Pamplona-traceroute: topology discovery and alias resolution to build router level Internet maps

Santiago Garcia-Jimenez, Eduardo Magaña, Daniel Morató and Mikel Izal
Public University of Navarre, Campus Arrosadia, 31006 Pamplona, Spain
{santiago.garcia, eduardo.magana, daniel.morato, mikel.izal}@unavarra.es

Abstract—**An Internet topology map at the router level not only needs to discover IP addresses in Internet paths (traceroute) but also needs to identify IP addresses belonging to the same router (IP aliases). Both processes, discovery and IP alias resolution, have traditionally been independent tasks. In this paper, a new tool called Pamplona-traceroute is proposed to improve upon current results in a state of the art for Internet topology construction at the router level. Indirect probing using TTL-scoped UDP packets, usually present in the discovery phases, is reused in IP alias resolution phases, providing high identification rates, especially in access routers.**

I. Introduction

The measurement of Internet topologies can be performed at different levels of aggregation depending on the final application. These levels include the Autonomous System (AS) level and the router level (IP level). In all cases, a first phase is in charge of collecting real data about *topology configuration*. This phase is called the topology collection phase. A second phase is in charge of processing previous data and obtaining extra information to provide the topological view at the desired aggregation level. This second phase is called *topology construction*.

One of the most interesting aggregation levels for Internet topology is the router or IP-level because routers are the main interconnection elements between networks. This router level map should provide information about routers, the interfaces of those routers and the links between routers. The router level map has a multitude of applications in networking problems: routing protocols, delay measurements, protocol performance, network management and, of course, network simulation.

However, the lack of public and systematic information about routers and router interconnections in the Internet makes the measurement of Internet topologies at the router level a challenging task. The topology collection phase has traditionally been based on the traceroute tool [1] and, to a much lesser extent, the record route option in the IP header.

To obtain a representative router level map of the Internet, the traceroutes have to be launched between the maximum number of end points [2]. Additionally, the paths and the routers can change over time so the traceroutes have to be launched periodically. Several projects have focused on this part of topology collection based on traceroutes. ARK (archipelago) [3] uses dedicated measurement boxes, and others such as iPlane [4] or Rocketfuel [5] utilize Planetlab nodes [6]. DIMES [7], a software agent installable in any personal computer, can be used to greatly increase the number of vantage points in the measurement.

The traceroute is affected by load balancing, so different probing packets can be forwarded through different paths, so the tool called Paris-traceroute [8] instead of classic traceroute is used to avoid their effect.

Routers usually have several interfaces, each of them identified by an IP address. To obtain a router level topology map, the topology construction phase has to be applied. One of the main tasks in this phase is the identification of IP addresses belonging to the same router. This task is called *IP alias resolution* [9]. Several IP alias resolution methods are available and they will be reviewed in the following section. The general idea is to obtain extra information about router behaviour for each IP address to identify similar behaviours that could indicate that two IP addresses belong to the same physical router.

The two phases, collection and construction, are traditionally decoupled. In this paper, a modified traceroute is proposed that integrates the collection and construction phases. This method takes advantage of traceroute probing packets for IP alias resolution. The method takes the perspective of another type of probes (indirect ones) instead of using the typical probes in alias resolution (direct ones). To do so, a new and decoupled set of responses is added to the set of responses given by the rest of the alias resolution methods. The benefits of this proposal are related to the scalability, speed and improved identification ratios compared to the decoupled alternatives.

The rest of the paper is organized as follows. Section II presents the current state of the art alternatives for IP alias resolution and topology measurement projects. A new proposal is presented in Section III. The network scenarios used in the analysis are presented in section IV. Section V presents the evaluation. Conclusions and future work are presented in section VI.

II. Related work

The traceroute is one of the main tools that are currently used for router level topology discovery. Once the IP addresses of the routers and the relationships between neighboring routers are collected, alias resolution methods are applied to identify all of the IP addresses that belong to the same router. For a pair of IP addresses, the result of an alias resolution method can be positive or negative (also called true positive or true negative) when those IP addresses belong or do not belong, respectively, to the same router. Additionally, those methods can obtain erroneous results: false positive (when the method provides a positive result but it is not true) and false negative (when the method provides a negative result but it is not true).

Usually, the alias resolution methods are classified as active probing-based and inference-based. Active probing-based methods send probing packets to routers and analyze the response packets. Inference-based methods use only information provided by traceroutes to analytically infer aliasing.

Active probing-based methods are intrusive, but they provide the best identification results [10]. The most well-known methods of this type are Mercator [9][11] and Ally [5]. Mercator was used in the Skitter project [12]. It is based on sending UDP probing packets from the same vantage point to all IP addresses that could belong to the same router. The ICMP error packets responses are generated from the interface with the best path to the destination (vantage point), as indicated by the routing table. Therefore, if all of the probed IP addresses belong to the same router, all of the ICMP error packets will have the same source IP address.

The Ally method also uses UDP probing packets, but in this case the IP identification field (IPID) in the IP header is used to check for aliases. Typical TCP/IP implementations use an incremental counter to generate the IPID for each packet created by a host. Therefore, several IP packets received from the same host and close in time will have close IPID values. Several variations on the Ally scheme are available. Ally-based alias resolution methods provide improvements over the standard Ally [13]. The RadarGun tool [14] applies a technique called velocity modeling to IPIDs to obtain IP alias resolution with a linear cost and are oriented toward massive measurement campaigns. Midar [15] is an improvement of the Radargun technique. It solves problems related with accuracy and distributability observed in the RadarGun tool. Midar uses an strategy called sliding-window to select the IP addresses to be measured together.

