

# Metodologia para estimar o atraso em um sentido e experimentos na Internet\*

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

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

**Resumo.** Determinar as características de uma rede tais como a distribuição de retardos e o processo de perda de pacotes, não é uma tarefa trivial. Entretanto, essas medidas são essenciais para a construção de modelos que visem, por exemplo, estudar a eficácia de mecanismos de controle para aumentar a eficiência de uma transmissão, ou ainda técnicas para compensar a baixa qualidade de serviço oferecida por uma rede com características de melhor esforço. Algumas características como o retardo de ida e volta (round trip time - RTT) podem ser facilmente obtidas. Porém, outras são difíceis de serem estimadas, como o atraso dos pacotes em um único sentido (one-way delay). Sem o emprego de equipamentos tais como GPS, o cálculo dessa medida requer um tratamento especial dos dados devido à falta de sincronismo entre os relógios das máquinas envolvidas na medição. Neste trabalho, foram avaliadas técnicas existentes na literatura para remoção de skew e offset de relógios, necessárias para o cálculo do atraso em sentido único, e identificados dois problemas que causam erros na estimativa desta medida de interesse. Uma solução para os problemas encontrados é proposta e avaliada. A metodologia para o cálculo do retardo em sentido único foi implementada em uma ferramenta que foi usada na realização de uma série de experimentos adicionais visando caracterizar a distribuição do retardo entre três pontos na Internet: dois no Brasil e um no exterior. Estes resultados são úteis na escolha dos valores de parâmetros de um modelo de desempenho que inclua os atrasos inerentes à rede.

**Abstract.** Obtaining performance measures of a communication network such as delay distributions or packet loss probabilities is not an easy task. However, these measures are essential to build models in order to study, for instance, the efficacy of control mechanisms to increase the throughput of a connection or yet to study techniques aiming at compensating the poor quality of service offered by best-effort networks. Some measures such as the round trip time (RTT) can be easily obtained in most cases. On the other hand, other measures such as the one-way delay may be difficult to estimate. Without the use of special hardware, such as GPS equipment, it is necessary to employ techniques to compensate for the lack of synchronism between the clocks involved in the measurement. In this work we evaluate existing techniques to remove the clock skew and offset which are necessary to calculate the one-way delay. We also identify two problems that introduce errors in the calculation of this measure. A solution to overcome these problems is proposed and evaluated. The methodology to calculate the one-way delay was implemented in a tool that was used in a series of additional experiments to characterize the one-way delay distribution between two out of three points in the Internet: two in Brazil and another in the USA. These results are useful to choose the proper parameter values of a performance model that includes packet transmission delays.

---

\*Este trabalho é parcialmente financiado pelo CNPq e FAPERJ.

## 1. Introdução

Aplicações multimídia possuem estreitos requisitos de desempenho, tais como retardo, latência, *jitter*, probabilidade de perda de pacotes. Porém a Internet foi concebida para fornecer apenas um serviço do tipo *melhor esforço*, não provendo garantias para que uma aplicação tenha a mínima qualidade de serviço (QoS) necessária ao seu bom funcionamento. Em conseqüência, é importante que se construam modelos de forma a poder entender a eficácia dos mecanismos de compartilhamento de recursos, prever a QoS obtida pelas aplicações, e ainda elaborar mecanismos a nível de aplicação visando compensar a ausência de garantias de QoS da rede. Para que os modelos possam ser corretamente parametrizados, é necessário medir e caracterizar uma variedade de parâmetros da rede. O retardo entre dois pontos está dentre os mais importantes destes parâmetros.

Existem basicamente duas formas de se fazer medições na rede: passiva e ativa [8]. As medições passivas necessitam de equipamentos especiais de coleta de medidas. Esses equipamentos observam os pacotes que trafegam em um determinado canal de comunicação e coletam as informações relevantes. As técnicas de medição ativa são baseadas no envio de sondas (*probes*) a partir de determinadas fontes e na coleta destas sondas pelas próprias fontes ou por outras máquinas receptoras. A grande vantagem destas técnicas é que não é necessário o uso de equipamentos específicos para a coleta de dados da rede. No entanto, as medidas obtidas podem não ter a precisão adequada devido a problemas como a falta de sincronização entre os relógios das máquinas envolvidas na medição e a interferência do tráfego dos usuários sobre as sondas geradas para a medição. Este trabalho versa sobre medições ativas e, em particular, aquelas voltadas à obtenção do retardo (ou atraso) de um ponto a outro na rede (chamado de atraso em sentido único ou *one way delay*), devido a sua importância e aos desafios técnicos para a obtenção desta métrica.

A dificuldade para a estimativa do atraso em sentido único tem motivado diversos estudos na literatura. O cálculo deste atraso é conceitualmente simples, e igual a diferença entre os instantes de tempo coletados nas máquinas receptora e emissora. Esta diferença produz o retardo correto se os relógios destas máquinas se encontram *perfeitamente* sincronizados. No entanto, relógios distintos nunca estão *perfeitamente* sincronizados, a não ser que sejam empregados equipamentos especiais de sincronização como o *GPS (Global Positioning System)* nas máquinas envolvidas na medição.

Diversos algoritmos tem sido propostos para tratar da falta de sincronia entre dois relógios [3, 5, 6, 7, 10, 11]. Estes algoritmos podem, a princípio, ser usados na estimativa do atraso em sentido único sem a necessidade de uma infra-estrutura especial. Entretanto, outros problemas existem. Por exemplo, em [7, 11] foi observado que as técnicas propostas para estimar o atraso em sentido único podem obter resultados incorretos caso os relógios das máquinas envolvidas sejam atualizados durante o processo de medição. (O que não é incomum devido ao uso de protocolos de atualização de relógio, por exemplo o NTP.) Estes trabalhos propõem também soluções para resolver este problema. No entanto, estas novas soluções ainda podem fornecer estimativas incorretas do atraso, como será detalhado nas seções subseqüentes deste trabalho. Além da atualização de relógios, experimentos que realizamos revelaram a existência de um outro evento que pode causar erro na estimativa do atraso em sentido único. A aplicação coletora de sondas pode deixar de receber temporariamente, do sistema operacional, as sondas enviadas, causando erro na estimativa da medida atraso em sentido único.

São vários os objetivos deste trabalho. Inicialmente são avaliadas as técnicas existentes na literatura de remoção de *Skew* e *Offset* com o propósito de calcular o retardo em sentido único, baseado no paradigma de medições ativas. São então identificados dois problemas que resultam em erros na estimativa desta medida. Em seguida é proposta uma solução para esses problemas, e avaliada a sua robustez no cálculo de retardo em sentido único, através de uma série de experimentos. Os algoritmos necessários para o cálculo desta medida (incluindo os propostos neste

trabalho) foram implementados na ferramenta TANGRAM-II [1, 2], no módulo *Traffic Generator* [8, 9], e com esta ferramenta realizados os experimentos apresentados. Finalmente, experimentos complementares foram realizados entre o nosso laboratório (LAND-COPPE/UFRJ), a University of Massachusetts at Amherst (UMass - CS Department), e a laboratório NUPERC/UNIFACS, visando caracterizar a distribuição do retardo entre esses pontos. Esses resultados servem como exemplo de aplicação da ferramenta de forma a obter os parâmetros necessários à construção de um modelo de desempenho, que inclua os atrasos inerentes à Internet.

