Estimando a média e variância do atraso em um sentido utilizando o IPID da máquina remota*

Antonio A. de A. Rocha, Rosa M. M. Leão, Edmundo de Souza e Silva

COPPE/Prog. de Engenharia de Sistemas e Computação
e Departamento de Ciência da Computação do IM
Universidade Federal do Rio de Janeiro
Caixa Postal 68511 – Rio de Janeiro, RJ – 21941-972
{arocha, rosam, edmundo}@land.ufrj.br

Resumo. *A área de medições ativas é de extrema importância para inferir características importantes da Internet. Dentre as mais úteis variáveis de interesse está o retardo em um sentido entre dois nós da rede. No caso em que o analista tem acesso a máquinas localizadas no vértice do caminho a ser estudado, e ainda se essas máquinas possuírem seus relógios perfeitamente sincronizados, o cálculo desta medida é trivial. Porém, o processo de estimação requer algoritmos sofisticados caso as máquinas envolvidas na medição não possuam hardware de GPS. O problema torna-se ainda mais complexo quando o analista não tem acesso à máquina localizada no final do caminho. Neste trabalho propomos uma nova técnica para lidar com ambos os problemas de falta de acesso e falta de sincronismo, de forma a estimar a média e a variância da variável retardo em um sentido. A técnica faz uso do campo IPID do cabeçalho de pacotes IP.*

Abstract. *Active measurements are an extremely useful tool to infer important characteristics of the Internet. Among the most important variables of interest is the one way packet delay. If the analyst has access to machines located at the vertices of the studied path and these machines have their clocks perfectly synchronized (for instance with GPS equipment), then calculating the one way delay statistics is trivial. However, the estimation process requires sophisticated algorithms if the machines involved in the measured process have no GPS hardware. The issue is even more complex if the analyst has no access to the machine located at the end of the path. In this work we propose a novel technique to handle both of these problems in order to estimate the average and the variance of the one-way delay variable. The technique takes advantage of the IPID field in the header of the IP packet.*

1. Introdução

A Internet é um sistema complexo, a serviço de milhões de usuários, com variadas demandas por aplicações de diferentes requisitos de serviços. Aplicações multimídia, por exemplo, possuem estreitos requisitos para medidas de desempenho fim-a-fim, tais como atraso, *jitter* e probabilidade de perda de pacotes. A crescente demanda por tais aplicações torna o conhecimento das características da rede fundamental para seu uso na Internet. Para que se possa entender melhor essas características e com isso prover melhores

*Este trabalho é parcialmente financiado pela Finep, CNPq e Faperj.

serviços para as aplicações é importante a realização de medições e a criação de modelos baseados nestas medições.

As técnicas de medição ativa são baseadas no envio de sondas a partir de fontes escolhidas, e na coleta destas sondas pelas próprias fontes ou por uma ou mais máquinas receptoras. No caso em que a máquina receptora não é a origem, as métricas estimadas são referentes ao caminho da rede percorrido pelas sondas "em um sentido". Quando as sondas enviadas não são coletadas pela máquina alvo e sim replicadas de volta à máquina de origem, as métricas estimadas são relacionadas aos caminhos de "ida e volta" percorridos pelos pacotes.

Na Internet atual, os caminhos de ida e volta entre duas máquinas podem ser assimétricos. Isto é, as capacidades dos roteadores em um sentido podem ser diferentes das capacidades dos roteadores no sentido oposto, ou ainda, a seqüência de roteadores percorridos pode ser diferente. Mesmo quando a seqüência de roteadores for a mesma e a capacidade deles simétrica, os caminhos podem apresentar características de desempenho completamente diferentes devido a assimetria do tráfego (e consequentemente do tamanho das filas) dos roteadores. Por isso, medir os caminhos de forma independente permite identificar o desempenho da rede em cada um dos sentidos. No entanto, as técnicas para a medição dos caminhos de ida e volta quando comparadas às técnicas de medição em um sentido, em geral, são mais simples. Estimar o atraso e a taxa de perda na ida e volta dos pacotes na rede, por exemplo, é trivial utilizando ferramentas como PING. Isso porque é comum nas máquinas conectadas à Internet estar habilitada a função de *ICMP echo reply* em resposta ao recebimento de um *ICMP echo request*.

A estimativa do atraso e a taxa de perda em um sentido, normalmente necessitam da execução de processos na máquina remota para coletar as sondas recebidas. Informações como chegadas com sucesso e instantes de chegada devem ser computadas para cada pacote coletado. A não ser que dispositivos específicos para sincronização de relógios como *GPS(Global Positioning System)* sejam utilizados, o cálculo do atraso em um sentido requer um tratamento especial às diferenças existentes entre os relógios envolvidos na medição [Rocha et al. 2004]. Portanto, as ferramentas existentes na literatura que estimam essas métricas em um sentido necessitam de permissão para execução do processo coletor na máquina remota onde são computadas as informações referentes às chegadas das sondas.

Recentes trabalhos propõem novas técnicas que possibilitam a um usuário final, sem privilégios especiais, medir algumas características de desempenho dos caminhos de rede em um sentido. A partir de informações contidas no campo de identificação do cabeçalho IP *(IPID)*, as técnicas permitem identificar a taxa de perda [Mahajan et al. 2003, Savage 1999], chegadas fora de ordem [Mahajan et al. 2003, Bellardo and Savage 2002], e as diferenças entre os atrasos de duas máquinas fonte para uma máquina alvo [Chen et al.].

Neste trabalho é proposta uma técnica para inferir a média e a variância da distribuição do atraso em um sentido, utilizando apenas processos executados em máquinas geradoras de sondas. Como a proposta utiliza o campo IPID ao invés do instante de chegada das sondas, processos de coleta na máquina remota não são requeridos. Esta técnica possibilita que medições de atraso em um sentido sejam executadas na Internet independente de acesso e permissão de execução de coletas na máquina remota. A técnica assume que as

sondas são geradas a partir de duas máquinas fonte para uma mesma máquina alvo. Essas sondas não são coletadas quando chegam à máquina remota, elas são replicadas de volta às máquinas de origem. A implementação desta técnica pode ser facilmente feita utilizando mensagens de *echo request* e *echo reply* do protocolo *ICMP*. Inicialmente iremos supor que as duas máquinas de origem estão com seus relógios sincronizados através de GPS. Em seguida demonstraremos como a técnica pode ser aplicada mesmo sem o uso dos dispositivos de sincronização nas máquinas geradoras.

A organização deste trabalho é feita da seguinte forma. Na Seção 2 é feita uma breve descrição dos métodos que usam o campo IPID para se obter as medidas de interesse. A definição dos problemas gerais para a estimativa do atraso em um sentido utilizando valores do IPID e a técnica proposta neste trabalho são apresentadas na Seção 3. Os problemas resultantes da falta de sincronismo entre os relógios das duas máquinas de origem na estimativa do atraso são descritos na Seção 4. Nesta seção é ainda demonstrada a extensão da técnica para evitar o uso de equipamentos de sincronização. A Seção 5 é dedicada à validação do método proposto: resultados de simulações e de experimentos realizados na Internet serão mostrados. Na Seção 6 é apresentado um sumário das contribuições e trabalhos futuros.

