IP addresses distribution in Internet and its application on reduction methods for IP alias resolution

S. García-Jiménez, E. Magaña, M. Izal and D. Morató
Public University of Navarre, Campus de Arrosadia s/n, E-31006 Pamplona, Spain
e-mail: {santiago.garcia, eduardo.magana, mikel.izal, daniel.morato}@unavarra.es

Abstract

Discovery of Internet topology is an important and open task. It is difficulted by the high number of networks and internetworking equipments, and even by the dynamic of those interconnections. Mapping Internet at router-level needs to identify IP addresses that belong to the same router. This is called IP address alias resolution and classical methods in the state of the art like Ally need to test IP addresses in pairs. This means a very high cost in traffic generated and time consumption, specially with an increasing topology size. Some methods have been proposed to reduce the number of pairs of IP addresses to compare based on the TTL or IP identifier fields from the IP header. However both need extra traffic and they have problems with the probing distribution between several probing nodes. This paper proposes to use the peculiar distribution of IP addresses in Internet Autonomous Systems in order to reduce the number of IP addresses to compare. The difference between pairs of IP addresses is used to know a priori if they are candidates to be alias with certain probability. Performance evaluation has been made using Planetlab and Etomic measurement platforms. The paper justifies the reduction method, obtaining high reduction ratios without injecting extra traffic in the network and with the possibility to distribute the process for alias resolution.

I. Introduction

Internet topology information discovery at router-level is possible thanks to alias resolution methods. These methods are able to associate IP addresses that belong to the same router. Some of these methods need to generate probing traffic (active probing methods) while others analyze already available information (inference methods). The first ones provide better results in alias resolution but with the overhead of extra traffic to be injected into the network [1].

An example of active probing method is Ally [2]. It uses UDP probing packets sent to random destination ports in target IP addresses in order to provoke ICMP notifications of port unreachable. The method uses the IP identification field (IPID) in the returned IP header to check for aliases. This IPID value is originally used in the procedures of fragmentation and reassembly. This field has the same value for all fragments belonging to an original IP datagram before fragmentation, so it is used to reassemble the original IP datagram in destination. Typical TCP/IP implementations of IP identifier use a counter which is incremented by one for each packet created in the host, independently of destination, protocol or service. Therefore, several IP packets received from the same host and near in time will have close values in the IP identification field. The differences in the counter will be caused by other IP traffic generated in between by that host to other destinations.

The Ally tool sends two probe packets almost back-to-back to two IP addresses (potential aliases), receiving two ICMP error packets with type "destination port unreachable" and IP identifiers x and y respectively. One second later, a third probe packet is sent to the IP address that sent first the previous ICMP error. Then a third ICMP error with IP identifier z is received. The two IP addresses will be alias if $x < y < z$ with $|z - min(x, y)| < 200$. If there were not IP traffic generated by the router in between: $x + 2 = y + 1 = z$. It must be noted that IP traffic generated by a router is related mainly with management tasks (routing protocols, SNMP, ping, traceroute, etc.). It does not take into account the packets forwarded by the router, which keep their original IPID value. The threshold of 200 sequence numbers in one second is chosen taking this into account [2].

Best results with alias resolution are obtained with Ally and related methods, for example, changing the protocol of probing packets (using TCP or ICMP Echo Request) [3][4] or increasing the number of probing packets [4]. Other proposals, like Mercator [5], provide a lower success rate on IP alias resolution mainly because of packet filtering and firewalls on the Internet.

The problem with Ally method is that tests have to be made in pairs of IP addresses and this means a complexity $O(n^2)$ with n the number of IP addresses. This is a very costly task specially for big networks and finally the whole Internet. *Reduction methods* will try to select the IP addresses that have more probability to be alias in order to make tests only between them. The final objective is to improve the efficiency of IP alias resolution methods. Several alternatives for reduction methods in alias resolution have been proposed in the state of the art:

- TTL-based [5]: it is based on the TTL (Time-to-Live) field of the IP header. If two IP addresses are alias it is very probable that the distance between them ($TTL_1 - TTL_2$) was 0 (the TTL distance is the same from the probing station).

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)

However, this is not always true because the path and consequently the number of hops can be different for each interface at a certain router. In that case, larger TTL distances (for example up to 3) can be considered.

- IPID-based [6]: like in alias resolution methods, the IPID field from the IP header is used to identify pairs of IP addresses that are good candidates to be alias. If both IP addresses belong to the same router, the IPID value in each packet response must be close because the IPIDs are generated incrementally for all IP packets in a router independently of the outgoing interface.
- AS-based [7]: the requisite to consider a pair of IP addresses as candidates to be alias is that both IP addresses belong to the same Autonomous System (AS). AS internal routers have IP addresses registered in that AS. However AS border routers, that interconnect different ASs, have IP addresses in each interface that could belong to different ASs. Therefore the method is not very precise, and it is usually completed with other reduction methods like that TTL distance was exactly 0 and that IPIDs were close enough. In this case the reduction is more important, losing some completeness in the alias identification.

For the TTL-based reduction method, TTL data can be obtained from the original traceroutes used in IP address discovery, but only if they were made from a unique probing station, which is not usual. In large topology maps it would be needed to distribute the traceroute measurement collection among several probing stations. However, as the TTL information is needed from the same probing station, extra traffic would be needed. This would mean increasing the probing traffic proportionally to the number of probing stations.