A organização deste trabalho é feita da seguinte forma. Na seção 2 são apresentados os problemas gerais para estimativa do atraso em sentido único causados pela falta de sincronia dos relógios e os trabalhos relacionados. Outros problemas para estimar esta medida são apresentados na seção 3. Nesta seção é ainda proposto um método robusto para solucionar esses problemas adicionais. A seção 4 é dedicada à implementação e validação do método proposto. A ferramenta desenvolvida é aplicada em experimentos visando caracterizar a distribuição de retardo entre a UFRJ e UMass e UFRJ e UNIFACS, e os resultados são apresentados na seção 5. A seção 6 apresenta nossas conclusões.

## 2. Algoritmos para tratar dos problemas de *Skew* e *Offset*

Inicialmente serão descritos nesta seção os problemas de *Skew* e *Offset* e algumas soluções existentes na literatura serão apresentadas. Em uma segunda etapa serão descritos em maior detalhe os algoritmos com melhor desempenho para tratar destes problemas.

### 2.1. Definição do problema e algumas soluções

O atraso experimentado por um pacote na rede pode ser feito de forma trivial quando as máquinas envolvidas na medição estão ligadas a um equipamento GPS para sincronizar seus relógios, ou quando a origem e o destino do pacote são a mesma máquina. Neste caso,  $D = T_r - T_s$ , onde  $T_s$  é o instante de envio do pacote e  $T_r$  o instante de chegada. Entretanto, nem sempre equipamentos de GPS estão disponíveis nas máquinas origem-destino dos pacotes. Portanto, os relógios não estão *perfeitamente* sincronizados e, em consequência, o valor absoluto dos relógios, assim como as taxas de incremento de cada um, são diferentes. A solução via GPS, embora simples, não é de baixo custo e ainda restringe possíveis experimentos ao conjunto de máquinas com esses equipamentos.

A Figura 1 ilustra o problema decorrente da falta de sincronia entre os relógios. Foram feitas medições entre duas máquinas com sondas geradas nos dois sentidos. Para cada sentido, uma seqüência  $\Omega := [v_i = (i, d_i) : i = 1, \dots, N]$  foi calculada a partir das sondas enviada, onde  $i$  equivale ao número de seqüência da  $i$ -ésima sonda enviada e  $d_i$  ao atraso obtido pela simples diferença entre os tempos de envio e recebimento da sonda  $i$ . As seqüências são mostradas na Figura 1.

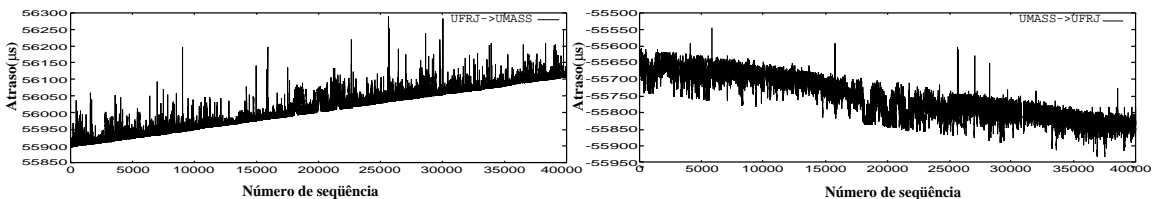


Figura 1: Atraso de pacotes entre máquinas com relógios diferentes.

O primeiro problema, chamado de *Offset*, aparece em consequência dos relógios das máquinas envolvidas na medição possuírem valores distintos no início da medição. O valor dessa diferença pode estar sendo somado ou diminuído do valor total do atraso, resultando até mesmo em valores negativos para as estimativas  $d_i$ . O segundo, chamado de *Skew*, é resultante da diferença na

taxa de crescimento dos relógios das máquinas. Considerando que os relógios não são atômicos, o intervalo de um microssegundo em uma máquina pode ser maior ou menor do que o intervalo de um microssegundo em 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.

Existem algumas propostas na literatura para possibilitar o cálculo do atraso em um sentido. Uma possível opção é o uso de protocolos tipo NTP(*Network Time Protocol*) [4] para sincronizar os relógios. A configuração desses protocolos requer acesso de administrador sobre as máquinas que ainda devem estar sincronizadas a um mesmo servidor NTP. Outro problema desta abordagem são os retardos na transmissão das informações de sincronismo enviadas do servidor para cada uma das máquinas envolvidas, uma vez que estes retardos podem ser bastante diferentes. A freqüência no envio das informações de sincronismo pelo servidor constitui ainda outro problema. Esta freqüência não é alta o suficiente para permitir a sincronização necessária ao cálculo preciso da medida atraso em sentido único. Uma outra proposta é a aplicação de algoritmos para remover o *Skew* e *Offset* dos dados coletados.

## 2.2. Algoritmos para remoção do *Skew*

Várias são as propostas para solucionar o problema do *Skew* [5, 7, 11]. O objetivo de todas elas é estimar uma função linear que esteja abaixo e mais próxima possível de todos os pontos em  $\Omega$  para representar a tendência de crescimento ou decrescimento do atraso. A função linear é obtida através da solução de um problema de otimização. Um exemplo de função objetivo definida em [5] é dado por: minimizar a soma das distâncias verticais entre os vértices  $v_i$  e a reta da função linear.

Em [5], Moon e Towsley, propõem o uso de um algoritmo de programação linear para estimar a função linear. Além de proporem o novo método, fazem uma comparação entre esse e o proposto por Paxson [7]. Na avaliação dos algoritmos é demonstrado um fraco desempenho no quesito robustez por parte da proposta de Paxson, sendo verificado que em caso de altos valores do *Skew* o algoritmo falha na estimativa desse parâmetro.

Uma avaliação dos algoritmos e uma nova proposta é também apresentada por Liu e Zhang [11]. Os autores provam que sua proposta possui uma menor complexidade computacional do que a feita por Paxson, e menor ou igual do que a proposta de Moon e Towsley, além de mais intuitiva do que esta última. Portanto, a abordagem considerada neste trabalho para remoção de *Skew* foi a de [11] e será detalhada a seguir.

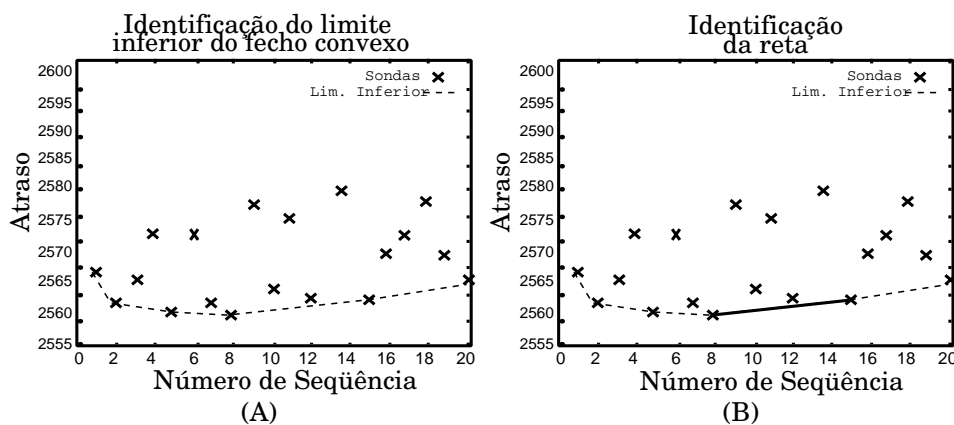


Figura 2: Funcionamento dos algoritmos para remoção do *Skew*.

