

# Medição, análise e modelagem de tempos de ida-e-volta na Internet

Everthon Valadão<sup>1</sup>, Assahaf Geçary<sup>1</sup>, Thiago Silva<sup>1</sup>,  
Dorgival Guedes<sup>1</sup>, Ricardo Duarte<sup>2</sup>

<sup>1</sup> Departamento de Ciência da Computação — Universidade Federal de Minas Gerais  
Belo Horizonte, MG.

<sup>2</sup> Departamento de Engenharia Eletrônica — Universidade Federal de Minas Gerais  
Belo Horizonte, MG.

{evaladao, assahaf, thiagozs, dorgival}@dcc.ufmg.br

ricardoduarte@ufmg.br

**Abstract.** Round-trip times (RTTs) are an important metric for the operation of many protocols in the Internet: they are taken into account when choosing servers or peers in streaming systems, they impact the operation of fault detectors and congestion control algorithms. Therefore, detailed knowledge about RTTs in the Internet is important for application and protocol developers. In this work we present results on measuring RTTs between 81 PlanetLab nodes every ten seconds, for ten days. The resulting dataset has over 550 million measurements. Our analysis of that dataset gives us a profile of delays in the network and identifies a Gamma distribution as the model that best fits our data. We observe that asymmetry is prevalent in the measurements. By using a clustering technique we observe that links can be divided in five groups based on the distribution of RTTs over time and observed losses.

**Resumo.** Tempos de ida-e-volta (round-trip times, ou RTT) são uma métrica importante para diversos tipos de protocolos na Internet: eles são considerados na escolha de servidores ou pares em sistemas de troca de arquivos e de streaming, têm impacto na operação de mecanismos de detecção de falhas e de controle de congestionamento. Por esse motivo, um conhecimento detalhado do comportamento dos RTTs na rede é importante para o projeto de protocolos e aplicações. Neste trabalho apresentamos os resultados da medição de tempos de ida-e-volta entre 81 nós da rede PlanetLab a cada dez segundos, por dez dias. O conjunto de dados resultante possui cerca de 550 milhões de medições. A análise desse conjunto nos permite traçar um perfil detalhado do atraso na rede e identificar uma distribuição Gamma como a que melhor aproxima os dados coletados. Verificamos que a assimetria nas conexões é frequente. Aplicando uma técnica de clusterização verificamos que os enlaces podem ser classificados em cinco grupos distintos com base na distribuição dos RTTs, sua variação e as perdas observadas.

## 1. Introdução

Na Internet atual, muitas aplicações dependem de informações sobre os tempos de ida-e-volta pela rede (*Round-Trip Times*, ou RTT) para certas decisões. Esses tempos são importantes para aplicações que operam segundo o princípio de requisição resposta, como

DNS e aplicações baseadas em chamadas de procedimentos remotos (RPC). Além disso, essa informação também é usada quando se objetiva reduzir as latências de transmissão e dimensionar *buffers* na rede [Wille et al. 2008]. Por exemplo, aplicações de *multicast* de áudio/vídeo dependem dessa informação para montar suas árvores de distribuição; aplicações *peer-to-peer* como Gnutella e KAD escolhem como vizinhos nós que respondem mais rápido a consultas.

Um estudo sobre a distribuição de atrasos (unidirecionais ou de ida-e-volta) é importante para servir como base para auxiliar projetistas de protocolos e aplicações em seu trabalho, já que fornece dados de referência para a tomada de decisões. Além disso, os dados coletados podem servir como entrada realista para simulações de sistemas distribuídos e como base para o desenvolvimento de modelos sintéticos a serem usados em simuladores.

Tendo esses objetivos em mente, neste trabalho apresentamos um estudo sobre tempos de ida-e-volta e atrasos unidirecionais na Internet. Medidas de RTT foram coletadas entre aproximadamente 100 pontos da rede a cada dez segundos durante dez dias. Os dados coletados foram analisados utilizando ferramentas estatísticas e de mineração de dados, buscando fornecer uma caracterização detalhada dos dados coletados. Analisamos, entre outros aspectos, a distribuição de atrasos e perdas observadas, o nível de assimetria observado durante as medições e a existência de agrupamentos de canais de comunicação com características distintas. Além disso, desenvolvemos um modelo sintético para a distribuição de RTTs.

Para uma medida precisa de tempos de ida-e-volta e de atrasos unidirecionais é necessário acesso aos dois extremos do canal de comunicação que se deseja medir, o que descarta o uso de aplicações usuais que utilizam o protocolo ICMP, como o comando *ping*. Sendo assim, utilizamos nós da rede PlanetLab em nossas medições, tendo o cuidado de selecionar máquinas com uma ampla cobertura geográfica, localizadas tanto em redes acadêmicas/de pesquisa quanto em redes comerciais.

Nossos resultados mostram que os RTTs variam significativamente entre pares de máquinas na Internet, como seria de se esperar, mas também podem variar significativamente entre duas máquinas definidas, mesmo em escalas de tempo reduzidas. Apesar da variabilidade, nossa análise indica que há alguns grupos de comportamento similar em função da estrutura da rede. Perdas, em geral, são baixas e tendem a se concentrar nos canais mais longos (aqueles com maiores RTTs).

O restante deste trabalho é organizado da seguinte forma: a seção a seguir discute conceitos e trabalhos relacionados, para então discutirmos o processo de medição e coleta de dados na seção 3. Os resultados são apresentados na seção 4, com sub-seções para os principais elementos da nossa análise. Finalmente, concluímos com algumas observações e uma discussão de possíveis trabalhos futuros.

## 2. Trabalhos relacionados

Choi e Yoo propuseram uma técnica para derivar atrasos unidirecionais, analisada pelos autores através de simulações [Choi and Yoo 2005], porém, não chegaram a aplicá-la na análise de assimetria ou em um estudo abrangente na Internet. Pela necessidade de medição nos dois extremos da comunicação e da existência de mecanismos de sincronização

de relógios nesses extremos, estudos sistemáticos sobre RTT (*round-trip time*) na Internet só se tornaram viáveis recentemente, como o advento do PlanetLab <sup>1</sup>.

