

Caracterização de Interrupções na Internet usando RIPE Atlas

Henrique Duarte Moura¹, Alisson R. Alves¹, Julio H. Soto¹, Luis H. C. Reis¹, Alex Borges Vieira², Daniel F. Macedo¹, José M. S. Nogueira¹

¹ Departamento de Ciência da Computação, UFMG
{henriquemoura, alissonralves, julio.soto, luiscantelli}@dcc.ufmg.br
{damacedo, jmarcos}@dcc.ufmg.br

² Departamento de Ciência da Computação, UFJF
alex.borges@ufjf.edu.br

Resumo. *Deteção e classificação de interrupções na Internet é um tema amplamente estudado. Entretanto, apesar de existir diversas técnicas e estudos para deteção de interrupções, essa não é realizada com foco nas interrupções para o cliente final da Internet. Apresentamos uma avaliação sistemática das interrupções da Internet a partir de dados históricos de um sistema de monitoração da Internet, o RIPE Atlas. Os resultados mostram que os parâmetros geoeconômicos utilizados nesta pesquisa têm pouca correlação com a quantidade de interrupções na Internet. Contudo, observamos que existem diferenças na duração e frequência de interrupções com a localização das sondas. Mostramos que um modelo simples de interrupções pode ser utilizado, como no caso do Furacão Harvey, para identificar problemas na rede. Identificamos um incremento de 10% na quantidade de interrupções e de 44% no tempo das interrupções no período do furacão na região de Houston, Texas. Os resultados podem ser utilizados no futuro como substrato para o desenvolvimento de novos preditores de interrupções.*

Abstract. *Detection and classification of Internet interruptions is a widely studied topic. However, although there are several techniques and studies for detecting interruptions, this is not done with a focus on the interruptions to the end user. This paper presents a systematic evaluation of Internet interruptions from historical data of an Internet monitoring system, called RIPE Atlas. The results show that the geoeconomic parameters used in this research have little correlation with the amount of interruptions in the Internet. However, we observe that there are differences in the duration and frequency of interruptions with the location of the devices. We show that a simple interruption model can be used to identify network problems, as in the case of Hurricane Harvey. We identified a 10% increase in the number of interruptions and 44% in interruption time in Houston (Texas) during the hurricane. The results can be used in the future as a substrate for the development of new interrupt predictors.*

1. Introdução

As redes de computadores são partes integrantes das vidas e negócios das pessoas. A IHS Markit [Machowinski 2016] estimou em 2016 que o tempo de inatividade da rede custa

cerca de 700 bilhões de dólares por ano para as empresas norte-americanas de tecnologia da informação. De acordo com o mesmo estudo, cada empresa de médio porte (100 a 1000 funcionários) perdeu em média 1 milhão de dólares de receita em um ano por tempo de inatividade, enquanto 60 milhões de dólares por ano são perdidos por cada grande empresa em receita. As interrupções de rede podem ainda prejudicar a qualidade da experiência do usuário.

Espera-se que os operadores de serviços na Internet possuam um certo nível de qualidade, e portanto podem estar sujeitos a multas governamentais se este nível não for atingido. Por exemplo, no Brasil, cada cliente cujo tempo de inatividade seja maior que trinta minutos deve ser compensado com uma redução na assinatura mensal equivalente ao custo do tempo de inatividade [Anatel 2013]. Ao mesmo tempo, a satisfação do cliente com a Internet está diminuindo nos EUA¹. Operadores e provedores de internet estão entre as empresas que têm o menor índice de satisfação entre os clientes. Apesar de existir ferramentas comerciais (e.g., SpeedTest²) ou patrocinadas por governos (e.g., Brasil Banda Larga³) para medir o rendimento dos provedores, não existem muitos estudos aprofundados avaliando a frequência e a duração das interrupções de conectividade para o cliente final. De fato, criar modelos de falhas para usuários finais de Internet é desafiador, uma vez que os dados para determinar tais características são essenciais para as regras de negócios de empresas e ISPs.

Este artigo apresenta uma avaliação sistemática da frequência e duração de interrupções da Internet pelo mundo. Apresentamos a evolução temporal das interrupções e avaliamos se países com índice de desenvolvimento mais elevado possuem Internet mais confiável do que países subdesenvolvidos. Esta avaliação é realizada usando dados históricos obtidos de sondas do RIPE Atlas (<https://atlas.ripe.net/>) espalhadas por todo o globo. Avaliamos ainda se existem relações entre as condições geoeconômicas dos países com a frequência de interrupções. Apesar de não programarmos uma sondagem ativa da Internet, utilizamos resultados de sondagens ativas já disponíveis na base de dados do RIPE Atlas, uma vez que todos os dispositivos utilizados já realizam periodicamente sondagens utilizando ping, traceroute, requisições de DNS, conexões SSH, entre outros.

Os resultados mostram que os parâmetros geoeconômicos utilizados nesta pesquisa têm pouca correlação com a quantidade de interrupções na Internet. Contudo, observamos que existem diferenças na duração e na frequência de interrupções conforme a localização dos dispositivos. Utilizando um modelo simples de interrupções, mostramos que existem diferenças em períodos de normalidade e anormalidade, utilizamos dados coletados durante o Furacão Harvey. Durante o furacão a duração e frequência das interrupções é superior àquelas observadas antes dele.

O restante deste artigo está organizado da seguinte forma. A Seção 2 discute os trabalhos relacionados ao tema. Na seção seguinte apresentamos o modelo que orientou as nossas avaliações do conjunto de dados. O conjunto de dados analisado é apresentado e discutido na Seção 3 e também mostramos a sanitização dos dados. A seção 4 caracteriza os dados de acordo com condições geoeconômicas e com o modelo apresentado. A seção 5 conclui o artigo e lista os trabalhos futuros.

