Borders and Gateways: Measuring and Analyzing National AS Chokepoints

Kirtus G. Leyba
Arizona State University
kleyba@asu.edu

Benjamin Edwards
Cyentia Institute
BJEdwards@gmail.com

Cynthia Freeman
University of New Mexico
cynthiaw2004@gmail.com

Jedidiah R. Crandall
University of New Mexico
crandall@cs.unm.edu

Stephanie Forrest
Arizona State University
steph@asu.edu

ABSTRACT

Internet topology reflects economic and political constraints that change over time. Although autonomous systems (AS) topology has been measured and modeled for many years, focusing primarily on economic relationships, earlier studies have not quantified how topology is changing with respect to nation-state boundaries. National boundaries are natural points of control for surveillance, censorship, tariffs and data localization. This paper introduces a measure, *national chokepoint potential (NCP)*, to characterize how a country's AS topology is organized in terms of BGP paths that can carry traffic across international borders. To study country-level chokepoints, we developed BGP-SAS, an open source, cross platform, efficient set of tools for simulating BGP routing and calculating national chokepoint measures. We use these tools to assess how AS topologies have changed over a ten-year span, finding significant variability among countries, with some increasing their chokepoint potential and others remaining constant, fluctuating, and in some cases declining. Overall, however, most national Internet boundaries have either become more pronounced or remained constant, despite new infrastructure buildouts and increased Internet usage. When compared to independent measures of Internet freedom, we find statistically significant relationships between NCP and Internet freedom.

ACM Reference Format:
Kirtus G. Leyba, Benjamin Edwards, Cynthia Freeman, Jedidiah R. Crandall, and Stephanie Forrest. 2019. Borders and Gateways: Measuring and Analyzing National AS Chokepoints. In *ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS) (COMPASS '19), July 3–5, 2019, Accra, Ghana.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3314344.3332502

1 INTRODUCTION

Over time, the Internet has evolved from a borderless collection of networks that spanned geo-political boundaries and promoted the free flow of information into a network that reflects political and economic constraints, providing more opportunity for control. As routine tasks, communications, entertainment, and information move online and are mediated by the Internet, most of us have little choice about whether or not to rely on the Internet. According to the International Telecommunication Union (ITU), the number of individual Internet users increased from 1.024 billion in 2005 to 3.578 billion in 2017 [3]. The majority of these users operate in an environment that restricts Internet freedom in some way. For example, the 2017 Freedom on the Net report from Freedom House reports that 64% of Internet users belong to a nation with Internet that is not free or partly free. Beyond censorship and surveillance, the EU is considering content rules surrounding copyright, net neutrality rules in the U.S. were recently overturned, and large content services are under enormous pressure to control fake news and bad actors. Other forms of Internet intervention such as code injection [32, 33] have been observed within the Internet's backbone. Taken together, these trends will likely restructure the Internet in unforeseen ways, as organizations respond to new challenges and realities, particularly those imposed by legal and regulatory regimes.

Today, most such control is exercised at the country level, as governments have recognized both the threat and the opportunity that is posed by ubiquitous online communication. This is natural, because governments properly mediate many aspects of human society that have moved on-line. However, the Internet provides these opportunities in new ways and at unprecedented scale while the core Internet architecture and protocols were not designed with such considerations in mind. Ease of control, whether to censor or spy on one's own citizens, protect citizens' data from surveillance as it passes through other countries, collect tariffs on on-line transactions or even to provide resilience in the case of a global catastrophe, depends to a large extent on the number and diversity of connections to the external Internet. Previous work has investigated country-level AS topology [26], focusing on how the global AS graph layout, together with interdomain routing protocols, implies that Internet communications pass through the infrastructure of other nations. In addition, the transnational Internet paths of nations have been shown to detour into other nations, indicating that some nations can interfere with a disproportionate number of paths [16].

Here we ask a related but different question. Do the Internet borders of nations contain many ports of access, or are there only a few powerful chokepoints? By studying the number of paths intercepted by border ASes, we can compare nations according to

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
COMPASS '19, July 3–5, 2019, Accra, Ghana
© 2019 Association for Computing Machinery.
ACM ISBN 978-1-4503-6714-1/19/07…$15.00
https://doi.org/10.1145/3314344.3332502

the diversity of AS-level paths crossing each nation's borders. This approach allows us to measure the strength of political boundaries in the Internet and how they have changed over time, for any country.

ASes are groups of routers under common management, such as the network of a university or an ISP. The total size of the AS graph (the global network of all ASes) has grown from about 10,000 ASes in the early 2000s to over 60,000 today. We are not only interested in the current state of the AS graph; the historical dynamics of the AS graph are of interest as well. The structural properties of national subgraphs have evolved differently from nation to nation, whether from economic decision making, infrastructural necessities, or efforts to build a powerful censorship and surveillance network [11, 12]. This rapid expansion, together with an increasing interest by national governments, points to the importance of understanding properties like path robustness, AS hierarchies, and the potential for organizations to control information as it flows in and out of their networks.

We focus on border ASes, i.e. ASes that can connect directly to at least one AS from another nation, and introduce a measure called *chokepoint potential (CP)*, which quantifies the percentage of BGP paths into or out of a country that pass through the AS. At the country level, we aggregate chokepoint potential in different ways to quantify the tendency of all BGP paths crossing the border to pass through a small (or large) number of chokepoints (border ASes). We refer to this as *national chokepoint potential (NCP)*. With these measures we ask how the AS-level topology has changed over time with respect to national boundaries. We find significant diversity among countries, both in absolute chokepoint potential and in their trends over time, but overall our analysis confirms the general belief that Internet topology is evolving to reflect national boundaries.

We developed a suite of tools, called BGP Simulation, Analysis, and Storage (BGP-SAS), for studying national chokepoints on the AS graph efficiently. To illustrate these ideas and tools, we study how NCP correlates with two independent measures of civil liberty, finding that a significant relationship exists between NCP and each measure. The paper extends earlier research on AS topology in several ways: it introduces chokepoint potential, it describes our open-source cross-platform tools and datasets for simulating Border Gateway Protocol (BGP) paths, determining NCP, and analyzing changes over time. In addition, we report and analyze data for several countries of interest.

