

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Thiago Boubée Cardozo

Evolução da conectividade fim-a-fim da Internet

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Ciências Exatas da Universidade Federal de Juiz de Fora como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Orientador: Ana Paula Couto da Silva

Coorientador: Alex Borges Vieira

Coorientador: Artur Ziviani

Juiz de Fora

2013

Ficha catalográfica elaborada através do Programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Cardozo, Thiago Boubée.

Evolução da conectividade fim-a-fim da Internet / Thiago Boubée Cardozo. -- 2013.
58 p.

Orientadora: Ana Paula Couto da Silva

Coorientador: Alex Borges Vieira

Coorientador: Artur Ziviani

Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Programa de Pós-Graduação em Ciência da Computação, 2013.

1. Redes de computadores. 2. Metrologia de redes. 3. Bufferbloat. I. da Silva, Ana Paula Couto, orient. II. Vieira, Alex Borges, coorient. III. Ziviani, Artur, coorient. IV. Título.

Thiago Boubée Cardozo

Evolução da conectividade fim-a-fim da Internet

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Ciências Exatas da Universidade Federal de Juiz de Fora como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Aprovada em 27 de Maio de 2013.

BANCA EXAMINADORA

Profa. D.Sc. Ana Paula Couto da Silva - Orientador
Universidade Federal de Juiz de Fora

Prof. D.Sc. Alex Borges Vieira
Universidade Federal de Juiz de Fora

Prof. D.Sc. Artur Ziviani
Laboratório Nacional de Computação Científica

Prof. D.Sc. Marcelo Ferreira Moreno
Universidade Federal de Juiz de Fora

Prof. D.Sc. Antonio Augusto de Aragão Rocha
Universidade Federal Fluminense

*“We are always slow in admitting
any great change of which we do
not see the intermediate steps.”*

- Charles Darwin

RESUMO

A Internet é um sistema em contínua evolução. Nesta dissertação, é caracterizada e analisada a recente evolução da conectividade fim-a-fim da Internet, comparando métricas importantes ao desempenho fim-a-fim de dois períodos distintos separados por 5 anos. Os resultados mostram que a distribuição do tamanho médio dos caminhos se mantém pouco alterada de 2006 para 2011, mas a distribuição do atraso acabou piorando, com um aumento de 45% no atraso dos caminhos de 2006 para 2011. Isso afeta diretamente o desempenho e degrada a experiência do usuário. Além disso, é mostrado que a diversidade de caminhos diminuiu e, com isso, os caminhos distintos ficaram um pouco mais similares. Esse resultado tem um impacto direto nos algoritmos de roteamento que tentam explorar a diversidade de caminhos para uma maior tolerância a falhas. Depois foi estudado mais a fundo o fenômeno conhecido como *bufferbloat*, que é uma das possíveis causas do aparente aumento da latência. Os resultados mostraram que filas muito grandes podem impactar fortemente no desempenho da rede, mas também indicaram que provavelmente o problema não irá ocorrer em máquinas Unix Like com as configurações padrão.

Palavras-chave: Tamanho dos caminhos. Evolução do atraso. Diversidade de caminhos. Geodispersão. Bufferbloat.

ABSTRACT

The Internet is a system under continuous evolution. In this dissertation, it is characterized and analyzed the recent end-to-end connectivity evolution of the Internet, comparing key end-to-end performance metrics from two distinct periods separated by five years. The findings show that the average path length distribution remains virtually unchanged from 2006 to 2011, but the delay distribution actually became worse, with a 45% increase in path delay from 2006 to 2011. This directly affects network performance and degrades user experience. Furthermore, it is shown that path diversity decreased, and accordingly, distinct paths became slightly more similar. This result has a direct impact on routing algorithms that try to explore path diversity to become more fault-tolerant. Later it was studied more deeply the phenomenon known as bufferbloat, which is one of the possible causes of the apparent increase of latency. The results show that large buffers can strongly degrade the network performance, but they also show that the problem is not likely to occur in Unix Like machines with default settings.

Keywords: Delay evolution. Sizes of the paths. Path diversity. Geodispersion. Bufferbloat.

LISTA DE FIGURAS

2.1	CDF - número de saltos.	20
2.2	CDF - Número de saltos dos K-menores caminhos.	20
2.3	CDF - Comprimento dos K-menores caminhos.	21
2.4	CDF - Atraso dos caminhos.	22
2.5	CDF - Diferença dos K-menores caminhos.	23
2.6	CDF - Diferença dos K-menores caminhos.	24
2.7	Histograma - Dispersão geográfica dos K-menores caminhos.	25
2.8	CDF - Dispersão geográfica dos K-menores caminhos (K variante).	25
2.9	Atraso e comprimento dos K-menores caminhos.	26
2.10	Dispersão do atraso e geográfica dos K-menores caminhos.	27
2.11	Atraso e o atraso mínimo dos K-menores caminhos.	28
3.1	O efeito de filas grandes em gargalos [Gettys e Nichols (2012)].	30
3.2	Bufferbloat testbed.	35
3.3	Buffers na pilha de rede no sentido de transmissão. ¹	36
3.4	Destaque do <i>qdisc</i> na Figura 3.3.	40
3.5	<i>qdisc</i> com tamanho 200.	40
3.6	<i>qdisc</i> com tamanho 1.000.	41
3.7	<i>qdisc</i> com tamanho 100.000.000.	41
3.8	Trens de pacotes UDP com <i>qdisc</i> em 200.	42
3.9	Trens de pacotes UDP com <i>qdisc</i> em 1.000.	43
3.10	Trens de pacotes UDP com <i>qdisc</i> em 100.000.000.	43
3.11	Destaque do <i>ringbuffer</i> na Figura 3.3.	44
3.12	Ringbuffer com tamanho 500.	44
3.13	Ringbuffer com tamanho 1.000.	45
3.14	Ringbuffer com tamanho 5.000.	45
3.15	Ringbuffer com tamanho 15.000.	46
3.16	Ringbuffer com tamanho 100.	47
3.17	Ringbuffer com tamanho 350.	47
3.18	Ringbuffer com tamanho 1.700.	48

3.19 Ringbuffer com tamanho 3.850.	48
3.20 Ringbuffer com tamanho 13.000.	49

LISTA DE SÍMBOLOS

AS - Autonomous System

AQM - Active Queue Management

BGP - Border Gateway Protocol

CAIDA - The Cooperative Association for Internet Data Analysis

CDF - Cumulative Distribution Function (Função de distribuição acumulada)

ICMP - Internet Control Message Protocol

NTP - Network Time Protocol

OSPF - Open Shortest Path First

RTT - Round Trip Time (Tempo de ida e volta)

TCP - Transmission Control Protocol

UDP - User Datagram Protocol

SUMÁRIO

1	INTRODUÇÃO	11
1.1	CONTRIBUIÇÕES DO TRABALHO	12
1.2	ORGANIZAÇÃO DA DISSERTAÇÃO	12
2	CARACTERIZAÇÃO	13
2.1	TRABALHOS RELACIONADOS	13
2.2	METODOLOGIA	15
2.2.1	Conjuntos de dados.....	15
2.2.2	Metodologia de amostragem	16
2.2.3	Métricas.....	18
2.3	RESULTADOS	19
2.3.1	Tamanho dos caminhos.....	20
2.3.2	Atraso dos caminhos	22
2.3.3	Diferença entre caminhos.....	23
2.3.4	Dispersão geográfica.....	24
2.3.5	Correlação entre as métricas.....	25
2.4	CONCLUSÃO DO CAPÍTULO	28
3	BUFFERBLOAT	30
3.1	TRABALHOS RELACIONADOS	31
3.2	METODOLOGIA	33
3.2.1	Ambiente de experimentação	34
3.2.2	Filas analisadas	37
3.2.3	Métricas de interesse	38
3.2.4	Cenários.....	38
3.3	RESULTADOS	39
3.3.1	Experimentos com o qdisc.....	39
3.3.2	Experimentos com o ringbuffer	43
3.3.3	Experimentos com o qdisc e ringbuffer	45
3.4	CONCLUSÃO DO CAPÍTULO	49

4	CONCLUSÃO	51
4.1	TRABALHOS FUTUROS	52
	REFERÊNCIAS	54

1 INTRODUÇÃO

A Internet atualmente tem um papel crucial na infraestrutura de comunicação e colaboração internacional. Tal sistema está em contínua evolução ao longo dos anos, com muitos nós e enlaces sendo criados e suprimidos. Além disso, o desempenho fim-a-fim da rede na Internet é crítico para as comunicações e serviços *on-line* (LEE et al., 2010a). Do entretenimento caseiro a sistemas de negócios, as métricas da Internet, tais como atraso e contagem de saltos, impactam a satisfação dos usuários e o lucro das empresas. O objetivo deste trabalho é entender o comportamento da Internet nos últimos anos e analisar tendências, em particular como um conjunto de métricas relevantes mudaram nos últimos anos, para o melhor desenvolvimento de modelos de rede e sistemas.

Na primeira parte deste trabalho, é estudada a evolução da conectividade fim-a-fim da Internet ao longo de anos recentes. Isso permite uma caracterização com granularidade mais fina, em contraste com a maioria dos estudos anteriores que focam em termos de Sistemas Autônomos (DHAMDHERE; DOVROLIS, 2011; SIGANOS et al., 2002). Métricas-chaves de dois períodos distintos separados por cinco anos são comparadas entre si. São considerados dados de 2006 e 2011, incluindo a latência dos caminhos menores fim-a-fim, o número de saltos desses caminhos e a dispersão geográfica desses caminhos. Além disso, é utilizado um método baseado na distância de Levenshtein (LEVENSHTEIN, 1966) (também conhecida com distância de edição (WAGNER; LOWRANCE, 1975)) para caracterizar a diversidade de caminhos e medir o quão diferentes são os caminhos nos dois períodos.

Na segunda parte deste trabalho, é estudada mais a fundo uma das possíveis causas do aumento da latência percebido no período estudado na primeira parte do trabalho, que é o fenômeno conhecido como *bufferbloat* (GETTYS; NICHOLS, 2012). Esse fenômeno está associado ao armazenamento excessivo de pacotes nas filas dos roteadores, o que pode causar um grande aumento na latência e reduzir a vazão. A latência é possivelmente a métrica de maior interesse do usuário final, uma vez que a sua deterioração impacta diretamente na experiência do uso de aplicativos em rede cotidianos. Páginas *web* com elevado tempo de carregamento, vídeos de sites como Youtube ¹ ficam sendo continuamente

¹<http://www.youtube.com/>

interrompidos, e uma imensa redução no tempo de reação em jogos *on-line* são algumas das situações percebidas pelos usuários em geral. Nesta análise do *bufferbloat*, foram analisadas como diversos tamanhos de fila impactam a latência da rede. Foi percebido que filas muito grandes podem deteriorar significativamente o desempenho da rede.

1.1 CONTRIBUIÇÕES DO TRABALHO

As principais contribuições do trabalho são:

- Caracterização da evolução da conectividade fim-a-fim da Internet em um intervalo de 5 anos.
- Melhor compreensão dos sintomas causados pelo fenômeno *bufferbloat*.

1.2 ORGANIZAÇÃO DA DISSERTAÇÃO

O restante do documento está organizado como segue. O Capítulo 2 apresenta toda a parte de análise da evolução da conectividade fim-a-fim da Internet, incluindo como se dá a nossa representação da Internet e todas as métricas estudadas. No capítulo 3 é realizada a primeira investigação mais aprofundada para melhor entender o grande aumento de latência apresentado no Capítulo 2. Finalmente, o Capítulo 4 conclui o trabalho e discute trabalhos futuros.

2 CARACTERIZAÇÃO

Neste capítulo é apresentada a parte inicial do trabalho que consiste no estudo da evolução da conectividade fim-a-fim da Internet em um intervalo de 5 anos entre 2006 e 2011. Essa é uma etapa fundamental, pois a análise permite detectar anomalias e comportamentos não óbvios que estimulam estudos que buscam obter melhor compreensão da evolução da rede e propor soluções para os problemas encontrados.

Há vários estudos sobre a modelagem e medição da topologia da Internet (HADDADI et al., 2008; OLIVEIRA et al., 2007; ZHOU, 2006). Entretanto, a maioria deles se baseia em sondas enviadas por um período limitado de tempo para visualizar a topologia da Internet a partir de uma perspectiva de Sistemas Autônomos (AS), ignorando relações de conectividade fim-a-fim. Sobretudo, a maioria dos trabalhos anteriores que alegam caracterizar e modelar a dinâmica da Internet focam na evolução de propriedades topológicas, tais como distribuição de grau, coeficiente de clusterização ou diâmetro do grafo, ignorando métricas centradas no usuário, tais como latência fim-a-fim ou tamanho do caminho. Essas são métricas importantes consideradas neste estudo dado que impactam a satisfação do usuário e possivelmente os custos de tráfego inter-domínio.

Este estudo revelou algumas características contra-intuitivas, como o fato de que, apesar da rede estar se tornando mais densa, está ocorrendo um aumento considerável na latência. Outra informação muito importante é o fato de que a diversidade de caminhos está diminuindo.

Na Seção 2.1 são apresentados diversos esforços para o entendimento da estrutura e evolução da Internet. Na Seção 2.2 é descrito como os dados foram obtidos, processados e quais as métricas calculadas. Os resultados e os seus significados são discutidos na Seção 2.3. O capítulo é fechado na Seção 2.4, onde os resultados mais interessantes são revisados.

2.1 TRABALHOS RELACIONADOS

A caracterização da evolução e o comportamento da conectividade da Internet atraiu diversos esforços de pesquisa na última década. A maioria desses trabalhos foca em como a Internet mudou na conectividade em nível de AS, focando na caracterização de

propriedades topológicas, sem focar em medidas fim-a-fim.

Siganos et al. (2002) mostraram que o crescimento de ambos o número de nós e arestas no nível de AS foi exponencial entre 1997 e 2001. Esse resultado, juntamente com a observação de que a Internet segue uma estrutura “small world”, indica a densificação da rede. Características importantes encontradas foram que o grau de um nó cresce em proporção ao grau atual, ou seja quanto maior o grau maior será o crescimento, e que novas conexões preferem nós com alto grau de conectividade. Esses resultados vão ao encontro com o resultado encontrado na Seção 2.3.1, onde mostramos que apesar do crescimento da rede, os caminhos na Internet possuem comprimento estável.

Mais recentemente, Dhamdhere e Dovrolis (2011), quase uma década após o trabalho de Siganos et al. (2002), estudaram como a evolução do ecossistema da Internet ocorreu de 1998 até 2010. Foi mostrado que o crescimento da rede no nível de AS agora é linear, e que novos *links* estão mais associados a ASes existentes do que a novos. Outro resultado é que atualmente o tamanho médio dos caminhos se mantém praticamente constante. Há concordância com nosso resultado na Seção 2.3.1, onde observamos uma diferença de 2 saltos no tamanho médio dos caminhos de 2006 para 2011. Também notaram que a economia e interesses particulares em AS possuem forte influência na dinâmica da topologia. Um aspecto chave estudado em nossa análise, porém pouco explorado nesses trabalhos, é como essas mudanças topológicas afetam a diversidade de caminhos.

Lee et al. (2010a) apresentam resultados que mostram aparente aumento na latência da rede de 2004 para 2009. Enquanto nós obtemos o atraso diretamente dos *logs* (Seção 2.2.1) e a partir de caminhos reconstruídos a partir de um grafo (Seção 2.2.3), Lee et al. (2010a) utilizam um algoritmo chamado “path stitching” (LEE et al., 2010b). Esse algoritmo reconstrói caminhos entre ASes utilizando dados de tabelas BGP. O RTT é então calculado a partir desses caminhos. O atraso médio encontrado para 2004 foi de 166ms e 213ms para 2009. Esse atraso foi atribuído ao crescimento do número de ASes ao redor do mundo, muitos dos quais não existiam em 2004, o que causou maior distribuição geográfica das amostras.

Em (SCHWARTZ et al., 2010), foi estudada a simetria, estabilidade e diversidade das rotas da Internet em 2006 e depois em 2009. O estudo utiliza dados coletados em um período de 96 horas pela projeto DIMES (SHAVITT; SHIR, 2005), que consiste em uma plataforma de medição distribuída baseada em computação voluntária. Para a análise

foram selecionados cerca de 100 pares de nós para 2006 e 2009. De acordo com este trabalho, 25% dos pares de nós em ambos os anos possuem exatamente uma rota dominante, e cerca de 30% dos pares utilizaram mais de 10 rotas diferentes. Os resultados mostraram que as rotas fim-a-fim da Internet são mais diversas em 2009 do que em 2006, mas que os caminhos são relativamente estáveis em ambos os períodos. Foi encontrado que o tipo de rede age como definidor do comportamento, onde redes acadêmicas são mais estáveis que as comerciais. A explicação dada para esse comportamento foi o aumento de balanceadores de carga junto com o aumento da topologia da rede, além do uso de *tunneling* que resulta em rotas mais estáveis. Nossa análise contradiz esse aumento de diversidade, indicando justamente o contrário. Isso pode ser pelo fato de utilizarmos uma outra metodologia que não envolve a observação das diferentes rotas encontradas nos *logs*, mas na inferência de possíveis rotas não visíveis em traçados de rota tradicionais (Seção 2.2.2). Além do fato de utilizarmos de outra fonte de dados (CAIDA), que não permite a observação dos mesmos pares de nós ao longo do tempo, e pelo fato de utilizarmos uma amostra de um intervalo de tempo muito maior (1 ano inteiro).