¹<https://goo.gl/C47MLP>

²<http://www.speedtest.net/pt/>

³<http://www.brasilbandalarga.com.br/speedtest>

2. Trabalhos Relacionados

Desde os estágios anteriores da Internet, a análise de interrupções é importante. Apesar do notório interesse por medição da Internet e de seus serviços, historicamente, têm-se realizado más medições, más interpretações e pouca instrumentação [Paxson et al. 1998]. Até mesmo os pontos de coleta, se mal selecionados, influenciam negativamente nas medições realizadas [Chatzis et al. 2013]. De fato, 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 monitorar o comportamento dinâmico da rede [Nucci and Papagiannaki 2009]. Atualmente, a Internet é um vasto conjunto de redes interconectadas, porém operadas por organizações distintas que, por vezes, são concorrentes entre si [Clark et al. 2005]. Como uma consequência disso, muitos domínios não cooperam com iniciativas externas de medição de desempenho.

De maneira geral, existem duas formas antagônicas para a realização de medições de tráfego na Internet. A primeira refere-se às medições passivas de tráfego, enquanto a segunda, refere-se à abordagem ativa de coleta de informações [Finamore et al. 2011, Paxson et al. 1998]. A abordagem passiva avalia apenas o tráfego gerado pelas aplicações de interesse. Geralmente, tal tráfego é armazenado em determinados pontos de coleta, e a análise é feita após um período dessa coleta. A abordagem ativa, ao contrário da primeira, é aquela em que um tráfego é injetado na rede (ou na aplicação de interesse), o que induz efeitos que podem ser medidos. Existem uma série de compromissos entre as duas abordagens de medição. Por exemplo, a abordagem ativa interfere na aplicação a ser estudada, enquanto a captura do tráfego realizada pela abordagem passiva apresenta problemas de segurança e privacidade.

Nesse sentido, [Bush et al. 2009] explora as propriedades topológicas da Internet usando dupla sondagem e envenenamento de rotas. Embora estes autores não se concentrem em interrupções, seu trabalho forneceu inspiração para outros trabalhos nesta área. Por exemplo, [Quan et al. 2012] utiliza sondagem por ICMP ativo que examina alguns (ou todos) os endereços em alguns (ou todos) os blocos de endereço /24 analisáveis no espaço de endereço IPv4. Com este método, os autores conseguem detectar interrupções da Internet a partir de um único computador. Eles mostram que essa abordagem pode fornecer resultados significativamente mais precisos e com menos falsos positivos.

Mais recentemente, Trinocular usa sondas ativas para entender a disponibilidade da rede [Quan et al. 2013]. Trinocular captura informações de sondas ICMP e apreende qual é o estado atual da rede. Este modelo usa inferência bayesiana para encontrar correlações entre os dados de um evento. A inferência bayesiana apresenta duas fases: a) solicitação de prováveis eventos das sondas ativas nos mesmos locais; e b) análise dessas probabilidades para confirmar algum evento. Nossa abordagem não usa sondagem ativa. Utilizamos resultados de sondagens ativas, já efetuadas pela plataforma RIPE Atlas, em muitos pontos de observação implantados em toda a Internet, melhorando a cobertura dos eventos de interrupção. Além disso, as sondas do RIPE Atlas fornecem acesso a informações IPv6. Nosso trabalho também se beneficia da infraestrutura fornecida pelo RIPE Atlas. Usamos as informações IPv4 e IPv6, além de acessibilidade para os DNS raiz.

Uma série de trabalhos caracteriza e modela as interrupções da Internet. Por exem-

plo, [Dainotti et al. 2014] utiliza geolocalização para entender as interrupções na Internet no Egito e na Líbia em 2011. Os autores selecionaram dados como endereços IP, prefixos BGP e sistemas autônomos para monitorar ocorrências em ambos os países. Os dados são separados em dois aspectos: plano de controle e tráfego não solicitado no plano de dados. Os dados analisados mostram como ocorreu a interrupção do acesso à Internet, e que ataques de DoS (*Denial of Service*) geraram interrupções em ambos os países.

Além disso, existem sistemas on-line que detectam interrupções na Internet. Por exemplo, Disco é um sistema de diagnóstico baseado no modelo de rajadas de Kleinberg para detectar desconexões TCP usando a infraestrutura RIPE Atlas [Shah et al. 2017]. Neste caso, os autores limitaram-se a apresentar apenas um caso de uso de uma queda de energia em Amsterdã, assim como não forneceram a frequência e o tempo de desconexão dos eventos. Nossa abordagem também usa o mesmo banco de dados, mas nossos dados usam outros tipos de informações para caracterizar os eventos de interrupção. IODA⁴ monitora a Internet em tempo integral, detectando e analisando interrupções que afetam grandes redes distribuídas em diversas áreas geográficas. O sistema faz inferências sobre interrupções da Internet usando séries temporais e combina informações de plano de controle BGP com plano de dados e informações de sondagem ativa. Atualmente este sistema fornece alertas de redes quando as interrupções acontecem, mas não combina informações econômicas e sociais com eles.

As medidas feitas no plano de dados são diretas e podem ser mais precisas do que as do plano de controle [Quan et al. 2012]. Assim, este trabalho utiliza um modelo simples de interrupções para monitorar a Internet utilizando informações obtidas de um conjunto de pontos de presença distribuídos mundialmente pela plataforma RIPE Atlas. Estes pontos estão localizados normalmente em sistemas autônomos (AS) clientes. Nossa abordagem não utiliza sondagem ativa, contudo pode se beneficiar de sondagens ativas já programadas nas sondas do RIPE Atlas, uma vez que os dispositivos já realizam sondagens a intervalos regulares para pontos específicos utilizando, por exemplo, ping, trace-route e requisições DNS.

3. Metodologia de Coleta de Dados

Os dados utilizados nesta pesquisa foram coletados do RIPE Atlas, que é um conjunto de dispositivos de sondagem – sondas e âncoras – instalados em todo o mundo. A infraestrutura é mantida pelo RIPE NCC, que é o registro regional da Internet para Europa, Oriente Médio e partes da Ásia Central. Estes dispositivos executam uma série de medições utilizando ferramentas comuns na Internet, como ping, traceroute, consultas DNS, etc. Além disso, pesquisadores podem usar créditos (obtidos ao hospedar uma sonda ou patrocinar o RIPE Atlas) para executar suas próprias medições no RIPE Atlas.

