POSTER: Why Are You Going That Way? Measuring Unnecessary Exposure of Network Traffic to Nation States

Jordan Holland
University of Tennessee
jholla19@vols.utk.edu

Max Schuchard
University of Tennessee
mschucha@utk.edu

ABSTRACT

In this work, we examine to what extent the Internet's routing infrastructure needlessly exposes network traffic to nations *geographically* irrelevant to packet transmission. We quantify what countries are *geographically logical* to see on a network path traveling between two nations through the use of convex hulls circumscribing major population centers, and then compare that to the nation states observed in utilized paths. Our preliminary results show that the majority of paths, 52%, unnecessarily expose traffic to at least one nation. We also explore which nation states are disproportionately allowed to observe and manipulate a larger fraction of Internet traffic than they otherwise should.

1 INTRODUCTION

The Internet is comprised of independent networks called Autonomous Systems (ASes), which depend on each other for internetwork connectivity. Network traffic must often traverse multiple ASes in order to reach its final destination. Any adversarial transit AS situated between sender and receiver can degrade network availability, violate data integrity, and undermine confidentiality and anonymity properties. We term such an adversary a *path based adversary*. However, a more powerful class of adversary also exists, **the nation states where the utilized network infrastructure is physically located**. Revelations in recent years about the extent to which countries such as the United States, Great Britain, and other members of the so called Five Eyes intelligence alliance have integrated dragnet surveillance into core Internet transit links that reside within their borders [4] only underscores the importance of understanding such nation state level path based adversaries.

Inter-domain routing decisions, which are made at the AS level, typically result in paths that generally expose traffic to as few ASes as possible, and reduces the capability of any one AS level path based adversary. However, a path that *appears* low risk, spanning only a single transit AS, might actually involve a large number of nation states in the process of traversing that lone AS. This means that network traffic is exposed to potentially a much broader collection of actors than the AS level path suggests. Additionally, since inter-AS routing focuses on the *logical* topology rather than the *geographic* topology, routing decisions can result in exposing traffic to nations which do not lie between the geographic locations of the sender and receiver. This additional exposure to nations not necessary for transmission of data needlessly increases the power of nation state level path based adversaries.

In this work, we examine to what extent the Internet's routing infrastructure exposes network traffic to nations that do not lie along the *geographically logical* path between sender and receiver. In order to do this, we must first quantify what countries we *geographically expect* to see on a network path traveling between two nations. We accomplish this by establishing which nations reside inside of the population biased convex hull between two countries. When then compare the set of nation states data actually traverses by examining traceroutes conducted on the RIPE Atlas [5] measurement infrastructure. We present preliminary results of what fraction of paths contain nations unnecessary to the transmission of data between source and destination, finding that more than 52% of tested paths involved at least one extraneous nation. We also explore which nation states are the benefactors of these "bad" (in the geographic sense) paths, allowing them to observe and manipulate a larger fraction of Internet traffic than they otherwise should.

2 APPROACH

Our goal is to accurately measure the fraction of paths which do not expose their traffic to nations not required for the actual transmission of data. To do this we first must establish a set of expected countries traffic could be exposed to during transit between a particular source/destination pair, what we term *geographically normal* or simply *normal*. After establishing this, we can then compare the normal set to the observed set of countries traffic is *actually* exposed to.

Defining the normal path from one country to another was done using the *convex hull* between a set of points that define the country containing the source and a set of points that define the country containing the destination. The *convex hull* of a set of points S in n dimensions is the intersection of all convex sets containing S. For N points $p1, ..., pN$, the convex hull C is then given by the expression:

$$C = \sum_{j=1}^{N} \lambda_j \rho_j : \lambda_j \geq 0 \; \forall j \; and \; \sum_{j=1}^{N} \lambda_j = 1 \quad (1)$$

CCS '17, October 30-November 3, 2017, Dallas, TX, USA

ACM ISBN 978-1-4503-4946-8/17/10.
https://doi.org/10.1145/3133956.3138842

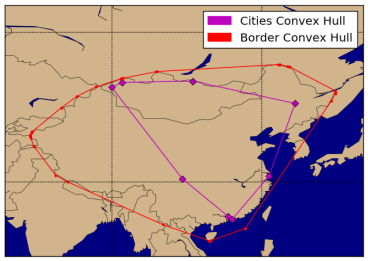


Figure 1: A comparison of the border based convex hull and population biased convex hull between China and Mongolia. Note that over 83% of China's population lives on its eastern coast.

