Ann. Telecommun. (2011) 66:31–43
DOI 10.1007/s12243-010-0197-8

# On the performance and improvement of alias resolution methods for Internet core networks

**Santiago Garcia-Jimenez · Eduardo Magaña · Daniel Morató · Mikel Izal**

Received: 8 January 2010 / Accepted: 17 August 2010 / Published online: 5 September 2010


**Abstract** The Internet is a huge interconnection of thousands of networks with different technologies, equipment, configurations, and administrative owners. This, added to the lack of public information about those individual infrastructures, makes it a difficult task to provide a so-called Internet map: a topological map with information of routers, interconnections between routers, and IP addressing configuration. Traditional topology discovery methods based on traceroutes only provide IP addresses in the path between end-nodes. Some of those IP addresses can belong to the same router, and this identification is made by alias resolution methods. Therefore, alias resolution allows to provide router-level map of the Internet with important applications in network simulation, protocol design, network management, network security, network service design, and geolocation. In this paper, alias resolution methods are analyzed in Internet core networks (GlobalNOC, Canet4, and Geant). This allows to identify peculiar behaviors in these core networks, improving alias resolution methods. Simultaneously, reduction methods are used to decrease the number of probing packets in alias resolution methods.

**Keywords** Internet topology · IP alias resolution · Reduction methods · Router identification

## 1 Introduction

The creation of a router-level Internet map is still a research challenge. This kind of map could provide invaluable help in order to analyze the behavior of the Internet. Measurements of delay, congestion, routing, or protocol performance could take advantage of this kind of map. The lack of public information about router configuration and network connectivity turns this into a challenging task.

The traceroute tool brings to the user a set of IP addresses in the path from the user to another destination host. Usually, the user gets this information as a path of connected IP addresses until reaching the destination host. At the moment, the majority of efforts in Internet topology have been focused on obtaining a huge set of traceroutes from the maximum number of probing stations and updating them periodically [1]. The resulting maps are graphs where each node is a single IP address and not a single router, so interfaces of the same router are drawn as different nodes.

However, this procedure is not very accurate because it can connect neighboring IP addresses which do

The authors thank the partial support of the EU ICT MOMENT Collaborative Project (Grant Agreement No.215225) and Spanish MEC project STRRONG (TEC2007-62192/TCM).

S. Garcia-Jimenez (✉) · E. Magaña · D. Morató · M. Izal
Universidad Publica de Navarra Depto. de Automatica y Computacion Campus Arrosadia,
31006 Pamplona, Navarra, Spain
e-mail: santiago.garcia@unavarra.es

E. Magaña
e-mail: eduardo.magana@unavarra.es

D. Morató
e-mail: daniel.morato@unavarra.es

M. Izal
e-mail: mikel.izal@unavarra.es

not share the link at all; both IP addresses can be at the same TTL distance but they can belong to different routers in different real paths. This kind of situation is caused by load balancing, MPLS paths, or topological changes. In load balancing, different packets in the same traceroute are sent thought different paths to the destination. Therefore, there is no absolute guarantee that IP addresses obtained from a traceroute belong to the same path, and therefore, it is impossible to infer a proximity relationship between those IP addresses. In MPLS paths, the behavior is similar to flow-based balancing, where packets belonging to different flows can be forwarded through different paths [2]. Topological changes happen in time scales of several days or weeks, larger than those considered for obtaining a topology (hours) [3, 4], so their effects are negligible. Therefore, the main effect will be load balancing (packet- or flow-based).

To solve this situation, there is a technique called Paris-traceroute [5]. This technique is based on the fact that the majority of load balancing in Internet is flow-based. Packets belonging to the same flow are forwarded through the same path, avoiding, for example, problems related with packet disordering or jitter. Paris-traceroute will send probing packets keeping certain header fields to make the routers think that they belong to the same flow (IP addresses, ports, and protocol).

In order to provide a more close-to-reality network graph, we need to identify nodes with routers and not with individual interfaces. Several studies already offer techniques to identify interfaces belonging to the same router from traceroute data sets. Those IP addresses that belong to the same router are called *aliases*, and the procedure is named *alias resolution method*. The most well known alias resolution methods are Mercator [6] and Ally [7] and are reviewed in Section 2. These alias resolution methods help in providing a router-level map of the Internet that will have application in network simulation, protocols design, routing protocols design, network management, security, service design, and geolocation.

Network simulation provides valuable performance indicators for several aspects of networks. Usually, network simulation needs to reproduce Internet topology, specially router placement and connectivity between routers. Traditionally, synthetic models have been used [8, 9]; however, they provide topological information that is unclear to be representative [7].

In [10], the impact in protocol design using synthetic and experimental Internet topologies is reviewed. It concludes that topology information can be used to make smarter and faster protocols. For example, in [11], topology maps at router level are used to test new traffic engineering protocols. Other applications of router-level maps of the Internet can be found in P2P balancing schemes [12] and routing protocols [13]. In P2P overlay networks, the topology information may be important in order to send the data using the shortest path in a router level point of view [14].

These maps can help the final user to know his placement in the network and, for example, to choose the right ISP connectivity each time. Similarly, they can be applied to choose the best place for servers [15]. Router-level topology maps are also important in network management tasks like detecting bottlenecks or misconfigurations [16]. In [4], prediction of the paths and latencies in the network is performed thanks to a previous topology discovering process combined with alias resolution methods.

Router-level topology maps can be used in security issues too. For example, the paper [17] presents a new way to discover the source network from where a denial-of-service attack is generated without the help of intermediate ISPs. A backtrace of the attack is made thanks to the router-level topology map.

Some IP geolocation techniques depend on the topology between some land points and the IP addresses to be located. The technique described in [18] needs the router-level map in order to offer more accuracy in the location of the IP addresses. This technique requires the geolocation of every router in the path from the IP address to be located and the chosen land points. Each new located IP in the path means a new constraint to be used in the final solution. If some aliases are found, they will help in the better location of the intermediate router.

Therefore, router-level topology maps are important, but they are also difficult to obtain. This paper focuses on the performance analysis of alias resolution methods. The following metrics can be used to compare alias resolution methods [19]: accuracy, completeness, efficiency, and distributability. Accuracy is related with the quantity of aliases resolved without error. The different alias resolution methods are usually applied for each pair of IP addresses in order to verify aliasing. For simplicity, the term *pair* will be used instead of pair of IP addresses. Over each pair, the result of the method can be positive or negative when those IP addresses belong or do not belong, respectively, to the same router. The methods can also provide incorrect information. When a method gives a positive result but it is not true, it could be said that we have a false positive. The same is true for the inverse situation, if a method provides a negative result and the pair of IP addresses belong to the same router, it is a false

negative. The difficulty of this metric is related with the lack of public information about details of most network topologies, so the debug phase is manual and with the possibility of wrong identifications. Besides, error conditions can arise when no response to probing packets is obtained from the IP addresses under analysis. Finally, an unknown condition is produced when there are incoherences or the impossibility to produce a final decission.