The inference methods based on graph analysis use several heuristics to join information from expansion trees obtained by different traceroutes. Analytical Alias Resolver (AAR) [16] is based on finding /30 and /31 network masks from the information obtained in the traceroutes. A link is detected by looking for a pair of IP addresses obtained from opposite traceroutes that verify one of the previous network masks. Analytical and Probe-based Alias Resolver (APAR) [17] uses larger network masks that are accompanied by a low rate of probing packets based on PING to verify wheter aliases have been detected correctly.

TraceNET [18] is a recent proposal that attemps to discover the subnetworks attached to the routers instead of the IP addresses, as a conventional technique would make by probing ranges of IP addresses.

Palmtree [19] incurs a linear probing overhead to identify IP aliases. It follows a method similar to Mercator but uses indirect TTL-limited probes to generate the ICMP error responses.

In method [20], the prespecified timestamp option of the IP header is used. The method exploits the presence and the order of those timestamps, making the probing in both directions for each path. The IP addresses of the same router will fill up prespecified timestamp fields with very similar timestamps, allowing for the identification of aliases.

In this paper we propose a new method based on indirect probing that is able to perform the construction phase by using information obtained in the collection phase. Therefore, the IP alias resolution is performed in the construction phase without sending additional probes. In this proposal, indirect probing is used, which provides extra information compared to what is provided by the direct probing used by the majority of IP alias resolution schemes. Because of the similarities with the traceroute tool, this method is called Pamplona-traceroute.

III. All-in-one topology measurement tool: Pamplona-traceroute

Pamplona-traceroute uses indirect TTL-scoped probe packets to collect IP addresses of routers and, at the same time, to collect information about the particular behaviour of routers to apply IP alias resolution. Specifically, the Ally-based techniques are applied: the IPID of returning ICMP packets are used to identify IP aliases. The destination IP address in the probing packets is not the IP address under analysis (as in Ally, direct probing), rather the IP address of the end node in the path (indirect probing). Some routers are configured to be responsive to indirect probes, but they are configured to be non-responsive to direct probes. Therefore, the percentage of answers from this method is different compared to the standard Ally. The percentage of responses is expected to increase because this type of ICMP error in indirect probes is usually enabled in routers [21].

The full process of IP alias discovery in Pamplona-traceroute is composed of three phases. In the first phase, the data are collected. The second phase pre-processes and validates the data. The third phase analyzes the data to identify IP aliases. The details about the three phases are provided in the following subsections.

A. Collection phase

For a given scenario, the general procedure in the collection phase is the following. The Pamplona-traceroute instances are launched simultaneously at all of the end nodes (vantage points) in the topology. For each end node, all of the other end nodes will be considered destinations for the Pamplona-traceroute. Only one probing packet is sent per TTL and destination, up to the number of hops in the path (H). An ICMP error response of ”TTL expired in transit” caused by those indirect probes is obtained for each router in the path to the destination. If the destination IP address does not answer, the probes are sent up to a maximum number of 30 hops by default. The end nodes have to be clock-synchronized because the responses received from them are mixed. Network Time Protocol (NTP) is used for this purposse because this protocol achieves a synchronization between two distant points in the Internet on the order of hundreds of milliseconds [22].

To obtain enough information on IPID evolution and perform IP alias resolution, several responses (N) are needed for each IP address in the scenario. A set of indirect probes are sent back-to-back in the same path, and they provide N samples of IPID for each IP address in the path. This N will have a value of approximately 10, depending on the type of probing packet used as explained later.

For each probing packet in each Pamplona-traceroute instance, the following information is recorded:

1) The source IP address of the associated ICMP response: the IP address of the router in the path to the destination.
2) TTL distance: the number of hops from the vantage point to obtain the associated ICMP response.
3) The IPID of the associated ICMP response.
4) Timestamp of the associated ICMP response.

In Pamplona-traceroute, three types of probing packets are used: UDP (as the original traceroute does), TCP and ICMP Echo. The different probing packets are treated differently by each router and, obtain different results with regard to the number of ICMP responses.

To avoid load balancing, Pamplona-traceroute implements the Paris-traceroute strategy. The source/destination ports in the UDP/TCP probing packets and the type/code/checksum in the ICMP Echo Requests are maintained as constants for packets belonging to the same instance of Pamplona-traceroute.

B. *Pre-processing phase*

Once all of the data are collected, the processing can occur in a distributed or centralized way. In the distributed version, the data collected by each end node can be shared with other processing nodes. This scheme has to guarantee that all of the data collected for a given pair of IP addresses are made available to a certain processing node. Each pair of IP addresses is assigned to a given node for processing. These processing nodes can be the end nodes used in the data collection phase or any others.

The other possibility is to upload all of the data to a central point where processing is performed to identify the IPID sequences that could imply IP aliases. In this pre-processing phase, the distributed or centralized processing is chosen, and as a first step, the data are distributed accordingly.

Before checking for aliasing by pairs of IP addresses, the IPIDs have to be verified because different behaviours in the generation of IPIDs in routers that depend on the model and the manufacturer have been observed [13]: incremental, random, randomT and reset. In IP alias resolution, the desired behaviour for the IPID is incremental per router: each packet generated in a router increases by one unit a general IPID counter common to all of the interfaces in the router. This allows the identification of IP addresses belonging to the same router because the IPID values from the ICMP responses are correlated.

