Inferring Country-Level Transit Influence of Autonomous Systems

Alexander Gamero-Garrido[*], Esteban Carisimo[‡*], Shuai Hao[*], Bradley Huffaker[*], kc claffy[*], Alex C. Snoeren[†], Alberto Dainotti[*], and Amogh Dhamdhere[*]

[*]CAIDA/UC San Diego

[‡]Universidad de Buenos Aires, CONICET
[†]UC San Diego

ABSTRACT

We tackle the problem of identifying the most influential transit providers in each country that may have the potential to observe, manipulate or disrupt Internet traffic flowing towards that country. We develop two new Internet cartography metrics to overcome several challenges with making such inferences using BGP data. The *transit influence* (TI) metric estimates the share of addresses of an origin AS served by the transit AS. The Aggregate Transit Influence (ATI) captures the aggregate of all fractions of each country's origin ASes' addresses that the transit AS serves. We apply these two metrics to identify the most influential ASes in each country, and the origin ASes in those countries that heavily depend on transit ASes. We include extended case studies of the transit ecosystems of countries in Latin America, Africa and Europe, and we also investigate the role of state-owned ASes in the Internet ecosystem of their home country and in foreign countries. We believe these metrics advance our ability to characterize structural weaknesses in the global Internet topology.

1 INTRODUCTION

The central question of this work is the automatic identification of the most influential transit providers in each country, those who potentially have the largest capability to observe, manipulate or disrupt Internet traffic, or whose accidental misconfiguration would affect the connectivity of many users and organizations (*e.g.,* [1, 2]). This transit influence characterization requires studying the Internet global routing ecosystem, including its Border Gateway Protocol (BGP) routing infrastructure, the system relied upon by operators to announce and implement their routing policies. The largest compendia of publicly-available BGP routing data are collected by `RouteViews` [7] and `RIPE RIS` [5], who aggregate BGP messages from actual operational routers (BGP monitors) at cooperating Autonomous Systems (*monitor ASes*). In this paper, we develop novel analysis techniques to infer country-level transit influence from these BGP measurements and address four major technical challenges.

The first challenge is that BGP data collection is heavily biased towards paths seen from the (small sample of) monitor ASes. As monitors are not distributed uniformly across and between countries, and many countries and most ASes have none, the inferences of transit influence built with these measurements will *heavily* oversample paths towards monitor ASes. We mitigate this sample bias, and improve on the state of the art [9] by implementing novel filters. We concentrate on the transit influence of inferred providers of each origin AS, allowing us to determine who serves as direct or indirect transit providers of the organizations in each country for their international connectivity. We also prioritize the diversity of ASes hosting observation points in our computation, limiting the oversampling of BGP paths towards ASes who host multiple monitors. Finally, we limit our analysis to paths going from monitors which we infer to be outside each country to prefixes in the country, resulting in more consistency of our study of individual countries.

A second major challenge is that there is no direct way to map the IP addresses in a BGP prefix (block of addresses) to a geographic location [8]. Without accounting for the geographic presence of a prefix, it is impossible to determine which ASes are most influential in each country: paths reaching ASes in Central Asia may have little if any relevance for the connectivity of Central Africa, for example. We tackle this issue by leveraging commercial geolocation datasets from *Netacuity* [4] along with a study of delegated IP blocks published by Regional Internet Registries [6], to identify the set of prefixes that are relevant to each country. We also develop analysis techniques to determine the primary country of operation of transit ASes from these datasets, and find countries overwhelmingly served by foreign providers for their international connectivity. As a consequence, these countries may be in a vulnerable position with little leverage

to audit the practices of foreign ASes (*e.g.,* determining if traffic flowing towards the country is being observed).

The third major obstacle to inferences of transit influence is the massive scale of the global Internet, which has tens of thousands of ASes and links connecting them, combined with the dearth of publicly-available topological information at the country-level. While previous work has tackled this challenge for the global AS-level topology (*e.g.,* [10–14]), there is a gap in methods to find the most influential ASes in each country. Our study addresses this gap by building a bottom-up view of influence starting from the addresses geolocated to each country and originated by each AS. We take into account BGP's longest prefix matching (as operators typically prefer more specific prefixes) when assigning influence to ASes transiting overlapping prefixes. Then, we find ASes who are influential on many origin ASes in each country, to capture the exposure of the country's *organizations* (as opposed to addresses) to traffic observation, manipulation or disruption by those transit providers.

