RAN: Routing Around Nation-States

Paper #332 – 12 Pages + References

Abstract

Many countries now engage in interference, degradation, blocking, or surveillance of Internet traffic. In response, individuals, organizations, and even entire countries are taking steps to control the geographic regions that their traffic traverses. For example, some countries are building local Internet Exchange Points (IXPs) to prevent domestic traffic from detouring through other countries. Unfortunately, our measurements reveal that many such ongoing efforts are futile, for two reasons: local content is often hosted in foreign countries, and networks within a country often fail to peer with one another. Yet, our work offers hope: we also find that routing traffic through strategically placed relay nodes can reduce transnational routing detours, in the best case, from 85% of studied paths to 38% of studied paths. Based on these findings, we design and implement RAN, a lightweight system that routes a client's web traffic around specified countries with no modifications to client software (and in many cases with little performance overhead). Anyone can use RAN today; we have deployed long-running RAN Web proxy relays around the world, released the source code, and provided instructions for configuring a client to use the system.

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. For example, an increasing number of countries have passed laws that facilitate mass surveillance of their networks [23, 29, 32, 37]. Governments and citizens alike may want to divert their Internet traffic from countries that perform surveillance (notably, the United States [18, 19, 45]).

Many countries are taking impressive measures to reduce the likelihood that Internet traffic transits the United States [9–12, 27]. For example, Brazil is 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, 13]. Brazil is also building Internet Exchange Points (IXPs) [14], now has the largest national ecosystem of public IXPs in the world [16], and the number of internationally connected Autonomous Systems (ASes) continues to grow [15]. And, 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 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 [53]; migrating a website to HTTPS is challenging, and 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 [30]. DNS traffic is also revealing and is essentially never encrypted [53]. Third, ISPs often terminate TLS connections, conducting man-in-the-middle attacks on encrypted traffic for network management purposes [24]. And, of course, encryption offers no solution to interference, degradation, or blocking of traffic that a country might perform on traffic that crosses its borders. Circumventing surveillance and interference 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 governments and citizens better control transnational Internet paths? In contrast to previous work [31], which simulates Internet paths, we *actively measure* and analyze the paths originating in five different countries: Brazil, Netherlands, Kenya, India, and the U.S. 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 analyzed Border Gateway Protocol (BGP) routes [31, 46]. Although BGP routing can offer some 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 routes from RIPE Atlas probes [43] in five countries to the Alexa Top 100 domains for each country; we directly measure the paths not only to the websites corresponding to 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 3, 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 4 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.

We first *characterize the current state of transnational Internet routing detours* (Section 4). We 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; in many cases, that country is one 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 certain countries [9–12, 27], their traffic still does so. For example, 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 the same country, yet transits an intermediate country; for example, about 13% of the paths that we explored from Brazil tromboned through the United States.

Next, we *explore the potential effectiveness of building a network of overlay relays to help clients certain countries* (Section 5). We find that this technique can be effective for clients in certain countries, yet the effectiveness depends on the country. 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. 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. 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.

Finally, we *design, implement, and deploy RAN, a system that allows a client to access web content without traversing a specified country* (Section 7). RAN uses a series of overlay network relays to automatically route a client's traffic around a specified country. We design and implement RAN for country avoidance, usability, and scalability. Our evaluation shows that RAN is effective for avoiding many different countries and introduces minimal performance overhead.

2 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) [46]. 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. Obar and Clement analyzed traceroutes that started and ended in Canada, but tromboned through the United States, and argued that this is a violation of Canadian network sovereignty [39]. Karlin et al. developed a framework for country-level routing analysis to study how much influence each country has over interdomain routing [31]. This work measures the centrality of a country using BGP routes and AS-path inference; in contrast, our work uses active measurements and measures avoidability of a given country.

Mapping national Internet topologies. Roberts et al. described a method for mapping national networks of

ASes, identifying ASes that act as points of control [44]. Also, several studies have measured and classified the network within a country, including Germany [50, 51] and China [54], or a country's interconnectivity within a region or with the rest of the world [7, 22, 25].

