Metrologia na Internet

Artur Ziviani [1] **e Otto Carlos M. B. Duarte** [2]

[1]Laboratório Nacional de Computação Científica (LNCC)
Av. Getúlio Vargas, 333 – 25651-075 – Petrópolis, RJ

[2]Grupo de Teleinformática e Automação (GTA) – COPPE/Poli
Universidade Federal do Rio de Janeiro (UFRJ)
Caixa Postal 68504 – 21945-970 – Rio de Janeiro, RJ

ziviani@lncc.br, otto@gta.ufrj.br

Resumo. *A Internet apresenta grandes desafios para a caracterização de sua estrutura e comportamento. Diferentes razões contribuem para essa situação, incluindo a imensa comunidade de usuários, a diversidade de aplicações, a heterogeneidade de equipamentos, a administração distribuída, a vasta cobertura geográfica e o dinamismo típicos da Internet atual. Para enfrentar essas dificuldades, diversas abordagens baseadas em medições vêm sendo propostas recentemente para estimar e melhor compreender o comportamento, a dinâmica e as propriedades da Internet. O conjunto dessas técnicas baseadas em medições é denominado neste texto de Metrologia na Internet. Este texto aborda a temática de Metrologia na Internet de forma abrangente apresentando ferramentas e métodos baseados em medições que influenciam diretamente outras áreas convencionais, tais como o projeto e planejamento de redes, engenharia de tráfego, qualidade de serviço e gerenciamento de redes.*

Abstract. *The Internet presents great challenges to the characterization of its structure and behavior. Different reasons contribute to this situation, including the huge user community, the large range of applications, the equipment heterogeneity, the distributed administration, the vast geographic coverage, and the dynamism that are typical of the current Internet. In order to deal with these challenges, several measurement-based approaches have been recently proposed to estimate and better understand the behavior, dynamics, and properties of the Internet. The set of these measurement-based techniques is what we call Internet Measurements. This text covers the Internet Measurements area in a comprehensive way by presenting measurement-based tools and methods that directly influence other conventional areas, such as network design and planning, traffic engineering, quality of service, and network management.*

1. Introdução

Há cerca de uma década a Internet iniciou sua transformação de um instrumento restrito à comunidade científica para um componente fundamental da sociedade de informação. Possivelmente, a conseqüência mais importante do sucesso da Internet é que o propósito comum que norteava os seus componentes não mais se mantém. Usuários, provedores comerciais de acesso, governos, operadores de telecomunicações e fornecedores de conteúdo possuem interesses que podem ser contraditórios entre si, conduzindo a uma convivência em disputa [Clark et al., 2002]. Um exemplo dessa convivência em disputa é a relação entre provedores comerciais que precisam estar interconectados para obterem conectividade universal, mesmo que freqüentemente estes sejam fortes concorrentes.

A heterogeneidade e a administração distribuída decorrentes desse cenário, aliadas à vasta cobertura geográfica e ao dinamismo típicos da Internet atual, dificultam a caracterização da estrutura e do comportamento da rede como um todo [Floyd e Paxson, 2001].

A Internet possui atualmente uma enorme comunidade de mais de 800 milhões de usuários em fevereiro de 2005 que se encontra em franca expansão com uma taxa de crescimento de 126% entre os anos 2000 e 2005 [Internet World Stats, 2005]. Esse crescente número de usuários se serve de uma grande variedade de aplicações. Essas aplicações geram uma grande diversidade de tipos de tráfego e requerem novos serviços com qualidade. Em função dessa diversidade, provedores, usuários e operadores conscientizam-se da necessidade de melhor compreender a estrutura e o comportamento dinâmicos da rede.

O trabalho seminal de Paxson [Paxson, 1997] introduziu uma abordagem baseada em medições para caracterizar a dinâmica do tráfego na Internet. Também houve a caracterização baseada em medições da natureza auto-similar do tráfego de rede em ambientes locais [Leland et al., 1994] e de longas distâncias [Paxson e Floyd, 1995], assim como no tráfego da *World Wide Web* [Crovella e Bestavros, 1997]. A consideração dos conceitos de dependência de longa duração e de auto-similaridade influenciaram decisivamente a modelagem do tráfego na Internet na última década [Karagiannis et al., 2004, Figueiredo et al., 2005]. O trabalho dos irmãos Faloutsos [Faloutsos et al., 1999] também causou grande impacto ao constatar que a aparente aleatoriedade da topologia da Internet na verdade segue leis de potência. Isso implica na possibilidade de estimação de importantes parâmetros como o número médio de vizinhos e influencia o projeto e a análise de protocolos. Essa característica também pode ser usada para gerar sinteticamente topologias mais realistas para simulações. Seguindo esses primeiros trabalhos, diversas abordagens baseadas em medições vêm sendo propostas para estimar e caracterizar diferentes aspectos da Internet, buscando tornar o seu comportamento mais observável [Chen, 2001, Varghese e Estan, 2004]. Essa maior capacidade de observação do comportamento da rede vem ajudando os pesquisadores a desvendar alguns mitos sobre as características e propriedades da Internet [Claffy, 2002, Shannon et al., 2002, Spring et al., 2003]. O conjunto dessas técnicas baseadas em medições concebidas para observar e inferir diferentes características da rede compõe o que neste texto é denominado de Metrologia na Internet.

Este texto aborda a temática de Metrologia na Internet de forma abrangente, apresentando ferramentas e métodos recentemente propostos para inferir e melhor compreender o comportamento, a dinâmica e as propriedades da Internet atual [Brownlee e Claffy, 2004]. Essas ferramentas e métodos têm influência direta em outras áreas convencionais, como o projeto e o planejamento de redes, a engenharia de tráfego, o provimento de qualidade de serviço (QoS) e o gerenciamento de redes.

A Metrologia na Internet tem por base a medição de aspectos específicos, é difícil discutir seus desafios sem considerar problemas específicos. Portanto, após introduzir os fundamentos da área de Metrologia na Internet na Seção 2, nós discorremos sobre diversos problemas atuais, onde existem propostas baseadas em medições para enfrentá-los. A Seção 3 apresenta técnicas para a estimação de banda passante. Na Seção 4, é discutida a inferência de matrizes de tráfego. O diagnóstico de anomalias é o assunto da Seção 5. Técnicas para estimar a proximidade entre dois nós arbitrários na rede são discutidas na Seção 6. Outras metodologias recentes para diferentes serviços baseados em medições são apresentadas na Seção 7. A Seção 8 discute duas plataformas para medições e experimentação. Na Seção 9, é descrito o desenvolvimento de um serviço baseado em medições para estimar a localização geográfica de nós na Internet. Finalmente, a Seção 10 apresenta nossas considerações finais, incluindo referências a projetos nacionais e internacionais que utilizam medições, e perspectivas da área de Metrologia na Internet.

2. Fundamentos da área de Metrologia na Internet

O funcionamento básico da Internet foi concebido com o objetivo de minimizar a complexidade dos mecanismos em seu interior, concentrando o controle e a adaptação nas extremidades de transmissão. Esse princípio permitiu a expansão da Internet para as suas dimensões atuais, porém também limitou a possibilidade de monitoração do comportamento dinâmico da rede [Habib et al., 2004, Mao, 2005]. Atualmente, a Internet é um vasto conjunto de redes interconectadas, porém operadas por organizações distintas que em geral são concorrentes entre si. Como uma conseqüência disso, muitos domínios não cooperam com iniciativas externas de medição de desempenho.

A capacidade limitada de observação da Internet foi adequada para o serviço de melhor esforço (*best-effort*). No entanto, a Internet evolui para serviços mais avançados, tais como serviços integrados e diferenciados, com maior expectativa em relação ao desempenho e à qualidade de serviço oferecidos [Larrieu e Owezarski, 2005]. Tornar a rede mais observável é essencial para a verificação e a melhoria do desempenho da rede frente a aplicações mais exigentes. Além disso, uma melhor maneira de monitorar a rede se faz necessária para lidarmos com a crescente complexidade da Internet, representada por um enorme crescimento em extensão, diversidade, velocidades de transmissão e volume de tráfego. A Figura 1 ilustra esse crescimento vertiginoso ao mostrar a evolução ao longo dos últimos 15 anos do número de entradas ativas utilizadas pelo protocolo BGP (*Border Gateway Protocol*) [Huitema, 2000], o protocolo de roteamento inter-domínio que interliga os diferentes sistemas autônomos na Internet.

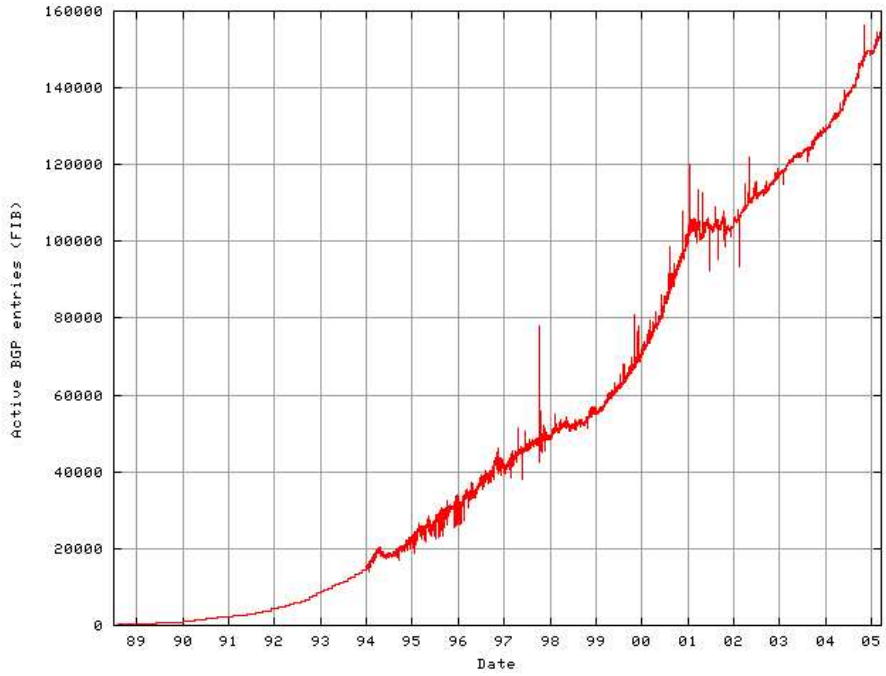


Figura 1: Evolução das entradas ativas no BGP (fonte: [Smith, 2005]).

O gerenciamento de redes provê uma função complementar à metrologia ao monitorar o estado de nós individuais na rede. O protocolo SNMP (*Simple Network Management Protocol*) [Case et al., 1990] permite a um gerente de rede centralizado inquerir dados de componentes da rede. Esse gerente também pode ser avisado sobre a ocorrência de eventos pré-definidos. Em termos de desempenho, o gerente limita-se a coletar medições simples individualmente de cada equipamento gerenciável. Embora os roteadores sejam os pontos ideais para medições de tráfego, em geral, eles não estão equipados para uma monitoração detalhada. Fabricantes de roteadores evitam a adição de características de

metrologia devido ao impacto no desempenho de encaminhamento de pacotes. A ferramenta NetFlow [Cisco, 1999] alcança grande sucesso entre operadores de rede e fornecedores de acesso. Essa ferramenta faz uma amostragem de fluxos capaz de fornecer dados sobre o tráfego presente na rede. Embora popular, o NetFlow possui vários pontos que podem ser melhorados como uma taxa de amostragem adaptativa e uma melhor capacidade de amostrar fluxos que não sejam TCP [Estan et al., 2004]. Como conseqüência dos problemas dos métodos existentes, diversos métodos indiretos de medição são desenvolvidos. O grupo de trabalho IPPM (*IP Performance Metrics*) do IETF (*Internet Engineering Task Force*) se dedica à definição das métricas relevantes à qualidade, desempenho e confiabilidade dos serviços de rede [Uijterwaal e Zekauskas, 2003].

2.1. Características dos métodos de medição

As abordagens baseadas em medições para a investigação de problemas relacionados com redes de computadores utilizam técnicas de medição passiva ou ativa [Barford e Sommers, 2004]. A monitoração de tráfego tipicamente consiste do registro passivo de pacotes em um enlace, enquanto que medições ativas de desempenho envolvem o envio de pacotes de medição.

Medições passivas referem-se ao processo de monitorar o tráfego de rede sem injetar algum novo tráfego ou modificar o tráfego na rede. Isso se realiza em um ou mais pontos da rede. Medições passivas podem fornecer um conjunto detalhado de informações sobre os pontos da rede onde as medições são realizadas e sobre o tráfego de passagem por estes pontos [Jaiswal et al., 2004]. Exemplos são a amostragem dos cabeçalhos dos pacotes de passagem pelo ponto de monitoração ou o registro do número de perda de pacotes em um determinado intervalo de tempo. Para uma monitoração passiva de alto desempenho, equipamento dedicado é necessário, sendo que o equipamento mais usado em medições passivas atualmente são as placas DAG [DAG, 2001], desenvolvidas originalmente na Universidade de Waikato, Nova Zelândia. Há um grupo de trabalho do IETF chamado *Packet Sampling* (PSAMP) [Bierman e Quittek, 2001] dedicado à definição de métodos padronizados para a amostragem de pacotes em dispositivos de rede. Buscam-se métodos simples o suficiente para que estes possam ser implementados de forma ubíqua sem degradar significativamente as taxas de encaminhamento de pacotes dos dispositivos de rede atuais. Um exemplo com pontos de medição passiva encontra-se retratado na Figura 2.

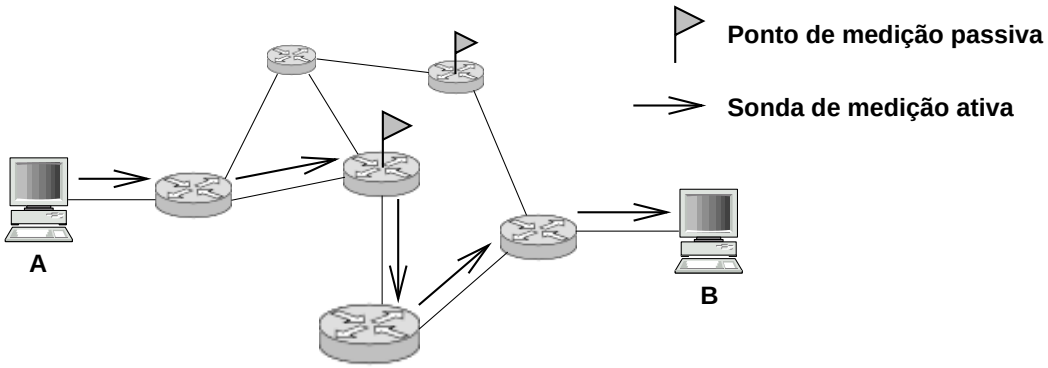


Figura 2: Exemplo de medições ativas e passivas.

Em contraste com as medições passivas, as medições ativas transmitem pacotes dedicados, chamados de sonda, e o resultado da travessia destes pacotes pela rede é monitorado para a inferência de características da rede. Medições ativas obtêm em geral pouca informação sobre pontos isolados da rede, mas podem fornecer uma representação do caminho entre dois pontos da rede. Na Figura 2, sondas de medição ativa enviadas

do nó A para o nó B fornecem informação sobre o caminho que conecta estes dois pontos finais. Nesse exemplo, supõe-se extremidades na rede sincronizadas. Do ponto de vista das medições passivas, somente um dos pontos de medição passiva registra a passagem das sondas ativas desse exemplo. Deve-se sempre considerar se o volume de tráfego introduzido por um método ativo e seu efeito no comportamento da rede estão influenciando significativamente, ou não, os resultados obtidos. Cenários híbridos podem ser concebidos onde ambas as medições ativa e passiva são combinadas para a estimação de características da rede [Ishibashi et al., 2004].

Além da classificação em ativos ou passivos, os métodos de medição também podem ser diferenciados por outras características [Chen, 2001]. Assim, medições podem ser:

- ligadas a um fluxo específico de pacotes ou concebidas para monitorar o comportamento da rede de forma mais genérica. No caso de estarem ligadas a um fluxo específico, as medições podem ser internas ao fluxo monitorado, onde campos adicionais no cabeçalho dos pacotes de dados são usados, ou externas ao fluxo monitorado, onde sondas adicionais aos pacotes de dados são adotadas;
- realizadas continuamente ou sob demanda;
- diretas ou indiretas. Por exemplo, pode-se medir diretamente uma determinada característica da rede ou usar algum dado de medição coletado de forma direta para estimar um outro aspecto indiretamente;
- unidirecionais ou bidirecionais;
- compostas de um ou múltiplos pontos de coleta de dados ou lançamento de sondas.

O projeto de métodos de metrologia precisa considerar compromissos entre essas diferentes escolhas. Por exemplo, métodos sob demanda podem ser preferíveis a métodos contínuos de medição para economizar banda passante. Em outro exemplo, estampas de tempo podem ser inseridas nos cabeçalhos de pacotes (interno) ou em pacotes dedicados (externo) para a monitoração do atraso de pacotes.

O desempenho da rede é geralmente associado à velocidade, correção e confiabilidade da entrega de um pacote IP ao seu destinatário. A velocidade é medida por parâmetros como o valor máximo, médio e a variação do atraso de um pacote. O atraso fim-a-fim de um pacote inclui o acúmulo dos atrasos de transmissão, propagação, permanência em filas e processamento em cada roteador intermediário. O atraso máximo é importante para aplicações interativas e a variação de atraso afeta a quantidade de memória necessária nas aplicações receptoras de fluxos multimídia para reprodução simultânea à transmissão. A correção contempla pacotes entregues com erros binários de transmissão, seja no cabeçalho, seja na carga útil. A correção não é motivo de grande preocupação no nível IP, pois supõe-se que erros binários na carga útil podem ser corrigidos pelos protocolos das camadas superiores, se a integridade de dados for importante para a aplicação. A confiabilidade refere-se à perda de pacotes ou à fração de pacotes não entregues. Existem diversas razões que podem causar a perda ou a incapacidade de entrega de um pacote: a saturação de filas em roteadores intermediários, erros binários no cabeçalho do pacote, a expiração do campo TTL (*Time-To-Live*), o endereço de destinatário não ser reconhecido ou alcançável, e a incapacidade de fragmentação, se esta for necessária. Aplicações que não retransmitem dados perdidos devido a restrições de tempo real podem sofrer degradação com a perda excessiva de pacotes. Mesmo aplicações que podem recuperar dados perdidos através de retransmissão podem ser ineficientes devido a múltiplas retransmissões.

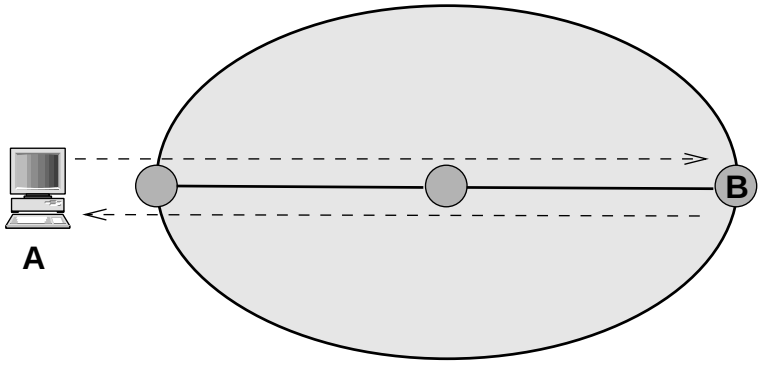
A ferramenta mais básica e tradicional para a monitoração na Internet é o popular `ping`. O `ping` envia um pedido de resposta (*echo request*) do protocolo ICMP (*Internet*

Control Message Protocol) para um determinado equipamento da rede, o qual por sua vez retorna uma mensagem ICMP de resposta ao pedido (*echo response*) [Postel, 1981]. Além do teste de conectividade, uma série de pacotes de `ping` também oferece uma estimação simples do desempenho genérico de um caminho na rede em termos de atraso e perda de pacotes. Em seu uso corriqueiro, a freqüência de transmissão de pacotes em um `ping` não é suficiente para afetar o desempenho da rede, porém o fato do intervalo entre envio de pacotes em uma mesma sessão ser fixo pode causar distorções nos dados observados. Se houver um comportamento periódico na rede, sondas periódicas como as do `ping` podem não observar esse comportamento corretamente. De maneira similar, uma amostragem periódica pode estar sincronizada com algum fenômeno imprevisível e, como conseqüência, o desempenho observado na rede será pior do que o real. Por essas razões, uma amostragem poissoniana, ou seja, com intervalos de tempo aleatórios seguindo uma distribuição exponencial entre amostras, é recomendada [Paxon et al., 1998].

