8th International Wireless Communications and Mobile Computing Conference, IWCMC 2012, IEEE, 2012. 53-58

DOI: http://dx.doi.org/10.1109/IWCMC.2012.6314177

On the Real Impact of Path Inflation in Networks Under Production

Felipe Mata, Roberto González-Rey, José Luis García-Dorado, and Javier Aracil
High Performance Computing and Networking, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Spain
Email: {felipe.mata, jl.garcia, javier.aracil}@uam.es; roberto.gonzalezr@estudiante.uam.es

***Abstract*—The research community has proved the existence and studied the root causes of Path Inflation on the Internet—end-to-end paths significantly longer than necessary. However, it has been typically ignored that the popularity of traffic destinations and, consequently, of network paths, is clearly heterogeneous—some destinations are popular while others are barely accessed. In this paper, we propose a trace-driven methodology to measure the Path Inflation accounting for the popularity of Internet destinations from a given network, thus evaluating the implications that Path Inflation exerts on real networks under production. This information is important for network operators because it allows them to objectively stand out those destinations whose connection analysis must be prioritized. The results of applying this methodology to the Spanish academic network show that the most critical regions to focus on are Spain's closest countries, which either are very popular or have large Path Inflation as a consequence of the use of transatlantic links as intermediate nodes, or both.**

***Index Terms*—Path Inflation; Traffic Patterns; Network Measurement; Routing Policy; Topology; Traceroute**

I. Introduction

The existence of Path Inflation (PI)—end-to-end routes that are significantly larger than necessary—on the Internet network is a well-known fact since almost 15 years ago [1], [2]. The Internet community has largely studied the existence of this phenomenon and its root causes [3], [4], motivated by the impact that this circuitousness exerts on the network performance. Specifically, the one way delay is the parameter that is principally affected by inflation of paths, which could be considerably smaller in case less inflated paths between end-hosts are used. As a result, the optimal throughput in TCP connections is reached later as a consequence of the slow start congestion control strategy [5]. Moreover, the error rate increases because the more time a packet spends on the network, the higher the chances are that any problem may affect it. However, not only the existence and causes of PI have been analyzed, but also some procedures for reducing it have been devised [3]. Among these solutions, it outstands the proposal of including effective mechanisms to achieve optimal paths directly to BGP, such as appending geographic coordinates to route advertisements. With this information, a trade-off between hop-count metrics and geographic distance could be used in order to improve the latency in the network. As a consequence, network operators and service providers pay special attention to the PI in their networks, and take actions to reduce such inflation as much as possible.

978-1-4577-1379-8/12/$26.00

However, the above mentioned studies rely on a narrow snapshot of the network: they base their analysis on a small set of predefined network hosts—up to several thousands of hosts. This limitation directly applies to the representativeness of the results they provide, given that it is assumed that each analyzed connection pairs are equally likely. Previous work [6] shows that this is far to be the case, essentially it is pointed out that there is a small set of destinations with large popularity, and a large set of destinations which is barely accessed. So why ISPs should pay attention to the inflation of such unpopular routes? As it turns out, these results call for including network traffic analysis into PI metrics, in order to obtain representative information regarding what to measure and how to appropriately weight the obtained measurements. Our work fill in this gap, leveraging on a network trace analysis to infer what connections are in fact conducted in the network under study, and which is their popularity among the population of customers of the network. Thus, the priority of an ISP should be those destinations that combine popularity and high PI.

This information (inflation of paths and their popularity) is of paramount importance for network operators and service providers. On one hand, knowledge of the inflated paths allows the operators and providers to identify the locations that are poorly connected, whereas knowledge of the popularity of the traffic destinations serves to focus on the most demanded destinations by their customers, and both tasks eventually result in similar traffic engineering tasks: improving current traffic inter-exchange relationships or establishing new ones. Consequently, merging PI information and remote locations popularity knowledge allows for setting priorities to these traffic engineering tasks that the operators and providers should eventually take action on in a short time span.

