UNIVERSITY OF CALIFORNIA SAN DIEGO

Transit Influence of Autonomous Systems: Country-Specific Exposure of Internet Traffic

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Computer Science

by

Alexander M. Gamero-Garrido

Committee in charge:

Professor Alberto Dainotti, Co-Chair
Professor Alex C. Snoeren, Co-Chair
Professor Margaret Roberts
Professor Stefan Savage
Professor Aaron Schulman

2021

The Dissertation of Alexander M. Gamero-Garrido is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

To my mom, dad, brother and husband.
It really does take a village.
You are mine.

EPIGRAPH

Caminante, son tus huellas el camino y nada más;
Caminante, no hay camino, se hace camino al andar.
Al andar se hace el camino, y al volver la vista atrás
se ve la senda que nunca se ha de volver a pisar.
Caminante, no hay camino sino estelas en la mar.

(Traveler, your footprints are the only road, nothing else.
Traveler, there is no road; you make your own path as you walk.
As you walk, you make your own road, and when you look back
you see the path you will never travel again.
Traveler, there is no road; only a ship's wake on the sea.)

Antonio Machado
(Translation by Mary G. Berg and Dennis Maloney)

We should be taught not to wait for inspiration to start a thing. Action always generates inspiration. Inspiration seldom generates action.

Frank Tibolt

You may shoot me with your words,
You may cut me with your eyes,
You may kill me with your hatefulness,
But still, like air, I'll rise.

Maya Angelou

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

If I were to express my gratitude towards everyone who has made this doctoral journey possible, the full document would no doubt occupy multiple volumes. I must, however, acknowledge as many folks as I can.

Let me begin by thanking my advisors, Alex Snoeren and Alberto Dainotti, for their invaluable insights and relentless support. Alex has been one of the few constants in my graduate career and has provided much needed stability, along with helping me develop my scientific intuition and thinking. Alberto also made my dissertation possible with his deep technical dives, as well as his patient explanations and thoughtful feedback. I could never have aspired to finding mentors of their analytical and scientific caliber, capacity for sincere empathy, generous advice on both professional and personal matters, and with whom I have enjoyed many hours of banter, but I am extremely grateful that they took a chance on me.

Stefan Savage, kc claffy, David D. Clark and Amogh Dhamdhere (at UCSD CSE, CAIDA, MIT and Amazon, respectively) have also been exceptional mentors, and I must express my gratitude for their co-authorship with me on matters of Internet measurement and technology policy. I am deeply impressed with their ability to articulate complex technical ideas using accessible language, and with their long-term commitment to improving the Internet. My gratitude also extends to two additional members of my committee for their generous feedback, Aaron Schulman Shalev and Margaret Roberts, as well as to my soon-to-be postdoctoral supervisor, David Choffnes.

Esteban Carisimo, the peer with whom I have collaborated most closely, deserves a special mention for his large reserves of patience and his always dependable assistance with technical questions. Muchas gracias amigo. I am also very grateful for the support from co-authors Kirill Levchenko, Shuai Hao, Bradley Huffaker, Matthew Luckie, Ricky KP Mok, Gautam Akiwate, Kabir Gogia, and Vaibhav Bajpal.

My time at UCSD would have not been nearly as enjoyable without the friendship of Graduate Women in Computing folks: Ariana Mirian, Steven Rick, Ailie Fraser and Tricia Ngoon among many others. My time with GradWIC is among my most cherished memories from grad school. I also must thank my CSE compatriots Peter Edge, Brian and Mary Johannesmeyer, Rob McGuiness, Sunjay Cauligi and Alex Sanchez-Stern, and also Canaan Breiss (in Los Angeles), for bringing much needed levity and entertainment that played no small part in helping me get through tough times.

Finally, none of this would have been possible without the unwavering support of my family—Arelis Garrido, Remigio Gamero, Wladimir Gamero Garrido, and David Arias—and my MIT friends Erin Leidy and Justin Montgomery. Thank you for always believing in my scientific career.

Following UC San Diego’s formatting guidelines, I proceed to acknowledge the projects that have contributed to my dissertation:

Chapters 2, 3, 4, and 5, in part, are currently being prepared for submission for publication of the material. Gamero-Garrido, Alexander; Carisimo, Esteban; Hao, Shuai; Huffaker, Bradley; Snoeren, Alex C.; Dainotti, Alberto. The dissertation author was the primary investigator and author of this paper.

Chapter 6, in part is currently being prepared for submission for publication of the material. Gamero-Garrido, Alexander; Carisimo, Esteban; Snoeren, Alex C.; Dainotti, Alberto. The dissertation author was the primary investigator and author of this material.

This dissertation was supported in part by the National Science Foundation, Grant No. CNS 1705024, and by the Microsoft Research Dissertation Grant (2019).

VITA

2009–2010	Exchange Student, Lund University School of Engineering, Sweden
2011	Electronics Engineer, Simon Bolivar University, Caracas, Venezuela
2011–2013	Positioning Engineer, Schlumberger
2015	Masters of Science, Massachusetts Institute of Technology
2021	Doctor of Philosophy, University of California San Diego

ABSTRACT OF THE DISSERTATION

Transit Influence of Autonomous Systems: Country-Specific Exposure of Internet Traffic

by

Alexander M. Gamero-Garrido

Doctor of Philosophy in Computer Science

University of California San Diego, 2021

Professor Alberto Dainotti, Co-Chair
Professor Alex C. Snoeren, Co-Chair

Computer networks play a central role in the transmission of information across the world. Autonomous systems (administrative domains or “ASes”) are the building blocks of such wide-area networks and are responsible for delivering traffic to their individual subscribers as well as to other networks to which they are connected. So-called transit ASes—who sell access to the rest of the Internet to customer ASes for a fee—are mostly invisible to end users but may be able to operate on their traffic, for instance by observing unencrypted traffic or metadata, or by tampering with specific network flows serving popular applications. In many countries, transit ASes serve as the principal intermediaries between domestic access ASes and the global Internet.

In this dissertation, we introduce the concept of transit influence, which quantifies the exposure of an AS or groups thereof, such as in an industrial sector, or nation, to observation and tampering by a specific transit network. We hypothesize that there are countries where transit agreements are the dominant form of international connectivity, and where specific ASes have significant degrees of transit influence (TI) over the country as a whole as well as over individual organizations within them. We quantify TI by developing three metrics at distinct granularities: at the country level (CTI), at the AS level (ATI), and at the sectoral level (W-ATI). In order to apply these methods, we first identify 75 countries—with approximately 1 billion Internet users, on aggregate—where transit providers are the dominant mode of inbound connectivity, using analyses of existing interconnection data and our own large-scale measurement campaign. Applying CTI, we find 32 nations that have transit ecosystems with concerning topological features: traffic destined to over 40% of their IP addresses is exposed to a single network. We further study the AS topologies of three nations in South America and find a small number of transit ASes that exert out-sized influence in several sectors, including finance, utilities, and education. We validate our findings with in-country network operators at 123 ASes in 19 countries, who confirm that our results are consistent with their understanding of their countries' networks.

Chapter 1

Introduction

The electronic transmission of information is among the most essential services in the modern world, enabling data communication at global scale. In comparison to the physical domain of the shipment of goods, where ports of entry and transit stations are well-defined—such as airports, mail sorting facilities, etc.—and often operated by tightly regulated public-sector institutions, digital data traverses computing and telecommunications infrastructure—such as data centers, submarine cables—that is both opaque and operated by a complex array of public-sector organizations and private, for-profit companies. This infrastructure delivers traffic to individual Internet subscribers, as well as to a wide array of enterprises including those operating critical services such as power grids.

The Internet has grown rapidly during the last three decades, with each successive wave of innovation digitizing key functions of an ever-increasing set of industries, a trend that was only accelerated and expanded by the onset of the COVID-19 pandemic. Yet, almost all popular Internet services are hosted in a small set of countries, forcing other nations to rely on international connectivity to access them. These large-scale digital communication services result in potentially vast opportunities for observation and tampering. For instance, observation (of unencrypted traffic and metadata) may be performed by domestic or foreign actors with the purpose of conducting surveillance or espionage, respectively. Selective tampering—for instance, with individual network flows carrying popular-application traffic—has been reported by both domestic actors (*e.g.,* government censorship [30, 62]) and foreign perpetrators (*e.g.,* dis-information campaigns [75, 93]). Critical organizations—such as power grid operators, government agencies and universities—may be exposed to these calamities, as they often rely on the Internet to send and receive content internationally.

Understanding the exposure of traffic to observation and tampering—both in terms of quantifying the potential exposure as well as to whom the information may be exposed—has important implications for Internet architecture, technology policy, national security and international relations. For instance, revealing that a state-owned company has capabilities to observe a large fraction of Internet traffic entering a country may allow activists and regulators to demand more transparency in the technical operations of that company. Conversely, disclosing dependencies on foreign companies for the transmission of information into a country may prompt the national government to invest into the creation of alternative routes and therefore mitigate such dependencies, for instance by deploying additional submarine cables.

Understanding this exposure requires measuring the extent to which the operator of any network delivers traffic towards a country and organizations within it, since traffic recipients include both individual consumers who purchase connectivity from a telecommunications operator, as well as institutions who operate their own networks. The goal of this dissertation is to create a technical framework that quantifies the exposure of nations—and individual organizations within them—to observation and tampering of their Internet traffic by specific networks.

1.1 High-Level Challenges

The Internet is not a monolithic entity administered by a centralized organization, but rather a massive distributed system maintained by tens of thousands of autonomous systems (AS), who are the Internet service providers (ISPs) "and other large organizations that are responsible for routing traffic both within the larger Internet and within their own network" [88]. Identifying the ASes to which a country's Internet traffic is exposed requires overcoming several technical challenges. Many of these obstacles stem from the implementation of the central protocol enabling wide-area interconnection across network operators, the Border Gateway Protocol (BGP). Few ASes have a truly global infrastructure footprint, meaning that, in general, ISPs (also referred to as access networks) must rely on third parties to deliver their content. The agreements between ISPs and these third-party ASes vary considerably in their scope and complexity, the details of which are almost always proprietary. Therefore, researchers interested in studying how ASes connect to each other—the AS-level topology—at scale must rely on inferences from public datasets that are voluntarily disclosed by a select few operators primarily based in North America and Europe [23, 84].

Business agreements between network operators are generally classified in two broad categories: *(i)* transit, where a (typically larger) provider connects a customer (and its customers) to the rest of the Internet for a fee, or *(ii)* peering, where two networks (generally of comparable size) agree to exchange traffic towards each other and their customers, usually free of charge as long as the traffic volumes across them are roughly symmetric. The datasets used by academic researchers have good coverage of transit agreements [45, 74] because BGP typically allows for the propagation of network routes from customers to providers (and vice versa), recursively. By contrast, peering agreements are only visible by the networks involved (the peers) and their respective customers, meaning that public datasets are poorly suited to observing these connections.

1.2 Transit Network Visibility

ISPs in different countries rely on varying combinations of transit and peering agreements to connect their individual subscribers to the rest of the world, resulting in varying degrees of visibility of each country's Internet topology on public datasets. Peering interconnections often occur at physical sites such as Internet exchange points (IXPs), where networks exchange routes and traffic. In countries where these facilities are in early stages of development (or do not exist), ISPs must rely on transit providers, often foreign companies with the financial gravitas to deploy long-range terrestrial or submarine cables. Further, since transit providers generate revenue from their customers, they often do not have an incentive to establish peering agreements with potential customers in underserved markets.

Unlike access networks, which charge subscribers a recurring fee established in a contract, transit networks are frequently invisible to end users but may be able to similarly observe or tamper with Internet traffic. This opacity may allow both domestic and foreign actors to observe or tamper with traffic without facing diplomatic or political backlash from governments, activists or consumer groups. In this dissertation, we aim to bring transparency to the public regarding outsized observation and tampering capabilities available to specific transit networks in a large group of nations.

1.3 Thesis

We define *transit influence* as the set of capabilities that allows transit networks to observe or tamper with traffic flowing towards specific countries and organizations within them. We hypothesize that **there are countries where transit agreements are the dominant form of international connectivity, and where specific ASes have significant degrees of transit influence over the country as a whole as well as over individual organizations within them.**

1.4 Approach

Our goal is to study the exposure of each nation's traffic to observation and tampering of Internet traffic. Because actual traffic information is difficult to obtain at a global scale, we instead determine the fraction of IP addresses belonging to a country (or a specific organization) that are exposed to tampering and observation by specific networks. IP addresses are often used as a proxy for traffic, *e.g.,* in [92], and previous work has found strong correlations between number of IP addresses observed in BGP and traffic volume for ASes that provide either access or transit service [72].

Studying this exposure requires a quantitative model of the reliance of the country's access networks, in aggregate, on specific transit networks. The model must factor in the size of the address space originated by each AS with presence in the country. Intuitively, the greater the share of a country's IP addresses that are served by a particular transit AS, the higher the potential exposure of the nation's inbound traffic to observation or tampering by that AS. The model must then produce a country-level metric of exposure for each transit network serving the nation. To that end, we determine the frequency at which transit networks appear on routes towards the country's IP addresses.

At the AS level, the model considers only those IP addresses that belong to that specific AS, rather than those belonging to any AS in that country. In particular, this dissertation includes a study of critical organizations' international connectivity, such as that of power grid operators, government agencies and universities, which frequently rely on the Internet to send and receive content internationally. This reliance means that their traffic, which may be of especially sensitive nature, is potentially exposed to observation or tampering by domestic or foreign actors. Understanding the connectivity of these organizations, then, is likely of interest to governments and citizens alike.

This dissertation employs a two-stage approach to *(i)* identify transit-dominant countries, and *(ii)* quantify transit influence of the networks serving each country, as well as those serving critical organizations within them. We validate our findings from both stages with in-country network operators at 123 ASes in 19 countries, who confirm that our results are consistent with their understanding of their countries' networks. At the core of our approach there are four distinct metrics summarized in Table 1.1.

Table 1.1. Metrics at the core of this dissertation's approach.

Metric	Purpose
Country-Level Transit Influence (CTI)	Quantify the transit influence a particular network exerts on a nation's traffic
AS-Level Transit Influence (ATI)	Quantify the transit influence a particular network exerts on an access network's traffic
Weighted AS-Level Transit Influence (W-ATI)	Quantify the transit influence a particular network exerts on a set of networks within the same critical sector
Country-Level Transit Fraction $T(C)$	Quantify how frequently a transit provider-customer link is traversed when crossing the AS-level national boundary into a country

1.4.1 Country-Level Transit Influence

In order to reveal crucial, nation-level topological features of exposure to observation and tampering by specific ASes, we develop the country-level transit influence (CTI) metric. CTI quantifies the transit influence a particular network exerts on a nation's traffic. CTI is based on an analysis of a large compendium of BGP data [23, 84] and includes both topological and geographic filters aimed at extracting transit influence inferences from incomplete and biased data [45, 52, 74].

1.4.2 AS-Level Transit Influence

CTI allows us to quantify the influence of a transit network on a country's IP addresses, but not all IP addresses within a country are created equal. For instance, some IP addresses belong to apparel retailers, while others are owned by national defense contractors. If any given network has transit influence over the latter, the implications are more concerning from the standpoint of national security than those of a transit network serving a retailer. In order to begin addressing this conceptual gap in CTI, we develop the AS-Level Transit Influence (ATI). ATI quantifies how prevalent a transit AS is on observed paths towards IP addresses belonging to a critical organization, with higher prevalence signaling greater observation exposure. This goal has the prerequisite of identifying which access ASes belong to critical organizations in the first place, a task which we accomplish using manual classification and confirmation with network operators.

1.4.3 Weighted ATI (W-ATI)

In order to study critical sectors as a whole, rather than single organizations within them, we derive an additonal metric from ATI: Weighted AS-Level Transit Influence (W-ATI). W-ATI quantifies the transit influence of a specific network on an entire industrial sector composed of multiple ASes. W-ATI is an average of a transit network's ATI over all the origin ASes in a sector, weighted by the number of IP addresses originated by each AS.

1.4.4 Country-Level Transit Fraction

We focus on countries where international connectivity is dominated by transit inter-domain relationships ("transit dominance"), as they are easier to identify from public data sources. Our methodology relies on information self-reported by network operators as well as on traceroutes [27] we collect ourselves. In order to quantify transit dominance, we present the country-level transit fraction $T(C)$. $T(C)$ measures how frequently inbound traceroutes cross the AS-level national boundary into a country through a transit link.

CTI is particularly salient in transit-dominant countries, which often lack peering facilities such as Internet exchange points (IXPs) at which access networks might connect directly with networks of other nations. In these nations, transit networks—often a select few based in geographically distant nations [33, 51, 55, 88]—serve as the dominant form of connectivity to the global Internet. Moreover, the lack of domestic co-location facilities places these nations at further risk of exposure to observation and tampering, because popular content is generally hosted abroad [34, 47, 60, 77, 90].

1.5 Contributions

Our contributions include:

1. We confirm this dissertation's thesis using four new Internet cartography metrics:

   - The Country-level Transit Influence (CTI) metric captures the extent to which a given network influences a nation's inbound traffic. We find that, among 75 transit-dominant countries, many have topologies significantly exposing them to observation or tampering: in the median case, the most influential transit network manages traffic towards 34% of the nation's IP addresses.

   - The AS-level Transit Influence (ATI) metric, which captures the extent to which a given network influences an access network's inbound traffic, and the Weighted ATI (W-ATI), which captures the transit influence of a network on an industrial sector. We find that large telecommunication companies—including several foreign companies and state-owned providers—have a broad transit footprint on critical organizations in three countries.

   - We identified 75 transit-dominant countries using the Country-Level Transit Fraction $T(C)$, capturing how frequently provider-customer links are traversed when crossing the national boundary into a country. These nations have, in aggregate, $\approx$1 billion Internet users (26% of the world [7]).

2. In these 75 nations, by studying the influential transit networks inferred by CTI, we identify two classes of ASes that are frequently influential: those who operate submarine cables and companies owned by national governments.

3. We apply both ATI and W-ATI in three of these 75 countries, and present an understanding of the international connectivity of some critical organizations and sectors at the logical layer.

1.6 Ethical disclaimer

We acknowledge several ethical implications of this dissertation. Our mass survey of operators for validation purposes (Ch. 6) was classified as exempt by the UC San Diego Institutional Review Board[1]. Our reporting of available paths to repressive countries—for instance, those studied by Freedom House [54]—in Ch. 5 might trigger government intervention to remove such paths. Another potential issue is the identification of networks (and specific submarine cables) that would yield the most expansive observation/tampering capabilities in a country (and its critical organizations, Ch. 6), which is potentially useful information for a malicious actor. We believe most governments and sophisticated attackers already have access to this information and that our study may lead to mitigation of these concerning topological features; thus, we judge the benefits to significantly exceed the risks.

1.7 Organization

The rest of this dissertation is organized as follows. Chapter 2 includes an overview of the necessary technical background as well as the related work. Chapter 3 presents our identification of transit-dominant countries, including both our methodology and the findings from its application. Then, Chapter 4 defines our Country-Level Transit metric, whereas Chapter 5 includes the findings of applying CTI in 75 countries primarily served by transit networks. Chapter 6 presents our identification of the transit ASes serving critical organizations by applying our AS-Level Transit Influence metric. Finally, Chapter 7 presents our conclusions and potential future lines of research.

[1] UC San Diego IRB Project #202135XX Identifying the Organizational Purpose of Computer Networks.

Chapter 2

Background and Related Work

In this chapter we define the necessary technical background for this dissertation, and discuss the external datasets that we use for our analyses. We also review the related work in the relevant fields of study.

2.1 Fundamentals of Country-Level Routing

We begin by defining basic terminology and describing the foundations of wide-area network measurments.

2.1.1 Autonomous Systems and Country-Level Topology

AS topology is the study of how ASes connect to each other. For the purposes of this dissertation, we are particularly interested in finding which ASes are most relevant to the external connectivity of an entire country, or the country-level topology. These central nodes are the most likely to expose many other ASes and their traffic to traffic observation or selective tampering.

Conceptually, international Internet traffic crosses a nation's border at some physical location, likely along a link connecting two routers. For our purposes, we are not interested in the physical topology, but the logical one: in which autonomous system(s) does international traffic enter a nation on its way to access networks in that country (i.e., origin ASes). Topologically, these ASes can have two different types of relationship with the first domestic AS encountered: transit (provider-to-customer or *p2c*) or peering (peer-to-peer or *p2p*).

2.1.2 Border Gateway Protocol

An IPv4 BGP announcement[1] consists of two main components: the set of 32-bit IP addresses it refers to, which are aggregated into *prefixes*, or a block of IP addresses sharing the same initial (8 to 24) bits; and the AS *path*, which is the chain of networks, identified by their AS Number (ASN), through which traffic towards those IP addresses can flow. ASes along the path can be either the origin AS or a transit/peer AS. An origin AS is the last AS in the path, responsible for delivering the traffic to the destination IP address. There can be more than one path towards the same prefix, though the origin AS is generally the same for a given prefix.

To illustrate how BGP announcements work, say that `AS1` announces that it is the origin (or that it "originates") for the prefix $1.2.3.0/24$, and that this prefix can be reached through the following AS-level path: $AS3$-$AS2$-$AS1$. In this scenario, `AS3` would know it can forward any packets towards any IP address in the range $1.2.3.X$ (where X is a number between 1 and 255) to `AS2`. Then, `AS2` would forward them along to `AS1`. Finally `AS1`, the origin AS, would route these packets internally towards the destination machine in its network with the address $1.2.3.X$. A useful abstraction to derive understanding of the network as a whole, and gain the ability to identify ASes that may pose topological weaknesses given their appearance along many paths, is to consider the ASes themselves as nodes and the connections between them as links in a large AS network. In our toy example, the nodes would be `AS1`, `AS2` and `AS3`; and there would be two links: one between `AS2` and `AS3`, and another one between `AS1` and `AS2`. With this abstraction we can use graph theory to understand the most important nodes in the network.

2.1.3 BGP Monitors and Measurement Infrastructure Bias

Our study of country-level routing relies on a collection of real BGP announcements. A BGP *monitor* is an operational border router that forwards announcements to a collection database. These routers are hosted by cooperative ASes who voluntarily disclose their routing information.

[1] We limit our discussion to IPv4 announcements.

We do not have access to BGP announcements from all ASes. Trade secrets and business agreements are among the reasons why many ASes are limited in their ability to share routing information. The most comprehensive databases of BGP announcements are maintained by RouteViews (USA) and RIPE NCC (EU). While this information is extremely valuable for the purpose of understanding the AS topology, these databases rely on measurement infrastructure consisting of BGP monitors located inside cooperative ASes, a set that is not uniformly distributed around the world. Rather, it is concentrated on a small set of countries, and a tiny fraction of the world's ASes. While these monitors, on aggregate, have visibility towards most of the routed prefixes in the Internet, the set of *paths* they observe is biased towards those reaching the ASes hosting the monitor themselves. Further, the BGP monitors in the infrastructure might not see all the announcements of important domestic ASes in other countries, particularly those that do not have any monitors.