As sondas são pequenos dispositivos de medição que geram uma quantidade pequena de tráfego de medição. Elas são instaladas em qualquer lugar, incluindo residências, de forma que qualquer pessoa possa hospedar uma sonda. As âncoras, por outro lado, são dispositivos mais potentes que requerem mais largura de banda e, desta forma, são instalados em locais selecionados, como por exemplo, provedores de rede e institutos de pesquisa. As sondas são geralmente entregues aos participantes em conferências e, por isso, sua distribuição está mais concentrada em países desenvolvidos ou naqueles com

⁴<https://www.caida.org/projects/ioda/>

participação mais pronunciada em simpósios científicos. Até setembro de 2017, o RIPE Atlas forneceu mais de 23 mil sondas e 287 âncoras. Essas sondas cobrem 212 países, cerca de 6% dos ASN IPv4 e quase 10% dos ASNs IPv6.

3.1. Coleta e sanitização dos dados

O RIPE Atlas fornece uma Interface de Programação de Aplicação (API) que permite aos pesquisadores consultar os dados coletados. Utilizamos servidores do laboratório WINET (*Wireless Network*) no Departamento de Ciência da Computação da UFMG para solicitar à API, os dados das sondas. As solicitações dos dados foram feitas exaustivamente e os mesmos foram armazenados em um banco de dados local. A partir disso, realizamos as análises deste trabalho nesse banco, sendo solicitado esporadicamente, ao RIPE Atlas, outras informações complementares.

Foram gerados até o momento 34.024 números de identificação para as sondas. Contudo a partir dos dados obtidos pela API RIPE Atlas identificamos um total de 23.254 sondas em seus registros. Do total de 23.254 sondas, descobrimos que cerca de 43% são ativas (10.098 sondas conectadas), 36,5% são marcadas como abandonadas e 6,9% estão desconectadas. As sondas abandonadas não fornecem dados para nossa pesquisa. O conjunto de dados utilizado neste trabalho consiste em 84 meses (7 anos) de dados oriundos de 19.657 sondas, correspondendo a 10.114.430 eventos de rede (731 MB). Não utilizamos âncoras, pois nosso objetivo focou na caracterização das interrupções do ponto de vista do usuário. As âncoras tipicamente estão localizadas em PTTs e em ISPs devido à sua maior demanda por banda, e por isso, foram excluídas.

Interrupções consistem nos períodos identificados como um bloco cheio no modelo apresentado na Seção 4.1. Interrupções com duração inferior a 1 segundo não foram computadas em função do próprio processo de coleta e aquelas acima de três meses, que configuram abandono pelo RIPE Atlas, foram descartadas. Das 23.254 sondas, descartamos aquelas marcadas como “Nunca conectada” por não fornecerem dados. Desconsideramos também 517 sondas marcadas como “Abandonadas” e 28 sondas “Desconectadas” pelo mesmo motivo. Isso resultou em 12.594.048 de dias de tempo de atividade e 2.489.521 de dias de tempo de inatividade, quando somados os tempos para todos os dispositivos.

3.2. Sumário dos dados coletados

A Figura 1 mostra a distribuição cumulativa (CDF) dos tempos de desconexão, como definido na seção 4.1, para todos os eventos registrados no conjunto de dados (mais de 10 milhões de registros analisados). Na figura destacamos o tempo de desconexão para 90%, 92,6% e 95% dos eventos. Observamos que 90% dos eventos ocorrem por menos de 24,5 minutos e que 95% dos eventos ocorrem por menos de uma hora. Ainda na figura marcamos o percentual para 30 minutos correspondendo a 92,6% dos casos.

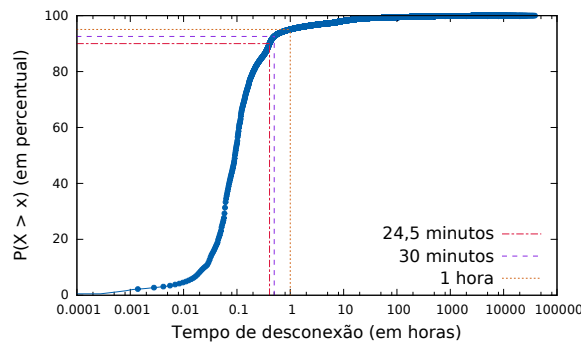


Figura 1. Distribuição cumulativa dos eventos de desconexão

A Figura 2 mostra um gráfico com a distribuição de sondas entre os continentes. Ela mostra 21.092 sondas porque no registro no RIPE Atlas algumas sondas não apresentam país associado. Vemos ainda que a maioria das sondas estão na Europa (62%), seguido pela Ásia (17%) e América do Norte (12%). Na Europa, a Dinamarca é o país com mais sondas (2.254) - também é o país com mais sondas no mundo. Seguem-se a França (1.690), a Inglaterra (1.304), a Rússia (1.209) e a Holanda (1.104). Esses cinco países representam 58% das sondas europeias. Na América do Norte e no Caribe, a maioria das sondas estão nos Estados Unidos - cerca de 75,9% das sondas do continente. Os Estados Unidos são também o segundo país (9,4%) no mundo em número de sondas.

Podemos ver na Figura 3 a quantidade de sondas ativas a cada mês, durante o período avaliado. Representamos na mesma figura a linha de regressão para as quantidades mensais e obtivemos um coeficiente de correlação ajustado de 84%. Isto implica que o número de sondas adicionadas está realmente aumentando mensalmente, apesar das variações abruptas no número de sondas ativas.

Quando o RIPE Atlas foi implantado existiam ativos cerca de 35.000 Sistemas Autônomos (AS) na Internet. Um AS é um grupo de redes de endereços IP que é gerenciado por um ou mais operadores de rede que possuem uma clara e única política de roteamento. Cada AS tem associado um número (ASN – Número de Sistema Autônomo) que é utilizado como um identificador do AS para troca de rotas com outros sistemas externos. Um dos objetivos do RIPE Atlas é ter uma sonda em cada ASN [Staff 2015] e para isso é necessário cerca de 100.000 sondas.