Circumvention and Routing Systems. There has been research into circumvention systems, particularly for censorship circumvention, which is complementary to our work, but not sufficient for surveillance circumvention. Existing circumvention systems generally rely on encryption, which does not prevent surveillance; prior research has shown that websites can be fingerprinted based on size, content, and location of third party resources, which reveals information about the content a user is accessing [53]. Additionally, ISPs often execute a man-in-the-middle attacks on TLS connections to perform network management functions [24]. Therefore, while encryption can be used for censorship circumvention, additional measures must be taken for surveillance circumvention. Some existing circumvention tools are Tor and VPNGate. Tor is an anonymity system that uses three relays and layered encryption to allow users to communicate anonymously [20]. VPNGate is a public VPN relay system aimed at circumventing national firewalls [38]. Unfortunately, VPNGate does not allow a client to choose any available VPN, which makes surveillance avoidance harder. Another system, Alibi Routing, is a peer-to-peer system that uses round-trip times to prove that that a client's packets did not traverse a forbidden country or region [34]; our work differs by measuring which countries a client's packets would (and does) traverse. Our work then uses active measurements to determine the best path for a client wishing to connect to a server. RON, Resilient Overlay Network, is an overlay network that routes around failures, whereas our overlay network routes around countries [1].

3 State of Surveillance and Interference

We focused on traffic *originating* in five countries:

Brazil. Brazil is actively trying to avoid having their traffic transit the U.S. They have been building IXPs, deploying underwater cables to Europe, and pressuring large U.S. companies to host content within Brazil [8–14, 27]. These efforts to avoid a specific country led us to investigate whether they have been successful.

Netherlands. First, the Netherlands is beginning to emerge as a site where servers are located for cloud services, such as Akamai, Second, the Netherlands is where a large IXP is located (AMS-IX). Third, they are drafting a mass surveillance law [37]. Analyzing the Netherlands will allow us to see what effect AMS-IX and the emergence of cloud service hosting has had on traffic.

Kenya. Previous research on the interconnectivity of Africa [22, 25] led us to explore the characterization of an African country's interconnectivity. We chose Kenya for three reasons: 1) it terminates many submarine cable landings; 2) it has relatively high Internet access and usage; and 3) it has more than one IXP [3, 48].

India. India has one of the highest number of Internet users in Asia, second only to China, which has already been well-studied [49, 52].

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.

When analyzing which countries Internet traffic *traverses*, we gave additional attention to countries that have known laws and practices inolving surveillance of or interference with Internet traffic. These countries include, the "Five Eyes" [21, 33] (the United States, Canada, United Kingdom, New Zealand, and Australia), as well as France, Germany, Poland, Hungary, Russia, Ukraine, Belarus, Kyrgyzstan, and Kazakhstan. Countries such as China, Iran, and Russia, are censoring, blocking, and interfering with traffic that crosses their borders. We have studied surveillance states in more detail, and that information is available in our technical report [2].

4 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.

4.1 Measurement Pipeline

Figures 1 and 2 summarize our measurement process. We analyze traceroute measurements to discover which countries are on the path from a client in a particular country to a popular domain. Because we conduct active measurements, which are limited by our resources, we study just five countries. We report on measurements conducted on January 31, 2016.

4.1.1 Resource Limitations

The iPlane [35] and Center for Applied Internet Data Analysis (CAIDA) [17] projects maintain large repositories of traceroute data, neither of which are suitable for our study. Unfortunately, because iPlane uses PlanetLab [41] nodes, which mostly use the Global Research and Education Network (GREN), iPlane measurements are not be representative of typical Internet users' traffic paths [6]. 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.

Instead, 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 [43]. RIPE Atlas

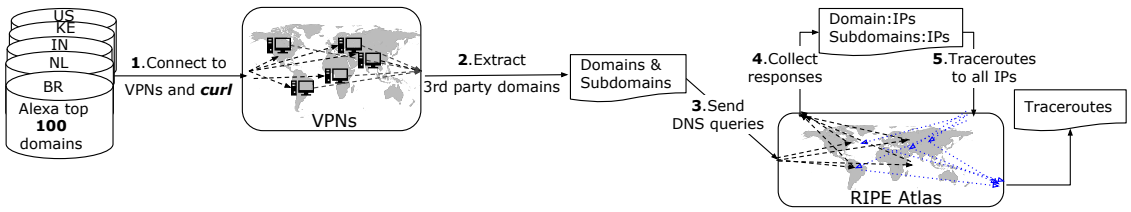


Figure 1: *Measurement pipeline to study Internet paths from countries to popular domains.*

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 addresses 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.

4.1.2 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 [26]. 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

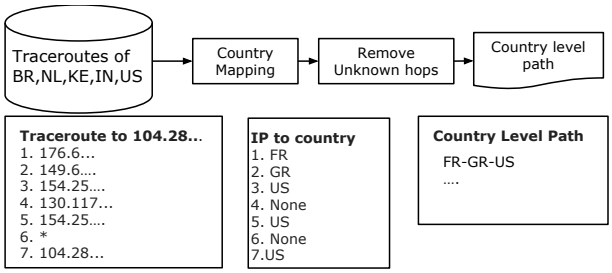


Figure 2: *Mapping country-level paths from traceroutes.*

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.

