INSTITUTO FEDERAL
MINAS GERAIS
Reitoria

Pró-Reitoria de Pesquisa, Inovação
e Pós-Graduação

SEMINÁRIO DE
INICIAÇÃO CIENTÍFICA

Resumo Expandido

Título da Pesquisa: Caracterização de comunicação na Internet com análise das rotas utilizadas pelos nós da rede PlanetLab		
Palavras-chave: Atraso de Comunicação, RTT, Traceroute, PlanetLab, Internet		
Campus: Formiga	**Tipo de Bolsa:** PIBITI	**Financiador:** CNPq
Bolsista: Christiano Eduardo Dutra e Silva		
Professor Orientador: Prof.º M.Sc. Everthon Valadão		
Áreas de Conhecimento: Redes de Computadores / Ciência da Computação / Sistemas de Computação / Sistemas Distribuídos		

Resumos: Muitas aplicações em rede dependem de informações sobre os tempos de ida-e-volta para tomada de decisão. Por isso, caracterizar a comunicação na Internet é essencial para o projeto de novos protocolos e aplicações. Neste trabalho de iniciação científica, pesquisou-se a comunicação distribuída em escala global, com a utilização de centenas de nós espalhados em 40 países em 5 continentes para coletar informações sobre atrasos de comunicação, perda de mensagens e carga de trabalho das máquinas. Também foram coletadas informações sobre as rotas pelas quais as mensagens trafegaram e a localização geográfica aproximada dos nós. Foi realizada a coleta destes dados utilizando nós da rede PlanetLab, a intervalos de 10 segundos e durante 12 dias, obtendo um registro com mais de 500 milhões de medições. De posse dos dados coletados, foram analisadas as rotas e observado que a maioria dos nós tem acesso a enlaces pouco congestionados, porém no núcleo da rede há roteadores sobrecarregados, apresentando altas latências por interconectar e receber tráfego de diversos países e continentes. Percebeu-se também a existência de pontos críticos de falhas, havendo dezenas de nós dependentes de um mesmo roteador, sem alternativas de roteamento em caso de falha. O código-fonte do software coletor e analisador, bem como os resultados da análise foram disponibilizados na Web.

INTRODUÇÃO

A Internet caracteriza-se por uma tecnologia democratizante e o desempenho da mesma deve ser monitorado, de maneira que seja possível acompanhar seu crescimento e evolução e planejar melhorias na infraestrutura da grande rede (SCHULZE & MOCHALSKI, 2009). Tempos de ida-e-volta (RTTs) são uma métrica importante para diversas aplicações na Internet. Por exemplo, eles afetam a escolha de servidores ou pares em sistemas de troca de arquivos e de *streaming* (KIM et al., 2006; SILVA et al., 2009), a operação de mecanismos de detecção de falhas (VALADÃO, 2009) e de controle de congestionamento. Assim, caracterizar a comunicação na Internet é essencial para o projeto de novos protocolos e aplicações, uma vez que fornece dados de referência para a tomada de decisões.

Trabalhos de medição de características da rede só se tornaram realmente viáveis recentemente, com o advento do PlanetLab. O All-Pairs Ping (APP) foi um projeto que mediu RTTs no PlanetLab a cada 15 minutos por 2 anos (STRIBLING, 2005; YOSHIKAWA, 2006), porém com a utilização do comando ping, o que não permitia o casamento preciso entre as informações nas duas extremidades do canal de comunicação. O serviço CoMon realiza a monitoração do RTT no PlanetLab, além de diversas métricas de carga, mas se limita a um intervalo de 5 minutos (PARK; PAI, 2006). Numa análise do progresso evolutivo do RTT médio do

PlanetLab, Tang e outros observaram uma considerável variação no RTT entre duas medições para os mesmos pares de máquinas, ao realizar medições apenas a cada 15 minutos (TANG et al., 2007). O uso do PlanetLab para experimentos sobre a Internet é abordado por Pathak et al. (PATHAK et al., 2008) e por Pucha et al. (PUCHA; HU; MAO, 2006).