2.2 METODOLOGIA

Nesta seção são descritas as diversas métricas obtidas e a metodologia da obtenção dos dados para a análise.

2.2.1 CONJUNTOS DE DADOS

Este trabalho conta com um conjunto de dados coletados pelo projeto CAIDA¹, que é uma organização empenhada em realizar medições na Internet. Para tal, foram espalhados em diversas instituições ao redor do mundo diversos coletores de dados chamados de monitores. Esses monitores diariamente fazem rastros de rota para milhares de endereços IP desde 1998. Os dados coletados são disponibilizados para o público. Esses dados registram resultados como os da ferramenta *traceroute*² de onde é possível retirar dados sobre caminhos fim-a-fim e o atraso da Internet. São usados dados de um conjunto de monitores da América do Norte e da Europa dado que a maioria dos monitores de 2006 se encontravam localizados nesses continentes nos períodos de realização dos experimentos.

¹<http://www.caida.org/>

²Ferramenta para o mapeamento do caminho entre duas máquinas

Os conjuntos de dados da CAIDA obviamente não fornecem o mapa completo da Internet. Apesar disso, esses dados ainda assim fornecem um ponto apropriado para basear a investigação realizada nesta dissertação, com uma representativa distribuição do atraso na Internet. Estes estão entre as maiores coleções disponíveis publicamente, sendo atualizados continuamente.

Foram recuperados todos os rastros da rede (*network traces*) do projeto CAIDA para todos os meses de 2006 (projeto Skitter) e de 2011 (projeto Ark). Em tais rastros, há 19 e 39 monitores rotulados como fonte em 2006 e em 2011, respectivamente. Desses, 13 e 18 fontes estão na América do Norte e 6 e 21 na Europa em 2006 e 2011, respectivamente. Para que o aumento no número de monitores em 2011 não impacte nos resultados, foram utilizados apenas os monitores localizados na Europa e América do Norte, pois são as regiões com mais monitores em 2006. Cada ano contém um número expressivo de rastros de rota. Foram obtidos cerca de 800 milhões de rastros de 2006 e mais de 1 bilhão de rastros de 2011.

O Skitter é o projeto original da CAIDA para o mapeamento da Internet que começou em 1998. O Ark é uma evolução do Skitter que entrou em operação em 2007, onde as principais diferenças são o suporte ao IPv6 e a utilização de uma lista aleatória de destinos, ao contrário do Skitter que seguia sequencialmente a lista.

2.2.2 METODOLOGIA DE AMOSTRAGEM

Para a análise dos dados, primeiramente foi realizado um processamento para a remoção de caminhos inconsistentes. Neste trabalho, define-se caminhos inconsistentes como aqueles que estão incompletos ou com mais de trinta saltos e *loops*, o que indica roteadores configurados fora do padrão ou erros de roteamento.

Os conjuntos de dados são usados de duas formas. Primeiro, é processado um subconjunto dos dados disponíveis pela CAIDA no formato padrão de *traceroute* para calcular métricas que podem ser inferidas diretamente das sondas. Sondagens são pacotes enviados na rede com o propósito específico da obtenção de informações da mesma. Com as sondagens são obtidos, por exemplo, o número de saltos e o atraso de cada caminho. Tais métricas são discutidas mais adiante na Seção 2.2.3. Segundo, é montada uma visão parcial da Internet reconstruindo um grafo a partir de um subconjunto de rastros coletados da CAIDA. Isso permite a reconstrução de caminhos fim-a-fim entre dois pontos e também permite inferir

novos caminhos fim-a-fim não presentes nos dados originais disponibilizados pela CAIDA. Neste caso, pode-se analisar a diversidade de caminhos entre dois pontos.

Foram selecionados aleatoriamente n rastros de rota únicos disponíveis nos dados originais da CAIDA. Espera-se que essas n medidas fim-a-fim únicas qualitativamente representem um conjunto de dados muito maior, dado que *hosts* no mesmo bloco de endereçamento /24 estão inclinados a experimentar desempenho similar, como atraso na rede e perda de pacotes. Neste trabalho, foi escolhido $n = 200$ milhões como espaço amostral para aprimorar a acurácia da análise da conectividade fim-a-fim.

A visão parcial da Internet é representada por um grafo $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, onde \mathcal{V} é o conjunto de nós encontrados em todos os rastros de rota selecionados e \mathcal{E} o conjunto das arestas entre um par de nós. \mathcal{G}_{2006} e \mathcal{G}_{2011} são referenciados como grafos que modelam a topologia da Internet nos anos 2006 e 2011, respectivamente. O grafo da Internet é reconstruído de $m = 100.000$ rastros de rota aleatoriamente selecionados da amostra n . A escolha deste valor se baseia no artigo de Lee et al. (2010a), que indica que uma amostra de tamanho $n = 50.000 \sim 60.000$ apresenta pequenos erros na estimação de média para um intervalo de confiança de 99%. Como a Internet é muito maior do que a visão parcial, e a criação de um grafo com todos os dados obtidos é inviável, foram gerados 5 grafos a partir de 3 diferentes amostras m para alcançar maior confiabilidade no tamanho da amostra. Observou-se que há pouca variabilidade entre os resultados obtidos a partir dos grafos gerados, logo o tamanho da amostra se mostrou adequado para a nossa análise. Os grafos \mathcal{G}_{2006} têm ~ 178.000 nós e ~ 475.000 arestas; e os grafos \mathcal{G}_{2011} têm ~ 163.000 nós e ~ 455.000 arestas. Após a construção dos grafos, verificou-se que todos são conexos.

Na maioria dos casos, a caracterização apresentada é baseada nas propriedades encontradas nos K -menores caminhos da visão parcial da Internet.

Seja P_{st} o conjunto de todos os caminhos de s para t e c uma função de custo para cada caminho $p \in P_{st}$. O problema dos K -menores caminhos consiste na determinação de um conjunto $\{p_1, \dots, p_k\}$ de caminhos entre um dado par de nós quando a função objetivo do problema dos caminhos mais curtos é considerada de tal forma que $c(p_k) \leq c(p)$ para qualquer $p \in P_{st} - P_{st}(k-1)$, onde $P_{st}(k) = \{p_1, \dots, p_k\}$ e $P_{st}(0)$ é um conjunto vazio. O caminho mais curto é o primeiro a ser determinado, seguido pelo segundo mais curto e assim por diante. Se o caminho mais curto tiver 5 saltos, o primeiro caminho a ser encontrado será de 5 saltos e o segundo caminho obrigatoriamente será maior ou igual a

5 saltos.

A importância da análise baseada nos K -menores caminhos está relacionada ao fato da maioria dos protocolos de roteamento tenderem a seguir o caminho mais curto entre dois nós em uma rede (e.g. OSPF). Caminhos mais longos, apesar de existirem, são pouco usados na prática. São selecionados 10 origens aleatórias e 10.000 destinos para gerar os K -menores caminhos usados na caracterização. Os K -menores caminhos não representam um roteamento completamente real pois o roteamento entre ASs é normalmente feito com o protocolo BGP (Border Gateway Protocol), que é influenciado por políticas comerciais. Como tais políticas são de difícil inferência, o roteamento entre ASs é abstraído.

2.2.3 MÉTRICAS

Este estudo se baseia nas métricas mais comuns para caracterizar a conectividade fim-a-fim. Por exemplo, é possível descrever um caminho da Internet em termos do seu tamanho (contagem de saltos) e latência (atraso do caminho). Também é caracterizada a diversidade dos caminhos na Internet. A seguir, o conjunto de métricas selecionadas para a análise da evolução da conectividade da Internet é formalizada:

- *Tamanho do caminho* τ (ou número de saltos): é a distância entre dois *hosts*. São coletados o número de saltos de $n = 200$ milhões rastros de rota dos dados da CAIDA e de visão parcial da rede reconstruída com $m = 100.000$ rastros de rota da CAIDA. Também é analisado o número de saltos dos K -menores caminhos obtidos da visão parcial da Internet.
- *Atraso do caminho* η : é o tempo de ida e volta (RTT) entre dois *hosts*. Um grande atraso pode indicar, por exemplo, um grande caminho (alta contagem de saltos) ou um caminho congestionado. O atraso do caminho é analisado usando o RTT obtido de $n = 200$ milhões de rastros de rota da CAIDA. Também é usado o atraso inferido a partir da visão parcial ($m = 100.000$ rastros de rota da CAIDA) da Internet para correlacionar propriedades do caminho.
- *Diversidade de caminhos* μ : é uma métrica que reflete o número de rotas alternativas disponíveis entre um par origem-destino. A diversidade de caminhos permite quantificar a riqueza da conectividade da topologia da Internet. As características

da diversidade de caminhos são obtidas analisando as diferenças entre os K -menores caminhos, usando os grafos \mathcal{G}_{2006} e \mathcal{G}_{2011} criados a partir da amostra m .

- *Diferença entre caminhos Δ* : a diferença entre dois caminhos que ligam dois pontos dados é quantificada usando a distância Levenshtein (distância de edição) (LEVENSHTEIN, 1966; WAGNER; LOWRANCE, 1975). Em outras palavras, a diferença entre os caminhos é a quantidade mínima de operações (inserção, remoção e substituição) necessárias para transformar um caminho em outro. O método original foi utilizado para diferenciar texto. Aqui uma frase corresponde a um caminho, onde cada palavra é o endereço IP de um roteador. Assim é possível quantificar quantos nós foram removidos, adicionados ou trocados de posição entre os caminhos entre 2 nós.
- *Dispersão geográfica do caminho D_g* de um caminho entre dois pontos: é definida como a razão entre a distância geodésica total D_p , considerando a sequência de diferentes pontos intermediários que compõem um caminho fim-a-fim, e a distância geodésica direta real D_r entre os dois extremos do caminho, i.e.

$$D_g = \frac{|D_p - D_r|}{D_r}.$$

A distância geodésica é obtida a partir da Lei de Cossenos, i.e

$$d = \text{acos}(\sin(\varphi_1) \cdot \sin(\varphi_2) + \cos(\varphi_1) \cdot \cos(\varphi_2) \cdot \cos(\Delta\lambda) \cdot R)$$

Onde d é a distância geodésica, φ a latitude, λ a longitude e R o raio da Terra. É esperado que quanto mais rica for a conectividade observada, menor será a dispersão geográfica, dado que mais caminhos diretos estão disponíveis. Os K -menores caminhos da visão parcial da Internet (grafos \mathcal{G}_{2006} e \mathcal{G}_{2011}) são usados para analisar a dispersão geográfica dos caminhos.

2.3 RESULTADOS

Nesta seção, são analisados os resultados da caracterização estudada usando as métricas apresentadas na Seção 2.2.3.

2.3.1 TAMANHO DOS CAMINHOS

Como primeiro conjunto de resultados, foi analisado o tamanho dos caminhos da amostra de $n = 200$ milhões de rastros de rota da CAIDA de 2006 e de 2011. A Figura 2.1 mostra que a CDF (Cumulative Distribution Function) do número de saltos dos caminhos muda levemente de 2006 para 2011, com as curvas mantendo a mesma inclinação. Apesar de novos investimentos na infraestrutura da rede no intervalo de 5 anos, os caminhos fim-a-fim tiveram uma redução marginal de um salto. Em ambos os casos, 80% dos caminhos têm 17 saltos ou menos.

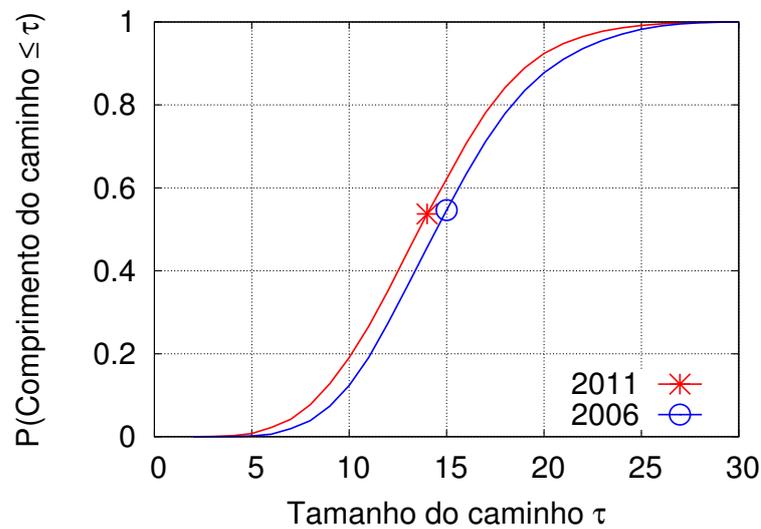


Figura 2.1: CDF - número de saltos.

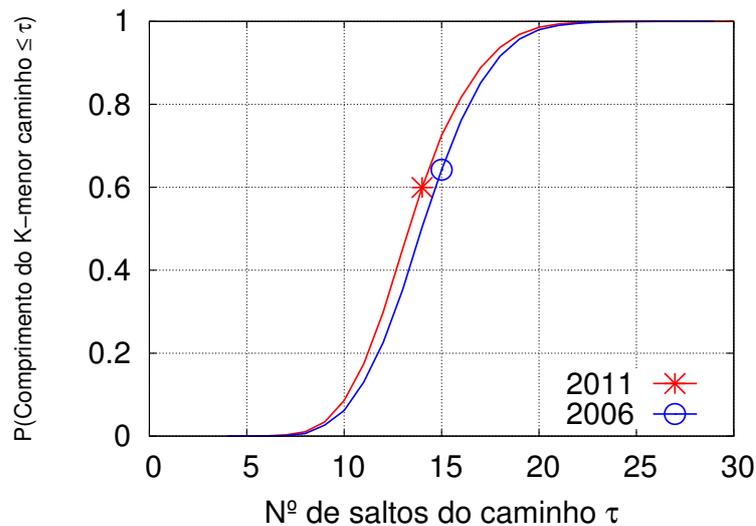


Figura 2.2: CDF - Número de saltos dos K-menores caminhos.

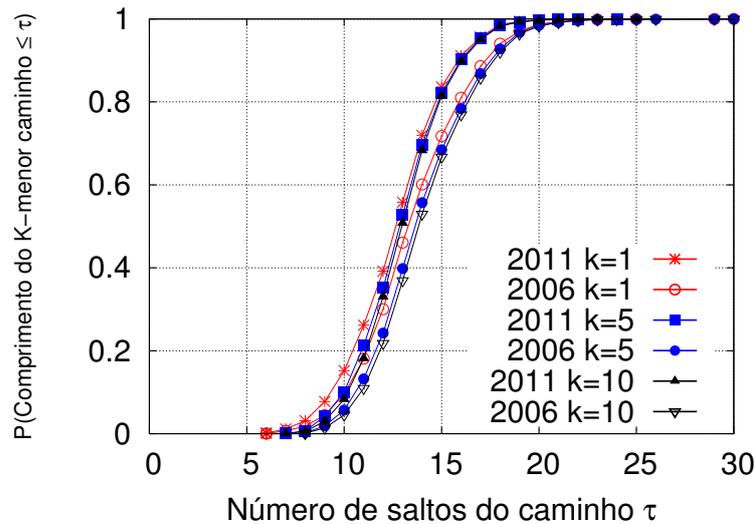


Figura 2.3: CDF - Comprimento dos K -menores caminhos.

Como a rede cresce e o comprimento dos caminhos se mantém praticamente inalterado, há o indício de que a rede está se tornando mais densa, o que é corroborado por investigações recentes (LABOVITZ et al., 2010; AGER et al., 2012). Em uma rede mais densa, espera-se ter uma conectividade mais rica, dado que devem existir mais caminhos disponíveis entre dois extremos. De forma similar, em uma rede mais densa, intuitivamente seriam esperados atrasos menores. Entretanto, este fato não é comprovado experimentalmente. Os atrasos, em realidade, se tornaram maiores entre 2006 e 2011. Os resultados são contra-intuitivos em relação ao fato da rede estar se tornando mais densa. Esta aparente contradição será discutida em mais detalhes na Seção 2.3.2.

