

Evolução da conectividade fim-a-fim da Internet

Thiago B. Cardozo¹, Ana Paula C. Silva², Alex B. Vieira¹, Artur Ziviani³

Departamento de Ciência da Computação
Universidade Federal de Juiz de Fora (UFJF)
Juiz de Fora – MG – Brasil

²Departamento de Ciência da Computação
Universidade Federal de Minas Gerais (UFMG)
Belo Horizonte – MG – Brasil

³Laboratório Nacional de Computação Científica (LNCC)
Petrópolis – RJ – Brasil

thiago.boubee@ice.ufjf.br, ana.coutosilva@dcc.ufmg.br

alex.borges@ufjf.edu.br, ziviani@lncc.br

Resumo. A Internet é um sistema em contínua evolução. Neste artigo, 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.

Abstract. The Internet is a system under continuous evolution. In this paper, we characterize and analyze 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. Our 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, we show 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.

1. Introdução

A Internet atualmente tem um papel crucial na nossa infraestrutura de comunicação e colaboração. 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., 2010]. 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. Portanto, para o melhor desenvolvimento de modelos de rede e sistemas, é importante entender o comportamento da Internet e analisar tendências, em particular como métricas relevantes mudam ao longo do tempo.

Neste artigo, 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 (AS) [Dhamdhere e 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 fim-a-fim, o número de saltos dos caminhos e a dispersão geográfica dos caminhos. Além disso, é utilizado um método baseado na distância de Levenshtein [Levenshtein, 1966] (também conhecida com distância de edição [Wagner e Lowrance, 1975]) para caracterizar a diversidade e medir o quão diferentes são os caminhos nos dois períodos.

Há vários estudos sobre a modelagem e medição da topologia da Internet [Haddadi et al., 2008, Oliveira et al., 2007, Zhou, 2006]. 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 AS, ignorando relações 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.

Os resultados mostram que aspectos cruciais da Internet estão piorando nos últimos anos. A média da distribuição do tamanho dos caminhos se mantém quase idêntica de 2006 para 2011, porém a distribuição do atraso tornou-se pior. Nesse caso, é observado um aumento de 45% no atraso dos caminhos de 2006 para 2011. Isso afeta diretamente o desempenho da rede e degrada a experiência do usuário. Adicionalmente, a diversidade dos caminhos diminuiu e, com isso, caminhos distintos se tornaram um pouco mais similares. Dois caminhos distintos para o mesmo ponto final em 2006 costumavam ser 63% diferentes. Em 2011 essa diferença caiu para 54%. Esse resultado tem um impacto direto em algoritmos de roteamento que tentam explorar a diversidade de caminhos para uma maior tolerância a falhas. De forma geral, este estudo ajuda na compreensão da recente evolução da conectividade da Internet e a consequente experiência do usuário.

O restante do artigo está organizado como segue. A Seção 2 introduz os conjuntos de dados utilizados e a metodologia de amostragem adotada. Na Seção 3, são apresentadas as métricas avaliadas e os resultados obtidos. A Seção 4 brevemente revê os trabalhos relacionados. Finalmente, a Seção 5 conclui o trabalho e discute trabalhos futuros.

2. Amostragem dos caminhos da Internet

Esta seção descreve os conjuntos de dados utilizados e a metodologia de amostragem adotada para a nossa análise.

2.1. Conjunto de dados

Este trabalho conta com um conjunto de dados coletados pelo projeto CAIDA¹. Esses dados registram resultados como os do traceroute 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 já que a maioria dos monitores de 2006 se encontravam localizados nesses continentes.

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 com uma representativa distribuição do atraso na Internet e estão entre as maiores coleções disponíveis, sendo atualizados continuamente.

Foram adquiridos todos os rastros da rede (*network traces*) do projeto CAIDA realizados em 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. 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.

2.2. Metodologia de amostragem

Antes da utilização dos dados, realizamos um processamento para a remoção de inconsistências. Neste trabalho, defini-se caminhos inconsistentes aqueles que estão incompletos e/ou com mais de trinta saltos e *loops*. 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. Por exemplo, são caracterizados o número de saltos e o atraso de cada caminho. Tais métricas são discutidas mais adiante na Seção 3.1. 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. Lee et al. [Lee et al., 2010] 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%.

