estas informações podem ser coletadas também com um analisador de protocolos tradicional, como o utilizado na coleta dos pacotes.

A rede é composta por assinantes pré-pago (80% do total de assinantes) e pós-pago (20% do total de assinantes).

Tabela 3.1: Período de coleta dos dados

Data	Horário	Quantidade de chamadas
03/08/2009	08:00 às 16:00	46820
26/10/2009	10:00 às 17:00	52181

Para a coleta dos dados referentes às características dos pacotes foi utilizada uma porta espelhada de um switch que está conectado diretamente na saída VoIP de uma *Media Gateway*. Neste trabalho, para esta coleta, foi utilizado um PC Pentium 4 com a aplicação de uso livre Wireshark [30]. Como o tráfego total da rota é superior à capacidade de processamento do equipamento que faz a coleta, foram capturados apenas os primeiros 48 bytes de cada pacote. Isto não causará impacto neste trabalho, pois os dados de interesse estão no cabeçalho dos pacotes e os dados desprezados tratam-se apenas do conteúdo do pacote de áudio, além de preservar a privacidade dos usuários. Quanto à utilização da rota, sabe-se que o backbone não está congestionado, indicando que não há perda de pacotes. Para se confirmar este fato, será analisado se o número de sequência dos pacotes RTP é contínuo, o que indica que realmente não há perda de pacotes.

Para se coletar as informações referentes à duração da sessão, foram capturadas as mensagens BICC entre os dois servidores, designados como Servidor 1 e Servidor 2 na Figura 3.2. As coletas referentes ao tráfego dos pacotes foram obtidas através da coleta dos pacotes RTP entre as *Media Gateways* MGW1 e MGW2 da mesma figura.

Para se fazer a análise, foi utilizado o software de uso livre R [31], que é um software para ambiente de análise e gráficos estatísticos.

## 3.2 Topologia da rede em estudo

A Figura 3.2 ilustra a topologia da rede onde foram coletados os dados. Nesta figura, a *Media Gateway* (MGW) celular é onde estão conectadas as estações rádio-base e esta encaminha as chamadas para a MGW1, que é uma *Media Gateway* intermediária que faz o trânsito das chamadas para a MGW2, que por sua vez, encaminha as chamadas para as outras operadoras celulares e para a rede fixa. O fluxo de voz chega da MGW celular convertida em VoIP na MGW 1 com o codec AMR e transportado pelo protocolo RTP através de switches e roteadores Gigabit Ethernet. Os dados de RTP foram coletados no trecho entre as MGW1 e MGW2, indicado na figura como VoIP RTP.

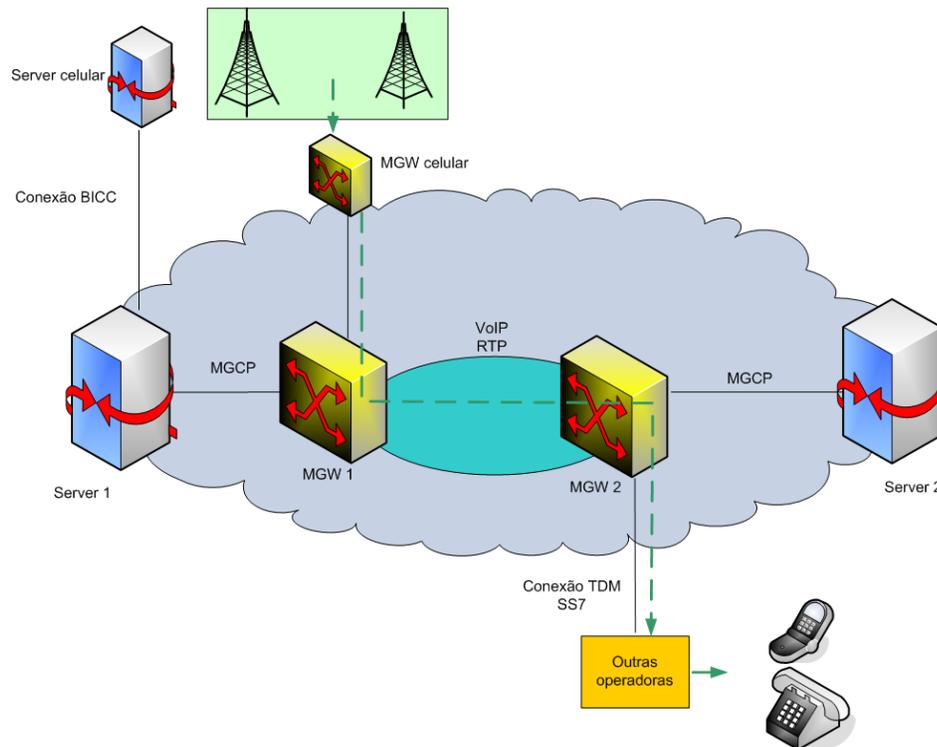


Figura 3.2: Topologia da rede em estudo

## 4 Resultados

Este capítulo irá apresentar os resultados obtidos na modelagem dos parâmetros envolvidos, verificando se os dados coletados na rede estudada podem ser representados segundo alguma distribuição de probabilidade conhecida.

Inicialmente obtém-se, a partir dos dados empíricos, os valores dos parâmetros da distribuição que se supõe que se ajuste. A partir destes valores, obtém-se o gráfico da função de distribuição acumulada dos dados teóricos e faz-se uma comparação com o gráfico dos dados teóricos.

O gráfico QQPlot também será utilizado para se verificar se as duas populações de dados (empíricos e teóricos) possuem a mesma distribuição. O QQPlot é uma ferramenta gráfica que mostra os dados ordenados, comparando com os quantis de uma certa distribuição. Se as duas amostras tiverem características semelhantes, os pontos estarão dispostos em torno da bissetriz dos eixos  $x$  e  $y$  [32].