4.1.3 Traceroute Origin 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 domains, but we must do so from within the country of interest. 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.

4.1.4 Country Mapping

Accurate IP geolocation is challenging. We use MaxMind's geolocation service to map IP addresses to their respective countries [36], which is known to contain inaccuracies. Fortunately, our study does not require high-precision geolocation; we are more interested in providing

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 1: *Fraction of paths terminating in a 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 2: *Fraction of paths that a country transits by default.*

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 [28]. 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.

4.2 Results

Table 1 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 U.S. is the endpoint of 77%

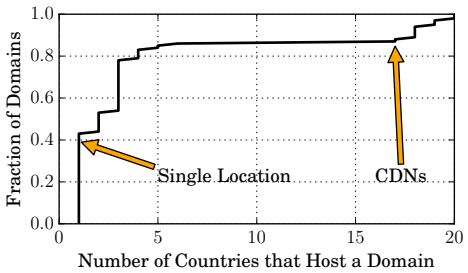


Figure 3: *The number of Alexa Top 100 US Domains hosted in different countries.*

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 2 shows the fraction of paths that transit (or end in) certain countries, with a row for each country that is transited.

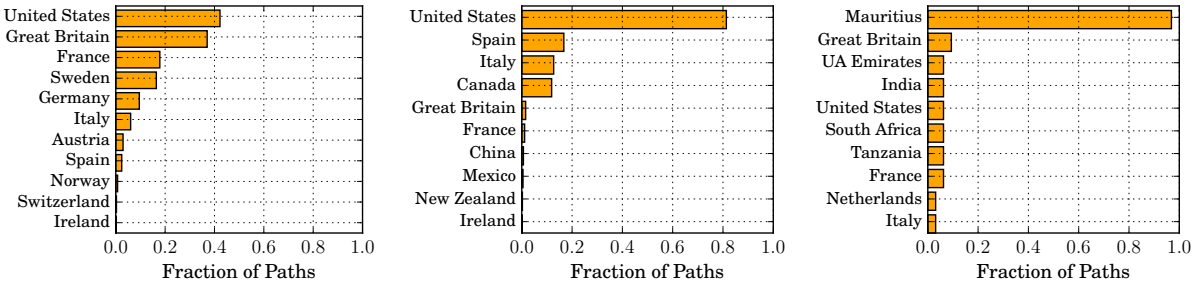
Finding 4.1 (Hosting Diversity)**:** *About half of the top domains in each of the five countries studied are hosted in a single country. The other half are located in two or more different countries.*

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, in geographically diverse locations. Then we mapped the IP addresses in the DNS responses to countries to determine how many unique countries host a domain. Figure 3 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.

Finding 4.2 (Domain Hosting)**:** *The most common destination among all five countries studied is the U.S.: 77%, 45%, 63%, 44%, and 97% of paths originating in Brazil, Netherlands, India, Kenya and the U.S., respectively, are currently reaching content located in the U.S.*

Table 1 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 4.3 (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 coun-*



(a) *The Netherlands.* **(b)** *Brazil.* **(c)** *Kenya.*

Figure 4: *The countries that tromboning paths from the Netherlands, Brazil, and Kenya transit.*

try; 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, but only 5 of the 24 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 4.4 (Transit Traffic)**:** *Surveillance states (specifically the U.S. 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 [47]. 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 4.5 (Tromboning Traffic)**:** *Brazilian and Netherlands paths often trombone to the United States, despite the prevalence of IXPs in both countries.*

Figures 4 shows 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, their paths still trombone to the U.S. 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 [42]. It is likely that Brazilian providers see one another as competitors and therefore as a threat at IXPs, which causes 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 [5].

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 4.6 (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 U.S. The results from studying paths that originate in the U.S. 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 U.S. hosts 97% of the content that is accessed from within the country. Only 13 unique countries are ever on a path from the U.S. 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.

4.3 Limitations

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

Traceroute accuracy and completeness. Our study is limited by the accuracy and completeness of traceroute. Anomalies can occur in traceroute-based measurements [4], 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. 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 [40]. While there are inaccuracies and incompleteness in MaxMind's data [28], the focus of this work is on measuring and avoiding surveillance. We use Maxmind to map IP to country, which provides a lower bound on the amount of surveillance.

IPv4 vs. IPv6 connectivity. We collect and analyze only IPv4 paths. IPv6 paths likely differ from IPv4 paths as not all routers that support IPv4 also support IPv6. Future work includes studying IPv6 paths.