A proposta de Liu e Zhang é baseada na estimativa do fecho convexo da seqüência co-

letada  $\Omega$ . No primeiro passo do algoritmo é calculado o limite inferior do fecho convexo de  $\Omega$  conforme ilustrado na Figura 2(A). É mostrado então que, retas que cobrem determinados pontos do limite inferior do fecho convexo são a solução dos problemas de otimização formulados naquele trabalho. Por exemplo, a reta que cobre o ponto médio da coleta é a solução do seguinte problema de otimização: minimizar a área entre a curva formada pelos vértices  $v_i$  e uma função linear qualquer. No caso de sondas CBR o ponto médio é igual  $N/2$ , onde  $N$  é o maior número de seqüência gerado. A Figura 2(B) ilustra a reta obtida para a coleta  $\Omega$ . Uma vez obtida a solução do problema de otimização, estimar a inclinação da função linear é trivial. Seja  $y = f(x)$  a reta estimada e  $v_i$  e  $v_j$  dois pontos desta reta onde  $v_i$  é o vértice inicial. A inclinação desta reta em relação ao eixo das abscissas é dada por  $\alpha = (d_j - d_i)/(j - i)$ .

Uma vez estimada a diferença entre as taxas de crescimento dos relógios envolvidos na medição, o valor do atraso com o *Skew* removido pode ser calculado da seguinte forma:  $atraso\_sem\_Skew_i = d_i - ((e_i - e_1) * \alpha)$  onde  $e_i$  e  $e_1$  são os instantes de envio da sonda  $i$  e da primeira sonda, respectivamente.

Uma nova seqüência  $\gamma$  é então gerada após o cálculo do atraso sem *Skew* para todas as  $N$  sondas recebidas. É importante perceber que, como os relógios não se encontram sincronizados no início da medição, a seqüência  $\gamma$  contém o *Offset* inicial da coleta. Na próxima seção iremos apresentar o algoritmo considerado neste trabalho para remover o *Offset*.

### 2.3. Algoritmos para remoção do *Offset*

Existem algumas propostas na literatura para resolver o problema de *Offset*. No entanto, todas exceto a de [3, 10] assumem que os caminhos entre os pontos de medição são simétricos, ou seja as capacidades de transmissão ao longo do *caminho de ida* são iguais as do *caminho de volta*. Portanto, é o algoritmo de [3, 10] o escolhido para este trabalho.

O algoritmo de [3, 10] requer o envio de uma seqüência de sondas de diferentes tamanhos e em sentidos opostos, e supõe que não existe *Skew* nos valores coletados para o atraso, e ainda que a distância percorrida pelas sondas enviadas por uma e outra máquina são aproximadamente as mesmas. As sondas que obtiverem o menor atraso, para cada tamanho usado, devem ser selecionadas. A partir dos valores dos atrasos das sondas selecionadas é obtida uma reta como mostra a Figura 3. As amostras de atraso selecionadas equivalem a sondas que supostamente não entraram em fila durante todo o caminho percorrido. Neste caso, o atraso de uma sonda é igual ao tempo de propagação no caminho somado ao seu tempo de transmissão.

Considerando que o tamanho das sondas obedece a uma função linear é possível estimar o atraso sofrido por uma sonda de tamanho nulo, caso tal sonda pudesse ser enviada, que é o ponto de interseção entre o eixo das ordenadas e a reta obtida usando os menores valores de atraso em cada sentido. Sejam  $t_{AB}$  e  $t_{BA}$  os atrasos de uma sonda de tamanho nulo enviada da máquina A para a B e da máquina B para a A, respectivamente. Então,  $t_{AB} = O + t_{prop}$ , e  $t_{BA} = -O + t_{prop}$  onde,  $t_{prop}$  é o tempo de propagação entre o transmissor e receptor (supostamente igual nos dois sentidos de transmissão), e  $O$  é o valor do *Offset*. Logo,  $t_{AB} - t_{BA} = 2O$ . Portanto, o *Offset* é obtido subtraindo os valores  $t_{AB}$  e  $t_{BA}$  estimados, e dividindo o resultado por dois.

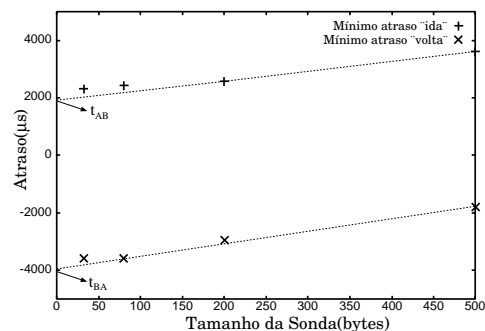


Figura 3: Atraso das sondas de tamanhos variados.

### 3. Outros problemas para o cálculo do atraso em sentido único

Nesta seção serão apresentados dois outros problemas encontrados para o cálculo do atraso em sentido único. Esses problemas foram causados por eventos de atualização de relógio das máquinas utilizadas para a medição, ou por eventos que causaram interrupções temporárias na execução do processo de medição. Em seguida serão comentadas as soluções da literatura para os eventos de atualização de relógio e serão apresentadas novas propostas para tratar esses problemas.

#### 3.1. Definição dos problemas

Os dois problemas citados acima são ilustrados na Figura 4 através de um experimento realizado entre a UFRJ e a UMass. Está representado na Figura 4(A) o atraso no sentido da UFRJ para a UMass e na Figura 4(B) o atraso no sentido da UMass para a UFRJ. É possível determinar que houve uma atualização no relógio de uma das máquinas logo no início do experimento, o que causa uma alteração constante nos valores estimados para os atrasos nos dois sentidos.

Na Figura 4(A), podemos observar também picos devido a valores de atrasos bastante altos. Estes picos foram causados pela interrupção na execução do processo de recebimento das sondas na máquina que estava em UMass. Este processo teve sua execução temporariamente interrompida por outro processo de maior prioridade sendo executado nesta mesma máquina. Quando este experimento foi realizado, a máquina de UMass estava executando uma aplicação para transmissão de vídeo em tempo real e a ferramenta *TANGRAM-II Traffic Generator* para estimar o atraso em um sentido. (Maiores detalhes sobre este experimento serão abordados na seção 4.) A aplicação para transmissão de vídeo em tempo real possui uma função de sincronização com a placa de captura de imagem. Quando essa função era executada, os processos da ferramenta de medição para envio e recebimento de sondas eram interrompidos por poucos microssegundos. O evento não é perceptível ao usuário, porém, significativo para uma ferramenta que estima grandezas na ordem de microssegundos. Após o término da interrupção do processo que executa a medição, várias das sondas enviadas pela máquina na UFRJ já se encontravam na memória da interface de rede para serem capturadas. É claro então que o atraso calculado para estas sondas é superior ao *real* pois inclui o tempo de espera das sondas para entrega à aplicação de medição.

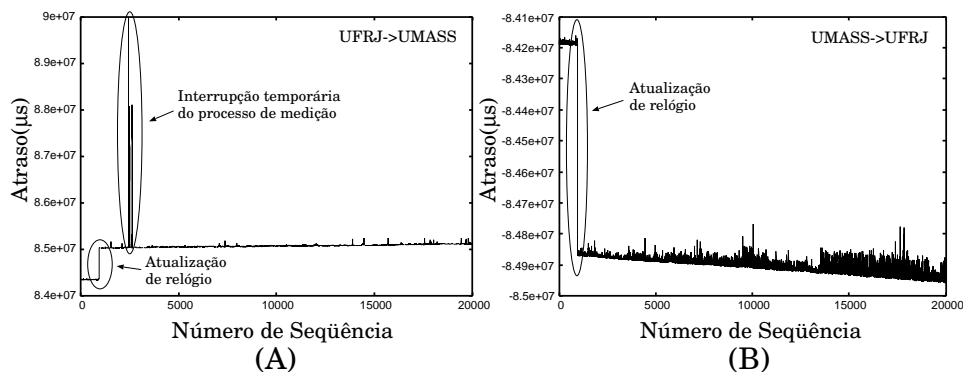


Figura 4: Atraso de pacotes entre máquinas que sofreram atualização nos relógios e interrupções temporárias na execução dos processos de medição.