The completeness metric is the percentage of aliases (positives) and not aliases (negatives) that have been discovered. In a perfect identification scenario, both percentages should sum up to 100% to be sure that the network has been identified completely: all pairs have been indentified as aliases or not aliases. However, the presence of false positives, false negatives, and errors can reduce the final completeness.

Efficiency metric tries to determine how intrusive an alias resolution method is. Some methods need several probing packets for each pair of IP addresses, incurring in a sensible overload. Besides, this traffic, if significant, can be identified as network attacks because it is addressed to routers that usually are not destinations of Internet traffic. In large scale identifications, this parameter is critical, not only because of the traffic generated but also due to the time needed to complete the alias resolution method. The process should be limited to a feasible period of time. One way to reduce this time is by applying alias resolution methods not to the full set of pairs but only to selected pairs of IP addresses with more probability to be aliases. These techniques are called *reduction methods* [20], and they are a step previous to the application of alias resolution methods.

The last metric, distributability, is related to the possibility of a method of being distributed between several probing stations. A method is distributable if part of the alias identification or reduction tasks can be made from different probing stations simultaneously.

Therefore, three different phases can be identified in the process of getting a topology map at router level: discovery (obtaining IP addresses and adjacency information from traceroutes), reduction, and alias resolution. We will concentrate on the last two phases and their evaluation in real networks.

The rest of the paper is structured as follows. In Sections 2 and 3, reviews of alias resolution methods and reduction methods are presented. Then, core networks used in the study are presented in Section 4. Analyses of alias resolution methods and reduction methods in selected core networks are presented in Sections 5 and 6. Finally, conclusions are presented.

## 2 Previous works in alias resolution methods

During the past few years, a lot of effort has been done to discover new and more efficient alias resolution methods. Alias resolution methods can be divided in two main groups. The first one, active-probing-based, needs to send probing packets to the network and analyze the response packets. The second one, inference methods, use only the information provided by traceroutes to analytically infer the aliasing.

The first group of methods is intrusive, but at the moment, they are the ones that provide the best results in completeness and accuracy. The most known methods in this group are Mercator [6] and Ally [7].

Mercator was designed by the Co-operative Association for Internet Data Analysis group (CAIDA), and it was used in the Skitter project [21]. This method uses the usual behavior of routers when they send ICMP error messages: routers return ICMP error response messages from the interface with the shortest path to the destination. This ICMP error message (port unreachable) is triggered by sending UDP packets to random destination ports on the candidate IP addresses to be aliases. Two IP addresses are aliases if the ICMP error messages returned from both have the same source IP address. As can be seen in Fig. 1, UDP packets are sent to two IP addresses of the same router, and the ICMP error responses are sent by the same interface of the router, this means, with the same source IP address.

Mercator can be used only to identify aliases and not to identify non-aliases because the commented behavior is not present in all routers, or even worse, it can apply to some interfaces of the router and not to other interfaces of the same router. Therefore, if two different target IP addresses return packets from the same IP address, it is for sure that both belong to the same router, but if responses come from different IP addresses, it cannot be said that both belong to different routers.

Ally method is based on another characteristic behavior of routers. Most IP stack implementations have an incremental counter to provide different identifiers for each IP packet generated by the router independently of destination, protocol, or service. This is the IP identification (IPID) field of IP header, and it is used by the fragmentation and reassembly mechanisms. Each packet from the same IP address must have differentiable IPIDs in certain time windows [22]. Therefore, several IP packets received from the same router and near in time will have close values in the IP identifier field. The difference in the counter will be caused by other IP traffic generated in between by that router to other destinations. Ally method checks

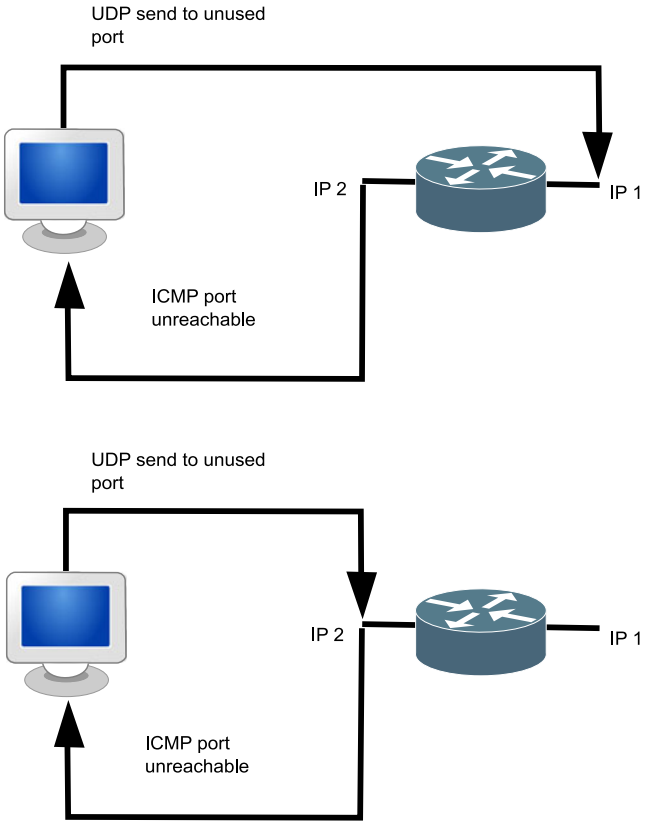


**Fig. 1** Mercator alias resolution method

two candidate IP addresses by sending three UDP probes with random ports to trigger again an ICMP error message whose IP identifiers can be analyzed (Fig. 2). The first packet is sent to one IP address, and back-to-back, the second packet is sent to the other IP address. One second later, the third packet is sent to the IP address that answered in the first place. The IPID values contained in the three response packets are analyzed. Both target IP addresses will be aliases if the distance between IP identifiers in the responses is in between a threshold of 200 sequence numbers [7]. Ally provides the best results for completeness and accuracy in alias resolution [23].

An example of IPIDs received can be seen in Fig. 3. In this figure, IPIDs have an incremental sequence, and the distance between the IPIDs does not exceed 200. Therefore, the two IP addresses are considered to be aliases. Ally method needs an incremental behavior of IPID, but some IP stack implementations use random values or other strategies. However, this incremental behavior is the most usual. For example, in the analysis of [23], 82% of routers had an incremental counter for IPID using UDP as probing packets.

