Characterizing and Avoiding Routing Detours Through Surveillance States

Abstract

An increasing number of countries are passing laws that facilitate the mass surveillance of Internet traffic. In response, governments and citizens are increasingly paying attention to the countries that their Internet traffic traverses. In some cases, countries are taking extreme steps, such as building new Internet Exchange Points (IXPs), which allow networks to interconnect directly, and encouraging local interconnection to keep local traffic local. We find that although many of these efforts are extensive, they are often futile, due to the inherent lack of hosting and route diversity for many popular sites. By measuring the country-level paths to popular domains, we characterize transnational routing detours. We find that traffic is traversing known surveillance states, even when the traffic originates and ends in a country that does not conduct mass surveillance. Then, we investigate how clients can use overlay network relays and the open DNS resolver infrastructure to prevent their traffic from traversing certain jurisdictions. We find that 84% of paths originating in Brazil traverse the United States, but when relays are used for country avoidance, only 37% of Brazilian paths traverse the United States. Using the open DNS resolver infrastructure allows Kenyan clients to avoid the United States on 17% more paths. Unfortunately, we find that some of the more prominent surveillance states (e.g., the U.S.) are also some of the least avoidable countries.

1 Introduction

When Internet traffic enters a country, it becomes subject to that country's laws. As a result, users have more need than ever to determine—and control—which countries their traffic is traversing. An increasing number of countries have passed laws that facilitate mass surveillance of their networks [24, 32, 35, 39], and governments and citizens are increasingly motivated to divert their Internet traffic from countries that perform surveillance (notably, the United States [17, 18, 48]).

Many countries—notably, Brazil—are taking impressive measures to reduce the likelihood that Internet traffic transits the United States [9–11, 14, 30] including building a 3,500-mile long fiber-optic cable from Fortaleza to Portugal (with no use of American vendors); pressing companies such as Google, Facebook, and Twitter (among others) to store data locally; and mandating the deployment of a state-developed email system (Expresso) throughout the federal government (instead of what was originally used, Microsoft Outlook) [8, 12]. Brazil is also building Internet Exchange Points (IXPs) [7], now has the largest national ecosystem of public IXPs in the world [15], and the number of internationally connected Autonomous Systems (ASes) continues to grow [13]. Brazil is not alone: IXPs are proliferating in eastern Europe, Africa, and other regions, in part out of a desire to "keep local traffic local". Building IXPs alone, of course, cannot guarantee that Internet traffic for some service does not enter or transit a particular country: Internet protocols have no notion of national borders, and interdomain paths depend in large part on existing interconnection business relationships (or lack thereof).

Although end-to-end encryption stymies surveillance by concealing URLs and content, it does not by itself protect all sensitive information from prying eyes. First, many websites do not fully support encrypted browsing by default; a recent study showed that more than 85% of the most popular health, news, and shopping sites do not encrypt by default [57]; migrating a website to HTTPS is challenging doing so requires all third-party domains on the site (including advertisers) to use HTTPS. Second, even encrypted traffic may still reveal a lot about user behavior: the presence of any communication at all may be revealing, and website fingerprinting can reveal information about content merely based on the size, content, and location of third-party resources that a client loads. DNS traffic is also quite revealing and is essentially never encrypted [57]. Third, ISPs often terminate TLS connections, conducting man-in-the-middle attacks on encrypted traffic for network management purposes [27]. Circumventing surveillance thus requires not only encryption, but also mechanisms for controlling where traffic goes in the first place.

In this paper, we study two questions: (1) Which countries do *default* Internet routing paths traverse?; (2) What methods can help increase hosting and path diversity to help governments and citizens better control transnational Internet paths? In contrast to previous work [34], which simulates Internet paths, we *actively measure* and analyze the paths originating in five different countries: Brazil, Netherlands, Kenya, India, and the United States. We study these countries for different reasons, including their efforts made to avoid certain countries, efforts in building out IXPs, and their low cost of hosting domains. Our work studies the router-level forwarding path, which differs from all other work in this area, which has focused on analyzing Border Gateway Protocol (BGP) routes [34, 50]. Although BGP routing can offer useful information about paths, it does not necessarily reflect the path that traffic actually takes, and it only provides AS-level granularity, which is often too coarse to make strong statements about which countries that traffic is traversing. In contrast, we measure traffic routes from RIPE Atlas probes in five countries to the Alexa Top 100 domains for each country; we directly measure the paths not only to the websites corresponding to the themselves, but also to the sites hosting any third-party content on each of these sites.

Determining which countries a client's traffic traverses is challenging, for several reasons. First, performing direct measurements is more costly than passive analysis of BGP routing tables; RIPE Atlas, in particular, limits the rate at which one can perform measurements. As a result, we had to be strategic about the origins and destinations that we selected for our study. As we explain in Section 2, we study five geographically diverse countries, focusing on countries in each region that are making active attempts to thwart transnational Internet paths. Second, IP geolocation—the process of determining the geographic location of an IP address—is notoriously challenging, particularly for IP addresses that represent Internet infrastructure, rather than end-hosts. We cope with this inaccuracy by making conservative estimates of the extent of routing detours, and by recognizing that our goal is not to pinpoint a precise location for an IP address as much as to achieve accurate reports of *significant* off-path detours to certain countries or regions. (Section 3 explains our method in more detail; we also explicitly highlight ambiguities in our results.) Finally, the asymmetry of Internet paths can also make it difficult to analyze the countries that traffic traverses on the reverse path from server to client; our study finds that country-level paths are often asymmetric, and, as such, our findings represent a lower bound on transnational routing detours.

The first part of our study (Section 3) characterizes the current state of transnational Internet routing detours. We first explore hosting diversity and find that only about half of the Alexa Top 100 domains in the five countries studied are hosted in more than one country, and many times that country is a surveillance state that clients may want to avoid. Second, even if hosting diversity can be improved, routing can still force traffic through a small collection of countries (often surveillance states). Despite strong efforts made by some countries to ensure their traffic does not transit unfavorable countries [9–11,14,30], their traffic still traverses surveillance states. Over 50% of the top domains in Brazil and India are hosted in the United States, and over 50% of the paths from the Netherlands to the top domains transit the United States. About half of Kenyan paths to the top domains traverse the United States and Great Britain (but the same half does not traverse both countries). Much of this phenomenon is due to "tromboning", whereby an Internet path starts and ends in a country, yet transits an intermediate country; for example, about 13% of the paths that we explored from Brazil tromboned through the United States. Infrastructure building alone is not enough: ISPs in respective regions need better encouragements to interconnect with one another to ensure that local traffic stays local.

The second part of our work (Section 4) explores potential mechanisms for avoiding certain countries, and the potential effectiveness of these techniques. We explore two techniques: using the open DNS resolver infrastructure and using overlay network relays. We find that both of these techniques can be effective for clients in certain countries, yet the effectiveness of each technique also depends on the county. For example, Brazilian clients could completely avoid Spain, Italy, France, Great Britain, Argentina, and Ireland (among others), even though the default paths to many popular Brazilian sites traverse these countries. Additionally, overlay network relays can keep local traffic local: by using relays in the client's country, fewer paths trombone out of the client's country. The percentage of tromboning paths from the United States decreases from 11.2% to 1.3% when clients take advantage of a small number of overlay network relays.

