A First Look into Transnational Routing Detours

Anne Edmundson
Princeton University

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).

1 Problem and Motivation

When Internet traffic enters a country, it becomes subject to those countries' laws. As a result, users have more need than ever to determine—and control—which countries their traffic is traversing. As an increasing number of countries pass laws that facilitate mass surveillance of their citizens [15], governments and citizens are increasingly motivated to divert their Internet traffic from countries that perform surveillance (notably, the United States [4]).

Many countries—notably, Brazil—are taking impressive measures to reduce the likelihood that Internet traffic transits the United States [4] 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 switching its dominant email system (Microsoft Outlook) to a state-developed system called Expresso [4]. Brazil is also building Internet Exchange Points (IXPs), now has the largest national ecosystem of public IXPs in the world, and the number of internationally connected ASes continues to grow. 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).

Defending against these activities requires not only encryption, but also mechanisms for controlling where traffic goes in the first place: end-to-end encryption conceals some information content, but it does not protect all sensitive information. 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 [21]; migrating a website to HTTPS can be 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 [11]. DNS traffic is also revealing and is almost never encrypted [21]. Third, ISPs often terminate TLS connections, conducting man-in-the-middle attacks on encrypted traffic for network management purposes [8]. 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. Finally, a nation-state may collect and store encrypted traffic; if the encryption is broken in the future, a nation-state may be able to discover the contents of previous communications.

Our work tackles two questions: (1) Which countries do *default* Internet routing paths traverse?; (2) What types of methods can we use to take advantage of hosting and path diversity to help governments and citizens better control transnational Internet paths?

Previous work has analyzed Border Gateway Protocol (BGP) routes [12, 20], but these provide an indirect estimate of country-level paths at best. Although BGP routing can offer some information about paths, it does not necessarily reflect the path that traffic actually takes, and it only provides Autonomous System (AS) -level granularity, which is often too coarse to make strong statements about which countries that traffic is traversing. We take a different approach and actively measure *direct* paths to popular websites to characterize the current state of transnational routing detours.[1]

Next, we explore the extent to which a network of overlay relays could help clients avoid certain countries to popular destinations. Due to promising measurement results, we design,

[1] This work was partially published in [6].

implement, and deploy *RAN*, a system that allows a client to access web content while avoiding the traversal of a specified country.[2] We evaluate *RAN* to assess its ability to avoid certain countries, as well as the effect on end-to-end performance.

2 Background and Related Work

Nation-state routing analysis. Shah and Papadopoulos recently measured international routing detours—paths that originate in one country, cross international borders, and then return to the original country—using public Border Gateway Protocol (BGP) routing tables [20]. The study discovered 2 million detours each month out of 7 billion paths. Our work differs by *actively* measuring traceroutes, yielding a more precise measurement of the 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 [17]. Karlin *et al.* developed a framework for country-level routing analysis to study how much influence each country has over interdomain routing [12]. This work measures country centrality using BGP routes and AS-path inference; in contrast, our work uses active measurements and measures avoidability of a given country.

Several studies have also characterized network paths *within* a country [3, 7, 9]; these studies focus on intra-country paths, as opposed to focusing on transnational paths.

Routing overlays and Internet architectures. Alibi Routing uses round-trip times to prove that that a client's packets did not traverse a forbidden country or region [13]; our work differs by measuring which countries a client's packets would (and do) 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 [2], whereas our overlay network routes around countries. ARROW introduces a model that allows users to route around ISPs [18], but requires ISP participation, making it considerably more difficult to deploy than *RAN*. ARROW also aims to improve fault-tolerance, robustness, and security, rather than explicitly attempting to avoid certain countries; ARROW provides mechanisms to avoid individual ISPs, but such a mechanism is at a different level of granularity, because an ISP may span multiple countries. Zhang *et al.* presented SCION, a "clean-slate" Internet architecture that provides route control, failure isolation, and explicit trust information for communication [22]; SCION, however, requires fundamental changes to the Internet architecture, whereas *RAN* is deployable today.