2.1.4 Geolocation

There is no straightforward method to determine which BGP announcements refer to prefixes used to connect machines physically present in any given country. The set of mechanisms to solve that mapping and assign IP addresses to a physical location is referred to as geolocation. Several commercial providers sell databases linking IP addresses to physical locations for a fee; these "geolocation databases often successfully geolocate IP addresses at the country-level" [82], which is the required geographic granularity for the purposes of finding topological weaknesses in each country.

AS registration information has also been used by researchers to determine the physical location of either the AS itself or of its blocks of IP addresses (*e.g.,* [88]). This was the main geolocation method available before the accuracy of commercial geolocation databases was rigorously characterized by [82]. IP addresses belonging to each AS are allocated by ICANN, the Internet Corporation for Assigned Names and Numbers, through its regional subdivisions, Regional Internet Registries (RIRs). "These registries maintain the authoritative lists of the autonomous system number and the IP address blocks associated with each autonomous system. They also keep the country in which each autonomous system was registered.” [88]

2.1.5 AS- and IP-Level Centrality

We study the exposure a country’s to observation and selective tampering by a specific network—the topic of Chapters 4 and 6—using an evaluation of the frequency at which a specific transit provider appears along paths towards an origin (destination) AS_o. Mathematically, for a transit AS_t serving inbound international traffic to any origin AS_o in country C from any third AS_n, we define AS_t’s centrality, ASC, as

$$ASC(AS_t, C) \in [0,1] = \frac{\sum_{AS_o \in C} S(AS_t, AS_o)}{\sum_{AS_n} S(AS_n, AS_o),} \tag{2.1}$$

where $S(AS_t, AS_o)$ is the number of paths starting outside C towards AS_o where AS_t is present as a transit provider, and $S(AS_n, AS_o)$ is the total number of inbound paths towards AS_o from transit AS AS_n. This definition of ASC is directional (paths going from outside the country towards the origin in the country) and excludes paths where no provider-customer relationship is present.

There are two major shortcomings of the model expressed in Eq. 2.1 for our purposes: first, it provides no mechanism to compute AS centrality on a country, *i.e.,* a collection of IP addresses originated by multiple ASes. Second, Eq. 2.1 treats paths towards each AS_o equally, regardless of how many IP addresses each of them originates. Since we are interested in measuring the exposure to observation and tampering of a country’s address space (*not* origin ASes), we instead quantify the IP-level centrality (IPC) of a transit AS AS_t along paths towards addresses originated in each nation C, which we define as

$$IPC(AS_t, C) \in [0,1] = \sum_{p \mid \text{onpath}(AS_t, p)} \frac{a(p, C)}{A(C)}, \tag{2.2}$$

where onpath(AS_t, p) is true if AS_t is present on a BGP path towards a prefix p, $a(p,C)$ is the number of addresses in prefix p geolocated to country C, and $A(C)$ is the total number of IP addresses geolocated to country C. Note that this is a departure from established models of AS centrality, such as Betweenness Centrality [2]. This is because the leaves of the graph are geographically-annotated nodes, or a single origin AS abstraction per country, representing the union of each of the origin ASes' addresses present in the country. Further, IP addresses determine the weight of the edges terminating in them, and the core of the graph being ASes who carry traffic towards the edge only. In practice, a single AS may serve as an origin AS and a transit AS in the same country, a case in which our model creates two abstractions for the same AS: one as a transit provider for other origin ASes, and the other as a component of the country's compendium of originated addresses.

In theory, IPC is a measure of the IP-level centrality of an *AS*, with the highest values of IPC for any set of transit ASes serving a country, providing an indication of how concentrated the exposure of a nation's inbound routing ecosystem is. A country where many transit ASes have similar values of IPC (for example, 10 ASes with $IPC = 0.1$) will likely be less exposed than another nation where an AS has high IPC (for example, a single AS with $IPC = 1.0$) and all other transit ASes are marginal ($IPC \leq 0.01$). Therefore, comparing the distribution of IPC values across countries gives us an indication of the relative exposure to observation and tampering of a country's inbound routing infrastructure compared to other nations.

2.2 Datasets and Filters

Our goal is to identify transit ASes that are influential at the country level. In order to do so, we first need to identify the set of transit ASes serving each country, the IP prefixes each AS originates, and the nationality of our measurement infrastructure. In this section we describe how we obtain this information.

2.2.1 BGP Data

We begin our analyses with the 848,242 IPv4 prefixes listed in CAIDA's Prefix-to-Autonomous System mappings derived from RouteViews [37], excluding the 6,861 (0.8%) prefixes with length greater than 24, and the 9,275 (1.1%) originated by multiple ASes. We find those prefixes in the 274,520,778 IPv4 AS-level paths observed in BGP table dumps gathered by AS-Rank [24] from RouteViews [23] and RIPE RIS [21] during the first five days of March 2020. We consider the set of prefixes and the ASes that originate them on each observed path in combination with the 377,879 inferred AS-level relationships published by CAIDA [19] for March of 2020.[2]

2.2.2 Definitions of Nationality

Our study hinges on carefully assigning nationality to IP address prefixes, BGP monitors and ASes (the latter solely to study transit dominance). We devise distinct methods for each. For our purposes, nationality is assigned to one of the 193 United Nations member states[3], either of its two permanent non-member observer states, or Antarctica.

Address prefixes

We first geolocate each IP address in every observed BGP prefix to a country using Netacuity [26]. (While geolocation databases are known to be unreliable at fine granularities, previous work has found them to be accurate at the country level [66, 82], with Netacuity in particular having accuracy between 74–98% [57].) Then, on a country-by-country basis, we count how many addresses in each prefix are geolocated to that country. If the number is less than 256 (a /24), we round up to 256. If Netacuity does not place any of a prefix's IP addresses in a

[2] In the 75 countries where we study transit influence, no path contained any of: unallocated ASes, loops, poisoned paths (where a non-clique AS is present between two clique ASes, clique being the AS-level core of the Internet inferred by [19]); additionally, all paths towards these countries are seen at least once per day across all five days.

[3] We assign subnational jurisdictions (such as Puerto Rico) to the sovereign nation-state of which they are part (the United States).

country, we attempt to find a delegation block from the March 2020 RIR delegation files [22] that covers the entirety of the prefix. If there is one we assign all of the delegated prefix's addresses to the indicated country. Hence, while Netacuity can place a prefix in multiple countries, at most one country will receive addresses through the RIR process, and only if it was not already associated with the prefix through Netacuity. We denote the number of IP addresses assigned to a country C (by either Netacuity or RIR) as $A(C) = \sum a(p,C)$, where $a(p,C)$ is the number of IP addresses in a given prefix p assigned to C. In this dissertation, Netacuity accounts for 95.1% of all prefix-to-country mappings, while delegation-derived geolocation accounts for the rest.

A particularly pressing concern with geolocation is the correct assignment of IP addresses belonging to large transit ASes with a presence in many countries. We compute the fraction of a country's address space that is originated by ASes that have at least two thirds of their addresses in that country. The output of this analysis is (25th perc.,median,mean,75th perc.) = 69%, 87%,77%,77%. We find, then, In the vast majority of countries, the address space is dominated by ASes that are primarily domestic.

Monitor locations

Our study is focused on measuring inbound country-level connectivity, so we limit our analysis to paths going towards addresses in the target country from a BGP monitor located outside that country. Hence, we confirm the BGP monitor locations listed by RouteViews [89] and RIPE RIS [87] through a set of active measurements.

We begin with the 685 monitors in RIPE and RouteViews. We discard (91) monitors aggregated at multi-hop collectors and monitors that are not full-feed, so we are left with 350 monitors in 209 ASes. We determine the location of each remaining BGP monitor as follows. First, we find the locations of RouteViews and RIPE RIS BGP collectors. We build a first set of locations by finding RIPE Atlas probes co-located at Internet Exchange Points (IXPs), by searching the list of peers for the IXP name, and assign that probe to the country where the (single-location) IXP is present, *e.g.,* BGP RRC01 – LINX / LONAP, London, United Kingdom.

We confirm the BGP monitor location by running ping measurements from RIPE Atlas probes hosted at an IXP to the BGP monitor's IP address and conclude that the BGP monitor is in the same city as the IXP if the RTT is lower than 5 ms. For the remaining BGP monitors we look for available RIPE Atlas probes in the ASes that peer with the same BGP collector and similarly run ping measurements towards both the BGP monitor's IP address and a RIPE Atlas probe located in the same city as the one listed for the monitor. We conclude that the BGP monitor and RIPE Atlas probe are in the same city if both sets of RTTs are under 5 ms.

We exclude 118 monitors at this stage because there is no available RIPE Atlas probe hosted at the IXP (in the city where the monitor is listed) nor at any of the other peers of the collector aggregating announcements from the BGP monitor. We discard remote peers from our set, *i.e.,* those that have ping RTTs higher than 30 ms from the RIPE Atlas probe in the BGP monitor's listed city. For monitors with an RTT between 5–30 ms, we infer them to be at the listed location if we get confirmation using DNS records—*i.e.,* we find a geographical hint such as a three-letter city or airport code, or the full name of the city, using a reverse lookup with the BGP monitor's IP address—or a matching country of the BGP monitor's peer_asn record in the RIPE RIS or RouteViews collector list [87, 89]. Our final set of located monitors is M. M consists of 214 monitors in 145 ASes and 19 countries.

Autonomous Systems

Our definition of AS nationality (which is only relevant for our identification of transit-dominant countries, Ch. 3) is built on the intuition that an AS will use its IP addresses in countries where it operates, and that the country with most of its addresses will therefore be the primary country of operation. For transit providers, we include the IP addresses originated by direct customers, as they are part of their transit footprint, or to the set of addresses in a given country that the transit AS serves.

By contrast, we exclude indirect customers (*e.g.,* customers of customers) as these do not have a direct relationship with the transit provider, and as a consequence it is possible the two ASes have a peering relationship we do not observe—in which case the indirect customers' addresses would not be part of the transit provider's transit footprint.

In practice, we classify each autonomous system *AS* operating in a country *C* as being *domestic*, $AS \in \mathrm{dom}(C)$, when the AS has at least two thirds of its addresses in the country (according to Netacuity [26]), and *foreign* otherwise; this last group includes ASes that are primarily international in nature, which are classified as foreign to every country. The vast majority (97.4%) of ASes are classified as domestic in one country, with the remaining small fraction being classified as foreign in every country. In fact, 89.8% of ASes have all of their address in a single country, and 98.6% have a strict majority of addresses in one country.

2.3 Related Work

The most relevant body of related work comes from the study of country-level routing. Several previous studies have focused on country-level routing, both for the identification of topological bottlenecks [68, 88] and to evaluate the impact of specific countries' ASes on routes towards other countries [65]. All of these studies have used RIR delegation data to map an entire AS to a country; these inferences are prone to inaccuracies when compared with more accurate and granular data such as IP-level geolocation, as important transit ASes may span multiple or many countries, or operate in a country different from their registration. Most recently, Leyba *et al.* [68] addressed the identification of topological bottlenecks, a framework that would also help in quantifying exposure to observation (as CTI aims to address), but with some methodological differences, including: they identify transnational links towards each country using delegation records, and they define bottleneck ASes as those serving the most paths (rather than IP addresses).

Further, both CTI and Leyba *et al.* [68] have as a goal the identification of international inbound—and, in their case, also outbound—*chokepoints* (*i.e.,* topological bottlenecks) in each country, based on simulations of plausible BGP paths towards each origin AS. However, their work does not try to capture the fraction of the country's addresses served by a transit provider, but rather the fraction of paths that a border AS (*i.e.,* an AS which is registered to the same country as the origin, but which has a neighbor that is registered to another country) may be able to intercept. Our work is more narrowly focused on the specific case of a transit provider serving traffic towards a transit-dominant country, taking into account the address space of the direct or indirect customers. At a high level, weighting by paths enhances the influence—or potential, in Leyba *et al.*'s terminology—of ASes frequently serving a broad share of the country's networks, whereas weighting by IPs yields higher influence to ASes frequently serving a large fraction of the country's end hosts.

Other previous work focused on the topologies of specific countries (Germany [102] and China [105]) and relied on country-specific methods and data sets that do not generalize to automatic inference of AS influence in any given country. Fanou *et al.* [51] studied the interdomain connectivity of intracontinental paths in Africa, using a large traceroute campaign (rather than BGP paths).

Country-level routing has also been studied in developing nations, where transit ASes—often a select few based in geographically distant nations [33, 51, 55, 88]—serve as the dominant form of connectivity to the global Internet. Moreover, the lack of domestic co-location facilities places these nations at further risk of exposure to observation and tampering because popular content is generally hosted abroad [34, 47, 60, 77, 90].

2.3.1 AS and Country Centrality

Hegemony [52] aims to identify the transit ASes that are most prevalent on paths towards origin ASes, weighted by the IP address space they serve, taking into account the limited topological scope of the BGP measurement infrastructure. Hegemony is derived from Betweenness Centrality [2], which "captures how much a given node is in-between others." Hegemony can be applied either to the global AS-level graph, or to a "Local graph: ... made only from AS paths with the same origin AS" [52]. The latter application is closest to CTI (and ATI), as this analysis is limited to paths reaching a single origin AS. Indeed, CTI uses some of Hegemony local's filtering techniques in our analysis (see 4.2.4). CTI is a better fit for our purposes than Hegemony, both methodologically and conceptually, due to three unique features. First, Hegemony cannot find ASes that serve traffic towards most of a country with multiple ASes operating in it. Second, unlike Hegemony, CTI is designed to study the international routes reaching ASes in a particular country, and therefore can handle cases where an origin AS has its address space split across countries. Third, CTI is designed to take into account the extremely limited visibility of peering links (p2p) [45], and focuses on transit links (in a set of transit-dominant countries).

Edmundson et al. [47] have also looked into the transit prevalence of specific countries (as opposed to ASes) when accessing popular content from five nations. They find that, despite efforts by regulators in some countries, the U.S. is still frequently transited when accessing top domains.

2.3.2 AS Classification

Dhamdhere et al. [45] use decision trees to classify ASes—based on their peering and customer degree—into enterprise customers, small transit providers, large transit providers, and content/hosting/access providers[4]. The categories they create are based on their role in the transit ecosystem: enterprise customers, small transit providers, large transit providers, and content/hosting/access providers. This classification allows the authors to study the evolution of each group independently. For instance, they find that enterprise customers increasingly prefer to connect to smaller transit providers. They have a similar goal as our study, then, in characterizing the connectivity of categories of ASes. A similar study by Dimitropoulos et al. [46] also apply machine learning to the problem of classifying ASes in the global topology. It relies on data published by Internet Routing Registries (for RIR AS description strings) and routing tables from Routeviews [23] for a variety of network metrics, including size of address space originated.

[4] "Degree" refers to the number of neighbors of an AS; "peering degree" and "customer degree", conversely, refer to the number of neighbors of an AS that are either its peers or its customers, respectively.

Wahlisch et al. [102] present a classification of the economic purpose of an AS by extending a taxonomy created by the German Government, and using partially-verified assignment using keyword search on AS names, descriptions, and address fields. The verification is by manual inspection. This mechanism allows them to go beyond AS topology and compare real sectors of the German economy. Our analysis of critical-sector ASes does not rely on pre-existing taxonomies of any given country, but rather on manual classification of all ASes in the country coupled with partial operator validation.

Chapter 2, in part, is currently being prepared for submission for publication of the material. Gamero-Garrido, Alexander; Carisimo, Esteban; Hao, Shuai; Huffaker, Bradley; Snoeren, Alex C.; Dainotti, Alberto. The dissertation author was the primary investigator and author of this paper.

Chapter 3

Quantifying the Prevalence of Inbound Transit

This chapter presents our identification of countries where provider-customer (p2c) relationships are likely the dominant mode of inbound international connectivity. Our tools look for nations where origin ASes—*e.g.,* consumer-serving access networks—do not tend to have foreign peers, nor the ability to establish peering agreements with foreign counterparts due to lack of physical access to facilities where such interconnections happen (*e.g.,* Internet Exchange Points or IXPs based in other countries). This lack of peering infrastructure suggests that transit providers are frequently traversed by traffic entering the country. We confirm this hypothesis using our own large-scale measurement campaign and discussions with in-country network operators.

Countries that are transit dominant may be more exposed to observation and tampering by transit providers than that of countries where peering agreements are prevalent: the latter can receive some traffic from other countries through such peering agreements and bypass transit providers. We define foreign peering as a (logical) link between two ASes that: *(i)* include an AS that is domestic to the country and a non-domestic AS (Sec. 2.2.2), and *(ii)* where that link is not an inferred provider-customer link.

3.1 Approach

We take a multi-step approach to identifiy countries where provider-customer links are an important inbound modality. Since peer-to-peer and provider-customer links are conceptually[1] disjoint, meaning that they cannot both exist at the same time for the same pair of ASes, we first find a set of candidate countries to study based on the *observed absence* of foreign p2p links (Sec. 3.2). That absence suggests that transit providers are serving an important fraction of inbound traffic; we later confirm this hypothesis using active measurements (Sec. 3.2) and operator validation (Sec. 3.3).

The first three boxes in Fig. 3.1 summarize our approach to identifying the observed fraction of a country's address space with neither *existing* nor *potential* foreign peering. We start by identifying countries for which public datasets of Internet Exchange Points (IXPs) and Private Colocation facilities (Colo) show no evidence of foreign peering (Sec. 3.2). Ethiopia is the candidate country in the figure, and the United Kingdom is an example of a country where a potential or observed foreign peer might be present. Potential peering refers to the ability of an origin AS to acquire a foreign peer, *e.g.,* by virtue of being a member of an IXP where another member is foreign (even if no link between them is actually observed).

Based on the results of this analysis of public datasets (Sec. 3.2), we conduct an active measurement campaign to confirm the absence of foreign peering (rightmost box in Fig. 3.1). This second stage based on traceroutes is necessary because peering datasets are incomplete, particularly when it comes to membership lists at IXPs in developing countries [72]. We consider the prevalence of transit links being used to reach each of our target countries from probes distributed worldwide (Sec. 3.2.2) to select a set of transit-dominant countries.

[1] A small fraction of AS links do not fall into either category, reflecting more complex relationships between interconnecting ASes [59].

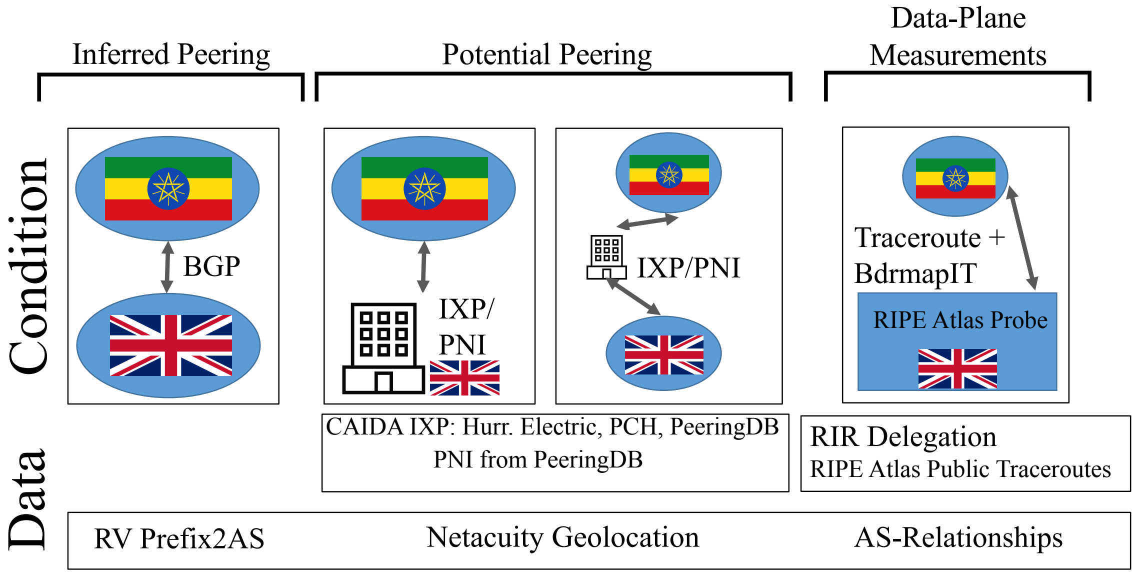

Figure 3.1. Tests and datasets involved in inference of transit-dominant candidate countries (three leftmost boxes), and active campaign to confirm candidates (rightmost box). The bottom data row includes inputs used by all conditions, whereas the top data row is only used by the conditions directly above them.

3.2 Identifying Transit-Dominant Countries

We begin with the set of ASes that originate addresses in each country. This set includes origin ASes that we classify as foreign to that country, but that originate BGP prefixes entirely geolocated in the country. We look for these origin ASes in CAIDA's IXP dataset[2] (from October 2019 [35]), PeeringDB Colo dataset (from March 1st, 2020 [38]), and inferred AS-Relationships from BGP (March 2020 [19]). A "Colo" is an industry term for typically-private facilities where AS interconnections occur[3].

[2] "This dataset provides information about Internet eXchange Points (IXPs) and their geographic locations, facilities, prefixes, and member ASes. It is derived by combining information from PeeringDB, Hurricane Electric, Packet Clearing House (PCH), and GeoNames." [35]

[3] Other services such as cloud computing are also often on offer at these facilities.

We classify an origin AS as a *candidate* if the following three conditions are all true:

1. the origin AS has no foreign or international peers in BGP [19];
2. the origin AS is not a member of any IXPs or Colos based in another country [35, 38]; and
3. the origin AS is not a member of any IXPs or Colos where any member AS is based in a different country than the origin AS [35, 38].

The first test is related to directly observed peering, while the latter two tests are related to potential peering: if an AS is a member of an IXP/Colo in another country (2), or a member of an IXP/Colo where another member is from a different country (3), the origin AS is at least capable of establishing peering relationships with those other ASes.

Fig. 3.2 shows the percentage of a country's address space originated by candidate ASes. We select the top-100 countries as candidates for active measurements; each is a country where at least 25% of addresses are originated by candidate ASes. These 100 candidate countries are colored in Fig. 3.3.