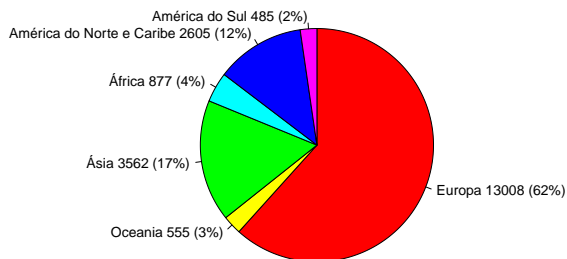


Figura 2. Distribuição das sondas por continente

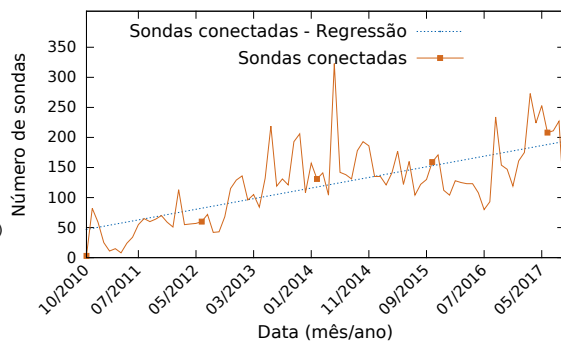


Figura 3. Evolução da quantidade de sondas ativas por mês

Tabela 1. Cobertura de ASN pelas sondas do Ripe Atlas.

Continentes	ASN IPv4		ASN IPv6	
	Cobertura	%	Cobertura	%
Europa	2.961	56,7	1.162	66,1
Ásia	1.180	22,6	247	14,1
América do Norte e Caribe	549	10,5	216	12,3
África	264	5,1	49	2,8
Oceania	183	3,5	69	3,9
América do Sul	167	3,2	59	3,4
Total	5.221	100	1.757	100

A Tabela 1 mostra o número total de ASN IPv4 e IPv6 cobertos no RIPE Atlas e quanto este número representa do total de ASN presentes na região (%). As colunas mostram quantos ASN são cobertos por sondas, e ainda se as sondas possuem conexão IPv4 ou IPv6 no ASN. Observamos que a quantidade de ASN cobertos é menor que a quantidade de ASN ativos atualmente (cerca de 14%).

A maioria das sondas está conectada a uma rede IPv4 (78%). Somente 25% dos dispositivos estão conectados em redes IPv6. Observamos que a adoção do IPv6 pela plataforma RIPE Atlas é baixa, mas a proporção de IPv6 no RIPE Atlas é maior do que a observada por outras entidades. Por exemplo, o Google registra 20,5% de conexões IPv6 em 11 de setembro de 2017⁵, enquanto APNIC⁶ mostra que há 15,69% sistemas que suportam IPv6. Isto se deve ao fato que as sondas são executadas em locais de trabalho e residências de pessoas que trabalham em áreas relacionadas às redes de computadores, pois as sondas são entregues majoritariamente em conferências científicas da área.

4. Caracterização das interrupções na Internet

Nesta seção apresentamos os resultados obtidos pela aplicação do modelo da Seção 4.1 e o cruzamento de informações geoconômicas. Apresentamos, ainda, como o nosso modelo pode ser utilizado para caracterizar interrupções em situações de catástrofes.

4.1. Modelo

A Figura 4 mostra o modelo simples que utilizamos na interpretação e análise dos dados sobre interrupções obtidos das sondas do RIPE Atlas. Este modelo é formado por um conjunto $P = \{p_1, p_2, \dots, p_n\}$ de sondas. Cada sonda é caracterizada por dois intervalos: (i) **tempo de queda** – que representa o período onde a sonda não responde a solicitações e portanto podem ter ocorrido problemas de *hardware*, queda de energia, problemas nos enlaces que conectam esta sonda à Internet, problemas de configuração, entre outros; e (ii) **tempo entre quedas** – que corresponde ao tempo onde a sonda é responsiva e está conectada à Internet. A cada falha, associamos um rótulo de tempo (*timestamp*), e as sondas possuem metadados relacionados, tais como posição geográfica, código de país, código de AS à qual o endereço IP recebido está associado. Nem todos estes metadados podem estar preenchidos, como veremos a seguir.

Apesar de simples, este modelo pode ser empregado para obter dados de interesse ao realizarmos agregações nas sondas a partir de metadados comuns. Por exemplo, ao agregar as sondas de um AS é possível ordenar os ASs por confiabilidade, ou uma agregação por país pode indicar a evolução da Internet neste país.

⁵<https://www.google.com/intl/pt-BR/ipv6/statistics.html>

⁶Asia-Pacific Network Information Centre – <https://stats.labs.apnic.net/ipv6/>

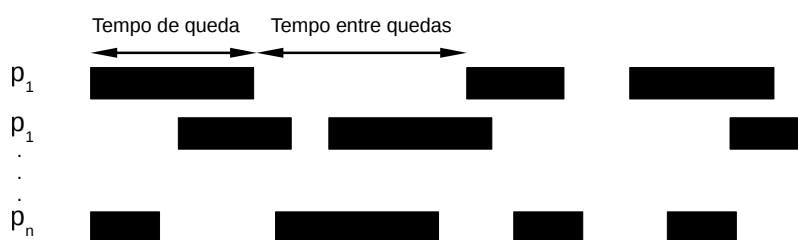


Figura 4. Modelo simples de interrupções no RIPE Atlas.