A atualização dos relógios pode causar estimativas incorretas de *Skew* e *Offset* dado que os métodos para estimar esses valores são baseados no crescimento linear do atraso. Já os eventos de interrupção na execução do processo de envio e recebimento das sondas, têm influência insignificante sobre o cálculo do atraso médio em sentido único. No entanto, esses eventos afetam a distribuição desta medida pois existirão valores altos de atraso com probabilidade muito baixa. (Estes são os valores do atraso obtidos para as sondas que não foram tratadas imediatamente após a sua chegada.) Esta alteração na distribuição pode indicar de maneira incorreta que uma distribuição obtida possui cauda longa.

### 3.2. Proposta de um novo método

O evento de atualização dos relógios já havia sido considerado em outros trabalhos. Paxson [7] propôs um método para compensar o atraso calculado quando existirem atualizações dos relógios. Lui e Zhang [11] tratam deste problema no caso da estimativa do *Skew*. Os dois métodos acima são baseados na detecção de uma variação abrupta nos atrasos calculados no receptor. O problema dessas propostas é que uma alteração significativa nos atrasos na rede sofridos por pacotes nos dois sentidos, e em instantes próximos de tempo, pode ser considerada erroneamente como uma atualização nos relógios, levando a resultados incorretos.

O método proposto neste trabalho é capaz de estimar corretamente o atraso em sentido único mesmo que ocorram os eventos descritos acima. Ele requer que as sondas sejam geradas a intervalos constantes e se baseia na avaliação tanto da variação do atraso como do intervalo entre envio de sondas. Como as sondas são emitidas a intervalos regulares, a curva  $\varepsilon(i)$ , onde  $i$  é o número de seqüência das sondas e  $\varepsilon(i)$  o instante de envio da  $i$ -ésima sonda deve ser crescente e linear conforme ilustrado na Figura 5(A). Denominamos  $\varepsilon(i)$  de reta de envio.

Considere a Figura 5(B) onde é apresentado o gráfico de  $\varepsilon(i)$  obtido para o mesmo experimento ilustrado na Figura 4. Neste gráfico pode-se perceber uma alteração em  $\varepsilon(i)$  (isto é, na taxa de envio de sondas) próxima a sonda de número de seqüência 800. Porém, nem sempre uma alteração na taxa de envio caracteriza a atualização dos relógios. Caso um dos emissores não envie sondas temporariamente, em consequência de uma interrupção na execução do processo de medição, a taxa também terá uma variação. Por isso os valores coletados para os atrasos na máquina receptora também devem ser analisados. Caso haja alterações também nos atrasos da máquina receptora, ficará caracterizado que houve o evento de atualização de relógio.

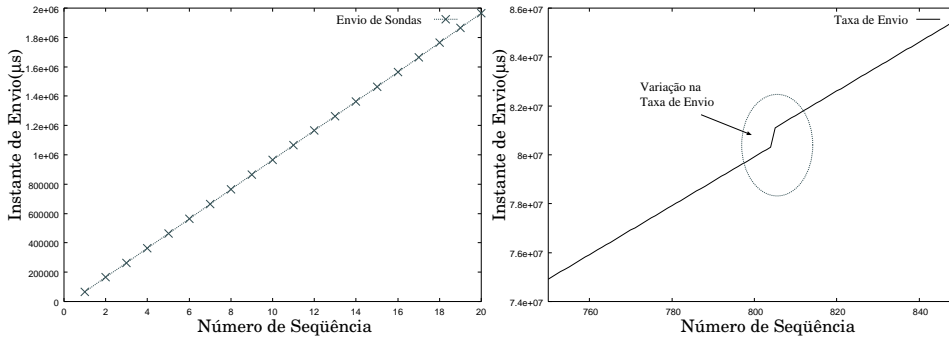


Figura 5: Taxa de envio das sondas.

Suponha que o objetivo é medir o atraso no sentido de um computador A para outro computador B, e sondas estejam sendo enviadas para coletar os dados necessários ao cálculo. Seja  $d_i^{AB}$  o atraso da  $i$ -ésima sonda enviada de A para B e  $d_i^{BA}$  o atraso da sonda  $i$  no sentido oposto. (Tanto  $d_i^{AB}$  como  $d_i^{BA}$  são os atrasos calculados sem a remoção de *skew* ou *offset*, calculados pela simples subtração do instante de envio da sonda e instante de recebimento.) As máquinas A e B enviam sondas a intervalos constantes  $T_e$ . Seja  $e_i^{AB}$  e  $e_i^{BA}$  o instante de envio da  $i$ -ésima sonda de A para B e de B para A, respectivamente. Logo a máquina A que mede o atraso B-A armazena os seguintes dados: número de seqüência  $i$ , instante de envio  $e_i^{BA}$  e atraso da sonda  $d_i^{BA}$ . Para efeito de descrição do algoritmo, suponha que a máquina B sofreu atualização do seu relógio. Neste caso, conforme ilustrado na Figura 6 caso (1), podemos notar que houve alteração na reta de envio armazenada em A e nos atrasos estimados nos dois sentidos. O algoritmo é composto dos passos descritos a seguir.

**Passo 1:** Verificar se  $|e_{i+1}^{BA} - e_i^{BA}| > T_e * (1 + \epsilon)$ , para  $i = 1, N - 1$ , onde  $\epsilon$  é uma dada tolerância para o intervalo  $T_e$  entre o envio de sondas. (Isto é, é tolerado que  $T_e$  sofra pequenas variações.)

**Passo 2:** Suponha que, como resultado do passo 1 acima, seja detectada alteração no intervalo de

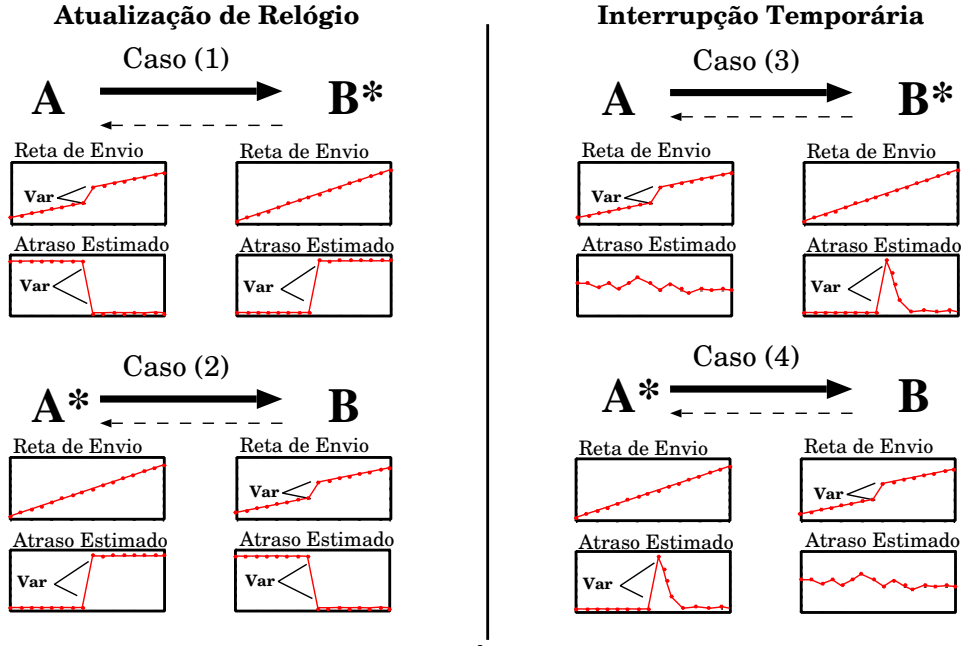


Figura 6: Cenários possíveis.