Tipicamente, projetos desenvolvidos para medir atrasos na Internet foram baseados em períodos longos de amostragem. O *All-Pairs Ping*<sup>2</sup> (APP) foi um projeto que mediu RTTs no PlanetLab a cada 15 minutos por 2 anos. Nesse caso o comando `ping` foi utilizado, o que não permite o casamento preciso entre as informações nas duas extremidades do canal de comunicação. Já o serviço CoMon [Park and Pai 2006] realiza no PlanetLab a monitoração do RTT, além de diversas métricas de carga, mas não armazena a relação entre os pares de nós utilizados e se limita a um intervalo de 5 minutos. Numa análise do progresso evolutivo do RTT médio do PlanetLab [Tang et al. 2007] os autores observaram uma considerável variação no RTT, mesmo entre curtos períodos de tempo. Entretanto, o intervalo de amostragem por eles utilizado foi de 15 minutos, enquanto nosso objetivo neste trabalho é o acompanhamento de RTTs com amostras muito mais frequentes, a cada 10 segundos.

Em uma análise mais recente da assimetria dos atrasos na Internet, Pathak e outros realizaram um estudo abrangente no qual mediram a severidade da assimetria no atraso entre diversos pares de nós do PlanetLab [Pathak et al. 2008]. Seus resultados estão de acordo com os de Tang e outros, sendo observado que a assimetria dos atrasos prevalece na grande maioria dos nós, podendo ser em parte atribuída à assimetria nas rotas e em parte a congestionamentos transientes. Um resumo comparativo dos principais trabalhos de análise dos tempos de ida-e-volta em cenários da Internet e no PlanetLab é apresentado na Tabela 1.

<i>Trabalho</i>	<i>Método</i>	<i>Intervalo / Duração</i>	<i>Varição Atraso</i>	<i>Assimetria</i>	<i>Modelo Sintético</i>	<i>Escala</i>
[Choi and Yoo 2005]	<i>one-way</i>	~1 ms / 100 seg.	✓	-	-	2 nós
[Tang et al. 2007]	<i>ping</i>	15 min. / 2 anos	✓	-	-	600 nós
[Pathak et al. 2008]	<i>one-way</i>	20 min. / 10 dias	✓	✓	-	94 nós
<b>Este trabalho</b>	<i>two-way</i>	<b>10 seg. / 10 dias</b>	✓	✓	✓	<b>81 nós</b>

**Tabela 1. Sumário comparativo das pesquisas sobre atrasos na Internet**

O uso do PlanetLab para experimentos sobre a Internet é abordado por Pathak e outros, que observaram uma diferença de conectividade entre nós da Internet acadêmica e da Internet comercial [Pathak et al. 2008]. Levamos esse aspecto em consideração ao incluir nós localizados em redes comerciais em nossas medições. Também, neste trabalho, utilizamos uma técnica derivada de Pathak e outros para medir atrasos unidirecionais, com base no uso do protocolo NTP <sup>3</sup> para sincronização dos relógios das máquinas e avaliação dos erros máximos envolvidos. Rocha e outros discutem as limitações dessa técnica [Rocha et al. 2004], especialmente os problemas de atualização de relógio, que foram consideradas ao se excluir nós problemáticos conforme veremos adiante.

### 3. Medição de tempos de ida-e-volta

O RTT (*round-trip time*) entre dois nós A e B é definido como sendo a soma do atraso direto de A para B e do atraso inverso de B para A. Ele tem sido utilizado largamente como

<sup>1</sup><http://www.planet-lab.org/>

<sup>2</sup>[http://pdos.csail.mit.edu/~strib/pl\\_app/](http://pdos.csail.mit.edu/~strib/pl_app/)

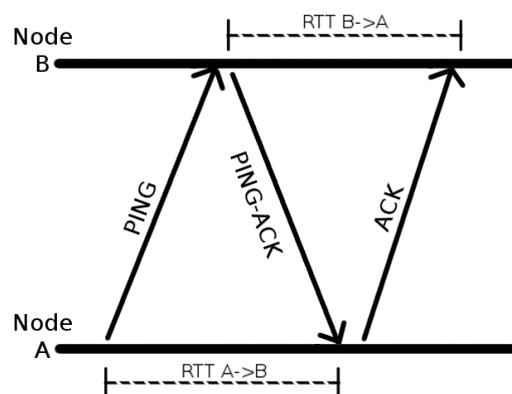
<sup>3</sup><http://www.ietf.org/rfc/rfc1305.txt>

Todos os endereços foram acessados pela última vez em 30/03/2010.

métrica em diversas aplicações de rede que necessitam inferir atrasos e/ou localização dos nós. O RTT é comumente utilizado para derivar tais atrasos individuais, onde normalmente assume-se seu valor como sendo  $1/2$  RTT. Porém a existência de assimetrias nesses atrasos individuais prejudica a precisão e desempenho das aplicações que se utilizam deste pressuposto [Pathak et al. 2008]. Portanto, se possível, a utilização dos atrasos individuais de cada direção do enlace é desejável por trazer consigo benefícios decorrentes de sua maior precisão.

Para a realização da coleta dos atrasos de rede e dados de utilização de recursos dos nós no PlanetLab foi necessário realizar o projeto e implementação de um software próprio, visto que o *ping* (*ICMP echo request*) não nos proveria as informações necessárias para inferir os atrasos individuais de cada direção da comunicação. Optamos por utilizar uma abordagem “*Three-Way Ping*”, de maneira a ter acesso ao RTT nos dois sentidos em momentos próximos.

O *Three-Way Ping* (TWP) funciona da seguinte maneira: o nó A envia periodicamente solicitações ao nó B, através de mensagens UDP. A solicitação inicial é denominada PING, ao passo que o nó B responderá um PING-ACK. Ao receber a resposta do nó B, o nó A devolve um ACK para fechar a rodada de troca de mensagens deste protocolo. Assim, o TWP possibilita que, em um nó A qualquer, calculemos o RTT de A para B como sendo o intervalo de tempo entre a emissão um PING e o recebimento da resposta PING-ACK. Para um momento imediatamente posterior, é também possível calcular em B o RTT de B para A como sendo o intervalo entre a emissão de um PING-ACK e o recebimento do ACK, conforme ilustra a Figura 1.