Um resumo comparativo dos principais trabalhos de análise do desempenho da comunicação em cenários da Internet e no PlanetLab é apresentado na tabela 1. Como pode nela ser observado, a maior parte dos projetos desenvolvidos para medir atrasos na Internet foram baseados em períodos longos entre amostragens, da ordem de minutos, ou realizaram medições por intervalos curtos de tempo. Entretanto, alguns trabalhos indicam que as amostras muito espaçadas perdem detalhes da variação dos atrasos em intervalos de tempo menores (TANG et al., 2007), enquanto medições muito próximas (da ordem do valor do RTT) podem sofrer de interferências (BOLOT, 1993). Neste trabalho, optou-se por adotar uma resolução temporal intermediária (uma medição a cada 10 segundos) para evitar os dois problemas mencionados.

Trabalho	Método	Duração	Atraso	Assimetria	Distribuição	Escala
(BOLOT, 1993)	NetDyn	8-500 ms / 400s	√	-	-	2 nós
(CHOI; YOO, 2005)	TCP ACKs	~1 ms / 100s	√	-	-	2 nós
(PUCHA; HU; MAO, 2006)	TCP ACK/RST	-	√	-	-	180 nós
(YOSHIKAWA, 2006)	Ping ICMP	15 min. / 2 anos	√	-	-	600 nós
(PARK; PAI, 2006)	Ping ICMP	5 min. / -	√	-	-	-
(TANG et al., 2007)	Ping ICMP	15 min. / 2 anos	√	-	-	600 nós
(PATHAK et al., 2007)	One-way Ping	20 min. / 10 dias	√	√	-	94 nós
(VALADÃO et al., 2010b)	Three-way Ping	10 seg. / 10 dias	√	√	√	100 nós

Tabela 1: Sumário comparativo das pesquisas sobre atrasos de comunicação na Internet

Os objetivos deste trabalho são (1) implementar um mecanismo de coleta das rotas utilizadas nas comunicações entre os nós do PlanetLab, (2) realizar uma coleta geograficamente abrangente de dados de comunicações, (3) a partir das rotas, visualizar a topologia da rede e caracteriza-la. Também, visa-se (4) aplicar uma técnica de clusterização, buscando classificar os enlaces em grupos com base na distribuição dos atrasos, sua variação e perdas, analisar sua distribuição geográfica e buscar regras de associação que expliquem desempenhos discrepantes.

METODOLOGIA

Para a obtenção de dados sobre a caracterização de ida e volta na internet, inicialmente foi realizada uma abrangente revisão bibliográfica sobre o assunto. Após, foi realizado um levantamento das ferramentas que seriam utilizadas na pesquisa, bem como conduzido o projeto e implementação do coletor de informações de rota. Por fim foi realizado o planejamento e execução da coleta no ambiente distribuído da rede PlanetLab. Todas as ferramentas que foram utilizadas durante o projeto são gratuitas e/ou de código-fonte aberto, utilizando o ambiente GNU/Linux.

Visando realizar uma nova coleta mais abrangente de dados, era necessário primeiramente escolher em quais nós da rede PlanetLab, seria executado o experimento. O PlanetLab conta com mais de 1100 máquinas espalhadas em 544 localidades do globo (vide figura 1). Como estas máquinas são cedidas e administradas por diferentes organizações, muitas delas podem estar desligadas ou em manutenção. Obteve-se uma lista relativamente estável de 106 nós ativos e comunicantes, com uma boa cobertura geográfica (40 países e nos 5 continentes) e evitando repetir mais do que duas máquinas de um mesmo local. Conectou-se a este conjunto de 106 nós através do software de gerenciamento PLMan, visando a instalação dos softwares necessários para rodar o programa de coleta, tais como a máquina virtual Java e o serviço NTP de sincronização dos relógios. Também, foi enviado aos nós os softwares de coleta e os arquivos de configuração (descrevendo os nós participantes do experimento e os intervalos de coleta).