The IPID-based reduction method is near to being a full Ally resolution method. In the reduction method only the two first probing packets are sent to each IP address and if they are close enough (200 threshold) the third packet for Ally is sent. This means applying a partial Ally resolution method to all pairs of IP addresses (2 from the 3 probing packets), and therefore it is not a good improvement in the reduction of probing traffic or time to complete the alias resolution.

Finally, the AS-based reduction alternative is the option used in iPlane measurements [7]. iPlane is deployed as an application-level overlay network with the task of generating and maintaining a topology map of Internet. For this task, an alias resolution method based on Ally is used, updating the measurements each 2 months. The proposed reduction method gets a good reduction percentage but losing completeness in alias identification.

In this paper the distribution of IP prefixes in the paths through the Internet is characterized. This characterization is used to provide a reduction method that improves the efficiency of alias resolution methods. The reduction method will be based on the offsets between IP addresses. The rest of the paper is organized as follows. Section II presents the peculiarities in the distribution of IP addresses in routes through Internet. Then, this characterization is used to justify a reduction method in section III. Next section IV presents the evaluation of the proposal in different scenarios. Finally, conclusions are presented.

II. IP ADDRESSES IN ROUTES BETWEEN ENDPOINTS IN INTERNET

Internet topology discovery has been traditionally based on discovering IP addresses in the path between two endpoints. This process is automated with the well-known tool called *traceroute* [8]. A variation called *paris-traceroute* [9] provides better results when applied on routers with flow balancing. In traceroute, the IP addresses at the input interfaces from the routers in the path to destination endpoint are obtained for each direction. This means that, besides the set of IP addresses, the neighboring relation is obtained.

Networks and routers in Internet belong to ASs [10] operated by different administrative domains such as Internet Service Providers, research organizations and companies. Links between routers provide information about the relations between ASs. Each AS has allocated one or several IP subnetworks with common addressing schemes. This means that IP addresses are aggregated in contiguous blocks sharing the same prefixes. In this section, addressing allocation in ASs will be reviewed to demonstrate that it is not distributed uniformly along the 2^{32} possible IP addresses and that their peculiar distribution can be exploited.

ASs are organized hierarchically, with Tier-1 ASs that provide the root interconnection in Internet, and other Tier-2,3 ASs that need an upper-level tier AS to get full connectivity to Internet [11]. Routers in Tier-1 ASs have entries in its routing tables to all possible networks in Internet. ASs can make interconnection agreements of client-provider type if ASs are in different tier level or peering agreements between ASs at the same tier level [12].

The hypothesis that will be reviewed in this work is the following. An internal router interconnecting different networks in the same AS will have IP addresses in each interface sharing the same IP prefix because IP addressing allocation is usually quite near for certain AS. A border router interconnecting networks in different ASs will have very different IP addresses in each interface without sharing prefixes [13]. At first, AS relationships do not follow a certain IP allocation scheme.

Figure 1 shows typical Internet paths traversing several routers and ASs. Internal router R2 in AS C has IP addresses with similar prefixes because it interconnects networks that belong to the same AS and therefore with a certain allocated addressing scheme. This means that IP addresses for all interfaces in an internal router will be close in distance with high probability. A border router can interconnect different ASs like R1 in figure 1 that interconnects AS A and AS C. In this case, IP addresses in different interfaces of that border router belong to different ASs and therefore with different IP prefixes.

The different IP addressing in each interface of a router can be related with the tier-level of the AS. At first, with the original classful addressing scheme in Internet, ASs at Tier-1 would have addressing prefixes of A-class typically while other ASs at

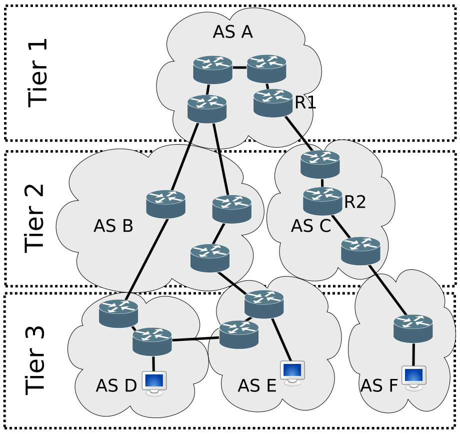


Fig. 1. Internet path traversing several routers and ASs

Tier-2 o 3 will have addressing prefixes around B and C-class. However since 1993 with the application of CIDR (Classless Interdomain Routing) [13], subnetting and supernetting is in use and it is possible to announce subnets in the Internet, making the distinction in addressing between tiers not so clear.