**Figura 1. Sequência de Troca de Mensagens no TWP**

Caso ambos os nós A e B estejam com seus relógios sincronizados, é possível também calcular os atrasos individuais se estes forem de ordem de grandeza superior ao escorregamento dos relógios (*clock-drift*). Os relógios dos nós do PlanetLab não escorregam muito de seus servidores NTP: 60% apresentam um escorregamento de menos de 2 milissegundos e 40% apresentam um erro máximo estimado pelo NTP de menos de 10 milissegundos [Pathak et al. 2008]. O atraso direto de A para B é calculado subtraindo-se o tempo de chegada da solicitação PING em B do seu tempo de envio em A. O atraso inverso também poderá ser calculado ao subtrair-se o tempo de chegada da resposta PING-ACK em A do seu tempo de envio em B. Ademais, uma estimativa da precisão do atraso direto pode ser realizada se for considerado também o atraso direto disponibilizado pe-

los tempos de envio e chegada da confirmação final ACK. Estes atrasos individuais serão utilizados na análise de assimetria (conforme será visto na seção 4.4).

Implementamos o TWP em Java, como um software para coleta dos atrasos de rede no PlanetLab. Cada mensagem do protocolo TWP consiste num datagrama UDP com a especificação do tipo de mensagem (PING, PING-ACK ou ACK) e um número de sequência, para ordenação e identificação única de cada sequência de TWP. Para os identificadores, que cobriram todo o espaço de endereçamento a eles dedicado algumas vezes durante o experimento, foi utilizado um mecanismo idêntico ao utilizado pelo TCP no tratamento de seus números de sequência. No ato de envio e recebimento das mensagens, cada nó armazena em seu *trace* a quintupla: (*timestamp*, número de sequência, tipo de mensagem, ID do nó fonte, ID do nó destino). Para reduzir a quantidade de mensagens trocadas e evitar redundância, cada nó inicia TWPs para metade do conjunto restante de nós e responde às solicitações da outra metade do conjunto.

A operação de todo o sistema é controlada por um nó coordenador, que centraliza as informações de inicialização de todos os nós participantes e as repassa aos demais, determinando como cada participante deverá atuar durante o experimento. Os nós participantes armazenam o *trace* localmente com um mecanismo de rotação periódica. Os arquivos que já foram rotacionados são comprimidos e enviados para o servidor por uma tarefa de fundo (*background*), possuindo um sistema de escalonamento de transmissões para evitar que os nós sobrecarreguem o coordenador com acessos simultâneos. Ao final do experimento, o coordenador envia um comando para os nós participantes finalizarem a troca de mensagens e enviarem a ele os resultados restantes.

A coleta foi realizada por 10 dias em 2009, do dia 27 de maio ao dia 05 de junho. Nesse período, 14 nós foram reiniciados pelo menos uma vez. Quando um nó era reiniciado ele comunicava-se com o coordenador para obter a lista dos demais nós participantes e voltava a trocar requisições TWP com eles. Entretanto, durante a reinicialização desses 14 nós houve o problema de nova sincronização de relógios pelo NTP e de identificadores pelos programas de monitoração. Outros 5 nós apresentaram problemas de contato ao servidor NTP associado. A premissa da sincronização de relógios não seria importante para uma análise local dos tempos em uma máquina, mas é essencial para podermos correlacionar as medições de diversas máquinas ao longo do tempo para as análises com base temporal. Por esse motivo precisamos descartar os 19 nós que tiveram algum problema de sincronização durante o período. Assim, os traces utilizados neste trabalho referem aos 81 nós restantes.

Nos 81 *traces* analisados, ignoramos as três primeiras e as três últimas horas do experimento para evitar valores discrepantes nas perdas e atrasos, devidos ao início e encerramento dos diversos programas envolvidos e a sobrecarga do coordenador com as transferências finais. Acerca da intrusividade da ferramenta, a monitoração gerou um volume de tráfego de cerca de 2,2 Kbps por máquina (com 10 segundos entre TWPs). Tal volume é pouco significativo para a rede do PlanetLab, onde os enlaces de última milha raramente são inferiores a 1 Mbps. Quanto à carga de processamento gerada nos nós, essa foi sempre inferior a 1 % (*CPU utilization*).

#### 4. Análise dos dados coletados

O tamanho médio dos *traces* colhidos foi de 85000 ( $\approx 10 \times 24 \times 3600s/10s$ ) *Three-Way Pings/nó*, o que leva a cerca de 255.000 mensagens trocadas por cada par de nós (considerando-se que cada TWP gera três mensagens: PING, PING-ACK e ACK). Como cada nó troca TWP com os (N-1) outros nós, ao todo os *traces* coletados nos 81 nós considerados resultam em  $81 \times 80 = 6480$  enlaces de comunicações entre dois pares de nós. Porém como vimos na seção 3, cada nó envia TWP a N/2 nós e responde da outra metade. Assim, ao todo coletamos cerca de  $85000 \times 3240 \approx 275$  milhões de TWPs, ou 826 milhões de mensagens trocadas durante os 10 dias da coleta. Pela definição do *Three-Way Ping*, cada TWP nos permite calcular dois RTTs, chegando a um total aproximado de 551 milhões de medidas de RTT. Consideramos como perda de mensagem os PINGs (e PING-ACKs) enviados mas não respondidos (identificados pelo número de sequência de cada mensagem), quando um RTT não pode ser calculado. Ao todo, os 81 *traces* binários ocupam 20 GiB de espaço em disco (4 GiB com compressão 7Zip). O *trace* completo, bem como os programas de medição e coleta dos TWP podem ser obtidos em <http://www.speed.dcc.ufmg.br/>.

##### 4.1. Evolução diária

Como mencionado anteriormente, a coleta foi realizada dos dias 27/05/2009 a 05/06/2009, incluindo portanto um final de semana (dias 4 e 5 da coleta). A Figura 2 apresenta as médias diárias para algumas grandezas medidas a partir do *trace*. Nos quatro casos, as barras verticais indicam os intervalos com 99% de confiança para a grandeza calculada.

