

Dificuldades em Simular a Internet

Marita Maestrelli marita@cbpf.br

Janeiro 2007

PREFÁCIO

Este estudo originou do curso de Redes de Computadores, ministrado pelo prof. Ronaldo Moreira Salles do IME em 2006.

O tema foi entregue a mim para ser desenvolvido e apresentado em um seminário realizado durante o curso.

Durante alguns meses estudei alguns artigos a respeito e me baseei no dos cientistas seniors do ICSI-LBNL¹, Sally Floyd e Vern Paxson[3].

A simulação é um meio de obter e/ou prever resultados mais rápidos sem a necessidade de altos custos e grandes deslocamentos, além de manipular dados heterogêneos e muito variáveis.

A Internet é uma união de redes que interliga milhões de computadores pelo mundo permitindo, assim, o acesso a informações de todo tipo e transferência de dados, e isso envolve um alto grau de complexidade.

¹ *International Computer Science Institute (ICSI) in Berkely and Lawrence Berkeley National Laboratory(LBNL)*

Índice

1 – Introdução	04
2 – Simulação	05
3 – Propriedades da Internet	07
4 – Dificuldades de modelar a Internet	08
4.1 – Heterogeneidade	08
4.2 – Dimensão	11
4.3 – Mudanças Drásticas	12
5 – Estratégias	12
6 – Simuladores	14
7 – Conclusões	16
8 – Referências	17

1 - Introdução

A simulação é um recurso muito utilizado em diversas áreas por proporcionar a representação artificial de um mundo real, sem que sejam necessários altos custos e grandes deslocamentos.

Neste trabalho, nosso mundo é a Internet com alto grau de complexidade devido a sua grande heterogeneidade e as rápidas mudanças em seu meio - o tráfego de pacotes na Rede.

A heterogeneidade se caracteriza pelos links individuais onde passam o tráfego da rede, aos protocolos que operam nesses links, a variedade de aplicações utilizadas em sites, e o nível de congestionamento de cada link.

Assim, a simulação é uma técnica utilizada com muita frequência, pela flexibilidade em testar cenários variados, incluindo o comportamento de protocolos e novas tecnologias e efeito de diferentes topologias.

Entretanto, simular o comportamento global da Internet é um imenso desafio que tem feito especialistas desenvolverem diversos estudos a respeito. Sally Floyd é uma dessas especialistas.

O foco central desse trabalho é a publicação de Sally Floyd e Vern Paxson sobre as dificuldades em simular a Internet [3]. Como escrito em seu artigo: Em face da complexidade da rede, simulação tem um papel vital na caracterização do comportamento da Internet atual e nos possíveis efeitos das mudanças propostas para a sua operação. Sendo a modelagem e simulação da Internet uma tarefa difícil, o objetivo do artigo é a discussão de algumas questões e dificuldades na modelagem do tráfego na Internet, topologias, e protocolos.

No item 2, será abordado o papel da simulação com uma comparação entre as outras técnicas, medição, experimento e análise.

As propriedades da Internet em face de suas dificuldades são comentadas no item 3 e no item 4 tais dificuldades são discutidas.

O item 5 apresenta algumas estratégias para lidar com as dificuldades expostas neste trabalho.

O NS (Network Simulator), um simulador de rede que é muito utilizado no meio acadêmico, além de ser gratuito, inclusive seu código fonte.

Finalmente no item 7, discutimos as conclusões de Floyd & Paxson e de outros trabalhos.

2 – Simulação

O comportamento da Internet é pesquisado utilizando quatro técnicas conhecidas como: medição, experimento, simulação e análise. Todas são necessárias, embora cada uma delas tenha os seus benefícios e problemas.

A medição e o experimento oferecem meios para explorar o mundo real, enquanto a simulação e a análise são restritas a explorar modelos de mundo abstrato. Em algumas áreas a interpolação entre essas quatro técnicas pode ser óbvia, mas a pesquisa na Internet acrescenta alguns pontos incomuns, em face da grande escala e da rapidez de evolução da “Internet Global”.

Medição é o ponto chave para uma verificação da realidade na Internet atual. Serve frequentemente para desafiar nossas suposições implícitas ([3], sec. II). Medições podem ser usadas, para tentar inferir o tamanho da Internet (redes e usuários), assim como padrões de tráfego, protocolos mais utilizados e volume de tráfego. É possível obter informações sobre o tamanho dos pacotes, quantidade de roteadores entre pontos distintos na Internet, além de parâmetros de desempenho, como atraso fim a fim, perda de pacotes e vazão. Como exemplo, a medição feita no backbone acadêmico do Rio de Janeiro pela CEO/RedeRio- Coordenação de Engenharia de Operações da RedeRio de Computadores:

- Uma das principais tarefas do administrador de uma rede metropolitana é a monitoração de anomalias no comportamento do tráfego nos roteadores e comutadores (switches). Essas podem se mostrar como mudanças de configuração dos equipamentos, desvio abrupto nos dados armazenados (caracterizando rajadas e abusos). O reconhecimento e a identificação desses comportamentos anômalos são geralmente baseados em métodos aperfeiçoados durante anos de experiência em gerenciamento de redes.

