Inferring Geography from BGP Raw Data

Enrico Gregori*, Alessandro Improta†, Luciano Lenzini‡, Lorenzo Rossi*, Luca Sani†

*Institute of Informatics and Telematics, Italian National Research Council, Pisa, Italy
†Information Engineering Department, University of Pisa, Pisa, Italy
‡IMT, Institute for Advanced Studies, Lucca, Italy

Abstract—Research to date has analyzed the Internet AS-level topology at a worldwide level of detail. Every inference found for an AS is extrapolated from the global set of AS paths gathered from public monitors, independently of the geographic location of the ASes. This approach is useful when the Internet is analyzed at a very coarse level. However, it may be misleading if the analysis is more focused on a specific geographical region. The risk is that the particular characteristics that the Internet has in that region may be lost. An AS connection that has been identified in a global analysis may hide multiple connections located in different geographic regions, each with its own characteristics. Moreover, a couple of ASes may establish different economic relationships in each geographic region where they are connected.

In this paper we propose an innovative technique to geolocate the AS connections retrieved from BGP raw data, in order to highlight the Internet regional characteristics. The analyses that we performed revealed some regional characteristics, in terms of graph properties and inter-AS economic relationships, that should be taken into account in a future analysis of the Internet.

Index Terms—Internet topology, Geography, Autonomous system, BGP

I. INTRODUCTION

The Internet is a complex system that evolved over the last few decades from a small network confined to the U.S. (i.e. ARPANET, 1969) to the current worldwide network of networks. It now consists of thousands and thousands of networks, under the administrative control of about 40,000 Autonomous Systems (ASes). This pervasive evolution did not occur homogeneously around the world for obvious historical, economic and political reasons. The result is that the Internet today is the composition of loosely connected groups of networks identifiable by some geographic boundaries, each with its particular pricing models, business contexts and regulatory environments [1]. Typically each AS has a particular role and specific economic behavior in each region of the world where it is present, which strictly depend on the connectivity and performance that it can provide for its customers in that region. For example, an intercontinental AS may be widespread in its home region – in terms of the number of connections and services offered – while it may be not competitive outside that region. This different level of pervasiveness may lead the same AS to establishing economic relationships in those regions with different criteria. Most research in the Internet topology analysis have considered ASes as homogeneous entities, each with a global set of metrics and characteristics, regardless of their heterogeneity. Those works (e.g. [2], [3], [4]) that tried to get a better insight in the geographic distribution of ASes relied on traceroute and packet delays to infer geographic information but, to the best of our knowledge, none of them explicitly focused on inferring regional AS topologies. Moreover from traceroute data is not trivial to obtain an AS-level topology, due to dealiasing and router-to-AS mapping issues [5], and it is almost impossible to have a precise view of the dynamic evolution of each AS path, that is fundamental to correctly infer the economic nature of each connection [6].

These problems can be easily solved exploiting the BGP data collected by Route Collectors (RCs) deployed by projects such as Route Views [7] and RIPE RIS [8]. The main source of information in BGP data about the AS-level topology is represented by the well-known mandatory AS_PATH attribute. This attribute contains the sequence of ASes that the traffic crosses to reach the announced subnet, but no information on the geographical regions where the traffic actually flows. Our aim is to search for a methodology able to infer geographic information from AS paths and to analyze local properties of the Internet, with a special focus on AS topology properties and economic relationships. Initially we infer the geolocation of each AS by exploiting the geolocation of its prefixes, which can be collected from BGP data. These data are exploited to produce a set of geographically tagged AS paths in which each connection is geolocated. This set is used to derive regional AS-level topologies that are analyzed both from a statistical and economic perspective. The economic analysis requires an economic tagging algorithm (e.g. [6], [9] and [10]), but none of the available algorithms can deal with geographic information without any modification. In this paper we show how to adapt the algorithm proposed in [6] to deal with geographic AS paths. We found that the Internet actually consists of regional and independent ecosystems, which differ greatly in terms of both topological and economic properties and introduce particular characteristics that are hidden in a global-level analysis. These differences should be taken into serious consideration by all research based on the AS-level topology of the Internet – e.g. Internet modeling evolutions,

This paper was partly supported by the MOTIA European Project (JLS/2009/CIPS/AG/C1-016). This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.
protocol analyses, worm spread analyses – in order to rely on a more realistic structure of the Internet, instead of on a coarse and potentially-misleading representation.

