Avaliação do Impacto de Falhas na Rede Nacional de Ensino e Pesquisa no Tráfego de um Campus Universitário

Rodrigo Duarte[1] **Alex B. Vieira**[1] **Ítalo Cunha**[2] **Jussara Almeida**[2]

[1]Departamento de Ciência da Computação, Universidade Federal de Juiz de Fora
[2]Departamento de Ciência da Computação, Universidade Federal de Minas Gerais

{rodrigo.duarte, alex.borges}@ufjf.edu.br {cunha, jussara}@dcc.ufmg.br

Abstract. *In this paper we characterize the impact of failures in Brazil's national education and research network on traffic at a client university. In particular, we study the impact of failures on traffic, user, and application behavior. The failures are interesting in that they persist for several hours and impact only international links, so destinations hosted in Brazil remain reachable. Our results show that although failures in international links have negligible impact on the performance of national traffic, users do adapt their behavior to the unavailability of services hosted abroad. For example, entertainment traffic migrates from Facebook to YouTube, which remains reachable during the analyzed failures; and the fraction of interactive traffic gradually decreases during failures, indicating that users may leave the campus early. We also show that asynchronous applications hosted abroad, like Dropbox and SMTP, queue up tasks during the failure and cause traffic bursts when the failure is restored.*

Resumo. *Neste trabalho caracterizamos o impacto de falhas na Rede Nacional de Ensino e Pesquisa (RNP) no tráfego da Universidade Federal de Juiz de Fora. Nós estudamos o impacto das falhas no comportamento do tráfego, dos usuários e das aplicações na rede da universidade. As falhas estudadas são relevantes pois persistem por várias horas e afetam apenas enlaces internacionais da RNP, sem impedir acesso a destinos no Brasil. Nossos resultados mostram que embora falhas nos enlaces internacionais da RNP tenham impacto desprezível no desempenho de conexões nacionais, usuários modificam seu comportamento em função da indisponibilidade de serviços hospedados fora do Brasil. Por exemplo, o tráfego de entretenimento migra do Facebook para o YouTube, que permanece ativo durante as falhas; e a fração de tráfego interativo reduz gradativamente durante a falha, indicando evasão dos usuários da rede. Mostramos também que aplicações assíncronas com servidores fora do Brasil, como Dropbox e SMTP, acumulam tarefas durante a falha e causam rajadas de tráfego quando a falha é restaurada.*

1. Introdução

Falhas de comunicação na Internet podem ser causadas por problemas de *hardware*, como o rompimento de cabos, ou erros de *software*, como configuração inadequada de um roteador. A maioria destas falhas passa despercebida graças às mudanças automáticas de roteamento que as contornam usando rotas alternativas [Markopoulou et al. 2008]. Porém, falhas causadas por erros de configuração ou ausência de rotas alternativas requerem intervenção humana e suas soluções pode levar horas [Kompella et al. 2007].

Embora a literatura seja rica em esforços para *caracterizar* anomalias e falhas na Internet (e.g., [Markopoulou et al. 2008, Turner et al. 2010, Kompella et al. 2007, Zhang et al. 2008, Lakhina et al. 2004]), o *impacto* desses problemas no tráfego e no comportamento dos usuários ainda é pouco conhecido [Markopoulou et al. 2008]. Além disso, o comportamento dos usuários e o tráfego da Internet mudaram ao longo dos anos. O volume de tráfego P2P, antes fração dominante do tráfego total, reduziu frente à popularização da distribuição de vídeo sob demanda via HTTP [Sandvine 2013]. Aplicações interativas, como redes sociais, ferramentas colaborativas de edição de documentos e serviços bancários, são cada vez mais utilizadas pelos usuários de Internet. Várias novas aplicações estão disponíveis na nuvem e só podem ser utilizadas com acesso à rede. Em geral, a maior dependência da Internet para realização de várias tarefas diárias agrava o impacto de falhas de conectividade nos usuários.

Neste trabalho nós investigamos o impacto de falhas em enlaces internacionais no tráfego de um importante campus universitário. Para tal, nós caracterizamos o tráfego capturado do roteador de borda da Universidade Federal de Juiz de Fora (UFJF), conectada à Internet por meio da Rede Nacional de Ensino e Pesquisa (RNP). Nossa base de dados contém sumários de todos os fluxos de entrada e saída da rede da universidade durante o período letivo (entre janeiro e março de 2013), cobrindo mais de 40 TB de dados (seção 3). Em particular, caracterizamos mudanças no *comportamento*—volume, padrões e características do tráfego—dos usuários e aplicações durante períodos em que, sabidamente, houve queda dos enlaces internacionais da RNP. Nossos objetivos específicos são caracterizar o impacto das falhas no comportamento do tráfego (seção 4), dos usuários (seção 5) e das aplicações utilizadas no campus universitário (seção 6). Um dos diferenciais deste trabalho é que as falhas analisadas são parciais: apesar de destinos fora do Brasil terem ficado inacessíveis, destinos no Brasil praticamente não foram afetados.

Nossos resultados mostram que falhas nos enlaces internacionais da RNP não afetam o desempenho do tráfego nacional, possivelmente devido ao bom provisionamento dos enlaces nacionais. Nossa análise indica redução gradativa de tráfego interativo durante as falhas, o que pode ser resultante de usuários deixando o campus prematuramente devido aos problemas de conectividade. Outro padrão identificado é a migração das aplicações utilizadas, particularmente aplicações de redes sociais, de sites indisponíveis (hospedados no exterior) para sites hospedados nacionalmente (e.g., do Facebook para o YouTube). Mostramos também que aplicações assíncronas que executam em plano de fundo, e.g., Dropbox e SMTP, acumulam tarefas durante a falha e causam rajadas de tráfego após a restauração da falha.

Nossos resultados, mesmo limitados a uma universidade, representam mais um passo na direção da melhor compreensão do impacto de falhas no comportamento dos usuários. Acreditamos que os resultados apresentados são particularmente interessantes para a comunidade de pesquisa em redes de computadores e podem motivar mudanças práticas do uso e do gerenciamento dos recursos de rede existentes.

Por exemplo, de acordo com nossos resultados, rajadas de tráfego geradas por aplicações assíncronas podem comprometer o desempenho da rede após restauração de falhas. Assim, o uso de modeladores de tráfego (*traffic shaping*) ou priorização de tráfego pode ser necessário para manter os serviços de rede. Mais ainda, os resultados podem ser motivadores para potenciais mecanismos para redução da evasão dos usuários da rede

durante falhas. Por exemplo, o estabelecimento de parcerias de troca de tráfego com redes que hospedam serviços de produtividade.