In this paper, we provide the methodology for merging both metrics into a new one that is able to determine which connections should receive attendance first, and apply this analysis to the Spanish academic network (RedIRIS) as study case. The measurement of PI entails the identification of the intermediate nodes in a network path and the geographical mapping of IP addresses to measure PI in terms of distance. The selection of representative nodes is based on a trace analysis, which should be at least 35 days long in order to obtain stable destination patterns according to the results in [6]. Our findings after applying the proposed metric show that the most critical regions to pay attention to are the closest ones to Spain, which either are very popular or have large PI as a consequence of the use of transatlantic links as intermediate nodes, or both. Our results show that *a byte from our network travels more than 8000 extra kilometers* on average in the Internet to reach its destination.

After reviewing the related work in Section II, the rest of the paper is organized as follows. Section III describes the network analyzed in this study, while Section IV is devoted to describe the methodology. Then, Section V presents the results and a discussion of the main findings. Finally, Section VI summarizes the achievements and concludes the paper.

II. Related Work

In this work we study the inflation of paths from the Spanish academic network, estimating the routing distance as the sum of the geographical distances between each router of a given path, as reported by means of the `tcptraceroute` [7] tool. Consequently, let us divide this section into these two areas, first *Path Inflation* and then *geolocation*.

A. Path Inflation

The Path Inflation phenomenon has received much attention by the Internet community, since that, almost 15 years ago, the authors in [1], [2] found that the routes in the Internet are clearly longer than necessary. Since then, the Internet community has tried to characterize the PI, explain the causes of such phenomenon, and study its correlation with the performance experienced by users.

The authors in [8] explain that there are both technical and economic reasons to expect suboptimal Internet routes. Specifically, in that paper it is found that between 30-80% of the paths are not optimal. On one hand, wide-area routing protocols do not incorporate performance measurements into their decisions. On the other hand, the administrator of a given AS may refuse to carry traffic of another ISP because of competitive reasons or simply because the lack of contractual agreements.

The authors in [9], [10] focus on the signification of such economic reasons. These papers show that about 20% of Internet paths are inflated by more than 50% in terms of number of hops from the source to the destination with regards to the optimal route path. However, the authors point out that they are assuming that all the routes between each pair of studied hosts are equally likely, and this is not true [6].

Similarly, the authors in [3] wonder why Internet paths are sometimes absurdly long. They analyzed this fact from the intra- and inter-domain ISP points of view as well from the ISP peering relationship. They found that the intra-domain routing is the most significant factor in the Path Inflation phenomenon, because routers typically use minimum AS-hop count ignoring other metrics. They concluded that almost 50% of the paths were longer that the shortest available path because of intra-domain routing. In addition, they remark that according to their measurements geography is a good indicator of latency for most of the studied ISPs. However, the authors notice that their study is assuming, not in a totally realistic way, that all nodes are equally important regardless of traffic volumes, and point out that it would be more interesting to study the fraction of packets that suffer Path Inflation rather than studying the fraction of paths.

Padmanabhan and Subramanian in [11] worked further to extend the characterization of the PI phenomenon. In that study, the authors measure the PI as the ratio between the linearized distances, i.e., the sum of distances in kilometers between each of the nodes of a path, and the linear geographic distance between the end-hosts. They evaluated the PI from 20 institutions (placed in the U.S., Sweden, Italy, and Hungary) and two home cable modem networks to an extensive set of pre-defined destination hosts. Such a set included essentially web servers, some of them located on U.S. campuses, and public libraries which were easily geographically placed. They again found that PI is a common phenomenon in the Internet, and that it strongly depends on the geographic location of the end-hosts. This was explained by the fact that the connectivity of the different parts of the world is far from being homogeneous. That is, the paper takes an arbitrary set of end-hosts and no distinctions on their popularity were performed, however, as the authors showed, there are significant differences between the PI from some geographic areas to others—specifically, they pointed out that PI in the San Francisco bay area was significant smaller than in other places.

Bearing all these previous works in mind, we note that the *real* impact of the PI is not currently well known, that is, it is proven that PI would be large in an Internet in which all the destinations were equally popular and all the places were equally connected. As this is not the case, in this paper we take a step further and try to fill such a gap, i.e., appropriately account for PI leveraging on end-host location popularity.