O sistema utiliza a técnica de “FlowScan” [6] que é um programa de código aberto, para processar e analisar o fluxo de dados na rede. O “FlowScan” processa os dados coletados pelo Cflowd [7] que comunica com os roteadores da rede, através do pacote Netflow [8] da empresa americana CISCO (modelo de 5 variáveis presentes na rede). Outro pacote em uso é o Flow-Tools [9]. Com a funcionalidade dos programas instalados e o desenvolvimento de uma ferramenta para integrá-los é possível monitorar vários tipos de anomalias no backbone [10]. Na figura 1 é observado o fluxo de tráfego medido em Mbits por segundo, em intervalos de 5 minutos, no roteador de borda do backbone.

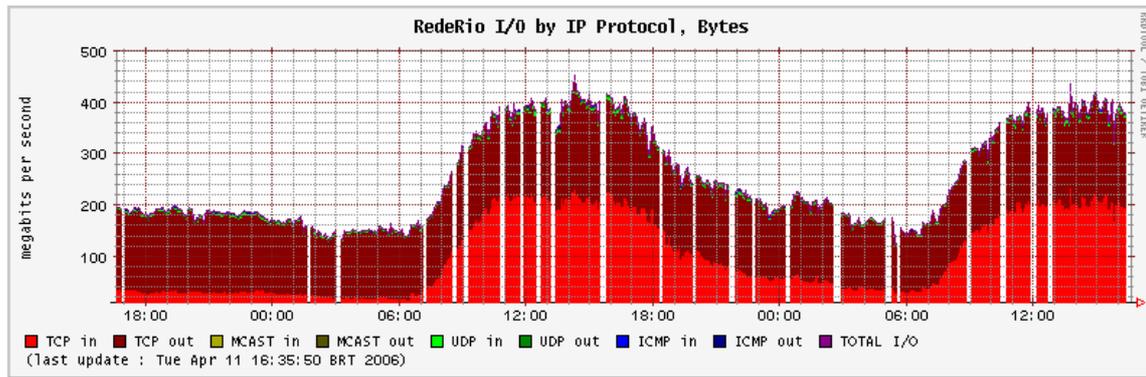


Figura 1: Gráfico do tráfego em Mbits/s do roteador de borda da RR em 11/04/2006

Experimento é uma técnica importante para conhecer e avaliar questões de implementação de aplicações e protocolos. O experimento pode ser realizado com novos protocolos/aplicações na Internet atual, ou então em ambientes de teste(laboratórios), onde se pode ter um maior controle sobre os elementos a serem estudados.

Entretanto, medição e experimento são técnicas limitadas somente a explorar a Internet atual. Elas não podem ser usadas para explorar diferentes arquiteturas possíveis num cenário futuro da Internet. Além disso, nem sempre é possível ter acesso a redes e equipamentos, e essas experiências práticas geralmente têm um custo elevado.

Um problema que acontece em pesquisa na Internet e que não aparece na maioria das áreas, é a possibilidade de desastre, ou seja, depois de uma nova aplicação ter sido testada e implementada, no mundo real multiplica-se devido a utilidade básica da nova funcionalidade. Um exemplo é o protocolo http que é utilizado na *World Wide Web*.

Análise possibilita explorar um modelo da Internet sobre o qual se tem controle completo. Através de modelos matemáticos imensamente simplificados é possível obter resultados rapidamente. Por isso, é também uma técnica barata, pois em princípio não necessita de equipamentos especiais. O risco de confiar em resultados analíticos é que os pressupostos e simplificações são tantos que o comportamento original da Internet pode ser perdido. Obviamente, trabalhar com análise requer um forte embasamento matemático.

Simulações são complementos para a análise, não para provar se a análise foi correta, mas para explorar cenários complicados que seriam difíceis ou impossíveis de analisar. Um importante papel da simulação é desenvolver a intuição das pessoas sobre o comportamento de aplicações, protocolos e tecnologias de rede, principalmente o controle de congestionamento do TCP.

No artigo de Floyd & Paxson [3], o argumento desenvolvido é que devido a heterogeneidade e as rápidas mudanças na Internet, não existe um único conjunto simulado de cenário que seja suficiente para demonstrar que um protocolo ou sistema proposto terá bom desempenho no futuro na Internet.

A simulação tem um papel limitado no exame de determinados aspectos de mudanças propostas ou no comportamento da Internet, e fica a nosso cargo o entendimento da dinâmica subjacente.