Although some IPID schemes are not useful for identifying positive aliases in a pair of IP addresses, they can be used to identify negative aliases. This will be the case for pairs of IP addresses in which both IP addresses use a different IPID scheme.

C. *Alias resolution phase*

The IP addresses are verified for aliasing in pairs. The IP addresses that are discovered by different end nodes and are validated in the pre-processing phase as following an incremental IPID schema are mixed to consider all possible combinations of pairs. The IPIDs for each pair of IP addresses are sorted by timestamp. If both IP addresses belong to the same router, the IPIDs form an incremental sequence.

To identify aliases, similar strategies to Ally-based alias resolution methods [13] are used. The following criteria are considered over the sequence of the IPIDs to reduce the rate of false positives and false negatives:

1) The sequence of the IPIDs must follow an incremental pattern. At least 2 mixed increments are considered necessary. A mixed increment refers to incremental IPIDs between two IP addresses pairs, the first of which (the responding interface with the lower timestamp) changes alternately.
2) Consecutive samples of the IPIDs (belonging both to each IP address of the pair) within the same second are not considered.
3) Consecutive samples of the IPIDs (belonging both to each IP address of the pair) separated more than 3 seconds are not considered.
4) In the case of IPID counter wrapping, a maximum threshold of 200 units of IPID difference is considered for aliases.

Criterion number 1 is the basic consideration for all IP alias resolution methods derived from Ally. The decremental steps are allowable when the IPID counter wraps (criterion 4). In addition, they can occur because of packet disordering when the ICMP response packets follow through different paths. In this case, criterion number 2 allows those decrements to be ignored. To guarantee that both IP addresses (IP_1, IP_2) are aliases, the incremental sequence has to mix both IP addresses in a way so that IP_1 and IP_2 do not form a parallel incremental sequence by themselves.

The IPID increments have to be mixed between both IP addresses on at least 2 occasions. For example, in the following sequence of IPIDs for a pair of IP addresses $(IP_1\,5, IP_2\,10, IP_2\,12, IP_1\,14, IP_2\,16, IP_1\,18)$ the following pairs and increments can be derived: $(IP_1\,5, IP_2\,10)$, $(IP_2\,12, IP_1\,14)$, $(IP_1\,14, IP_2\,16)$ and $(IP_2\,16, IP_1\,18)$. As can be observed, there are three mixed increments: the first IP address of the pairs changes three times in this sequence of increments. Due to lags between the different responses from the interfaces, not every IPID can be used and compared with the previous one (Criterion 3). The number of mixed increments defines the increment threshold.

An increment threshold value of 2 mixed increments does not provide any false positives in our analysis. However, it could be expected that this effect would appear in a more extensive measurement. In that case, it would be enough to increase the increment threshold to reduce the rate of false positives.

Depending on the rate of the packets generated by a router, too much time between IPID samples could imply that wrapping occurred in the IPID counter. In [13], the IPID increment speed is reviewed, and the rates of 100 IPIDs per second is typical for 90% of 434 tested routers. In a maximum time of 3 seconds between IPID samples (as indicated in criterion number 3), an increase of 300 IPID numbers is assumed to occur in the worst case. This IPID interval is short enough compared to the IPID range ($2^{16} = 65,536$), and therefore, there is a low probability that two consecutive wrapping effects that would distort the analysis would occur.

Criterion number 4 is applied in the case of IPID counter wrapping [14]. Two hundred units of IPID difference is adequate for the probing time used by criteria 3 in the case of a decrement in the IPIDs.

IV. Network scenarios and methodology

Two main scenarios were considered using the Etomic [23] and Planetlab [6] measurement infrastructures. The measurement campaign was conducted from September to November 2010 and November 2012. All of the data sets and software developed for this paper are available at [24].

In Etomic, six nodes were chosen around Europe as end nodes. This number provided 91 IP addresses of routers in the paths between all of the end nodes. In Planetlab, 50 end nodes were chosen distributed worldwide. Only those 50 nodes were found to be stable and available over the two months of the first measurement campaign, although we tried to obtain a larger number of nodes. The paths between all 50 end nodes contained 1,123 router IP addresses. Different subsets were defined in this scenario. Subsets 1, 2 and 3 are composed of 15 random end nodes with 370, 306 and 141 router IP addresses respectively. Subset 4 is the complete set of 50 end nodes and 1,123 router IP addresses.

The routing testbed presented in figure 1, was also used. This testbed is composed of 9 Cisco routers and 1 Linux router in our labs. It contains 25 router IP addresses that imply 300 pairs of IP addresses to check for aliasing. Only 18 pairs of IP addresses are real aliases. In this case, there are 4 vantage points that would allow the measurements to be perform. This testbed allows for checking the identification rates of different alias resolution methods in a controlled scenario.

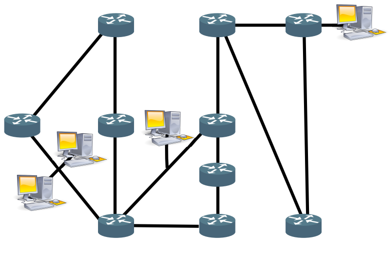

Fig. 1. Testbed used in verification of alias resolution methods

V. Evaluation of Pamplona-traceroute

The following metrics were analyzed to evaluate the proposal [25]: accuracy, completeness, efficiency and distributability.