Também foi analisado o tamanho de todos os K -menores caminhos da visão parcial da rede. A Figura 2.2 mostra que os caminhos de 2011 são ligeiramente mais curtos que os caminhos de 2006. Quase 80% dos caminhos não possuem mais que 16 saltos e não foi encontrado nenhum K -menor caminho com mais de 25 saltos em 2011.

Na Figura 2.3, é exibida uma visão mais próxima de cada K -menor caminho individual. O comportamento é praticamente o mesmo. Como esperado, o caminho $K = 1$ é o mais curto, com mediana de 14 saltos para 2006 e 13 saltos para 2011. O comprimento do caminho se mantém estável de acordo com o crescimento do K até ~ 10 , com os caminhos de 2011 tendo ~ 1 salto a menos do que os de 2006.

2.3.2 ATRASO DOS CAMINHOS

A Figura 2.4 apresenta a CDF do atraso dos caminhos em escala logarítmica. O atraso dos caminhos na Internet está se tornando *pior* comparando 2011 com 2006. De fato, isso mostra que em 2006, cerca de 50% (mediana) dos caminhos apresenta um atraso de pelo menos 122ms, enquanto em 2011 a mediana do atraso cresceu cerca de 37% para quase 168ms. Considerando o percentil 95%, o atraso aumentou de cerca de 352ms em 2006 para cerca de 512ms em 2011 (um aumento de 45% em cinco anos). Esse é um resultado contra-intuitivo aos estudos recentes que apontam o aumento densidade da rede nos últimos anos (LABOVITZ et al., 2010; AGER et al., 2012).

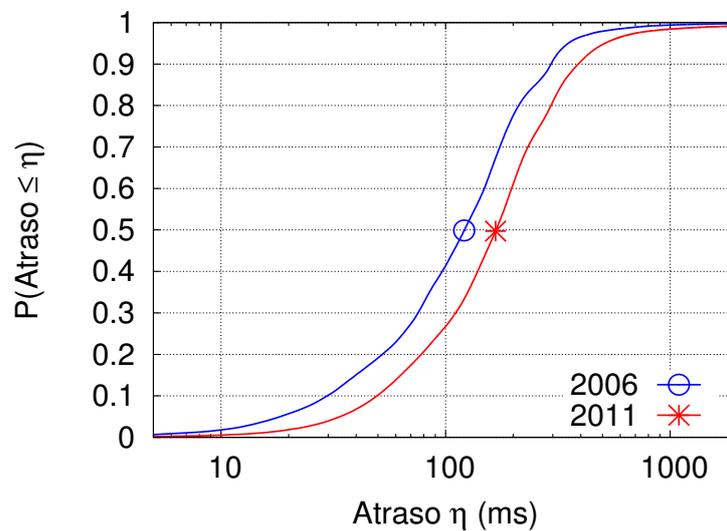


Figura 2.4: CDF - Atraso dos caminhos.

Os resultados aqui discutidos reforçam que o desempenho da Internet está se tornando pior, como mostrado em (LEE et al., 2010a). Uma possível explicação pode ser o fenômeno conhecido como *bufferbloat*, recentemente relatado por Gettys e Nichols (2012), onde a *bufferização* exagerada de pacotes na rede causa grande aumento na latência e no *jitter*. Além disso, nota-se que o número de redes virtualizadas está em franco crescimento. O acesso a servidores em máquinas virtuais pode introduzir atrasos de até 100ms ao RTT (tempo de ida e volta de um pacote na rede) (WHITEAKER et al., 2011). Como consequência, a latência geral percebida da rede também aumenta.

2.3.3 DIFERENÇA ENTRE CAMINHOS

A Figura 2.5 mostra a CDF da diferença entre os K -menores caminhos entre dois extremos da rede considerada. A diferença entre dois caminhos é calculada com base na distância de Levenshtein. Os resultados mostram que a diferença entre caminhos está diminuindo durante este período analisado de 5 anos. Em 2011, menos de 10% dos caminhos apresentam uma diferença maior do que 50%, enquanto em 2006, esse número chega a 15% dos caminhos. Uma possível explicação é que a diversidade cresceu em torno dos caminhos mais curtos disponíveis. Sendo assim, os caminhos compartilham muitos nós em comum, havendo sobreposições ao longo do caminho, tornando esses caminhos mais similares. Outra possibilidade é a prevalência de acordos financeiros entre provedores, que pode estar forçando o tráfego a passar por certos pontos específicos dentro dos ASs envolvidos.

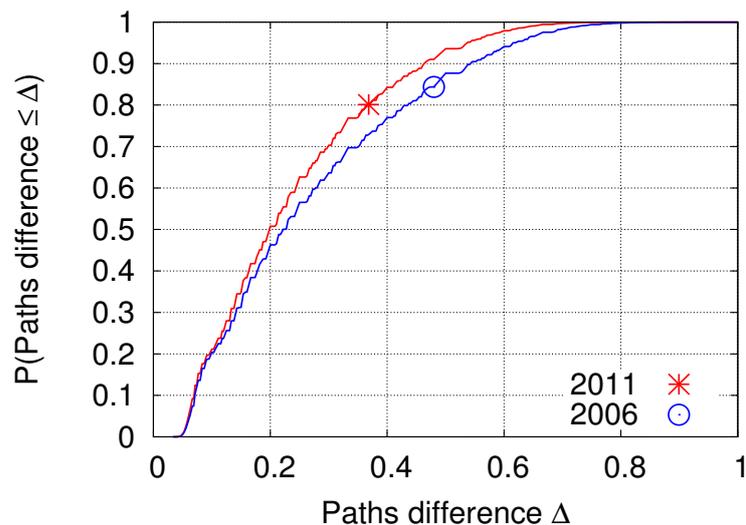


Figura 2.5: CDF - Diferença dos K -menores caminhos.

A Figura 2.6 mostra que conforme K aumenta, maior é a diferença entre os caminhos. Por exemplo, 80% das diferenças entre os caminhos de 2011 são menores do que 15% para $K = 2$, enquanto a diferença entre os caminhos é maior em 36% para $K = 10$. Portanto, protocolos de roteamento que consideram diversidade de rotas (e.g. (WANG; GAO, 2009)), ou seja, mais caminhos distintos do que os mais curtos, podem utilizar melhor esta diversidade e obter maior tolerância a falhas, por exemplo.

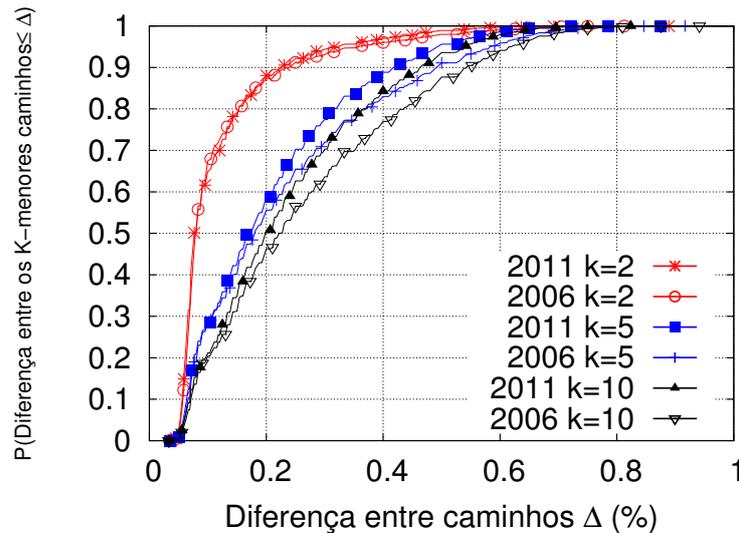


Figura 2.6: CDF - Diferença dos K-menores caminhos.

2.3.4 DISPERSÃO GEOGRÁFICA

A localização geográfica dos *hosts* foi obtida utilizando uma base de dados gratuita disponibilizada pela MaxMind³. Essa base permite obter a localização geográfica de cada *host* fornecendo detalhes como a cidade, latitude e longitude.

A Figura 2.7 mostra que a maioria dos caminhos tem dispersão geográfica de 0 até 3, onde 0 é uma linha geodésica direta. Enlaces longos, como cabos transoceânicos, podem ter maior impacto na dispersão geográfica. Então, um único enlace transoceânico entre todos os outros impacta a dispersão geográfica geral.

Finalmente, dados de 2011 tem cerca de 10% (Figura 2.7) dos caminhos com dispersão geográfica maior que a de 2006. Isso pode sugerir que pequenos provedores de Internet estão mudando seus enlaces locais para enlaces de longa distância. A maioria dos operadores das grandes redes podem tipicamente desviar seu tráfego de rede por distâncias maiores por questões econômicas.

A Figura 2.8 mostra a CDF da dispersão geográfica para cada K -ésimo caminho. Nota-se que com o aumento de K para 10, a dispersão geográfica se mantém estável. Nesse caso, quase 90% dos caminhos tem dispersão menor que 4.

³<http://www.maxmind.com/>

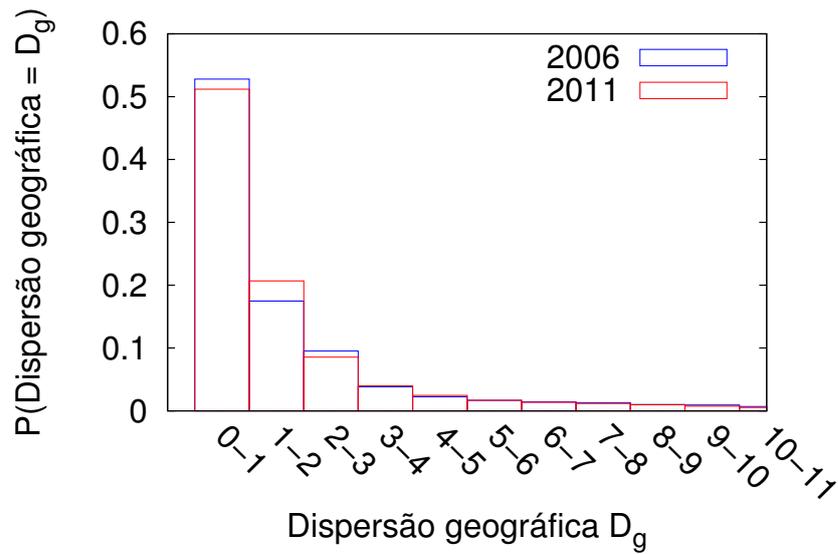


Figura 2.7: Histograma - Dispersão geográfica dos K-menores caminhos.

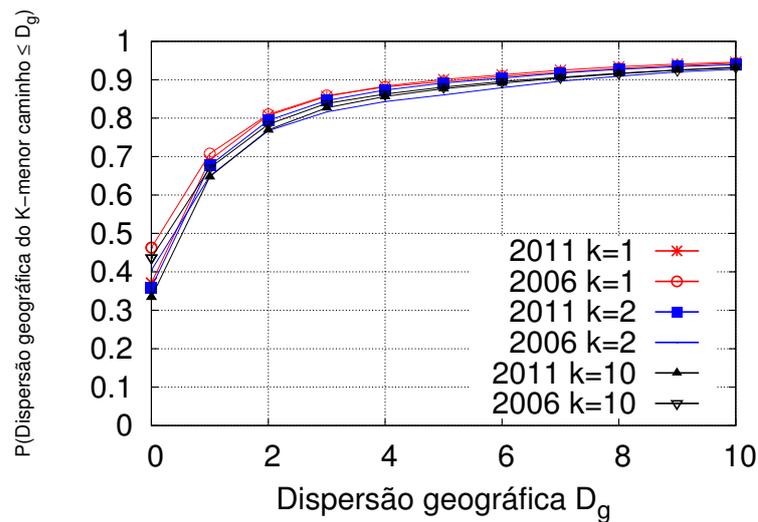
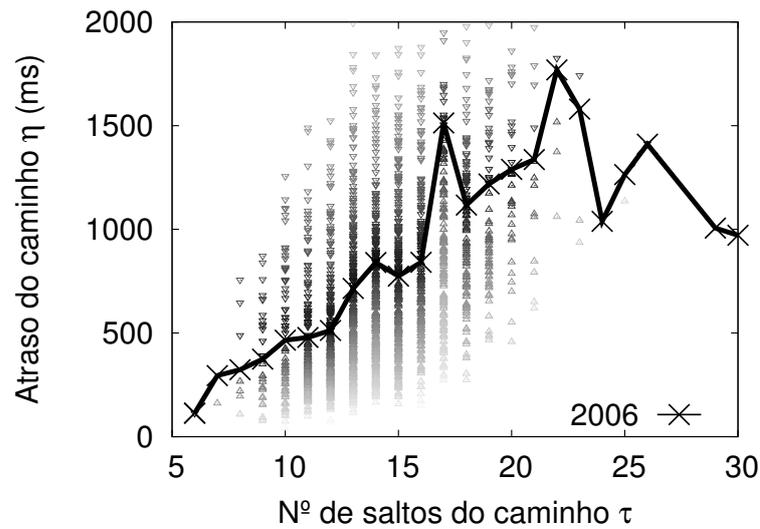


Figura 2.8: CDF - Dispersão geográfica dos K-menores caminhos (K variante).

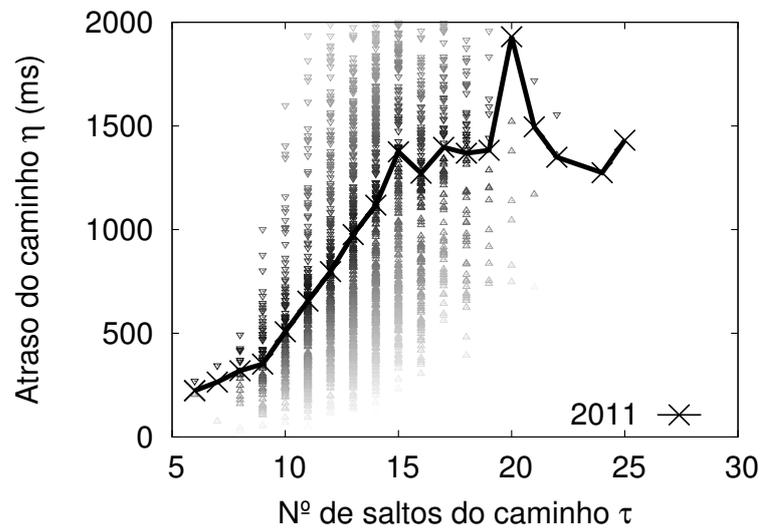
2.3.5 CORRELAÇÃO ENTRE AS MÉTRICAS

Para investigar mais profundamente a evolução recente da conectividade fim-a-fim da Internet, a correlação entre algumas métricas foi calculada.

Em nossa análise, usamos o coeficiente de correlação de Spearman (SPEARMAN, 1904). O coeficiente de Spearman é uma medida não paramétrica que mede o quão relacionadas são duas variáveis. Como o método não utiliza os valores observados, mas apenas a ordem, há menor influência de pontos discrepantes. A variação do coeficiente fica entre -1 e 1, e quanto mais próximo desses dois valores, maior a correlação.



(a) 2006



(b) 2011

Figura 2.9: Atrazo e comprimento dos K-menores caminhos.

A Figura 2.9 mostra um gráfico de dispersão correlacionando o atrazo do caminho com o seu comprimento (em número de saltos). O tom de preto dos triângulos indica a quantidade de pontos na coordenada, onde quanto mais escuro mais pontos. A média dos pontos está destacada pela linha. É claramente notável a alta correlação entre os caminhos com comprimento menor que 20 para ambos os anos analisados. Foi encontrado uma correlação de Spearman de 0,8 para os dados de 2006 e de 0,88 para os dados de 2011. Uma alta correlação como essa mostra que quanto maior um caminho, maior a chance de aumento de latência.

Para caminhos com mais de 20 saltos, observa-se que a latência permanece razoa-

velmente estável. Isso pode ocorrer, pois, em caminhos longos, um número pequeno de caminhos dominam o atraso e, portanto, aumentar o caminho fim-a-fim com poucos enlaces de baixa latência não impacta significativamente a latência total. Sobretudo, essa porção da correlação (caminhos maiores que 20 saltos) representa uma pequena quantidade de caminhos. Vale ressaltar que, de acordo com a Figura 2.2, o número de caminhos com mais de 20 saltos é desprezível.

A correlação entre a dispersão geográfica e a dispersão do atraso também foi analisada. Define-se a dispersão do atraso ρ como a razão entre o atraso medido e o atraso mínimo (i.e. apenas o atraso de propagação) esperado em um caminho usando um enlace de fibra ótica. De acordo com Percacci e Vespignani (2003), a informação digital viaja através da fibra ótica em cerca de quase exatamente $2/3$ da velocidade da luz no vácuo. Assim, pode-se estimar o atraso total usando um cabo de fibra ótica conectando dois extremos diretamente.