Improvements over Ally method are proposed in [23]. The improvements do not suppose a huge change

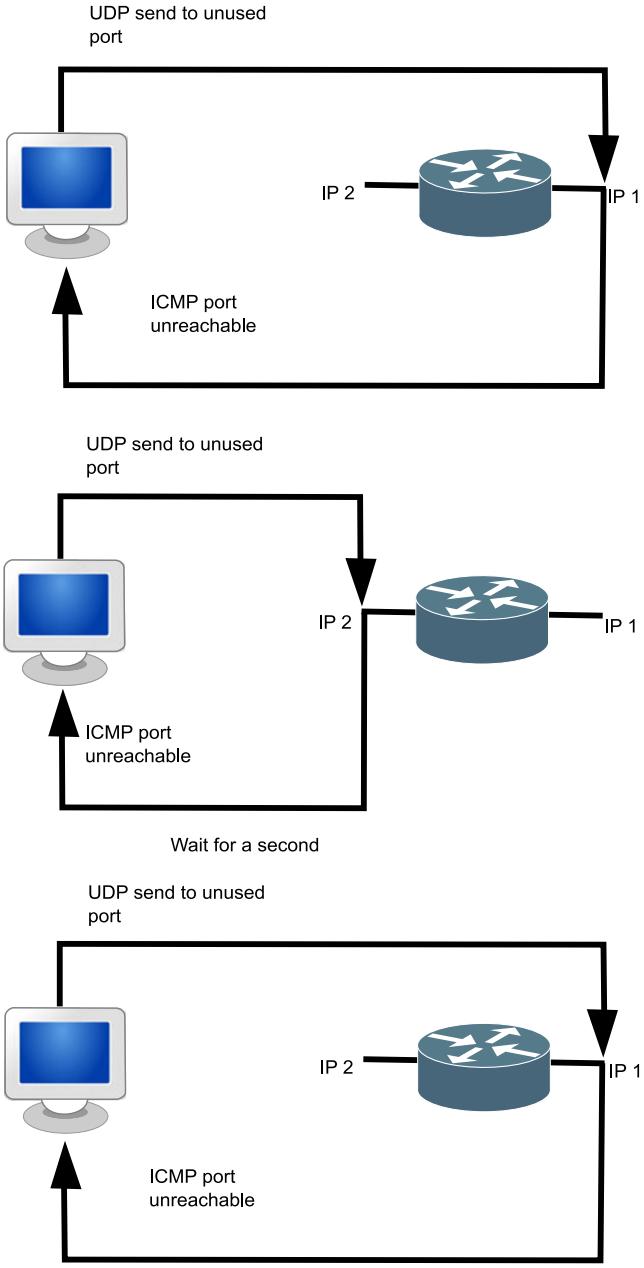


**Fig. 2** Ally alias resolution method

in Ally method but provide a clear increment in the identification percentage. Part of the modifications are related to the type of probing packets being used. In Ally method, only UDP probing packets and ICMP error responses are used. In [23], new probing packets are used:

- IPID_TCP: TCP SYN packets are sent to random ports in destination IP address, and IPID will be got from reset packets generated by the router.
- IPID_ECHO: ICMP Echo Request packets are sent and IPID will be got from the ICMP Echo Reply generated by the router.

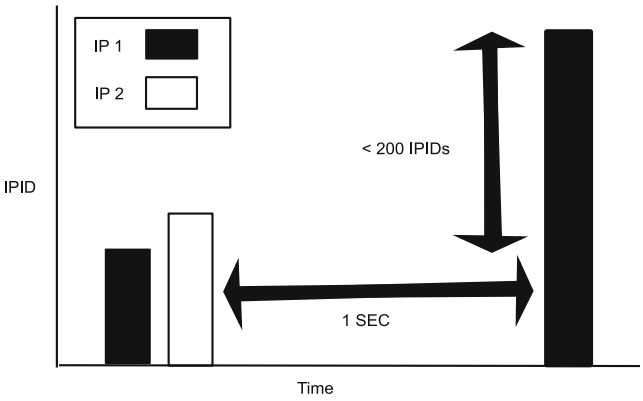


**Fig. 3** IPID threshold in received packets for Ally alias resolution method

- IPID_TIME: ICMP Timestamp Request packets are sent and the IPID obtained from the ICMP Timestamp Reply generated by the router.
- IPID_UDP: It will use the same probing packet as Ally, but with an increase in the number of probing packets.

These new types of probing packets provide better results, avoiding some of the more strict filtering rules present nowadays in routers [23]. Another modification is related with the timing. Some routers discard back-to-back Ally packets, treating them as some kind of flooding attack. In other cases, only one of the two first packets is replied. In the proposal [23], a static offset of 0.6 s is used between the three packets (Fig. 4). This offset is chosen to guarantee that most of the routers have time to reply to the probing packet.

As analyzed in [23], there is a probability around $10^{-4}$ of false positives in Ally because of routers that occasionally share the same IPIDs in the probing interval. Increasing the number of probing packets reduces the probability of wrong identification (there is a lower probability to share the same IPIDs during an extended probing interval) but also reduces the efficiency (more probing packets are needed).

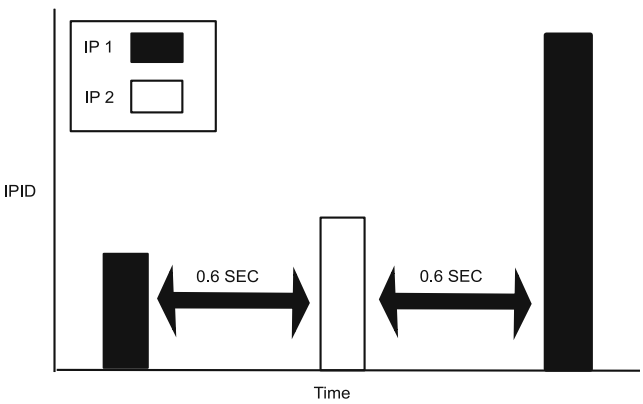


**Fig. 4** Timing for modified Ally alias resolution method

In the second group of alias resolution methods, those called inference methods, Analytical Alias Resolver (AAR) [24] and Analytical and Probe-based Alias Resolver (APAR) [25] are the most representative. Both are based on the same premise: routers have dedicated IP subnetworks to configure their interconnection links. These methods are based on finding the /30 and /31 network masks from the information obtained in traceroutes. A link is detected looking for a pair of IP addresses obtained from opposite traceroutes that verify one of the previous network masks. The IP address before and after this link will be used to identify an alias. The APAR method uses larger network masks, but this time, accompanied by a low rate of probing packets. Probing packets are based on PING (ICMP Echo Request/Reply) and the TTL returned in the reply is used to verify if the alias has been detected correctly.

## 3 Previous works in reduction methods

In order to improve the performance of alias identification, reduction methods are able to decrease the number of pairs to test for aliasing by using any extra information available before performing the alias resolution. This allows to reduce the probing traffic and the time needed to complete the alias resolution (improving efficiency). However, it can reduce completeness of alias resolution because some pairs of IP addresses (possible aliases) are discarded before applying any alias resolution method. The objective will be to keep some balance between those metrics: completeness and efficiency.

The first well-known reduction method uses the Time-to-Live (TTL) field of IP headers. This method determines that the alias resolution method has to be applied only for those pairs where TTL difference between IP addresses is less than a predefined threshold [26]. TTL information could not be obtained from traceroute because it has to be measured from the same probing station in order to be able to compare TTL values for different target IP addresses. A router is located at a certain distance in hops from a specified probing station, so if two IP addresses belong to the same router, they should be approximately at the same number of hops. Probing packets can go through different paths to different IP addresses even if both IP addresses belong to the same router. Therefore, the method defines a threshold for these cases.