4.2. Resultados

Tabela 2. Comparação entre tempo de atividade total e o tempo de inatividade, por continente

Continente	Número de sondas	% Inatividade por sonda	Média (em dias)	
			Ativo	Inativo
América do Norte e Caribe	2.605	0,22	678,3 ± 21,2	1,4 ± 0,1
Europa	13.008	0,34	692 ± 10,6	2,3 ± 0,2
Ásia	3.562	0,40	505,7 ± 18,3	2 ± 0,2
Oceania	555	0,58	606,9 ± 45,9	3,5 ± 1,5
América do Sul	485	0,64	437,9 ± 44,2	2,8 ± 0,9
África	877	0,66	427,4 ± 32,6	2,8 ± 0,5
Total	21.092	0,35	642,4 ± 8,1	2,2 ± 0,1

A Tabela 2 resume nossas descobertas para o tempo de atividade total (soma dos tempos entre quedas, segundo nosso modelo) e o tempo de inatividade (soma dos tempos de queda) por continente. O primeiro resultado importante encontra-se na avaliação da disponibilidade do sistema. Considerando que os SLAs (*Service Level Agreements*) são medidos em classes de “noves”, isto é, a disponibilidade do serviço é referida pelo número de nove nos dígitos, então a Internet, como medida pelo RIPE Atlas, é um sistema de “dois noves” (99%). Este valor é equivalente à disponibilidade de planos de acesso Internet nível SCM9 na Anatel [Anatel 2016].

Os resultados obtidos para a Europa são consistentes, o que pode ser explicado pelo forte compromisso com os objetivos do RIPE Atlas. Embora a Europa tenha o maior número de sondas, ela tem o segundo menor tempo de inatividade entre os continentes (0,34%). A América do Norte mostrou a menor variação de intervalo encontrada entre todos os continentes nos resultados obtidos para cada sonda (com um intervalo de confiança de 95%). Por outro lado, os continentes formados principalmente por países subdesenvolvidos, como a Oceania, a África e a América do Sul, são aqueles que apresentam as maiores porcentagens de tempo de inatividade por sonda e apresentam também alta variabilidade dos resultados.

A Figura 5 mostra a evolução do número médio de interrupções em cada continente. Os valores foram divididos pelo número de sondas em cada continente. Desconsideramos nesta análise as sondas que não indicam o país onde foram instalados. Observamos que o número médio de interrupções vem diminuindo desde 2011 para todos os continentes. Em 2010, a variação indicada pelos intervalos de confiança é discrepante devido à pequena quantidade de dados e à grande variação dos valores observados. Por exemplo, o valor médio para a África foi de 52,5 interrupções com um limite superior indicado pelo intervalo de confiança a 95% de 560 interrupções. Calculamos a regressão

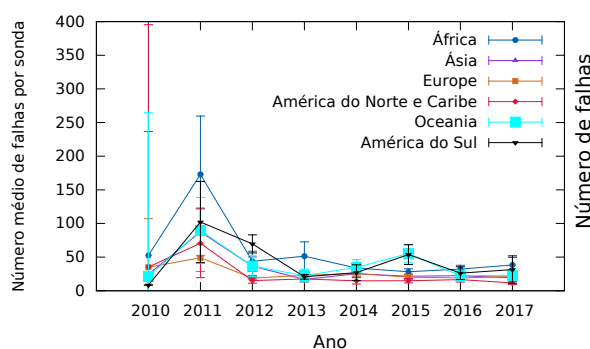


Figura 5. Evolução da média de interrupções por continente

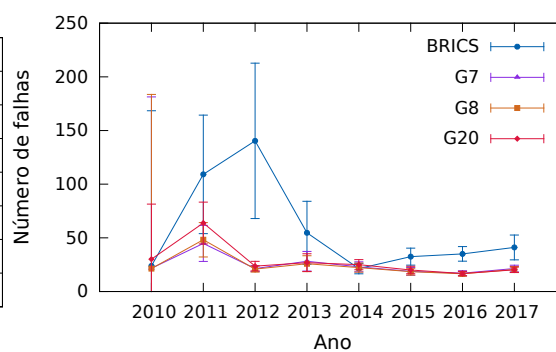


Figura 6. Evolução no número médio de interrupções por grupos de países

linear dos valores obtidos pelo ano de observação para todos os continentes. Observamos que a inclinação para todas as regressões é negativa, indicando que a média das interrupções está diminuindo.

Para avaliar se existe correlação entre as interrupções e grupos de países, agrupamos alguns países com base em fatores macroeconômicos. Para tal, consideramos os grupos econômicos G7, G8, G20 e BRICS. O G7 é composto por Canadá, França, Alemanha, Itália, Japão, Reino Unido e Estados Unidos. Esses países representam mais de 64% da riqueza global líquida. O G8 é formado por países do G7 e Rússia. O G20 engloba as 20 maiores economias globais (todos os países do G8, mais Argentina, Austrália, Brasil, China, Índia, Indonésia, México, Arábia Saudita, África do Sul, Coreia do Sul e Turquia). Já o BRICS representa as cinco principais economias emergentes: Brasil, Rússia, Índia, China e África do Sul.

A Figura 6 mostra a evolução da média anual do número de interrupções nesses grupos com um intervalo de confiança de 95%. Vemos que em 2010 a variação indicada pelo intervalo de confiança é grande para que possamos considerar que o valor mensal médio das interrupções nos grupos é equivalente. A partir desta data, os valores para o grupo BRICS se destacam dos outros grupos, apresentando maiores valores médios. Isto é devido à participação da Índia, China e África do Sul. Os intervalos de confiança para G7, G8 e G20 se sobrepõem de 2012 a 2017. Como os resultados para G7, G8 e G20 não foram muito diferentes, concentramos nossa análise no G7 e BRICS, que apresentaram as maiores diferenças.

Avaliamos ainda se existe correlação do número de interrupções com indicadores de bem-estar e de riqueza das nações. Estes dois indicadores são: IDH (Índice de Desenvolvimento Humano⁷) e PIB (Produto Interno Bruto⁸) por país, respectivamente. O IDH mede a saúde, a educação e a riqueza da população. Já o PIB representa a soma de todos os bens e serviços finais produzidos no ano, em dólares⁹. As Figuras 7 e 8 mostram a distribuição dos pontos representados pela quantidade média de interrupções e a métrica

⁷Fonte World Bank: <https://data.worldbank.org/country>

⁸Fonte ONU - Human Development Reports: <http://hdr.undp.org/en/data>

⁹O valor do PIB é atualizado em dólares para todos os países contudo não está calculado como *Purchasing power parity*.

indicada – IDH ou PIB. Utilizamos as interrupções que ocorreram no período de 2010 até 2015, e para minimizar o efeito da distribuição não homogênea das sondas pelos países, utilizamos na análise a quantidade de interrupções dividida pelo total de sondas do país. Mostramos ainda, em cada figura, três retas de regressão linear entre o número de interrupções das sondas e as métricas. Na curva identificada como “Todos” a regressão abrange todos os países. As retas “G7” e “BRICS” são as regressões das interrupções somente nos países do G7 e dos BRICS, respectivamente.