The accuracy metric measures the number of errors performed in IP alias resolution (ratio of false positives and false negatives). The completeness metric measures the ratio of aliases (positives) and non aliases (negatives) that are identified compared to the total number of pairs of IP addresses. The efficiency metric is related to how intrusive the method is (volume of probing traffic needed). The distributability metric indicates whether data collection and processing can be distributed between different vantage points.

TABLE I. Percentage of IP addresses with IPID incremental behaviour in Pamplona-traceroute

Network	*ICMP: %*	*UDP: %*	*TCP: %*
Etomic	40.00	42.04	40.24
Planetlab subset 1	32.86	18.32	38.67
Planetlab subset 2	47.76	31.57	52.23
Planetlab subset 3	29.77	19.14	47.45
Planetlab subset 4	32.19	30.27	35.97

TABLE II. Percentage of IP addresses with IPID incremental behaviour in Ally-based aliases resolution methods

Network	*ICMP: %*	*UDP: %*	*TCP: %*
Etomic	19.98	22.82	24.45
Planetlab subset 1	22.11	57.75	54.12
Planetlab subset 2	17.30	48.78	37.37
Planetlab subset 3	12.72	60.00	29.09
Planetlab subset 4	15.74	55.60	42.55

The proposal was compared with the most advanced IP alias resolution methods, specifically the Ally-based alias resolution methods, the RadarGun tool and the Midar tool.

A. *Completeness*

The accuracy and the completeness achieved using Pamplona-traceroute are determined mainly by the amount of useful data obtained from the collection phase. One aspect to analyze is the number of responsive routers in relation to the type of probing packet used. For example, in subset 4 from planetlab scenario, the percentage of non-answering routers is 3.78% with ICMP, 6.04% with UDP and 3.28% with TCP.

However, a responsive router is not enough to obtain a correct IP alias resolution. The desired IPID behaviour for IP alias resolution is the incremental scheme. In tables I and II, the percentages of the IP addresses with incremental IPID behaviour is presented for Pamplona-traceroute and Ally-based alias resolution methods, respectively. The rest of the percentage to reach 100% is related to the schemes random, randomT, reset, reset0, destination-dependent, as well as unresponsive IP addresses that do not generate ICMP responses. In those tables, percentages are distinguished by the type of probing packet (ICMP Echo, UDP and TCP) that obtains similar results for all of them. The results indicate that Pamplona-traceroute achieves a larger percentage of the desired IPID incremental behaviour for some of the scenarios and types of probing packets, but it also achieves worse results for other combinations. For ICMP probes, Pamplona-traceroute provides a larger percentage of the desired IPID incremental behaviour. For UDP and TCP probes, the results depend on the specific scenario.

A large percentage of routers without incremental IPID behaviour are core routers. Those routers correspond to intermediate hops in the path between end points. Therefore, the results in table I are not distributed uniformly for all access and core routers. Figure 2 presents the normalized histogram for the number of routers that answer with IPID behaviour following the reset0 scheme depending on the hop number measured from the vantage point in the Etomic scenario. The plotted data are obtained in the collection phase of Pamplona-traceroute. The value for each hop represents the ratio of routers with the reset0 scheme compared with the total number of routers observed in that hop. The vast majority of the paths in our scenarios have approximately 16 hops. The reset0 scheme is concentrated in the routers in the middle of the path. In those core routers (hops 6-7 in figure 2), the ratio of routers with the reset0 behaviour is up to 95%. This ratio is almost independent of the type of probing packet used.

In figure 3, the corresponding survival (complementary cumulative distribution) function is plotted for the Etomic scenario. The large percentage of non-incremental behaviour for the IPID is caused by core routers answering with the reset0 scheme. This is the most limiting factor for completeness. However, in this respect, there exists good complementarity between Pamplona-traceroute and Ally-based aliases resolution methods. Figure 4 presents the survival function of the reset0 scheme in Ally-based alias resolution methods. In this case, the reset0 scheme is present uniformly in all hop numbers, not only in core routers. Additionaly, in Ally-based techniques, the UDP direct probes provide a smaller percentage of reset0 behaviours.

Therefore, Pamplona-traceroute is good at identifying access routers, and Ally-based alias resolution methods are good at identifying core routers. A full IP alias resolution methodology with the best results can be designed with a first phase using Pamplona-traceroute and a second phase with Ally-based alias resolution methods for those pairs of IP addresses not previously resolved.

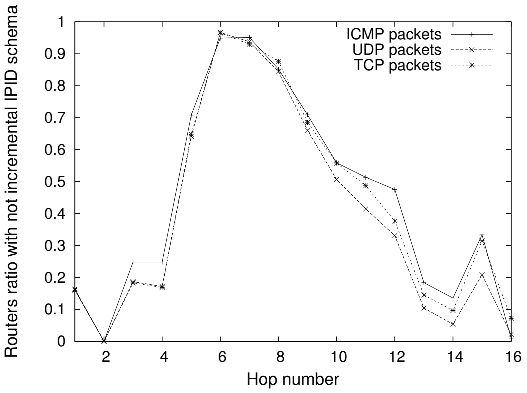

Fig. 2. Normalized histogram for the number of routers with not incremental IPID schema per hop in Pamplona-traceroute