This paper is organized as follows. Section II details the process of geolocating each AS from raw BGP data. Section III introduces the methodology aimed at producing geographically tagged AS paths. Section IV shows the graph properties found by the undirected analysis of the regional topologies extracted from geographically tagged AS paths. Section V shows how to modify the economic aggregation algorithm described in [6]. Finally, section VI presents the economic analysis of the economic regional topologies extracted using the economic tagging algorithm and section VII concludes the paper.

II. AS GEOLOCATION

Knowledge regarding the geographic range of an AS is one of the fundamental parameters for decisions concerning the establishment of a settlement-free peering or a transit type of relationship between ASes. Several Tier-1 (T1) ASes include in their peering requirements at least one geographic constraint for candidate peers that need to be fulfilled. Just to name a few, AT&T requires a list of the countries served by the candidate peer in the peering request submission; Verizon requires a minimum number of served countries in the region where the peering is requested and that candidates own a "geographically-dispersed network"; and TeliaSonera requires that the candidate peer is present and able to exchange traffic and to be interconnected in a minimum number of cities in two out of three regions (Europe, North America and/or Asia Pacific/Oceania). To geolocate each AS we start from its formal definition, given in RFC 1930:

"an AS is a connected group of one or more IP prefixes run by one or more network operators which has a single and clearly defined routing policy."

Given this definition, it is straightforward that an AS is geolocated if its own prefixes are geolocated. The list of all the active prefixes of a given AS can be collected by parsing the BGP raw data provided by Route Views and RIPE RIS. Each prefix can be geolocated in turn by geolocating each IP address inside it, using one of the IP geolocation databases available [11]. Consider a generic route \( x.y.z.0/24 - A \) B C D. It is possible to claim that the last element of the AS path owns at least a network – and thus it may or may not be present – in the region(s) where the prefix is geolocated. This approach is correct for any given geographic region (e.g. countries, continents) iff the granularity of the geolocation tool is fine enough and iff the route does not carry the AGGREGATOR and the ATOMIC_AGGREGATE attribute. The AGGREGATOR is an optional discretionary attribute and the ATOMIC_AGGREGATE is a well-known discretionary attribute of the BGP protocol and may be included in UPDATE messages by a BGP speaker which performs route aggregation. If one of these attributes is present, it is possible that part of the real AS path is missing, hidden by the aggregating router. In this case, it is not possible to state that the considered prefix belongs to the last element of the AS path, but additional confirmation is needed from the WHOIS service provided by the Internet Routing Registries: the prefix is considered to belong to the last AS of the AS path if that AS is the owner of the announced prefix also according to the WHOIS response. For example consider the route shown in Figure 1 – which is the textual representation in MRT data of a route collected by the RC rrc12 of RIPE RIS – where the prefix is entirely geolocated in Europe. Given the presence of the AGGREGATOR attribute, we need to query the WHOIS service. Since the response state that the prefix belongs to AS 2597, we can conclude that AS 2597 is present in Europe.

Results presented in this paper are obtained using the Maxmind GeoIPLite database [12] and the following regional division: 1) Africa, 2) Asia-Pacific (Asia and Oceania), 3) Europe, 4) Latin America (the Caribbean, Central America, Mexico and South America) and 5) North America (Bermuda, Canada, Greenland, Saint Pierre and Miquelon, USA).

III. INTRODUCTION OF GEOGRAPHY IN BGP DATA

Geolocation of ASes by itself is not enough to extract geographic information from AS paths. An AS can have a geographic range that spread across multiple regions, thus it is not possible to infer where each AS connection forming an AS path is located. To overcome this problem we propose a three-step algorithm which, based on the geolocation of each AS, is able to geolocate each AS connection of the AS paths.