We also find that some of the most prominent surveillance states are also some of the least avoidable countries. For example, many countries depend on ISPs in the United States, a known surveillance state, for connectivity to popular sites and content. Brazil, India, Kenya, and the Netherlands must traverse the United States to reach many of the popular local websites, even if they use open resolvers and network relays. Using overlay network relays, both Brazilian and Netherlands clients can avoid the United States for about 65% of paths; yet, the United States is completely unavoidable for about 10% of the paths because it is the only country where the content is hosted. Kenyan clients can only avoid the United States on about 55% of the paths. On the other hand, the United States can avoid every other country except for France and the Netherlands, and even then they are avoidable for 99% of the top domains.

2 State of Surveillance

We focused our study on five different countries, and for each, we actively measured and analyzed traffic that originated there. These five countries were chosen for specific reasons and we present them here. We also discuss countries that currently conduct surveillance; this exploration is not exhaustive, but highlights countries that are passing new surveillance laws and countries that have strict surveillance practices already.

2.1 Studied Countries

We selected Brazil, Netherlands, Kenya, India, and the United States for the following reasons.

Brazil. It has been widely publicized that Brazil is actively trying to avoid having their traffic transit the United States. They have been building IXPs, deploying underwater cables to Europe, and pressuring large U.S. companies to host content within Brazil [7–12, 14, 30]. This effort to avoid traffic transitting a specific country led us to investigate whether their efforts have been successful or not.

Netherlands. We selected to study the Netherlands for three reasons: 1) the Netherlands is beginning to emerge as a site where servers are located for cloud services, such as Akamai, 2) the Netherlands is where a large IXP is located (AMS-IX), and 3) they are drafting a mass surveillance law [39]. Analyzing the Netherlands will allow us to see what effect AMS-IX and the emergence of cloud service hosting has on their traffic.

Kenya. Prior research on the interconnectivity of Africa [22, 28] led us to explore the characterization of an African country's interconnectivity. We chose Kenya for a few reasons: 1) it is a location with many submarine cable landing points, 2) it has high Internet access and usage (for the East African region), and 3) it has more than one IXP [1, 52].

India. India has one of the highest number of Internet users in Asia, second only to China, which has already been well-studied [53, 56]. With such a high number of Internet users, and presumably a large amount of Internet traffic, we study India to see where this traffic is going.

United States. We chose to study the United States because of how inexpensive it is to host domains there, the prevalence of Internet and technology companies located there, and because it is a known surveillance state.

2.2 Surveillance States

When analyzing which countries Internet traffic traverse, special attention should be given to countries that may be unfavorable because of their surveillance laws. Some of the countries that are currently conducting surveillance are the "Five Eyes" [21, 36] (the United States, Canada, United Kingdom, New Zealand, and Australia), as well as France, Germany, Poland, Hungary, Russia, Ukraine, Belarus, Kyrgyzstan, and Kazakhstan.

Five Eyes. The "Five Eyes" participants are the United States National Security Agency (NSA), the United Kingdom's Government Communications Headquarters (GCHQ), Canada's Communications Security Establishment Canada (CSEC), the Australian Signals Directorate (ASD), and New Zealand's Government Communications Security Bureau (GCSB) [21]. According to the original agreement, the agencies can: 1) collect traffic; 2) acquire communications documents and equipment; 3) conduct traffic analysis; 4) conduct cryptanalysis; 5) decrypt and translate; 6) acquire information about communications organizations, procedures, practices, and equipment. The agreement also implies that all five countries will share all intercepted material by default. The agencies work so closely that the facilities are often jointly staffed by members of the different agencies, and it was reported "that SIGINT customers in both capitals seldom know which country generated either the access or the product itself." [36].

A number of other countries are passing laws to facilitate mass surveillance. These laws have differing levels of intensity, which can be seen in Table 1; the countries with the least intense surveillance laws are listed at the top of the table, and those with the more intense laws are listed at the bottom. These countries, along with the "Five Eyes" participants should be flagged when characterizing transnational detours in the following section.

	Collecting Metadata (Phone, Internet)	Requiring ISPs to Participate	No Need for Court Order	Targeted Surveillance
France	✓ [20, 23]	✓ [23]		
Germany	✓ [26]			
UA Emirates				✓ [25]
Bahrain				✓ [3]
Australia	✓ [21]			
New Zealand	✓ [21]			
Canada	✓ [21]			
United States	✓ [21]			
Great Britain	✓ [21]			
Poland	✓ [20]		✓ [20]	
Hungary	✓ [20]		✓ [20]	
Ukraine	✓ [20]	✓ [42, 49]		
Belarus	✓ [20]	✓ [42, 49]		
Kyrgyzstan	✓ [20]	✓ [42, 49]		
Kazakhstan	✓ [20]	✓ [42, 49]		
Russia	✓ [20]	✓ [42, 49]		

Table 1: *Some countries that actively conduct surveillance.*

3 Characterizing Transnational Detours

In this section, we describe our measurement methods, the challenges in conducting them, and our findings concerning the transnational detours of default Internet paths.

3.1 Measurement Pipeline

Figure 2 summarizes our measurement process, which the rest of this section describes in detail. We analyze traceroute measurements to discover which countries are on the path from a client in a particular country to a popular domain. Using traceroutes to measure transnational detours is new; prior work used BGP routing tables to *infer* country-level paths [34]. Because we conduct active measurements, which are limited by our resources, we make a tradeoff and study five countries, as opposed to all countries' Internet paths. We report on measurements that we conducted on January 31, 2016.

3.1.1 Resource Limitations

The iPlane [37] and Center for Applied Internet Data Analysis (CAIDA) [16] projects maintain two large repositories of traceroute data, neither of which turn out to be suitable for our study. iPlane measurements use PlanetLab [44] nodes and has historical data as far back as 2006. Unfortunately, because iPlane uses PlanetLab nodes, which mostly use the Global Research and Education Network (GREN), the traceroutes from PlanetLab nodes will not be representative of typical Internet users' traffic paths [5]. CAIDA runs traceroutes from different vantage points around the world to randomized destination IP addresses that cover all /24s; in contrast, we focus on paths to popular websites from a particular country.

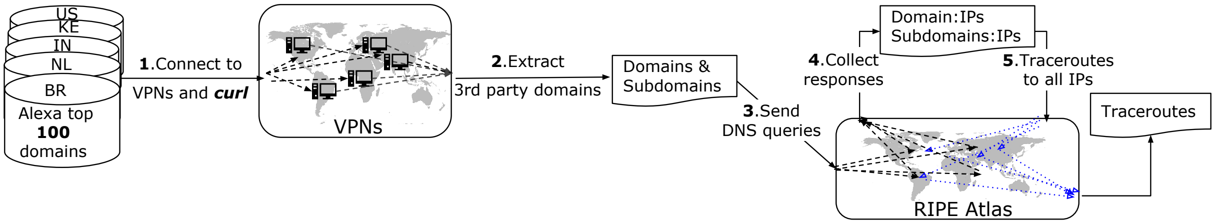


Figure 1: *Measurement pipeline to study Internet paths from countries to popular domains.*

In contrast to these existing studies, we run active measurements that would represent paths of a typical Internet user. To do so, we run DNS and traceroute measurements from RIPE Atlas probes, which are hosted all around the world and in many different settings, including home networks [46]. RIPE Atlas probes can use the local DNS resolver, which would give us the best estimate of the traceroute destination.

Yet, conducting measurements from a RIPE Atlas probe costs a certain amount of "credits", which restricts the number of measurements that we could run. RIPE Atlas also imposes rate limits on the number of concurrent measurements and the number of credits that an individual user can spend per day. We address these challenges in two ways: (1) we reduce the number of necessary measurements we must run on RIPE Atlas probes by conducting traceroute measurements to a single IP address in each /24 (as opposed to all IP address returned by DNS) because all IP addresses in a /24 belong to the same AS, and should therefore be located in the same geographic area; (2) we use a different method—VPN connections—to obtain a vantage point within a foreign country, which is still representative of an Internet user in that country.