envio entre a  $j$ -ésima e a  $(j + 1)$ -ésima sondas, isto é:  $|e_{j+1}^{BA} - e_j^{BA}| = T_e + \nu > T_e * (1 + \epsilon)$ . É necessário então verificar se o atraso calculado para a sonda  $(j + 1)$  é maior que o esperado, ou seja,  $d_{j+1}^{BA} > e_{j+1}^{BA} - e_j^{BA} - T_e + \bar{S}^{BA}(j)$ , onde  $\bar{S}^{BA}(j)$  é a média amostral dos atrasos coletados para as sondas de 1 até  $j$  no sentido B-A. Note que em geral  $e_{j+1}^{BA} - e_j^{BA} = T_e$ , mas devido ao passo 1,  $|e_{j+1}^{BA} - e_j^{BA}| = T_e + \nu$ . Portanto,  $d_{j+1}^{BA} > \nu + \bar{S}^{BA}(j)$ , logo este passo verifica se  $d_{j+1}^{BA}$  é ainda maior que a variação sofrida no intervalo de envio mais o retardo esperado.

**Passo 3:** Suponha que  $d_{j+1}^{BA} > \nu + \bar{S}^{BA}(j)$ . Então é feita uma atualização nos valores de atraso das sondas  $k = j + 1, \dots, N$ , ou seja,  $d_k^{BA} = d_k^{BA} - (e_j^{BA} - e_{j-1}^{BA} - T_e)$ . Note que um evento de atualização de relógio na máquina B é detectado pela variação na taxa de envio de B (que deveria ser constante) e de atraso estimado de B para A. Logo, o atraso de A para B também deve ser atualizado. Seja  $\tau_k^B$  o instante, medido na máquina B, de envio da  $k$ -ésima sonda, e  $k'$  a sonda recebida de A logo após  $\tau_k^B$ . A seqüência recebida por B deve ser atualizada a partir de  $k'$ :  $d_{k'}^{AB} = d_{k'}^{AB} - (e_j^{BA} - e_{j-1}^{BA} - T_e)$ . O caso (1) da Figura 6 ilustra este cenário.

O caso (2) da Figura 6 ilustra o cenário onde ocorreu atualização do relógio da máquina A. O algoritmo para o caso (2) é idêntico ao descrito acima, aplicado a seqüência A-B.

Conforme citado anteriormente, o evento de interrupção na execução do processo de envio e recebimento das sondas, também ocasiona uma alteração na taxa de envio de sondas. O que o diferencia do evento de atualização de relógio é que somente o atraso coletado na máquina onde ocorreu o evento, possui uma alteração significativa. Iremos descrever o algoritmo proposto para tratar deste evento considerando que ele ocorreu na máquina B. Pode-se notar através do caso (3) da Figura 6 que houve alteração na reta de envio armazenada em A e no atraso estimado de A para B. Os passos 1 e 2 deste algoritmo são idênticos aos passos do algoritmo para tratar do evento de atualização de relógio, portanto só descreveremos os passo 3 e 4.

**Passo 3:** Caso não seja detectada variação maior que a esperada no atraso de B para A, então é verificado se o atraso da  $(j + 1)$ -ésima sonda no sentido de A para B foi alterado, ou seja, se  $d_{j+1}^{AB} > e_{j+1}^{BA} - e_j^{BA} - T_e + \bar{S}^{AB}(j)$ .

**Passo 4:** Suponha que  $d_{j+1}^{AB} > e_{j+1}^{BA} - e_j^{BA} - T_e + \bar{S}^{AB}(j)$ . Então as amostras coletadas durante



um período  $\xi$  serão descartadas. O valor escolhido para este período é:  $\xi = 2 * (|e_{j+1}^{BA} - e_j^{BA}| - T_e)$ , ou seja,  $\xi$  equivale ao dobro do tempo que o processo de medição ficou sem tratar as sondas já recebidas pela máquina. Com isso garante-se que as sondas com amostras “ruins” são descartadas.

O caso (4) da Figura 6 ilustra o cenário onde ocorreu interrupção do processo de medição da máquina A. O algoritmo para o caso (4) é idêntico ao descrito acima.

É importante ressaltar que a detecção de atualização de um relógio é feita pela variação da taxa de envio da máquina deste relógio, e não pela variação abrupta de atrasos na rede, como nos algoritmos propostos na literatura. Como as taxas de envio são constantes, com grande precisão, o algoritmo proposto é robusto.

## 4. Implementação na ferramenta TANGRAM-II

A proposta de [11] para remoção do *Skew*, de [3] para remoção do *Offset*, e os algoritmos propostos na seção 3.2 para tratar os problemas de atualização de relógios e interrupção temporária do processo de medição foram implementados e integrados ao módulo *TANGRAM-II Traffic Generator* do ambiente *TANGRAM-II*. Do nosso conhecimento, a ferramenta *TANGRAM-II Traffic Generator* é a única que permite a estimativa do atraso em um sentido sem uso de equipamentos GPS.

O tráfego gerado pela ferramenta deve seguir os padrões definidos pelos algoritmos. A geração das sondas é determinística, nas duas direções e com tamanho de pacotes variando entre: 40, 80, 200, e 500 (*bytes*). Esses valores foram escolhidos para prevenir a segmentação dos pacotes, evitando a possibilidade de erros nos cálculos da métrica. Apesar da grande maioria dos pacotes nas redes IP’s manterem o tamanho máximo de 1500 *bytes*, existem *MTU*’s bem menores que podem chegar a 576 *bytes*.

Após o término da coleta os algoritmos são executados na seguinte ordem: (i) tratamento de atualizações de relógios, (ii) tratamento de interrupções temporárias do processo de medição, (iii) remoção de *Skew* e (iv) remoção de *Offset*. A Figura 7(A) ilustra valores de atraso obtidos após a execução dos algoritmos (i) e (ii). O passo seguinte é a remoção do *Skew* (Figura 7(B)) e do *Offset* (Figura 7(C)) com os algoritmos (iii) e (iv).