Ao final será feita uma simulação de várias sessões superpostas com os parâmetros obtidos e verificar se o tráfego agregado simulado tem boa aderência com o tráfego agregado original.

### 4.1 Intervalo entre sessões

Para todo o período de coleta, não foi detectado nenhum tipo de tendência, caracterizando a estacionariedade da série. Isto significa que esta série possui as características estatísticas semelhantes, como a média e o desvio padrão, independente do tempo em que a observação foi feita (Tabela 4.1).

Para se modelar o intervalo entre sessões, foi analisado o gráfico da função de distribuição acumulada (Figura 4.1) que ilustra a aderência entre os dados empíricos a uma

Tabela 4.1: Média e desvio padrão do intervalo entre sessões (em ms)

Chamadas	Média	Desvio padrão	erro (h) para 99% de confiança
0:10000	41,88	39,94	1,03
10000:20000	43,63	41,29	1,06
20000:30000	44,76	41,67	1,07
30000:40000	43,57	41,15	1,06
40000:50000	42,85	41,22	1,06

distribuição exponencial, cuja média vale  $43ms$ . A realização de teste Kolmogorov-Smirnov para todo o intervalo indica um valor aproximado de 0,5 para todo o intervalo, o que mostra que esta distribuição é adequada para os valores empíricos obtidos, já que para se rejeitar a hipótese, este valor deveria estar abaixo de 0,05 [33]. A aderência também pode ser comprovada pelo gráfico QQPlot ilustrado na Figura 4.2, comparando-se os dados empíricos com uma distribuição teórica exponencial. Nesta mesma figura observa-se que as comparações com outras distribuições não mostram aderência adequada. O tempo médio do intervalo entre sessões obtido foi de  $43ms$  e o desvio-padrão  $41ms$ .

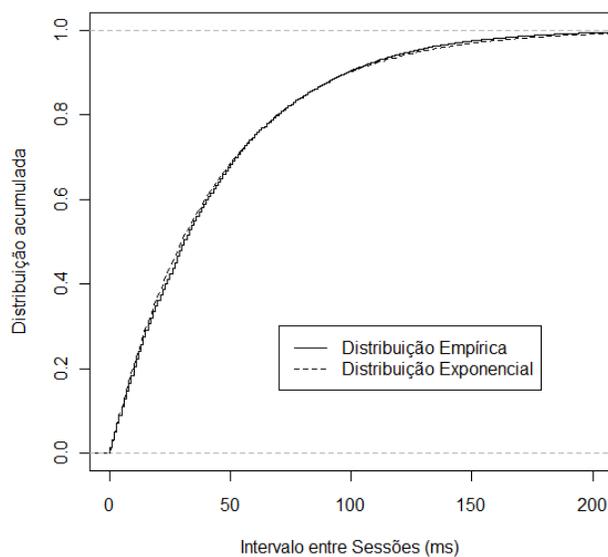


Figura 4.1: Distribuição acumulada do intervalo entre sessões

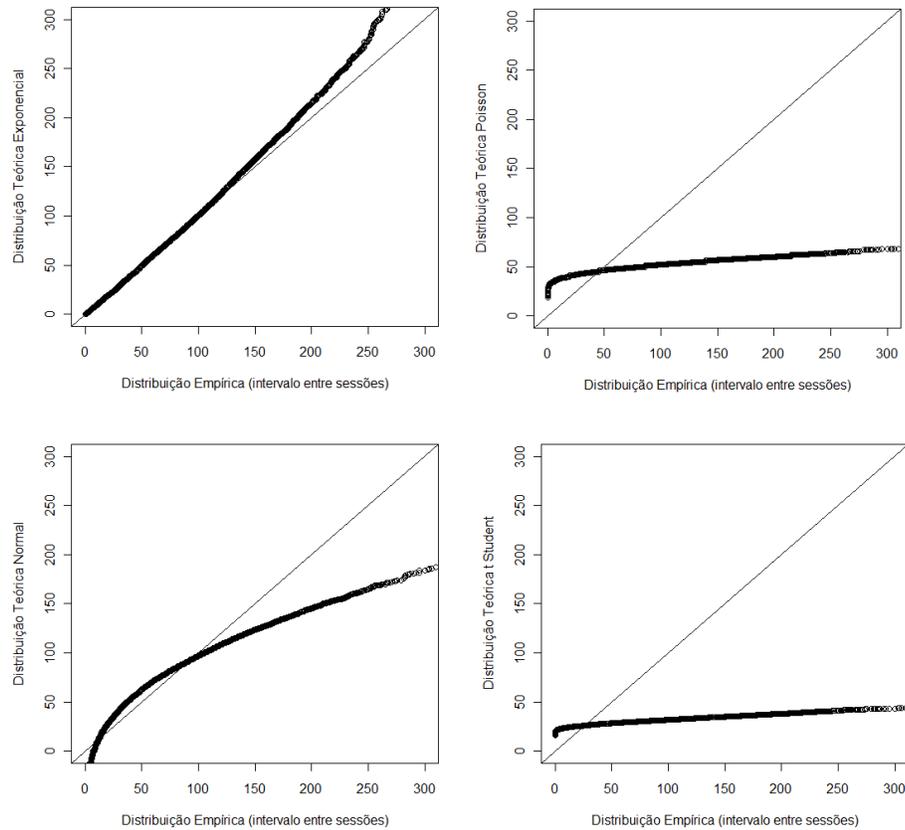


Figura 4.2: Comparação do QQPlot do intervalo entre sessões com outras distribuições

## 4.2 Tempo de duração da sessão

Para o tempo de duração da sessão o estudo mostrou que uma distribuição de cauda pesada foi ajustada nos dados coletados, com uma boa aderência à distribuição de pareto tipo II, com  $\alpha = 2,78$  e  $\beta = 37$ .