3.1.2 Path Asymmetry

The reverse path is just as important as (and often different from) the forward path. Previous work has shown that paths are not symmetric most of the time—the forward path from point A to point B does not match the reverse path from point B to point A [29]. Most work on path asymmetry has been done at the AS level, but not at the country level. Our measurements consider only the forward path (from client to domain or relay), not the reverse path from the domain or relay to the client.

We measured path asymmetry at the country granularity. If country-level paths are symmetric, then the results of our measurements would be representative of the forward *and* reverse paths. If the country-level paths are asymmetric, then our measurement results only provide a lower bound on the number of countries that could potentially conduct surveillance. Using 100 RIPE Atlas probes located around the world, and eight Amazon EC2 instances, we ran traceroute measurements from every probe to every EC2 instance and from every EC2 instance to every probe. After mapping the IPs to countries, we analyzed the paths for symmetry. First, we compared the set of countries on the forward path to the set of countries on the reverse path; this yielded about 30% symmetry. What we wanted to know is whether or not the reverse path has more countries on it than the forward path. Thus, we measured how many reverse paths were a subset of the respective forward path; this was the case for 55% of the paths. This level of asymmetry suggests that our results represent a lower bound on the number of countries that transit traffic; our results are a lower bound on how many unfavorable countries transit a client's path. It also suggests that while providing lower bounds on transnational detours is feasible, designing systems to completely prevent these detours on both forward and reverse paths may be particularly challenging, if not impossible.

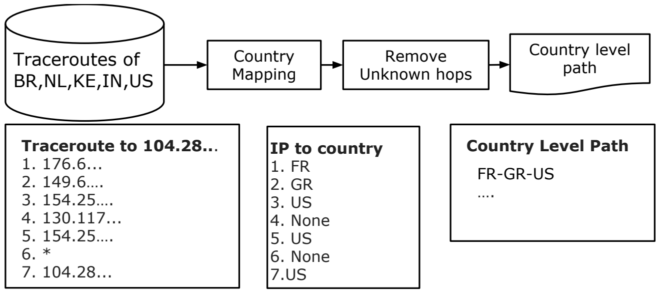


Figure 2: *Mapping country-level paths from traceroutes.*

3.1.3 Traceroute Origination and Destination Selection

Each country hosts a different number of RIPE Atlas probes, ranging from about 75 probes to many hundreds. Because of the resource restrictions, we could not use all probes in each of the countries. We selected the set of probes that had unique ASes in the country to get the widest representation of origination (starting) points.

For destinations, we used the Alexa Top 100 domains in each of the respective countries, as well as the third-party domains that are requested as part of an original web request. To obtain these 3rd party domains we `curl` (*i.e.*, HTTP fetch) each of the Top 100 domains, but we must do so from within the country we are studying. There is no current functionality to `curl` from RIPE Atlas probes, so we establish a VPN connection within each of these countries to `curl` each domain and extract the third-party domains; we `curl` from the client's location in case web sites are customizing content based on the region of the client.

Terminating in Country	Brazil	Netherlands	India	Kenya	United States
Brazil	.169	-	-	-	-
Canada	.001	.007	.015	.006	-
United States	.774	.454	.629	.443	.969
France	.001	.022	.009	.023	.001
Germany	.002	.013	.014	.028	.001
Great Britain	-	.019	.021	.032	.002
Ireland	.016	.064	.027	.108	.001
Netherlands	.013	.392	.101	.200	.024
Spain	.001	-	-	-	-
Kenya	-	-	-	.022	-
Mauritius	-	-	-	.004	-
South Africa	-	-	-	.021	-
United Arab Emirates	-	-	-	.011	-
India	-	-	.053	.002	-
Singapore	-	.002	.103	.027	-

Table 2: *Fraction of paths that terminate in each country by default.*

Transiting Country	Brazil	Netherlands	India	Kenya	United States
Brazil	1.00	-	-	-	-
Canada	.013	.007	.016	.008	.081
United States	.844	.583	.715	.616	1.00
France	.059	.102	.104	.221	.104
Germany	.005	.050	.032	.048	.008
Great Britain	.024	.140	.204	.500	.006
Ireland	.028	.106	.031	.133	.006
Netherlands	.019	1.00	.121	.253	.031
Spain	.176	.004	-	-	-
Kenya	-	-	-	1.00	-
Mauritius	-	-	-	.322	-
South Africa	-	-	-	.334	-
United Arab Emirates	-	-	-	.152	-
India	-	-	1.00	.058	-
Singapore	-	.002	.270	.040	.003

Table 3: *Fraction of paths that each country transits by default.*

3.1.4 Country Mapping

Accurate IP geolocation is challenging. We use MaxMind's geolocation service to map IP addresses to their respective countries [38], which is known to contain inaccuracies. Fortunately, our study does not require high-precision geolocation; we are more interested in providing accurate lower bounds on detours at a much coarser granularity. Fortunately, previous work has found that geolocation at a country-level granularity is more accurate than at finer granularity [31]. In light of these concerns, we post-processed our IP to country mapping by removing all IP addresses that resulted in a 'None' response when querying MaxMind, which causes our results to provide a conservative estimate of the number of countries that paths traverse. It is important to note that removing 'None' responses will *always* produce a conservative estimate, and therefore we are *always* underestimating the amount of potential surveillance. Figure 2 shows an example of this post-processing.

3.2 Results

Table 2 shows the five countries we studied along the top of the table, and the countries that host their content along in each row. For example, the United States is the endpoint of 77% of the paths that originate in Brazil. A "-" represents the case where no paths ended in that country. For example, no Brazilian paths terminated in South Africa. Table 3 shows the fraction of paths that transit certain countries, with a row for each country that is transited.

Finding 3.1 (Hosting Diversity)**:** *About half of the top domains in each of the five countries studied are hosted in a single country. The other half are located in two or more different countries.*

First we analyze hosting diversity; this shows us how many unique countries host a domain. The more countries that a domain is hosted in creates a greater chance that the content is replicated in a favorable country, and could potentially allow a client to circumvent an unfavorable country. We queried DNS from 26 vantage points around the world, which are shown in Figure 3; we chose this set of locations because they are geographically diverse. Then we mapped the IP addresses in the DNS responses to their respective countries to determine how many unique countries a domain is hosted in. Figure 4 shows the fraction of domains that are hosted in different numbers of countries; we can see two common hosting cases: 1) CDNs, and 2) a single hosting country. This shows that many domains are hosted in a single unique country, which leads us to our next analysis—where are these domains hosted, and which countries are traversed on the way to reach these locations.