B. Geolocation

There are several ways to find the physical location of IP addresses, which can be classified as active or passive [12]. The former class includes mechanism based on the delays between a set of reference nodes—landmarks—and the target node. Examples of this are [11], [13], [14]. This approach is based on the linear correlation between the delay in the networks and the geographical distance between the objective and the set of landmark nodes. Basically, it is expected that hosts placed in a similar geographical distance present similar delays measured typically by means of the *ping* tool. Such correlation has been found in some regions of the Internet, essentially North America and west Europe [13], but the precision is limited in the rest of the world. As the target of our work is to span all the possible destinations in the world, this precision depending on the area represents a significant caveat.

On the other hand, the passive mechanism to locate hosts is typically based on *i)* the identification of some pattern in the routers' DNS name—essentially, names and codes of cities or airports—that allows to infer their location or, at least, their AS and *ii)* the use of databases, typically commercial applications, which directly relate IP addresses and geographical locations. An example of the use of DNS name patterns is GeoTrack [11]. GeoTrack estimates the geographical location of the objective node as that of the last identifiable router in a given path. Its precision tends to be notable but the number of routers whose name follow some recognizable policy on its naming is limited, albeit according to the authors these are more than 70%; in addition GeoTrack is designed to locate routers but unlikely it could locate final hosts.

According to the literature a more accurate option is the database approach, whose implementation is poorly known because they are typically commercial applications. The performance of this approach has been studied and reported ([15], [16]), resulting in median error around 60 kilometers. In this study, we have used this latter approach, specifically the free version of the *GeoIP Country* database of *MaxMind*, i.e., *GeoLite Country*, which has an accuracy of 99.5% according to the company [17] and outperforms other equivalent approaches [16]. Such database has entries for the country code, country name, and continent data.

Finally, for a better understanding of geolocation procedures, the reader is referred to [18, Section 5.3.6] and references therein.

III. Description of the Network

This work aims to characterize the PI from the Spanish academic network RedIRIS, paying special attention to the connections that are commonly established from it. RedIRIS network comprises more than 350 institutions, mainly universities and research centers, and kindly provides us with flow traffic measurements for research purposes[1]. Figure 1 graphically describes the network. Our premises are located under the Point of Presence of Madrid, which is at one hop distance from the RedIRIS external gateway that connects to the rest of the Internet through commercial links (TeliaSonera, Global Crossing, Espanix, etc.) and with the European research network, GÉANT.

IV. Methodology

A. Selection of Representative Destinations

Based on the results from [6], at least 35 days of traffic measurements from a subnetwork should be aggregated to make stable the traffic distribution of the geographic destinations. Consequently, we have gathered 35 days of flow measurements from the Point of Presence in Madrid, which allow us to calculate the number of bytes that are destined to each foreign country, and the IP addresses that are requested. The traffic traces used in this analysis partially comprise April-May 2010. From this dataset, we have ruled out those countries that received less than 0.005% of the total sent traffic. Overall, there are more than 31 million different IP addresses receiving almost 1 PB of traffic after the filtering process.

This is a vast dataset compared to others analyzed in previous works that were in the order of thousands IP addresses [4]. The distribution of traffic among the different countries that comprise the dataset is depicted in Fig. 2 in logarithmic scale. A logarithmic scaling is necessary in order to enhance visibility provided the power-law shape of the traffic distribution [6]. To draw the maps, we used Google's Visualization API[2] that can be used directly as a gadget from Google docs.

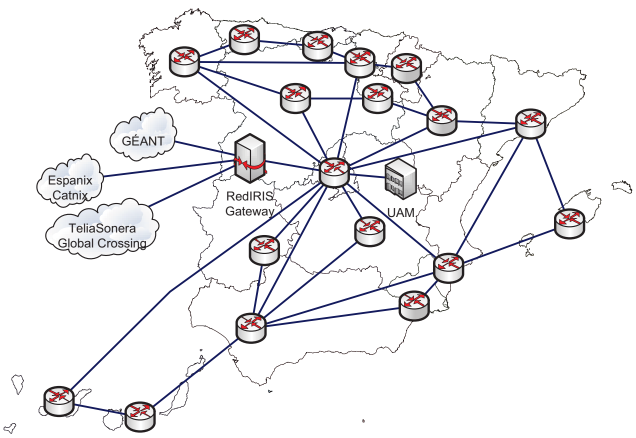


Fig. 1. Map of Spain and RedIRIS Points of Presence, showing the logical connections with the RedIRIS external gateway and UAM premises.