a) Enhanced routes from BGP raw data: In this step we obtain an enhanced route – defined as the triplet \{SOURCE, ASPATH, DESTINATION\} – for each route available in the BGP data. SOURCE is the region where the BGP AS Border Router that announced that route to the RC is located and can be obtained by geolocating its IP address (FROM field in Figure 1). ASPATH is the content of the homonym BGP attribute cleaned of private/reserved/unallocated ASNs and the ASN 23456, i.e. AS_TRANS used by 4-octet capable BGP speakers to communicate with 2-octet capable BGP speakers (RFC 4893). DESTINATION is the region where the prefix announced is located. Since a prefix could be geolocated in more than one geographic region, more than one enhanced route could be created from a single route, one for each region where the destination is found to be located. Consider the route reported in Figure 1. Both the IP address of the BGP speaker (80.81.192.98) and the prefix announced (192.12.193.0/24) are located in Europe, thus we obtain the enhanced route \{EU, 9189 8422 3356 2597, EU\}.

b) Detection of Single Region Located Transit Points (SRLTPs) in each enhanced route: In this step we extract from each enhanced route the set of SRLTPs. This set contains regional intermediate points where the traffic needs to flow. The SOURCE and the DESTINATION of each enhanced route are by definition part of this set, since they are both geolocated in a single region. This set also includes two classes of ASes that can be found in the ASPATH. The first class of ASes that fits in this definition is represented by ASes that own prefixes

1http://www.corp.att.com/peering/

2http://www.verizonbusiness.com/terms/peering/


4The last AS of an AS path is the AS that has originated the BGP UPDATE.

5A list of available WHOIS locations can be found at http://www.irr.net/
The first important result is that the regional topologies extracted overlap only slightly, as highlighted by the Jaccard similarity indices computed between pairs of regions for nodes (\(J_{\text{nodes}}\)) and connections (\(J_{\text{edges}}\)) and reported in Table I. This poor overlap is confirmed by the fact that only about 4.5% of ASes are located in more than one region and only about 1% in more than two. This evidence show that the Internet consists of regional ecosystems interconnected by...
just a very small number of inter-regional ASes. These ASes guarantee full Internet reachability, since only one AS owning an inter-regional IP network infrastructure can handle inter-regional traffic. This poor overlap also means that the regional principle has been applied by the algorithm has only on a small set of connections and thus the largest part of connections is correctly geolocated.

Further evidence regarding the differences between regional topologies is provided by the graph properties summarized in Table II and by the CCDFs of the node degree $k$ and the normalized average neighbor degree $\frac{k_{NN}}{\max(k)}$ depicted in Figures 3 and 4 respectively. The sizes of the topologies differ greatly in terms of nodes, reflecting the different degrees of economical and technological development of the regions. It is particularly interesting to compare North American and European topologies, which have a quite similar number of nodes but differ significantly in terms of edges. The CCDF of the node degree shows that the European region is more densely connected than the North American region where, on the other hand, there are ASes with a larger degree and where the number of ASes with a small degree is higher. This suggests quite a hierarchical structure in North America versus a flatter structure in Europe. This is confirmed by the CCDF of the normalized Average Neighbor Degree, which shows that in Europe, ASes tend to connect to ASes with a similar degree, while in North America they tend to connect to ASes with a very large degree. The differences between these ecosystems reflect the Internet’s historical evolution in the respective regions. In North America, especially in the U.S., a small set of large ISPs (e.g. AT&T, Centurylink and Verizon Communications) provide connectivity to all the states. In Europe on the other hand, each country is typically characterized by the presence of a national telco (e.g. Deutsche Telekom and Telecom Italia) which usually own a large part of the national Internet infrastructure, and by the presence of at least one Internet eXchange Point (IXP) that encouraged the establishment of settlement-free peering connections among local ISPs. More details on the role of IXPs in the development of the Internet can be found in [13], [14] and [15].

V. GEOGRAPHY AND INTER-AS BUSINESS RELATIONSHIPS

The analysis of the undirected graph of the Internet highlighted just part of the extreme complexity of the Internet ecosystem, since it completely lacks the real business model of each element. The Internet consists of a network managed by thousand of different organizations. Some of these organizations, e.g. small or large ISPs, live on the sale of Internet transit to other organizations, others, such as CDNs and search engines, just aim to offer content to end users. Most organizations, though, just care about Internet connectivity. In order to highlight that heterogeneity and to get a better insight into the Internet, some research studies ([9] and [16]) have proposed to annotate the Internet undirected graph with some economic tags that reflect the real business relationships established between different ASes, typically identified as provider-to-customer (p2c), customer-to-provider (c2p), peer-to-peer (p2p) and sibling-to-sibling (s2s). Several existing economic tagging algorithms can be used to obtain an economic tagged AS-level graph. The common approach of these algorithms