2. Detecção, Caracterização e Impacto de Falhas na Internet

Nosso trabalho estuda o impacto que falhas na Internet causam no comportamento do tráfego, dos usuários e das aplicações. Nesta seção contextualizamos nosso trabalho apresentando técnicas de detecção e características de falhas na Internet disponíveis na literatura. Por último discutimos trabalhos relacionados sobre o impacto de falhas.

Detecção. A maioria das falhas são detectadas por equipamentos de rede e automaticamente reportadas a operadores via ferramentas como SNMP e *syslogs* [Turner et al. 2010]. Infelizmente, apenas operadores de rede têm acesso ao equipamento de rede e a essas ferramentas de detecção. Clientes de acesso, usuários finais e operadores de outras redes possuem pouca ou nenhuma visibilidade sobre falhas. Outro problema é que algumas falhas, como as causadas por *bugs* no *software* e erros de configuração, não são reportadas pelo equipamento de rede [Kompella et al. 2007].

Devido à dificuldade de se detectar alguns tipos de falhas, várias ferramentas de monitoramento foram propostas para complementar alarmes de equipamentos de rede. Essas ferramentas de monitoramento correlacionam informações de vários dispositivos na rede e injetam sondas de medição para obter informações mais precisas sobre o estado da rede. Além disso, essas ferramentas frequentemente detectam falhas num ambiente específico, limitando o escopo das falhas e do monitoramento realizado. Por exemplo, o SLAM é uma ferramenta que clientes de acesso à Internet podem utilizar para verificar se seus provedores estão provendo conectividade com a qualidade contratada [Sommers et al. 2007]; o NICE é uma ferramenta para detecção e diagnóstico de falhas intermitentes em *backbones* IP [Mahimkar et al. 2008]; e o NetMedic é uma ferramenta que coleta informações nos servidores e estações de trabalho de uma rede empresarial para detectar falhas [Kandula et al. 2009]. Outra frente para detecção de falhas é a análise do tráfego da rede em busca de anomalias. Estes métodos constroem um modelo do tráfego normal da rede ou enlace e depois detectam desvios deste modelo (e.g., [Lakhina et al. 2004], [Silveira and Diot 2010]).

Neste trabalho não fazemos detecção de falhas pois caracterizamos falhas reportadas pela RNP. Porém, nossos resultados podem servir de base para melhorias em sistemas de detecção de anomalias causadas por falhas parciais.

Caracterização. Trabalhos de caracterização de falhas de rede são menos comuns que os de detecção de falhas porque operadores em geral não publicam detalhes sobre falhas em suas redes. Pesquisadores já caracterizaram falhas na rede da CENIC (*Corporation for Education Network Initiatives in California*) e da Sprint [Turner et al. 2010, Markopoulou et al. 2008]. Nestas redes, dois tipos comuns de falha são resultantes de manutenção programada da rede e falhas intermitentes causadas por *hardware* defeituoso. Estes trabalhos também mostram que nenhum enlace está totalmente livre de falhas mas que alguns enlaces são mais propensos a falhas que outros (e.g., enlaces de tecnologia diferente). Outros trabalhos caracterizaram falhas detectadas com medições ativas na Internet (e.g., [Katz-Bassett et al. 2008, Quan et al. 2013]). Estes trabalhos enviam sondas de medição para vários destinos na Internet e inferem uma falha quando um prefixo antes acessível torna-se inacessível. Estes trabalhos encontram várias falhas dentre os prefixos monitorados, ge-

ralmente em redes menores longe do núcleo da Internet. Por exemplo, [Quan et al. 2013] encontraram que, na média, 0.15% dos prefixos normalmente acessíveis da Internet estão inacessíveis em um dado instante. Em geral, estes trabalhos indicam que a maioria das falhas na Internet duram poucos minutos, mas que poucas falhas de longa duração são responsáveis pela maior parte do tempo sem conectividade (*downtime*).

O foco deste trabalho não é caracterizar falhas, mas estudar o impacto de falhas no tráfego, usuários e aplicações. Ressaltamos que as falhas que estudamos estão entre as poucas falhas de longa duração e impacto prolongado no tráfego.

Impacto. Vários trabalhos caracterizaram o impacto de falhas em mudanças de roteamento na Internet [Feamster et al. 2003, Zhang et al. 2007, Li and Brooks 2011], mostrando que falhas são frequentemente seguidas (e às vezes precedidas) por mensagens BGP. Nosso trabalho é similar a outros trabalhos que tentam estimar o impacto de falhas e problemas de desempenho no comportamento do usuário em sítios Web (e.g., [Stefanov 2008]). Enquanto esses trabalhos mostram que usuários podem desistir de acessar um serviço devido a problemas de desempenho, nosso trabalho estuda mudanças no comportamento do usuário entre vários serviços durante falhas prolongadas. Não conhecemos nenhum trabalho que avalie o impacto de falhas parciais no comportamento do usuário e no tráfego de aplicações. Uma possível explicação para a existência de poucos trabalhos sobre o impacto de falhas é a necessidade de coletar dados em local que tenha visibilidade sobre o tráfego *durante* a falha, o que nem sempre é possível. Por exemplo, [Dainotti et al. 2011] caracterizaram o impacto do bloqueio de tráfego nos enlaces internacionais do Egito, em 2012. Como os autores não tinham dados coletados no Egito, eles usaram pacotes não requisitados recebidos em prefixos IP inativos (*Internet telescopes*) e mensagens BGP disponíveis publicamente. Infelizmente, como a falha no Egito bloqueou tráfego internacional, uma análise de mudança de comportamento do usuário com dados coletados externamente é impossível.

3. Conjunto de Dados de Tráfego e Descrição das Falhas Estudadas

Neste trabalho nós avaliamos o impacto de falhas na Rede Nacional de Ensino e Pesquisa (RNP) em dados trafegados na Universidade Federal de Juiz de Fora (UFJF). A figura 1 apresenta uma visão geral do ambiente de coleta de dados. Instalamos um chaveador (*switch*) entre o roteador de borda e o *firewall* da UFJF para espelhar o tráfego do enlace entre esses dispositivos. O roteador de borda e o *firewall* são responsáveis por rotear e filtrar, respectivamente, todo o tráfego de entrada e de saída do campus. O *firewall* da UFJF também faz tradução de endereços (NAT) para alguns nós da rede interna.

O tráfego espelhado é encaminhado a um servidor de coleta. Devido à quantidade de dados trafegados, aproximadamente 15,5 TB por mês, o servidor sumariza informações sobre o tráfego usando o TSTAT [Finamore et al. 2011]. TSTAT é uma ferramenta de código livre que coleta 111 métricas sobre as conexões, incluindo endereços IP e portos de origem e destino, horários de início e fim, número de pacotes, tráfego total, latência e um identificador da aplicação ou protocolo que criou a conexão. Os pacotes espelhados são descartados após sumarização, preservando a privacidade quanto aos dados dos usuários.