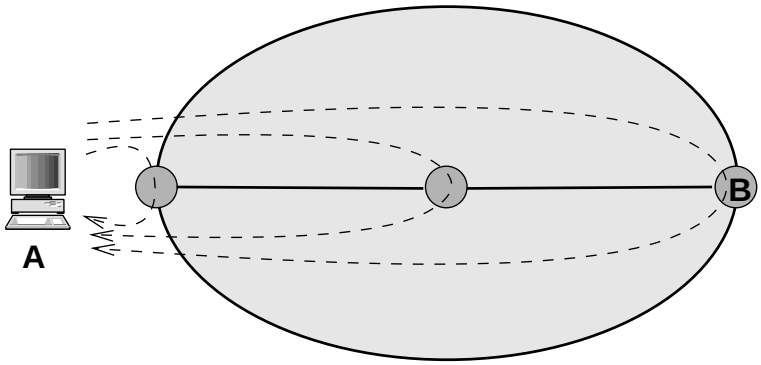
O `ping` fornece o RTT (*Round Trip Time*, ou seja, o tempo de ida e volta da fonte ao destino), mas o atraso unidirecional é um parâmetro importante para várias aplicações. Para que o atraso unidirecional seja medido, necessita-se de sincronização entre a fonte e o destinatário da transmissão. Uma alternativa é sincronizar a fonte e o destino das medições com servidores NTP (*Network Time Protocol*). No entanto, como os pacotes NTP são distribuídos junto ao tráfego comum da rede, os erros de sincronização são da ordem dos atrasos na rede [Paxson, 1998], o que pode comprometer a realização de medições precisas. Alguns trabalhos [Moon et al., 1999, Rocha et al., 2004, Wang et al., 2004] propõem métodos para estimar e remover a discrepância (*offset*) e a diferença entre as taxas de crescimento dos relógios (*skew*) entre os relógios dos nós finais sincronizados por NTP. O objetivo é viabilizar medições de atraso unidirecional. Uma alternativa direta a esse problema de sincronização é a adoção de placas GPS (*Global Positioning System*) [Enge e Misra, 1999] para a sincronização, porém estas necessitam estar ao alcance dos sinais de satélite e o seu custo pode também limitar uma adoção em larga escala. Em [Pásztor e Veitch, 2002], os autores propõem um relógio alternativo em *software* para aumentar a precisão de medições sem uso de placas GPS. A instalação de placas GPS é, pelo menos no momento, a solução adotada por diversos projetos de medição (ver Seção 10.2). Outro requisito para medições precisas é a capacidade de escrever estampas de tempo imediatamente antes da transmissão do pacote e de lê-las logo após a recepção. Essa característica pode ser obtida com equipamento próprio.

Os resultados de um `ping` mostram a rede como uma caixa preta, não havendo informações sobre os roteadores intermediários. Informações sobre os roteadores intermediários podem ser obtidas através de outra ferramenta popular: o *traceroute*. O `traceroute` utiliza engenhosamente a mensagem do ICMP gerada por roteadores intermediários obrigados a descartar um pacote devido à expiração do campo TTL. Ao enviar pacotes com TTL limitado e incrementando este limite a cada etapa, o `traceroute` identifica roteadores intermediários. No entanto, devido ao fato do IP não ser orientado à conexão, não há garantias de que o caminho identificado por essa ferramenta seja o mesmo seguido por um pacote de dados. A Figura 3 ilustra as diferenças de funcionamento entre as ferramentas `ping` e `traceroute`. Na Figura 3(a), o nó A ao fazer um `ping` no roteador B tem suas mensagens de `echo request` encaminhadas diretamente ao roteador B que responde com mensagens de `echo response`. Assim, o nó A obtém informações sobre a sua conectividade, o atraso e a taxa de perda de pacotes no caminho ao roteador B. Na Figura 3(b), por sua vez, ao realizar um `traceroute`, as mensagens enviadas ao roteador B pelo nó A com TTL limitado vão sendo descartadas pelos roteadores intermediários sucessivamente. Esses roteadores geram uma mensagem de erro ao nó A que pode então progressivamente inferir o caminho esperado ao rotea-

dor B, supondo-se que não haja mudanças de rotas no tempo da experiência nem a atuação de mecanismos de balanceamento de carga que podem encaminhar pacotes com o mesmo par origem-destino por rotas diferentes para distribuir a carga da rede.



(a) `ping`



(b) `traceroute`

Figura 3: Ilustração do funcionamento das ferramentas `ping` e `traceroute`.

2.2. Classificação e caracterização de tráfego

A monitoração mais comum para o gerenciamento de redes contabiliza o volume de tráfego em uma determinada interface ou a taxa de perda de pacotes em um determinado roteador intermediário. Esses parâmetros ao nível de pacotes são úteis para uma visão genérica do comportamento da rede, mas em muitos casos uma análise mais detalhada é necessária. O tráfego presente na Internet pode ser observado segundo diferentes níveis de granulosidade. Ao observarmos os pacotes que passam por um determinado ponto de medição de forma mais detalhada, podemos identificar o protocolo de transporte adotado e o protocolo da camada superior atendido. Dessa forma, podemos identificar quais aplicações estão mais presentes na rede. A Figura 4 apresenta a vazão de diferentes aplicações em um POP (*Point of Presence*) servindo usuários conectados por ADSL (*Asynchronous Digital Subscriber Line*) da FranceTelecom localizado na região parisiense (fonte: Projeto Metropolis [Metropolis, 2001]). Esse exemplo mostra claramente o domínio de aplicações de navegação (`http`) e compartilhamento de arquivos em *peer-to-peer* (P2P).

Um enlace da Internet pode transportar uma coleção de fluxos de uma variedade de aplicações, transmitidos por diversos protocolos de transporte, especialmente TCP e UDP [Brownlee et al., 1999]. As primeiras análises e simulações do comportamento do TCP eram focadas no comportamento de estado estacionário, usando fontes de carga

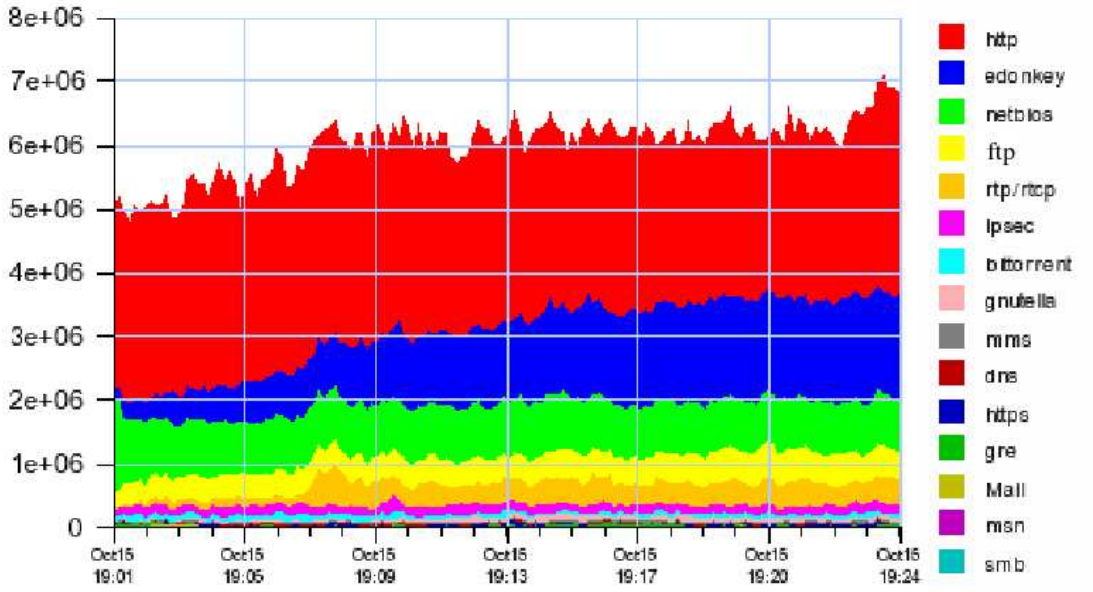


Figura 4: Vazão em Mbps em um POP da FranceTelecom.

"infinita", ou seja, grandes transferências de arquivo. Essas análises também supunham que fluxos TCP de grande volume não seriam significativamente afetados por fluxos TCP pequenos. A literatura de Metrologia na Internet faz uso freqüente de analogias com animais para a classificação de fluxos [Soule et al., 2004b, Brownlee, 2005]. Quanto ao seu tamanho, os fluxos de grande volume, como transferência de arquivos, são chamados de elefantes. Por sua vez, os fluxos de pequeno volume, como os pedidos `http`, são conhecidos como camundongos. Os fluxos elefante podem ser de duas a três ordens de magnitude maiores do que os camundongos [Estan e Varghese, 2002, Papagiannaki et al., 2004a].

A diferença fundamental entre os fluxos de rede ditos elefantes e camundongos se refere ao fato que uma sessão TCP caracterizada como elefante ultrapassa a fase de *slow start* do TCP. Logo, o comportamento de um fluxo elefante, incluindo a sua interação com outras sessões TCP concorrentes, é condicionada pelos algoritmos realimentados de controle de congestionamento do TCP. Simultaneamente, os fluxos camundongos não podem ser controlados por essa realimentação, pois eles são transmitidos em sua integridade antes que o TCP seja capaz de aplicar seu mecanismo de controle de congestionamento.

Modelos mais recentes do comportamento do TCP estão crescentemente considerando a interação entre fluxos elefantes e camundongos na rede. Por exemplo, em [Joo et al., 1999] é analisada a vazão esperada de fluxos TCP e como estes interagem com fluxos concorrentes. Eles constataram que múltiplos elefantes podem sincronizar entre si, o que pode causar o descarte de pacotes em roteadores. Outra constatação é que embora os elefantes sejam responsáveis pela maior parte dos octetos presentes na rede, o número de pacotes transportados por fluxos camundongos pode ser suficiente para gerar perdas esporadicamente. Os autores também examinaram a dinâmica da perda de pacotes e concluíram que os fluxos camundongo são capazes de romper efeitos de sincronismo, levando a uma utilização mais eficiente dos recursos da rede. Julga-se que este efeito de rompimento de sincronismo possa justificar o motivo pelo qual o serviço de melhor esforço tenha alcançado tanto sucesso como o serviço básico da Internet.

Como uma alternativa à classificação de fluxos por tamanho (em número de octetos), isto é, em elefantes ou camundongos, pode-se também classificar os fluxos em termos de seu tempo de vida (em segundos). Em [Shaikh et al., 1999], os autores identificaram fluxos com duração de até 2.000 s que representam uma grande proporção dos

octetos nos enlaces observados. Baseados nessa informação, eles propuseram roteadores sensíveis à carga para buscar melhores rotas para esses fluxos de longa duração, portanto incrementando a utilização da rede.

Em [Brownlee e Claffy, 2002], são propostos novos critérios de classificação baseados no tempo de duração do fluxo. Por um lado, os autores identificam uma grande quantidade de fluxos bastante rápidos, com menos de 2 s de duração. Esses fluxos rápidos, chamados de libélulas, representam pelo menos 45% dos fluxos presentes nos enlaces observados. Aproximadamente 98% dos fluxos observados têm menos de 15 minutos de duração. Por outro lado, os remanescentes 2% dos fluxos alcançam duração de horas ou até mesmo dias. Esses fluxos de longa duração são chamados de tartarugas. Entretanto, mesmo representando apenas 2% dos fluxos, os fluxos ditos tartaruga carregam de 40 a 50% do total de octetos nos enlaces observados.

Portanto, os fluxos presentes na Internet podem ser classificados não somente pelo seu tamanho em elefantes ou camundongos, mas também pelo seu tempo de duração em libélulas ou tartarugas. Além disso, em [Brownlee e Claffy, 2002], demonstra-se que o tamanho do fluxo em octetos e o seu tempo de duração são dimensões independentes, sendo cada uma de interesse para a compreensão do comportamento da rede.

3. Estimação de banda passante

Administradores de rede que possuam acesso privilegiado a um roteador ou comutador conectado a um enlace de interesse podem medir alguns parâmetros relacionados à banda passante diretamente. Por exemplo, esses parâmetros podem ser a capacidade nominal do enlace, a sua utilização média e a quantidade de pacotes ou octetos transmitidos em um período de tempo. Isso pode ser feito através do protocolo SNMP. Entretanto, esse acesso é tipicamente disponível somente aos administradores e não aos usuários finais. Os usuários finais podem somente *estimar* a banda passante de enlaces ou do caminho fim-a-fim com base em medições. Mesmo administradores de rede, com acesso privilegiado a alguns roteadores, precisam determinar a banda passante entre os roteadores sob seu controle e roteadores externos. Nesse caso, esses administradores também utilizam a estimação de banda passante baseada em medições fim-a-fim.

A estimação de banda passante é de grande utilidade para a comunicação em redes de pacotes, pois ela indica o volume de dados que um enlace ou caminho de rede pode transportar por unidade de tempo. Para diversas aplicações, a banda passante disponível influencia diretamente o desempenho destas. Alguns casos onde uma estimação acurada de banda passante contribui significativamente são a otimização do desempenho fim-a-fim ao nível de transporte, o roteamento de sobrecamada (*overlay*) e a distribuição de arquivos em sistemas P2P. Técnicas para a estimação acurada de banda passante também são importantes para o suporte à engenharia de tráfego e ao planejamento de capacidade da rede. Mesmo aplicações interativas, que normalmente são mais sensíveis ao atraso do que à banda passante disponível, podem beneficiar-se dos baixos atrasos fim-a-fim associados a enlaces de alta capacidade e baixa latência de transmissão de pacotes.

Em [Prasad et al., 2003b] são definidas métricas associadas à estimação de banda passante. Primeiro, diferencia-se entre banda passante de um enlace e a banda passante de uma seqüência de enlaces, ou o caminho fim-a-fim. Segundo, as métricas são capacidade e banda passante disponível. Capacidade é a banda passante máxima que pode-se alcançar em um enlace ou caminho. Banda passante disponível é a banda passante máxima ociosa em um enlace ou caminho. Identificar a menor capacidade disponível ao longo do caminho, ou seja, o enlace de gargalo (*bottleneck*) em um caminho também recebe grande interesse de pesquisa [Hu et al., 2005].

Existem três técnicas principais para a estimação de banda passante. São elas: a sondagem com pacotes de tamanho variável, a dispersão de par de pacotes e o uso de fluxos periódicos auto-carregáveis. A primeira técnica infere a capacidade dos enlaces individualmente. A segunda estima a capacidade fim-a-fim. A terceira técnica estima a banda passante disponível fim-a-fim. Em geral, essas técnicas supõem que, durante o processo de medição, o caminho fim-a-fim permanece o mesmo e que o tráfego é estacionário. Alterações dinâmicas no roteamento ou na carga podem criar erros em qualquer dessas metodologias. Em [Jain e Dovrolis, 2004], são apontados diversos problemas que precisam ser examinados com mais cuidado na área de estimação de banda passante disponível.

A primeira ferramenta baseada na sondagem com pacotes de tamanho variável foi o `pathchar` [Jacobson, 1997]. Melhorias e refinamentos a essa técnica foram posteriormente propostos em [Downey, 1999, Lai e Baker, 2000]. A idéia básica desta técnica é medir o RTT de uma fonte até cada salto no caminho como uma função do tamanho do pacote-sonda adotado. Para tanto, usa-se o campo TTL do cabeçalho IP para forçar o descarte de pacotes em um salto determinado, de forma semelhante a utilizada pela ferramenta `traceroute`. Assim, a fonte utiliza as mensagens de erro ICMP enviadas pelos roteadores intermediários para avaliar o RTT até cada roteador intermediário. O RTT medido dessa maneira consiste, para os caminhos de ida e de volta, dos atrasos de transmissão, de propagação e de enfileiramento. Enviando vários pacotes-sonda de cada tamanho, a técnica supõe que pelo menos um pacote e a sua respectiva mensagem de erro ICMP não sofrerão atrasos em filas. Portanto, o menor RTT medido para cada tamanho de pacote consistirá de dois termos: um provocado pelos atrasos de propagação que são independentes do tamanho do pacote e outro proporcional ao tamanho do pacote devido ao atraso de transmissão em cada enlace do caminho. A Figura 5 ilustra esse processo para o primeiro roteador do caminho utilizando-se 6 amostras para 8 diferentes tamanhos de pacotes-sonda. Ao realizarmos a interpolação linear entre os mínimos RTTs dos diferentes tamanhos de pacote, obtemos a linha tracejada mostrada na Figura 5. A interceptação dessa linha tracejada com o eixo dos y corresponde ao envio de um pacote de tamanho nulo, ou seja, o ponto de interseção estima o atraso de propagação. A capacidade do enlace medido é estimada pelo inverso da inclinação da linha tracejada mostrada na Figura 5. Erros de subestimação podem surgir em decorrência da presença de comutadores de nível 2 [Prasad et al., 2003a], pois estes introduzem atraso de transmissão sem gerar mensagens de erro ICMP por não implementarem a camada IP.

A estimação da capacidade fim-a-fim é o objetivo das técnicas baseadas em sondagem por par de pacotes. Essa técnica possui suas origens no trabalho de [Jacobson, 1988]. Uma fonte envia múltiplos pares de pacotes ao receptor, onde cada par de pacotes consiste de dois pacotes do mesmo tamanho. A dispersão medida entre esses dois pacotes em um enlace específico do caminho é a distância temporal entre o último bit de cada pacote. Busca-se medir a menor capacidade entre os enlaces intermediários, ou seja, o gargalo do caminho fim-a-fim. Recentemente, uma técnica semelhante foi proposta como um meio de classificar o tipo de acesso à rede utilizado em três categorias: Ethernet, rede local sem fio e conexão de baixa capacidade (cabo, ADSL ou linha discada) [Wei et al., 2005].

A Figura 6 mostra a dispersão de um par de pacotes antes e depois deste par atravessar um enlace de capacidade C_i, supondo-se que não haja tráfego concorrente. A dispersão no primeiro enlace é igual a $\Delta_1 = L/C_1$, onde L é o tamanho de cada pacote que compõe o par. Se a dispersão antes de um enlace de capacidade C_i é Δ_x, a dispersão Δ_y posterior ao enlace é dada por:

$$\Delta_y = \max\left(\Delta_x, \frac{L}{C_i}\right). \quad (1)$$

Conforme o par de pacotes atravessa os enlaces de um caminho livre, a dispersão Δ_R medida no receptor é dada por:

$$\Delta_R = \max_{i=1,\ldots,H}\left(\frac{L}{C_i}\right) = \frac{L}{min_{i=1,\ldots,H}(C_i)} = \frac{L}{C}\,, \quad (2)$$

onde H é o número de enlaces atravessados e C é a capacidade de gargalo fim-a-fim. Portanto, um receptor pode estimar a capacidade de um caminho usando $C = L/\Delta_R$.

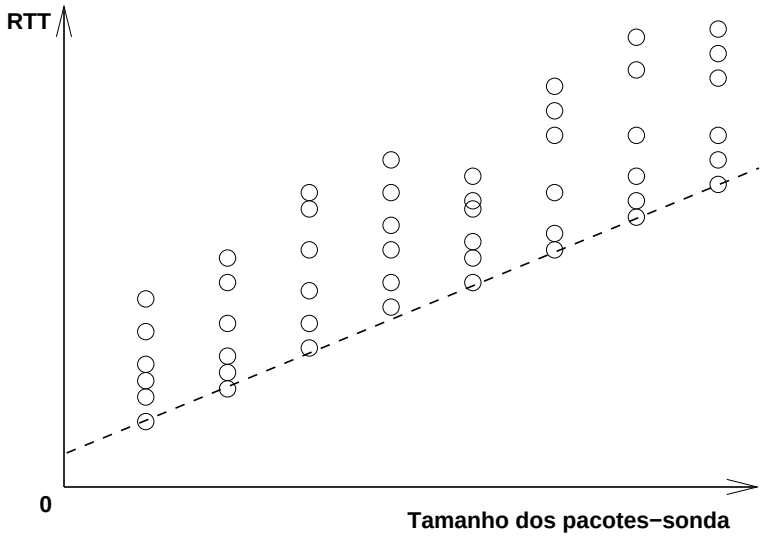


Figura 5: Funcionamento da sondagem com pacotes de tamanho variável.