In figure 2, the complementary cumulative distribution function (CCDF, survival function) of IP prefix assigned to Tier-1, Tier-2 and Tier-3 ASs are presented. The following ASs are considered as Tier-1 [14]: AboveNet, AT&T, Global Crossing, Level 3, Verizon, NTT Communications, SAVVIS, Sprint and VSNL. Also 52 ASs are chosen randomly from those in Tier-2 like NTL, Telefonica or TELLCOM, and 2500 ASs in Tier-3. Data have been obtained from BGP information available at [15]: ASs and their allocated addressing. Each AS has allocated some subnetworks with specific prefixes and masks. For an IP address chosen randomly for an AS in certain Tier-x, Figure 2 shows the CCDF of the corresponding prefix. Prefixes are represented as a 32 bits unsigned integer. Each subnetwork contributes with a number of IP addresses depending on its mask size. At first, the addressing is very similar for all tiers, so nowadays the previous distinction between tiers based on addressing is not correct. However, the figure shows two clear steps around prefixes 1.2e+09 and 3.4+e09. This indicates that IP addresses are mainly concentrated in these two zones, containing around 70-80% of addresses. This peculiar behaviour can anticipate the typical offsets between pairs of IP addresses that are alias. For an internal router, both addresses could belong to one of both zones, resulting an IP offset around 0. In a border router, both IP addresses could belong to different zones, resulting an IP offset around 2.15+09.

Another characteristic that has to be noted from figure 2 is related with the central zone where the decrement in CCDF for Tier-2 and Tier-3 is significant. This means that Tier-2 and Tier-3 ASs have allocated a big percentage of addresses along this zone. Also it must be noted that for small prefixes, Tier-3 ASs are predominant.

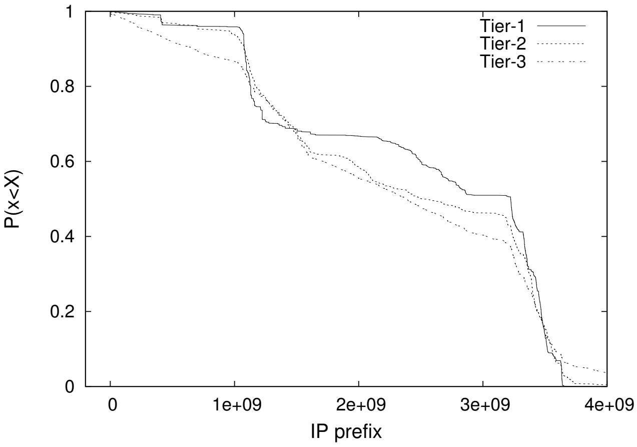


Fig. 2. CCDF of IP prefixes corresponding to Tier-1, Tier-2 and Tier-3 ASs

This organization in ASs, with internal and border routers, and the specific addressing scheme used in each AS, can be used to infer the characteristics in IP addressing distribution for interfaces that belong to the same router. This will be studied in the next section.

III. Relationship between IP addresses belonging to the same router (aliases)

In this section, the IP address distribution for interfaces that belong to the same router (aliases) is analyzed. In order to apply classical IP address alias resolution methods like Ally, probing is made in pairs of IP address to check if both addresses belong to the same router. We want to identify the characteristics of IP addresses that are alias. Both IP addresses in a pair are considered as two 32 bits unsigned integer numbers. We will consider *IP offset* as the absolute value of subtracting one IP address from the other $|IP_1 - IP_2|$.

In order to analyze the behavior of this IP offset related with alias resolution, experimental measurements have been made using Planetlab [16] measurement infrastructure. 50 planetlab end-nodes around the world have been used to obtain the IP addresses in the paths between them, resulting in 1708 IP addresses discovered using paris-traceroute [9]. This means 1,440,971 possible pairs of IP addresses to test with Ally as alias resolution method. In this scenario 1,036 pairs of addresses are found to be alias. The alias have been found applying Mercator [5], Ally [2] and other related methods [4]. As end-nodes are placed around the world, the results can be representative of the general Internet. The data traces and software used in this study are available online in [17].

The distribution of distances between IP addresses (IP offsets) that belong to the same router is related to the type of router as defined in previous section: internal o border router. Figure 3 shows the CCDF of IP offset for pairs of IP addresses that are alias depending on whether both IP addresses belong to the same or to different AS. It shows that aliases where both IP addresses are in the same AS result in small IP offsets. Being both IP addresses in the same AS, there is a high probability of having a common IP prefix. There is a second "frequent" offset for this "same AS" curve, at much higher values, but with a much lower probability. For IP addresses considered as alias in different ASs, there are three main steps in the curve. The IP offsets are concentrated around 0, 1.1e+09 and 2.15e+09. This third zone is related with the difference between the two main steps shown in figure 2. For all aliases in the scenario, 768 correspond to aliases in the same AS and 265 to aliases in different AS. This means that 75% of full set of aliases correspond to the same AS. Therefore, the range of IP offset around 0 is the most important in the final contribution to aliases.