5 Preventing Transnational Detours

In light of our analysis in Section 4, we now explore how much techniques and systems can help clients prevent transnational routing detours. We explore how to (i) increase path diversity with the use of overlay nodes and (ii) discover additional website replicas by diverting DNS queries through global open DNS resolvers. Due to space limitations, our measurement methods and results on the use of open DNS resolvers can be found at [2]. Here, we discuss our measurement method, develop an avoidance metric and algorithm, and present our results for the use of overlay nodes to prevent transnational detours.

5.1 Measurement Approach

Using an overlay network may help clients route around unfavorable countries or access content that is hosted in a different country. Figure 5 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 EC2 instances, one in each geographic region (U.S., Ireland, Germany, Singapore, South Korea, Japan, Australia, Brazil), as well as four Virtual Private Server (VPS) machines (France, Spain, Brazil, Singapore), which are virtual machines. Combining these two sets of machines allow us to evaluate surveillance avoidance with a geographically diverse set of relays.

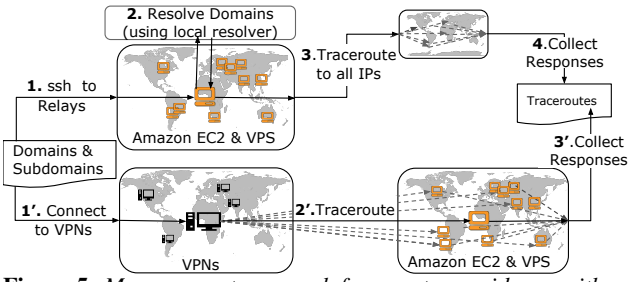


Figure 5: *Measurement approach for country avoidance with overlay network relays.*

5.2 Avoidability Metrics

We introduce a new metric and algorithm to measure how often a client in Country X can avoid a specific country Y.

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 4.1.4, 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. Measuring the avoidability of 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.

Upper bound on avoidability. Although the avoidability metric provides a way to quantify how avoidable Country Y is for a client in Country X, some domains may be hosted only in Country Y, so the avoidance value would

Algorithm 1 Avoidability Algorithm

```
1:  function CalcAvoidance(set paths1, set paths2, string c)
2:      set suitableRelays
3:      for each (relay, path) in paths1 do
4:          if c not in path then
5:              suitableRelays ← path
6:      set accessibleDomains
7:      for each (relay, domain, path) in paths2 do
8:          if relay in usableRelays then
9:              if c not in path then
10:                 accessibleDomains ← domain
11:     D ← number of all unique domains in paths2
12:     A ← length of accessibleDomains
13:     return A/D
```

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. 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.

5.3 Results

We examine the effectiveness of relays for country avoidance, as well as for keeping local traffic local. Table 3 shows avoidance values; the top row shows the countries we studied and the left column shows the country that the client aims to avoid. As seen in Table 3, 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 5.1 (Relay Effectiveness)**:** *For 84% of the (Country X, Country Y) pairs shown in Table 3 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 5.2 (Relays Achieve Upper Bound)**:** *Clients in the U.S. can achieve the upper bound of avoidance for all countries—relays help clients in the U.S. avoid all other Country Y unless the domain is 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 5.3 (Surveillance States are Less Avoidable)**:** *The ability for any country to avoid the U.S. is significantly lower than its ability to avoid any other country in all three situations: without relays, with relays, and the upper bound.*

Despite increasing the ability to avoid the U.S., relays are less effective at avoiding the U.S. compared to all other Country Y. Clients in India can avoid the U.S. more often than clients in Brazil, Netherlands, and Kenya, by avoiding the U.S. for 65% of paths. Even using relays, Kenyan clients can only avoid the U.S. 40% of the time. Additionally, the upper bound for avoiding the U.S. is significantly lower in comparison to other countries.

Finding 5.4 (Keeping Local Traffic Local)**:** *Using relays decreased both the number of tromboning paths, and the number of countries involved in tromboning paths.*

Where there were relays located in one of the five studied countries, we evaluated how well the relays kept local traffic local. This evaluation was possible for the U.S. and Brazil. Tromboning Brazilian paths decreased from 13.2% without relays to 9.7% with relays; when relays are used, all tromboning paths go only to the U.S. With the relays, we see only 1.3% tromboning paths for a U.S. client, compared to 11.2% without relays. The 1.2% of paths that trombones from the U.S. go only to Ireland.

6 Design Goals

Our measurement results motivate the design and implementation of a relay-based avoidance system, RAN, with the following design goals.

Country Avoidance. The primary goal of RAN is to avoid a given country when accessing web content. RAN should provide clients a way to route around a specified country, when accessing a domain. This calls for the role of measurement in the system design, and systematizing the measurement methods discussed earlier in the paper.