Para o tempo de duração da sessão uma distribuição de cauda pesada foi ajustada a partir dos dados coletados usando máxima verosimilhanca (funcao *gpd* no software R). Os dados mostraram boa aderência a distribuição de pareto tipo II, ccom  $\alpha = 2,78$  e  $\beta = 37$ .

Uma distribuição de cauda pesada possui valores extremos, ou seja, alguns de seus valores distanciam-se consideravelmente da média, e a distribuição de pareto possui esta característica de cauda pesada. Esta aderência pode ser observada no gráfico da distribuição complementar na Figura 4.3 e também pode ser observado no gráfico QQPlot na Figura 4.4.

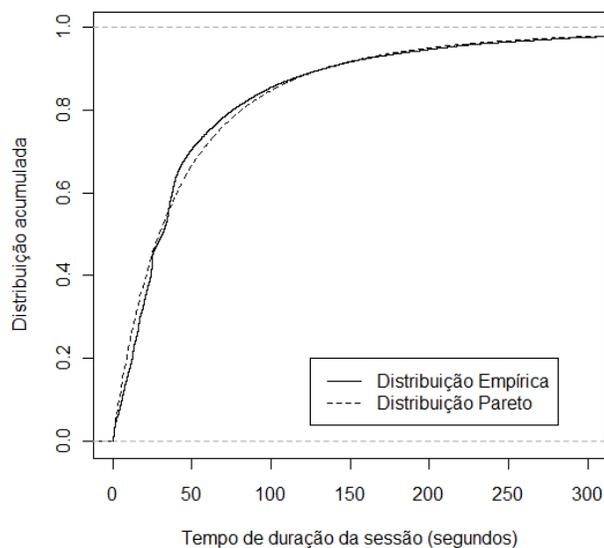


Figura 4.3: Distribuição acumulada do tempo de duração da sessão

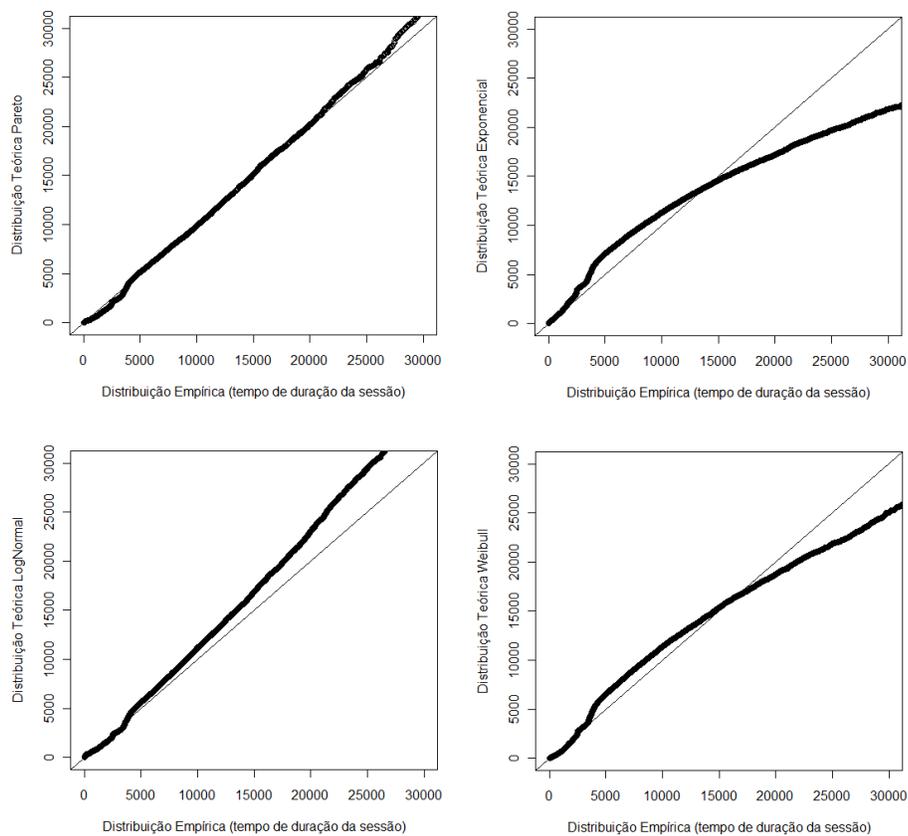


Figura 4.4: Comparação do QQPlot do tempo de duração da sessão com outras distribuições

O fato que pode explicar este ajuste à uma distribuição de cauda pesada é a forma de tarifação utilizada pela operadora. Há planos de minutos ilimitados entre números de mesma operadora, e também planos com franquia de grande quantidade de minutos para conversação. Estes planos fazem com que várias chamadas extrapolem a média do tempo de conversação, caracterizando a cauda pesada.

No gráfico da distribuição complementar, observa-se que a curva obtida com a distribuição teórica de Pareto sobrepõe-se à curva obtida com os dados empíricos (linha contínua). Isto significa que os parâmetros estimados de  $\alpha$  e  $\beta$  são valores que tem boa representação para caracterizar a distribuição dos dados empíricos à uma distribuição de Pareto.

A presença de distribuição de cauda pesada nos dados analisados pode ser avaliada de uma forma simples [22], verificando-se visualmente a linearidade do gráfico resultante da distribuição complementar (CCDF) em escala logarítmica e obtendo-se o coeficiente angular deste trecho linear (equação 4.1).

$$\lim_{x \rightarrow +\infty} \frac{d \log F(x)}{d \log x} \quad (4.1)$$

A Figura 4.5 mostra o gráfico da distribuição complementar em escala logarítmica para o tempo de duração da sessão. Este gráfico mostra que a partir de 200s, o gráfico torna-se uma reta, com um ângulo aproximado de  $-69^\circ$ . O módulo do coeficiente angular desta reta é um estimador do parâmetro de forma da distribuição de Pareto  $\alpha = 2,7$ . O tempo médio de duração da sessão obtido foi de 58,25s e o desvio-padrão 96,31s.