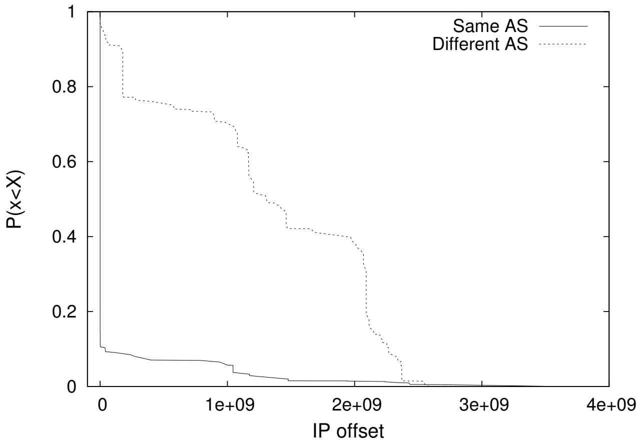
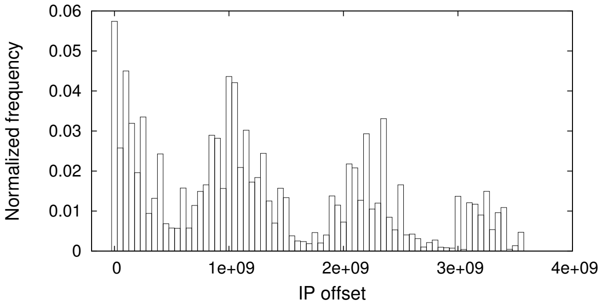


Fig. 3. CCDF of IP offset for aliases in the same and different AS

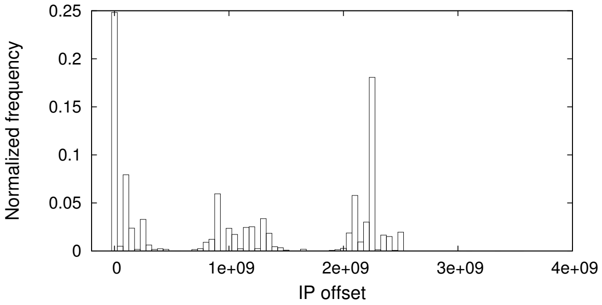
The change of AS in the path transversal between end-nodes in Internet is in part cause of the different IP offsets. This distribution of aliases depending on the AS can be translated to how pair of IP addresses are distributed in function of IP offset.

Figure 4 presents the histogram of IP offsets for all pairs of IP addresses (subfigure a) and for those pairs of IP addresses that are alias (subfigure b). The IP offset for a certain pair of IP addresses is considered as many times as those IP addresses appear in the traceroutes. This means that those core routers that are common to several traceroutes will receive more importance in this preliminary study. Later, in the study of aliases, all IP addresses will have the same importance. In the histogram of IP offsets for all pairs of IP addresses, the IP offset values are distributed along almost all the IP offset space. Therefore, it is indicating that the full set of IP addresses assigned to routers is distributed along all the IP addressing space. In Figure 4.b), in the case of alias pairs, IP offsets are concentrated mainly around 0 and in much lower percentage around 2.15e+09 (half the number of possible IP addresses, 2^{32}). Also a mid zone is present around IP offset with 1.1e+09, but much less concentrated than previous zones. We will consider this three zones in order to search for pairs of IP addresses with more probability to be aliases: zone 0 (around 0), zone 1 (around 1.1e+09) and zone 2 (around 2.15e+09).

These three zones of IP offset are important for discovering IP alias. The CCDF for both histograms in Figure 4 are presented in Figure 5 where the steps present in the 'Aliases' curve correspond to the points of interest mentioned before for the three zones. The CCDF for all IP address space is distributed along the IP offset axis while the CCDF for aliases has three main steps. This means that IP offsets within these steps have higher probability to be aliases.



(a)



(b)

Fig. 4. Histogram of IP offsets for all pairs (a) and for pairs that are aliases (b)

Taking the experiments between 18 nodes available in the Etomic [18] measurement platform, an overview of the IP offset effect in the european addressing scheme is obtained. In this case all nodes are distributed around Europe. The results are very similar to those obtained in Planetlab case as can be shown in Figure 6. However, this time the importance of zone 1 is much lower than before because the second step is less appreciable. Again, the largest percentage of aliases is concentrated around zone 0 and zone 2.

Figure 7 shows the percentage of aliases found in the different IP offset zones 0, 1 and 2 depending on the TTL from end-nodes in the Planetlab scenario. An occurrence is considered for each IP address in the alias pair, and the minimum TTL observed in all traceroutes is considered for each IP. As it can be observed, the distribution of aliases between zones is almost independent of the TTL distance from end-nodes in the topology. The main percentage of aliases is concentrated in zone 0, this means with an IP offset around 0. This indicates that IP addresses are very close, probably in the same AS. Only two peaks are present in zone 2 for TTL 5 and TTL 14 (near to starting and finishing end-node respectively) indicating those TTLs where a change of AS is more probable and then an IP offset in the range of 2.15e+09. Anyway, the dependence of TTL on IP offset is not so clear to make the IP offset zonification dependent on TTL.

IP addresses used in Internet ASs have a peculiar distribution. Here, the specific addressing distribution in our experimental scenario is analyzed.

In Figure 8.a), the matrix of all IP addresses in Planetlab scenario is presented. It represents all possible combinations of pairs of IP addresses that is the input to an Ally-like alias resolution method. In order to appreciate the concentration of IP addresses in different zones, a dispersion technique has been applied in the representation: a x-y small random offset is applied to each point in order to visualize zones with more concentration of occurrences. IP addresses are concentrated in ranges associated with ASs transversed in the scenario, covering a big percentage of the addressing space. The addresses not present in Figure 8.a) correspond to private and reserved addressing, and mainly to addresses reserved by APNIC (Asia-Pacific Network Information Center). If we plot only the pairs of IP addresses that are alias, the results are shown in Figure 8.b). This is a subset of previous figure with some peculiarities. First, a large percentage of aliases are concentrated around the diagonal of Figure 8.b). The diagonal implies an IP offset near to 0 value, and therefore, IP addresses very close together, mainly both IP addresses being part of the same AS. This would correspond to previously defined zone 0.