<table>
<thead>
<tr>
<th></th>
<th>Africa</th>
<th>Asia Pacific</th>
<th>Europe</th>
<th>Latin America</th>
<th>North America</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>-∞</td>
<td>0.03,0.04</td>
<td>0.02,0.01</td>
<td>0.03,0.04</td>
<td>0.02,0.02</td>
</tr>
<tr>
<td>Asia Pacific</td>
<td>0.03,0.04</td>
<td>-∞</td>
<td>0.03,0.04</td>
<td>0.04,0.04</td>
<td>0.04,0.08</td>
</tr>
<tr>
<td>Europe</td>
<td>0.02,0.01</td>
<td>0.03,0.04</td>
<td>-∞</td>
<td>0.02,0.02</td>
<td>0.04,0.06</td>
</tr>
<tr>
<td>Latin America</td>
<td>0.05,0.04</td>
<td>0.03,0.04</td>
<td>0.02,0.02</td>
<td>-∞</td>
<td>0.03,0.03</td>
</tr>
<tr>
<td>North America</td>
<td>0.02,0.02</td>
<td>0.03,0.08</td>
<td>0.04,0.06</td>
<td>0.03,0.03</td>
<td>-∞</td>
</tr>
</tbody>
</table>

Table I: Jaccard similarities indices $J = (J_{nodes}, J_{edges})$

<table>
<thead>
<tr>
<th></th>
<th>Africa</th>
<th>Asia Pacific</th>
<th>Europe</th>
<th>Latin America</th>
<th>North America</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodes</td>
<td>815</td>
<td>6,427</td>
<td>17,101</td>
<td>2,453</td>
<td>15,894</td>
</tr>
<tr>
<td>Edges</td>
<td>2,002</td>
<td>18,040</td>
<td>72,581</td>
<td>8,329</td>
<td>42,610</td>
</tr>
<tr>
<td>Unique edges</td>
<td>1,047</td>
<td>12,580</td>
<td>64,962</td>
<td>6,504</td>
<td>33,905</td>
</tr>
<tr>
<td>$(k)$</td>
<td>4.90</td>
<td>5.61</td>
<td>8.49</td>
<td>6.79</td>
<td>5.36</td>
</tr>
<tr>
<td>$\max k$</td>
<td>110</td>
<td>433</td>
<td>1,818</td>
<td>357</td>
<td>2,542</td>
</tr>
</tbody>
</table>

Table II: Regional topology statistics

$\langle \cdot \rangle = \text{average, } k = \text{degree}$
is to infer economic relationships by exploiting the \textit{valley-free} rule introduced in [9]. For example, [6], [9] and [10] infer economic relationships by searching for AS(es) with no provider in the AS path and then applying this rule. On the other hand, [17] and [18] try to find a tag assignment that maximizes the number of valley-free AS paths. Although these approaches were devised for the global Internet, they are still valid in a regional analysis, since the Internet is nothing more than the sum of the regional economic ecosystems that it is made up of. We now present a methodology to infer the regional economic relationships between ASes, by adapting the algorithm proposed in [6].

The original algorithm requires a list of AS paths as input together with their maximum lifespan, defined as the time interval during which there is at least one active route in a RC that includes that AS path in its attributes. It consists of three steps. In the first step all the possible tags for each connection \((A,B)\) are computed by applying the knowledge of the list of Tier-1 ASes to infer economic relationships in each of the AS connection of the path. In the second step, all the tags inferred for each connection \((A,B)\) are merged to obtain a single economic relationship for each connection. Finally, in the last step, the economic relationship obtained in the second step for the specular connections \((A,B)\) and \((B,A)\) are merged to infer a single economic relationship between AS \(A\) and AS \(B\). It should be noted that both the merging phases are based on the parameter \(N_{MAG}\): economic tags with lifespans that differ by more than \(N_{MAG}\) orders of magnitude are not merged together in order to avoid transient – and potentially wrong [19] – AS paths from distorting the results. The lower the \(N_{MAG}\) value, the lower the probability that transient information will affect the results. On the other hand, a low \(N_{MAG}\) value also reduces the number of two-way validated\(^8\) economic relationships.