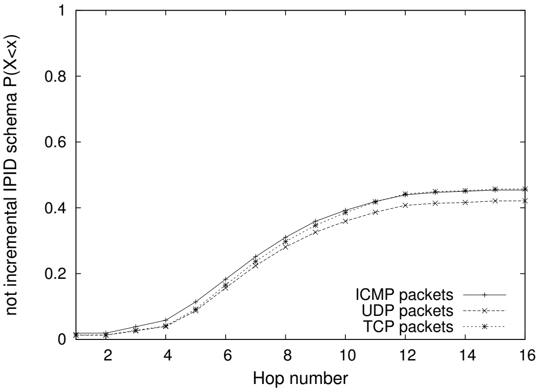

Fig. 3. Survival function for the number of routers with not incremental IPID schema per hop in Pamplona-traceroute

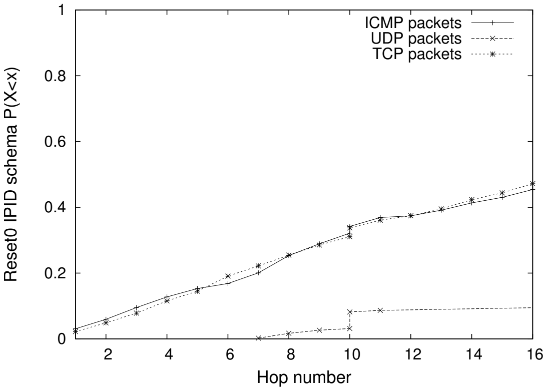

Fig. 4. Survival function for the number of routers with reset0 IPID schema per hop in Ally-based aliases resolution methods

B. Accuracy

The well-known scenario to check for the accuracy of alias resolution methods is the lab testbed presented in section IV. The results of the identification for the Pamplona-traceroute and other alias resolution methods are presented in table III for this testbed. In this table, the percentage of positives, negatives, false positives and false negatives with respect to the total number of pairs of IP addresses are presented. The column titled *Identified* is the sum of the correctly identified pairs (positives and negatives). The column titled *Aliases* refers to the percentage of positive aliases identified over the real number of aliases in the testbed. Larger values in those two columns indicate a better alias resolution method. The Pamplona-traceroute provides the best results, even improving upon results of Ally-based methods. In addition, the Pamplona-traceroute has neither false positives nor false negatives. Radargun and the Prespecified-timestamp provide the worst identification results.

Although the size of the testbed is very limited, interesting results were obtained. In the following discussion, larger scenarios in which the ground truth of the real topology is unknown were evaluated. In these scenarios, only Ally-based, Radargun, Midar and Pamplona-traceroute were considered. In Etomic and Planetlab scenarios, Pamplona-traceroute was launched in each end node with a destination of all the other end nodes. It was iterated 30 times and repeated for each type of probing packet available: ICMP Echo, UDP and TCP. The IP alias resolution results are summarized in tables IV, V, VI and VII. Because the real topology information is not available for these network scenarios, positives, negatives, false positives and false negatives were verified using as reference the topology that was generated by the Ally-based alias resolution methods. This method was chosen as the reference because it provides the best identification results of the methods [26].

Tables IV, V and VI present identification rates for each scenario and for each type of probing packet, ICMP Echo, UDP and TCP, in Pamplona-traceroute. These tables show the number of IP addresses discovered, the percentage of identification (sum of positive and negative aliases), and the

TABLE III. COMPARATIVE OF ALIASES RESOLUTION METHODS IN THE TESTBED

Method	*Positives*	*Negatives*	*False*	*False*	*Identified*	*Aliases*
	%	*%*	*positives %*	*negatives %*	*%*	*%*
Pamplona-traceroute	6.15	83.69	0	0	89.85	94.44
Ally-based methods	4.71	44.56	0	0	49.27	72.22
Palmtree	3.62	-	00.72	-	3.62	55.55
Prespecified-timestamps	0.36	5.07	0	00.72	5.43	5.55
Radargun	0.36	9.78	0	0	10.14	5.55
Midar	4.71	-	0	-	4.71	72.22
TraceNET	3.26	-	00.72	-	3.26	50.00

TABLE IV. IDENTIFICATION RATES FOR ICMP PROBING PACKETS IN PAMPLONA-TRACEROUTE

Network	*IP*	*Identification*	*False*	*False*
	addresses	*%*	*positives %*	*negatives %*
Etomic	85	66.32	0.00	0.00
Planetlab 1	286	59.31	0.00	0.18
Planetlab 2	224	62.31	0.00	0.10
Planetlab 3	131	50.31	0.00	0.00
Planetlab 4	963	55.31	0.00	0.04

TABLE V. IDENTIFICATION RATES FOR UDP PROBING PACKETS IN PAMPLONA-TRACEROUTE

Network	*IP*	*Identification*	*False*	*False*
	addresses	*%*	*positives %*	*negatives %*
Etomic	88	66.02	0.00	0.00
Planetlab 1	322	60.38	0.00	0.00
Planetlab 2	266	62.07	0.00	0.05
Planetlab 3	141	56.19	0.00	0.00
Planetlab 4	1004	53.38	0.00	0.01

TABLE VII. AGGREGATING RESULTS OF IDENTIFICATION RATES FOR ALL TYPES OF PROBING PACKETS IN PAMPLONA-TRACEROUTE

Network	*Number of*	*Total IP*	*Total Identified*
	vantage points	*addresses*	*%*
Etomic	6	91	70.0
Planetlab 1	15	370	66.83
Planetlab 2	15	306	72.86
Planetlab 3	15	141	81.85
Planetlab 4	50	1123	62.83

TABLE VIII. IDENTIFICATION RATES IN RADARGUN

Network	*Identification*	*False*	*False*
	%	*positives %*	*negatives %*
Etomic	13.74	0.00	0.16
Planetlab 1	25.80	7.89	0.01
Planetlab 2	23.19	44.76	0.04
Planetlab 3	18.33	0.00	0.00
Planetlab 4	26.74	25.85	0.00

percentage of false positives and false negatives (errors of the method). The percentages are measured with respect to the total number of pairs of IP addresses discovered in each case. The success and errors are obtained from those previously obtained with the Ally-based alias resolution method used as the reference.