Simulação como uma ferramenta de pesquisa em redes tem seus perigos e armadilhas. Pode haver uma considerável dificuldade em verificar se o simulador de fato executa o modelo pretendido. É geralmente mais fácil verificar a exatidão de uma análise matemática do que verificar a exatidão da execução do programa de um modelo subjacente extensivo e complexo.

Existem casos de haver extrapolações a partir de resultados de simulação que freqüentemente se mostram incorretas. Por exemplo, dadas as limitações de topologias e cenários avaliados em simulações, não é correto utilizar simulação para concluir que determinada técnica A tem um desempenho 23% melhor que a técnica B quando ambas são utilizadas na Internet. O motivo é que pequenas modificações na topologia, nos cenários, nos parâmetros ou mesmo no nível de detalhe que o modelo está implementado no simulador podem produzir resultados surpreendentemente contrários a essa afirmação. Um erro comum e muito freqüente é optar pelas situações nas quais a proposta que está sendo defendida apresenta um desempenho melhor que as outras propostas e omitir as outras situações. Por isso, é importante fazer um bom planejamento de simulação e realizar análises estatísticas sólidas.

Se as simulações são utilizadas para obter resultados quantitativos ou explorar relacionamentos gerais entre parâmetros de rede e dinâmicas de rede, simulações são mais úteis se outros pesquisadores puderem confirmar que as mudanças ligeiras no cenário da rede não mudam significativamente os resultados, e que os resultados da simulação não são realmente devido aos erros na execução do simulador. O melhor caminho para validar a simulação é que os pesquisadores disponibilizem seus simuladores e certificados publicamente, de modo que outros pesquisadores possam facilmente verificar se existe efeito na mudança de suposições subjacentes do cenário da rede. Por esse motivo, uma das recomendações do **1999 DARPA/NIST [NSVW99] Network Simulation Validation Workshop [11]** é que os pesquisadores façam públicos os seus certificados de simulação.

3 – Propriedades da Internet

- Internet, um imenso alvo móvel, Floyd & Paxson.

A Internet tem diversas propriedades-chaves que dificultam bastante a sua caracterização tanto quanto a simulação. A princípio seu grande sucesso se deve em grande parte a funcionalidade principal da arquitetura do protocolo Internet(IP) que é unificar as diversas tecnologias de rede e os domínios administrativos. IP permite uma vastidão de redes diferentes administradas por vastas políticas diferentes funcionar sem emendas. Embora o IP mascare

essas diferenças pela perspectiva do usuário isso não os faz desistir. Em vista da diversidade, o IP consegue uma conectividade uniforme, mas não um comportamento uniforme. Certamente o maior sucesso do IP é unificar redes diversas, mas duro é entender o comportamento de uma grande rede IP.

Uma segunda propriedade da Internet é o seu imenso tamanho. Em 2000 estimavam 99,8 milhões de computadores. Seu tamanho traz dois problemas: um é a vasta heterogeneidade, e o outro, o do escalonamento – funciona perfeitamente com pequeno número de computadores, mas não em grande escala.

A terceira propriedade é devido às mudanças drásticas. Como exemplo, o crescimento do número de computadores na Internet: em 1997 eram 15 milhões e em 2000 eram 100 milhões; isso representa um crescimento anual de 60%. Esse crescimento exagerado nos leva a uma pergunta: Qual será o tamanho em mais 5 anos, ou 10 anos?

Infelizmente, o crescimento em algum tempo não é o único responsável de a Internet ser um alvo móvel. O que assumimos deve ser sólido, pois as propriedades estatísticas podem ser modificadas em um breve intervalo de tempo. Por exemplo, em outubro de 1992, o tamanho médio de uma conexão FTP (transferência de arquivos) observada no LBNL era de 4500 B, e cinco meses depois foi de 2100 B (menos da metade do que foi observado antes). A média era considerada uma estatística altamente firme. Repetindo a mesma análise 7 anos depois, em março de 1998, o tamanho médio da conexão foi de 10900 B, enquanto 9 meses mais tarde caiu pela metade, 5600B.

Assim devemos ser cautelosos em assumir que a observação feita em um determinado espaço de tempo tem a mesma propriedade em outro espaço de tempo.

Em resumo: a diversidade técnica e administrativa da Internet, reforçada pelo crescimento em tempo total, e as imensas variações temporais quanto as aplicações utilizadas e em que momento, levam as imensas dificuldades de tentar simulá-la com a meta de obter resultados gerais.

4 – Dificuldades de modelar a Internet

4.1 – Heterogeneidade

Heterogeneidade é a propriedade chave na difícil tarefa de modelagem e simulação da Internet. Ela não deve ser ignorada sob pena de deteriorar os resultados da simulação.

Topologia e Link:

Uma pergunta básica para uma simulação de rede é que topologia usar para a rede que está sendo simulada, especificamente como os computadores estão conectados (direta ou indiretamente) com cada outro, e as propriedades dos enlaces que provêm as interconexões.