3.2.1 Active Measurement Campaign

In order to confirm the prevalence of provider-customer links on inbound routes towards candidate countries, we target them with an active traceroute campaign. By analyzing the output of the campaign, we quantify the frequency at which provider-customer links are traversed when entering each candidate country. To that end, we ran a traceroute campaign to the 100 candidate countries for 14 days starting on May 2nd, 2020. Additionally, we use all publicly available IPv4 traceroutes on RIPE Atlas during the same period—on the order of several million per hour—in order to take advantage of other measurements towards the same ASes.

We design our traceroute campaign guided by two principles. First, we want to select a geographically and topologically diverse set of probes. Second, we have to operate within the rate limits of RIPE Atlas[4] [85], particularly regarding concurrent measurements and credit

[4] Which RIPE Atlas generously relaxed for this study upon direct request.

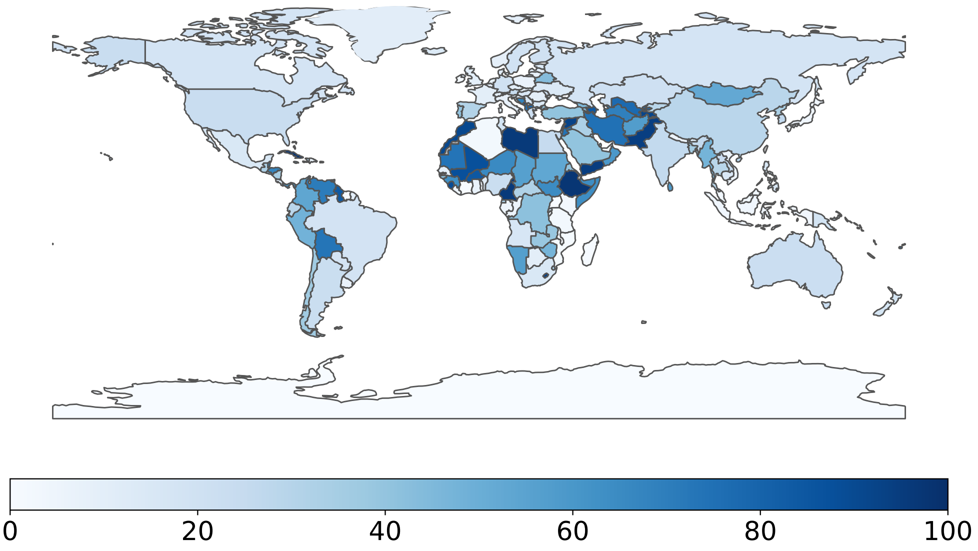

Figure 3.2. Non-peering observed fraction on passive datasets.

expenditure. Therefore, we remove marginal ASes that originate a very small fraction of the country's address space (less than 0.05%).

Within these constraints, we launch ICMP traceroutes[5] from 100 active—shown as "connected" during the previous day [86]—RIPE Atlas probes towards a single destination in each AS, twice daily[6]. Probing at this frequency gives us 28 opportunities to reach an AS during the two-week period from each RIPE Atlas probe. Since we are studying inbound international connectivity, we make sure no probe is located in any candidate country, *i.e.,* our traceroutes start in countries outside the target set. This filter excludes 8.5% of active probes).

We target an IP address for each origin AS in each candidate country by looking for any prefix originated by that AS that is entirely geolocated or delegated within the candidate country (see Sec. 6.1). Our final dataset is comprised of 33,045,982 traceroutes, including those launched

[5] Using all default RIPE Atlas values save for number of packets to send, which we reduce to 1 to stay within our measurement credit budget.

[6] Because of limits on user-defined measurements [85] we space traceroutes an hour apart in 800-target IP blocks, which also allows time for the RIPE Atlas server to process and execute each request.

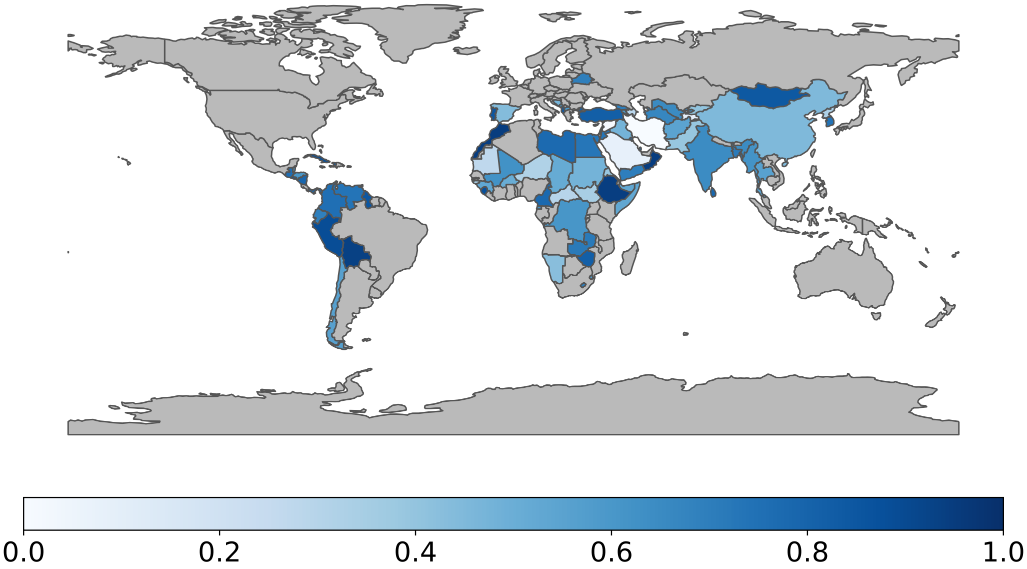

Figure 3.3. Scaled country-level transit fraction $T(C)$ in candidate countries.

by other RIPE users that meet our constraints. The distribution of the number of traceroutes reaching each country has the following properties: (Minimum, 25th Percentile, Median, Mean, 75th Percentile, Max) = (36, 13k, 46k, 330k, 250k, 3.3m). That is, the median country received 46k traceroutes. Only three countries received fewer than a thousand traceroutes: Eritrea (667), Nauru (154), and Tuvalu (36).

We use BdrmapIT [76] to translate our traceroutes into AS-level interconnections. This tool requires a number of external datasets in its operation, which we specify as follows: inferred AS-Level customer cone [74] from March 2020; *AS2Org*, which infers groups of ASes who belong to the same organization[7] from January 2020; and datasets we mention in other sections—prefix-to-Autonomous System mappings (Sec. 6.1), *PeeringDB* records (Sec. 3.2), and RIR delegation records (Sec. 2.2.2). From traceroutes and external datasets, BdrmapIT infers AS-level interconnections and the IP addresses (interfaces) at which they occur. Each interface inferred by BdrmapIT has an AS "owner" assignment. We reconstruct the AS-level path observed on the

[7] This dataset is published quarterly.

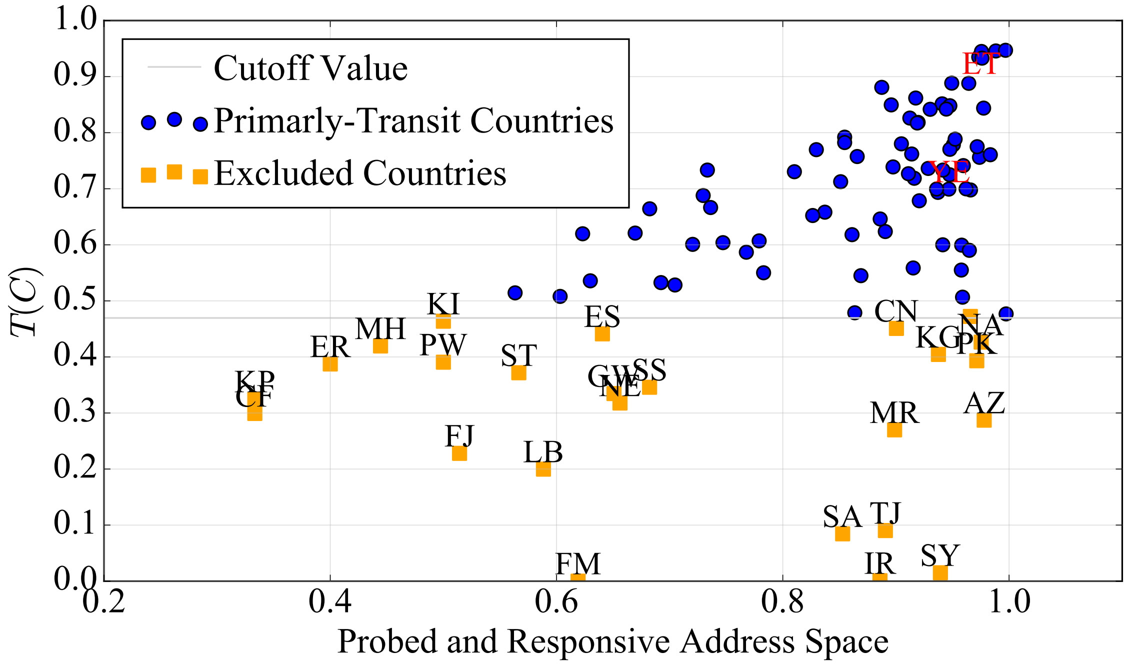

Figure 3.4. Country-level transit fractions $T(C)$ for countries in our sample.

traceroute using such assingments. Then, in order to determine if two consecutive ASes have a provider-customer relationship, we use CAIDA's (April 2020) AS relationship inferences [19].

3.2.2 Country-Level Transit Fraction

We have so far built a set of AS-level paths taken from the traceroute source to the destination AS. We now need a quantitative analysis technique to infer the prevalence of transit links on inbound traces towards each country.

We calculate how frequently, in the inbound traceroutes we process with BdrmapIT, the AS-level national border crossing occurs on a provider-customer link for each origin AS. A border crossing occurs when we observe a foreign or international AS followed by a domestic AS, with the latter closest to the origin (or itself the origin); this nationality assignment is described in Sec. 2.2.2. We scale this fraction to take into account the size of the address space originated by each AS using the *country-level transit fraction*:

$$T(C) = \sum_{AS_o, AS_c \in \mathrm{dom}(C)} \sum_{AS_t \notin \mathrm{dom(C)}} \frac{R(AS_o, AS_t, AS_c)}{R(AS_o)} \cdot \frac{a^*(AS_o, C)}{A(C)}, \tag{3.1}$$

where $R(AS_o, AS_t, AS_c)$ is the number of traceroutes destined toward a prefix originated by AS_o that traverse a transit link between a foreign provider AS_t and a domestic customer AS_c in country C; $R(AS_o)$ is the total number of traceroutes where AS_o is the last observed AS; and $a^*(AS_o, C)/A(C)$ is the fraction of country C's address space originated by AS_o. For instance, if an AS originates 50% of the country's origin addresses, and 50% of the traces towards it traverse a foreign transit provider AS, the contribution of that AS to the country-level transit fraction becomes 0.25. Note that AS_c and AS_o are not necessarily the same, as the border crossing may occur at the link between (direct and/or indirect) providers of AS_o. The values of $T(C)$ for each candidate country are represented in Fig. 3.3: countries in darker shades of blue have both a large probed and responsive fraction and a large fraction of traceroutes from outside the country traversing transit providers. The closer the fraction is to 1, the more evidence we have that the country relies on transit providers for its international inbound connectivity.

3.2.3 Final Selection

Finally, in order to identify a set of transit-dominant countries, we evaluate the values of $T(C)$ across countries, shown in Fig. 3.4. At one extreme of Fig. 3.4 and Fig. 3.3 are countries such as Ethiopia (ET) and Yemen (YE), $T(C)$ = 0.95 and 0.70, respectively, where all available evidence points towards transit links as the main inbound modality. At the other extreme are countries such as Syria (SY) and Iran (IR), $T(C) \leq 0.01$, where we rarely observe AS-level national borders being crossed using transit links.

Outside the upper and lower extremes in Fig. 3.4, where the decision of whether to include a country in our study is obvious, the middle results (most countries) do not offer clear

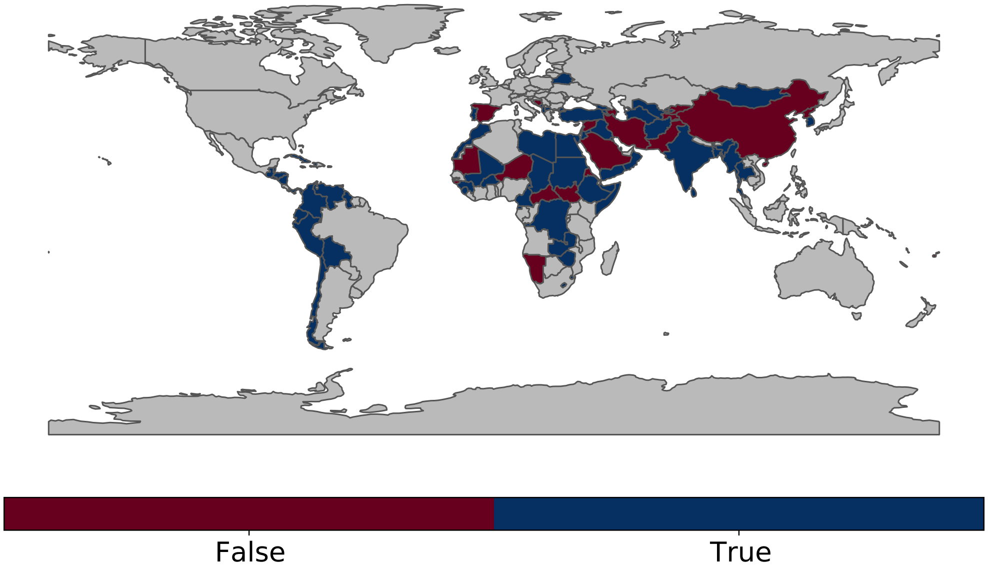

Figure 3.5. Final set of transit-dominant countries (nations in red excluded).

dividing points. We set the threshold for $T(C)$ to classify a country as transit-dominant based on our validation with operators (presented later in this chapter); in particular, we use the value of $T(C)$ for Sudan (0.48) as a lower bound, which is the lowest $T(C)$ in any country that we were able to confirm relies on transit links for its inbound connectivity.

The final transit-dominant countries we identify are colored blue in Fig. 3.5, and also shown as blue circles in Fig. 3.4, 75 of the 100 candidates. Darker shades of blue signal higher values of $T(C)$. (Countries in red are excluded from further analysis, as at this time we lack sufficient evidence to support that they are primarily using transit providers for inbound connectivity.) These results confirm part of this dissertation's thesis, as they show that a large number of countries are dominated by transit providers in their external connectivity.

3.2.4 Limitations

Any active campaign launched using publicly available infrastructure will be limited in its effectiveness to reveal peering links by the location of vantage points (VPs) from which the traceroutes are launched. Our campaign is no exception: our VPs are located in a small subset of the world's ASes, and primarily in Europe and North America. However, we argue that our measurements form a sufficient basis to infer that, in the countries we have identified, provider-customer links are an important inbound modality. While our measurements are launched primarily from the U.S. and Europe because that is where the majority of RIPE Atlas probes are located, these regions do serve as important content sources and transit hubs (including for intracontinental traffic) for countries in Latin America, the Caribbean and Africa [33, 55, 67, 56, 50], where most of the nations we identify are located. Another important limitation comes from our use of BdrmapIT [76], the tool we use to translate traceroutes to AS-level paths. BdrmapIT may produce between 1-8% inaccurate inferences [76], which may impact our assessment of transit dominance.

3.3 Operator Validation

We discussed our findings with employees or contractors of two types of organizations: commercial network operators and non-profits who conduct networking research (universities, registrars, and non-commercial network operators). Discussions with all of these organizations are anonymized to protect their privacy. These discussions took place in the spring of 2020, unless otherwise specified. Our findings are largely consistent with each operator's view of the transit ecosystem of the countries discussed with them: we did not find any false positives in our identification of transit-dominant countries.

Commercial Network Operators. We emailed a former and eight current employees at nine companies operating transit and/or access networks primarily in Africa and Latin America. Eight of them responded, but two declined to participate in our discussions, and another two declined to comment on country-level prevalence of peering. The four remaining operators confirmed that three countries (two in Africa and one[8] in Latin America) are transit dominant.

Networking Researchers at Non-Profits. We contacted sixteen researchers in ten countries in Africa and Latin America. In five of those countries (four in Africa, one in Latin

[8] This conversation took place in the fall of 2020.

America), six researchers confirmed the countries are transit dominant. Two other researchers declined to comment altogether, and eight did not respond.

Chapter 3, in part, is currently being prepared for submission for publication of the material. Gamero-Garrido, Alexander; Carisimo, Esteban; Hao, Shuai; Huffaker, Bradley; Snoeren, Alex C.; Dainotti, Alberto. The dissertation author was the primary investigator and author of this paper.

Chapter 4

Country-Level Transit Influence

We develop a metric to quantify the influence a transit AS exerts on a given country, which we term *Country-level Transit Influence (CTI)*. Intuitively, CTI captures the fraction of the addresses originated in the country that are served by that AS for transit. We begin by describing the problem we aim to solve, then define the metric and describe in detail the heuristic filters we apply in its use. CTI ranges over $[0,1]$ and captures the extent to which a given AS influences a nation's (inbound) international transit service. In later chapters, we use CTI to study various features of country-level topology—in particular, how networks connect to one another in order to carry international traffic towards a country—to quantify how exposed a nation's inbound traffic might be to observation or selective tampering by a particular AS.

There are considerable terchnical challenges with this approach stemming from the limited visibility of AS interconnection in public datasets on which we rely. We describe the most important such limitations in Sec. 4.3. Despite the challenges, our approach is lent credibility by our validation with network operators (Sec. 5.4).

4.1 Approach

We start our model by building a graph where nodes are ASes and edges are connections between them, weighted by address space originated by the leaf (origin) AS on the path. Then, a metric of node prominence on said graph provides a quantitative assessment of how frequently a (transit) node AS_t is traversed when delivering traffic from any given node to edge (origin) nodes. The higher the value of this metric for any AS_t in a given country, the more exposed the transit ecosystem is. At one extreme (most exposed) are countries with a single transit provider (*e.g.,* a legally-mandated monopoly) connecting every network in the country to the rest of the Internet; at the other end are countries with many transit providers, each delivering traffic to a small fraction of the nation's IPs.

4.1.1 Reachability

While BGP dumps reveal potential paths toward destination ASes, they may not reflect the actual routes packets traverse, both because ASes may have alternative routes they do not export (e.g., based on peering relationships) and because the destination network may not, in fact, be reachable. Hence, we conduct a two-week-long active measurement campaign (see Ch. 3) in May 2020 to determine which ASes in a country are actually reachable, and the set of ASes traversed by our probe packets (as inferred by BdrmapIT [76]).

Our CTI metric combines BGP data with reachability information for each origin AS included in our large-scale `traceroute` campaign. If at least one of our traceroutes received a response, the origin AS is included in the potentially exposed set of origin ASes. All origin ASes' addresses (including unresponsive networks) in a country are factored in the calculation of the country's total address space. This method yields a conservative estimate of the capabilities for observation and tampering of any given transit AS (an exposure lower bound). In the countries we study using CTI, addresses originated by responsive ASes represent, in aggregate, a median of 88% and an average of 92% of the country's addresses; that figure is over 60% in all but one country (Chad, 56%).

4.1.2 Indirect Transit

As the number of AS-level hops from the origin increases, so too does the likelihood that there exist alternative paths towards the same origin AS of which we have no visibility (*e.g.,* backup links, less-preferred paths). Fig. 4.1 shows this limitation in visibility for a toy example with a single origin AS. There, given the location of BGP monitor *C* we see the AS-level chain in black, erroneously concluding that the origin AS has a single direct transit provider and two indirect transit providers. In reality, there exists another set of both direct and indirect transit providers (the AS-level chain in light gray). We miss all these paths given that we do not have a monitor in any neighbor of a light-gray AS (such as that marked with a plus sign). In this example we miss backup links of the origin AS, as well as preferred links of the origin's direct transit provider, and a backup link of both indirect transit providers.

As a coarse mechanism aimed at mitigating this limited visibility, we discount the influence of transit providers in proportion to the AS-level distance from the origin: we apply a discount factor as $1/1, 1/2, ..., 1/k$, where k is the number of AS-level hops from the origin AS. In practice, that means we do not discount the measurements of direct transit providers. We note that this heuristic yields a conservative estimate of the observation opportunities of an indirect transit provider over traffic flowing towards a country.

4.2 Transit Influence

Formally, the transit influence of autonomous system *AS* on country *C* is the fraction of *C*'s address space for which *AS* is present on announced, preferred paths toward prefixes originated in *C* by a responsive AS that we are able to observe from our set of monitors. The transit influence $CTI_M(AS,C) \in [0,1]$ is calculated using a set of monitors M as

$$\sum_{m \in M} \left(\frac{w(m)}{|M|} \cdot \sum_{p|\text{onpath}(AS,m,p)} \left(\frac{a(p,C)}{A(C)} \cdot \frac{1}{d(AS,m,p)} \right) \right), \tag{4.1}$$

where $w(m)$ is monitor m's weight (Sec. 4.2.1) among the set of monitors (Sec. 4.2.4); $\text{onpath}(AS,m,p)$ is true if *AS* is present on a preferred path observed by monitor m to a prefix p originated by a probed and responsive origin network, and m is not contained within *AS* itself (Sec. 4.2.4); $a(p,C)$ is the number of addresses in prefix p geolocated to country C that are

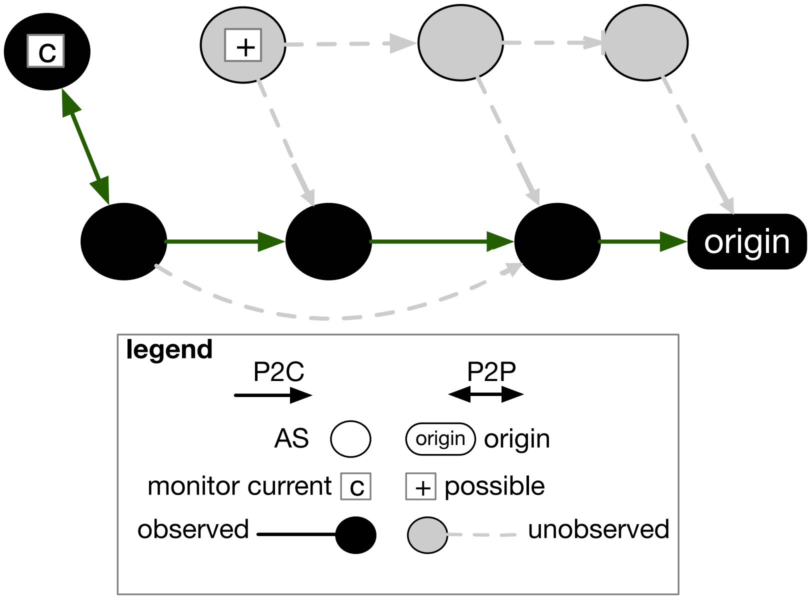

Figure 4.1. Unobserved paths in BGP.

not referenced by a more specific prefix (Sec. 4.2.2); $A(C)$ is the total number of IP addresses geolocated to country C; and $d(AS, p, m)$ is the number of AS-level hops between AS and prefix p as viewed by monitor m (Sec. 4.1.2).