Usability. RAN should require as little effort as possible from clients. Clients should not have to download or install software, collect any measurements, or understand how the system works. This calls for a way for clients to automatically and seamlessly multiplex between relays (proxies) based on different destinations. The Proxy Autoconfiguration (PAC) file supports this function.

Scalability. This country avoidance system should be able to scale to large numbers of users. Therefore, RAN should be able to handle the addition of relays, as well as be cost-effective in terms of resources required. This requires clever measurement vantage points, such that each

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

Table 3: *Avoidance values for different country-avoidance techniques. 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 U.S. is significantly lower than for any other country; .886, .790, .844, and .765 are the upper bounds on avoidance of the U.S. for paths originating in Brazil, Netherlands, India, and Kenya, respectively.*

vantage point is representative of more than one client. The PAC file allows the system to grow with the number of clients and also supports incremental deployment.

Non-goals. There are some challenges that RAN does not attempt to solve. The system does not address the notion of anonymity; it routes around countries (for reasons such as avoiding mass surveillance), but it does not attempt to keep users anonymous. RAN, of course, does not address domestic surveillance (for example, a client in the U.S. attempting to avoid surveillance by the U.S.).

7 Design

RAN comprises an overlay network of relays and an oracle that directs clients to the appropriate relays, as shown in Figure 6. RAN's relays are TCP proxy servers, which allow clients to access web content without installing software. RAN uses the measurement methods described earlier to learn paths between clients, relays, and domains; the results are stored at the oracle. The oracle uses the data to decide which relay a client in some location should use for accessing a certain domain while avoiding a certain country. The oracle periodically computes paths for many combinations of client AS, destination, and country.

7.1 Periodic Path Measurement

All of the paths are measured using `traceroute`, which is then mapped to the country level using the same methods as described in Section 4 and shown in Figure 2. The paths we measure are the: forward paths from the client to each relay, forward paths from each relay to each domain, forward paths from the client to each domain, and reverse paths from each relay to the client. RAN measures paths from clients to relays, clients to domains (servers), relays

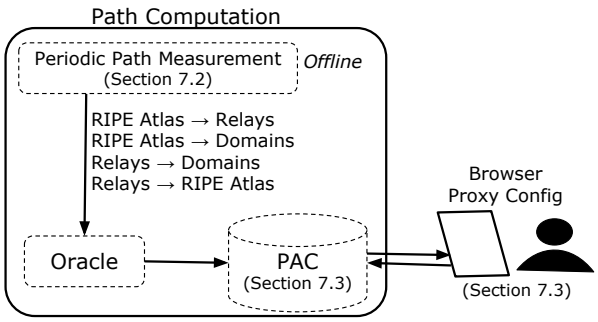


Figure 6: *RAN architecture.*

to domains, and relays to clients; the reverse path from the domains to the relays is challenging to measure due to a lack of vantage points in ASes of common destinations. Despite the inability to measure this part of the path, it would be difficult for the country being avoided to perform traffic analysis because it is *at most* only on the reverse path from the server to the relay.

Client-to-Relay Paths. To avoid requiring the client to install custom software, RAN measures client-to-relay paths from RIPE Atlas probes that serve as vantage points for potential client ASes. RAN selects probes that are geographically close the client (*e.g.*, in the same country). The oracle triggers the probe to run traceroutes to each relay. After collecting the responses, the oracle maps the IP-level paths to country-level paths and stores the results.

Relay-to-Client Paths. The relays perform traceroutes to the IP addresses of RIPE Atlas probes, which represent

Configuration 1: *Example PAC file.*

```
function FindProxyForURL(url, host){
    if ((shExpMatch(host, "*.google.com")))
        return "PROXY 1.2.3.4:3128";
    if ((shExpMatch(host, "*.twitter.com")))
        return "PROXY 5.6.7.8:3128";
    return "DIRECT";
}
```

client ASes. They then derive country-level paths; the oracle learns these paths from each relay.

Relay-to-Server Paths. Relays perform traceroutes to each domain. As with paths to clients, relays derive country-level paths and send them to the oracle.

Client to Server Paths. In a path from a client to a domain does not pass through the country specified to avoid *by default*, then none of the proxies should be used. These paths are measured using the RIPE Atlas probes in similar locations as the clients, and the oracle triggers traceroutes from each of them to each of the domains. Corresponding country-level paths are stored at the oracle.

These paths must be re-computed as paths may change. To our knowledge, there has not been any previous work on how often country-level paths change; prior work has explored how often AS-level paths change. To measure how often country-level paths change, we computed the paths from relays to domains once every two hours and once every hour. Fewer than five paths changed every two hours; the results were similar for one-hour increments. As it takes approximately 30 minutes to compute all paths, RAN re-computes the paths every one hour to incorporate the most recent country-level paths.