Fig. 2. Received traffic by country in logarithmic scale.

B. Intermediate Nodes Identification

To identify the intermediate nodes between our premises and the target IP addresses in our dataset, we leverage on `traceroute` and `tcptraceroute` tools. Such tools provide an equivalent approach to identify the intermediate nodes in a path within two IP addresses, but based on different network protocols. `traceroute` sends out either UDP or ICMP echo request packets, whereas `tcptraceroute` uses TCP SYN packets to circumvent the widespread use of firewalls. These tools allow us to identify the IP addresses of the intermediate nodes, which we map to their geographic coordinates by means of the geolocation method described in Section II-B.

[1]Data are stored and analyzed in full compliance with Spanish regulation concerning privacy of electronic communications

[2]http://code.google.com/intl/es-ES/apis/visualization/documentation/gallery/intensitymap.html

Because our study is trace-driven, our results are limited to the lack of visibility of some Internet hosts that do not reply to `traceroute` or `tcptraceroute` messages. We have found that `tcptraceroute` outperforms `traceroute` given the widespread deployment of firewalls. Consequently, we selected `tcptraceroute` as the path-analysis tool for this study, and limit our initial set of target IP addresses to the subset that answer to `tcptraceroute` queries. This set is still very large compared to previous ones used in the literature, containing more than *5 million different IP addresses*. In order to reduce the impact of path fluctuations in our analysis, we coordinate the path-analysis tools with the trace collection, in a way that what we observe from the `tcptraceroute` tool are the paths that were used by the connections in the trace during their lifetime. Further work will be needed to determine the extent to which our results generalize to other periods of time and other set of destinations. In addition, we have ruled out some instances of our dataset that lead to incongruent data, such as network paths traversed at higher speed than the speed of light.

C. Path Inflation Metric versus Weighted Path Inflation Metric

Other studies existing in the literature have faced the Path Inflation analysis leveraging on different metrics, such as distance, time, or number of hops. In this paper, we focus on a distance metric, such as the one used in [4]. The authors of [4] define the PI metric as the ratio between the routing and the geographical distances, where the routing distance is estimated as the sum of the geographical distances between each pair of consecutive intermediate nodes. Defining the unidirectional path $\{X_j\}_{j=0}^{N_{a,b}}$ between the IP addresses a (IP_a) and b (IP_b) as $\mathrm{IP}_a = X_0 \to X_1 \to X_2 \to \cdots \to X_{N_{a,b}} = \mathrm{IP}_b$, where $N_{a,b}$ equals the number of intermediate nodes plus one, we obtain:

$$PI(\mathrm{IP}_a, \mathrm{IP}_b) = \frac{d_r(\mathrm{IP}_a, \mathrm{IP}_b)}{d_g(\mathrm{IP}_a, \mathrm{IP}_b)} = \frac{\sum_{j=0}^{N_{a,b}-1} d_g(X_j, X_{j+1})}{d_g(\mathrm{IP}_a, \mathrm{IP}_b)}, \tag{1}$$

where $d_g(X, Y)$ is the geographical distance between the locations of IP addresses X and Y, and we have used $d_r(\cdot,\cdot)$ to denote the routing distance.

Consequently, the larger the circuitousness of the path, the larger the PI metric, which is interpreted as the number of times the path is larger than what would be necessary if a straight route would be possible.