The main contributions of the paper are:

(1) We define *chokepoint potential* a novel measure of AS-level national chokepoints, both for individual border ASes and for an entire country.
(2) A study of how national AS-level chokepoints have changed over the past ten years, demonstrating that the Internet has evolved to facilitate stronger AS-level chokepoints for some countries and led to the more open flow of information across borders for others.
(3) The open-source tool, BGP-SAS, and datasets for simulating BGP paths and evaluating chokepoints efficiently at different time points for the entire Internet.
(4) A study of the relationship between chokepoint potential and Internet freedom, as measured by two independent sources. We find a statistically significant relationship, suggesting that chokepoint potential is correlated with a country's tendency to conduct censorship or surveillance of its international Internet traffic.

The remainder of the paper is organized as follows: Section 2 provides relevant background for the problems we are investigating; Section 3 introduces and defines the measures of chokepoint potential and national chokepoint potential; Section 4 describes BGP-SAS for simulating and evaluating BGP networks; Section 5 explains the experimental setup and data sources; Sections 6, 7, and 8 contain experimental results, related work, and discussion, respectively. Section 9 concludes the paper.

2 BACKGROUND

2.1 BGP Data

The Border Gateway Protocol (BGP) is used to route traffic among ASes [24]. AS-level routes are known as *paths* and are stored in routing tables locally by each AS. The paths in these tables are selected to forward traffic based on local preferences configured for each AS.

There is no single accepted method for inferring the topology of the AS graph from BGP data. Several strategies have been developed, each with advantages and drawbacks. One can directly measure AS paths via *traceroute*, but collecting sufficient data requires many routing resources [31]. Collecting routes from publicly hosted routing tables such as from the Route Views Project [5] provides sufficient data, but many paths are malformed, including cycles or false paths. However, these data can be used to infer business relationships between ASes, and a routing model can be used to find likely paths between source and destination pairs [30].

A popular routing model for BGP is the Gao-Rexford model (GR routing), pioneered by Gao in [19]. This model is not without error, however. Gill *et al.* [21] showed in a survey that some network operators violate the rules of this model. For the sake of this project, we adopt GR routing, as most operators in the survey followed expectations, and our objective is a broad analysis of global trends and not fine-grained detail that a *traceroute* measurement study might provide. The business relationships between ASes have been inferred by the CAIDA team [30] and we use their provided relationship annotations [1].

We apply GR routing to the annotated business relationships to generate a list of likely paths between every AS. An efficient algorithm to do this was introduced as BGPSim by Gill *et al.*[20] . BGPSim takes as input a set of AS relationships and returns routing trees, represented as directed-acyclic-graphs (DAGs). A routing tree is a union of all paths found using a modified breadth-first search (BFS) algorithm. The BFS adds edges to routing trees first according to local preference (LP), then shortest path (SP), and finally tiebreak (TB). We adopted the basic algorithm, but reimplemented it so we could efficiently generate and save routing trees to disk, allowing us to easily extract ordered paths.

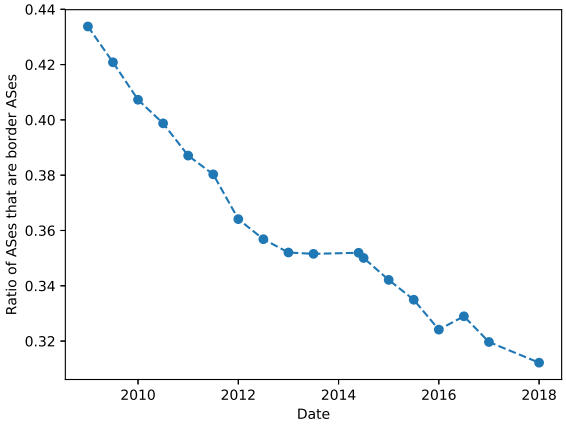

Figure 1: The ratio of border ASes to all ASes plotted over ten years.

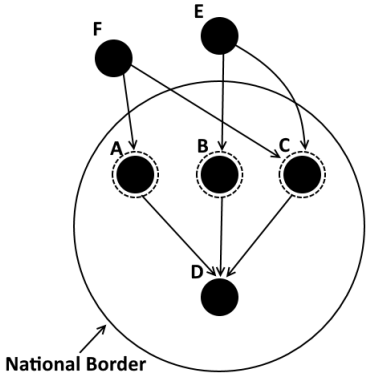

Figure 2: Chokepoint potential example. ASes A,B, and C are all border ASes. AS D is an internal AS. ASes E and F are both external ASes. The out-to-in chokepoint potentials of A,B, and C are 0.25, 0.25, and 0.5 respectively.

2.2 Internet Borders

The chokepoints in a nation's Internet border can change for a variety of reasons. For instance, China conducts keyword filtering in both border ASes and internal provincial ASes [39], while Iran routes its Internet traffic through a centralized facility [11]. A nation can attempt to restructure its Internet to prevent other nations from having access to their domestic traffic, concerns that have been voiced by China, Brazil, and Russia [6, 7, 25].

Nations can take advantage of powerful Internet chokepoints for national level interference. A canonical example is the "Great Firewall of China," which was first documented in 2006 in the technical research literature by Clayton *et al.* [13]. Large-scale Internet surveillance is also prominent and has a long history, but surveillance is much more difficult to measure than censorship, and information dumps such as the 2013 Snowden revelations provide only a partial view of its prevalence and duration. National-scale manipulation of packets for reasons other than censorship, for example to inject malware or carry out distributed denial-of-service attacks, was publicly identified when China's "Great Cannon" was detected in 2015 [32]. More recently, Egypt conducted similar injections using a commercial product [33]. This phenomenon is more targeted, and therefore more stealthy, than national-scale censorship, and it may well extend beyond these two publicly documented examples.

3 BORDER ASES AND CHOKEPOINT POTENTIAL

In this section we motivate the choice to focus our analysis on border ASes and then introduce and define the measure of chokepoint potential. We additionally introduce the aggregate measure of chokepoint potential, national chokepoint potential (NCP) for ranking and comparing countries according to the strength of their border AS chokepoints.

3.1 Border ASes

Border ASes are ASes that lie adjacent to an AS registered to a different country along a routing path. More formally, we define $g(\cdot)$ as the mapping from ASes to the country of registration, and say an AS a_u is a border AS if $g(a_u) \neq g(a_v)$, where a_v is a neighboring AS. Border ASes must be transited by any BGP path including multiple nations. We see in Figure1 that the ratio of border ASes to all ASes has decreased over time, indicating that these border ASes likely intercept more transnational paths. Because AS-level paths are generated by the distributed BGP, a useful way to understand the relative strengths of chokepoints is by comparing what fraction of paths they intercept. We propose an intuitive measure of AS chokepoints called *chokepoint potential*. We compare countries using an aggregate form of this measure called *national chokepoint potential*.