## 4.3 Análise do tráfego de pacotes

### 4.3.1 Tempo de transmissão dos pacotes

O codec AMR utilizado gerou pacotes RTP que se distribuíram em 3 valores discretos ilustrados na Tabela 4.2. Nesta mesma tabela estão ilustrados os tempos que estes

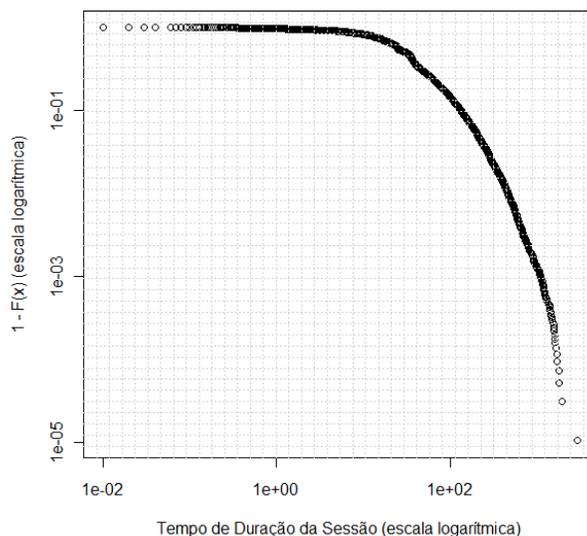


Figura 4.5: Distribuição complementar em escala log do tempo de duração da sessão

pacotes levam para serem transmitidos, sabendo-se que o tempo é a razão entre a quantidade de bytes a serem transmitidos e a velocidade do enlace. Verificou-se que após iniciada a sessão, não houve variação no tempo entre os pacotes. A interpretação deste fato, foi de que embora o CODEC AMR possa operar com taxa variável, a taxa de operação nesta rede era determinada apenas no início da sessão, em função das condições do sinal.

Tabela 4.2: Tempo de transmissão dos pacotes

Tamanho dos pacotes	%	Tempo de Transmissão
89 bytes	75%	0,712 $\mu$ s
98 bytes	22%	0,784 $\mu$ s
63 bytes	3%	0,504 $\mu$ s

### 4.3.2 Intervalo entre pacotes

Para o período de intervalo entre pacotes, coletou-se os cabeçalhos dos pacotes RTP de uma mesma sessão e obteve-se a diferença entre os valores do *time-stamp* dos pacotes para cada sessão. A análise destes dados mostraram que a distribuição de Pareto é a que

melhor se ajusta a estes dados.

Na Figura 4.6 pode-se ver o gráfico da distribuição acumulada do intervalo entre pacotes. Houve um bom ajuste para a distribuição de pareto, com  $\alpha = 3,84$  e  $\beta = 337$ . A comparação do gráfico QQPlot do intervalo entre pacotes (Figura 4.7) com outras distribuições mostra que o ajuste escolhido é adequado.

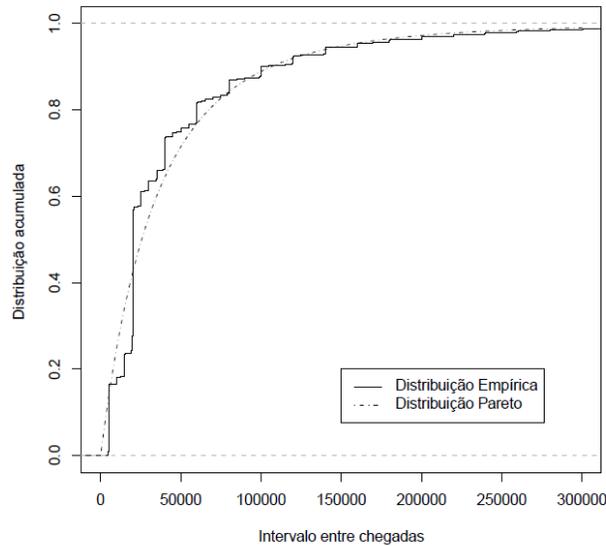


Figura 4.6: Distribuição acumulada do tempo de intervalo entre pacotes

## 4.4 Simulação

Para se verificar a qualidade do modelo proposto, foi feita uma simulação com os parâmetros e distribuições obtidos e os resultados foram comparados com os valores empíricos.

Esta simulação foi feita em uma aplicação desenvolvida em Java, por parceiros de pesquisa [34]. A topologia da simulação (Figura 4.8) é composta por várias fontes UDP geradora de pacotes e dois nós, um enviando os pacotes em direção ao outro. Estas aplicações simulam a geração de um tráfego agregado de pacotes de várias sessões simultâneas, simulando um tráfego com as características de distribuição determinadas nos estudos realizados nos itens anteriores deste capítulo. A aplicação irá gerar um valor aleatório para cada parâmetro do modelo (tempo de duração da sessão, intervalo entre sessões, intervalo entre pacotes e