Infelizmente a topologia da Internet é difícil de caracterizar. Primeiro, está sempre mudando. Segundo, a topologia é projetada por empresas concorrentes, nem todos dispostos a fornecer informações topológicas. Como não existe uma topologia “típica” da Internet, a simulação de protocolos e tecnologias cujo desempenho depende de questões topológicas pode no máximo chegar a conclusões locais. Não é possível extrapolar esses resultados para a Internet global.

Embora as propriedades de diferentes tipos de conexões usadas na rede sejam geralmente conhecidas, elas são em grande escala. podem variar de linhas antigas de 9600 Kbps para até 10 Gbps em algumas redes avançadas de pesquisa.

Alguns exemplos de enlaces discrepantes:

- modem x link de fibra ótica
- rede com fio (*Traverse copper, glass wire*) x rede sem fio (radio, infravermelho)
- ponto a ponto x *broadcast*(multi-acesso)
- conexão terrestre x conexão de satélite

As características de atraso e perda de pacotes também variam consideravelmente.

Finalmente, as rotas são na maioria assimétricas, como a rota do computador A para o computador B através da rede difere em *hops* (pulos) da rota reversa de B para A. Roteamento assimétrico pode levar a assimetria em tráfegos padrões tanto quanto a latência e a largura de banda. Uma interessante faceta do roteamento assimétrico é que surge freqüentemente e somente em topologias grandes. Isso provê um bom exemplo de como o escalonamento pode levar a prevenção de problemas.

Protocolos Diferentes:

Uma vez que todos esses problemas quanto aos protocolos e aos enlaces foram expostos, o responsável de conduzir uma simulação terá o desafio de especificar os protocolos para o estudo. Em alguns casos, versões simples de protocolos relevantes da Internet devem funcionar bem. Mas para estudos que são sensíveis aos detalhes dos protocolos, pesquisadores e engenheiros terão uma dura tarefa. Enquanto, conceitualmente a Internet utiliza um conjunto de protocolos único, na realidade cada protocolo foi implementado por muitas comunidades diferentes, freqüentemente com características e problemas diferentes.

Um exemplo é o TCP (Transmission Control Protocol) que sofreu mudanças, de grande porte, em sua evolução. O TCP tinha em 2001, aproximadamente 400 implementações e versões. As implementações para controle de congestionamento, Reno e Tahoe não são mais dominantes na família TCP, e foram substituídas New Reno e SACK.

Enfim, fica a cargo dos pesquisadores definirem quais as características e peculiaridades do mundo-real devem ser incluídas em seu estudo, e o que pode ser ignorado. Aplicações diferentes têm características principais diferentes, e essas características podem variar consideravelmente de site para site.

Geração de Tráfego:

A geração de tráfego também representa um dos grandes problemas para modelar e simular a Internet. Em primeiro lugar, deve-se utilizar a proporção correta de tráfego de cada tipo de protocolo. Esses dados podem ser obtidos através de medições, mas são sempre específicos de alguma localidade. Além disso, cada aplicação e/ou protocolo gera tráfego de uma maneira diferente.

A questão básica é como introduzir diferentes origens de tráfego na simulação, e manter o controle do congestionamento fim a fim. O gerador de tráfego *Trace-Driven* (utilização de resultados da análise dos *traces* do tráfego atual) aparece como uma solução. Se pudermos coletar vários e suficientes *traces*, a princípio podemos capturar toda a diversidade. O problema é a adaptação da taxa do controle de congestionamento fim a fim causar uma forma, pois em outro contexto esse modelo não servirá, porque cada conexão apresenta um comportamento diferente.

Como exemplo: uma conexão observada num enlace de alta velocidade sem carga pode enviar pacotes a uma taxa muito mais baixa que o link pode sustentar porque em algum lugar pelo caminho, os recursos insuficientes estão disponíveis.

A solução é a simulação *Trace-Driven* nível-fonte ao invés de *Trace-Driven* nível-pacote, porque o volume dos dados e o padrão da aplicação-nível não são modelados pela propriedade atual da rede.

Floyd & Paxson notaram que nem todas as fontes podem ser caracterizadas com confiança pelos *traces* do tráfego. Por exemplo, usuários remotos de *login* que enfrentam pesado congestionamento podem terminar suas sessões prematuramente, ou podem obter modesta saída ao inserirem comandos. Um caso geral de exceção é de aplicações inerentemente adaptáveis, como algumas formas de vídeo na Internet. Estas aplicações têm suas características de nível de pacote modeladas pelo tráfego atual tanto quanto o comportamento de seu nível de aplicação. Por exemplo, apesar do nível de congestionamento atual simplesmente determinar a taxa de transmissão para um tamanho fixo de dados de vídeo, pode ser determinado o conteúdo ou nível de detalhe para o vídeo transmitido.