The input of the algorithm can no longer be a AS path together with its lifespan. An AS path may be gathered from multiple BGP routers located in different parts of the world and may be used to reach different locations. In this case, some of the AS connections that make up the AS path are likely to refer to different physical links and the lifespan of each AS path may differ depending on the destination region. For example, the AS Path \(A\ B\ C\) used to reach subnets in Asia may be affected by the failure in the link that interconnects \(B\) and \(C\) in Tokyo, while the same AS Path used to reach subnets in Europe may not, since \(B\) and \(C\) are also interconnected in Paris. To overcome this problem, we exploit the concept of geographic AS path introduced in Section III which, together with its lifespan, represents the enhanced algorithm input. The lifespan of a geographic AS path is computed similarly to the lifespan of an AS path, i.e. it is the maximum period of time in which there is at least one active route that includes the related ASPATH that is announced from a router in SOURCE and reaches at least one subnet in DESTINATION.

To deal with the new input, we enhance part of the first step of the original algorithm in order to obtain only information concerning a specific region \(R\), as is shown in Figure 5. The economic tagging of each AS connection that make up the considered Geographic AS Path is initially performed using exactly the same methodology proposed in [6] (line 4-9) exploiting the presence of T1 ASes in the AS path and their provider-free property. Note that the initial tagging phase needs to be performed on the full AS paths, since the presence of a T1 AS need to affect every connection of the AS path, irrespectively of any geographic information.

Once the classic tagging phase is completed, we introduce an additional discarding phase (see line 10), in which all the economic tags related to AS connections not geolocated in \(R\) are removed from the partial results. As a consequence of the new version of the first step, the remaining steps of the original algorithm are fed only with economic tags related to AS connections established in \(R\) and, thus, the resulting economic tagged topology is related to \(R\).

VI. \textbf{ECONOMIC ANALYSIS} 

The application of the enhanced economic tagging algorithm to the sets of geographic AS paths allows deeper insights of each regional ecosystem which reveal the real nature of the regional differences that were only deduced from the undirected analysis of the Internet (Section IV). Table III shows the results obtained by applying the enhanced economic algorithm with \(N_{MAG} = 1\), listing the number of economic tags inferred for each region. The choice of \(N_{MAG} = 1\) is justified by the higher reliability of the economic tags inferred, as shown in [6]. We also performed the analysis for other \(N_{MAG}\) values, finding that the following inferences still hold.

The most relevant characteristic highlighted by the economic analysis is the large proliferation of potential p2p connections in the European ecosystem, representing the 54.76\% of the total. This feature is in contrast with the peering behaviors of other regions, where the amount of p2c connections is larger (around 70\% of the total) than the amount of p2p connections. Together with the conclusions drawn in Section IV, this allows to understand the real nature of the flat structure of the European Internet ecosystem. This joint analysis shows

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|}
\hline
Region & Africa & Asia & Europe & Latin America & North America & World \\
\hline
P2C & 1.436 & 12.808 & 32.471 & 4,514 & 31,820 & 80,095 \\
P2P & 492 & 5,012 & 39,747 & 3,719 & 10,164 & 58,040 \\
S2S & 21 & 102 & 297 & 37 & 350 & 1,743 \\
\hline
\end{tabular}
\caption{Economic Statistics}
\end{table}

\(^8\) An economic relationship between AS \(A\) and \(B\) is considered to be \textit{two-way validated} if an economic tag for the connection \((A, B)\) and for the reverse connection \((B, A)\) has been found.
that Europe is rich in small/medium transit providers that, in addition to offering transit to end-users and stub ASes\(^9\), tend to establish settlement-free p2p connections among them. The establishment of these BGP connections allows the ISPs to minimize the amount of traffic directed to their provider, thus reducing their transit costs. The proliferation of these small/medium providers is also the reason for the development in Europe of largely-populated IXPs (e.g. AMS-IX, LINX and DE-CIX) which in turn facilitated the establishment of settlement-free relationships among ASes, helping to create the large amount of p2p connections just described.