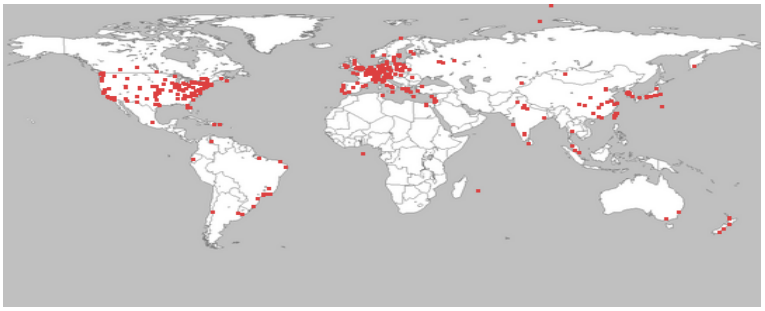

Figura 1: A rede PlanetLab, com mais de 1100 nós espalhados em mais de 540 locais

Para a realização da coleta dos atrasos de rede foi utilizada uma ferramenta com um protocolo próprio, visto que o ping (ICMP echo request) não provê as informações necessárias para inferir os atrasos individuais de cada sentido da comunicação. A ferramenta foi desenvolvida e disponibilizada por Valadão et al., na qual optou-se por utilizar uma abordagem TWP (*Three-Way Ping*) de maneira a ter acesso ao atraso de comunicação nos dois sentidos em momentos próximos (VALADÃO et al., 2010b). O TWP é um ping de três vias e foi implementado em Java, como um framework para coleta de dados de comunicação em rede. Além dos dados acerca dos atrasos, o coletor também registra os seguintes dados de utilização de recursos, informados pelo CoTop: carga média dos últimos 1, 5 e 15 minutos; % de utilização de CPU; % de utilização de memória principal; vazão de banda de *upstream* e *downstream*.

Como a primeira coleta (VALADÃO, 2009) registrava apenas o TWP (Three-Way-Ping) e a carga das máquinas, era necessário que o código do coletor fosse alterado de modo a registrar também a rota pela qual as mensagens trafegavam. Realizou-se uma pesquisa sobre os métodos existentes para descobrir o caminho pelo qual os pacotes trafegam de um nó para o outro. Até onde pôde-se levantar, são majoritariamente utilizados os softwares *tracepath* e *traceroute*, que enviam pacotes "sonda" com tempo de vida (TTL) incremental para descobrirem os roteadores intermediários entre a comunicação de duas máquinas. Foram realizados testes com pares de nós aleatoriamente escolhidos, visando verificar a eficiência da descoberta e o tempo gasto por cada método. O *traceroute* apresentou resultados mais detalhados e com um tempo igual ou inferior ao *tracepath*, motivo pelo qual ele foi definido como o mecanismo para descoberta das rotas.

Foi mantida a mesma metodologia utilizada na coleta originalmente conduzida por Valadão, onde a operação de todo o sistema é controlada por um nó coordenador, que centraliza a identificação de todos os nós no início da coleta e redistribui as informações de inicialização e finalização para todos os nós participantes (VALADÃO et al., 2010a). O experimento foi iniciado no dia 02 de Agosto de 2012 e terminou no dia 14 de Agosto de 2012. O experimento foi conduzido por 2 dias adicionais aos 10 dias originalmente definidos no trabalho precursor, visando possibilitar que fosse obtido um maior intervalo para a análise porém sem transpor as limitações de armazenamento dos dados no PlanetLab. A análise destes dados obtidos é apresentada na próxima seção.