TTL field is filled up with an initial value that could be different for each IP stack implementation, so the differences between two IP addresses belonging to different routers can be higher. As stated before, to use this method, it is needed to make the probing from one unique point to all the IP addresses in the set under identification. Therefore, this method requires to inject extra probing traffic, and it is not distributable.

In [20], another reduction method is proposed. It is based on the IPID field in IP headers. The idea is similar to the proposal in Ally alias resolution method. If we compare the distance between IPIDs in return packets from two target IP addresses, the pairs with less distance are more likely to be aliases than the ones that have a higher distance. The probing traffic needed by this method could be distributed into different probing stations if we would want to complete the probing in a shorter time window. The method needs the IPIDs generated by all IP addresses approximately in the same short time interval in order to calculate distances between all possible pairs. However, in large networks, it is not practical to distribute the measuring because the probing can take several minutes and, therefore, makes it impossible to calculate a realistic IPID distance between IP addresses. For example, it is usual to find routers in the Internet whose IPID counter increments 300 IPIDs per second (the router generates 300 paq/s to some destinations; it is not a forwarding rate). It means that, in only $2^{16}/300 = 218$ s, the IPID counter restarts, and the IPID increment is enough to be considered as another router in only some seconds.

The last reduction method presented in [27] does not need to inject probing traffic in the network, and therefore, it is efficient and easily distributable. The method is based on the characteristic distribution of IP addressing in Internet routers presented in [28]. IP-offset parameter is defined as the absolute value of subtracting one IP address from another $|IP1 - IP2|$, considered both as 32-bits unsigned integers. Due to the Internet organization in AS and the addressing allocated to each AS, IP addresses that belong to the same router have concrete IP-offset ranges that can be used to predict which IP addresses have more probability to be aliases. IP-offsets with more probability to be aliases are centered around 0 and $2.14 * 10^9$ (a half of IP addresses range $2^{32}$) as stated in [28]. A clustering algorithm is used to find those and other IP-offset intervals that imply significant probability of aliasing.

In order to select the intervals, clustering algorithms based on IP-offset parameter are used. K-means and Expectation Maximization (EM) clustering algorithms are used to group IP-offset distances in clusters [27]. Afterwards, those clusters are ordered, choosing first those with more probability to indicate aliasing (those IP-offset intervals that have a larger proportion of aliases compared to the total).

One advantage of the IP-offset reduction method is that IP-offset ranges with more probability to indicate aliases are maintained for different kinds of networks. This means that IP-offset ranges with more probability to be aliases can be pre-calculated in certain networks, and those intervals are valid in another totally different network. In [28], the training network has close to 100 thousand pairs, and clusters generated are perfectly applicable to another network with close to 1.5 billion pairs. Calculating the clustering in the same network where aliasing is going to be performed only provides a marginal improvement. Therefore, the way to proceed in order to obtain those intervals is to make the training in a relatively small network.

The results in IP-offset reduction method improve those obtained by other reduction methods, with the added advantages of efficiency (no probing traffic is needed) and distributability (without probing traffic, the distribution of the IP-offset reduction method between different probing stations is trivial). In the three reduction methods, subsets of pairs of IP addresses with more probability to be aliases are identified without a priori knowledge of the actual topology.

## 4 Core networks used in the study

One of the big problems in verifying topology maps is the lack of public information about network configuration, addressing, equipments, interfaces, and links present in each administrative unit in which the Internet is organized: the so called AS. Previous analyses have been made using testbeds [23] that provided results with limited applicability. Nodes in Etomic [29] and Planetlab [30] platforms have been used to discover routers in the path between end-nodes and to apply identification and reduction methods. However, information about networks that interconnect those nodes is not available, so there is no certainty about the accuracy of the results.

A few core networks, mainly National Research and Education Networks (NRENs), make publicly available information about its network configuration. GlobalNOC, Canet4 and Geant are examples of these. They provide public information about their routers and their configuration. For our research, the provided "show interfaces" functionality will specify IP addressing configured for all interfaces in a router. All those IP addresses for certain routers are aliases, and they

should be identified as aliases when running the alias resolution methods.

We will use this real addressing information per router to check completeness of alias resolution methods. However, these networks are not representative of the whole Internet because they are core networks that do not include information about access networks. They are suitable for the analysis of accuracy and completeness of alias resolution methods. For the analysis of performance of reduction methods, they could offer biased results, specially for those methods that take into account how IP-addressing is organized in the Internet. In those cases, we will use information from Etomic and Planetlab that provide end-to-end addressing data, including access networks.

GlobalNOC [31] is the Global Research Network Operations Center, located in the USA. This center provides network coordination, engineering, and installation services. It has a huge set of partners, like Internet2, CIC OmniPOP, I-Light, Indiana GigaPOP, IP Grid, Man LAN, National LamndaRAil, and TransPAC2. GlobalNOC provides information about their routers. Using a web application called GlobalNOC Router Proxy [32], a user can get a lot of information about the state of the network in real time. For example, information about the IP interfaces of the routers can be obtained. Internet2 is one of the main partners in GlobalNOC. Its topology map is shown in Fig. 5. This study takes into account all the networks in GlobalNOC. This represents 16 routers and 593 IP addresses.

Canet4 [33] is Canada's Research and Education Network. The CANARIE group has developed this ultra high-speed optical network across Canada using links at 10 Gbps. The information of IP addresses for some routers in this network can also be accessed using a web application [34]. A global map of this network is shown in Fig. 6, with six routers and 103 IP addresses.

**Fig. 6** Canet4 map

Geant [35] is the pan-European data network dedicated to the research and education community. In conjunction with the European National Research and Education Networks (NRENs), this network is supposed to connect 40 million users. The information about their routers can be accessed using a web-based looking glass. The locations of the routers, which can be queried through the looking glass, are shown in Fig. 7, with 19 routers and 309 IP addresses.