3.2 Chokepoint Potential

We define a BGP path of length n from source AS a_1 to destination AS a_n as a sequence $p_{1 \rightsquigarrow n} = \{a_i : i \in [1, n]\}$. We define a *routing tree* to be a tree rooted at a single destination AS and composed of every acceptable BGP path from the various source ASes that can reach the destination AS. Note that if multiple paths exist from a source to a destination, the routing tree is technically a DAG, but we use the traditional convention and call all such data structures routing trees. The chokepoint potential of a border AS is the ratio of the number of paths in (or out of) the country that contain the border AS to the total number of paths in (or out of) the same country. Formally, for country c, let $P = \{p_{i \rightsquigarrow j} : g(a_i) = c \text{ and } g(a_j) \neq c\}$, i.e. the set of all paths originating in c and ending outside c. Then the chokepoint potential, $cp(\cdot)$, of a border AS a_u with $g(a_u) = c$ is

$$cp(a_u) = \frac{|\{p : p \in P \text{ and } a_u \in p\}|}{|P|} \quad (1)$$

The above definition applies to paths originating in country c and routing to a different country, which we will refer to as *outgoing* chokepoint potential or *cpo*. We similarly define *incoming* chokepoint potential or *cpi*, by redefining P as $\{p_{i \rightsquigarrow j} : g(a_i) \neq c \text{ and } g(a_j) = c\}$. That is all paths originating outside of c but terminating in c and recomputing Equation 1.

The incoming chokepoint potential for a countries is illustrated in Figure 2. We define both *cpo* and *cpi* because past work has revealed that country-level paths are often asymmetric, meaning that the forward path from AS a to AS b does not necessarily match the reverse path [15].

Given a set of BGP paths, chokepoint potential is an intuitive way to compare individual ASes. Note that the sum of the chokepoint potentials for all border ASes for a given country is 1.0. We define the aggregate *national chokepoint potential* as the number of border ASes required to control a particular fraction f of paths. Formally, given a country,c, with a set of k border ASes $\{a_1, a_2, \cdots, a_k\}$ with $cp(a_i) > cp(a_j) \forall i < j$ (i.e. sorted in descending order), we find the smallest j such that

$$\sum_{i=1}^{j} cp(a_i) > f \tag{2}$$

and define national chokepoint potential as $CP(c, f) = 1/j$. The multiplicative inverse preserves a clearer semantic meaning i.e. higher chokepoint potential implies more control over paths. The fewer border ASes required to control f paths, the easier it is to perform censorship or surveillance, e.g., by restricting ISPs, placing filtering hardware, etc. Previous work used $f = 0.90$ as a threshold indicating strong control of information [8], so we use this percentage as well. When used here, however, 90% path control indicates the percentage of paths into (or out of) a country. National chokepoint potential is calculated as the smallest number of border ASes needed to control f of the paths.[1]

We measure the absolute number of ASes required to control a specific fraction rather than a percentage within a country to make comparisons between countries more equitable. Consider, for instance, the U.S., which has many more ASes than most countries, and China, which has a much smaller number of ASes. If the United States and China were to both require the same percentage of border ASes to intercept 90% of paths, the similarity of these nations would be misleading because the cost of controlling the larger absolute number of ASes would be significantly higher. Our technique provides a comparison more likely to reflect reality.

4 BGP-SAS

This section describes BGP-SAS, how it relates to earlier work, and how it is used to calculate chokepoint potential.

BGP-SAS determines a set of plausible BGP paths through simulation and calculates AS chokepoint potential and national chokepoint potential. BGP-SAS takes as input AS relationship data, such as that provided by CAIDA [1]. It also takes as input a set of AS country codes (that identifies which nation an AS belongs to) for determining chokepoint potential. BGP-SAS uses an algorithm based on BGPSim [20] to generate a set of routing trees. BGP-SAS calculates both AS and national chokepoint potential for any country.

Our experiments used country codes returned from Team Cymru's IP to ASN whois service [4] to determine which ASes were registered to which nation. This service maintains only the most recent registration of each AS. To control for this possible source of error, we retain ASes in the path simulation, even if they were registered more recently than the current timestamp being studied. This is because the relationships of these ASes and paths through these ASes are still valid. We do not assign these ASes chokepoint potentials, as it is not clear which nation these ASes are registered to. We took this approach because the datasets of AS registrations for the various ASN registries have uneven coverage, with more missing data at specific timestamps. We do not expect this to have a significant effect on our results because few ASes cannot be assigned definitively (in the best case in the most recent test this is less than 1% of ASes, in the worst case in the earliest test it is around 12% of ASes.) Additionally, some ASes that are unassigned are internal ASes, and those will not affect our results at all.

To calculate the routing trees, we use an extended version of the BGPSim algorithm developed by Gill et. al in [20]. This component of BGP-SAS addresses several limitations that prevented us from using BGPSim directly: (1) BGPSim was not used to test statistics on paths, so it doesn't return ordered paths for calculating AS-level statistics; (2) Once routing trees are determined, they cannot be accessed later without recalculation; (3) BGPSim relies on the outdated parallelization framework DryadLinq for C#. To address these issues, we implemented our routing simulation in Python. BGP-SAS returns ordered paths from its routing trees, saves routing trees to disk after calculation, and is parallelized with MPI via the mpi4py library. These improvements provide a cross-platform routing tree algorithm that is portable to most hardware. The source code for BGP-SAS and all datasets are publicly available[2].

Once BGP-SAS generates the routing trees, they can be processed to determine chokepoint potentials or by researchers interested in other questions. To calculate chokepoint potential, BGP-SAS iterates over every path between each AS-pair. Because we use the same random tie-break method as BGPSim, this process returns exactly one path between each AS-pair considered, even if multiple options exist in the routing tree. Tiebreaks do not have a noticeable effect on our results. For instance, the maximum standard deviation for the chokepoint potential of any AS in 5 runs of one of our timestamps was ≤ 0.005, meaning the chokepoint potentials for each AS barely changed based on the tiebreaks. Once the path has been determined, it is traversed to identify border nodes and increment their counts, both for outgoing and incoming paths. Thus, only one traversal is conducted per path. Additionally, the number of paths of each type that belong to each nation is tallied. BGP-SAS takes the resulting chokepoint potentials and generates national chokepoint potentials for each nation.