7.2 PAC File Generation

The oracle follows four steps to decide which relay a client should use to access a specific domain: (1) If the default path from the client to the domain does not pass through the specified country, then do not use any of the relays. (2) Otherwise, for all the paths from the client to the relays, select suitable relays, which are relays where the country to avoid is not on the forward or reverse path between the client and relay. (3) From this set, if there is a path from a suitable relay to the domain that does not include the specified country, then use that relay for that domain. (4) If there is no path from the client through any of the relays to the domain that does not pass through the specified country, then select the relay that provides the most avoidance (measured by how many other domains that avoid the specified country). The oracle applies this decision process to each domain, which results in a mapping of domains to relays that can be used to avoid the given country. To facilitate automatic multiplexing between relays, RAN utilizes Proxy Auto-configuration (PAC) files, which define how browsers should choose a proxy when fetching a URL. In the example PAC file in Configuration 1, proxy 1.2.3.4:3128 should be used when accessing `www.google.com`, but proxy 5.6.7.8:3128 should be used when accessing `www.twitter.com`. The oracle uses the mapping of domains to relays to generate a PAC file, which specifies which domains should be accessed through which proxy. The PAC file is published online to a URL of the format <client_country>_<country_to_avoid>_pac.pac. The client uses this URL to specify their proxy configuration. Paths are re-computed every hour, so the contents of the PAC file are also updated every hour.

7.3 Scalability and Fault Tolerance

Adding relays to RAN is straightforward. Additionally, RAN is resilient to failures of system components.

Adding relays and oracles. To add a relay, the system operator must set up a machine as a proxy server, install the relay software, and update the oracle's list of relays. From that point onward, paths will be computed to and from the new relay, and clients will begin using the new proxy. Adding an oracle requires installing the oracle software on a different machine, and specifying the client locations handled by that oracle (*e.g.*, one oracle handles clients in North America and Europe, and another handles clients elsewhere). Both oracles will publish the PAC files to the same server, which causes no changes for the client.

Failed relays and oracles. Unresponsive relays are handled by the PAC file. The PAC file allows the oracle to specify multiple proxies in a sequential order, such that if the the first proxy fails, then the client users the second proxy (and so on). This feature can be used to specify all of the relays that have a path to the domain. Among other mechanisms, we can detect a failed oracle by determining that its PAC file is older than one hour. Detecting a failed oracle could trigger a backup oracle to re-compute the PAC files periodically. Because oracles are stateless, failover is straightforward. Without backup oracles, clients can still use the system when the oracle fails. The clients will simply be using stale paths, which are likely (but not guaranteed) to be functional, since country-level paths change infrequently.

8 Implementation

Our implementation of RAN includes relays, an oracle, and a client. RAN is open source.[1]

Relays. We established nine relays, one in each of the following countries: Brazil, Germany, Singapore, Japan, Australia, France, United States, United Kingdom, and Canada; these are shown, along with their corresponding AS number in Figure 7. They are running as Ubuntu Virtual Private Servers (VPSs) with Squid as the proxy server. They are also running the RAN Relay software.

[1] Please contact the PC chairs for the source code.

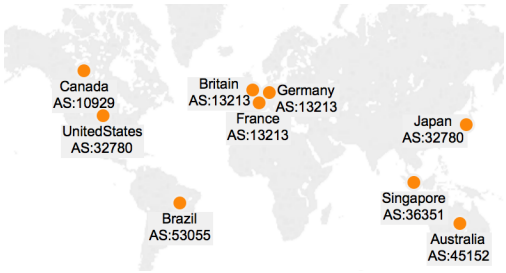


Figure 7: *The locations and ASNs for RAN relays.*

Oracle. The oracle software runs on a Fujitsu RX200 S8 server with dual, eight-core 2.8GHz Intel Xeon E5 2680 v2 processors with 256GB RAM running the Springdale distribution of Linux.

Client. To evaluate RAN, we set up a client machine in the Netherlands, which simply accesses web content and uses the PAC file generated by the oracle.

9 Evaluation

We evailate RAN's ability to avoid a given country, its performance, and its storage and measurement costs.

9.1 Country Avoidance

We measured RAN's effectiveness in achieving country avoidance. We did so by first calculating the number of *default* paths that avoid a given country. Then we added a single relay, and calculated how many domains the client could access without traversing through the given country. We repeated this approach for the remaining relays. We conducted the evaluation under the condition that the client wished to avoid different countries when accessing the Netherlands top 100 domains; Figure 8 shows these results. Each line represents the fraction of domains accessible while avoiding the country that the line represents. For example, 46% of domains are accessible without traversing the U.S. when RAN is not being used (zero relays), and if RAN is used, then 63% of domains are accessible without traversing the U.S.