A rede da UFJF integra 22 unidades e provê conectividade para aproximadamente 6.000 computadores, conectados por rede cabeada em laboratórios de pesquisa, escritórios

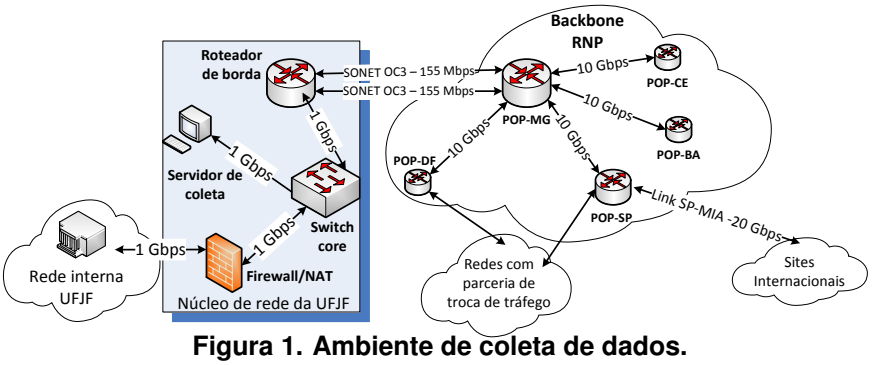


Figura 1. Ambiente de coleta de dados.

de administração, salas de aula e pontos de acesso sem fio. A UFJF possui aproximadamente 19.000 alunos, 1.500 funcionários e 1.400 professores. A coleta realizada não captura o tráfego interno da universidade. Apesar do tráfego interno ser interessante para análise, coletá-lo exigiria uma infraestrutura de coleta significativamente mais extensa, com um servidor de coleta em cada rede local.

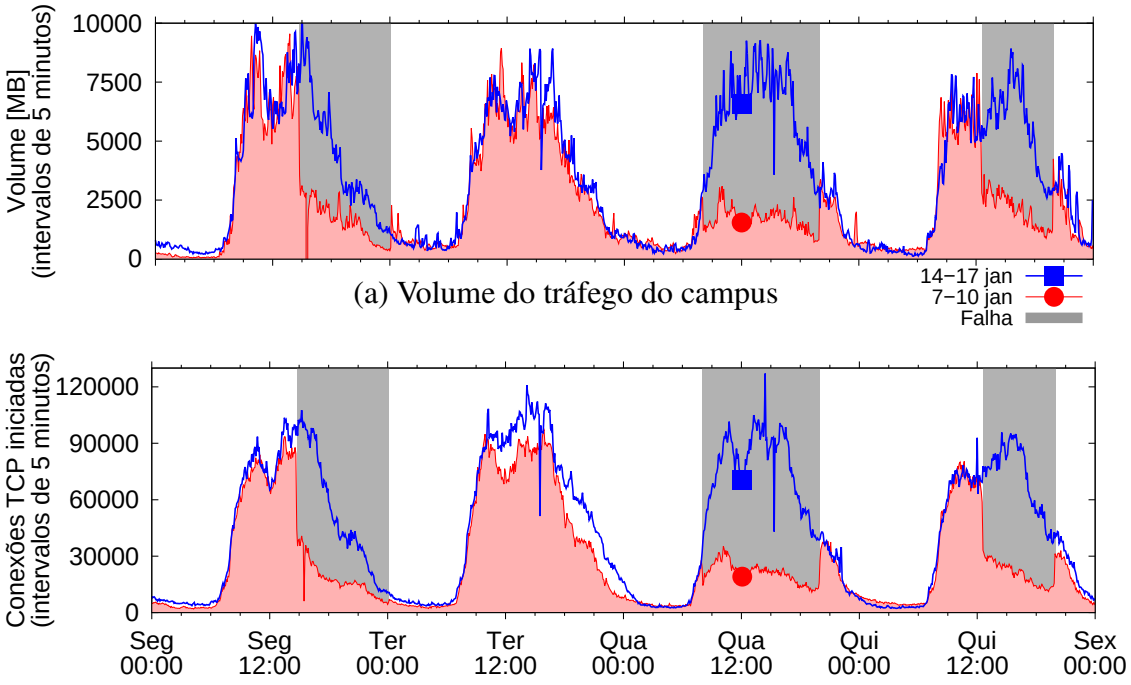
Todo o tráfego de dados da UFJF é enviado ao PoP-MG, em Belo Horizonte, por 2 enlaces ponto-a-ponto OC-3 com total de 310 Mbps de banda. A partir do PoP-MG, o tráfego da UFJF entra na RNP, que encaminha o tráfego ao seu destino. A RNP interliga praticamente todas as instituições públicas de ensino do Brasil, bem como algumas instituições governamentais (e.g., EMBRAPA). A infraestrutura da RNP é gerenciada em colaboração com pontos de troca de tráfego regionais operados por universidades, como o PoP-MG em Belo Horizonte. Empresas e redes comerciais podem se ligar à RNP nos pontos de troca de tráfego regionais, na maioria das vezes, via acordos de troca livre de tráfego (*peering*). O tráfego na RNP destinado a computadores fora do Brasil passa por enlaces internacionais. Atualmente, o mais importante desses enlaces liga São Paulo a Miami e tem banda de 20 Gbps. O tráfego da RNP destinado a empresas e redes conectadas aos pontos de troca de tráfego não utiliza os enlaces internacionais.

Focamos nossas análises em dias de falhas anunciadas pela RNP. Em particular, estudamos falhas parciais onde o acesso à Internet não é totalmente interrompido. De acordo com relatórios publicados pela RNP,[1] nos dias 7, 9 e 10 de janeiro de 2013 ocorreram falhas na infraestrutura de fibra óptica de algumas operadoras de telecomunicação. Essas falhas impossibilitaram o acesso aos enlaces internacionais da RNP. Consequentemente, destinos e serviços hospedados fora do Brasil ficaram inacessíveis, mas destinos conectados aos pontos de troca de tráfego da RNP continuaram acessíveis. Argumentamos que falhas globais têm impacto forte mas simples no tráfego—interrompendo-o por completo—e que falhas parciais são mais interessantes de analisar.

4. Impacto das Falhas no Tráfego

Nesta seção discutimos o impacto das falhas no tráfego da UFJF. Esta discussão servirá de base para compreensão dos resultados mais detalhados nas seções 5 e 6.

[1] `http://www.rnp.br/backbone/weblog/arquivo/arquivo_2013-m01.php`



(a) Volume do tráfego do campus

(b) Número de conexões TCP no campus

Figura 2. Visão geral do tráfego de dados na rede da UFJF.