Eq. 4.1 only considers prefixes originated by networks that we probe and are responsive, yet divides by $A(C)$ (*i.e.,* all addresses originated from the country). As a result, the actual CTI of AS might be higher if other origin networks that we do not probe are also reached through AS. Moreover, because CTI is computed with respect to the entire country regardless of the amount of probing, it is possible to compare CTIs across countries. Note that originating addresses directly does not grant an AS transit influence, as our focus is on identifying ASes that carry traffic towards destinations outside of their network.

Fig. 4.2 shows CTI values for a toy example with three transit ASes and four origin ASes, in a country with eight /24 prefixes: the transit AS on the right (TAS_3) has the highest CTI, since it serves the most addresses (half of the country), followed by TAS_1 $(3/8)$ and TAS_2 $(1/8)$. Note

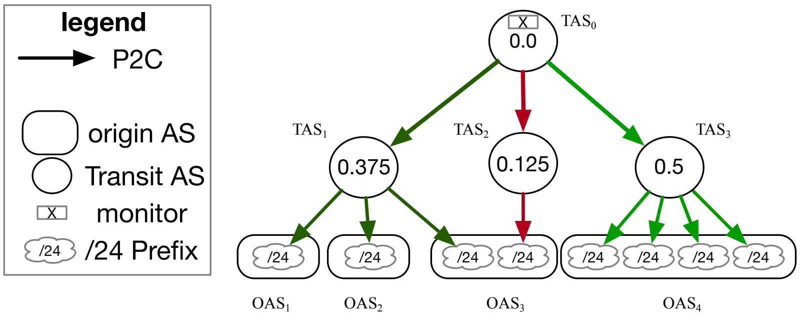

Figure 4.2. Example of Country-Level Transit Influence.

that TAS_0 has a CTI of 0, because it hosts the BGP monitor from which the set of routes used in this toy example are learned—hence, onpath(AS_t, m, p) is always false for that AS. Should that AS not be the host of the BGP monitor (or be seen on these routes through another monitor), it would have a CTI of 0.5—transit influence over the entire country as an indirect transit provider (distance 2 from the prefixes). We explain the rationale for the various factors in Eq. 6.2.2 in the following subsections.

4.2.1 Prioritizing AS diversity

ASes can host more than one BGP monitor. In fact, more than 20 ASes in RIPE RIS and RouteViews host multiple monitors; for instance, AS3257-GTT hosts five. In order to favor a topologically-diverse view (given the available observations), if more than one monitor from the same AS sees an announcement for the same prefix, we discount their observations to limit the influence of monitor ASes with multiple monitors. Formally, the weight for each monitor m's observation of a prefix is $w(m) = 1/n$, where n is the number of BGP monitors in a single AS that see an announcement of that prefix.

Table 4.1. Toy example of prefixes collected by an AS hosting two BGP monitors, with preferred and less preferred paths towards prefixes of different length in country C.

Monitor	Prefix	Length	Path
X	1.0.0.0	/8	A D
X	1.1.0.0	/16	J E D
Y	2.0.0.0	/8	B D
Y	2.1.1.0	/24	G D

4.2.2 Overlapping prefixes

Consistent with BGP's longest-prefix matching, when prefixes overlap, we compute the number of a country's addresses influenced by a transit AS by counting more-specific announcements first. We subtract the addresses of the more-specific prefix from the less specific. Our procedure ensures that we build a non-overlapping count of influence by preventing multiple transit providers announcing prefixes covering the same address space at different granularities from being over- or under-counted.

4.2.3 Example of calculating CTI

A toy example helps illustrate the transit influence calculation. Consider the BGP announcements described in Table 4.1, and graphically represented in Fig. 4.3. AS D, the only origin AS in country C in this example, originates two /8 blocks, or $2 \cdot 2^{24}$ IP addresses. ASes A, B, E and G are direct transit providers of AS D. AS J is an indirect transit provider of AS D. ASes A, J and E carry traffic towards either a subset or all of prefix 1.0.0.0/8. With regards to the prefixes observed by monitor X, we begin by computing the transit influence of the ASes appearing along the path towards the more specific prefix, in this case 1.1.0.0/16. AS E is a direct provider of the origin AS, and therefore we do not apply a discount factor based on its position along the path. In this example, since there are no two monitors seeing an announcement for a single prefix, $w(X) = w(Y) = 1.0$. Since AS E provides transit for 2^{16} of the $2 * 2^{24}$ IP addresses originated by AS D, the resulting influence metric $CTI(E,C) = 2^{16}/(2*2^{24}) = 0.00195$. Further, as AS J

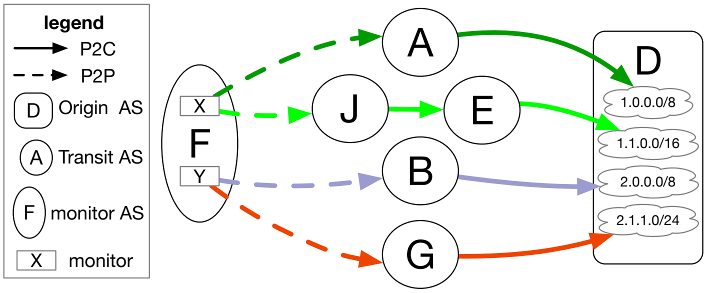

Figure 4.3. Toy example of overlapping address spaces originated by AS D (the only AS in country C) seen by two BGP monitors in the same AS.

provides transit for the same fraction of AS D's addresses, but it is an indirect provider with a position discount of 2, we compute the resulting transit influence as $CTI(J,C) = CTI(E,C)/2 = 0.00977$.

To compute the transit influence of AS A, we subtract the addresses of the more specific prefix (1.1.0.0/16) from the /8 prefix transited by AS A: $CTI(A,C) = (2^{24} - 2^{16})/(2 * 2^{24}) = 0.49805$. Using similar arithmetic on the prefixes seen by monitor Y we have that the CTI for the remaining transit providers is as follows: $CTI(G,C) = 2^{8}/(2 * 2^{24}) = 0.00001$; $CTI(B,C) = (2^{24} - 2^{8})/(2 * 2^{24}) = 0.49999$. Intuitively, AS B appears along preferred paths to slightly less than half of AS D's (and therefore country C's) addresses.

4.2.4 Filtering ASes

To correct for the limited, non-uniform coverage of the BGP monitors that collect our table dumps, we apply a number of filters to the set of paths over which we compute CTI.

Provider-customer AS filter

BGP monitors by definition collect paths from the AS hosting the monitor to the origin AS. Therefore, we always exclude the AS hosting the BGP monitor from the path to avoid inflating their transit influence. In general, as the number of AS-level hops from the origin AS increases, so does the likelihood that we are seeing ASes upon which the AS hosting the BGP monitor—as opposed to the origin—relies for connectivity. Hence, we employ a heuristic that attempts to consider only the portion of the path relevant to the origin prefix, and ignore the portion dictated by the monitor's topological location.

The intuition behind our filter is that, from the perspective of the origin AS, there is a "hill" above it capped by the last observed provider-customer (p2c, *i.e.,* transit) link, with traffic flowing from the hill's peak down towards the origin. The transit AS in that link is the highest point in the path we want to keep, as it directs traffic towards its customer (and its customer's customers, if applicable). After reaching that topological peak, we discard any other AS present in the path. The remaining path would then include the origin AS, its direct or indirect transit provider at the topological peak, and any other ASes appearing between the origin AS and the direct or indirect transit provider.

Formally, for the analysis presented in this dissertation, we refine onpath(AS_t, m, p) to be true only if the path observed at monitor m has at least one inferred p2c link where the customer is either the origin of p or closer to it than AS_t, *i.e.,* we discard paths where there is no topological peak from the perspective of the origin. This heuristic discards 4.3% of the paths observed by our monitors. In the median country we discard 3.4% of paths using this filter, with 6.0% being the average case. In all countries we keep over 78% of paths.

We call this mechanism the *p2c filter*, and it ensures that at least one AS (the inferred customer of the transit AS) relies on at least one other AS (the inferred transit provider) for transit from and towards the core of the Internet. As we aim to measure transit influence, these business relationships are an important source of information: merely being directly connected to an AS path that reaches the origin AS in a given country does not necessarily make an AS influential; being a direct provider of the origin, or of an AS closer to the origin, lends more confidence to our inference of influence.

CTI outlier filtering

We aim to further filter BGP-monitor location noise by removing outlier estimates of transit influence—both overestimates and underestimates resulting from the AS hosting a

BGP monitor being topologically too close or too far from the origin AS—to get an accurate assessment of transit influence towards that origin. Consider the example in Fig. 4.4. Based on the view of monitor X at OAS_1, TAS_1 is the only relevant transit provider because the p2c links between (TAS_2,OAS_3) and (TAS_3,OAS_4) are not visible[1] to TAS_1. The resulting CTI values underestimate the influence of both TAS_2 and TAS_3.

We implement a filter recently proposed for another AS-topology metric (Hegemony [52]). Specifically, we compute the CTI of each transit provider AS_t using BGP monitors from each monitor-hosting AS_h independently, as $CTI_{m(AS_h)}(AS_t,C)$, where $m(AS_h)$ is the set of monitors within AS_h. We determine which potentially-biased AS_h have gathered observations producing $CTI_{m(AS_h)}(AS_t,C)$ values in the bottom and top 10% of all values for that transit provider in that country and define $B(AS_t,C) := \{m(AS_h)|CTI_{m(AS_h)}(AS_t,C)$ is in the top or bottom 10% for AS_t and $C\}$. We then disregard all paths observed by monitors hosted in these potentially-biased AS_h by computing the $CTI_{M \setminus B(AS_t,C)}(AS_t,C)$; *i.e.,* CTI as observed by monitors not in $B(AS_t,C)$. As in the proposed filter by Fontugne at al. [52], we only implement outlier filtering where we have observations of $CTI_{m(AS_h)}(AS_t,C)$ from 10 or more AS_h, which occurs for 52.9% of transit AS-country pairs in our sample (a single AS can operate in multiple countries). In Sec. 5.5 we build a benchmark country-level metric by adapting Hegemony, and compare CTI to that benchmark.

4.3 Limitations

At a high level, CTI assumes all ASes and IP addresses are equivalent, which is certainly not the case. At the AS level, it is possible that one, dominant AS provides stronger security than a multitude of smaller ASes with tighter budgets. From the perspective of an attacker, though, a single AS having high CTI creates an opportunity; in the case of sophisticated attackers such as nation-states, the possibility of infiltration of any network cannot be discarded, but compromising

[1] In general, BGP announcements are propagated from customers to providers and vice versa, but peers only exchange announcements involving their own customers; as a result, links (TAS_2,OAS_3) and (TAS_3,OAS_4) are visible to TAS_0, who does not propagate them to TAS_1.

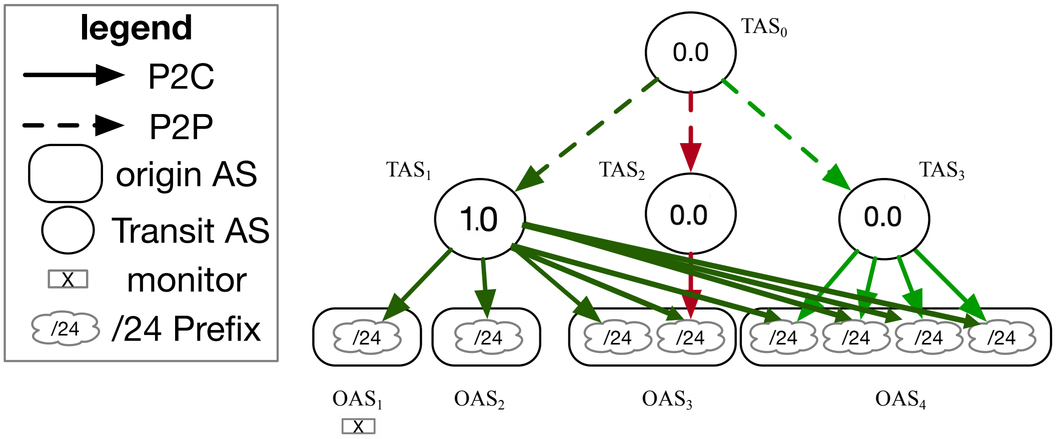

Figure 4.4. Example of a biased view of CTI due to location of monitor.

many ASes simultaneously (in order to observe traffic towards countries where no AS has high CTI) may be more challenging. As such, ASes with very high CTI still present a concerningly large observation footprint, regardless of their level of security against infiltration.

Similarly, IP addresses can represent vastly different entities. Both access and transit ASes may deploy carrier-grade network access translation (CGNAT) [83]. Since our model treats all routed IPs equally, it does not currently take into account the number of hosts multiplexing a single IP address. We leave this to future work, but note that an additional weight may be added to CTI: one that scales up the number of IP addresses in a given prefix by the number of hosts connected to those IPs, on aggregate. Even within a given network, however, individual hosts are unlikely to be equally important as some (e.g., those belonging to governmental organizations or power-grid operators) may have more sensitive traffic. Conversely, some networks might not even actually use all their IP addresses—although the latter issue is likely less of a concern in the countries we have studied as their allocation of IPv4 addresses tends to be constrained [44]. In addition to this fundamental conceptual limitation, there are a variety of technical details that could have out-sized impact on our conclusions, described in the following paragraphs.

We acknowledge that the BGP paths we observe and use to compute CTI are incomplete given the location of BGP monitors. Given the serious implications for countries that appear highly exposed to external observation and selective tampering by an AS, we argue that it is important to study such exposure with available data. Further, we note that there are two important factors aiding the credibility of our CTI findings: *(i)* our validation with network operators (Sec. 5.4), who confirm that the set of transit ASes identified in their countries is largely consistent with their own understanding of the country's routing ecosystem. *(ii)* There is greater visibility over provider-customer links in the AS-level topology [45, 74], which enables our analysis as we are studying exposure to observation or selective tampering by transit ASes, in particular.

A further potential source of inaccuracy is IP geolocation, as assigning prefixes to a geographic area is challenging and the commercial providers who sell such information use proprietary methods. However, since we limit our analysis to the country level, this source of inaccuracy is unlikely to significantly impact CTI, as geolocation databases are typically reliable at that granularity [66, 57, 82]. Further, while determining the location of prefixes originated by large transit providers with a global presence is problematic because of its dynamic nature and wide geographic spread, most networks are much smaller and will have limited geographic presence beyond their primary country of operation [106] (where most or all of their addresses will be located).

As a result of the bias and incompleteness of available BGP data (described in Ch. 2), the relationship inference that CTI relies on [74] is fed an incomplete input and is therefore prone to inaccuracies. Besides the possibility of incorrect inferences, these business relationships evolve over time, with customers switching between providers for cost- or performance-related reasons [45], and further a fraction of them are inadequately captured by the simple peer-to-peer vs. provider-customer distinction [59].

Finally, we note that our model does not apply to IPv6 addresses. For the purposes of this dissertation, a scope covering only IPv4 is appropriate, given that IPv6 deployment is far from wide in many developing regions, including Africa [29, 71].

Chapter 4, in part, is currently being prepared for submission for publication of the material. Gamero-Garrido, Alexander; Carisimo, Esteban; Hao, Shuai; Huffaker, Bradley; Snoeren, Alex C.; Dainotti, Alberto. The dissertation author was the primary investigator and author of this paper.

Chapter 5

CTI-Based Case Studies

In this chapter we present the findings of our CTI analysis for the set of countries identified in Ch. 3. We provide a high-level characterization of the transit ecosystem in each country by comparing the CTI scores of the top-5 ASes ranked by CTI (Sec. 5.1), as well as a set of ASes that appear in the top 5 of many countries (at least 10). Investigating the companies operating the ASes with high CTI, we find two prominent groups of organizations: submarine cable operators (Sec. 5.2) and state-owned providers (Sec. 5.3).

5.1 CTI Distribution Across Countries

In this we present an overview of the CTI distribution across countries. These countries show different transit profiles, ranging from high exposure to observation, where one AS is the most influential transit provider and others are very marginal, to less exposed countries with an ensemble of ASes with similar values of CTI. This dissertation's thesis aims at identifying countries with a top-heavy distribution of CTI values, which are particularly exposed to specific networks. Other nations, with a more flat distribution, signal an ecosystem that is less exposed to prominent transit ASes. Fig. 5.1 shows the distribution of CTI values for ASes ranked in the top 5 by CTI in each country. In 51 countries, the top-ranked AS has CTI ≥ 0.3, signaling high exposure to observation and tampering by that specific network.

The distribution of CTI rapidly declines across AS rank, with the median halving from the first to the second position. In 54 of 75 countries, CTI declines by over 30% from the top-ranked AS to its successor; the average and median decline across all countries are 46% and 49%. This suggests that in the vast majority of countries in our sample, a single AS is particularly prominent in terms of its capabilities to observe or tamper with traffic. Results for the full set of countries we study are included in Table 5.1.

5.1.1 Most exposed countries

Only five countries have a top-ranked AS with a CTI over 0.75: Cuba (in Fig. 5.1), Libya, Sierra Leone, Solomon Islands and Cape Verde. The latter two are small island nations. Among the remaining countries, Cuba appears to have the most-exposed transit ecosystem[1], in which the top-ranked AS has CTI of 0.96. Because CTI discounts indirect transit—and the top AS monopolizes observed, direct connectivity—the CTI of Cuba's remaining ASes declines rapidly (81% from the top-ranked AS to the second).

5.1.2 Countries around the median

The median of the leftmost bar in Fig. 5.1 consists of countries that are still considerably exposed to observation and tampering, with CTI values ranging from 0.34 to 0.37. These countries include Egypt, Tonga, Equatorial Guinea (shown in Fig. 5.1), Belize and Thailand. In Eq. Guinea, the median country, the top-two ASes each have a CTI over 0.3; these ASes have a provider-customer relationship with each other. Egypt and Belize have more skewed distributions, with a 62–79% decline from the top AS to its successor.

5.1.3 Least exposed countries

At the other end of the spectrum are five countries where the top-ranked has CTI values under 0.2: Chad, Bangladesh (in Fig. 5.1), Belarus, Turkey and North Macedonia. These

[1] This is consistent with previous work that focused exclusively on Cuba, finding its international connectivity to be constrained [33].

Table 5.1. CTI values of the five top-ranked ASes in each country (colored rows match Fig. 5.1). The table is sorted by the Herfindal-Hichman Index (HHI) [6] of the (normalized) top-5 CTI values. HHI reflects both the exposure of the country to the ASes, in aggregate, and the skew of the exposure's distribution.

Country name	CC	AS1	AS2	AS3	AS4	AS5	HHI
CAPE VERDE	CV	0.76	0.12	<0.01	<0.01	-	0.69
MALI	ML	0.7	0.13	0.07	0.03	0.02	0.46
SRI LANKA	LK	0.53	0.05	0.05	0.05	0.05	0.45
CUBA	CU	0.96	0.18	0.1	0.07	0.04	0.42
ST VINCENT	VC	0.67	0.11	0.1	0.07	0.01	0.41
LIBYA	LY	0.95	0.41	0.05	0.01	0.01	0.4
SAINT LUCIA	LC	0.59	0.1	0.1	0.08	0.02	0.34
SOLOMON ISLANDS	SB	0.78	0.39	0.05	0.03	0.02	0.34
BARBADOS	BB	0.59	0.12	0.11	0.06	0.05	0.31
MOROCCO	MA	0.59	0.14	0.07	0.07	0.06	0.31
ZAMBIA	ZM	0.58	0.15	0.07	0.07	0.04	0.3
IRAQ	IQ	0.57	0.14	0.09	0.08	0.05	0.28
GUYANA	GY	0.54	0.3	0.06	0.04	0.03	0.27
SIERRA LEONE	SL	0.81	0.37	0.11	0.08	0.06	0.25
BELIZE	BZ	0.34	0.07	0.07	0.06	0.03	0.24
YEMEN	YE	0.48	0.31	0.08	0.05	0.04	0.21
EL SALVADOR	SV	0.46	0.18	0.11	0.05	0.04	0.21
TURKMENISTAN	TM	0.33	0.24	0.07	0.03	0.02	0.21
TRINIDAD AND TOBAGO	TT	0.56	0.15	0.11	0.09	0.09	0.21
OMAN	OM	0.51	0.15	0.13	0.1	0.04	0.21
BOLIVIA	BO	0.54	0.27	0.11	0.08	0.05	0.19
PERU	PE	0.42	0.14	0.1	0.06	0.05	0.19
JORDAN	JO	0.55	0.15	0.13	0.1	0.09	0.19
LUXEMBOURG	LU	0.3	0.11	0.08	0.06	0.04	0.16
NAURU	NR	0.55	0.28	0.17	0.09	0.05	0.16
TUVALU	TV	0.6	0.3	0.2	0.13	0.04	0.14
JAMAICA	JM	0.51	0.17	0.12	0.12	0.11	0.14
ST KITTS	KN	0.33	0.12	0.07	0.07	0.07	0.14
PANAMA	PA	0.44	0.19	0.14	0.08	0.07	0.13
LESOTHO	LS	0.48	0.2	0.19	0.11	0.06	0.13
EGYPT	EG	0.37	0.14	0.1	0.1	0.07	0.12
ETHIOPIA	ET	0.58	0.26	0.25	0.1	0.08	0.12
ESWATINI	SZ	0.25	0.08	0.07	0.07	0.06	0.12
EQUATORIAL GUINEA	GQ	0.34	0.32	0.19	0.09	0.01	0.12
NICARAGUA	NI	0.31	0.14	0.09	0.07	0.06	0.11
MONGOLIA	MN	0.46	0.16	0.14	0.11	0.11	0.11
SUDAN	SD	0.46	0.15	0.14	0.13	0.1	0.11
MONTENEGRO	ME	0.44	0.18	0.14	0.11	0.09	0.11
ZIMBABWE	ZW	0.38	0.21	0.13	0.08	0.07	0.11
MYANMAR	MM	0.3	0.13	0.13	0.07	0.05	0.11
CAMEROON	CM	0.44	0.17	0.17	0.11	0.09	0.11
DRC	CD	0.27	0.15	0.09	0.06	0.06	0.1
BAHAMAS	BS	0.32	0.2	0.17	0.07	0.04	0.1
QATAR	QA	0.27	0.19	0.09	0.07	0.06	0.09
BURKINA FASO	BF	0.32	0.28	0.22	0.05	0.05	0.09
UZBEKISTAN	UZ	0.46	0.24	0.16	0.13	0.09	0.09
ARMENIA	AM	0.33	0.18	0.12	0.09	0.07	0.09
KUWAIT	KW	0.32	0.14	0.14	0.13	0.05	0.08
HAITI	HT	0.37	0.17	0.16	0.1	0.09	0.08
MALTA	MT	0.31	0.18	0.18	0.08	0.06	0.07
KOREA	KR	0.21	0.15	0.08	0.07	0.06	0.07
GUATEMALA	GT	0.22	0.22	0.17	0.07	0.04	0.07
THAILAND	TH	0.34	0.19	0.17	0.14	0.06	0.06
GEORGIA	GE	0.25	0.16	0.11	0.08	0.07	0.06
PALESTINE, STATE OF	PS	0.22	0.18	0.14	0.1	0.03	0.06
COLOMBIA	CO	0.33	0.18	0.15	0.11	0.1	0.06
INDIA	IN	0.25	0.18	0.17	0.12	0.03	0.06
SAN MARINO	SM	0.27	0.14	0.12	0.09	0.09	0.06
PORTUGAL	PT	0.26	0.23	0.13	0.13	0.05	0.06
SOMALIA	SO	0.26	0.17	0.14	0.1	0.08	0.05
TIMOR-LESTE	TL	0.27	0.27	0.23	0.13	0.07	0.04
TONGA	TO	0.35	0.24	0.17	0.15	0.13	0.04
GUINEA	GN	0.23	0.21	0.21	0.1	0.07	0.04
AFGHANISTAN	AF	0.22	0.15	0.12	0.1	0.07	0.04
VENEZUELA	VE	0.31	0.19	0.15	0.15	0.13	0.03
CHILE	CL	0.26	0.22	0.17	0.14	0.08	0.03
HONDURAS	HN	0.21	0.16	0.1	0.1	0.09	0.03
ECUADOR	EC	0.2	0.19	0.13	0.12	0.07	0.03
ALBANIA	AL	0.22	0.16	0.13	0.12	0.09	0.03
SAMOA	WS	0.27	0.23	0.19	0.18	0.09	0.02
NORTH MACEDONIA	MK	0.19	0.19	0.13	0.11	0.09	0.02
TURKEY	TR	0.16	0.13	0.12	0.09	0.08	0.01
CHAD	TD	0.14	0.12	0.11	0.09	0.08	0.01
BANGLADESH	BD	0.15	0.12	0.11	0.09	0.08	0.01
BELARUS	BY	0.15	0.13	0.11	0.11	0.09	0.01

countries have flatter distributions, with CTI declining at most 19% (or 13% on average) between the top-two ASes. As a result, we find no evidence of these nations being particularly exposed to a single network (unlike most of their peer countries in our sample). India, the country with the most Internet users in our sample, is in the bottom quintile (close to the other nations mentioned in this paragraph) with a top-AS CTI of 0.25, declining by 28% between the top 2 ASes.