Using the definition of a convex hull allowed us to mathematically define and generate a "normal" path between two countries given two sets of points that define the countries. A more intuitive way to think about the definition of a convex hull is: given a set of points, what is the shape a stretched rubber band takes when encompassing all of them.

One option for defining the set of points that make up a country is to utilize the nation's political borders. In order to accomplish this, we utilized shapefiles which contain points that define polygons of the actual borders of each country. However, the political borders of a country does not necessarily reflect where bulk the Internet infrastructure of the country is located; as this generally lies in the more populated areas. To address, this we built a separate definition of each country using the latitude and longitude of the top 15 most populous cities in each country [3].

Figure 1 shows an example of the two construction techniques for the path between China and Mongolia. The population based convex hull results in a stricter version of a normal path between two countries and accurately reflects the fact that 83% of China's population, including all of its major cities, reside in the eastern portion of the country. The border based convex hull includes countries in the wrong cardinal direction, such as India and Vietnam, a result of China's concave shape. We chose to use the city based construction of a convex hull for the measurements contained inside this work. For each pair of countries, we build the set of expected nations on paths between the two countries by building the convex hull between them, taking into account the spherical nature of the Earth, and enumerating all nations that either partially or completely reside inside the convex hull.

Establishing the utilized path from one IP address to another was achieved using Ripe Atlas traceroutes data [5] from March 2016 to April 2017. In order to expand the number of source/destination pairs, we inferred the path from each hop contained in the traceroute to the destination, rather than simply that of the originating node. The result was a data set of over 26 million different paths to test against our definitions of normal. Each IP address in the path was mapped back to the country and AS it was located. The correct country was done using the geolite IP geolocation database [2], while the accuracy of geolocation is at times limited in its precision, it has been show to be accurate at a country level [6]. To build the mapping between IP address and owning AS, we consulted routing tables from the CAIDA infrastructure [1]. Using this information we parsed our IP level paths into AS/Nation State tuples and compressed repeated instances of the same tuple down to a single instance. Establishing if a path was considered **normal** was achieved with simple set comparison between the expected set of nations and the set of observed nations.

3 RESULTS

All together, we examined over 26,000,000 traceroute paths involving 77,187 different ASes and 249 different countries. As a metric of normalcy, we have defined **degree of normality (DoN)** as:

$$DoN = \frac{total\ "normal\ paths"\ seen}{total\ paths\ seen} \quad (2)$$

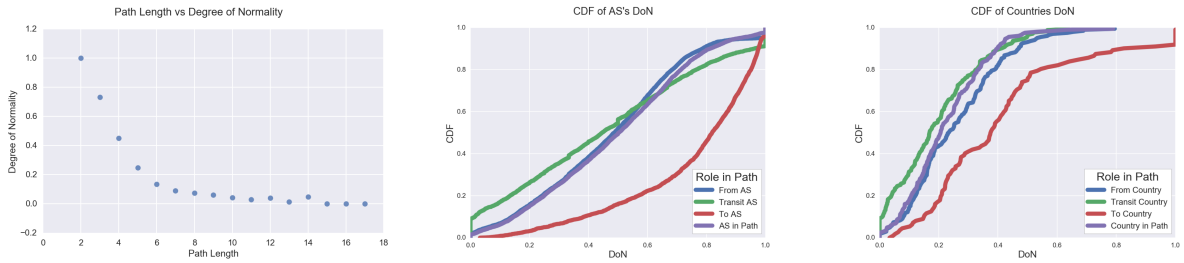
Over the entirety of the paths we examined, the total DoN was .473. Additionally, figure 2a shows that as the length of a path grows, the DoN immediately drops below .5, and continues to degrade as the path length grows.

We split the overall summary of DoN into three levels: AS, country, and regional. Additionally, we split scenarios for AS and country levels into the following based on if the entity is: the data source, the destination, neither the source or destination (a transit entity), and all three. At the AS level, figure 2b shows that in general, the DoN for paths transiting most ASes and starting in most ASes follow the same curve. However, the curve for paths to ASes demonstrates a trend of higher DoN, suggesting that a minority of the destinations, by AS, contribute to poor DoN. When examining the paths at a country level, we see the same trends as at the AS level. The curves for seeing a country in a path, a country transiting data, and beginning from a country all almost mirror each other. Following the AS level data, the curve for paths to countries is shifted right, suggesting that many countries are more difficult to get to while staying inside our defined normal path. Further examination needs to be done to determine if many of the ASes that have a low DoN coincide with the countries that have a low DoN.