Calculating routing trees for the entire AS graph takes time proportional to $|V|^2$ where $|V|$ is the total number of ASes; calculating chokepoint potential takes $|V|^2 D$ where D is the average depth of a routing tree. For example, it took about 12 wall clock hours to compute all of the routing trees for the Jan. 2017 dataset, running on 4 Intel Xeon E5-2680 V4 CPUs with 25 threads running per CPU.

[1] We note that a dual measure could be defined as the fraction of paths controlled by j ASes. Our results are substantively the same for either definition.

[2] BGP-SAS and the routing tree dataset are available at https://kirtusleyba.github.io/routingtrees

For each dataset, this is a one-time cost because the routing trees are stored externally.

5 EXPERIMENTAL SETUP

We used BGP-SAS to ask three research questions related to national borders and the BGP network: (1) To what extent do BGP networks today reflect national boundaries and how do they vary across nations? (2) How has this structure evolved over time? and (3) How do topological chokepoints correlate with independent measures of Internet and press freedom?

For each point in time that we studied, we used the closest CAIDA AS relationship to that time. These contain inferred relationships between ASes, which the simulation framework uses to generate routing trees and calculate chokepoint potentials. We use the relationships to identify border ASes, and we use the Team Cymru ASN lookup service to map each AS to the country with which it is registered. BGP-SAS returns both the set of routing trees built via simulation and the chokepoint potential that was determined for each AS. These data form the basis for multiple levels of analysis: comparing chokepoint potential between countries, assessing how many ASes are required to control most international BGP paths, and comparing results across multiple time points.

Using the chokepoint data from our experiments, we tested the statistical relationship between national chokepoint potential and two relevant qualitative evaluations of Internet freedom. First, we used Freedom House's Freedom On The Net (FOTN) report [2]. FOTN scores quantify the level of Internet freedom in countries. Each country receives a numerical score from 0 (the most free) to 100 (the least free). We compare our results to FOTN scores (for years 2014-2017) by ranking nations according to the number of border ASes required to intercept 90% of paths for each nation. FOTN includes only 65 nations in the most recent report (2017), so we also considered the Freedom of the Press evaluation (2014-2017), which assesses 201 nations and territories.

6 EXPERIMENTAL RESULTS

This section reports our experimental results for each of the three research questions described in Section 5.

6.1 National Chokepoint Potential (NCP)

To analyze national chokepoint potential we selected eight nations to study in depth, although our datasets and tools support exploration for all countries. We chose countries with varying levels of Internet freedom, according to [2]: The United States, Great Britain, Germany, Brazil, India, Russia, Egypt, and China. Of these, the United States, Great Britain, and Germany are labeled as free; Brazil and India as partly free; and Russia, Egypt, and China as not free. Additionally, all of these countries practice censorship at some level [37] and they represent a spectrum of AS topology sizes (number of ASes).

Figure 3 shows national chokepoint potential for 2018 for different values of f (x-axis). The y-axis reports the inverse of the number of ASes required to control fraction f of paths, i.e. $NCP(f)$. We report only in-to-out results because the results for out-to-in are very similar for all experiments. Figure 3 shows the diverse nature of national topologies. China is one extreme, where over 80% of paths can be controlled by a single AS, Other nations exhibiting strong chokepoints are Egypt, Turkey, and the United Kingdom. The nations of the US, France, Germany, and Russia show considerably lower NCP for 2018. These results are as expected, except possibly for Russia. As an early adopter of the Internet, Russia's AS topology developed before the current era of concern over data localization and censorship, and we speculate that this early history dominates our measurements. Russia's current efforts to 'temporarily disconnect from the Internet,' stand up its own DNS, and reroute traffic through government-controlled routers suggest that the NCP we observe in the data reflects in part a domestic network structure that is problematic for current governmental policies.

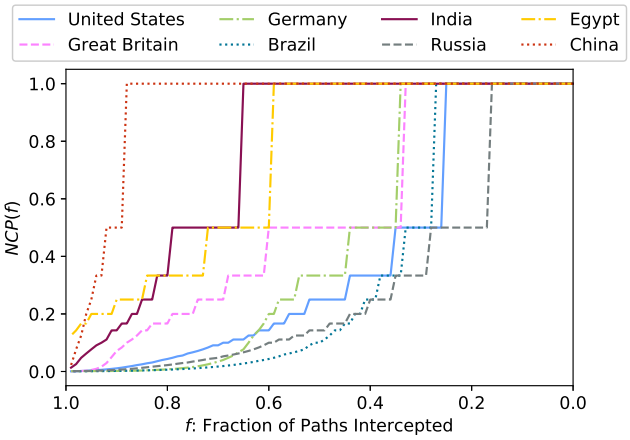

Figure 3: The national chokepoint potentials of various nations in January of 2018. f is the fraction of paths used to calculate the national chokepoint potential, i.e. the ratio of border ASes selected to count intercepted paths.

6.2 Topology Over Time

To study changes over time, we first plotted the national chokepoint potential of each of the eight selected countries for each year (Figure 4. In these results we see that the dynamics of NCP vary by country. For instance China has grown dramatically in NCP, indicating that the most powerful border ASes in China now intercept even more paths than in the past, having an NCP > 0.5, which translates to needing only two carefully chosen ASes to intercept more than 90 percent of its international paths. Other countries have a more stable NCP, such as the U.S. and India, although the U.S. NCP decreased slightly. A stable NCP, in the presence of overall growth in the number of ASes, suggests that new source and destination AS pairs from new infrastructure tend to reuse border gateways instead of creating new ones, but the new paths are relatively evenly distributed. Russian border ASes, for instance, likely intercept more paths overall than in the past, but maintain their ratio of all the transnational paths. NCP has decreased steadily for Germany and Brazil, as border ASes in these nations have developed a more even share of international paths. This could be for many reasons, including heavy reliance on paths that traverse foreign networks to access popular sites [16], which influences connections to foreign