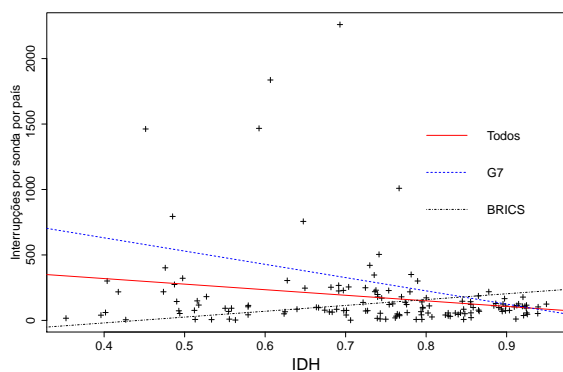


Figura 7. Correlação entre interrupções e IDH

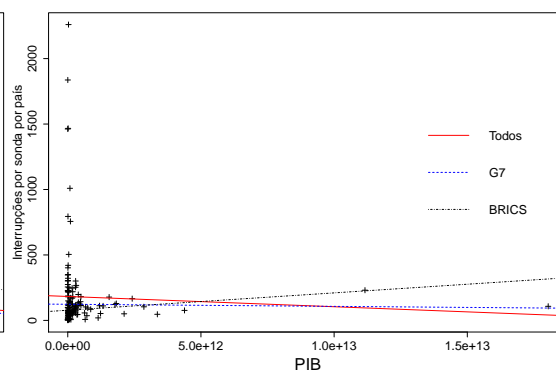


Figura 8. Correlação entre interrupções e PIB (em US\$)

A Figura 7 avalia se existe relação entre o bem-estar e a quantidade de interrupções. Observamos que as retas de regressão “Todos” e “G7” apresentam inclinações negativas, isto é, quanto maior o PIB, menor a quantidade média de interrupções, contudo destacamos que as correlações de Pearson para estes dois casos são baixas – 0,03343 e 0,08885, respectivamente. Já para os BRICS a regressão tem inclinação positiva, principalmente em função dos valores de China e Índia.

Observamos na Figura 8 que as regressões lineares com todas as interrupções e com aquelas do G7 apresentaram coeficientes de correlação de Pearson ajustado de 0,004973 e 0,1517, respectivamente, ou seja, não há correlação significativa para estes dois grupos. Esta afirmação é reforçada pelo fato das regressões não passarem pelo teste da tabela ANOVA com 95% de significância. Contudo é importante notar que ainda assim obtemos um resultado interessante, que indica que a quantidade de interrupções cai com o PIB do país. Outro resultado inusitado, mas que também apareceu em relação ao IDH, é que a regressão em relação ao PIB para o grupo dos BRICS apresentou um coeficiente ajustado de 0,7004, a regressão passa no teste ANOVA com 95% de significância e a inclinação da curva de regressão é positiva. Esta regressão apresenta portanto um comportamento oposto ao comportamento geral de melhoria da confiabilidade da Internet. Isto é influenciado principalmente pelo comportamento da China, que apresenta um PIB superior aos demais países do BRICS e ao mesmo tempo um número de interrupções também maior, tornando a inclinação positiva.

Fizemos as mesmas análise com PIB e IDH relacionando com o tempo médio das interrupções por países. Os coeficientes de correlação de Pearson ajustado para as interrupções “Todos”, “G7” e “BRICS” em relação ao PIB são 0,004175; 0,1985 e 0,299, respectivamente, ou seja, todas as correlações foram baixas. Neste caso é interessante destacar que a inclinação da reta de regressão é negativa para todos os três casos. Já os

coeficientes de correlação de Pearson ajustado em relação ao IDH são 0,06391, 0,1941 e 0,004091, respectivamente, mais uma vez baixos. A inclinação da reta de regressão linear entre o IDH e o tempo para os “BRICS” é positiva, repetindo o padrão da regressão com o número de interrupções e também em função dos valores da China.

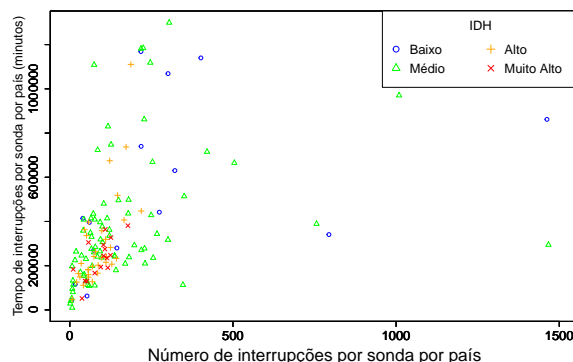


Figura 9. Agrupamentos por IDH.

O FMI agrupa os países em quatro categorias de acordo com o IDH: baixo, médio, alto e muito alto [Nielsen 2011]. A Figura 9 mostra a dispersão dos pontos em função do número de interrupções e do tempo de queda. Observamos que os países com IDH baixo estão dispersos na figura. Os países com IDH médio apresentam uma boa dispersão, principalmente, em relação ao tempo. Já os países com IDH alto e, em especial, com IDH muito alto concentram-se na parte esquerda e inferior da figura (baixo número de interrupções e baixo tempo por interrupção). Assim, vemos que países com maior IDH tendem a ter uma Internet mais confiável.