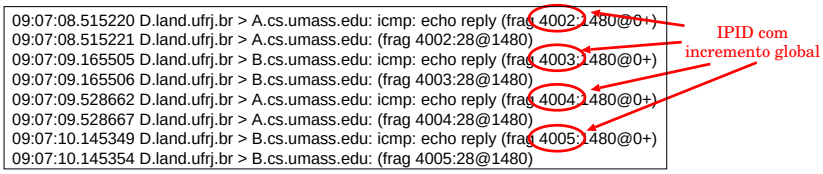
2. Medições utilizando campo IPID

O IPID é um campo de identificação existente no cabeçalho de pacotes do protocolo IP [Postel 1981b]. Este campo fornece uma identificação que é utilizada pelo processo de fragmentação e remontagem de datagramas na Internet. Ocupando 16-bits do cabeçalho IP, este identificador, juntamente com outras informações contidas também no cabeçalho IP, possibilitam identificar pacotes pertencentes a um mesmo datagrama que tenha sido fragmentado.

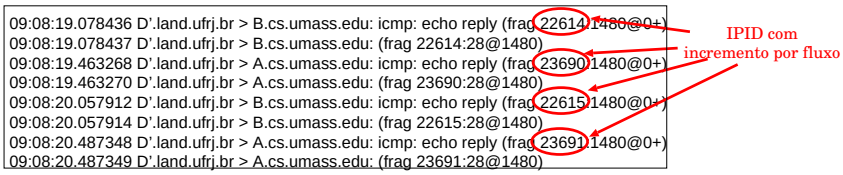
Embora a utilização do IPID na fragmentação e remontagem dos datagramas seja um padrão na Internet, o padrão não define uma regra ao uso deste campo. A forma como os valores de identificação do datagrama IP são incrementados depende da implementação do sistema operacional. Diversas máquinas na Internet programam o IPID com um simples contador global. Isso inclui as máquinas servidas com sistemas operacionais Windows, Freebsd, Mac OS e Linux até a versão 2.2 do kernel. As versões mais atuais do Linux, Solaris e Openbsd implementam um contador pseudo-aleatório para cada fluxo.

Um simples experimento com as gerações de sondas originadas de duas máquinas quaisquer para uma máquina remota permite identificar que tipo de implementação no IPID é utilizada pelo sistema operacional deste alvo. A Figura 1 ilustra dois logs obtidos com TCPDUMP executado no roteador de saída da rede do nosso laboratório (LAND-COPPE/UFRJ). (Para possibilitar o registro do campo IPID no log do TCPDUMP sondas foram geradas com tamanho superior a 1480 *bytes* forçando a fragmentação dos datagramas na fonte.) O primeiro log mostra pacotes de *ICMP echo reply* destinados a duas máquinas diferentes em resposta a sondas de *ICMP echo request* previamente enviadas a uma máquina com sistema operacional Windows XP. O outro log mostra os pacotes *echo reply* gerados por uma máquina com sistema operacional Linux com kernel 2.6. No primeiro log é possível verificar o crescimento global dos valores do IPID gerados pela máquina remota. Já no segundo log, existe um crescimento apenas nos valores do IPID relativos a cada fluxo. (Por uma questão de segurança, os nomes reais das máquinas foram

aqui substituídos por nomes fictícios.) Ferramentas para auditoria de segurança de rede utilizam técnicas semelhantes explorando essa característica do IPID para identificação do sistema operacional [Insecure.org 1998] e execução de *port scan* [Insecure.org 1997] em máquinas alvo.



(A) Log do TCPDUMP de pacotes enviados da máquina com Windows XP



(B) Log do TCPDUMP de pacotes enviados da máquina com Linux 2.6

Figura 1. Logs do TCPDUMP executado no roteador de saída da rede.

Outros trabalhos têm explorado os valores coletados do campo IPID para a obtenção de características da rede. [Chen et al.] faz um estudo de técnicas de inferência de várias medidas com uso do IPID. Em [Chen et al.] os autores definem três categorias de aplicações para as técnicas existentes: medição de atividade do tráfego [Insecure.org 1997]; agrupamento de fontes [Bellovin 2002, Insecure.org 1997]; e, identificação de perda, duplicação e chegada fora de ordem [Mahajan et al. 2003, Bellardo and Savage 2002]. Além desta classificação, o trabalho de [Chen et al.] propõe três novas técnicas para o uso do IPID, uma referente a cada uma destas classes.

Observando a variação do IPID de sondas recebidas de uma máquina fonte é possível estimar o tráfego em um dado intervalo de tempo, desde que a máquina destino implemente um contador global para o IPID. Em [Insecure.org 1997] é apresentada uma proposta para estimar o volume de tráfego para um servidor através de medições ativas. Sondas são enviadas para a máquina alvo (e.g., uma máquina receptora) e capturados os IPIDs dos pacotes de resposta. Seja $IPID(i)$ o valor de IPID capturado da sonda i e $T(i)$ o instante de chegada destas respostas. O número de requisições recebidas por um servidor entre os instantes de tempo $T(i)$ e $T(i+1)$ é $\Delta IPID(i)$ e equivale à diferença dos valores $IPID(i)$ e $IPID(i+1)$ se os valores de IPID pudessem crescer infinitamente. Entretanto, como o campo IPID possui um tamanho máximo de 16 bits, as técnicas que usam os valores do IPID devem levar em consideração que o incremento do valor deste campo volta a zero ao atingir 2^{16}.

Uma abordagem semelhante a [Insecure.org 1997] é apresentada em [Chen et al.] para estimar o volume de tráfego de um servidor. A diferença entre as propostas [Insecure.org 1997] e [Chen et al.] é que a segunda técnica utiliza medição passiva ao invés de medição ativa para observação do IPID gerado pelo servidor medido. A vantagem deste método em relação ao anterior é o menor *overhead*, uma vez que sondas não são geradas. No entanto, é necessário permissão para a coleta de dados no roteador do canal de saída deste servidor.

O campo IPID foi explorado também em propostas para identificar o número de servidores utilizados por um sistema de balanceamento de carga [Chen et al. , Insecure.org 1997] e o número de máquinas por detrás de um serviço *NAT(Network Address Translator)*

[Bellovin 2002]. Os métodos supõem que dois pacotes gerados por uma mesma máquina em um curto intervalo de tempo devem apresentar um valor pequeno para o $\Delta IPID$. Se cada servidor do sistema de balanceamento de carga possui um contador global independente, pacotes gerados por um servidor possuem uma seqüência do IPID diferente da seqüência dos pacotes gerados por outro servidor. Observando valores coletados do IPID, as técnicas de [Insecure.org 1997, Chen et al.] tentam identificar essas independências entre as seqüências e estimar o número de servidores utilizados para o balanceamento de carga. Embora essa técnica tenha sido sugerida em [Insecure.org 1997], apenas em [Chen et al.] foi definido um algoritmo. Técnica semelhante é utilizada em [Bellovin 2002] para detectar servidores NAT e contabilizar o número de máquinas em atividade por trás desses servidores.