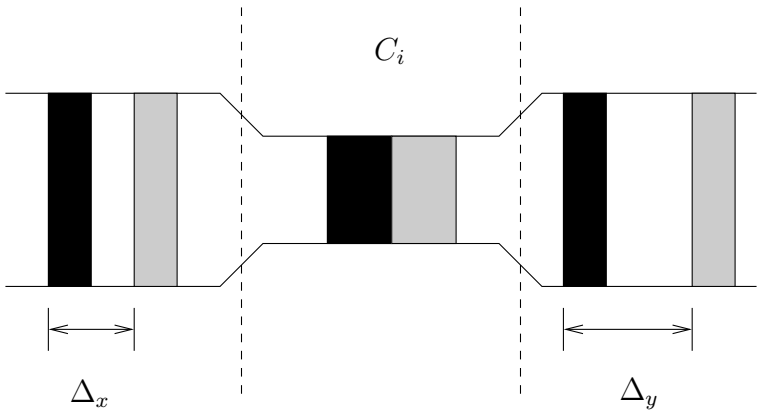


Figura 6: Dispersão em um par de pacotes.

A suposição de que o caminho encontra-se vazio, ou seja, sem qualquer outro tráfego é longe de ser realista. Pode haver subestimação da capacidade caso os pacotes de tráfego concorrentes sejam transmitidos entre os pacotes-sonda do par em um enlace específico, aumentando assim a dispersão além de L/C. Também pode haver superestimação caso o tráfego concorrente atrase o primeiro pacote de um par mais do que o segundo pacote em algum enlace posterior ao enlace de gargalo. Ao invés de um único par de pacotes, pode-se adotar uma seqüência, ou trem, de múltiplos pares de pacotes e calcular uma taxa de dispersão média para a inferência da capacidade do enlace

Tabela 1: Ferramentas de estimação de banda passante.

Ferramenta	Métrica medida	Metodologia	Referência
`pathchar`	capacidade por enlace	variação de pacotes	[Jacobson, 1997]
`clink`	capacidade por enlace	variação de pacotes	[Downey, 1999]
`bprobe`	capacidade fim-a-fim	pares de pacotes	[Carter e Crovella, 1996]
`nettimer`	capacidade fim-a-fim	pares de pacotes	[Lai e Baker, 2000]
`pathrate`	capacidade fim-a-fim	pares de pacotes	[Dovrolis et al., 2004]
`sprobe`	capacidade fim-a-fim	pares de pacotes	[Saroiu et al., 2002]
`cprobe`	banda disponível	trem de pacotes	[Carter e Crovella, 1996]
`pathload`	banda disponível	fluxo periódico	[Jain e Dovrolis, 2002]
`IGI/PTR`	banda disponível	fluxo periódico	[Hu e Steenkiste, 2003]
`pathchirp`	banda disponível	fluxo exponencial	[Ribeiro et al., 2004]
`pathneck`	banda disponível	trem de pacotes	[Hu et al., 2004]

de gargalo [Dovrolis et al., 2004]. Diversos autores propõem métodos para amenizar os efeitos do tráfego concorrente sobre a técnica de par de pacotes. Alguns desses trabalhos são [Carter e Crovella, 1996, Lai e Baker, 1999, Dovrolis et al., 2001].

Para a estimação da banda passante disponível fim-a-fim, existe a técnica de fluxos periódicos auto-carregáveis [Jain e Dovrolis, 2002]. Nessa técnica, uma fonte envia para um receptor um determinado número de pacotes de igual tamanho a uma taxa R fixa, constituindo assim um fluxo periódico de pacotes. A metodologia proposta monitora as variações dos atrasos unidirecionais dos pacotes-sonda. Se a taxa R de envio do fluxo é maior do que a banda passante disponível A ao longo do caminho, o fluxo de pacotes causa uma sobrecarga momentânea na fila do enlace de gargalo. Os atrasos unidirecionais continuarão a crescer conforme cada pacote do fluxo seja enfileirado no enlace de gargalo. Por outro lado, se a taxa R do fluxo for menor que a banda passante disponível A, os pacotes-sonda atravessam o caminho sem sofrer atrasos adicionais não aumentando os atrasos unidirecionais. Essa técnica procura aproximar a taxa de fluxo R à banda passante disponível A em um processo interativo semelhante à busca binária. Caso a banda passante disponível A varie ao longo das medições, o método pode detectar tal variação e retornar uma região referente à banda passante disponível estimada.

A Tabela 1 apresenta uma lista de ferramentas propostas para a estimação de banda passante, mostrando as diferentes métricas medidas e metodologias utilizadas. Uma análise recente de ferramentas de domínio público para a estimação de banda passante pode ser encontrada em [Shriram et al., 2005].

4. Estimação de matrizes de tráfego

Uma matriz que forneça os volumes de tráfego entre a origem e o destino em um domínio de rede possui uma grande utilidade potencial para o planejamento de capacidade e o gerenciamento de uma rede IP. O conhecimento da matriz de tráfego permite uma análise de confiabilidade para que em casos de falha em algum enlace do domínio seja possível ao operador prever as novas cargas nos enlaces restantes. Quando medições diretas ao nível de fluxos estão disponíveis, matrizes de tráfego acuradas podem ser derivadas seguindo as abordagens detalhadas em [Feldmann et al., 2001]. No entanto, matrizes de tráfego são freqüentemente difíceis de medir-se diretamente em grandes redes IP operacionais, pois o custo de obter medições diretas é proibitivo atualmente nessas grandes redes devido à infra-estrutura adicional necessária [Papagiannaki et al., 2004b]. Portanto,

em geral, matrizes de tráfego não estão disponíveis para os grandes operadores de rede, impedindo, por exemplo, que estes operadores possam quantificar o custo de provisão de qualidade de serviço em comparação com um super-aprovisionamento de recursos. Contudo, medições sobre a carga presente em cada enlace estão prontamente disponíveis em redes IP através das ferramentas comuns de gerenciamento. Logo, para estimar uma matriz de tráfego em um grande operador IP é necessário estimar as demandas fim-a-fim em um domínio a partir do conhecimento das cargas nos enlaces individuais. Esse problema, comumente chamado de tomografia de rede, de estimar uma matriz de tráfego a partir de informações parciais sobre as cargas individuais dos enlaces recebeu grande atenção dos pesquisadores nos últimos anos.

O problema de estimação de matrizes de tráfego pode ser formalizado da seguinte maneira [Medina et al., 2002]. Seja c o número de pares origem-destino (OD) em um domínio de rede. Se esse domínio tiver n nós de interesse em sua fronteira, então $c = n(n-1)$. A seguir, ordena-se os pares OD em um vetor $\mathbf{x}$ e define-se $x_j \in \mathbf{x}$ como sendo o volume de tráfego transmitido pelo par OD j. Seja $\mathbf{y} = [y_1, \ldots, y_r]^T$ o vetor que representa o volume de tráfego nos enlaces individualmente, onde y_l indica o volume de tráfego para o enlace l e r denota o número de enlaces na rede. Os vetores $\mathbf{x}$ e $\mathbf{y}$ estão relacionados através de uma matriz de roteamento $\mathbf{A}$ de dimensões r por c. A matriz $\mathbf{A}$ é composta por valores $\{0, 1\}$ com linhas representando os enlaces da rede e colunas representando os pares OD. O elemento $a_{ij} = 1$ indica que o enlace i pertence ao caminho associado ao par OD j, enquanto $a_{ij} = 0$ indica o contrário. Portanto, os fluxos OD estão relacionados aos volumes de tráfego nos enlace de acordo com a seguinte relação linear:

$$\mathbf{y} = \mathbf{A}.\mathbf{x} \tag{3}$$

Para melhor compreender a composição de $\mathbf{y}$, $\mathbf{A}$ e $\mathbf{x}$, observemos a Figura 7 que ilustra o problema de estimação de matrizes de tráfego. Nessa figura há três nós de interesse A, B e C interconectados pelo roteador R através dos enlaces 1, 2 e 3. As informações sobre a carga individual nesses enlaces estão disponíveis e compõem o vetor $\mathbf{y} = [y_1, y_2, y_3]^T$. Os pares OD que são os elementos da matriz de tráfego $\mathbf{x}$ são representados pelas linhas tracejadas na Figura 7. O problema consiste em estimar a matriz de tráfego $\mathbf{x}$ cujos elementos são x_1 que representa o par OD entre os nós A e B, x_2 que representa o par OD entre os nós B e C, e x_3 que representa o par OD entre os nós A e C. Deve-se observar que neste caso, por exemplo, $y_1 = x_1 + x_3$. Assim, a relação $\mathbf{y} = \mathbf{A}.\mathbf{x}$ para o caso ilustrado na Figura 7 é dada por:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{4}$$

A matriz de roteamento em redes IP pode ser obtida pela coleta de dados dos protocolos de roteamento e pelo cálculo dos caminhos mais curtos entre todos os pares OD. Os volumes de tráfego nos enlaces individuais estão disponíveis através do uso de SNMP. Logo, o problema é calcular o vetor $\mathbf{x}$, ou seja, encontrar um conjunto de fluxos OD que possa reproduzir os volumes de tráfego dos enlaces da maneira mais próxima possível. O problema associado à Equação (3) é altamente subdeterminado, pois em praticamente todas as redes o número de pares OD é muito maior do que o número de enlaces na rede, $r \ll c$. Isto significa que há um número infinito de soluções possíveis para o vetor $\mathbf{x}$.

Ao considerar-se a existência de diversos (K) períodos de medição, denota-se os volumes de tráfego em cada enlace como y_l^k para indicar a carga média no enlace l no

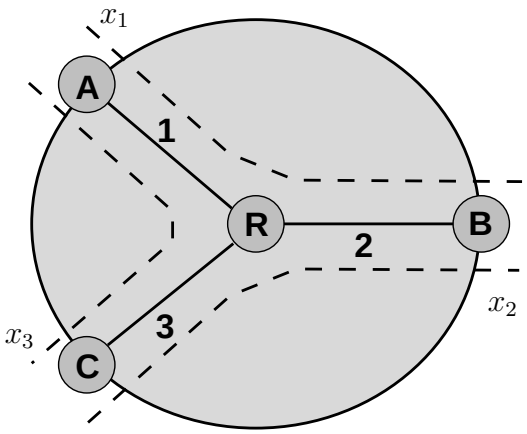


Figura 7: Exemplo de estimação de matriz de tráfego.

período de medição k, $k = 1, \ldots, K$. De forma similar, x_j^k representa as demandas do tráfego, indicando a demanda de tráfego para o par OD j no período de medição k. Os fluxos OD e os volumes de tráfego dos enlaces estão relacionados através da matriz de roteamento $\mathbf{A}$ da maneira seguinte:

$$\mathbf{y}^k = \mathbf{A}.\mathbf{x}^k \tag{5}$$

Em [Medina et al., 2002] são comparados três métodos para a estimação de matrizes de tráfego. O primeiro método aplica diretamente uma abordagem de programação linear. O segundo utiliza técnicas de inferência bayesiana. O terceiro método adota um algoritmo de maximização de expectativas para calcular estimativas de máxima verossimilhança (*maximum likehood*). Algumas lições desta análise comparativa são:

1. basear-se nos contadores SNMP dos enlaces como única fonte de informação sobre a rede gera um conjunto de informações parciais que pode ser extremamente limitante. Os operadores de rede comumente possuem uma grande quantidade de conhecimento e informações sobre a sua própria rede. Um passo importante para melhorar as técnicas de estimação de matrizes de tráfego é a elaboração de métodos capazes de incorporar informações específicas da rede. Essa informação pode ser obtida por medições isoladas ao nível de pacote ou de fluxo, ou ainda, pelo conhecimento da própria rede com informações como o tamanho dos POPs em termos de capacidade total, número de clientes ou *peers* por POP;
2. supor modelos para o comportamento dos fluxos OD pode gerar resultados com acurácia limitada se esses modelos não refletirem a verdadeira natureza do comportamento dos fluxos. Os modelos comumente adotados preenchem a matriz de tráfego com estimativas da média de demanda de tráfego entre pares OD. Se o verdadeiro comportamento do tráfego é, por exemplo, multimodal, então calcular a média levará os pares OD para algum valor central que não representa nenhum dos modos. Um outro passo importante para melhorar as técnicas de estimação de matrizes de tráfego é a elaboração de modelos que representem melhor o comportamento real do tráfego presente no domínio da rede. Usando medições diretas pode-se obter componentes da matriz de tráfego que ajudam na compreensão do comportamento de alguns elementos da matriz de tráfego real. Fazendo uso dessas propriedades, pode-se construir modelos mais realistas;
3. o desempenho de técnicas de inferência estatística em termos dos seus erros é altamente dependente da qualidade da informação prévia usada como entrada. Portanto, a geração de melhores matrizes iniciais é outro passo importante para a melhoria da estimação de matrizes de tráfego.

Ao obter-se uma estimação da matriz de tráfego, esta estimação pode conter erros que a afastam da matriz real. Em [Roughan et al., 2004] são avaliados os efeitos de estimações inexatas de matrizes de tráfego na engenharia de tráfego. Um problema comum em engenharia de tráfego é otimizar o roteamento de forma a minimizar o congestionamento. Basicamente, ajusta-se os parâmetros de roteamento em um domínio de rede de acordo com a matriz de tráfego de forma a minimizar o congestionamento nesta matriz de tráfego. Os parâmetros de roteamento determinam, para cada par OD, a fração de tráfego que segue por diferentes caminhos da origem ao destino. Dado este problema, os autores avaliam os efeitos da utilização de matrizes de tráfego estimadas para alimentar este processo de otimização do roteamento. Os autores demonstram que a combinação da técnica de otimização de rotas para OSPF proposta em [Fortz e Thorup, 2000] e da estimação de matrizes de tráfego descrita em [Zhang et al., 2003a] apresenta os melhores resultados de desempenho.

Outras abordagens para o problema de estimação de matrizes de tráfego foram propostas recentes, tais como uma abordagem baseada em teoria da informação [Zhang et al., 2003b] e outra que adota um estimador da variância das matrizes de tráfego [Soule et al., 2004a]. O efeito de mudanças no roteamento na variação das matrizes de tráfego é investigado em [Teixeira et al., 2005]. Uma avaliação do equilíbrio entre a adoção de medições, o uso de inferência e modelagem para a estimação de matrizes de tráfego pode ser encontrada em [Soule et al., 2005].

5. Amostragem de tráfego e diagnóstico de anomalias

Anomalias em redes são definidas como mudanças significantes e pouco comuns nos padrões de tráfego em um ou múltiplos enlaces da rede [Barford et al., 2002]. O diagnóstico dessas anomalias envolve a detecção, a identificação e a quantificação desses fenômenos. Esse procedimento pode ser essencial para os operadores e para os usuários finais. Independente das anomalias presentes na rede terem sido causadas intencionalmente ou não, a sua análise é importante por duas razões. Primeiro, anomalias podem causar congestionamento na rede e esgotar recursos dos roteadores, o que torna a sua detecção crucial do ponto de vista dos operadores. Segundo, algumas anomalias não necessariamente afetam o desempenho da rede, mas elas podem ter um grande impacto em clientes ou usuários finais. O diagnóstico de anomalias apresenta grandes desafios, pois é necessário extrair padrões anômalos de grandes volumes de dados e as causas das anomalias podem ser bastante variadas. Como exemplos de causas de anomalias podemos listar ataques distribuídos de negação de serviço, enganos na configuração de roteadores ou resultados de modificações nas políticas de roteamento BGP.

Ao detectar-se uma anomalia, uma propriedade interessante é a capacidade de rastrear a trajetória dos pacotes que compõem um determinado tráfego em seu caminho no interior de um domínio. Esse tipo de capacidade baseada em medições torna uma rede mais resistente a falhas e à presença de anomalias. Em [Duffield e Grossglauser, 2001] é proposto um método de amostragem das trajetórias de pacotes em uma rede. A metodologia de amostragem seleciona um subconjunto dos pacotes presentes na rede, mas se um pacote é selecionado em um enlace, ele o será em todos os outros enlaces que o pacote atravessar. Ao atravessar a rede, cada pacote indica implicitamente se ele deve ser alvo da amostragem ou não pela sua parte invariante, ou seja, aqueles bits de informação que não mudam de um enlace a outro. Um valor de dispersão (*hash*) é calculado em cada roteador para esses bits de informação invariantes. Então, somente os pacotes cujo valor de dispersão esteja em um determinado intervalo são selecionados para serem amostrados. Dessa forma, se a mesma função de dispersão for utilizada através do domínio para sele-

cionar pacotes para amostragem, então há a garantia de que ou um pacote é selecionado em todos os enlaces do domínio que ele atravesse ou o pacote não é nunca selecionado. Portanto, o método permite coletar amostras das trajetórias de um subconjunto de pacotes. Claramente, a escolha da função de dispersão é decisiva para que o subconjunto amostrado não seja de maneira alguma tendencioso. Para tanto, o processo de amostragem, embora uma função determinística do conteúdo de cada pacote, deve se assemelhar a um processo de amostragem aleatório.

Para se obter as amostras de trajetória é necessário realizar a etiquetagem dos pacotes amostrados. Para isso, é suficiente gerar um identificador, ou etiqueta, único por pacote para cada pacote amostrado durante um período de tempo. Como a etiqueta é única, pode-se saber o conjunto de enlaces percorrido por um determinado pacote, pois estes terão relatado a passagem da mesma etiqueta. Em [Duffield e Grossglauser, 2001], os autores propõem para a identificação dos pacotes uma segunda função de dispersão que gere etiquetas únicas no período de monitoração com alta probabilidade. O tamanho das etiquetas dos pacotes pode ser relativamente pequeno, como 20 bits. Como o tráfego de medições coletadas dos nós pertencentes ao domínio consiste somente dessas etiquetas (além de uma pequena informação adicional), a sobrecarga para a coleta das amostras de trajetória é relativamente pequena.

A Figura 8 apresenta um exemplo de amostragem de trajetória. As setas sólidas representam o percurso através do domínio observado de um pacote cujo conteúdo invariável ativa o processo de amostragem. Com isso, todos os roteadores, ao aplicarem a mesma função de dispersão, selecionam esse pacote para amostragem. Usando a segunda função de dispersão para a identificação do pacote, esses mesmos roteadores enviam a etiqueta gerada para o sistema centralizador de medidas, como indicado pelas setas tracejadas. Embora isso baste para identificar as trajetórias dos pacotes amostrados no interior do domínio, alguma informação adicional pode ser necessária para diversos propósitos de monitoração. Essa informação adicional pode incluir os endereços da fonte e do destino do pacote, assim como o seu tamanho. Entretanto, é suficiente coletar esse tipo de informação uma vez por pacote amostrado. Logo, os nós de ingresso no domínio podem ser configurados para recuperar essa informação adicional além das etiquetas de identificação, enquanto os demais nós somente recolhem as etiquetas, como ilustrado na Figura 8. Vale ressaltar que pacotes multicast não requerem nenhum tratamento adicional. Simplesmente, nesse caso, a trajetória associada a um pacote multicast é uma árvore ao invés de um caminho. Um abordagem semelhante à amostragem de trajetórias de fluxos é adotada em [Snoeren et al., 2002] para rastrear um ataque em curso.