RESULTADOS E DISCUSSÕES

Utilizando o software Weka e os dados da coleta original (VALADÃO, 2009), foram conduzidas análises de regras de associação com o algoritmo “PredictiveApriori”, buscando verificar possíveis causas para o comportamento de grupos de nós que apresentaram desempenho discrepante. Na tabela 2 (a), o Par de Continentes refere-se à localização dos nós de origem e destino. Pelas regras de associação encontradas, pode-se observar que em 77,9% dos casos de pares de nós localizados na Ásia - América do Sul, eles foram alocados no grupo 3 que é caracterizado por baixo atraso e pouca variação (rtt = 167 ms, cv = 0,33). O mesmo pode ser dito para pares de nós localizados no Oriente Médio - América do Sul, porém em 61,3% dos casos. Para casos de pares de nós localizados no Oriente Médio - Oriente Médio, observa-se que 47,2% foram alocados no grupo 1, o qual apresenta ótimo desempenho porém com alta variação (rtt = 49 ms, cv = 1,12). Esse ótimo desempenho provavelmente se dá pelo fato de que as distâncias entre um nó e outro serem pequenas, potencialmente estando esses pares de nós em um mesmo ambiente.

Par de Continentes	Grupo	Aceitação
Ásia - América do Sul	3	0,779
Oriente Médio - América do Sul	3	0,613
Oriente Médio - Oriente Médio	1	0,472

Tabela 2 (a) Associação dos pares de continente com os grupos;

Média do RTT	Associação	Aceitação
Acima de 500ms	Stdev > 115	0,99
Acima de 500ms	Loss > 39%	0,59

Tabela 2 (b) Associação da média do atraso com desvio padrão;

Pela tabela 2 (b), observa-se que na maioria das vezes em que a média foi alta, o desvio padrão também o foi, ou seja, atrasos médios altos tipicamente não correspondem a atrasos altos consistentes, mas sim a valores discrepantes que induzem a elevação da média. Em boa parte das vezes em que foi observada uma média de atraso acima de 500ms, a perda de mensagens ficou acima de 39% (tipicamente, observaria-se perdas de menos de 1%), demonstrando que enlaces com altos atrasos tendem também a gerar altas taxas de perda de mensagens.

Para melhor compreender os relacionamentos entre as rotas que possibilitam a comunicação dos nós do PlanetLab e também de muitos outros dispositivos com acesso à Internet, os dados do *traceroute* de

cada nó foram representados em um grafo. Optou-se pelo formato GEXF (Graph Exchange XML Format), uma linguagem para descrever estruturas complexas de redes, seus dados associados e sua dinâmica. O formato GEXF é aberto, extensível e intercambiável, sendo utilizado pelo software Gephi, utilizado manipulação e visualização de grafos. Foi construído um *script* para converter os dados coletados do *traceroute* para o formato XML estipulado pelo GEXF. Nele, os nós e arestas do grafo são descritos de maneira que o peso da aresta corresponde ao atraso médio observado nas comunicações do nó de origem até o roteador em questão.

Foram selecionados dois nós para ilustrar e analisar a representação visual do grafo das rotas de comunicação. Nas imagens que serão apresentadas a seguir, os nós são representados por círculos numerados, tendo a sua cor e tamanho variando de acordo com o seu grau de conectividade. As arestas ligam os nós através de linhas, cuja cor e espessura representam a magnitude do atraso médio observado nestes enlaces. Assim, linhas mais grossas e vermelhas representam enlaces congestionados enquanto que linhas mais finas e azuis representam enlaces de desempenho mais rápido. Para a distribuição dos nós, foi utilizado o algoritmo de Yifan Hu (Gasner et al., 2010), que empurra para a periferia nós com baixa conectividade. Portanto, o tamanho das linhas não tem nenhuma correlação com os atrasos, servindo somente para distanciar suficientemente os nós e facilitar sua visualização, ao evitar sobreposições.