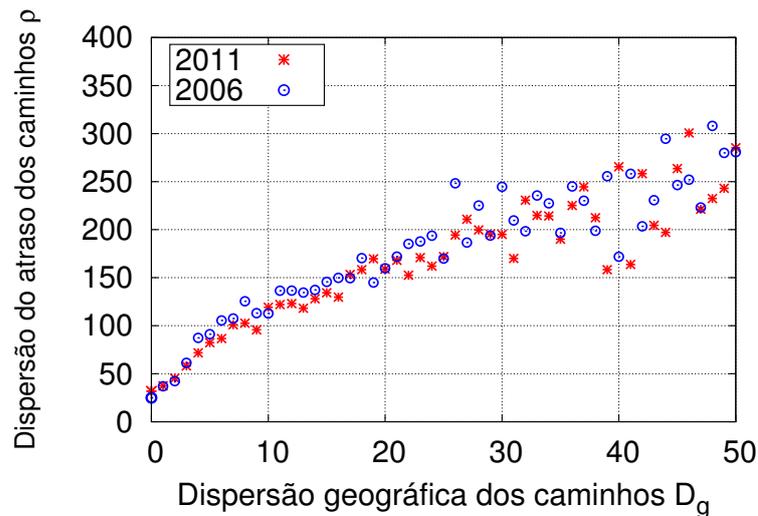


Figura 2.10: Dispersão do atraso e geográfica dos K -menores caminhos.

A Figura 2.10 mostra que a média da dispersão do atraso e a média da dispersão geográfica do caminho são altamente correlacionadas para 2011 e 2006, especialmente para a porção dos caminhos com geodispersão entre 0 e 30. O coeficiente de Spearman foi 0,93 para 2006 e 0,84 para 2011. Esse resultado mostra que quanto menos direto um caminho entre dois pontos for, maior será o atraso. Um ponto importante é que em 2011 ainda existem caminhos com elevada dispersão geográfica. Um dos principais motivos são as regras de roteamento BGP resultantes de acordos comerciais entre operadoras.

Finalmente, analisamos a correlação entre a média do atraso do caminho e a média do

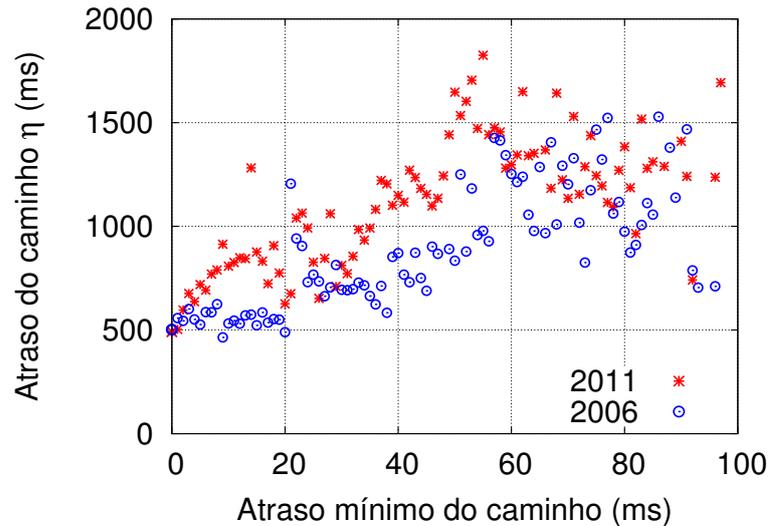


Figura 2.11: Atraso e o atraso mínimo dos K-menores caminhos.

atraso mínimo esperado em um enlace de fibra ótica. A Figura 2.11 mostra que em 2011 há uma maior dispersão entre o atraso mínimo e o atraso real. Caminhos mais curtos podem ter seus enlaces em fibra local ou cabos. Na direção oposta, caminhos mais longos possuem tipos de enlaces que podem explicar a má correlação encontrada em caminhos mais longos com atraso. É possível que esses caminhos mais longos com atraso estejam conectados por um número menor de enlaces com alta latência e, como consequência, o atraso total do caminho não é proporcional ao atraso mínimo esperado. Observamos uma melhor correlação para 2006, com coeficiente de Spearman de cerca de 0,78 e uma correlação média para 2011, com coeficiente de Spearman de cerca de 0,67.

2.4 CONCLUSÃO DO CAPÍTULO

Alguns dos resultados contra-intuitivos encontrados na análise realizada indicam que o desempenho da Internet está piorando. Por exemplo, a distribuição do tamanho médio dos caminhos muda pouco de 2006 para 2011. Durante esse período, entretanto, era esperado que investimentos (novos equipamentos e enlaces) na rede tornassem os caminhos atuais mais curtos e com menos atraso. Entretanto, a análise também mostra que o atraso fim-a-fim se tornou claramente pior. Mais uma vez, novos equipamentos e enlaces deveriam prover uma melhor conectividade com enlaces mais rápidos, resultando em uma experiência de usuário aprimorada. Contudo, nesse caso, é mostrado que está ocorrendo o aumento perceptível do atraso. Conjectura-se que os novos equipamentos na Internet,

sobretudo nas pontas, que é onde se encontram os roteadores caseiros que atuam como gargalos, não são configurados para usar gerenciamento ativo de fila, também conhecido como AQM (Active Queue Management). AQMs são algoritmos que empregam técnicas para impedir o uso completo das filas dos roteadores. A falta de AQMs pode levar ao fenômeno recentemente nomeado de *bufferbloat* (GETTYS; NICHOLS, 2012), que é o excesso de armazenamento de pacotes na rede, causando alta latência.

Um outro ponto importante que deve ser considerado ao se considerar os motivos do aumento do atraso é como o perfil do tráfego tem mudado na última década. Junto com o enorme crescimento da Internet, é esperado que também haja grande aumento do tráfego, uma vez que há muito mais máquinas participantes da rede. Esse aumento de tráfego não está associado apenas ao tamanho da rede, mas também ao tipo de tráfego. Nos últimos anos houve grande aumento do uso de aplicativos multimídia como Netflix⁴, Youtube e sistemas de compartilhamento de arquivo P2P (Peer-to-Peer), como o Bittorrent⁵.

Adicionalmente, mostrou-se que a diversidade de caminhos diminuiu, e similarmente, caminhos mais distintos ficaram mais similares. Esse resultado tem impacto direto nos algoritmos de roteamento que tentam explorar a diversidade de caminhos para a obtenção de maior tolerância a falhas. Esse resultado indica que os caminhos da Internet estão crescendo em torno do caminho mínimo e, como consequência, o impacto de novos enlaces é marginal na diferença geral entre dois caminhos fim-a-fim alternativos. Uma redução na diversidade de caminhos pode afetar o desempenho da rede, uma vez que pode levar a sobrecarga dos nós que compõem esses caminhos mais utilizados.

Finalmente, foi encontrado que o atraso está altamente correlacionado ao número de saltos na maior porção dos caminhos fim-a-fim. A dispersão geográfica tem alta correlação com o atraso do caminho, especialmente para valores baixos na dispersão geográfica (onde encontram-se quase todos os caminhos). Também é notado que o atraso dos caminhos tem uma correlação com padrão exponencial com o atraso mínimo esperado (para atraso abaixo de 50 *ms*). Esses padrões de correlação, para a grande porção de caminhos, indica que pode-se inferir métricas utilizando-se de uma métrica bem simples, como o tempo de ida e volta.

No capítulo seguinte é estudado mais a fundo o fenômeno *bufferbloat* e a sua relação com o atraso encontrado nos dados de rede analisados.

⁴www.netflix.com

⁵www.bittorrent.com

3 BUFFERBLOAT

Neste capítulo é realizada uma análise mais detalhada sobre o fenômeno conhecido como *bufferbloat*, que é uma das possíveis causas do aumento de latência percebido nos resultados apresentados no capítulo 2. Esta análise consiste na observação do impacto do fenômeno em diferentes cenários de experimentação.

Bufferbloat é, em essência, o aumento exagerado das filas nos roteadores, principalmente nos que atuam como gargalos. Quando a taxa de entrada de tráfego em um nó excede a taxa de saída, ocorrerá naturalmente o aumento do preenchimento da fila. Quanto maiores forem as filas, mais tempo novos pacotes deverão esperar para serem re-enviados e, conseqüentemente, ocorrerá um aumento contínuo na latência, como ilustrado na Figura 3.1. Essas filas que nunca diminuem de tamanho são chamadas de filas ruins. Filas boas são aquelas que mesmo quando absorvem grandes rajadas de pacotes tendem a serem desocupadas com o passar do tempo.

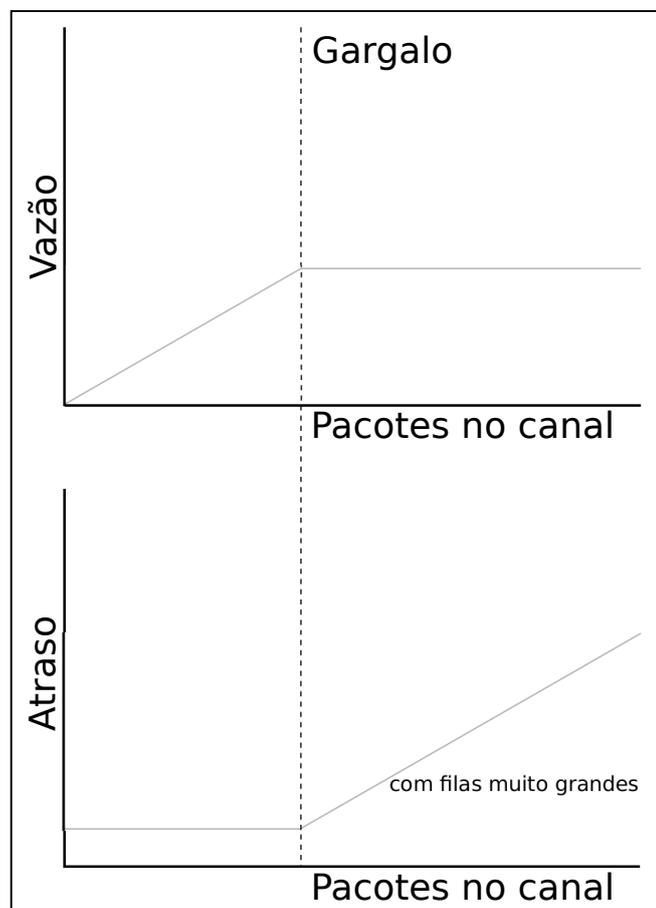


Figura 3.1: O efeito de filas grandes em gargalos [Gettys e Nichols (2012)].

O problema de filas muito grandes em roteadores não é um problema novo. Nagle (1985) já havia mostrado que em uma rede com filas infinitas onde os pacotes possuem tempo de vida limitado, a vazão tende a zero. Isso porque os pacotes ficam tempo demais na fila e acabam expirando antes ou logo após serem retransmitidos.

Esse aumento das filas nos roteadores se deve primariamente ao barateamento de *hardware*, a cultura de que descarte de pacotes é algo ruim e a dificuldade de definir tamanhos de fila em um cenário dinâmico como a Internet. O barateamento do preço do *hardware*, em particular, permitiu que os fabricantes de roteadores expandissem a capacidade de armazenamento dos seus equipamentos de uma forma imprevisível há duas décadas atrás. Um dos principais efeitos colaterais desse demasiado aumento é o impedimento do correto funcionamento do algoritmo de controle de congestionamento do TCP (Transmission Control Protocol).

O TCP é o protocolo de transporte mais utilizado na Internet. É usado no envio de e-mails, transferências de arquivos e qualquer outra troca de informação que exija confiabilidade e integridade na entrega. Uma das diferenças para o outro protocolo da camada de transporte, o UDP (User Datagram Protocol), é que o UDP deixa todo o controle da transferência para a aplicação que faz o seu uso. Já o TCP possui um algoritmo de controle de congestionamento (JACOBSON, 1988) que ajusta a taxa de transferência frente ao aumento da latência, evitando o saturamento desnecessário do caminho. Contudo, esse algoritmo considera primariamente o descarte de pacotes como indicador de congestionamento. Com o aumento considerável das filas, o congestionamento ocorre e não há sinalização para o TCP. Com isso, não ocorre o reajuste da janela de transmissão do TCP.

Este capítulo está estruturado como se segue. Na Seção 3.1 são descritos os trabalhos relacionados. Os métodos de experimentação e a configuração do ambiente de testes é mostrado na Seção 3.2. Os resultados desses testes são discutidos na Seção 3.3. Por fim a Seção 3.4 conclui o capítulo.

3.1 TRABALHOS RELACIONADOS

De acordo com Gettys e Nichols (2012), o problema de filas é algo generalizado e encontrado por toda a rede. O artigo propõe a adoção de AQM, que reúne algoritmos para o gerenciamento ativo de filas, que conseguem detectar quando uma fila está crescendo demais (fila ruim). O primeiro esforço para gerar um algoritmo para o controle do pre-

enchimento das filas foi feito por (FLOYD; JACOBSON, 1993), resultando no algoritmo *Random Early Detection* (RED) que ao longo dos anos seguintes recebeu algumas aprimorações e outras versões. O problema com esse algoritmo é ser de difícil configuração e todo o processo é manual, e não está disponível em todos os dispositivos.

Nichols e Jacobson (2012), motivados pelo trabalho de Gettys e Nichols (2012), desenvolveram um novo método de AQM que resolve os problemas de configuração manual que atrapalhavam a implantação de algoritmos de gerenciamento de fila. Esse novo algoritmo monitora continuamente o tempo que os pacotes ficam em uma fila, tentando distinguir filas boas de filas ruins, e caso seja ruim inicia-se uma fase de descarte. A grande vantagem está no fato de que, ao contrário das propostas anteriores, a configuração do AQM é automática e auto adaptativa. Uma limitação em adotar esse algoritmo é que sua implantação deve ocorrer em todos os pontos na rede, uma vez que apenas a fila local é observada. Os resultados de prova de conceito do seu funcionamento se limitam a simulações realizadas no NS2¹, e ainda não foi analisado o desempenho do algoritmo proposto em cenários de experimentação reais.

Chirichella e Rossi (2013) realizaram experimentos com o campo de *timestamp* do cabeçalho do LEBDAT, que é um protocolo proposto pelo BitTorrent em alternativa ao TCP, para inferir o atraso causado por filas grandes. Os resultados mostraram que 90% dos *peers* testados possuem enfileiramento com menos de 100 ms de espera. Isso mostra que o sistema de controle de congestionamento baseado em atraso do LEBDAT é bem resiliente a filas grandes para o caso de transferências de arquivos na arquitetura P2P. De fato, algoritmos de controle de congestionamento orientados a RTT são conhecidos por serem imunes ao *bufferbloat*, contudo também são conhecidos por serem incapazes de realizarem um uso ótimo da rede (JIANG et al., 2012b), pois as variações no RTT levam esses algoritmos a realizar uma subestimação da capacidade de transmissão. Ao contrário de Chirichella e Rossi (2013) que foca nas redes de BitTorrent, nesta dissertação é estudado o caso de transferência de dados utilizando o protocolo padrão TCP.

Gijsbers e Akkoorath (2011) observaram os efeitos do *bufferbloat* no simulador de redes Omnet++². Foi notado que com o aumento das filas, menor era a quantidade de descartes e maior se tornava o atraso. Este trabalho não realiza experimentos em ambientes mais próximos do real como o explorado nesta dissertação, nem observa o comportamento do

¹Simulador de redes (<http://www.isi.edu/nsnam/ns/>)

²Simulador de redes (www.omnetpp.org/)

TCP em termos de retransmissões.

Em (JIANG et al., 2012b,a), os autores exploram os efeitos do uso excessivo de *buffers* em redes móveis. Devido a natureza das redes móveis, dispositivos sem fio possuem filas grandes para absorver rajadas de tráfego e variabilidade de canal(3G/4G, Wi-Fi). Para conter o eventual super preenchimento, os dispositivos anunciam uma fila de recebimento de tamanho fixo muito menor do que a capacidade real. O problema com essa abordagem é que o uso de um tamanho de fila fixo se torna sub-ótimo em algumas ocasiões. O trabalho então propõe um algoritmo (DRWA - Dynamic Receiving Window Adjustment) que dinamicamente ajusta a fila de recebimento do TCP, realizando troca de informações bidirecionalmente entre a origem e o destinatário. Os experimentos mostram uma redução de 24 ~ 49% na latência em comparação a implementação padrão do TCP. Esse tipo de abordagem que leva em conta apenas as pontas tem a vantagem de ser de fácil implantação.