It is evident that RAN helps a client avoid a foreign country, as the fraction of domains accessible without traversing the specified country without RAN is lower than with RAN. Additionally, adding the first relay provides the greatest benefit, while subsequent relays offer diminishing returns. Figure 8 clearly shows that avoiding the U.S. is much more difficult (or impossible) than any other country. Only 63% of domains can be accessed while avoiding the U.S., whereas almost all domains can be accessed while avoiding any other given country.

9.2 Performance

To measure the performance of RAN, we measure both the throughput and latency.

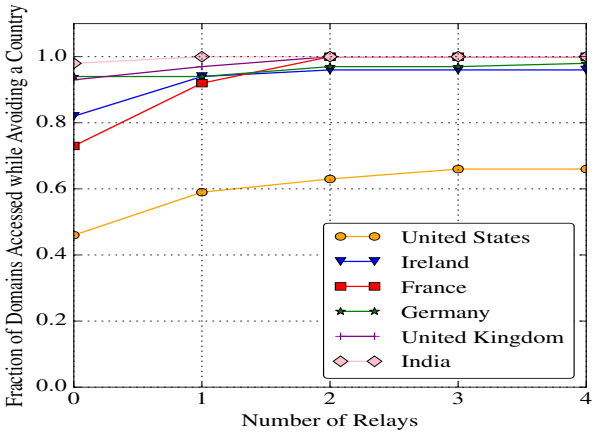


Figure 8: *The effect of the number of relays on avoidance, for a client in the Netherlands. We tested RAN with up to nine relays.*

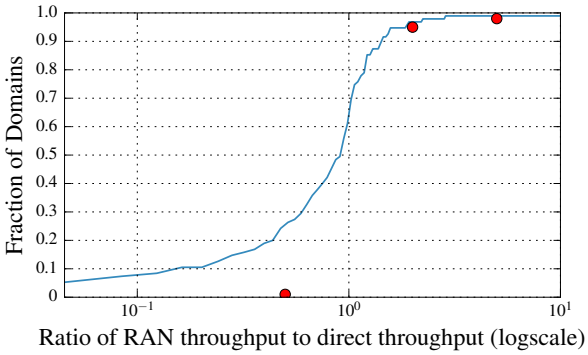


Figure 9: *The ratio of RAN throughput to direct throughput. The points on the graph are taken from the RON study and represent a "normal" overlay network's performance.*

To measure throughput, we ran `wget` for each of the top 100 domains from the client machine in the Netherlands using an oracle-generated PAC file. Because different relays could have been used to avoid a single domain, the oracle selected a random relay from those that would allow the client to avoid the country. The oracle generated ten PAC files for a client in the Netherlands who wishes to avoid the United States, randomly selecting a relay for domains that could have used different relays, and `wget` was used for the top 100 domains for each PAC file generated. Based on the `wget` output, we calculate the number of seconds to access content using our system and take the average across the ten experiments.

Figure 9 shows a CDF of the ratio of RAN throughput to direct throughput. The throughput of RAN is not significantly worse than that of default paths. In some cases the performance of RAN is *better* than that of default paths. Such improvements could be a result of the relays keeping local traffic local, or due to a closer content replica being selected. These results show that RAN's performance is comparable to the performance of accessing domains

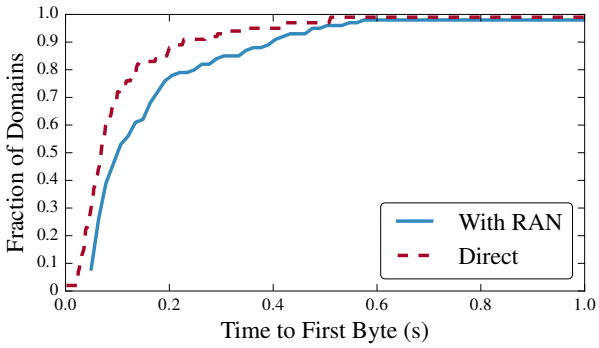


Figure 10: *Time to First Byte for RAN and direct paths.*

without RAN. Figure 9 also compares RAN's throughput to RON's throughput, illustrated with the red dots. RAN performs worse than RON ($x < 1$), which is expected, as the detours that RAN introduces inherently inflates paths. Interestingly, both RON and RAN improve throughput for a similar fraction of samples ($x > 1$).

To measure the latency of RAN, we ran `curl` to each of the top 100 domains from the client in the Netherlands, while using the ten oracle-generated PAC files. This provided the time to first byte (TTFB); we found the average TTFB when accessing content using RAN and found the TTFB when using direct paths; the results are shown in Figure 10. The median TTFB for direct paths is 68.5 ms; for RAN paths the median is 100.8 ms; 90th percentile TTFB is 22.5 ms and 40.4 ms, respectively.