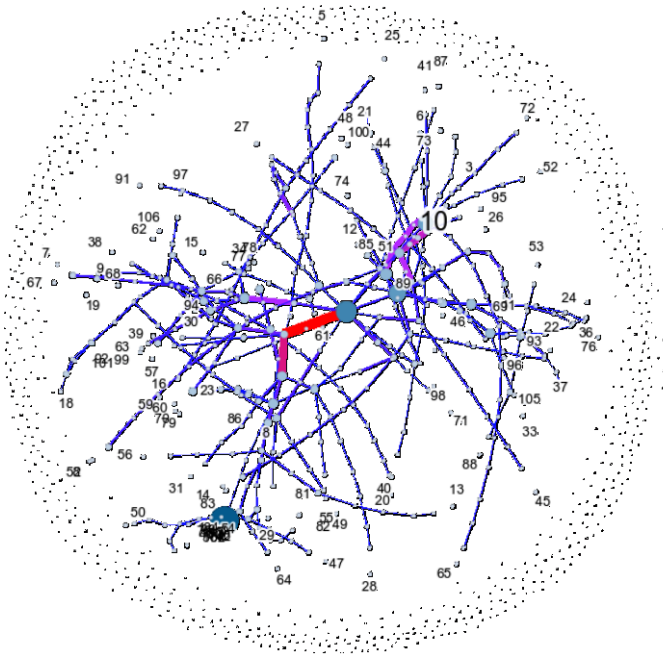

Figura 2 (a) - Topologia: saturn.planetlab.carleton.ca (ID #23)

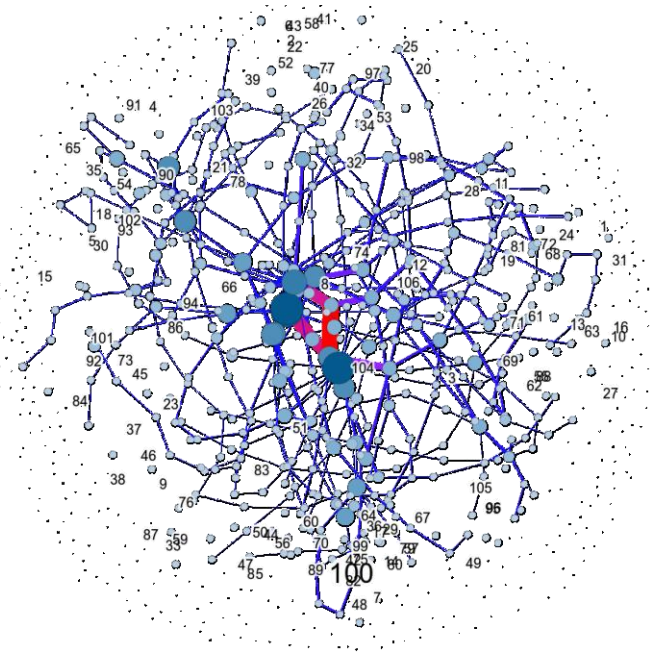

Figura 2 (b) - Topologia: saturn.planetlab.carleton.ca (ID #23)

No total, o nó #23 registrou 217.520 entradas com informações sobre as rotas utilizadas para comunicar-se com os 105 demais nós. Considerando que tipicamente as rotas não são alteradas frequentemente, nestas entradas foram verificadas 26.323 rotas únicas, das quais estão representadas no grafo apresentado na Figura 2 (a) aquelas que são utilizadas com maior frequência. O nó #23 está localizado um pouco à esquerda da área central, onde é possível observar que para o nó #23 comunicar-se com a maioria dos nós ele deve utilizar-se de um enlace congestionado, localizado na área central do grafo em cor vermelha, próximo ao nó 61 (localizado em Massachusetts, EUA). Observe também que o nó #10, apesar de ter uma grande importância devido ao seu alto grau de conectividade, encontra-se envolto por rotas relativamente congestionadas (cor roxa). Por fim, no canto inferior esquerdo, pode-se observar dezenas de nós aglomerados em uma mesma rota. Caso o roteador em questão apresente problemas de comunicação, todos estes nós terão que buscar rotas alternativas para continuarem a se comunicarem com os demais nós que estão do outro lado do grafo.