4.1. Impacto das Falhas no Tráfego Agregado

A figura 2 apresenta uma visão geral do tráfego na rede da UFJF durante dois períodos distintos, cada um cobrindo quatro dias consecutivos. A figura 2(a) mostra o tráfego total, agregado em intervalos de cinco minutos. Como o TSTAT grava apenas o total de bytes trafegados e a duração de cada fluxo, nós distribuímos os bytes de um fluxo uniformemente ao longo de sua duração. A figura 2(b) mostra o número de conexões TCP iniciadas em intervalos de cinco minutos. As linhas azuis, marcadas com um quadrado, mostram o tráfego total e o número de conexões iniciadas no período de 14 a 17 de janeiro de 2013, quando nenhuma falha foi reportada pela RNP. As linhas vermelhas, marcadas com um círculo, mostram o tráfego e o número de conexões iniciadas entre os dias 7 e 10 de janeiro de 2013, quando a RNP reportou três falhas. Ambos períodos cobrem dias de semana de segunda a quinta-feira. As falhas ocorreram de 14:45 do dia 7 às 00:05 do dia 8 de janeiro, de 08:05 à 20:00 do dia 9 de janeiro e de 12:35 à 19:55 no dia 10 de janeiro (horários de verão de Brasília); nós sombreamos estes períodos na figura 2. Note que, devido à greve de 2012 e consequente adaptação do calendário acadêmico, as aulas na UFJF foram retomadas dia 7 de janeiro.

Em geral, o tráfego apresenta o típico padrão de uso diurno, com pico entre 10 e 16 horas e mínimo durante a madrugada. O aumento do tráfego a partir das 7 horas é mais acentuado que sua redução a partir das 18 horas. Isso ocorre devido à existência de cursos noturnos com quantidade de alunos menor que os cursos diurnos. Note que o impacto das falhas no volume total é imediato devido à interrupção do tráfego internacional. O volume de tráfego aumenta imediatamente após a restauração das falhas.

O comportamento do número de conexões TCP iniciadas é qualitativamente similar. Tentativas de conexões a destinos fora do Brasil durante a falha nunca são completadas com sucesso, são marcadas como conexões incompletas pelo TSTAT e não são contabilizadas na figura 2(b). O comportamento do número de conexões TCP ativas (ao

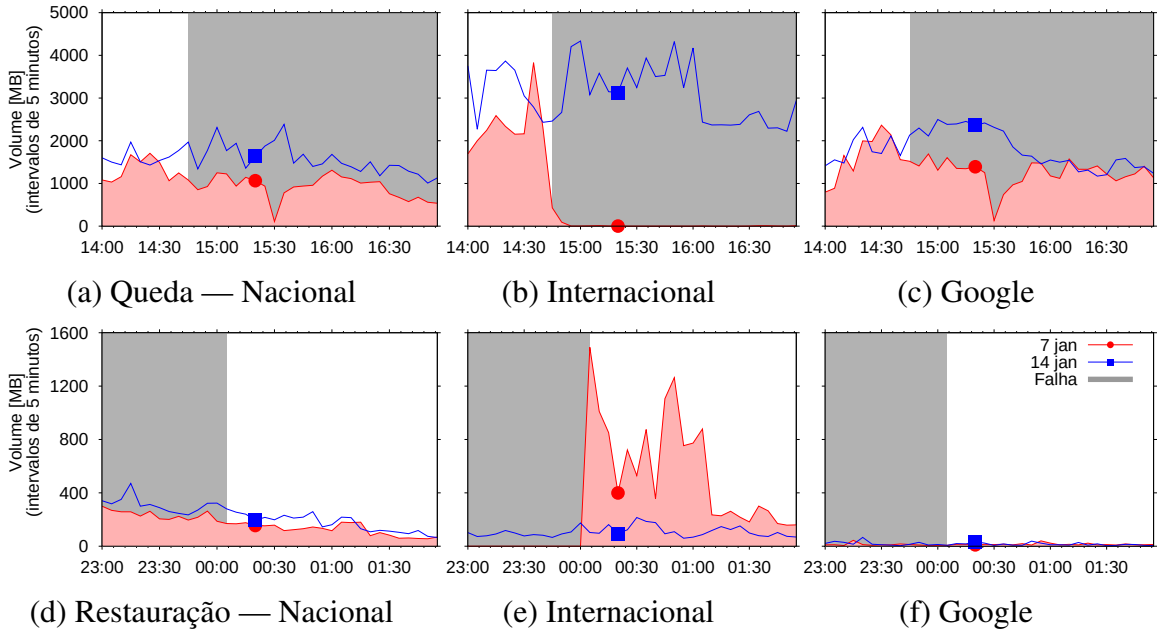


Figura 3. Detalhe do impacto da falha no tráfego durante o início da falha (linha superior) e durante a recuperação da falha (linha inferior) no dia 7 de janeiro.

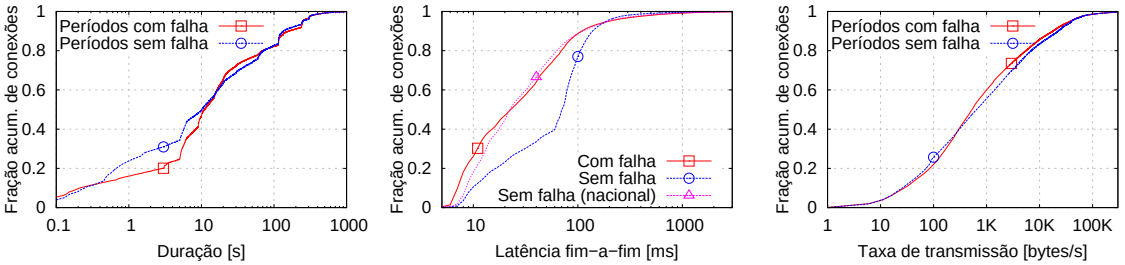
contrário de iniciadas), também é qualitativamente similar pois as conexões internacionais também são interrompidas pelas falhas (não mostrado).

4.2. Impacto das Falhas por Geolocalização dos Destinos

Nós separamos o tráfego total em três conjuntos: tráfego nacional, com destino no Brasil, tráfego internacional, com destino fora do Brasil, e tráfego para serviços do Google. Nós separamos tráfego para serviços do Google porque eles continuam acessíveis durante as falhas através do Ponto de Troca de Tráfego em São Paulo; além disso, grande parte do tráfego do campus é do YouTube (seção 5). Nós classificamos tráfego entre nacional e internacional usando a base de dados livre do MaxMind. Apesar das limitações conhecidas para bases de dados de geolocalização, a precisão é suficiente para a granularidade da nossa classificação [Poese et al. 2011]. Para classificar o tráfego do Google, nós resolvemos os endereços IPs associados a domínios de serviços do Google (e.g., youtube.com, gmail.com, google.com) na UFJF. Nós classificamos como tráfego para o Google qualquer conexão com endereço de origem ou destino nos prefixos referentes a estes IPs nas tabelas de roteamento do PTT-Metro em São Paulo.