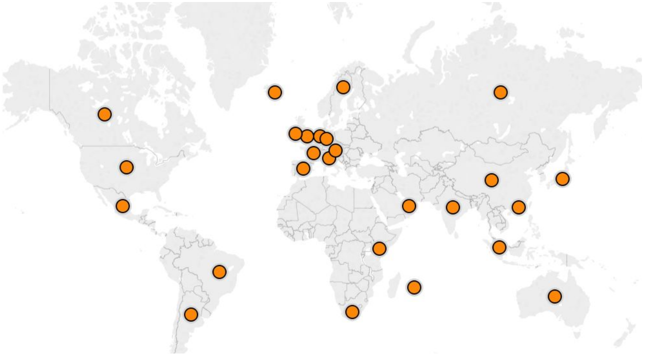
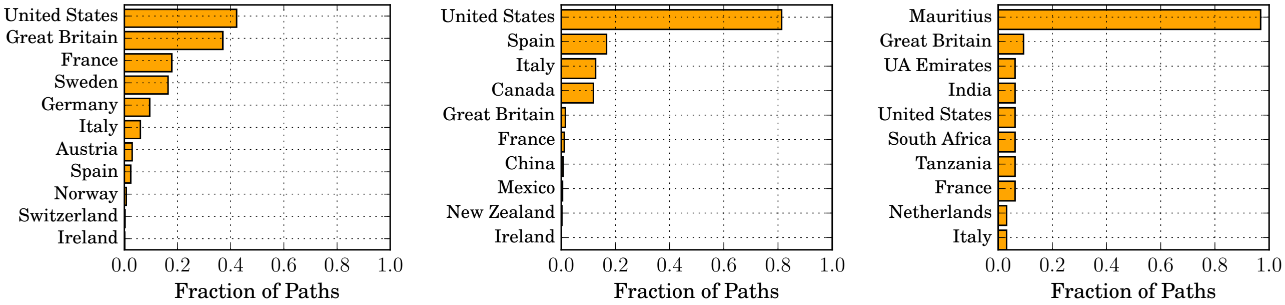


Figure 3: *The locations of vantage points in measuring hosting diversity.*

Finding 3.2 (Domain Hosting)**:** *The most common destination among all five countries studied is the United States: 77%, 45%, 63%, 44%, and 97% of paths originating in Brazil, Netherlands, India, Kenya and the United States, respectively, are currently reaching content located in the United States.*



(a) *The Netherlands.*

(b) *Brazil.*

(c) *Kenya.*

Figure 5: *The countries that tromboning paths from the Netherlands, Brazil, and Kenya transit.*

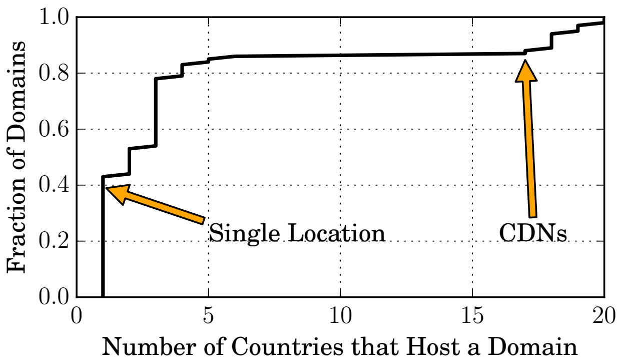


Figure 4: *The number of Alexa Top 100 US Domains hosted in different countries.*

Table 2 shows the fraction of paths that are hosted in various countries. Despite the extent of country-level hosting diversity, the majority of paths from all five countries terminate in a single country: the United States, a known surveillance state. Our results also show the Netherlands is a common hosting location for paths originating in the Netherlands, India, and Kenya.

Finding 3.3 (Domestic Traffic)**:** *All of the countries studied (except for the United States) host content for a small percentage of the paths that originate in their own country; they also host a small percentage of their respective country-code top-level domains.*

Only 17% of paths that originate in Brazil also end there. Only 5% and 2% of Indian and Kenyan paths, respectively, end in the originating country. For Kenya, 24 out of the Top 100 Domains are .ke domains, and of these 24 domains only 5 are hosted within Kenya. 29 out of 40 .nl domains are hosted in the Netherlands; 4 of 13 .in domains are hosted in India; 18 of 39 .br domains are hosted in Brazil. Interestingly, all .gov domains were hosted in their respective country.

Finding 3.4 (Transit Traffic)**:** *Surveillance states (specifically the United States and Great Britain) are on the largest portion of paths in comparison to any other (foreign) country.*

84% of Brazilian paths traverse the United States, despite Brazil's strong efforts to avoid United States surveillance. Although India and Kenya are geographically distant, 72% and 62% of their paths also transit the United States.

Great Britain and the Netherlands are on the path for a significant percentage of paths originating in India and Kenya: 50% and 20% of paths that originate in Kenya and India, respectively, transit Great Britain. Many paths likely traverse Great Britain and the Netherlands due to the presence of large Internet Exchange Points (*i.e.*, LINX, AMS-IX). Mauritius, South Africa, and the United Arab Emirates transit 32%, 33%, and 15% of paths from Kenya. There are direct underwater cables from Kenya to Mauritius, and from Mauritius to South Africa [51]. Additionally, there is a cable from Mombasa, Kenya to Fujairah, United Arab Emirates, which likely explains the large fraction of paths that include these countries.

Finding 3.5 (Tromboning Traffic)**:** *Brazilian and Netherlands paths often trombone to the United States, despite the prevalence of IXPs in both countries.*

Figures 5a, 5b, and 5c show the fraction of paths that trombone to different countries for the Netherlands, Brazil, and Kenya. 24% of all paths originating in the Netherlands (62% of domestic paths) trombone to a foreign country before returning to the Netherlands. Despite Brazil's strong efforts in building IXPs to keep local traffic local, we can see that their paths still trombone to the United States. This is due to IXPs being seen as a threat by competing commercial providers; providers are sometimes concerned that "interconnection" will result in making business cheaper for competitors and stealing of customers [45]. It is likely that Brazilian providers see other Brazilian providers as competitors and therefore as a threat at IXPs, which cause them to peer with international providers instead of other local providers. Additionally, we see Brazilian paths trombone to Spain and Italy. We have observed that MaxMind sometimes mislabels IP addresses to be in Spain when they are actually located in Portugal. This mislabelling does not affect our analysis of detours through surveillance states, as we do not highlight either Spain or Portugal as a surveillance state. We see Italy often in tromboning paths because Telecom Italia Sparkle is one of the top global Internet providers [4].

Tromboning Kenyan paths most commonly traverse Mauritius, which is expected considering the submarine cables between Kenya and Mauritius. Submarine cables also explain South Africa, Tanzania, and the United Arab Emirates on tromboning paths.

Finding 3.6 (United States as an Outlier)**:** *The United States hosts 97% of the content that is accessed from within the country, and only five foreign countries—France, Germany, Ireland, Great Britain, and the Netherlands—host content for the other 3% of paths.*

Many of the results find that Brazilian, Netherlands, Indian, and Kenyan paths often transit surveillance states, most notably the United States. The results from studying paths that originate in the United States are drastically different from those of the other four countries. The other four countries host very small amounts of content accessed from their own country, whereas the United States hosts 97% of the content that is accessed from within the country. Only 13 unique countries are ever on a path from the United States to a domain in the top 100 (or third party domain), whereas 30, 30, 25, and 38 unique countries are seen on the paths originating in Brazil, Netherlands, India, and Kenya, respectively.

3.3 Limitations

This section discusses the various limitations of our measurement methods and how they may affect the results that we have reported.

Traceroute accuracy and completeness. Our study is limited by the accuracy and completeness of traceroute. Anomalies can occur in traceroute-based measurements [2], but most traceroute anomalies do not cause an overestimation in surveillance states. The incompleteness of traceroutes, where a router does not respond, causes our results to underestimate the number of surveillance states, and therefore also provides a lower bound on surveillance.