Another regional feature is highlighted by tag changes. In Table IV are summarized the most relevant tag changes from the worldwide to the regional scenarios, i.e. peering (p2p) to transit (p2c, c2p, s2s) and vice versa. Although the number of these connections may look not relevant at a first glance (around 4-10% of the total in each region), it should be considered that these tags are referred to AS connections that compose the core of the region. This is highlighted by the large number of tag changes that involve only non-stub ASes and that involve only multi-regional ASes. Most of the changes consists in shifting from transit to peering connections, showing that an AS may establish multiple economic relationships with another AS, depending on the location of the interconnection. This means that inter-regional providers may decide to establish regional p2p connections – exchanging only routes of regional customers – in those region where their pervasiveness is similar, while they may decide to establish a p2c agreement elsewhere. The presence of this regional type of relationship is confirmed by the existence of BGP communities dedicated to regional peers in some of the largest ASes, but our methodology is still too coarse-grained to detect their presence and to distinguish it from the general peering, where the ASes exchange the routes of all the customers.

### Table IV: Economic relationships changes

<table>
<thead>
<tr>
<th></th>
<th>Africa</th>
<th>Asia Pacific</th>
<th>Europe</th>
<th>Latin America</th>
<th>North America</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tag changes</td>
<td>202</td>
<td>1,094</td>
<td>3,011</td>
<td>464</td>
<td>1,557</td>
</tr>
<tr>
<td>Peering to transit</td>
<td>12</td>
<td>86</td>
<td>325</td>
<td>36</td>
<td>219</td>
</tr>
<tr>
<td>Transit to peering</td>
<td>165</td>
<td>824</td>
<td>2,304</td>
<td>361</td>
<td>1,136</td>
</tr>
<tr>
<td>Among multi-reg. ASes</td>
<td>155</td>
<td>594</td>
<td>725</td>
<td>277</td>
<td>703</td>
</tr>
<tr>
<td>Among non-stub ASes</td>
<td>175</td>
<td>846</td>
<td>2,279</td>
<td>317</td>
<td>1,101</td>
</tr>
</tbody>
</table>

Another regional feature is highlighted by tag changes. In Table IV are summarized the most relevant tag changes from the worldwide to the regional scenarios, i.e. peering (p2p) to transit (p2c, c2p, s2s) and vice versa. Although the number of these connections may look not relevant at a first glance (around 4-10% of the total in each region), it should be considered that these tags are referred to AS connections that compose the core of the region. This is highlighted by the large number of tag changes that involve only non-stub ASes and that involve only multi-regional ASes. Most of the changes consists in shifting from transit to peering connections, showing that an AS may establish multiple economic relationships with another AS, depending on the location of the interconnection. This means that inter-regional providers may decide to establish regional p2p connections – exchanging only routes of regional customers – in those region where their pervasiveness is similar, while they may decide to establish a p2c agreement elsewhere. The presence of this regional type of relationship is confirmed by the existence of BGP communities dedicated to regional peers in some of the largest ASes, but our methodology is still too coarse-grained to detect their presence and to distinguish it from the general peering, where the ASes exchange the routes of all the customers.

### VII. Conclusions and Future Work

In this paper we proposed an innovative tagging algorithm in order to geolocate AS connections starting from BGP data. This algorithm allowed us to infer regional AS-level topologies, that we have analyzed both from an undirected graph and economic perspective. From this analysis, we found that the study of the Internet at a global level fails to take into account several characteristics that, on the other hand, play a fundamental role in the regional Internet connectivity. In particular we found evidence of structural differences between the European and the North American Internet topologies, that reflect different historical developments of the Internet in those regions. The same methodology proposed can be applied at other regional levels, such as nations. However, it must be noted that some studies (e.g. [11], [20]) have found that the more the geographic scope is refined, the lower the precision of the IP geolocation database used is.

There is plenty of room for improvements in this study. For example, the geographic tagging algorithm proposed assumes that two ASes establish a BGP session in each region where they are co-located and, consequently, the connection between these ASes may be not correctly geolocated. Moreover, this coarse-grained assumption does not allow to distinguish between regional and worldwide peer-to-peer connections. To overcome these problems, we are currently working to enhance our methodology with a traceroute-based tool. This exploits the free availability of looking glasses owned by more than 200 different ASes, which may discover the real location of connection of currently unknown or uncertain location, allowing to refine our algorithm.

### REFERENCES


\(^9\) An AS is defined as stub if it does not transit traffic for any other AS.