

Uma Proposta de Infra-Estrutura para a Medição Passiva de Tráfego na Rede RNP2

Eduardo Leivas Bastos
elbastos@ieee.org

Luís Felipe Balbinot
hades@inf.ufrgs.br

Programa de Pós-Graduação em Computação (PPGC)
Universidade Federal do Rio Grande do Sul (UFRGS)
Av. Bento Gonçalves, 9500 – Bloco IV – 91501-970 – Porto Alegre, RS
Pesquisa em Redes de Alta Velocidade (PRAV)
Universidade do Vale do Rio dos Sinos (UNISINOS)
Av. Unisinos, 950 – Sala 6008 – 93022-000 – São Leopoldo, RS

1 Introdução

Apesar de seu crescimento exponencial verificado nos últimos anos e de sua importância cada vez maior na economia e na sociedade, a Internet caracteriza-se por ser uma rede fracamente monitorada e instrumentalizada [1]. Salvo algumas exceções, o planejamento de capacidade, a engenharia de novas redes, e o oferecimento de garantias de desempenho são processos que baseiam-se, na maioria das vezes, mais na intuição e premissas duvidosas do que em dados concretos dos níveis de utilização e desempenho da rede. Infra-estruturas de medição de tráfego são cada vez mais necessárias na Internet a fim de torná-la mais previsível, gerenciável e escalável.

Existem vários usos para uma infra-estrutura de medição de tráfego [1]: 1) diagnosticar problemas de desempenho no interior da rede através da coleta e análise de tráfego em diversos pontos; 2) medir as propriedades de vários caminhos da rede (*paths*) a fim de facilitar pesquisas sobre o comportamento e a evolução da rede [2, 3]; e 3) certificar o desempenho fim-a-fim oferecido por diferentes provedores de acesso à Internet.

Vários esforços de pesquisa têm sido feitos no sentido de desenvolver infra-estruturas de medição de tráfego na Internet [4, 5]. Entre eles, vale destacar a atuação de dois organismos americanos que trabalham em conjunto na coleta e análise de dados de tráfego e no desenvolvimento de ferramentas de captura e visualização: o CAIDA (*Cooperative Association for Internet Data Analysis*) [6] e o NLANR (*National Laboratory for Applied Network Research*) [7].

Esse artigo propõe a criação de uma infra-estrutura de medição de tráfego, similar à desenvolvida pelo CAIDA e pelo NLANR, no backbone RNP2. É proposto um modelo de monitoramento passivo (sem interferência na operação normal da rede) onde diversos *probes* (ou coletores de dados) são dispostos em enlaces estratégicos do backbone para a coleta de tráfego. A análise dos dados coletados é feita em um nodo central e os resultados, bem como os dados originais, são disponibilizados em um servidor central e acessíveis via interface Web. O objetivo principal nesse estágio é a geração de *traces* que possam ser utilizados para análise de perfis de tráfego e utilização dos enlaces. O uso de monitoramento ativo não está inserido nessa proposta, mas é esperado que venha a ser utilizado no futuro para a medição de latência, estabilidade de roteamento e outras métricas que exigem a cooperação fim-a-fim.

2 A Arquitetura *CoralReef*

CoralReef é um conjunto de ferramentas flexível e de alto desempenho para a coleta e análise de tráfego da Internet através dos cabeçalhos dos pacotes capturados. Esse conjunto inclui *drivers*,

bibliotecas e aplicações de análise para monitores de interfaces STM-x ATM e PoS e arquivos do tipo `pcap`. A metodologia de coleta de tráfego é não intrusiva e passiva, ou seja, não requer que tráfego extra seja inserido na rede para realizar as medições, bem como não interfere no funcionamento dos equipamentos. Não é necessária uma infra-estrutura adicional na rede para a sua operação.

Em enlaces ATM é feita a captura da primeira célula de cada quadro AAL5 e em redes PoS é feita a captura dos primeiros 64 bytes de cada pacote (em ambas capturando os cabeçalhos TCP/IP e UDP/IP). Em casos especiais pode-se realizar a captura de todas as células de um enlace. Uma marcação de *timestamps* bastante precisa é feita pelo relógio interno das placas de captura, que pode ser sincronizado com uma fonte GPS, dando uma precisão melhor do que 1 μ s. Em níveis mais acima, bibliotecas realizam o desencapsulamento de IP sobre ATM e IP sobre SONET/SDH. Pouco – ou nenhum – dado de usuário é carregado nos *traces*, mas esses dados podem ser anonimizados, juntamente com as informações dos cabeçalhos, para que os *traces* possam ser disponibilizados ao público sem comprometer a segurança e a privacidade.

As estações de coleta são computadores comuns (Pentium II 400Mhz, 256MB RAM, 20+ GB HD) com placas de captura ATM. O sistema operacional pode ser Linux ou FreeBSD. Uma estação completa custa em torno de R\$ 5.000,00 para enlaces OC-3 e R\$ 15.000,00 para enlaces OC-3/OC-12. Para a captura podem ser utilizadas as placas Fore 200E ou as placas DAG, desenvolvidas pela Universidade de Waikato, que são específicas para a captura de tráfego.