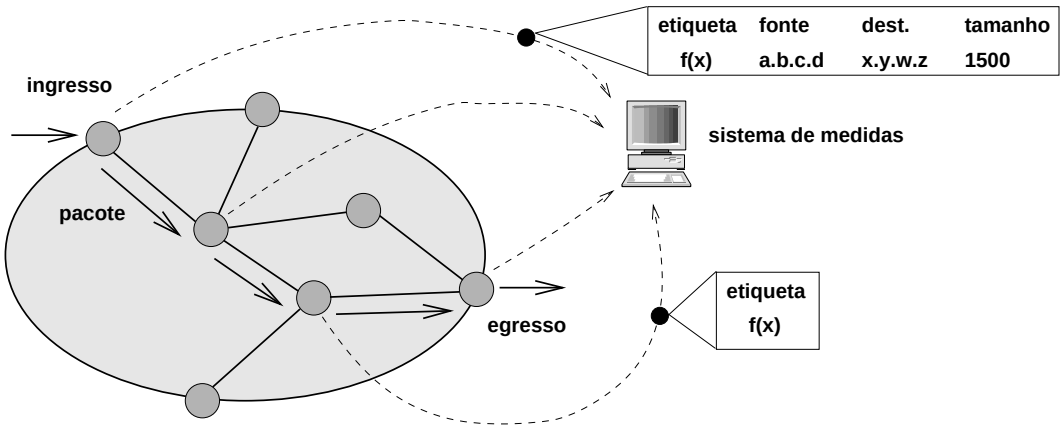


Figura 8: Amostragem de trajetórias.

Além das anomalias decorrentes de ataques em curso, outra classe comum de anomalias envolve o encaminhamento IP, como as anomalias causadas por falhas de equipamento, erros (*bugs*) de implementação ou erros de configuração. Essas anomalias podem degradar significativamente e até mesmo interromper o serviço de rede. A detecção robusta e confiável de tais anomalias é essencial para a identificação rápida do problema e para a tomada de ações que o corrijam. Em [Roughan et al., 2004], os autores propõem um sistema para a detecção de anomalias de encaminhamento cuja principal contribuição é a combinação de dados de tráfego e de roteamento para prover uma detecção confiável dessas anomalias com baixa taxa de falsos positivos.

Um método genérico para a detecção, identificação e quantificação de anomalias é proposto em [Lakhina et al., 2004]. A detecção consiste em determinar os pontos no tempo nos quais a rede enfrenta uma anomalia. A identificação envolve a classificação da anomalia em investigação a partir de um conjunto de anomalias conhecidas. A quantificação mede a importância da anomalia ao estimar a quantidade de tráfego anômalo de um determinado tipo presente na rede. Portanto, para um diagnóstico bem sucedido de uma anomalia no volume de tráfego, é necessário detectar o momento de ocorrência desta anomalia, identificar a sua causa e quantificar o seu tamanho. Nesse trabalho, os autores propõem uma metodologia para separar o tráfego da rede em um componente considerado normal que é dominado pelo tráfego previsível e um componente anômalo caracterizado por picos significativos de tráfego. Mostra-se que essa separação é possível através da Análise de Componentes Principais (*Principal Components Analysis* – PCA) [Bryant e Yarnold, 1998].

6. Proximidade em redes

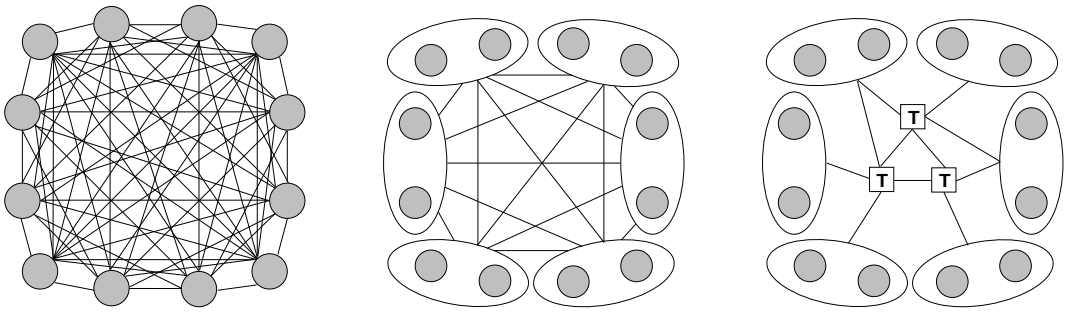
Há uma crescente necessidade de se estimar distâncias entre nós na Internet de maneira rápida e eficiente [Huffaker et al., 2002]. Distância neste contexto refere-se a alguma métrica de desempenho da rede como atraso ou banda passante. A consideração da distância entre nós da rede em termos de atraso é importante para aplicações e serviços, tais como serviços de hospedagem distribuídos, busca de servidores mais próximos, redes de multicast aplicativo, redes de distribuição de conteúdo e sistemas de compartilhamento de arquivos em P2P.

Embora os nós na rede possam medir características dos caminhos com ferramentas como `ping` ou `traceroute`, a realização de medições de desempenho antes de cada interação na Internet inevitavelmente levaria a uma alta sobrecarga tanto para os nós finais quanto para a própria rede. Portanto, um serviço útil para a Internet é o fornecimento de um meio para um nó qualquer estimar a distância entre dois outros nós na Internet de maneira rápida e eficiente. A idéia é fornecer um serviço capaz de estimar a proximidade de rede entre nós em termos de atraso de forma escalável, porém sem a necessidade de realização de medições diretas entre estes nós.

6.1. Estimação do atraso entre dois pontos da rede

IDMaps [Francis et al., 2001] foi a proposta pioneira de um arquitetura global para a estimação de distância entre nós na Internet. A forma de distância entre nós mais acurada que poderia ser medida por IDMaps consiste nas distâncias entre qualquer par de nós globalmente alcançáveis por um endereço IP. A distância de um endereço IP a outro seria então determinada para cada par de nós. A imensa escala dessa informação, da ordem de H^2 onde H é o número de nós na Internet, o que pode alcançar centenas de milhões, torna esta forma simples de distância inviável. Não somente o acompanhamento periódico da distância entre os nós de tal número de pares é inviável, mas também a identificação des-

ses nós em uma rede que está em constante mudança. A Figura 9(a) mostra essa situação para 12 nós, onde cada linha representa a distância de rede medida entre estes nós.



(a) Malha (b) Prefi xos (c) *Tracers*

Figura 9: Diversas formas da informação de distância.

Uma alternativa à medição das distâncias entre os nós de forma individual é medir a distância entre cada prefixo de endereços alcançável globalmente na Internet até cada outro prefixo. Essa configuração está ilustrada na Figura 9(b), onde novamente as linhas representam as distâncias medidas. Um prefixo de endereços é uma faixa de endereços IP consecutivos na qual todos os nós com endereços nesta faixa podem ser considerados eqüidistantes, com alguma tolerância, ao resto da Internet. No entanto, a escala dessa informação é ainda proibitiva dada a dimensão atual da Internet. O número de blocos CIDR atribuídos [Smith, 2005] ultrapassa os 100.000 em março de 2005 e espera-se que haja mais prefixos de endereços distintos do que blocos CIDR. Descobrir, disseminar e armazenar as distâncias para a lista completa de (P^2) pares de distâncias prefixo-prefixo é certamente extremamente custoso, mesmo sabendo-se que $P \ll H$.

Claramente há a necessidade de se buscar uma maneira de reduzir o volume de informação requerida para a estimação de distância entre nós. Uma maneira alternativa seria manter a lista de distâncias entre cada Sistema Autônomo (SA) a todos os outros SAs. Os diferentes SAs são interligados pelo protocolo de roteamento inter-domínio BGP. O BGP também pode mapear blocos de endereços IP aos respectivos AS quando estes os anunciam. Isto reduz o tamanho do conjunto de distâncias para $A^2 + P'$, onde A ($A \ll P$) é o número de SAs e P' é o número de blocos de endereço IP anunciados ao BGP. Esses blocos de endereço anunciados pelos SAs através do BGP não correspondem aos prefixos de endereço definidos anteriormente, mas são da mesma ordem de magnitude desses. Embora ainda seja uma lista grande de distâncias, a manutenção desta lista passa a ser viável. Há cerca de 19.000 SAs e uma média de aproximadamente 8 blocos de endereço anunciados por SA em março de 2005 [Smith, 2005]. No entanto, aproximar a distância entre dois nós pela distância entre os seus respectivos SAs pode ser questionável. Muitos SAs possuem cobertura praticamente global e múltiplos SAs podem cobrir a mesma região geográfica. Como conseqüência, pode ser freqüente o caso em que alguns nós estejam muito próximos tanto em termos geográficos quanto em atraso, mas pertençam a SAs diferentes. Da mesma forma, pode ser comum outros nós estarem bastante distantes e pertencerem ao mesmo SA.

A arquitetura IDMaps [Francis et al., 2001] propõe ainda uma outra forma de representar a informação de distância que inclui o agrupamento de alguns prefixos de endereços, porém em uma unidade menor do que a dos SAs. A arquitetura estabelece alguns sistemas, chamados de *tracers*, a serem distribuídos pela Internet de forma que todo conjunto formado por um prefixo de endereços esteja relativamente próximo a um

ou mais *tracers*. As distâncias entre os *tracers* são medidas, assim como a distância entre esses conjuntos de prefixos de endereços e o *tracer* mais próximo (ver Figura 9(c)). A distância entre dois prefixos de endereços quaisquer pode então ser calculada como a soma das distâncias entre cada prefixo ao *tracer* mais próximo, acrescida da distância entre os dois *tracers*. A qualidade do resultado em distância depende do número de *tracers* adotado e de onde estes estão localizados. Existe, portanto, um compromisso entre o aumento da qualidade do resultado ao custo de mais medições. Esta abordagem reduz o tamanho do conjunto de distâncias para $B^2 + P$, onde B é o número de *tracers*. O número de prefixos de endereços é da ordem de 150.000 em março de 2005 [Smith, 2005]. Portanto, se o número B de *tracers* adotado se limitar a algumas centenas, o volume total de distâncias a ser gerenciado mantém-se viável. O sistema funciona em uma arquitetura cliente-servidor onde servidores HOPS (*HOst Proximity Service*) fornecem a distância entre dois nós arbitrários usando as medições da arquitetura IDMaps.

Para efeitos de avaliação, a arquitetura IDMaps foi utilizada para o problema de seleção do servidor espelho (*mirror*) mais próximo em um cenário de simulação com topologias geradas sinteticamente. O desempenho do sistema melhora significativamente com o uso de IDMaps quando comparado com uma seleção aleatória. Esse resultado é obtido usando-se heurísticas para a escolha da localização dos *tracers* que não exigem conhecimento total da topologia da rede. Também é mostrado que o número de *tracers* necessário para a obtenção de resultados satisfatórios é relativamente pequeno, pois o uso de somente 0.2% dos nós como *tracers* fornece uma resposta correta em 90% dos casos.

6.2. Abordagens baseadas em sistemas de coordenadas

Como uma alternativa à arquitetura cliente-servidor proposta por IDMaps, outras propostas surgiram para a predição da proximidade de rede com base em um modelo de operação P2P. O modelo P2P possui um maior potencial de escalabilidade quando comparado ao modelo cliente-servidor. Gargalos de desempenho são evitados pela ausência de servidores que podem estar muito distantes. Além do mais, esse modelo é consistente com as aplicações P2P, tais como compartilhamento de arquivos, redes de distribuição de conteúdo e redes de multicast aplicativo que podem se beneficiar de forma significativa de informações sobre a proximidade entre nós na rede.

A proposta P2P pioneira para a estimação de distância de rede entre dois nós na Internet foi o sistema GNP (*Global Network Positioning*) [Ng e Zhang, 2002]. O GNP propõe uma abordagem baseada em sistemas de coordenadas para a predição da distância de redes entre dois nós arbitrários usando uma arquitetura P2P. A idéia fundamental do GNP é manter nos nós participantes coordenadas que representem as suas posições relativas na Internet de tal maneira que as distâncias de rede possam ser preditas avaliando uma função de distância sobre as coordenadas dos nós.

A primeira etapa da arquitetura GNP usa um pequeno conjunto de nós de referência distribuídos, chamados de *landmarks*, para fornecer o conjunto de coordenadas de referência para orientar outros nós no espaço abstrato resultante. Esses nós de referência medem a distância entre si periodicamente para corrigir suas coordenadas se necessário. As distâncias podem ser medidas como o RTT mínimo de diversas medições usando a ferramenta `ping`. Então esses nós transformam a distância medida entre si em coordenadas no espaço abstrato. Dessa forma modela-se a Internet em um espaço geométrico abstrato, como ilustrado na Figura 10 para um espaço abstrato hipotético em três dimensões. Vale destacar que a distância entre os nós nesse espaço abstrato no contexto da proposta GNP representa alguma métrica de rede como atraso e não a distância física entre as localizações geográficas reais dos nós. Deve-se também notar que podem haver infinitas soluções para as coordenadas nos nós de referência no espaço abstrato,

pois qualquer rotação ou translação de um conjunto solução de coordenadas preserva a distância relativa entre os nós de referência. No entanto, para os objetivos do GNP, somente a distância relativa entre os nós de referência é importante, então qualquer solução adotada é suficiente.

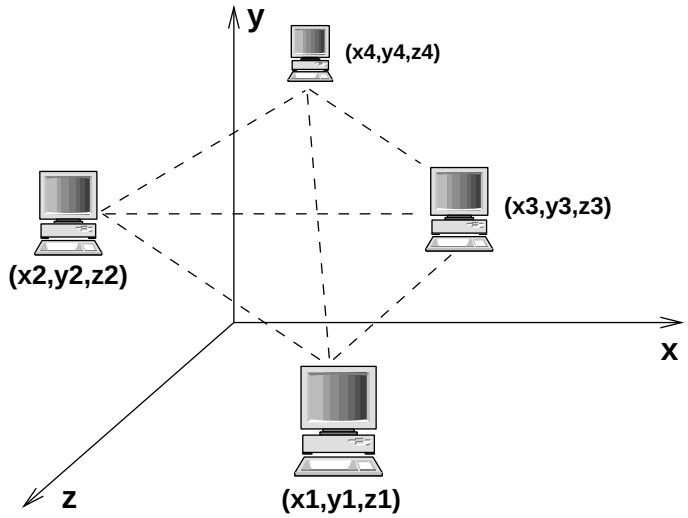


Figura 10: Internet modelada como um espaço geométrico abstrato.

Uma vez calculadas as coordenadas de cada nó de referência, estas são disseminadas junto com o identificador do espaço abstrato utilizado e a função de distância correspondente a qualquer outro nó que pretenda participar do sistema. O mecanismo e o protocolo de disseminação dessas informações não é especificado na proposta do GNP.

Na segunda etapa da arquitetura GNP, nós comuns podem participar do sistema. Usando as coordenadas dos nós de referência no espaço abstrato, cada nó comum pode determinar suas próprias coordenadas. Para tanto, o nó comum mede seu RTT até os nós de referência usando por exemplo a ferramenta `ping` e considera o RTT mínimo de diversas medições para cada caminho como a distância. Nesta etapa, os nós de referência são completamente passivos e somente respondem às mensagens ICMP vindas do nó comum que pretende integrar o sistema. Usando as suas distâncias aos nós de referência, o nó comum pode então calcular suas próprias coordenadas que minimizem o erro entre as distâncias medidas e as calculadas no espaço abstrato de coordenadas. Esse procedimento é ilustrado na Figura 11.

Em [Ng e Zhang, 2002], a proposta GNP é comparada diretamente à arquitetura IDMaps. Os resultados demonstram que o GNP supera significativamente o IDMaps em desempenho e robustez. O ganho em desempenho é especialmente significativo na predição de distâncias curtas. Para explicar a razão dessa diferença em desempenho no caso de distâncias curtas, consideremos a situação da Figura 12, onde X e Y são nós de referência no GNP ou *tracers* no IDMaps, e A e B são dois nós finais que são próximos um do outro, mas distantes de X e Y. IDMaps fornece uma predição de distância pessimista como $(A, X) + (B, Y) + (X, Y)$ ou se ambos os nós A e B usarem o *tracer* X: $(A, X) + (B, X)$. GNP, por sua vez, é capaz de estimar diretamente a distância entre A e B usando o espaço abstrato de coordenadas. Dessa forma, GNP possui um melhor desempenho pois ele explora as relações entre as posições dos nós de referência e as dos nós finais ao invés de depender do posicionamento dos *tracers* na topologia da rede.

Duas propostas semelhantes para melhorar o desempenho do GNP na acurácia da conversão das distâncias medidas em coordenadas de um espaço abstrato de menor dimensão usando a Análise de Componentes Principais (*Principal Components*

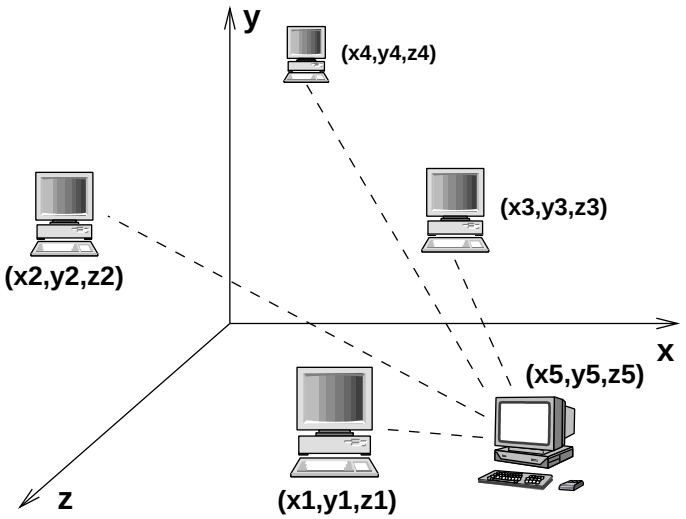


Figura 11: Estabelecimento das coordenadas de um nó qualquer.

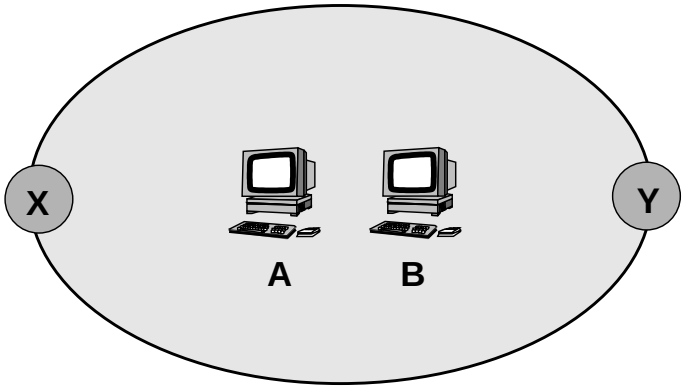


Figura 12: Predição de distâncias curtas.

Analysis – PCA) [Bryant e Yarnold, 1998] são investigadas nos sistemas ICS (*Internet Coordinate System*) [Lim et al., 2003] e *Virtual Landmarks* [Tang e Crovella, 2003]. Nos últimos anos, diversas outras propostas se inspiraram nas idéias básicas inicialmente introduzidas pela arquitetura GNP, como o uso do modelo P2P e de sistemas de coordenadas para a estimação de proximidade entre nós na rede. Essas propostas incluem os sistemas *King* [Gummadi et al., 2002], *Lighthouses* [Pias et al., 2003], *Big-Bang Simulation* (BBS) [Shavitt e Tankel, 2003], *Practical Internet Coordinates* (PIC) [Costa et al., 2004] e Vivaldi [Dabek et al., 2004].

7. Outras metodologias baseadas em medições

Nesta seção, nós apresentamos algumas metodologias recentes que se baseiam em medições e abrem novas áreas de pesquisa.

7.1. Projeção de tráfego

O planejamento das necessidades futuras em termos de capacidade e dimensionamento em uma rede IP é uma tarefa desafiadora. Em geral, esse planejamento se apoia na experiência e na intuição dos operadores de rede. Usando dados como o número projetado de clientes em diferentes localidades e em suposições sobre o tráfego gerado por estes, os operadores estimam o efeito que esses clientes adicionais possam ter na carga do domínio de rede como um todo. A escolha dos pontos onde ocorrerá um aumento da capacidade dos enlaces é baseada na experiência e no estado atual da rede. Por exemplo,

enlaces que no momento transportam grandes volumes de tráfego são mais suscetíveis de terem sua capacidade incrementada primeiro.