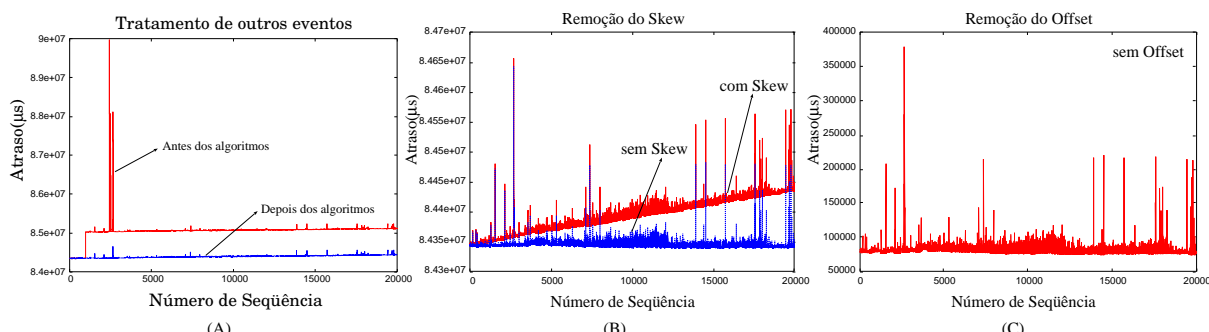


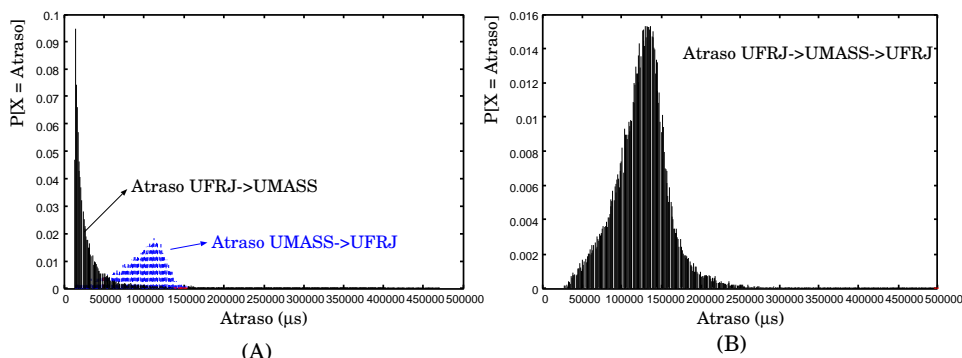
Figura 7: Atraso estimado por uma medição da ferramenta *TANGRAM-II Traffic Generator*.

### 4.1. Validação das implementações na ferramenta

Uma forma de validar os algoritmos seria com o uso de dois pares de computadores: um par em um ponto A, e outro em um ponto B. Um dos computadores em cada ponto deveria ter equipamento GPS e atuar em modo promíscuo lendo pacotes endereçados para o seu par. Os computadores sem o equipamento de sincronização calculariam o atraso com os algoritmos acima, e os demais o atraso com GPS.

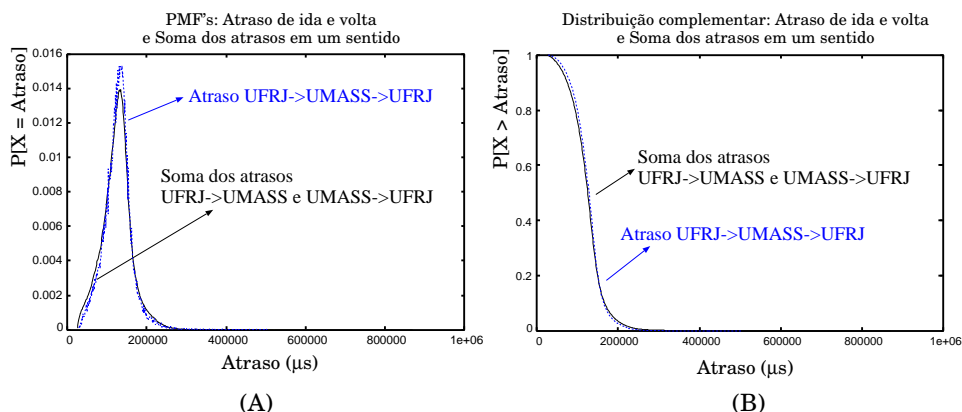
Uma outra forma de validação, sem depender de equipamento de sincronização é feita pela comparação da distribuição do atraso de ida e volta (RTT) com a distribuição da soma do atraso em cada uma das direções da medição. Esta foi a forma usada neste trabalho uma vez que não dispomos do primeiro cenário. Seja  $d^{AB}$  e  $d^{BA}$  a variável aleatória que representa o atraso de uma sonda enviada de A para B, e de B para A, respectivamente. Seja  $RTT^{AB}$  a variável aleatória que representa o RTT entre os pontos A e B. Para que a comparação seja válida é necessário que o processo de medição dos retardos de A-B, B-A e  $RTT^{AB}$  seja iniciado e terminado a instantes próximos de tempo.

Diversos experimentos foram realizados com o objetivo de validar os algoritmos. Selecionamos um experimento realizado entre a UFRJ e a UMass para ilustrar a validação. As duas curvas da Figura 8(A) representam as funções probabilidade de massa (PMF) do atraso em sentido único e o gráfico da Figura 8(B) representa a PMF do atraso de ida e volta. Podemos observar que as PMF's obtidas para o atraso em um sentido são bem diferentes. Este resultado ilustra a importância da estimativa da distribuição do atraso em cada um dos sentidos da comunicação. É possível verificar também, que a média do atraso de ida e volta dos pacotes é aproximadamente igual a soma das médias do atraso em um sentido.



**Figura 8: PMF's estimadas nas medições: (A) PMF's do atraso em um sentido; (B) PMF do atraso de ida e volta.**

A Figura 9(A) ilustra a PMF da soma de  $d^{AB}$  e  $d^{BA}$  e também a PMF de  $RTT^{AB}$ . Podemos notar que a PMF da soma dos atrasos é praticamente idêntica a PMF do atraso de ida e volta. Para facilitar a comparação na Figura 9(B) é plotada a função distribuição cumulativa dessas duas variáveis.



**Figura 9: Comparação entre os cálculos dos atrasos.**

## 4.2. Módulo para cálculo do atraso de aplicações

Os algoritmos definidos nas seções anteriores exigem que as sondas sejam geradas segundo o padrão descrito acima, isto é, a intervalos constantes e sondas com tamanhos pré-definidos. No

entanto, muitas vezes pode ser bastante útil a obtenção do atraso em sentido único para pacotes de uma determinada aplicação, por exemplo software de transmissão de vídeo, que gere pacotes sem obedecer o padrão necessário ao cálculo desta medida de interesse.

Foi desenvolvido um módulo para a ferramenta *TANGRAM-II Traffic Generator* que possibilita a estimativa do atraso em sentido único para os pacotes de uma aplicação. Para que seja possível esta estimativa, é necessário que se faça uma medição auxiliar em paralelo e durante toda a execução da aplicação, com o padrão de geração de sondas exigido pelos algoritmos. A partir desta medição auxiliar serão obtidos os valores para o *Skew* e o *Offset* que serão usados para obtenção do atraso dos pacotes da aplicação. (Atualmente o módulo implementado para estimativa do atraso dos pacotes de uma aplicação não trata dos problemas apresentados na seção 3. Encontra-se em fase de implementação um módulo que trate destes problemas.)

O algoritmo executado neste módulo possui como parâmetros de entrada os seguintes dados da medição auxiliar: o (*Offset*), o parâmetro  $\alpha$  obtido pelo algoritmo para remoção de *Skew*, e o instante inicial da medição auxiliar. Os dados fornecidos pela aplicação são: o instante de envio e recebimento dos pacotes, e portanto medidos com relógios de máquinas diferentes.

A Figura 10 mostra um experimento onde foi calculado o atraso para uma transmissão de vídeo em tempo real da UMass para a UFRJ. Nesta figura está ilustrada **parte** da medição auxiliar que foi iniciada poucos instantes antes da transmissão de vídeo começar. Na parte (A) da figura são ilustrados os atrasos coletados para as sondas da medição auxiliar e para os pacotes da aplicação de vídeo. Como a transmissão de vídeo foi realizada da UMass para a UFRJ, podemos notar que o atraso da medição auxiliar nesta direção possui a mesma tendência de decrescimento do atraso dos pacotes enviados pela ferramenta de vídeo. A parte (B) da figura mostra os atrasos obtidos após a remoção de *Skew* e *Offset*.