The limitation of this metric is that it does not take into account the amount of traffic that is destined to each destination host. To take into account the amount of traffic, we group the PI metrics by destination country, taking the mean value as a representative, namely $\overline{PI}$. Such mean PI metric by country is then weighted with the logarithm to base 10 of traffic that is destined to such country, in order to provide larger weights to the popular destinations, which we define as WPI:

$$WPI(\mathrm{country}_c) = \overline{PI}(\mathrm{country}_c)\log_{10}(B_c), \tag{2}$$

where B_c is the amount of bytes destined to $\mathrm{country}_c$. This metric measures the PI taking into account the connection patterns in the network under study. Popular countries $\overline{PI}$ is penalized, whereas the impact of $\overline{PI}$ in those countries which barely receive traffic is reduced. We have chosen the logarithmic scale to weight the country average Path Inflation based on our experience with power law data [6]. The selection of the number of bytes as the measurement item for the weighting function, instead of the number of packets or flows, is because we have found the number of bytes to be more representative than the other measurable items, such as the number of packets or flows, since the number of bytes in fact accounts for the real network usage of the connections.

Fig. 3. Mean PI metric by country, in logarithmic scale

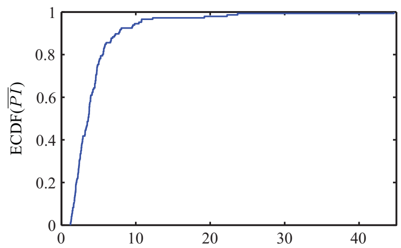
V. RESULTS

A. Path Inflation Results

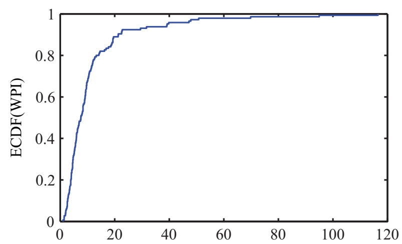
In this section, we present the results of measuring the $\overline{PI}$ in the set of destinations that are fully characterized by `tcptraceroute`. As the PI metric has been deeply analyzed and characterized in previous works, we present such results here just as a benchmark for comparison with the results obtained when the traffic weights are introduced in the metric, as presented in the next section.

In Fig. 3 we present the mean value of the PI metric when grouped by country. As can be observed in the figure, the largest values of $\overline{PI}$ are found in the countries surrounding Spain. Although this may be counterintuitive at first glance—if the distance between two locations is not large, there should be less alternatives to choose a path within them, and consequently the circuitousness should be smaller—, this phenomenon is explained by the common usage of transatlantic routes (via the U.S.), even for connecting pairs of locations within Europe. As a consequence of the popularity of the transatlantic routes, American countries suffer low values of $\overline{PI}$ when measured from our premises. In addition, we observe that Far East and Australian countries also have low values of $\overline{PI}$, which is a consequence of the common usage of a direct link connecting Europe and China.

On the other hand, we present this information summarized in a Cumulative Distribution Function (CDF) plot in Fig. 4(a), where we can observe that approximately 80% of the analyzed countries have paths larger than twice the distance measured in a straight line. This situation has consequences for instance



(a) $\overline{PI}$



(b) WPI

Fig. 4. Empirical Cumulative Distribution Function of the (a): $\overline{PI}$ and (b): WPI by country as observed from the central node of RedIRIS.

in the minimum one way delay, which is a key performance indicator usually related with quality of service/experience in multimedia services, such as voice conversations.

B. *Weighted Path Inflation Results*

In this section, we present the results of weighting the average PI metric by country with the base 10 logarithm of the number of bytes that are destined to such country, and compare them with the ones previously presented as benchmark in the preceding section. Analogously as in Section V-A, we present a world map with the WPI results in Fig. 5 and the data summarized in a CDF in Fig. 4(b).

In the world map figure (Fig. 5) we observe many differences with regards to Fig. 3. On one hand, we find that American countries now have negligible values of WPI, which is a consequence of the low average PI of the U.S. ($\approx$1.5) and that most of the connections to America go through the U.S. Consequently and despite the popularity within the users of our network—consider that most popular web services and contents are located in the U.S., and our customers share the same language with most of the population in South America—, America can be regarded as well connected to RedIRIS. Similarly, we observe low values of WPI in South Africa, the Far East and Australia. However, the reasons are quite different. On one hand, the popularity of South African countries is scarce within RedIRIS users, whereas it is the average PI to the Far East and Australian countries what is low on the other hand. Anyhow, the connections to such countries should not require attention from RedIRIS network managers.

On the other hand, we observe that there are countries that have barely experienced variation in the $\overline{PI}$ and WPI maps. Those countries are mainly located in Europe and

Fig. 5. WPI metric by country, in logarithmic scale after an axis rescaling to enhance visibility.