O último dimensionamento para a geração de tráfego seria descobrir a que nível o tráfego congestionaria o enlace da rede. Virtualmente, todos os graus de congestionamento, incluindo nenhum, são observados na Internet com probabilidade significativa.

Estimar a evolução futura do congestionamento na Internet é tão difícil quanto caracterizar o nível de congestionamento na Internet global em determinado período.

Enquanto é impossível definir um tráfego misto típico na Internet ou um nível de congestionamento típico, nós podemos utilizar a simulação para

explorar o comportamento da rede em função do tráfego, da topologia, das propriedades do enlace, e etc. O ponto crucial é que devemos ter em mente que devemos considerar um espectro de cenários ao invés de um cenário particular. Mas isso aumenta muito o trabalho.

Utilizar modelos fonte de conexões individuais para gerar tráfego agregado para simulações pode também apresentar escalada de questões. Simular tráfego agregado para cada fonte individual fica caro em termos de tempo de processamento.

4.2 - Dimensão

A Internet é muito grande, sob qualquer parâmetro que se queira observar. Seu tamanho pode ser medido em número de usuários, redes, computadores, interconexões, tráfego, mensagens de correio eletrônico etc.

Para responder a pergunta de Floyd em 2000, a figura 2 mostra o gráfico do crescimento na Internet em número de hosts [12] no intervalo de julho de 1994 a julho de 2005, onde se observa um crescimento exponencial.

O tamanho da Internet traz duas dificuldades. A primeira é que se apenas uma pequena parcela dos computadores apresentarem um comportamento atípico, difícil de ser capturado em simulações, mesmo assim a Internet pode ter milhares desses computadores. A segunda dificuldade se refere ao problema de escala. Alguns protocolos ou mecanismos de rede que funcionam bem em redes pequenas, médias e até grandes (milhares de computadores), podem se tornar impraticáveis quando a rede for várias ordens de magnitude maior que isso (a Internet atual). Isso significa que mesmo estudos de simulação bem planejados e com topologias grandes, podem produzir conclusões incorretas, quando extrapoladas para a Internet.

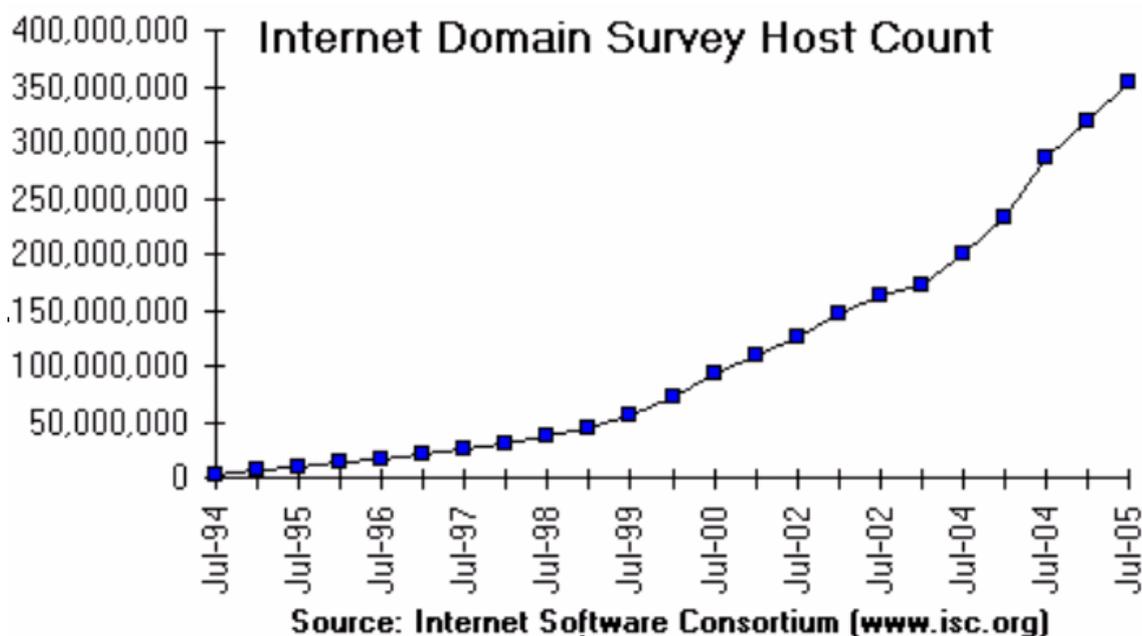


Figura 2: Gráfico do número de *hosts* na Internet

4.3 – Mudanças drásticas

Outra grande dificuldade para simular a Internet são mudanças drásticas que ocorrem de maneira rápida e imprevisível. Muitos estudos revelam resultados surpreendentes, como crescimento súbito do tráfego em determinada situação e depois uma diminuição também expressiva a níveis semelhantes aos do início do crescimento. Um exemplo de crescimento imprevisível foi o aparecimento de aplicações de transferência de arquivos MP3 (como Napster e Gnutella), ocorrido em 1999. Em alguns casos, o crescimento do tráfego foi tão intenso que muitos administradores de rede limitaram a quantidade de tráfego gerado por essas aplicações. Atualmente, essas aplicações já não parecem estar com o mesmo nível de crescimento. Até mesmo houve uma queda, com os problemas enfrentados pelo Napster.