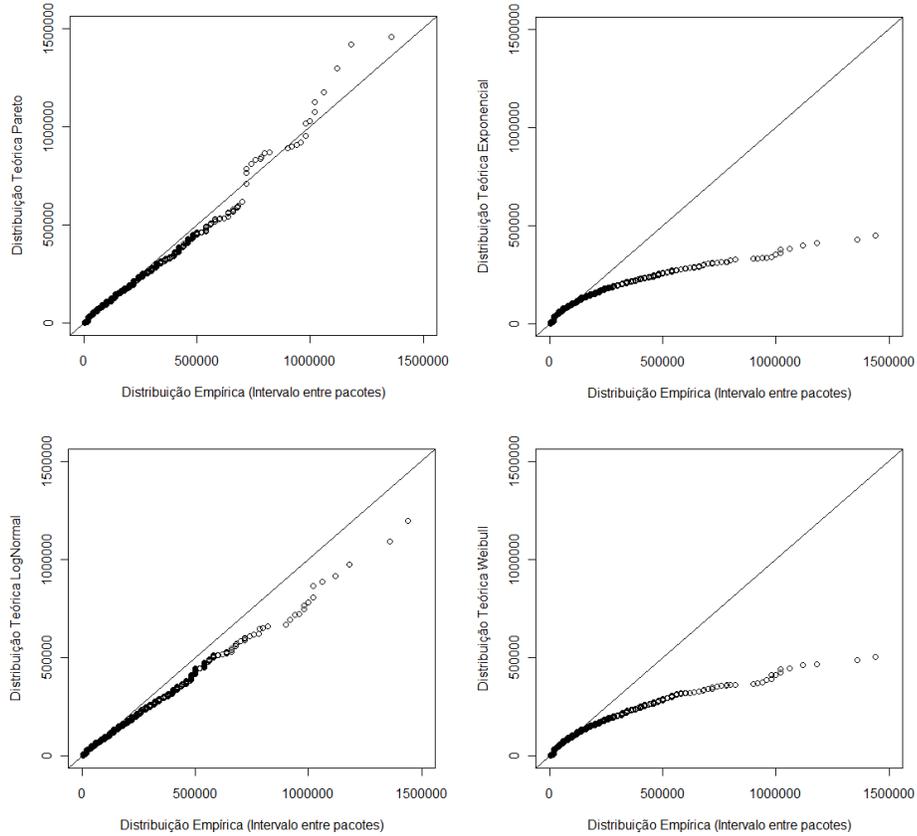


Figura 4.7: Comparação do QQPlot do intervalo entre pacotes com outras distribuições

tamanho do pacote), seguindo a distribuição determinada. Com estes valores determinados, a aplicação irá informar o tempo e o tamanho que cada pacote é gerado, fazendo o mesmo processo para várias sessões simultâneas.

Os parâmetros determinados neste estudo e utilizados na simulação estão listados na Tabela 4.3. Nos casos onde a distribuição utilizada foi a distribuição de Pareto, utilizou-se nas simulações a distribuição de Pareto tipo II.

A primeira comparação foi realizada através dos gráficos da distribuição acumulada do intervalo de tempo entre pacotes do tráfego agregado (Figura 4.9). Nesta figura pode-se verificar que a curva da simulação obteve uma boa aderência aos dados empíricos. Na Figura 4.10 ilustra-se a comparação do gráfico QQPlot dos dados reais com os dados da simulação verificando-se também uma boa aderência.

Além disso, foram obtidos também os gráficos da função de auto-correlação (ACF)

Tabela 4.3: Resultados da modelagem

Parâmetro	Distribuição	Valor
Intervalo entre sessões	Exponencial	Média=43ms
Duração da sessão	Pareto	$\alpha = 2,78; \beta = 37$
Transmissão dos pacotes	Constante para cada sessão	$t=0,712\mu s$
Intervalo entre pacotes	Pareto	$\alpha = 3,84; \beta = 337$

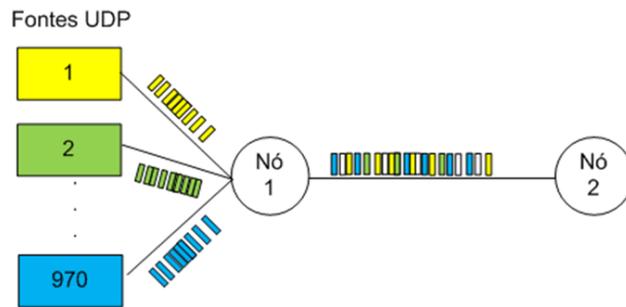


Figura 4.8: Topologia da simulação

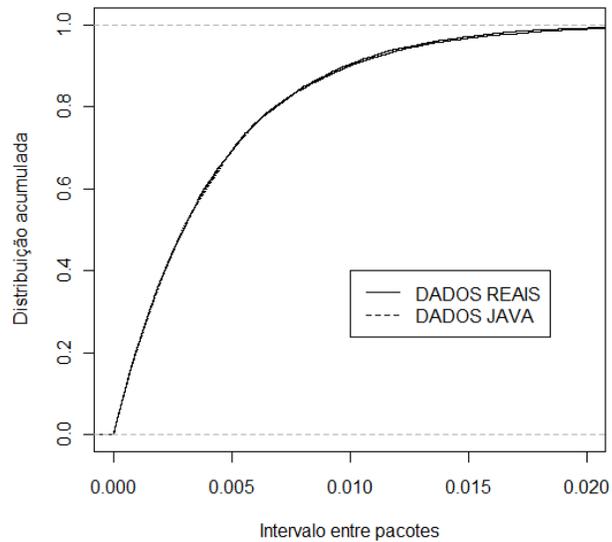


Figura 4.9: Distribuição acumulada do tempo de intervalo entre pacotes do tráfego agregado para os dados da simulação

do tráfego agregado (Figura 4.11) em várias escalas de tempo. Nestes gráficos verifica-se também a similaridade entre os dados reais e os dados gerados pelo simulador, para a mesma

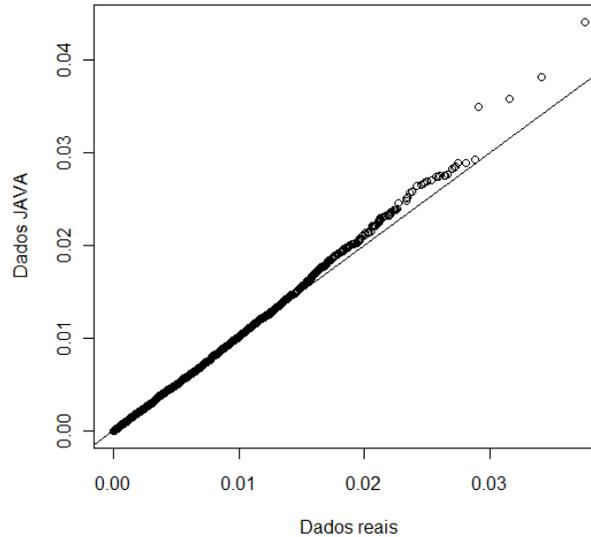


Figura 4.10: Comparação do QQPlot para os dados da simulação do tempo de intervalo entre pacotes do tráfego agregado