Recentemente alguns trabalhos propuseram novas técnicas que possibilitam medir características de desempenho da rede, a partir dos IPID. Essas técnicas permitem identificar, dentre outras medidas, a taxa de perda e chegadas fora de ordem [Mahajan et al. 2003, Bellardo and Savage 2002]. Embora as sondas utilizadas pelas técnicas sejam geradas e coletadas na mesma máquina, os valores do IPID obtidos da máquina remota permitem a estimativa destas métricas em um sentido.

Em [Chen et al.] foi proposta uma técnica para determinar a diferença entre os atrasos de sondas enviadas de máquinas fontes para uma máquina alvo. Na técnica proposta máquinas com relógios sincronizados com GPS enviam sondas para uma máquina remota. A máquina alvo, que não precisa estar com seu relógio sincronizado com as demais, replica as sondas para as máquinas de origem incluindo no campo IPID os valores referentes ao contador global desta máquina. Intuitivamente se duas sondas enviadas de máquinas diferentes retornarem às máquinas de origem com valores próximos de IPID, essas sondas chegaram à máquina remota em instantes muito próximos de tempo.

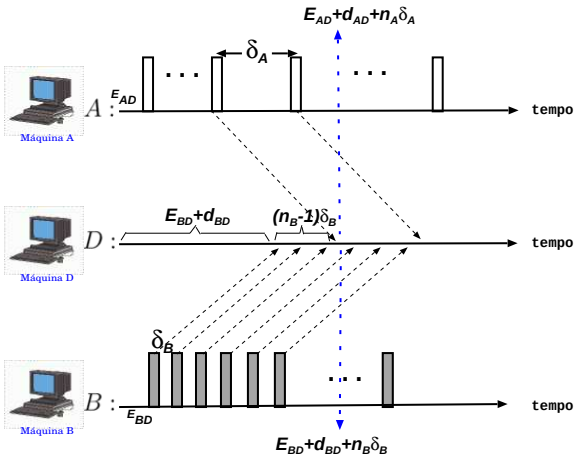


Figura 2. Figura utilizada em [Chen et al.] para ilustrar a técnica.

A Figura 2 retirada de [Chen et al.] ilustra a técnica. Considere duas máquinas A e B com relógios sincronizados gerando sondas para a máquina alvo D. As sondas são geradas por A e B a intervalos constantes iguais a δ_A e a δ_B, respectivamente. Para um δ_B próximo de zero, dois pacotes enviados de B chegarão à máquina D praticamente no mesmo instante. É possível supor então que esses dois pacotes experimentaram o mesmo atraso na rede. Suponha que o n_A-ésimo pacote enviado por A chegue a D entre os pacotes $n_B - 1$ e n_B enviados por B. Sejam d_{AD} e d_{BD} os atrasos experimentados pelos pacotes de A para D e de B para D respectivamente. Então:

$$E_{BD} + d_{BD} + (n_B - 1)\delta_B \leq E_{AD} + d_{AD} + (n_A)\delta_A \leq E_{BD} + d_{BD} + (n_B)\delta_B$$

$$E_{BD} - E_{AD} + n_B\delta_B - n_A\delta_A - \delta_B \leq d_{AD} - d_{BD} \leq E_{BD} - E_{AD} + n_B\delta_B - n_A\delta_A$$

Sendo: $E_{BD}(n_B) = E_{BD} + n_B\delta_B$ e $E_{AD}(n_A) = E_{AD} + n_A\delta_A$

$$E_{BD}(n_B) - E_{AD}(n_A) - \delta_B \leq d_{AD} - d_{BD} \leq E_{BD}(n_B) - E_{AD}(n_A)$$

Note que os limites máximo e mínimo dependem de δ_B. Logo, quanto menor o valor de δ_B mais estreita é a diferença entre os limites inferior e superior. Dessa forma, para δ_B pequeno em [Chen et al.] a diferença entre os atrasos em um sentido pode ser estimada pelos instantes de envio das sondas:

$$d_{AD} - d_{BD} = E_{BD}(n_B) - E_{AD}(n_A)$$

3. Proposta para estimar a média e variância do atraso

Embora a medida da diferença entre os atrasos estimada em [Chen et al.] seja útil para algumas aplicações, a medida "atraso em um sentido" encontra um número maior de aplicações. Por outro lado, essa medida é bem mais difícil de ser estimada, a não ser no caso em que as máquinas envolvidas estão sincronizadas e se tem acesso a todas elas. Este então é o problema que abordamos e será o foco desta seção. Apresentaremos uma nova técnica para inferir a média e a variância da distribuição do atraso em um sentido baseando-se nos resultados da Seção 2.

Como na Seção 2, supomos que as sondas são geradas a partir de duas (ou mais) máquinas fonte para uma mesma máquina alvo. As sondas não são coletadas pela máquina remota e são replicadas de volta às máquinas de origem. Como a proposta utiliza o campo IPID ao invés do instante de chegada das sondas, processos de coleta na máquina remota não são requeridos possibilitando que a medida seja feita para qualquer máquina na Internet que implemente um contador global para o campo IPID.

Inicialmente supomos que as duas máquinas de origem estão com seus relógios sincronizados através de GPS. Na Seção 4 estenderemos os resultados de forma a dispensar o uso de dispositivos de sincronização.

3.1. Definição do problema

Considere as máquinas A e B com relógios sincronizados gerando sondas para a máquina alvo D, conforme ilustra a Figura 3. As sondas que chegam muito próximas umas das outras à máquina alvo apresentam valores próximos para o IPID. Para cada amostra coletada em A e em B que chegaram de D com valores próximos de IPID podemos montar o seguinte sistema de equações:

$$\begin{cases} d_{AD} + d_{DA} = RTT_{ADA} \\ d_{BD} + d_{DB} = RTT_{BDB} \\ d_{AD} - d_{BD} = \Psi_{AD-BD} \\ d_{DA} - d_{DB} = \Psi_{DA-DB} \end{cases}$$

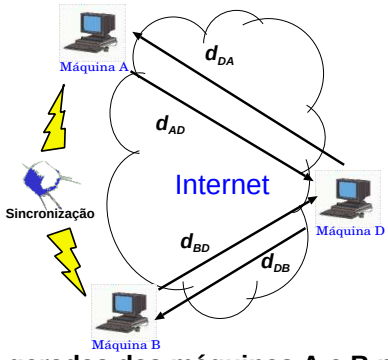


Figura 3. Sondas geradas das máquinas A e B para a máquina D.

onde, Ψ_{AD-BD} é a diferença entre os atrasos de A para D e de B para D; Ψ_{DA-DB} é a diferença entre os atrasos de D para A e de D para B; e, RTT_{ADA} e RTT_{BDB} são os atrasos de ida e volta computados para as amostras enviadas de A e de B, respectivamente.