¹www.caida.org/

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 : o grafo \mathcal{G}_{2006} tem 172.926 nós e 459.498 arestas; e o grafo \mathcal{G}_{2011} tem 162.119 nós e 448.614 arestas. Após a construção dos grafos, verificou-se que ambos são conexos.

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

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.

3. Evolução da conectividade

Nesta seção, são revisadas as métricas usadas para a caracterização realizada (Seção 3.1). Também são relatados os resultados encontrados durante a análise de um intervalo de cinco anos na conectividade fim-a-fim da Internet (Seção 3.2).

3.1. 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. No que se segue, é formalizado o conjunto de métricas selecionadas para a análise da evolução da conectividade da Internet:

- *Tamanho do caminho* (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 nossa 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* é o tempo de ida e volta (RTT) entre dois *hosts*. Um grande atraso pode indicar 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* é uma métrica que reflete o número de rotas alternativas disponíveis entre um par origem-destino. A diversidade de caminhos per-

mite 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} .

- *Diferença entre caminhos*: a diferença entre dois caminhos que ligam dois pontos dados é quantificada usando a distância Levenshtein (distância de edição) [Levenshtein, 1966, Wagner e Lowrance, 1975]. Em outras palavras, a diferença entre os caminhos é a quantidade mínima de operações necessárias para transformar um caminho em outro.
- *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.

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

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}) também são usados para analisar a dispersão geográfica dos caminhos.

3.2. Resultados

Nesta seção, são analisados os resultados da caracterização estudada usando as métricas apresentadas na Seção 3.1.

3.2.1. Tamanho do caminho

Como primeiro conjunto de resultados, foi analisado tamanho dos caminhos entre dois extremos da amostra de $n = 200$ milhões de rastros de rota da CAIDA de 2006 para 2011. A Figura 1 mostra que a função de distribuição acumulada (*cumulative distribution function* – *CDF*) 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 um aprimoramento marginal de um salto. Em ambos os casos, 80% dos caminhos têm cerca de 17 saltos. Resultados similares foram encontrados nas redes em termos de AS [Dhamdhare e Dovrolis, 2008].

Como a rede cresce e o comprimento dos caminhos se mantém 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, já que devem haver mais caminhos disponíveis entre dois extremos. De forma similar, em uma rede mais densa, intuitivamente, seriam esperados atrasos menores. Entretanto, isso não é comprovado experimentalmente com atrasos em realidade se tornando maiores entre 2006 e 2011 de maneira contra-intuitiva ao fato da rede estar se tornando mais densa, como será discutido em mais detalhes na Seção 3.2.2.

Também foi analisado o tamanho de todos os K -menores caminhos da visão parcial da rede. A Figura 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.

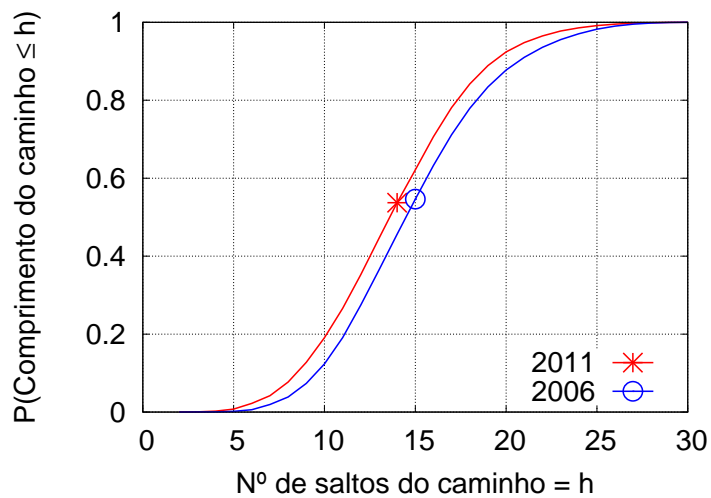


Figura 1. CDF - número de saltos.