9.3 Storage and Measurement Costs

As the number of clients increase, and subsequently the number of paths being computed increases, the amount of storage must remain reasonable. The storage used by paths can be calculated as $DR + 2CR + CD$ where D is the number of domains; R is the number of relays; and C is the number of ASes from which RAN measures. The storage required for a single client, 100 domains, and nine relays is 480 KB. Because there is a single PAC file for all clients in a country, C will grow much slower than if there was a different PAC file for each individual client. There are 196 countries; if RAN computed paths and a PAC file for each country, with 100 domains, and three relays required storage would be only 94 MB, making it feasible to increase the number of relays and domains.

RIPE Atlas credits are also a limited resource. Cost is proportional to $C \cdot (R + D)$. Each traceroute costs 60 RIPE Atlas credits, so one set of measurements for one client, 100 domains, and nine relays costs 6,180 credits; because these paths are updated each hour, then the daily credit cost is 148,320 credits. In return for hosting a RIPE Atlas probe, we earn 216,000 credits per day, which will support our existing prototype. In order to provide for more clients, more domains, or more resources, we can tune the system to re-compute paths less frequently, as we discuss in Section 10.

10 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.

Evolution 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 should explore how paths evolve over time, and analyze the effectiveness of different country-avoidance strategies.

Additional features. Additional features can be implemented at the relay to help preserve client privacy. An example would be to use the relay as a mix, or to send out fake traffic to confuse an attacker that may be trying to perform traffic analysis at the relay. The oracle could add additional steps in the decision chain introduced in Section 7.2 that take into account relay and path loads. For example, if multiple relays provide a path to a domain that does not traverse the specified country, then the decision between the suitable proxies could be determined based on current relay load or performance. Our current implementation of RAN re-computes all paths once per hour; we could only re-compute paths when necessary. For example, a BGP monitoring system could be implemented that alerts the oracle to a routing changes that affects any path currently in the system.

11 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 step towards a remedy, we have investigated how clients can use overlay network relays to prevent routing detours through unfavorable jurisdictions. This method gives clients the power to avoid certain countries, as well as help keep local traffic local.

We make country avoidance accessible to Internet users by designing and implementing RAN, which employs overlay network relays to route Internet traffic around a given country. Our evaluation shows that RAN is successful at avoiding countries while performing as well, if not better, than taking default routes.

References

[1] D. Andersen, H. Balakrishnan, F. Kaashoek, and R. Morris. Resilient overlay networks. In *ACM Symposium on Operating Systems Principles (SOSP)*, volume 35. ACM, 2001.

[2] Anonymized ArXiv Tech Report, May 2016. Ask program chairs for anonymized version.

[3] 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.

[4] 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.

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

[6] 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.

[7] 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.

[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'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.

[13] Brazil to press for local Internet data storage after NSA spying. https://www.rt.com/news/brazil-brics-internet-nsa-895/, 2013.

[14] Brasil Internet Exchange Participants Diversity. http://ix.br/doc/nic.br.ix.br.euro-ix-27th-berlin.20151027-02.pdf, 2015.

[15] Brazil Winning Internet. http://research.dyn.com/2014/07/brazil-winning-internet/#!prettyPhoto/1/, 2015.

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

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

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

[19] 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.

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

[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 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/.

[24] 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.

[25] 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.

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

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

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

[29] 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.

[30] A. Johnson, C. Wacek, R. Jansen, M. Sherr, and P. Syverson. Users Get Routed: Traffic Correlation on Tor by Realistic Adversaries. In *CCS*. ACM, 2013. http://www.ohmygodel.com/publications/usersrouted-ccs13.pdf.

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

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

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

[34] D. Levin, Y. Lee, L. Valenta, Z. Li, V. Lai, C. Lumezanu, N. Spring, and B. Bhattacharjee. Alibi Routing. In *The 2015 ACM Conference on Special Interest Group on Data Communication*, pages 611–624. ACM, 2015.

[35] 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.

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

[37] Netherlands New Proposal for Dragnet Surveillance Underway. https://edri.org/netherlands-new-proposals-for-dragnet-surveillance-underway/, 2015.

[38] 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.

[39] 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.

[40] 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.

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

[42] 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, 2012.

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

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

[45] Russia Needs More Internet Security Says Putin. http://www.wsj.com/articles/russia-needs-more-internet-security-says-putin-1412179448, 2014.

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

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

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

[49] 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.

[50] 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.

[51] 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.

[52] 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.

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

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