O sistema formado pelas equações é linearmente dependente, possível e indeterminado. Logo, esse sistema apresenta infinitas soluções. A técnica definida neste trabalho consiste em restringir o espaço de soluções deste sistema estimando o atraso em um sentido quando as sondas enviadas por A ou as sondas enviadas por B não encontrarem fila nos caminhos de ida e volta. Desta forma é possível redefinir esse sistema removendo das variáveis d_{AD}, d_{DA}, d_{BD}, e d_{DB} os valores referentes a atrasos constantes na rede, tais como o tempo de transmissão e propagação. Neste caso, quando os tempos em filas nos caminhos de ida e volta entre as máquinas B e D forem iguais a zero, podemos resolver o sistema e estimar o atraso sofrido pelas sondas em cada um dos sentidos(d_{AD}, d_{DA}, d_{BD} e d_{DB}). Repetindo-se os cálculos quando os tempos em filas nos caminhos entre as máquinas A e D forem iguais a zero é possível encontrar a solução procurada, conforme será detalhado abaixo.

3.2. Estimando os tempos em fila

O atraso sofrido por um pacote na rede é formado basicamente pela soma dos tempos de transmissão (T^{tx}), propagação (T^{prop}), processamento (T^{proc}) e fila (T^{fila}). Considerando que o tempo de processamento é desprezível em relação aos demais, então o atraso d_{AD} é:

$$d_{AD} = T^{tx}_{AD} + T^{prop}_{AD} + T^{fila}_{AD}$$

Assumindo que o mesmo caminho de rede é utilizado por todas as sondas durante a medição entre as máquinas A e D e que todas as sondas sejam de mesmo tamanho, é possível afirmar que: (i) Os tempos de transmissão e propagação serão idênticos para cada sonda; (ii) Aquelas que obtiverem o menor atraso equivalem as sondas que supostamente não entraram em fila durante todo o caminho percorrido.

Seja $d_{AD}(n)$ o atraso obtido pela n-ésima sonda e seja $d_{m,AD}$ o menor valor de atraso computado dentre todas as amostras. O tempo em fila sofrido pela n-ésima sonda pode ser estimado através da diferença entre o atraso desta amostra e o menor atraso computado dentre todas as amostras do experimento.

$$T^{fila}_{AD}(n) = d_{AD}(n) - d_{m,AD}$$

Sejam $RTT_{m,iji}$ o tempo mínimo de ida-e-volta de i-j-i obtido considerando-se todas as amostras. A partir de $RTT_{m,ADA}$ e $RTT_{m,BDB}$, respectivamente, podemos re-

escrever as equações $d_{AD} + d_{DA} = RTT_{ADA}$ e $d_{BD} + d_{DB} = RTT_{BDB}$ da seguinte forma:

$$T_{AD}^{fila} + T_{DA}^{fila} = T_{ADA}^{fila} \qquad e \qquad T_{BD}^{fila} + T_{DB}^{fila} = T_{BDB}^{fila}$$

Para reescrever as equações $d_{AD} - d_{BD} = \Psi_{AD-BD}$ e $d_{DA} - d_{DB} = \Psi_{DA-DB}$ em função dos tempos em fila são necessários os tempos de transmissão e propagação em cada um dos sentidos. Porém, o menor valor do atraso em um sentido não é conhecido. Como as capacidades de transmissão dos caminhos de ida e de volta podem ser assimétricas, os tempos de transmissão e propagação em um dos sentidos não podem ser obtidos com os menores valores de atraso de ida e volta computados entre A e D e entre B e D, denotados respectivamente por $RTT_{m,ADA}$ e $RTT_{m,BDB}$.

Uma abordagem que permite tratar a assimetria das capacidades nos caminhos de ida e volta e estimar os tempos de transmissão em cada um dos sentidos foi utilizada. Essa técnica considera que as capacidades de transmissão em cada um dos sentidos são diferentes, mas a propagação nos caminhos são aproximadamente iguais. Para estimar o tempo de transmissão e o tempo de propagação em cada sentido, as sondas geradas devem seguir alguns padrões definidos. Para uma medição feita entre as máquinas A e D sondas são enviadas por A e replicadas por D com tamanho igual a por exemplo 50 *bytes*, sondas são enviadas por A e replicadas por D com tamanho igual a 500 *bytes* (número bem maior que 50 *bytes*), e sondas são enviadas por A com tamanho igual a 500 *bytes* e replicadas por D com tamanho igual a 50 *bytes*. Considerando que o menor atraso para cada tamanho diferente de sonda obedece a uma função linear, é possível estimar os tempos de transmissão em cada sentido.

A implementação desta técnica pode ser feita utilizando o protocolo *ICMP* através das mensagens do tipo *echo request* e *echo reply*. A especificação deste protocolo apresentada em [Postel 1981a] define que mensagens do tipo *ICMP echo request* oriundas de outras máquinas devem ser respondidas com uma mensagem do tipo *ICMP echo reply*. De acordo com as especificações, para formar uma mensagem de *echo reply* uma máquina deve simplesmente inverter os endereços de origem e destino, alterar o código do tipo da mensagem *ICMP* de 8 (*echo request*) para 0 (*echo reply*) e recalcular o *checksum*. Os dados originais devem ser mantidos preservando assim o tamanho da mensagem de resposta. Dessa forma, sondas de mesmo tamanho podem ser enviadas e recebidas.

Como as especificações do protocolo *ICMP* não permitem ao emissor da mensagem de *echo request* um controle do tamanho das mensagens de *echo reply* que devem ser enviadas pelo receptor, um método utilizando pares de pacotes pode ser aplicado para emular o efeito do envio de um pacote de 500 *bytes* e o recebimento de uma resposta de 50 *bytes*. O método consiste na emissão de dois pacotes separados por um intervalo de tempo próximo de zero, o primeiro com 500 *bytes* e o segundo com 50 *bytes*. Os pacotes atravessam o mesmo caminho de rede até chegarem a um único destino. Supondo que esses dois pacotes seguiram juntos ao longo de todo o caminho, no sentido de ida o segundo pacote será atrasado a cada salto pelo tempo de transmissão de um pacote de 500 *bytes* e portanto permanecerá bem atrás deste. Ao chegarem à máquina destino, o primeiro pacote será descartado e o segundo será enviado de volta para a máquina de origem. Para que o primeiro pacote seja descartado, um endereço IP de origem falso é utilizado pela máquina emissora na mensagem de *echo request*. Com isso, no sentido de volta o pacote não sofrerá o retardo do pacote maior. Dessa forma, podemos assumir que o menor atraso

experimentado por uma sonda com esta técnica será igual ao tempo de propagação nos dois sentidos somado ao tempo de transmissão de um pacote de 500 *bytes* no caminho de ida e ao tempo de transmissão de um pacote de 50 *bytes* no caminho de volta.

Sejam $RTT_{m,ADA}^{50-50}$, $RTT_{m,ADA}^{500-500}$ e $RTT_{m,ADA}^{500-50}$ os menores valores de atraso de ida e volta estimados para os experimentos com sondas dos tamanhos especificados. Considerando que os tempos de propagação são aproximadamente iguais nos dois sentidos $T_{AD}^{prop} \approx T_{DA}^{prop}$). É fácil verificar que:

$$\begin{cases} T_{AD}^{tx} + T_{DA}^{tx} + 2T_{AD}^{prop} = RTT_{m,ADA}^{50-50} \\ 10T_{AD}^{tx} + 10T_{DA}^{tx} + 2T_{AD}^{prop} = RTT_{m,ADA}^{500-500} \\ 10T_{AD}^{tx} + T_{DA}^{tx} + 2T_{AD}^{prop} = RTT_{m,ADA}^{500-50} \end{cases}$$

Onde, o valor "10" é devido ao tamanho do pacote de 500 *bytes*, 10 vezes maior que o de 50 *bytes*.