Circumvention systems. Certain tools, such as anonymous communications systems or virtual private networks, may use a combination of encryption and overlay routing to allow clients to avoid surveillance. Tor is an anonymity system that uses three relays and layered encryption to allow users to communicate anonymously [5]. In contrast, *RAN* does not aim to achieve anonymity; instead, its aim is to ensure that traffic does not traverse a specific country, a goal that Tor cannot achieve. Even tools like Tor do not inherently thwart surveillance: Tor is vulnerable to traffic correlation attacks and some attacks are possible even on encrypted user traffic. VPNGate is a public VPN relay system aimed at circumventing national firewalls [16]. Unfortunately, VPNGate does not allow a client to choose any available VPN, which makes it more difficult for a user to ensure that traffic avoids a particular part of the Internet. Neither of these systems explicitly avoid countries; thus, they may not be able to avoid surveillance or the laws or jurisdiction of a particular country.

3 Approach and Uniqueness

Here we discuss our measurement approach to characterize transnational routing detours (Section 3.1), as well as our approach to measuring how avoidable different countries are relative to where a client is located (Section 3.2). Lastly, we introduce *RAN*, which employs an overlay network to provide end-users some control over which countries their Internet paths traverse (Section 3.3). All results for these measurements are presented in Section 4.

Our analysis focuses on five countries of interest—Brazil, Netherlands, Kenya, India, and the United States—for a variety of reasons. For example, Brazil has made a concerted effort to avoid traversing certain countries such as the United States through extensive buildout of Internet Exchange Points (IXPs). The Netherlands has one of the world's largest IXPs and relatively inexpensive hosting. Kenya is one of the most well-connected African countries, but it is still thought to rely on connectivity through Europe and North America for many destinations, even content that might otherwise be local (*e.g.*, local newspapers).

3.1 Characterizing Default Internet Paths

To measure which countries *default* Internet paths traverse, we use RIPE Atlas [19] probes in the country of interest to locally resolve each domain in the Alexa Top 100 [1] and run traceroutes to the IP addresses resolved from the DNS queries. This measurement method is shown in Figure 1. The measurements were run using Paris traceroute and each (probe, destination IP) pair was used twice: once using ICMP traceroute and once using TCP traceroute. Using MaxMind [14], each IP address was geolocated at a country granularity, and with the resulting set of country-level paths, we analyzed which countries host and/or transit the traffic.

Accurate IP geolocation is challenging, but our study does not require high-precision geolocation; we are interested in providing accurate lower bounds on detours at coarse granularity, and previous work has found that geolocation at a country-level granularity is more accurate than at finer granularity [10].

3.2 Measuring Country Avoidability

One method a client can use to control their transnational Internet paths is to use an overlay network; this helps discover georeplication of a site or service that can facilitate the avoid-

[2] This work is under submission.

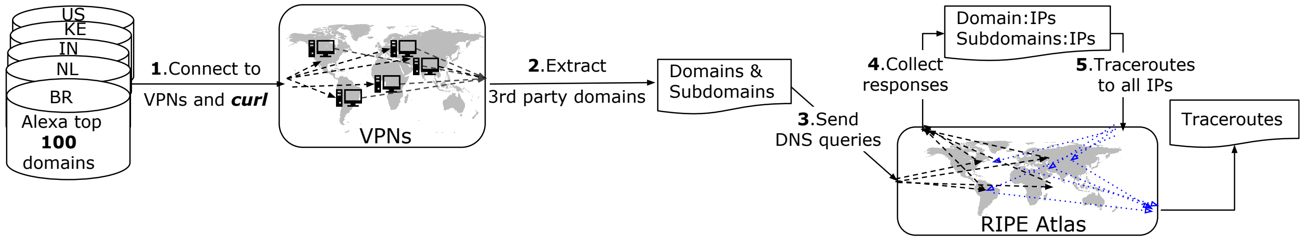

Figure 1: *Measurement pipeline to study Internet paths from countries to popular domains.*