A Figura 10 é composta por dois gráficos, onde na parte superior mostramos a distribuição cumulativa do número de interrupções para as sondas, enquanto na parte inferior mostramos os valores, entretanto restringindo o período (eixo X) para 1000 minutos. Em cada gráfico da figura mostramos três curvas: Total, que considera todos os dispositivos do RIPE Atlas, os dispositivos localizados nos BRICS, e os dispositivos em países do G7. Destacamos que 84% dos dispositivos tiveram 500 ou menos interrupções no período avaliado. Observando o gráfico inferior da Figura 10 vemos que a duração das interrupções nos BRICS é maior que no G7. Por exemplo, para 50% dos dispositivos dos BRICS as interrupções são inferiores a 112 minutos, enquanto no total e no G7 elas são inferiores a 26,3 minutos. A figura 11 mostra a incidência de interrupções total, no G7 e nos BRICS. Apresentamos dois CDFs nesta figura – no gráfico superior mostramos todas as interrupções para os três casos, enquanto no gráfico inferior mostramos somente as interrupções com incidência inferior a 1000 minutos. Observamos que a quantidade de interrupções é maior nos países que compõem o BRICS do que os outros dois grupos. Verificamos pela Figura que os países do G7 verificaram 112 interrupções ou menos, contudo os BRICS registraram 155 interrupções.

4.3. Caracterização e recuperação de interrupções em situações de catástrofe

Nesta seção utilizamos o RIPE Atlas como um “termômetro” da saúde das redes em eventos de catástrofes. O objetivo é investigar se é possível monitorar eventos de catástrofes e a recuperação da infra-estrutura após tais eventos. Para isto analisamos as interrupções durante a ocorrência do furacão Harvey.

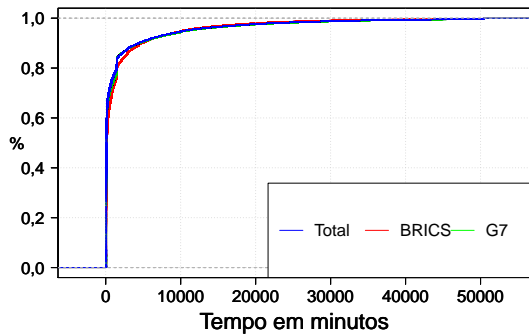


Figura 10. Duração das interrupções - separação por grupo de país.

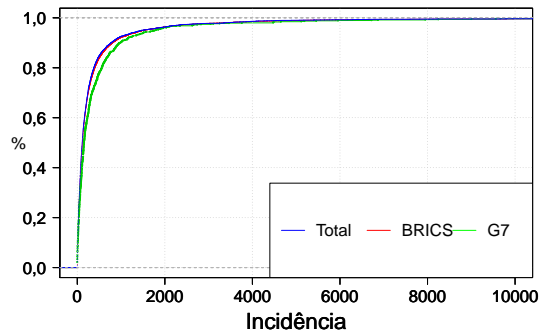


Figura 11. Quantidade de interrupções - separação por grupo de país.

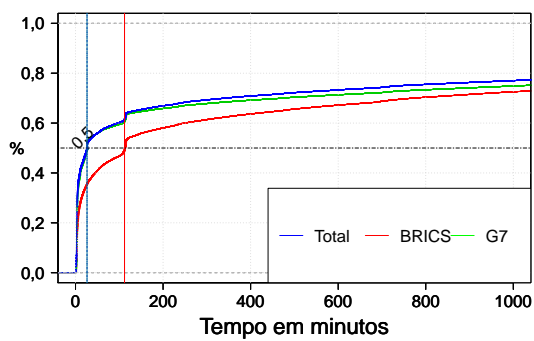


Figura 12. Quantidade das interrupções

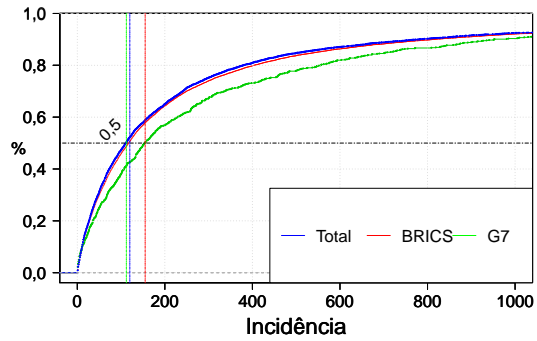
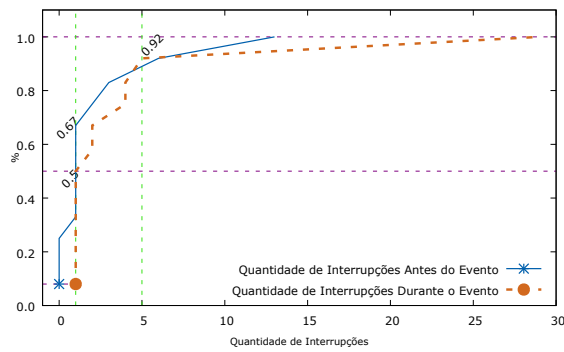
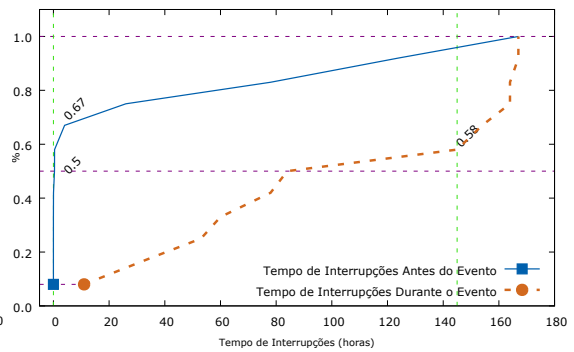


Figura 13. Duração das interrupções



O furacão Harvey é um ciclone tropical que atingiu a América do Norte e o Caribe entre os dias 17 de agosto e 3 de setembro de 2017. Neste sentido, selecionamos os EUA devido à sua quantidade de sondas, assim analisamos o comportamento das interrupções regionais por existir mais sondas distribuídas pela área do seu território. Em nossa análise caracterizamos as interrupções em 12 sondas localizadas exclusivamente na região de Houston (Texas) pois o Harvey atingiu a categoria 4 nessa região. A análise considerou a semana anterior e a semana da ocorrência do ciclone. Nesse período, contabilizamos



a quantidade e a duração das interrupções, que são mostradas nas Figuras 12 e 13, respectivamente. Nestas figuras exibimos duas curvas, uma representando o período anterior e outra durante o evento. Na Figura 12 observamos que a quantidade de interrupções durante o evento é superior à quantidade da semana anterior em 10%. Na Figura 13 as curvas se diferenciam com maior destaque. Identificamos que a duração das interrupções durante o período do evento é prolongada, gerando um acréscimo em 44% na duração média das interrupções com a ocorrência do ciclone.