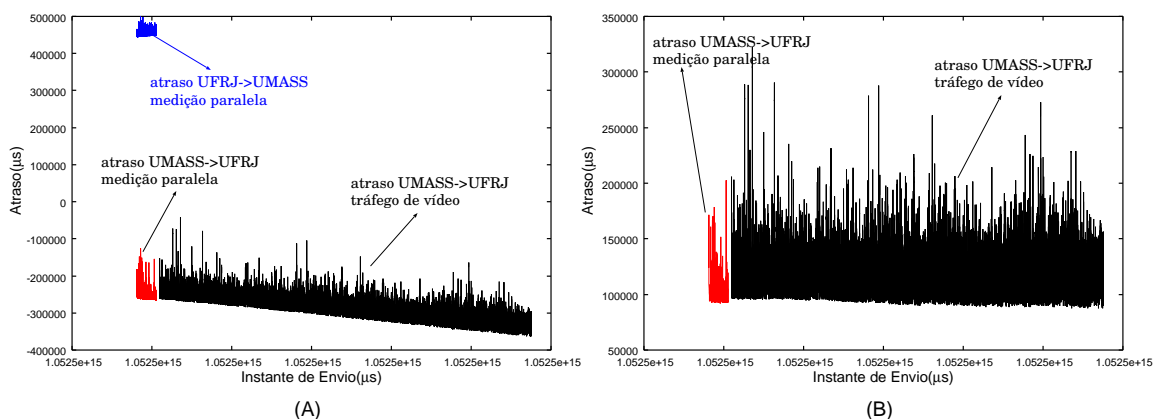


Figura 10: Atraso calculado para uma aplicação de transmissão de vídeo

## 5. Avaliação dos resultados dos experimentos

A ferramenta *TANGRAM-II Traffic Generator* foi utilizada em uma série de experimentos executados entre a UFRJ, UNIFACS e UMass. Com o objetivo de poder identificar a distribuição que melhor caracteriza o atraso obtido nos experimentos, foi adicionada uma funcionalidade na ferramenta que permite ao usuário visualizar um conjunto de distribuições da literatura parametrizadas com estatísticas do atraso. Seis são as distribuições disponibilizadas: Exponencial, Gaussiana, Lognormal, Pareto, Gamma e Weibull. Os parâmetros dessas distribuições foram calculados através do método de momentos e foi utilizado o *Mean Square Error (MSE)* para auxiliar na estimativa da melhor distribuição. O MSE foi calculado a partir da equação:

$$MSE = \frac{\sum_0^M |F(x) - \hat{F}(x)|^2}{M}, \text{ onde } M \text{ é o número de amostras.}$$

Foram realizados experimentos durante uma semana entre a UFRJ e a UNIFACS. As medições foram realizadas diariamente na parte da tarde. Estes experimentos deram origem a 14 distribuições para o atraso, sete para o atraso entre a UFRJ e a UNIFACS e sete para o atraso no outro sentido. Das seis distribuições parametrizadas com os descritores estatísticos obtidos para as 14 coletas, quatro apresentaram *MSE* de ordem menor ou igual a  $10^{-3}$ . São elas: Gaussiana, Lognormal, Exponencial e Pareto. A distribuição Gaussiana mostrou-se a mais adequada para caracterizar o atraso estimado em nove das coletas, as distribuições Exponencial e Pareto foram mais adequadas em duas coletas cada, e a Lognormal em apenas uma coleta.

A Figura 11(A) apresenta a distribuição complementar do atraso entre a UNIFACS e a UFRJ obtida em um dos experimentos. São mostradas também as distribuições parametrizadas. Pelo gráfico da Figura 11(A) não é possível identificar facilmente a distribuição que melhor caracteriza o atraso. Já através da Figura 11(B) onde são ilustradas as distribuições em escala logarítmica, é possível identificar que a cauda da distribuição do atraso possui decaimento semelhante à cauda das distribuições Exponencial e Lognormal. Além da visualização gráfica, o *MSE* foi obtido para cada uma das distribuições. Pelos valores obtidos pelo *MSE* podemos concluir que a distribuição exponencial é a que melhor caracteriza o atraso medido neste exemplo.

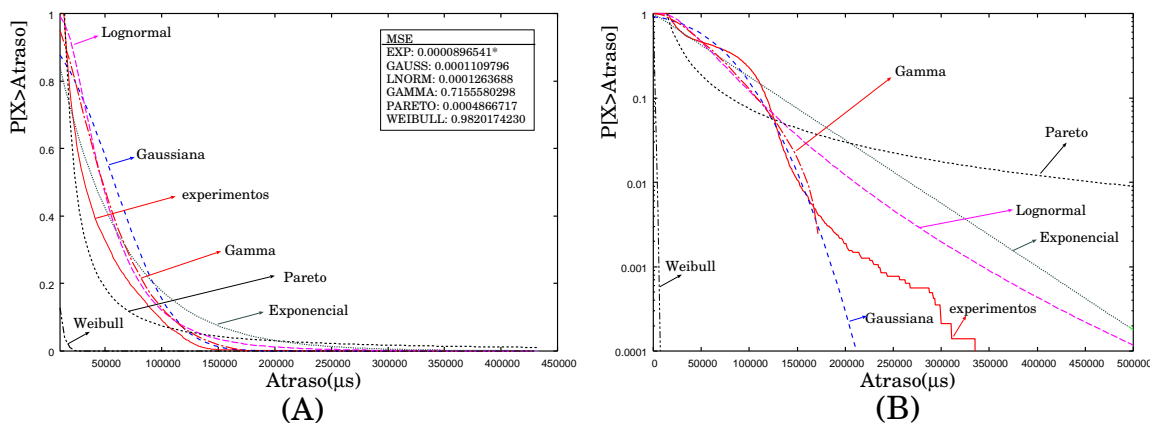


Figura 11: Estimativa da distribuição do atraso entre a UNIFACS e a UFRJ

Além das medições com a UNIFACS, durante seis meses foram executados experimentos entre a UFRJ e a UMass. Eles foram realizados ao longo do segundo semestre de 2002. As características dos experimentos são as seguintes: foram realizadas medições duas vezes na semana, iniciados as 12:00h com duração de 1:30h e executados em paralelo a uma sessão de videoconferência. O objetivo da execução em paralelo a uma transmissão de vídeo, foi caracterizar o atraso quando uma aplicação que possui altos requisitos de banda estivesse sendo executada. Mais de 40 coletas obtidas nos experimentos foram analisadas.

As PMF's do atraso em um único sentido obtidas em três dos experimentos são mostradas na Figura 12. Podemos observar que elas são bem diferentes para cada um dos sentidos da comunicação e que o atraso no sentido UMass-UFRJ possui uma variabilidade maior do que no sentido contrário. A Figura 13 mostra a distribuição do atraso em cada um dos sentidos da comunicação e as distribuições parametrizadas para um dado experimento. Neste caso, a distribuição indicada pelo *MSE* para caracterizar o atraso UFRJ-UMass é a Pareto. No sentido oposto a distribuição indicada é a Gaussiana.

Após análise dos atrasos coletados nos experimentos, podemos obter algumas conclusões a respeito da caracterização do atraso entre a UFRJ e a UMass. Tanto no sentido UFRJ-UMass quanto no sentido oposto, a distribuição que obteve o menor *MSE* em 80% dos casos foi a Pareto. As distribuições Gaussiana e Lognormal foram apontadas como mais apropriadas para caracterizar o atraso em 10% das coletas, cada uma das duas.