Allman (2012) realizou um estudo empírico buscando avaliar a prevalência do fenômeno *bufferbloat*. Para tal, foi observado o comportamento do tráfego do projeto Case Connection Zone (CCZ)³. O CCZ é um projeto experimental que fornece cabeamento de fibra ótica para as residências próximas a universidade Case Western Reserve University. Os fluxos de dados foram analisados durante 14 meses, entre 2011 e 2012. Os resultados mostraram que apesar de ocorrer um certo enfileiramento, menos de 1% dos pacotes tiveram um aumento de mais de 1 segundo no RTT. Outro experimento realizado foi a análise de uma proposta de aumentar a janela de congestionamento inicial do TCP, de cerca de 3 segmentos para 10. Foram coletados dados de 3 pontos da rede CCZ durante 1 semana. O experimentos concluem que para a maioria das situações o aumento da janela de congestionamento não causa impacto no enfileiramento. O estudo realizado nesta dissertação não realiza uma análise passiva do tráfego na rede, mas sim uma experimentação de como diversos tamanhos de fila impactam no desempenho da rede.

3.2 METODOLOGIA

Nesta seção é apresentada a configuração dos experimentos realizados em nosso estudo sobre o *bufferbloat*.

³<http://www.caseconnectionzone.org/>

3.2.1 AMBIENTE DE EXPERIMENTAÇÃO

Para observar o *bufferbloat* será utilizada a técnica de medição ativa (CROVELLA; KRISHNAMURTHY, 2006), que consiste no envio de pacotes criados especificamente para serem enviados pela rede, partindo da origem e passando pelas máquinas que compõem o ambiente de experimentação até o destino. Esses pacotes são chamados de sondas. As sondas permitem a obtenção de informações, não apenas do destino, mas de todo o caminho. As sondas funcionam em um esquema de envio e resposta, onde é primeiramente enviada uma sonda, e a consecutiva só será enviada após a chegada da resposta da primeira. Quando esses pacotes são agrupados em fluxos da mesma origem para o mesmo destino, são chamados de *trens de pacotes* (JAIN; ROUTHIER, 1986). No caso de trens de pacotes, as sondas são enviadas consecutivamente tão rápido quanto permitido pelo ambiente. Enviado o primeiro trem, o consecutivo só será enviado após o recebimento das respostas relativas as sondas do trem anterior. Esses trens podem ser utilizados, por exemplo, para medir como a rede se comporta durante um aumento expressivo de tráfego.

A Figura 3.2 mostra a configuração do ambiente de testes. São ao todo 6 MacMinis⁴ executando o GNU/Linux Ubuntu 10.04 Server⁵ conectados por um *switch*. A fim de que cada nó se comporte como um roteador, a rede que conecta as máquinas está dividida em 5 VLANs. O uso de VLANs separa as máquinas em diferentes redes, de tal forma que a comunicação só pode ser feita através dos nós de conexão entre as VLANs utilizando algoritmos de roteamento. Como exemplo, as sondas enviadas do nó A para o nó E precisam passar pelos nós C e D, formando um caminho de três saltos. As VLANs estão configuradas como segue:

- VLAN 1: nós A e C
- VLAN 2: nós B e C
- VLAN 3: nós C e D
- VLAN 4: nós E e D
- VLAN 6: nós F e D

⁴<http://www.apple.com>

⁵<http://www.ubuntu.com>

Para que haja comunicação entre todos os integrantes da rede, todos os nós possuem o software de roteamento Quagga⁶, que é uma implementação código livre de algoritmos de roteamento. Para os experimentos realizados neste trabalho, é utilizado o protocolo de roteamento intra-domínio OSPF (Open Shortest Path First). O OSPF foi o escolhido pois é protocolo de roteamento padrão para os roteadores domésticos.

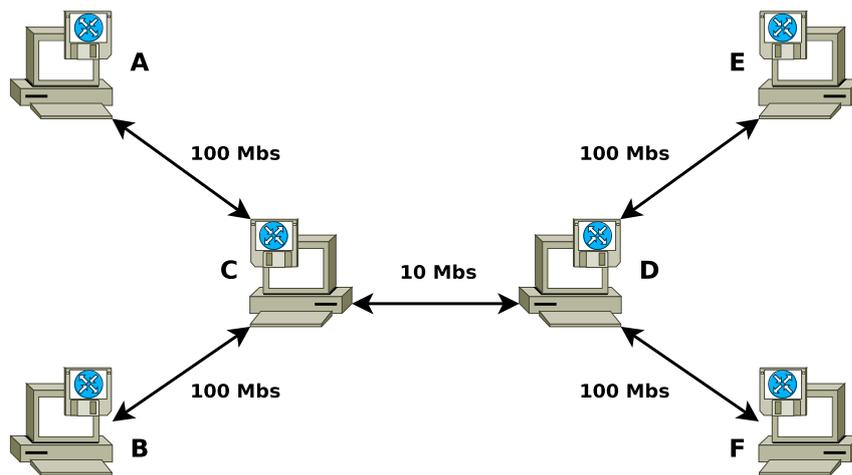


Figura 3.2: Bufferbloat testbed.

Todas as interfaces estão limitadas a 100Mbps, com exceção da conexão de C para D, que está limitada a 10Mbps. Essa limitação de 10Mbps se dá pelo fato de não dispormos de máquinas com interfaces de rede com maior velocidade.

Durante os testes, foram realizados envios de sondas em paralelo a cópia de um arquivo, e então foi observado como as sondas se comportam durante o preenchimento das filas. Para que ocorra o *bufferbloat*, o nó C foi selecionado como gargalo. Uma outra possível escolha para gargalo é o nó D.

Para tornar a simulação mais realista, é gerado tráfego de fundo para simular o uso cotidiano da rede de uma empresa. Para a geração desse tráfego de fundo foi utilizada a ferramenta *bittwist*⁷, que é capaz de reproduzir arquivos no formato *pcap* (packet capture). Pcap⁸ é uma API que fornece ferramentas para a observação do tráfego nas interfaces de rede, que pode ser armazenado em arquivos para posterior uso. O arquivo *pcap* foi obtido do projeto *LBNL/ICSI Enterprise Tracing Project*⁹, que disponibiliza um extenso repositório de captura de tráfego empresarial. Os arquivos são reproduzidos continuamente,

⁶<http://www.nongnu.org/quagga/>

⁷<http://bittwist.sourceforge.net/>

⁸<http://www.tcpdump.org/>

⁹<http://www.icir.org/enterprise-tracing/>

tendo o nó B como origem e o nó F como destino.

Na Figura 3.3 é mostrado o caminho que os dados percorrem desde a aplicação até o canal. No topo da pilha temos a aplicação escrevendo no *socket* TCP, que armazenará os pacotes no *buffer* de envio do TCP. Desde a versão 2.6 do Linux este valor é configurado automaticamente. Os pacotes então são passados para a camada IP, onde serão preparados para transmissão e armazenados na fila de transmissão (*qdisc* na Figura 3.3). Se a fila de transmissão estiver cheia, será iniciado um evento de congestionamento local, que irá bloquear a fila até que a mesma fique completamente vazia. Uma vez dentro da fila de transmissão, o pacote irá esperar a sua vez para ser enviado para o *ringbuffer*, de onde será copiado pelo dispositivo de rede para o canal.

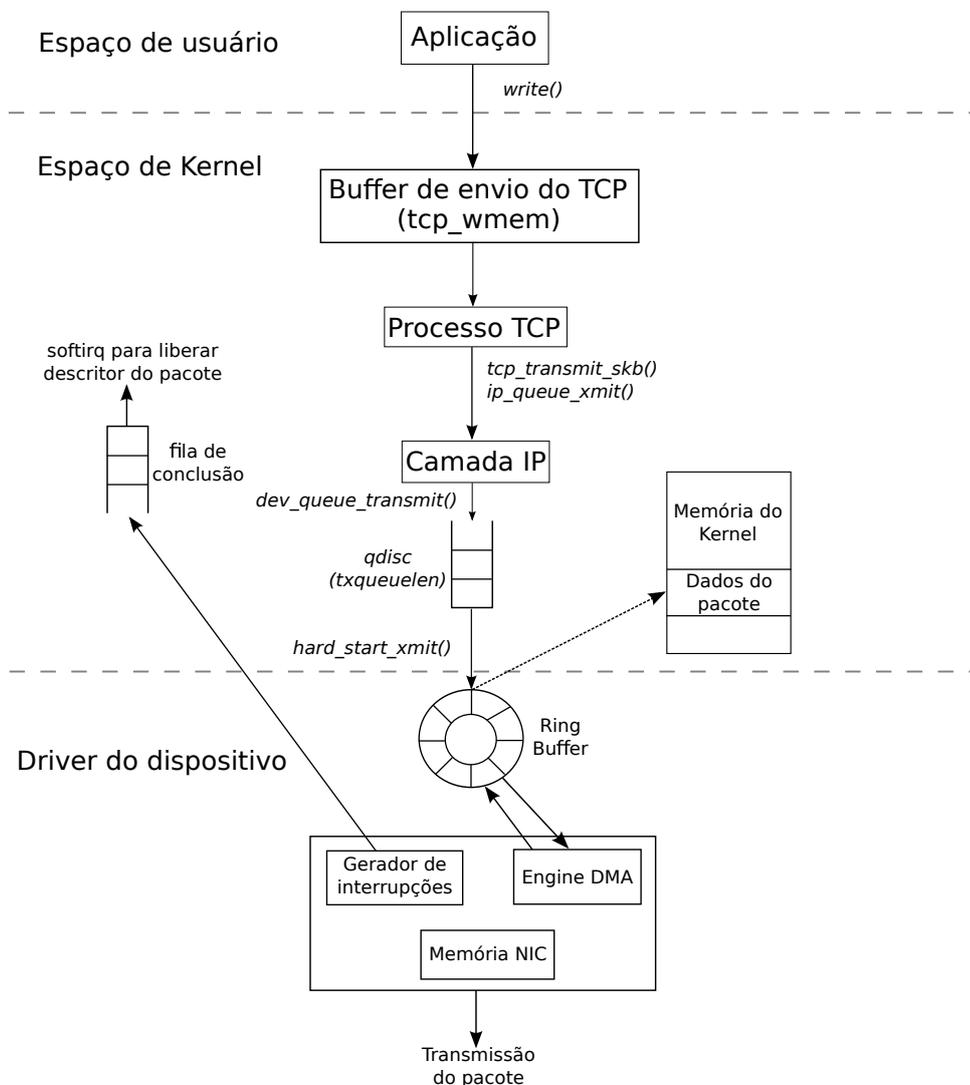


Figura 3.3: Buffers na pilha de rede no sentido de transmissão.¹⁰

¹⁰<http://www.cnblogs.com/ggjucheng/archive/2012/11/01/2750217.html>

Nos testes foram utilizadas diversas configurações para fila de transmissão na camada de rede (*qdisc*), e para a fila de transmissão na camada de enlace (*ringbuffer*). O posicionamento dessas filas na camada de rede do Linux pode ser visto na Figura 3.3. Apenas essas duas filas foram estudadas devido ao fato de que nos roteadores em geral a pilha de rede não possui as camadas de transporte e aplicação. As filas de recebimento não foram estudadas devido ao fato do gargalo estar no canal de saída. Para que ocorra o *bufferbloat* nas filas de recebimento seria necessário que o gargalo fosse no processamento dos pacotes.

3.2.2 FILAS ANALISADAS

Ringbuffer é um *buffer* de espera na camada de enlace para a entrada no canal de transmissão (sentido de saída, Figura 3.3). Como é gerenciado pelo *driver* do dispositivo, os detalhes de implementação variam para cada interface de rede. O tamanho padrão do *ringbuffer* em distribuições GNU/Linux é de 256 pacotes, e os limites inferior e superior são definidos pelo *driver*. Para alterar o tamanho em um sistema Linux pode-se utilizar o comando *ethtool*¹¹, caso o *driver* do dispositivo dê suporte a essa mudança.

A fila de transmissão (*qdisc*) é um *buffer* de interface entre a camada de rede e a camada de enlace. As distribuições GNU/Linux por padrão possuem esta fila configurada para armazenar 1.000 pacotes, que ficam em espera para passar para o *ringbuffer* na camada de enlace abaixo. O tamanho da fila de transmissão pode ser definido desde 0 até o quanto a máquina suportar em memória RAM. Em sistemas Linux, a fila de transmissão pode ser facilmente alterada pelo comando *ifconfig*¹² através do parâmetro *txqueuelen*. O *qdisc* segue alguma disciplina de fila, que no caso do Linux o padrão é *pfifo_fast*. Essa disciplina é baseada na FIFO (First In First Out), onde os pacotes não recebem nenhum tipo de priorização e são desempilhados da fila na ordem em que chegaram. No caso particular da *pfifo_fast*, a fila é dividida em 3 bandas de prioridade, onde primeiro são desempilhados os pacotes da banda 1, seguido pelos da banda 2 e por último os pacotes da banda 3. O preenchimento de cada banda é definido pela tabela *priomap*, que atribui as prioridades de acordo com o campo ToS (Type of Services) dos pacotes. O *tc* é um comando para a manipulação da configuração do controle de tráfego. Através desse comando pode-se

¹¹Comando para configurar o *driver* da interface de rede.

¹²Comando para a configuração de interfaces de rede

definir qual será a disciplina de fila utilizada no *qdisc*.

3.2.3 MÉTRICAS DE INTERESSE

As métricas analisadas são formalizadas a seguir:

- *Atraso do caminho η* : é o tempo de ida e volta (RTT) entre dois *hosts*. Diferentemente da métrica descrita na seção 2.2, aqui é esperado que o atraso esteja relacionado diretamente ao tempo de espera na fila. Quando a taxa de chegada em um fila for maior que a taxa de saída, maior será o seu crescimento e, conseqüentemente, maior será o tempo de espera dos pacotes recém chegados.
- *Uso do *qdisc* Ω* : é a quantidade de pacotes que estão na fila de espera entre a camada de rede e enlace. É esperado que um aumento considerável no preenchimento do *qdisc* afete diretamente o RTT observado.
- *Taxa de transferência Λ* : é a taxa de transmissão das cópias dos arquivos. Uma vez que o TCP possui controle de congestionamento, é de grande interesse estudar se ocorre alguma adaptação da taxa de envio ou quando o RTT aumenta consideravelmente ou quando pacotes são descartados.
- *Retransmissões Θ* : é a quantidade de pacotes retransmitidos pelo TCP devido a descartes ou devido a falta de resposta de recebimento por parte do destino (ack). O algoritmo de retransmissão rápida do TCP opera sobre a expectativa de resposta dentro de um certo intervalo de tempo estimado, e caso não haja resposta, é iniciada a retransmissão. Pacotes TCP em espera em uma fila muito grande podem ter os seus temporizadores esgotados antes mesmo de serem processados, o que pode acarretar no aumento da taxa de retransmissões.
- *Descartes β* : é a contagem de pacotes descartados diretamente na máquina escolhida para operar como gargalo. Neste caso em particular, ocorrem dado a falta de espaço na fila de armazenamento da máquina.

3.2.4 CENÁRIOS

Os cenários a seguir foram observados durante os experimentos. Para a maior confiabilidade dos resultados, cada cenário foi testado 5 vezes.

No cenário 1 é analisado o comportamento da fila de transmissão de forma exclusiva. Para essa análise foram testados os tamanhos de fila de 200, 1.000 e 100.000.000 pacotes. Para cada tamanho definido foi realizada uma cópia de arquivo de 512Mb do nó A para o nó E via *scp*¹³. Neste cenário, foram enviadas sondas ICMP de 1450 bytes em intervalos de 1s durante cópia, de B para F. Para o envio das sondas foi utilizado o programa *ping*¹⁴.

No cenário 2 é experimentado como a fila de transmissão se comporta durante um tráfego sem controle de congestionamento. Foram testados os tamanhos de fila de 200, 1.000 e 100.000.000 pacotes. Para cada tamanho de fila definido foram enviados *trens* de 200.000 pacotes, com cada pacote tendo 1450 bytes. Foi definido A como origem e E como destino. O intervalo entre cada *trem* é de 1s. Neste cenário foram enviadas sondas ICMP de 1450 bytes em intervalos de 1s durante cópia, de B para F.

No cenário 3 é analisado comportamento do *ringbuffer* de forma exclusiva. A fila de transmissão foi fixada no valor padrão de 1.000 pacotes. Os tamanhos do *ringbuffer* testados foram 500, 1.000, 5.000, 10.000 e 15.000. Para cada um desses tamanhos foi realizada uma cópia de arquivo de 512Mb do nó A para o nó E. As sondas ICMP de 1450 bytes em intervalos de 1s durante cópia foram enviadas de de B para F.