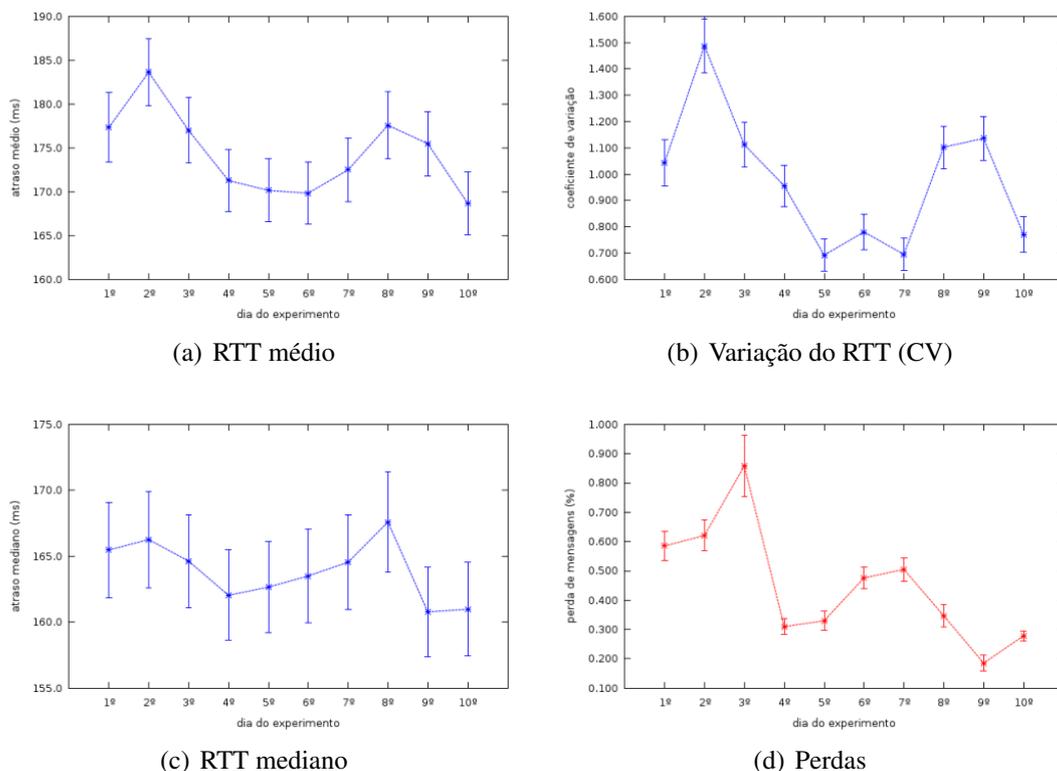


Figura 2. Evolução diária das medições

A Figura 2(a) exibe a evolução diária do RTT médio ao longo do *trace*. Nela é possível observar que do 2º ao 6º dia houve um período de queda no RTT médio, que foi de 183 ms a 169 ms, incluindo um fim de semana. Após este período observa-se uma breve subida do 7º ao 8º dia para o valor de 177 ms e uma nova queda até o fim da coleta no 10º dia, alcançando então o menor valor para o RTT médio: 168 ms. Já a mediana, exibida na Figura 2(c), manteve-se ligeiramente inferior à média, indicando uma concentração maior em valores mais baixos.

A variabilidade das medições é ilustrada na Figura 2(b), que exibe a evolução diária da variabilidade média dos atrasos ao longo do *trace*. Nela é possível observar que, tal como observado no RTT médio, do 2º ao 8º dia houve um período de queda na variabilidade, seguido de uma subida do 8º ao 9º dia. Ou seja, ao mesmo tempo que o RTT médio decaiu, observou-se uma maior homogeneidade no coeficiente de variação dos atrasos do referido período.

Finalmente, a Figura 2(d) exibe a evolução diária da perda média de mensagens ao longo do *trace*. Observamos que a perda de mensagens apresentou uma maior variabilidade nos 3 primeiros dias da coleta, passando a demonstrar valores mais homogêneos a partir do 4º dia. O 3º dia apresentou o maior valor global de 0,86% para a perda média de mensagens e o 9º dia apresentou o menor valor global de 0,18%.

É interessante observar que as médias diárias, calculadas considerando todos os 81 pares, apresentam quedas durante o final de semana para todas as grandezas ilustradas na Figura 2. Se aceitarmos que o tráfego na Internet pode ser mais reduzido durante os finais de semana, a variação observada pode ser atribuída à redução das filas nos roteadores naquele período (e, conseqüentemente, a redução da fração da latência associada).

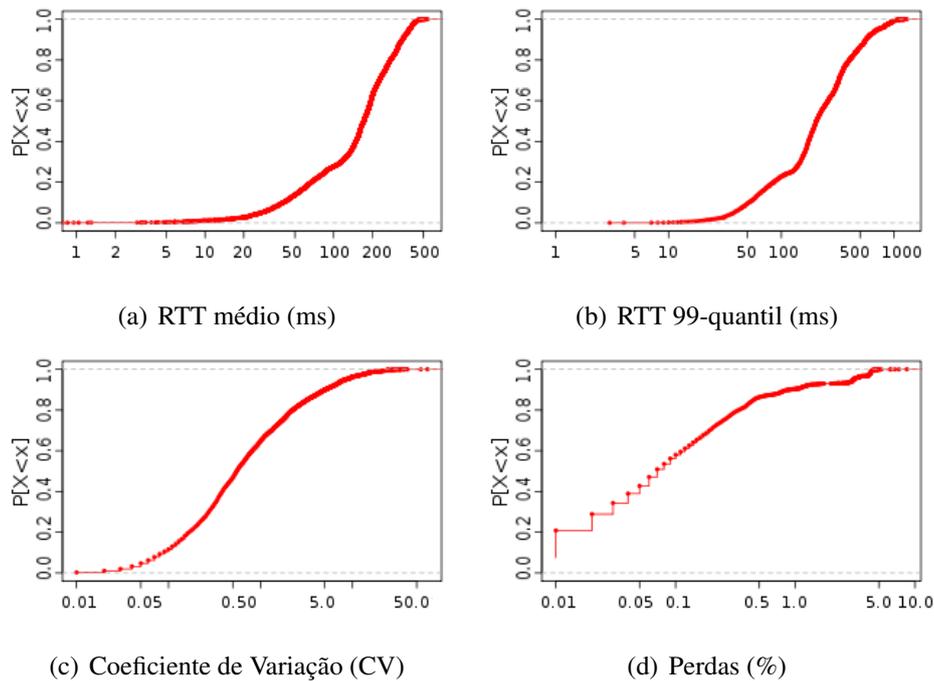
## 4.2. Distribuição Acumulada dos RTTs

Uma outra forma de interpretar os dados é determinando como os tempos de ida-e-volta e as perdas se comportam para cada par de máquinas participantes. Para isso, calculamos as métricas de interesse separadamente para cada par e apresentamos aqui as distribuições acumuladas obtidas.