North Africa. Again, this is due to different reasons. On one hand, Spanish surrounding countries have the largest values of average PI due to the use of transatlantic links. In this group we include North African countries as well as Andorra and Portugal, which do not have great popularity, but the connections to them are very poor. On the other hand, we found the remaining European countries, which have a mix of great popularity and middle-poor connections given the usage in some cases of the transatlantic links. In both situations, RedIRIS network managers should pay special attention to the connections to such countries, and improve them given that, taken into account the traffic destinations popularity, there is not a good connection between them and RedIRIS.

Finally, in Fig. 4(b) we observe the CDF of the WPI. Compared to the CDF of the $\overline{PI}$, we observe that both distributions are much alike. The major differences cannot be appreciated in the summarizing statistics, because they are in the form of a reordering of the countries. There were some countries with large average PI in Fig. 4(a) that now have small value of WPI, and the same in the opposite way—large WPI value despite a low average PI given the great popularity of the country. In any case, we observe a high clustering of countries in small values of WPI, and a flattening of the distribution for values of WPI larger than 20, which roughly coincides with the 90% percentile. We believe this could be treated as a *knee point*, and the RedIRIS network manager should inspect the countries which have WPI values larger than such knee point.

C. *Discussion*

So far we have motivated that network operators and service providers not only should pay attention to the PI in their networks, but weight its relevance with the popularity of their destinations. Let us now show the impact that this exerts in the destination priority order of an ISP such as RedIRIS. Table I shows the comparison of the critical countries when only the average PI is taken into account, and the critical countries when this average PI is weighted with the amount of traffic that is destined to such country.

As can be observed in the table, when only the average PI is taken into account there appear countries that, although

TABLE I
COMPARISON OF THE CRITICAL COUNTRIES WITH THE AVERAGE PI AND WPI METRICS.

Rank	Average PI	WPI
1	Andorra	Andorra
2	Portugal	Portugal
3	Morocco	Morocco
4	Algeria	Italy
5	San Marino	France
6	Luxembourg	Algeria
7	Italy	Belgium
8	Liechtenstein	United Kingdom
9	France	Germany
10	Libyan Arab Jamahiriya	Luxembourg
11	Belgium	Netherlands
12	Montenegro	Denmark
13	United Kingdom	Russian Federation
14	Germany	Czech Republic
15	Tunisia	Poland
16	Malta	Switzerland

have large values of PI, are not of interest from the network operator point of view, such as San Marino, Liechtenstein, Libyan Arab Jamahiriya or Montenegro, since they do not reach a 1% of the traffic share. However, when the popularity of the countries in terms of received bytes is considered, we can observe that such countries are filtered out. Consequently, leveraging on destinies popularity is of paramount interest for network operators before deciding which actions regarding improving the network connections take first.

Finally, we have observed that the most critical regions according to the WPI metric are the nearest countries to Spain. The reasons are mainly the countries' large popularity and/or the use of transatlantic links. We believe that similar results would be obtained if the study is carried in other European countries, for the same reasons. On the contrary, we believe that the situation would be fairer if the analysis is performed from America, since the use of transatlantic links would not be so representative in the PI metric, accordingly with previous results [11].

VI. SUMMARY AND CONCLUSIONS

This paper puts on perspective the importance of the PI in the current Internet. Whereas the previous studies detected the existence of such phenomenon in the Internet, we have determined to what extent such inflation is critical, taken into account the amount of the traffic that is destined to each location. We have proposed a new methodology to study the PI, essentially a trade-off between the popularity of the destination and the PI that suffer the traffic volumes sent to a given destination. Such methodology permits network operators and service providers to really identify those paths that deserve to be improved because they suffer PI and, at the same time, much traffic is carried through them.

We present the case study of the Spanish academic network, which has shown that a set of geographically close countries is not as well connected as desired, yet being very popular. On the other hand, the PI metric weighted by destination popularity, WPI, has proven to be useful to filter out unpopular destinations, which according solely to the PI would have required special attention to the detriment of popular destinations which affect a large number of users. These results encourage the Spanish academic network managers to pay attention to the international relationships with ISPs located at these areas. These actions will reduce the amount of extra distance that a byte travels in average, which is larger than 8000 kilometers according to our results.