Algumas áreas que podem apresentar mudanças imprevisíveis são: estruturas de tarifação, escalonamento de pacotes, tecnologias de redes sem fio, dispositivos móveis com menor poder computacional, multicast nativo, qualidade de serviço (QoS) e *cache* de Web. A implicação dessas mudanças na simulação é que previsões futuras baseadas em simulações podem ficar totalmente inválidas com uma mudança brusca de comportamento da Internet.

5 – Estratégias

O foco desse trabalho é sobre os fatores que fazem a simulação na Internet exigir um difícil esforço. Nesta parte, discutimos algumas estratégias para lidar com essas dificuldades.

Busca por Invariantes:

Invariantes são algumas facetas do comportamento que foi mostrada empiricamente para suportar uma grande variedade de ambientes. O projeto dos sistemas de telecomunicações foi construído em cima de propriedades invariantes, como: características de tráfego, chamada de processo de chegada, duração de sessões, e etc.

Em relação às propriedades da Internet a invariância tem recebido pouca atenção, mas para Floyd & Vaxson isso não foi conduzido sistematicamente. Relacionamos a seguir alguns candidatos promissores a invariantes:

- **Padrões diurnos de atividade** – há mais de 30 anos que o padrão de atividades na rede segue um padrão diário. Com relação as pessoas, as atividades aumentam a partir das 8:00 horas da manhã, reduzindo na hora do almoço, retornando após 1:00 hora da tarde, e diminuindo na hora que encerra o expediente comercial, ou seja, 5:00 horas da tarde. O padrão também aponta um aumento as 8:00

horas da noite indo até meia noite. No início, supunha-se que essa atividade noturna era o efeito de *hackers* noturnos, mas agora percebe-se que são atividades de usuários caseiros.

Uma invariante relacionada é a diminuição das atividades nos fins de semana e férias. Existem variações significantes nos padrões diurnos, tais como:

1 – padrões diferentes para protocolos diferentes (NNTP);

2 – padrões diferentes para mesmos protocolos, como trabalhos relacionados com navegar na Web versus laser relacionado a navegar na Web por horas seguidas;

3 – Efeitos geográficos devido a comunicação através da zona de tempo;

- **Auto-similaridade** – as correlações de período mais longo na chegada do pacote vistas no tráfego agregado na Internet são bem descritas nos termos de processos de auto-similaridade (ou fractal).
- **Chegadas de sessão de Poison** – chegadas de sessão de um usuário da rede são descritas utilizando processos de Poison. Uma chegada de sessão de um usuário corresponde ao tempo que alguém decide utilizar a rede para uma tarefa. Exemplos são conexões remotas, como o início do diálogo da transferência de um arquivo, ou como o início de uma sessão de navegação na Web. A invariância se dá quando existem múltiplas conexões na rede comprometendo cada sessão. Isso não se comporta como Poison, e sim uma invariante.
- **Tamanhos de log-normal de conexões** – log-normal é a distribuição logarítmica dos tamanhos ou durações aproximada com uma distribuição Gaussiana. Mas em alguns casos nenhum modelo analítico ou empírico se ajusta bem, devido a grande variação nas características das conexões de *site* para *site* e sobre o tempo.
- **Distribuições *heavy-tailed* (também conhecida como distribuição de lei da potência)** – quando caracterizamos distribuições em associação com atividade na rede é esperado encontrar *heavy-tailed*. Entende-se por *heavy-tailed* uma distribuição de Pareto com forma de parâmetro $\alpha < 2$. É evidente que *heavy tails* é comum, incluindo tempo de CPU utilizada por processos Unix, tamanho de arquivos Unix, *frames* de vídeo comprimidos, rajadas da ethernet, etc.
- **Distribuição de invariância pela geração de pacotes via Telnet** - o padrão de pacotes gerados na rede por usuários digitando num teclado (utilizando uma aplicação Telnet) tem uma distribuição de invariância.
- **Características de invariância da topologia global** – enquanto muitas das propriedades da topologia global são prováveis de mudar

drasticamente ao longo do tempo, existirão algumas características invariantes da topologia global, sabendo que a terra é dividida em dois continentes, e que a velocidade da luz não se altera. Em outras palavras, existem 5850 km de New York a Paris, e um sinal levará 2 ms entre esse dois pontos.