For each type of probing packet, the number of discovered IP addresses in the paths (routers that answer with ICMP response) are quite similar and are slightly better in UDP probes. The identification rates are also quite similar for all types of probing packets; however, the UDP probing packets provide better results in Planetlab scenarios. In all cases, there are no false positives, and there is a very low rate of false negatives in the identification.

Table VII presents the aggregated identification results obtained by combining the data from the different types of probing packets. The total number of IP addresses identified grew by up to 8% in the Planetlab scenario if when compared with the single packet type Pamplona-traceroute tests. Identification rates were calculated with respect to the total number of IP addresses discovered by any type of probing packet. The "total identified" column aggregates the identification results of all of the probing packets types with respect to the total number of IP addresses. This aggregation provides some improvement compared with individual cases, increasing identification rates from values near 50-60% to 60-70%.

When compared with RadarGun in the same scenarios (see table VIII), the results for Pamplona-traceroute are significantly improved, not only in the percentage of identification (completeness) but also in the absence of errors in the identification (accuracy). Even compared to the identification results provided by the Ally-based alias resolution methods (see table IX), Pamplona-traceroute improves the identification results up to 25% for some Planetlab scenarios. In this table, the false positives and false negatives can not be calculated, because the Ally-based alias resolution methods are used as a reference.

Table X presents the identification results for Midar in the main scenarios. In Midar technique completeness is reduced compared with other IP alias resolution techniques mainly because this technique only provides positive alias identification. False positives have not been found when the results are compared with the Ally-based ones. The positive identification rates are very similar in Midar and Ally-based techniques, but Midar introduces much less probing traffic into the network. Midar and Pamplona-traceroute does not share some of the identified positive pairs and therefore both techniques can be complementary.

The identification results of Pamplona-traceroute can be improved if the method is complemented by Ally-based or Midar alias resolution methods. Pamplona-traceroute is first

TABLE VI. IDENTIFICATION RATES FOR TCP PROBING PACKETS IN PAMPLONA-TRACEROUTE

Network	*IP*	*Identification*	*False*	*False*
	addresses	*%*	*positives %*	*negatives %*
Etomic	82	65.83	0.00	0.00
Planetlab 1	287	56.25	0.00	0.03
Planetlab 2	268	50.82	0.00	0.02
Planetlab 3	118	59.82	0.00	0.37
Planetlab 4	970	43.86	0.00	0.09

TABLE IX. IDENTIFICATION RATES IN ALLY-BASED ALIASES RESOLUTION METHODS

Network	*Identification*	*False*	*False*
	%	*positives %*	*negatives %*
Etomic	68.46	-	-
Planetlab 1	45.45	-	-
Planetlab 2	31.17	-	-
Planetlab 3	58.74	-	-
Planetlab 4	51.39	-	-

TABLE X. IDENTIFICATION RATES IN MIDAR

Network	*Identification %*	*False positives %*	*False negatives %*
Etomic	0.087	0.00	-
Planetlab 1	0.083	0.00	-
Planetlab 2	0.091	0.00	-
Planetlab 3	0.102	0.00	-
Planetlab 4	0.078	0.00	-

TABLE XI. IDENTIFICATION RATES COMBINING PAMPLONA-TRACEROUTE AND ALLY-BASED ALIAS RESOLUTION METHODS

Network	*Pamplona-traceroute identif. %*	*% of pairs to check with Ally-based methods*	*Combined identification %*
Etomic	70.08	29.91	93.06
Planetlab 1	66.83	33.17	81.26
Planetlab 2	72.86	27.14	72.96
Planetlab 3	81.85	18.15	88.11
Planetlab 4	62.83	37.17	81.27

used to discover the IP addresses and to provide a first phase of IP alias identification. Those pairs of IP addresses that are not identified as positive or negative aliases are checked in a second phase using Ally-based alias resolution methods. The results are shown in table XI in the following columns: identification percentage using only Pamplona-traceroute, percentage of pairs of IP addresses to check with Ally-based alias resolution methods and total identification percentage combining both methods. The percentage of IP addresses pairs to be checked in this second phase is approximately 20-40%, but that percentage can be reduced using reduction methods to 3-4%. With this full process, the identification rates of 60-70% in stand-alone Pamplona-traceroute increase to 80-90% with the combination of Pamplona-traceroute and Ally-based alias resolution methods.

C. Efficiency

The probing packets used in Pamplona-traceroute have the minimum size (Ethernet 64 bytes). Each Pamplona-traceroute instance has to send N packets per hop and destination end node. Assuming that there are M end nodes and that each path is composed of H hops on average, the total network traffic generated by an end node is $(M-1)*H*N*64$ bytes. The value of H depends on the chosen network scenario.