IP Geolocation vs. country mapping. Previous work has shown that there are fundamental challenges in deducing a geographic location from an IP address, despite using different methods such as DNS names of the target, network delay measurements, and host-to-location mapping in conjunction with BGP prefix information [43]. While it has been shown that there are inaccuracies and incompleteness in MaxMind's data [31], the focus of this work is on measuring and avoiding surveillance. We use Maxmind to map IP to country (as described in Section 3.1.4), which provides a lower bound on the amount of surveillance, as we have described.

IPv4 vs. IPv6 connectivity. The measurements we conducted only collect and analyze IPv4 paths, and therefore all IPv6 paths are left out of our study. IPv6 paths likely differ from IPv4 paths as not all routers that support IPv4 also support IPv6. Future work includes studying IPv6 paths and which countries they transit, as well as a comparison of country avoidability between IPv4 and IPv6 paths.

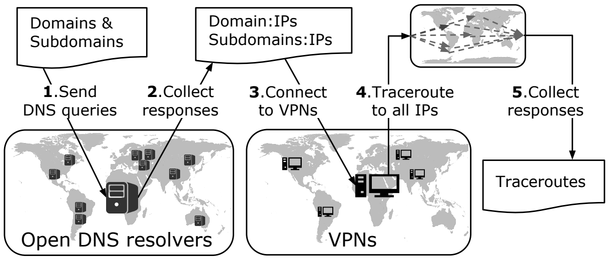


Figure 6: *Measurement approach for country avoidance with open DNS resolvers.*

4 Preventing Transnational Detours

In light of our analysis of the state of default Internet paths from Section 3, we now explore the extent to which various techniques and systems can help clients in various countries prevent unwanted transnational routing detours. We explore two different mechanisms for increasing path diversity: discovering additional website replicas by diverting DNS queries through global open DNS resolvers and creating additional network-layer paths with the use of overlay nodes. We discuss our measurement methods, develop an avoidance metric and algorithm, and present our results.

4.1 Measurement Approach

Country Avoidance with Open Resolvers. If content is replicated on servers in different parts of the world, open DNS resolvers located around the world may also help clients discover a more diverse set of replicas.

Figure 6 illustrates our measurement approach for this study, which differs slightly from that described in Section 3.1: instead of using RIPE Atlas probes to query local DNS resolvers, we query open DNS resolvers located around the world [33]. These open DNS resolvers may provide different IP addresses in the DNS responses, which represent different locations of content replicas. The measurement study in Section 3.1 used RIPE Atlas probes to traceroute to the IP addresses in DNS response; in contrast, for this portion of the study we initiate a VPN connection to the client's country and traceroute (through the VPN connection) to the IP addresses in the DNS responses returned by the open resolvers.

Country Avoidance with Relays. Using an overlay network may help clients route around unfavorable countries or access content that is hosted in a different country. Figure 7 shows the steps to conduct this measurement. After selecting relay machines, we run traceroute measurements from Country X to each relay and from each relay to the set of domains. We then analyze these traceroutes using the pipeline in Figure 2 to determine country-level paths.

We use eight Amazon EC2 instances, one in each geographic region (United States, Ireland, Germany, Singapore, South Korea, Japan, Australia, Brazil), as well as 4 Virtual Private Server (VPS) machines (France, Spain, Brazil, Singapore), which are virtual machines that are functionally equivalent to dedicated physical servers. The conjunction of

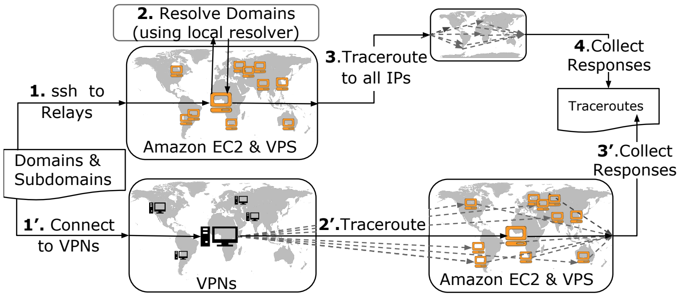


Figure 7: *Measurement approach for country avoidance with overlay network relays.*

these two sets of machines allow us to evaluate surveillance avoidance with a geographically diverse set of relays. By selecting an open resolver in each country that also has a relay in it we can keep the variation in measurement methods low, leading to a more accurate comparison of country avoidance methods.

4.2 Avoidability Metrics

We introduce a new metric and algorithm to measure how often a client in Country X can avoid a specific country Y. Using the proposed metric and algorithm, we can compare how well the different methods achieve country avoidance for any (X, Y) pair.

Avoidability metric. We introduce an avoidability metric to quantify how often traffic can avoid Country Y when it originates in Country X. Avoidability is the fraction of paths that start in Country X and do not transit Country Y. We calculate this value by dividing the number of paths from Country X to domains that do not traverse Country Y by the total number of paths from Country X. The resulting value will be in the range [0,1], where 0 means the country is unavoidable for all of the domains in our study, and 1 means the client can avoid Country Y for all domains in our study. For example, there are three paths originating in Brazil: (1) $BR \rightarrow US$, (2) $BR \rightarrow CO \rightarrow None$, 3) $BR \rightarrow *** \rightarrow BR$. After processing the paths as described in Section 3.1.4, the resulting paths are: (1) $BR \rightarrow US$, (2) $BR \rightarrow CO$, (3) $BR \rightarrow BR$. The avoidance value for avoiding the United States would be 2/3 because two out of the three paths do not traverse the United States. This metric represents a lower bound, because it is possible that the third path timed out ($***$) because it traversed the United States, which would make the third path: $BR \rightarrow US \rightarrow BR$, and would cause the avoidance metric to drop to 1/3.

Avoidability algorithm with open resolvers. Recall from the measurement pipeline for avoidance with open resolvers, described in Section 4.1, that the resulting data are traceroutes from the client in Country X to *all* IP addresses in *all* open DNS resolver responses. To measure avoidability, there must exist at least one path from the client in Country X to the domain for the client to be able to avoid Country Y when accessing the domain. The country avoidance value is the fraction of domains accessible from the client in Country X without traversing Country Y.

Algorithm 1 Avoidability Algorithm

```
function CALCAVOIDANCE(set $paths1$, set $paths2$, string c)
    set $usableRelays$
    for each $(relay, path)$ in $paths1$ do
        if $c$ not in $path$ then
            $usableRelays \leftarrow path$
    set $accessibleDomains$
    for each $(relay, domain, path$ in $paths2$ do
        if $relay$ in $usableRelays$ then
            if $c$ not in $path$ then
                $accessibleDomains \leftarrow domain$
    $D \leftarrow$ number of all unique domains in $paths2$
    $A \leftarrow$ length of $accessibleDomains$
    return $A/D$
```

Avoidability algorithm with relays. Measuring the avoidability of a Country Y from a client in Country X using relays has two components: (1) Is Country Y on the path from the client in Country X to the relay? (2) Is Country Y on the path from the relay to the domain? For every domain, our algorithm checks if there exists at least one path from the client in Country X through any relay and on to the domain, and does not transit Country Y. The algorithm (Algorithm 1) produces a value in the range [0,1] that can be compared to the output of the avoidability metric described above.

Upper bound on avoidability. Although the avoidability metric and algorithm provide a method to quantify how avoidable Country Y is from a client in Country X, it may be the case that a number of domains are only hosted in Country Y, so the avoidance value for these countries would never reach 1.0. For this reason, we measured the *upper bound* on avoidance for a given pair of (Country X, Country Y) that represents the best case value for avoidance. Algorithm 2 shows the pseudocode for computing this metric. The algorithm analyzes the destinations of all domains from all relays and if there exists at least one destination for a domain that is not in Country Y, then this increases the upper bound value. An upper bound value of 1.0 means that every domain studied is hosted (or has a replica) outside of Country Y. This value puts the avoidance values in perspective for each (Country X, Country Y) pair.