5.1.4 Frequently Top-Ranked ASes

Of the 170 ASes present in Fig. 5.1, 129 of them are in the top-5 for only one country, with a further 34 ASes in the top-5 of at most 10 countries. There are some notable exceptions, however: 1299*-Telia (top-5 in 26 countries), 174*-Cogent (25), 3356*-Lumen (formerly Level3/CenturyLink) (22), 6939-HE (17), 6762*-T. Italia (14), 23520-C&W (14), and 6453*-Tata (12). Nearly all of these networks (marked with *) are in the inferred clique at the top of the global transit hierarchy [28]. C&W is only present in our analysis for countries in the Caribbean. HE has a very broad footprint, with countries in Africa (7), the Mid. East (3), W. Europe (2), Southeast Asia (2), South Pacific (2) and E. Asia (1).

5.2 Submarine Cable Operators

Submarine cables are known to be an important part of the global Internet infrastructure [32, 49, 70] and appear in the top-5 ASes of most countries we study. Nicaragua, Guatemala, and Guyana are the only three nations where none of the top-5 ASes are associated with the submarine cables landing in the country. In this section, for each country, we find the highest-ranked AS by CTI where there is evidence of an institutional connection between the AS and an owner or operator of a submarine cable. We define an AS as a submarine cable operator if we find a direct match between the AS Name, the AS Organization [36], or a corporate parent organization (*e.g.,* CenturyLink for Level3, the Government of Sierra Leone for Sierra Leone Cable Company) and the owners of a submarine cable operator according to TeleGeography [97] and Infrapedia [63].

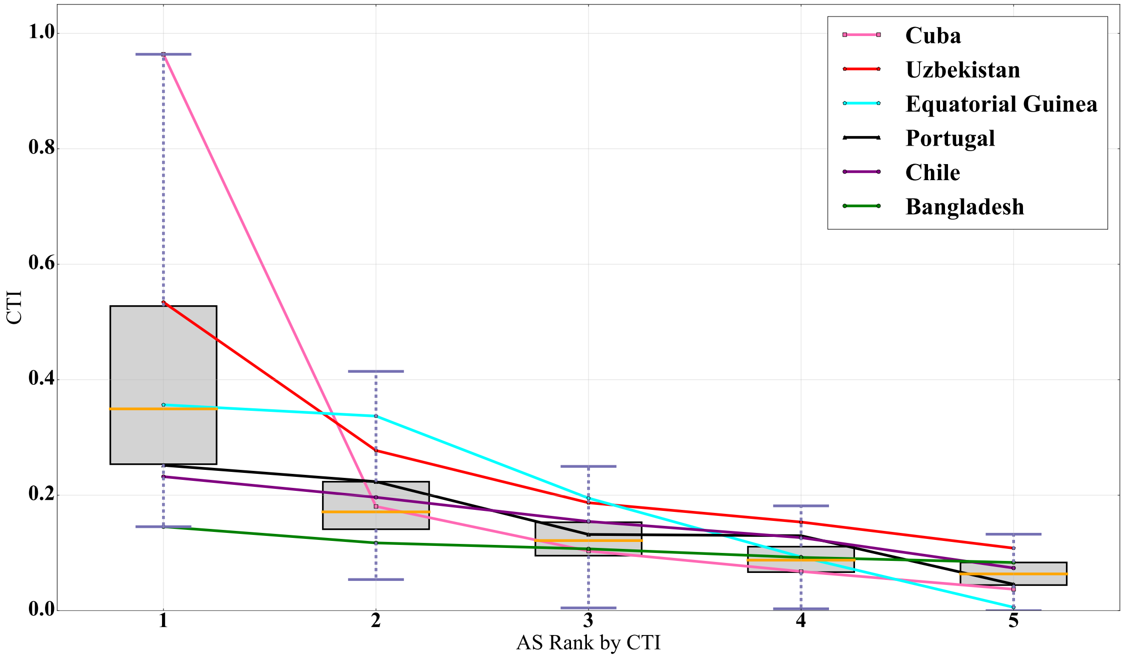

Figure 5.1. Boxplot of CTI distributions for the top-5 ASes. Countries highlighted are representative cases ranging from those that are most exposed (Cuba) to least exposed (Bangladesh). The findings confirm part of this dissertation's thesis, as the median country is significantly exposed (34%) to the top AS.

This process yields submarine cable ASes in 46 countries out of 51 possible, as 19 of the 75 countries are landlocked, and 5 have no submarine cable connectivity according to the operator databases. In three additional countries (Myanmar [94], Solomon Islands [43], and Dem. Rep. of Congo [69]) only TeleGeography provides an AS to submarine cable match, which we confirm with information from the cited sources (the operators themselves, the government of Australia, and a submarine cable news source). In the remaining two countries (Thailand [99] and Samoa [95]) where we were not able to find an AS to submarine cable from TeleGeography, we rely on the cited sources (from the operator and a Samoan news outlet) to find a match. Note that only operators of submarine cables who appear as an AS on the BGP path can be identified using this method, so our findings may be a lower bound of the influence of submarine cable operators in some countries.

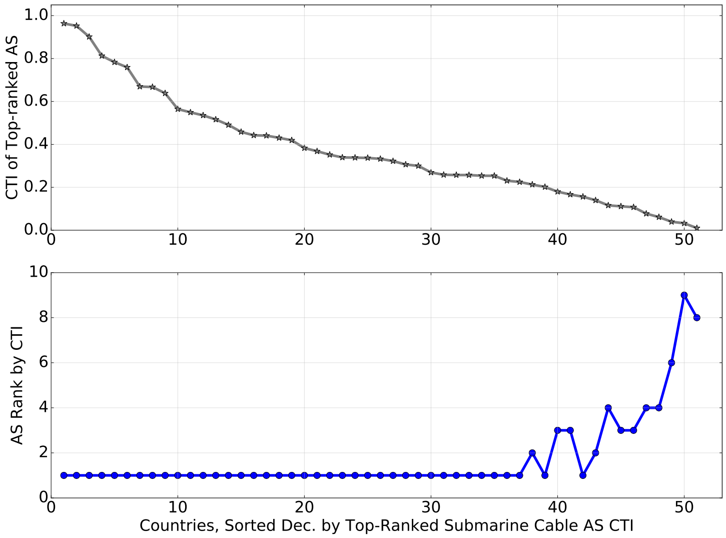

Figure 5.2. Top: CTI of top-ranked submarine cable AS. Bottom: CTI rank of top-ranked submarine cable AS.

Our findings are shown in Fig. 5.2, with the CTI of the top cable-owning AS in each of the 51 countries shown in the upper portion, and the ordinal ranking of that AS in its country's ecosystem in the bottom portion (the order of countries is the same in both plots, and sorted by the CTI of the top cable-owning AS). In 39 countries, a submarine cable AS is ranked at the top by CTI, with an average rank of 1.8.

Note that being the top operator by CTI means different things in different countries, as the underlying potential exposure to observation affects the CTI of the top AS. For instance, in Turkey and South Korea a cable-owning AS ranks first by CTI, but these ASes (9121-Turk Telecom and 6939-Hurricane Electric) have CTI scores of 0.16 and 0.21, respectively. By contrast, in Cuba and Libya, a submarine cable operator (11960-ETECSA and 37558-LIT) is also ranked first but with CTIs of 0.96 and 0.95, respectively. As a result, Turkey and South Korea are much less exposed to a single AS than Cuba and Libya.

Because submarine cable operators are inherently linked to physical infrastructure, it is possible to construct an alternative view of the findings in Fig. 5.2 based on the actual submarine cables being operated. If a single cable is linked to an AS with high CTI in multiple countries, an event affecting that cable (which may include weather-related or "anchoring and fishing activities” [64], as well as targeted attacks) may have serious consequences in multiple countries [104]; such cables and the associated ASes are listed in Table 5.2. Most notably, C&W is among the top providers in 10 countries in Central America and the Caribbean thanks to its ownership of the ECFS and ARCOS-1 cables. Its CTI in those countries ranges from a dominant position (0.56–0.67 in all the islands save for Saint Kitts and Nevis, and Bahamas), to a more marginal position in Central America (0.03–0.11 in Nicaragua and Honduras). Tata, Telefonica and Bharti Airtel also have an important transit presence in West Africa, Western South America, and South Asia respectively. Table 5.2 also highlights the critical nature of the SeaMeWe family of submarine cables, with top ASes by CTI being identified as a co-owner of one of the cables in eight countries. Cables in this family have received attention in previous studies [42].

Table 5.2. Submarine cables and their top AS operators by CTI. ASes listed match the countries from left.

Submarine Cable	ASes (# of countries, if more than one)	Countries
SeaMeWe-4	9498-BHARTI Airtel (2), 8452-TE	India, Bangladesh and Egypt
EIG	37558-LIT, 9498-BHARTI Airtel, 8452-TE	Libya, India, Egypt
SeaMeWe-5	45489-SL Tel., 9121-Turk Tel., 8452-TE	Sri Lanka, Turkey, Egypt
AAE-1	15412-Reliance, 8452-TE, 8781-Ooredoo, 38040-TOT, 8529-Omantel	Yemen, Egypt, Qatar, Thailand, Oman
EASSy	16637-MTN, 37662-WIOCC	Sudan, Somalia
SeaMeWe-3	6762-TIS, 8452-TE, 45558-MPT, 9121-Turk Tel., 6939-HE[1]	Morocco, Egypt, Myanmar, Turkey and South Korea
ACE	327903-Ministry I&C, 37529-GITGE, 8346-Sonatel	Sierra Leone, Eq. Guinea, Guinea
WACS	6453-TATA, 30844-Liquid Tel., 15964-Camtel	Cape Verde, Congo DRC, Cameroon
ECFS	23520-C&W (5)	S.V.G., S.K.N., St. Lucia, Barbados, Trinidad & Tobago
ARCOS-1	23520-C&W (6)	Honduras, Nicaragua, Venezuela, Belize, Bahamas, Panama
Pan-Am	12956-Telefonica (3)	Ecuador, Peru, Bolivia[2]
AMX-1	14754-Telgua, 14080-Telmex Colombia	Guatemala, Colombia

[1] Partnership with Telecom Malaysia [98].

[2] Not included in Fig. 5.2 as it is landlocked.

Submarine cable operators in Nicaragua and Guyana

We discuss Nicaragua and Guyana, the only two countries where we were unable to identify a submarine cable operator in the top 5 by CTI. In the former, Infrapedia [63] lists a terrestrial link operator (UFINET) which is ranked fifth, so it is possible that that terrestrial link to Honduras and Costa Rica is frequently used by Nicaraguan operators. The CTI top 4 of Nicaragua is composed of well-known international transit providers: Cogent, Telia, Tata and Telefonica. Telefonica states that it operates fiber routes connecting Managua, Nicaragua's capital, with both northbound and southbound international terrestrial lines [96]. We speculate that the remaining carriers operate some combination of leased capacity on terrestrial links (to reach Mexico or another country in the region with better submarine cable connectivity) and leased capacity on ARCOS—the sole submarine cable with a landing station in the country [97, 63]—in order to deliver traffic to the nation. Another possibility is that an owner or operator of ARCOS delivers traffic to the country but is not visible on BGP announcements.

In Guyana, a submarine cable operator AS (19863-Guyana T&T) originates 78% of the country's addresses and is ranked 8th by CTI, which follows our definition of influence: an AS originating addresses cannot have transit influence over them. Further, the country seems to be connected only to relatively short cables with landing points in Trinidad & Tobago, Barbados and Suriname (which is not in our CTI study). C&W, a major international cable operator in the region, is the top AS by CTI in Trinidad and Barbados and is ranked second in Guyana (its parent company, Liberty, is ranked third), so while C&W seems to have no landing points in the country it may still be an AS with observation capabilities over traffic there flowing. Another AS with presence in Trinidad is ranked fourth in Guyana, AS5639-T.S. T&T, with the remaining two ASes in the top 5 being Cogent (ranked first) and Telia (ranked fifth). The aforementioned transit ASes in Trinidad may therefore be among Guyana's most influential, which is not entirely surprising from a geopolitical perspective, as Guyana *(i)* has a contested border with both Venezuela [101] (which claims much of its territory) and Suriname [61], the latter of which would not connect

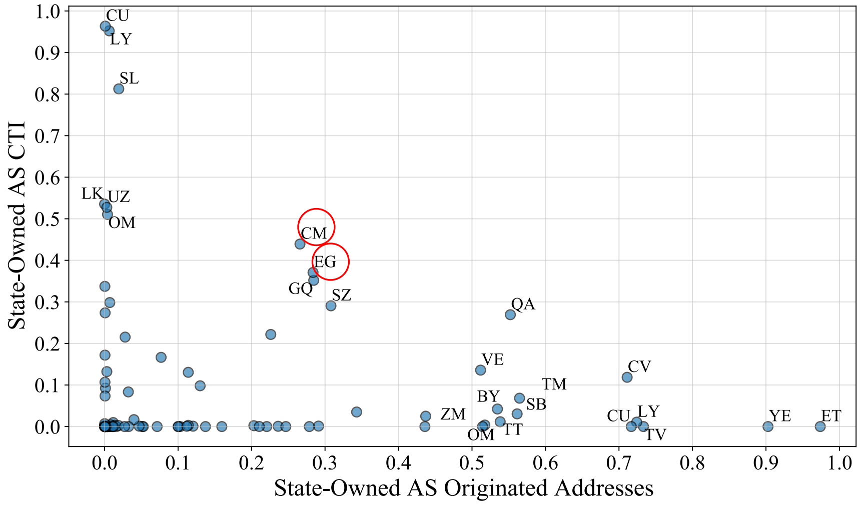

Figure 5.3. CTI and fraction of addresses originated by domestic, state-owned ASes in our study. ASes providing both transit and Internet access include those in Cameroon (CM) and Egypt (EG).

Guyana to additional cables [63, 97]; and *(ii)* is constrained by the Amazon forest to the south (difficulting a connection to Brazil).

5.3 State-Owned Transit Providers

In more than a third (26) of nations, we find that at least one of the top-5 ASes is state-owned (according to a recent study [41]), motivating us to further examine the total influence of a country's government on its Internet connectivity. In particular, we adapt CTI to quantify the influence of state-owned conglomerates—as some nations have more than one state-owned AS—and apply it to the 75 countries in our sample. We use as input a list of ASes that are majority-owned by sovereign states [41]. The list was manually verified and encompasses both access and transit ASes. The dataset includes major telecommunication providers as well as its sibling networks and subsidiaries. Using this list, we find 100 state-owned ASes who operate domestically (*i.e.,* where the state owner and the country of operation are the same) in 41 countries.

5.3.1 Conglomerate Footprint

Our initial exploration of the influence of state-owned ASes concerns the role each AS plays in the ecosystem of its country, as shown in Fig. 5.3. We find that state-owned ASes tend to provide either transit or access, usually not a combination of both. (Most points in Fig. 5.3 line up along an axis, rather than towards the middle.) As a consequence, meaningfully estimating the footprint of the state requires combining the two kinds of influence as well as aggregating data for AS conglomerates. Two exceptions where a state-owned AS provides both Internet access (*i.e.,* as an origin AS) and serves transit to other ASes are Cameroon and Egypt; in the former, Camtel has both a high CTI (0.44, ranked first) and originates 27% of the country's addresses (second only to Orange Cameroon). Egypt's TE has a CTI of 0.37 and originates 28% of the country's addresses.

We begin that estimation by computing CTI for not just a single AS, but a set of ASes, while not "double counting" influence over the same address space, *i.e.,* if two of the state's ASes originate and provide transit to the same addresses, we add those addresses to the state-owned conglomerate's footprint once. We call this derived metric $CTIn$. Intuitively, $CTIn$ reflects the "pure-transit" footprint of the ASes, crediting only the addresses where state-owned ASes serve exclusively as transit providers. For instance, if AS A and AS B (both of which operate in country C) respectively originate and provide transit to the same /24 prefix, $CTIn$ says that the conglomerate $S_C = \{A, B\}$ does not have transit influence over the /24 prefix. Formally, $CTIn_M(S_c, C) \in [0,1]$ is calculated as

$$\text{‘}\sum_{m\in M}\left(\frac{w(m)}{|M|}\cdot\sum_{p|\text{onpath}^*(S_c,m,p)}\left(\frac{a(p,C)}{A(C)}\cdot\frac{1}{d^*(S_c,m,p)}\right)\right), \tag{5.1}$$

which is essentially identical to Eq. 6.2.2, except that S_c is a set containing all of the ASes in the conglomerate; $\text{onpath}^*(S_c, m, p)$ is true if $\text{onpath}(AS_t, m, p)$ is true for some $AS_t \in S_c$ and p is *not* originated by any AS in S_c; and $d^*(S_c, m, p) = \min_{AS_t \in S_c} d(AS_t, m, p)$, *i.e.,* the AS-level distance from p to the closest AS in the conglomerate.

Finally, we define the total footprint of the conglomerate, *i.e.*, the address space that is either originated or for which transit is served by a conglomerate AS. The state's footprint $F(C) \in [0,1]$ is calculated as

$$F(C) = CTIn_M(S_c, C) + \sum_{AS_o \in S_c} \frac{a^*(AS_o, C)}{A(C)}, \tag{5.2}$$

where $a^*(AS_o, C)/A(C)$ is the fraction of addresses in country C originated by AS_o. The first term of the sum is the pure-transit footprint and the second term is the addresses directly originated by the state-owned conglomerate S_c.

5.3.2 Findings

Fig. 5.4 shows our findings for the state-owned footprint (F, bar height), the originated fraction by state-owned ASes (orange bar), and pure-transit footprint of state-owned ASes ($CTIn$, blue bar). Our results suggest that domestic state influence exists on a spectrum where some countries, such as Ethiopia, Cuba, Libya and Yemen, rely overwhelmingly on the state for the provision of Internet access and (F between 0.90–0.97), whereas others, such as Colombia, Turkey, Mongolia and Ecuador have relatively marginal state-owned enterprises (F between 0.01–0.12).

Regarding the mode of influence that states use, in many countries in Fig. 5.4, most of the bar height is contributed by the orange portion, meaning that the footprint of the state comes from addresses directly originated. However, in some countries the state punches above its access network weight by deploying an influential transit provider, *i.e.,* those where the bar height is not dominated by the origin contribution in orange. The countries where pure-transit influence is largest (0.2 or more, or pure-transit influence over at least a fifth of the country's addresses)

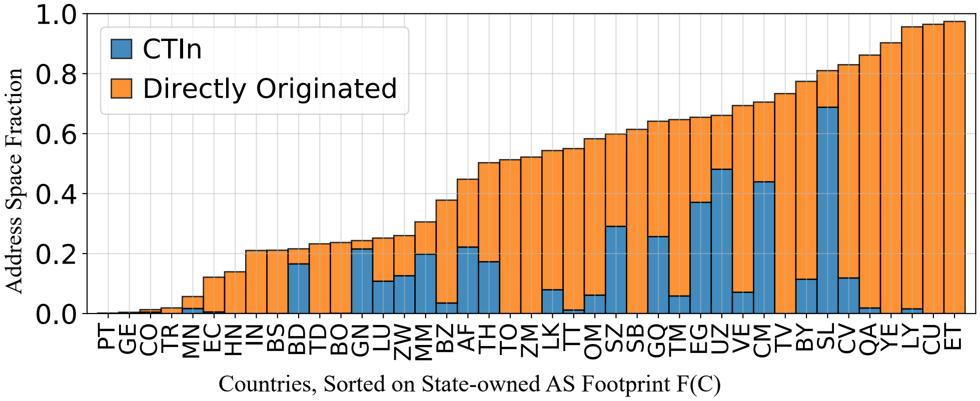

Figure 5.4. State-owned originated address space a^* (orange bars), $CTIn$ (blue bars), and state footprint F (bar height) for countries in our study (X-Axis, sorted by F).

are shown in Tab. 5.3. In these countries, all of which are in Africa or Central Asia, providing transit considerably increases the influence of the state.