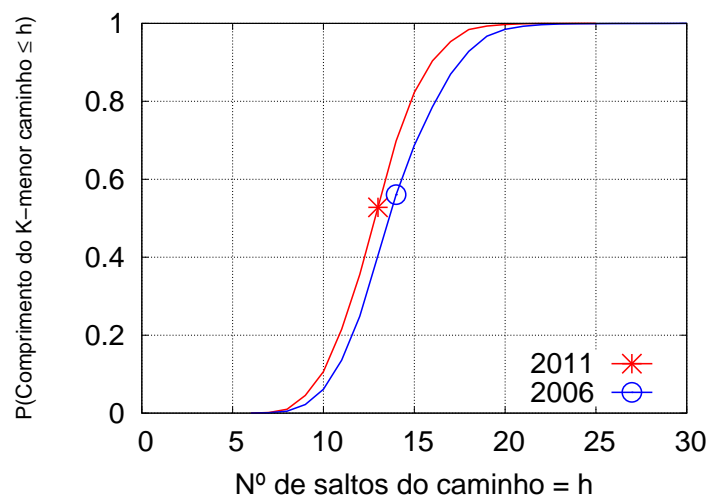


Figura 2. CDF - Número de saltos dos K-menores caminhos.

Na Figura 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 e o comprimento do caminho se mantém estável de acordo com o crescimento do K até ~ 10 , com os caminhos de 2011 sendo um pouco menores que os de 2006.

3.2.2. Atraso dos caminhos

A Figura 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 122 ms , enquanto em 2011 a mediana do atraso cresceu cerca de 37% para quase 168 ms . Considerando o percentil 95%, o atraso aumentou de cerca de 352 ms em 2006 para cerca de 512 ms em 2011 (um aumento de 45% em cinco anos). Esse é um resultado

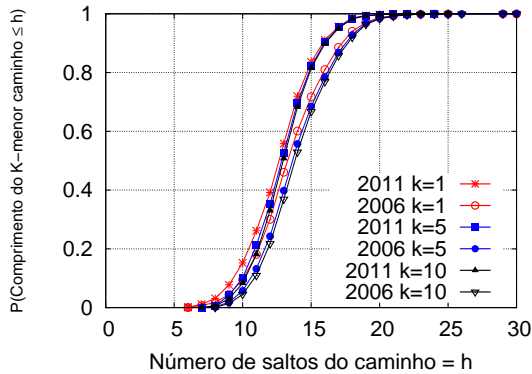


Figura 3. CDF - Comprimento dos K-menores caminhos.

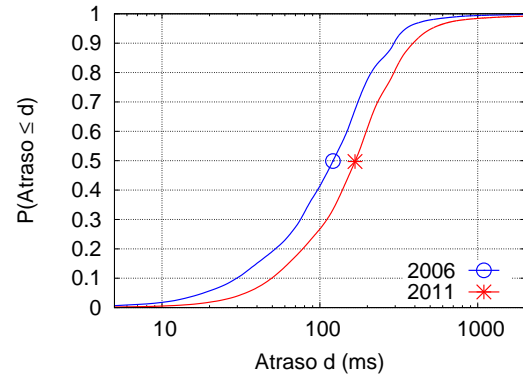


Figura 4. CDF - Atraso dos caminhos.

contra-intuitivo aos estudos recentes que apontam a maior densidade da rede nos últimos anos [Labovitz et al., 2010, Ager et al., 2012].

Os resultados aqui discutidos reforçam que o desempenho da Internet está se tornando pior ultimamente, como mostrado em [Lee et al., 2010]. Uma possível explicação pode ser o fenômeno conhecido como *bufferbloat*, recentemente relatado por Gettys e Nichols [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 uso pesado da rede em máquinas virtuais pode introduzir atrasos de até 100 ms ao RTT [Whiteaker et al., 2011]. Como consequência, a latência geral da rede também aumenta.

3.2.3. Diferença entre caminhos

A Figura 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 questão crucial que surge desse resultado é “*se a rede está ficando mais densa, por que os menores caminhos estão mais similares?*”. Uma 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.

A Figura 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, ou seja, mais caminhos distintos do que os mais curtos, podem utilizar melhor esta diversidade e obter maior tolerância a falhas, por exemplo.

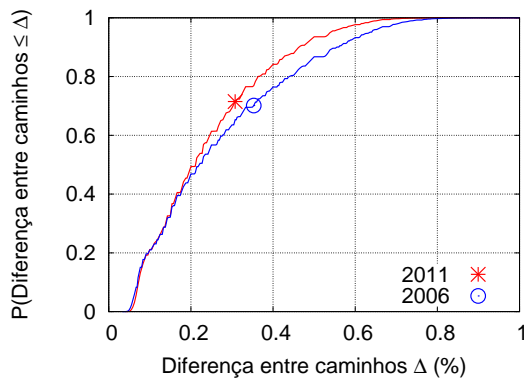


Figura 5. CDF - Diferença dos K-menores caminhos.