Para o RTT, foi obtida uma média global de 181 ms. Conforme pode ser observado na Figura 3(a), apenas 40% dos pares apresentaram médias acima de 181 ms, com um máximo de 537 ms para o RTT médio. Se observarmos os resultados para o 99-quantil na Figura 3(b), veremos que em 90% dos pares de máquinas considerados, apenas um por cento dos RTTs medidos ficou acima de 500 ms.

O coeficiente de variação (CV) é apresentado na Figura 3(c) e indica a variabilidade dos dados. É calculado pelo desvio-padrão dividido pela média, de maneira que quanto menor o CV, mais homogêneo é o conjunto de dados. Foram observadas altas variabilidades nos atrasos das mensagens para a maioria dos pares. Uma fração pequena dos pares (menos de 5%) apresentaram CVs abaixo de 0,05; 50% ficaram abaixo de 0,56, e 50% ficaram entre 0,22 e 1,74.

Na Figura 3(d) podemos observar que, em geral, a taxa de perda de mensagens foi baixa. A taxa de perda média foi de 0,41% apenas, cerca de 60% dos pares de máquinas experimentaram perdas inferiores a 0,1% e 90% dos nós apresentaram perdas abaixo de 0,90%, com um máximo de 8,6% de perda de mensagens.



**Figura 3. Distribuições acumuladas das medições**

Quanto aos valores extremos (máximos e mínimos) do RTT, cujos gráficos não foram incluídos por questões de espaço, observamos que a média dos RTTs mínimos foi de 145 ms, com 75% dos nós abaixo de 198 ms e 95% abaixo de 330 ms. O mínimo absoluto foi de 1 ms (entre nós de uma mesma LAN) e o maior valor observado para o RTT mínimo foi de 409 ms. Já a média dos RTTs máximos foi de 27 segundos, com 75% dos nós abaixo de 32 segundos e 95% abaixo de 112 segundos. O máximo absoluto foi de 170 segundos, ou seja, 2,8 minutos de atraso. Esses casos de valores muito altos não se encontravam em rajadas, mas, como mensagens que aparentemente se separaram do fluxo usual e acabaram chegando ao destino muito mais tarde que aquelas enviadas imediatamente antes e depois delas.

### 4.3. Caracterização da Distribuição dos Atrasos

Um dos objetivos deste trabalho era identificar a melhor distribuição estatística que modelasse os atrasos medidos. Tal distribuição poderia ser utilizada em trabalhos que necessitem gerar valores sintéticos de RTT que acompanhem uma distribuição realista. Para determinar a distribuição utilizamos o software Minitab<sup>4</sup>. Devido a suas limitações da quantidade máxima de linhas permitidas e do tempo necessário para o processamento, dos 551 milhões de medidas de RTT coletados obtivemos uma amostra aleatória de 0,1%, ou 551 mil medidas.

A distribuição estatística dos dados pode ser encontrada através de testes padrões para determinar o grau de adequação (*goodness of fit*): Chi-square, Kolmogorov-Smirnov (KS) e Anderson-Darling (AD) [Stephens 1974], dos quais o Minitab utiliza esse último. Há uma grande quantidade de distribuições estatísticas que consistiriam em candidatos a modelo de distribuição dos dados, das quais a Tabela 2 lista as distribuições que se

<sup>4</sup><http://www.minitab.com/>

mostraram mais apropriadas. Na segunda coluna, o coeficiente AD fornece-nos uma medida da qualidade do ajuste. Ao compararmos o ajuste das distribuições para o conjunto de dados, a distribuição com o menor valor de Anderson-Darling oferecerá o melhor ajuste (*best fit*). Porém é necessário que ela rejeite a hipótese nula de que os dados têm distribuição normal, de maneira que o *p-value* deve ser menor ou igual a 0,05.

<i>Distribuição</i>	<i>AD</i>	<i>P-Value</i>	<i>Distribuição</i>	<i>AD</i>	<i>P-Value</i>
3-Par Lognormal	202	*	3-Par Weibull	272	< 0,005
Gamma	207	< 0,005	Weibull	471	< 0,010
3-Par Loglogistic	214	*	Normal	985	< 0,005
Lognormal	267	< 0,005	2-Par Exponential	2376	< 0,010

**Tabela 2. Qualidade do Ajuste da Distribuição dos Atrasos**

Ao observar a Tabela 2 verificamos que as distribuições Lognormal de 3 parâmetros e Gamma foram as que melhor se ajustaram aos atrasos observados. Porém, com a Lognormal de 3 parâmetros não foi possível rejeitar a hipótese de normalidade. Além disso, das duas, a distribuição Gamma foi a que melhor se ajustou aos extremos (atrasos <100 ms e >500 ms). Portanto, propomos a utilização da distribuição Gamma como um modelo de distribuição apropriado para a modelagem de atrasos na rede, provendo um balanço entre melhor ajuste e simplicidade.

**Tabela 3. Estimativas dos Parâmetros da Distribuição dos Atrasos**

<i>Distribuição</i>	<i>Localização <math>\mu</math></i>	<i>Forma <math>\lambda</math></i>	<i>Escala <math>\Theta</math></i>	<i>Limiar</i>
Gamma	-	4,63062	43,16537	-
3-Par Lognormal	5,34111	-	0,41280	-27,12915

A Tabela 3 exibe as estimativas dos parâmetros das distribuições para modelagem do atraso de mensagens. Através dela é possível configurar as distribuições para exibirem atrasos de rede compatíveis com o PlanetLab. Desta maneira, este trabalho fornece tanto uma carga real quanto as distribuições estatísticas mais apropriadas para a modelagem dos atrasos de rede através de uma carga sintética.