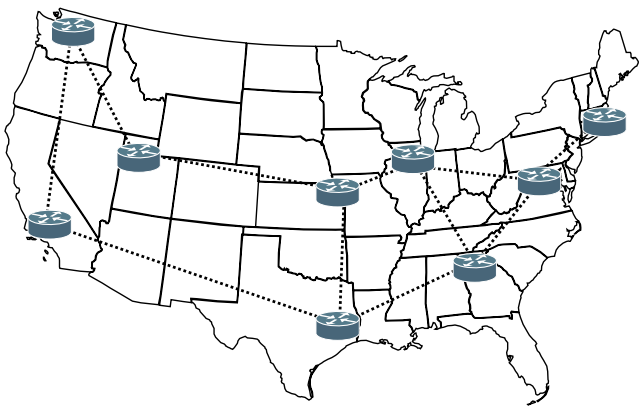

**Fig. 5** Internet2 map in GlobalNOC

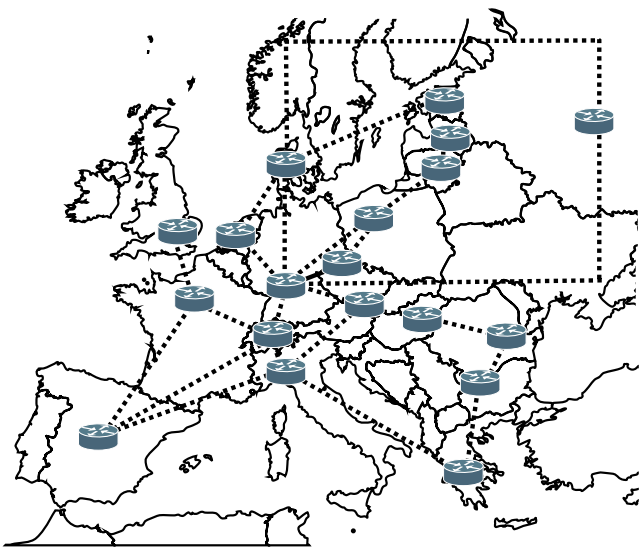

**Fig. 7** Geant map

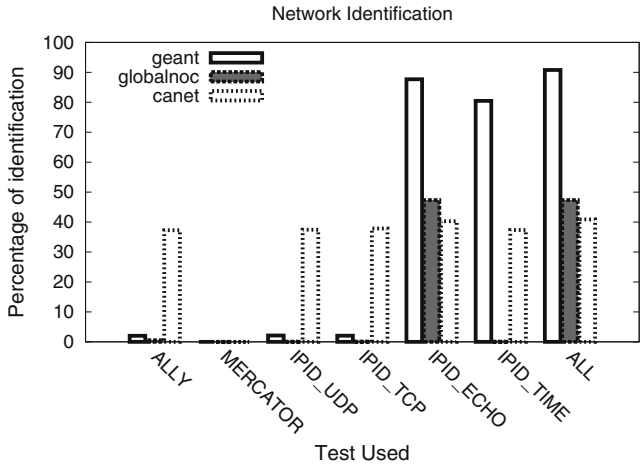


**Fig. 8** Alias resolution identification results for different core networks

These well-known networks will be used in the following sections. First, they will provide information to verify the accuracy and completeness of alias resolution methods. These networks provide all the information about the IP addresses that belong to the same router (aliases). This information will allow to verify performance of alias resolution methods. Second, they will be used to analyze performance of reduction methods in core networks and, with an extension, in full end-to-end networks. As we have a priori knowledge of topology information and we are focusing our effort in alias resolution and reduction methods, discovering phase, based usually in doing traceroutes from some vantage points, is not needed. All the IP addresses for routers in each network are obtained directly from topology information known a priori.

## 5 Analysis of alias resolution methods

Most alias resolution methods described in Section 2 need to check aliasing between pairs of IP addresses. Therefore, identification tests are run over all possible pairs of IP addresses present in each core network (datasource). Each datasource has been treated as a different network, and pairs of IP addresses are calculated independently for each datasource.

The probing task for each method has been distributed among several Planetlab nodes. A special software for the distribution of the pairs of IP addresses to test and to make the identification part was developed. Besides, specific attention has been made in order to avoid simultaneous tests to the same IP address. This will avoid router overload and interference between tests. All data sets and software developed for this paper are available in [36].

The results are presented in Fig. 8. The figure shows the comparison of alias resolution methods and the percentage of identification (completeness) provided in each core network. All the identification responses have been counted in the graph (positives and negatives responses). In *x* axis, all alias resolution methods are plotted. In the *y* axis, the percentage of identification per method and per network is shown. The column named ALL refers to the combination of the results from all methods, providing the best identification ratio.

Combining all methods, around 90% of identification is obtained in the Geant network, but this identification lowers down to 50% or even less for GlobalNOC and Canet4. These low percentages of identification are mainly produced by unresponsive interfaces and the filtering of some kinds of packets in the networks. The specific reasons will be provided in the following paragraphs.

Detailed information for identification in each core network is presented in Tables 1, 2, and 3. The first columns present the percentage of positives, negatives, false negatives, unknowns (not enough data obtained by the method to provide a conclusion), and errors (none or not enough number of response packets received) for each method. IPID_ECHO provides the best results in the alias resolution. For the last two columns, number of nodes and percentage of total identification by alias resolution, results of each method

**Table 1** Details of alias resolution results over Geant network

| Method | Positive % | Negative % | False negatives % | Unknown % | Error % | Number of nodes | Total identified % |
|---|---|---|---|---|---|---|---|
| Mercator | 0 | 0 | 0 | 100 | 0 | 318 | 0 |
| Ally | 0.089 | 1.944 | 0.000065 | 0 | 97.966 | 288 | 2.033 |
| IPID_UDP | 0.104 | 2.003 | 0 | 0 | 97.892 | 287 | 2.217 |
| IPID_TCP | 0.111 | 1.939 | 0 | 0 | 97.948 | 286 | 2.348 |
| IPID_ECHO | 4.947 | 82.745 | 0 | 0 | 12.306 | 48 | 87.869 |
| IPID_TIME | 4.532 | 75.940 | 0 | 0 | 19.526 | 47 | 90.790 |

**Table 2** Details of alias resolution results over Canet4 network

| Method | Positive % | Negative % | False negatives % | Unknown % | Error % | Number of nodes | Total identified % |
|---|---|---|---|---|---|---|---|
| Mercator | 0 | 0 | 0 | 99.952 | 0.047 | 104 | 0 |
| Ally | 6.566 | 40.186 | 0.0002 | 0.0477 | 53.199 | 34 | 46.752 |
| IPID_UDP | 7.211 | 39.708 | 0 | 0 | 53.080 | 31 | 47.540 |
| IPID_TCP | 7.521 | 39.947 | 0 | 0 | 52.531 | 30 | 50.214 |
| IPID_ECHO | 8.213 | 42.335 | 0 | 0 | 49.450 | 30 | 51.289 |
| IPID_TIME | 6.661 | 35.792 | 0 | 0 | 57.545 | 30 | 51.289 |

are aggregated with the previous ones: several methods can be used simultaneously to improve the alias resolution.

In the results, false positives are not present for all the methods and core networks. This is a very interesting result, showing that different routers can present quite different behaviors in relation to the indicators used by the alias resolution methods. Those indicators are enough to distinguish when two IP addresses belong to different routers. However, in extensive tests, a very low rate of false positives can be found due to IPID synchronization between routers and random behaviors in IPIDs generation [23]. Therefore, for small networks, any method can be used without having a significant effect of false positives because this probability is quite low.

Problems related to false negative aliases have appeared in the identification stage. In Geant and GlobalNOC networks, three false negatives are provided by Ally method. In Canet network, 25 false negatives are provided by Ally method. In all cases, the other alias resolution methods do not provide false negatives. This evaluation has been made by checking for coherence between different alias resolution methods, without using public information provided by network administrators.