4.3 Results

We compared avoidance values when using open resolvers, when using relays, and when using no country avoidance tool. First, we discuss how effective open resolvers are at country avoidance. We then examine the effectiveness of relays for country avoidance, as well as for keeping local traffic local. Table 4 shows avoidance values; the top row shows the countries we studied and the left column shows the country that the client aims to avoid.

	No Relay	Open Resolvers	Relays	No Relay	Open Resolvers	Relays	No Relay	Open Resolvers	Relays	No Relay	Open Resolvers	Relays	No Relay	Open Resolvers	Relays
Country to Avoid	*Brazil*			*Netherlands*			*India*			*Kenya*			*United States*		
Brazil	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Canada	.98	1.00	1.00	.99	1.00	1.00	.98	.98	.98	.99	.99	.99	.92	1.00	1.00
United States	.15	.19	.62	.41	.57	.63	.28	.45	.65	.38	.55	.40	0.00	0.00	0.00
France	.94	.98	1.00	.89	.96	.99	.89	.98	1.00	.77	.89	.98	.89	.99	.99
Germany	.99	.99	1.00	.95	.98	.99	.96	.97	.99	.95	.99	1.00	.99	.99	1.00
Great Britain	.97	.97	1.00	.86	.87	.99	.79	.79	1.00	.50	.71	.97	.99	.99	1.00
Ireland	.97	.98	.99	.89	.97	.99	.96	.99	.99	.86	.98	.99	.99	.99	.99
Netherlands	.98	.98	.99	0.00	0.00	0.00	.87	.98	.99	.74	.98	.99	.97	.99	.99
Spain	.82	1.00	1.00	.99	.99	.99	1.00	.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Kenya	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	1.00	1.00	1.00
Mauritius	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.67	.97	.99	1.00	1.00	1.00
South Africa	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.66	.87	.66	1.00	1.00	1.00
United Arab Emirates	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.84	1.00	.99	1.00	1.00	1.00
India	1.00	1.00	1.00	.99	1.00	1.00	0.00	0.00	0.00	.94	.94	1.00	.99	1.00	1.00
Singapore	.99	.99	1.00	.99	.99	1.00	.73	.92	.94	.96	.96	1.00	.99	.99	1.00

Table 4: *Avoidance values for different techniques of country avoidance. The upper bound on avoidance is 1.0 in most cases, but not all. It is common for some European countries to host a domain, and therefore the upper bound is slightly lower than 1.0. The upper bound on avoidance of the United States is significantly lower than the upper bound on avoidance for any other country; .886, .790, .844, and .765 are the upper bounds on avoidance of the United States for paths originating in Brazil, Netherlands, India, and Kenya, respectively.*

Algorithm 2 Avoidance Upper Bound Algorithm

```
function CALCUPPERBOUND(set $relayDomainPaths$, string $c$)
    $zeros(domainLocations)$
    for each $(r, d, p)$ in $relayDomainPaths$ do
        $dest \leftarrow$ last item in $p$
        $domainLocations[d] \leftarrow dest$
    set $accessibleDomains$
    for each $domain$ in $domainLocations$ do
        if $domainLocations[domain] \neq \text{set}[c]$ then
            $accessibleDomains \leftarrow domain$
    $D \leftarrow$ all unique domains in $relayDomainPaths$
    $A \leftarrow$ length of $accessibleDomains$
    return $A/D$
```

4.3.1 Avoidance with Open Resolvers

A given country is more avoidable (higher avoidance value) when open resolvers are used as a tool for country avoidance.

Finding 4.1 (Open Resolver Effectiveness)**:** *Using open DNS resolvers for country avoidance achieves more country avoidance than using local resolvers and less (or equal) avoidance than using relays for clients in most countries.*

For Brazilian paths, open resolvers only achieve 4% more avoidance than using local resolvers when avoiding the United States, whereas relays achieve 47% more avoidance. On the other hand, open resolvers are about as effective as relays are for avoidance for paths originating in the United States.

Finding 4.2 (Kenya as an Outlier)**:** *For clients in Kenya, open DNS resolvers are significantly more effective than relays for avoiding the United States, South Africa, and the United Arab Emirates.*

Clients in Kenya should use open DNS resolvers when avoiding specific countries, as they can avoid these specific countries more often than when using relays. Kenyan clients can avoid the United States for 55% of paths when using open resolvers, whereas they can only avoid the United States for 40% of paths when using relays. The difference in how often the United States can be avoided can be attributed to the lower amount of DNS diversity when using relays as compared to using open resolvers. For a client in Kenya trying to avoid the United States, the client can only use the relay located in Ireland (because all paths from the client to the other relays traverse the United States), and therefore only gets DNS responses from locally resolving domains on the Ireland relay. When using open resolvers, the client gets more DNS diversity as he gets DNS responses from all open resolvers located in different countries.

The amount of avoidance Kenyan clients can achieve for avoiding South Africa is the same, regardless of whether the client is using relays, because all paths between the client and the relays traverse South Africa. Fortunately, clients can avoid South Africa for significantly more paths when using open resolvers, likely as a result of the fact that open DNS resolvers can better uncover hosting diversity.

4.3.2 Avoidance with Relays

As seen in Table 4, there are two significant trends: 1) the ability for a client to avoid a given Country Y increases with the use of relays, and 2) the least avoidable countries are surveillance states.

Finding 4.3 (Relay Effectiveness)**:** *For 84% of the (Country X, Country Y) pairs shown in Table 4 the avoidance with relays reaches the upper bound on avoidance.*

In almost every (Country X, Country Y) pair, where Country X is the client's country (Brazil, Netherlands, India, Kenya, or the United States) and Country Y is the country to avoid, the use of an overlay network makes Country Y more avoidable than the default routes. The one exception we encountered is when a client is located in Kenya and wants to avoid South Africa, where, as mentioned, all paths through our relays exit Kenya via South Africa.

Finding 4.4 (Relays Achieve Upper Bound)**:** *Clients in the United States can achieve the upper bound of avoidance for all countries—relays help clients in this country avoid all other Country Y in all cases that the domain is not hosted in Country Y.*

Relays are most effective for clients in the United States. On the other hand, it is much rarer for (Kenya, Country Y) pairs to achieve the upper bound of surveillance, showing that it is more difficult for Kenyan clients to avoid a given country. This is not to say that relays are not effective for clients in Kenya; for example, the default routes to the top 100 domains for Kenyans avoid Great Britain 50% of the time, but with relays this percentage increases to about 97% of the time, and the upper bound is about 98%.

Finding 4.5 (Surveillance States are Less Avoidable)**:** *The ability for any country to avoid the United States is significantly lower than it's ability to avoid any other country in all four situations: without relays, with open resolvers, with relays, and the upper bound.*

Despite increasing clients' ability to avoid the United States, relays are not as effective at helping clients avoid this country as compared to the effectiveness of the relays at avoiding all other Country Y. Clients in India can avoid the United States more often than clients in Brazil, Netherlands, and Kenya, by avoiding the United States for 65% of paths. Kenyan clients can only avoid the United States 40% of the time even while using relays. Additionally, the upper bound for avoiding the United States is significantly lower in comparison to any other country.