Em [Papagiannaki et al., 2003], os autores introduzem uma metodologia para predizer quando e onde se faz necessário a adição ou a atualização de um enlace em um rede IP de *backbone*. Trabalhos anteriores [Sang e Li, 2000] tipicamente focavam na projeção do tráfego na Internet sobre pequenas escalas de tempo, como segundos ou minutos, que são relevantes para a alocação dinâmica de recursos. Por sua vez, ao visar o planejamento das necessidades futuras em termos de capacidade e dimensionamento, o método apresentado em [Papagiannaki et al., 2003] projeta o volume de tráfego na rede meses no futuro. Essa metodologia baseia-se em medições históricas da rede recolhidas com SNMP. A intuição por trás dessa abordagem consiste no uso de métodos matemáticos para processar as informações históricas e extrair tendências na evolução do tráfego em diferentes escalas de tempo. Nessa abordagem, é utilizada a análise de multiresolução por *wavelets* para isolar as tendências de longo-termo e analisar a variabilidade em múltiplas escalas de tempo. Os resultados revelam que o maior volume de variabilidade no tráfego ocorre em flutuações na escala de tempo de 12 horas. Essa metodologia combinada com medições reais do tráfego de *backbone* leva a diferentes modelos de projeção para diferentes partes da rede. Os resultados indicam que pares de POPs distintos apresentam diferentes taxas de crescimento e sofrem diferentes tipos de flutuações. Essa constatação reforça a importância do estabelecimento de uma metodologia capaz de derivar modelos específicos ao invés de desenvolver-se um modelo único para o tráfego agregado entre POPs. No entanto, essa abordagem requer uma coletânea de dados de medição na rede sobre longos períodos de tempo.

Métodos para a projeção de tráfego são avaliados em [Silva et al., 2004]. Essa avaliação demonstra os problemas encontrados nos métodos analisados, tais como a necessidade de uma grande quantidade de amostras e a garantia de periodicidade. Esses problemas motivaram uma nova proposta com complexidade menor e poucas restrições quanto à coleta de dados. O novo método prevê o volume de tráfego futuro em um enlace baseando-se no histórico do crescimento dos volumes máximo e mínimo do tráfego medidos no enlace de interesse. A idéia básica é inferir uma curva que represente o crescimento médio entre os volumes máximo e mínimo e adotar esta curva para estimar o tráfego futuro. Essa proposta foi comparada com as anteriores com dados da Rede Nacional de Ensino e Pesquisa (RNP). Os resultados obtidos demonstram que o método proposto supera os demais na qualidade da projeção e possui um menor grau de complexidade na sua parametrização, fornecendo uma projeção de tráfego em tempo futuro da ordem de meses.

Modelos para predizer o comportamento futuro com base em medições também podem ser adotados para outros fins, além de planejamento de capacidade. Em [Bremler-Barr et al., 2003], os autores investigam metodologias para prever degradações no serviço oferecido a usuários finais. Esse trabalho se concentra na previsão da ocorrência dessas degradações de serviço do ponto de vista dos usuários finais, onde as degradações são eventos que podem afetar de forma significativa a qualidade percebida pelo usuário, especialmente em aplicações interativas.

7.2. Identificação e caracterização de aplicações

Medições são comumente usadas para identificar e classificar o tráfego presente na rede em termos das aplicações que o geram. O conhecimento da composição do tráfego carregado por uma rede IP é de interesse de projetistas e administradores de rede para efeitos de planejamento e aprovisionamento de recursos. Essa monitoração permite aos operadores detectar o surgimento e o crescimento no uso de novas aplicações

que podem mobilizar uma grande população de usuários em um espaço de tempo relativamente curto dado o dinamismo permitido pela Internet atual. Por exemplo, os últimos anos testemunharam um crescimento acelerado na utilização de sistemas P2P para o compartilhamento de arquivos, tais como Napster, Gnutella, e-Donkey, Kazaa, entre outros [Ripeanu et al., 2002, Fraleigh et al., 2003]. Essas aplicações P2P alteraram significativamente a composição de tráfego na rede. Além do tráfego de aplicações P2P, contribuem crescentemente para a diversidade na composição do tráfego aplicações de *streaming* de áudio e vídeo, voz sobre IP ou videoconferência multimídia.

O método clássico para a identificação das aplicações de rede através de monitoração de tráfego utiliza os números de porta bem conhecidos (*well-known port numbers*). No entanto, esses números de porta não mais indicam de forma confiável a aplicação utilizada. Diversas novas aplicações na Internet não usam portas bem conhecidas ou usam outros protocolos como `http` para passar através de *firewalls*. Como conseqüência, o simples exame do número de porta no cabeçalho do pacote pode levar a uma classificação errônea. Uma alternativa recente é o exame do conteúdo dos pacotes para identificar as aplicações em uso [Moore e Papagiannaki, 2005]. Os resultados desse estudo mostram que o exame das portas bem conhecidas pode identificar corretamente cerca de 70% do tráfego e o restante necessitaria da análise baseada em conteúdo para uma correta classificação. O trabalho futuro dos autores prevê a investigação de métodos para a sua implementação em tempo real. No entanto, alguns casos, como o surgimento de novas aplicações cujo comportamento não é de conhecimento comum, podem exigir algum nível de intervenção manual.

Uma aplicação P2P que vem encontrando crescente sucesso recentemente é o Skype [Zennström e Friis, 2003], uma aplicação de voz sobre IP. Essa aplicação permite comunicação gratuita entre seus usuários e comunicação com telefones convencionais de forma global a baixo custo. Com esses atrativos, desde o seu lançamento em setembro de 2003, o Skype atingiu a marca de 1 milhão de usuários conectados simultaneamente em outubro de 2004 e em março de 2005 ultrapassa a marca de 2 milhões de usuário simultâneos. Dado o crescimento vertiginoso em sua utilização, pesquisadores e operadores começam a debater se não estão diante de uma aplicação que poderá responder por um volume significativo do tráfego em um futuro próximo. Mesmo estando disponível gratuitamente para instalação, o modo de funcionamento do Skype não é de domínio público. Monitorando clientes Skype em laboratório, um trabalho recente [Baset e Schulzrinne, 2004] busca inferir a organização da estrutura e o funcionamento do protocolo de comunicação do Skype. O estudo contempla os mecanismos usados pelo Skype para *login*, a travessia transparente de dispositivos de NAT (*Network Address Translation*) e *firewalls*, estabelecimento de chamadas, transferência de dados de voz, codificação, entre outros aspectos. Essa análise é possível devido à medição e à investigação cuidadosas do tráfego de rede gerado por clientes Skype.

7.3. Inferência de topologia

O conhecimento da topologia da Internet é importante para a pesquisa e o desenvolvimento de novas abordagens para a rede. No entanto, em geral, as topologias reais que compõem a Internet não são de domínio público, pois os provedores de acesso e administradores de Sistemas Autônomos (SA) vêem as suas topologias como uma informação confidencial. Alguns provedores de acesso publicam versões simplificadas de suas topologias, mas que não revelam detalhes ou estão desatualizadas. Esse conhecimento é crucial para a validação de geradores de topologias para a simulação de redes, tais como o GT-ITM [Zegura et al., 1999] e o Brite [Medina et al., 2001], o que levanta questões quanto a representatividade dos atuais geradores de topologia [Tangmunarunkit et al., 2002].

A topologia da Internet pode ser representada ao nível de SAs e de roteadores. Para inferir a topologia da Internet no nível de SAs, usa-se dados das tabelas de roteamento BGP [Chang et al., 2004, Dimitropoulos et al., 2005]. Ao utilizar múltiplos pontos de observação, pode-se estimar as relações entre os diferentes SAs [Subramanian et al., 2002]. Os irmãos Faloutsos [Faloutsos et al., 1999] utilizaram dados como esses para propor que a topologia da Internet pode ser representada por leis de potência. Entretanto, a qualidade das informações nas quais esse modelo se baseou é alvo de debate recente. Em [Lakhina et al., 2003b] mostra-se por simulações que em uma rede cuja distribuição do grau de conectividade não segue uma lei de potência, os resultados de medições por `traceroute` nesta rede conduzidas a partir de um pequeno número de pontos de medição tende a indicar um subgrafo que segue uma lei de potência. Em [Amini et al., 2004], é investigado o efeito da qualidade dos dados fornecidos por `traceroutes` na inferência da topologia da Internet ao nível de SAs.

A inferência da topologia da Internet no nível de roteamento exige mais informações. A ferramenta `traceroute` é adotada para a identificação dos roteadores em vários caminhos a partir de diversos pontos de medição. A informação desses diversos pontos é então combinada para inferir a topologia ao nível do roteamento. Esse procedimento é a base de projetos como o Skitter [Claffy et al., 1999] e o Mercator [Govindan e Tangmunarunkit, 2000], que buscam mapear a Internet como um todo através de um grande volume de medições.

Uma alternativa para a inferência de topologia ao nível de roteamento foi proposta através da ferramenta Rocketfuel [Spring et al., 2004]. O Rocketfuel propõe mapear a topologia de provedores de serviço individuais, ao invés de coletar informação ao nível de roteadores de toda a Internet. Dessa forma, essa ferramenta é capaz de gerar topologias com mais qualidade e usando menos medições. Em [Barford et al., 2001], os autores também pretendem reduzir o número de medições, mas buscando minimizar o número de pontos de medição utilizados ao buscar um melhor posicionamento destes. Duas técnicas são propostas pelo Rocketfuel para escolher as medições que contenham a informação mais valiosa para a inferência de topologia. A primeira técnica usa a informação de roteamento BGP para selecionar somente os `traceroutes` que devem transitar pelo provedor de serviço de interesse. A segunda técnica suprime `traceroutes` que tenham grande chance de gerar resultados sobre caminhos que já tenham sido atravessados. A idéia principal é escolher `traceroutes` que forneçam a máxima informação para a inferência da topologia, enquanto omite-se aqueles `traceroutes` que provavelmente iriam fornecer informação redundante. O uso combinado dessas duas técnicas permite ao Rocketfuel reduzir o número de medições necessárias para mapear a topologia de um provedor de serviço em três ordens de magnitude quando comparado aos métodos de força bruta anteriores.

Recentemente, uma nova abordagem chamada Doubletree [Donnet et al., 2005] foi proposta para reduzir o número de medições para a inferência de topologia ao nível de roteamento. Essa abordagem evita que esforços de medição sejam duplicados pelos pontos de monitoração ou monitores. Esforços duplicados podem ser de dois tipos: medições feitas por um monitor individual que replica o seu próprio trabalho e medições realizadas por múltiplos monitores que replicam o trabalho uns dos outros. Esses esforços duplicados são chamados de redundância intra-monitor e inter-monitor. Os resultados mostram que somente 10,9% das sondas de medição descobrem uma nova interface e, portanto, o restante gera algum nível de informação redundante. Também é observado que uma grande parte das interfaces é visitada por todos os monitores. Ao combater a redundância nas medições, Doubletree pode reduzir significativamente o impacto em roteadores e alvos de medição, mantendo uma alta taxa de descoberta de enlaces e interfaces.

8. Plataformas de medições e experimentação

8.1. NIMI

A plataforma NIMI (*National Internet Measurement Infrastructure*) [Paxson et al., 1998] foi a primeira plataforma genérica de medições em larga escala disponível para a comunidade de pesquisa. A base da estrutura NIMI é inspirada no *Network Probe Daemon* (NPD), desenvolvido por Paxson [Paxson, 1997]. O acesso à plataforma é feito sob solicitação por questões de segurança. A plataforma é genérica, pois não determina as ferramentas de medição disponíveis, sendo permitida a integração de novas ferramentas para medições específicas quando necessário.

Existem quatro componentes na arquitetura NIMI:

- Sonda NIMI – é o ponto de medição que realiza as medições solicitadas;
- CPOC (*Configuration Point Of Contact*) – é o sistema que autoriza usuários a realizar medições;
- MC (*Measurement Client*) – é o cliente das medições;
- DAC (*Data Analysis Client*) – é o cliente que recebe os dados das medições.

O funcionamento da arquitetura NIMI é ilustrado na Figura 13. Suponhamos que um cliente da arquitetura NIMI queira realizar um experimento em que é necessário coletar dados referentes a quatro alvos usando a ferramenta `traceroute`. O cliente (MC) envia a solicitação de execução de um `traceroute` em direção a quatro alvos também determinado no pedido (passo 1). Ao receber o pedido do MC, a sonda NIMI consulta o CPOC para que este conceda uma autorização ao cliente que se utiliza do MC (passo 2). Para a autorização são utilizadas chaves geradas com o programa `gen_keys`. O MC passa a chave do cliente junto ao pedido, a sonda NIMI repassa ao CPOC e este autoriza a sonda NIMI a realizar o experimento solicitado, se for o caso. No passo 3, a sonda NIMI realiza os `traceroutes` até os alvos determinados. O resultado de um experimento pequeno pode ser enviado de volta ao MC diretamente, mas para isto o MC precisa manter uma conexão aberta com a sonda NIMI. Para experimentos grandes, pode-se agendar medições junto à sonda NIMI no pedido MC e assim o cliente pode desconectar-se. Conforme os experimentos agendados vão sendo concluídos pela sonda NIMI, esta envia os resultados de cada experimento ao DAC designado pelo MC no pedido original, como ilustrado no passo 4 da Figura 13.

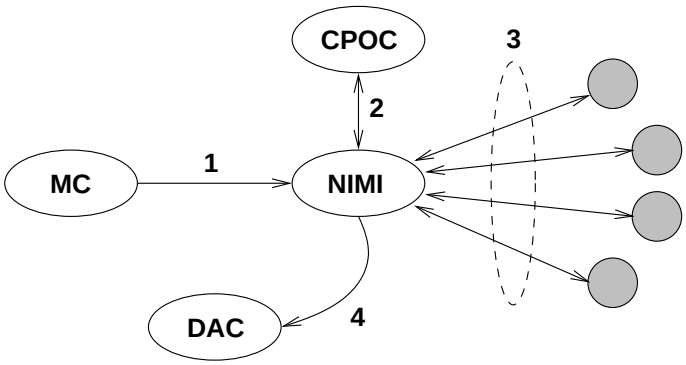


Figura 13: Ilustração do funcionamento da arquitetura NIMI.

A arquitetura NIMI permitiu a implementação progressiva de dezenas de sondas e a realização de experimentos de grande porte. O gerenciamento das sondas é local, mas o controle de acesso às sondas NIMI remotas para execução de medições é centralizado por questões de segurança. Um controle centralizado e não-automatizado é realizado quando um usuário requer a inclusão em alguma sonda NIMI de uma ferramenta de medição que ainda não está disponível.

8.2. PlanetLab

Há alguns anos, a Internet testemunha o surgimento de uma nova classe de serviços de rede que são distribuídos geograficamente, tais como redes de distribuição de conteúdo e aplicações P2P. A plataforma PlanetLab [Peterson et al., 2002] nasceu de um esforço cooperativo das comunidades de pesquisa em sistemas distribuídos e em redes para criar uma maneira eficiente de projetar, avaliar e implementar serviços distribuídos. O projeto PlanetLab fornece uma rede de sobrecamada (*overlay*) que se presta tanto a uma plataforma de pesquisa quanto a uma plataforma de experimentação para o desenvolvimento de protótipos. Ao contrário das plataformas de pesquisa anteriores, o PlanetLab com essa dualidade de uso pretende atender ambos os pesquisadores que pretendem desenvolver novos serviços de rede e os clientes que querem usar tais serviços.

A arquitetura PlanetLab foi concebida com quatro princípios de projeto: (i) execução das aplicações continuamente; (ii) controle distribuído dos recursos; (iii) gerenciamento repartido; e (iv) interfaces centradas nas aplicações. A execução contínua das aplicações é possível pelo fato de que a cada aplicação no PlanetLab é atribuída uma fatia (*slice*) da rede sobrecamada para a execução da aplicação. Cabe a um mecanismo distribuído de virtualização requerer a cada nó a multiplexação de serviços em competição. O controle distribuído de recursos é essencial para permitir a convivência de dois tipos de usuários, pesquisadores desenvolvendo serviços e clientes os consumindo. O gerenciamento repartido envolve a divisão do gerenciamento da rede de sobrecamada em serviços de gerenciamento independentes, cada um sendo executado em sua própria fatia da rede de sobrecamada. Alguns serviços são primordiais e devem estar sempre presentes, como o gerenciamento das contas de usuário. No entanto, ao ser executado em fatias, os serviços podem ser aprimorados independentemente e novos serviços podem ser acrescentados. Assim, a estrutura de gerenciamento favorece a evolução e a inovação. Finalmente, a plataforma oferece uma interface simples que incentiva o desenvolvimento, a avaliação e a implementação de novas aplicações.

Devido aos seus princípios inovadores, a rede PlanetLab alcança grande sucesso na comunidade de pesquisa nos dias atuais. Em março de 2005, a rede conta com mais de 500 nós distribuídos ao redor do mundo.

9. Geolocalização: um exemplo de serviço baseado em medições

O desenvolvimento de uma maneira eficiente de inferir a localização geográfica de nós na Internet abre perspectivas para novas aplicações conscientes da localização dos usuários [Zook, 2001, Lakhina et al., 2003a]. A disponibilidade de informação de localização permite o desenvolvimento de aplicações conscientes de localização que podem ser úteis tanto aos usuários corporativos quanto aos particulares. Por exemplo, essas aplicações podem incluir:

- Publicidade direcionada em páginas *web* – usuários podem ter diferentes preferências regionais. Ser capaz de adaptar regionalmente produtos, serviços, estratégias de propaganda e conteúdo, provê um diferencial de atratividade;
- Distribuição restrita de conteúdo – seguindo alguma regulamentação regional, um serviço de geolocalização fornece subsídios para a definição de quais usuários estão autorizados, ou não, a receber um determinado conteúdo;
- Verificação de segurança baseada em localização – se localizações autorizadas são conhecidas, uma transação de comércio eletrônico que for requisitada de algum outro lugar pode gerar avisos sobre um comportamento atípico ou não autorizado de um cliente.

No entanto, a inferência da localização geográfica de nós na Internet a partir de seus endereços IP constitui um problema desafiador, pois não há uma relação direta entre o endereço IP de um nó e a sua localização geográfica. Esta seção descreve o desenvolvimento de serviços de localização na Internet baseados em medições de atraso para a inferência da localização geográfica de nós na Internet. Vale ressaltar que diferentemente das propostas de estimação de proximidade em redes discutidas na Seção 6 onde as distâncias eram medidas em atraso, as abordagens apresentadas nesta seção buscam localizar geograficamente um nó na Internet, então as distâncias se referem à distância física entre os nós. Nesta seção, apresentamos duas abordagens, uma discreta e outra contínua, para o uso das medições coletadas com o propósito de estimar a localização geográfica de um nó na Internet.

9.1. Geolocalização discreta

Em um sistema discreto, a localização dos nós é inferida pela comparação de padrões de atraso de nós de referência, de localização conhecida, com o padrão de atraso do nó alvo a ser localizado, como na técnica GeoPing [Padmanabhan e Subramanian, 2001]. Nesse sistema, adota-se uma abordagem empírica baseada na observação que nós que compartilham atrasos semelhantes em relação a outros nós fixos na rede tendem a estar próximos uns dos outros geograficamente.

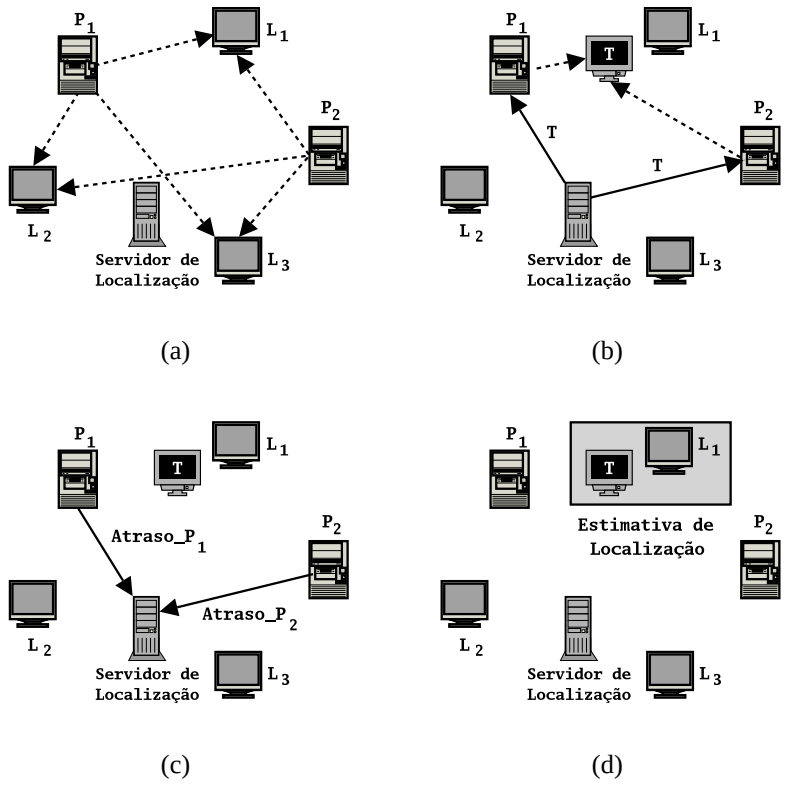


Figura 14: Estimando a localização de um nó alvo T em um sistema discreto.