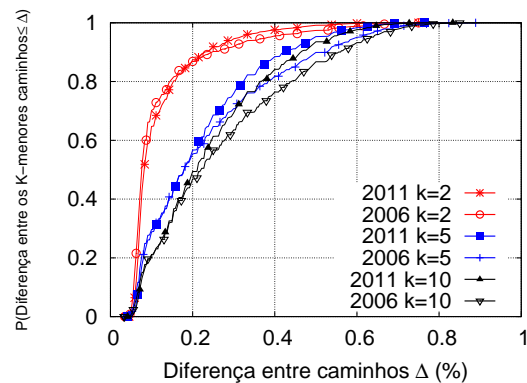


Figura 6. CDF - Diferença dos K-menores caminhos.

3.2.4. Dispersão geográfica

A localização geográfica dos *hosts* foi obtida utilizando uma base de dados gratuita disponibilizada pela MaxMind². Essa base permite obter a localização geográfica de cada *host* fornecendo detalhes como a cidade sede, latitude e longitude.

A Figura 7 mostra que a maioria dos caminhos tem dispersão geográfica de 0 até 3, onde 0 é praticamente uma linha geodésica direta. Essa dispersão geográfica tende a ser relativamente pequena porque a maioria dos enlaces em um caminho tende a seguir uma linha reta. Além do mais, enlaces longos que podem ter maior impacto na dispersão geográfica (como cabos transoceânicos), ocorrem em caminhos longos. Então, um único enlace transoceânico entre todos os outros não impacta a dispersão geográfica geral.

Finalmente, dados de 2011 tem cerca de 10% (Figura 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.

Na Figura 8 mostra-se 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.

3.2.5. 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*, que é mais indicado para relações não lineares dado que este avalia correlações entre valores ordenados. Assim, este coeficiente também lida melhor com valores discrepantes.

A Figura 9 mostra um gráfico de dispersão correlacionando o atraso do caminho com o seu comprimento (em número de saltos). É claramente notável a alta correlação

²<http://www.maxmind.com/>

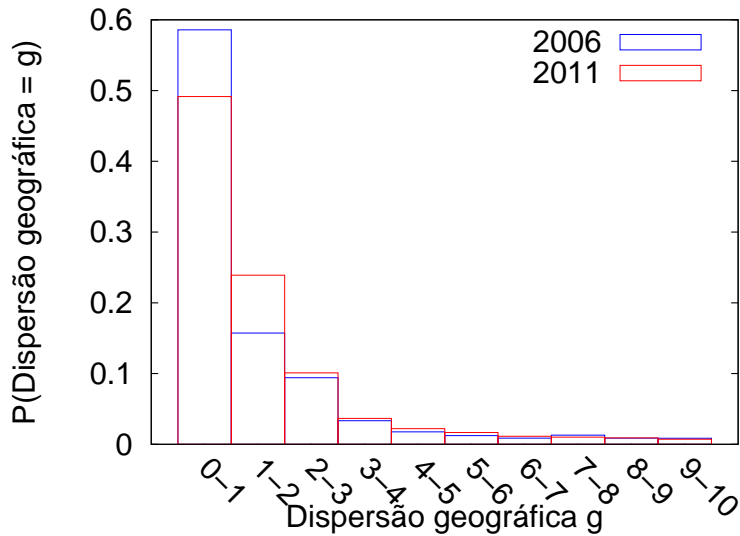


Figura 7. PDF - Dispersão geográfica dos K-menores caminhos.

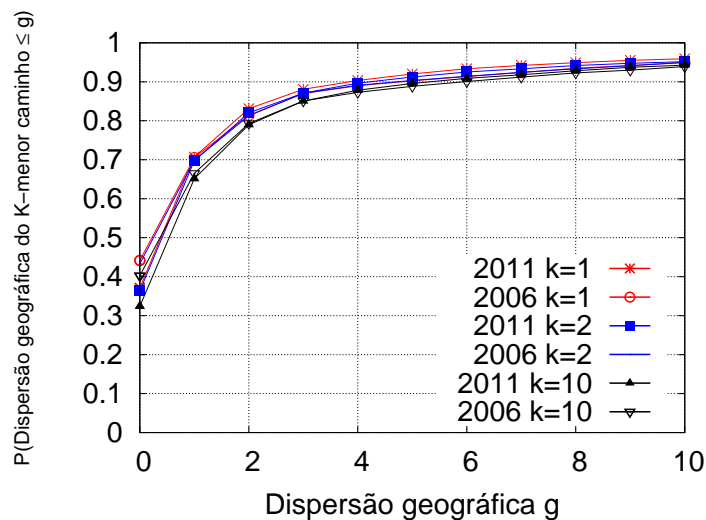


Figura 8. CDF - Dispersão geográfica dos K-menores caminhos (K variante).