5.3.3 Pure-transit footprint of state-owned ASes

In countries where state-owned ASes have a large value of $CTIn$ (Tab. 5.3), it is possible that providing Internet access directly is beyond the capabilities of the state (at least in some of each country's regions) which would explain the relatively low footprint contribution of addresses directly originated. In these countries, building an influential transit network may be a cost effective way to expand the purview of the state, be it for monetary gain (improved tax collection), infrastructure improvement (increasing the country's available international bandwith), or surveillance (expanding the fraction of the country's traffic that traverses a state-owned organization). We note that the mere existence of these influential transit ASes does not signal willingness of the state to engage in surveillance or selective tampering, but rather that the government may have opportunities to do so. For instance, Myanmar's state-owned *Myanma Posts and Telecommunications (MPT)*, included in our analysis (see Tab. 5.3), appears to have been involved in the disruption of the country's Internet service during the recent coup [58].

Table 5.3. Top countries by *CTIn*.

Country	*CTIn*	*F*
Sierra Leone	0.69	0.81
Uzbekistan	0.48	0.66
Cameroon	0.44	0.71
Egypt	0.37	0.65
Swaziland	0.29	0.60
Eq. Guinea	0.26	0.64
Afghanistan	0.22	0.45
Guinea	0.22	0.24
Myanmar	0.20	0.31

5.3.4 Impact of public policy

In our CTI results, we find anecdotal evidence of the impact of policy on each country's telecommunications ecosystem in two ways. First, while the underlying motives for centralized and state-owned operation of national networks is outside the scope of this study, it is worth noting that the four countries where $F > 0.9$ (Ethiopia, Cuba, Lybia and Yemen) are all labeled as authoritarian countries by the Democracy Index [16], so the national government's extensive footprint may allow for effective surveillance capabilities. Second, two countries where public policy has generally favored the diversification of international routes (Bangladesh [78]), and the establishment of a strong domestic peering mesh (Chile [79, 31, 40]), have among the lowest values for CTI of the top-ranked AS (0.15 and 0.26, respectively, or the bottom quintile of the countries in our sample). These trends, if emulated in other nations, might mitigate the risks imposed by concentration of inbound routes on a few ASes.

5.4 Validation

In this section we describe our discussion of CTI findings with operators, as well as an analysis of CTI's temporal stability. We summarize our discussions with operators regarding ASes identified by CTI as highly influential in their countries. *Our findings are largely consistent* *with each operator's view of the transit ecosystem of the countries discussed with them*: the per-country rate of true positives—in terms of influential transit ASes confirmed by the operators in 6 countries—was 83%, on average[2].

5.4.1 Operator Validation

We discussed our findings with employees or contractors of two types of organizations: commercial network operators and non-profits who conduct networking research (universities, registrars, and non-commercial network operators). Discussions with all but one of these organizations are anonymized following their requests. These discussions took place in the spring of 2020, unless otherwise specified.

Commercial Network Operators

We emailed one former and eight current employees at nine companies operating transit and/or access networks primarily in Africa and Latin America. An operator confirmed that they operate a large transit network in two specific countries in Sub-Saharan Africa. Two operators in Africa (one of which is Liquid Telecom[3], the sole non-anonymized conversation we report) broadly confirmed being a transit provider for traffic flowing towards the countries we indicated. One former and one current operator in a single Sub-Saharan African country were sent the set of top 7 ASes[4] by CTI in that country. An operator responded with four ASes that they state are the only direct upstreams of a large access network; these are ranked 1, 2, 5 and 6 by CTI in that country. The second operator confirmed the top 5 of the 7 we sent; these are ranked 1, 3, 4, 5 and 6 by CTI in that country. One operator in LACNIC confirmed—in Oct. 2020—12 of the 15 top ASes identified by CTI in that nation as influential operators. Two additional operators never responded, while a third one declined our request.

[2] The per-country true positive rates are, sorted increasingly: 66%, 70%, 80%, 90%, 90%, 100%.

[3] Which operates in Zimbabwe, Zambia, Lesotho, Somalia, and D.R. Congo.

[4] Unfortunately due to the timing of the validation process, we sent a set of ASes—we did not include actual CTI values in the message, just the set of top ASes—to these operators that was produced before updating our CTI methodology to its current form; 6 of 7 ASes are present in both our final CTI top 7 and in the outdated list we sent them.

Networking researchers at non-profits

We contacted 16 researchers in 10 countries in Africa and Latin America. Of these, one (from Latin America) declined to comment on the list of top ASes we sent them. Two other researchers declined to comment altogether, and eight did not respond. Two researchers in two different Sub-Saharan African countries confirmed 8 of the 10 ASes in the top 10 we sent[5]. A researcher in a country in Sub-Saharan Africa confirmed 5 of the top 7 ASes by CTI[6]. Two researchers in a single country in North Africa responded to our set of top ASes by CTI. The first researcher was able to confirm that 7 of the 10 ASes are transit providers of access ASes operating in this country. The second researcher's response in this regard was to dispute an AS in the top 10 by CTI, and suggesting that we investigate the transit providers of the country's access networks[7]; the disputed AS is an inferred transit provider of one of those access networks, a relationship which was directly confirmed by the first researcher.

ASes with Prefixes Geolocated

We sent a mass email request to the WHOIS `abuse` address registered by ASes that had prefixes geolocated in 10 countries[8] (with IRB approval): BO, CO, VE, CM, BD, GT, CL, HN, SV and ZW. These were selected as a mix of large and small (by #ASes) countries where English or Spanish are among the primary languages. We received 111 responses in 9 of these countries (all but ZW). Of these, 107 confirmed they operate primarily in the country that we geolocated their prefixes to[9]. Additionally, 108/111 were willing to discuss which type

[5] See previous footnote; in this case 9/10 ASes remained constant across both sets for both countries, including the 8 the researchers confirmed. The operators also confirmed an additional AS from the outdated set in each country, which are ranked 12 and 13, respectively, in the final CTI tally.

[6] See previous footnotes. In this case 5/7 ASes are constant across both sets; the operator confirmed two additional ASes from the outdated set, which are ranked 9th and 10th by CTI in the final tally.

[7] Which they listed in their response; we found that these networks originate 99.7% of addresses in that country, an anecdotal but encouraging sign regarding the correctness of our assessment of which ASes originate addresses in this particular country.

[8] We only contacted ASes who had at least 1% of their addresses in the country, and since this survey took place in 2021, we use the addresses geolocated in Jan. of that year.

[9] In 3 cases, they stated that the country was among their primary places of operation, but that they also operated in other countries.

of business relationship dominated their inbound international traffic. Of these, 83 stated that transit relationships are the primary modality.

5.4.2 CTI Temporal Stability

We apply our CTI methodology to a set of BGP paths from Feb. 2020 and compare the output to that discussed in the results sections (from Mar. 2020). Specifically, we compute the absolute value of the difference in CTI across both months for ASes listed in the top 5 for each country in Mar. 2020. We compute the absolute difference in CTI for a total of 374 AS-country pairs[10], or 172 ASes in 75 countries. The 25th percentile, mean, median, and 75th percentile of this absolute difference are 0.002, 0.003, 0.008 and 0.025, so the CTI values are relatively stable across these months.

5.5 Comparison with an Alternative Country-Level Metric

In this section, we build a country-level alternative metric based on Hegemony [52] and compare CTI to it. The reason for the comparison is to determine if CTI is too aggressive in its filters, discarding too much input data. For that purpose, we build a benchmark using local hegemony, a metric of centrality of any AS (including both transit providers and peers) on paths towards a single origin. Hegemony consists mostly of a single filter on input BGP data, making it an appropriate benchmark. This benchmark was not trivial to build, as "hegemony local" produces a bilateral metric of influence between a transit AS and an origin AS on the global topology. While Hegemony is concerned with extracting the most accurate estimate of centrality on an existing graph, and not with estimating country-level inbound route diversity as CTI, it is possible to build a metric that serves a similar purpose as CTI, which we call *country-level hegemony* (*CLH*) as follows:

$$CLH(AS_t,C) \in [0,1] = \sum_{AS_o \in (C)} H(AS_t,AS_o) \cdot \frac{a^*(AS_o,C)}{A(C)}$$

[10] Cape Verde only has four transit ASes, which is why there are 374 AS-country pairs instead of 385.

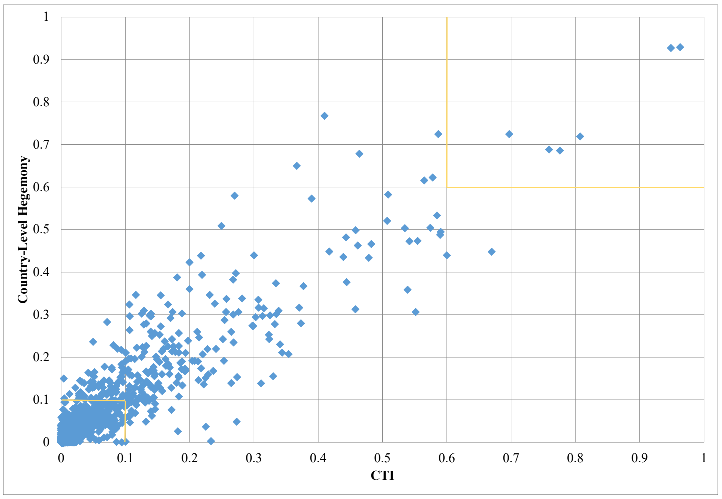

Figure 5.5. CTI and CLH scores for the 6,428 AS-country pairs in our study.

where $H(AS_t, AS_o)$ is the hegemony score of AS_t on AS_o during the same period[11] in March 2020 when we applied CTI, (all the other terms have been previously introduced in Eq. 3.1).

In other words, CLH is a conceptually equivalent metric to CTI. For each AS-country pair (a transit AS serving a country) in our study of 75 countries, we show both CTI and CLH in Fig. 5.5. These metrics tend to agree on a score for a given transit AS in a given country: a linear regression has a slope of 0.9988, intersection of 0.002, and R^2 of 0.87. The takeaway is that the heuristics of CTI do not introduce unnecessary noise to our analysis because, on aggregate, a country-level alternative based on Hegemony—which applies a single filter to BGP data in order to estimate AS centrality, and excludes considerably fewer BGP monitor than CTI does—tends to agree with CTI's assessment.

Further, the metrics tend to produce qualitatively similar assessments of each transit AS in each country: they either assess the AS as marginal—both metrics assign it a score lower than 0.1 (6,124 of 6,428 AS-country pairs); or they assess the AS as having a very high score—both

[11] As Hegemony is published in 15-min intervals [25], we take the 5-day average score.

assign a value greater than 0.6 (6 AS-country pairs). Hence, we find that out heuristics are not overly aggressive nor unduly excluding input data.

There is, however, disagreement among the metrics in the middle section of Fig. 5.5 (remaining 298 AS-country pairs); we study the data points in the area where either metric has a score between 0.1 and 0.6.) In particular, we investigate the data points where the two metrics disagree the most in the remainder of this section.

We manually inspected 15 AS-country pairs where CLH produces a much higher score than CTI ($CLH - CTI > 0.2$), or vice versa. We stress that the true score for each AS-country pair is unknowable, so what we intend to evaluate is the impact of the individual components of CTI on these data points. To that end, we compute an alternative set of CTI scores, where the indirect transit discount (a coarse heuristic defined in Sec. 4.1.2) is not applied. For the 11 AS-country pairs where CLH produces a higher score than CTI, the indirect transit filter has a meaningful impact on our estimate of CTI: not applying this filter would have increased the CTI score by a median and average of 93% and 102%, respectively. Indeed, for these 11 AS-country pairs, the indirect transit filter causes 95% (median) and 96% (metrics) of the gap between the metrics. Since our purpose was to produce a conservative estimate of the transit influence of indirect transit providers, we find that this CTI heuristic is working as intended, given the possibility that we are not observing alternative links further away from the origin (and, therefore, overestimating the transit influence of indirect transit providers). Further, one of these data points is a transit AS serving a country in our validation set; the operator rejected the claim that the AS is influential in that country, which is one case where the indirect transit discount moved the CTI score towards a model that is consistent with that operator's understanding of their country's inbound routing ecosystem.

All AS-country pairs where CLH produces a meaningfully lower score than CTI measure influence on island nations: Nauru, Samoa, East Timor, and St. Vincent & Grenadines. The indirect transit filter has a minor impact on CTI for these 4 AS-country pairs: between 0-2.5%. The disagreement between the metrics, then, primarily stems from the core analysis each does over BGP paths. One of these AS-country pairs involve C&W (a previously introduced influential AS in the Caribbean, Sec. 5.1), an operator which owns or co-owns submarine cables landing in St. Vincent & Grenadines, which suggests that the operator is likely influential on these island, potentially justifying a high CTI score. One other AS-country pair relates to a small island in the South Pacific, Nauru, reportedly relying primarily on satellite connections for international connectivity [17]. Estimating inbound route diversity in this nation may be particularly challenging for any metric, but they are nonetheless likely exposed in their external connectivity (and therefore their inclusion in our study is justified given our goal of identifying exposed nations).

Chapter 5, in part, is currently being prepared for submission for publication of the material. Gamero-Garrido, Alexander; Carisimo, Esteban; Hao, Shuai; Huffaker, Bradley; Snoeren, Alex C.; Dainotti, Alberto. The dissertation author was the primary investigator and author of this paper.

Chapter 6

Understanding the Exposure of Critical Networks

We tackle the issue of identifying transit providers that serve traffic to organizations in three critical sectors—education, financial services, as well as government and utilities—and determine when the connectivity of these sectors is more dependent on particular transit providers than the nation as a whole. In this chapter, we introduce our methodology to identify origin ASes belonging to critical organizations and the transit networks that serve them.

Our first step is to identify critical organizations in each country that operate ASes, since there is no central repository listing that information. We rely on manual inference and operator validation for this task. Because of the laborious inference process we rely on, we limit our scope to three countries: Bolivia, Chile and Venezuela. Then, we proceed to study the transit providers that serve the critical-sector ASes in these countries. A challenge in this space is the lack of a consistent definition of "critical" organization. The critical nature of an institution depends on many economic, social, political and cultural factors. For instance, universities may be considered critical by the residents of one country, but more marginal by those in another nation. We introduce a working definition of critical organization as those operating in industries that tend to be tightly regulated (education, finance, government and utilities). Regulation itself is a signal that policymakers—and often also the general public—have a strong interest in the operations of these sectors.

CTI is a country-level model, and highlights the most influential transit providers towards all (reachable) access networks in the country. By definition, CTI will assign more influence to transit providers who serve large ISPs, who are typically consumer-serving companies. Critical organizations also own ASes, and are often much smaller (in terms of number of IP addresses originated) than ISPs. In order to identify the transit ASes that are most influential on critical organizations, then, we need a transit influence metric at a different granularity: the AS level. To that end, in this chapter we present the AS-Level Transit Influence (ATI) metric, which determines which transit providers are most influential over a single origin AS.

Finally, in order to draw inferences that allow for comparisons across sectors composed of multiple origin ASes in a single industry, we present the Weighted ATI (W-ATI) metric. W-ATI allows us to identify the transit providers serving each critical sector as a whole.

6.1 Preliminaries

In this section we describe our country sample as well as external datasets we use to study interconnection.

Country sample. Our AS-classification method partly relies on local knowledge. We are familiar with networks on three continents: North America, Europe and South America. We adapt CTI to study critical-sector interconnection, a tool which is particularly in *transit-dominant* countries, and the first two continents are not well represented among them. As a result, we develop our methodological prototype by studying South America.

The following South American countries are not included in the 75 transit-dominant nations: Brazil, Argentina, Uruguay, Paraguay, and Suriname. From the remaining 7 countries, we attempt to build a balanced sample in terms of small and large countries by number of ASes with presence in these nations, and also in terms of their location in the continent. Chile is the largest country (with over 300 ASes), so we include it. Bolivia and Suriname are the smallest, with under 100 ASes; we choose Bolivia as it is landlocked. Finally, Venezuela and Peru have the median-high and median-low number of ASes: 149 and 142, respectively. We choose Venezuela from these two, as it is more geographically dissimilar from Chile.

BGP and AS relationships. We begin our analyses with the prefixes listed in CAIDA's Prefix-to-Autonomous System mappings derived from RouteViews [37]. We then limit our analysis to prefixes that are either commercially geolocated [26] or delegated by Regional Internet Registries [22] to a country: 5,332 prefixes in Chile, 1,933 in Venezuela, and 847 in Bolivia. We find those prefixes in the IPv4 AS-level paths observed in BGP table dumps from RouteViews [23] and RIPE RIS [21] during the first five days of March 2020 (Bolivia and Venezuela) or January 2021 (Chile). The final set has 185,279 unique AS-level paths towards 8,027 prefixes (a prefix may be split across multiple countries). We consider the set of prefixes and the ASes that originate them on each observed path in combination with the inferred AS-level relationships published by CAIDA [19] for March of 2020 or January 2021, depending on the country. We then find the set of ASes that originate either geolocated addresses or delegated IP blocks in the country in either March of 2020 (Venezuela, Bolivia) or January 2021 (Chile[1]).

6.2 Methodology

In this section we describe our prototypical methodology to identify critical-sector ASes and study the transit networks serving them.

6.2.1 AS Classification

We identify critical organizations that operate ASes by manually inspecting the following sources, sequentially, until we identify the AS' sector: *(i)* top-level websites (and their "about us" sections) parsed from registration records published by Regional Internet Registries (RIRs), specifically domains of abuse-contact emails (excluding known email providers), *(ii)* institutional websites published by the company listed in the AS Name or its parent organization (AS2Org [36],

[1] In Chile, in order to reduce the size of the input to our manual classification, we exclude 107 ASes (24% of the total) that have less than 1% of their originated addresses in Chile, *i.e.,* we exclude ASes with a small fraction of their addresses in the country. We argue that these ASes are unlikely to belong to Chilean critical organizations.

which is based on RIR delegation records) or third-party financial databases (*e.g.,* Bloomberg [11]), *(iii)* PeeringDB [80] record for either the AS or its parent organization, *(iv)* Spam, Abuse or other networking databases listing the ASNumber (*e.g.,* CleanTalk [10]), *(v)* keywords on the ASName, such as “Telecomunicacion”, as well as the commercial brands of well-known entities in the region, such as “Telmex”.

Our categories to classify the organizational purpose of ASes are as follows:

- *Telecommunications, Industrial and Professional.* A broad category including the vast majority of non-critical ASes: access and IP transit providers, content delivery networks and hosting companies, and professional services (*e.g.,* management consulting).
- *Education and Academic.* Universities, research institutes, and other academic-oriented institutions.
- *Government and Public Utilities.* National or local government organizations, along with utilities such as electric grid operators.
- *Financial Services.* Banks, credit unions, insurance companies and other financial organizations.
- *Other Non-Critical.* Miscellaneous organizations not falling under any of the preceding categories, such as retailers.

In the rest of this chapter, we restrict our analysis to the three middle categories above, *i.e.,* critical organizations.

Findings and validation. We classify the organizational purpose of 551 ASes. The results are included in Table 6.1. The vast majority of ASes in these countries are in the telecommunications, industrial and professional services sectors. The subject of the analysis of international connectivity in this chapter, critical ASes, represent between 16-20% of the total in each country.

Table 6.1. The number of ASes in each category.

Category	Chile	Bolivia	Venezuela
Telecommunications, Industry, Professional Services	262	53	120
Financial Services	24	6	7
Education and Academic	18	3	12
Government & Util.	12	5	5
Other (Non-Critical)	16	3	5
Total	332	70	149

We validate our findings by requesting that network operators disclose the organizational purpose of their AS (by emailing registered email addresses in RIR registration records—with IRB approval). We find 39 cooperative ASes: 31 in Chile, 5 in Venezuela and the rest in Bolivia. Then, we compared their responses with our classification. In 38 cases[2], our classification was correct (97.4% accuracy).

6.2.2 AS-Level Transit Influence (ATI)

We study transit influence at the AS level, as that is the granularity that is suitable for the purpose of understanding the connectivity of ASes operating critical organizations. To that end, we introduce the AS-Level Transit Influence (ATI) metric. We restrict ATI's input data to prefixes originated by each individual critical-sector AS—in other words, while CTI considers all prefixes (partially) geolocated to a country, ATI treats each origin AS in a country separately, and considers only the subset of prefixes originated by that AS. This method yields a bilateral metric of transit influence of transit provider AS_t on a single origin AS_o, in country C, or the AS-Level Transit Influence $ATI(AS_t, AS_o, C) \in [0,1]$.

ATI depends on the fraction of the origin AS' IPs that the transit AS services. Higher ATI values signal that the transit AS is more often present on the path towards the origin, and therefore has greater potential to observe or selectively tamper with its traffic.

[2] The exception was a Chilean critical AS, which we reclassified in the correct category.

Recall that $CTI_M(AS,C) \in [0,1]$ is computed using a set of BGP monitors M as

$$\sum_{m \in M} \left(\frac{w(m)}{|M|} \cdot \sum_{p|\text{onpath}(AS,m,p)} \left(\frac{a(p,C)}{A(C)} \cdot \frac{1}{d(AS,m,p)} \right) \right),$$

where $w(m)$ is monitor m's weight among the set of monitors; onpath(AS,m,p) is true if AS is present on a preferred path observed by monitor m to a prefix p, and m is not contained within AS itself; $a(p,C)$ is the number of addresses in prefix p geolocated to country C; $A(C)$ is the total number of IP addresses geolocated to country C; and $d(AS,p,m)$ is the number of AS-level hops between AS and prefix p as viewed by monitor m.

We similarly define $ATI_M(AS_t,AS_o,C) \in [0,1]$, the ATI of transit network AS_t on critical-sector origin network AS_o in country C, using a set of BGP monitors M as

$$\sum_{m \in M} \left(\frac{w(m)}{|M|} \cdot \sum_{p \in AS_o|\text{onpath}(AS_t,m,p)} \left(\frac{a(p,C)}{A(AS_o,C)} \cdot \frac{1}{d(AS_t,m,p)} \right) \right),$$

where $p \in AS_o$ means p is originated by AS_o and $A(AS_o,C)$ is the total number of IP addresses geolocated to country C that are originated by AS_o.