A figura 3 mostra a variação dos tráfegos nacional, internacional e para serviços do Google durante períodos de três horas que cobrem o início (figuras 3(a–c)) e o término (figuras 3(d–f)) da falha do dia 7 de janeiro. Para fins de comparação, a figura mostra curvas correpondentes também para o dia 14 de janeiro, mesmo dia da semana mas sem falha reportada. Como na figura 2, nós mostramos o volume em intervalos de cinco minutos e o período de falha está sombreado.

As figuras 3(a) e 3(c) mostram que o início da falha não causa impacto imediato significativo nos tráfegos nacional e para o Google. Porém, o tráfego internacional rapidamente cai para zero. O vale mostrado em ambas as curvas por volta das 15:30 do dia 7 de janeiro foi causado por uma falha local na UFJF (reinício do roteador de borda). No momento de restauração da falha ocorre uma rajada de tráfego internacional (figura 3(e))



(a) Duração (b) Latência fim-a-fim (c) Taxa de transmissão

Figura 4. Comparação do desempenho das conexões TCP durante a falha do dia 7 de janeiro com o mesmo período do dia 14 de janeiro (sem falha).

gerada por aplicações assíncronas, como detalharemos na seção 6. Observamos impacto semelhante no número de conexões TCP (não mostrado).

4.3. Impacto das Falhas em Características e Desempenho de Conexões

Durante períodos de falha, as conexões TCP podem apresentar características diferentes das encontradas em períodos sem falha. A figura 4 mostra a distribuição acumulada da duração, latência fim-a-fim e taxa de transmissão das conexões durante um período que inclui do início à recuperação de uma falha e, para fins comparativos, um período de igual duração sem falhas. Os períodos mostrados são de 14:45 às 00:05 dos dias 7 (linha com quadrado) e 14 (linha com círculo) de janeiro.

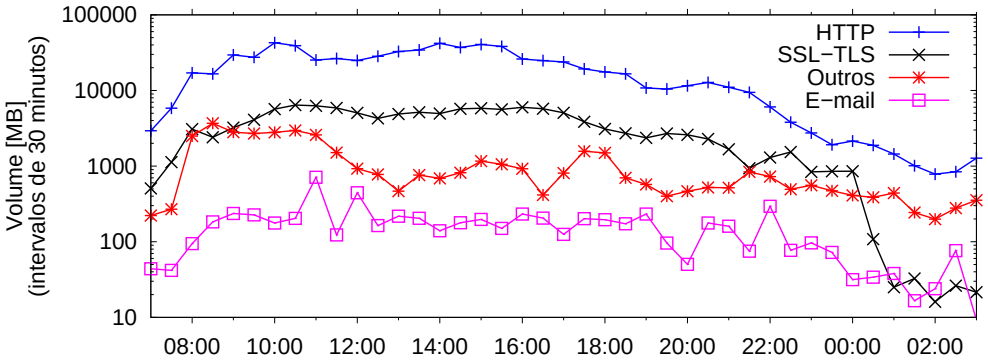
A Figura 4(a) mostra as distribuições acumuladas das durações das conexões TCP durante os períodos considerados. As distribuições são, em geral, similares exceto pela redução (em 54%) da fração de conexões com duração entre 0,5 e 3 segundos durante falhas e pelo aumento (em 16%) da fração de conexões com duração maior que 3 segundos. Como discutiremos na seção 5, uma causa para esse efeito está na mudança de comportamento dos usuários.

A figura 4(b) mostra as distribuições acumuladas da latência fim-a-fim das conexões TCP nos mesmos períodos. A diferença na latência entre os períodos com e sem falha resulta dos destinos acessados durante a falha estarem localizados em redes no Brasil e geograficamente mais próximos. Para uma comparação mais justa, mostramos também a latência fim-a-fim das conexões para destinos no Brasil durante o dia 14 (sem falha, linha marcada com um triângulo). Vemos que a falha não tem impacto na latência fim-a-fim das conexões com destino no Brasil.

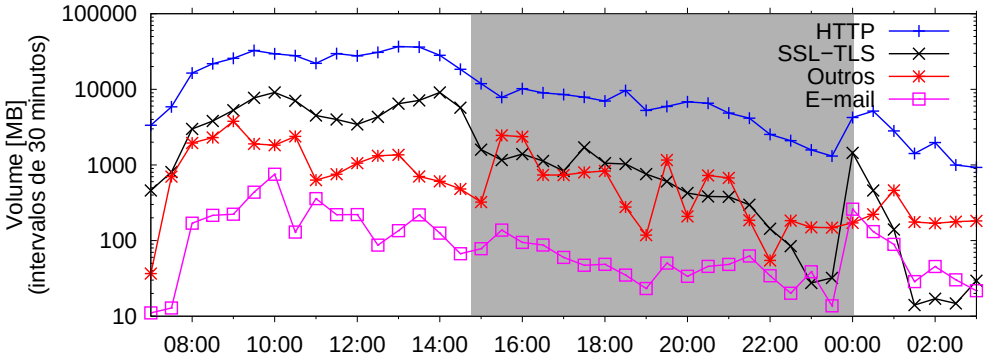
A figura 4(c) mostra que a distribuição acumulada da taxa de transmissão das conexões TCP durante o período considerado não é significativamente impactada pela falha. Isto indica que os enlaces não ficaram sobrecarregados durante a falha. Analisamos também a taxa de perda de pacotes das conexões TCP e não observamos diferenças significativas entre períodos com e sem falhas. Em ambos os casos, a porcentagem de tráfego transferido em retransmissões é menos de 1,5% do tráfego total.

5. Impacto das Falhas no Comportamento dos Usuários

Nesta seção analisamos o impacto das falhas em características do tráfego para inferir modificações no comportamento dos usuários da rede. Discutimos redução da quantidade de tráfego interativo e modificações da mistura de aplicações utilizadas durante falhas.



(a) Dia sem falha, 14 de janeiro



(b) Dia com falha, 7 de janeiro

Figura 5. Volume de tráfego por protocolo.

5.1. Redução de Tráfego Interativo Durante Falhas

A figura 5 mostra o volume total de tráfego de diferentes protocolos, segundo classificação do TSTAT, em intervalos de trinta minutos. A linha "E-mail" combina os protocolos POP, IMAP e SMTP. Todos os demais protocolos são agrupados na linha "Outros".