Second, the other zone with an important contribution in aliases is IP address pairs with a distance of approximately 2.15e+09

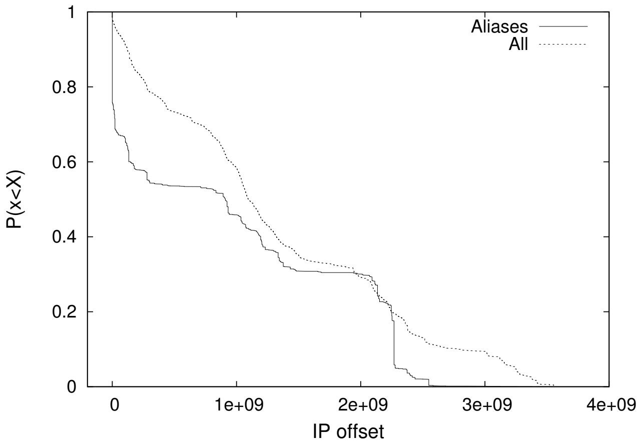


Fig. 5. IP offsets CCDF for all pairs and only aliases in Planetlab infrastructure

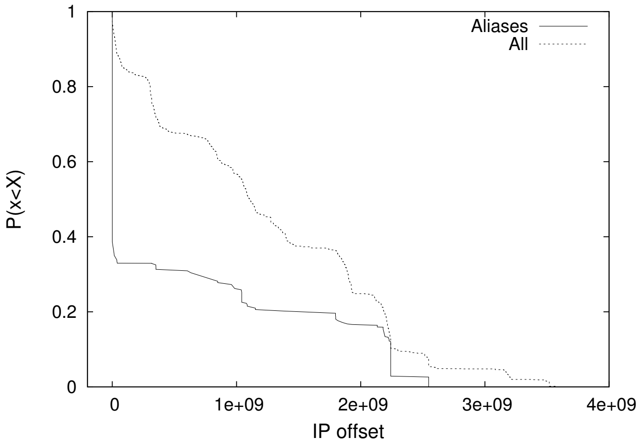


Fig. 6. IP offsets CCDF for all pairs and only aliases in Etomic infrastructure

from the diagonal (previously defined zone 2). These IP addresses correspond to border routers that interconnect different ASs. Because of the hierarchy in tiers for ASs, it is usual the interconnection between ASs at different tier-level. ASs at different level can allocate IP addresses belonging to different prefixes of the addressing space. This distance from the diagonal coincides with the distance between the two steps observed in Figure 2.

Third, the other zones are much less important, with a distance from the diagonal of half the previous value. These points would correspond with previously defined zone 1.

This localization of IP addresses which are alias compared with the full set of IP addresses can be used to know in advance if two IP addresses are good candidates to be aliases. This will correspond to pairs of IP addresses with an IP offset present in zone 0, 2 and 1 in order of importance.

Similar results are shown in Etomic scenario, even with more concentration in zones 0 and 2 previously described. This result is shown in Figure 9. In this case, because all the nodes are related with European networks, IP addresses are localized in some parts of the addressing scheme, making more important the zone 2 with IP offset around 2.15e+09.

As a conclusion, IP offset metric is related with the aliasing property of pairs of IP addresses. The zones 0, 1 and 2 define ranges of IP offset with more probability to have pairs of IP addresses that are aliases. Therefore, IP offset can be used as a reduction method as presented in next section.

IV. Reducing the number of IP addresses to compare

The IP offset metric, as a relationship between IP addresses belonging to the same router, can be used as a reduction method in techniques for IP address alias resolution. The idea is to consider the ranges of IP offset where there is more probability to find aliases. These ranges of IP offset have been previously defined as zones 0, 1 and 2.

This reduction method has characteristics that improve those in the state of the art. Extra probing traffic is not needed as the metric is calculated directly from the IP addresses. The method is not time-consuming. This is a good advantage compared with TTL and IPID methods that in both cases need to inject extra probing packets to get the information to apply the reduction. Besides, as no extra probing traffic is needed, the reduction method can be applied in a fully distributed way and, consequently,

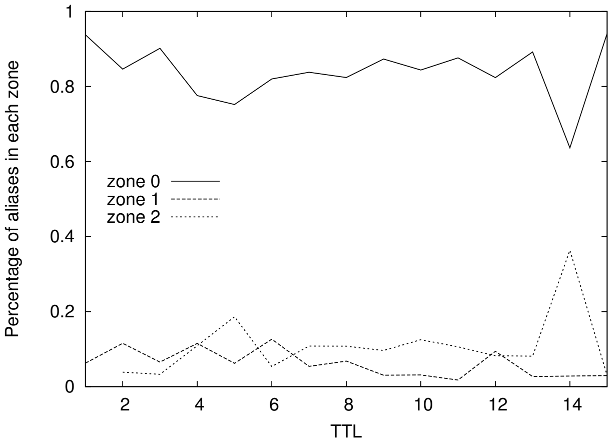


Fig. 7. Percentage of aliases found in different IP offset zone depending on the TTL distance