O cenário 4 explora o comportamento do *ringbuffer* juntamente com comportamento da fila de transmissão. Para o *ringbuffer* são testados valores discretos de 100 até 16.000, e para a fila de transmissão de 100 até 100.000.000. São realizadas 10 cópias de arquivo de 32Mb em concorrência do nó A para o nó E. Neste cenário foram enviadas sondas ICMP de 1450 bytes em intervalos de 1s durante cópia, de B para F.

3.3 RESULTADOS

Nesta seção são apresentados os resultados sobre a análise mais aprofundada dos efeitos do *bufferbloat* na experiência dos usuários da rede congestionada.

3.3.1 EXPERIMENTOS COM O QDISC

Inicialmente foi analisado o impacto que a fila de transmissão (qdisc, Figura 3.4) causa na latência. De acordo com a literatura, aumentos no *buffer* deveriam incorrer em aumentos na latência. Entretanto, no presente experimento, apenas a fila de transmissão foi

¹³Programa para cópia de arquivos pela rede.

¹⁴Programa que envia mensagens ICMP ECHO_REQUEST para hosts na rede.

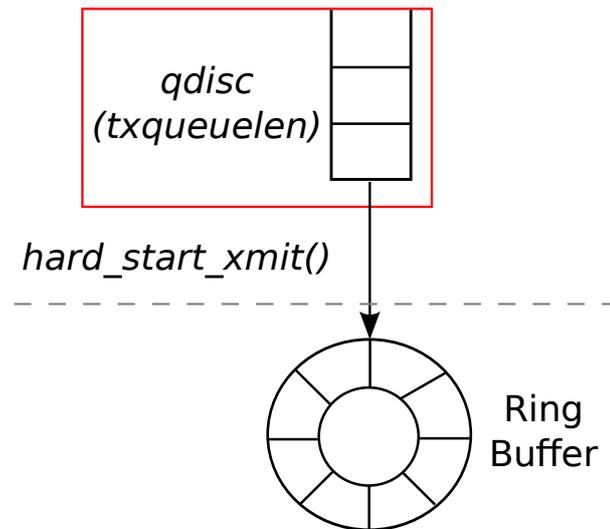


Figura 3.4: Destaque do *qdisc* na Figura 3.3.

analisada. Para esse primeiro experimento foi utilizado o cenário 1.

As figuras 3.5, 3.6 e 3.7 mostram o tamanho médio da fila em cada instante de tempo durante a cópia do arquivo. A média foi obtida a partir de 5 execuções do experimento. De acordo com a Figura 3.5, ao se utilizar um tamanho de fila 200 ($qdisc = 200$ pacotes), o tempo de ida e volta de um pacote (RTT) é de 317ms em média. O tamanho médio da fila ficou em torno de 118 pacotes.

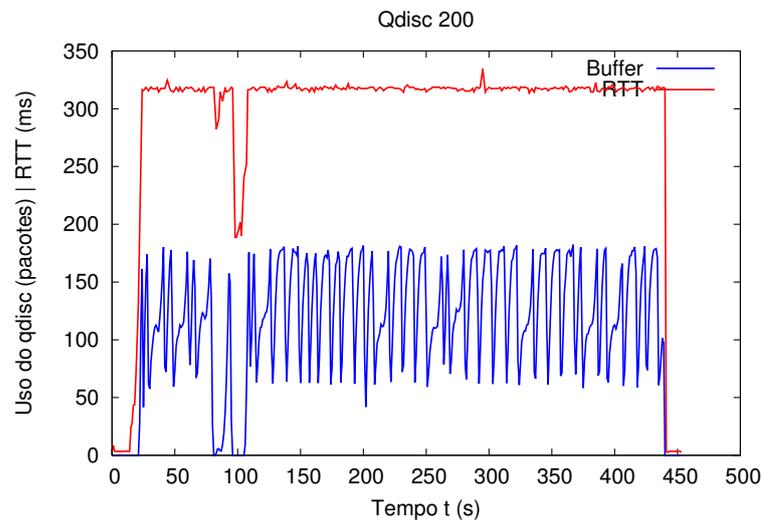


Figura 3.5: *qdisc* com tamanho 200.

Ao se aumentar a fila para um tamanho de 1.000 ($qdisc = 1.000$ pacotes), como mostrado na Figura 3.6, os atrasos médios não tiveram alteração significativa, apresentando valor médio de 316ms. A quantidade de pacotes em fila aumenta consideravelmente para 704 pacotes.

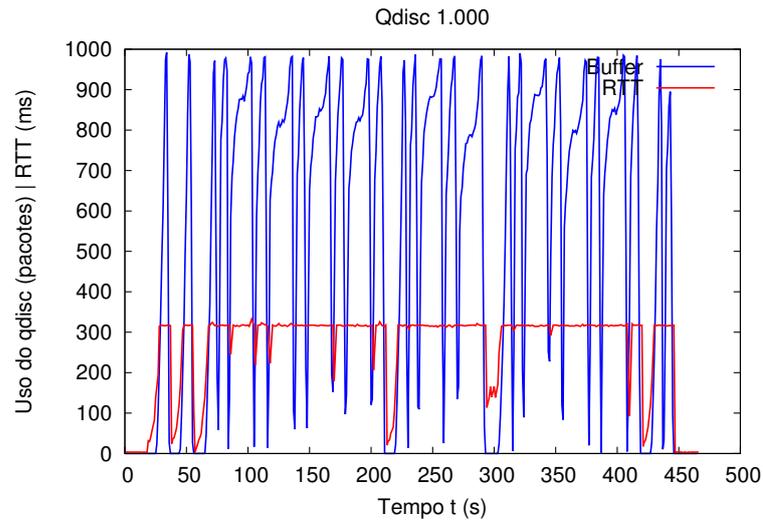


Figura 3.6: qdisc com tamanho 1.000.

No último caso analisado, apresentado na Figura 3.7, foi avaliado o cenário 1, com fila de tamanho 100 milhões de pacotes. O RTT médio teve um decremento de 14ms, sendo 301ms e o preenchimento médio da fila foi 1.221 pacotes.

O aumento do uso do *qdisc* no caso particular do cenário 1 causou uma pequena redução do RTT das sondas.

As médias das taxas de transferência observadas nos 3 experimentos anteriores não se alteram de forma significativa. Em todos os casos, o processo de cópia atingiu 1,1 Mbs em média, que é a capacidade de transmissão máxima do gargalo.

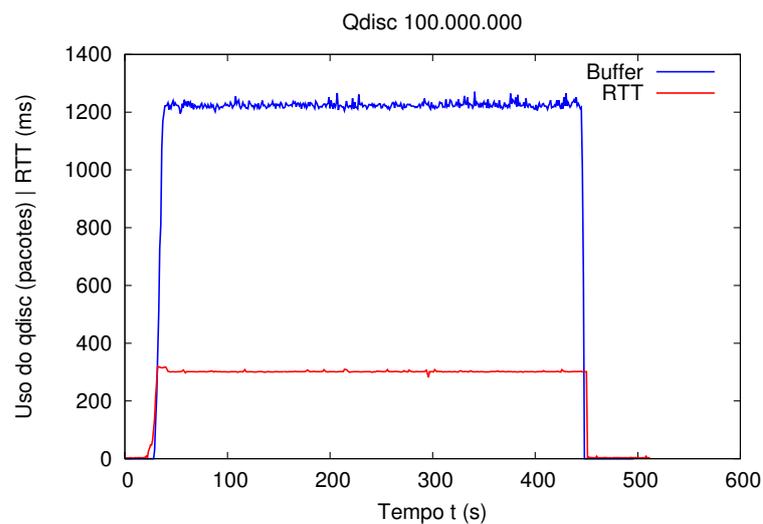


Figura 3.7: qdisc com tamanho 100.000.000.

Nesse mesmo cenário 1, também foram experimentadas sondas diferentes da camada

de transporte. Nesse sentido, espera-se avaliar o impacto causado pelo protocolo de transporte em si, ao invés das sondas ICMP. Foram utilizadas sondas TCP e depois UDP, com tamanho de 1450 bytes e enviadas em intervalos de 1 segundo. Para o envio das sondas TCP e UDP foi utilizada a ferramenta FLAME (ZIVIANI et al., 2012). Essa ferramenta permite a rápida prototipagem de ferramentas de medição ativa. Os resultados obtidos não apresentaram diferença aos testes com uso de sondas ICMP. Este resultado mostra que apenas uma fila de transmissão grande não necessariamente afetará a latência para um único fluxo TCP.

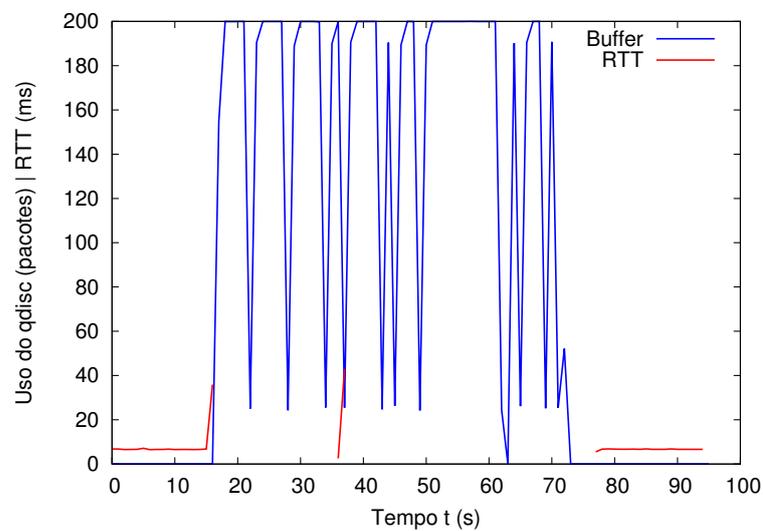


Figura 3.8: Trens de pacotes UDP com qdisc em 200.

Em seguida foi analisado como o *qdisc* afetaria o desempenho frente a uma carga de pacotes sem intervalo e qualquer tipo de controle de congestionamento. Para esse fim, foi utilizado o cenário 2, que envolve rajadas de pacotes UDP.

A Figura 3.8 mostra ausência de pontos em determinados pontos da curva azul (Buffer), próximo ao limite de 200. Significa que houve o preenchimento total da fila, que passou a descartar novos pacotes. Com os descartes, o RTT não pode ser verificado. Esse fenômeno pode ser visto pela intermitência da curva vermelha (RTT). No caso das figuras 3.9 e 3.10, ocorreu o fim do tempo de espera pela resposta das sondas, que é no máximo 5 segundos.

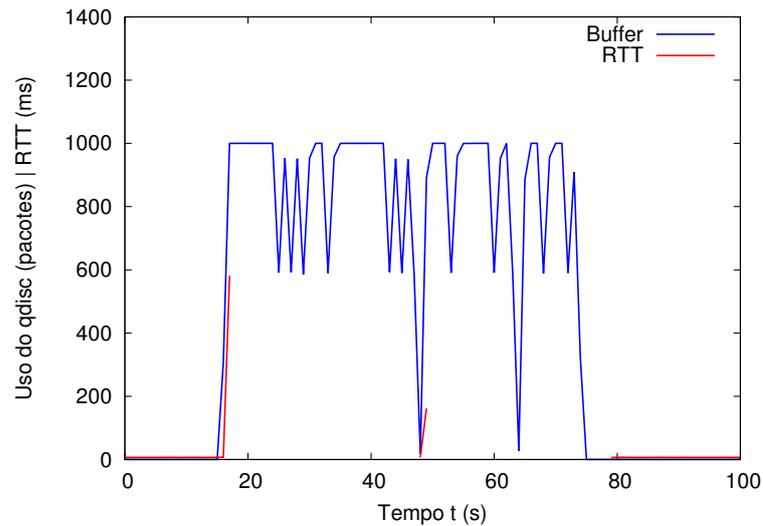


Figura 3.9: Trens de pacotes UDP com qdisc em 1.000.

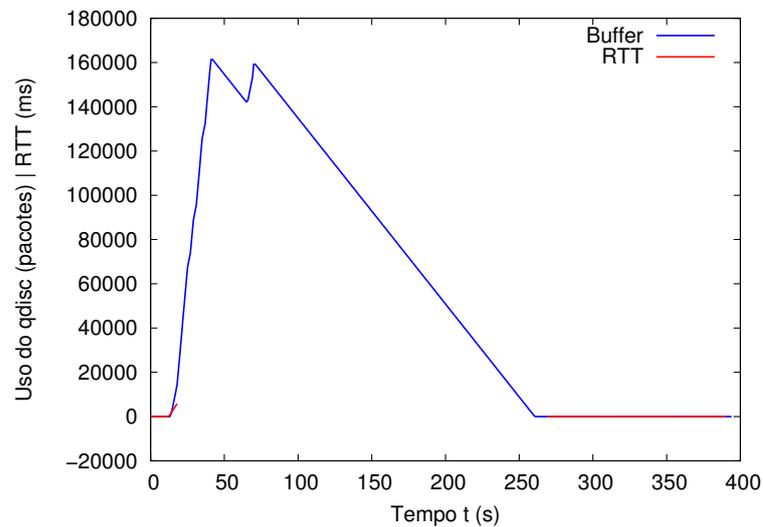


Figura 3.10: Trens de pacotes UDP com qdisc em 100.000.000.

3.3.2 EXPERIMENTOS COM O RINGBUFFER

Neste experimento foi analisado o impacto que alterações no tamanho do *ringbuffer* (3.11) causam na latência. O tamanho do *qdisc* foi fixado em 1.000 pacotes, que é o valor tamanho padrão para os sistemas Linux que compõem o *testbed*. Esse experimento foi realizado usando as configurações do cenário 3, que envolve o envio de sondas ICMP durante uma cópia de arquivo de 512Mb, com diferentes tamanhos para o *ringbuffer* no gargalo.

Na Figura 3.12, com o *ringbuffer* de tamanho 500 pacotes, o tempo de ida e volta de um pacote (RTT) ficou na média de 615ms.

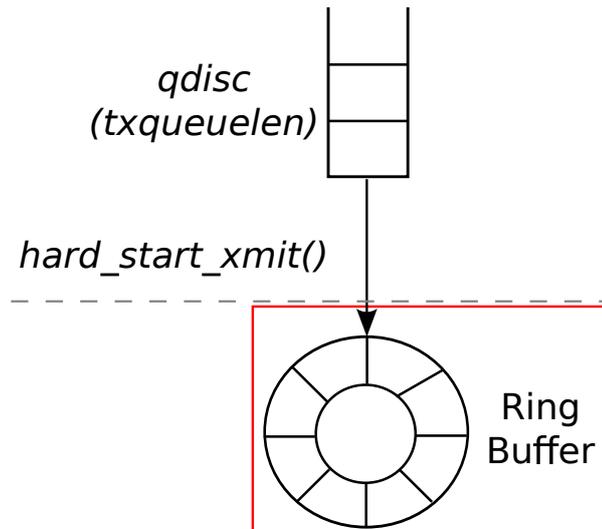


Figura 3.11: Destaque do *ringbuffer* na Figura 3.3.

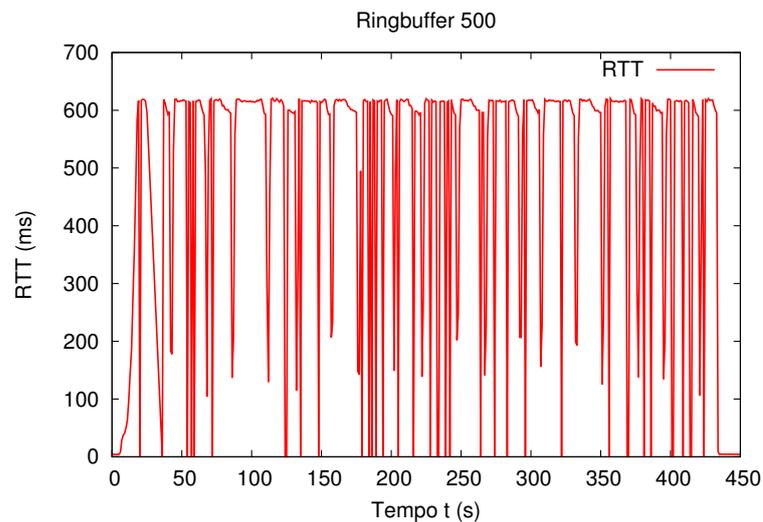


Figura 3.12: Ringbuffer com tamanho 500.

Aumentando o tamanho do *ringbuffer* para 1.000 pacotes, conforme mostrado na Figura 3.13, houve um aumento sustentado da média do RTT para 1.172ms.

O RTT atingiu o ponto máximo com tamanho do *ringbuffer* em 5.000 pacotes, chegando a média de 1.765ms, conforme a Figura 3.14. A fila com o tamanho de 15.000 pacotes apresentou a mesma média de tempo de ida e volta, como pode ser visto na Figura 3.15.