Peculiar router behaviors have produced some incoherences that were located and solved. To check the reliability of an alias resolution method, if we use the public information provided by network administrators, we can find more false negatives than those described before: 126 false negatives. From all these pairs, 90 belong to Internet2 (globalNOC). This happens when one or both IP addresses of the pair correspond to interfaces with *private peering* tag. This tag is configured to some interfaces in point-to-point links between routers, making those addresses not reachable outside the same router. Therefore, this tag makes it impossible to reach the interfaces by the usual traceroute and probing packets in alias resolution methods. Private peering is not a problem in alias resolution. However, as those interfaces are not discovered, the topology map will not include links with private peering. In our study, the presence of private peering in Internet2 (GlobalNOC) network is only 2.9% of the total number of links. For this kind of configuration, usually private addressing is used, but in some cases, public IP addresses are used instead of private IP addressing.

The way to check if we have any pair badly cataloged as negative (false negatives) and using public addressing is to search for IP addresses that are aliases of those false negatives and not belonging to the routers under study. Remember that we have full information for each network scenario obtained from the "show interfaces" functionality at each router. These false negatives indicate that we are reaching different routers in the pair and that the test is correct: the IP addresses we are reaching do not belong to the same router. They are true negatives. Also, we have observed that all false negatives found in globalNOC network are caused by this behavior and they are not really false negatives, except three cases obtained by Ally method. For these three cases, other alias resolution methods provide the right identification.

**Table 3** Details of alias resolution results over GlobalNOC network

| Method | Positive % | Negative % | False negatives % | Unknown % | Error % | Number of nodes | Total identified % |
|---|---|---|---|---|---|---|---|
| Mercator | 0 | 0 | 0 | 100 | 0 | 574 | 0 |
| Ally | 0.030 | 0.038 | 0.0007 | 0 | 99.930 | 559 | 0.069 |
| IPID_UDP | 0.031 | 0.036 | 0 | 0 | 99.931 | 559 | 0.069 |
| IPID_TCP | 0.033 | 0.048 | 0 | 0 | 99.918 | 557 | 0.084 |
| IPID_ECHO | 4.659 | 48.321 | 0 | 0 | 47.018 | 103 | 52.985 |
| IPID_TIME | 0.060 | 0.498 | 0 | 0 | 99.440 | 103 | 52.985 |

**Table 4** Percentage of duplicated IP addresses present in scenarios

| Network | Duplicated IP addresses (%) |
|---|---|
| Geant | 0 |
| GlobalNOC | 4.21 |
| Canet | 4.85 |

Another peculiar behavior observed in the study has been the existence of *duplicated addresses*. This means that different routers share the same IP address for some of their interfaces. This practice, present for example in anycast deployments [37], can lead to a mis-identification of aliases. There is no way to know which is the real router we are receiving the probing packet from. Depending on several factors, false negatives and false positives are obtained because, with one IP address, we can be referring to different routers. From an identification point of view, it is not a failure of the alias resolution method. The problem is that, from the beginning, alias resolution methods assume the premise of one unique IP address per interface, and this premise, as we see, is not always true.

Duplicated addresses are not abundant in the Internet. They are present mainly in core networks. For the networks under study, the percentage of duplicated IP addresses is presented in Table 4.

## 6 Analysis of reduction methods

Reduction methods reviewed in Section 3 are analyzed for the proposed core networks. The main idea is to reduce the number of pairs to check for aliases without losing accuracy and completeness in the result. TTL-based and IPID-based methods are implemented with different thresholds. In order to get better alias resolutions results, more pairs are added to the identification task.

In TTL-based reduction methods, a different set of IP addresses will be considered for each TTL threshold chosen. This means that we will check for aliasing only in those IP addresses with a difference in the number of hops (measured from the probing station) up to TTL threshold. The larger the TTL threshold, the more pairs of IP addresses are considered to check for aliasing, and therefore, it increases the probing packets injected in the network, the time needed to complete the analysis and the computational cost. However, at the same time, better results in completeness are obtained as TTL threshold is increased. In Fig. 9, results are presented for the three core networks assuming a perfect alias resolution and, therefore, only the effect of the reduction method in completeness. Not all IP addresses answer the probing packets, so there is no information about reduction for a significant number of interfaces. This and the following figures have been plotted for the total number of pairs in order for it to be easy to compare them.

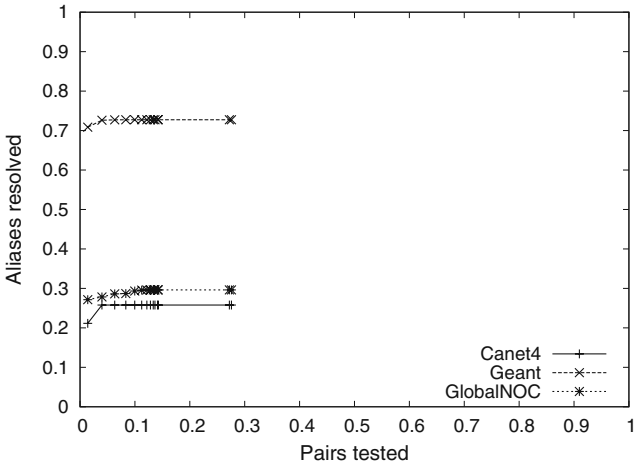


**Fig. 9** TTL-based reduction method over core networks

In IPID-based methods, the idea is similar. In this case, the IPID threshold will limit the maximum difference between IPIDs for packets generated from IP addresses to be considered to check for aliasing. As IPID threshold is increased, the alias resolution method has to be applied to more pairs of IP addresses, increasing again the cost of the identification. However, we can choose a low IPID threshold, reducing the number of pairs to check, and reducing minimally the effect on completeness. In Fig. 10, results are presented supposing a perfect alias resolution

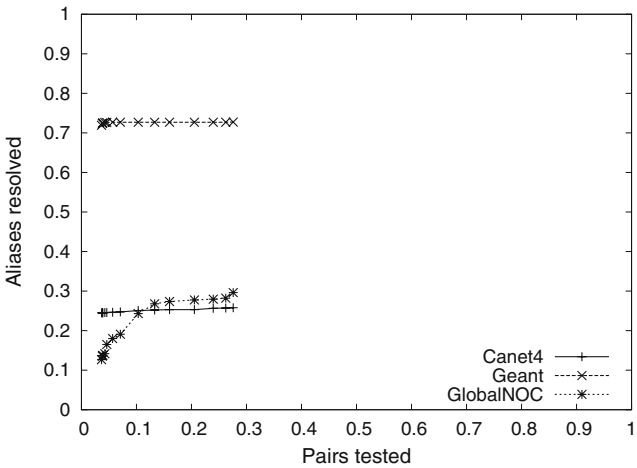


**Fig. 10** IPID-based reduction method over core networks

and, therefore, only the effect of the reduction method in completeness. As before, the effect of lack of responses for a part of probing packets is reflected in the figure.