6.2.3 Weighted AS-level Transit Influence

In order to measure the influence of transit ASes on entire sectors composed of multiple ASes, we need a quantity that reflects the composite ATI of each AS_t on the origin ASes of the sector. To this end, we compute a weighted ATI (W-ATI). W-ATI is the weighted average of all the ASes' ATIs in a sector, where each origin AS is weighted by the amount of address space it originates:

$$\sum_{AS_o \in S} ATI(AS_t,AS_o,C) \cdot O(AS_o)$$

where $O(AS_o)$ is the fraction of the sector's total address space originated by AS_o in critical sector S.

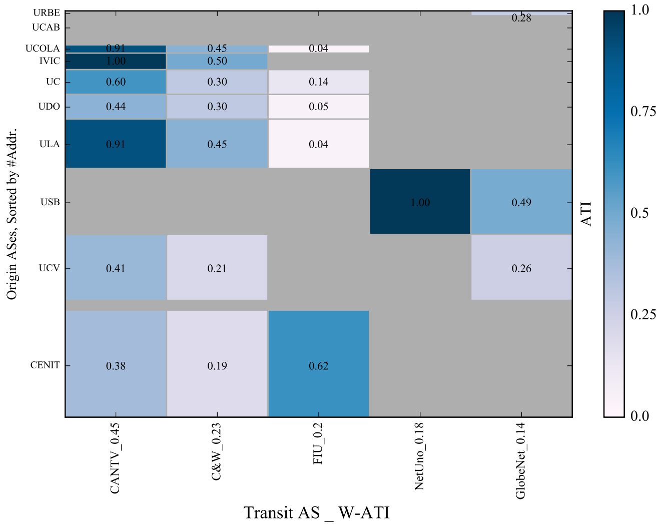

Figure 6.1. Heatmap of ATI values, academic sector in Venezuela. A transit network's W-ATI is the sum of the products between each cell's ATI (in its column) multiplied by the row's width (addresses originated).

We illustrate an application of W-ATI by focusing on the 12 ASes in the academic sector in Venezuela (chosen as it has close to the average number of ASes in critical sectors in our sample). Fig. 6.1 shows this illustration for the top 5 transit ASes in the sector. CANTV is influential on most ASes, with ATI ≥ 0.4 on all but five ASes, which causes the transit network to become the sector's leader with a W-ATI of 0.45. Note that the W-ATI of CANTV approaches its ATI on the biggest ASes, CENIT and UCV (ATI = 0.38 and 0.41), rather than the transit network's ATI on the thinner rows above them, *e.g.,* ULA and UC (ATI = 0.91 and 0.60).

6.3 Results

We present an overview of our findings first, and then describe each country's results in more detail. We show the results of our analysis in Fig. 6.2 for all three critical sectors[3] in Chile, Bolivia and Venezuela. There, we also show as a baseline the country-wide ATI, or the ATI of the transit network over all ASes in the country. Critical sectors exhibit varying degrees of

[3] When two ASes with the same name appear in the top 5, we show in the chart only the top such AS by W-ATI. This happens for Telmex (Chile-education) and Lumen (Chile-education and Venezuela-finance).

exposure to the top transit AS, with their W-ATI ranging from 0.22 for Venezuela-finance, to 0.89 for Bolivia-education and academic In all sectors, the top-ranked transit AS overperforms their countrywide W-ATI, meaning that critical-sector ASes are particularly exposed to that transit network. Further, in all but three sectors (Venezuela-Finance, Venezuela-education and academic, Chile-government and utilities), the top transit AS overperforms the countrywide baseline by at least a factor of 5. This may signal a concerning topological feature, as the specialization of these transit ASes might hide their influence from consumer-advocacy groups and regulators.

We find two salient groups of transit ASes: *(i)* large and foreign telecommunications companies in all three countries (an AS is "foreign" if less than a third of its addresses are in the country), *(ii)* state-owned domestic operators in Bolivia and Venezuela [48, 39]. These two groups are also frequently owners of subsea or long-range terrestrial cables in the region. We describe the ATI of these two groups of ASes and other influential transit networks in the following paragraphs.

Chile, which has spent considerable efforts [31, 79, 40] building a domestic peering infrastructure, has a less exposed ecosystem in the academic sector, where the top AS has a W-ATI of 0.37. There, Lumen has a W-ATI only 20% higher than its baseline countrywide. However, two other sectors appear highly exposed: the country's financial sector, as well as its government and utilities. The former is dominated (W-ATI = 0.78) by AS27986-ENTEL[4], while the latter is primarily serviced (W-ATI = 0.81) by AS7004-CTC. CTC overperforms its countrywide baseline in the government and utilities sector by a factor of 2.3, whereas ENTEL-Chile does so in the financial sector by a factor of 2.9. Both companies were formerly state-owned and privatized in the 1980s and early 90s [53]. CTC is a subsidiary of Telefonica [81], the large telecom conglomerate based in Spain, Chile's former colonial ruler. Telefonica itself (AS12956) is also highly influential in the gov. & utils. sector (ranked second).

[4] The Spanish-language acronym means "National Telecommunications Company", which explains why two unrelated companies in Chile and Bolivia have the same name. We differentiate them in our prose by using all caps for the Chilean AS.

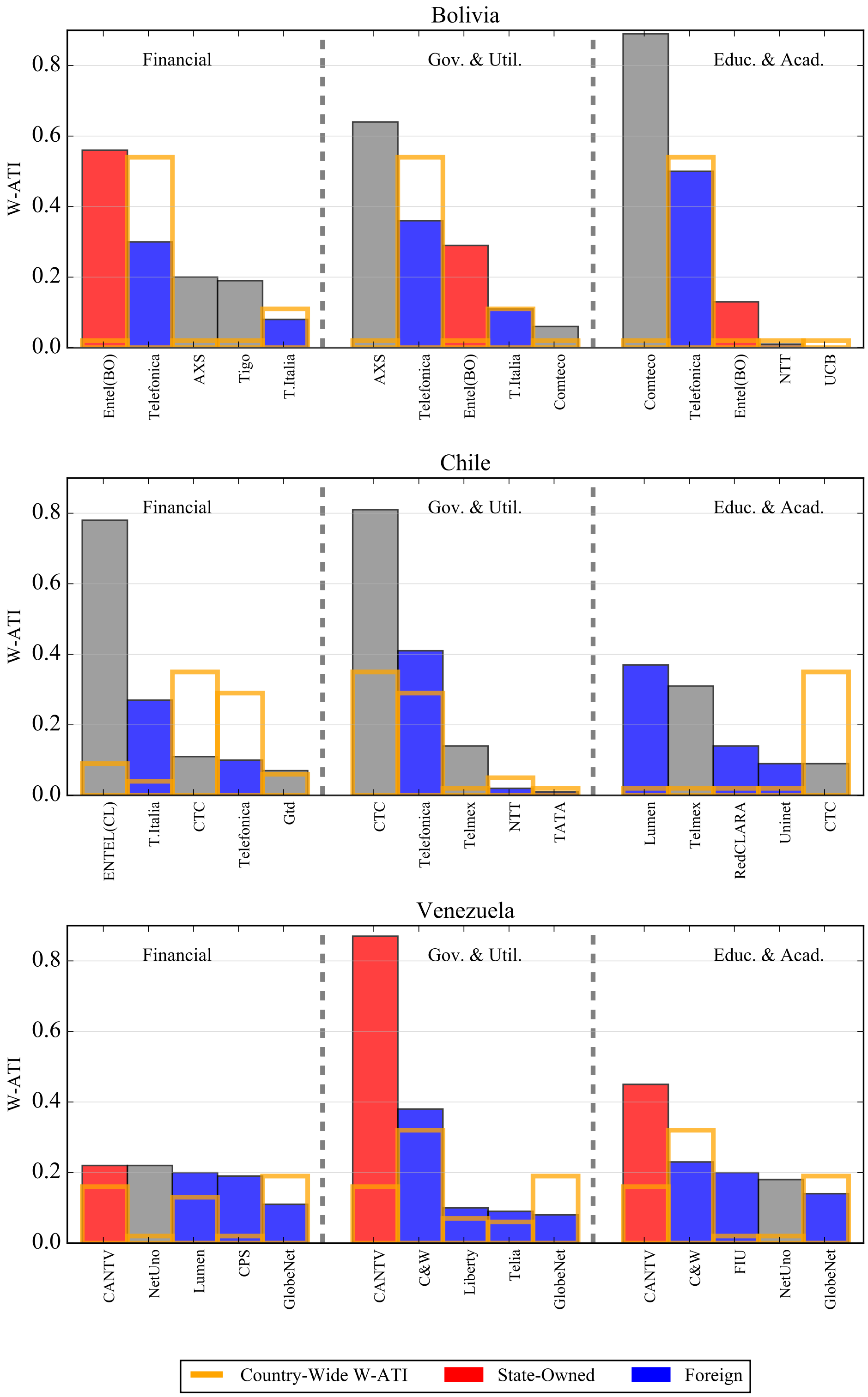

Figure 6.2. Weighted ATI of top ASes in each critical sector. Critical sectors often are dominated by 1-2 transit ASes, and more exposed to those ASes than the country as a whole. This exposure is often to (colored) ASes that are either state-owned or foreign.

We speculate that there are two geographic factors impeding Chile's domestic providers from deploying their own long-range infrastructure, and therefore relying on foreign conglomerates: the country's distance to both the US and EU, as well as its location on the Andes mountain range which complicates reaching the rich interconnection mesh on the Atlantic [40, 63, 97]. As a result, traffic towards the country's critical sectors may be overly exposed to large telecommunications providers with the financial means to invest in long-range subsea connectivity, despite the country's mature domestic peering infrastructure [31, 40, 79]. There is evidence of this trend continuing, such as the projected cable deployment with Telefonica and America Movil (AM) as co-onwers of *AM-Telxius West Coast* [63]. AM is the parent company of Telmex [13], which is ranked in the top 3 of Chile's financial and government sectors (Fig. 6.2).

Bolivia has a concentrated transit ecosystem in all critical sectors, with a single AS_t providing transit for over half of the sectors' IP addresses. W-ATI for the top AS in critical sectors ranges between 0.56-0.89, which is in all cases higher than the W-ATI of the top AS countrywide, Telefonica (0.54). We note that the top AS in Bolivian critical sectors is, in all cases, marginal countrywide (W-ATI ≤ 0.01).

The four ASes we identified in the top 2 of each Bolivian sector have been previously named by the International Telecommunication Union—an agency of the United Nations–as the "main" operators in the country [100]: 6568-Entel, 12956-Telefonica, 27839-COMTECO and 26210-AXS[5] (pronounced "Access"). Given that the top ASes are different in each sector, exposure is distributed across various networks, and not just concentrated on the state-owned provider (unlike in Venezuela, discussed later on).

State-owned Entel, in Bolivia, recently deployed terrestrial fiber to reach the Pacific ocean and a subsea cable along Peru's coast [3, 4, 9, 97] (one of the operator databases lists a different owner for the subsea portion [63]). According to a consulting firm operating in the region [3], an executive of the state-owned company stated that "Bolivia will have absolute

[5] With a "concentrated" presence in the country's three largest Departments [18], including the capital La Paz, where many of the ASes in the government and utilities sector—AXS' strongest sector—may be based.

sovereignty in its [data] exit to the Pacific", signaling that the government is indeed concerned with reliance on foreign transit networks.

Venezuela has a critical-sector transit ecosystem dominated by the country's state-owned provider, CANTV. CANTV is influential over the country as a whole (countrywide W-ATI = 0.16, ranked third), but it overperforms the countrywide baseline on critical sectors by a factor of 1.4–5.4. The state-owned company is ranked first in all three sectors.

CANTV has a W-ATI of 0.9 over the Government & Utilities Sector. This overperformance may give the national government capabilities to observe or selectively tamper with sensitive traffic towards third-parties, as it reportedly does to traffic towards its own subscribers [5]. Transit is an additional lever for the company to enact content filters that may be overlooked by Internet rights groups (*e.g.,* Freedom House [14]). Non-state-owned providers in Venezuela, such as C&W, NetUno and Lumen, mostly lag behind in critical sectors.

6.4 Traceroutes

In this section we describe the results of a traceroute campaign we conducted in an attempt to confirm some of our BGP-based findings in the previous section. The purpose of our active campaign is to measure the AS-Level path taken by probes sent from outside these countries, with 100 probes from RIPE Atlas [84] selected at random. This campaign is identical in design to that described in Ch. 3, except for two important differences: first, we select a set of probes for each country, rather than a single set for all countries; second, the campaign lasts one week starting Aug. 6th, 2020.

Our traceroute targets are critical organizations in Venezuela and Bolivia[6]. The output of this campaign includes 9,815 traceroutes successfully reaching 10 critical ASes. We also use traceroutes launched by other users—or RIPE itself—during the same timeframe, as long as they terminate at one of the origin ASes in the critical sectors (including in Chile), and are launched from a probe located outside the target country. This larger set includes 12,450 traceroutes

[6] At campaign-launch time we had not yet finished classifying the numerous Chilean ASes.

successfully reaching 16 ASes (six of them in Chile). Given that the traceroute campaign reached only a fraction of critical ASes, we limit our use of traceroutes to the confirmation of select ATI-inferred results in the following sections. We translate IP-level traceroutes to AS-level paths, which is necessary to compare the output of the active campaign with BGP inferences, by applying state-of-the-art tool *BdrmapIT* [76].

While much progress has been made recently [73, 76], the issue of translating traceroutes to AS-level paths is not fully resolved. This makes a comprehensive validation of BGP-inferred findings using active measurement campaigns challenging. We find evidence of this challenge in our active campaign and analysis, as we do not always observe the same ASes we inferred using BGP paths, and when we do observe them they sometimes appear more (or less) frequently than the BGP analysis would suggest. This may be due to a number of factors: *(i)* Traceroute vantage points are (as observed BGP paths) biased towards nodes in certain regions and therefore the visible IP-level paths also are biased. *(ii)* Our campaign was insufficient in scope—duration and scale—resulting in a suboptimal probing scheme that fails to reveal all available AS-level paths into an AS, as well as paths towards a larger set of origin ASes in each sector. *(iii)* IP-to-AS translation is missing topological information and therefore produces sometimes incorrect inferences. *(iv)* The BGP inference was inaccurate in the first place, *e.g.,* there was an incorrectly inferred provider-customer relationship.

In the remainder of this section, we present our findings from analyzing these traceroutes. We partially confirmed Telefonica's prominent role in Bolivia as it frequently appears on traceroute paths towards three critical ASes: two in government and utilities, AS61458 (93% of traces) and AS265779 (78%); and one in academia, AS27828 (86%). AXS also appears on path 83% and 100% of the time towards government ASes 265779 and 61458. Traceroutes did not confirm our BGP-derived findings with regards to Entel and COMTECO. The former only appears on path towards educ. & acad. AS27828 5% of the time. The latter appears between 1-7% of the time on path towards two critical ASes (gov. & util. AS265779, educ. & acad. AS27828).

With regards to Venezuela, CANTV appears on traceroutes towards AS19192-UCV 39% of the time. We did not observe CANTV on traceroutes towards the other academic ASes. We also observed AS23520-C&W on traceroutes between 81–100% of the time towards five of seven origin ASes on which it has ATI (ranging between 0.2–0.5). Also in Venezuela, NetUno appears on traceroutes towards USB 99% of the time (ATI = 1.0), confirming the importance of the provider-customer relationship.

In Chile, our analysis of RIPE traceroutes confirms CTC's frequent appearance towards six critical ASes: two in government and utilities, AS52226 (CTC appears on 80% of traceroutes) and AS17147 (62%); three academic ASes, AS11340 (35%), AS16742 (29%) and AS23140 (3%); and one in finance, AS16780 (84%). We also observed AS3549-Lumen on traceroutes towards three academic ASes (Lumen is ranked first in the sector): AS16742 (64%), AS23140 (84%) and AS11340 (1%). We did not observe AS27986-ENTEL on any traceroutes towards Chilean ASes in finance, which is not entirely surprising since our campaign did not actively target ASes in Chile.

6.5 Limitations

In this section we describe the major limitations and lessons learned from developing a methodology to identify the transit networks serving critical-sector ASes.

AS Classification: Scalability. We note the open challenge of the scalability of this work. Manual classification is not feasible for the entire Internet (tens of thousands of ASes). We explored crowdsourcing as a potential avenue for scaling our method. We used Amazon M-Turk [1], and quickly realized that the nature of labeling our dataset is too technical for most of the platform's workers. The output of the initial experiment was a set of wildly inaccurate labels, and an incomplete submission of labels.

Our second M-Turk experiment was more selective in recruiting workers, requiring over 1,000 previously-approved tasks and an approval rate over 99%. In this case, one worker did submit a complete and reasonably accurate (80% correct) set of labels. (We obtain ground truth of AS classification directly from operators, Sec. 6.2). The worker mislabeled two ASes belonging as a telecommunications service provider instead of an educational institution, which is a reasonable label using the organization's *index.html* page [8].

Despite the more encouraging results of this second experiment, two challenges remain: first, recruiting is an issue, as we only received a submission from a single worker. Second, by setting up this experiment we realized that the total cost might be prohibitive, at about US$ 6 per AS (we used an effective compensation around the U.S. federal minimum wage). An open challenge is the exploration of machine learning techniques to increase the scalability of our method, such as automated content extraction and labeling, rather than relying on crowdsourced avenues.

AS Classification: Non-AS Organizations. We identify ASes operated by critical organizations, which by definition will exclude those that do not own an AS. There are also ASes in a single country belonging to other sectors that arguably should be classified as critical, including airlines (*e.g.,* AS27746-LAN CHILE). We have excluded these sectors in order to keep the categories consistent across countries.

Study of Interconnection: Traffic, Purpose of ASes, and Topology Evolution. We note three limitations to our W-ATI approach. First, we do not study (nor have access to) traffic volumes, so we restrict our analysis to metrics based on the number of IP addresses served by a transit provider. Second, we do not know for what purpose these ASes are actually used by the organizations that registered them. As a consequence, we are unable to determine the types of traffic that these ASes would receive from other countries. Third, our approach presents a point-in-time analysis of interconnection. AS interconnection evolves over time [45], some of the findings in this chapter may no longer hold in the future. For example, Telefonica is in the process of selling its South American subsidiaries this year [12, 15]. As a result, some of our findings based on data from Mar. 2020 and Jan. 2021 may no longer apply.

Chapter 6, in part is currently being prepared for submission for publication of the material. Gamero-Garrido, Alexander; Carisimo, Esteban; Snoeren, Alex C.; Dainotti, Alberto. The dissertation author was the primary investigator and author of this material.

Chapter 7

Conclusion

In this dissertation, we present a comprehensive set of tools to evaluate the exposure to observation and selective tampering of Internet traffic flowing towards specific countries, as well as towards specific organizations and critical sectors within those countries. We confirm this dissertation's thesis by identifying 75 transit-dominant countries, the majority of which are significantly exposed to observation and tampering by a single transit network (median exposure is 34% of the country's IP addresses). Our tools produce inferences on three distinct phenomena: *(i)* the likelihood that a country's traffic is delivered primarily by transit providers (Ch. 6), *(ii)* the fraction of a country's IP addresses that is exposed to observation and tampering by a specific transit provider (Chs. 4 and 5), *(iii)* the fraction of IPs belonging to a specific AS or sector that is similarly exposed (Ch. 6).

Our method to study *(i)* transit dominance consists of an evaluation of each country's AS-level topology, which we approach by analyzing existing interconnection datasets and our own large-scale traceroute campaign. Our approach to study *(ii)* and *(iii)* has distinct metrics at the country level (CTI), at the AS level (ATI), and at the sector level (W-ATI); all three metrics quantify transit influence, or the capabilities of a particular transit provider to observe or tamper with Internet traffic.

7.1 Transit Dominance

We identify 75 countries where transit-provider customer relationships are still the dominant inbound modality for international traffic, *i.e.,* where international peering is uncommon. We identify these nations using analyses of peering databases, our own large-scale measurement campaign, and validation with operators. Countries in all five RIRs are represented in the 75 nations we identify, including some high-income countries such as South Korea, though the majority are in developing regions. This group is composed of both large countries (*e.g.,* India, Turkey, Egypt) and smaller nations (*e.g.,* Guyana, Tuvalu).

7.2 Country-Level Transit Influence

We design the CTI metric to quantify the exposure of entire nations' traffic to observation and tampering by specific networks. CTI aims to overcome several challenges with making transit influence inferences using BGP data. We apply CTI in these 75 countries, the majority of them in either Africa or Latin America, to identify the most influential transit ASes. By studying the CTI values for the top ASes in each country, we find that 32 nations have transit ecosystems that render them particularly exposed, with traffic destined to over 40% of their IP addresses served by single AS. In the nations where we are able to validate our findings with in-country operators, we obtain 83% accuracy on average. In the countries we examine, CTI reveals two classes of networks that play a particularly prominent role: submarine cable operators and state-owned ASes.

7.3 AS-Level Transit Influence (ATI) and Weighted ATI

We study the transit ASes serving each AS in a country's critical sectors (ATI) as well as the transit ASes serving the sector as a whole (W-ATI). In order to apply this methodology, we first identify 83 ASes in critical sectors in three South American countries: Bolivia, Chile and Venezuela. (We partially validate our classification methodology with operators, with a resulting accuracy of 97%.) We find that a small number of transit ASes that exert out-sized influence in several sectors, including finance, utilities, and education. In four cases, just one AS carries traffic towards 80–90% of the IP addresses in a critical sector in a particular country. Our study reveals two kinds of transit networks in particular: state-owned companies and large, foreign providers.

7.4 Future Directions and Final Thoughts

We conclude this dissertation by presenting a set of research directions that would further illuminate the questions we explore. We also describe lessons learned from our research.

7.4.1 Future Directions

We would like to develop measurement and analysis techniques that can be applied to study the exposure of countries that are not primarily served by transit providers, but rather by a dense mesh of bilateral and multilateral peering agreements. This group includes the majority of industrialized nations such as the United States, Japan and Germany. Our preliminary exploration of this question for a large (commercial) access network in the US revealed potential peering agreements with hundreds of international and foreign ASes. A comprehensive study of this topic, particularly one that relies on an expanded measurement footprint such as by leveraging crowdsourced traceroutes, might reveal a set of interconnections of an even larger scale. We also plan on applying our ATI and W-ATI methodology to ASes and critical sectors in other countries.

While our analysis focuses on the country level, our tools may be applied to jurisdictions of any size, given appropriate validation of geolocation inputs. This finer granularity may enable the quantification of social inequities in exposure to observation and tampering, for instance across states or regions in the same country, or even across specific groups. Future work in this area may reveal whether demographic variables—such as race or ethnicity—correlate with the exposure of Internet traffic flowing towards users in different groups. Previous work has shown regional and ethnic disparities in access to a high-quality access network [20, 91, 103], so differences in routing configuration may also exist.