Tabela 4.4: Estatísticas da quantidade de bytes enviados dos dados simulados

Escala	Dados	Média	Desvio Padrão	Erro (h) para
				99% de confiança
50ms	Original	1100,254	298,4485	30,6774
	Java	1091,154	310,2141	34,0962
500ms	Original	10254,54	1063,349	335,1646
	Java	10032,85	1004,526	337,4076
1s	Original	20374,41	1580,315	699,2362
	Java	20017,93	1204,99	567,6002

escala de tempo, onde se conclui que os dados simulados foram coerentes com os dados reais. A Tabela 4.4 mostra a comparação da média e desvio padrão do tráfego agregado real e simulado, onde pode-se comprovar que estes valores são compatíveis para os conjuntos de valores na mesma escala.

Os resultados mostram que o tráfego simulado com os parâmetros obtidos através da análise dos dados capturados da rede estudada, possuem características semelhantes aos

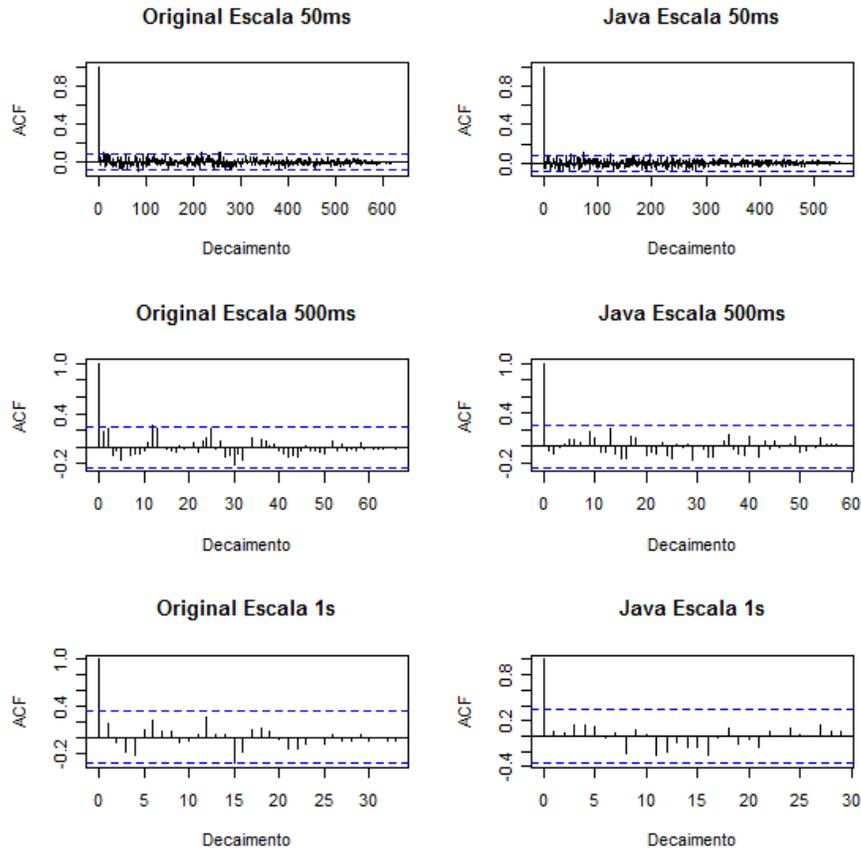


Figura 4.11: Comparação da função de auto-correlação do tráfego agregado da simulação em diferentes escalas

dados empíricos. Pelos resultados obtidos, verifica-se que os dados coletados da rede estudada não mostram lento decaimento, indicando que este tráfego não possui dependência de longa duração.

Para comprovar que este estudo também pode ser aplicado em tráfego com características auto-similares, fez-se simulações com parâmetros que geram tráfegos com dependência de longa duração. Foi simulada a geração de tráfego com a distribuição de Pareto tipo II no simulador, com os parâmetros de duração da chamada com valor fixo de  $\beta$  em 37 e  $\alpha$  variando entre 2 e 6. Feitas as simulações, calculou-se a quantidade de bytes enviados a cada 100ms, para cada simulação, e obteve-se os gráficos da função de auto-correlação (ACF). Verificou-se que este gráfico do tráfego agregado mostra um lento decaimento, indicando que este tráfego tem dependência de longa duração (Figura 4.12). Os gráficos mostram que o

decaimento é mais rápido quanto maior o valor de  $\alpha$ . A Tabela 4.5 mostra o parâmetro de Hurst para as três simulações.