Finding 4.6 (Keeping Local Traffic Local)**:** *Using relays decreased both the number of tromboning paths, and the number of countries involved in tromboning paths.*

For the cases where there were relays located in one of the five studied countries, we evaluated how effectively the use of relays kept local traffic local. This evaluation was possible for Brazil and the United States. Tromboning Brazilian paths decreased from 13.2% without relays to 9.7% with relays; when relays are used, all tromboning paths goes only to the United States. With the use of relays, there was only 1.3% tromboning paths for a United States client, whereas without relays there was 11.2% tromboning paths. For the 1.2% of paths that trombones from the United States, it goes only to Ireland.

5 Discussion

Avoiding multiple countries. We have studied only the extent to which Internet paths can be engineered to avoid a single country. Yet, avoiding a single country may force an Internet path into *other* unfavorable jurisdictions. This possibility suggests that we should also be exploring the feasibility of avoiding multiple surveillance states (*e.g.*, the "Five Eyes") or perhaps even entire regions. It is already clear that avoiding certain combinations of countries is not possible, at least given the current set of relays; for example, to avoid the US, Kenyan clients rely on the relay located in Ireland, so avoiding both countries is often impossible.

The evolution of routing detours and avoidance over time. Our study is based on a snapshot of Internet paths. Over time, paths change, hosting locations change, IXPs are built, submarine cables are laid, and surveillance states change. Future work can and should involve exploring how these paths evolve over time, and analyzing the relative effectiveness of different strategies for controlling traffic flows.

Isolating DNS diversity vs. path diversity. In our experiments, the overlay network relays perform DNS lookups from geographically diverse locations, which provides some level of DNS diversity in addition to the path diversity that the relays inherently provide. This approach somewhat conflates the benefits of DNS diversity with the benefits of path diversity and in practice may increase clients' vulnerability to surveillance, since each relay is performing DNS lookups on each client's behalf. We plan to conduct additional experiments where the client relies on its local DNS resolver to map domains to IP addresses, as opposed to relying on the relays for both DNS resolution and routing diversity.

6 Related Work

Nation-state routing analysis. Recently, Shah and Papadopoulos measured international BGP detours (paths that originate in one country, cross international borders, and then return to the original country) [50]. Using BGP routing tables, they found 2 million detours in each month of their study (out of 7 billion total paths), and they then characterized the detours based on detour dynamics and persistence. Our work differs by actively measuring traceroutes (actual paths), as opposed to analyzing BGP routes. This difference is fundamental as BGP provides the AS path announced in BGP update messages, which is not necessarily the same as the actual path of data packets. Obar and Clement analyzed traceroutes that started and ended in Canada, but "boomeranged" through the United States ("boomerang" is another term for tromboning), and argued that this is a violation of Canadian network sovereignty [41]. Most closely related to our work, Karlin et al. developed a framework for country-level routing analysis to study how much influence each country has over interdomain routing [34]. This work measures the centrality of a country to routing and uses AS-path inference to measure and quantify country centrality, whereas our work uses active measurements and measures avoidability of a given country.

Mapping national Internet topologies. In 2011, Roberts et al. described a method for mapping national networks of ASes, identifying ASes that act as points of control in the national network, and measuring the complexity of the national network [47]. There have also been a number of studies that measured and classified the network within a country. Wahlisch et al. measured and classified the ASes on the German Internet [54, 55], Zhou et al. measured the complete Chinese Internet topology at the AS level [58], and Bischof et al. characterized the current state of Cuba's connectivity with the rest of the world [6]. Interconnectivity has also been studied at the continent level; Gupta et al. first looked at ISP interconnectivity within Africa [28], and it was studied later by Fanou et al. [22].

Circumvention Systems. There has been research into circumvention systems, particularly for censorship circumvention, that is related this work, but not sufficient for surveillance circumvention. Tor is an anonymity system that uses three relays and layered encryption to allow users to communicate anonymously [19]. VPNGate is a public VPN relay system aimed at circumventing national firewalls [40]. Unfortunately, VPNGate does not allow a client to choose any available VPN, which makes surveillance avoidance harder.

7 Conclusion

We have measured Internet paths to characterize routing detours that take Internet paths through countries that perform surveillance. Our findings show that paths commonly traverse known surveillance states, even when they originate and end in a non-surveillance state. As a possible step towards a remedy, we have investigated how clients can use the open DNS resolver infrastructure and overlay network relays to prevent routing detours through unfavorable jurisdictions. These methods give clients the power to avoid certain countries, as well as help keep local traffic local. Although some countries are completely avoidable, we find that some of the more prominent surveillance states are the least avoidable.

Our study presents several opportunities for follow-up studies and future work. First, Internet paths continually evolve; we will repeat this analysis over time and publish the results and data on a public website, to help deepen our collective understanding about how the evolution of Internet connectivity affects transnational routes. Second, our analysis should be extended to study the extent to which citizens in one country can avoid groups of countries or even entire regions. Finally, although our results provide strong evidence for the existence of various transnational data flows, factors such as uncertain IP geolocation make it difficult to provide clients guarantees about country-level avoidance; developing techniques and systems that offer clients stronger guarantees is a ripe opportunity for future work.

References

[1] Assessment of the Impact of Internet Exchange Points – Empirical Study of Kenya and Nigeria. http://www.internetsociety.org/sites/default/files/Assessment%20of%20the%20impact%20of%20Internet%20Exchange%20Points%20%E2%80%93%20empirical%20study%20of%20Kenya%20and%20Nigeria.pdf.

[2] B. Augustin, X. Cuvellier, B. Orgogozo, F. Viger, T. Friedman, M. Latapy, C. Magnien, and R. Teixeira. Avoiding traceroute anomalies with Paris traceroute. In *The 6th ACM SIGCOMM Internet Measurement Conference*, pages 153–158. ACM, 2006.

[3] Bahraini Activists Hacked by Their Government Go After UK Spyware Maker. https://www.wired.com/2014/10/bahraini-activists-go-after-spyware-source/.

[4] A Baker's Dozen, 2015 Edition. http://research.dyn.com/2016/04/a-bakers-dozen-2015-edition/.

[5] S. Banerjee, T. G. Griffin, and M. Pias. The interdomain connectivity of PlanetLab nodes. In *Passive and active network measurement*, pages 73–82. Springer, 2004.

[6] Z. S. Bischof, J. P. Rula, and F. E. Bustamante. In and Out of Cuba: Characterizing Cuba's Connectivity. In *The 2015 ACM Internet Measurement Conference*, pages 487–493. ACM, 2015.

[7] Brasil Internet Exchange Participants Diversity. http://ix.br/doc/nic.br.ix.br.euro-ix-27th-berlin.20151027-02.pdf.

[8] Brazil Builds Internet Cable To Portugal To Avoid NSA Surveillance. http://www.ibtimes.com/brazil-builds-internet-cable-portugal-avoid-nsa-surveillance-1717417.

[9] Brazil conference will plot Internet's future post NSA spying. http://www.reuters.com/article/us-internet-conference-idUSBREA3L1OJ20140422.

[10] Brazil Looks to Break from US Centric Internet. http://news.yahoo.com/brazil-looks-break-us-centric-internet-040702309.html.

[11] Brazil to host global internet summit in ongoing fight against NSA surveillance. https://www.rt.com/news/brazil-internet-summit-fight-nsa-006/.

[12] Brazil to press for local Internet data storage after NSA spying. https://www.rt.com/news/brazil-brics-internet-nsa-895/.

[13] Brazil Winning Internet. http://research.dyn.com/2014/07/brazil-winning-internet/#!prettyPhoto/1/.