A solução deste sistema é simples e fornece uma estimativa para os tempos de transmissão em cada um dos sentidos entre as máquinas A e D. De forma semelhante o tempo de transmissão em cada sentido pode ser calculado para os caminhos BD e DB.

3.3. Estimando a média e variância da distribuição do atraso em um sentido

Utilizando a abordagem descrita é possível rever o problema definido anteriormente para estimar o atraso em um sentido. (Apenas para manter uma compatibilidade com a subseção anterior, assumimos que as sondas têm tamanho igual a 50 *bytes*.)

Seja Ψ_{AD-BD}^{fila} a diferença dos tempos em fila sofridos pelas sondas no caminho de A para D e de B para D e seja Ψ_{DA-DB}^{fila} a diferença dos tempos em fila de D para A e de D para B. Estimados os tempos de transmissão e propagação sofridos pelas sondas em cada um dos caminhos de rede, esses valores podem ser estimados por:

$$\Psi_{AD-BD}^{fila} = \Psi_{AD-BD} - (T_{AD}^{tx} + T_{AD}^{prop} + T_{BD}^{tx} + T_{BD}^{prop})$$

$$\Psi_{DA-DB}^{fila} = \Psi_{DA-DB} - (T_{DA}^{tx} + T_{DA}^{prop} + T_{DB}^{tx} + T_{DB}^{prop})$$

O sistema previamente definido pode assim ser reformulado da seguinte forma, agora com o espaço de soluções mais reduzido:

$$\begin{cases} T_{AD}^{fila} + T_{DA}^{fila} = T_{ADA}^{fila} \\ T_{BD}^{fila} + T_{DB}^{fila} = T_{BDB}^{fila} \\ T_{AD}^{fila} - T_{BD}^{fila} = \Psi_{AD-BD}^{fila} \\ T_{DA}^{fila} - T_{DB}^{fila} = \Psi_{DA-DB}^{fila} \end{cases}$$

onde, T_{ADA}^{fila} e T_{BDB}^{fila} são os retardos experimentados pelas sondas nas filas nos caminhos de ida e volta ADA e BDB, respectivamente.

Quando os tempos em filas nos caminhos de ida e volta ADA ou BDB forem iguais a zero, é possível resolver o sistema e estimar o atraso sofrido pelas sondas em cada um dos sentidos (d_{AD}, d_{DA}, d_{BD} e d_{DB}). Isto é, se $T_{ADA}^{fila} = 0$, então $T_{BD}^{fila} = \Psi_{AD-BD}^{fila}$ e $T_{DB}^{fila} = \Psi_{DA-DB}^{fila}$. Da mesma forma, caso $T_{BDB}^{fila} = 0$, $T_{AD}^{fila} = \Psi_{AD-BD}^{fila}$ e

$T_{DA}^{fila} = \Psi_{DA-DB}^{fila}$. Somados os tempos em fila aos tempos de transmissão e propagação calculados previamente são obtidos os atrasos em cada sentido.

Para inferir a média e a variância da distribuição do atraso em um sentido, diversas amostras deste atraso são estimadas. Supondo que, de todas as sondas geradas entre as máquinas A e D e entre B e D, i amostras originadas de A e B chegaram a D com valores de IPID muito próximos; e que, dessas i amostras, o atraso em cada sentido foi estimado para j sondas. Sejam $d_{AD}(n)$, $d_{DA}(n)$, $d_{BD}(n)$ e $d_{DB}(n)$ os atrasos em um sentido estimados para a n-ésima dessas j amostras, a média e a variância amostral da distribuição do atraso em cada sentido são calculadas por:

$$\overline{d}_{sentido} = \frac{1}{j}\sum_{n=1}^{j} d_{sentido}(n) \quad e \quad Var(d_{sentido}) = \frac{1}{j-1}\sum_{n=1}^{j}\left(d_{sentido}(n) - \overline{d}_{sentido}\right)^2$$

onde, "sentido" representa o caminho desejado da métrica: AD, DA, BD ou DB

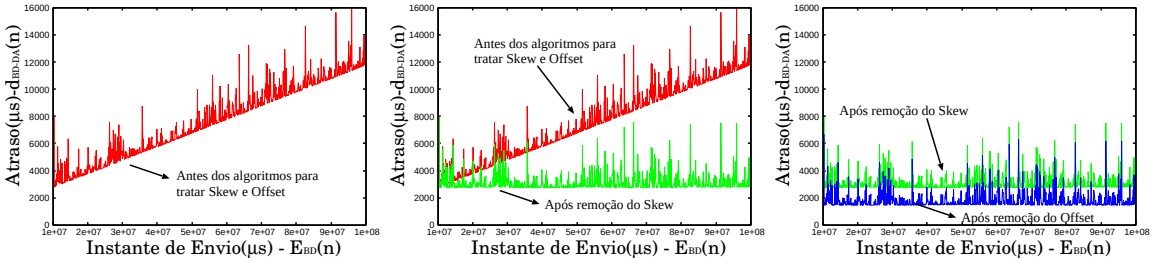
4. Extensão da técnica sem o uso de GPS

A técnica descrita acima pressupõe o uso de sondas geradas por máquinas com relógios sincronizados. Nesta seção será demonstrado como a técnica proposta neste trabalho pode ser estendida para o caso em que não exista a sincronia entre os relógios. Os problemas para estimar o atraso em um sentido de pacotes, quando não é garantida a sincronia dos relógios, já foram amplamente discutidos na literatura, assim como soluções já foram propostas [Loung and Biro 2000, Moon et al. 1999, Paxson 1998, Rocha et al. 2004, Tsuru et al. 2002, Zhang et al. 2002].

O primeiro problema, chamado de *Offset*, surge em conseqüência dos relógios das máquinas envolvidas na medição possuírem valores distintos no início da medição. O valor dessa diferença é somado ou diminuído do valor real do atraso. O segundo, chamado de *Skew*, é resultante da diferença na taxa de crescimento dos relógios das máquinas. Considerando que os relógios não são atômicos, a taxa do relógio em uma máquina pode ser maior ou menor do que na outra. Em conseqüência, o resultado do cálculo do atraso entre duas máquinas sofre um crescimento ou decrescimento constante. Quando o experimento é executado por um tempo maior que poucos segundos, o erro causado pela diferença nas taxas de crescimento dos relógios é significativo.

Uma abordagem nova foi definida para tratar os problemas de *Skew* e *Offset* entre os relógios nas coletas das máquinas A e B. A técnica é uma extensão de [Rocha et al. 2004], uma vez que naquele artigo sondas são geradas de uma máquina A para B e o algoritmo proposto remove o *Skew* e *Offset* entre A e B. Entretanto, neste trabalho, não são geradas sondas de A para B ou vice-versa, mas sim de A e B para uma máquina alvo D.