4.4. Discussão

Em função das atividades necessárias para a obtenção dos resultados apresentados na seção 4, verificamos que o RIPE Atlas pode indicar a confiabilidade da Internet em países com uma maior distribuição de sondas. Vimos que a Internet mundial está se tornando mais confiável, com uma redução na quantidade e tempo médio de interrupções.

O modelo simples utilizado neste trabalho permite ainda obter informações relevantes de interrupções na Internet, como mostramos no caso do furacão Harvey. Contudo nosso trabalho atual precisa ser expandido para que possamos identificar quais são os períodos específicos de interrupções. Em trabalhos futuros pretendemos explorar as diferenças nas características entre os períodos de normalidade e de anormalidade. Tal estudo é relevante pois pode indicar métodos para identificar eventos de médio e grande porte na rede, e com isso informar aos administradores de ASs para que estes tomem medidas para evitar as suas consequências. Como trabalhos futuros, iremos investigar como prever tais eventos, de forma que a rede possa adaptar-se ao evento antes que ele aconteça.

5. Conclusões e Trabalhos Futuros

Este trabalho apresenta como as interrupções na Internet evoluem no tempo e avalia se os países com maior índice de desenvolvimento têm Internet mais confiável do que os países subdesenvolvidos. Mostramos que é possível usar dados existentes em plataformas de monitoramento e teste, como o RIPE Atlas, para obter informações sobre como a Internet funciona. Validamos nossa abordagem comparando eventos de interrupção durante o furacão Harvey e observamos que neste período nosso método é capaz de identificar, usando apenas os dados do RIPE Atlas, diferenças no funcionamento da Internet quando comparamos os períodos de normalidade com os de anormalidade.

A sondagem ativa complementa esse trabalho e é fundamental para detectar pequenas interrupções e fornecer estatísticas em toda a Internet. Também podemos utilizar informações de plano de controle para validar a detecção de uma interrupção. Além disso, a API do RIPE Atlas deve ser melhorada para permitir consultas mais seletivas, economizando a largura de banda do pesquisador e aumentando a velocidade de resposta.

Referências

- [Anatel 2013] Anatel (2013). Resolution nº 614, of may 28 of 2013. Technical report, National Telecommunications Agency.
- [Anatel 2016] Anatel (2016). Qualidade da Banda Larga Fixa (SCM) - Relatório de Indicadores de Desempenho Operacional. Technical report, Anatel.
- [Bush et al. 2009] Bush, R., Maennel, O., Roughan, M., and Uhlig, S. (2009). Internet optometry: assessing the broken glasses in internet reachability. In *Proceedings of the*

- 9th ACM SIGCOMM conference on Internet measurement conference*, pages 242–253. ACM.
- [Chatzis et al. 2013] Chatzis, N., Smaragdakis, G., Böttger, J., Krenc, T., and Feldmann, A. (2013). On the benefits of using a large IXP as an Internet vantage point. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 333–346. ACM.
- [Clark et al. 2005] Clark, D. D., Wroclawski, J., Sollins, K. R., and Braden, R. (2005). Tussle in cyberspace: defining tomorrow’s internet. *IEEE/ACM transactions on networking*, 13(3):462–475.
- [Dainotti et al. 2014] Dainotti, A., Squarcella, C., Aben, E., Claffy, K. C., Chiesa, M., Russo, M., and Pescapé, A. (2014). Analysis of country-wide internet outages caused by censorship. *IEEE/ACM Transactions on Networking*, 22(6):1964–1977.
- [Finamore et al. 2011] Finamore, A., Mellia, M., Meo, M., Munafo, M. M., and Rossi, D. (2011). Experiences of Internet Traffic Monitoring with Tstat. volume 25, pages 8–14. IEEE.
- [Machowinski 2016] Machowinski, M. (2016). The Cost of Server, Application & Network Downtime Survey & Calculator - 2016. Technical report, IHS Inc.
- [Nielsen 2011] Nielsen, L. (2011). IMF Working Paper – Classifications of Countries Based on Their Level of Development: How it is Done and How it Could be Done. Technical report, International Monetary Fund.
- [Nucci and Papagiannaki 2009] Nucci, A. and Papagiannaki, K. (2009). *Design, Measurement and Management of Large-Scale IP Networks*. Cambridge University Press.
- [Paxson et al. 1998] Paxson, V., Mahdavi, J., Adams, A., and Mathis, M. (1998). An architecture for Large Scale Internet Measurement. *IEEE Communications Magazine*, 36(8):48–54.
- [Quan et al. 2012] Quan, L., Heidemann, J., and Pradkin, Y. (2012). Detecting internet outages with precise active probing (extended). *USC/Information Sciences Institute, Tech. Rep.*
- [Quan et al. 2013] Quan, L., Heidemann, J., and Pradkin, Y. (2013). Trinocular: Understanding internet reliability through adaptive probing. In *ACM SIGCOMM Computer Communication Review*, volume 43, pages 255–266. ACM.
- [Shah et al. 2017] Shah, A., Fontugne, R., Aben, E., Pelsser, C., and Bush, R. (2017). Disco: Fast, good, and cheap outage detection. In *Network Traffic Measurement and Analysis Conference (TMA), 2017*, pages 1–9. IEEE.
- [Staff 2015] Staff, R. (2015). Ripe atlas: A global internet measurement network. *Internet Protocol Journal*, 18(3).