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.

Para caminhos com mais de 20 saltos, observa-se que a latência permanece razoavelmente 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, 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 como a razão entre o atraso medido e o atraso

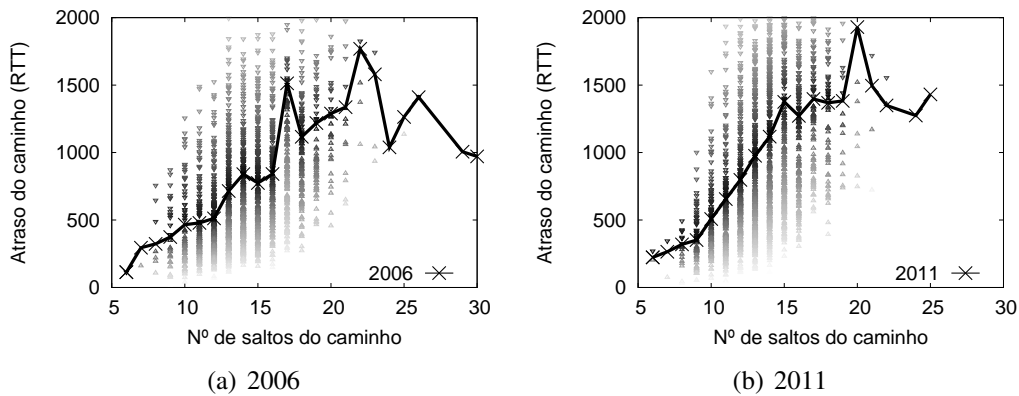


Figura 9. Atraso e comprimento dos K-menores caminhos.

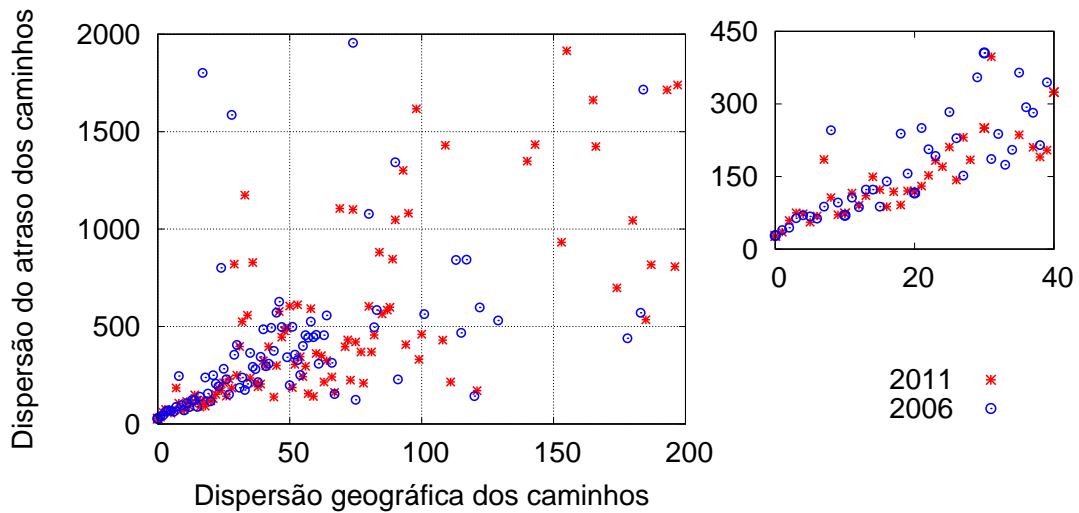


Figura 10. Dispersão do atraso e geográfica dos K-menores caminhos.

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. A Figura 10 mostra que a média da dispersão do atraso e a média da dispersão geográfica do caminho são altamente correlacionadas, especialmente para a maior porção dos caminhos (com baixa dispersão geográfica). O coeficiente de Spearman nesse caso é de 0,86 para 2006 e 0,88 para 2011.