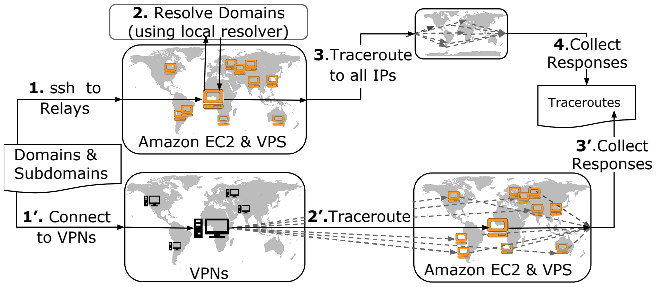

Figure 2: *Measurement approach for country avoidance with overlay network relays.*

ance of specific countries. This approach can also prevent the client from traversing an unfavorable country by introducing a path that detours from the default.

To evaluate the feasibility of this approach, we establish 12 relays in geographically diverse locations around the world, run traceroutes from a client's country (the countries of interest) to each relay, as well as from each relay to the Alexa Top 100 domains. The measurement pipeline for this is shown in Figure 2.

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 originate in Country X and do not transit Country Y; this 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.

After mapping the traceroutes to country-level paths, we use the avoidability metric to measure how avoidable each country is; the results are discussed in Section 4.2 and shown in Table 1.

3.3 *RAN*: Routing Around Nation-States

As our results showed that we can benefit from a system of overlay network relays, we designed and implemented this system, called *RAN*. *RAN* comprises (1) an overlay network of relays; and (2) an oracle that directs clients to the appropriate relays. *RAN*'s relays are TCP proxy servers that allow clients to access web content without installing custom software. *RAN* uses the measurement methods described in Sections 3.1 and 3.2 to learn paths between clients, relays, and domains; these results are stored at the oracle, which 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 to avoid. A client can then query the oracle to determine the appropriate relay to use to avoid a certain country en route to a particular destination.

4 Results

Here we highlight some of our findings; Table 1 shows avoidance values. The top row shows the countries we studied and the left column shows the country that the client aims to avoid.

4.1 Characterizing Default Internet Paths

We discuss three main findings on the current state of transnational routing detours.

Finding 4.1 (Domain Hosting)**:** *The most common destination, regardless of originating country, 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.*

Despite the extent of country-level hosting diversity, the majority of paths from all of the countries we studied 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 4.2 (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, but only 5 of the 24 are hosted within Kenya. 29 out of 40 .nl domains are hosted in the Netherlands; four 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 4.3 (Transit Traffic)**:** *The United States and Great Britain are on the largest portion of paths in comparison to any other (foreign) country.*

85% 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.

4.2 Country Avoidability

Based on our measurement method in Section 3.2, we present some findings on country avoidability.

Finding 4.4 (Relay Effectiveness)**:** *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 4.5 (U.S. is Least 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.

4.3 *RAN* Evaluation

Due to the promising results we saw in the previous subsection, we built *RAN*, which uses the proposed country avoidance technique of overlay network relays. Here we discuss how well *RAN* circumvents unfavorable countries and its performance impact.

We conducted the evaluation under the condition that the client is located in the Netherlands and she wishes to avoid different countries when accessing the Netherlands top 100 domains.

To measure *RAN*'s effectiveness in achieving country avoidance, we first calculated 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 the approach for the remaining relays, and the results are shown in Figure 3. We can see 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 3 clearly shows that avoiding the U.S. is much more difficult (or impossible) than any other country.