Pode-se formalizar o sistema discreto para inferir a localização geográfica de um nó na Internet a partir de medições de atraso da seguinte maneira [Ziviani et al., 2005a]. Consideremos um conjunto $\mathcal{L} = \{L_1, L_2, \ldots, L_K\}$ de K nós de referência, também chamados de *landmarks*. A localização geográfica desses nós de referência é conhecida. Consideremos também um conjunto $\mathcal{P} = \{P_1, P_2, \ldots, P_N\}$ de N pontos de medição. A Figura 14 ilustra os passos do sistema discreto para obter uma estimativa de localização

de um nó alvo T usando um conjunto de nós de referência (L_1, L_2 e L_3) e de pontos de medição (P_1 e P_2). Linhas tracejadas representam as medições realizadas pelos pontos de medição enquanto linhas sólidas indicam a troca de informações. Os pontos de medição periodicamente determinam o atraso de rede, que na realidade é obtido como o mínimo de várias medições, para cada nó de referência (Figura 14(a)). Portanto, cada ponto de medição P_x, $1 \leq x \leq N$ mantém um vetor de atraso $\mathbf{d}_x = (d_{1x}, d_{2x}, \ldots, d_{Kx})$, onde d_{ix} é o atraso entre o ponto de medição P_x e o nó de referência $L_i \in \mathcal{L}$. Suponhamos que se queira determinar a localização geográfica de um certo nó alvo T. Um servidor de localização que conhece o conjunto de nós de referência $\mathcal{L}$ e o conjunto de pontos de medição $\mathcal{P}$ é então contactado. O servidor de localização solicita aos N pontos de medição a realização de medições de atraso até o nó T (Figura 14(b)). Cada ponto de medição P_x, $1 \leq x \leq N$ retorna um vetor de atraso $\mathbf{d}'_x = (d_{1x}, d_{2x}, \ldots, d_{Kx}, d_{Tx})$, ou seja, o vetor de atraso $\mathbf{d}_x$ acrescido da medição de atraso recém-realizada até o nó T (Figura 14(c)). Depois de receber os vetores de atraso dos N pontos de medição, o servidor de localização pode construir a matriz de atraso $\mathbf{D}$ de dimensões $(K+1) \times N$:

$$\mathbf{D} = \begin{bmatrix} d_{11} & d_{12} & \ldots & d_{1N} \\ d_{21} & d_{22} & \ldots & d_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ d_{K1} & d_{K2} & \ldots & d_{KN} \\ d_{T1} & d_{T2} & \ldots & d_{TN} \end{bmatrix} \tag{6}$$

Os vetores de atraso coletados pelo servidor de localização dos pontos de medição correspondem às colunas da matriz de atraso $\mathbf{D}$. O servidor de localização então compara as linhas da matriz de atraso $\mathbf{D}$ para estimar a localização do nó T. Para inferir a localização do nó T, o nó de referência L que apresenta o padrão de atraso mais semelhante ao padrão de atraso do nó T é determinado. A localização correspondente ao nó de referência L é a estimativa de localização do nó T (Figura 14(d)). A matriz de atraso $\mathbf{D}$ combinada com o conhecimento da localização dos nós de referência do conjunto $\mathcal{L}$ compõe um mapeamento que armazena a relação entre atraso na rede e localização geográfica. Resultados práticos de medições para a localização geográfica de nós na Internet usando a plataforma NIMI (ver Seção 8.1) são apresentados em [Ziviani et al., 2004].

Em [Ziviani et al., 2005a] são propostas técnicas para aprimorar a localização geográfica de nós na Internet através do sistema discreto. Investiga-se primeiro a correlação encontrada na rede entre a distância geográfica e o atraso na rede. Essa correlação é de fraca a moderada se considerada globalmente, porém é observado que ela torna-se mais forte em regiões de rica conectividade. O termo rico, ou pobre, para conectividade representa a variedade de conectividade e as opções de trânsito encontradas em determinadas regiões tanto ao nível de roteadores quanto de sistemas autônomos. De um ambiente com conectividade rica espera-se mais opções de rotas que possam se aproximar do caminho geográfico direto entre uma fonte e o destino. Dois pontos-chave que influenciam a precisão do sistema discreto são identificados. A precisão depende basicamente do posicionamento dos nós de referência e dos pontos de medição, e da eficiência na avaliação da similaridade entre os padrões de atraso. Assim, em [Ziviani et al., 2005a], busca-se aumentar o desempenho do sistema discreto de estimação da localização geográfica de nós na Internet de duas maneiras: (i) posicionando estrategicamente nós de referência e pontos de medição; (ii) selecionando modelos para melhor avaliar a similaridade entre os padrões de atraso dos nós de referência e do nó alvo.

9.2. Geolocalização contínua

Trabalhos anteriores [Padmanabhan e Subramanian, 2001, Ziviani et al., 2005a] usam a posição de nós de referência, que possuem localização geográfica bem conhecida, como possíveis estimativas de localização para o nó-alvo. Isto leva a um espaço discreto de respostas, ou seja, o número de respostas equivale ao número de nós de referência, o que pode levar a resultados imprecisos porque o nó de referência mais próximo pode estar afastado do alvo.

Para superar a limitação de um espaço discreto de respostas, é proposta a abordagem CBG (*Constraint-Based Geolocation*) [Gueye et al., 2004, Ziviani et al., 2005b] para inferir a localização geográfica de nós na Internet usando multilateração. Multilateração refere-se ao processo de estimar uma posição usando um número suficiente de distâncias a alguns pontos fixos. Como resultado, a multilateração estabelece um espaço contínuo de respostas no lugar de um espaço discreto. Nós utilizamos um conjunto de nós de referência para estimar a localização de outros nós na Internet. A idéia fundamental é que dadas as distâncias geográficas até um determinado nó-alvo a partir dos nós de referência, uma estimativa de localização do nó-alvo seria viável usando multilateração, assim como faz o sistema GPS.

Um elemento-chave de CBG é a sua habilidade em transformar de forma acurada medições de atraso em restrições de distância. O ponto de partida consiste na constatação de que a informação digitalizada trafega por cabos de fibra ótica a quase exatamente 2/3 da velocidade da luz no vácuo [Bovy et al., 2002, Percacci e Vespignani, 2003]. Isso significa que qualquer medição de atraso fornece imediatamente um *limite superior* na distância entre os pontos finais. Esse limite superior é a medição de atraso multiplicada pela velocidade da luz na fibra. Do ponto de vista de um par de pontos finais, nós argumentamos que há algum atraso mínimo teórico para a transmissão de pacotes que é ditado pela distância geográfica entre eles. Portanto, o atraso real medido entre estes pontos envolve somente uma distorção *aditiva*.

Contudo, se CBG usasse as medições de atraso para inferir diretamente as restrições de distância, a proposta não seria muito acurada. Para resultados acurados, é importante estimar e remover o tanto quanto for possível da distorção aditiva. CBG realiza essa tarefa auto-calibrando as medições de atraso tomadas de cada ponto de medição. Após a auto-calibração, CBG é capaz de transformar de forma mais acurada um conjunto de medições de atraso até um alvo em restrições de distância. CBG então usa multilateração com essas restrições de distância para estabelecer uma região geográfica que contenha o nó-alvo. Determinada essa região, uma estimativa razoável da localização do nó-alvo é o centróide desta região, o que é usado por CBG como estimativa pontual da posição do alvo. Deve-se ressaltar que, em contraste com outras propostas, CBG associa uma região de confiabilidade para cada estimativa de localização. Isso permite a uma aplicação consciente de localização avaliar se a qualidade da estimativa fornecida é suficiente às suas necessidades.

A Figura 15 ilustra o processo de multilateração usado por CBG com um conjunto de nós de referência $\mathcal{L} = \{L_1, L_2, L_3\}$ na presença de alguma distorção aditiva de distância devido a medições imprecisas. Cada nó de referência L_i deve inferir a sua restrição geográfica de distância ao nó-alvo τ, cuja localização é desconhecida. No entanto, a restrição de distância geográfica estimada é na realidade determinada por $\hat{g}_{i\tau} = g_{i\tau} + \gamma_{i\tau}$, ou seja, a distância geográfica real $g_{i\tau}$ acrescida de uma distorção aditiva de distância geográfica representada por $\gamma_{i\tau}$. Essa distorção de distância puramente aditiva $\gamma_{i\tau}$ resulta da presença eventual de alguma distorção aditiva de atraso. Como uma conseqüência da existência da distorção aditiva de distância, a estimativa de localização

do nó-alvo τ encontra-se em alguma parte no interior da área acinzentada (ver Figura 15) que corresponde à interseção das restrições de distância geográfica super-estimadas dos nós de referência ao nó-alvo.

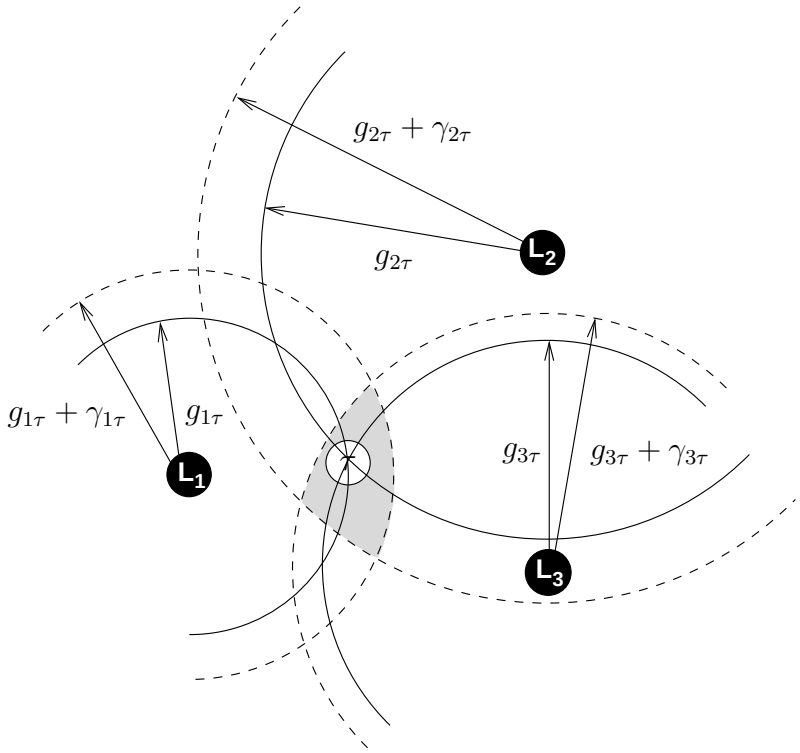
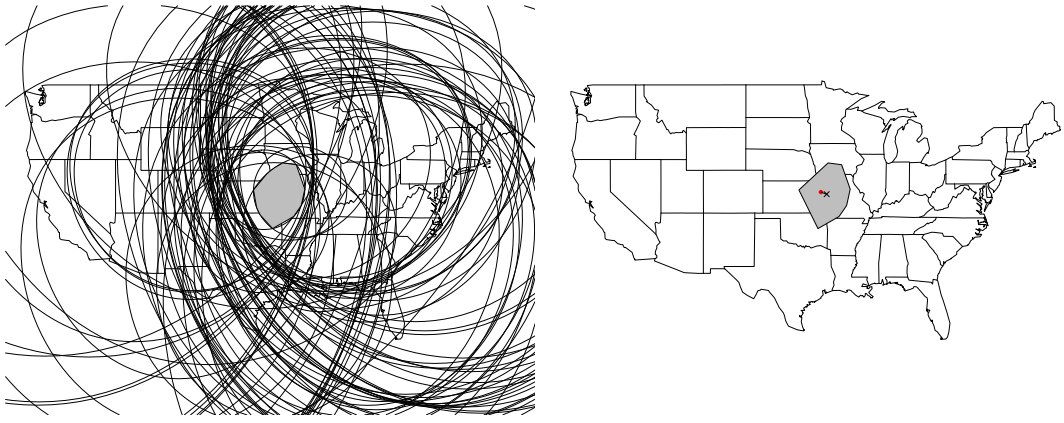


Figura 15: Multilateração com restrições de distância geográfica.

Para ilustrar a metodologia proposta por CBG, a Figura 16(a) mostra um exemplo de um conjunto de curvas fechadas extraído do estudo experimental que avalia a proposta [Gueye et al., 2004, Ziviani et al., 2005b]. A área da região de interseção $\mathcal{R}$, ou seja, a área acinzentada na Figura 16(a), indica a região de confiabilidade que CBG associa a essa estimativa de localização. Deve-se ressaltar que a maioria das regiões de confiabilidade observadas possuem uma área relativamente pequena, não visíveis em ilustrações semelhantes com todas as curvas fechadas presentes. Esse exemplo possui uma região de confiabilidade maior do que o tamanho típico, mas foi selecionada exatamente por possuir uma região suficientemente visível para ilustrar a metodologia de CBG.



(a) Determinaç ˜ao da regi˜ao de confi abilidade (b) Estimativa de localizaç ˜ao do n´o-alvo

Figura 16: Exemplo do procedimento de geolocalização usando CBG.

A proposta CBG é avaliada usando bases de dados com nós que estão geograficamente distribuídos pelos EUA e pela Europa Ocidental. Os resultados experimentais são promissores e mostram que CBG supera em desempenho outras técnicas de geolocalização. A mediana do erro em distância está abaixo de 25 km para os dados da Europa e abaixo de 100 km para os dados dos EUA. Para a maioria dos nós-alvo analisados, as regiões de confiabilidade obtidas permitem uma resolução em nível regional, ou seja, aproximadamente o tamanho de estados brasileiros relativamente pequenos em extensão territorial, como Rio de Janeiro ou Santa Catarina.

10. Considerações finais

Nesta seção, nós descrevemos diversos projetos envolvendo medições em andamento no Brasil e no mundo. A lista não pretende ser exaustiva, mas objetiva fornecer indicações para que os interessados estejam cientes de projetos importantes existentes na área. Finalmente, na Seção 10.3, discutimos algumas perspectivas para a área de Metrologia na Internet.

10.1. Projetos nacionais

10.1.1. GT-Medições

O GT-Medições [Monteiro, 2004] é o grupo de trabalho de medições em rede da RNP. O objetivo é a implementação de uma infra-estrutura de monitoração para a rede da RNP que contemple tanto medições ativas quanto passivas. As atividades propostas nesse grupo de trabalho incluem a implantação de um ambiente de acesso aos dados de medições espelhado no piPEs (ver Seção 10.2.8), a interoperabilidade com outros ambientes de medição e a criação do Observatório de Redes da RNP. Esse observatório objetiva centralizar em um único ambiente a visualização e o acompanhamento das características do *backbone* da RNP.

Este grupo de trabalho possui estreitas relações com o projeto GIGA da RNP chamado GigaIQoM, que prevê uma infra-estrutura de medições para a rede GIGA. O objetivo dessa infra-estrutura de medições é estender o GT-Medições para dar suporte à monitoração e à configuração da rede, essenciais para atender à QoS esperada pelas aplicações avançadas previstas, tais como grades computacionais, videoconferência e ambientes virtuais colaborativos.

10.1.2. LAND

O laboratório LAND[1] localizado na UFRJ também atua na área de Metrologia na Internet. Alguns de seus trabalhos foram mencionados ao longo deste texto: a proposição de métodos para a projeção de tráfego futuro em grandes redes IP [Silva et al., 2004] e para o cálculo de atraso unidirecional fim-a-fim [Rocha et al., 2004].

10.1.3. Metrologia de redes e grades computacionais

Este projeto do LNCC objetiva investigar a interação da área de Metrologia na Internet com a área de grades computacionais. Um dos objetivos é criar uma infra-estrutura, através de por exemplo um serviço de *middleware*, que viabilize a utilização

[1] `http://www.land.ufrj.br`

das informações providas por medições da rede na alocação de recursos entre agrupamentos computacionais geograficamente distribuídos [Ziviani e Schulze, 2004].

Deve-se ressaltar que o serviço de *middleware* proposto possui dupla função. De um lado, as grades computacionais podem utilizar os resultados oriundos de medições na Internet para melhorar o gerenciamento de recursos. De outro lado, a área de Metrologia na Internet requer uma infra-estrutura distribuída e uma grande capacidade de processamento para lidar com grandes quantidades de dados coletados em diferentes pontos estratégicos da rede. A alta capacidade computacional oferecida pelos agrupamentos computacionais adequa-se elegantemente a essa tarefa. Assim, o objetivo do serviço proposto é fornecer subsídios para o melhor gerenciamento de recursos para as grades computacionais e uma plataforma de alta capacidade computacional para o processamento de grande volume de dados obtidos por medições na Internet.

A monitoração da qualidade momentânea dos enlaces da rede constitui outro serviço visado que é realizável através de técnicas de medição. Esse serviço pode servir de base para o fornecimento de dados que habilitem a inferência da qualidade a ser esperada pelas aplicações executadas na grade.

10.2. Projetos internacionais

10.2.1. PingER

Como o nome indica, o projeto PingER [Matthews e Cottrell, 2000] se apoia no uso da ferramenta `ping` para a coleta de estatísticas da rede. Esse projeto foi implementado por diferentes laboratórios de física nuclear e de partículas de alta energia ao redor do mundo. O objetivo consiste em monitorar o desempenho do compartilhamento e da distribuição de um volume potencialmente muito grande de dados experimentais para análise pelos diferentes laboratórios participantes.

Cada ponto de monitoração do PingER envia 11 pacotes *ping* com uma carga de 100 octetos e intervalo entre pacotes de 1 s para nós remotos pré-estabelecidos. Em seguida, outra série de 10 pacotes com carga de 1.000 octetos e intervalo entre pacotes de 1 s é enviada para o mesmo conjunto de nós remotos. Cada ponto de monitoração envia a seqüência de *pings* a cada meia-hora. Portanto, a carga injetada na rede permanece baixa. As principais métricas obtidas são as estatísticas típicas do `ping`: taxa de perda de pacotes, RTT e conectividade.

Os dois pontos fracos da metodologia simples do projeto PingER são a amostragem periódica e o uso de pacotes ICMP. Para evitar as possíveis falhas em observar o comportamento real da rede oriundas da amostragem periódica, recomenda-se o uso de intervalos aleatórios com distribuição exponencial entre amostras, como já discutido na Seção 2.1. O uso de pacotes ICMP pode levar a observações errôneas, pois estes podem receber uma baixa prioridade de encaminhamento em alguns domínios quando comparados com pacotes TCP ou UDP para a melhoria do desempenho de QoS destes. Da mesma forma, pacotes ICMP podem ser utilizados em um ataque, então alguns domínios podem reduzir a prioridade dos pacotes ICMP para reduzir os efeitos de um eventual ataque. Em ambos os casos uma abordagem de monitoração baseada exclusivamente no uso de `ping` pode ser levada a subestimar o desempenho e o comportamento da rede monitorada.

10.2.2. MINC

A metodologia MINC (*Multicast Inference of Network Characteristics*) [Adams et al., 2000] identifica características internas da rede usando medições

fim-a-fim em multicast. Para a inferência de parâmetros como taxa de perdas de pacotes e atraso, a metodologia MINC se baseia na inerente correlação do desempenho da rede observado pelos diferentes receptores presentes na árvore multicast. Essas medições não requerem acesso administrativo aos nós internos da rede e podem monitorar grandes redes devido ao uso eficiente da banda passante pela adoção de multicast.