Explorando o parâmetro do espaço:

A segunda estratégia para lidar com a grande heterogeneidade e as mudanças na arquitetura da Internet é explorar o comportamento da rede como uma função de parâmetros modificados.

Variáveis relevantes que podem mudar são: protocolos específicos, gerenciamento de fila do roteador ou tabela de pacotes, topologias de rede e propriedades de enlace, ou tráfego misto. Uma regra de manuseio é considerar ordens de magnitude em limites de parâmetros.

Uma armadilha deve ser evitada, pois a Internet tem mecanismos não lineares de alimentação que podem causar erros pelos artefatos de simulação não presentes no mundo real. Uma correta interpretação dos resultados da simulação é a solução.

6 – Simuladores

Vários simuladores surgiram nos últimos anos para permitir a avaliação de cenários de rede antes da sua implantação. Entre os simuladores mais utilizados estão OPNET, Parsec, SSF e *ns*. Os quatro simuladores têm características semelhantes, tais como: utilizam simulação discreta baseada em eventos no nível de pacotes e modelam uma grande quantidade de protocolos da família TCP/IP. Porém, cada simulador tem um foco diferente.

O simulador de redes NS

O NS (Network Simulator-2) [13] é um simulador baseado em eventos discretos e orientado a objetos para a simulação de redes. O objetivo do NS é proporcionar um ambiente para o desenvolvimento de pesquisas em torno dos protocolos que constituem a Internet, isto é, que utilizam a pilha TCP/IP, tanto no contexto das redes fixas quanto móveis, com fio e sem fio (redes locais e de satélite).

NS é um simulador multiprotocolo que implementa algoritmos de roteamento *unicast* e *multicast*, transporte e sessão de protocolos, reservas e serviços integrados, e protocolos de nível de aplicação, como o http.

O NS tem sua origem a partir do simulador REAL (*Realistic and Large*) [23], que por sua vez derivou do NEST (*Network Simulation Testbed*) da universidade de Columbia, NY. REAL foi desenvolvido por S. Keshav da Universidade de Cornell com o intuito de ser uma ferramenta para o estudo do comportamento dinâmico do controle de fluxo e congestionamento em redes

comutadas por pacotes. Posteriormente, o NS surgiu como uma variante do REAL e em 1995 seu desenvolvimento foi suportado como parte do projeto VINT (*Virtual InterNetwork Testbed*) (netweb.usc.edu/vint). Atualmente seu desenvolvimento e distribuição são mantidos pelo ISI (*Information Sciences Institute* - www.isi.edu), financiado pela DARPA e NSF. A distribuição do NS é gratuita, inclusive o código fonte, que pode ser alterado para refletir a pesquisa que está sendo desenvolvida [13].

O NS foi desenvolvido (portado) para várias plataformas computacionais, sendo possível instalar o pacote em vários sabores Unix como, FreeBSD, Linux, SunOS e Solaris, além da plataforma Windows. A biblioteca de protocolos e mecanismos implementados no *ns* é bastante vasta, abrangendo implementação dos protocolos TCP, UDP, IP além de disciplinas de serviços, como, WFQ (*Weight Fair Queueing*), protocolos para redes móveis, como o IP móvel, tecnologias de redes sem fio locais e de longa distância, como, 802.11, Bluetooth e GPRS (*General Packet Radio Service*). Alguns destes não são distribuídos diretamente no pacote do NS, sendo contribuições disponibilizadas como *patches* na página que hospeda as informações sobre o NS e que podem ser baixadas e adicionadas ao módulo básico através de recompilação do núcleo do simulador. O NS fornece também bibliotecas de funções para a geração de alguns tipos de tráfego como: CBR (*Constant Bit Rate*) utilizado para simular tráfego constante e voz, ON-OFF para tráfego em rajada e voz comprimida, FTP para gerar tráfego correspondente a aplicações de transferência de arquivos e VBR (*Variable Bit Rate*) para tráfego com taxa de dados variável. Além das bibliotecas com os módulos específicos de protocolos, tecnologias e geração de tráfego, o NS possui funções específicas de simulação e geração de números aleatórios.

Vantagens do NS

- Disponibilidade de um simulador padrão para a comunidade acadêmica e científica;
- Simulador de código aberto e gratuito;
- Boa infra-estrutura para desenvolver novos protocolos;
- Grande quantidade de protocolos e tecnologias existentes (*multicast*, redes *ad-hoc*, MPLS (*Multiprotocol Label Switching*), redes de satélites etc);
- Oportunidade para estudar interações de protocolos em um ambiente controlado.