We see in Table 1 that certain regions have better Degree of Normality when the path is to them than from them. For instance, the Americas have a higher than average DoN when the path ends or starts there, but a much lower DoN when they are found transiting a traceroutes message. Part of this could be explained by the smaller number of countries in the Americas, particularly North America. When having more adjacent countries, such as in Europe, there are more choices of countries to route through, and could naturally bring down the DoN for the region.

Table 1: Regional Degree of Normality

	Africa	Americas	Asia	Europe	Oceania
DoN From	.2003	.6214	.2933	.4853	.2817
DoN To	.2057	.6348	.2399	.4756	.2035
DoN Transit	.1799	.2849	.1666	.3806	.1507
DoN In	.1937	.3565	.2319	.4484	.2062



(a) Path Length DoN **(b) AS DoN** **(c) Country DoN**

Figure 2: Examining DoN at different levels

Table 2: Region to Region Degree of Normality

To \ From	Africa	Americas	Asia	Europe	Oceania
Africa	.4491	.3536	.0821	.1861	.0337
Americas	.2354	.7756	.4383	.6411	.4666
Asia	.0834	.3733	.3159	.2065	.1476
Europe	.1970	.6042	.2547	.5061	.1417
Oceania	.0420	.2265	.1554	.0550	.8477

Table 2 examines DoN on a region to region basis. When staying inside a region, all but Africa and Asia have above the overall average DoN. Interestingly, a traceroute traveling from Europe to the Americas has a better chance of following a normal path than a path staying inside Europe. Table 2 also shows that the DoN from one region to another is highly symmetrical; the DoN traversing *from* region 1 to region 2 is typically close to the DoN when traversing from region 2 *to* region 1.

Finally, we present a case study of one of the most interesting countries in our measurements: the United States. Of the over 26,000,000 examined paths, the United States showed up in roughly 52% of them, with a DoN of .345, well below the overall average of .473. However, paths *to* the United States have an average DoN of .725. This discrepancy in the degree of normality between the paths it is found *in* and the paths *ending there* can be found when examining who benefits from "bad" paths the most, shown in figure 3. The United states is the largest benefactor in seeing data it should not, showing up 6,796,688 paths that it should not have been in. Further examining this trend, we see in Table 3 that the US shows up in many European countries paths that it should not, but encompasses all regions when looking at the top 9 countries it benefits from the most. This is made more confusing given that paths contributing to this will transit across the Atlantic, to the United States, and then return back across the Atlantic to Europe.

Future Work In the future, we plan to expand our measurements to look at countries that see temporary, but marked, changes in their DoN, and attempt to establish the root cause of such changes. We are interested in examining if adversarial actions could result in a temporarily reduced DoN for nations, or if particular nations could inordinately benefit from adversarial reductions in DoN. Lastly, we wish to examine if nations can adjust their routing policies in an effort to increase their DoN, effectively reducing their exposure to nation state level path adversaries.

Table 3: Countries the United States Benefits from Most

Country	Number of Paths Ruined
Great Britain	695,534
Denmark	513,468
France	457,116
Australia	272,705
Netherlands	271,090
Russia	266,742
Japan	235,016
Italy	219,951
Spain	208,898

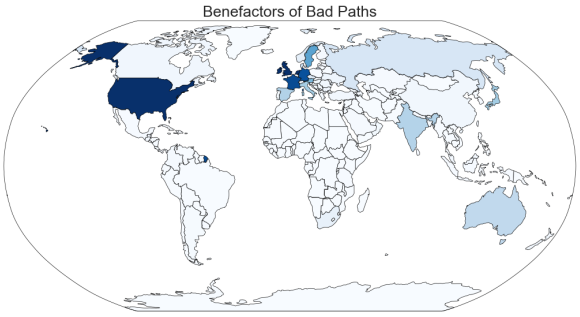


Figure 3: A visual representation of who benefits the most from "bad" paths

REFERENCES

[1] CAIDA AS relationship dataset. http://www.caida.org/data/active/as-relationships/index.xml.
[2] Geolite database. https://dev.maxmind.com/geoip/legacy/geolite/.
[3] The geonames database. http://download.geonames.org/export/dump/.
[4] The NSA uses powerful toolbox in effort to spy on global networks. http://www.spiegel.de/international/world/the-nsa-uses-powerful-toolbox-in-effort-to-spy-on-global-networks-a-940969.html.
[5] The ripe atlas dataset. https://atlas.ripe.net/.
[6] Y. Shavitt and N. Zilberman. A geolocation databases study. *IEEE Journal on Selected Areas in Communications*, 29(10):2044–2056, 2011.