Finalmente, analisamos a correlação entre a média do atraso do caminho e a média do atraso mínimo esperado em um enlace de fibra ótica. A Figura 11 mostra que, para ambos os casos de 2006 e de 2011, a correlação segue uma curva exponencial quando os valores do atraso são menores que 50 ms . 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

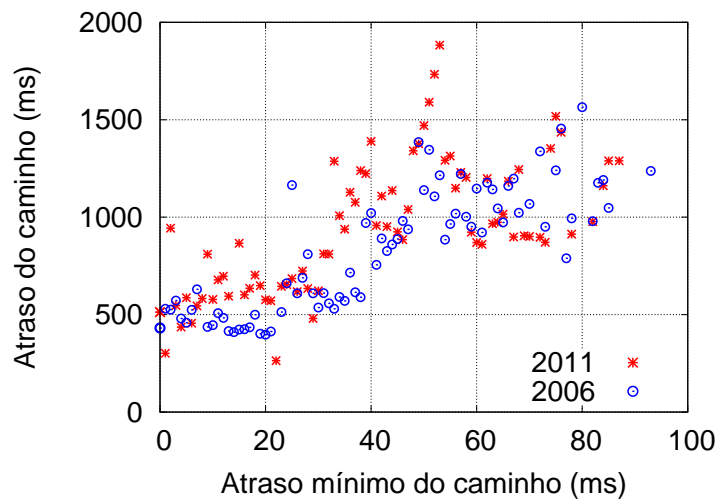


Figura 11. Atraso e o atraso mínimo dos K-menores caminhos.

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.

4. Trabalhos relacionados

A caracterização da evolução e o comportamento da Internet atraiu diversos esforços de pesquisa na última década [Pastor-Satorras et al., 2001, Zhou, 2006, Oliveira et al., 2007, Haddadi et al., 2008, Borgnat et al., 2009, Edwards et al., 2012]. A maioria desses trabalhos foca em como a Internet mudou no nível de AS. Além disso, esses trabalhos focam na caracterização de propriedades topológicas, sem focar em medidas fim-a-fim.

Em particular, Siganos et al. [Siganos et al., 2002] mostrou que em 2002 o crescimento de ambos o número de nós e arestas no nível de AS tem sido exponencial. Esse resultado, juntamente com a observação de que a Internet segue uma estrutura “small world”, indica a densificação da rede. Mais recentemente, Dhamdhare e Dovrolis [Dhamdhare e Dovrolis, 2011], quase uma década após o trabalho de Siganos et al., mostraram que o crescimento no nível de AS agora é linear. Eles também notam que a economia e interesses particulares de AS possuem forte influência na dinâmica da topologia. Um aspecto chave, mas pouco explorado nesses trabalhos, é como essas mudanças topológicas afetam a diversidade de caminhos.

Trabalhos recentes de Edwards et al. [Edwards et al., 2012] e Lee et al. [Lee et al., 2010] também apresentam resultados que mostram aparente piora na eficiência da rede. O caso particular do aumento perceptível da latência, também observado neste nosso artigo, pode estar relacionado ao fenômeno de *bufferbloat*, recentemente investigado por Gettys et al. [Gettys e Nichols, 2012]. Whiteaker et al. [Whiteaker et al., 2011] também mostra que o aumento do atraso na rede possa ser uma consequência do aumento do uso de redes virtualizadas. O uso pesado da rede em máquinas virtuais pode introduzir atrasos de até 100 ms.

O trabalho mais próximo ao nosso é [Schwartz et al., 2010], em que foi estudada a simetria, estabilidade e diversidade das rotas da Internet em 2006 e 2009. O estudo utiliza

dados coletados em um período de 96 horas pela projeto DIMES [Shavitt e Shir, 2005], que consiste em uma plataforma de medição distribuída baseada em computação voluntária. Uma vantagem desta plataforma sobre a CAIDA é a maior quantidade de máquinas que enviam sondas. A grande vantagem da CAIDA são as coletas realizadas diariamente durante todo o ano desde 1998, o que permite análises em maior espaço temporal. Os resultados mostraram que as rotas fim-a-fim da Internet são diversas mas relativamente estáveis, tendo o tipo de rede como ditador do comportamento, onde redes acadêmicas são mais estáveis que as comerciais.

5. Conclusões

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 em particular para o planejamento de enlaces inter-rede e para um melhor ajuste de protocolos de roteamento. Além disso, os resultados podem ser úteis para apoiar visões que alimentem o desenvolvimento de novos protocolos, especialmente protocolos de roteamento.