[14] Brazil's President Tells U.N. That NSA Spying Violates Human Rights. http://www.usnews.com/news/articles/2013/09/24/brazils-president-tells-un-that-nsa-spying-violates-human-rights.

[15] S. Brito, M. Santos, R. Fontes, and D. Perez. Dissecting the Largest National Ecosystem of Public Internet eXchange Points in Brazil. 2016.

[16] CAIDA: Center for Applied Internet Data Analysis. http://www.caida.org/home/.

[17] Chinese Routing Errors Redirect Russian Traffic. http://research.dyn.com/2014/11/chinese-routing-errors-redirect-russian-traffic/.

[18] Deutsche Telekom to Push for National Routing to Curtail Spying. http://www.businessweek.com/news/2013-10-14/deutsche-telekom-to-push-for-national-routing-to-curtail-spying.

[19] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The second-generation onion router. Technical report, DTIC Document, 2004.

[20] EU Disdains U.S. Surveillance, but Seeks Easier Access. http://www.bna.com/eu-disdains-us-n57982070518/.

[21] Eyes Wide Open. https://www.privacyinternational.org/sites/default/files/Eyes%20Wide%20Open%20v1.pdf.

[22] R. Fanou, P. Francois, and E. Aben. On the diversity of interdomain routing in africa. In *Passive and Active Measurement*, pages 41–54. Springer, 2015.

[23] France Has a Powerful and Controversial New Surveillance Law. http://www.recode.net/2015/11/14/11620670/france-has-a-powerful-and-controversial-new-surveillance-law.

[24] France Must Reject Law that Gives Carte Blanche to Mass Surveillance Globally. https://www.amnesty.org/en/press-releases/2015/09/france-must-reject-law-that-gives-carte-blanche-to-mass-surveillance-globally/.

[25] Freedom on the Net: United Arab Emirates. https://freedomhouse.org/report/freedom-net/2015/united-arab-emirates.

[26] German Bundestag Passes New Data Retention Law. https://lawfareblog.com/german-bundestag-passes-new-data-retention-law.

[27] Gogo Inflight Internet serves up 'man-in-the-middle' with fake SSL. http://www.csoonline.com/article/2865806/cloud-security/gogo-inflight-internet-serves-up-man-in-the-middle-with-fake-ssl.html.

[28] A. Gupta, M. Calder, N. Feamster, M. Chetty, E. Calandro, and E. Katz-Bassett. Peering at the internet's frontier: A first look at ISP interconnectivity in Africa. In *Passive and Active Measurement*, pages 204–213. Springer, 2014.

[29] Y. He, M. Faloutsos, S. Krishnamurthy, and B. Huffaker. On routing asymmetry in the Internet. In *Global Telecommunications Conference. IEEE*, volume 2. IEEE, 2005.

[30] How Brazil Crowdsourced a Landmark Law. http://foreignpolicy.com/2016/01/19/how-brazil-crowdsourced-a-landmark-law/.

[31] B. Huffaker, M. Fomenkov, and K. Claffy. Geocompare: a comparison of public and commercial geolocation databases. *Proc. NMMC*, pages 1–12, 2011.

[32] Investigatory powers bill: snooper's charter lacks clarity, MPs warn. http://www.theguardian.com/law/2016/feb/01/investigatory-powers-bill-snoopers-charter-lacks-clarity-mps-warn.

[33] Internet-Wide Scan Data Repository. https://scans.io/study/washington-dns.

[34] J. Karlin, S. Forrest, and J. Rexford. Nation-state routing: Censorship, wiretapping, and BGP. *arXiv preprint arXiv:0903.3218*, 2009.

[35] Kazakhstan will require internet surveillance back doors. http://www.engadget.com/2015/12/05/kazakhstan-internet-back-door-law/.

[36] S. S. Lander. International intelligence cooperation: an inside perspective 1. *Cambridge Review of International Affairs*, 17(3):481–493, 2004.

[37] H. V. Madhyastha, T. Isdal, M. Piatek, C. Dixon, T. Anderson, A. Krishnamurthy, and A. Venkataramani. iPlane: An information plane for distributed services. In *The 7th Symposium on Operating Systems Design and Implementation*, pages 367–380. USENIX Association, 2006.

[38] MaxMind. https://www.maxmind.com/en/home.

[39] Netherlands New Proposal for Dragnet Surveillance Underway. https://edri.org/netherlands-new-proposals-for-dragnet-surveillance-underway/.

[40] D. Nobori and Y. Shinjo. VPN gate: A volunteer-organized public vpn relay system with blocking resistance for bypassing government censorship firewalls. In *The 11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*, pages 229–241, 2014.

[41] J. A. Obar and A. Clement. Internet surveillance and boomerang routing: A call for Canadian network sovereignty. In *TEM 2013: The Technology & Emerging Media Track-Annual Conference of the Canadian Communication Association (Victoria)*, 2012.

[42] Once a Defender of Internet Freedom, Putin Is Now Bringing China's Great Firewall to Russia. http://www.huffingtonpost.com/andrei-soldatov/putin-china-internet-firewall-russia_b_9821190.html.

[43] V. N. Padmanabhan and L. Subramanian. An investigation of geographic mapping techniques for Internet hosts. In *ACM SIGCOMM Computer Communication Review*, volume 31, pages 173–185. ACM, 2001.

[44] PlanetLab. http://planet-lab.org/.

[45] Promoting the use of Internet Exchange Points (IXPs): A Guide to Policy, Management and Technical Issues. https://www.internetsociety.org/sites/default/files/Promoting%20the%20use%20of%20IXPs.pdf.

[46] RIPE Atlas. https://atlas.ripe.net/.

[47] H. Roberts, D. Larochelle, R. Faris, and J. Palfrey. Mapping local internet control. In *Computer Communications Workshop (Hyannis, CA, 2011), IEEE*, 2011.

[48] Russia Needs More Internet Security Says Putin. http://www.wsj.com/articles/russia-needs-more-internet-security-says-putin-1412179448.

[49] Russia's Surveillance State. http://www.worldpolicy.org/journal/fall2013/Russia-surveillance.

[50] A. Shah and C. Papadopoulos. Characterizing International BGP Detours. Technical Report CS-15-104, Colorado State University, 2015.

[51] TeleGeography Submarine Cable Map. http://www.submarinecablemap.com/.

[52] The East African Marine System. http://www.teams.co.ke/.

[53] L. Tsui. The panopticon as the antithesis of a space of freedom control and regulation of the internet in china. *China information*, 17(2):65–82, 2003.

[54] M. Wählisch, S. Meiling, and T. C. Schmidt. A framework for nation-centric classification and observation of the internet. In *The ACM CoNEXT Student Workshop*, page 15. ACM, 2010.

[55] M. Wählisch, T. C. Schmidt, M. de Brün, and T. Häberlen. Exposing a nation-centric view on the German internet–a change in perspective on AS-level. In *Passive and Active Measurement*, pages 200–210. Springer, 2012.

[56] S. S. Wang and J. Hong. Discourse behind the forbidden realm: Internet surveillance and its implications on china's blogosphere. *Telematics and Informatics*, 27(1):67–78, 2010.

[57] What ISPs Can See: Clarifying the Technical Landscape of the Broadband Privacy Debate. https://www.teamupturn.com/reports/2016/what-isps-can-see.

[58] S. Zhou, G.-Q. Zhang, and G.-Q. Zhang. Chinese Internet AS-level topology. *Communications, IET*, 1(2):209–214, 2007.