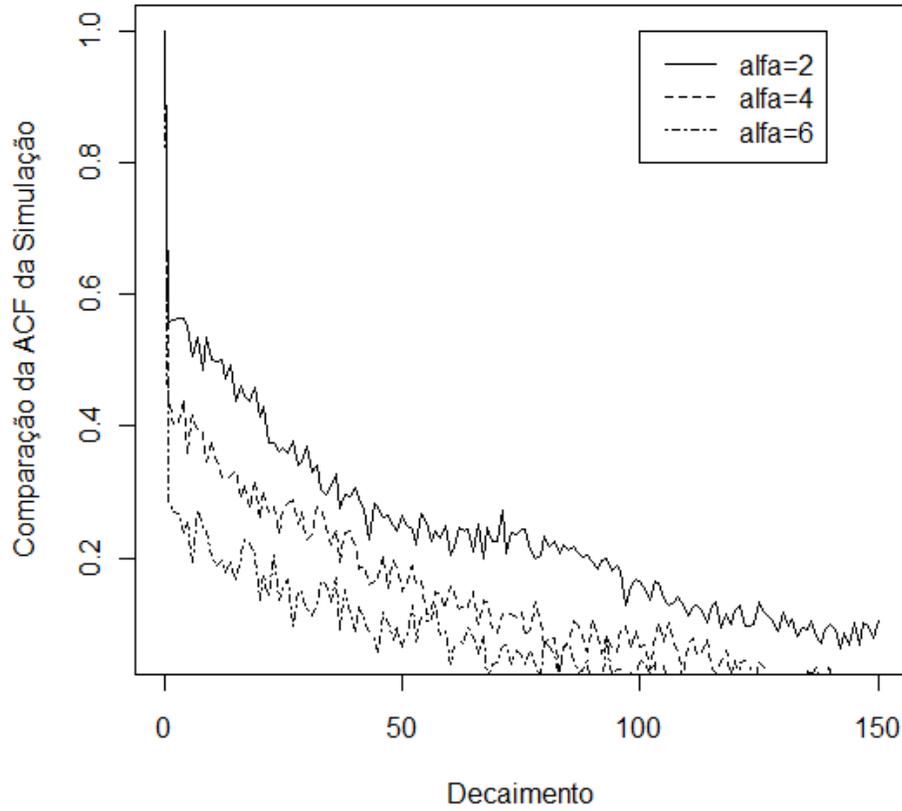


Figura 4.12: Comparação da ACF do tráfego agregado da simulação para diferentes valores de alfa

Tabela 4.5: Parâmetro de Hurst para as simulações

$\alpha$	Parâmetro de Hurst
2	0,7329
4	0,6907
6	0,5760

## 5 Considerações Finais

### 5.1 Análise dos resultados e discussão

Neste trabalho foi apresentado um modelo para o tráfego VoIP com o qual analisam-se dados de uma rede comercialmente ativa. As chamadas são originadas em uma rede celular com destino a rede fixa e outras operadoras celulares. Este tipo de tráfego representa grande parte do tráfego total de uma grande operadora de telecomunicações no Brasil.

Como era esperado, o intervalo de tempo entre sessões foi modelado com uma distribuição exponencial, com média igual a  $43ms$ . O tempo de duração da sessão foi caracterizado por uma distribuição de cauda pesada. No caso estudado, ajustou-se a uma distribuição de Pareto, com os parâmetros  $\alpha = 2,78$  e  $\beta = 37$ .

A cauda pesada, identificada nos tempos de duração da sessão pode ser explicada pelos vários tipos de tarifação utilizada pela operadora. A tarifação utilizada, com ligações sem limite de tempo entre celulares geram chamadas com tempo de duração longo, com a ocorrência frequente de ligações com tempo de duração com valores extremos.

O tempo de transmissão dos pacotes é caracterizado por três tamanhos de pacotes (de 63, 89 e 98 bytes), gerados pelo codec AMR. Estes tamanhos geraram pacotes com tempo de transmissão de, respectivamente,  $0,504\mu s$ ,  $0,712\mu s$  e  $0,784\mu s$ .

O intervalo de tempo entre pacotes também foi ajustado a uma distribuição de cauda pesada, também aderindo a uma distribuição de Pareto tipo II, com  $\alpha = 3,84$  e  $\beta = 337$ .

Para a topologia estudada, o modelo conseguiu representar com uma boa aproximação as características da rede, o que pode ser comprovado com as semelhanças visuais obtidas da função de auto-correlação do tráfego agregado. O fato de não se ter uma função de auto-correlação com um lento decaimento pode ser explicado pelo valor de  $\alpha$  observado

não ser muito baixo e pelas características do codec utilizado (AMR), que, apesar de ser um codec de taxa variável, na topologia estudada não houve variação suficiente para se obter esta característica.

As distribuições de cauda pesada são consideradas como a razão principal para que ocorra a dependência de longa duração no tráfego de redes [22]. Uma análise de qualquer processo temporal possui uma dependência de longa duração se, e somente se, existir uma distribuição de cauda pesada. Por esta razão, este estudo deve ser complementado com uma nova coleta de dados com outros tipos de codecs com maior variabilidade na taxa de compressão.

Mostrou-se ainda que o modelo pode também simular satisfatoriamente um tráfego auto-similar que exhibe dependência de longa duração conforme verificado no longo decaimento da função de auto-correlação (quando  $\alpha = 2$ ).

## 5.2 Conclusão e estudos futuros

As informações levantadas pelo modelo apresentado podem mostrar para a operadora qual é o perfil de tráfego de seus clientes em determinado cenário. O modelo pode ser utilizado pela operadora também para se analisar qual o impacto causado na rede estudada quando é feita alguma alteração nos parâmetros analisados, como por exemplo, qual o impacto na ocupação da rede se a operadora fizer uma promoção nas tarifas. Neste caso, o resultado desta promoção pode ser um aumento no número de chamadas e na duração das mesmas. Com esta informação e a utilização deste modelo é possível se fazer uma simulação com os novos parâmetros e observar qual o impacto no tráfego da rede. De posse destes dados, faz-se os ajustes adequados na rede, de forma a se manter a qualidade do serviço e a rentabilidade.

Como estudos futuros, sugere-se a aplicação deste modelo em aplicações multi-mídia, principalmente utilizando codecs com taxa variável. Como outros serviços, pode-se sugerir aplicações em TV digital.