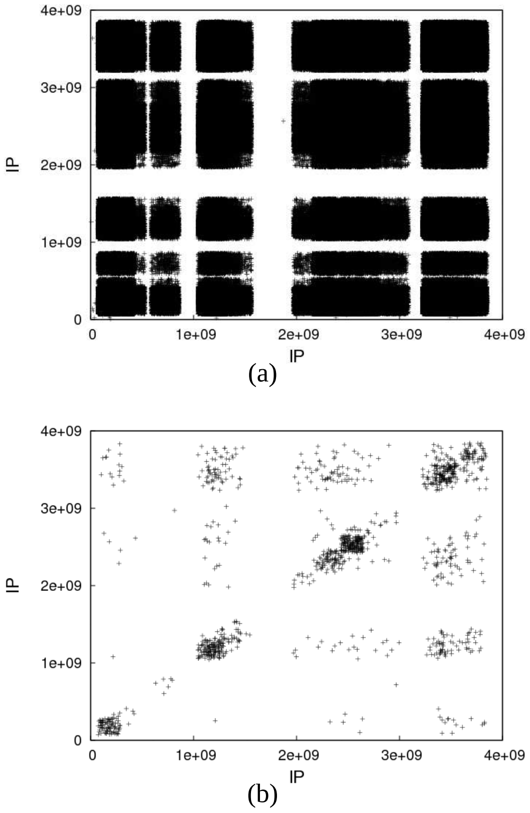
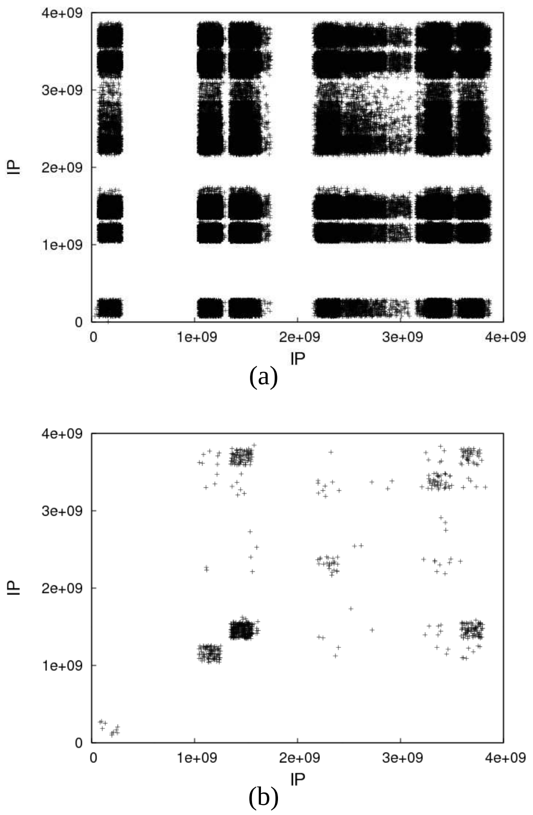


Fig. 8. IP pairs fulls space (a) and only alias (b) in Planetlab scenario

the whole alias resolution technique can be distributed. Different sets of IP address pairs can be distributed between different probing nodes. This is different from TTL and IPID based methods where probing must be made from the same node to each target IP address. Each probing node needs to get the metric for all IP addresses, and then, there will be repeated probing packets from different probing nodes.

Clustering algorithms are used to find the specific range of IP offsets to consider around zones 0-2. Several clustering methods like K-means [19] and Expectation Maximization (EM) [20] have been tested using different training sets. EM has been checked to provide the best results [21]. The resulting clusters indicate the ranges of IP offset where there is more probability to get aliases.

Depending on the number of clusters considered, and therefore, the number of pairs of IP addresses to check for aliases, we can get the percentage of completeness desired in alias resolution. The results of the reduction method are shown in Figure 10. Percentage in alias resolution is presented compared with the percentage of number of pairs of IP addresses to test. The results correspond to the whole set of pairs of IP addresses in Planetlab scenario using three types of training sets: the set of IP addresses in the Etomic scenario, a subset of Planetlab with 18 end-nodes, and the Planetlab full scenario. As observed in



(a)

(b)

Fig. 9. IP pairs fulls space (a) and only alias (b) in Etomic scenario

the figure, the results are almost independent of the training set. This is a good characteristic of the method that allows to use precalculated clusters and to apply those clusters to almost any common network scenario.

In Figure 10, the first point corresponds to a one cluster scenario, and the successive points correspond to adding one cluster at a time. Adding a cluster will imply an increase in the number of pairs to test with the alias resolution method, but at the same time, it will imply an increase in the percentage of aliases resolved positively. The number of clusters can be chosen in order to provide the percentage of aliases needed. For example, testing only 10% of IP address pairs, around 73% of aliases are resolved.