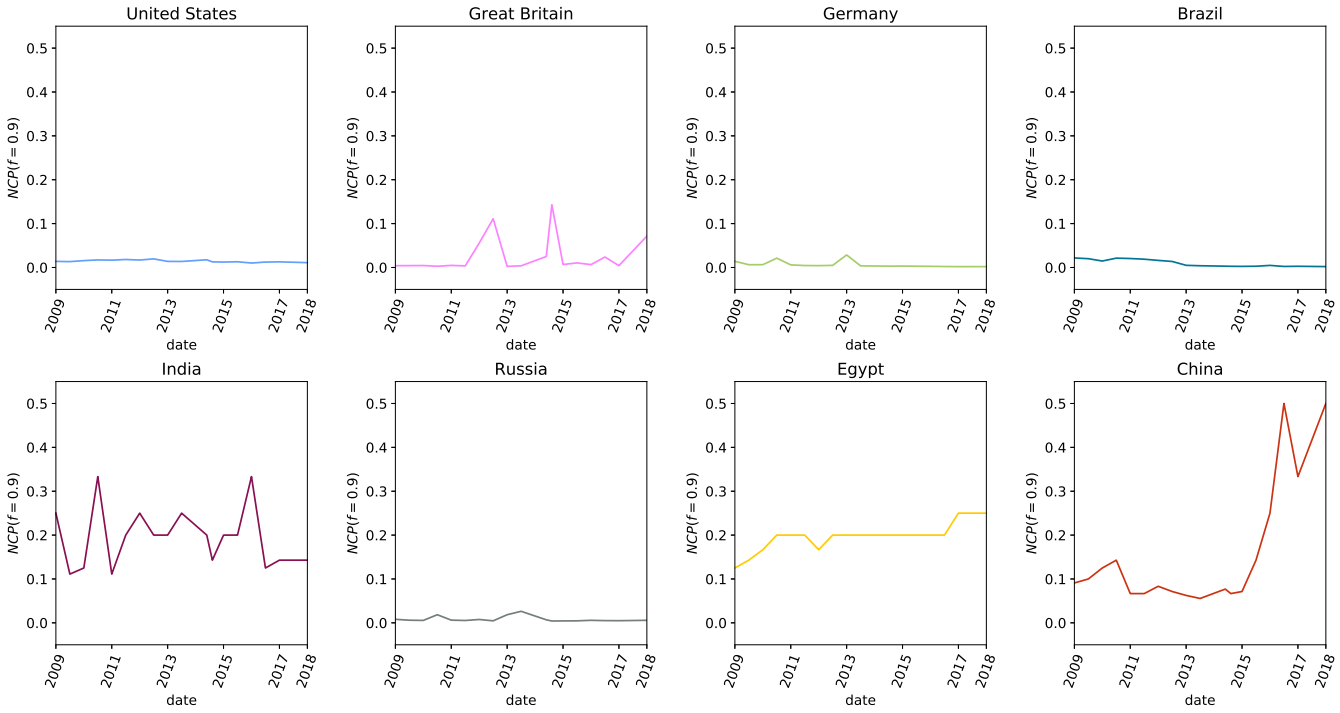

Figure 4: The time evolution of NCP for multiple nations. The x-axis is the year of the observation and the y axis is the calculated NCP for $f = 0.9$.

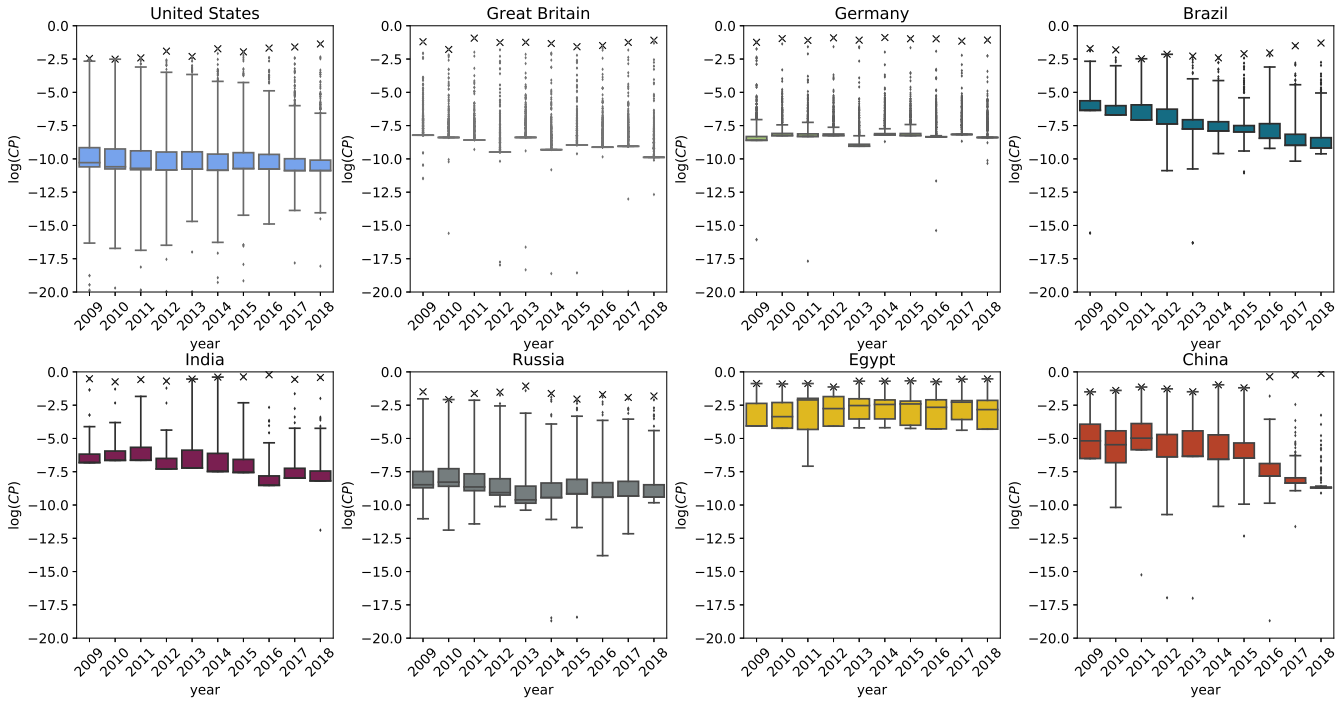

Figure 5: Distribution of the $log(CP)$ values for each of the 8 selected nations over multiple years. For each year, a box-and-whisker plot shows the quartiles and outliers for a particular nation's set of border ASes. The AS with maximum $log(CP)$ is indicated by an x in each timestamp.

routers. Regardless of the cause, these nations must intercept traffic at more ASes now to intercept an equal fraction of paths as before.