Para estimar o *Skew* a seguinte seqüência deve ser considerada. Seja $\Omega := [v_n = (E_{BD}(n), d_{BD-DA}(n)) : n = 1, \ldots, i]$ uma seqüência obtida das coletas das i sondas que chegaram à máquina D em instantes próximos, onde $E_{BD}(n)$ equivale ao instante de envio da n-ésima sonda por B e $d_{BD-DA}(n)$ equivale à diferença entre o instante de recebimento da n-ésima sonda pela máquina A e o instante de envio da n-ésima sonda pela máquina B. A diferença básica entre o problema tratado em [Rocha et al. 2004] e o deste artigo é que neste último é preciso trabalhar com a seqüência $d_{BD-DA}(n)$. Note que



(A) (B) (C)

Figura 4. Tratamento dos problemas de Skew e Offset nas coletas.

os instantes de tempo utilizados na seqüência são referentes aos instantes nos relógios das respectivas máquinas e estes não estão sincronizados.

Na Figura 4(A) é possível verificar a tendência de crescimento causada pelas diferenças nas taxas dos relógios das coletas de um experimento que será descrito na próxima seção. A seqüência gerada desta forma permite identificar um limitante inferior para os valores de $d_{BD-DA}(n)$. Esse limite é definido pela soma dos tempos de transmissão e propagação nos caminhos de B para D e de D para A, acrescido dos valores causados pelo *Skew* e *Offset*. Assim com em [Rocha et al. 2004] o objetivo é estimar uma função linear que esteja abaixo e mais próxima possível de todos os pontos em Ω para representar a tendência de crescimento ou decrescimento entre os relógios das máquinas.

Tratado o problema da diferença entre as taxas de crescimento dos relógios, uma nova seqüência γ é então gerada após o cálculo do atraso sem *Skew* (d^s_{BD-DA}) para todas as i sondas. Esta seqüência está ilustrada na Figura 4(B). É importante perceber que, como os relógios não se encontram sincronizados no início da medição, os valores estimados de d^s_{BD-DA} na seqüência γ contém o *Offset* inicial da coleta. Portanto, podemos assumir que $d^s_{BD-DA}(n) = T^{tx}_{BD}(n) + T^{prop}_{BD}(n) + T^{fila}_{BD}(n) + T^{tx}_{DA}(n) + T^{prop}_{DA} + T^{fila}_{DA}(n) + O_{AB}$

O algoritmo apresentado em [Tsuru et al. 2002] e utilizado em [Rocha et al. 2004] poderia ser utilizado para estimar e remover o *Offset* da coleta. No entanto, sondas deveriam ser geradas da máquina A para a máquina B e vice-versa. Evitando que sondas extras sejam geradas, a estimativa do *Offset* pode ser feita a partir da diferença entre os menores valores computados para RTT_{BDB} e d_{BD-DA} dentre todas as i amostras. Se considerarmos que os menores valores destas amostras representam o caso em que estas sondas não experimentaram fila ao longo dos seus caminhos de rede, podemos definir $d^s_{m,BD-DA}$ como sendo o menor valor de d^s_{BD-DA} existente entre as i sondas da seqüência γ e $RTT_{m,BDB}$ como sendo o menor valor do atraso de ida e volta computado para as sondas enviadas de B para D. Assim,

$$RTT_{m,BDB} = T^{tx}_{BD} + T^{prop}_{BD} + T^{tx}_{DB} + T^{prop}_{DB}$$

$$d^s_{m,BD-DA} = T^{tx}_{BD} + T^{prop}_{BD} + T^{tx}_{DA} + T^{prop}_{DA} + O_{AB}$$

A diferença entre $RTT_{m,BDB}$ e $d_{m,BD-DA}$ pode ser definida por:

$$RTT_{m,BDB} - d_{m,BD-DA} = (T^{tx}_{DB} + T^{prop}_{DB}) - (T^{tx}_{DA} + T^{prop}_{DA}) + O_{AB}$$

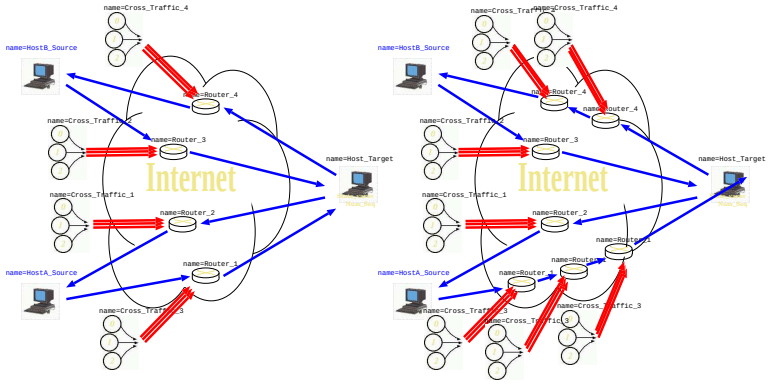
Como os valores dos tempos de transmissão e propagação em cada sentido são conhecidos, independente da existência ou não de problemas como *Skew* e *Offset*. Então, é possível estimar o O_{AB} da seguinte forma:

$$O_{AB} = (RTT_{m,BDB} - d_{m,BD-DA}) - (T^{tx}_{DB} + T^{prop}_{DB}) + (T^{tx}_{DA} + T^{prop}_{DA})$$

A Figura 4(C) ilustra os valores da seqüência γ após removidos os valores de *Offset*. Estimados os valores para O_{AB} e α_{AB} é possível calcular as diferenças entre os atrasos em um sentido de máquinas não sincronizadas.

5. Validação

A fim de validar a técnica proposta e avaliar a sua eficácia, desenvolvemos modelos de simulação e realizamos experimentos na Internet. Os resultados obtidos serão apresentados nesta seção.



(A) Cenário 1 (B) Cenário 2

Figura 5. Cenários de modelos utilizados nas simulações.

Dois modelos de simulação foram desenvolvidos no ambiente *TANGRAM-II* [Carmo et al. 1998, de Souza e Silva and Leão 2003]. A Figura 5 ilustra os dois cenários modelados. A diferença entre os dois modelos é o número de roteadores nos caminhos entre as máquinas. No primeiro cenário quatro roteadores são definidos, no segundo cenário dois novos roteadores foram inseridos no caminho AD e um no caminho DB.

Nos modelos os objetos *Host_A* e *Host_B* representam as máquinas geradoras de sondas. Cada máquina envia 100 sondas por segundo. As sondas são enviadas pelos caminhos da rede à máquina alvo, representada no modelo pelo objeto *Host_Target*. Quando recebida pela máquina alvo, as sondas são replicadas e enviadas pela rede às máquinas de origem contendo o valor atual do contador global (*IPID*) do *Host_Target*. No objeto *Host_Target* traces são gerados contendo o instante de chegada das sondas para que sejam comparados os valores reais dos atrasos com os valores estimados pela técnica desenvolvida neste trabalho.