## Referências Bibliográficas

- [1] M. CROVELLA and P. BARFORD, “Generating representative web workloads for network and server performance evaluation,” *Proceedings of the 1998 ACM SIGMETRICS International Conference*, pp. pag. 151–160, 1998.
- [2] H. SCHULZRINNE, S. CASNER, R. FREDERICK, and V. JACOBSON, “Rtp: A transport protocol for real-time applications,” *RFC3550*, 2003.
- [3] R. G. P. PROJECT, “Technical specification group services and system aspects; mandatory speech codec speech processing functions; adaptive multi-rate (amr) speech codec; transcoding functions; (release 8),” *3GPP*, 1999.
- [4] ETSI, “Adaptive multi-rate (amr); speech processing functions,” *European Standard*, 1999.
- [5] A. ADAS, “Traffic models in broadband networks,” *IEEE Communications Magazine*, vol. Vol. 35, pp. pag. 82–89, 1997.
- [6] A. ESTEPA, R. ESTEPA, and J. VOZMEDIANO, “A new approach for voip traffic characterization,” *IEEE COMMUNICATIONS LETTERS*, vol. VOL. 8, pp. pp. 644 – 646, 2004.
- [7] L. MUSCARIELLO, M. MELLIA, M. MEO, and M. AJMONE-MARSAN, “An mmp-based hierarchical model of internet traffic,” *IEEE International Conference on Communications*, vol. Vol. 4, pp. pag. 2143 – 2147, 2004.
- [8] S. KIM, M. LEE, and M. KIM, “Sigma-matching technique for mmp modeling of heterogeneous on-off sources,” *Global Telecommunications Conference, 1994. GLOBECOM '94. 'Communications: The Global Bridge'.*, *IEEE*, vol. Vol. 2, pp. pag. 1090–1094, 1994.

- [9] C. M. PEDROSO, J. CALDEIRA, K. FONSECA, and M. CRUZ, “Análise da evolução de características de tráfego voip,” *XXVI SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES - SBRT08*, 2008.
- [10] C. M. PEDROSO, E. P. RIBEIRO, and E. M. MUNETIKO, “Modelagem de sistemas voip: Análise de características da sessão,” *XXVII SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES SBRT09*, Outubro 2009.
- [11] W. CHEN, H. HUNG, and Y. LIN, “Modeling voip call holding times for telecommunications,” *IEEE Network*, November/December 2007.
- [12] J. WENYU and H. SCHULZRINNE, “Analysis of on-off patterns in voip and their effect on voice traffic aggregation,” *IEEE*, 2000.
- [13] M. MENTH, A. BINZENHÖFER, and S. MÜHLECK, “Source models for speech traffic revisited,” *IEEE/ACM TRANSACTIONS ON NETWORKING*, vol. VOL. 17, pp. pag. 1042–1051, 2009.
- [14] W. WILLINGER, M. TAQQU, R. SHERMAN, and V. DANIEL, “Self-similarity through high-variability : Statistical analysis of ethernet lan traffic at the source level,” *IEEE/ACM Trans. on Networking*, vol. 5, 1997.
- [15] P. BRADY, “A technique for investigating on-off patterns of speech,” *Bell System Technical Journal*, vol. Vol. 1, p. pag. 44, 22, January 1965.
- [16] V. PARETO, *Cours d ‘ Economique Politique*. Macmillan, 1897.
- [17] M. E. CROVELLA and A. BESTRAVOS, “Self-similarity in world wide web traffic: Evidence and possible causes,” *IEEE/ACM TRANSACTIONS ON NETWORKING*, vol. Vol. 5, pp. pag. 835–846, 1997.
- [18] H. NETO, J. ALMEIDA, L. ROCHA, W. MEIRA, P. GUERRA, and V. ALMEIDA, “A characterization of broadband user behavior and their e-business activities,” *Latin American Web Conference*, pp. 121 – 128, 2008.

- [19] S. A. KLUGMAN, H. H. PANJER, and G. E. WILLMOT, *Loss Models, From Data to Decisions*. Wiley, New York, 2004.
- [20] B. B. MANDELBROT, “The fractional geometry of nature,” 1982.
- [21] W. E. LELAND, M. S. TAQQU, W. WILLINGER, and D. V. WILSON, “On the self-similar nature of ethernet traffic (extended version),” *IEEE/ACM Trans. On Networking*, vol. Vol. 2, pp. pag. 1–15, 1994.
- [22] W. WILLINGER and K. PARK, *Self-similar network traffic and performance evaluation*. John Wiley & Sons, 2000.
- [23] H. E. HURST, “Long-term storage capacity of reservoirs,” *Transactions of the American Society of Civil Engineers*, pp. pages 770–808, 1951.
- [24] M. S. TAQQU and V. TEVEROVSKY, “Robustness of whittle type estimators for time series with long-range dependence,” *Stochastic Models*, vol. 13, pp. 723–757, 1997.
- [25] M. S. TAQQU, V. TEVEROVSKY, and W. WILLINGER, “Estimators for long-range dependence: an empirical study,” *Fractals*, vol. 3, pp. 785–788, 1995.
- [26] H. SCHULZRINNE, G. CAMARILLO, A. JOHNSTON, J. PETERSON, R. SPARKS, M. HANDLEY, and E. SCHOOLER, “Sip: Session initiation protocol,” *RCF: 3261*, 2002.
- [27] ITU, *H.323 Packet-based multimedia communications systems*. ITU, 2006.
- [28] ITU, *Bearer Independent Call Control CS1*. ITU Q.1901, 1990.
- [29] ITU, *Bearer Independent Call Control CS2*. ITU Q.1902, 1991.
- [30] A. OREBAUGH, G. RAMIREZ, J. BURKE, and L. PESCE, *Wireshark & Ethereal Network Protocol Analyzer Toolkit (Jay Beales Open Source Security)*. Syngress Publishing, 2006.
- [31] R. D. C. Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, <http://www.R-project.org>, 2005. ISBN 3-900051-07-0.

- [32] M. N. MAGALHÃES and A. C. P. LIMA, *Noções de Probabilidade e Estatística*. IME-USP, 2010.
- [33] R. JAIN, *The art of computer systems performance analysis: techniques for experimental design, measurement, simulation and modeling*. John Wiley & Sons, 1991.
- [34] C. M. PEDROSO, C. MATTOS, and W. BESSA, “Software simulador de modelo voip.”