Para desviar a luz para as interfaces de captura são utilizados *splitters* ópticos. Em enlaces monomodo, esses *splitters* fazem uma divisão de 90/10, ou seja, 90% dos fótons continuam fluindo normalmente até a outra ponta do enlace, enquanto que os 10% restantes são desviados para as interfaces de captura. Essa divisão pode ser feita em ambas direções. A atenuação resultante permite que o sinal monomodo seja inserido nas interfaces multimodo das placas de captura. Em enlaces multimodo utiliza-se *splitters* multimodo com divisão 50/50. Um par de *splitters* monomodo 90/10 custa em torno de R\$ 2.000,00.

3 Coleta e Disponibilização de *Traces*

A coleta e disponibilização de dados proposta é em grande parte baseada na infra-estrutura de análise do *Measurement & Operations Analysis Team* (MOAT) do NLANR, que é apresentada em [7]. A infra-estrutura base de operação é ilustrada na Figura 1.

3.1 Coleta

A captura de células em um enlace OC-3 em pleno uso gera grandes quantidades de dados. A longo prazo isso exige muito espaço de armazenamento para manter um histórico relativamente recente. Para reduzir a quantidade de dados que são transportados entre os pontos de monitoração e a estação central de armazenamento seria necessário fazer uma abstração desses dados em locais próximos aos pontos de monitoração. O problema é que essa abstração não conseguiria gerar informações valiosas o suficiente para que estudos mais avançados fossem feitos sobre os pacotes capturados e limitaria muito o estudo do tráfego. É importante que esses dados sejam armazenados na íntegra por várias razões, sendo as principais:

- Para se realizar uma caracterização detalhada do comportamento e tendências da rede é necessário que se mantenha a base de dados original em um histórico;
- Correlacionamento de tráfego em enlaces diferentes;
- No futuro, novas análises podem ser feitas (p. ex., para estudar o impacto de um novo protocolo introduzido na rede);
- Existe uma carência de modelos reais de tráfego para a realização de simulações em universidades (p. ex., estudos de tráfego auto-similar);

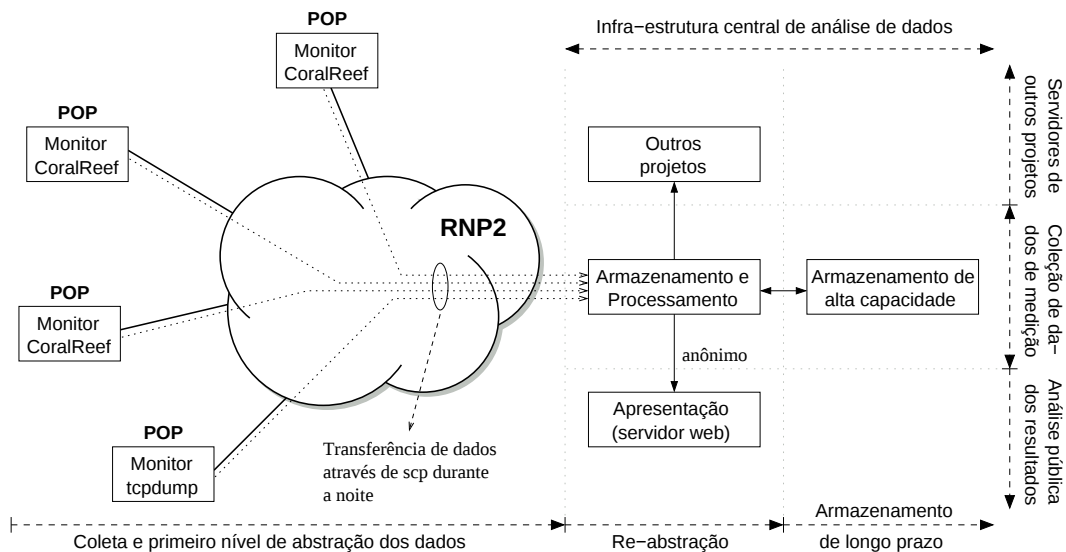


Figura 1: Infra-estrutura da proposta de medição de tráfego

- Para compartilhar os *traces* com grupos de pesquisa no Brasil e no exterior.

Baseados em dados e experiência de outros grupos de pesquisa ([7] e [9]), propomos a captura de 90 segundos de tráfego, 8 vezes ao dia, em intervalos de 3 horas. O momento exato de início de captura seria determinado por um valor pseudo-randômico dentro da hora, para evitar a colisão com eventos regulares na rede. Os dados podem ser capturados e compactados em tempo real. Logo após o final da coleta são gerados relatórios (primeiro nível de abstração) sobre o *trace* recém capturado, para evitar que esse processamento seja delegado à estação central.

3.2 Disponibilização

Em horários predeterminados (durante a noite), seria feita a transferência dos *traces* e relatórios para uma estação central de armazenamento (p. ex., através de *scp*) onde os arquivos recebem um pré-processamento (anonimização, catalogação, etc.) e são disponibilizados ao público. Algum processamento para a coleta de estatísticas também pode ser realizado nessa estação. Uma interface de busca no estilo do *Data cube* [8] pode ser desenvolvida para facilitar a busca e a obtenção dos dados catalogados. Outros projetos de pesquisa podem fazer uso dos *traces* públicos e disponibilizar essas informações independentemente.