A fourth challenge in this space is that collections of BGP table dumps are aggregated in limited time windows (limiting the computational burden of analyzing prefix-level data) which are prone to missing backup or less preferred links that are only announced under disturbances, *e.g.,* if the preferred link is overloaded. This issue is exacerbated when considering indirect providers (without an inferred direct transit agreement with the origin AS) as the likelihood of missing a backup link increases with the number of AS-level hops from the origin: the origin itself may have backup links, its direct provider may *also* have backup links, and so on, so the inferences of transit influence become noisier for transit ASes farther away from the origin. A related issue is that, given our limited measurement footprint, we will also miss *preferred* paths towards the same prefixes, as they may be served by a different indirect provider that is located farther away from a BGP monitor AS. To limit the impact of these missing edges between ASes in our inference of transit influence, we develop a novel filter which takes into account both the number of paths a transit AS appears along *as well as* its AS-level distance from the origin.

In tackling these challenges we develop two metrics that we believe advance our ability to characterize structural weaknesses of the global Internet topology. Our contributions are as follows.

(1) We develop two new metrics for Internet cartography. The *transit influence* (TI) metric estimates the share of addresses of an origin AS served by the transit AS. The Aggregate Transit Influence (ATI), captures the aggregate of all fractions of each country's origin ASes' addresses that the transit AS serves. We use the ATI metric to rank ASes based on their country-level transit influence and determine who are the ASes most likely to have the capability to capture, manipulate or disrupt traffic towards origin ASes located in that country.
(2) We use these metrics to infer the most influential transit providers in 194 countries in March 2018, quantifying the share of each country's address space that is primarily served by domestic, foreign and global transit providers. We find that the exposure of traffic towards a country's addresses to observation, manipulation or disruption by *domestic* transit providers correlates with the Democracy Index [3], so countries with stronger democracies are less exposed. In contrast, GDP Per Capita, a measure of a country's wealth, does not correlate with the share of a country's address space primarily served by domestic providers.
(3) We identify 148 state-owned providers in 107 countries (that may give national governments a more direct mechanism to observe, manipulate or disrupt traffic towards the country) and evaluate their transit influence. We find that countries with a track record of engaging in Internet restrictions tend to have state-owned providers with high ATI.
(4) We provide extended case studies of the transit ecosystems of several countries in Africa, Latin America and Europe, including the five-year evolution of the transit ecosystem in South Africa; the inference of conglomerates who are highly influential in multiple South American countries; and the identification of transit providers who primarily serve South American origin ASes handling sensitive traffic (*e.g.,* large banks).

REFERENCES

[1] 2008. How Pakistan knocked YouTube offline. https://www.cnet.com/news/how-pakistan-knocked-youtube-offline-and-how-to-make-sure-it-never-happens-again/. (2008).
[2] 2015. Latin America and Caribbean Region Network Operators Group (LACNOG) Mailing List. https://mail.lacnic.net/pipermail/lacnog/2015-December/004262.html. (2015).
[3] 2017. Democracy Index 2017: Free speech under attack. https://www.eiu.com/public/topical_report.aspx?campaignid=DemocracyIndex2017. (2017).
[4] 2019. Netacuity. http://info.digitalelement.com/. (2019).
[5] 2019. RIPE Routing Information Service (RIS). https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris. (2019).
[6] 2019. RIR Delegation Files. https://ftp.ripe.net/pub/stats/ripencc/. (2019).
[7] 2019. RouteViews. http://www.routeviews.org/routeviews/. (2019).
[8] Bradley Huffaker and Marina Fomenkov and kc claffy. 2011. Geocompare: a comparison of public and commercial geolocation databases - Technical Report. Cooperative Association for Internet Data Analysis (CAIDA), May 2011. (2011).

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

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

[11] Ricardo V. Oliveira, Beichuan Zhang, and Lixia Zhang. 2007. Observing the Evolution of Internet As Topology. In *ACM SIGCOMM Conference*.

[12] Oliveira, Ricardo and Pei, Dan and Willinger, Walter and Zhang, Beichuan and Zhang, Lixia. 2010. The (in)Completeness of the Observed Internet AS-level Structure. In *IEEE/ACM Trans. Netw. (TON)*, Vol. 18. Issue 1.

[13] Oliveira, Ricardo V. and Pei, Dan and Willinger, Walter and Zhang, Beichuan and Zhang, Lixia. 2008. In Search of the Elusive Ground Truth: The Internet's As-level Connectivity Structure. In *ACM SIGMETRICS*.

[14] Zhang, Beichuan and Liu, Raymond and Massey, Daniel and Zhang, Lixia. 2005. Collecting the Internet AS-level Topology. *ACM SIGCOMM Comput. Commun. Rev.* 35, 1 (Jan. 2005).