Estes resultados evidenciam que o tamanho do *ringbuffer* deve ser definido com muita cautela, uma vez que não é comum *drivers* de interface de rede implementarem algoritmos de gerenciamento ativo de fila. Normalmente as disciplinas de gerenciamento de fila são implementadas no *qdisc*, uma vez que a implementação do *driver* está mais associada ao

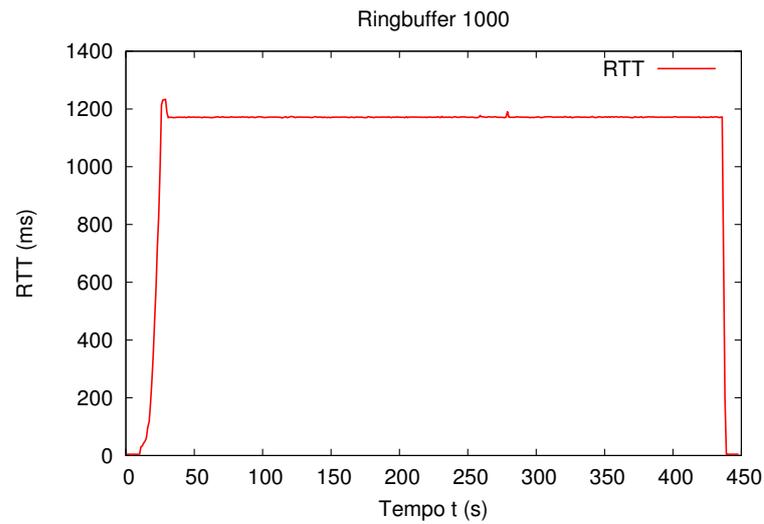


Figura 3.13: Ringbuffer com tamanho 1.000.

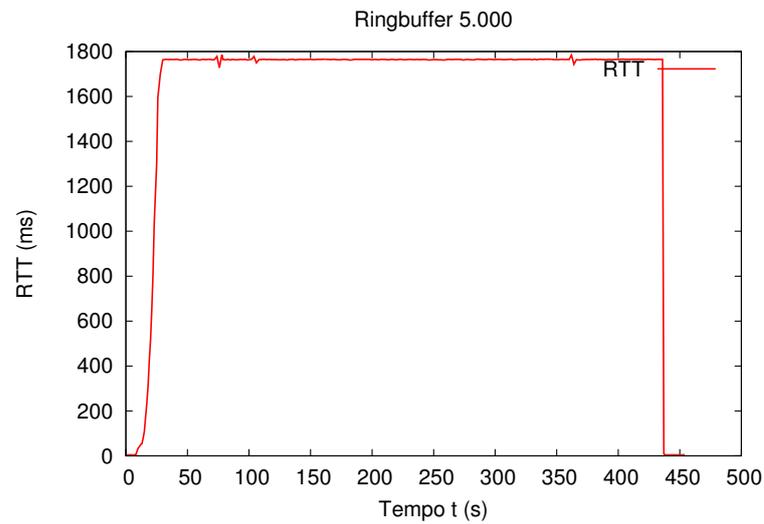


Figura 3.14: Ringbuffer com tamanho 5.000.

fabricante do que a gerência da rede.

3.3.3 EXPERIMENTOS COM O QDISC E RINGBUFFER

O próximo teste utiliza o cenário 4, que envolve 10 fluxos TCP em paralelo para maior saturação das filas. Neste experimento é observado mais a fundo o que acontece com várias configurações de tamanho tanto para o *qdisc* quanto para o *ringbuffer*. O esperado é que o TCP não consiga se adaptar ao rápido aumento da fila, uma vez que não ocorrerão descartes sinalizando congestionamento.

Dentre os detalhes observados, são analisados a quantidade de pacotes descartados e

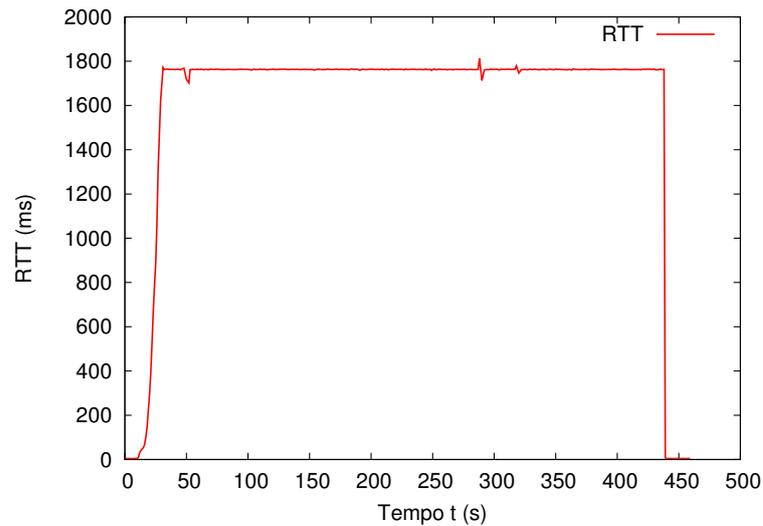


Figura 3.15: Ringbuffer com tamanho 15.000.

retransmitidos, a taxa de transferência e o RTT das sondas durante as cópias. As barras de erro nas figuras 3.16, 3.17, 3.18, 3.19 e 3.20 mostram o desvio padrão do item em questão. De acordo com a literatura (GETTYS; NICHOLS, 2012), os sintomas do *bufferbloat* são o grande aumento da latência, aumento das retransmissões, redução dos descartes e *time out* dos pacotes TCP.

Na Figura 3.16, com o *ringbuffer* com tamanho 100 pacotes, pode-se ver que houve mais descartes e retransmissões quando o tamanho do *qdisc* estava em 2.500 pacotes. Com o *qdisc* em 2.500 pacotes a média dos descartes ficou em 7.440 pacotes, e a média de retransmissões foi 7.421 pacotes. O tempo de ida e volta dos pacotes (RTT) ficou na média de 126ms. Neste caso as barras do intervalo de confiança das colunas do RTT estão sobre a média.

Aumentando o *ringbuffer* para 350 pacotes, como mostrado na Figura 3.17, resultou igualmente em mais descartes e retransmissões quando o tamanho do *qdisc* estava em 2.500 pacotes. A média dos descartes para o *qdisc* em 2.500 foi 7.479 pacotes, e a média de retransmissões ficou em 7.466 pacotes. O tempo médio de ida e volta dos pacotes (RTT) ficou na média de 417ms.

Com o *ringbuffer* com tamanho em 1.700 pacotes, novamente os picos de retransmissões e descartes ocorreram quando o *qdisc* estava em 2.500 pacotes, como mostrado na Figura 3.18. Para o *qdisc* em 2.500 pacotes, a média de descartes e retransmissões foram 8.259 e 8.244 pacotes respectivamente, e o RTT dos pacotes ficou na média de 2.089ms.

Aumentando novamente o *ringbuffer*, agora com o tamanho de 3.850 pacotes, o *qdisc*

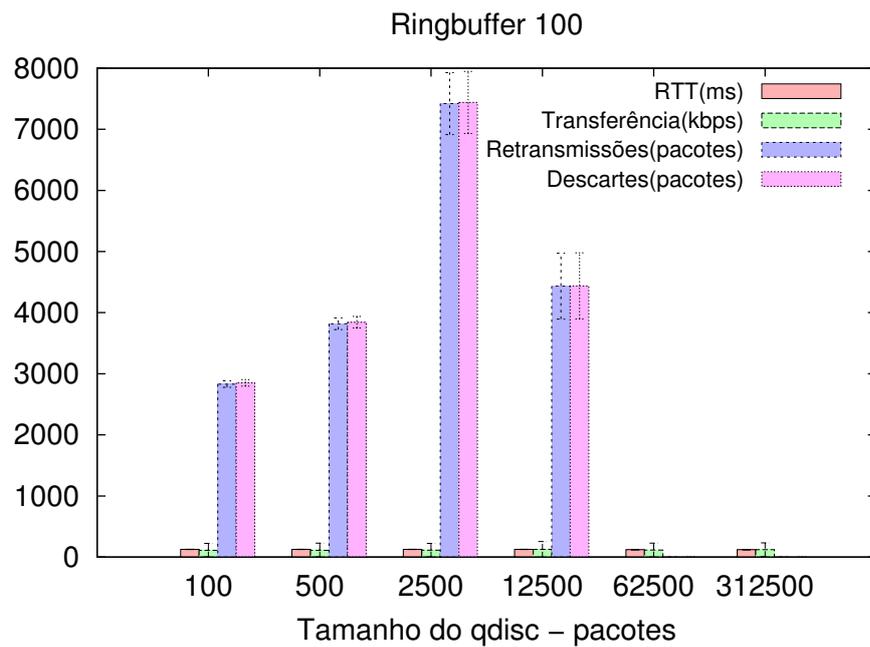


Figura 3.16: Ringbuffer com tamanho 100.

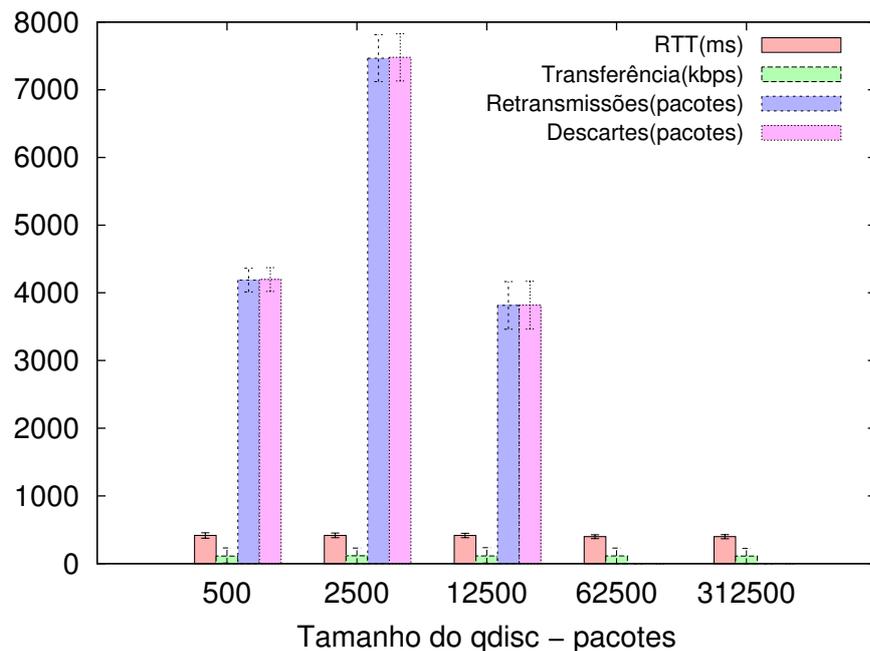


Figura 3.17: Ringbuffer com tamanho 350.

com tamanho de 2.500 pacotes deixou de apresentar os picos de retransmissões e descartes. Além disso, o número de descartes deixou de ser equivalente ao de retransmissões, como pode ser visto na Figura 3.19. Para o caso do *qdisc* em 2.500 pacotes, a diferença passou a ser de 7.627 pacotes. O RTT médio para esse caso foi 2.539ms.

O último tamanho testado para o *ringbuffer* foi o de 13.000 pacotes, conforme a Figura 3.20. O ponto mais importante a ser notado é a ausência de descartes para todos os

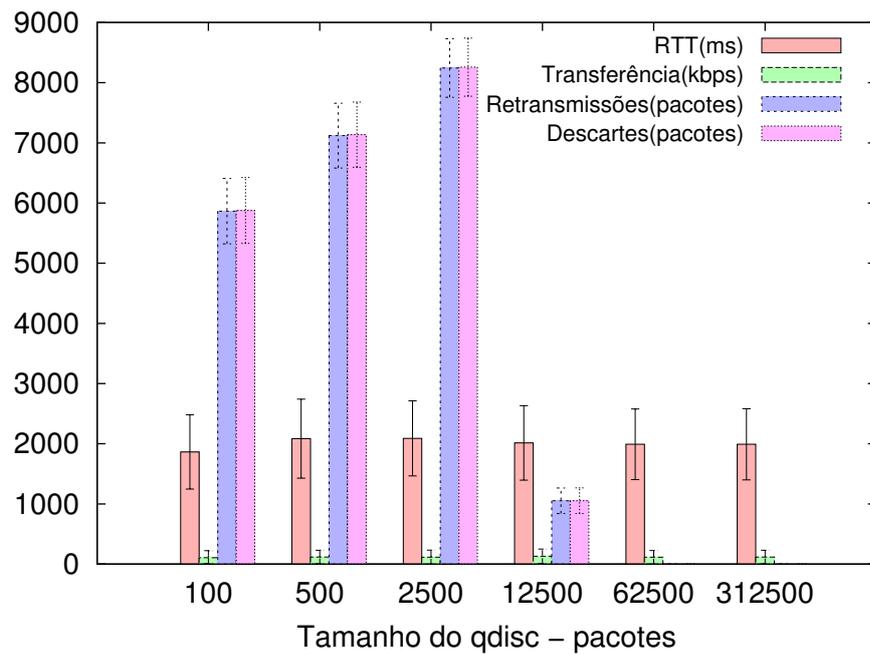


Figura 3.18: Ringbuffer com tamanho 1.700.

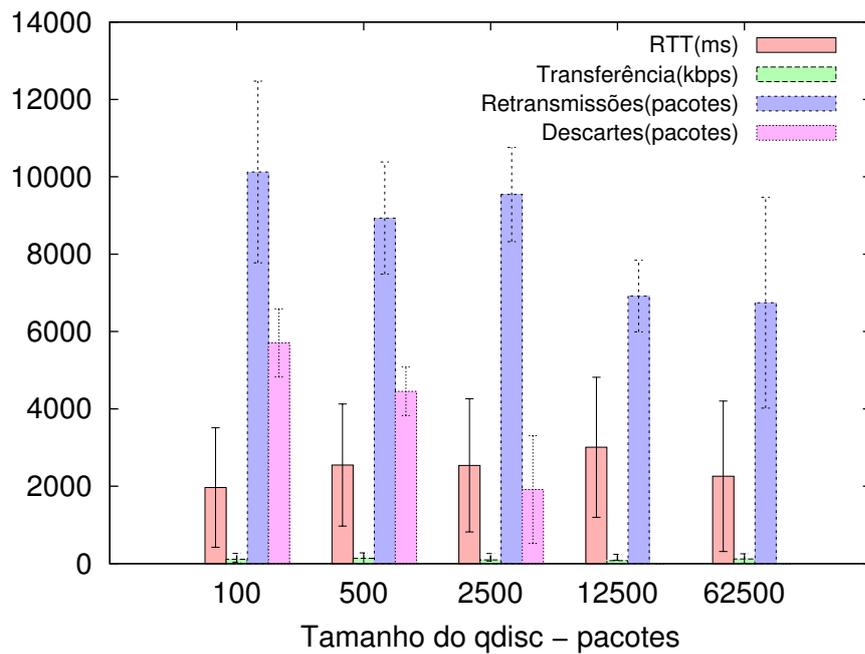


Figura 3.19: Ringbuffer com tamanho 3.850.

tamanhos testados para o *qdisc*. Nota-se também uma grande oscilação nas retransmissões, causadas pelo algoritmo de retransmissão rápida do TCP, que acaba por reenviar os pacotes que estão tendo o seu tempo de vida esgotado enquanto esperam nas filas. Isso acarreta em uma grande carga de tráfego desnecessário. Para o *qdisc* com tamanho 2.500 pacotes, a média de retransmissões ficou em 11.068, e o RTT médio em 2.428ms.