Alguns dos resultados contra-intuitivos encontrados 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 links deveriam providenciar uma melhor conectividade com enlaces mais rápidos, resultando em uma experiência de usuário aprimorada. Contudo, nesse caso, é mostrado que o resultado oposto com atrasos mais perceptíveis está aumentando. Conjectura-se que os novos equipamentos na Internet não são corretamente configurados para usar gerenciamento ativo de fila. Isso leva ao fenômeno recentemente nomeado de *bufferbloat* [Gettys e Nichols, 2012], onde o excesso de *bufferização* de pacotes na rede causa alta latência e *jitter*.

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.

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.

Como trabalho futuro, pretendemos investigar mais a fundo o fenômeno *buffer-bloat* e a sua relação com o atraso encontrado nos dados de rede analisados.

Agradecimentos

Este trabalho foi parcialmente suportado pela CAPES, CNPq, FAPEMIG e FAPERJ.

Referências

- Ager, B., Chatzis, N., Feldmann, A., Sarrar, N., Uhlig, S., e Willinger, W. (2012). Anatomy of a large european IXP. In *Proc. of the ACM SIGCOMM*, Helsinki, Finland.
- Borgnat, P., Dewaele, G., Fukuda, K., Abry, P., e Cho, K. (2009). Seven years and one day: Sketching the evolution of Internet traffic. In *Proc. of the IEEE INFOCOM*, Rio de Janeiro, Brazil.
- Dhamdhere, A. e Dovrolis, C. (2008). Ten years in the evolution of the Internet ecosystem. In *Proc. of the Internet Measurement Conference (IMC)*, Vouliagmeni, Greece.
- Dhamdhere, A. e Dovrolis, C. (2011). Twelve years in the evolution of the internet ecosystem. *IEEE/ACM Transactions on Networking (TON)*, 19(5):1420–1433.
- Edwards, B., Hofmeyr, S., e Stelle, G. (2012). Internet topology over time. *Arxiv preprint arXiv:1202.3993*.
- Gettys, J. e Nichols, K. (2012). Bufferbloat: Dark buffers in the Internet. *Communications of the ACM*, 55(1):57–65.
- Haddadi, H., Uhlig, S., Andrew Moore, A., Mortier, R., e Rio, M. (2008). Modeling Internet topology dynamics. *ACM SIGCOMM Computer Communication Review*, 38(2):65–68.
- Labovitz, C., Iekel-Johnson, S., McPherson, D., Oberheide, J., e Jahanian, F. (2010). Internet inter-domain traffic. In *Proc. of the ACM SIGCOMM*, New Delhi, India.
- Lee, D., Cho, K., Iannaccone, G., e Moon, S. (2010). Has Internet delay gotten better or worse? In *Proc. of the 5th International Conference on Future Internet Technologies (CFI)*, Seoul, Korea.
- Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10:707–710.
- Oliveira, R., Zhang, B., e Zhang, L. (2007). Observing the evolution of Internet AS topology. *ACM SIGCOMM Computer Communication Review*, 37(4):313–324.
- Pastor-Satorras, R., Vázquez, A., e Vespignani, A. (2001). Dynamical and correlation properties of the Internet. *Physical Review Letters*, 87(25):258701.
- Percacci, R. e Vespignani, A. (2003). Scale-free behavior of the Internet global performance. *The European Physical Journal B*, 32:411–414.
- Schwartz, Y., Shavitt, Y., e Weinsberg, U. (2010). On the diversity, stability and symmetry of end-to-end Internet routes. In *Proc. of the Global Internet (GI)*, San Diego, CA, USA.
- Shavitt, Y. e Shir, E. (2005). DIMES: let the Internet measure itself. *ACM SIGCOMM Computer Communication Review*, 35(5):71–74.

- Siganos, G., Faloutsos, M., e Faloutsos, C. (2002). The Evolution of the Internet: Topology and Routing. Technical report, Univ. of California, Riverside.
- Wagner, R. A. e Lowrance, R. (1975). An extension of the string-to-string correction problem. *Journal of the ACM*, 22(2):177–183.
- Whiteaker, J., Schneider, F., e Teixeira, R. (2011). Explaining packet delays under virtualization. *ACM SIGCOMM Computer Communication Review*, 41(1):38–44.
- Zhou, S. (2006). Understanding the evolution dynamics of Internet topology. *Physical Review E*, 74(1):016124.