The number of iterations N of the Pamplona-traceroute can be adjusted depending on the desired accuracy and completeness. With a greater number of iterations, more IPID values are collected per IP address, and therefore, the method is better for identification. In figure 5, the results of the IP alias resolution are presented for ICMP probing packets. For each N value on the x-axis, the corresponding identification ratio is plotted on the y-axis. The increment threshold, explained in Section III-C, was 2, and the identification was performed in the Planetlab subset 4 scenario. For the UDP and TCP probing packets, the curves are similar and are not shown for simplification. A number of iterations N of approximately 20 provides the best identification results. As observed in figure 5, the plot reaches the stabilized maximum of 0.45 for the identification ratio. A considerably close rate is reached using 20 iterations

Figure 6 shows the survival function for the duration of the Pamplona-traceroute instances in the Planetlab subset 4 scenario. More than 90% of the Pamplona-traceroute instances last for less than 50 seconds for the ICMP and TCP probing packets. With the UDP probing packets, this time is much shorter. The reason for the difference is that with the ICMP and TCP probing packets, if the destination end node does not answer (for example, if it is blocked by the typical local firewalls), it is not possible to determine when to finish Pamplona-traceroute. Therefore, all hops are checked up to 30. Depending on the number of iterations N needed to obtain the desired ratios of identification, the collection phase can take several minutes.

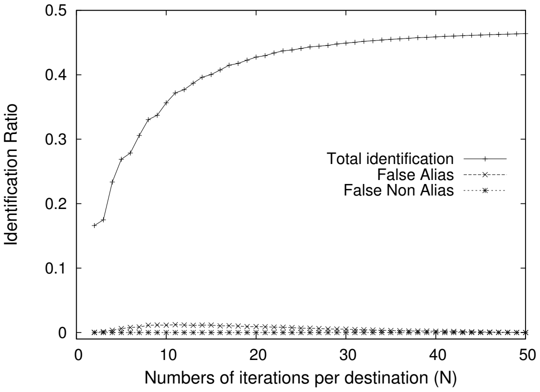

Fig. 5. IP alias results per number of iterations of the Pamplona-traceroute and ICMP probing packets

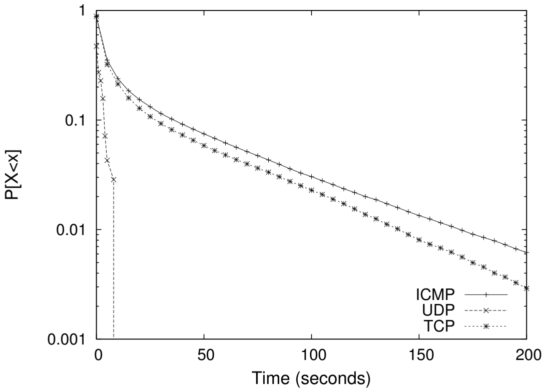

Fig. 6. Survival function of the time needed for running each Pamplona-traceroute instance

D. Distributability

In Pamplona-traceroute, the collection and alias resolution phases can be distributed. In the collection phase, the measurements have to be performed from different vantage points to clarify as much of the intermediate topology as possible. This process provides the IP addresses of the routers in the paths between the end points and IPID information per IP address.

The alias resolution phase can be centralized or distributed. All possible combinations of IP address pairs are checked for IP aliases. Subsets of these pairs can be processed in a central processing point, or they can be shared out between different nodes. Therefore, the processing can be distributed, for example, between the same end nodes that made the measurements. This allows for an easyscale up of the number of end nodes in the size of the full topology to map.

In RadarGun, the measurement and processing are conducted from a unique vantage point. Therefore it can be limited by bandwidth consumption and the control rate disciplines present in routers [27]. Additionally, the IP address discovery is a separate process that is usually performed by traceroutes. In last version of Midar (Sep. 2012), the identification can be distributed.

VI. Conclusions

Topology discovery is a complex and costly task. The traditional topology strategies use an IP address discovery step via traceroute in addition to an IP alias resolution step that usually needs extra probing traffic. Our proposal, Pamplona-traceroute, conducts a specific probing in the discovery phase to obtain data simultaneously while performing IP address discovery and IP alias resolution. Additionally, the probing packets are indirect, unlike the majority of IP alias resolution methods. These probing packets generate an ICMP error by TTL exceeded packets, improving the responsiveness of a set of routers compared with Ally-based schemes. The best results are obtained for access routers because of their IPID incremental behaviour. Therefore, the results can be complemented with Ally-based schemes that behave better for core routers. By combining both, the Pamplona-traceroute and Ally-based schemes, better identification ratios are obtained, and the probing traffic overhead is reduced compared to a stand-alone Ally-based scheme.

Pamplona-traceroute improves the accuracy, completeness, efficiency and distributability of the complete process of topology discovery and alias identification. Identification rates of approximately 60-70% are achieved with an almost insignificant presence of errors in our testbed scenario where the ground truth is known. These identification rates can approach to 80-90% if combined with Ally-based schemes.

Future work will be related to the integration of Pamplona-traceroute in already existing topology measurement infrastructures. As a first step, Pamplona-traceroute is already integrated in Etomic periodic measurements. In addition, new types of probing packets can be analyzed. The evaluation of methods to reduce the volume of probing traffic and the post-processing load is also of high interest.

References

[1] V. Jacobson, ftp://ftp.ee.lbl.gov/traceroute.tar.gz (October 1989).

[2] Y. Shavitt, U. Weinsberg, Quantifying the importance of vantage points distribution in internet topology measurements., Infocom 2009.

[3] CAIDA, ARK, Archipelago Measurement Infrastructure, http://www.caida.org/projects/ark/ (2002).