Para exemplificar a metodologia MINC, consideremos o problema de inferir a taxa de perda de pacotes em algum enlace da árvore multicast. A intuição em que se apoia a metodologia MINC é que a chegada de um pacote a um nó intermediário da árvore pode ser inferida pela chegada deste pacote em um ou mais receptores que estejam abaixo do mesmo nó intermediário na árvore multicast. Ao criar-se uma condição a esse último evento, pode-se determinar a probabilidade de uma transmissão bem sucedida até ou além de um determinado nó. Por exemplo, consideremos a Figura 17 que apresenta uma árvore multicast simples composta de uma fonte F, dois receptores, $R1$ e $R2$, um enlace compartilhado da fonte até um roteador multicast intermediário e um enlace deste roteador para cada um dos receptores. A fonte envia uma série de pacotes numerados em multicast aos dois receptores. Se um pacote chegar a pelo menos um dos receptores, pode-se inferir que este pacote chegou ao roteador intermediário. Logo, a razão entre o número de pacotes que alcançam ambos os receptores e o número total de pacotes que alcançam apenas o receptor $R2$ fornece uma estimativa da probabilidade de sucesso na transmissão no enlace entre o roteador intermediário e o receptor $R1$. O mesmo raciocínio pode ser aplicado para estimar a probabilidade de transmissão com sucesso nos demais enlaces. A formulação da inferência de taxa de perda de pacotes em árvores multicast com a metodologia MINC podem ser encontrada em [Cáceres et al., 1999].

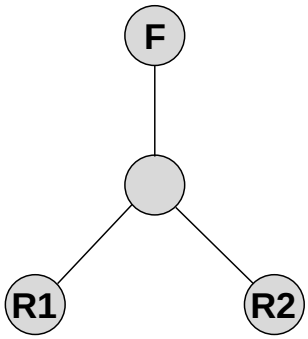


Figura 17: Exemplo da metodologia MINC.

Em [Cáceres et al., 2002], os autores propõem a combinação da metodologia MINC com o protocolo de transporte RTP (*Real-Time Transport Protocol*) para a criação de uma plataforma alternativa de medições. A idéia é utilizar o RTP e o seu protocolo de controle RTCP (*Real-Time Control Protocol*) como sondas e mensagens relatando o resultado da transmissão multicast, respectivamente. Essa abordagem pode ser facilmente incorporada em aplicações multicast que se utilizem de RTP/RTCP para transmissão e assim elas podem obter um meio de identificar enlaces problemáticos em sua árvore de distribuição.

10.2.3. NLANR

O laboratório NLANR (*National Laboratory for Applied Network Research*) é responsável por uma extensa infra-estrutura de análise de redes [McGregor et al., 2000]. A tarefa do grupo de análise e medições desse laboratório é monitorar a rede da comunidade americana HPC (*High Performance Connection*), financiada pela NSF (*National Science Foundation*). A plataforma da NLANR mantém projetos de medições

ativas e passivas. O projeto de medições ativas, chamado AMP (*Active Measurement Project*), possui atualmente cerca de 140 pontos de medição nos EUA, que se monitoram em malha de forma contínua. Pesquisadores do NLANR propuseram o IPMP (*IP Measurement Protocol*) [Luckie e McGregor, 2002] como um substituto ao uso de protocolos como ICMP, TCP e UDP para medição ativa, já que estes apresentam limitações para o uso em medições por não terem sido concebidos com esta finalidade. O projeto de medições passivas, chamado PMA (*Passive Measurement and Analysis*), possui 17 pontos de monitoração nos EUA. Dados coletados nesses projetos estão disponíveis aos pesquisadores da área e são usados amplamente pela comunidade de pesquisa.

Vale ressaltar também que o NLANR é responsável pelo suporte e desenvolvimento de ferramentas populares de medição, tais como o `iperf` [Tirumala et al., 2004]. O objetivo do `iperf` é avaliar o desempenho de transmissões TCP e UDP, medindo parâmetros como banda passante, perda de pacotes e variação do atraso. Recentemente, o NLANR lançou uma nova ferramenta chamada Advisor [Lattner et al., 2005] para integrar em uma única aplicação a medição, a análise e a apresentação de estatísticas sobre o desempenho da rede.

10.2.4. IPMON

O IPMON (*IP Monitoring*) [Fraleigh et al., 2003] é um sistema de monitoração passiva capaz de capturar estatísticas ao nível de pacotes em enlaces de alta velocidade em uma rede de *backbone*. Essa infra-estrutura de monitoração foi inovadora por duas razões. Primeiro, ela é capaz de coletar simultaneamente informações com alta granulosidade em múltiplos enlaces geograficamente dispersos. Segundo, toda a informação coletada recebe uma estampa de tempo de relógios sincronizados por GPS. Essas características permitem a realização de análises detalhadas do enfileiramento de pacotes e do comportamento do tráfego na Internet.

Esta infra-estrutura de monitoração permitiu identificar recentes mudanças em características do tráfego na Internet. Foi observado que as características de carga nos enlaces freqüentemente variam de um enlace a outro e que estas variações se correlacionam usualmente com o tipo de usuários conectados ao POP. Em alguns enlaces, foi constatado que o tráfego `http` não era mais o tráfego dominante. Nesses enlaces, aplicações de compartilhamento de arquivos em P2P e *streaming* multimídia chegavam a picos de 80% do tráfego total. Outras observações foram possíveis como a baixa taxa de desordem de pacotes em fluxos TCP e que o atraso é dominado nos *backbones* pela velocidade da luz.

10.2.5. CAIDA

A CAIDA (*Cooperative Association for Internet Data Analysis*) [CAIDA, 1997] é uma associação colaborativa com participantes dos setores comercial, governamental e acadêmico. O principal objetivo dessa associação é a investigação da estrutura e do comportamento da Internet através de medições para uma melhor compreensão da rede para a sua extensão a níveis globais de forma robusta. Dessa forma, os objetivos se dividem em medições e análise da infra-estrutura da Internet, investigação de novas tecnologias para melhorar o desempenho da rede, caracterização do comportamento do tráfego presente na rede através de medições passivas e ativas, e desenvolvimento de ferramentas de análise e visualização de características da rede.

Atualmente, a CAIDA vem se concentrando no desenvolvimento de ferramentas para a medição, análise e visualização de dados da Internet. Alguns exemplos de ferra-

mentas são o CoralReef para a análise de dados gerados por pontos de medição passiva do tráfego da Internet e o Skitter para a coleta de dados para o mapeamento da estrutura da Internet, onde mais de 23.000 destinações são monitoradas a partir de 17 pontos distribuídos nos EUA, Europa e Ásia.

10.2.6. RIPE TTM

RIPE é a organização responsável pelos registros de nomes nas regiões da Europa, Oriente Médio, Ásia Central e de alguns países africanos. Essa organização gerencia o projeto TTM (*Test Traffic Measurements*) [RIPE, 2000] para fornecer métricas unidirecionais padronizadas para o atraso e a taxa de perda de pacotes entre pontos de medição dedicados. Medições unidirecionais são importantes, pois o roteamento na Internet apresenta significativos níveis de assimetria, ou seja, pacotes entre dois nós seguem por caminhos diferentes dependendo da direção do tráfego. Para a medição de métricas unidirecionais, os pontos de medição dedicados são equipados com placas GPS para sincronização. Os pontos de medição monitoram o estado da rede entre eles em malha de forma contínua, a uma taxa de 2 pacotes por minuto. O objetivo é prover medições ativas de desempenho na rede como um serviço regular para os fornecedores de serviço.

10.2.7. Metropolis

O projeto Metropolis [Metropolis, 2001] é um projeto nacional francês sobre Metrologia na Internet, sendo atualmente o maior projeto de medições fora dos EUA. Os objetivos principais do projeto são a medição e a verificação da QoS oferecida pela rede, e o desenvolvimento de modelos realistas para a interpretação das medições. Para tanto, adotam-se medições passivas e ativas em diversas plataformas com pontos de medição RIPE, NIMI e PlanetLab atuando em conjunto. Para alcançar seus objetivos, o projeto Metropolis atua em diferentes áreas ligadas à Metrologia na Internet, tais como classificação de tráfego e dimensionamento de redes, análise de redes, desenvolvimento de métodos de medição e amostragem, modelagem de tráfego, tarifação e estabelecimento de SLAs (*Service Level Agreements*) e avaliação das plataforma de medição.

10.2.8. E2E piPEs

O projeto E2E piPEs (*End-to-End Performance Initiative Performance Environment System*) [Boyd et al., 2004] é uma nova infra-estrutura de medições para a Internet2 com 4 grandes objetivos. O primeiro objetivo é permitir aos usuários finais e aos operadores de rede determinar as capacidades de desempenho fim-a-fim, localizar problemas e contatar a pessoa correta para resolver este problema. O segundo objetivo é possibilitar o lançamento remoto de testes de desempenho em caminhos parciais. O terceiro objetivo é disponibilizar os dados de desempenho em caminhos parciais publicamente. Finalmente, o quarto objetivo consiste em promover a interoperabilidade do sistema E2E piPEs com outras arquiteturas de medição.

O sistema E2E piPEs se baseia no uso de medições realizadas através de diversas ferramentas para determinar o desempenho fim-a-fim pela agregação de informações sobre diferentes segmentos do caminho completo. Dessa forma, caminhos parciais problemáticos podem ser identificados e relatados, com o apoio de dados, para o administrador de redes apropriado.

10.2.9. CoMo

O projeto CoMo (*Continuous Monitoring*) [Iannaccone et al., 2004] é o mais recente sistema de monitoração da rede. O objetivo do sistema CoMo é fornecer um bloco básico para uma infra-estrutura de monitoração aberta que permita pesquisadores e operadores de rede processar e compartilhar facilmente estatísticas de tráfego em múltiplos pontos. O núcleo do sistema funciona separadamente dos módulos de medição. O núcleo transporta os dados da rede para armazenamento e gerencia os recursos do sistema. Os módulos de medição são responsáveis somente pela amostragem de pacotes e coleta de dados de interesse para o cálculo de alguma métrica determinada. A arquitetura é aberta de forma que os módulos possam ser implementados independentemente por diferentes desenvolvedores e então integrados ao sistema dinamicamente. Também está disponível uma interface para permitir aos usuários induzir o sistema a exportar os resultados das medições realizadas.

10.3. Perspectivas em Metrologia na Internet

Neste texto foi realizada uma revisão de métodos, técnicas e projetos existentes na área de Metrologia na Internet. Apesar dos avanços nessa área nos últimos anos, a coleta, interpretação e modelagem dos dados empíricos na Internet permanecem áreas desafiadoras de pesquisa. O primeiro grande obstáculo é a constante mudança de vários aspectos da Internet. Por exemplo, o tráfego `http` cresceu de zero em 1995 para mais de 80% do tráfego total em grande parte dos enlaces em 2000. Neste momento, a proporção de tráfego `http` parece estar caindo na maioria dos enlaces em benefício de uma presença crescente de tráfego P2P ao passo que novas aplicações baseadas neste modelo surgem.

A escala global da Internet também impõe grandes dificuldades aos projetos de medição, pois muitas vezes a composição do tráfego e o seu comportamento são dependentes da localização e de características específicas de grupos de usuários. Como conseqüência, resultados apurados em uma localização podem não ser representativos da Internet global. Portanto, as medições precisam ser realizadas a partir de múltiplos pontos para a obtenção de uma visão mais representativa.

Os protocolos da Internet não foram concebidos originalmente para suportar medições detalhadas de desempenho. Por essa razão, pesquisadores e desenvolvedores precisam elaborar meios de medir indiretamente diversos fenômenos da rede.

Ao longo deste texto, nós descrevemos diferentes problemas presentes na rede junto com diversas soluções propostas para enfrentar estes problemas. Em nenhuma das áreas investigadas há uma solução definitiva de consenso, havendo espaço para a melhoria das técnicas existentes. O aumento da capacidade de se observar o comportamento da Internet pode ser apenas o primeiro passo na direção de uma monitoração mais eficiente da rede. Apenas coletar um enorme volume de dados não é eficiente sem o desenvolvimento de ferramentas avançadas para processar esta quantidade de dados e fornecer novas bases para o projeto de aplicações e de serviços de rede avançados. O surgimento recente de novas plataformas, tais como PlanetLab, piPEs e CoMo (ver Seção 10.2), que se apoiam nas lições aprendidas nos últimos anos na área de Metrologia na Internet abre perspectivas promissoras para novas pesquisas baseadas em medições.

Agradecimentos

Este trabalho recebeu suporte financeiro do LNCC, UFRJ, FAPERJ, CAPES e CNPq. Os autores também são gratos a diversas pessoas por diferentes razões que expressamos a seguir. Antonio Tadeu Azevedo Gomes (LNCC) comentou versões preliminares deste texto. José Augusto Suruagy Monteiro (UNIFACS) disponibilizou a descrição

das propostas do GT-Medições e do projeto GigaIQoM. Lisandro Zambenedetti Granville (UFRGS) indicou o projeto E2E piPEs. John Towns (NCSA) indicou a recente ferramenta Advisor da NLANR. Timur Friedman (LIP6/CNRS), Andrew Adams (PSC) e Vern Paxon (ICIR) possibilitaram o uso da plataforma NIMI e forneceram dicas de sua utilização. Bruno Schulze (LNCC), José Ferreira de Rezende (COPPE/UFRJ), Bamba Gueye (LIP6/CNRS), Serge Fdida (LIP6/CNRS) e Mark Crovella (Boston University) são co-autores em alguns de nossos artigos na área de Metrologia na Internet e as lições assimiladas na concepção destes artigos foram de grande valia na redação deste texto.

Referências

Adams, A., Bu, T., Friedman, T., Horozitz, J., Towsley, D., Cáceres, R., Duffield, N., Lo Presti, F., Moon, S. B., e Paxson, V. (2000). The use of end-to-end multicast measurements for characterizing Internet network behavior. *IEEE Communications Magazine*, 8(5):152–158.

Amini, L., Shaikh, A., e Schulzrinne, H. (2004). Issues with inferring Internet topological attributes. *Computer Communications*, 27(6):557–567.

Barford, P., Bestravos, A., Byers, J., e Crovella, M. (2001). On the marginal utility of network topology measurements. In *Proc. of ACM/SIGCOMM Internet Measurement Workshop – IMW 2001*, San Francisco, CA, EUA.

Barford, P., Kline, J., Plonka, D., e Ron, A. (2002). A signal analysis of network traffic anomalies. In *Proc. of ACM/SIGCOMM Internet Measurement Workshop – IMW 2002*, Marselha, França.

Barford, P. e Sommers, J. (2004). Comparing probe-based and router-based packet-loss measurement. *IEEE Internet Computing*, 8(5):50–56.

Baset, S. A. e Schulzrinne, H. (2004). An analysis of the skype peer-to-peer Internet telephony protocol. Technical Report CUCS-039-04, Columbia University.

Bierman, A. e Quittek, J. (2001). *Packet Sampling (PSAMP)*. `http://www.ietf.org/html.charters/psamp-charter.html`.

Bovy, C. J., Mertodimedjo, H. T., Hooghiemstra, G., Uijterwaal, H., e van Mieghem, P. (2002). Analysis of end-to-end delay measurements in Internet. In *Proc. of the Passive and Active Measurement Workshop - PAM'2002*, Fort Collins, CO, EUA.

Boyd, E. L., Boote, J. W., Shalunov, S., e Zekauskas, M. J. (2004). *The Internet2 E2E piPEs Project: An Interoperable Federation of Measurement Domains for Performance Debugging*. `http://e2epi.internet2.edu/e2epipes`.

Bremler-Barr, A., Cohen, E., Kaplan, H., e Mansour, Y. (2003). Predicting and bypassing end-to-end Internet service degradations. *IEEE Journal on Selected Areas in Communications*, 21(6):961–978.

Brownlee, N. (2005). Some observations of Internet stream lifetimes. In *Proc. of the Passive and Active Measurement Workshop - PAM'2005*, Boston, MA, EUA.

Brownlee, N. e Claffy, K. C. (2002). Understanding Internet traffic streams: dragonflies and tortoises. *IEEE Communications Magazine*, 40(10):110–117.

Brownlee, N. e Claffy, K. C. (2004). Internet measurement. *IEEE Internet Computing*, 8(5):30–33.

Brownlee, N., Mills, C., e Ruth, G. (1999). Traffic flow measurement: Architecture. *RFC 2722*.

Bryant, F. B. e Yarnold, P. R. (1998). *Reading and Understanding Multivariate Statistics*, chapter Principal-Components Analysis and Exploratory and Confirmatory Factor Analysis, pages 99–136. APA Press.

Cáceres, R., Duffield, N., e Friedman, T. (2002). Impromptu measurement infrastructures using RTP. In *Proc. of the IEEE INFOCOM'2002*, Nova Iorque, NY, EUA.

Cáceres, R., Duffield, N., Horowitz, J., e Towsley, D. (1999). Multicast-based inference of network-internal loss characteristics. *IEEE Transactions on Information Theory*, 45(7):2462–2480.

CAIDA (1997). Cooperative association for Internet data analysis (caida). `http://www.caida.org`.

Carter, R. L. e Crovella, M. (1996). Measuring bottleneck link speed in packet-switched networks. *Performance Evaluation*, 27-28:297–318.

Case, J. D., Fedor, M., Schoffstall, M. L., e Davin, J. R. (1990). Simple network management protocol (SNMP). *RFC 1157*.

Chang, H., Govindana, R., Jamin, S., Shenker, S., e Willinger, W. (2004). Towards capturing representative AS-level Internet topologies. *Computer Networks*, 44(6):737–755.

Chen, T. M. (2001). Increasing the observability of Internet behavior. *Communications of the ACM*, 44(1):93–98.

Cisco (1999). *NetFlow*. `http://www.cisco.com/warp/public/732/Tech/nmp/netflow/`.

Claffy, K. C. (2002). Internet measurement: myths about Internet data. Talk at NANOG24 Meeting. `http://www.caida.org/outreach/presentations/Myths2002/`.

Claffy, K. C., Monk, T. E., e McRobb, D. (1999). Internet tomography. *Nature*.

Clark, D. D., Wroclawski, J., Sollins, K. R., e Braden, R. (2002). Tussle in cyberspace: Defining tomorrow's Internet. In *Proc. of the ACM SIGCOMM'2002*, Pittsburgh, PA, EUA.

Costa, M., Castro, M., Rowstron, A., e Key, P. (2004). PIC: Practical Internet coordinates for distance estimation. In *Proc. of the IEEE International Conference on Distributed Computing Systems – IEEE ICDCS'2004*, Tóquio, Japão.

Crovella, M. E. e Bestavros, A. (1997). Self-similarity in world wide web traffic: evidence and possible causes. *IEEE/ACM Transactions on Networking*, 5(6):835–846.

Dabek, F., Cox, R., Kaashoek, F., e Morris, R. (2004). Vivaldi: A decentralized network coordinate system. In *Proc. of the ACM SIGCOMM'2004*, Portland, OR, EUA.

DAG (2001). DAG cards – Endace measurement systems. `http://www.endace.com`.

Dimitropoulos, X., Krioukov, D., e Riley, G. (2005). Revisiting Internet AS-level topology discovery. In *Proc. of the Passive and Active Measurement Workshop - PAM'2005*, Boston, MA, EUA.

Donnet, B., Raoult, P., Friedman, T., e Crovella, M. (2005). Efficient algorithms for large-scale topology discovery. In *Proc. of the ACM SIGMETRICS'05*, Banff, Canadá.

Dovrolis, C., Ramanathan, P., e Moore, D. (2001). What do packet dispersion techniques measure? In *Proc. of the IEEE INFOCOM'2001*, Anchorage, AK, EUA.

Dovrolis, C., Ramanathan, P., e Moore, D. (2004). Packet dispersion techniques and a capacity estimation methodology. *IEEE/ACM Transactions on Networking*, 12(6):963–977.

Downey, A. B. (1999). Using pathchar to estimate Internet link characteristics. In *Proc. of the ACM SIGCOMM'99*, Cambridge, MA, EUA.

Duffield, N. e Grossglauser, M. (2001). Trajectory sampling for direct traffic observation. *IEEE/ACM Transactions on Networking*, 9(3):280–292.