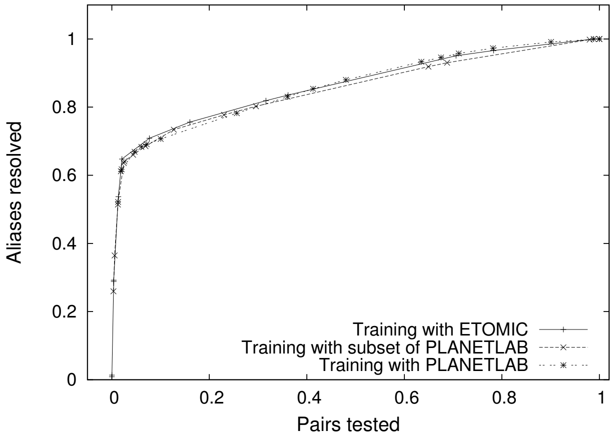


Fig. 10. Alias resolution results with reduction in the number of pairs of IP addresses to test using IP offset

A detail of resulting clusters for Planetlab full scenario is shown in Figure 11. It shows the range of IP offset covered by each of the resulting clusters, ordered from the cluster 0 with more number of aliases to cluster 13 with minor contribution. The clusters with higher relative contribution are those around zones 2 and 1, and all of them with limited range of IP offset (low number of IP addresses to test). The details of contribution for each cluster is presented in table 1. In this table, the percentage of pairs of IP addresses used in each cluster with respect to the total number of pairs to test is presented. The

Cluster *number*	Alias in Cluster %	*Pairs Used* %	*Alias Resolved* %
1	6.923	0.0090	0.8712
2	5.362	0.3691	27.4927
3	4.528	0.0184	1.1616
4	1.918	0.8649	23.0396
5	1.211	0.0863	1.4520
6	1.086	0.7311	11.0358
7	0.078	5.6013	6.0987
8	0.038	8.3742	4.4530
9	0.028	15.293	6.2923
10	0.026	4.5432	1.6456
11	0.023	33.7891	11.2294
12	0.014	7.7265	1.5488
13	0.012	21.0741	3.6786

TABLE 1

DETAILS FOR EACH CLUSTER

percentage of contribution to the total alias recognition is also shown. The first clusters use only a fraction of pairs in the test but with a high percentage of alias resolution, indicating the localization of aliases around the zones covered by those clusters.

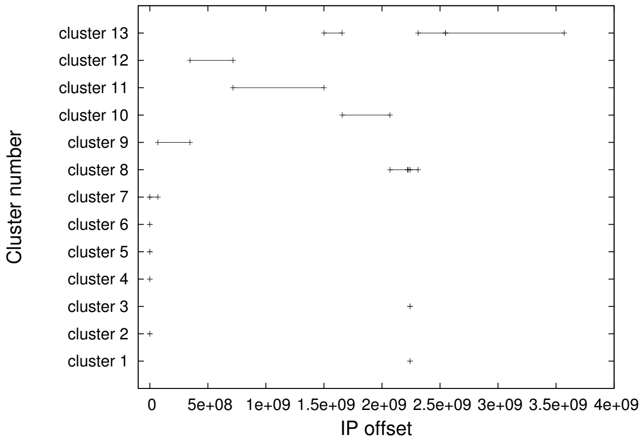


Fig. 11. IP offsets covered by each cluster

Although some advantages of using the IP offset method have been explained compared with methods in the state of the art, another parameter to compare reduction methods is the percentage of aliases positively resolved depending on the reduction applied in the number of IP address pairs to check. This can be an important factor in improving efficiency of alias resolution methods. Figure 12 presents the percentage of aliases resolved positively compared with the percentage of pairs of IP addresses tested for several reduction methods: IPID-based, TTL-based, AS-based and our proposal IPoffset-based. The AS-based reduction method used in iPlane has only one scenario (it is not possible to vary the number of pairs to test), with a great percentage of reduction but with very bad results in alias identification as shown in Figure 12. The TTL-based reduction method provides results near to those in the IP offset method but only when testing more than 30% of the IP address pairs. For a lower percentage of pairs tested, the IP offset method outperforms the rest of the methods. For example, the AS-based reduction method obtains 32.04% of aliases resolved when 1.55% of the pairs are tested. The proposal in this paper, the IP offset-based reduction method, obtains more aliases resolved (53.72%) even when a lower number of IP address pairs are tested (1.34%).

The IPID-based method uses only two UDP probing packets, one to each IP address, and it counts those pairs of IP addresses that answer with similar IPIDs. It does not provide good results, as explained in [21], mainly because of filtering, the variation of traffic generated by a router (the 200-threshold does not always work) and sometimes because of randomly generated IPIDs at the routers.

The reduction method based on IP offset has been demonstrated to be a valid option with very good results specially in order to avoid extra probing traffic. Also the possibility to distribute the information to apply the alias resolution method from different probing nodes is important mainly in large network scenarios. Finally, the results in percentage of reduction and alias identification are promising, specially for requirements with a high reduction in the number of pairs to test for aliasing.