#### 4.4. Assimetria dos Atrasos

Em uma análise do progresso evolutivo do RTT (*round-trip time*) médio do PlanetLab, Tang e outros observaram uma considerável variação no RTT entre pares de nós (de 150 a 550 ms), mesmo entre curtos períodos de tempo (15 minutos) [Tang et al. 2007]. Isso apresenta um desafio significativo aos modelos topológicos que pressupõem uma latência fim-a-fim estática entre pares de nós. Em recente análise da assimetria dos atrasos na Internet [Pathak et al. 2008] os autores realizaram um compreensivo estudo no qual mediram a severidade da assimetria no atraso entre diversos pares de nós do PlanetLab. Seus resultados estão de acordo com os de Tang e outros, sendo observado que a assimetria dos atrasos prevalece na grande maioria dos nós, podendo ser em parte atribuída à assimetria nas rotas e em parte a congestionamentos transientes. Observaram também que a assimetria é dinâmica, ou seja, à medida que o tempo progride, a assimetria dos atrasos varia. Além disso, notaram que redes comerciais exibem níveis mais altos de assimetria que redes educacionais e de pesquisa.

Visto que a assimetria dos atrasos na Internet é uma realidade, conduzimos uma análise da assimetria do *trace* coletado. Utilizamos a métrica *assimetria relativa*, que consiste no módulo da diferença entre os atrasos direto e inverso normalizado pelo atraso mínimo, conforme representado na equação 1. Nesta equação, quanto mais próximo de zero for o resultado, mais simétrico será e conseqüentemente, quanto maior for o valor encontrado, maior será a assimetria.

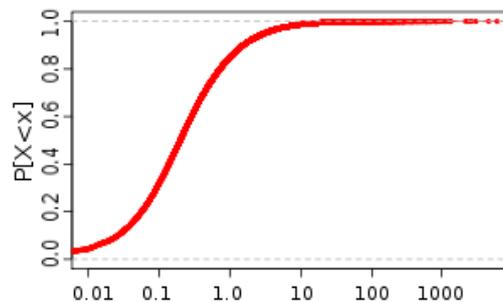
$$assimetria = \frac{|atraso_{AB} - atraso_{BA}|}{\min(atraso_{AB}, atraso_{BA})} \quad (1)$$

Por exemplo, considere que em determinado momento o RTT observado num par de nós foi de 150 ms e o atraso direto de 60 ms. Portanto, o atraso inverso foi de 90 ms o que representa uma assimetria de 30 ms entre esses dois atrasos. Com 60 ms de atraso direto e 90 ms de atraso inverso, temos uma assimetria relativa de  $\text{abs}(60-90)/60=0,50$ . Isso significa que a direção com maior atraso apresentou um acréscimo de 50% em relação ao valor do atraso da outra direção.

<i>Média</i>	<i>Desvio Padrão</i>	<i>CV</i>	<i>Q<sub>25</sub></i>	<i>Mediana</i>	<i>Q<sub>75</sub></i>	<i>Q<sub>90</sub></i>	<i>Q<sub>95</sub></i>	<i>Q<sub>99</sub></i>
0,947	8,78	9,27	0,08	0,20	0,56	1,62	3,08	12,07

**Tabela 4. Estatísticas da Assimetria Relativa dos Atrasos**

Os resultados obtidos para esta métrica estão expressos na Tabela 4 e na Figura 4, onde podemos observar que a assimetria média do PlanetLab é alta (0,947), porém com uma alta variabilidade (9,27). Mesmo a média calculada com a exclusão de 5% dos valores discrepantes ainda pode ser considerada alta, com um valor de 0,409. Metade dos nós apresentou uma assimetria relativa entre 0,08 e 0,56 e 95% deles apresentaram uma assimetria abaixo de 3,08.



**Figura 4. Resumo Estatístico da Assimetria Relativa dos Atrasos**

#### 4.5. Clusterização dos Enlaces

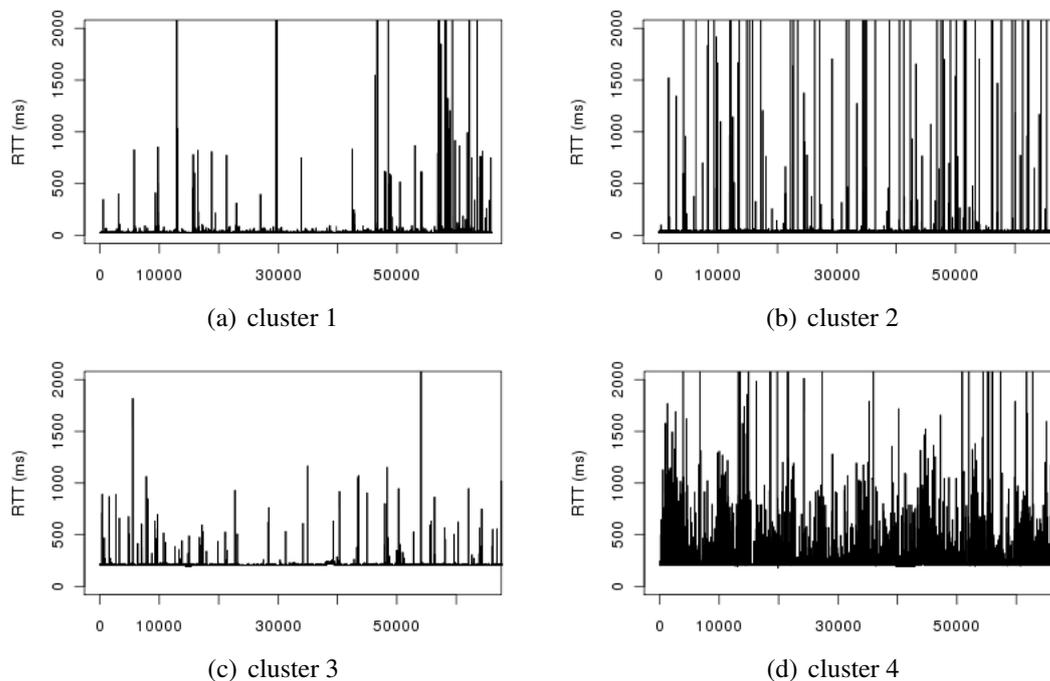
Realizamos também uma clusterização dos enlaces de acordo com suas estatísticas descritivas (atraso médio, seus quantis, variabilidade do atraso e % de perda), buscando separá-los em grupos que, para estas métricas, possuam um comportamento aproximado. Os canais entre duas máquinas A e B foram considerados separadamente em cada sentido, isto é, as medidas de RTT de A para B foram consideradas como um canal de medição e as medidas de B para A como outro canal independente.