In IP-offset reduction method, an EM clustering algorithm [38] is used to classify IP addresses with more probability to be aliases [27]. The IP addressing organization provides a specific behavior in this sense, and it is independent of the network under study [28]. In our case, we use Etomic to obtain the cluster definition. Etomic measurement infrastructure is composed by 18 end nodes around Europe that are used to obtain traceroute information between them. EM clustering algorithm is applied over pairs of IP addresses that are aliases and that are characterized by certain IP-offsets. Resulting clusters are used on the proposed core networks.

Results of applying IP-offset reduction method for each core network are presented in Fig. 11. The $x$ axis indicates the percentage of pairs to be considered to check for aliasing (the lower the better), and the $y$ axis indicates the proportion of alias resolution identification (the larger the better). The combination of all alias resolution methods reviewed in Fig. 8 is used this time. The first impression is that the results are not good enough. Only in GlobalNOC is a better identification rate provided with less percentage of pairs. This has an explanation: those routers in core networks that usually share a common addressing scheme make it difficult to use IP addressing to infer those IP addresses with more probability to be aliases. In a real topology discovering process, routers in access networks and different core networks should be traversed, providing more variability in the IP addressing schemes used and

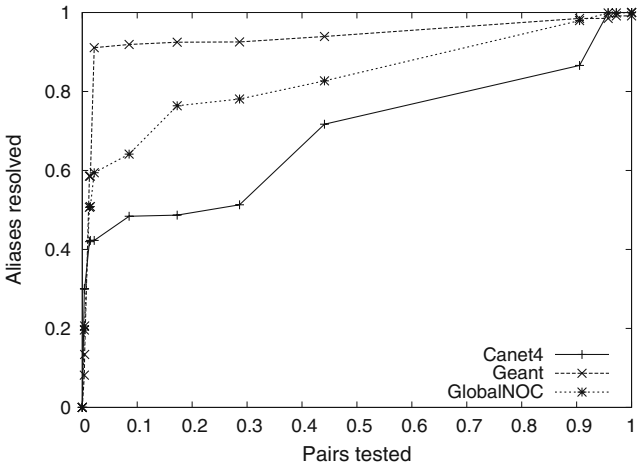


**Fig. 12** IP-offset reduction method results using Etomic clustering over access+core networks

therefore making useful this reduction method [28]. Using 50 Planetlab nodes as probing stations, traceroutes were made between all of them. The obtained IP addresses in first hops (access routers) are merged with those in the core networks under study, obtaining the results presented in Fig. 12. The results are improved for the three core networks, specially for Canet4, being a more realistic full scenario. With around only 20% of pairs, we achieve alias resolution from 42% to 91%.

In Fig. 13, the results for IP-offset reduction method with clusters obtained in Etomic scenario are compared to specific clustering results obtained from each core network. In the specific clustering, the EM clustering algorithm is run using the data from the network scenario under study. As explained before, the benefits of specific clustering are not meaningful, so a general

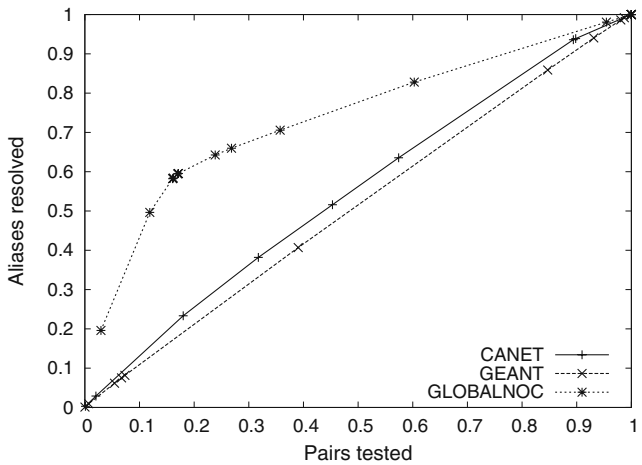


**Fig. 11** IP-offset reduction method results using Etomic clustering over core networks

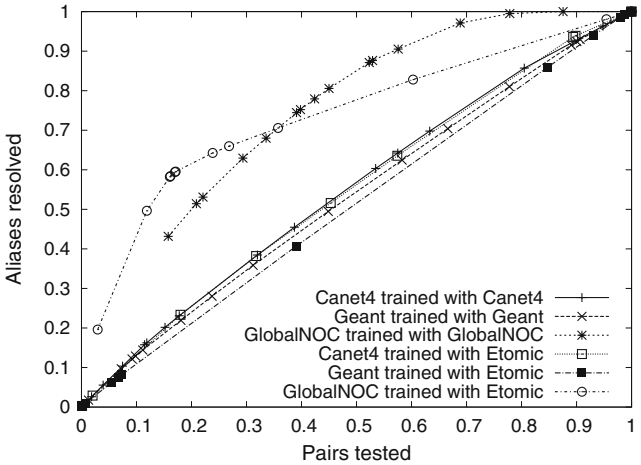


**Fig. 13** IP-offset reduction method results using different clustering strategies

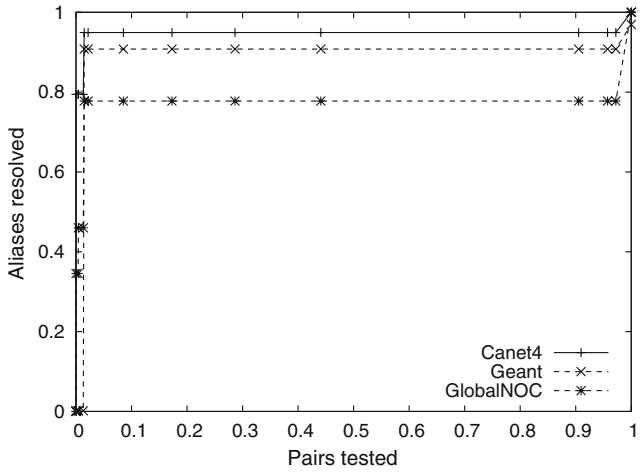


**Fig. 14** IP-offset reduction method results using Etomic clustering over core networks and perfect alias resolution

strategy of using a pre-calculated clustering is possible in the IP-offset reduction method. This will allow the reduction of the overall computational cost. In our study, the pre-calculated clustering comes from the Etomic network.

Lack of completeness in Fig. 12 is caused by alias resolution methods and by effects of the reduction method. As we have perfect information of addressing for each router under study, we can simulate the effect of considering a hypothetical perfect alias resolution method (100% accuracy). The results are shown in Fig. 14. In this figure, the lack of completeness is caused only by the reduction method (IP-offset reduction method in this case). With only 1.47% of pairs, maximum achievable alias resolution rate is 94.93% in Canet4, 90.78% in Geant, and 77.75% in GlobalNOC.

## 7 Conclusions

This paper has analyzed the results of alias resolution and reduction methods over three well-known core networks (GlobalNOC, Canet4, and Geant). Those networks provide information about their topology, making them useful to verify how alias resolution procedures work. Identification methods, especially those improvements over Ally method, provide a high percentage of completeness and good accuracy if we compare them with Mercator and Ally methods. Actually, if only one method should be selected in order to get the best results in alias resolution, the one should be ICMP_ECHO method. In this case, ICMP Echo Request probes are sent and ICMP Echo Response packets are analyzed to check for IP identifier values. For example, in Geant network, Ally method obtains only a 2.03% of identification and ICMP_ECHO method obtains a 87.86%.