Nós mostramos na figura 5(a) tráfego normal no dia 14 de janeiro, sem falha, e na figura 5(b) o tráfego no dia 7 de janeiro, com falha. O comportamento da quantidade de conexões TCP por protocolo ao longo do tempo é qualitativamente similar (não mostrado). Apesar de mostrarmos apenas resultados para a falha do dia 7, os resultados apresentados também são válidos para as falhas dos dias 9 e 10 de janeiro.

Durante o período sem falhas, o volume de tráfego de cada protocolo se mantém relativamente estável entre 7:00 e 22:00 horas (figura 5(a)). O volume de tráfego criptografado (protocolo "SSL/TLS") diminui significativamente durante a madrugada. Em períodos de falhas, o volume de tráfego criptografado diminui gradativamente a partir do início da falha (figura 5(b)); compare, por exemplo, a quantidade de tráfego criptografado entre 20:00 e 20:30 dos dias 7 e 14 de janeiro (425 MB e 1.3 GB, respectivamente). Discutiremos os picos de tráfego criptografado e de e-mail após a restauração da falha (figura 5(b)) na seção 6.

Como o tráfego criptografado é resultante primariamente de atividades dos usuários na rede,[2] acreditamos que os *usuários saem do campus ou de suas estações de trabalho*

[2]Por exemplo, um dos destinos com maior quantidade de conexões criptografadas em nossos dados é o

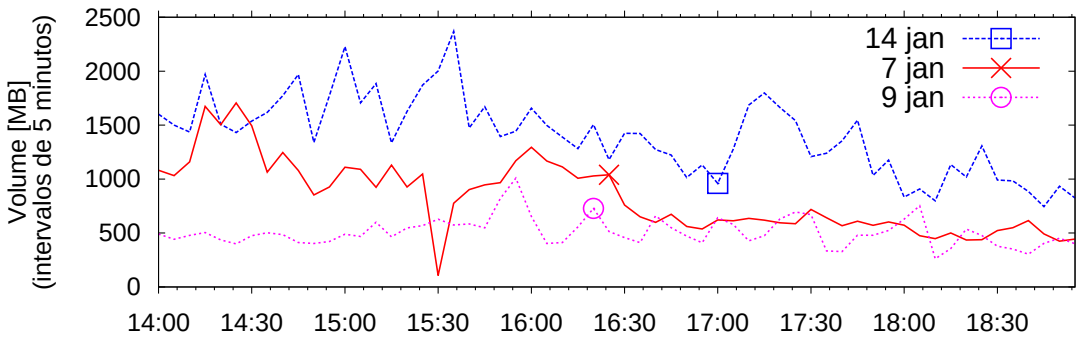


Figura 6. Modificação no volume de tráfego nacional devido às falhas.

prematuramente em função da falha de conectividade. Para testar a hipótese acima nós estimamos o número de usuários pela quantidade de conexões realizadas a serviços do Google (não mostrado). Não consideramos serviços como Twitter e Facebook porque estão hospedados fora do Brasil e, diferentemente de serviços do Google, ficam inacessíveis durante falhas. Às 17:00 horas do dia 7 de janeiro, aproximadamente 2 horas após o início da falha, a quantidade de conexões para serviços do Google era 45% menor que durante o mesmo período no dia 14, sem falha.

A figura 6 compara o volume de tráfego nacional nos dias 7 e 9 de janeiro (com falha) e no dia 14 de janeiro (sem falha). As falhas começaram às 14:45 e 8:05 nos dias 7 e 9 de janeiro, respectivamente (nesta figura não sombreamos as falhas). Vemos que mesmo o tráfego nacional, cujos serviços continuam acessíveis durante a falha, diminui no dia 7; mais um indicativo que usuários deixam o campus. Além disso, no dia 9 o tráfego nacional nunca alcança o volume normal em dias sem falha; indicando que usuários talvez nem vão ao campus ao receber notícia da falha de rede.

5.2. Modificações da Mistura de Aplicações Utilizadas

A tabela 1 compara tráfego das aplicações em um período de falha (entre 14:45 à 00:05 dos dias 7 e 8 de jan.) com período equivalente dos dias 14 e 15 de janeiro. O tráfego foi classificado em aplicações pelo TSTAT. Sumarizamos os resultados agregando aplicações como Twitter, MSN e Flickr em "Social"; MegaUpload, HotFile e RapidShare em "Hospedagem de Arquivos"; BitTorrent e eDonkey em "P2P" e provedores de anúncios em "Propaganda". Para cada um desses conjuntos, mostramos: a fração de suas conexões relativas ao total de conexões no período, a fração de seu tráfego relativo ao tráfego total no período, o volume de tráfego e o volume médio de suas conexões.

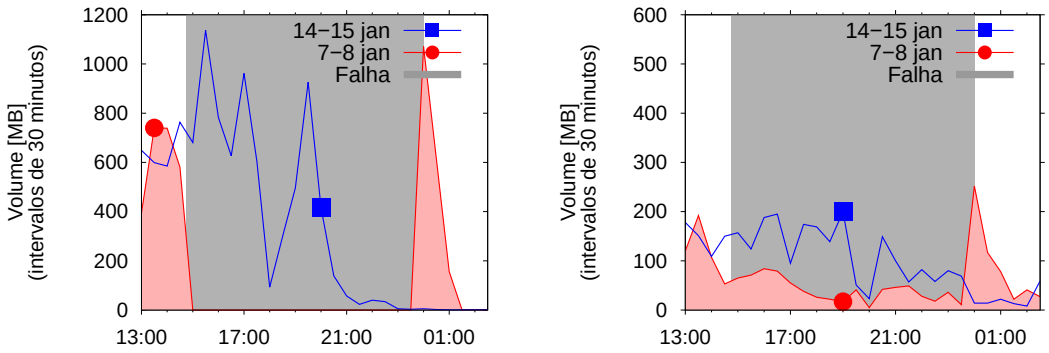
Como o Facebook é hospedado fora do Brasil, seu tráfego é reduzido a zero durante a falha. Em contrapartida, a fração de tráfego e conexões para o YouTube aumenta significativamente. Isso indica adaptabilidade dos usuários e uma migração das conexões de entretenimento para serviços que continuam disponíveis durante a falha. Notamos que a fração de conexões e tráfego nem sempre aumenta para aplicações que continuam ativas durante a falha (e.g., Google Maps, não mostrado).

Notamos que a fração do tráfego para serviços de hospedagem de arquivos é pe-

site da Prefeitura de Juiz de Fora (http://pjf.mg.gov.br).

Tabela 1. Comparação do tráfego de aplicações entre 14:45 e 00:05 dos dias 7 e 8 de janeiro (período de falha) e o mesmo período nos dias 14 e 15 de janeiro (sem falha).