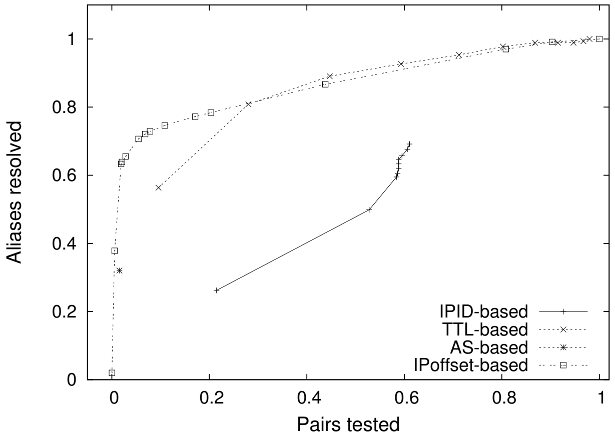


Fig. 12. Comparison of reduction methods in alias resolution

V. Conclusions

The characteristic distribution of IP addresses in paths through the Internet can be used to identify those pairs of IP addresses with a higher probability to be aliases. For this task, a new metric called IP offset has been considered. The results indicate that it is possible to know the ranges of IP offsets with more probability to be aliases and to use this ranges to reduce the number of pairs of IP addresses to test in classical alias resolution methods as Ally.

The IP offset reduction method provides an efficiency improvement better than previous methods in the state of the art. Besides, the proposed reduction method has some good characteristics. First, no extra probing traffic is needed; the IP offset is calculated directly from the IP addresses. Second, the reduction method can be calculated in a distributed way, for example at different nodes in charge of checking alias resolution for subsets of IP addresses of the network.

The clustering could have some improvements that will be studied in future. For example, the cluster in zone 0 includes the IP offset with value equal to 1. This could correspond mainly to both ends of a point-to-point line with a /30 or /31 mask. However, the output from traceroute does not provide mask information and therefore the solution is not as easy as discarding this specific IP offset.

References

[1] N. Spring, M. Dontcheva, M. Rodrig, and D. Wetherall, “How to resolve ip aliases,” Tech. Report 04-05-04, Washington Univ. Computer Science, 2004.

[2] N. Spring, R. Mahajan, and D. Wetherall, “Measuring ISP topologies with Rocketfuel,” in *Proc. ACM SIGCOMM*, 2002.

[3] Adam Bender, Rod Sherwood, and Neil Spring, “Fixing ally’s growing pains with velocity modeling,” *IMC 08*, October 2008.

[4] Santiago Garcia-Jimenez, Eduardo Magaña, Daniel Morato, and Mikel Izal, “Techniques for better alias resolution in internet topology discovery,” in *To be published in 11th IFIP/IEEE International Symposium on Integrated Network Managemen miniconference*, New York, USA, June 2009.

[5] Jean-Jacques Pansiot and Dominique Grad, “On routes and multicast trees in the internet,” ACM SIGCOMM Computer Communication Review, 1998.

[6] Hal Burch, “Measuring an IP network in situ,” Carnegie Mellon University, PhD thesis, ISBN 0-542-01549-8, 2005.

[7] Harsha V. Madhyastha, Tomas Isdal, Michael Piatek, Colin Dixon, Thomas Anderson, Arvind Krishnamurthy, and Arun Venkataramani, “iPlane: An information plane for distributed services,” in *7th USENIX Symposium on Operating Systems Desing and Implementation*, Nov. 2006.

[8] V. Jacobson, “Traceroute ftp://ftp.ee.lbl.gov/traceroute.tar.gz,” October 1989.

[9] Brice Augustin, Xavier Cuvellier, Benjamin Orgogozo, Fabien Viget, Matthieu Latapy Timur Friedman, Clemence Magnien, and Renata Teixeira, “Avoiding traceroute anomalies with paris traceroute,” in *6th ACM SIGCOMM*, October 2006, pp. 153–158.

[10] R. Govindan and H. Tangmunarunkit, “Heuristics for internet map discovery,” in *Proc. IEEE INFOCOM*, 2000.

[11] Lixin Gao and Feng Wang, “The extent of as path inflation by routing policies,” *7th IEEE Global Internet Symposium (Taipei, Taiwan)*, Nov. 2002.

[12] Feng Wang and Lixin Gao, “On inferring and characterizing internet routing policies,” *Internet Measurement Conference (Miami, Florida, USA)*, October 2003.

[13] Olaf Maennel and Anja Feldmann, “Realistic bgp traffic for test labs,” *ACM SIGCOMM Computer Communication Review*, vol. 32, no. 4, pp. 31–44, Oct. 2002.

[14] Mark Winther, “Tier 1 isps: What they are and why they are important,” *IDC White Paper, NTT Communications*, May 2006.

[15] “A launch pad for network & internet management related resources,” http://www.netconfigs.com.

[16] “Plantelab: An open platform for developing, deploying, and accessing planetary-scale services,” http://www.planet-lab.org.

[17] Santiago Garcia-Jimenez et al., “Tools and data sets used in this paper,” http://www.tlm.unavarra.es/∼santi/research.

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

[19] J. B. MacQueen, “Some methods for classification and analysis of multivariate observations,” in *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*. 1967, vol. 1, pp. 281–297, University of California Press.

[20] N. M. Laird D. B. Rubin A. P. Dempster, “Maximum likelihood from incomplete data via the em algorithm,” *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[21] Santiago Garcia-Jimenez, Eduardo Magaña, Daniel Morato, and Mikel Izal, “Improving efficiency of ip alias resolution based on offets between ip addresses,” in *To be published in 21st International Teletraffic Congress (ITC 21)*, Paris, France, Sept. 2009.