In Great Britain, major legislation to force ISPs to block pornography was enacted in 2015 and 2017 [17, 35] (note the two peaks in Figure 4). During this time period Great Britain was in the EU, which was discussing net neutrality legislation [34]. These competing pressures, combined with the fact that Great Britain's IXP-based infrastructure simplifies the creation and destruction of chokepoints via the BGP system, is a possible explanation for the significant fluctuations observed in the Great Britain data during this time period.

Next, we consider the visualization of chokepoint dynamics shown in Figure 5. In each subplot, the distribution of chokepoint values of border ASes is shown for that country via box-and-whisker plots. In all subplots we observe that one or a very few ASes intercept many paths. This can be seen by noting that for each year, in each country, the highest outlier is close to 0.0. The strength of these number-one chokepoints varies substantially, however. For instance, the highest Chinese AS with the highest chokepoint potential intercepts the vast majority of transnational paths, while for other nations like Brazil, the U.S. and Russia, the number one border AS only intercepts around 10% of such paths. Another common feature is the tendency for most ASes to intercept a small ratio of paths (mean is well separated from the least negative outlier). Because Egypt has a very small AS graph to begin with, its relatively high mean chokepoint potential (closer to zero on the y-axis) is an artifact of its small size rather than an open border. Turning to Germany and Brazil, which Figure 4 suggests have declining NCPs, this can be explained by considering Figure 5, which shows that these two countries have evolved a more even distribution of paths across border ASes.

6.3 Internet Freedom

To test the correlation between Internet Freedom and NCP, we compared FOTN scores against the number of border ASes required to intercept 90% of outgoing paths, $NCP(f = 0.9)$, and computed the Ordinary Least Squares (OLS) fits, finding that the relationship is statistically significant (e.g., p-value ≤ 0.0001 for 2017). The relationship was significant for the other years studied. As mentioned previously, FOTN scores rank only a small number of nations, so we also compared Freedom House's Freedom of the Press (FOTP) rankings and found a significant relationship with a p-value ≤ 0.001.

The trend holds for all years investigated: Countries that are not free or partly free according to FOTN and FOTP tend to have high NCP, while free nations tend to have low NCPs. There are interesting outliers for both situations, however. Countries like Estonia and Iceland are very free but require few border ASes to control most of their paths. This is likely because their overall AS counts are so low. As we saw earlier, Russia is considered not free by FOTN, but requires a large number of ASes to control most of its paths. This suggests that censorship in Russia is implemented despite relatively open Internet borders at the AS-level.

These results points to an important question: Why does a relationship exist between the Internet freedom the national chokepoint potential of many nations? We cannot infer any causal relationship using our data, but we can hypothesize that the technology used for censorship practices is more easily deployable on a national network with powerful chokepoints. It also stands to reason that in countries which conduct extensive censorship or surveillance it would be more difficult to establish new connections in and out of the nation, particularly for those where the primary ISPs are tightly controlled by the government.

7 RELATED WORK

Researchers have investigated transnational Internet routes as they relate to national Internet hegemony in the past. Edmundson *et al.* used a measurement approach to study national level paths. They particularly focused on a phenomenon they refer to as "tromboning paths," where a path to a domestic domain takes a detour to another nation. These detours usually pass through the United States, and Edmundson *et al.* mention that this could show that the United States has disproportionate ability to interfere with Internet paths [16]. We have presented an alternative point of view in this paper, highlighting the fact that some nations can intercept more paths on fewer ASes. While tromboning paths might lead to a loss of the sovereignty of a nation's domestic Internet traffic, they do not explicitly assist or prevent a national government from interfering with its own paths to foreign destinations. We argue that both the interference by foreign powers, such as packet sniffing on tromboning paths, and chokepoint control at national borders are relevant to the Internet freedoms of a particular nation. As such, these viewpoints taken together broaden current understanding of the global Internet control dynamics.

Previous studies also used BGP path models to find ASes that intercept a high fraction of paths, e.g., [8] which reported that 90% of paths on the Internet could be intercepted by only 30 or so ASes. This work generated paths starting with paths to top websites as defined by the Alexa top websites project, and then appending additional edges from the full AS graph according to the Gao algorithm [19]. This work was extended in [22], which showed how ASes that intercept many paths could be used for decoy routing. Our paper extends this work in several ways. First, many of the websites in the Alexa dataset are in China, due to its large Internet population, potentially biasing results, while we consider paths between each source-destination AS pair. More importantly, our work studies how chokepoints have changed over time, rather than considering a single snapshot. Finally, we compare chokepoints quantitatively between different countries by defining a measure, rather than simply identifying those countries with ASes that intercept the most paths.

In [39], Xu et al. investigated the AS level topology of China to identify where keyword filtering occurred. They found that the most effective ASes are those in the backbone of the Chinese AS topology. A relevant contribution of [39] is that, while most filtering occurs in border ASes, some filtering occurs in provincial ASes. China had a diverse strategy for Internet censorship at the time, targeting both chokepoints and the Chinese provincial network, but this may have changed since. The potential for various forms of censorship in regards to various AS level topologies motivates the question: Is centralized censorship or decentralized censorship more common? Instead of directly identifying censorship devices on

the AS graph, we instead have quantified the chokepoint potential of ASes on borders, and then compared that to Internet freedom measures.

Similar techniques to those reported here have been used to classify nations according to the connectivity of their ASes [36] for a single snapshot in time. Our work presented here is more general, allowing each AS to be quantified in regards to chokepoint potential and studying the evolution of chokepoints over time.

Routing and its interplay with Internet censorship, surveillance, and related issues is a general research area with a broad set of research questions. Karlin *et al.* [26] considered the centrality of countries with respect to routing of other nations' traffic, a related problem that is distinct from chokepoints to monitor/control traffic into and out of a government's own country. Dainotti *et al.* [14] analyze two specific large-scale Internet disruptions at the routing level. Khattak *et al.* [28] performed a detailed analysis of censorship at the ISP level in Pakistan.

While there have been attempts to characterize methods for censoring Internet content [27, 38] there is no comprehensive list of all the different ways a state actor can manipulate traffic. For some methods, the AS graph is relevant, such as IP address blacklisting, traffic throttling [10], URL or packet filtering, packet injection, physically shutting down infrastructure, and BGP attacks. For other methods, the AS graph is less relevant, such as creating internal national networks [9], portal censorship such as search engine filtering and social media post deletion, propaganda campaigns, and manipulation of the DNS system.