Limitações do NS

O NS tem algumas limitações que são desvantagens em alguns casos. O fato de ter sido originalmente desenvolvido para o estudo da Internet, especificamente, de redes IP, limita sua aplicação para o estudo de protocolos/tecnologias sub-IP. Outra limitação é a curva de aprendizado difícil e a falta de documentação adequada para iniciantes. O desenvolvimento de modelos simples requer conhecimento em Otcl que não é complexa, mas pouco conhecida. À medida que se necessita desenvolver novos protocolos ou mecanismos não presentes no NS, deve-se ter não apenas intimidade com o

desenvolvimento em C++, mas desvendar a estrutura de dados do NS e seu esquema de hierarquias de classes para saber de onde novas classes podem ser derivadas e encaixadas, o que não é uma tarefa simples para um iniciante no NS.

Apesar da distribuição do NS possuir uma versão para a plataforma Windows, preferencialmente, adota-se a plataforma de desenvolvimento FreeBSD/Linux, na qual o simulador apresenta um funcionamento mais robusto, pelo fato de ser a plataforma nativa de desenvolvimento do NS.

7 – Conclusão

A Internet é formada por uma grande diversidade de redes interconectadas pelo protocolo IP. Compreender o seu funcionamento, prever o seu crescimento futuro e propor novos protocolos e mecanismos são atividades complexas, mas que podem ser auxiliadas em grande parte com a realização de simulações adequadamente conduzidas. Simulação é uma técnica que pode ser utilizada tanto para obter resultados de pesquisa sobre a Internet quanto para auxiliar a compreensão de aspectos relacionados aos seus protocolos e mecanismos internos.

Embora existam muitas dificuldades para a aplicação de simuladores em estruturas de rede, sua utilização não é desestimulada, muito ao contrário, deve ser estudada pelos pesquisadores e divulgada para que a comunidade possa interagir com os trabalhos e conseqüentemente depurar erros e acertos, evoluindo para a melhora dos simuladores de rede.

O trabalho de Floyd & Paxson data de 2001, e a Internet já cresceu muito, inclusive seus problemas e suas funcionalidades. Mas uma coisa que eles enfocam é a meticulosidade dos resultados das simulações. Pesquisadores devem ter grande cuidado na interpretação dos resultados das simulações e na extração de suas conclusões. Um pesquisador utilizando simulação deve também confiar em outras ferramentas, quando possível, que incluam medições, experimentos, e análises, tão bem como a intuição e um bom julgamento.

Um ponto pouco explorado no trabalho de Floyd & Vaxson foi o tráfego gerado por ataques de robôs na Internet, até porque em 2001 os *worms* (vírus que se atualizam automaticamente, aumentando seu potencial de ataque) não eram presentes. Simular tais ataques é quase que impossível, mas deve ser pensada.

Referências

- [1] - V. Paxson and S. Floyd. **Why We Don't Know How To Simulate The Internet**, In Proceedings of the 1997 Winter Simulation Conference, December 1997.
- [2] - L. Breslau, D. Estrin, K. Fall, S. Floyd, J. Heidemann, A. Helmy, P. Huang, S. McCanne, K. Varadhan, Y. Xu, and H. Yu. **Advances in Network Simulation**, IEEE Computer, 33 (5), pp. 59-67, May 2000
- [3] – S. Floyd and V. Paxson. **Difficulties in simulating the Internet**. IEEE/ACM Transactions on Networking, 9(4):392-403, 2001
- [4] - Kamienski, C.A., Sadok, D., Cavalcanti, D. A. T., Souza, D. M. T., Dias, K. L. **Simulando a Internet: Aplicações na Pesquisa e no Ensino**, 21ª JAI (Jornada de Atualização em Informática), Congresso da SBC, Florianópolis/SC, julho de 2002.
- [5] – Alves Jr., N., Albuquerque, Marcio P., Albuquerque, Marcelo P., Assis, J. T., Oliveira, J. C. C. **Simulações da Rede de Conexões da Internet Brasileira**, CBPF-NT-016/04, dezembro de 2004.
- [6] – FlowScan, <http://net.doit.wisc.edu/~plonka/FlowScan/>
- [7] – Cflowd, <http://www.caida.org/tools/measurement/cflowd/>
- [8] – NetFlow, <http://www.cisco.com/warp/public/732/Tech/nmp/netflow/index.shtml>
- [9] – FlowTools, <http://www.splintered.net/sw/flow-tools/>
- [10] – SysNetNix, <https://sysnetnix.rederio.br>
- [11] - L. Breslau, D. Estrin, K. Fall, S. Floyd, J. Heidemann, A. Kelmy, P. Huang, S. McCanne, K. Varadhan, Y. Xu, and H. Yu, **Advances in network simulation**, IEEE Computer, pp. 59-67, May 2000
- [12] – Internet Growth Charts, <http://www.navigators.com/stats.html>
- [13] - **Network Simulator**, <http://www.isi.edu/nsnam/ns/>
- [14] - **NAM: Network Animator**, <http://www.isi.edu/nsnam/nam/>