Distintas capacidades de transmissão foram atribuídas aos canais ligados aos roteadores. Pelos canais que ligam os roteadores trafegam sondas e tráfego concorrente. As sondas geradas pelas máquinas fonte e replicadas pela máquina alvo são encaminhadas aos seus destinos ou ao próximo roteador no caso do segundo cenário. Já os pacotes de tráfego concorrente são roteados para outros caminhos da rede.

O tráfego concorrente injetado em cada roteador da rede é gerado por diversas fontes *On-Off*. O tempo de permanência nos estados *On* e *Off* dessas fontes é modelado por uma distribuição Pareto com parâmetro $\alpha < 2$. Em [Taqqu et al. 1997] foi mostrado que a agregação destas fontes produz um tráfego com características de dependência de longa duração e que este modelo é adequado para caracterizar o tráfego real de uma rede.

Diversas simulações foram executadas variando-se os parâmetros das fontes. O

objetivo foi avaliar o comportamento da técnica proposta para diferentes utilizações dos roteadores nos caminhos de rede. Por questões de limitação de espaço, três resultados de simulações serão apresentados, dois do primeiro cenário e um do segundo.

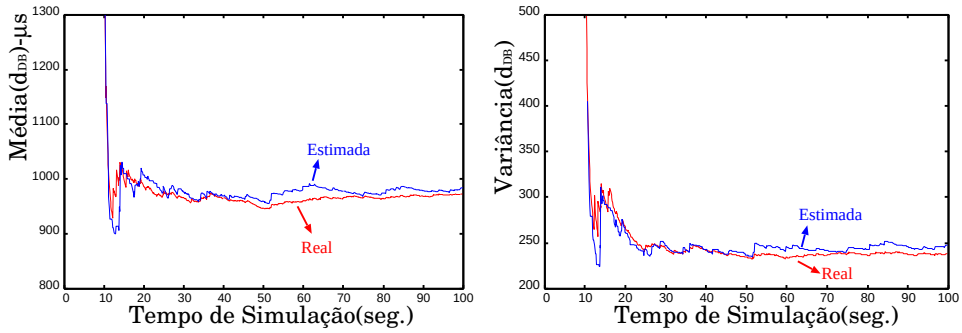


Figura 6. Média e variância do caminho DB - primeiro cenário - utilização entre 40% e 50%.

Em uma das simulações executadas para o primeiro cenário, a utilização dos canais ao longo do tempo de simulação variou de 40% a 50% (intervalo típico de operação de uma rede). As Figuras 6 (A) e (B) ilustram a média e variância amostral estimadas pela técnica proposta e a real do caminho DB, após 100 segundos de simulação. As estimativas da média e variância para os demais caminhos encontram-se na tabela abaixo. Os resultados obtidos demonstram que as estimativas obtidas da média e variância estão bem próximas do real ao final da simulação.

Caminho	Média	Variância
	Estimativa/Real	Estimativa/Real
AD	492(μs)/455(μs)	148/110
DA	1023(μs)/932(μs)	229/175
BD	728(μs)/727(μs)	285/253
DB	984(μs)/972(μs)	248/238

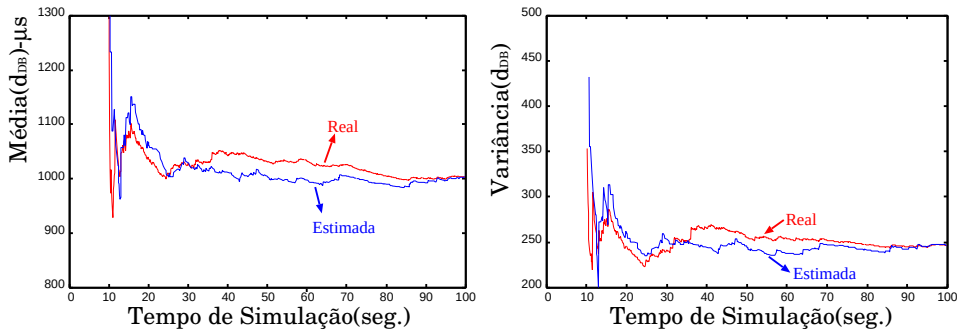
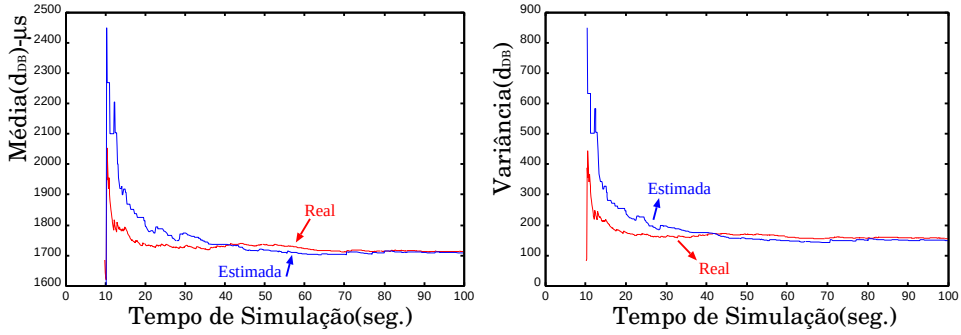


Figura 7. Média e variância do caminho DB - primeiro cenário - utilização entre 50% e 60%.

Os gráficos ilustrados na Figura 7, demonstram os resultados obtidos para o primeiro cenário considerando a carga nos roteadores em torno de 50 e 60%. As demais estimativas encontram-se na tabela abaixo. Com os resultados apresentados na figura é possível verificar que as medidas estimadas continuam convergindo para os valores reais

mesmo para uma carga maior na rede.

Caminho	Média	Variância
	Estimativa/Real	Estimativa/Real
AD	579(μs)/517(μs)	187/147
DA	968(μs)/918(μs)	208/167
BD	973(μs)/945(μs)	412/358
DB	1004(μs)/1003(μs)	246/246



(A) (B)

Figura 8. Média e variância do caminho DB - segundo cenário - utilização entre 30% e 50%.

A Figura 8 refere-se ao segundo cenário e as utilizações dos canais variaram entre 30% e 50%. A Figura ilustra as estimativas feitas para o caminho DB. Para este segundo cenário, existem dois roteadores entre as máquinas D e B. As estimativas para os demais caminhos são descritos na tabela.

Caminho	Média	Variância
	Estimativa/Real	Estimativa/Real
AD	1366(μs)/1338(μs)	129/137
DA	942(μs)/929(μs)	173/175
BD	765(μs)/755(μs)	124/143
DB	1707(μs)/1712(μs)	155/150

Experimentos utilizando a Internet foram realizados entre o nosso laboratório (UFRJ), a University of Massachusetts at Amherst (UMass), e um laboratório na UNIFACS, visando avaliar a técnica desenvolvida neste trabalho em um cenário real. Para estes experimentos, as máquinas localizadas na UNIFACS e na UMass foram utilizadas como geradoras de sondas para a máquina alvo localizada em nosso laboratório. Um módulo da ferramenta *TANGRAM-II Traffic Generator* foi adaptado para emular o protocolo *ICMP* e as mensagens de *echo request* e *echo reply*. Sondas *UDP* são enviadas pelo gerador da ferramenta que ao chegarem na máquina alvo são replicadas de volta à máquina de origem. Como os relógios das máquinas não estão em sincronia, os problemas de *Skew* e *Offset* foram tratados.