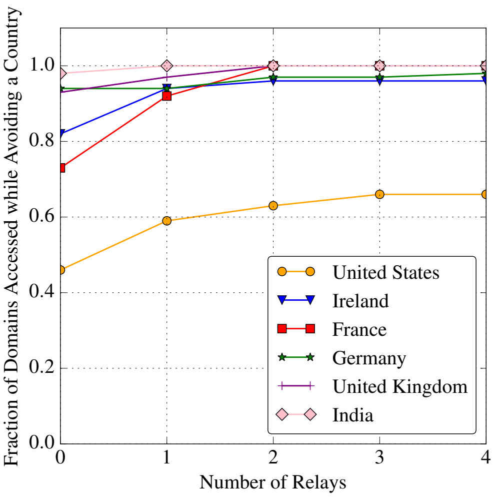

Figure 3: *The effect of the number of relays on avoidance for a client in the Netherlands. We tested RAN with up to nine relays.*

To measure the latency of *RAN*, we ran `curl` to each of the top 100 domains ten times from the client in the Netherlands. This experiment allowed us to measure the time to first byte (TTFB) for web downloads; we found the average TTFB when accessing content using *RAN* and found the TTFB when using direct paths; Figure 4 shows these results. 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.

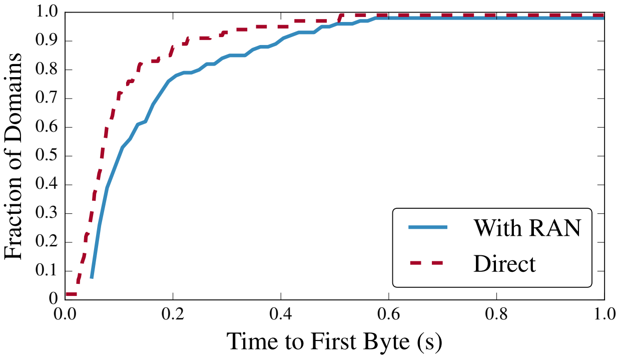

Figure 4: *Time to First Byte for* RAN *and direct paths.*

5 Contributions

We have characterized routing detours that take Internet paths through foreign countries, which may make clients susceptible to foreign surveillance, performance degradation, and increased costs. We find that paths commonly traverse known surveillance states, even when they originate and end in a non-surveillance state. As a first step towards a remedy, we have investigated how clients, ISPs, and governments 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.

Country to Avoid	*Brazil*		*Netherlands*		*India*		*Kenya*		*United States*	
	No Relay	Relays	No Relay	Relays	No Relay	Relays	No Relay	Relays	No Relay	Relays
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 1: *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.*

We have designed, implemented, and deployed *RAN*, which employs overlay network relays to route traffic around a given country. Our evaluation shows that *RAN* can in many cases avoid certain countries while performing nearly as well, if not better, than taking default routes.

References

[1] Alexa Top Domains. http://www.alexa.com/topsites.

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

[3] 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.

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

[5] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The Second-Generation Onion Router. Technical report, DTIC Document, 2004.

[6] A. Edmundson, R. Ensafi, N. Feamster, and J. Rexford. A first look into transnational routing detours. In *Proceedings of the 2016 conference on ACM SIGCOMM 2016 Conference*, pages 567–568. ACM, 2016.

[7] 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.

[8] 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.

[9] A. Gupta, M. Calder, N. Feamster, M. Chetty, E. Calandro, and E. Katz-Bassett. Peering at the Internets Frontier: A First Look at ISP Interconnectivity in Africa. In *Passive and Active Measurement*, pages 204–213. Springer, 2014.

[10] B. Huffaker, M. Fomenkov, and K. Claffy. Geocompare: A Comparison of Public and Commercial Geolocation Databases. *Proc. NMMC*, pages 1–12, 2011.

[11] 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.

[12] J. Karlin, S. Forrest, and J. Rexford. Nation-state Routing: Censorship, Wiretapping, and BGP. *arXiv preprint arXiv:0903.3218*, 2009.

[13] 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.

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

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

[16] 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.

[17] 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.

[18] S. Peter, U. Javed, Q. Zhang, D. Woos, T. Anderson, and A. Krishnamurthy. One tunnel is (often) enough. *ACM SIGCOMM Computer Communication Review*, 44(4):99–110, 2015.

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

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

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

[22] X. Zhang, H.-C. Hsiao, G. Hasker, H. Chan, A. Perrig, and D. G. Andersen. Scion: Scalability, control, and isolation on next-generation networks. In *2011 IEEE Symposium on Security and Privacy*, pages 212–227. IEEE, 2011.