[4] H. V. Madhyastha, T. Isdal, M. Piatek, C. Dixon, T. Anderson, A. Krishnamurthy, A. Venkataramani, iplane: an information plane for distributed services, in: Proceedings of the 7th symposium on Operating systems design and implementation, OSDI '06, USENIX Association, Berkeley, CA, USA, 2006, pp. 367–380.
URL http://dl.acm.org/citation.cfm?id=1298455.1298490

[5] N. Spring, R. Mahajan, D. Wetherall, T. Anderson, Measuring isp topologies with rocketfuel, IEEE/ACM Trans. Netw. 12 (1) (2004) 2–16. doi:10.1109/TNET.2003.822655.
URL http://dx.doi.org/10.1109/TNET.2003.822655

[6] B. Chun, D. Culler, T. Roscoe, A. Bavier, L. Peterson, M. Wawrzoniak, M. Bowman, Planetlab: An overlay testbed for broad-coverage services, ACM SIGCOMM Computer Communications Review 33 (2003) 3–12.

[7] Y. Shavitt, E. Shir, Dimes: let the internet measure itself, SIGCOMM Comput. Commun. Rev. 35 (5) (2005) 71–74. doi:10.1145/1096536.1096546.
URL http://doi.acm.org/10.1145/1096536.1096546

[8] B. Augustin, X. Cuvellier, B. Orgogozo, F. Viget, M. L. Timur Friedman, C. Magnien, R. Teixeira, Avoiding traceroute anomalies with paris traceroute, in: 6th ACM SIGCOMM, Rio de Janeiro, Brazil, 2006, pp. 153–158.

[9] J. J. Pansiot, D. Grad, On Routes and Multicast Trees in the Internet, ACM SIGCOMM Comput. Commun. Rev. 28 (1998) 41–50.

[10] K. Keys, Internet-Scale IP Alias Resolution Techniques, ACM SIGCOMM Computer Communication Review (CCR) 40 (1) (2010) 50–55.

[11] R. Govindan, H. Tangmunarunkit, Heuristics for internet map discovery, in: INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, Vol. 3, 2000, pp. 1371–1380 vol.3. doi:10.1109/INFCOM.2000.832534.

[12] B. Huffaker, D. Plummer, D. Moore, K. Claffy, Topology discovery by active probing, in: Proc. the Symposium on Applications and the Internet (SAINT), 2002.

[13] S. Garcia-Jimenez, E. Magaña, D. Morato, M. Izal, 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, 2009, pp. 513–520.

[14] A. Bender, R. Sherwood, N. Spring, Fixing Ally's Growing Pains with Velocity Modeling, in: (IMC 08) 8th ACM SIGCOMM conference on Internet measurement, ACM, New York, NY, USA, 2008, pp. 337–342.

[15] K. Keys, Y. Hyun, M. Luckie, K. Claffy, Internet-scale IPv4 alias resolution with MIDAR, Networking, IEEE/ACM Transactions on 21 (2) (2013) 383–399. doi:10.1109/TNET.2012.2198887.

[16] M. Gunes, K. Sarac, Analytical IP alias resolution, in: ICC '06. IEEE International Conference on Communications, Istanbul, 2006, pp. 459–464.

[17] M. H. Gunes, K. Sarac, Resolving IP aliases in building traceroute-based Internet maps, IEEE/ACM Transactions on Networking 17 (2009) 1738–1751.

[18] K. S. M. Engin Tozal, Tracenet: An internet topology data collector, Internet Measurement Conference IMC (2010) 356–368.

[19] K. Sarac, M. E. Tozal, Palmtree: An ip alias resolution algorithm with linear probing complexity, Computer Communications 34 (5) (2011) 658–669.

[20] J. Sherry, E. Katz-Bassett, M. Pimenova, H. V. Madhyastha, T. Anderson, A. Krishnamurthy, Resolving IP aliases with prespecified timestamps, in: Proceedings of the 10th annual conference on Internet measurement, IMC '10, ACM, New York, NY, USA, 2010, pp. 172–178.

[21] M. H. Gunes, K. Sarac, Analyzing router responsiveness to active measurement probes, in: Proceedings of the 10th International Conference on Passive and Active Network Measurement, PAM '09, Springer-Verlag, Berlin, Heidelberg, 2009, pp. 23–32.

[22] D. Mills, A brief history of NTP time: confessions of an internet timekeeper., ACM Computer Communications Review 33.

[23] D. Morato, E. Magaña, M. Izal, J. Aracil, F. Naranjo, F. Astiz, U. Alonso, I. Csabai, P. Haga, G. Somin, J. Seger, G. Vattay, The European Traffic Observatory InfraestruCture (ETOMIC): A testbed for universal active and passive measurements, in: Proc. TRIDENTCOM, 2005, pp. 283–289.

[24] S. G.-J. et al., Tools and data sets used in this paper, http://www.tlm.unavarra.es/~santi/research/paper9.html.

[25] N. Spring, M. Dontcheva, M. Rodrig, D. Wetherall, How to resolve IP aliases, Tech. Rep. UW-CSE-TR 04-05-04, Washington Univ. Computer Science (2004).

[26] S. Garcia-Jimenez, E. Magaña, D. Morato, M. Izal, Probing distribution in time and space for IP aliases resolution, in: Published in IFIP Networking 2012 conference, Prague, Czeck Republic, 2012.

[27] S. Garcia-Jimenez, E. Magaña, M. Izal, D. Morato, 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, 2009, pp. 1079 – 1086.