As future work we plan to extend the work to commercial networks, and also focus on network performance metrics besides the popularity of the destinations. Furthermore, we will explore the variance of the avergage PI and the causes of such variability.

ACKNOWLEDGMENTS

The authors would like to thank the support of the Spanish Ministerio de Ciencia e Innovación (*MICINN*) to this work, under project *ANFORA* (TEC2009-13385) and the FPU fellowship program that has funded this research work.

REFERENCES

[1] C. Labovitz, G. Malan, and F. Jahanian, "Internet routing instability," *IEEE/ACM Trans. Netw.*, vol. 6, no. 5, pp. 515 –528, 1998.
[2] V. Paxson, "Measurements and analysis of end-to-end Internet dynamics," *Ph.D. dissertation*, 1997.
[3] N. Spring, R. Mahajan, and T. Anderson, "Quantifying the causes of path inflation," in *Proceedings of ACM SIGCOMM*, 2003, pp. 113–124.
[4] L. Subramanian, V. N. Padmanabhan, and R. H. Katz, "Geographic properties of Internet routing," in *Proceedings of USENIX Annual Technical Conference*, 2002, pp. 243–259.
[5] W. Stevens, "RFC 2001: TCP Slow Start, Congestion Avoidance, Fast Retransmit, and Fast Recovery Algorithms," 1997.
[6] F. Mata, J. L. García-Dorado, J. E. López de Vergara, and J. Aracil, "Factor analysis of Internet traffic destinations from similar source networks," *Internet Research*, vol. 22, no. 1, pp. 29–56, 2012.
[7] M. C. Toren, "tcptraceroute," http://michael.toren.net/code/tcptraceroute.
[8] S. Savage, A. Collins, E. Hoffman, J. Snell, and T. Anderson, "The end-to-end effects of internet path selection," in *Proceedings of ACM SIGCOMM*, 1999, pp. 289–299.
[9] H. Tangmunarunkit, R. Govindan, S. Shenker, and D. Estrin, "The impact of routing policy on Internet paths," in *Proceedings of IEEE INFOCOM*, vol. 2, 2001, pp. 736–742.
[10] H. Tangmunarunkit, R. Govindan, and S. Shenker, "Internet path inflation due to policy routing," in *Proceedings of SPIE International Symposium on Convergence of IT and Communication*, 2001, pp. 188–195.
[11] V. N. Padmanabhan and L. Subramanian, "An investigation of geographic mapping techniques for internet hosts," in *Proceedings of ACM SIGCOMM*, 2001, pp. 173–185.
[12] D. Chatzopoulou and M. Kokkodis, "IP geolocation," Computer Science and Engineering Dept, UC Riverside, Tech. Rep., 2007.
[13] A. Ziviani, S. Fdida, J. de Rezende, and O. Duarte, "Improving the accuracy of measurement-based geographic location of Internet hosts," *Comput. Networks ISDN*, vol. 47, no. 4, pp. 503–523, 2005.
[14] E. Katz-Bassett, J. John, A. Krishnamurthy, D. Wetherall, T. Anderson, and Y. Chawathe, "Towards IP geolocation using delay and topology measurements," in *Proceedings of ACM SIGCOMM Internet Measurement Conference*, 2006, pp. 71–84.
[15] B. Gueye, S. Uhlig, and S. Fdida, "Investigating the Imprecision of IP Block-Based Geolocation," in *Proceedings of International conference on Passive and Active Network Measurement*, 2007, pp. 237–240.
[16] I. Poese, S. Uhlig, M. A. Kaafar, B. Donnet, and B. Gueye, "Ip geolocation databases: unreliable?" *SIGCOMM Comput. Commun. Rev.*, vol. 41, pp. 53–56, 2011.
[17] L. L. C. MaxMind, "GeoLite free country database."
[18] M. Crovella and B. Krishnamurthy, *Internet measurement: infrastructure, traffic and applications*. New York (USA): John Wiley and Sons Inc., 2006.