We propose three additional lines of inquiry: *(i)* studying the evolution of transit influence over time, which may reveal the impact of regulatory, economic or commercial forces on the wide-area network configuration of a country, *(ii)* quantifying the sensitivity of transit influence inferences to changes in BGP monitor sets (both additions and removals), which may assist in future inquiries researching countries with more complex topologies (where it is more challenging to study country-level exposure), and *(iii)* measuring the transit influence of organizations composed of multiple ASes, such as multinational telecommunications conglomerates (*e.g.,* those identified by [36]).

7.4.2 Final Thoughts

We are able to validate our findings from both stages with in-country network operators at 123 ASes in 19 countries who confirm that our results are consistent with their understanding of their countries' networks. The scale of this validation effort suggests that operators are often willing to reveal valuable information to academic researchers, though they almost always prefer to do so anonymously. This willingness creates opportunities for computer scientists that engage in "soft" inquiries such as emails, phone interviews, and large-scale surveys.

While this dissertation does not delve into the underlying causes for the disparities in exposure that we observe across countries and across organizations or sectors, future work in this area may systematically reveal the impact of public policy and international relations on such exposure. We find anecdotal evidence of this impact with our own work, such as the dominance of state-owned companies in countries ruled by authoritarian regimes (*e.g.,* Cuba, Myanmar, Venezuela), and the relatively low exposure of countries that have prioritized the development of their peering infrastructure (*e.g.,* Chile, Bangladesh). These anecdotes suggest that the routing ecosystem of many countries might be influenced by non-technical factors that have implications for digital rights, government surveillance, and international conflicts.

Bibliography

[1] Amazon mechanical turk. https://www.mturk.com/. (Accessed in May 2021).

[2] Betweenness Centrality - an overview — sciencedirect topics. https://www.sciencedirect.com/topics/computer-science/betweenness-centrality. (Accessed in May 2021).

[3] Bolivia inaugura conexion propia de red de fibra optica al pacifico - bnamericas. https://www.bnamericas.com/es/noticias/bolivia-inaugura-conexion-propia-de-red-de-fibra-optica-en-el-pacifico. (Accessed in May 2021).

[4] Dentro de 60 dias, Entel empalmara su fibra optica submarina con Bolivia - Diario Pagina Siete. https://www.paginasiete.bo/economia/2020/1/12/dentro-de-60-dias-entel-empalmara-su-fibra-optica-submarina-con-bolivia-243205.html. (Accessed in May 2021).

[5] Global Internet Censorship — SpringerLink. https://vpn-2.ucsd.edu/+CSCO+0075676763663A2F2F797661782E66636576617472652E70627A++/chapter/10.1007/978-94-007-1245-4_3. (Accessed on 05/22/2021).

[6] Herfindahl-Hirschman Index. https://www.justice.gov/atr/herfindahl-hirschman-index. (Accessed in May 2021).

[7] Internet users - The World Factbook. https://www.cia.gov/the-world-factbook/field/internet-users/country-comparison. (Accessed in May 2021).

[8] Nic chile, somos el punto cl - nic chile. https://www.nic.cl/index.html. (Accessed in May 2021).

[9] Paraguay expreso interes en comprar servicios del cable submarino de fibra optica de Entel. https://www.ofep.gob.bo/index.php/comunicacion/noticiasplataforma/item/1263-paraguay-expreso-interes-en-comprar-servicios-del-cable-submarino-de-fibra-optica-de-entel. (Accessed in May 2021).

[10] Spam stats for AS8048 CANTV. https://cleantalk.org/blacklists/as8048. (Accessed on 05/04/2021).

[11] Synapsis Soluciones y Servicios It Ltda - Company Profile and News - Bloomberg Markets. https://www.bloomberg.com/profile/company/6982081Z:CI. (Accessed on 05/04/2021).

[12] Telefonica bautiza como Telxius su nueva filial de infraestructuras — empresas — cinco dias. https://cincodias.elpais.com/cincodias/2016/02/10/empresas/1455103059_743536.html. (Accessed in May 2021).

[13] Telmex reports revenue, net profit drops in 2q 2011. https://www.commsupdate.com/articles/2011/07/20/telmex-reports-revenue-net-profit-drops-in-2q-2011/. (Accessed in May 2021).

[14] Venezuela: Freedom on the net 2019 country report — freedom house. https://freedomhouse.org/country/venezuela/freedom-net/2019. (Accessed in May 2021).

[15] Why Telefonica is slimming down in South America. https://www.ft.com/content/3065e7ec-11d9-11ea-a7e6-62bf4f9e548a. (Accessed on 05/18/2021).

[16] Democracy Index 2017: Free speech under attack. https://www.eiu.com/public/topical_report.aspx?campaignid=DemocracyIndex2017, 2017.

[17] Nauru. Measuring the Information Society Report Volume 2. ICT Country Profiles. https://www.itu.int/en/ITU-D/LDCs/Documents/2017/Country%20Profiles/Country%20Profile_Nauru.pdf, 2017.

[18] ITU - Challenges and Opportunities on the matter of Bolivia's Connectivity (Spanish). https://www.itu.int/en/ITU-D/LDCs/Documents/2018/Publication/D012A0000DE3301PDFS.pdf, 2018.

[19] CAIDA AS-Relationships. http://data.caida.org/datasets/as-relationships/, 2019.

[20] Digital gap between rural and nonrural America persists. https://www.pewresearch.org/fact-tank/2019/05/31/digital-gap-between-rural-and-nonrural-america-persists/, 2019.

[21] RIPE Routing Information Service (RIS). https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris, 2019.

[22] RIR Delegation Files. https://ftp.ripe.net/pub/stats/ripencc/, 2019.

[23] RouteViews. http://www.routeviews.org/routeviews/, 2019.

[24] CAIDA's AS-Rank. http://as-rank.caida.org/, 2020.

[25] Internet Health Report API. Hegemony. https://ihr.iijlab.net/ihr/api/hegemony/, 2020.

[26] Netacuity. http://info.digitalelement.com/, 2020.

[27] traceroute(8) Linux manual page. https://man7.org/linux/man-pages/man8/traceroute.8.html, 2020.

[28] CAIDA's AS-Rank. http://as-rank.caida.org/, 2021.

[29] EC Agbaraji, FK Opara, and MI Aririguzo. Ipv6 deployment status, the situation in Africa and way out. *International Journal of Advances in Engineering & Technology*, 2(1):315, 2012.

[30] Simurgh Aryan, Homa Aryan, and J. Alex Halderman. Internet censorship in Iran: A first look. In *3rd USENIX Workshop on Free and Open Communications on the Internet (FOCI 13)*, Washington, D.C., August 2013. USENIX Association.

[31] Biblioteca del Congreso Nacional de Chile. FIJA PROCEDIMIENTO Y PLAZO PARA ESTABLECER Y ACEPTAR CONEXIONES ENTRE ISP. leychile.cl/Navegar?idNorma=146170, 2020.

[32] Zachary S. Bischof, Romain Fontugne, and Fabián E. Bustamante. Untangling the world-wide mesh of undersea cables. In *Proceedings of the 17th ACM Workshop on Hot Topics in Networks*, HotNets '18, page 7884, New York, NY, USA, 2018. Association for Computing Machinery.

[33] Zachary S. Bischof, John P. Rula, and Fabián E. Bustamante. In and out of Cuba: Characterizing Cuba's connectivity. In *Proceedings of the 2015 Internet Measurement Conference*, IMC '15, page 487493, New York, NY, USA, 2015. Association for Computing Machinery.

[34] Xue Cai, M. Rey, CA xuecai, J. Heidemann, CA johnh, and Walter Willinger Niksun. A holistic framework for bridging physical threats to user QoE USC / ISI Technical Report. 2013.

[35] CAIDA. CAIDA Internet eXchange Points (IXPs) Dataset. https://www.caida.org/data/ixps/, 2020.

[36] CAIDA. Mapping Autonomous Systems to Organizations: CAIDA's Inference Methodology. https://www.caida.org/research/topology/as2org/, 2020.

[37] CAIDA. Routeviews Prefix-to-AS mappings (pfx2as) for IPv4 and IPv6. http://data.caida.org/datasets/routing/routeviews-prefix2as/, 2020.

[38] CAIDA - PeeringDB. CAIDA's PeeringDB dumps. http://data.caida.org/datasets/peeringdb/, 2020.

[39] CANTV. La empresa. https://www.cantv.com.ve/la-empresa/, 2021.

[40] Esteban Carisimo, Julián M Del Fiore, Diego Dujovne, Cristel Pelsser, and J Ignacio Alvarez-Hamelin. A first look at the Latin American IXPs. *ACM SIGCOMM Computer Communication Review*, 50(1):18–24, 2020.

[41] Esteban Carisimo, Alexander Gamero-Garrido, Alex C. Snoeren, and Alberto Dainotti. Identifying ASes of State-Owned Internet Operators. In *(to appear at) ACM SIGCOMM Conference on Internet Measurement*, IMC '21, New York, NY, USA, 2021. ACM.

[42] Edmond W. W. Chan, Xiapu Luo, Waiting W. T. Fok, Weichao Li, and Rocky K. C. Chang. Non-cooperative Diagnosis of Submarine Cable Faults. In *Passive and Active Measurement*, 2011.

[43] Coral Sea Cable System. Coral Sea Cable System. https://www.coralseacablesystem.com.au/about/, 2020.

[44] Dainotti, A. and Benson, K. and King, A. and claffy, k. and Glatz, E. and Dimitropoulos, X. and Richter, P. and Finamore, A. and Snoeren, A. Lost in Space: Improving Inference of IPv4 Address Space Utilization, Oct 2014.

[45] Amogh Dhamdhere and Constantine Dovrolis. Ten Years in the Evolution of the Internet Ecosystem. In *Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement*, IMC '08, pages 183–196, New York, NY, USA, 2008. ACM.

[46] X. Dimitropoulos, D. Krioukov, G. Riley, and k. claffy. Revealing the Autonomous System Taxonomy: The Machine Learning Approach. In *Passive and Active Network Measurement Workshop (PAM)*, Adelaide, Australia, Mar 2006. PAM 2006.

[47] 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, New York, NY, USA, 2018. Association for Computing Machinery.

[48] ENTEL. 13 aniversario de la nacionalizacion. https://institucional.entel.bo/inicio3.0/files/Presentacion_institucional/Brochure%20Entel%202021.pdf, 2021.

[49] R. Fanou, B. Huffaker, R. Mok, and k. claffy. Unintended consequences: Effects of submarine cable deployment on Internet routing. In *Passive and Active Measurement Conference (PAM)*, Mar 2020.

[50] R. Fanou, F. Valera, P. Francois, and A. Dhamdhere. Reshaping the African Internet: from scattered islands to a connected continent. *Computer Communications*, 113:25 – 42, 2017.

[51] Roderick Fanou, Pierre Francois, and Emile Aben. On The Diversity of Interdomain Routing in Africa. In *PAM*, 2015.

[52] Fontugne, Romain and Shah, Anant and Aben, Emile. The (thin) Bridges of AS Connectivity: Measuring Dependency using AS Hegemony. In *Passive and Active Measurement Conference (PAM)*, 2018.

[53] Organization for Economic Cooperation and Development (OECD). Regulation, competition and privatisation. https://www.ft.com/content/3065e7ec-11d9-11ea-a7e6-62bf4f9e548a, 1998. (Accessed in May 2021).

[54] Freedom House. Freedom of the Network. https://freedomhouse.org/report-types/freedom-net, 2019.

[55] Hernan Galperin. Connectivity in Latin America and the Caribbean: The role of Internet exchange points. 2013.

[56] Gustavo Garcia. Why Miami is Latin America's center of interconnection - interconnections - the Equinix blog. https://blog.equinix.com/blog/2018/05/01/why-miami-is-latin-americas-center-of-interconnection/, May 2018. (Accessed on 02/03/2021).

[57] Gharaibeh, Manaf and Shah, Anant and Huffaker, Bradley and Zhang, Han and Ensafi, Roya and Papadopoulos, Christos. A look at router geolocation in public and commercial databases. In *ACM Internet Measurement Conference (IMC)*, 2017.

[58] Christopher Giles. Myanmar coup: How the military disrupted the Internet - BBC News. https://www.bbc.com/news/world-asia-55889565. (Accessed on 02/05/2021).

[59] Vasileios Giotsas, Matthew Luckie, Bradley Huffaker, and kc claffy. Inferring complex AS relationships. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, IMC '14, pages 23—30, New York, NY, USA, 2014. Association for Computing Machinery.

[60] Arpit Gupta, Matt Calder, Nick Feamster, Marshini Chetty, Enrico Calandro, and Ethan Katz-Bassett. Peering at the Internet's frontier: A first look at ISP interconnectivity in Africa. In Michalis Faloutsos and Aleksandar Kuzmanovic, editors, *Passive and Active Measurement*, pages 204–213, Cham, 2014. Springer International Publishing.

[61] GuyanaJournal. Guyana-Suriname Border Issue - from the 1960s to 2004. http://www.guyanajournal.com/guyana_suriname_border.html, 2006.

[62] Freedom House. Venezuela Freedom of The Net 2015. https://freedomhouse.org/sites/default/files/resources/FOTN%202015_Venezuela.pdf, 2015.

[63] Infrapedia. Infrapedia. https://www.infrapedia.com/app, 2020.

[64] International Cable Protection Committee. Shark Bites and Cable Faults. https://www.iscpc.org/documents/?id=1489, 2020.

[65] Josh Karlin, Stephanie Forrest, and Jennifer Rexford. Nation-State Routing: Censorship, Wiretapping, and BGP. In *CoRR*, 2009. http://arxiv.org/abs/cs/0608082, 2009.

[66] kc claffy, Marina Fomenkov and Bradley Huffaker. Geocompare: a comparison of public and commercial geolocation databases - Technical Report. Cooperative Association for Internet Data Analysis (CAIDA), May 2011.

[67] D. Kiedanski and E. Grampín. Understanding Latin America IPv6 connectivity: A preliminary exploration. In *2017 36th International Conference of the Chilean Computer Science Society (SCCC)*, pages 1–6, 2017.

[68] Kirtus G. Leyba, Benjamin Edwards, Cynthia Freeman, Jedidiah R. Crandall, and Stephanie Forrest. Borders and Gateways: Measuring and Analyzing National AS Chokepoints. In *COMPASS*, 2019.

[69] Liquid Telecom. Network. https://www.liquidtelecom.com/about-us/our_network, 2020.

[70] Shucheng Liu, Zachary S. Bischof, Ishaan Madan, Peter K. Chan, and Fabián E. Bustamante. Out of sight, not out of mind: A user-view on the criticality of the submarine cable network. In *Proceedings of the ACM Internet Measurement Conference*, IMC '20, pages 194–200, New York, NY, USA, 2020. Association for Computing Machinery.

[71] Ioana Livadariu, Ahmed Elmokashfi, and Amogh Dhamdhere. Measuring IPv6 Adoption in Africa. In Victor Odumuyiwa, Ojo Adegboyega, and Charles Uwadia, editors, *e-Infrastructure and e-Services for Developing Countries*, pages 345–351, Cham, 2018. Springer International Publishing.

[72] A. Lodhi, N. Larson, A. Dhamdhere, C. Dovrolis, and k. claffy. Using PeeringDB to Understand the Peering Ecosystem. In *ACM SIGCOMM Computer Communication Review (CCR)*, 2014.

[73] M. Luckie, A. Dhamdhere, B. Huffaker, D. Clark, and k. claffy. bdrmap: Inference of Borders Between IP Networks. In *ACM Internet Measurement Conference (IMC)*, pages 381–396, Nov 2016.

[74] Matthew Luckie, Bradley Huffaker, Amogh Dhamdhere, Vasileios Giotsas, and Kc Claffy. AS relationships, customer cones, and validation. In *ACM Internet Measurement Conference (IMC)*, 2013.

[75] Josephine Lukito. Coordinating a Multi-Platform Disinformation Campaign: Internet Research Agency Activity on Three U.S. Social Media Platforms, 2015 to 2017. *Political Communication*, 37(2):238–255, 2020.

[76] A. Marder, M. Luckie, A. Dhamdhere, B. Huffaker, J. Smith, and k. claffy. Pushing the Boundaries with bdrmapIT: Mapping Router Ownership at Internet Scale. In *Internet Measurement Conference (IMC)*, pages 56–69, Nov 2018.

[77] Babacar Mbaye, Assane Gueye, Desire Banse, and Alassane Diop. Africa's online access: What data is getting accessed and where it is hosted? In Ghada Bassioni, Cheikh M.F. Kebe, Assane Gueye, and Ababacar Ndiaye, editors, *Innovations and Interdisciplinary Solutions for Underserved Areas*, pages 50–61, Cham, 2019. Springer International Publishing.

[78] Ministry of Posts and Telecommunications. INTERNATIONAL LONG DISTANCE TELECOMMUNICATIONS SERVICES (ILDTS) POLICY, 2010. http://www.btrc.gov.bd/sites/default/files/ildts_policy_2010_english_0.pdf, 2010.

[79] NAP Chile. Acerca de NAP. http://www.nap.cl/f_acerca.html, 2020.

[80] PeeringDB. https://www.peeringdb.com, 2019.

[81] PeeringDB. PeeringDB record for CTC Transmisiones Regionales S.A. https://www.peeringdb.com/net/11934, 2021.

[82] Ingmar Poese, Steve Uhlig, Mohamed Ali Kaafar, Benoit Donnet, and Bamba Gueye. IP geolocation databases: Unreliable? *SIGCOMM Comput. Commun. Rev.*, 41(2):5356, April 2011.

[83] Philipp Richter, Florian Wohlfart, Narseo Vallina-Rodriguez, Mark Allman, Randy Bush, Anja Feldmann, Christian Kreibich, Nicholas Weaver, and Vern Paxson. A multi-perspective analysis of carrier-grade NAT deployment. In *Proceedings of the 2016 Internet Measurement Conference*, IMC '16, page 215229, New York, NY, USA, 2016. Association for Computing Machinery.

[84] RIPE NCC. Probes. https://atlas.ripe.net/probes/, 2020.

[85] RIPE NCC. RIPE Atlas - User-Defined Measurements. https://atlas.ripe.net/docs/udm/, 2020.

[86] RIPE NCC. RIPE Atlas Probe Archive. https://ftp.ripe.net/ripe/atlas/probes/archive/, 2020.

[87] RIPE NCC. RIS - RIPE Network Coordination Center. http://www.ris.ripe.net/peerlist/all.shtml, 2020.

[88] Roberts, Hal and Larochelle, David and Faris, Rob and Palfrey, John. Mapping Local Internet Control. Technical Report, Berkman Center for Internet & Society, Harvard University, 2011.

[89] RouteViews. Collectors - RouteViews. http://www.routeviews.org/routeviews/index.php/collectors/, 2020.

[90] Anant Shah, Romain Fontugne, and Christos Papadopoulos. Towards characterizing international routing detours. In *Proceedings of the 12th Asian Internet Engineering Conference*, AINTEC '16, page 1724, New York, NY, USA, 2016. Association for Computing Machinery.

[91] Esther Showalter, Nicole Moghaddas, Morgan Vigil-Hayes, Ellen Zegura, and Elizabeth Belding. Indigenous Internet: Nuances of Native American Internet Use. In *Proceedings of the Tenth International Conference on Information and Communication Technologies and Development*, ICTD '19, New York, NY, USA, 2019. Association for Computing Machinery.

[92] F. Soldo and A. Metwally. Traffic anomaly detection based on the IP size distribution. In *2012 Proceedings IEEE INFOCOM*, pages 2005–2013, 2012.

[93] Indigo J. Strudwicke and Will J. Grant. #junkscience: Investigating pseudoscience disinformation in the Russian Internet Research Agency tweets. *Public Understanding of Science*, 29(5):459–472, 2020. PMID: 32597365.

[94] Submarine Cable Networks. MPT, China Unicom Plan International Cable to Boost Internet Connectivity. https://www.submarinenetworks.com/news/mpt-china-unicom-plan-international-cable-to-boost-internet-connectivity, 2013.

[95] Talanei. ASH Cable buys bandwidth from Tui Samoa. https://www.talanei.com/2018/05/10/ash-cable-buys-bandwidth-from-tui-samoa/, 2020.

[96] Telefonica. Telefonica International Network. https://www.wholesale.telefonica.com/media/2946/uk-network-map-12-2019-download.pdf, 2019.

[97] TeleGeography. Submarine Cable Map. https://www.submarinecablemap.com/, 2020.

[98] TM. TM AND HURRICANE ELECTRIC ESTABLISH STRATEGIC PARTNERSHIP FOR HIGH-SPEED IP BROADBAND SERVICES. https://www.tm.com.my/AboutTM/NewsRelease/Pages/TM-AND-HURRICANE-ELECTRIC-ESTABLISH-STRATEGIC-PARTNERSHIP-FOR-HIGH-SPEED-IP-BROADBAND-SERVICES.aspx, 2016.

[99] TOT Public Company Limited. TOT: INTERNATIONAL SUBMARINE CABLE. https://www.boi.go.th/upload/content/tot_5d254fe992f21.pdf, 2020.

[100] International Telecommunication Union. Connectivity Challenges and Opportunities - Bolivia. https://www.itu.int/myitu/-/media/Publications/2018-Publications/BDT-2018/Landlockeddeveloping-countries-LLDCs-in-the-Americas-region--Connectivity-challenges-Bolivia.pdf, 2018.

[101] United Nations. Agreement to resolve the controversy over the frontier between Venezuela and British Guiana. https://treaties.un.org/doc/Publication/UNTS/Volume%20561/volume-561-I-8192-English.pdf, 2020.

[102] Wahlisch, Matthias and Schmidt, Thomas and de Brun, Markus and Haberlen, Thomas. Exposing a Nation-Centric View on the German Internet - A Change in Perspective on AS-level. In *Passive and Active Measurement Conference (PAM)*, 2012.

[103] Nils B. Weidmann, Suso Benitez-Baleato, Philipp Hunziker, Eduard Glatz, and Xenofontas Dimitropoulos. Digital discrimination: Political bias in Internet service provision across ethnic groups. *Science*, 353(6304):1151–1155, 2016.

[104] Wired. What Would Really Happen If Russia Attacked Undersea Internet Cables. https://www.wired.com/story/russia-undersea-internet-cables/, 2018.

[105] Shi Zhou, Guoqiang Zhang, and Guoqing Zhang. Chinese Internet AS-Level Topology. *IET Communications*, 2(1), April 2007.

[106] Ran Zhuo, Bradley Huffaker, kc claffy, and Shane Greenstein. The Impact of the General Data Protection Regulation on Internet Interconnection. *Telecommunications Policy*, 45 (2), 2021.