O alto desvio padrão observado está relacionado ao comportamento do sistema durante

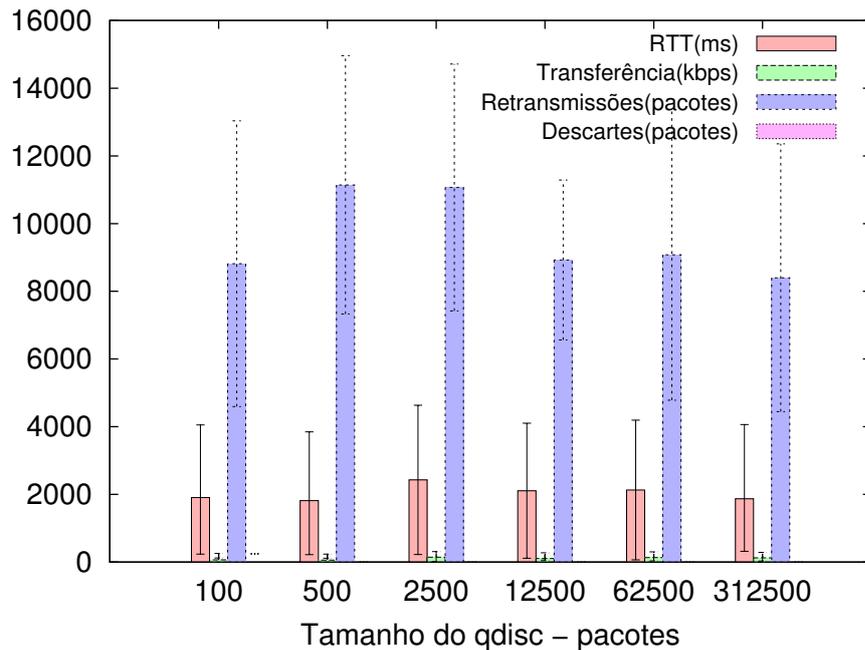


Figura 3.20: Ringbuffer com tamanho 13.000.

a cópia do arquivo. Por exemplo, o RTT em um primeiro instante começa baixo pois as cópias ainda estão se ajustando. Depois o RTT vai aumentando com o passar do tempo, e então vai gradualmente diminuindo durante o estágio final do experimento, quando as cópias estão terminando. Outro motivo para o alto desvio padrão é que taxa de transferência dos arquivos não foi distribuída igualmente entre todas as sessões *scp*. Ao invés disso, algumas das cópias foram momentaneamente paradas, e a taxa de transferência das cópias remanescentes foi aumentada. A medida que as cópias foram sendo completadas, as que tinham sido interrompidas continuaram, mas como havia menos tarefas em paralelo as taxas foram incrementadas, o que compensou o tempo de pausa. Por isso a taxa de transferência média é praticamente a mesma.

Quando o *ringbuffer* começou a ser testado com tamanhos acima de 1.000 pacotes, observou-se que ocorreu uma grande queda do preenchimento do *qdisc*. Isso está relacionado ao fato de que os pacotes não precisam mais esperar para passar da camada de rede para a camada de enlace.

3.4 CONCLUSÃO DO CAPÍTULO

Neste capítulo foram apresentados os experimentos relacionados ao fenômeno conhecido como *bufferbloat* (GETTYS; NICHOLS, 2012), motivados pelos resultados apresentados

no capítulo 2, que indicam grande aumento da latência na Internet nos últimos anos.

Esses experimentos consistiram na observação da latência no ambiente de teste durante várias cópias de arquivos utilizando vários tamanhos de fila. As filas trabalhadas foram o *qdisc* e o *ringbuffer*, que são duas filas diretamente relacionadas, estando respectivamente uma na camada de rede e a outra na interface com camada de enlace.

Os resultados mostraram que o *bufferbloat* é um problema real e pode afetar a experiência final do usuário. Foi observado que filas muito grandes afetam diretamente aplicações que utilizam de pequenas mensagens em intervalos de tempo, assim como é o caso das sondas utilizadas. Com isso, a navegação web, serviços de DNS, SSH, Telnet, NTP, jogos e outros se tornam inviáveis durante um fluxo de dados contínuo como uma simples cópia de arquivo pela rede.

Outro ponto a ser notado é que o impacto do *ringbuffer* é muito maior que o do *qdisc*, onde mesmo que o *qdisc* seja marcado como 0, um *ringbuffer* alto pode acarretar em enormes latências caso seja configurando com um valor muito grande. Uma questão importante mas de difícil análise é o fato das interfaces de rede (NIC) possuírem memória interna muitas vezes de tamanhos desconhecidos pelo usuário.

O efeito do *bufferbloat* não ocorreu naturalmente com as configurações padrão das máquinas. Para que o problema ocorra, as filas devem ser intencionalmente configuradas em tamanhos altos.

Definitivamente, filas grandes acarretam em problemas de desempenho, contudo sistemas diferentes precisam de configurações diferentes, não apenas por causa do comportamento do ambiente, mas por questões de capacidade de *hardware*. O nosso estudo se limita a uma pequena rede semelhante a uma LAN de uma empresa, longe da realidade dos roteadores dos provedores de acesso e dos *backbones*. Para esses casos são aconselhados o uso de algoritmos de gerenciamento ativo de filas, como o proposto em (NICHOLS; JACOBSON, 2012). Isso não significa que com o uso de AQMs o tamanho das filas possa ser completamente negligenciado. O dimensionamento correto das filas levando em consideração não apenas as necessidades da rede, mas também questões de orçamento ainda é algo difícil de ser realizado dada a grande quantidade de combinações desses 2 elementos. Dentre os trabalhos futuros, temos justamente a intenção de analisar tamanhos de fila mais próximos do ideal para diferentes tipos de sistemas.

4 CONCLUSÃO

Neste trabalho foi caracterizada e analisada a recente evolução da conectividade fim-a-fim da Internet. Foram comparadas métricas de desempenho importantes entre dois períodos distintos separados por cinco anos. Foram analisados dados de 2006 e 2011, incluindo a latência dos caminhos fim-a-fim, número de saltos, dispersão geográfica e diversidade de caminhos. Entender essa recente evolução da conectividade da Internet, assim como as tendências associadas, é importante, por exemplo, para o planejamento de enlaces inter-rede e para um melhor ajuste de protocolos de roteamento.

Os resultados da análise da evolução da conectividade fim-a-fim da Internet no capítulo 2 mostraram que, ao contrário do que era esperado, o desempenho da Internet, principalmente em termos de atraso, tem aparentemente piorado nos últimos anos.

Resultados anteriores encontrados na literatura indicam que a rede está se tornando mais densa. Duas consequências esperadas a partir desse resultado são a maior diversidade de caminhos possíveis entre dois pontos da rede, e a redução do número de saltos desses caminhos. Estas características tornam a rede mais robusta e resiliente a eventuais falhas que venham a ocorrer. Contudo, os resultados encontrados indicam que os caminhos entre dois pontos na rede estão se tornando mais parecidos entre si, o que indica que a diversidade da rede não está sendo propriamente explorada. Em 2006 a diferença entre os diversos caminhos disponíveis entre dois pontos era de 63%, e em 2011 a diferença caiu para 54%.

Outros resultados também indicam que os caminhos entre dois pontos não estão diminuindo de tamanho. Esses dois resultados mostram que está ocorrendo uma possível priorização de rotas em certos roteadores. A característica mais importante encontrada foi o fato de a latência da Internet estar aumentado recentemente. Houve de 2006 para 2011 um aumento de 45% na latência dos caminhos. Há diversas causas possíveis para essa ocorrência, e acreditamos que não seja um caso em particular, mas uma combinação desses eventos prejudiciais. Dentre as causas que poderiam acarretar em um grande aumento de atraso, analisamos o aumento não gerenciado das filas dos roteadores, que causa o fenômeno conhecido como *bufferbloat*.

Para estudar como o fenômeno do *bufferbloat* poderia impactar a latência em uma rede,

foram realizados diversos experimentos no capítulo 3. Esses experimentos consistiram na observação do comportamento de um roteador em diferentes cenários. Em cada um desses cenários foram testados diferentes tamanhos de fila para duas filas, uma situada na camada de rede e outra na camada de enlace. O estudo foi limitado a essas camadas porque normalmente os roteadores não implementam as camadas acima da de redes.

Os resultados da observação desses cenários mostraram que o *bufferbloat* pode tornar a experiência do usuário final muito desconfortável devido ao grande aumento da latência. Atividades do dia a dia como a navegação *web* podem ficar comprometidas se realizadas com uma configuração errônea dos roteadores. Não apenas o usuário final sofre com o problema, mas a infraestrutura geral de serviços da Internet encontra problemas, principalmente aqueles serviços sensíveis a atraso e de troca periódica de mensagens. Um ponto importante é que o *testbed*, que é composto por *desktops*, não apresentou nenhum problema com as *configurações de fábrica*. O fenômeno do *bufferbloat* teve que ser induzido via uma série de configurações. Com isso, o problema não parece se encontrar nas máquinas finais dos usuários, uma vez que os tamanhos de fila padrão parecem estar adequados. O problema pode estar ocorrendo em roteadores caseiros ou de pequeno porte, localizados entre a máquina do usuário e o provedor de acesso.

4.1 TRABALHOS FUTUROS

Como trabalhos futuros, há outros itens a serem explorados mais a fundo:

- É de interesse a caracterização da evolução da conectividade fim-a-fim da Internet nos períodos anteriores aos estudados, para mostrar se os problemas encontrados são mais antigos ou se são mais recentes.
- Outros fatores associados ao aumento de latência que serão estudados futuramente:
 - Virtualização.
 - Roteamento por software.
 - Influência de políticas de roteamento.
 - Influência do tráfego multimídia e P2P.

- Também serão analisados outros cenários que envolvem diferentes sistemas operacionais, diferentes dispositivos de roteamento e a influência dos algoritmos de gerenciamento de fila.

REFERÊNCIAS

- AGER, B.; CHATZIS, N.; FELDMANN, A.; SARRAR, N.; UHLIG, S.; WILLINGER, W. Anatomy of a large european IXP. In: **Proceedings of the ACM SIGCOMM 2012 conference on Applications, technologies, architectures, and protocols for computer communication**, 2012. (SIGCOMM '12), p. 163–174. ISBN 978-1-4503-1419-0. Disponível em: <<http://doi.acm.org/10.1145/2342356.2342393>>.
- ALLMAN, M. Comments on bufferbloat. **SIGCOMM Comput. Commun. Rev.**, ACM, New York, NY, USA, v. 43, n. 1, p. 30–37, jan. 2012. ISSN 0146-4833. Disponível em: <<http://doi.acm.org/10.1145/2427036.2427041>>.
- CHIRICHELLA, C.; ROSSI, D. To the Moon and back: are Internet bufferbloat delays really that large? **IEEE INFOCOM Workshop on Traffic Measurement and Analysis**, p. 14–19, 2013. Disponível em: <<http://perso.telecom-paristech.fr/drossi/paper/rossi13tma-a.pdf>>.
- CROVELLA, M.; KRISHNAMURTHY, B. **Internet Measurement: Infrastructure, Traffic and Applications**. New York, NY, USA: John Wiley & Sons, Inc., 2006. ISBN 047001461X.
- DHAMDHARE, A.; DOVROLIS, C. Twelve years in the evolution of the internet ecosystem. **IEEE/ACM Transactions on Networking (TON)**, v. 19, n. 5, p. 1420–1433, 2011.
- FLOYD, S.; JACOBSON, V. Random early detection gateways for congestion avoidance. **IEEE/ACM Trans. Netw.**, IEEE Press, Piscataway, NJ, USA, v. 1, n. 4, p. 397–413, ago. 1993. ISSN 1063-6692. Disponível em: <<http://dx.doi.org/10.1109/90.251892>>.
- GETTYS, J.; NICHOLS, K. Bufferbloat: Dark buffers in the Internet. **Communications of the ACM**, v. 55, n. 1, p. 57–65, 2012.
- GIJSBERS, B.; AKKOORATH, D. D. **Performance simulation of buffer bloat in routers**. Technical report, Universiteit van Amsterdam, mar. 2011.

HADDADI, H.; UHLIG, S.; MOORE, A.; MORTIER, R.; RIO, M. Modeling internet topology dynamics. **SIGCOMM Comput. Commun. Rev.**, ACM, New York, NY, USA, v. 38, n. 2, p. 65–68, mar. 2008. ISSN 0146-4833. Disponível em: <<http://doi.acm.org/10.1145/1355734.1355745>>.

JACOBSON, V. Congestion avoidance and control. **SIGCOMM Comput. Commun. Rev.**, ACM, New York, NY, USA, v. 18, n. 4, p. 314–329, ago. 1988. ISSN 0146-4833. Disponível em: <<http://doi.acm.org/10.1145/52325.52356>>.

JAIN, R.; ROUTHIER, S. Packet Trains—Measurements and a New Model for Computer Network Traffic. **Selected Areas in Communications, IEEE Journal on**, v. 4, n. 6, p. 986–995, 1986. ISSN 0733-8716.

JIANG, H.; LIU, Z.; WANG, Y.; LEE, K.; RHEE, I. Understanding bufferbloat in cellular networks. In: **Proceedings of the 2012 ACM SIGCOMM workshop on Cellular networks: operations, challenges, and future design**, 2012. (CellNet '12), p. 1–6. ISBN 978-1-4503-1475-6. Disponível em: <<http://doi.acm.org/10.1145/2342468.2342470>>.

JIANG, H.; WANG, Y.; LEE, K.; RHEE, I. Tackling bufferbloat in 3G/4G networks. In: **Proceedings of the 2012 ACM conference on Internet measurement conference**, 2012. (IMC '12), p. 329–342. ISBN 978-1-4503-1705-4. Disponível em: <<http://doi.acm.org/10.1145/2398776.2398810>>.

KENNEY, J. F.; KEEPING, E. S. Linear Regression and Correlation. In: **Mathematics of Statistics**. 3. ed., 1962. v. 1, cap. 15, p. 252–285.

LABOVITZ, C.; IEKEL-JOHNSON, S.; MCPHERSON, D.; OBERHEIDE, J.; JAHANIAN, F. Internet inter-domain traffic. **SIGCOMM Comput. Commun. Rev.**, ACM, New York, NY, USA, v. 40, n. 4, p. 75–86, ago. 2010. ISSN 0146-4833. Disponível em: <<http://doi.acm.org/10.1145/1851275.1851194>>.

LEE, D.; CHO, K.; IANNACCONE, G.; MOON, S. Has Internet Delay gotten Better or Worse? In: **Proc. of the 5th International Conference on Future Internet Technologies (CFI)**, 2010.

- LEE, D. K.; JANG, K.; LEE, C.; IANNACCONE, G.; MOON, S. Path stitching: internet-wide path and delay estimation from existing measurements. In: **Proceedings of the 29th conference on Information communications**, 2010. (INFOCOM'10), p. 201–205. ISBN 978-1-4244-5836-3. Disponível em: <<http://dl.acm.org/citation.cfm?id=1833515.1833556>>.
- LEVENSHTTEIN, V. Binary codes capable of correcting deletions, insertions, and reversals. **Soviet Physics Doklady**, v. 10, p. 707–710, 1966.
- NAGLE, J. On packet switches with infinite storage. RFC Editor, United States, 1985.
- NICHOLS, K.; JACOBSON, V. Controlling Queue Delay. **Queue**, ACM, New York, NY, USA, v. 10, n. 5, p. 20:20–20:34, maio 2012. ISSN 1542-7730. Disponível em: <<http://doi.acm.org/10.1145/2208917.2209336>>.
- OLIVEIRA, R.; ZHANG, B.; ZHANG, L. Observing the evolution of Internet AS topology. **ACM SIGCOMM Computer Communication Review**, v. 37, n. 4, p. 313–324, 2007.
- PERCACCI, R.; VESPIGNANI, A. Scale-free Behavior of the Internet Global Performance. **The European Physical Journal B**, v. 32, p. 411–414, 2003.
- SCHWARTZ, Y.; SHAVITT, Y.; WEINSBERG, U. On the Diversity, Stability and Symmetry of End-to-End Internet Routes. In: **Proc. of the Global Internet (GI)**, 2010.
- SHAVITT, Y.; SHIR, E. DIMES: let the Internet measure itself. **ACM SIGCOMM Computer Communication Review**, v. 35, n. 5, p. 71–74, out. 2005. ISSN 0146-4833.
- SIGANOS, G.; FALOUTSOS, M.; FALOUTSOS, C. **The Evolution of the Internet: Topology and Routing**. Technical report, Univ. of California, Riverside, 2002.
- SPEARMAN, C. The proof and measurement of association between two things. **American Journal of Psychology**, v. 15, p. 72–101, 1904.
- WAGNER, R. A.; LOWRANCE, R. An Extension of the String-to-String Correction Problem. **Journal of the ACM**, ACM, New York, NY, USA, v. 22, n. 2, p. 177–183, abr. 1975. ISSN 0004-5411. Disponível em: <<http://doi.acm.org/10.1145/321879.321880>>.

- WANG, F.; GAO, L. Path Diversity Aware Interdomain Routing. In: **IEEE INFOCOM**, 2009. p. 307–315.
- WHITEAKER, J.; SCHNEIDER, F.; TEIXEIRA, R. Explaining Packet Delays Under Virtualization. **ACM SIGCOMM Computer Communication Review**, v. 41, n. 1, p. 38–44, 2011.
- ZHOU, S. Understanding the Evolution Dynamics of Internet Topology. **Physical Review E**, v. 74, n. 1, p. 016124, 2006.
- ZIVIANI, A.; CARDOZO, T. B.; GOMES, A. T. A. Rapid prototyping of active measurement tools. **Comput. Netw.**, Elsevier North-Holland, Inc., New York, NY, USA, v. 56, n. 2, p. 870–883, fev. 2012. ISSN 1389-1286. Disponível em: <<http://dx.doi.org/10.1016/j.comnet.2011.11.008>>.