À primeira vista, alguém pode questionar se a transferência dos arquivos de *trace* não gera uma quantidade muito grande de informações na rede. Considerando-se grosseiramente que, em média, os pacotes IP capturados em um enlace E3 ATM tenham 400 bytes de tamanho, um *trace* de 90 segundos ocuparia aproximadamente 22 MB em disco (compactado). Um arquivo com esse tamanho levaria menos de 20 segundos para ser transferido nesse mesmo enlace. Estudos publicados em [9] revelam que esses *traces* possuem tipicamente uma taxa de compactação de 55% (usando-se *gzip* com nível 9 de compactação).

Para montar uma estrutura organizada para a disponibilização dos *traces* é preciso estabelecer uma regra para a nomenclatura dos arquivos. Nossa proposta é o uso do estilo `XXXXX-YYYYYYYYY.EEE.{enc}.gz`, onde `XXXXX` é uma sigla de até 5 caracteres identificando o enlace de onde o tráfego é proveniente, `YYYYYYYYY` é um *timestamp* UNIX indicando a hora e data de captura e `EEE` indica o formato do arquivo de captura, que é `“crl”` no caso de *traces* do *CoralReef*. A extensão opcional `“enc”` fica reservada para *traces* que foram anonimizados. Por exemplo, um *trace* proveniente do POP-RS poderia ser representado por `POPRS-985703842.crl1.enc.gz`.

Os *traces* podem ser obtidos diretamente de um servidor FTP, onde existe toda uma hierarquia de diretórios (ordenados por data), sendo que em cada um desses diretórios existe todo o conjunto de *traces* capturados pelos *probes* naquele dia específico. Também existe a possibilidade de se utilizar uma interface Web, para facilitar a busca por esses *traces*. Outros projetos que possuem servidores localizados dentro das instalações da central de medição de tráfego podem fazer acesso direto a esses *traces* ou aos *traces* originais não anonimizados (com as autorizações devidas). Esses acessos podem ser feitos por protocolos mais eficientes, se necessário.

4 Conclusão

A implementação de uma infra-estrutura de medição de tráfego no backbone RNP2 trará um entendimento maior do comportamento observado em redes de alta velocidade, além de propiciar maior visibilidade, gerenciabilidade e previsibilidade ao backbone. Os dados coletados constituirão uma base de dados rica e extensa a respeito de perfis de tráfego, graus de utilização de enlaces e outras métricas de desempenho da rede. A colocação de novos *probes* deverá ser incentivada sempre que possível, de forma a ampliar a malha de cobertura e aumentar a precisão e cobertura das análises. A criação de um organismo, similar ao CAIDA, seria de inestimável valor para a Internet brasileira. Tal organismo coordenaria os esforços de monitoramento da Internet brasileira e serviria como ponto central para discussões, encontros, debates e disseminação de informações sobre o assunto. Além disso, a existência de um organismo oficial facilitaria a inclusão da Internet brasileira em outras infra-estruturas de medição de tráfego e abriria novas áreas de pesquisa, tais como o desenvolvimento de ferramentas de monitoramento e visualização.

Agradecimentos

Os autores gostariam de agradecer à Universidade do Vale do Rio dos Sinos (UNISINOS), participante do projeto MetroPoa/RMAV através da Pesquisa em Redes de Alta Velocidade (PRAV), pelo apoio e pelos recursos oferecidos para a montagem do analisador de tráfego *CoralReef*, hoje em funcionamento em um dos enlaces da rede MetroPoa. Agradecemos também à RNP, pelos recursos destinados a montagem da Internet2 brasileira.

Referências

- [1] V. Paxson e J. Mahdavi, “An Architecture for Large-Scale Internet Measurement”, *IEEE Communications Magazine*, agosto de 1998.
- [2] V. Paxson, “End-to-End Routing Behavior in the Internet”, *IEEE/ACM Trans. Networking*, 5(5):601–615, 1997.
- [3] V. Paxson, “End-to-End Packet Dynamics”, SIGCOMM’97, Cannes, França, setembro de 1997.
- [4] C. Labovitz et al, “The Internet Performance and Analysis Project”, <http://www.merit.edu/ipma/docs/team.html>, 1997.
- [5] C. Huitema et al, “Project Felix: Independent Monitoring for Network Survivability”, <ftp://ftp.bellcore.com/pub/nwg/felix/index.html>, 1997.
- [6] CAIDA. <http://www.caida.org>.
- [7] A. McGregor, H-W. Braun e J. Brown, “The NLANR Network Analysis Infrastructure”, *IEEE Communications Magazine*, 38(5):122–128, 2000.
- [8] MOAT/NLANR *Data cube*. <http://moat.nlanr.net/Datacube>.
- [9] J. Meehl, I. Graham e N. Brownlee, “The Auckland data set: an access link observed”. Em *Proc. of the 14th ITC Specialists Seminar on Access Networks and Systems*, Barcelona/Gerona, Espanha, abril de 2001.