**Tabela 5. Estatísticas da Clusterização**

<i>Cluster</i>	<i>RTT Média</i>	<i>Taxa de Perda</i>	<i>RTT Desvio Padrão</i>	<i>RTT CV</i>	<i>% Enlaces</i>
c1	49 ms	0,22%	46 ms	1,12	21%
c2	131 ms	0,40%	673 ms	6,37	21%
c3	167 ms	0,30%	55 ms	0,33	24%
c4	269 ms	0,12%	266 ms	0,96	21%
c5	358 ms	1,40%	158 ms	0,44	13%
Global	181 ms	0,41%	240 ms	1,89	100%

Para a clusterização utilizamos o software *Weka*<sup>5</sup> com o algoritmo de clusterização *Expectation-Maximization* (EM), donde chegamos às 5 instâncias de *clusters* listadas na Tabela 5. Observamos que o cluster 1 apresentou um baixo RTT médio e baixa perda de mensagens, ao contrário do cluster 5, que apresentou atrasos relativamente bem maiores e uma maior taxa de perda de mensagens. O cluster 2 mostrou uma grande variabilidade nos dados, dado seu alto CV, ao passo que o cluster 3 teve a menor variabilidade. Acerca do cluster 4, podemos dizer que ele apresentou um comportamento próximo da média global do PlanetLab.

A fim de exibir o comportamento típico dos RTTs medidos por um nó, selecionamos aleatoriamente uma conexão de cada cluster e geramos a série temporal dos RTTs medidos para aquela conexão. A Figura 5 traz os resultados deste procedimento.

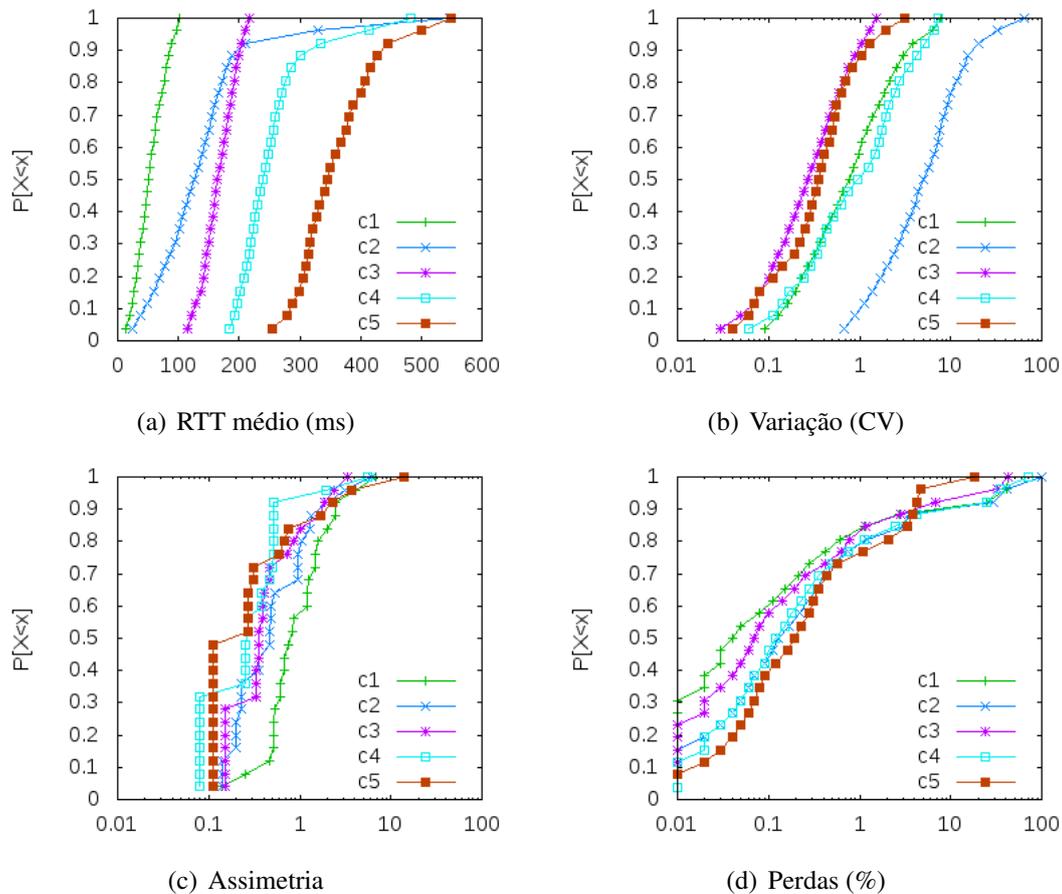
**Figura 5. RTTs ao longo do tempo para os diferentes *clusters***

Pela Figura 5, nos exemplos pode-se observar algumas das características que identificam cada cluster: em (a) os valores baixos da maioria das medidas de RTT para o

<sup>5</sup><http://www.cs.waikato.ac.nz/ml/weka/>

par do cluster 1, porém com muitos picos elevados (alta variabilidade); em (b) a variabilidade ainda mais elevada do par do cluster 2, ainda com um limite inferior baixo; em (c) o menor número de variações nas medidas do cluster 3, porém já com um limite inferior mais alto e, por fim, em (d), para o cluster 4, um valor médio ainda mais alto com significativas variações. Um exemplo de série temporal para o cluster 5 não foi incluído por limitações de espaço, mas também apresenta padrão visualmente diferente dos demais.

Para ilustrar mais claramente as diferenças entre os grupos exibimos abaixo as distribuições acumuladas, já apresentadas para o *trace* completo mas agora com uma curva para cada um dos *clusters* encontrados.



**Figura 6. Análise comparativa dos *clusters***

Na Figura 6, notamos que as curvas de RTT médio (fig. 6(a)) são as que melhor separam os *clusters*. Nas demais, alguns *clusters* apresentam distribuições semelhantes, mas ainda assim é possível observar as separações. As distribuições de perdas, apesar de parecerem próximas, se diferenciam mais claramente nas duas extremidades, pela quantidade de nós que experimentam poucas perdas em cada cluster e pelos valores máximos de perdas em cada caso.