8 DISCUSSION

8.1 Routing Trees Dataset

Despite the use of efficient simulation techniques, collecting BGP path data remains challenging. For a global, multiple timestamp study such as this one, computation times can be a limiting factor to researchers without access to supercomputing resources. We have publicly released the routing tree datasets and chokepoint potential calculations produced for this study generated with BGP-SAS. By releasing these datasets we hope that researchers looking for a particular set of routing trees will find working with these simpler than recalculating them, or creating new datasets by extensive measurement or inference.

8.2 Other Chokepoint Measures

There are several ways that a country can create chokepoints, either intentionally or accidentally: taking advantage of Internet Exchange Points (IXPs), limiting the number of physical connections that cross the border, centralizing the DNS infrastructure within the country, or working with—and supporting—a set of ISPs that have international connections. These methods may reduce the number of organizations controlling Internet paths and the number of physical locations at which traffic needs to be intercepted. By choosing the AS graph we focus on the virtualized layer of Internet connectivity instead of physical locations. Physical chokepoints need further study as another element of Internet route control. A benefit of studying the AS-level of Internet chokepoints is that the dramatic shifts in chokepoint potential over time depicted in our analysis are unrelated to constant geography. Business and political decisions influencing the control dynamic of Internet borders do not necessarily effect the geographical locations of Internet resources. Despite this, our chokepoint measure can quantify AS border chokepoints.

There are other aspects of the Internet where chokepoints can manifest. The DNS infrastructure, for instance, is a network where control over chokepoints can have political, security, and censorship ramifications [23]. DNS chokepoints are out of the scope of this study, and we leave such analysis for future work.

8.3 Value of Chokepoints

There are many different ways to implement Internet censorship, many of which require exerting influence on communications between two end hosts. A state actor conducting censorship on the network must position censorship devices in the path between the hosts, or at either host. Types of censorship that occur in the network include IP address blacklisting, network-based DNS tampering, web proxy-based filtering, deep packet inspection, and bandwidth throttling. The AS graph is directly relevant to all of these techniques because they require that Internet communications route through the censorship. Censorship on the end host, which can be on the server or the client, can include keyword-based content filtering, human monitors, post deletions, account suspensions and deletions, tampering within the DNS system itself, and any application-specific behavior that the application or server's maintainers program into the software. Although the AS graph is not directly related to these kinds of censorship, they are usually predicated on network-level censorship to remove uncensored alternatives. By making uncensored alternative applications, application servers, and DNS servers unavailable at the AS level, users are forced to opt into host-based censorship. To use China as an example, consider Weibo vs. Twitter and WeChat vs. WhatsApp: network-level blocking of Twitter's web service and WhatsApp's communications are prerequisite to censorship built into Weibo and WeChat. These domestic services are heavily censored [40], and the network level blocking prevents the use of international alternatives. Thus, because the AS-level graph governs routing, it transitively plays an important role in virtually all forms of Internet censorship.

In this paper, we have focused on how National Chokepoint potential is related to Internet freedom and the ability for a nation to control their local Internet. However, there may be other implications for countries with extremely high or low chokepoint potential. Countries with low chokepoint potential, that spread routing paths over a diverse number ASes arguably more robust to attacks on or failures of high chokepoint ASes. Conversely, high chokepoint countries may be able to react and coordinate responses to the same attacks or failures.

8.4 Limitations

The primary limitation of our results is the extent to which the set of paths calculated in our simulation framework agrees with reality. In practice, network operators do not always follow GR-routing policy when designing their routing preferences [21]. Unfortunately there is no perfect way to replicate the set of routes used on the Internet.

Collecting routes directly from publicly available routing collectors introduces errors due to path-poisoning, misconfigurations, and other sources of spurious links [29]. Additionally, limiting the dataset to measured data or public routing data would include only paths that are directly visible to the particular collectors we used. Simulation can generate an extensive set of paths reflecting the GR model using well-validated edge relationships as input, which has the benefit of identifying any possible GR model path between two ASes. We acknowledge the limitations of this approach but we believe it yields a practical estimation of network structure for evaluating chokepoint potential.

8.5 Future Work

While linking chokepoint potential to FOTN scores is a substantial contribution, FOTN is only a proxy for censorship. The Open Observatory of Network Interference, or OONI, [18] provides Internet users around the world with the ooniprobe. The ooniprobe lets users run a suite of tests to identify censorship anomalies of various types, and the results are recorded in the large OONI database. Comparing increased censorship campaigning in an authoritarian nation, with shifts in chokepoint potential would be a major step in understanding the interplay of censorship and AS-level chokepoints. This process involves designing a way to classify censorship events and chokepoint potential changes, and as such lies beyond the scope of this study.

9 CONCLUSION

It is generally believed that the Internet is becoming more national. This paper addresses the question of how Internet structure relates to international boundaries and how it has changed. Using our new measures of AS-level chokepoints, chokepoint potential and national chokepoint potential, we have shown that it is common for nations to consolidate paths through powerful chokepoints. We have shown that over the past decade some nations have increasingly closed off their Internet borders, while others seem to be becoming more open as new paths to the international Internet are opened. Additionally, we studied the relationship between national chokepoint potential and two evaluations of openness, Freedom on the Net and Freedom of the Press, finding statistically significant relationships in both cases.

Our technique for generating BGP paths and evaluating chokepoints using BGP-SAS provides public domain software and routing trees for the entire AS graph. We have taken advantage of efficient simulation, standard AS relationship datasets, and cross-platform design principles so that this tool will be readily deployable for future research.

REFERENCES

[1] Caida as relationships <date range: 01-2009 to 01-2018>. http://data.caida.org/datasets/as-relationships/.

[2] Freedom on the net 2017. https://freedomhouse.org/sites/default/files/FOTN_2017_Final.pdf.

[3] International telecommunication union internet statistics. https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx.

[4] Ip-asn mapping tool from team cymru. https://www.team-cymru.com/IP-ASN-mapping.html.

[5] University of oregon route views project.

[6] Brazil looks to break from us-centric internet. 2013.

[7] Russia considers 'unplugging' from internet. 2019.

[8] HB Acharya, Sambuddho Chakravarty, and Devashish Gosain. Few throats to choke: On the current structure of the internet. In *Local Computer Networks (LCN), 2017 IEEE 42nd Conference on*, pages 339–346. IEEE, 2017.

[9] Collin Anderson. The hidden internet of iran: Private address allocations on a national network. *CoRR*, abs/1209.6398, 2012.

[10] Collin Anderson. Dimming the internet: Detecting throttling as a mechanism of censorship in iran. *CoRR*, abs/1306.4361, 2013.