Enge, P. e Misra, P. (1999). Special issue on global positioning system. *Proceedings of the IEEE*, 87(1):3–15.

Estan, C., Keys, K., Moore, D., e Varghese, G. (2004). Building a better NetFlow. In *Proc. of the ACM SIGCOMM'2004*, Portland, OR, EUA.

Estan, C. e Varghese, G. (2002). New directions in traffic measurement and accounting. In *Proc. of the ACM SIGCOMM'2002*, Pittsburgh, PA, EUA.

Faloutsos, M., Faloutsos, P., e Faloutsos, C. (1999). On power-law relationships of the Internet topology. In *Proc. of the ACM SIGCOMM'99*, Cambridge, MA, EUA.

Feldmann, A., Greenberg, A., Lund, C., Reingold, N., Rexford, J., e True, F. (2001). Deriving traffic demands for operational IP networks: Methodology and experience. *IEEE/ACM Transactions on Networking*, 9(3):265–279.

Figueiredo, D. R., Liu, B., Feldmann, A., Misra, V., Towsley, D., e Willinger, W. (2005). On TCP and self-similar traffic. *Performance Evaluation*. Special issue on Long Range Dependence and Heavy Tail Distributions. No prelo.

Floyd, S. e Paxson, V. (2001). Difficulties in simulating the Internet. *IEEE/ACM Transactions on Networking*, 9(4):392–403.

Fortz, B. e Thorup, M. (2000). Internet traffic engineering by optimizing OSPF weights. In *Proc. of the IEEE INFOCOM'2000*, Tel Aviv, Israel.

Fraleigh, C., Moon, S., Lyles, B., Cotton, C., Khan, M., Moll, D., Rockell, R., Seely, T., e Diot, C. (2003). Packet-level traffic measurements from the Sprint IP backbone. *IEEE Network*, 17(6):6–16.

Francis, P., Jamin, S., Jin, C., Jin, Y., Raz, D., Shavitt, Y., e Zhang, L. (2001). IDMaps: A global Internet host distance estimation service. *IEEE/ACM Transactions on Networking*, 9(5):525–540.

Govindan, R. e Tangmunarunkit, H. (2000). Heuristics for internet map discovery. In *Proc. of the IEEE INFOCOM'2000*, Tel Aviv, Israel.

Gueye, B., Ziviani, A., Crovella, M., e Fdida, S. (2004). Constraint-based geolocation of Internet hosts. In *Proc. of ACM/SIGCOMM Internet Measurement Conference – IMC 2004*, Taormina, Itália.

Gummadi, K. P., Saroiu, S., e Gribble, S. D. (2002). King: Estimating latency between arbitrary Internet end hosts. In *ACM Internet Measurement Workshop 2002*, Marselha, França.

Habib, A., Khan, M., e Bhargava, B. (2004). Edge-to-edge measurement-based distributed network monitoring. *Computer Networks*, 44(2):211–233.

Hu, N., Li, L., Mao, Z. M., Steenkiste, P., e Wang, J. (2004). Locating Internet bottlenecks: Algorithms, measurements, and implications. In *Proc. of the ACM SIGCOMM'2004*, Portland, OR, EUA.

Hu, N., Li, L., Mao, Z. M., Steenkiste, P., e Wang, J. (2005). A measurement study of Internet bottleneck. In *Proc. of the IEEE INFOCOM'2005*, Miami, FL, EUA.

Hu, N. e Steenkiste, P. (2003). Evaluation and characterization of available bandwidth probing techniques. *IEEE Journal on Selected Areas in Communications*, 21(6):879–894.

Huffaker, B., Fomenkov, M., Plummer, D. J., Moore, D., e k claffy (2002). Distance metrics in the Internet. In *Proc. of the IEEE International Telecommunications Symposium - ITS'2002*, Natal, RN, Brasil.

Huitema, C. (2000). *Routing in the Internet*. Prentice Hall.

Iannaccone, G., Diot, C., McAuley, D., Moore, A., Pratt, I., e Rizzo, L. (2004). The CoMo white paper. Technical Report IRC-TR-04-017, Intel Research. `http://www.cambridge.intel-research.net/como/`.

Internet World Stats (2005). Internet usage and population statistics. `http://www.internetworldstats.com/stats.htm`.

Ishibashi, K., Kanazawa, T., Aida, M., e Ishii, H. (2004). Active/passive combination-type performance measurement method using change-of-measure framework. *Computer Communications*, 27(9):868–879.

Jacobson, V. (1988). Congestion avoidance and control. In *Proc. of the ACM SIGCOMM'88*, Stanford, CA, EUA.

Jacobson, V. (1997). Pathchar: A tool to infer characteristics of Internet paths. `http://www.caida.org/tools/utilities/others/pathchar/`.

Jain, M. e Dovrolis, C. (2002). End-to-end available bandwidth: Measurement methodology, dynamics, and relation with TCP throughput. In *Proc. of the ACM SIGCOMM'2002*, Pittsburgh, PA, EUA.

Jain, M. e Dovrolis, C. (2004). Ten fallacies and pittfalls on end-to-end available bandwidth estimation. In *Proc. of ACM/SIGCOMM Internet Measurement Conference – IMC 2004*, Taormina, Itália.

Jaiswal, S., Iannaccone, G., Diot, C., Kurose, J., e Towsley, D. (2004). Inferring TCP connection characteristics through passive measurements. In *Proc. of the IEEE INFOCOM'2004*, Hong Kong.

Joo, Y., Ribeiro, V., Feldmann, A., Gilbert, A., e Willinger, W. (1999). On the impact of variability on the buffer dynamics in IP networks. In *Proc. of the Allerton Conference on Communication, Control and Computing*, Urbana, IL, EUA.

Karagiannis, T., Molle, M., e Faloutsos, M. (2004). Long-range dependence: Ten years of Internet traffic modeling. *IEEE Internet Computing*, 8(5):57–64.

Lai, K. e Baker, M. (1999). Measuring bandwidth. In *Proc. of the IEEE INFOCOM'99*, Nova Iorque, NY, EUA.

Lai, K. e Baker, M. (2000). Measuring link bandwidths using a deterministic model of packet delay. In *Proc. of the ACM SIGCOMM'2000*, Estocolmo, Suécia.

Lakhina, A., Byers, J., Crovella, M., e Matta, I. (2003a). On the geographic location of Internet resources. *IEEE Journal on Selected Areas in Communications*, 21(6):934–948.

Lakhina, A., Byers, J., Crovella, M., e Xie, P. (2003b). Sampling biases in IP topology measurements. In *Proc. of the IEEE INFOCOM'2003*, San Francisco, CA, EUA.

Lakhina, A., Crovella, M., e Diot, C. (2004). Diagnosing network-wide traffic anomalies. In *Proc. of the ACM SIGCOMM'2004*, Portland, OR, EUA.

Larrieu, N. e Owezarski, P. (2005). Towards a measurement based networking approach for Internet QoS improvement. *Computer Communications*, 28(3):259–273.

Lattner, T., Engelhardt, S., Estabrook, J., Ferguson, J., Ko, S., Kutzko, M., Liu, J., Psaltoulis, D., Thompson, N., e Yang, C. (2005). Network performance advisor – version 2.0. `http://dast.nlanr.net/Projects/Advisor/`.

Leland, W., Taqqu, M., Willinger, W., e Wilson, D. (1994). On the self-similar nature of ethernet traffic. *IEEE/ACM Transactions on Networking*, 2(1):1–15.

Lim, H., Hou, J. C., e Choi, C.-H. (2003). Constructing Internet coordinate system based on delay measurement. In *ACM Internet Measurement Conference 2003*, Miami, FL, EUA.

Luckie, M. J. e McGregor, A. J. (2002). IPMP: IP measurement protocol. In *Proc. of the Passive and Active Measurement Workshop - PAM'2002*, Fort Collins, CO, EUA.

Mao, G. (2005). A real-time loss perfomance monitoring scheme. *Computer Communications*, 28(2):150–161.

Matthews, W. e Cottrell, L. (2000). The PingER project: Active Internet performance monitoring for the HENP community. *IEEE Communications Magazine*, 8(5):130–136.

McGregor, A., Braun, H.-W., e Brown, J. (2000). The NLANR network analysis infrastructure. *IEEE Communications Magazine*, 8(5):122–128.

Medina, A., Lakhina, A., Matta, I., e Byers, J. (2001). BRITE: An approach to universal topology generation. In *Proc. of the MASCOTS'2001*, Cincinnati, OH, EUA.

Medina, A., Taft, N., Salamatian, K., Bhattacharyya, S., e Diot, C. (2002). Traffic matrix estimation: Existing techniques and new directions. In *Proc. of the ACM SIGCOMM'2002*, Pittsburgh, PA, EUA.

Metropolis (2001). *Metropolis: METROlogie Pour l'Internet et ses services*. `http://www.telecom.gouv.fr/rnrt/rnrt/projets/res_01_57.htm`.

Monteiro, J. A. S. (2004). *Grupo de Trabalho de Medições em Redes*. `http://www.nuperc.unifacs.br/gtmed/`.

Moon, S., Skelly, P., e Towsley, D. (1999). Estimation and removal of clock skew from network delay measurements. In *Proc. of the IEEE INFOCOM'99*, Nova Iorque, NY, EUA.

Moore, A. e Papagiannaki, K. (2005). Toward the accurate identification of network applications. In *Proc. of the Passive and Active Measurement Workshop - PAM'2005*, Boston, MA, EUA.

Ng, T. S. E. e Zhang, H. (2002). Predicting Internet network distance with coordinates-based approaches. In *Proc. of the IEEE INFOCOM'2002*, Nova Iorque, NY, EUA.

Padmanabhan, V. N. e Subramanian, L. (2001). An investigation of geographic mapping techniques for Internet hosts. In *Proc. of the ACM SIGCOMM'2001*, San Diego, CA, EUA.

Papagiannaki, K., Taft, N., e Diot, C. (2004a). Impact of flow dynamics on traffic engineering design principles. In *Proc. of the IEEE INFOCOM'2004*, Hong Kong.

Papagiannaki, K., Taft, N., e Lakhina, A. (2004b). A distributed approach to measure IP traffic matrices. In *Proc. of ACM/SIGCOMM Internet Measurement Conference – IMC 2004*, Taormina, Italy.

Papagiannaki, K., Taft, N., Zhang, Z., e Diot, C. (2003). Long-term forecasting of Internet backbone traffic: Observations and initial models. In *Proc. of the IEEE INFOCOM'2003*, San Francisco, CA, EUA.

Pásztor, A. e Veitch, D. (2002). PC-based precision timing without GPS. In *Proc. of the ACM SIGMETRICS'02*, Los Angeles, CA, EUA.

Paxon, V., Almes, G., Mahdavi, J., e Mathis, M. (1998). Framework for IP performance metrics. *RFC 2330*.

Paxson, V. (1997). *Measurement and Analysis of End-to-end Internet Dynamics*. PhD thesis, University of California - Berkeley.

Paxson, V. (1998). On calibrating measurements of packet transit times. In *Proc. of the ACM SIGMETRICS'98*, Madison, WI, EUA.

Paxson, V. e Floyd, S. (1995). Wide area traffic: The failure of Poisson modeling. *IEEE/ACM Transactions on Networking*, 3(3):226–244.

Paxson, V., Mahdavi, J., Adams, A., e Mathis, M. (1998). An architecture for large-scale Internet measurement. *IEEE Communications Magazine*, 36(8):48–54.

Percacci, R. e Vespignani, A. (2003). Scale-free behavior of the Internet global performance. *The European Physical Journal B - Condensed Matter*, 32(4):411–414.

Peterson, L., Anderson, T., Culler, D., e Roscoe, T. (2002). A blueprint for introducing disruptive technology into the internet. In *Proc. of the 1st Workshop on Hot Topics in Networks (HotNets-I)*, Princeton, NJ, EUA. `http://www.planet-lab.org`.

Pias, M., Crowcroft, J., Wilbur, S., Harris, T., e Bhatti, S. (2003). Lighthouses for scalable distributed location. In *Proc. of the Second International Workshop on Peer-to-Peer Systems - IPTPS'03*, Berkeley, CA, EUA.

Postel, J. (1981). Internet control message protocol. *RFC 792*.

Prasad, R., Dovrolis, C., e Mah, B. (2003a). The effect of layer-2 store-and-forward devices on per-hop capacity estimation. In *Proc. of the IEEE INFOCOM'2003*, San Francisco, CA, EUA.

Prasad, R., Dovrolis, C., Murray, M., e Claffy, K. C. (2003b). Bandwidth estimation: Metrics, measurement tecniques, and tools. *IEEE Network*, 17(6):27–35.

Ribeiro, V., Riedi, R. H., e Baraniuk, R. G. (2004). Locating available bandwidth bottlenecks. *IEEE Internet Computing*, 8(5):34–41.

RIPE (2000). *RIPE Test Traffic Measurements*. `http://www.ripe.net/ttm/`.

Ripeanu, M., Iamnitchi, A., e Foster, I. T. (2002). Mapping the Gnutella network. *IEEE Internet Computing*, 6(1):50–57.

Rocha, A., Leão, R. M. M., 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, RS, Brasil.

Roughan, M., Griffin, T., Mao, M., Greenberg, A., e Freeman, B. (2004). IP forwarding anomalies and improving their detection using multiple data sources. In *Proc. of the ACM SIGCOMM'2004 Workshop on Network Troubleshooting*, Portland, OR, EUA.

Sang, A. e Li, S. (2000). A predictability analysis of network traffic. In *Proc. of the IEEE INFOCOM'2000*, Tel Aviv, Israel.

Saroiu, S., Gummadi, P. K., e Gribble, S. D. (2002). Sprobe: A fast technique for measuring bottleneck bandwidth in uncooperative environments. `http://sprobe.cs.washington.edu/`.

Shaikh, A., Rexford, J., e Shin, K. G. (1999). Long-sensitive routing of long-lived IP flows. In *Proc. of the ACM SIGCOMM'99*, Cambridge, MA, EUA.

Shannon, C., Moore, D., e Claffy, K. C. (2002). Beyond folklore: Observations on fragmented traffic. *IEEE/ACM Transactions on Networking*, 10(6):709–720.

Shavitt, Y. e Tankel, T. (2003). Big-bang simulation for embedding network distances in Euclidean space. In *Proc. of the IEEE INFOCOM'2003*, San Francisco, CA, EUA.

Shriram, A., Murray, M., Hyun, Y., Brownlee, N., Broido, A., Fomenkov, M., e Claffy, K. (2005). Comparison of public end-to-end bandwidth estimation tools on high-speed links. In *Proc. of the Passive and Active Measurement Workshop - PAM'2005*, Boston, MA, EUA.

Silva, E., Leão, R. M. M., Trindade, M. B., de A. Rocha, A. A., Ribeiro, B. F., Duarte, F. P., e Azevedo, J. A. (2004). Um método para projeção de tráfego usando wavelets e fecho convexo. In *XXI Simpósio Brasileiro de Telecomunicações*, Belém, PA, Brasil.

Smith, P. (2005). Cidr report. `http://www.cidr-report.org`.

Snoeren, A. C., Partridge, C., Sanchez, L. A., Jones, C. E., Tchakountio, F., Schwartz, B., Kent, S. T., e Strayer, W. T. (2002). Single-packet IP traceback. *IEEE/ACM Transactions on Networking*, 10(6):721–734.

Soule, A., Lakhina, A., Taft, N., Papagiannaki, K., Salamatian, K., Nucci, A., Crovella, M., e Diot, C. (2005). Traffic matrices: Balancing measurements, inference and modeling. In *Proc. of the ACM SIGMETRICS'05*, Banff, Canadá.

Soule, A., Nucci, A., Cruz, R., Leonardi, E., e Taft, N. (2004a). How to identify and estimate the largest traffic matrix elements in a dynamic environment. In *Proc. of the ACM SIGMETRICS'04*, Nova Iorque, NY, EUA.

Soule, A., Salamatian, K., Emilion, R., Taft, N., e Papagiannaki, K. (2004b). Flow classification by histograms or how to go on safari in the Internet. In *Proc. of the ACM SIGMETRICS'04*, Nova Iorque, NY, EUA.

Spring, N., Mahajan, R., Wetherall, D., e Anderson, T. (2004). Measuring ISP topologies with Rocketfuel. *IEEE/ACM Transactions on Networking*, 12(1):2–16.

Spring, N., Wetherall, D., e Anderson, T. (2003). Reverse-engineering the Internet. In *Proc. of the 2nd Workshop on Hot Topics in Networks (HotNets-II)*, Cambridge, MA, EUA.

Subramanian, L., Agarwal, S., Rexford, J., e Katz, R. H. (2002). Characterizing the Internet hierarchy from multiple vantage points. In *Proc. of the IEEE INFOCOM'2002*, Nova Iorque, NY, EUA.

Tang, L. e Crovella, M. (2003). Virtual landmarks for the Internet. In *Proc. of the ACM Internet Measurement Conference 2003*, Miami, FL, EUA.

Tangmunarunkit, H., Govindan, R., Jamin, S., Shenker, S., e Willinger, W. (2002). Network topology generators: Degree-based vs structural. In *Proc. of the ACM SIGCOMM'2002*, Pittsburgh, PA, EUA.

Teixeira, R., Duffield, N., Rexford, J., e Roughan, M. (2005). Traffic matrix reloaded: Impact of routing changes. In *Proc. of the Passive and Active Measurement Workshop - PAM'2005*, Boston, MA, EUA.

Tirumala, A., Qin, F., Dugan, J., Ferguson, J., e Gibbs, K. (2004). Iperf: Network performance testing – version 2.0.1. `http://dast.nlanr.net/Projects/Iperf/`.

Uijterwaal, H. e Zekauskas, M. (2003). *IP Performance Metrics (IPPM)*. `http://www.ietf.org/html.charters/ippm-charter.html`.

Varghese, G. e Estan, C. (2004). The measurement manifesto. *ACM Computer Communication Review*, 34(1):9–14.

Wang, J., Zhou, M., e Zhou, H. (2004). Clock synchronization for Internet measurements: A clustering algorithm. *Computer Networks*, 45(6):731–741.

Wei, W., Wang, B., Zhang, C., Kurose, J., e Towsley, D. (2005). Classification of access network types: Ethernet, wireless LAN, ADSL, cable modem or dialup? In *Proc. of the IEEE INFOCOM'2005*, Miami, FL, EUA.

Zegura, E. W., Calvert, K. L., e Bhattacharjee, S. (1999). How to model an internetwork. In *Proc. of the IEEE INFOCOM'99*, San Francisco, CA, EUA.

Zennström, N. e Friis, J. (2003). Skype. `http://www.skype.com/`.

Zhang, Y., Roughan, M., Duffield, N., e Greenberg, A. (2003a). Fast accurate computation of large-scale IP traffic matrices from link loads. In *Proc. of the ACM SIGMETRICS'03*, San Diego, CA, EUA.

Zhang, Y., Roughan, M., Lund, C., e Donoho, D. (2003b). An information-theoretic approach to traffic matrix estimation. In *Proc. of the ACM SIGCOMM'2003*, Karlsruhe, Alemanha.

Ziviani, A., Fdida, S., de Rezende, J. F., e Duarte, O. C. M. B. (2004). Toward a measurement-based geographic location service. In *Proc. of the Passive and Active Measurement Workshop - PAM'2004*, Lecture Notes in Computer Science (LNCS) 3015, pages 43–52, Antibes Juan-les-Pins, França.

Ziviani, A., Fdida, S., de Rezende, J. F., e Duarte, O. C. M. B. (2005a). Improving the accuracy of measurement-based geographic location of Internet hosts. *Computer Networks*, 47(4):503–523.

Ziviani, A., Gueye, B., Crovella, M., e Fdida, S. (2005b). CBG: Geolocalização na Internet usando medições de atraso. In *XXIII Simpósio Brasileiro de Redes de Computadores - SBRC'2005*, Fortaleza, CE, Brasil.

Ziviani, A. e Schulze, B. (2004). Measurement middleware service for grid computing. In *Poster in the 2nd International Workshop on Middleware for Grid Computing - MGC 2004*, Toronto, Canadá.

Zook, M. (2001). Connected is a matter of geography. *ACM NetWorker*, 5(3):13–17.