Reduction methods provide information to determine if two IP addresses are likely to be aliases. Using this information, the number of tests can be reduced. Depending on the desired reduction factor, different percentages of completeness will be obtained. IP-offset reduction methods provide the best results in the core networks under study, avoiding any additional probing packet generation. This means better efficiency (less time to complete the full identification process) and the possibility to make the identification in a fully distributed way. We can reach reduction rates with around 90% of completness by using only 1% of pairs.

Specific addressing schemes in core networks have their effect in alias resolution and reduction methods. Special behaviors have been identified as private peering, duplicated public IP addresses and private IP addressing in router interfaces. Low speed interfaces have appeared also in the study and they have caused wrong alias resolutions. Improvements have been made to alias resolution and reduction methods in order to consider these peculiar behaviors.

## References

1. McRobb D, Claffy K, Monk T (1999) Skitter: CAIDA's macroscopic Internet topology discovery and tracking tool. http://www.caida.org/tools/measurement/skitter/. Accessed July 2010
2. Zhao Z, Shu Y, Zhang L, Oliver W, Yang W (2005) Flow-Level mMultipath load balancing in MPLS network
3. Rexford J, Wang J, Xiao Z, Zhang Y (2006) Bgp routing stability of popular destinations. In: Proceedings of the 2nd ACM SIGCOMM workshop on internet measurment, pp 197–202
4. Madhyastha HV, Anderson T, Krishnamurthy A, Spring N, Venkataramani A (2006) A structural approach to latency prediction. In: Proc. USENIX internet measurement conference
5. Augustin B, Cuvellier X, Orgogozo B, Viget F, Latapy M, Friedman T, Magnien C, Teixeira R (2006) Avoiding traceroute anomalies with paris traceroute. In: 6th ACM SIGCOMM, pp 153–158
6. Govindan R, Tangmunarunkit H (2000) Heuristics for internet map discovery. In: Proc. IEEE INFOCOM
7. Spring N, Mahajan R, Wetherall D (2000) Measuring ISP topologies with Rocketfuel. In: Proc. ACM SIGCOMM
8. Calvert K, Doar M, Zegura EW (1997) Modeling internet topology. IEEE Commun Mag 35:160–163
9. Medina A, Matta I, Byers J (2000) On the origin of power-laws in internet topologies. ACM Computer Communication 30:18–28 Review

10. Radoslavov P (2001) The relationship between topology and protocol performance: case studies. Technical report, University of Southern California, Computer Science Department
11. Shi TJ, Mohan G (2006) An efficient traffic engineering approach based on flow distribution and splitting in MPLS networks. Comput Commun 29(9):1284–1291
12. Mirrezaei SI, Shahparian J, Ghodsi M (2009) A topology-aware load balancing algorithm for p2p systems. In: Digital Information Management ICDIM 2009. pp 1–6
13. castro M, Druchel P, Hu YC, Rowstron A (2002) Topology-aware routing in structured peer-to-peer overlay networks. Tech Rep
14. Garces-Erice L, Ross KW, Biersack EW, Felber PA, Urvoy-Keller G (2003) Topology-centric look-up service. In: Proc. COST264/ACM fifth international wORKSHOP on networked group communications
15. Radoslavov P, Govindan R, Estrin D (2001) Topology-informed internet replica placement. Sixth International Workshop on Web Caching and Content Distribution, pp 229–238
16. Siamwalla R, Sharma R, Keshav S (1998) Discovering internet topology. Tech Rep
17. Burch H, Cheswick B (2000) Tracing anonimous packets to theis approximate source. USENIX conference on System administration
18. Katz-Bassett E, John JP, Krishnamurthy A, Wetherall D, Anderson T, Chawathe Y (2006) Towards IP geolocation using delay and topology measurements. In: Proc. USENIX internet measurement conference
19. Pansiot J-J, Grad D (1998) On routes and multicast trees in the internet. ACM SIGCOMM Computer Communication Review
20. Burch H (2005) Measuring an IP network in situ. PhD thesis, Carnegie Mellon University. ISBN 0-542-01549-8
21. Huffaker B, Plummer D, Moore D, Claffy KC (2002) Topology discovery by active probing. In: Proc. the symposium on applications and the internet (SAINT)
22. Postel J (1981) Rfc 791—internet protocol
23. Garcia-Jimenez S, Magaña E, Morato D, Izal M (2009) Techniques for better alias resolution in internet topology discovery. In: Published in 11th IFIP/IEEE international symposium on integrated network managemen miniconference. New York, USA
24. Gunes M, Sarac K (2006) Analitycal IP alias resolution. ICC '06. IEEE International Conference
25. Gunes M, Sarac K (2006) Resolving IP aliases in building traceroute-based internet maps. Technical report, University of Texas at Dallas
26. Spring N, Dontcheva M, Rodrig M, Wetherall D (2004) How to resolve ip aliases. Tech. Report 04-05-04, Washington Univ. Computer Science
27. Garcia-Jimenez S, Magaña E, Morato D, Izal M (2009) Improving efficiency of ip alias resolution based on offets between ip addresses. In: Published in 21st international teletraffic congress (ITC 21). Paris, France
28. Garcia-Jimenez S, Magaña E, Izal M, Morato D (2009) Ip addresses distribution in internet and its application on reduction methods for ip alias resolution. In: Published in the 4th IEEE LCN workshop on network measurements (WNM 2009). Zurich, Switzerland
29. Morato D, Magaña E, Izal M, Aracil J, Naranjo F, Astiz F, Alonso U, Csabai I, Haga P, Somin G, Seger J, Vattay G (2005) The European traffic observatory infraestruCture (ETOMIC): a testbed for universal active and passive measurements. In: Proc. TRIDENTCOM 2005, pp 283–289
30. Plantelab (2010) An open platform for developing, deploying, and accessing planetary-scale services. http://www.planet-lab.org. Accessed July 2010
31. Globalnoc (2010) Official site. http://globalnoc.iu.edu/. Accessed July 2010
32. Globalnoc (2010) Looking glass tool. http://routerproxy.grnoc.iu.edu/. Accessed July 2010
33. Canarie (2010) Official site. http://www.canarie.ca/en/home. Accessed July 2010
34. Canet4 (2010) Looking glass web tool. http://dooka.canet4.net/lg/lg.php. Accessed July 2010
35. Geant (2010) Official site. http://www.geant.net/pages/home.aspx. Accessed July 2010
36. Garcia-Jimenez S et al (2010) Tools and data sets used in this paper. http://www.tlm.unavarra.es/~santi/research/paper6.html. Accessed July 2010
37. Abley L (2006) Rfc 4786—operation of anycast services
38. Laird NM, Rubin DB, Dempster AP (1977) Maximum likelihood from incomplete data via the em algorithm. J R Stat Soc, Ser B 39(1):1–38