Tipo de Tráfego	% das conexões		% do tráfego		Volume de tráfego (GB)		Volume por conexão (KB)	
	07/01/13	14/01/13	07/01/13	14/01/13	07/01/13	14/01/13	07/01/13	14/01/13
HTTP GET	63,86	44,93	41,03	53,51	317,08	592,85	34,98	48,68
HTTP POST	2,20	2,77527	0,44	1,12	3,41	12,46	10,97	16,57
Youtube	3,06	1,07	39,43	17,53	304,72	194,24	702,29	668,79
Propaganda	2,23	2,15	0,33	0,41	2,59	4,54	8,18	7,79
Social	0,057	0,50	0,03	0,08	0,25	0,89	31,03	6,52
Facebook	0,007	6,46	0	4,21	0	46,65	0,00	26,62
Hospedagem	0	0,01	0	2,76	0	30,62	0,00	0,00
P2P	0,47	1,25	3,34	1,72	25,81	19,06	662,19	119,57
Email	1,38	0,82	0,66	0,78	5,10	8,64	44,56	82,66
SSL/TLS	10,83	20,6	9,26	15,21	71,56	168,53	79,67	64,16
Demais	15,90	19,44	0,9	0,8	6,95	8,86	5,39	3,81



(a) Dropbox (b) SMTP

Figura 7. Impacto das falhas no volume de tráfego de aplicações assíncronas.

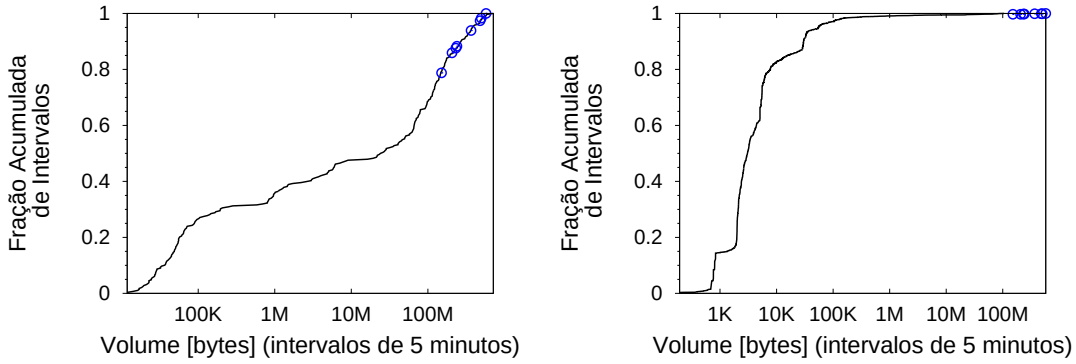
quena. O tráfego de aplicações P2P como Bittorrent e eDonkey também é pequeno, em parte devido a regras de bloqueio no *firewall* da UFJF. Como o tráfego dessas aplicações é pequeno com e sem falha, não conseguimos observar mudança de comportamento. Por último, a redução da proporção de conexões HTTP POST corrobora nosso resultado anterior de redução do tráfego interativo e possível evasão dos usuários da rede.

Também analisamos as frações de conexões e tráfego associadas a cada conjunto de aplicações 30 minutos após a restauração das falhas (não mostrado). Observamos que um intervalo de 30 minutos é suficiente para que usuários percebam a restauração da falha e retornem para o comportamento normal. Por exemplo, 30 minutos é suficiente para que a fração de conexões ao Facebook se normalize. Para as falhas analisadas, o volume de tráfego fica alterado por algumas horas, devido a modificações no comportamento de aplicações (seção 6). Infelizmente todas as falhas analisadas foram restauradas durante a madrugada e não pudemos analisar nenhuma falha restaurada em horário de pico.

6. Impacto das Falhas no Comportamento de Aplicações

Falhas têm impacto imediato no comportamento do tráfego e impacto gradativo no comportamento do usuário. Nesta seção avaliamos o impacto de falhas no comportamento de aplicações assíncronas que executam constantemente em plano de fundo.

A figura 7 mostra o volume de tráfego, em intervalos de trinta minutos para co-



(a) Tráfego no dia 17 de janeiro (b) Intervalos entre 19:55 e 20:35

Figura 8. Distribuição do volume de tráfego Dropbox em intervalos de 5 minutos em diferentes períodos. Pontos em destaque são o volume de tráfego Dropbox nos 40 minutos seguintes à falha do dia 10 de janeiro, em intervalos de 5 minutos.

nexões Dropbox[3] e SMTP (*Simple Mail Transport Protocol*), segundo classificação do TSTAT. Como anteriormente, as linhas azuis marcadas com um quadrado mostram o tráfego dos dias 14 e 15, sem falhas, e as linhas vermelhas marcadas com um círculo mostram o tráfego dos dias 7 e 8, com falha.

Como o Dropbox é hospedado na plataforma Amazon Web Services, todo o tráfego é interrompido durante a falha. De forma similar, parte do tráfego SMTP é internacional e interrompido durante a falha. Quando a falha é restaurada às 00:05 do dia 8, as tarefas acumuladas pelas duas aplicações ao longo da falha (i.e., arquivos criados e modificados no Dropbox bem como e-mails enfileirados) são disparadas. Para ambas aplicações, percebemos uma rajada de tráfego após a restauração da falha. A rajada aumenta o tráfego enviado pela aplicação e tem duração relativamente curta, entre 30 minutos e 1 hora.

A linha na figura 8(a) mostra a distribuição acumulada do volume de tráfego Dropbox durante o dia 17 de janeiro de 2013 (sem falha), em intervalos de cinco minutos. Nós também calculamos o tráfego Dropbox durante a rajada de tráfego após restauração de uma falha. Em particular, calculamos o tráfego em intervalos de cinco minutos durante os 40 minutos seguintes à restauração da falha do dia 10 de janeiro (i.e., oito intervalos de cinco minutos entre 19:55 e 20:35). Nós marcamos o tráfego relativo aos intervalos com rajada de tráfego com círculos azuis sobre a distribuição de tráfego Dropbox do dia 17 de janeiro. Focamos na falha do dia 10 de janeiro pois foi a falha restaurada mais cedo (às 19:55), de forma que uma comparação com o tráfego Dropbox em dias normais fosse mais realista. Os resultados mostram que os intervalos seguintes à falha têm volume de tráfego Dropbox maior que a maioria dos intervalos ao longo do dia, e comparável ao tráfego Dropbox em horários de pico.

A figura 8(b) é similar, mas a linha mostra a distribuição acumulado do volume de tráfego Dropbox apenas nos intervalos de 5 minutos entre 19:55 e 20:35 dos dias 14, 15, 16, 17 e 18 de janeiro (sem falhas). A figura 8(b) mostra que a rajada de tráfego Dropbox após restauração de uma falha é significativamente maior que o tráfego típico do horário, até 7 vezes maior no período de 40 minutos analisado.