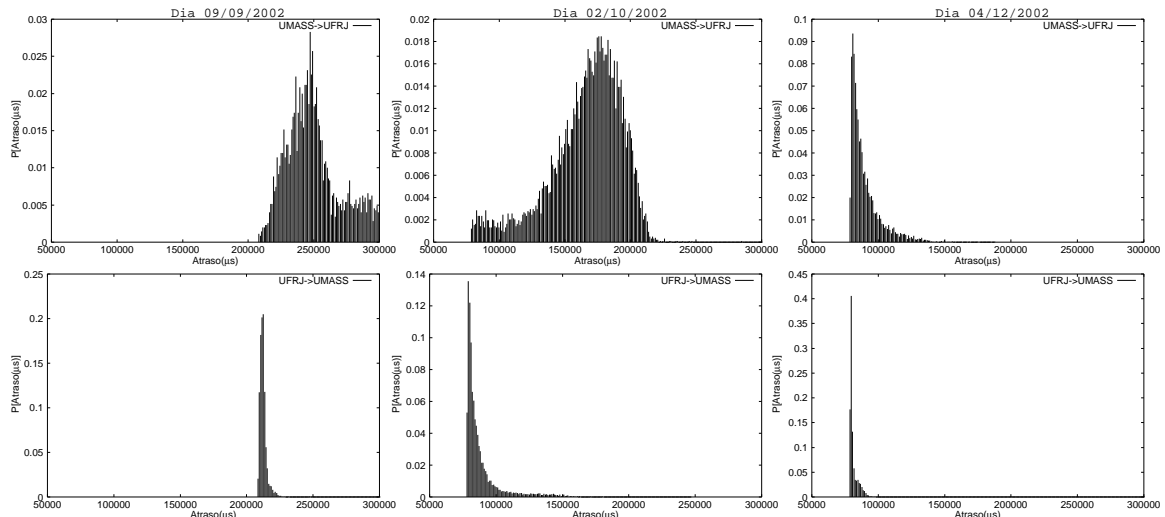


Figura 12: Resultados obtidos de experimentos entre a UFRJ e a UMass.

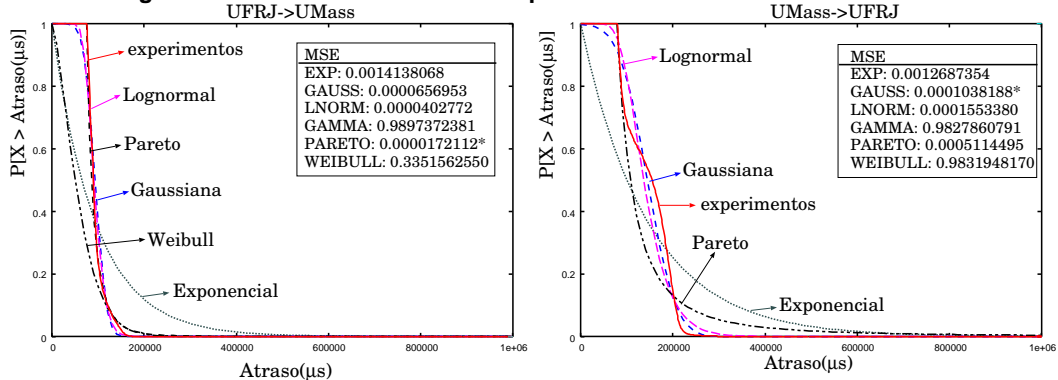


Figura 13: Caracterização do atraso de experimento entre a UFRJ e a UMass.

## 6. Conclusão

Realizar medições na Internet para caracterizar diversas das métricas que têm impacto significativo no desempenho das aplicações de rede é uma tarefa essencial. Do resultado obtido pode-se construir modelos mais precisos visando a estudar adequadamente uma variedade de técnicas, por exemplo àquelas para compensar a falta de um serviço que garanta um determinado nível de qualidade às aplicações. Este trabalho tem seu foco sobre uma medida das mais importantes, e por outro lado difícil de ser obtida sem o uso de uma infra-estrutura especial: o atraso em um único sentido.

Foram analisados algoritmos propostos na literatura para remover *skew* e *offset* dos relógios dos computadores envolvidos nas medições, e identificados problemas que resultam em erros nas medições. Um algoritmo robusto foi proposto para solucionar os problemas.

Uma ferramenta que implementa as técnicas necessárias ao cálculo do atraso em sentido único foi desenvolvida e validada. Em adição, foi desenvolvido um módulo para ajudar o analista no processo de caracterização da distribuição do atraso em sentido único. Foi mostrado ainda como é possível obter o atraso dos pacotes para uma aplicação genérica.

Vários experimentos foram realizados entre a UFRJ e UMass e UFRJ e UNIFACS, para mostrar a utilidade da ferramenta desenvolvida. Os resultados mostraram que pode ser significativamente diferente os atrasos em cada um dos sentidos. Além disso, embora as distribuições que melhor tenham representado os atrasos dos pacotes fossem a Pareto e a Gaussiana, fica claro que é necessário considerar várias distribuições em uma análise, e estudar a sensibilidade dos resultados em relação às distribuições.

## Agradecimentos

Agradecemos aos grupos dos professores José Augusto Suruagy Monteiro, da UNIFACS, e Don Towsley, de UMass, por disponibilizarem seus laboratórios para que pudéssemos realizar os experimentos relatados neste trabalho.

## Referências

- [1] R.M.L.R. Carmo, L.R. de Carvalho, E. de Souza e Silva, M.C. Diniz, and R.R. Muntz. Performance/Availability Modeling with the TANGRAM-II Modeling Environment. *Performance Evaluation*, 33:45–65, 1998.
- [2] E. de Souza e Silva and R.M.M. Leão. Modelagem e Análise de Redes com o Conjunto de Ferramentas TANGRAM II. In *XXI Simpósio Brasileiro de Redes de Computadores (Workshop de Ferramentas)*, pages 897–904, Natal, Brasil, Maio 2003.
- [3] D.D. Loung and J. Biro. Needed Services for Network Performance Evaluation. In *IFIP Workshop on Performance Modeling and Evaluation of ATM Networks*, Inglaterra, Julho 2000.
- [4] D. Mills. Network Time Protocol (Version 3) - Specification Implementation and Analysis, Março 1992. RFC 1305.
- [5] S.B. Moon, P. Skelly, and D. Towsley. Estimation and Removal of Clock Skew for Network Delay Measurements. In *IEEE/Infocom*, pages 227–234. New York, USA, Março 1999.
- [6] A. Pasztor and D. Veitch. PC Based Precision Timing Without GPS. In *ACM/Sigmetrics*, pages 1–10. Marina Del Rey, CA, USA, Junho 2002.
- [7] V. Paxson. On Calibrating Measurements of Packet Transit Times. In *ACM/Sigmetrics*, pages 11–21, Madison, Wisconsin, USA, Junho 1998.
- [8] A.A.A. Rocha. Medições ativas na Internet: algoritmos baseados em retardo fim-a-fim e experimentos. Master's thesis, UFRJ-COPPE/Sistemas, Agosto 2003.
- [9] A.A.A. Rocha, E. de Souza e Silva, and R.M.M. Leão. Uma ferramenta para estimar características fim-a-fim na Internet. In *WRNP'IV*, Natal, Brasil, Maio 2003.
- [10] M. Tsuru, T. Takine, and Y. Oie. 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, Fevereiro 2002.
- [11] L. Zhang, Z. Liu, and C.H. Xia. Clock Synchronization Algorithms for Network Measurements. In *IEEE/Infocom*, pages 160–169, New York, USA, Junho 2002.