Finalmente, como os enlaces em cada cluster possuem características próximas, seria de se esperar que as medidas dos canais de A para B e de B para A fossem classificados no mesmo cluster. Como as medidas são obtidas pelo *Three-Way Ping*, normalmente, um par de medidas (uma em cada direção) é obtido de cada vez. Usualmente, as duas

medições de RTT em um TWP tendem a ser semelhantes, dado que cada uma envolve uma mensagem em cada direção. Entretanto, oscilações rápidas da rede podem ocorrer durante a primeira ou a terceira mensagens, levando a valores de RTTs ligeiramente diferentes, ou a terceira mensagem pode se perder, o que afetaria apenas a segunda medição de RTT. Tais variações, quando frequentes, podem levar a casos onde as medidas de A para B sejam classificadas em um cluster diferente das de B para A. A Tabela 6 mostra, para cada cluster, de todos os pares de máquinas cujas medidas pertencem àquele cluster, qual a porcentagem daqueles que foram incluídos nas duas direções.

<i>Cluster</i>	c1	c2	c3	c4	c5
Pares mantidos juntos (%)	86	67	86	84	90

**Tabela 6. Distribuição dos canais entre pares de máquinas pelos *clusters***

O cluster 2, que teve a maior quantidade de pares com canais também atribuídos a outro grupo, foi aquele que teve a maior variabilidade (CV), muito acima dos demais *clusters*. Portanto, essa maior variabilidade pode ter contribuído para maior dispersão dos seus elementos.

## 5. Conclusão e trabalhos futuros

Tempos de ida-e-volta são importantes em diversas situações na Internet: eles podem, por exemplo, determinar a seleção de certos pares ou servidores em aplicações distribuídas, afetando a forma como participantes do sistema se organizam. Por esse motivo, informações sobre RTTs são importantes para orientar o projeto de protocolos e como entrada para sistemas de simulação, entre outros fatores. Neste trabalho apresentamos uma metodologia para medição de tempos de ida-e-volta pela Internet por um período de dez dias, com um intervalo entre medições baixo (10 segundos), utilizando 81 máquinas distribuídas pela Internet em quatro continentes.

As medições foram avaliadas em termos dos RTTs médios, sua variabilidade ao longo do tempo entre um mesmo par de máquinas e, em qualquer instante, entre pares diferentes, bem como as perdas observadas em cada canal de medição. O padrão de variação dos RTTs ao longo do tempo foi modelado como uma distribuição Gamma, cujos parâmetros foram identificados, podendo ser usada na geração de atrasos em ambientes de simulação, por exemplo. Com base no uso de relógios sincronizados avaliamos também o atraso unidirecional entre as diversas máquinas, encontrando uma prevalência significativa de assimetria nos canais. Finalmente, uma análise dos *traces* utilizando uma técnica de clusterização mostrou a existência de cinco grupos, caracterizados por diferentes combinações de valor médio de RTT, taxas de perda e variabilidade das medições, entre outros fatores. A inspeção de algumas amostras de cada grupo e a análise das distribuições acumuladas das diversas métricas em cada caso confirmam as diferenças entre os grupos e composições inter e intracontinentes.

Identificamos até o presente momento pelo menos duas linhas para trabalhos futuros. Por um lado, pretendemos investigar os fatores de rede que levaram à formação dos grupos identificados. Esperamos que isso leve a um modelo mais detalhado do comportamento da rede quanto aos tempos de ida-e-volta, com caracterização dos *clusters* separadamente. Paralelamente, pretendemos reproduzir o experimento em maior escala,

com mais pontos de medição e por períodos mais longos. A análise de intervalos maiores ou variáveis se encontram entre os trabalhos que estamos desenvolvendo no momento, junto com o projeto de uma versão adaptativa da malha de monitoração que permita a realização de medidas entre um número maior de máquinas. Para tal utilizaremos uma topologia do tipo par-a-par para minimizar a complexidade quadrática da quantidade de nós intercomunicantes.

### **Disponibilidade dos dados**

O *trace* completo, bem como os programas de medição e coleta dos TWP podem ser obtidos em <http://www.speed.dcc.ufmg.br/>.

### **Agradecimentos**

Esta pesquisa foi parcialmente financiada pelo Instituto Nacional de Ciência e Tecnologia para a Web - InWeb (MCT/CNPq 573871/2008-6), pelo Projeto REBU (CTInfo/CNPq 55.0995/2007-2), pela CAPES e pela FAPEMIG.

### **Referências**

- Choi, J. H. and Yoo, C. (2005). One-way delay estimation and its application. *SIGCOMM Computer Communication Review*, 28(7):819–828.
- Park, K. and Pai, V. S. (2006). Comon: a mostly-scalable monitoring system for planetlab. *SIGOPS Operating Systems Review*, 40(1):65–74.
- Pathak, A., Pucha, H., Zhang, Y., Hu, Y. C., and Mao, Z. M. (2008). A measurement study of internet delay asymmetry. In *Proceedings of the 9th International Conference on Passive and Active Network Measurement (PAM '08)*, volume 4979 of *Lecture Notes in Computer Science*, pages 182–191, Cleveland, OH, EUA. Springer.
- Rocha, A. R., Leão, R., and de Souza e Silva, E. (2004). Metodologia para estimar o atraso em um sentido e experimentos na Internet. In *Anais do XXII Simpósio Brasileiro de Redes de Computadores (SBRC 2004)*, pages 589–602, Gramado.
- Stephens, M. (1974). EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, pages 730–737.
- Tang, L., Chen, Y., Li, F., Zhang, H., and Li, J. (2007). Empirical study on the evolution of planetlab. In *Proceedings of the 2007 International Conference on Networking (ICN '07)*, page 64, Los Alamitos, CA, EUA. IEEE Computer Society.
- Wille, E., Yabcznski, E., and da Costa Bento, C. R. (2008). Algoritmos para solução do problema de dimensionamento de buffers em roteadores IP. In *Anais do XXVI Simpósio Brasileiro de Redes de Computadores (SBRC 2008)*, pages 217–229, Rio de Janeiro.