[3]Na figura 5 e na tabela 1 a maior parte do tráfego Dropbox está classificado como “SSL/TLS”.

Atualmente, aplicações de armazenamento de arquivos em nuvem (*cloud storage*) são responsáveis por uma fração não desprezível do tráfego do campus. Apenas o Dropbox, uma das aplicações mais populares para tal fim, consome na média 4% da banda da universidade. Rajadas de tráfego combinadas a um possível aumento do volume de tráfego destas aplicações podem comprometer o desempenho da rede em períodos pós-falha e degradar o desempenho de aplicações interativas como teleconferências.

7. Discussão e Trabalhos Futuros

Este trabalho apresentou uma caracterização do impacto de falhas na rede da RNP no tráfego da Universidade Federal de Juiz de Fora. As falhas estudadas são parciais: enlaces e destinos internacionais ficaram inacessíveis durante horas enquanto destinos no Brasil praticamente não foram afetados. Nós focamos no impacto das falhas no comportamento dos usuários e no comportamento de aplicações assíncronas que rodam em plano de fundo. Observamos indicativos de evasão dos usuários da rede da universidade durante falhas. Observamos também que o tráfego de entretenimento migra de aplicações indisponíveis para aplicações disponíveis (em particular, do Facebook para o YouTube). Por último, discutimos que aplicações assíncronas com servidores fora do Brasil, como Dropbox e SMTP, podem acumular tarefas durante a falha e gerar uma rajada de tráfego imediatamente após restauração da falha.

Um potencial mecanismo para redução da evasão dos usuários da rede durante falhas é o estabelecimento de parcerias de troca de tráfego com redes que hospedam serviços de produtividade. Em particular, conectividade com o Google, Akamai e Amazon Web Services em pontos de troca de tráfego nacionais tornaria vários serviços disponíveis durante falhas. Esta solução depende da criação de centros de processamento de dados no Brasil, o que requer investimentos. Uma alternativa mais imediata é usar outras redes como provedores de acesso a serviços críticos durante falhas; por exemplo, outras redes educacionais da América Latina podem prover conectividade internacional temporariamente durante falhas.

Rajadas de tráfego geradas por aplicações assíncronas após restauração de falhas podem comprometer o desempenho da rede. Apesar das falhas que analisamos terem sido restauradas em período de baixa carga (após 19:55), rajadas após uma restauração em horário de pico podem levar a congestionamento de enlaces internacionais e comprometer aplicações como voz sobre IP. Este problema pode ser mitigado por modificações no *software*, modeladores de tráfego (*traffic shaping*) ou priorização de tráfego.

Atualmente estamos procurando outras universidades parceiras para obter dados de períodos com falha e estender nossa análise. Queremos também explorar alterações no comportamento de usuários e aplicações para melhorar técnicas de detecção de falhas. Por exemplo, um período com ausência de tráfego Dropbox seguido de uma rajada pode ser indicativo de uma falha (potencialmente parcial) de acesso a servidores do Dropbox.

Referências

Dainotti, A., Squarcella, C., Aben, E., Claffy, K. C., Chiesa, M., Russo, M., and Pescapé, A. (2011). Analysis of Country-wide Internet Outages Caused by Censorship. In *Proc. IMC*.

Feamster, N., Andersen, D., Balakrishnan, H., and Kaashoek, F. (2003). Measuring the Effects of Internet Path Faults on Reactive Routing. In *Proc. ACM SIGMETRICS*.

Finamore, A., Mellia, M., Meo, M., Munafò, M. M., and Rossi, D. (2011). Experiences of Internet traffic monitoring with tstat. *IEEE Network*, 25(3):8–14.

Kandula, S., Mahajan, R., Verkaik, P., Agarwal, S., Padhye, J., and Bahl, P. (2009). Detailed Diagnosis in Enterprise Networks. In *Proc. ACM SIGCOMM*.

Katz-Bassett, E., Madhyastha, H., John, J. P., Krishnamurthy, A., Wetherall, D., and Anderson, T. (2008). Studying Black Holes in the Internet with Hubble. In *Proc. USENIX NSDI*.

Kompella, R., Yates, J., Greenberg, A., and Snoeren, A. (2007). Detection and Localization of Network Blackholes. In *Proc. IEEE INFOCOM*.

Lakhina, A., Crovella, M., and Diot, C. (2004). Diagnosing Network-wide Traffic Anomalies. In *Proc. ACM SIGCOMM*.

Li, J. and Brooks, S. (2011). I-seismograph: Observing and Measuring Internet Earthquakes. In *Proc. IEEE INFOCOM*.

Mahimkar, A., Yates, J., Zhang, Y., Shaikh, A., Wang, J., Ge, Z., and Ee, C. (2008). Troubleshooting Chronic Conditions in Large IP Networks. In *Proc. ACM CoNEXT*.

Markopoulou, A., Iannaccone, G., Bhattacharyya, S., Chuah, C. N., Ganjali, Y., and Diot, C. (2008). Characterization of Failures in an Operational IP Backbone Network. *IEEE/ACM Trans. Netw.*, 16(4):749–762.

Poese, I., , Uhlig, S., Kaafar, M. A., Donnet, B., and Gueye, B. (2011). IP Geolocation Databases: Unreliable? *SIGCOMM Comput. Commun. Rev.*, 41(2):53–56.

Quan, L., Heidemann, J., and Pradkin, Y. (2013). Trinocular: Understanding Internet Reliability Through Adaptive Probing. In *Proc. ACM SIGCOMM*.

Sandvine (2013). Global Internet Phenomena Report 2H2013. Available at: `http://www.sandvine.com/trends/global-internet-phenomena`.

Silveira, F. and Diot, C. (2010). URCA: Pulling out Anomalies by their Root Causes. In *Proc. IEEE INFOCOM*.

Sommers, J., Barford, P., Duffield, N., and Ron, A. (2007). Accurate and Efficient SLA Compliance Monitoring. In *Proc. ACM SIGCOMM*.

Stefanov, S. (2008). YSlow 2.0. In *CSDN SC2C*.

Turner, D., Levchenko, K., Snoeren, A., and Savage, S. (2010). California Fault Lines: Understanding the Causes and Impact of Network Failures. In *Proc. ACM SIGCOMM*.

Zhang, Y., Mao, Z., and Wang, J. (2007). A Framework for Measuring and Predicting the Impact of Routing Changes. In *Proc. IEEE INFOCOM*.

Zhang, Z., Zhang, Y., Hu, Y. C., Mao, Z. M., and Bush, R. (2008). iSPY: Detecting IP Prefix Hijacking On My Own. In *Proc. ACM SIGCOMM*.