[11] Simurgh Aryan, Homa Aryan, and J Alex Halderman. Internet censorship in iran: A first look. In *FOCI*, 2013.

[12] Hyunseok Chang, Sugih Jamin, and Walter Willinger. To peer or not to peer: Modeling the evolution of the internet's as-level topology. In *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*, pages 1–12. Citeseer, 2006.

[13] Richard Clayton, Steven J. Murdoch, and Robert N. M. Watson. Ignoring the great firewall of china. In *Proceedings of the 6th International Conference on Privacy Enhancing Technologies*, PET'06, pages 20–35, Berlin, Heidelberg, 2006. Springer-Verlag.

[14] Alberto Dainotti, Claudio Squarcella, Emile Aben, Kimberly C. Claffy, Marco Chiesa, Michele Russo, and Antonio Pescapé. Analysis of country-wide internet outages caused by censorship. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, IMC '11, pages 1–18, New York, NY, USA, 2011. ACM.

[15] Anne Edmundson, Roya Ensafi, Nick Feamster, and Jennifer Rexford. Characterizing and avoiding routing detours through surveillance states. *arXiv preprint arXiv:1605.07685*, 2016.

[16] Anne Edmundson, Roya Ensafi, Nick Feamster, and Jennifer Rexford. Nation-state hegemony in internet routing. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, COMPASS '18, pages 17:1–17:11, New York, NY, USA, 2018. ACM.

[17] Loulla-Mae Eleftheriou-Smith. Eu to block david cameron's plans on internet porn crackdown. 2015.

[18] Arturo Filasto and Jacob Appelbaum. Ooni: Open observatory of network interference. In *FOCI*, 2012.

[19] Lixin Gao. On inferring autonomous system relationships in the internet. *IEEE/ACM Transactions on networking*, 9(6):733–745, 2001.

[20] Phillipa Gill, Michael Schapira, and Sharon Goldberg. Modeling on quicksand: dealing with the scarcity of ground truth in interdomain routing data. *ACM SIGCOMM Computer Communication Review*, 42(1):40–46, 2012.

[21] Phillipa Gill, Michael Schapira, and Sharon Goldberg. A survey of interdomain routing policies. *Computer Communication Review*, 44(1):28–34, 2014.

[22] Devashish Gosain, Anshika Agarwal, Sambuddho Chakravarty, and HB Acharya. The devil's in the details: Placing decoy routers in the internet. In *Proceedings of the 33rd Annual Computer Security Applications Conference*, pages 577–589. ACM, 2017.

[23] Benjamin Greschbach, Tobias Pulls, Laura M. Roberts, Philipp Winter, and Nick Feamster. The effect of DNS on Tor's anonymity. In *NDSS*. The Internet Society, 2017.

[24] John W. Stewart III. *BGP4: Inter-Domain Routing in the Internet*. Addison-Wesley, 1999.

[25] Jiang Jie. No internet hegemony: Xi. 2015.

[26] Josh Karlin, Stephanie Forrest, and Jennifer Rexford. Nation-state routing: Censorship, wiretapping, and BGP. *CoRR*, abs/0903.3218, 2009.

[27] Sheharbano Khattak, Tariq Elahi, Laurent Simon, Colleen M. Swanson, Steven J. Murdoch, and Ian Goldberg. SoK: Making sense of censorship resistance systems. *Privacy Enhancing Technologies*, 2016(4):37–61, 2016.

[28] Sheharbano Khattak, Mobin Javed, Syed Ali Khayam, Zartash Afzal Uzmi, and Vern Paxson. A look at the consequences of internet censorship through an isp lens. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, IMC '14, pages 271–284, New York, NY, USA, 2014. ACM.

[29] Matthew Luckie. Spurious routes in public bgp data. *ACM SIGCOMM Computer Communication Review*, 44(3):14–21, 2014.

[30] Matthew Luckie, Bradley Huffaker, Amogh Dhamdhere, Vasileios Giotsas, et al. As relationships, customer cones, and validation. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 243–256. ACM, 2013.

[31] Zhuoqing Morley Mao, Jennifer Rexford, Jia Wang, and Randy H Katz. Towards an accurate as-level traceroute tool. In *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 365–378. ACM, 2003.

[32] Bill Marczak, Jakub Dalek, Sarah McKune, Adam Senft, John Scott-Railton, and Ron Deibert. Bad traffic: Sandvine's PacketLogic devices used to deploy government spyware in Turkey and redirect Egyptian users to affiliate ads? Citizen Lab Report, available at https://citizenlab.ca/2018/03/bad-traffic-sandvines-packetlogic-devices-deploy-government-spyware-turkey-syria/.

[33] Bill Marczak, Nicholas Weaver, Jakub Dalek, Roya Ensafi, David Fifield, Sarah McKune, Arn Rey, John Scott-Railton, Ron Deibert, and Vern Paxson. An analysis of china's "great cannon". In *5th USENIX Workshop on Free and Open Communications on the Internet (FOCI 15)*, Washington, D.C., 2015. USENIX Association.

[34] Miranda Prynne. Nine out of ten homes to have porn filters within two months. 2013.
[35] Jamie Rigg. How the digital economy act will come between you and porn. 2017.
[36] Rachee Singh, Hyungjoon Koo, Najmehalsadat Miramirkhani, Fahimeh Mirhaj, Phillipa Gill, and Leman Akoglu. The politics of routing: Investigating the relationship between AS connectivity and internet freedom. In *6th USENIX Workshop on Free and Open Communications on the Internet (FOCI 16)*, Austin, TX, 2016. USENIX Association.
[37] Ramesh Subramanian. The growth of global internet censorship and circumvention: A survey. 2011.
[38] Michael Carl Tschantz, Sadia Afroz, Anonymous, and Vern Paxson. SoK: Towards grounding censorship circumvention in empiricism. In *Symposium on Security & Privacy*. IEEE, 2016.
[39] Xueyang Xu, Z Morley Mao, and J Alex Halderman. Internet censorship in china: Where does the filtering occur? In *International Conference on Passive and Active Network Measurement*, pages 133–142. Springer, 2011.
[40] Tao Zhu, David Phipps, Adam Pridgen, Jedidiah R Crandall, and Dan S Wallach. The velocity of censorship: High-fidelity detection of microblog post deletions. In *Presented as part of the 22nd {USENIX} Security Symposium ({USENIX} Security 13)*, pages 227–240, 2013.