No total, o nó #100 registrou 98.477 entradas com informações sobre as rotas utilizadas para comunicar-se com os 105 demais nós. Considerando que tipicamente as rotas não são alteradas frequentemente, nestas entradas foram verificadas 7.979 rotas únicas, das quais estão representadas no grafo apresentado na Figura 2 (b) aquelas que são utilizadas com maior frequência. O nó #100 está localizado na parte inferior do grafo, numa região sem congestionamentos visíveis e sendo considerado um nó importante devido ao seu alto grau de conectividade. Compartilhando das mesmas rotas que o nó #100 estão os nós #99 (planetlab1.pop-mg.rnp.br) e #67 (planetlab1.pop-pa.rnp.br), ambos pertencentes ao backbone da RNP (Rede Nacional de Ensino e Pesquisa). Ainda, é possível observar importantes rotas no centro do grafo, com seus roteadores ilustrados em tamanho e cor mais fortes para representar sua maior conectividade. Verificando sua localização, nota-se que estes roteadores conectam os nós #104 (Califórnia, EUA), #8 (França), #74 (China), #66 (Uruguai), dentre outros, consistindo em pontos de troca de tráfego de diversos continentes, fato que talvez justifique o maior congestionamento nos enlaces destes roteadores.

CONCLUSÕES

Neste trabalho de iniciação científica, foi desenvolvida uma pesquisa envolvendo a comunicação distribuída em escala global, na qual foram utilizadas centenas de nós espalhados em 40 países em 5 continentes para coletar informações sobre atrasos de comunicação, perda de mensagens, carga de trabalho das máquinas e informações sobre as rotas pelas quais as mensagens trafegaram. Aplicou-se uma técnica de clusterização, classificando os enlaces em grupos com base na distribuição dos atrasos/variação/perdas e também na distribuição geográfica. A análise do trabalho anteriormente desenvolvido por Valadão *et al.* também foi complementada, ao verificar a relação entre o desempenho de grupos de nós e sua respectiva localização geográfica e para tal, realizou-se a extração de regras de associação entre os dados (VALADÃO et al., 2010b). O software de coleta foi revisado e foram implementadas melhorias como a obtenção e armazenamento de informações sobre a rota utilizada pelas mensagens. Estas rotas foram obtidas pelo sofware *traceroute*, paralelamente à coleta de *Three-Way-Ping* (TWP) e da carga de trabalho das máquinas, de maneira a possibilitar uma visão mais abrangente das possíveis causas dos atrasos e perdas de mensagens.

Uma vez implementado o novo coletor, foi realizada uma nova coleta de dados na rede PlanetLab, planejada durante meses e executada durante 12 dias, entre 106 nós espalhados em 40 países de 5 continentes. De posse dos dados coletados, as informações referentes às rotas pelas quais trafegaram as mensagens foram analisadas e nota-se que a maioria dos nós tem acesso a enlaces pouco congestionados. Porém, o núcleo da rede comumente está populado por roteadores sobrecarregados, apresentando altas latências por interconectar e receber tráfego de diversas redes. Também, foi possível observar pontos críticos de falha, com a existência de dezenas de nós dependentes de um mesmo roteador, sem alternativas de roteamento em caso de falha. O código-fonte do software coletor e analisador, bem como os resultados da análise foram disponibilizados no endereço eletrônico http://homepages.dcc.ufmg.br/~evaladao/dfd/.

Como trabalhos futuros pretende-se disponibilizar também o trace completo, mas para tal é necessário o acesso a um servidor Web com recursos suficientes tais como armazenamento para dezenas de Gigabytes e largura de banda adequada. Também, será dada continuidade à análise dos dados coletados, comparando-os com coletas anteriormente realizadas (VALADÃO, 2009). Assim, será possível verificar alterações no desempenho da grande rede de 2009 para 2012. Os dados coletados podem ser utilizados para análises de atrasos e perdas de mensagens na Internet, caracterização da distribuição estatística que melhor representa os dados, clusterização dos dados e de sua distribuição geográfica, simulações e análise de desempenho de algoritmos de comunicação em rede, dentre outros. Também, os dados coletados podem servir como entrada realista para simulações de sistemas distribuídos e como base para o desenvolvimento de modelos sintéticos de comunicação em rede.

REFERÊNCIAS BIBLIOGRÁFICAS

(Bolot, 1993) Bolot, J.-C. Characterizing end-to-end packet delay and loss in the internet. Journal of High Speed Networks, v. 2, n. 3, p. 289–298, 1993.