Para validar os resultados, experimentos adicionais foram feitos simultaneamente aos acima utilizando a técnica apresentada em [Rocha et al. 2004]. A média e variância estimada com os experimentos adicionais permitem uma comparação destas estimativas com os valores obtidos com a técnica desenvolvida aqui neste trabalho.

Os resultados obtidos estão descritos na tabela abaixo. Utilizamos na tabela os termos “Estimativa” para indicar os resultados obtidos com a técnica apresentada neste trabalho e “Real” para indicar aqueles obtidos com os experimentos adicionais com a

técnica [Rocha et al. 2004] e cuja precisão foi investigada naquele trabalho. Na tabela, não são apresentados os valores referentes a variância do retardo entre UFRJ e Unifacs, pois o número de amostras obtidas neste caminho foi pequeno (79) para que se tenha um resultado estatisticamente confiável. Mesmo assim os valores para a média são razoáveis. No caso do caminho UFRJ-UMass, o número de amostra foi pouco superior a 200 e neste caso a precisão do método foi excelente.

Caminho	Média Estimativa/Real	Erro Relativo da média	Variância Estimativa/Real
UMass-UFRJ	139.071(ms)/139.311(ms)	0.17%	2049.922/1565.246
UFRJ-UMass	123.253(ms)/122.726(ms)	0.42%	2445.292/1545.679
Unifacs-UFRJ	42.378(ms)/47.935(ms)	11.59%	X
UFRJ-Unifacs	34.577(ms)/31.136(ms)	11.05%	X

É importante enfatizar que a técnica deste trabalho não necessita de processo executando na máquina remota e, portanto, pode ser empregada para medir o atraso de uma máquina a qualquer outra da rede desde que a máquina destino tenha um sistema operacional que, como o Windows, implemente um contador global.

6. Resumo das contribuições e trabalhos futuros

Estimar métricas como média e variância da distribuição do atraso de um caminho de rede, em um determinado sentido, é de primordial importância para o dimensionamento de aplicações com estreitos requisitos de *QoS* e parametrizar modelos de forma a entender melhor as características das redes e o comportamento de novos métodos de gerenciamento dos recursos. Portanto, é essencial o desenvolvimento de técnicas para estimar com precisão essas métricas através de medições. Neste trabalho foi apresentada uma nova proposta para inferir a média e a variância da distribuição do atraso em um sentido usando medições ativas. A proposta baseia-se na utilização do campo IPID dos pacotes enviados pela máquina alvo, ao invés do instante de chegada das sondas. Desta forma, processos de coleta na máquina remota não são necessários. Essa técnica permite que medições de atraso em um sentido sejam executadas na Internet **independentemente** de se ter acesso e permissão de coleta na máquina remota alvo.

Inicialmente assumimos que as máquinas de origem estão com seus relógios sincronizados. Em seguida, uma solução para evitar equipamentos de sincronização foi apresentada.

Modelos de simulação foram elaborados para avaliar e validar a técnica proposta. Os resultados das simulações mostraram que as métricas média e variância do atraso em um sentido de um caminho de rede podem ser estimadas com precisão utilizando a metodologia deste trabalho. Através de simulações, mostramos que a técnica pode ser aplicada independente da sincronização entre os relógios das máquinas de origem das medições.

Foi também realizado um experimento visando obter uma estimativa preliminar da técnica proposta em um ambiente real. Um maior número de experimentos está programado para avaliar a precisão da proposta entre diversos pontos.

Referências

Bellardo, J. and Savage, S. (2002). Measuring Packet Reordering. In *2nd ACM SIGCOMM Internet Measurement Workshop(IMW)*, Marseille, France.

Bellovin, S. (2002). A Technique for Counting NATed Hosts. In *2nd ACM SIGCOMM Internet Measurement Workshop(IMW)*, pages 267–272, Marseille, France.

Carmo, R., de Carvalho, L., de Souza e Silva, E., Diniz, M., and Muntz, R. (1998). Performance/Availability Modeling with the TANGRAM-II Modeling Environment. *Performance Evaluation*, 33:45–65.

Chen, W., Huang, Y., Ribeiro, B., Suh, K., Zhang, H., de Souza e Silva, E., Kurose, J., and Towsley, D. Exploiting the IPID Field to Infer Network Path and End-System Characteristics. In *Lecture Notes in Computer Science*, volume 3431, pages 108–120.

de Souza e Silva, E. and Leão, R. (2003). Modelagem e Análise de Redes com o Conjunto de Ferramentas TANGRAM II. In *XXI SBRC - Workshop de Ferramentas*, pages 897–904, Natal, Brasil.

Insecure.org (1997). Idle Scanning and related IPID games. http://www.insecure.org/nmap/idlescan.html.

Insecure.org (1998). Remote OS detection via TCP/IP Stack FingerPrinting. http://www.insecure.org/nmap/nmap-fingerprinting-article.txt.

Loung, D. and Biro, J. (2000). Needed Services for Network Performance Evaluation. In *IFIP Workshop on Performance Modeling and Evaluation of ATM Networks*, Inglaterra.

Mahajan, R., Spring, N., Wetherall, D., and Anderson, T. (2003). User-level internet path diagnosis. In *19th ACM / Symposium on Operating Systems Principles (SOSP)*, pages 106–119, Bolton Landing, NY, USA.

Moon, S., Skelly, P., and Towsley, D. (1999). Estimation and Removal of Clock Skew for Network Delay Measurements. In *IEEE/Infocom*, pages 227–234. New York, USA.

Paxson, V. (1998). On Calibrating Measurements of Packet Transit Times. In *ACM/Sigmetrics*, pages 11–21, Madison, Wisconsin, USA.

Postel, J. (1981a). Internet Control Message Protocol. IETF RFC 792.

Postel, J. (1981b). Internet Protocol. IETF RFC 791.

Rocha, A., Leão, R., and de Souza e Silva, E. (2004). Metodologia para Estimar o Atraso em um Sentido e Experimentos na Internet. In *XXII Simpósio Brasileiro de Redes de Computadores*, Gramado, Brasil.

Savage, S. (1999). Sting: a TCP-based Network Measurement Tool. In *USENIX Symposium on Internet Technologies and Systems*, pages 71–79, Boulder, CO, USA.

Taqqu, M. S., Willinger, W., and Sherman, R. (1997). Proof of a Fundamental Result in Self-Similar Traffic Modeling. In *ACM/Computer Communications Review*, pages 5–23.

Tsuru, M., Takine, T., and Oie, Y. (2002). Estimation of Clock Offset from One-way Delay Measurement on Asymmetric Paths. In *SAINT International Symposium on Applications and the Internet*, pages 126–133, Nara, Japão.

Zhang, L., Liu, Z., and Xia, C. (2002). Clock Synchronization Algorithms for Network Measurements. In *IEEE/Infocom*, pages 160–169, New York, USA.