(Choi e Yoo, 2005) Choi, J.-H.; Yoo, C. One-way delay estimation and its application. SIGCOMM Computer Communication Review, ACM, New York, NY, EUA, v. 28, n. 7, p. 819–828, 2005.

(Gasner et al., 2010) E. Gansner; Y. Hu; e S. Kobourov. Gmap: Drawing graphs as maps. In Proceedings of IEEE Pacific Visualization Symposium, p. 201–208, 2010.

(Kim et al., 2006) Kim , T.; Han, Y; Lee, E. Analysis of delay times for the multimedia streaming services at the broadband convergence network (BcN). In: Proceedings of The Joint International Conference on Optical Internet and Next Generation Network, COIN-NGNCON 2006. [S.l.: s.n.], 2006. p. 219–221.

(Park e Pai, 2006) Park, K.; Pai, V. S. Comon: a mostly-scalable monitoring system for planetlab. SIGOPS Operating Systems Review, ACM, New York, NY, EUA, v. 40, n. 1, p. 65–74, 2006.

(Pathak et al., 2008) Pathak, A.; Pucha, H; Zhang, Y; Hu, Y.C.. A measurement study of internet delay asymmetry. In: Proceedings of the 9th International Conference on Passive and Active Network Measurement (PAM ’08). Berlin, Alemanha: Springer-Verlag, 2008. (Lecture Notes in Computer Science, v. 4979), p. 182–191.

(Pucha et al., 2006) Pucha, H.; Hu, Y. C.; Mao, Z. M. On the impact of research network based testbeds on wide-area experiments. In: Proceedings of the 6th ACM SIGCOMM conference on Internet measurement. New York, NY, EUA: ACM, 2006. (IMC ’06), p. 133–146.

(Schulze e Mochalski, 2010) Schulze, H. e Mochalski, K. *Internet study 2008/2009*. Ipoque, páginas 1–14. Disponível online em http://portal.ipoque.com/, último acesso em 29/11/2011.

(Silva et al., 2009) Silva, T. H. ; Mota, V. F. S. ; Valadão, E. ; Almeida, J. ; Guedes, D. O.. Caracterização do Comportamento dos Espectadores em Transmissões de Vídeo ao Vivo Geradas por Usuários. In: Anais do XXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC '09). Recife, PE, Brasil: 2009. v. XXVII. p. 613–626.

(Stribling, 2005) Stribling, J. All-Sites-Pings (APP). 2005. Disponível em http://pdos.csail.mit.edu/~strib/pl_app/, último acesso em 29/11/2011.

(Tang et al., 2007) Tang, L.; Chen, Y.; Li, F. Empirical study on the evolution of planetlab. In: Proceedings of the 6th International Conference on Networking (ICN '07). Los Alamitos, CA, EUA: IEEE Computer Society, 2007. p. 64–64.

(Valadão, 2009) Valadão, Everthon. FD-Sensi: um detector de falhas adaptativo e sua aplicação a um sistema distribuído em larga escala. Belo Horizonte: UFMG, 2009. 127 p. Dissertação (Mestrado) – Programa de Pós-Graduação em Ciência da Computação, Universidade Federal de Minas Gerais, Belo Horizonte, 2009.

(Valadão et al., 2010a) Valadão, E. ; Silva, T. H. ; Geçary, A. ; Guedes, D. O. ; Duarte, R. O. . Medição, análise e modelagem de tempos de ida-e-volta na Internet. In: Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos, 2010, Gramado, RS. Anais do XXVIII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC'2010), 2010. v. XXVIII. p. 407-420.

(Valadão et al., 2010b) Valadão, E. ; Guedes, D. O. ; Duarte, R. O. Caracterização de tempos de ida-e-volta na Internet. Revista Brasileira de Redes de Computadores e Sistemas Distribuídos, ISSN 1983-4217, v. 3, p. 21-34, 2010.

(Yoshikawa, 2006) Yoshikawa, C. PlanetLab - All Sites Pings. 2006. Disponível em http://ping.ececs.uc.edu/ping/, último acesso em 29/11/2011.